



Towards real-time STEM simulations through targeted subsampling strategies

Alex W. Robinson¹  | Jack Wells² | Daniel Nicholls¹ |
 Amirafshar Moshtaghpour^{1,3}  | Miaofang Chi⁴ | Angus I. Kirkland^{3,5} |
 Nigel D. Browning^{1,6,7}

¹Department of Mechanical, Materials and Aerospace Engineering, University of Liverpool, Liverpool, UK

²Distributed Algorithms Centre for Doctoral Training, University of Liverpool, Liverpool, UK

³Correlated Imaging Group, Rosalind Franklin Institute, Didcot, UK

⁴Chemical Science Division, Centre for Nanophase Materials Sciences, Oak Ridge National Laboratory, Oak Ridge, Tennessee, United States

⁵Department of Materials, University of Oxford, Oxford, UK

⁶Materials Sciences, Physical and Computational Science Directorate, Pacific Northwest National Laboratory, Richland, Washington, United States

⁷Research and Development, Sivananthan Laboratories, Bolingbrook, Illinois, United States

Correspondence

Alex W. Robinson, Department of Mechanical, Materials and Aerospace Engineering, University of Liverpool, Liverpool L69 3GH, UK.

Email: a.w.robinson@liverpool.ac.uk

Abstract

Scanning transmission electron microscopy images can be complex to interpret on the atomic scale as the contrast is sensitive to multiple factors such as sample thickness, composition, defects and aberrations. Simulations are commonly used to validate or interpret real experimental images, but they come at a cost of either long computation times or specialist hardware such as graphics processing units. Recent works in compressive sensing for experimental STEM images have shown that it is possible to significantly reduce the amount of acquired signal and still recover the full image without significant loss of image quality, and therefore it is proposed here that similar methods can be applied to STEM simulations. In this paper, we demonstrate a method that can significantly increase the efficiency of STEM simulations through a targeted sampling strategy, along with a new approach to independently subsample each frozen phonon layer. We show the effectiveness of this method by simulating a SrTiO₃ grain boundary and monolayer 2H-MoS₂ containing a sulphur vacancy using the abTEM software. We also show how this method is not limited to only traditional multislice methods, but also increases the speed of the PRISM simulation method. Furthermore, we discuss the possibility for STEM simulations to seed the acquisition of real data, to potentially lead the way to self-driving (correcting) STEM.

KEYWORDS

artificial intelligence, beam damage, compressive sensing, inpainting, stem simulation, subsampling

1 | INTRODUCTION

To fully identify the atomic scale structure and composition of complex materials, interfaces and defects

from experimental images, it is essential to use simulations to capture all the experimental parameters involved. The accurate simulation of scanning transmission electron microscopy (STEM) images^{1–4} is a computationally

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *Journal of Microscopy* published by John Wiley & Sons Ltd on behalf of Royal Microscopical Society.

expensive task due to the nature of the scattering and detection process. The most common method used to obtain simulations is the *multislice method*.^{5–29} In the multislice approach, the 3-dimensional atomic potential of a sample is first approximated by a series of 2-dimensional (2D) infinitely thin potential slices, V_s^{2D} , where s is the index of a slice and \vec{r} denotes a location in real space coordinates. For every probe position, the multislice approach involves the following steps. First, the incident wavefunction of the electron beam $\psi_{s+1}^i(\vec{r})$ for a slice $s + 1$, is computed from the exit wavefunction $\psi_s^e(\vec{r})$ of the previous slice s and the atomic potential of that layer:

$$\psi_{s+1}^i(\vec{r}) = \psi_s^e(\vec{r}) \cdot \exp[i\sigma V_s^{2D}(\vec{r})],$$

where σ denotes the beam-specimen interaction constant. Next, the exit wavefunction $\psi_{s+1}^e(\vec{r})$ for slice $s + 1$ is computed by propagating the incident wavefunction $\psi_{s+1}^i(\vec{r})$ using Fresnel propagation model:

$$\psi_{s+1}^e(\vec{r}) = \mathcal{F}^{-1}[\mathcal{F}[\psi_{s+1}^i(\vec{r})] \cdot \exp(-i\pi\lambda|\vec{q}|^2 t)],$$

where λ and t are the electron wavelength and slice depth, respectively.²⁷ Once the sample has been propagated through the sample, the exit probe is then determined in Fourier space. This must be repeated for all required STEM probe locations; hence the computation time of the STEM simulation is limited by computing multiple Fast Fourier Transforms (FFTs).

A more recent development through an algorithm called plane-wave reciprocal space interpolated scattering matrix (PRISM)²⁷ has led to much faster STEM simulations. This algorithm forms a basis set of plane waves (based on an interpolation factor where a higher interpolation factor introduces more error but faster simulation) which are independently propagated through the sample (in a multislice approach). After propagation, the plane waves form a scattering matrix, and can be superimposed with appropriate weighting to form a close approximation to the exit probe had it been calculated using the multislice approach. The set of plane waves are essentially shared between all probe locations, meaning that the multislice step must be completed only for the basis set of plane waves, greatly reducing the computational load. This leads to significant speed up in computation times, with minimal loss of information compared to the traditional multislice method (see Refs.^{27,28} for more details).

To account for all possible electron-phonon interactions, the frozen phonon approximation is also often used,^{21–23} where the intensity of multiple multislice calculations (layers) is averaged over various frozen phonon configurations (FPC) to achieve a final average output that represents a

more accurate time integrated-image of the sample. This series of calculations must then be solved at each probe position in STEM simulations.²⁹

Taking all the physics of these interactions into account at every beam location means the computation time of multislice STEM simulation scales with the number of required probe locations, the number of FPCs, and the number of reciprocal space sampling points. A typical multislice STEM simulation (for a sample 10 nm in thickness, 256×256 grid locations, using 20 FPCs and a reciprocal space sampling of 0.04 \AA^{-1}) operating on a system equipped with a graphics processing unit (GPU) for faster calculation can take of order hours, and even longer if a GPU is not available.²⁷ The PRISM method can perform the same simulation in significantly less time depending on the interpolation factor used (typically on the order of minutes or potentially seconds with a larger interpolation factor) at the expense of accuracy. For calculation of structures with a crystalline periodicity, a *tiling* method can be used to increase the number of effective probe locations. Furthermore, the maximum necessary scan positions required (for structures satisfying periodic boundary conditions) according to Shannon sampling theory is given by,

$$M_x \times M_y = \frac{4(a \times b)\alpha}{\lambda},$$

where $(a \times b)$ is the size of the orthogonal scan area, $M_x \times M_y$ is the scan dimension (pixels), α and λ are the probe convergence semiangle and electron wavelength respectively.²⁹

In our previous work,^{30,31} it was demonstrated that it is possible to effectively compress a simulation through spatial, reciprocal space, and FPC subsampling (Figure 1) to ultimately reduce the computation time and still obtain functionally identical results. A spatially compressed simulation (i.e. one that reduces the number of scan points below the Shannon limit) can be reconstructed using an *inpainting algorithm*, and the quality of the output reconstruction can be reported as a function of the sampling ratio, the maximum reciprocal vector, and the number of FPCs used. In this paper, we describe a new compression/reconstruction method which aims to reduce the computation time for simulations, irrespective of the simulation methodology. This method combines a new sampling strategy describing how each FPC layer is acquired to improve the efficiency of calculation and thus allows much lower sampling ratios to be used, while still leading to functionally identical results compared to the full simulations. The increase in speed means that simulations can in principle be used to seed the interpretation of experimental images using machine learning approaches,

