Using Multivariate Analysis of Scanning-Rochigram Data to Reveal Material Functionality

Stephen Jesse[1,2], Miaofang Chi[1,2], Albina Y. Borisevich[1,3], Alex Belianinov[1,2], Sergei V. Kalinin[1,2], Eirik Endeve[1,4], Richard K. Archibald[1,4], Christopher T. Symons[1,5], Andrew R. Lupini[1,3]

[1] Institute for Functional Imaging of Materials, Oak Ridge National Laboratory, Oak Ridge, TN 37831
[2] Center for Nanophase Materials Sciences, Oak Ridge National Laboratory, Oak Ridge, TN 37831
[3] Materials Science & Technology Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831
[4] Computer Science and Mathematics Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831
[5] Computational Sciences & Engineering Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831

The scanning transmission electron microscopy (STEM) is a powerful platform for studying materials and their many varied properties (including structural, electronic, magnetic, ferroic etc.) locally and at their fundamental length-scales. In its current standard implementation, STEM imaging is severely restricted by the instrumental linkage in which the information rich electron diffraction pattern (Ronchigram) is reduced to a single value (e.g. through the integrated intensity over a relatively large detector) at each beam location resulting in loss and distortion of information about important but subtle aspects of material properties. Previous theoretical and experimental work [1, 2, 3,4] suggests that full acquisition of the Ronchigram at each spatial location during a scan can enable super resolution, phase-contrast imaging, imaging of internal fields, and 3D sample reconstruction. Data acquisition and storage has evolved to a point now that it is possible to capture high resolution 4D ptychography data sets rapidly, and the use of large-scale compute facilities enables the processing and mining of these multi-GB data sets to distil the most salient aspects of the data while separating the statistically significant variations (signal) from noise.

In the current work we have utilized the DE-12 camera (Direct Electron, LP, San Diego, CA), equipped with a 4096 x 3072 pixels Direct Detection Device (DDD®) sensor [5] installed on an aberration corrected FEI Titan operating at 300 kV combined through a custom FPGA control system to synchronize frame capture and beam positioning to acquire 4D scanning-scattering data sets. Typical scans of 192X192 pixel images containing 384X384 Ronchigrams at every pixel were captured in less than 1 minute. A host of multivariate statistical methods can be brought to extract meaningful information from these large data sets. Statistical methods can be advantageous in that they are parallelizable and can be efficiently implemented on large-scale computing platforms, are model-free and operate with no pre-imposed expectation or bias, and are capable of elucidating inter-pixel correlations unlikely to be noticed by human observation alone.

As an example, Figure 1b shows the first 16 eigenvectors derived from principal component analysis (PCA) applied to the 4D data set, and figure 1c shows the results of clustering the first 128 PC's into 4 clusters. Figure 1d shows the spatial location of the 4 clusters and the clear identification of the ferroelectric sample into 2 distinct domains. Though PCA can be used to reduce the dimensionality of information within a data set, it may not always be the most efficient means to compress the data and the output of PCA is often not well suited for finding clusters that exist in subspaces of the data. In order to simultaneously address the difficulty of defining and finding neighbors in extremely high dimensions and the desire to find clusters that are locally defined, we utilize an approach that creates local, patch-based Multidimensional Spectral Hashing (MDSH) [6] codes for dimensionality reduction. MDSH is

used to generate small set of dimensions in which similarity among close points is preserved, while allowing distinctions between larger distances to become indistinguishable. This enables us to create a space with many fewer dimensions in which local clustering is preserved. Here, we will explore the use of high-dimensional techniques that automatically select an appropriate number of clusters [7], as well as explore how variations in the stride and receptive field of the patches impact the results. [8]

References:
[1] J. M. Rodenburg, B. C. McCallum and P. D. Nellist, Ultramicroscopy, **48**, (1993), p. 304.
[2] N. Shibata *et al*, Nature Physics, **8**, (2012), p. 611.
[3] T.J. Pennycook *et al*, Ultramicroscopy, 151, (2015), p. 160
[4] C. Ophus *et al*, Microsc. Microanal. **20**, 2014
[5] The authors would like to acknowledge Liang Jin and Benjamin Bammes for their assistance in allowing us to use the DE-12 system. http://www.directelectron.com/.
[6]Y. Weiss, R. Fergus, and A. Torralba. *Multidimensional Spectral Hashing*. in European Conference on Computer Vision. 2012.
[7]R.R. Vatsavai, *et al. GX-Means: A Model-Based Divide and Merge Algorithm for GeoSpatial Image Clustering.* in Procedia Computer Science. Vol. 4. 2011. p. 186-195.
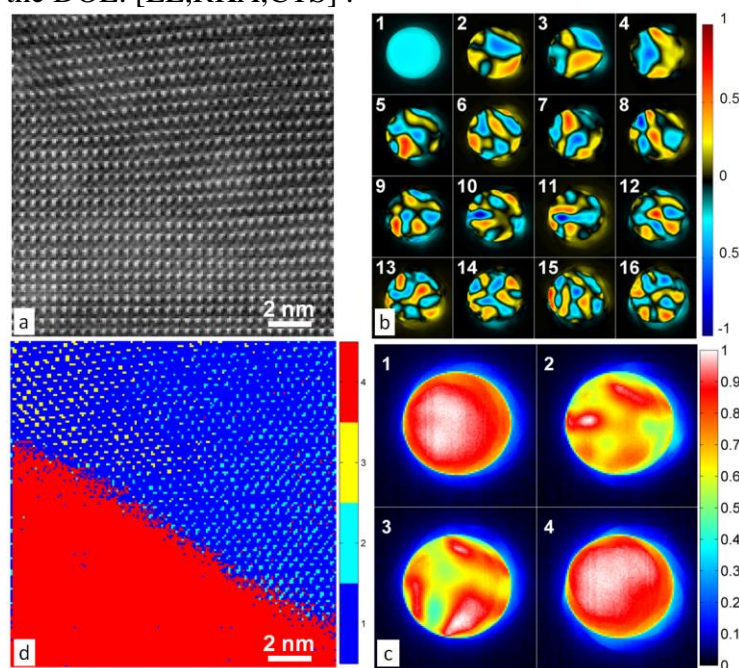
Figure 1. (a) Micrograph of tetragonal/rhombohedral phase boundary of thin film Bismuth Ferrite. (b) The first 16 PCA eigenvectors. (b) Results of clustering of the first 128 eigenvectors into 4 distinct groups (d) Map of cluster locations shows clear delineation between phases based entirely on statistical analysis of Ronchigrams.