

Framework of compressive sensing and data compression for 4D-STEM

Hsu-Chih Ni^{a,b}, Renliang Yuan^c, Jiong Zhang^c, Jian-Min Zuo^{a,b,*}

^a Department of Materials Science and Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

^b Materials Research Laboratory, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

^c Intel Corporation, Corporate Quality Network, Hillsboro, OR 97124, USA

ARTICLE INFO

Keywords:

4D-STEM
Compressive sensing
Data compression
Strain mapping

ABSTRACT

Four-dimensional Scanning Transmission Electron Microscopy (4D-STEM) is a powerful technique for high-resolution and high-precision materials characterization at multiple length scales, including the characterization of beam-sensitive materials. However, the field of view of 4D-STEM is relatively small, which in absence of live processing is limited by the data size required for storage. Furthermore, the rectilinear scan approach currently employed in 4D-STEM places a resolution- and signal-dependent dose limit for the study of beam sensitive materials. Improving 4D-STEM data and dose efficiency, by keeping the data size manageable while limiting the amount of electron dose, is thus critical for broader applications. Here we introduce a general method for reconstructing 4D-STEM data with subsampling in both real and reciprocal spaces at high fidelity. The approach is first tested on the subsampled datasets created from a full 4D-STEM dataset, and then demonstrated experimentally using random scan in real-space. The same reconstruction algorithm can also be used for compression of 4D-STEM datasets, leading to a large reduction (100 times or more) in data size, while retaining the fine features of 4D-STEM imaging, for crystalline samples.

1. Introduction

Scanning transmission electron microscopy (STEM) is a powerful tool for materials characterization. By scanning a focused electron probe across a sample region and collecting scattered electrons using a detector, a real-space image is formed in various forms of contrast, from atomic resolution Z-contrast to annular bright-field imaging of light atoms to differential phase contrast [1,2]. However, only a part of the scattered electrons is collected, and the electron momentum is not resolved, or only partially resolved. Most of diffraction information about the material is thus lost in STEM imaging. The emerging 4D-STEM collects electron diffraction patterns at each probe position using a pixelated detector. This technique combines the advantages of electron diffraction with STEM imaging, which is becoming more and more popular due to the advent of new detector technologies [3–6]. Studies have shown that a number of materials structural properties can be mapped by 4D-STEM, including but not limited to strain mapping [7–9], orientation mapping [10–12], and atomic arrangements using atomic resolution ptychography [13,14].

However, current 4D-STEM analysis is constrained by several factors. The first is the amount of beam dwell time to record a diffraction

pattern, which is milliseconds or more with the current generation of direct electron detectors [3,4,6]. The dwell time could be even longer using cameras with a large pixel format and a limited data readout rate. Typically, the amount of electron dose on the sample in a 4D-STEM experiment can be orders larger compared with conventional STEM. Second, for a fixed pixel resolution, the size of 4D-STEM datasets is proportional to the scan size, which determines the field of view, and thus the areas that can be analyzed by 4D-STEM is practically limited by the data size, in addition to long acquisition times. These limits must be overcome to improve the throughput of 4D-STEM analysis and to extend 4D-STEM applications to a broad range of materials, especially, to samples susceptible to beam damage.

The current implementation of 4D-STEM samples the real and reciprocal space uniformly using the STEM raster scan and a pixelated area detector. The amount of sampling follows the Nyquist sampling theorem at a rate that is twice the Nyquist frequency or higher. This sampling strategy ignores spatial correlations at different sample locations and spectral correlations among diffraction patterns. It creates massive amounts of data that are only to be used partially for image formation. This process is thus extremely inefficient and wasteful in the use of microscopy and data-storage resources and the electron dose.

* Corresponding author at: Department of Materials Science and Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA.
E-mail address: jianzuo@illinois.edu (J.-M. Zuo).

Compressive sensing [15–26] offers an alternative sampling strategy against the traditional Nyquist sampling. This strategy has attracted large interest in the electron microscopy community for overcoming electron dose limitations in the STEM spectroscopy analysis lately [16, 27, 19, 28]. The general idea is that signals obtained from samples are often sparse and its representation as a high-dimensional vector could be described by a small number of basis signal vectors (or endmembers). When the measurements taken are spread out across different sparse components of the signal, the sampling rate at much lower than the Nyquist sampling limit can be sufficient for recovering the original signal [22, 29]. Applications have demonstrated that compressive sensing could greatly reduce the amount of the electron dose and acquisition time for STEM imaging and spectroscopy mapping. For example, Andrew et al. developed a compressive sensing method based on the beta-process factor-analysis and applied it on the beam sensitive ZSM-5 zeolite sample [18]. Their results showed that with only 10 % of sampling, most of the information regarding the atomic columns and varying contrast in the ZSM-5 zeolite was retained in the reconstructed STEM images. Compressive sensing has also been explored for STEM tomography to reduce the number of sampling points and thus lowering the tilt-series acquisition time. A strategy for acquiring and restoring the sub-sampled STEM tomography datasets with large tilt steps and partially scanned STEM images was demonstrated by Saghi et al. [26]. For electron spectroscopy, Monier et al. reconstructed randomly sampled EELS maps by regularizing spatial smoothness and enforcing spectral sparsity by hard-thresholded principal component analysis [27]. An EELS spectra map of biological tissue was obtained with the method. Recently, the combination of compressive sensing and

4D-STEM has been explored, particularly in the context of ptychography. Robinson et al. applied Beta Process Factor Analysis to a 4D-STEM dataset, automatically learning a dictionary in which virtual images are sparse [30]. Moshtaghpour et al. utilized L0-regularization to reduce the number of sampling points for ptychography reconstructions, demonstrating that even with 4 % or lower subsampling, high-quality object phase and amplitude images can be obtained [31].

An added bonus of compressive sensing is dataset compression, which is naturally made. With fast detectors and opportunities for acquiring datasets from large field of view, a 4D-STEM experiment entails large volumes of data. The storage of these large datasets is problematic, especially when multiple analyzes are made. This emerging data issue has been discussed under the context of direct electron detectors and cryogenic electron microscopy [32], but has yet to be addressed regarding 4D-STEM. In compressive sensing, data can be decomposed into a small set of endmembers and their coefficients. The sizes of these two matrices can be significantly smaller than the original dataset, and thus greatly reduce the amount of stored data. To find the base vectors and their coefficients, a decompression algorithm is needed. This problem is the same as the reconstruction used in hyperspectral compressive imaging, for which different algorithms have been proposed [33–37].