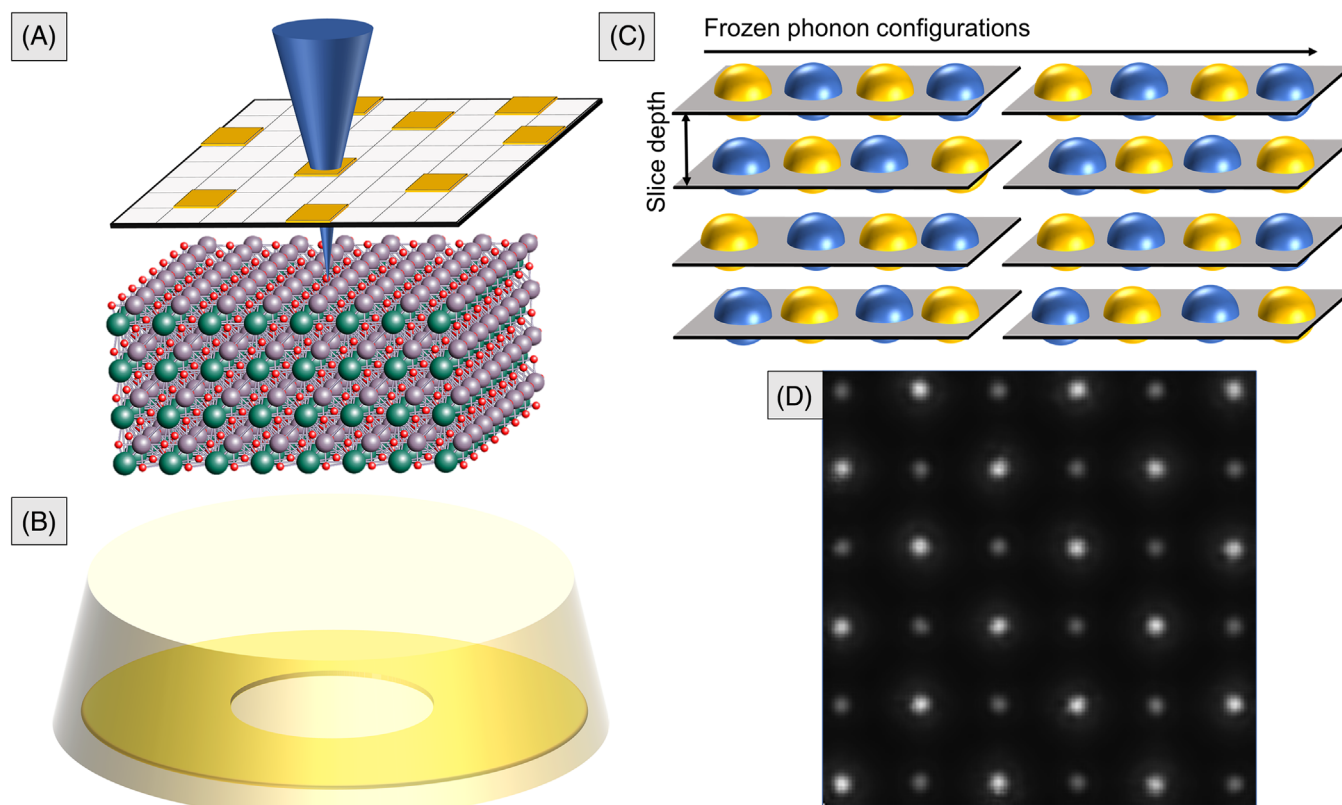


FIGURE 1 Compressed STEM simulations using (A) spatial subsampling (only acquiring highlighted pixels), (B) limiting the number of contributing reciprocal space vectors (to approximately the edge of the detector) and (C) reducing the number of frozen phonon configurations. An example reconstructed simulation of bulk strontium titanate (D) using the methods above and an inpainting algorithm.

leading to real-time analytics and the potential to develop a self-driving (correcting) microscope.

2 | METHODS: FROM RANDOM SAMPLING TO TARGETED SAMPLING IN STEM SIMULATIONS

Spatial subsampling^{30–39} has previously been applied to experimental STEM images^{32–34} and has shown a reduction in beam induced damage, faster acquisition speed, and reduced drift and image distortion without loss of signal at each measured probe location (i.e. the same probe dwell time and current).

Inspired by the success of CS theory for STEM (see an example in Figure 2), a method for spatially subsampling STEM simulations could be developed by only acquiring a subset of the simulated STEM exit probes. This is done by defining a subset of probe locations (similar to how experimental CS-STEM is performed) which forms a subsampled simulation, as in the case of Figure 2. The key difference is that simulations are not limited by experimental CS-STEM limitations such as scan-coil hysteresis, drift, or beam damage, permitting

an exploration of diverse sampling strategies for STEM simulations.

2.1 | Method of targeted sampling

We now turn our attention to developing a targeted sampling strategy for STEM simulations, with the intention of achieving high image recovery quality at low sampling ratios (<10%). When restricted to sparse signals (a signal is called sparse in a basis if its representation in that basis has only a few nonzero coefficients) in an orthonormal basis, for example, Fourier, the theory of CS recommends a purely random subsampling strategy,³⁵ referred to as Uniform Density Sampling (UDS). However, STEM simulations often involve structured images containing a significant amount of *vacuum* space. Although it is important to still understand the contrast in these vacuum regions for accurate analysis, they are often low frequency regions (i.e. with constant contrast or no structure), and hence fewer sampling points are required. Therefore, we aim to design an efficient sampling strategy, such that fewer measurements of complex signals would be sufficient for high quality recovery.

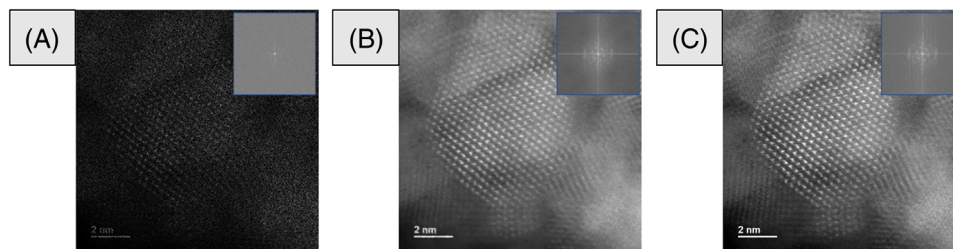


FIGURE 2 Artificial subsampling applied to an experimental STEM image of CeO_2 . Subsampled acquisition (A), reconstructed image (B) and fully sampled reference image (C). Each image has its respective FFT image overlaid.

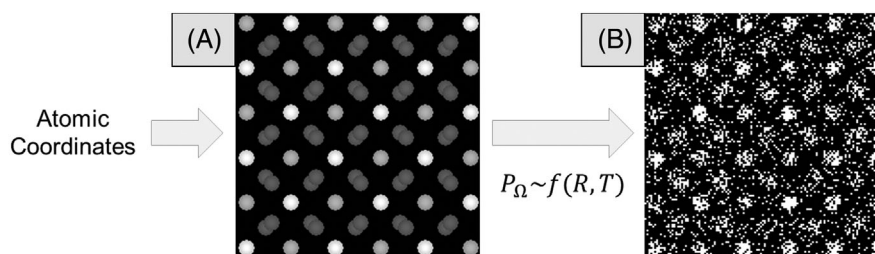


FIGURE 3 Formation of a targeted sampling mask for compressed simulations. The file for atomic locations is loaded, it is then passed through a code which generates a sampling space map \mathbf{T} (A) of atom locations where the relative radii is on the order of Angstroms, and the intensity of each atom's location is proportional to its atomic number (for this specific example). The space map is then used in conjunction with a random bias, \mathbf{R} , to form a targeted sampling mask (B) through a sampling function, $P_{\Omega} \sim f(\mathbf{R}, \mathbf{T})$. If the value of \mathbf{R} is 1, then the mask is purely random, if it is 0, then the mask is purely targeted.

A key feature of STEM simulations is that the atom locations are an input, which for abTEM²⁸ can be a file containing atomic coordinates, atomic number, occupancy and the root mean square displacement of the atom (for the frozen phonon approximation). Therefore, we propose that this file could also be used as a method to form a *targeted* sampling mask which prioritises sampling at atom locations (given we know this from the file provided), as opposed to a purely random approach (as in Refs.^{30,31}). However, the sampling is not purely targeted, but also includes some bias R , which allows any location to be sampled at random. This essentially makes the mask a layer of purely random sampling and purely targeted based on some targeted feature such as atomic number, edges, binary atom location, or inverse intensity. This is demonstrated in Figure 3.

2.2 | Applying targeted sampling to the frozen phonon model

The second step towards improving the efficiency of STEM simulation is to optimise how the frozen phonon model can be adapted through a targeted sampling method. The frozen phonon model is used to account for thermal diffuse scattering within the sample, by taking snapshots of the

sample at some given time where the atom locations are slightly displaced from their equilibrium position depending on the Debye–Waller factor (DFW) of the atom.²¹ Each snapshot of atom positions is known as a *frozen phonon configuration* (FPC) and as more configurations are considered, generally the more accurate the simulation is, given the final simulation is the average of simulations over all configurations (Figure 1C).

