doi:10.1017/S1431927619000795
Microsc. Microanal. 25 (Suppl 2), 2019


# Unsupervised Machine Learning to Distill Structural-Property Insights from 4D-STEM

Xin Li,[1,2] Ondrej Dyck,[1,2] Mark Oxley,[1,2] Andrew Lupini,[1,2] Stephen Jesse,[1,2] Sergei V. Kalinin,[1,2]

[1] Center for Nanophase Materials Science, Oak Ridge National Laboratory, Oak Ridge, TN
[2] Institute for Functional Imaging of Materials, Oak Ridge National Laboratory, Oak Ridge, TN

With recent advances in data acquisition techniques, scanning transmission electron microscopy (STEM), is now able to record the entire scattering distribution, generating a four-dimensional (4D) dataset that includes two real-space dimensions and two reciprocal-space dimensions for every probe position [1,2]. Other studies have shown that 4D-STEM enables probing electromagnetic fields by differential phase contrast (DPC) on segmented detectors [3,4] or by a simplified quantum theoretical interpretation of the pixelated detector [5].

Despite this progress, efficient processing and interpretation of large scale 4D-STEM datasets remain challenging, because of incomplete knowledge of the underlying physical models. Much information in 4D-STEM is currently unexploited. To address this fundamental issue, recently developed unsupervised learning algorithms based on graph-analytics and manifold learning [6,7] will allow subtle details, such local symmetry or defects, to be automatically identified, without any prior bias regarding the material structure and instrumental modality. With extracted patterns from unsupervised learning algorithms, follow-up experiment designs can be conducted for deeper investigation of the system under study.

As an example, we show how manifold learning on a simulated 4D-STEM dataset for single-layer graphene automatically reveals real-space neighbor effects on electron deflection patterns recorded on the pixelated detector, that relate to both individual atomic sites and sublattice structures. Additional examples and detailed study on experimental datasets can be found in [7]. In layman's terms, for straightforward visualization purpose, manifold learning will project the whole diffraction patterns into a 2D space based on pair-wise similarity, with a one-to-one mapping between the points in the 2D manifold plane and the diffraction patterns. Aggregating points in manifold space imply high similarities between the corresponding diffraction patterns. As a result, manifold learning projects the graphene diffraction datasets into the shape of a hexagon, consisting of 7 clusters (aggregating points), as shown in Figure 1a. Real-space positions of the cluster labels (Figure 1b) consist of two groups that are located around the two sublattice atom sites. Clusters whose average diffraction patterns have opposite electron deflection directions are located at opposite sides of the two mirrored atom sites over the two sublattices as illustrated in Figure 1c, (instead of radial symmetry centered at every atom site) revealing real-space neighbor effects on diffraction patterns.

Data-driven unsupervised learning methods are promising to enhance our understandings of such large-scale 4D-STEM datasets, providing local insights into the physical coupling between data-rich imaging mechanisms and material systems.

References:

[1] C Ophus *et al.*, Nat. Commun **7** (2016), p. 10719
[2] TJ Pennycook *et al.*, Ultramicroscopy **151** (2015), pp. 160.
[3] N Shibata *et al.*, Nat. Phys. **8** (2012), pp. 611.

[4] N Shibata *et al.*, Nat. Commun. **8 (**2017**)**, p. 15631.
[5] K Müller *et al.*, Nat. Commun.**5** (2014), p. 5653.
[6] X Li *et al.*, Nat. Commun. **9** (2018), p. 2428.
[7] X Li *et al.*, npj Comput. Mater. **5** (2019), p. 5
[8] Research was partially supported at the Center for Nanophase Materials Sciences, which is a US DOE Office of Science User Facility. Part of this work was supported by the Office of Basic Energy Sciences, Materials Sciences and Engineering Division, US Department of Energy.

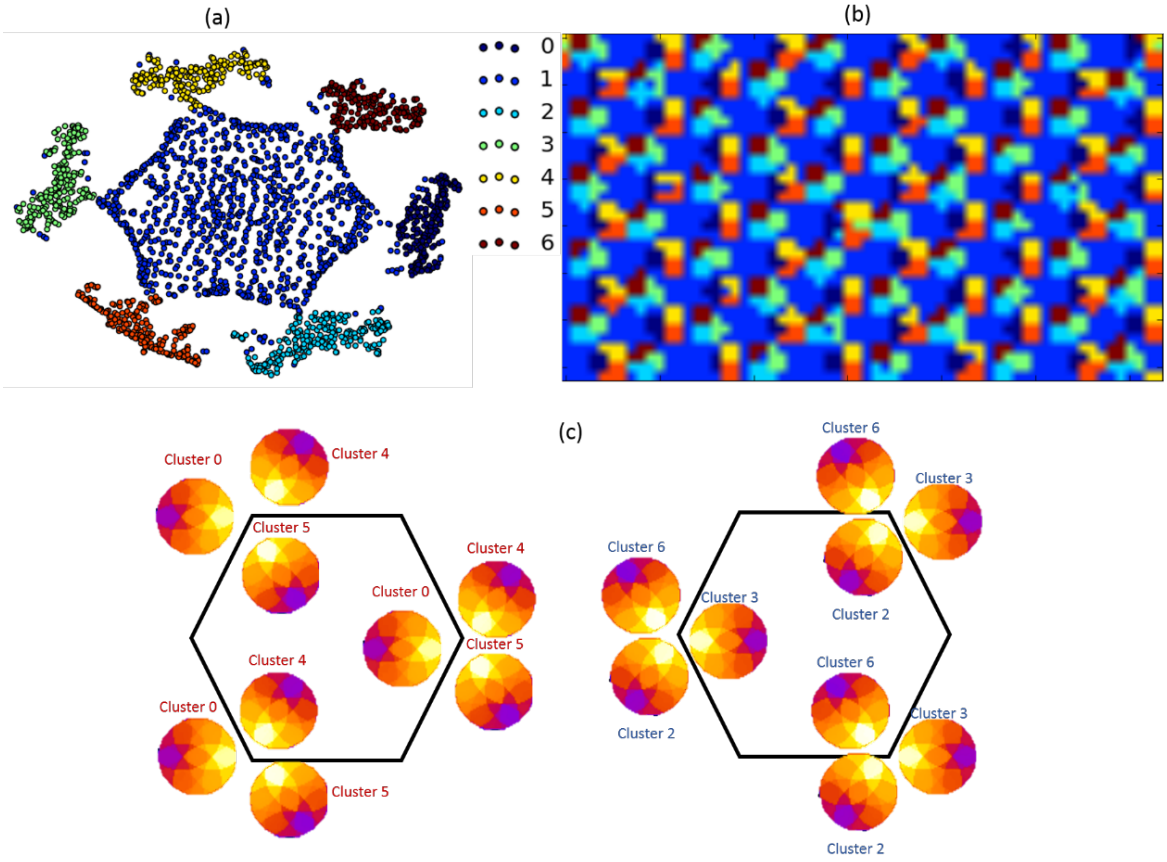


**Figure 1** a) Manifold projection of simulated graphene diffraction datasets, yielding seven clusters. b) Real space distribution of cluster labels. c) Illustration of average diffraction patterns of clusters and relative positions to the atom sites over two sublattices.

https://doi.org/10.1017/S1431927619000795 Published online by Cambridge University Press