doi:10.1017/S1431927616007935
Microsc. Microanal. 22 (Suppl 3), 2016

Deep Data Mining in a Real Space: Application to Scanning Probe Microscopy Studies on a "Parent" State of a High Temperature Superconductor

Maxim Ziatdinov[1,2], Artem Maksov[1,2], Athena Sefat[3], Peter Maksymovich[1,2], and Sergei Kalinin[1,2]

[1.] Institute for Functional Imaging of Materials, Oak Ridge National Laboratory, Oak Ridge, TN, USA
[2.] Center for Nanophase Materials Sciences, Oak Ridge National Laboratory, Oak Ridge, TN, USA
[3.] Materials Science and Technology Division, Oak Ridge National Laboratory, Oak Ridge, TN, USA

Nanoscale inhomogeneity of structural and electronic orders in a crystalline matter can have a powerful and non-random effect on the macroscopic properties of technologically relevant materials. Scanning tunneling microscopy and spectroscopy (STM/S), which probes topographic and electronic properties of the surfaces with a nanometer-scale resolution, constitutes an important experimental tool for exploring local inhomogeneity in these materials. The STS mode allows to acquire 3D $G(X,Y,U)$ datasets, where $G=\mathrm{d}I/\mathrm{d}U$ corresponds to a value of a differential conductance proportional to LDOS at the specific energy $E=\mathrm{e}U$ at each (X,Y) point. Unfortunately, in many strongly correlated materials, such as high-temperature superconductors (HTSC), there is usually no direct and simple connection between behavior of STS curves and the surface topographic ("structural") features. As the ever-increasing amount of STM/S data on strongly correlated systems makes the individual inspection of datasets highly impractical and, in many cases, nearly impossible, there is an urgent need for developing deep data mining tools for a reliable identification and spatial mapping of statistically significant electronic behaviors in an automated fashion of a full information extraction [1]. Here we present an approach based on *k*-means clustering, principle component analysis (PCA) and Bayesian linear unmixing (BLU) used to uncover "buried" electronic phases from the STS datasets of a parent magnetic state of iron-pnictide HTSC.

As an example, in Fig. 1 (a) we show the surface topographic image, on which the stack of 768 STS conductance maps were acquired with $X\times Y=50px\times 50px$ spatial resolution. To decorrelate the STS dataset we start with finding a minimum number of relevant electronic behaviors using the *k*-means algorithm. We found that the separation is best described by 3 clusters; the resultant spatial distribution of these clusters is shown in Fig. 1 (e-g). Mean curves associated with each of 3 clusters are shown in Fig. 1 (b-d). Analysis of a variance in the STS curves distributed over each cluster by means of PCA revealed a relatively moderate variance in clusters 2 and 3. It allow us to ascribe a physically-defined phase to these clusters, namely, the SDW phase whose behavior is altered between two clusters due to the defect-induced strain. The situation is quite different for cluster 1. Here, a stronger variance in the shape of STS curves, especially in the regions close to the Fermi level, does not allow assigning any physically well-defined phase. This indicates that the total number of relevant electronic behaviors in the STS dataset is larger than identified by the k-means method.

To perform a more thorough and detailed separation of electronic phases in the STS dataset we adopt the BLU algorithm which separates linear mixtures of spectral sources under non-negativity and full additivity constraints [2]. We assume that the total STM current at each pixel in the dataset can be represented as a linear combination of the currents flowing through each of the available "channels", so that the latter can be represented by the endmembers. The BLU results for k=6 endmembers are shown in Fig. 2. It becomes immediately clear that endmembers 4 and 5 collaborate the results on SDW-associated phase found earlier from *k*-means algorithm. In addition to phases already seen in the *k*-

means, the BLU analysis revealed new features in the electronic behavior that can be linked to the fundamental physics of the material. For example, the endmember 2 can be described in terms of a pseudogap state observed in recent spectroscopic measurements on a bulk sample, whereas the endmembers 1 and 6 are linked to the impurity-induced bound states, in a reasonable agreement with existing theoretical proposals. Overall, our deep data mining approach helped to uncover a wealth of a "hidden" information relevant to physical properties of iron-based superconductors [3].