In practice, beyond some number of configurations (depending on sample type, resolution, thickness and spatial density), the improvement in simulation quality diminishes; however, the computation time scales linearly. Therefore, for the purpose of speeding up simulations, it is better to limit the number of configurations to the point where the improvement begins to diminish.^{30,31}

As an example, consider sampling at a location where there is no atom. Here the number of configurations used makes an insignificant contribution to the intensity of the pixel at that location, and therefore to continue sampling this position would be time inefficient. Instead it would be better to sample at atom sites more frequently where the frozen phonon approximation has more effect. This can be achieved by using a different targeted mask for each FPC rather than using the same mask each time. This will also increase the net sampling of the final simulation as the pixel values are averaged in the final step (Figure 4), as

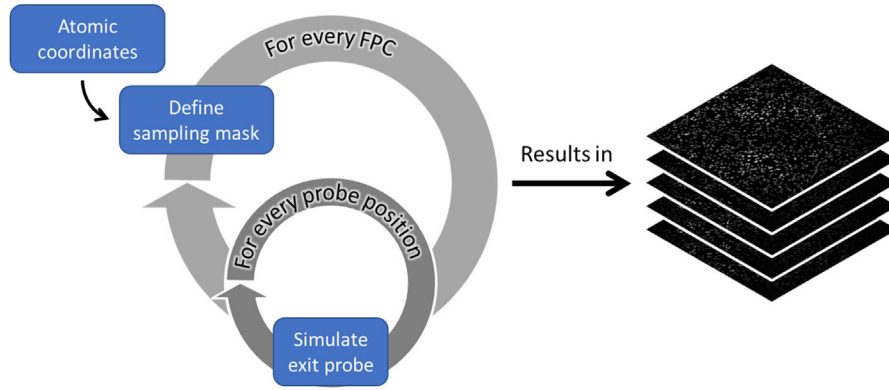


FIGURE 4 The method used for forming simulations for each different frozen phonon configuration layer. Each layer uses a different mask, in this case a *targeted mask*. This data is then weighted based on sampling frequency and passed to an inpainting algorithm for image recovery.

well as reducing the total sampling ratio required for each independent configuration.

These two methods, when used in conjunction, can yield a final simulation that still includes the frozen phonon approximation, but decreases the simulation run-time significantly.

2.3 | Sensing model and image recovery method

In this section we outline our sensing model and chosen image recovery method; describing Sections 2.1 and 2.2 mathematically.

2.3.1 | Sensing model

Let $\mathbf{X} \in \mathbb{R}^{N_1 \times N_2 \times N_3}$ be the discretised 3-dimensional representation of our simulation, with $\mathbf{X}_l \in \mathbb{R}^{N_1 \times N_2}$ representing each of our frozen phonon layers $l \in \{1, \dots, N_3\}$, and N_3 represents the total number of frozen phonon configurations. We assimilate each \mathbf{X}_l to its vectorised version $\mathbf{x}_l := \text{vec}(\mathbf{X}_l) \in \mathbb{R}^{\hat{N}}$ where $\hat{N} = N_1 N_2$. We now introduce our compressed STEM simulations to reduce simulation computation time. We do this by subsampling each of the frozen phonon layers where the number of acquired spatial locations in the l th layer is given as $M_l < \hat{N}$ acquired over the subsampling set $\Omega_l \subset \{1, \dots, \hat{N}\}$ with cardinality $|\Omega_l| = M_l$. This defines our acquisition model as,

$$\mathbf{y}_l = \mathbf{P}_{\Omega_l} \mathbf{x}_l + \mathbf{n}_l \in \mathbb{R}^{\hat{N}}, \quad (1)$$

for every frozen phonon layer where $\mathbf{P}_{\Omega_l} \in \{0, 1\}^{\hat{N} \times \hat{N}}$ is a mask operator with $(\mathbf{P}_{\Omega_l} \mathbf{x}_l)_j = \mathbf{x}_{l_j}$ if $j \in \Omega_l$ and $(\mathbf{P}_{\Omega_l} \mathbf{x}_l)_j =$

0 otherwise. We also assume that the same number of pixels are subsampled for different layers by setting $M_l = M$ for all layers and leave the general model for future study. The net subsampling set of pixels associated to simulating N_3 frozen phonon layers is the union of the layer-wise subsampled sets: $\Omega = \cup_{l=1}^{N_3} \Omega_l$. Let w_j be the number of times that a pixel $j \in \{1, \dots, \hat{N}\}$ is simulated, which reads

$$w_j = \left(\sum_{l=1}^{N_3} P_{\Omega_l} \right)_{j,j}. \quad (2)$$

The final, averaged simulation (over all frozen phonon layers) $\mathbf{y} \in \mathbb{R}^{\hat{N}}$ is computed as

$$y_j = \begin{cases} \frac{(\sum_{l=1}^{N_3} y_l)_j}{w_j}, & w_j > 0, \\ 0, & w_j = 0 \end{cases} \text{ for } j \in \{1, \dots, \hat{N}\}. \quad (3)$$

As can be seen in Equations (1) and (2), using more frozen phonon layers will yield a greater net sampling ratio (number of nonzero elements in w) if the sampling masks are nonidentical. In the case of identical masks ($P_{\Omega_l} = P_{\Omega}$ for all layers), Equation (3) is a subsampled version of the frozen phonon model where each probe is the average of all configurations.

2.3.2 | Image recovery with sparse dictionary learning

Now that we have defined the sensing model above, we turn our attention to the recovery of the full signal. Our choice of inpainting algorithm used to reconstruct the simulations consists of two key parts: a (blind) dictionary

learning algorithm, followed by a sparse coding algorithm. The image recovery problem is turned into a Bayesian dictionary learning problem based on the Beta Process Factor Analysis (BPFA) developed in Ref.³⁸ The underlying principle of dictionary learning is that a sparse signal can be represented by a linear combination of basis signals, where only a small number of these basis signals have a significant *weight* applied to them.⁴⁰ It has been shown that sufficiently sampled data also has these properties; therefore, the recovery is turned into a sparsity pursuit problem through determination of the weights of basis signals. The basis signals themselves are known as *dictionary atoms*, and these can be either inferred from the data (known as dictionary learning), or from a predetermined dictionary such as the *discrete cosine transform* (DCT).

In order to recover our full signal $\hat{X} \in \mathbb{R}^{N_1 \times N_2}$, we partition the subsampled (vectorised) acquisition $\mathbf{y} \in \mathbb{R}^{\tilde{N}}$ into a set of overlapping patches $\{\mathbf{y}_i\}_{i=1}^{N_b}$ of size $B \times B$ where the number of overlapping patches $N_p = (N_1 - B + 1) \times (N_2 - B + 1)$. Similarly, we can separate the mask operator, true signal, and noise into overlapping patches in a similar fashion, such that we have $\{\mathbf{P}_{\Omega_i}\}_{i=1}^{N_b}$, $\{\mathbf{x}_i\}_{i=1}^{N_b}$, and $\{\mathbf{n}_i\}_{i=1}^{N_b}$, respectively. This sets a patch-based sensing model as

$$\mathbf{y}_i = \mathbf{P}_{\Omega_i} \mathbf{x}_i + \mathbf{n}_i \in \mathbb{R}^{B^2}, \quad (4)$$

for each patch index $i \in \{1, \dots, N_p\}$. It is assumed that the image is sparse in a mutual basis such that $\mathbf{x}_i = \mathbf{D}\boldsymbol{\alpha}_i$ where $\mathbf{D} \in \mathbb{R}^{B^2 \times K}$ and $\boldsymbol{\alpha}_i \in \mathbb{R}^K$ for a dictionary with K atoms. Hence, the goal is to find the dictionary, weights, and noise statistics to recover the true signal based on the acquired, subsampled data. For this purpose, BPFA suits our problem as we can infer the dictionary and sparse weights simultaneously to recover the true signal. Nicholls et al.³⁹ provided a detailed analysis of the factors controlling the algorithm when applied to STEM data. We leave out the details of this specific method which is well documented in the literature.