In this paper, we introduce a dual-space (real and reciprocal spaces) compressive sensing (DSCS) scheme for 4D-STEM and demonstrate its use in compressive sensing for data collection and data compression of 4D-STEM datasets. The data collected by 4D-STEM is similar to hyperspectral imaging in remote sensing, in the sense that both form a stack of images using the spectral signals of specific wavelength as in

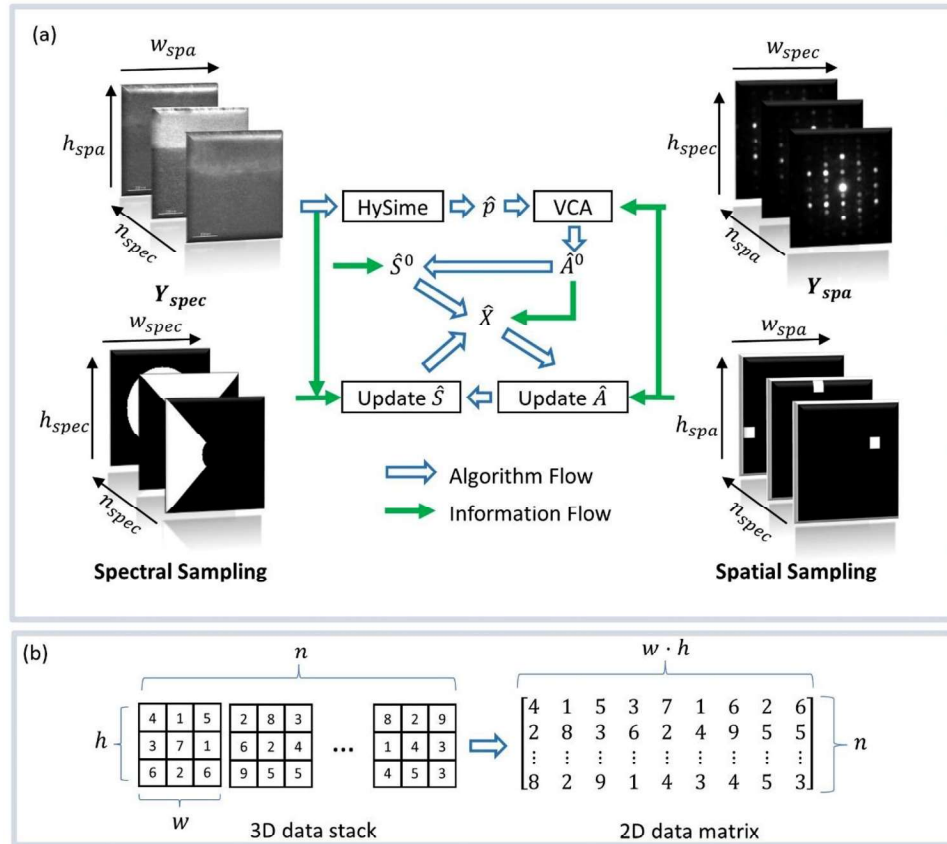


Fig. 1. (a) 4D-STEM CS and reconstruction algorithm based on SSCR-SU. The number of endmembers $\hat{p}$ is first estimated from Y_{spec} (the STEM images) stack using HySime. Then vertex component analysis is used with Y_{spa} to estimate the endmembers matrix $\hat{A}^0$. Abundance matrix $\hat{S}^0$ is then estimated. The multiplication of $\hat{S}^0$ and $\hat{A}^0$ gives $\hat{X}$, Which is the estimated full 4D-STEM dataset. $\hat{X}$ is subsequently optimized by updating $\hat{A}$ and $\hat{S}$ sequentially. (b) 3D data stacks are all vectorized into 2D data matrices before feeding into the algorithm.

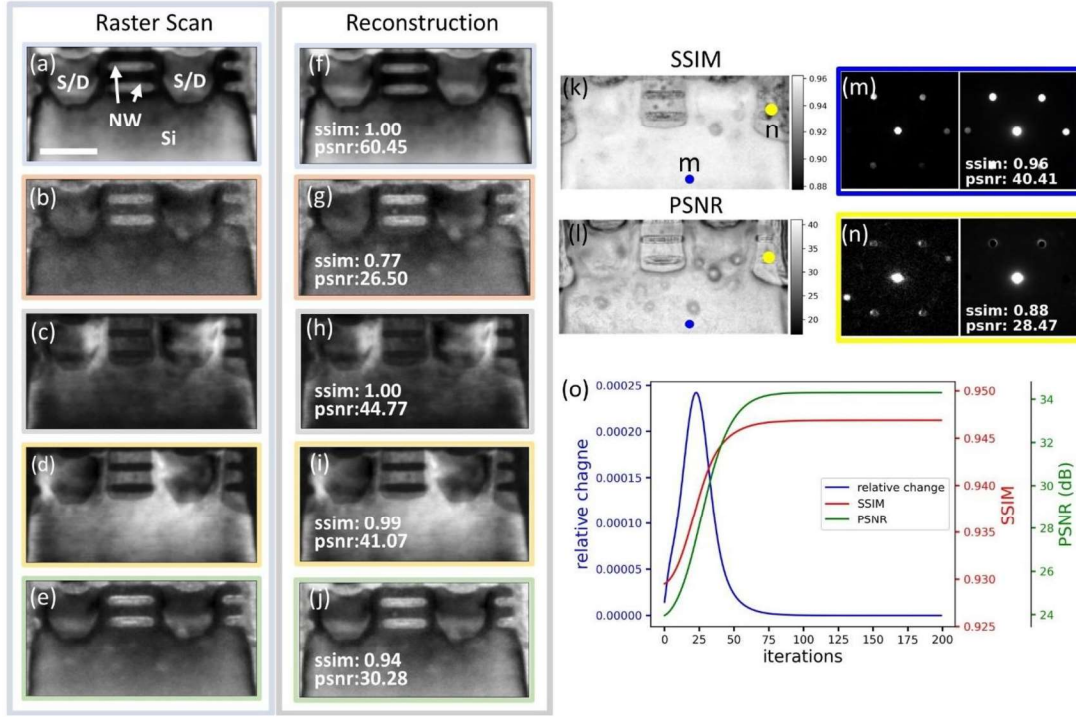


Fig. 2. STEM images created with virtual detectors. (a–e) virtual bright-field and dark-field images created from the original 4D-STEM dataset. The positions of source/drain (S/D), nanowire channel (NW), and Si subfin (Si) is labeled in (a). (f–j) virtual dark field images created from the reconstructed 4D-STEM dataset. The size of the detectors used for (a–j) corresponds to camera length of 4300 nm. (k, l) SSIM and PSNR map between the original dataset and the reconstructed dataset. The value of each point represents the similarity metrics value between the original (left) and reconstructed (right) diffraction pattern. (m, n) diffraction patterns extracted from original and reconstructed datasets. The spatial positions of these diffraction patterns are marked as yellow and blue dot in (k, l), respectively. (o) Change of average similarity metrics with respect to number of iterations of joint optimization.

hyperspectral images and of electron momentum in 4D-STEM. On the terminology, we use spectral interchangeably for diffraction space in line with the conventions in the signal processing community. The spatial-spectral compressed reconstruction based on the spectral unmixing (SSCR-SU) algorithm proposed by Wang et al. [33] for remote sensing is adapted and developed for 4D-STEM. This algorithm uses linear mixing model to decomposes hyperspectral images into the endmember and abundance matrices, whose sizes are much smaller than that of the original data and reconstruction from these can be done efficiently. The hyperspectral images are sampled both spectrally and spatially. The joint endmember extraction and abundance estimation problem is solved iteratively to obtain a reconstructed hyperspectral image. Applications to synthetic and real hyperspectral data have demonstrated that the algorithm is effective in obtaining the endmember and abundance information, and the performance is superior to other state-of-art algorithms in term of the accuracy of reconstructed hyperspectral images, as well as the computational efficiency. The DSCS method introduced here follows the concepts of SSCR-SU by taking STEM images and diffraction patterns on the designed spectral and spatial sampling or measurement matrices and uses them as inputs to reconstruct the full 4D-STEM dataset.