References:

[1] S.V. Kalinin, *et al*., Nature Mater. **14** (2015), p. 973.
[2] N. Dobigeon, *et al*., IEEE Transactions on Signal Processing **57** (2009), p. 4355.
[3] This research was sponsored by the Division of Materials Sciences and Engineering, BES, DOE. Research was conducted at the Center for Nanophase Materials Sciences, which is a DOE Office of Science User Facility.

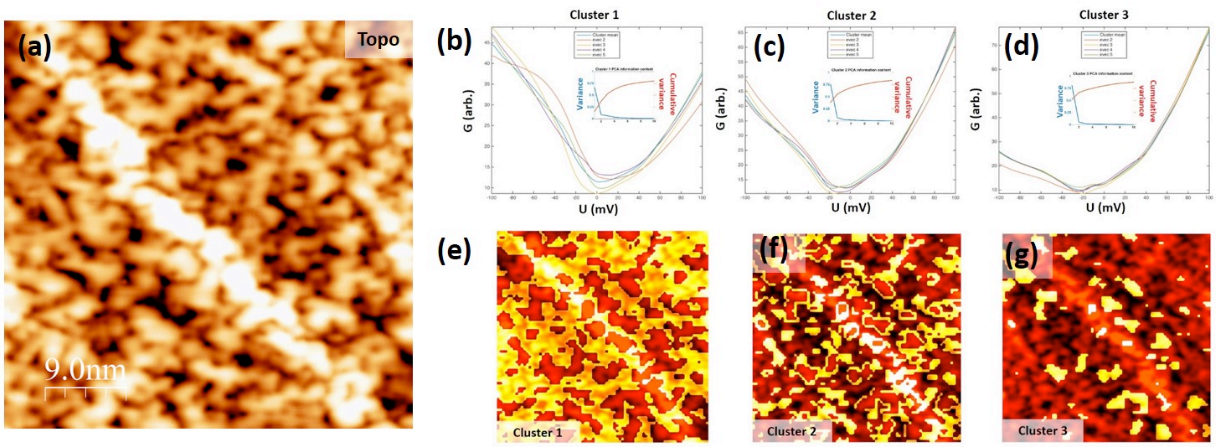

Figure 1. (a) STM topographic image of the area on which STS grid measurements were performed. (b-d) Mean STS curves for each of 3 *k*-means derived clusters are plotted with the PCA-derived deviation from the mean curve. (e-g) Corresponding spatial maps of clusters distribution.

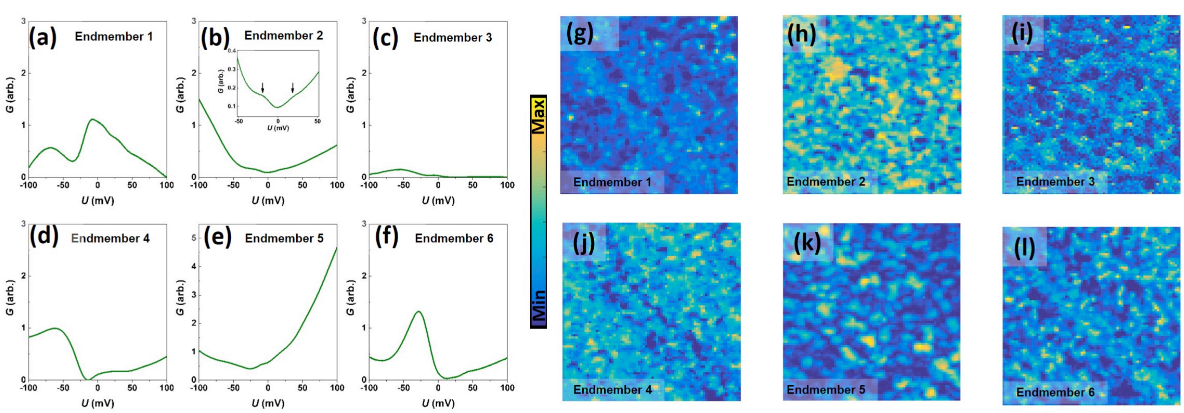

Figure 2. (a-f) 6 Bayesian endmembers. (g-l) Corresponding abundance maps.

https://doi.org/10.1017/S1431927616007935 Published online by Cambridge University Press