3 | RESULTS

To test subsampling of simulations and to compare with existing simulation methods, compressed high angle annular dark field (HAADF) STEM simulations using the multislice and PRISM methods were calculated. In all cases, the simulations were performed using abTEM (version 1.0.0 beta 31) on a desktop computer equipped with an AMD Ryzen 5 2400G with Radeon Vega Graphics CPU @ 3.40 GHz, and one NVIDIA GeForce RTX 3060Ti GPU running CUDA 11.8. As already noted, it is important to note that all computation times are relative to the capa-

bility of abTEM and hence it is much better to consider the relative performance of the methods than the absolute computation times, as these are transferable to any other STEM simulation algorithm. All image recoveries were performed using our version of the BPFA algorithm written using CUDA so that it can be parallelised for maximum speed. The run-time of reconstructions is negligible for images of the sizes quoted (<2 s per reconstruction). Finally, each simulation is compared to a ground truth simulation using the Structural Similarity Index Measure (SSIM)⁴¹ and Peak Signal to Noise Ratio (PSNR).⁴² As a rule of thumb, it is considered that a PSNR value greater than 20 dB is an acceptable reconstruction, anything over 25 dB is a very good reconstruction, and anything over 30 dB is indistinguishable from the ground truth (see Figures S3–S5).

3.1 | Strontium titanate grain boundary

One important test of this method is its application to the calculation of contrast in grain boundaries. These complex structures are inherently difficult to interpret due to contrast variations within the void/boundary. For this purpose, a simulation of a low energy strontium titanate grain boundary (SrTiO₃ 22.6° Σ13(510)/[100]) was performed (see Yang et al.⁴³ for more details on the structural model, calculation, and experimental set up).

According to Yang et al., the grain boundary energy of the rigid body shift structure can be calculated through first principles, where the nonstoichiometric model is relaxed and shows a significant reduction in energy, to 0.81 ± 0.01 J/m². This model (Figure 5A) was chosen and simulated as shown in Figure 6.

HAADF simulations were performed using both the multislice and PRISM methods through abTEM. The accelerating voltage was set at 200 kV, with a probe-forming aperture semiangle of 24.5 mrad, and a collection semiangle of 70–190 mrad. The sample had a maximum depth of 4.5 nm (including a 5 Å amorphous carbon layer on top), and both the fully sampled and subsampled simulations were performed with 10 frozen phonon configurations. The subsampled simulations were acquired at 5% sampling per frozen phonon layer, and each of the masks used was an atomic-number-based targeted sampling mask. All simulations had a reciprocal space sampling of 0.6 nm⁻¹. Therefore, the only compression was acquired through spatial subsampling of each frozen phonon layer. All simulations were taken with 256 × 256 probe locations and each of the atoms in the input had a Debye–Waller factor determined from the method in Ref.⁴⁴ where model coefficients are modelled through phonon density-of-state curves (we use a temperature of 300K).

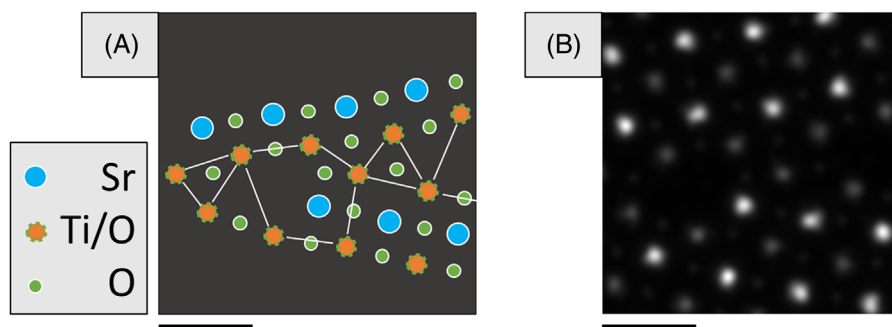


FIGURE 5 Model of the strontium titanate grain boundary (A) and a multislice simulation of the structure (B). The scale bar indicates 5 Å.

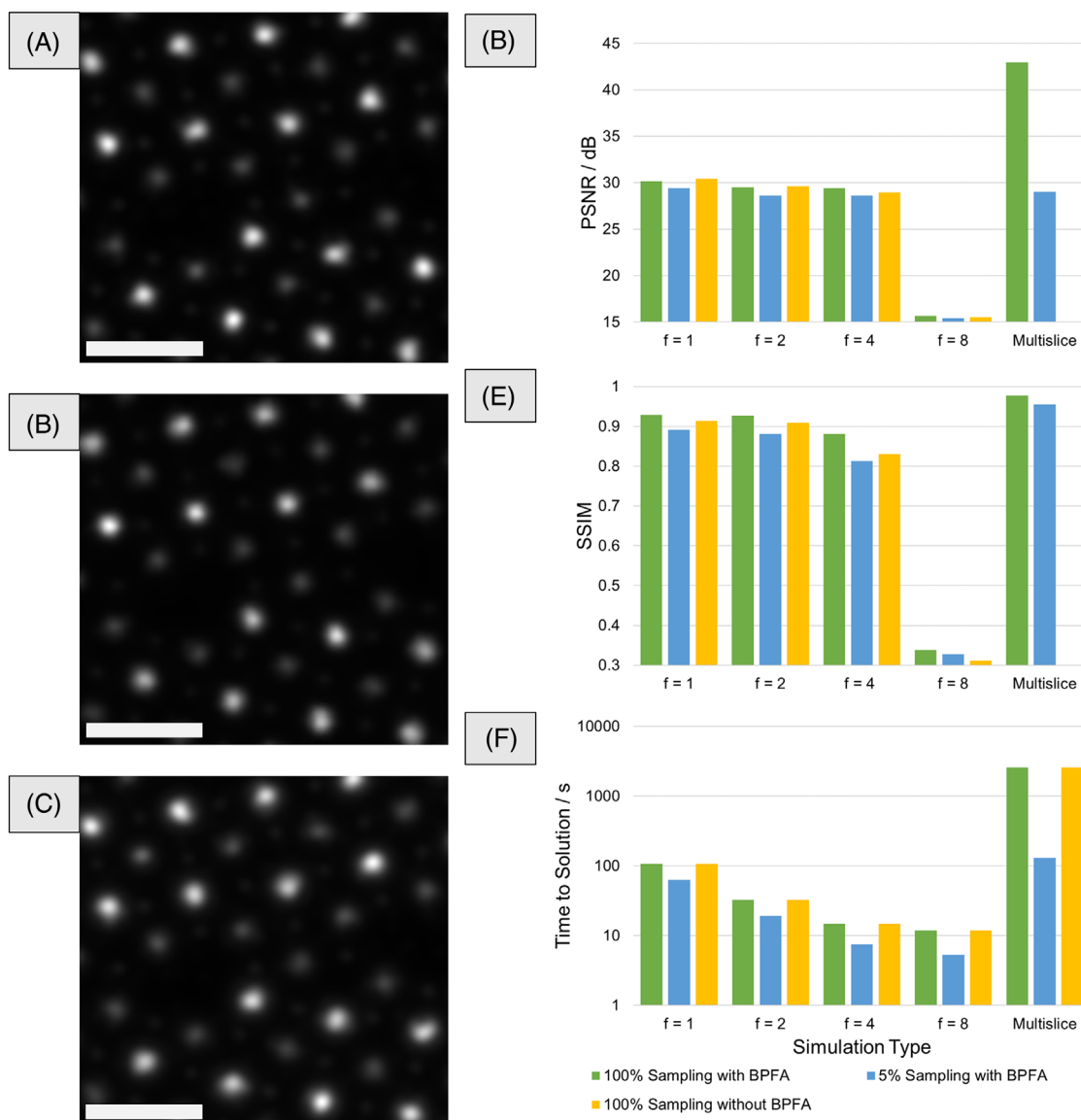


FIGURE 6 Reference simulation calculated using the multislice method (A), compressed simulation calculated using the multislice method (B), and simulation using the PRISM method with an interpolation factor of 2 (C) of the SrTiO₃ grain boundary structure. Plots of PSNR (D), SSIM (E) and computation times (F) of all the simulations. The term f refers to the interpolation factor. The scale bar in (A–C) indicates 0.5 nm. Full data sets can be found in Table S1 and Figure S1.