For the measurement matrices, we explored two sampling schemes. Real-space random scan based on Bernoulli sampling is used to obtain a set of diffraction patterns in both schemes. In the spectral (diffraction) space, one scheme uses images formed using segmented STEM detectors, and the other scheme uses random detector pixels for imaging. The diffraction patterns and images are collected separately. The use of segmented STEM detectors simplifies the DSCS implementation since it only requires a STEM capable of random scan. Once the DSCS datasets are acquired, endmembers are extracted from the collected diffraction patterns, while the abundance of each endmember is estimated from the recorded STEM images. We then combine the endmembers and their

abundance to reconstruct the full 4D-STEM data. We compare the reconstruction results using the structural similarity index (SSIM) and peak signal to noise ratio (PSNR). The effectiveness of the DSCS and SSCR-SU algorithms in the endmember and abundance estimations, and the accuracy of reconstructed 4D-STEM datasets, are demonstrated using strain mapping as an application test.

2. DSCS and reconstruction

A critical assumption for compressive sensing is that the data is sparse. For a 4D-STEM experiment, sparsity means that each diffraction pattern or virtual images can be separated into several “endmembers”, each endmember has the same dimension as the original diffraction patterns or virtual images [30]. Importantly, the number of endmembers is much smaller than the number of pixels in a diffraction pattern. That is, diffraction patterns can be described with only a handful of parameters and a set of endmembers. Also, we assume that diffraction patterns are spatially correlated so that these endmembers are shared between diffraction patterns within the field of view. Another assumption is that diffraction patterns can be represented as linear combinations of endmembers. The coefficients of linear combinations can be mapped in real-space and hereinafter referred as the abundance matrix S . The 4D-STEM dataset can be considered as the matrix multiplication of

$$X = SA \quad (1)$$

Where $X \in \mathbb{R}^{N \times L}$ is the full 4D-STEM dataset, $A \in \mathbb{R}^{P \times L}$ is the endmember matrix and $S \in \mathbb{R}^{N \times P}$ is the abundance matrix. $N = w_{spa} \cdot h_{spa}$ is the number of pixels in the field of view, $L = w_{spec} \cdot h_{spec}$ is the number of pixels per diffraction pattern and P is the number of endmembers.

We found the SSCR-SU algorithm developed by Wang et al. [33] can be adapted for 4D-STEM. SSCR-SU combines *spectral sampling* with

spatial sampling for compressive sensing. The concept of exploiting spatial or spectral correlation for compressive sensing is well-established in the remote sensing community [35,38,39]. The algorithm developed by Wang et al. [33] incorporates both spectral and spatial sampling. Our work demonstrates that this algorithm can be effectively extended to 4D-STEM with superior performances.

In our implementation of DSCS, spectral sampling is achieved by STEM imaging where diffraction intensity serves as the spectral signal using either STEM area detectors or randomly selected pixels in a 2D pixelated detector, which we call STEM spectral sampling and random spectral sampling respectively. The SSCR-SU reconstruction algorithm (Fig. 1) retrieves information from spatial sampling (real space) and spectral sampling (diffraction space). The rationale behind spectral sampling is to furnish the algorithm with supplementary information for efficiently estimating an initial reconstruction. Spectral sampling can also be leveraged as constraints, providing comprehensive real-space information. The reconstruction takes four inputs: 1) diffraction patterns of each sampled position ($Y_{spa} \in R^{n_{spa} \times L}$), 2) a stack of STEM images acquired with different STEM detector configurations and different camera lengths ($Y_{spec} \in R^{N \times n_{spec}}$), 3) the sampling positions ($\Phi_{spa} \in R^{n_{spa} \times N}$), and 4) the regions on the diffraction pattern which are spectrally sampled ($\Phi_{spec} \in R^{n_{spec} \times L}$). Here, n_{spa} and n_{spec} are the number of samples in real space and spectral space. The 3D data stacks are all vectorized into 2D data matrices as shown in Fig 1(b). The relationships between these inputs are

$$Y_{spa} = \Phi_{spa} X \quad (2)$$

and

$$Y_{spec} = X \Phi_{spec}^T \quad (3)$$

Where X is the full 4D-STEM dataset. Φ_{spa} is an identity matrix of size N for a raster scan when all pixels are probed. In DSCS, we randomly select n_{spa} sampling points with $n_{spa} \ll N$, through Bernoulli sampling, which is equivalent to randomly select n_{spa} rows from the identity matrix of size N to form Φ_{spa} . The random sampling in principle can be employed for spectral sampling as well, which works in the case of 4D-STEM data compression. The experimental implementation of random spectral sampling, however, requires the design of detectors capable of random pixel readout, which can be developed in future. DSCS can also be implemented using the available STEM detectors, for example, using a bright-field and 4 segmented annular dark-field STEM detectors (Fig. 2).

The reconstruction algorithm first estimates the number of endmembers ($\hat{p}$) in the stack of STEM images using the HySime algorithm [40], as shown in Fig. 1. HySime, which was developed by Bioucas-Dias et al. is an algorithm designed for determining the number of endmembers in hyperspectral images. This is achieved by estimating the signal and noise correlation matrices, which are used to define an error function, comprised of the power of the signal projection error and the power of the noise projected to the endmember subspace. The number of endmembers is determined by identifying the minimum of this error function, which balances the fidelity of subspace projection and the noise introduced in the projection. The stack of input STEM images contains information about the real-space distribution of the sampled spectral signals. It can be considered as a spectrally compressed 4D-STEM dataset, and thus it serves as a surrogate for the full 4D-STEM dataset. After obtaining $\hat{p}$, we perform vertex component analysis [41] to estimate the initial endmember matrix $\hat{A}^0$ from the sampled diffraction patterns. Next, a first estimation of the abundance matrix $\hat{S}^0$ is obtained by multiplying the stack of STEM images with the pseudo-inverse of the spectrally sampled endmember matrix.

$$\hat{S}^0 = Y_{spec} \left(\hat{A}^0 \Phi_{spec}^T \right)^T \left[\left(\hat{A}^0 \Phi_{spec}^T \right) \left(\hat{A}^0 \Phi_{spec}^T \right)^T \right]^{-1} \quad (4)$$

with the initial estimation of the full 4D-STEM dataset given by

$$\hat{X}^0 = \hat{S}^0 \hat{A}^0 \quad (5)$$

This estimation is further improved by updating S^k and A^k through the optimization of

$$A^k = \underset{A}{\operatorname{argmin}} \|\hat{X}^{k-1} - S^{k-1} A\|_2^2 + \lambda_1 \|Y_{spa} - \Phi_{spa} S^{k-1} A\|_2^2 \quad (6)$$

and

$$S^k = \underset{S}{\operatorname{argmin}} \|\hat{X}^{k-1} - S A^k\|_2^2 + \lambda_2 \|Y_{spec} - S A^k \Phi_{spec}^T\|_2^2 \quad (7)$$

where

$$\hat{X}^k = S^k A^k \quad (8)$$

and k denotes the number of iterations and $\|\cdot\|_2$ denotes the Frobenius norm of the matrix. S and A are optimized alternatively in an iterative loop. λ_1 and λ_2 are weighting parameters for the spatial and spectral constraints. The above step is called “joint optimization”. The optimization ends when a certain value or the relative change is smaller than a preset criterion. The relative change is defined as

$$\text{relative change} = \frac{\|\hat{X}^k - \hat{X}^{k-1}\|_2}{\|\hat{X}^k\|_2} \quad (9)$$

