Learning the Physics and Chemistry of Surfaces via Machine Vision and Deep Data Analysis

Maxim Ziatdinov[1,2], Artem Maksov[2,3], Sergei V. Kalinin[1,2]

[1]Institute for Functional Imaging of Materials, Oak Ridge National Laboratory, Oak Ridge TN 37831, USA
[2]Center for Nanophase Materials Sciences, Oak Ridge National Laboratory, Oak Ridge TN 37831, USA
[3]Bredesen Center for Interdisciplinary Research, University of Tennessee, Knoxville, TN 37996, USA

Email: ziatdinovma@ornl.gov

ABSTRACT

Advances in electron and scanning probe microscopies have opened the door to nanoscience and nanotechnology by enabling imaging and manipulation of the structure and functionality of materials. In many technologically relevant atomic and molecular systems, however, the information of interest is usually spatially distributed and may have a complex multi-dimensional nature. One of the critical issues, therefore, lies in being able to accurately identify all the individual building blocks in different atomic/molecular architectures on a scale of thousands of individual units in a fashion of full information extraction, and to link both minute deviations in local structure and large-scale assembly properties in statistically significant manner. Here we demonstrate a framework for automated and highly accurate analysis of structural and functional properties, as well as their spatially dependent relationships, in multi-modal microscopic imaging based on deep data analysis and machine vision tools.

Keywords: scanning probe microscopy, image analysis, machine learning, graphene, self-assembly.

SCANNING PROBE MICROSCOPY AND "BIG DATA"

The last decade has witnessed a dramatic increase in the size and quality of datasets produced by microscopic and spectroscopic experimental techniques. Particularly, the recent advances in scanning transmission electron (STEM) and scanning probe microscopies (SPM) have opened an unprecedented path towards probing the materials structural parameters and various functional properties in real space with a sub-nanometer precision.[1,2] For example, the subfields of SPM such as non-contact atomic force microscopy and scanning tunneling microscopy are known to provide an unprecedented, angstrom-resolved visual insight into a nature of chemical bonds[3] and spatial behavior of electronic density of states on a surface[4], respectively. Meanwhile, STEM experiments can produce picometer-resolved images of ferroelectric polarization[5], octahedral tilts[6], and chemical expansion strains[7]. Such experimental capabilities require adequate analytical methods for extracting a relevant physical and chemical parameters from the very large datasets in which the information of interest is spatially distributed and has a complex multi-dimensional nature. Here we demonstrate case studies involving our use of machine learning and deep data analysis for reading and recognizing complex molecular, atomic and electronic patterns on surfaces and elucidating physical and chemical mechanisms behind the derived structure-property correlations.

DATA MINING THE STRUCTURE-PROPERTY RELATIONSHIPS

The link between changes in the material crystal structure and its mechanical, electronic, magnetic, and optical functionalities – known as the structure-property relationship – is the cornerstone of the contemporary materials science research[8,9]. It allows scenarios in which relatively small changes in the material crystal structure may have a decisive impact on the physical properties of the system. Hence, it should be in principle possible to "break" the crystal lattice system down to the individual structural fragments and/or distortion modes and identify the geometric features that most influence the physical property of interest. To date, however, there has been no general and statistically-meaningful framework for realizing this vision for experimental data. To address this problem, we have developed a comprehensive approach for correlative analysis of information obtained in structure 'channel' and function

