Microscopy AND
Microanalysis

Original Article

Fast Grain Mapping with Sub-Nanometer Resolution Using 4D-STEM with Grain Classification by Principal Component Analysis and Non-Negative Matrix Factorization

Frances I. Allen[1,2]* 🆔, Thomas C. Pekin[1,2,†], Arun Persaud[3], Steven J. Rozeveld[4], Gregory F. Meyers[4], Jim Ciston[2], Colin Ophus[2] and Andrew M. Minor[1,2]

[1]Department of Materials Science and Engineering, UC Berkeley, Berkeley, CA 94720, USA; [2]National Center for Electron Microscopy, Molecular Foundry, LBNL, Berkeley, CA 94720, USA; [3]Accelerator Technology and Applied Physics Division, LBNL, Berkeley, CA 94720, USA and [4]Core R&D – Analytical Sciences, The Dow Chemical Company, Midland, MI 48674, USA

Abstract

High-throughput grain mapping with sub-nanometer spatial resolution is demonstrated using scanning nanobeam electron diffraction (also known as 4D scanning transmission electron microscopy, or 4D-STEM) combined with high-speed direct-electron detection. An electron probe size down to 0.5 nm in diameter is used and the sample investigated is a gold–palladium nanoparticle catalyst. Computational analysis of the 4D-STEM data sets is performed using a disk registration algorithm to identify the diffraction peaks followed by feature learning to map the individual grains. Two unsupervised feature learning techniques are compared: principal component analysis (PCA) and non-negative matrix factorization (NNMF). The characteristics of the PCA versus NNMF output are compared and the potential of the 4D-STEM approach for statistical analysis of grain orientations at high spatial resolution is discussed.

Key words: 4D-STEM, grain orientation mapping, NNMF, PCA, scanning nanobeam electron diffraction

(Received 10 March 2021; revised 24 April 2021; accepted 26 May 2021)

Introduction

The last decade has borne witness to a surge in scanning transmission electron microscopy (STEM) experiments implementing scanning nanobeam electron diffraction (NBED), fueled by the development of high-speed, high-efficiency, direct-electron detectors, as well as advanced computational methods. In scanning NBED, the electron probe is rastered over 2D spatial co-ordinates (x, y) and an NBED pattern with 2D reciprocal space co-ordinates (k_x, k_y) is acquired at every dwell point. The result is a data set with co-ordinates in four dimensions, hence the alternate name for the technique, "4D-STEM." 4D-STEM data sets contain a wealth of information that can be extracted by various means for a range of analytical purposes (Ophus, 2019). For example, virtual dark-field images can be reconstructed for any desired combination of diffraction peaks, enabling a multitude of virtual imaging experiments to be performed on the same data set post-acquisition (Gammer et al., 2015). Virtual imaging by 4D-STEM with atomic resolution has also been demonstrated (Kimoto &

Ishizuka, 2011). By analyzing the spacing between diffraction peaks, measurements of crystal lattice strain can be made (Béché et al., 2009), and by acquiring 4D-STEM data during *in situ* deformation experiments, strain evolution can be probed locally in time-discrete steps (Pekin et al., 2018). Converging the electron beam such that the Bragg diffraction disks significantly overlap enables electron ptychography experiments in which phase information is retrieved from the beam interference for resolution-enhanced imaging (Jiang et al., 2018). Electrostatic field mapping around individual atoms in 2D monolayers has also been demonstrated, via analysis of the center of mass of the diffraction pattern intensity distributions (Fang et al., 2019). In the work presented here, 4D-STEM is implemented in combination with fast electron detection to explore new possibilities for high-throughput grain mapping at high spatial resolution.

Existing electron-beam-based grain mapping techniques are primarily electron backscatter diffraction (EBSD) in the scanning electron microscope (SEM) and scanning precession electron diffraction (SPED) in STEM. EBSD is the method of choice for grains in the micron size range and above, and by using a thin sample and adopting a transmission geometry to instead collect forward-scattered electrons (which emanate from a much smaller interaction volume), mapping of grains down to 5–10 nm in size can be achieved (Keller & Geiss, 2012; Sneddon et al., 2016). This technique is also known as transmission Kikuchi diffraction (TKD). A key benefit of EBSD in the SEM is the ability to rapidly

*Author for correspondence: Frances I. Allen, E-mail: francesallen@berkeley.edu
†Current address: Institut für Physik & IRIS Adlershof, Humboldt Universität zu Berlin, 12489 Berlin, Germany.

CrossMark

scan and index over fields of view up to thousands of square microns, and the utility of TKD for high-throughput grain orientation mapping of nanoparticles and nanoprecipitates has been demonstrated (Bhattacharya et al., 2019; Mariano et al., 2020). However, the spatial resolution of SEM-based techniques is ultimately limited by the electron probe size. Grain mapping at high spatial resolution is, therefore, well-suited to STEM, where a sub-nanometer electron probe can be employed. In the precession-based SPED method (also a form of 4D-STEM), the STEM probe is rotated about a pivot point on the sample through a predefined tilt angle for each (x, y) co-ordinate in the scan. Hence, SPED gives diffraction patterns averaged over a range of orientations enabling robust indexing, albeit requiring long scan times (Rouviere et al., 2013; Midgley & Eggeman, 2015). In contrast, in 4D-STEM without precession, a single diffraction pattern is acquired per dwell point, reducing the total scan time (and thus total electron dose) considerably, in addition to greatly simplifying the experimental setup. While 4D-STEM without precession cannot offer the same level of indexing precision as SPED (in terms of the enhanced interpretability of the intensities in the indexed spots due to the quasi-kinematical nature of diffraction patterns), it has, for example, been shown to be well-suited for grain orientation mapping of beam-sensitive materials (Panova et al., 2016). Moreover, as we show here, 4D-STEM without precessing the beam presents a powerful technique for grain orientation mapping when both high throughput and high spatial resolution are required. 4D-STEM using an aperture to form multiple beams is now also poised to enhance the technique even further (Hong et al., 2020).

In the following, we apply 4D-STEM to an industrial catalyst comprising gold–palladium nanoparticles embedded in a silica support, using an aberration-corrected STEM equipped with a high-speed direct-electron detector to map sample sets with sub-nanometer resolution. 4D-STEM data sets are inherently large, especially when surveying wide fields of view with a sub-nanometer scan spacing. Hence, efficient methods for data processing with high levels of automation are essential. In our work, the diffraction peaks in the 4D-STEM data sets are first identified with sub-pixel precision using a disk registration algorithm developed previously (Pekin et al., 2017). Then, grain classification according to orientation is achieved *a priori* by unsupervised feature learning, using principal component analysis (PCA) and non-negative matrix factorization (NNMF), separately. The application of PCA to multi-dimensional (S)TEM spectroscopic data sets for phase analysis (electron energy loss and X-ray emission) has been demonstrated widely (Bosman et al., 2006; Yaguchi et al., 2008; Parish & Brewer, 2010; Allen et al., 2011). More recently, PCA of 4D-STEM data sets for strain mapping has also been demonstrated (Han et al., 2018). PCA is deterministic and computationally fast. However, a limitation of PCA is that the orthogonal matrices comprising the output contain both positive and negative values, which can make a physical interpretation of the results challenging. In contrast, NNMF is non-deterministic requiring many iterations to converge to a solution. This results in significantly longer computation times, yet a key benefit of NNMF is that the algorithm prohibits negative values and thus yields directly interpretable results (Lee & Seung, 1999). To date, NNMF has been employed most widely by the astrophysics community but is gaining momentum in the (S)TEM field, having been demonstrated for electron energy-loss spectrum-imaging (Nicoletti et al., 2013; Ringe et al., 2015; Shiga et al., 2016), SPED (Eggeman et al., 2015; Sunde et al., 2018; Martineau et al., 2019), and most recently also for 4D-STEM

(Savitzky et al., 2021; Uesugi et al., 2021). In this work, we compare PCA and NNMF in the context of grain mapping by 4D-STEM and discuss the characteristics and relative benefits of each classification approach.

