py4DSTEM: Open Source Software for 4D-STEM Data Analysis

Benjamin H. Savitzky[1*], Lauren Hughes[1], Karen C. Bustillo[1], Haitao D. Deng[2], Norman L. Jin[2], Eder G. Lomeli[2], William C. Chueh[2], Patrick Herring[3], Andrew Minor[1,4], and Colin Ophus[1]

[1.] National Center for Electron Microscopy, Molecular Foundry, Lawrence Berkeley National Laboratory, Berkeley, CA, USA
[2.] Department of Materials Science & Engineering, Stanford University, Stanford, CA, USA
[3.] Toyota Research Institute, Los Altos, CA, USA
[4.] Department of Materials Science & Engineering, University of California, Berkeley, CA, USA
* Corresponding author: ben.savitzky@gmail.com

py4DSTEM is a free, open source, python-based data analysis package for the emerging field of four-dimensional scanning transmission electron microscopy (4D-STEM). In 4D-STEM, the electron beam is rastered across a sample and a diffraction pattern is collected at each scan position, yielding highly information-rich datasets which can be used to virtually recreate images that map sample orientation, local lattice parameters, strain, and short/medium range ordering, or to reconstruct the electrostatic potential itself [1]. These datasets can easily reach tens or even hundreds of GB in size, and almost always require significant computational processing to extract useful results. The aim of py4DSTEM is to make 4D-STEM data analysis easy and accessible for everyone [2].

A sample workflow generating a strain map of an $FePO_4$ microplatelet is shown in Figs. 1-3. Figure 1a shows a virtual bright-field image generated from a 4D-STEM data hypercube. The three scan positions indicated by a blue, green, and red dot correspond to the three diffraction patterns shown in Fig. 1b. In the workflow demonstrated here, a template of the vacuum probe is generated (Fig. 1a, inset) and then cross correlated with the diffraction patterns to identify all the Bragg peaks in each diffraction pattern (Fig. 1c). The Bragg peaks are summed over all scan positions to generate the plots shown in Fig. 2a,b, showing the location of Bragg reflections in diffraction space; we refer to these plots as average deconvolutions. We find images of the real space microplatelet structure at each reciprocal lattice reflection (Fig. 2a) of the average deconvolution; these are caused by the translation of each diffraction pattern with the rastering of the beam. Fitting and correcting the diffraction shifts causes each Bragg reflection to collapse into a sharp peak (Fig. 2b). After removing the bright central beam (Fig. 2c) and taking a Radon transform, each reciprocal lattice peak becomes one of the eponymous sinusoidal curves of a sinogram (Fig. 2d). The angles where these curves converge immediately identifies the high symmetry zone axes (*red arrows and dashed lines*); a simple scoring function yields these lattice directions (Fig. 2e). The lattice vectors can then be extracted, and the Bragg peaks indexed (Fig. 2f). Using these indices and (average) lattice vectors as initial guesses, the lattice vectors corresponding to the detected Bragg peaks at each individual diffraction pattern can be extracted (Fig. 3a). Finally, the changes in the lattice vectors from one scan position to the next are used to calculate the infinitesimal strain matrix, ε_{ij} (Fig. 3b).

The py4DSTEM package includes functions for scripting, Jupyter notebooks with clearly worked examples including all of the processing described above, and a GUI for quick data visualization and interaction. py4DSTEM defines set structured data object types for 4D-STEM data processing, and establishes a set of HDF5-based file format conventions to facilitate open and reproducible processing of

4D-STEM data. The field of electron microscopy has greatly benefitted from recent efforts to make data analysis open, accessible, and reproducible [3]; py4DSTEM builds on and expands this effort [4].

References:
[1] C. Ophus et al., Microscopy and Microanalysis **20.S3** (2014), p. 62.
[2] https://github.com/bsavitzky/py4DSTEM. All code available under the 'copyleft' GPL v3 license.
[3] de la Peña, Francisco, et al., Microscopy and Microanalysis **23.S1** (2017), p. 214.
[4] The authors gratefully acknowledge support from the Toyota Research Institute. Work at the Molecular Foundry was also supported by the Office of Science, Office of Basic Energy Sciences, of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.
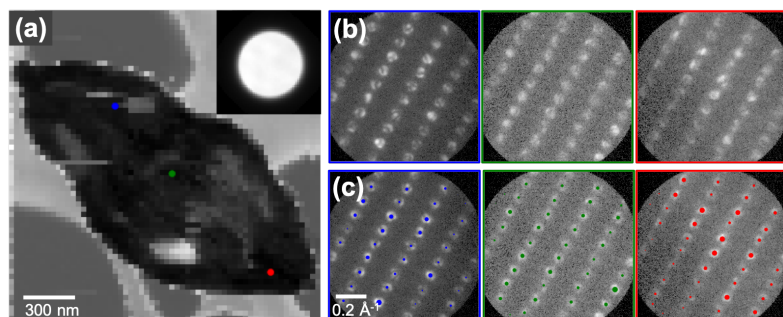
Figure 1. (a) A virtual bright-field image of a $FePO_4$ microplatelet generated from a 4D-STEM dataset. (*Inset*) The vacuum probe, generated by averaging the probe of many scan positions over the thin carbon support. (b) Diffraction patterns corresponding to the three scan positions shown in (a). (c) Bragg peak positions detected in each diffraction pattern by cross correlation with the vacuum probe, with dot sizes scaled by the cross correlation intensity.

Figure 2. (a) The average deconvolution of the 4D-STEM scan, i.e. the sum over real space of all detected Bragg peaks, weighted by intensity. The apparent images of the microplatelet (compare the inset to Fig. 1a) result from the scanning of the beam. (b) Correcting the diffraction shifts collapses each reciprocal lattice peak to a point; the inset shows the central beam. (c) The shift corrected, average deconvolution, with the central beam removed. (d) The Radon transform of (c), with a zone axis shown (*red arrow/dashed line*). (e) A cost function – the summed intensity of the maxima of a single Radon transform slice (e.g. the red dashed line in (d)) divided by the number of maxima – readily identifies the high symmetry directions. (f) The lattice vectors (*white and orange arrows*) and the reciprocal lattice indices of the average deconvolution maxima.
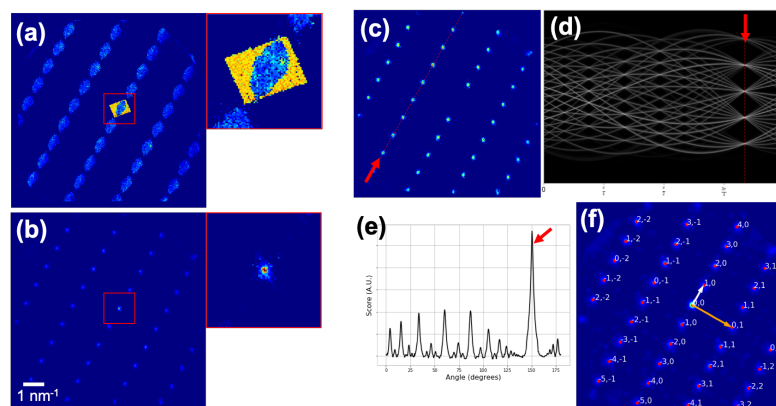

Figure 3. (a) The diffraction patterns shown in Fig. 1b,c, now overlaid with the lattice vectors for these diffraction patterns, calculated using the indexing and lattice vectors from Fig. 2f as initial inputs. (b) Maps of the infinitesimal strain matrix ε_{ij}. The map on the far right shows the rotation of the lattice vectors.