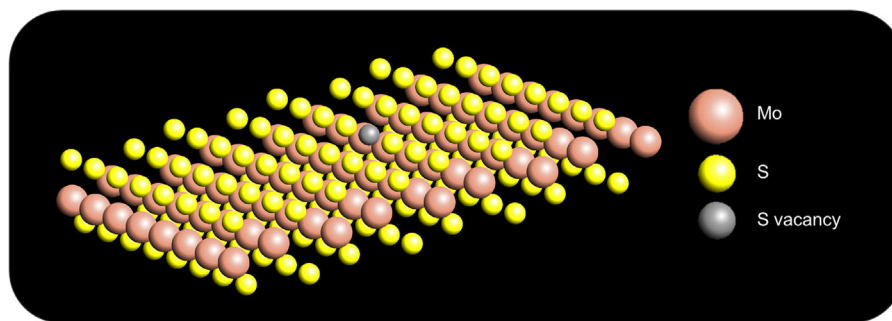


FIGURE 7 Schematic showing the monolayer 2H-MoS₂ structure with a V_S present. The model has been rotated such that the vacancy appears to sit on the top layer (for visibility); however, it was in fact removed from the bottom layer for the simulations. The graphic was rendered using the *OpenMX Viewer* toolbox.⁴⁸

The data in Figure 6 shows that subsampling a multislice simulation is only slightly slower (24s) than using the PRISM method with an interpolation factor of 1. Furthermore, the subsampled multislice simulation yields functionally results to PRISM ($f = 1$) with respect to the fully sampled multislice simulation (greater than 0.9 SSIM and greater than 28dB PSNR). Using the PRISM method with subsampling, we can achieve simulation times on the order of seconds (7.5 s) with an interpolation factor of 4 and achieve SSIM values greater than 0.8, and PSNR values greater than 28 dB. These values indicate a recovery that is not only functionally identical to the ground truth simulation, but of a high quality too.

3.2 | Monolayer molybdenum disulphide with monosulphur vacancies

2D materials are an active area of research within the electron microscopy community currently,^{45,46} so naturally their simulation is also important. 2D materials are popular for their use as semiconductors, and understanding their properties is important for the development of nano-electronic devices.

Our second test is therefore to determine whether our method can identify vacancies within the 2H phase of monolayer molybdenum disulphide (2H-MoS₂).⁴⁷ We specifically look at the monosulphur vacancy (V_S) case due to it having the lowest formation energy⁴⁷ (Figure 7).

Vacancies within the 2H-MoS₂ structure can dramatically change the mechanical and electrical properties of this semiconductor material.^{49–51} Hence, understanding the contrast through simulations is vital to determining the atomic structure to classify and statistically verify the abundance of defects and vacancies, not limited to just this specific example.

HAADF simulations were performed using both the multislice and PRISM methods through abTEM. The accelerating voltage was set at 60kV, with a 31 mrad

probe-forming aperture semiangle, and 86–200 mrad collection semiangle. Both the fully sampled, and subsampled simulations were performed with 10 frozen phonon configurations. The subsampled simulations were acquired at 5% sampling per frozen phonon layer, and each of the masks used was an atomic-number-based targeted sampling mask. All simulations had a reciprocal space sampling of 0.5 nm⁻¹, hence the only compression was acquired through spatial subsampling of each frozen phonon layer. All simulations were taken with 256 × 256 probe locations. The reference simulation refers to the fully sampled multislice simulation.

The results above are similar to those in Section 3.1 where the subsampled multislice method is faster than using PRISM with an interpolation factor of 1, and only slightly slower than an interpolation factor of 2. It also yields functionally identical results to both (approximately equal SSIM and PSNR values), as well as the fully sampled multislice simulation.

To validate that the monosulphur vacancy has been correctly simulated, an integrated line profile for simulations Figure 8A–C, as seen in Figure 9. The intensity profile across all three also shows functionally identical results, which agrees well with the experimentally observed results for contrast ratios of 0.5 : 1.0 : 2.3 for V_S:S:Mo, respectively.⁴⁷ The key difference however is that the PRISM method has a tailing effect in the vacuum regions, which is an artefact. This is due to the superposition of planewaves, where some of the frequencies required to compensate for this are missing from the scattering matrix. Subsampling the multislice simulation does not introduce such artefacts.

3.3 | Simultaneous theoretical and experimental recovery

We also consider the use of simulations to seed the recovery of subsampled experimental data. One aspect of faster

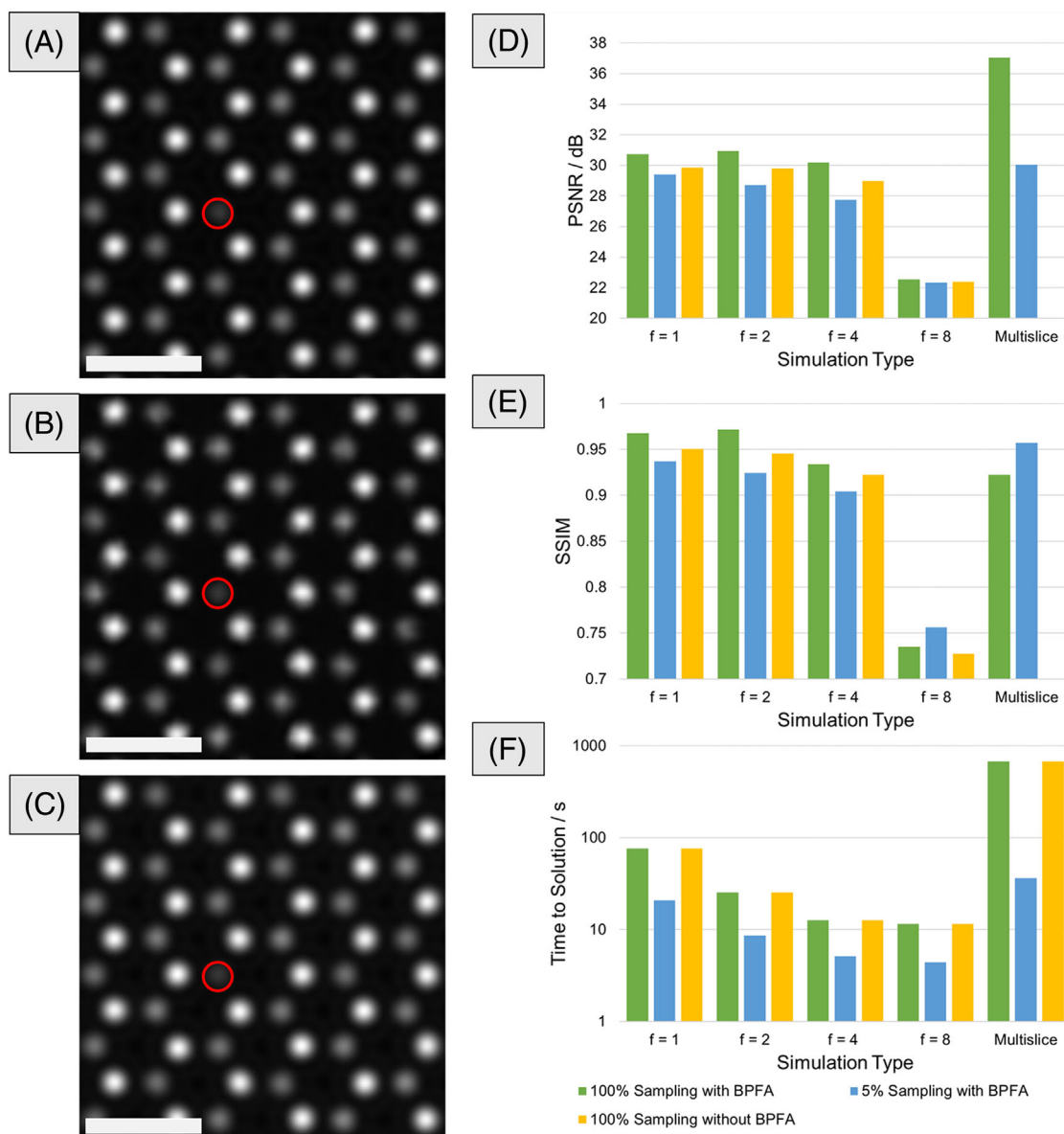


FIGURE 8 Reference simulation (A), compressed simulation calculated using the multislice method (B), and simulation using the PRISM method with an interpolation factor of 2 (C) of the $\text{H}_2\text{-MoS}_2$ structure with a V_s present (indicated by a red circle). Plots of PSNR (D), SSIM (E) and computation times (F) of all the simulations. Subsampling the multislice algorithm generally outperforms the PRISM method at full sampling in terms of quality (based on PSNR and SSIM values) and computation time. The term f refers to the interpolation factor. The scale bar in (A–C) indicates 5\AA . Full data sets can be found in Table S1 and Figure S2.

simulations is that matching simulation to experiment is more time efficient. This means more parameter testing can be performed in a shorter time frame. Given that the theory of electron scattering is very well understood,⁵² it would make sense to use simulations in a more practical aspect during acquisition. One of the ways this can be done is through (dictionary) *transfer learning*⁵³ where the dictionary from a simulation is used to seed the recovery of real subsampled data.