Materials and Methods

4D-STEM Experiments

Samples of a fresh silica-based catalyst containing gold–palladium nanoparticles were prepared by embedding a small amount of catalyst powder with LR White resin (medium) and curing overnight at 60°C. Samples were then sectioned at room temperature using a diamond knife with a Reichart Ultracut S ultramicrotome to a thickness of 60 nm. The sections were collected on a standard copper TEM grid (300 mesh) with a lacey–carbon support (Ted Pella 01883).

The 4D-STEM experiments were performed using a double aberration-corrected modified FEI Titan 80-300 microscope (TEAM I instrument at the Molecular Foundry, LBNL), equipped with a Gatan K2 IS direct-electron detector operating at 400 frames per second. The electron beam energy was 300 keV. For grain mapping at the highest spatial resolution, an electron probe with full width at half maximum (FWHM) of 0.46 nm was selected (convergence angle of 3 mrad, camera length of 230 mm) and a scan step size of 0.25 nm was used. Larger fields of view were mapped using a probe with a FWHM of 1.05 nm (convergence angle of 1.5 mrad, camera length of 285 mm) implementing a scan step size of 0.5 nm. Probe currents were <200 pA. Beam-induced damage of the nanoparticles was not observed. The largest data set comprised 32,000 diffraction patterns and surveyed an area of 80×100 nm^2 (acquisition time 80 s).

For reference, high-resolution high-angle annular dark-field (HAADF) STEM images of a subset of nanoparticles were also acquired using the TEAM I instrument at 300 keV. In addition, sample surveys using STEM-based X-ray energy-dispersive spectroscopy (XEDS) were performed at Dow Chemical using a Thermo-Fisher (FEI) Titan Themis operated at 200 keV using Bruker AXS XFlash XEDS detectors with an energy resolution of 137 eV/channel.

4D-STEM Data Analysis

Preprocessing

Analysis of the 4D-STEM data sets was performed in MATLAB (version R2019a). This involved a number of data preprocessing steps using custom scripts as outlined below, followed by grain classification by PCA and NNMF as described in the subsequent section. The MATLAB scripts are available from the authors upon request.

The data preprocessing steps are summarized in Figure 1. First, the original data acquired in proprietary .dm4 (or .dm3) Gatan Digital Micrograph format was read into MATLAB giving a 4D data stack with two dimensions defining the scanned areas and two dimensions defining the diffraction patterns. The diffraction patterns were binned by a factor of two in order to reduce file size and thus speed up the subsequent computations (note that this does not affect the spatial resolution of the grain maps, since only the NBED patterns are binned). Next, the 4D stacks were converted into 3D stacks by merging the scan co-ordinates to give a number sequence of diffraction patterns.

In order to quickly visualize the nanoparticle distribution in any given surveyed area, a virtual dark-field image was reconstructed by
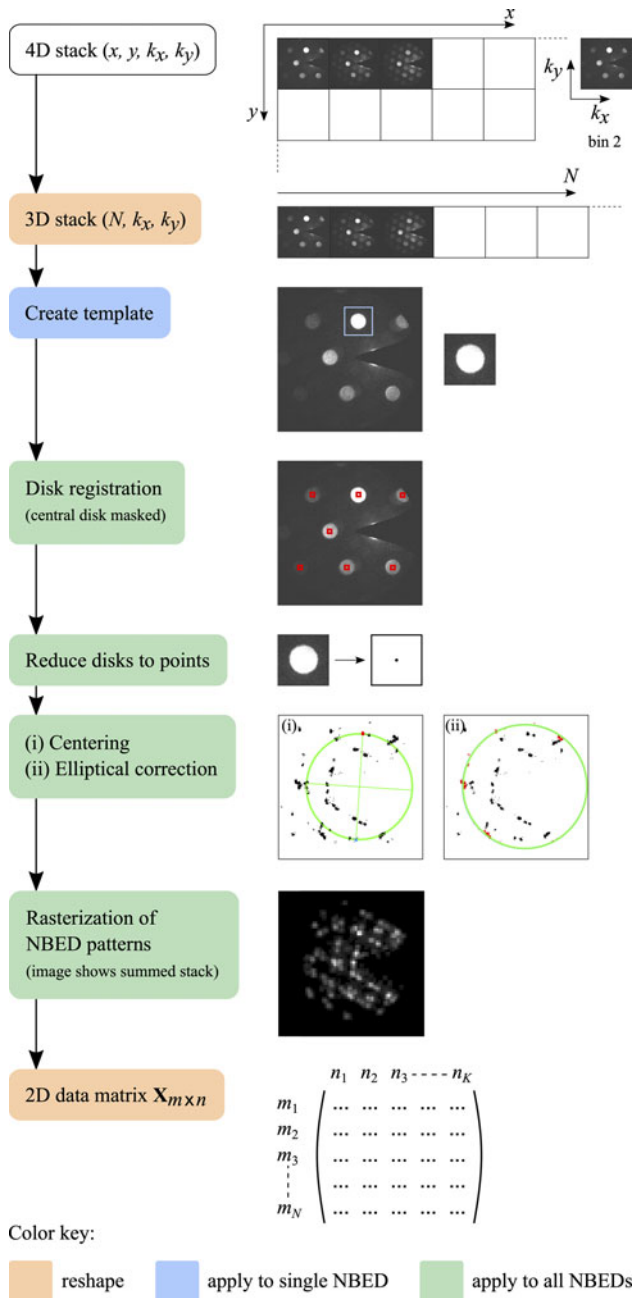
Fig. 1. Data preprocessing, starting from the 4D data stack acquired in the experiment and ending with a 2D data matrix, with m rows corresponding to the number sequence of NBED patterns and n columns corresponding to the number sequence of rasterized NBED mesh points.

summing the intensity values in each NBED pattern and plotting the result for each corresponding dwell point in the scan. Similarly, a quick overview of the diffraction data can be obtained by creating a single NBED pattern comprising the maximum intensity measured at each pixel (or alternatively, the mean).

The next step is disk registration, to automatically locate all the Bragg disks in the data set. For this, an intense Bragg disk was first cropped from the NBED pattern, corresponding to the brightest pixel in the virtual dark-field image, and used to create a template. Next, we implemented a disk registration algorithm developed previously, using a hybridized standard cross-correlation and phase correlation approach with a predefined cross-correlation

threshold, followed by Fourier upsampling (Pekin et al., 2017), to determine the positions of all Bragg disks in the stack with sub-pixel precision. For data sets acquired using a beam stop to block the intense nondiffracted central disk, a mask was used to crop out bright spots emerging from the sides of the beam stop before passing the data through the disk registration algorithm so as to prevent the registration of erroneous peaks. Each Bragg disk detected was then reduced to a point, thus significantly reducing the effects of variations in disk shape and structure due to dynamical effects from the subsequent analysis. Plotting all resulting Bragg peaks in a single image, the center of the diffraction pattern was determined by computing the center of mass of opposite peak clusters to enable re-centering of the NBED stack to $(k_x, k_y) = 0$. Any elliptical distortion of the diffraction pattern arising from the lens system of the microscope was then measured and corrected.