For the reconstruction performance evaluation, we use Peak signal to noise ratio [42] :

$$PSNR(I, \hat{I}) = -20 \log_{10} \left(\frac{\max(I)}{\frac{1}{N} \|I - \hat{I}\|_2} \right) \quad (10)$$

A higher PSNR value suggests the mean squared error is smaller, relative to the max intensity in the true image. Another performance index is the structural similarity index, which is defined by [43]

$$SSIM(I, \hat{I}) = \frac{(2\mu_I \mu_{\hat{I}} + C_1)(2\sigma_{I\hat{I}} + C_2)}{(\mu_I^2 + \mu_{\hat{I}}^2 + C_1)(\sigma_I^2 + \sigma_{\hat{I}}^2 + C_2)} \quad (11)$$

where μ_i is the average of i , $\sigma_{i\hat{i}}$ is the covariance of I and $\hat{I}$, $C_1 = (0.01L)^2$ and $C_2 = (0.03L)^2$, L is the dynamic range of the pixel value. The range of SSIM is -1 to 1, and higher SSIM means higher similarity between two images. Detail discussion of these metrics can be found in references [42, 43]

The DSCS algorithm described here is implemented in Python. All reconstructions and testing are done on a personal computer with an Intel i7-12700 CPU and 32 GB of RAM.

3. Results and discussions

3.1. Creation of sub-sampled 4D-STEM dataset

We use a 4D-STEM dataset collected using raster scan from a nanowire transistor to test our DSCS framework. The TEM sample was lift-out by focused-ion beam from a p-MOSFET device manufactured by Inter-university Microelectronics Centre (IMEC). The sample is oriented near the [110] zone axis with a thickness around 40–50 nm. The selected FOV contains a nanowire FET, including two nanowire channels, the epitaxial SiGe stressor and Si substrate body. The nanowire channel is covered with gate materials. (Fig. 2).

The 4D-STEM dataset tested here was acquired with EMPAD [3] on a Talos F200X S/TEM (Thermo-Fisher Scientific, USA, operated at 200 kV). We operated the microscope under the microprobe mode to form a probe of 0.6 mrad semi-convergence angle and diameter 2.2 nm in full-width half-maximum. The exposure time was 1 ms for each

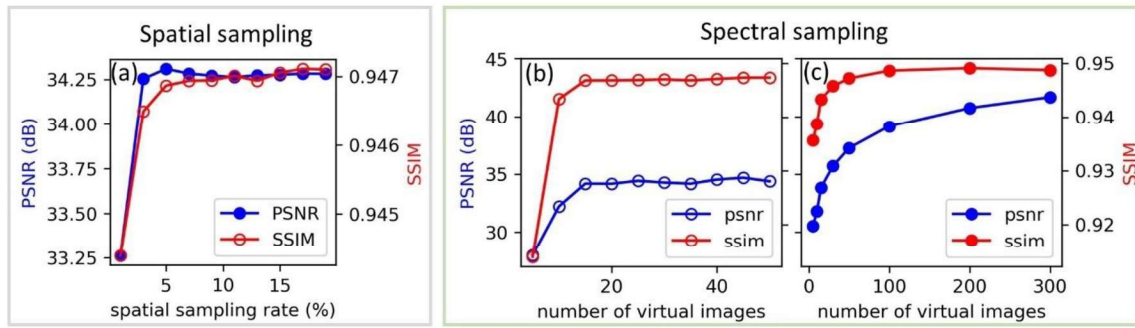


Fig. 3. Effects of spatial and spectral sampling on reconstruction performance. (a) influence of spatial sampling rate. Using (b) virtual STEM detector and (c) 10 % randomly selected pixels for sampling in diffraction space and the influence of number of virtual images.

diffraction pattern. The diffraction patterns were taken near the $[110]$ zone axis, and the camera length was adjusted so only $\{002\}$ and $\{1\bar{1}1\}$ diffraction disks were included in the recorded diffraction patterns.

We synthesized a subset of the full 4D-STEM dataset (X_{syn}) to test the DSCS reconstruction algorithm. The full 4D-STEM dataset has 200 by 100 pixels in the FOV, and each diffraction pattern contains 124 by 124 pixels. The X_{syn} was sampled with a spatial sample rate of 5 %. Virtual bright-field and dark-field images of the dataset were created by mimicking the standard STEM detectors with one for bright-field and four segmented annular dark-field detectors. Each virtual detector is scaled to three different sizes, corresponding to the camera length of 2850, 3400 and 4300 mm, respectively. With these detector and camera settings, different parts of the diffraction pattern are sampled, forming 15 virtual STEM images. Examples of virtual STEM images formed with different detector configurations are shown in Fig. 2(a)–(e). The diffraction contrast carried by virtual STEM images is encoded into the reconstruction dataset.

3.2. Reconstruction of the sub-sampled 4D-STEM dataset

The SSCR-RU reconstruction was performed using the synthesized dataset X_{syn} , producing a reconstructed 4D-STEM dataset $\hat{X}_{syn}$. Joint optimization was stopped when the relative change of Eq. 9 reached $1e-7$ with λ_1 and λ_2 set to 1 throughout this paper. The virtual STEM images formed using the reconstructed dataset as shown in Fig. 2(f)–(j) are visually identical to the original virtual STEM images [Fig. 2(a)–(e)]. The SSIMs of the two sets of images are all close to one. The lower PSNR and SSIM values for the pairs (b)(g) and (e)(j) result from the fact that the virtual detectors for these pairs didn't cover any Bragg peaks, leading to a significantly lower signal-to-noise ratio. Virtual STEM images dominated by more noise would naturally exhibit lower PSNR values since the reconstruction cannot accurately reproduce the noise. Fig. 2(k) and (l) map out the SSIM and PSNR value for each diffraction pattern in $\hat{X}_{syn}$, when compared with their counterpart in X_{syn} . The quality of the reconstructed diffraction pattern is demonstrated using two pairs of diffraction patterns from $\hat{X}_{syn}$ and X_{syn} , respectively [Fig. 2 (m), (n)]. The diffraction patterns taken away from the interface show very high similarity [Fig. 2(m)], with the intensity variations among the diffraction disks reproduced in high fidelity. The reconstructed diffraction patterns also appear smoother because of the denoising effect during the process of endmember extraction. In comparison, the reconstructed diffraction patterns closer to the interface of Si and SiO₂ [marked with a yellow dot in Fig. 2(k) and (l)], have obvious artifacts. We also noticed dark contrast around the interfaces between crystalline Si and amorphous SiO₂ and the black circular contrast in the crystalline Si region. Those contrasts are almost invisible in the virtual STEM images [(Fig. 2(a)–(e)), but very evident in the metrics maps. These circular regions are where electron beam was temporarily parked and were slightly contaminated. This introduces diffuse scattering in diffraction

patterns, which aren't completely captured by the reconstruction algorithm.