To test this method, we consider an yttrium silicide (Y_5Si_3) sample. The following paragraph is for complete-

ness and can be skipped if the reader is familiar with the properties of the sample.

Yttrium silicide is part of the electride class of compound materials. An electride is a framework composed cation and anion sublattices. These sublattices have a net positive electric charge which are balanced by loosely bonded, interstitial anionic electrons.⁵⁴ Y_5Si_3 has been well proposed as a low Schottky barrier material for n-type silicon semiconductors thanks to its low Schottky barrier height of 0.27 eV .⁵⁵ It has also been recently proposed as an encapsulation material to capture radioactive volatile products

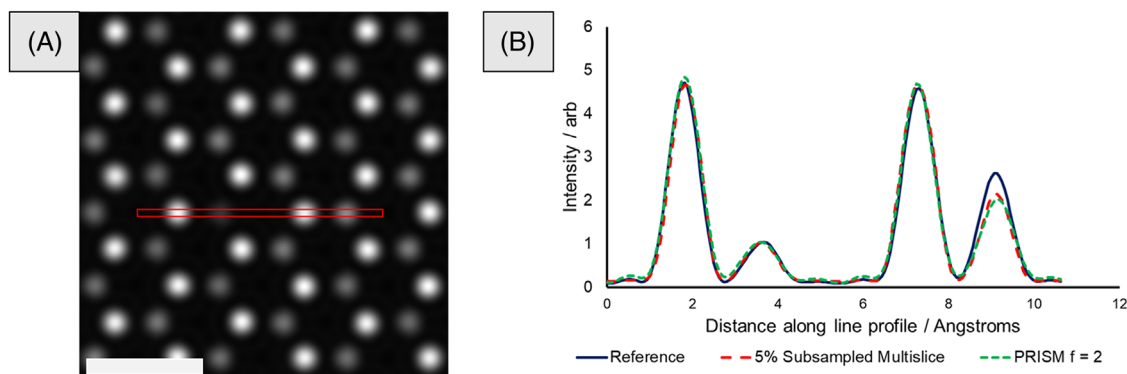


FIGURE 9 Integrated intensity line profiles (B) of the images Figure 7a–c over the region marked by the red box in (A). The first peak and third peak (from left to right) are molybdenum sites, the second peak is the sulphur vacancy site, and the fourth is a sulphur site. Scale bar indicates 5 Å.

within nuclear fission reactors.⁵⁶ All of this makes Y_5Si_3 a versatile material, hence understanding its properties are important. A recent paper from Q. Zheng et al.⁵⁴ looked at the local charge density of Y_5Si_3 by performing differential phase contrast experiments. They also used simulations to verify the contrast in their HAADF STEM images, specifically the missing contrast of the silicon atoms. While the Zheng paper performed a matching between the experiment and theory, in this example we are going to consider the simulation as the *seed* for recovering the real STEM data; that is, they are solved and matched simultaneously.

A simulation (matching experimental parameters; see Ref.⁵⁴ for details) of Y_5Si_3 was performed and then recovered using BPFA. The dictionary which is learned is then used to recover a subsampled acquisition of Y_5Si_3 through an orthogonal matching pursuit (OMP) algorithm.⁴⁰ We also show the result by passing the same acquisition through only BPFA for comparison.

Figure 10 shows that by using transfer learning we can improve the reconstruction quality by 2 dB. This is likely due to the simulation being free of noise, and therefore its dictionary is free of noise. Given that OMP approximates the correct weights to apply to each of the atoms, it can rescale the intensity if needed, matching directly to the real subsampled data. However, BPFA must learn a dictionary from the real subsampled data, which at low sampling rates can sometimes be challenging if the noise levels are also high, the dictionary transfer approach can show an advantage in the quality of reconstruction. This is demonstrated in the differences between Figure 10D and E where the yttrium columns are more refined using dictionary transfer than using BPFA alone. This is particularly important, as sampling at 3%^{57,58} using a scan generator⁵⁹ correlates to a 33× speed up in image acquisition and 33× less total electron dose during the experiment. For beam-sensitive materials this is excellent as the images acquired will be more representative of the pristine sample due to

knock-on and radiolysis damage occurring fewer times overall.⁶⁰ For less beam-sensitive materials, the speed up in acquisition means fewer artefacts in the image caused by stage drift.

In practice, however, it is found that *line-hop* sampling (i.e. random walk)^{61,62} is a better alternative sampling strategy for experimental data and it balances sparsity of acquisition and the limiting effects of hysteresis. UDS sampling is the optimal set up for image recovery,³⁵ however is limited in experiment due to hysteresis.⁶¹ Comparison of UDS and line-hop can be found in previous work⁵⁷ and readers are referred to Figure S6 for examples of different mask types.

4 | DISCUSSION AND CONCLUSIONS

Our method has been implemented using both the PRISM and multislice algorithms, demonstrating a robustness of the subsampling approach to increasing speed of simulations without significant loss of accuracy. It is also competitive in terms of its performance to the state-of-the-art method for faster simulation (PRISM). This highlights the effectiveness of sparse acquisition, in that much of the data requirements for STEM can be significantly reduced. We only require a subset of the data to recover a functionally identical result through image inpainting.

We have found with the experimental images that using a subsampling strategy need not be independent from other helpful approaches – other strategies for noise reduction, super resolution etc. can be applied to the inpainted image. We have found the same effect here with the simulations where the interpolation approach of PRISM can be used in conjunction with subsampling at a lower interpolation factor without significant loss of information. For example, a simulation with an interpolation factor of 4 and subsampling at 5% gives significantly better results

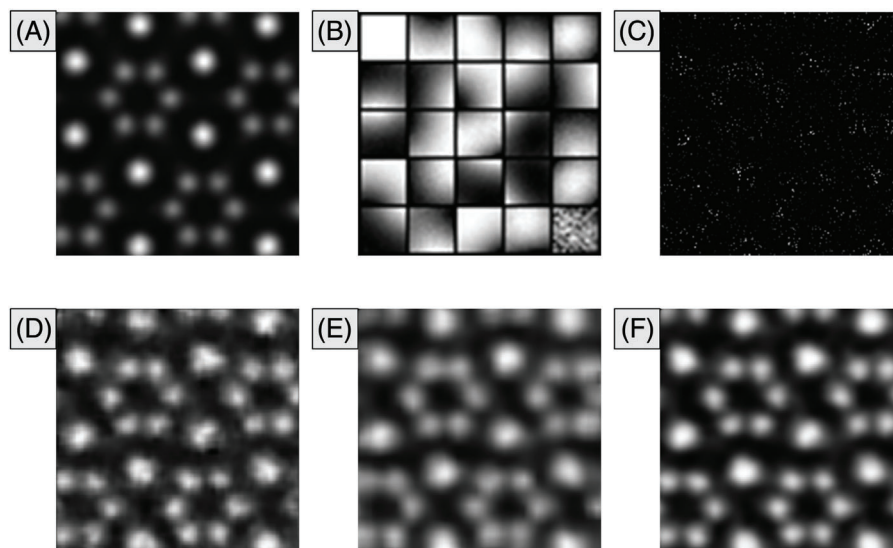


FIGURE 10 Recovery from 5% subsampled multislice simulation of Y_3Si_5 (A), and the dictionary determined by BPFA (B). This dictionary is then used to reconstruct a 3% subsampled acquisition of Y_3Si_5 (C) giving a reconstruction through OMP (D) with a PSNR of 24.8 dB and an SSIM of 0.87. Reconstruction using only BPFA to learn the dictionary and reconstruct at the same sampling rate (E) with a PSNR of 22.8 dB and an SSIM of 0.86. Both comparisons are made to the ground truth which was passed through BPFA at 100% sampling to denoise only (F).

than a simulation with an interpolation factor of 8 at 100% sampling, running almost twice as fast in the process.