In the final preprocessing step, the NBED stack was rasterized in diffraction space to generate an output array for the subsequent PCA/NNMF routines. For the rasterization, a square mesh grid was defined of appropriate mesh size (typically one quarter of the original Bragg disk diameter) and the intensity from each Bragg peak was distributed to the nearest four corners of its corresponding mesh square, weighted according to the distance. A 2D data matrix $\mathbf{X}$ was then generated, with m rows corresponding to the number sequence of diffraction patterns and n columns corresponding to the number sequence of rasterized NBED mesh points (essentially binned locations in diffraction space).

Grain Classification by PCA and NNMF

For grain classification by PCA and NNMF, MATLAB in-built functions were used. Various custom scripts were used for plotting and analyzing the output.

The basic principle of PCA is to reduce the dimensionality of a data set by finding a linear combination of uncorrelated variables which successively maximize statistical variance (Jolliffe & Cadima, 2016). Typically, the input data matrix (in our case $\mathbf{X}_{m \times n}$) is first centered at the mean, allowing PCA to be performed by singular value decomposition; these two steps are the default computations performed by the MATLAB *pca* function. The result of the matrix factorization can be written as

$$\mathbf{X}_{m \times n} = \mathbf{S}_{m \times n} \mathbf{V}_{n \times n}^{T},$$
$$\text{where} \quad \mathbf{S}_{m \times n} = \mathbf{U}_{m \times m} \mathbf{\Sigma}_{m \times n}. \tag{1}$$

$\mathbf{U}$ and $\mathbf{V}$ are square matrices with columns and rows composed of orthogonal unit vectors, T denotes a matrix transpose, and $\mathbf{\Sigma}$ is a rectangular diagonal matrix, effectively a scaling matrix. The principal components (basis vectors) are the columns of $\mathbf{S}$ listed in descending order of variance. The elements of $\mathbf{S}$ are known as the scores and the elements of $\mathbf{V}$ are known as the loadings (or coefficients). Due to the orthogonality constraint, the scores and loadings can have both positive and negative values—this can make interpretation of the individual components challenging, since negative values are nonphysical in diffraction patterns.

Once the matrix factorization has been performed, a truncated number of components, k, can then be selected and used to compute a reconstructed data matrix to approximate the original, that is,

$$\mathbf{X}_{m \times n} \approx \mathbf{S}_{m \times k} \mathbf{V}_{k \times n}^{T}, \quad \text{with } k < n. \tag{2}$$

A common method used to guide the choice of the reduced number of components k is to plot the variance (given by the squares

of the eigenvalues of $\mathbf{\Sigma}$, or equivalently the eigenvalues of the covariance matrix of $\mathbf{X}$) versus the component index. This is also known as a scree plot. Generally, there is first a sharp downwards trend and then the plot levels out, with the transition between these two regions representing the transition between statistically significant components and higher-order components corresponding to random noise. The value of k is often selected to include a subset of the first "noise" components, so as not to lose statistically significant information in the reconstruction. However, there is no general consensus on how exactly this selection should be performed, that is, how far beyond the elbow the series of components should be truncated. An alternative approach is to compute the percentage of the total variance accounted for by each component. This is also called the "percentage variance explained." The set of components for the reconstruction can then be selected based on a predefined amount of the cumulative variance explained.

The key distinction of dimensionality reduction by NNMF is that the initial data matrix $\mathbf{X}$, which must only contain positive values, is approximated by a matrix product whose elements are also constrained to be positive (Lee & Seung, 1999), that is,

$$\mathbf{X}_{m \times n} \approx \mathbf{W}_{m \times c}\mathbf{H}_{c \times n}, \quad \text{with } \mathbf{W}, \mathbf{H} \geq 0. \tag{3}$$

$\mathbf{W}$ and $\mathbf{H}$ are essentially equivalent to the $\mathbf{S}$ and $\mathbf{V}$ matrices of PCA, containing the basis vectors from the original data and their weights (also known as encoding coefficients), respectively, for each class identified. The rank c gives the total number of classes (corresponding to the components of PCA) and satisfies $0 < c < \min(m, n)$. The factorization proceeds iteratively, seeking to minimize the residual between $\mathbf{X}$ and $\mathbf{WH}$. In the MATLAB *nnmf* function, the residual is quantified by computing the Frobenius norm (square root of the sum of the squares of the matrix elements) of $\mathbf{X} - \mathbf{WH}$. Since the factorization may converge to a local minimum, a predefined number of repeat factorizations (known as replicates) are run with different initial values. The replicate that gives the smallest residual is selected for the final result. After experimenting with different numbers of replicates, 12 replicates were typically used for our analysis. An initial number of classes for the matrix factorization, c, must also be selected. We chose c to be two to five times greater than that expected for the data in hand, based on inspection of the scree plot obtained by PCA. A first pass of NNMF was then performed, using an alternating least-squares algorithm and random initial values for $\mathbf{W}$ and $\mathbf{H}$ (these are the default parameters in the MATLAB function).

The number of classes c was then reduced and subsequent passes of NNMF were performed as follows. First, the MATLAB *corr* function was used to compute a correlation matrix for $\mathbf{W}$ and $\mathbf{H}^T$, respectively, comprising the linear correlation coefficients (ranging from -1 to $+1$) for each pair of columns, that is, for each class. The correlation matrices were then combined and the classes merged for the cases where the correlation factors exceeded a predefined value (typically 0.4–0.5). Finally, the merged classes were used to compute new values for $\mathbf{W}$ and $\mathbf{H}$ and NNMF was rerun using these for the initial input values together with the new value for c. This sequence of merging classes and rerunning NNMF was repeated until the successive decrease in the number of classes leveled out to a plateau. For the data sets presented here, up to 18 passes of NNMF were performed.

The PCA and NNMF results were visualized by reshaping the product matrices from equations (2) and (3) to give image pairs for each method comprising a real-space image highlighting a particular region of the sample (i.e., a grain, calculated from $\mathbf{S}$ and $\mathbf{W}$, respectively) and a diffraction-space image of the corresponding pattern of Bragg peaks (calculated from $\mathbf{V}$ and $\mathbf{H}$, respectively). To aid interpretation of the peak patterns, a virtual central beam spot was added during the plotting sequence to compensate for the absence of a central disk due to the use of the beam stop during the data acquisition. The basis vectors from $\mathbf{S}$ and $\mathbf{W}$ were also used to compute a weighted average NBED pattern for each grain identified.
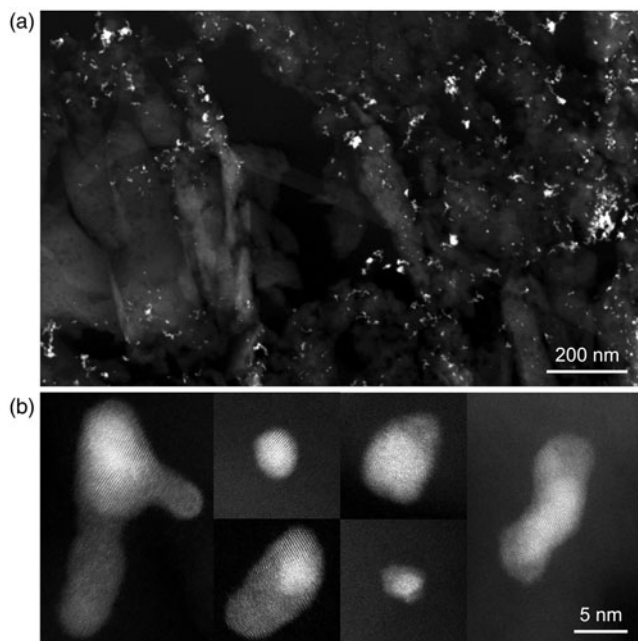
Since NNMF does not use eigen-decomposition, a PCA-like scree plot of component variance versus component index cannot be generated. Thus, in order to compare the ability of PCA and NNMF to capture the essence of the original data, we chose to plot the sum of squares of the residuals $\mathbf{X} - \mathbf{SV}^T$ and $\mathbf{X} - \mathbf{WH}$ versus the component/class index instead. In the case of PCA, the reconstructed matrix $\mathbf{SV}^T$ for increasing numbers of components can simply be computed by sequentially truncating the product matrices (which are already ordered in decreasing order of statistical importance). In contrast, in NNMF, the number of classes is one of the parameters of the factorization, thus the algorithm has to be rerun for each desired number of classes. One approach to obtain the reconstructed $\mathbf{WH}$ matrix for each NNMF class index is to run NNMF sequentially, increasing the value of c each time and using the product matrices from the previous run as the starting values for the next (Ren et al., 2018). However, in our analysis, multiple passes of NNMF were already performed as part of the class merging sequence described above, each NNMF pass being for a reduced number of classes as determined using the correlation function. Hence for the NNMF portion of our matrix residual comparison, the sparse data points from the merging sequence were plotted. In addition, single passes of NNMF spanning the range of classes up to $c = 100$ were also performed. As will be seen, these data points for the single NNMF passes follow the same trend as for the NNMF merging sequence.