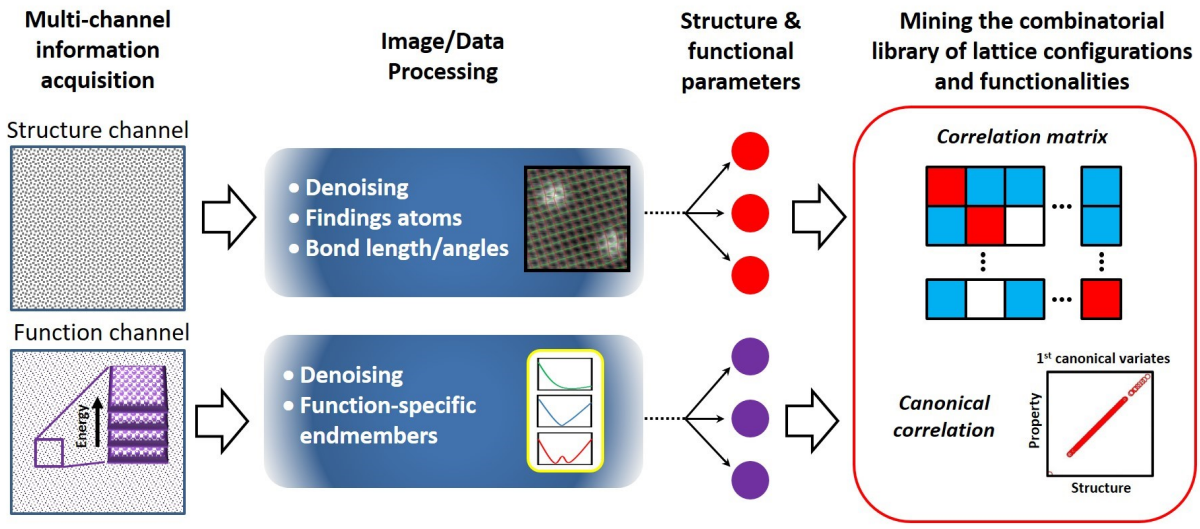

FIGURE 1. Schematic workflow for data mining the structure-property relationships of materials using data from multimodal electron or scanning probe microscopic measurements.

'channel' of atomically-resolved scanning probe microscopy experiments, and illustrated our approach using data obtained on graphene and iron-based superconductors.

Our approach for a two-channel microscopic imaging experiment is schematically outlined in Figure 1. It starts with recording 'structure' and 'function' information over the same sample area via two different acquisition channels. In this case, the first channel corresponds to 2D $r(X, Y)$ images in which r is a 'structural' variable used to calculate lattice parameters, such as bond lengths and angles. The second channel represents 3D $f(X, Y, V)$ dataset in which f is a 'function' variable, for example, differential conductance or electron energy loss. The data from both channels is cleaned from measurement noise using principal component analysis that allows to minimize the information loss. The central idea behind principal component analysis based image filtering is that the most relevant information (i.e., the 'signal') can be represented by a relatively small number of principal components with the largest variance, whereas the rest of the (low-variance) components correspond to 'noise'.The next step is to use cleaned data for constructing structural and functional descriptors from various atom finding tools (if atomic resolution is available), pattern recognition techniques, and blind source un-mixing/decomposition methods. Once exraction of all physically relevant descriptors is completed, we proceed to performing direct data mining of structure-property relationships via correlative analysis of the corresponding structural and functional descriptors.

We particularly focus on the analysis of a role of chemical and structural nanoscale disorder in the electronic properties of graphene and unconventional superconductors. In both of these materials, the presence of atomic vacancies, impurity atoms, and lattice strain creates a perturbation that can strongly disrupt the surrounding electronic environment and influence the overall macroscopic properties of the system.[10-13] Scanning tunneling microscopy and spectroscopy (STM/S), which probe topographic and electronic properties of the surfaces with a nanometer-scale resolution, constitutes an ideal experimental tool for probing such perturbation by measuring a spatial dependence of local density of electronic states (LDOS).

We first present an approach based on combination of blind linear unmixing, automated atom finding and local indicators of spatial association for highly-accurate separation, extraction, and correlation of structural and electronic behaviors in high resolution STM/S experiments on iron-based superconducting compounds FeSe and Au-doped $BaFe_2As_2$. In our analysis, we identify characteristic STS spectral features using physically robust Bayesian linear unmixing and non-negative matrix factorization, and show their direct relevance to the fundamental physical properties of the system, including electronic states associated with individual defects and impurities, as well as states linked to electron scattering from the defect sites. We also collect structural data from individual unit cells on the crystalline lattice, and calculate both global and local indicators of spatial correlation with electronic features, demonstrating, for the first time, a direct quantifiable connection between observed structural order parameters extracted from the STM data and electronic order parameters identified within the STS data. We were also able to identify different types of chemical disorder and to visualize a peculiar aggregation of dopants along the boundary between domains with drastically different dopant concentrations.

We next demonstrate our analysis of structure-property relationships for SPM data on graphene with hydrogenated and oxidized defects (Fig. 2). Specifically, we apply a combination of sliding window fast Fourier transform, Pearson correlation matrix, and linear and kernel canonical correlation, to study a relationship between lattice distortions and electron scattering. Our analysis revealed that the

