doi:10.1017/S1431927620015639
Microsc. Microanal. 26 (Suppl 2), 2020

Towards Automated Classification of Complex 4D-STEM Datasets

Benjamin Savitzky[1], Steven Zeltmann[2], Lauren Hughes[1], Kyle Seelman[1], Matthew Janish[1], Matthew Schneider[3], Chirranjeevi Gopal[4], Patrick Herring[4], Andrew Minor[1,2] and Colin Ophus[5]

[1]National Center for Electron Microscopy (NCEM), Molecular Foundry, Lawrence Berkeley National Laboratory, Berkeley, California, United States, [2]University of California - Berkeley, Berkeley, California, United States, [3]Los Alamos National Laboratory, Los Alamos, New Mexico, United States, [4]Toyota Research Institute, Los Altos, California, United States, [5]Lawrence Berkeley National Lab, Berkeley, California, United States

In four-dimensional scanning transmission electron microscopy (4D-STEM), a diffraction pattern is collected at each position of the electron beam in a 2D raster over the sample surface. Samples with many regions of differing order, such as samples with mixed regions of crystalline material in different orientations and amorphous regions, generate complex datasets. This data requires classification, or identification of the distinct regions, so that each region can be appropriately analyzed. Ideally, each class corresponds to a type of diffraction pattern, or to structurally meaningful features or motifs, such that a scan position will be included in a given class if and only if its diffraction pattern contains these features. Recent progress towards effective, efficient, and open source 4D-STEM classification algorithms has been rapid [1-5]. As 4D-STEM datasets commonly reach 100s of GB in size, it is necessary to judiciously reduce the data to a compressed representation before performing classification. Here, we define a feature vector calculated from measured Bragg disk positions, demonstrate its effectiveness at classifying complex 4D-STEM dataset, and discuss its potential for fully automating 4D-STEM classification.

The feature vector is derived from the positions of the Bragg disks. If high precision disk position measurement is combined with good calibrations, it is found that in crystalline regions the disk positions approach delta-like distributions, as expected – see Fig. 1. Even in complex data containing many distinct crystalline grains and orientations, it is easy to identify and enumerate the consequent deltas. For a dataset containing N sharp peaks in its distribution of measured Bragg disk positions, we construct for each scan position a length N vector of either Boolean values, describing whether or not the scan position contains the associated Bragg disk, or of floating point values describing the disk intensities.

These vectors, which reduce the data size by several orders of magnitude, are found to be highly effective for classification of crystalline diffraction patterns. Using non-negative matrix factorization we successfully classify 4D-STEM datasets containing many crystallites, as well as continuous structural variations within single crystals – see Fig. 2. We additionally classify distinct amorphous regions by masking off Bragg disks, then classifying based on elliptically corrected radial profiles. Results from several datasets will be presented, and prospects for automation will be discussed.

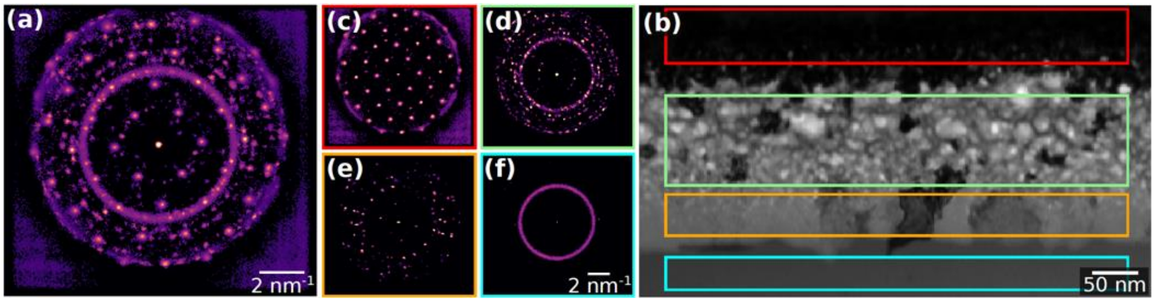

Figure 1. Measured Bragg disk positions, summed over (a) the entire image and (b-e) the subsets of scan positions within the colored boxes shown in (f) the bright field virtual image. We suggest classifying 4D-STEM data by identifying sharp peaks in the distribution of Bragg disk positions, labelling each diffraction pattern as containing or not containing each peak, then using these labels as classification feature vectors.

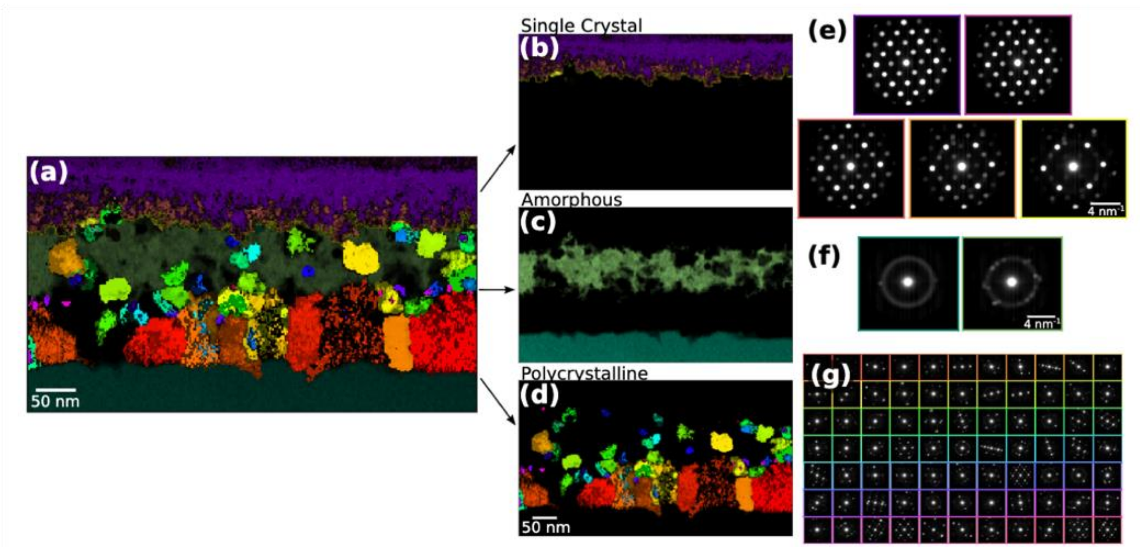

Figure 2. Classification of a model sample, a complex Gd2Ti2O7 dataset. (a) All of the structurally distinct phases identified in this system, using NNMF on our proposed feature vector. (b,e) The single crystal phases and their class diffraction patterns. (c,f) The amorphous phases and their class diffraction patterns. (d,g) The polycrystalline phases and their class diffraction patterns.

References
[1] https://github/LiberTEM/LiberTEM, DOI: 10.5281/zenodo.1478763
[2] https://github.com/pyxem/pyxem, DOI: 10.5281/zenodo.2649351
[3] https://hyperspy.org, DOI: 10.5281/zenodo.3396791
[4] https://pixstem.org/
[5] https://github.com/py4dstem/py4DSTEM, DOI: 10.5281/zenodo.3333960

https://doi.org/10.1017/S1431927620015639 Published online by Cambridge University Press