Results and Discussion

Sample Overview

Figure 2a shows a low-magnification HAADF-STEM image of the gold–palladium nanoparticles (bright contrast) embedded in silica that were used in this study. A part of the lacey carbon TEM grid support is also visible. A set of high-resolution HAADF-STEM images of individual nanoparticles is shown in Figure 2b. In the high-resolution images, we see that certain regions (or grains) appear brighter than others, due to diffraction contrast dictated by the orientation of the crystal lattice planes with respect to the incident electron beam. However, the precise locations of grain boundaries within particles are challenging to discern. Elemental maps obtained by STEM-XEDS show gold- as well as palladium-rich nanoparticles, but with no clear distinction between individual grains under the electron dose and dwell times used for these experiments (see Supplementary Figure S1). With these observations in mind, and given that both gold and palladium have the same crystal structure (fcc) and very similar lattice constants, it becomes clear that an alternative method to efficiently map the grains in this sample is needed, that is, 4D-STEM.

A schematic summarizing the 4D-STEM data acquisition method used in this study is presented in Figure 3a, indicating the collection of an NBED pattern for every scan point on the
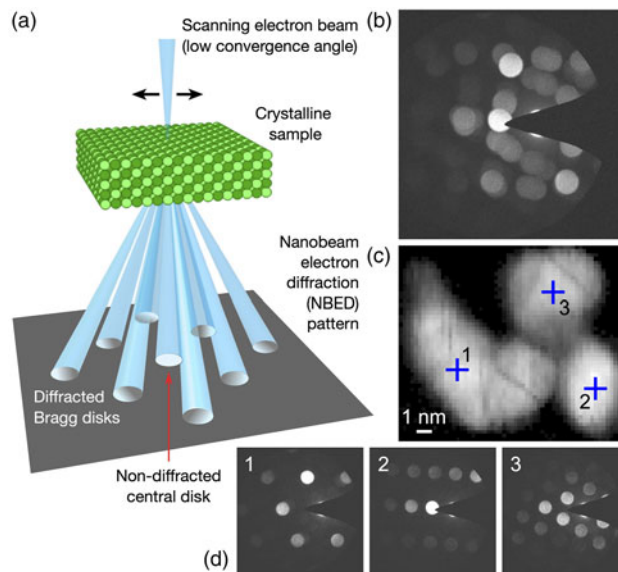
sample. Plotting the maximum intensity measured per diffraction pixel across the data set gives an NBED pattern representing the entire acquisition, as shown in Figure 3b. This data is for a single nanoparticle cluster scanned using a step size of 0.25 nm over a field of view of $15 \times 13 \text{ nm}^2$ (probe FWHM of 0.5 nm). The corresponding virtual dark-field image, computed by plotting the summed NBED intensity measured per scan point, is shown in Figure 3c. The three crosshairs in the virtual dark-field image mark individual scan points, with the respective NBED patterns recorded for these positions shown in Figure 3d. In the following, the grain mapping analysis for this small nanoparticle cluster is presented, followed by the results for a larger field of view capturing many clusters.

Grain Classification by PCA and NNMF

Single Nanoparticle Cluster

PCA and NNMF grain classification results for the single nanoparticle cluster are shown in Figures 4a and 4b, respectively, obtained following the data preprocessing and matrix factorization procedures described in the section "4D-STEM Data Analysis." The top row (i) of each subfigure shows the basis vectors (re-shaped) for the first eight classes identified, plotted in order of decreasing statistical significance. These images map the spatial regions of the nanoparticle corresponding to each class. The second row (ii) shows the respective coefficients (re-shaped) representing the diffraction "fingerprints" obtained for each class. The Bragg peaks appear as points rather than the original disks, since the disks were reduced to points prior to the disk registration in the data preprocessing step to enable grain mapping without influence from dynamical effects. To aid interpretation of the diffraction patterns, which were collected using a beam stop to mask the intense central beam, a virtual central beam spot has been added to these images. Finally, in the third row (iii), the reconstructed NBED pattern corresponding

to each component/class is plotted, obtained using the basis vectors to compute weighted sums from the NBED patterns in the original 4D-STEM data stack. We show the results for a total of eight components/classes, since this was the number of classes obtained after the NNMF class merging sequence (the initial number of classes had been set to 16). The scree plot analysis of the PCA results (Supplementary Figure S2) also indicates that eight components are a reasonable number.

The color bars inserted alongside the coefficients in Figure 4 show the range of intensity values that the coefficients possess. A comparison of these ranges for PCA (Fig. 4a) and NNMF (Fig. 4b) highlights one of the key distinguishing features of the two factorization techniques, that is, that the output from PCA can have both positive and negative values, whereas in the case of NNMF, only positive values are allowed. In this example, we see that the negative (blue) peaks in the PCA coefficients of Figure 4a row (ii) increasingly appear for the third component onwards. Negative output is also obtained for the PCA basis vectors, but for simplicity, only the positive values were used to create the PCA basis vector maps shown in the figure. The occurrence of both positive and negative values in PCA is a direct consequence of the orthogonal factorization constraint, which can yield non-physical results. That is, in diffraction data and in imaging in general, the physical measurement values are by nature all positive, hence interpretation of negative coefficients (in our case, negative Bragg peaks) is challenging. However, the weighted NBED patterns in Figure 4a row (iii) show diffraction patterns that are all "real," with positive peaks, since these were constructed by weighting the original NBED data.