We note that stopping criterion for joint optimization can influence the performance of the reconstruction algorithm. This dependence is measured using the relative change (Eq. 9) average SSIM and average PSNR for each iteration [Fig. 2(o)]. The reconstruction converges monotonically. The average SSIM and PSNR also increase, and they converge at ~ 80 iterations. The relative change negatively correlated with average SSIM and PSNR, indicating that relative change is a reasonable index for monitoring the optimization process. The criterion for stopping can be established as the point of convergence of relative change. Since the process of joint optimization is deterministic, the point of convergence can be determined by examining the relative change curve of a longer run. Here, the stopping criterion was set at relative change is equal to $1e-7$, where both SSIM and PSNR are converged.

3.3. More than 1000-fold reduction in data size

The abundance matrix and endmember matrix are saved after reconstruction. For the 4D-STEM data described in Section 3.1 with $w_{spa} = 200$ and $h_{spa} = 100$ in real space and diffraction patterns of 124×124 pixels ($w_{spec} = h_{spec} = 124$), the dimension of abundance matrix and endmember matrix are 20,000 by 6 and 6 by 15,376 ($N = 20,000$, $p = 6$ and $L = 15,376$). Comparing with the full dataset, which is a matrix of 20,000 by 15,376, there is a 1400-fold reduction in the volume of data. This data reduction is very attractive as an efficient means for storing and transferring 4D-STEM dataset.

3.4. Influence of the spatial sampling rate

To examine the sampling rate effect, we created additional 10 sub-sampled datasets with increasing sampling rates. Fig. 3 plots the averaged SSIM and PSNR between the reconstructed and original DPs in the 4D-STEM dataset. There is a large improvement in reconstruction at the 5 percent sampling rate. This minimum sampling rate is likely to be dependent on the complexity of the sample. For the nanowire device sample tested in this paper, the diffraction pattern stack is mainly composed of Si (110) zone axis pattern, so they are relatively uniform despite the presence of strain. For samples with multiple grains or phases, higher spatial sampling rate may be required.

3.5. Influence of spectral sampling

A benchmark for spectral sampling configurations is shown in Fig. 3 (b). The virtual images, up to 50 of them, were created using different STEM detector configurations and the camera lengths range from 2850 to 4300 nm. We evenly subdivided this range by the number of virtual images. At each camera length, we create 5 different virtual images. The improvements in the average SSIM and PSNR slow down after more than 15 virtual images.

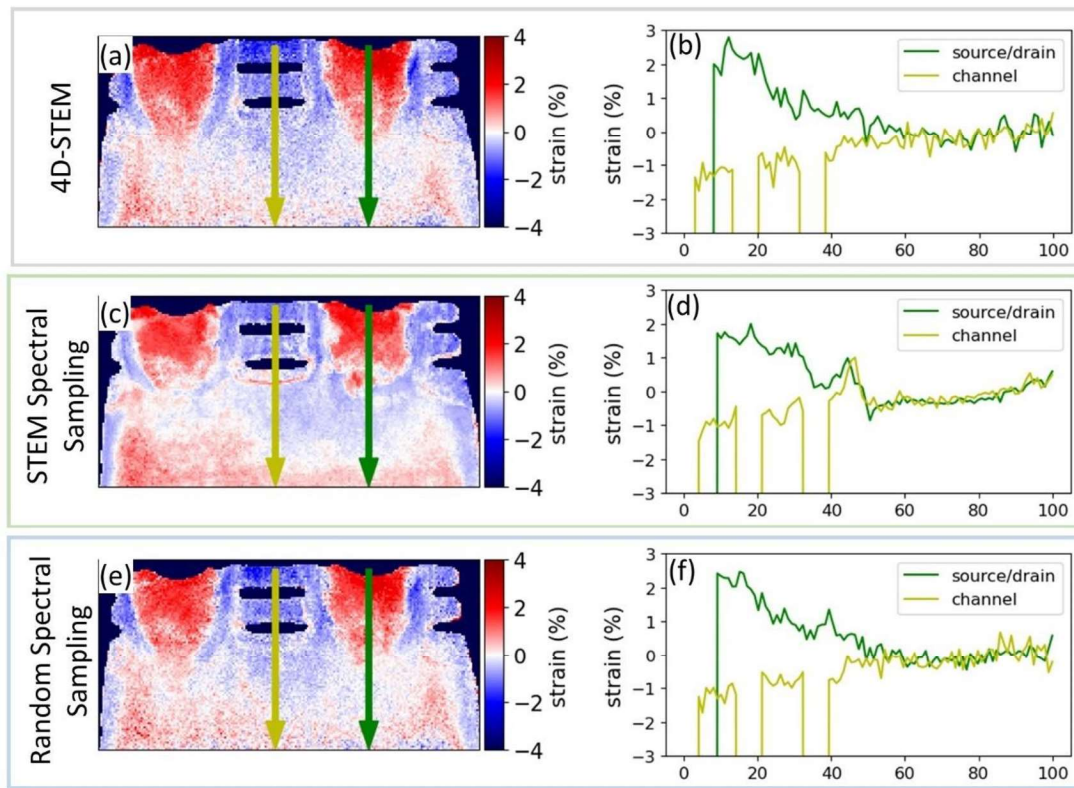


Fig. 4. strain analysis of datasets reconstructed with different spectral sampling method. (a) original dataset (c) STEM spectral sampling with 300 virtual images (e) random spectral sampling with 300 virtual images. (b), (d), and (f) are the line-profiles indicated in (a), (c) and (e). Unit for the x-axis in the line-profiles is in nanometer.

Alternatively, a very large number of virtual images can be created if random detector pixels are selected for spectral sampling. Each detector pixel creates a unique virtual image. We have also tested this spectral sampling scheme using our 4D-STEM test dataset. The virtual detectors are made of 10 percent of the total pixels in the diffraction pattern, which are randomly selected. To create multiple virtual images, pixels in diffraction pattern are randomly selected from the diffraction pattern multiple times. Fig. 3(c) shows that after having 30 or more detector configurations, random spectral sampling outperforms STEM spectral sampling.

To further explore the effect of spectral sampling, we mapped the strain along the $(2\bar{2}0)$ direction from the original and the reconstructed 4D-STEM datasets and compared the two spectral sampling methods with 300 virtual images in both cases. Strain analysis with 4D-STEM involves measuring the positions of Bragg diffraction disks from a local crystal volume. By fitting the disk positions, we can determine the reciprocal lattice vectors. Subsequently, strain can be calculated from these reciprocal lattice vectors. We adopt the approach outlined by Yuan et al. for performing the strain analysis [9]. Fig. 4 shows the results. Since strain analysis using 4D-STEM relies on accurately measuring the distances between diffraction disks, it provides a sensitive test of the diffraction pattern quality.

Fig. 4 shows while the strain maps obtained from the reconstructions using two different spectral sampling methods are similar and comparable to the strain map obtained from original 4D-STEM dataset, however, it is evident that random spectral sampling gives better results. By comparing the line profile of the strain map shown in Fig. 4(b), (d), and (f), we reached the same conclusion that random spectral sampling can reproduce the original result minus the difference that is comparable to noise fluctuation in original result.