Although the demonstrations here have only shown results for HAADF simulations, the methods described are applicable to all STEM imaging modes such as bright field, annular bright field, and 4D-STEM. For example, 4D-STEM can approximate sample thickness using position average convergent beam electron diffraction by matching simulation to experiment.⁵⁴

Furthermore, we have shown that STEM simulations can seed the dictionary for real image reconstruction, meaning faster ‘live’ reconstructions from experimental CS-STEM data. In the BPFA algorithm it is possible to begin the dictionary learning from a custom dictionary. By seeding the initial dictionary from the dictionary acquired from a simulation of the material in question, it could potentially speed up the convergence of the algorithm, or equally make the live reconstructions more accurate after fewer iterations. We also showed that transfer learning with an OMP can give functionally identical results to the ground truth. This can be even faster to recover than just using BPFA to reconstruct, indicating further speed improvements.

In conclusion, subsampling a multislice simulation can be both faster and a better representation of the full multislice simulation than the PRISM method at 100% sampling (depending on the image quality metric used). This is analogous to using fewer but more intense probes in real CS-STEM, as opposed to a full raster scan at low dose acquisition.⁵⁵ This shows that in some cases it is better

to have a more accurate calculation of individual probes than interpolating each probe estimate over the full scan area.


Transferring the dictionary learned from a simulation can also yield better results than blind inpainting the raw acquisition. This uses theory and experimental data together as opposed to just matching and comparing results. For microscopists applying this in a practical sense, it would speed up analysis of the imaging conditions, allowing for faster adjustment and hence less total electron dose on the sample. This could improve the final image quality for many materials by reducing the amount of sample damage, driving forward new science in beam-sensitive materials.

As noted, the purpose of faster simulation is not to improve their accuracy per se, but to improve the efficiency of calculation for faster determination of properties and characteristics in conjunction with experimental data. In the case where the accuracy of experimental data is not perfect, we propose that the simulation itself does not need to be perfect if time is a constraint. Of course, one could calculate an improved simulation for in-depth analysis, but if it is possible to realise real-time simulations then they could be used to assist a microscopist during acquisition. It may also be possible for fast STEM simulations (in conjunction with deep learning) to interpret the correct adjustments needed for real-time acquisition automatically. This would allow for both faster and more efficient alignment, which is of particular importance to studies of beam-sensitive materials.

ACKNOWLEDGEMENTS

This work was performed in the Albert Crewe Centre (ACC) for Electron Microscopy, a shared research facility (SRF) fully supported by the University of Liverpool. This work was also funded by the EPSRC Centre for Doctoral Training in Distributed Algorithms (EP/S023445/1), Sivananthan Labs, and Rosalind Franklin Institute. M.C. would like to acknowledge the support by the US DOE Office of Science Early Career project FWP# ERK CZ55 and the Center for Nanophase Materials Sciences (CNMS), a US DOE Office of Science User Facility. The authors would also like to recognise the efforts of Jacob Madsen and Toma Susi for their development of abTEM, as well as Colin Ophus and Alan Pryor Jr for their development of the PRISM algorithm.

ORCID

Alex W. Robinson  <https://orcid.org/0000-0002-1901-2509>

Amirafshar Moshtaghpour  <https://orcid.org/0000-0002-6751-2698>

REFERENCES

- Pennycook, S. J. (2012). Seeing the atoms more clearly: STEM imaging from the Crewe era to today. *Ultramicroscopy*, 123, 28–37.
- Crewe, A. V. (1984). An introduction to the STEM. *Journal of Ultrastructure Research*, 88(2), 94–104.
- Crewe, A. V. (1974). Scanning transmission electron microscopy. *Journal of microscopy*, 100(3), 247–259.
- Beck, V., & Crewe, A. V. (1975). High resolution imaging properties of the STEM. *Ultramicroscopy*, 1(2), 137–144.
- Cowley, J. M., & Moodie, A. F. (1957). The scattering of electrons by atoms and crystals. I. A new theoretical approach. *Acta Crystallographica*, 10(10), 609–619.
- Goodman, P., & Moodie, A. F. (1974). Numerical evaluations of N-beam wave functions in electron scattering by the multislice method. *Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography*, 30(2), 280–290.
- Ishizuka, K. (2002). A practical approach for STEM image simulation based on the FFT multislice method. *Ultramicroscopy*, 90(2–3), 71–83.
- Ishizuka, K., & Uyeda, N. (1977). A new theoretical and practical approach to the multislice method. *Acta Crystallographica Section A*, 33(5), 740–749.
- Kirkland, E. J. (2010). *Advanced computing in electron microscopy* (2nd ed.). New York: Springer.
- Kirkland, E. J., Loane, R. F., & Silcox, J. (1987). Simulation of annular dark field stem images using a modified multislice method. *Ultramicroscopy*, 23(1), 77–96.
- Stadelmann, P. A. (1987). EMS – A software package for electron diffraction analysis and HREM image simulation in materials science. *Ultramicroscopy*, 21(2), 131–145.
- Stadelmann, P. (2003). Image analysis and simulation software in transmission electron microscopy. *Microscopy and Microanalysis*, 9(S03), 60–61.
- Kilaas, R. (1992). MacTempas a program for simulating high resolution TEM images and diffraction patterns. <http://www.totalresolution.com/>
- Koch, C. T. (2002). *Determination of core structure periodicity and point defect density along dislocations*. Arizona.
- de Graef, M. (2003). *Introduction to conventional transmission electron microscopy*. Cambridge University Press.
- Zuo, J. M., & Mabon, J. C. (2004). Web-based electron microscopy application software: Web-EMAPS. *Microscopy and Microanalysis*, 10(S02), 1000–1001.
- Carlino, E., Grillo, V., & Palazzari, P. (2008). Accurate and fast multislice simulations of HAADF image contrast by parallel computing. In *Microscopy of Semiconducting Materials 2007* (pp. 177–180). Springer.
- Rosenauer, A., & Schowalter, M. (2008). STEMSIM – A new software tool for simulation of STEM HAADF Z-contrast imaging. In *Microscopy of Semiconducting Materials 2007*. Dordrecht: Springer Netherlands.
- Walton, S. K., Zeissler, K., Branford, W. R., & Felton, S. (2013). MALTS: a tool to simulate Lorentz transmission electron microscopy from micromagnetic simulations. *IEEE Transactions on Magnetics*, 49(8), 4795–4800.
- Bar-Sadan, M., Barthel, J., Shtrikman, H., & Houben, L. (2012). Direct imaging of single Au atoms within GaAs nanowires. *Nano Letters*, 12(5), 2352–2356.
- Barthel, J. (2018). Dr. Probe: A software for high-resolution STEM image simulation. *Ultramicroscopy*, 193, 1–11.
- Lobato, I., & van Dyck, D. (2015). MULTTEM: A new multislice program to perform accurate and fast electron diffraction and imaging simulations using Graphics Processing Units with CUDA. *Ultramicroscopy*, 156, 9–17.
- Lobato, I., van Aert, S., & Verbeeck, J. (2016). Progress and new advances in simulating electron microscopy datasets using MULTTEM. *Ultramicroscopy*, 168, 17–27.
- van den Broek, W., Jiang, X., & Koch, C. T. (2015). FDES, a GPU-based multislice algorithm with increased efficiency of the computation of the projected potential. *Ultramicroscopy*, 158, 89–97.
- Cosgriff, E. C., D'Alfonso, A. J., Allen, L. J., Findlay, S. D., Kirkland, A. I., & Nellist, P. D. (2008). Three-dimensional imaging in double aberration-corrected scanning confocal electron microscopy, Part I. *Ultramicroscopy*, 108(12), 1558–1566.
- Oelerich, J. O., Duschek, L., Belz, J., Beyer, A., Baranovskii, S. D., & Volz, K. (2017). STEMSalabim: A high-performance computing cluster friendly code for scanning transmission electron microscopy image simulations of thin specimens. *Ultramicroscopy*, 177, 91–96.
- Pryor, A., Ophus, C., & Miao, J. (2017). A streaming multi-GPU implementation of image simulation algorithms for scanning transmission electron microscopy. *Advanced Structural and Chemical Imaging*, 3(1), 1–14.
- Madsen, J., & Susi, T. (2021). The abTEM code: Transmission electron microscopy from first principles. *Open Research Europe*, 1(24), 24.