A further direct consequence of the orthogonal factorization of PCA is the tendency to produce mixed phases in the output (generally for the higher-order components), since in reality phases do not necessarily have to be linearly uncorrelated. For example, the weighted NBED pattern for the fourth PCA component in Figure 4a row (iii) clearly comprises two patterns indexing to
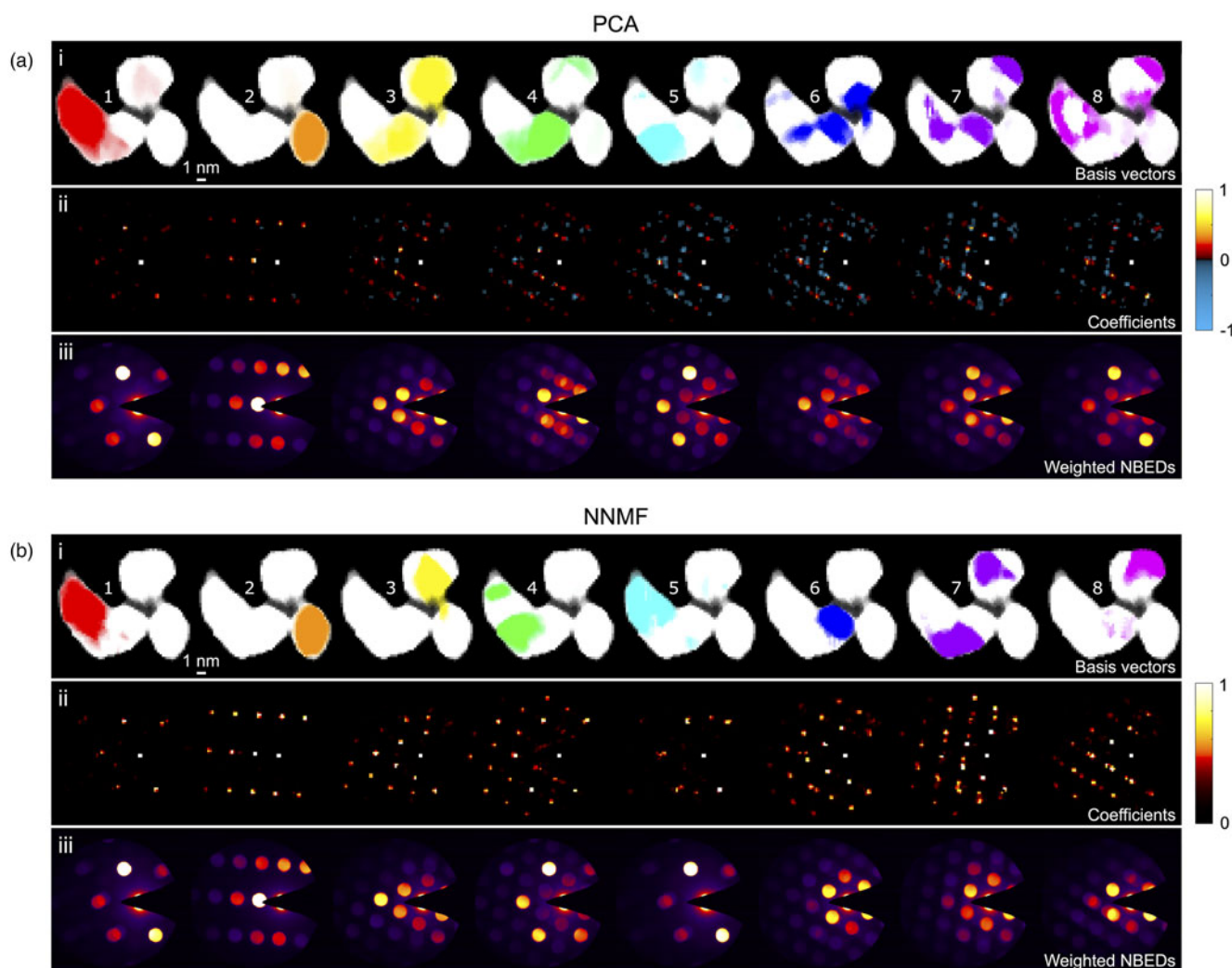
Fig. 4. Grain classification results for a single nanoparticle cluster using (**a**) PCA and (**b**) NNMF to analyze the 4D-STEM data. (i) Basis vectors (re-shaped) for the first eight components/classes. (ii) Respective coefficients (re-shaped) with color bars indicating the intensity values. (iii) Weighted NBED patterns reconstructed for each component/class.

the same lattice reflection with a slight rotational offset. It is worth noting that phase mixing in our results can also be real, resulting from overlapping grains in the beam direction. Indeed, both PCA and NNMF are sensitive to detecting mixed phases if present, since the reconstructed data are represented as a superposition of basis vectors, as for example shown in the analysis of various spectral data sets (Long et al., 2009; Kannan et al., 2018) and of SPED data sets (Sunde et al., 2018). However, in our example, the weighted NBED patterns from the NNMF analysis do not indicate overlapping phases to the same extent as is seen in the PCA results. Hence, the phase mixing observed in the PCA output would appear to largely be an artifact from the orthogonality-constrained PCA factorization. In the case of simple systems with no (or minimal) phase mixing, this artifact can be removed using methods like Varimax rotation after the PCA algorithm (Keenan, 2009), applying an orthogonal rotation in the spatial domain resulting in a relaxing of the orthogonal constraint in the diffraction domain. The Varimax rotation matrix is found by maximizing the contrast between the individual spatial features (Smentkowski et al., 2009), based on the assumption that each pixel in the scanned region corresponds to a single phase.

The NNMF results shown in Figure 4b give a similar classification of grains to the PCA results, but there are some important distinctions. First, the NNMF coefficients (and basis vectors) by definition contain positive values only, as reflected by the color bar in row (ii), that is, no negative peaks. Thus, each re-shaped coefficient matrix represents a real diffraction pattern. Second, upon inspection of the NNMF coefficients and weighted NBED patterns, we see that they are largely discrete for each class, with satellite peaks and overlapping disks less prevalent than in the PCA results. This highlights the ability of NNMF to identify discrete components with less tendency to capture mixed phases in a single class. We also note that the weighted NBED patterns for the first and fifth classes corresponding to the left-hand side of the cluster are very similar, suggesting that there may be two grains overlapping in depth with only a slight misorientation between them. Rather than classifying these as a mixed phase, NNMF discerned them as separate phases.

In terms of computation time (using a 2017 MacBook Pro Intel Core i5, 8 GB RAM), the PCA algorithm took 4 min to complete, whereas the NNMF algorithm, including the class merging sequence (8 merges) took 40 min. The size of the input 4D-STEM data set was 1 GB.
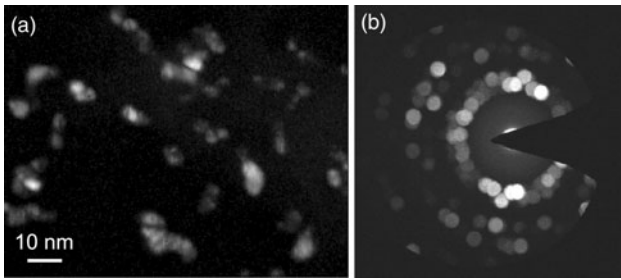
Fig. 5. (a) Virtual dark-field image for a distribution of nanoparticles surveyed in a 4D-STEM scan. (b) Summed NBED pattern plotting the maximum intensity recorded per pixel.