The outperformance with random sampling suggests that the sensing basis created by random spectral sampling is more effective in covering

different spectral features in a 4D-STEM dataset. Additionally, with its pixel-wise spectral resolution, compared to the use of STEM area detectors with fixed shapes and reduced spectral resolution. The major benefit of using STEM detectors is that it can be implemented within the scheme of current hardware setups and detector designs.

Random spectral sampling can also be used for the compression of 4D-STEM datasets. For the example using 300 virtual images sampling formed by randomly selected detector pixels, the reconstructed 4D-STEM dataset was decomposed into two matrices $S \in R^{N \times p}$ and $A \in R^{p \times L}$ where $N = 20,000$, $p = 83$ and $L = 15,376$, giving rise to a 100-fold smaller data volume. To implement such scheme, new detector technology with fast random readout is required to provide the capability of randomly sampling in diffraction space with scanning speed close to conventional STEM imaging.

The above performances were demonstrated for nanodiffraction with a small beam convergence angle. For a highly converged beam, as used in ptychography, the diffraction patterns contain complex features from the interference of transmitted and diffracted beams that vary with probe position. Nonetheless, common features and periodic repeats are expected and thus the assumption of sparsity still applies. It is conceivable that the DSCS algorithm would still be effective for ptychography. A potential benefit from using the proposed algorithm is to reduce the noise using the spectrally-unmixed endmembers, which can be integrated into ptychography reconstruction. The periodicity of the sample may influence the performance of the proposed method. For instance, samples with a large spatial period, such as twisted 2D materials, might not perform as well as single crystal samples.

3.6. DSCS implementation in a STEM

To explore experimental feasibility of DSCS, we implemented it on a Themis Z S/TEM (Thermo Fisher Scientific, USA, probe corrected and

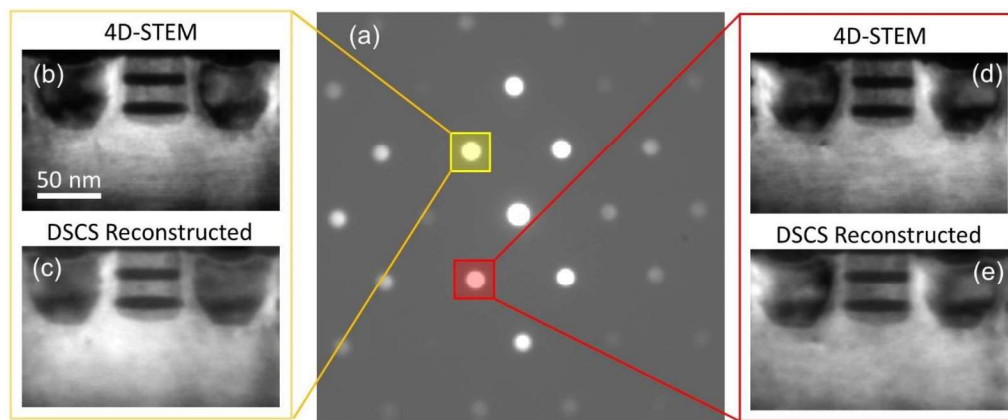


Fig. 5. Dual space compressive sensing applied for data acquisition in A STEM and comparison with the full 4D-STEM data acquired using EMPAD. The DSCS dataset was acquired using random scan and STEM images formed using segmented detectors. (a) shows an example diffraction pattern in the reconstructed 4D-STEM dataset using DSCS. Virtual dark-field images created with the yellow (top) and red (bottom) virtual detectors from (b,d) full 4D-STEM dataset and (c,e) DSCS reconstructed dataset.

operated at 300 kV) equipped with a Gatan Image Filter and an EMPAD detector at University of Illinois. The microscope was operated in STEM mode and the diffraction patterns were collected with the CCD mounted behind the Gatan Image Filter. First, a region of interest is selected for imaging, which is then scanned using the Bernoulli sampling discussed in Section 2 by modifying the scanning electron nanodiffraction setup described by Kim et al. [44]. A matrix, sized to match the region of interest, is created before scanning, and it is randomly initialized with values between 0 and 1. If a value is smaller than the sampling rate, the corresponding pixel will be sampled. At each sampling point, the diffraction pattern and the sampling coordinates are recorded and saved. The control of beam movement and the synchronization of camera and scanning was achieved by DigitalMicrograph scripting. A dataset with 5 % random spatial sampling was acquired with the custom script. Diffraction space sampling was done by taking STEM images using segmented detectors with camera lengths of 910 mm and 1150 mm to sample different parts of the diffraction pattern. The sampled diffraction patterns and STEM images were used to obtain a reconstructed 4D-STEM dataset. At the same region of interest, we also acquired a full 4D-STEM dataset using EMPAD for comparison.

Fig. 5 shows a set of virtual dark-field images generated from the reconstructed dataset and the full 4D-STEM dataset. Promisingly, the diffraction contrast is mostly reproduced. The small discrepancies in the images likely due to the experimental factors, such as the difference in the cameras' readout speeds, and the sample drift during diffraction pattern collection, which makes the spatial registration difficult. This further influence the registration between diffraction patterns and STEM images in DSCS. Additionally, the STEM detectors and CCD cameras have different response to electrons, correlating intensities from these two sources may also introduce noise and error. Lastly, when acquiring STEM images, the position of the diffraction pattern on the plane of STEM detectors was manually aligned so the spectral measurement matrix is estimated rather than precisely determined as in the synthetic compressive sensing dataset. Consequently, we are limited to providing qualitative results at this stage. These difficulties can be overcome by using the same detector for spatial and spectral sampling.

4. Conclusions

We have introduced an approach for 4D-STEM compressive sensing and data compression based on sub-sampling in both real and diffraction spaces [45]. The dual space compressive sensing scheme exploits the sparsity of diffraction data. We show that a full 4D-STEM dataset can be recreated from partially sampled 4D-STEM data. Using testing data created from 4D-STEM dataset We show that a factor of 100 data

reduction can be achieved in 4D-STEM data compression with good quality. The results also show that high-fidelity reconstruction is achieved with only 5 % of spatial sampling based on the measurement of the structural similarity index and peak signal to noise ratio between the original and reconstructed diffraction patterns. High fidelity of reconstructed diffraction patterns is demonstrated with strain analysis, which are sensitive small changes in diffracted beam positions. An initial attempt to implement compressive sensing 4D-STEM experimentally shows promising results. With developments in new detector and scan technologies, the dual space compressive sensing scheme can be fully implemented. The adaptation of compressive sensing in 4D-STEM will reduce the number of scanning points significantly so that the electron dose is greatly reduced. Applications previously hindered by the high dose and long acquisition time, such as the analysis of beam sensitive materials will become possible with the presented method.

CRedit authorship contribution statement

Hsu-Chih Ni: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. **Renliang Yuan:** Investigation, Methodology, Software. **Jiong Zhang:** Funding acquisition, Project administration, Resources, Supervision. **Jian-Min Zuo:** Conceptualization, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Hsu-Chih Ni reports financial support was provided by Intel Corp.

Data availability

Example code and data can be found in Ref. 45.