29. Dwyer, C. (2010). Simulation of scanning transmission electron microscope images on desktop computers. *Ultramicroscopy*, 110(3), 195–198.
30. Robinson, A., Nicholls, D., Wells, J., Moshtaghpour, A., Kirkland, A., & Browning, N. D. (2022). SIM-STEM Lab: Incorporating compressed sensing theory for fast STEM simulation. *Ultramicroscopy*, 242, 113625.
31. Robinson, A., Nicholls, D., Wells, J., Moshtaghpour, A., Kirkland, A., & Browning, N. D. (2022). Compressed STEM Simulations. *Microscopy and Microanalysis*, 28(S1), 3116–3117.
32. Stevens, A., Yang, H., Carin, L., Arslan, I., & Browning, N. D. (2014). The potential for Bayesian compressive sensing to significantly reduce electron dose in high-resolution STEM images. *Microscopy*, 63(1), 41–51.
33. Donati, L., Nilchian, M., Trépout, S., Messaoudi, C., Marco, S., & Unser, M. (2017). Compressed sensing for STEM tomography. *Ultramicroscopy*, 179, 47–56.
34. Kovarik, L., Stevens, A., Liyu, A., & Browning, N. D. (2016). Implementing an accurate and rapid sparse sampling approach for low-dose atomic resolution STEM imaging. *Applied Physics Letters*, 109(16), 164102.
35. Carin, L., Liu, D., & Guo, B. (2011). Coherence, compressive sensing, and random sensor arrays. *IEEE Antennas and Propagation Magazine*, 53(4), 28–39.
36. Pérez, E., Kirchhof, J., Krieg, F., & Römer, F. (2020). Subsampling approaches for compressed sensing with ultrasound arrays in non-destructive testing. *Sensors*, 20(23), 6734.
37. Arridge, S., Beard, P., Betcke, M., Cox, B., Huynh, N., Lucka, F., Ogunlade, O., & Zhang, E. (2016). Accelerated high-resolution photoacoustic tomography via compressed sensing. *Physics in Medicine & Biology*, 61(24), 8908.
38. Sertoglu, S., & Paisley, J. (2015). Scalable Bayesian nonparametric dictionary learning. In *2015 23rd European Signal Processing Conference (EUSIPCO)*, Nice, France, pp. 2771–2775.
39. Nicholls, D., Wells, J., Stevens, A., Zheng, Y., Castagna, J., & Browning, N. D. (2022). Sub-sampled imaging for STEM: maximising image speed, resolution and precision through reconstruction parameter refinement. *Ultramicroscopy*, 233, 113451.
40. Tropp, J. A., & Gilbert, A. C. (2007). Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Transactions on Information Theory*, 53(12), 4655–4666.
41. Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4), 600–612.
42. Horé, A., & Ziou, D. (2013). Is there a relationship between peak-signal-to-noise ratio and structural similarity index measure? *IET Image Processing*, 7(1), 12–24.
43. Yang, H., Lee, H. S., Sarahan, M. C., Sato, Y., Chi, M., Moeck, P., Ikuhara, Y., & Browning, N. D. (2013). Quantifying stoichiometry-induced variations in structure and energy of a SrTiO₃ symmetric Σ13 {510}/<100> grain boundary. *Philosophical Magazine*, 93(10–12), 1219–1229.
44. Sears, V. F., & Shelley, S. A. (1991). Debye–Waller factor for elemental crystals. *Acta Crystallographica Section A: Foundations of Crystallography*, 47(4), 441–446.
45. Mas-Balleste, R., Gomez-Navarro, C., Gomez-Herrero, J., & Zamora, F. (2011). 2D materials: To graphene and beyond. *Nanoscale*, 3(1), 20–30.
46. Novoselov, K. S., Mishchenko, O. A., Carvalho, O. A., & Castro Neto, A. H. (2016). 2D materials and van der Waals heterostructures. *Science*, 353(6298), aac9439.
47. Zhou, W., Zou, X., Najmaei, S., Liu, Z., Shi, Y., Kong, J., Lou, J., Ajayan, P. M., Yakobson, B. I., & Idrobo, J. C. (2013). Intrinsic structural defects in monolayer molybdenum disulfide. *Nano Letters*, 13(6), 2615–2622.
48. Lee, Y. T., & Ozaki, T. (2019). OpenMX Viewer: A web-based crystalline and molecular graphical user interface program. *Journal of Molecular Graphics and Modelling*, 89, 192–198.
49. Mak, K. F., Lee, C., Hone, J., Shan, J., & Heinz, T. F. (2010). Atomically thin MoS₂: A new direct-gap semiconductor. *Physical Review Letters*, 105(13), 136805.
50. Radisavljevic, B., Radenovic, A., Brivio, J., Giacometti, V., & Kis, A. (2011). Single-layer MoS₂ transistors. *Nature Nanotechnology*, 6(3), 147–150.
51. Yin, Z., Li, H., Li, H., Jiang, L., Shi, Y., Sun, Y., Lu, G., Zhang, Q., Chen, X., & Zhang, H. (2012). Single-layer MoS₂ phototransistors. *ACS Nano*, 6(1), 74–80.
52. Pennycook, S. J. (1989). Z-contrast STEM for materials science. *Ultramicroscopy*, 30(1–2), 58–69.
53. Pan, S. J., & Yang, Q. (2009). A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10), 1345–1359.
54. Zheng, Q., Feng, T., Hachtel, J. A., Ishikawa, R., Cheng, Y., Daemen, L., Xing, J., Idrobo, J. C., Yan, J., Shibata, N., Ikuhara, Y., Sales, B. C., Pantelides, S. T., & Chi, M. (2021). Direct visualization of anionic electrons in an electride reveals inhomogeneities. *Science Advances*, 7(15), eabe6819.
55. Isogai, T., Tanaka, H., Goto, T., Teramoto, A., Sugawa, S., & Ohmi, T. (2008). Formation and property of yttrium and yttrium silicide films as low Schottky barrier material for n-type silicon. *Japanese Journal of Applied Physics*, 47(4S), 3138.
56. Kuganathan, N., Chronos, A., & Grimes, R. W. (2021). One-dimensional yttrium silicide electride (Y₅Si₃: e[−]) for encapsulation of volatile fission products. *Journal of Applied Physics*, 129(24), 245105.
57. Nicholls, D., Robinson, A., Wells, J., Moshtaghpour, A., Bahri, M., Kirkland, A., & Browning, N. (2022). Compressive scanning transmission electron microscopy. In *ICASSP 2022–2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Singapore, pp. 1586–1590.
58. Nicholls, D., Wells, J., Robinson, A. W., Moshtaghpour, A., Kobylinska, M., Fleck, R. A., Kirkland, A. I., & Browning, N. D. (2022). A targeted sampling strategy for compressive cryo focused ion beam scanning electron microscopy. arXiv preprint, arXiv:2211.03494.
59. Direct Electron. (2021). Direct electron, LP. In *Microscopy today* (Vol. 29, pp. 12–12). Cambridge University Press.
60. Egerton, R. F. (2019). Radiation damage to organic and inorganic specimens in the TEM. *Micron*, 119, 72–87.
61. Browning, N. D., Castagna, J., Kirkland, A. I., Moshtaghpour, A., Nicholls, D., Robinson, A. W., Wells, J., & Zheng, Y. (2023). The advantages of sub-sampling and inpainting for scanning transmission electron microscopy. *Applied Physics Letters*, 122(5), 050501.
62. Nicholls, D., Lee, J., Amari, H., Stevens, A. J., Mehdi, B. L., & Browning, N. D. (2020). Minimising damage in high resolution

scanning transmission electron microscope images of nanoscale structures and processes. *Nanoscale*, 12(41), 21248–21254.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Robinson, A. W., Wells, J., Nicholls, D., Moshtaghpour, A., Chi, M., Kirkland, A. I., & Browning, N. D. (2023). Towards real-time STEM simulations through targeted subsampling strategies. *Journal of Microscopy*, 290, 53–66. <https://doi.org/10.1111/jmi.13177>