Finally, we found that the NNMF classification was much more sensitive to the mesh spacing that was chosen to rasterize the diffraction patterns in the data preprocessing step (described in the "Preprocessing" section). For example, mesh spacings of 5 or 6 pixels consistently yielded NNMF results in which the first (highest variance) component was the grain on the left-hand side of the cluster (i.e., the red grain in Figs. 4a and 4b), whereas using a mesh spacing of 7 or 8, the order in the NNMF output changed and the orange grain became the most statistically important. In contrast, in the PCA case, the order of components was not affected by these changes. This again points to the enhanced sensitivity of NNMF to subtle differences in the input

data set. For the data analysis presented here, a rasterization mesh size of 6 pixels was used in all cases.
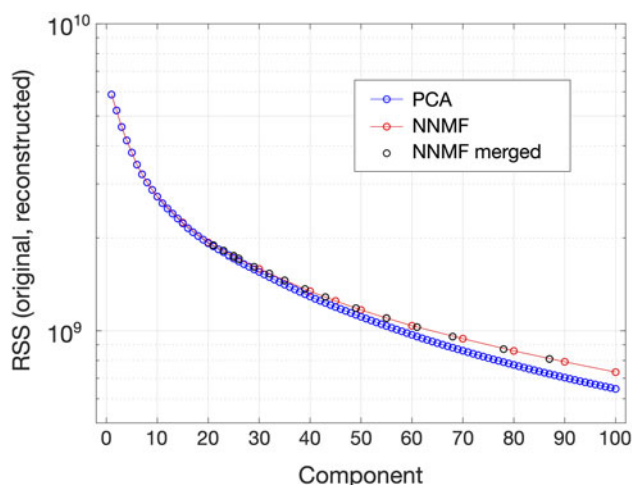
Distribution of Nanoparticles

The virtual dark-field image and the corresponding 4D-STEM NBED stack for the region of the sample surveyed over a larger field of view (80×100 nm^2) are shown in Figures 5a and 5b, respectively. For this data set, a scan step size of 0.5 nm and probe size of 1 nm (FWHM) were selected. The PCA and NNMF grain classification results are presented in Figure 6, showing composite maps of the (numbered) basis vectors in Figures 6a(i) and 6b (i), the individual coefficients for each component/class in Figures 6a(ii) and 6b(ii), and the weighted NBED patterns in Figures 6a(iii) and 6b(iii). The number of components/classes plotted is 21, since this was the number of classes obtained after the NNMF class merging sequence (the initial number of classes had been set to 100). The scree plot analysis of the PCA results (Supplementary Figure S3) also indicates that ~20 is a reasonable estimate for the number of distinct phases present in the surveyed area.

Comparing the PCA and NNMF results of Figure 6, it can be seen that essentially the grain maps in each case are very similar. There is some variation in the number order of the grains, and in the PCA case, there is a certain degree of phase mixing (most evident in the composite map, where the labels for the overlapping grains are comma-separated). These differences reflect those previously discussed in the PCA/NNMF analysis of the small nanoparticle cluster.
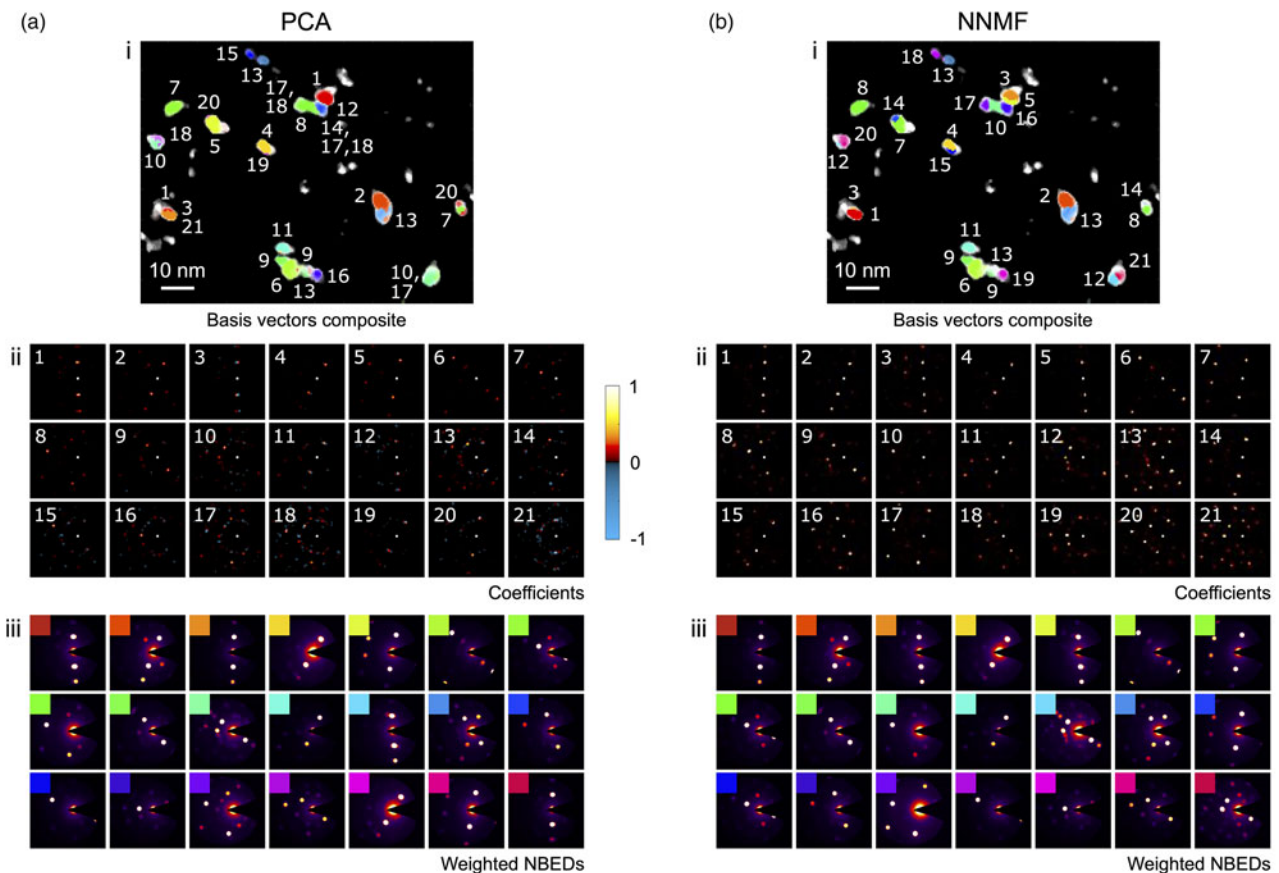
Fig. 6. Grain classification results for a distribution of nanoparticles using (a) PCA and (b) NNMF to analyze the 4D-STEM data. (i) Composite maps of the basis vectors obtained (numbered), (ii) coefficients for each component/class with color bars indicating the intensity values, and (iii) weighted NBED patterns reconstructed and color-coded for each component/class.

Fig. 7. Matrix residual-based comparison of the PCA and NNMF results for the nanoparticle distribution: Plot of the squared sums of the matrix elements of the residual matrices $X - SV^T$ (PCA) and $X - WH$ (NNMF) versus the component/class index. Data points for PCA in blue, for NNMF without merging in red, and for NNMF after merging (starting with $c = 100$) in black.

In order to further compare the PCA and NNMF results, plots of the sum of squares of the residual matrices, ((initial data matrix) − (reconstructed data matrix)) versus the component/class index have been generated, as shown in Figure 7 and described at the end of the section "Grain Classification by PCA and NNMF." The PCA data points (blue circles) have been obtained by sequentially truncating the reconstructed PCA matrix in increments of one component, whereas the data points for NNMF have been obtained from individual passes of NNMF using a different number of input classes *without* subsequent merging (red circles), and also using the $c = 100$ data set and plotting the data points corresponding to each class merging step (black circles). Similar analysis of the PCA and NNMF output for a 4D-STEM data set has been performed in recent work by Uesugi et al. (2021). Figure 7 shows that as the number of components increases, the matrix residuals rapidly decrease. This confirms that a larger number of components allows the PCA/NNMF reconstructions to capture the original data more effectively. Furthermore, as the NNMF classes are merged (moving from right to left in the plot), the residuals gradually increase. This is because each merging step introduces a small error (by definition the classes being merged were not perfect matches, but rather satisfied a predefined correlation threshold).

