Spectral Classification of Structurally Organized Adatom Configurations

Kevin Roccapriore, Zheng Gai, Bobby Sumpter, Mina Yoon and Sergei V. Kalinin

Oak Ridge National Laboratory, Oak Ridge, Tennessee, United States

In scanning tunneling microscopy (STM), structural organizations of atomic arrangements can have a direct impact on their electronic properties. Put another way, the electronic density of states which is probed in scanning tunneling spectroscopy (STS), will be modified due to different local atomic configurations. In particular, when cleaving a material, depending on the crystal structure and angle of the cleave, the resulting surface can be populated by adatoms which lie in a variety of different structural positions. The aim of this paper is to demonstrate a machine learning approach that classifies atoms based on their local structural environment, and subsequently extracts and statistically analyzes the associated spectral characteristics. This method uses hyperspectral datasets in which the structural data is acquired simultaneously and therefore the two are aligned in real space.

The classification problem can be tackled in many ways – all of which initially require accurate atomic coordinates. In the STM experimental datasets, there is a considerable variability in the intensity as well as the size of the adatoms within even a single image. This is a direct result of measuring the electronic density of states, therefore atoms with different nearest neighbor configurations will naturally have variability in their size and intensity. For this reason, we train and then use a convolutional neural network to retrieve the coordinates of both the adatoms and the underlying surface atoms. Here, we are interested only in the different adatom configurations and not the underlying surface, therefore, we use only the adatom coordinates to generate a new object consisting of 2D Gaussian blobs located at each found adatom coordinate. This essentially acts as a background filter which allows the subsequent structural classification to have better performance.

We seek to realize what spectral effect () the local atomic environment around each adatom causes. Many statistics are required to have a good representation of the spectral characteristics observed in a certain configuration. There exist several methods to perform a structural classification in an unsupervised manner, for example, a Gaussian Mixture Model (GMM). These GMMs exist in the form where the number of classes is either known or unknown[1]. The problem with this and other similar unsupervised methods is that even if the number of classes is known, structural classes are likely prioritized differently than what is desired. For example, consider a hexagonal lattice, where an atom has two nearest neighboring adatoms separated by 180 degrees, there are three different rotations that a human will understand to be the same configuration, but the unsupervised method may separate these based on their different rotations. We therefore turn to a supervised method, in which the classes are predefined. A simple method of analyzing the nearest neighbor of each 2D Gaussian is used in which if the intensity in specifically chosen regions of interest exceeds a threshold, a neighbor is said to be present in that location. After sorting atoms by their structure, we analyze the spectral characteristics within each class. The average spectra within a class serves to represent the entire class, while the standard deviation can provide a level of uncertainty.

Interestingly, this is taken a step further and additional, unexpected information likely relating to the structure beyond the first nearest neighbors is extracted. This is accomplished by a subsequent unsupervised learning step performed on the spectra of a single class. In an exploratory manner, Principal Component Analysis[2,3] (PCA) is applied to each structural class's spectral content, statistically significant components are separated, and a refined classification is made that is based on both first nearest neighbor structure and adatom spectra. Unexpectedly, the spectral behavior of adatoms in the same structural configuration, relative to their first nearest neighbor, wildly varies. It appears that non-local effects beyond the first nearest neighbors modify the

adatom spectral behavior and consequently their electronic density of states, and this discovery is made clear by the application of machine learning methods to further classify the structural classes. To this end, we begin to understand the reasons for spectral variation that are not attributed to the first nearest neighbors.

Ultimately, we find that a combination of informed supervised and unsupervised learning can reveal properties and statistical quantities that were initially a pursuit, as well as additional characteristics that were not predicted.
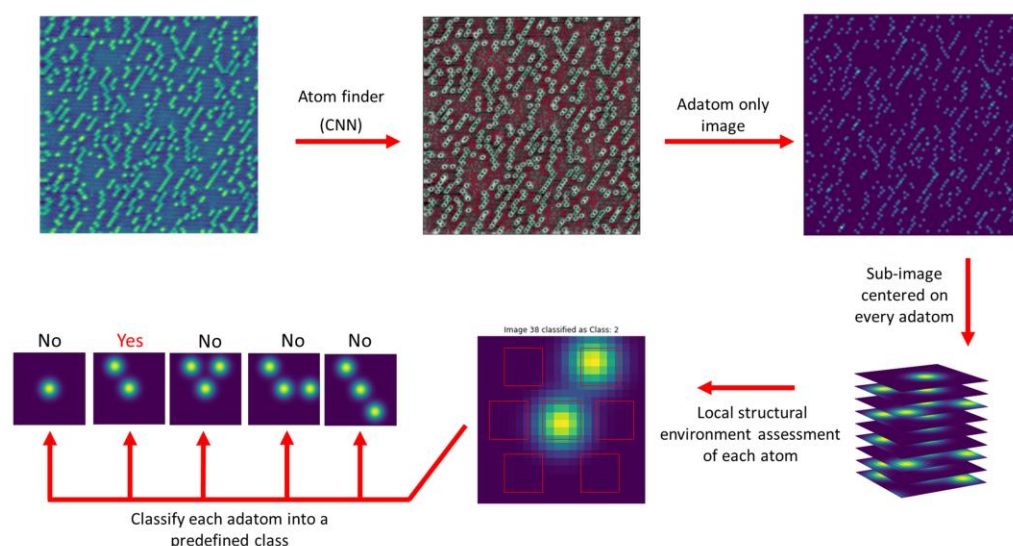

Figure 1. Diagram representing the process of extracting statistical information of the spectra of different structurally organized adatoms. First, atom coordinates are found by applying a trained convolutional neural network, then generating an image of strictly the features of interest (e.g., adatoms only). At each feature, a smaller "sub-image" is made showing the local structural environment up to first nearest neighbors of the feature, and this local environment is analyzed in specific regions of interest to classify into predetermined structural configurations. Finally, the spectra within each class are statistically analyzed.

References
1. Rasmussen, C. E. The Infinite Gaussian Mixture Model. in *Advances in Neural Information Processing Systems 12* (eds. Solla, S. A., Leen, T. K. & Müller, K.) 554–560 (MIT Press, 2000).
2. Jolliffe, I. T. & Cadima, J. Principal component analysis: a review and recent developments. *Philos. Transact. A Math. Phys. Eng. Sci.* **374**, (2016).
3. Jolliffe, I. T. Principal Component Analysis and Factor Analysis. in *Principal Component Analysis* (ed. Jolliffe, I. T.) 115–128 (Springer, 1986). doi:10.1007/978-1-4757-1904-8_7.