Acknowledgments

This work is supported by a grant from Intel Corporation (HCN) and by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Award Number DE-SC0024064 (JMZ). The authors acknowledge the use of facilities and instrumentation supported by NSF through the University of Illinois Materials Research Science and Engineering Center DMR-1720633.

References

- [1] N. Shibata, S.D. Findlay, Y. Kohno, H. Sawada, Y. Kondo, Y. Ikuhara, Differential phase-contrast microscopy at atomic resolution, *Nat. Phys.* 8 (2012) 611–615, <https://doi.org/10.1038/nphys2337>, 8 (2012).
- [2] N. Shibata, Y. Kohno, S.D. Findlay, H. Sawada, Y. Kondo, Y. Ikuhara, New area detector for atomic-resolution scanning transmission electron microscopy, *J. Electron. Microsc.* 59 (2010) 473–479, <https://doi.org/10.1093/JMICRO/DFQ014> (Tokyo).
- [3] M.W. Tate, P. Purohit, D. Chamberlain, K.X. Nguyen, R. Hovden, C.S. Chang, P. Deb, E. Turgut, J.T. Heron, D.G. Schlom, D.C. Ralph, G.D. Fuchs, K.S. Shanks, H. T. Philipp, D.A. Muller, S.M. Gruner, High dynamic range pixel array detector for scanning transmission electron microscopy, *Microsc. Microanal.* 22 (2016) 237–249, <https://doi.org/10.1017/S1431927615015664>.
- [4] M. Nord, R.W.H. Webster, K.A. Paton, S. McVitie, D. McGrouther, I. Maclaren, G. W. Paterson, Fast pixelated detectors in scanning transmission electron microscopy. Part I: data acquisition, live processing, and storage, *Microsc. Microanal.* 26 (2020) 653–666, <https://doi.org/10.1017/S1431927620001713>.
- [5] J. Ciston, I.J. Johnson, B.R. Draney, P. Ercius, E. Fong, A. Goldschmidt, J. M. Joseph, J.R. Lee, A. Mueller, C. Ophus, A. Selvarajan, D.E. Skinner, T. Stezelberger, C.S. Tindall, A.M. Minor, P. Denes, The 4D camera: very high speed electron counting for 4D-STEM, *Microsc. Microanal.* 25 (2019) 1930–1931, <https://doi.org/10.1017/S1431927619010389>.
- [6] H. Ryll, M. Simson, R. Hartmann, P. Holl, M. Huth, S. Ihle, Y. Kondo, P. Kotula, A. Liebel, K. Müller-Caspary, A. Rosenauer, R. Sagawa, J. Schmidt, H. Soltan, L. Strüder, A pnCCD-based, fast direct single electron imaging camera for TEM and STEM You may also like A pnCCD-based, fast direct single electron imaging camera for TEM and STEM, 2023. [10.1088/1748-0221/11/04/P04006](https://doi.org/10.1088/1748-0221/11/04/P04006).
- [7] B.H. Savitzky, S.E. Zeltmann, L.A. Hughes, H.G. Brown, S. Zhao, P.M. Pelz, T. C. Pekin, E.S. Barnard, J. Donohue, L. Rangel Dacosta, E. Kennedy, Y. Xie, M. T. Janish, M.M. Schneider, P. Herring, C. Gopal, A. Anapolsky, R. Dhali, K. C. Bustillo, P. Ercius, M.C. Scott, J. Ciston, A.M. Minor, C. Ophus, py4DSTEM: a software package for four-dimensional scanning transmission electron microscopy data analysis, *Microsc. Microanal.* 27 (2021) 712–743, <https://doi.org/10.1017/S1431927621000477>.
- [8] A. Béch , J.L. Rouvi re, L. Cl ment, J.M. Hartmann, Improved precision in strain measurement using nanobeam electron diffraction, *Appl. Phys. Lett.* 95 (2009) 123114, <https://doi.org/10.1063/1.3224886>.
- [9] R. Yuan, J. Zhang, J.M. Zuo, Lattice strain mapping using circular Hough transform for electron diffraction disk detection, *Ultramicroscopy* 207 (2019) 112837, <https://doi.org/10.1016/j.ultramic.2019.112837>.
- [10] R. Yuan, J. Zhang, L. He, J.M. Zuo, Training artificial neural networks for precision orientation and strain mapping using 4D electron diffraction datasets, *Ultramicroscopy* 231 (2021) 113256, <https://doi.org/10.1016/j.ultramic.2021.113256>.
- [11] E.F. Rauch, J. Portillo, S. Nicolopoulos, D. Bultreys, S. Rouvimov, P. Moeck, Automated nanocrystal orientation and phase mapping in the transmission electron microscope on the basis of precession electron diffraction, *Z. Kristallogr.* 225 (2010) 103–109, <https://doi.org/10.1524/ZKRI.2010.1205/MACHINEREADABLECITATION/RIS>.
- [12] M. Watanabe, D.B. Williams, Development of diffraction imaging for orientation analysis of grains in scanning transmission electron microscopy, *Microsc. Microanal.* 13 (2007) 962–963, <https://doi.org/10.1017/S1431927607075204>.
- [13] Y. Jiang, Z. Chen, Y. Han, P. Deb, H. Gao, S. Xie, P. Purohit, M.W. Tate, J. Park, S. M. Gruner, V. Elser, D.A. Muller, Electron ptychography of 2D materials to deep sub- ngstr m resolution, *Nature* 559 (7714 559) (2018) 343–349, <https://doi.org/10.1038/s41586-018-0298-5> (2018).
- [14] T.J. Pennycook, A.R. Lupini, H. Yang, M.F. Murfitt, L. Jones, P.D. Nellist, Efficient phase contrast imaging in STEM using a pixelated detector. Part 1: experimental demonstration at atomic resolution, *Ultramicroscopy* 151 (2015) 160–167, <https://doi.org/10.1016/J.ULTRAMIC.2014.09.013>.
- [15] A. Stevens, H. Yang, W. Hao, L. Jones, C. Ophus, P.D. Nellist, N.D. Browning, Subsampled STEM-ptychography, *Appl. Phys. Lett.* 113 (2018) 033104, <https://doi.org/10.1063/1.5040496>.
- [16] A. Stevens, L. Kovarik, H. Yang, Y. Pu, L. Carin, N.D. Browning, Compressive STEM-EELS, *Microsc. Microanal.* 22 (2016) 560–561, <https://doi.org/10.1017/S1431927616003652>.
- [17] D. Nicholls, J. Wells, A. Stevens, Y. Zheng, J. Castagna, N.D. Browning, Sub-Sampled Imaging for STEM: maximising Image Speed, Resolution and Precision Through Reconstruction Parameter Refinement, *Ultramicroscopy* 233 (2022) 113451, <https://doi.org/10.1016/j.ultramic.2021.113451>.
- [18] A. Stevens, H. Yang, L. Carin, I. Arslan, N.D. Browning, The potential for Bayesian compressive sensing to significantly reduce electron dose in high-resolution STEM images, *Microscopy* 63 (2014) 41–51, <https://doi.org/10.1093/jmicro/df042>.
- [19] E. Monier, T. Oberlin, N. Brun, X. Li, M. Tenc , N. Dobigeon, Fast reconstruction of atomic-scale STEM-EELS images from sparse sampling, *Ultramicroscopy* 215 (2020) 112993, <https://doi.org/10.1016/j.ultramic.2020.112993>.