Up to $c \approx 20$, the PCA and NNMF traces in Figure 7 essentially overlap, indicating close similarity between the reconstructions obtained using each technique up to this point (as also indicated by the comparison of PCA and NNMF output shown in Fig. 6). Then, for $c > 20$, where components start to represent noise, the two traces start to deviate with the PCA trace progressively reaching lower residual values than the NNMF trace. This can be explained by the tendency of PCA to overfit the data compared with NNMF (Ren et al., 2018).

Due to the file size of the large field of view data set (67 GB), these computations were performed using a workstation with two Intel Xenon Silver 4114 CPUs with 512 GB RAM. For the PCA algorithm, the computation time was 55 s, whereas for the NNMF algorithm, including the class merging sequence (16 merges), it was close to 44 h. In order to speed up the NNMF computation, the size of the input data set could be reduced by using more binning and/or by using thresholding to create a 2D probability distribution of all grains and then clustering the detected Bragg disks, for example, using Voronoi cells (Savitzky et al., 2021).

Conclusions

We have shown that 4D-STEM using a sub-nanometer probe with a beam overlap of 50% combined with fast direct-electron detection enables efficient grain mapping at high spatial resolution. Since relatively large fields of view (by STEM standards) can be probed in seconds to minutes, high-resolution mapping while maintaining minimal sample drift can be achieved. Beam and sample alignments for the 4D-STEM method used here are also greatly simplified compared with precession-based methods.

Developments in direct-electron detector technology continue to transform the field, with data acquisition speeds greater than ten thousand frames per second now realized (Ercius et al., 2020; Nord et al., 2020), that is, two orders of magnitude faster than the direct-electron detector used in this work. With this significant advance, 4D-STEM is now poised to enable efficient grain mapping over fields of view approaching the sizes surveyed by EBSD in the SEM, but with the spatial resolution provided by STEM. Thus, fast high-resolution grain mapping over large fields of view for high-throughput sub-nanometer grain analysis becomes possible. For the analysis of nanoparticle catalysts such as those surveyed in this work, this paves the way toward statistical analysis of catalyst grain size distributions and orientations through the screening of thousands of particles before and after use, providing crucial insight into the factors affecting catalyst efficiency and lifetime. Future work can also make use of multibeam electron diffraction, using an aperture to form multiple probes that increase the angular range surveyed per scan point and thus enable robust indexing due to the richer diffraction patterns obtained (Hong et al., 2020), as well as automated indexing based on template matching using image libraries of simulated diffraction patterns for all orientations of each crystal structure (Rauch & Véron, 2014).

Ever increasing data acquisition speeds mean ever increasing file sizes. Consequently, high levels of automation for parameter optimization in the data analysis routines will be essential. In the present work, we have demonstrated several critical semi-automated data preprocessing routines and shown how parameter choice can affect the final classification results. In the future, various open-source software routines that are currently being developed will be well-positioned to increase automation for enhanced multiparameter optimization (Savitzky et al., 2019; Clausen et al., 2020; Johnstone et al., 2021). Further benefit can be drawn from the use of patterned probes generated using a binary mask in the probe-forming aperture in order to facilitate the subsequent disk registration (Zeltmann et al., 2020). Finally, we have shown that feature learning using PCA offers fast classification of phases and is thus well-suited for a first pass analysis. NNMF computation is much more time-intensive, but offers direct interpretation of the matrix output due to the non-negativity constraint, delivers results free from orthogonality-induced phase-mixing, and is also more sensitive to subtle variations in the input data.

References

Allen FI, Watanabe M, Lee Z, Balsara NP & Minor AM (2011). Chemical mapping of a block copolymer electrolyte by low-loss EFTEM spectrum-imaging and principal component analysis. *Ultramicroscopy* **111**, 239. doi:10.1016/j.ultramic.2010.11.035

Béché A, Rouvière JL, Clément L & Hartmann JM (2009). Improved precision in strain measurement using nanobeam electron diffraction. *Appl Phys Lett* **95**, 123114. doi:10.1063/1.3224886

Bhattacharya A, Parish CM, Henry J & Katoh Y (2019). High throughput crystal structure and composition mapping of crystalline nanoprecipitates in alloys by transmission Kikuchi diffraction and analytical electron microscopy. *Ultramicroscopy* **202**, 33–43. doi:10.1016/j.ultramic.2019.03.015

Bosman M, Watanabe M, Alexandar DTL & Keast VJ (2006). Mapping chemical and bonding information using multivariate analysis of electron energy-loss spectrum images. *Ultramicroscopy* **106**, 1024. doi:10.1016/j.ultramic.2006.04.016

Clausen A, Weber D, Ruzaeva K, Migunov V, Baburajan A, Bahuleyan A, Caron J, Chandra R, Halder S, Nord M, Müller-Caspary K & Dunin-Borkowski RE (2020). LiberTEM: Software platform for scalable multidimensional data processing in transmission electron microscopy. *J Open Source Softw* **5**, 2006. doi:10.21105/joss.02006

Eggeman AS, Krakow R & Midgley PA (2015). Scanning precession electron tomography for three-dimensional nanoscale orientation imaging and crystallographic analysis. *Nat Commun* **6**, 7267. doi:10.1038/ncomms8267

Ercius P, Johnson I, Brown H, Pelz P, Hsu SL, Draney B, Fong E, Goldschmidt A, Joseph J, Lee J, Ciston J, Ophus C, Scott M, Paul D, Hanwell M, Harris C, Avery P, Stezelberger T, Tindall C, Ramesh R, Minor A & Denes P (2020). The 4D camera – An 87 kHz frame-rate detector for counted 4D-STEM experiments. *Microsc Microanal* **26**, 1896–1897. doi:10.1017/S1431927620019753

Fang S, Wen Y, Allen CS, Ophus C, Han GGD, Kirkland AI, Kaxiras E & Warner JH (2019). Atomic electrostatic maps of 1D channels in 2D semiconductors using 4D scanning transmission electron microscopy. *Nat Commun* **10**, 1127. doi:10.1038/s41467-019-08904-9

Gammer C, Ozdol VB, Liebscher CH & Minor AM (2015). Diffraction contrast imaging using virtual apertures. *Ultramicroscopy* **155**, 1. doi:10.1016/j.ultramic.2015.03.015

Han Y, Nguyen K, Cao M, Cueva P, Xie S, Tate MW, Purohit P, Gruner SM, Park J & Muller DA (2018). Strain mapping of two-dimensional heterostructures with subpicometer precision. *Nano Lett* **18**, 3746. doi:10.1021/acs.nanolett.8b00952

Hong X, Zeltmann SE, Savitzky BH, Rangel DaCosta L, Müller A, Minor AM, Bustillo KC & Ophus C (2020). Multibeam electron diffraction. *Microsc Microanal*, 1–11. doi:10.1017/S1431927620024770

Jiang Y, Chen Z, Han Y, Deb P, Gao H, Xie S, Purohit P, Tate MW, Park J, Gruner SM, Elser V & Muller DA (2018). Electron ptychography of 2D materials to deep sub-Ångström resolution. *Nature* **559**, 343. doi:10.1038/s41586-018-0298-5

Johnstone DN, Crout P, Nord M, Laulainen J, Høgås S, Opheim E, Martineau B, Francis C, Bergh T, Prestat E, Smeets S, Ross A, Collins S, Hjorth I, Danaie M, Furnival T, Jannis D, Cautaerts N, Jacobsen E, Herzing A, Poon T, Ånes HW, Morzy J, Doherty T, Iqbal A, Ostasevicius T, mvonlany & Tovey R (2021). pyxem 0.13.0. doi:10.5281/zenodo.4436723

Jolliffe IT & Cadima J (2016). Principal component analysis: A review and recent developments. *Philos Trans R Soc A* **374**, 20150202. doi:10.1098/rsta.2015.0202