- [20] A. Stevens, L. Luzzi, H. Yang, L. Kovarik, B.L. Mehdi, A. Liyu, M.E. Gehm, N. D. Browning, A sub-sampled approach to extremely low-dose STEM, *Appl. Phys. Lett.* 112 (2018) 43104, <https://doi.org/10.1063/1.5016192>.
- [21] J.M. Ede, Adaptive partial scanning transmission electron microscopy with reinforcement learning, *Mach. Learn. Sci. Technol.* 2 (2021) 45011, <https://doi.org/10.1088/2632-2153/abf5b6>.
- [22] T. Vogt, W. Dahmen, P. Binev, Modeling nanoscale imaging in electron microscopy, *Modeling Nanoscale Imaging in Electron Microscopy* (2012).
- [23] R. Leary, Z. Saghi, P.A. Midgley, D.J. Holland, Compressed sensing electron tomography, *Ultramicroscopy* 131 (2013) 70–91, <https://doi.org/10.1016/j.ultramic.2013.03.019>.
- [24] X. Li, O. Dyck, S. v. Kalinin, S. Jesse, Compressed sensing of scanning transmission electron microscopy (STEM) with nonrectangular scans, *Microsc. Microanal.* 24 (2018) 623–633, <https://doi.org/10.1017/S143192761801543x>.
- [25] A. B ch , B. Goris, B. Freitag, J. Verbeeck, Development of a fast electromagnetic beam blaster for compressed sensing in scanning transmission electron microscopy, *Appl. Phys. Lett.* 108 (2016) 93103, <https://doi.org/10.1063/1.4943086>.
- [26] Z. Saghi, M. Benning, R. Leary, M. Macias-Montero, A. Borras, P.A. Midgley, Reduced-dose and high-speed acquisition strategies for multi-dimensional electron microscopy, *Adv. Struct. Chem. Imaging* 1 (2015), <https://doi.org/10.1186/s40679-015-0007-5>.
- [27] E. Monier, T. Oberlin, N. Brun, M. Tence, M. de Frutos, N. Dobigeon, Reconstruction of partially sampled multiband images—application to STEM-EELS imaging, *IEEE Trans. Comput. Imaging* 4 (2018) 585–598, <https://doi.org/10.1109/TCI.2018.2866961>.
- [28] K.A. Hujsak, E.W. Roth, W. Kellogg, Y. Li, V.P. Dravid, High speed/low dose analytical electron microscopy with dynamic sampling, *Micron* 108 (2018) 31–40, <https://doi.org/10.1016/j.micron.2018.03.001>.
- [29] R.G. Baraniuk, Compressive sensing, *IEEE Signal Process. Mag.* 24 (2007), <https://doi.org/10.1109/MSP.2007.4286571>.
- [30] A.W. Robinson, A. Moshtaghpour, J. Wells, D. Nicholls, M. Chi, I. Maclaren, A.I. Kirkland, N.D. Browning, Simultaneous high-speed and low-dose 4-D STEM using compressive sensing techniques, (2023), <https://arxiv.org/abs/2309.14055v3> (accessed January 8, 2024).
- [31] A. Moshtaghpour, A. Velasco-Torrejon, A.W. Robinson, A.I. Kirkland, N. D. Browning, Exploring low-dose and fast electron ptychography using 10 regularisation of extended ptychographical iterative engine, *Microsc. Microanal.* 29 (2023) 344–345, <https://doi.org/10.1093/MICMIC/OZAD067.160>.
- [32] A. Datta, K.F. Ng, D. Balakrishnan, M. Ding, S.W. Chee, Y. Ban, J. Shi, N.D. Loh, A data reduction and compression description for high throughput time-resolved electron microscopy, *Nat. Commun.* 12 (1 12) (2021) 1–15, <https://doi.org/10.1038/s41467-020-20694-z> (2021).
- [33] L. Wang, Y. Feng, Y. Gao, Z. Wang, M. He, Compressed sensing reconstruction of hyperspectral images based on spectral unmixing, *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 11 (2018) 1266–1284, <https://doi.org/10.1109/JSTARS.2017.2787483>.
- [34] J. Bacca, C.v. Correa, H. Arguello, Noniterative hyperspectral image reconstruction from compressive fused measurements, *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 12 (2019) 1231–1239, <https://doi.org/10.1109/JSTARS.2019.2902332>.
- [35] G. Martin, J.M. Bioucas-Dias, Hyperspectral blind reconstruction from random spectral projections, *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 9 (2016) 2390–2399, <https://doi.org/10.1109/JSTARS.2016.2541541>.
- [36] C.H. Lin, J.M.B. Dias, T.H. Lin, Y.C. Lin, C.H. Kao, A new hyperspectral compressed sensing method for efficient satellite communications; a new hyperspectral compressed sensing method for efficient satellite communications, 2020.
- [37] C.C. Hsu, C.H. Lin, C.H. Kao, Y.C. Lin, DCSN: deep compressed sensing network for efficient hyperspectral data transmission of miniaturized satellite, *IEEE Trans. Geosci. Remote Sens.* 59 (2020) 7773–7789, <https://doi.org/10.1109/TGRS.2020.3034414>.
- [38] R. Arablouei, F. De Hoog, Hyperspectral image recovery via hybrid regularization, *IEEE Trans. Image Process.* 25 (2016) 5649–5663, <https://doi.org/10.1109/TIP.2016.2614131>.
- [39] Y. Jia, Y. Feng, Z. Wang, Reconstructing hyperspectral images from compressive sensors via exploiting multiple priors, *Spectrosc. Lett.* 48 (2015) 22–26, <https://doi.org/10.1080/00387010.2013.850727>.
- [40] J.M. Bioucas-Dias, J.M.P. Nascimento, Hyperspectral Subspace Identification, *IEEE Trans. Geosci. Remote Sens.* 46 (2008) 2435–2445, <https://doi.org/10.1109/TGRS.2008.918089>.
- [41] J.M.P. Nascimento, J.M.B. Dias, Vertex component analysis: a fast algorithm to unmix hyperspectral data, *IEEE Trans. Geosci. Remote Sens.* 43 (2005) 898–910, <https://doi.org/10.1109/TGRS.2005.844293>.
- [42] W. Zhou, A.C. Bovik, Mean squared error: love it or leave it? a new look at signal fidelity measures, *IEEE Signal Process. Mag.* 26 (2009) 98–117, <https://doi.org/10.1109/msp.2008.930649>.
- [43] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image quality assessment: from error visibility to structural similarity, *IEEE Trans. Image Process.* 13 (2004) 600–612, <https://doi.org/10.1109/TIP.2003.819861>.
- [44] K.H. Kim, H. Xing, J.M. Zuo, P. Zhang, H. Wang, TEM based high resolution and low-dose scanning electron nanodiffraction technique for nanostructure imaging and analysis, *Micron* 71 (2015) 39–45, <https://doi.org/10.1016/J.MICRON.2015.01.002>.
- [45] The python implementation of the compression algorithm, including test examples, can be found at https://github.com/swampni/Dual_Space_Compressive_Sensing.