Kannan R, Ievlev AV, Laanait N, Ziatdinov MA, Vasudevan RK, Jesse S & Kalinin SV (2018). Deep data analysis via physically constrained linear unmixing: Universal framework, domain examples, and a community-wide platform. *Adv Struct Chem Imaging* **4**, 6. doi:10.1186/s40679-018-0055-8

Keenan MR (2009). Exploiting spatial-domain simplicity in spectral image analysis. *Surf Interface Anal* **41**, 79–87. doi:10.1002/sia.2949

Keller RR & Geiss RH (2012). Transmission EBSD from 10 nm domains in a scanning electron microscope. *J Microsc* **245**, 245. doi:10.1111/j.1365-2818.2011.03566.x

Kimoto K & Ishizuka K (2011). Spatially resolved diffractometry with atomic-column resolution. *Ultramicroscopy* **111**, 1111–1116. doi:10.1016/j.ultramic.2011.01.029

Lee DD & Seung HS (1999). Learning the parts of objects by non-negative matrix factorization. *Nature* **401**, 788. doi:10.1038/44565

Long CJ, Bunker D, Li X, Karen VL & Takeuchi I (2009). Rapid identification of structural phases in combinatorial thin-film libraries using X-ray diffraction and non-negative matrix factorization. *Rev Sci Instrum* **80**, 103902. doi:10.1063/1.3216809

Mariano RG, Yau A, McKeown JT, Kumar M & Kanan MW (2020). Comparing scanning electron microscope and transmission electron microscope grain mapping techniques applied to well-defined and highly irregular nanoparticles. *ACS Omega* **5**, 2791–2799. doi:10.1021/acsomega.9b03505

Martineau BH, Johnstone DN, van Helvoort ATJ, Midgley PA & Eggeman AS (2019). Unsupervised machine learning applied to scanning precession electron diffraction data. *Adv Struct Chem Imaging* **5**, 3. doi:10.1186/s40679-019-0063-3

Midgley PA & Eggeman AS (2015). Precession electron diffraction—A topical review. *IUCrJ* **2**, 126. doi:10.1107/S2052252514022283

Nicoletti O, de la Peña F, Leary RK, Holland DJ, Ducati C & Midgley PA (2013). Three-dimensional imaging of localized surface plasmon resonances of metal nanoparticles. *Nature* **80**, 502. doi:10.1038/nature12469

Nord M, Webster RWH, Paton KA, McVitie S, McGrouther D, MacLaren I & Paterson GW (2020). Fast pixelated detectors in scanning transmission electron microscopy: Part I: Data acquisition, live processing, and storage. *Microsc Microanal* **26**, 653–666. doi:10.1017/S1431927620001713

Ophus C (2019). Four-dimensional scanning transmission electron microscopy (4D-STEM): From scanning nanodiffraction to ptychography and beyond. *Microsc Microanal* **25**, 563. doi:10.1017/S1431927619000497

Panova O, Chen XC, Bustillo KC, Ophus C, Bhatt MP, Balsara N & Minor AM (2016). Orientation mapping of semicrystalline polymers using scanning nanobeam diffraction. *Micron* **88**, 30. doi:10.1016/j.micron.2016.05.008

Parish CM & Brewer LN (2010). Multivariate statistics applications in phase analysis of STEM-EDS spectrum images. *Ultramicroscopy* **110**, 134. doi:10.1016/j.ultramic.2009.10.011

Pekin TC, Gammer C, Ciston J, Minor AM & Ophus C (2017). Optimizing disk registration algorithms for nanobeam electron diffraction strain mapping. *Ultramicroscopy* **176**, 170. doi:10.1016/j.ultramic.2016.12.021

Pekin TC, Gammer C, Ciston J, Ophus C & Minor AM (2018). In situ nanobeam electron diffraction strain mapping of planar slip in stainless steel. *Scr Mater* **146**, 87. doi:10.1016/j.scriptamat.2017.11.005

Rauch EF & Véron M (2014). Automated crystal orientation and phase mapping in TEM. *Mater Charact* **98**, 1–9. doi:10.1016/j.matchar.2014.08.010

Ren B, Pueyo L, Zhu GB, Debes J & Dûchene G (2018). Non-negative matrix factorization: Robust extraction of extended structures. *Astrophys J* **852**, 104. doi:10.3847/1538-4357/aaa1f2

Ringe E, DeSantis CJ, Collins SM, Duchamp M, Dunin-Barowski RE, Skrabalak SE & Midgley P (2015). Resonances of nanoparticles with poor plasmonic metal tips. *Sci Rep* **5**, 17431. doi:10.1038/srep17431

Rouviere JL, Béché A, Martin Y, Denneulin T & Cooper D (2013). Improved strain precision with high spatial resolution using nanobeam precession electron diffraction. *Appl Phys Lett* **103**, 241913. doi:10.1063/1.4829154

Savitzky BH, Zeltmann SE, Barnard E, DaCosta LR, Brown HG, Henderson M, Ginsburg D & Ophus C (2019). py4dstem. doi:10.5281/zenodo.3333960

Savitzky BH, Zeltmann SE, Hughes LA, Brown HG, Zhao S, Pelz PM, Pekin TC, Barnard ES, Donohue J, DaCosta LR, Kennedy E, Xie Y, Janish MT, Schneider MM, Herring P, Gopal C, Anapolsky A, Dhall R, Bustillo KC, Ercius P, Scott MC, Ciston J, Minor AM & Ophus C (2021). py4DSTEM: a software package for four-dimensional scanning transmission electron microscopy data analysis *Microsc Microanal*. doi: 10.1017/S1431927621000477.

Shiga M, Tatsumi K, Muto S, Tsuda K, Yamamoto Y, Mori T & Tanji T (2016). Sparse modeling of EELS and EDX spectral imaging data by

nonnegative matrix factorization. *Ultramicroscopy* **170**, 43–59. doi:10.1016/j.ultramic.2016.08.006

Smentkowski VS, Ostrowski SG & Keenan MR (2009). A comparison of multivariate statistical analysis protocols for ToF-SIMS spectral images. *Surf Interface Anal* **41**, 88–96. doi:10.1002/sia.2973

Sneddon GC, Trimby PW & Cairney JM (2016). Transmission Kikuchi diffraction in a scanning electron microscope: A review. *Mater Sci Eng R Rep* **110**, 1–12. doi:10.1016/j.mser.2016.10.001

Sunde JK, Marioara CD, van Helvoort ATJ & Holmestad R (2018). The evolution of precipitate crystal structures in an Al-Mg-Si(-Cu) alloy studied by a combined HAADF-STEM and SPED approach. *Mater Charact* **142**, 458. doi:10.1016/j.matchar.2018.05.031

Uesugi F, Koshiya S, Kikkawa J, Nagai T, Mitsuishi K & Kimoto K (2021). Non-negative matrix factorization for mining big data obtained using four-dimensional scanning transmission electron microscopy. *Ultramicroscopy* **221**, 113168. doi:10.1016/j.ultramic.2020.113168

Yaguchi T, Konno M, Kamino T & Watanabe M (2008). Observation of three-dimensional elemental distributions of a Si device using a 360°-tilt FIB and the cold field-emission STEM system. *Ultramicroscopy* **108**, 1603. doi:10.1016/j.ultramic.2008.06.003

Zeltmann SE, Müller A, Bustillo KC, Savitzky B, Hughes L, Minor AM & Ophus C (2020). Patterned probes for high precision 4D-STEM Bragg measurements. *Ultramicroscopy* **209**, 112890. doi:10.1016/j.ultramic.2019.112890