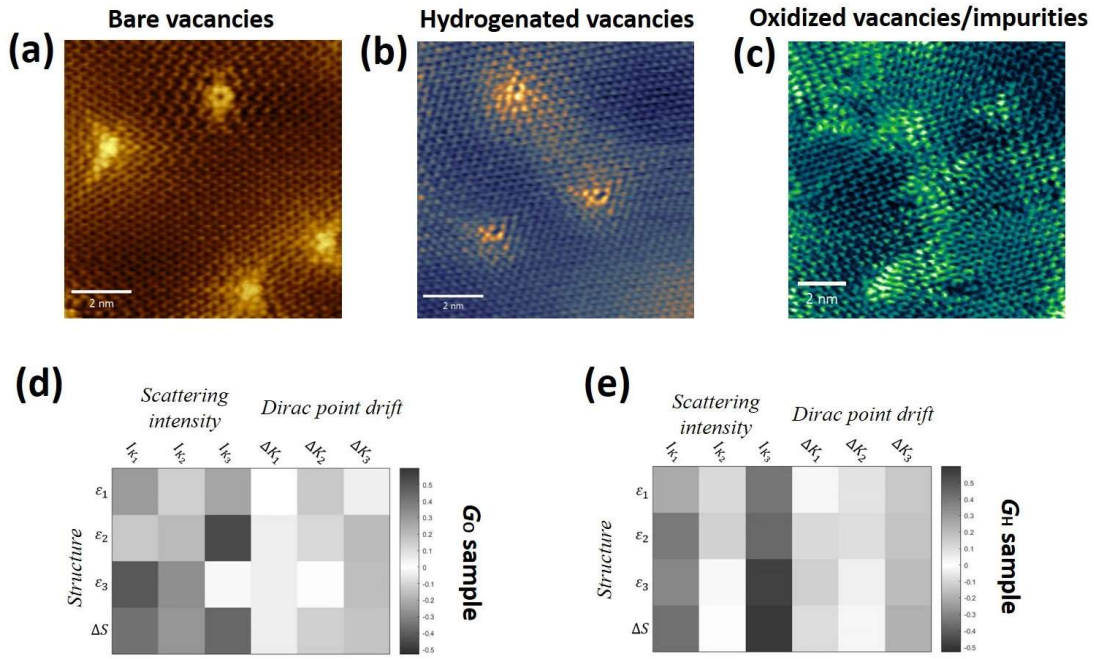

FIGURE 2. Correlative analysis of structural and electronic properties in graphene with different types of defects. (a-d) Scanning probe microscopy images of graphene with un-saturated vacancies (a), hydrogen-passivated vacancies (b), and oxidized vacancies and impurities (c). (d-e) Correlative matrix analysis for graphene samples with oxidized defects (d) and with hydrogenated defects (e)

strength of coupling to strain is altered between different paths in a momentum space, which can explain a peculiar coexistence of several electron interference patterns in nanoscale regions of interest. In addition, the application of kernel functions allowed us to extract a non-linear component of the relationship between the lattice strain and scattering intensity in graphene.

LEARNING COMPLEX ATOMIC, MOLECULAR, AND ELECTRONIC PATTTERNS VIA MACHINE VISION

We finally discuss applications of machine vision tools, such as Markov random field and convolutional neural networks, to analyze of 2D molecular self-assemblies. The ability to utilize molecular assemblies as data storage devices requires a development of methods to identify individual molecular states in non-strictly-periodic structures on a scale of thousands of molecules. We present a novel method of applying machine learning techniques for extraction of positional, structural, and rotational information from ultra-high vacuum molecular-resolved STM images and apply it to self-assembled monolayer of π-bowl nanocarbon molecules on gold. The molecule can preside in two different conformations (bowl-up and bowl-down), as well as in multible azimuthal rotational state on the metallic substrate. To ascertain the applicability and robustness of our machine learning and pattern recognition methods for general STM data, we start with constructing a synthetic dataset on a model system. Specifically, we use density functional theory-generated templates in conjunction with Markov Chain Monte Carlo sampler and noise modeling to create synthetic images representative of our model. We extract positional information of each molecule and use nearest neighbor criteria to construct a graph input to Markov Random Field model to identify molecular confromational states. We then train a convolutional Neural Network on a synthetic dataset and combine it with Markov random field model to classify molecules based on their azimuthal rotational state. The effectiveness of such approach compared to other, "standard", methods is discussed. Finally, we apply our approach to experimental images and achieve complete extraction of information assciated with different molecular conformations and rotations. The obtained full decoding at the nanoscale level allows us to directly construct both relevant pair density functions – a centerpiece in analysis of disorder-property relationship paradigm, as well as more complex structural descriptors. This in turn allows us to explore how individual blocks may form certain short-range orders, as well as to analyze potential (spatial) correlations between multiple order parameters, and to use the obtained information for constructing a reaction path for molecule conformational changes in the self-assembly.

FUTURE DIRECTIONS

Finally, we discuss future research directions that would allow incorporation of theoretical predictions/simulations *directly* into machine learning schemes. We forsee an application of the so-called domain-adversarial training of

neural networks[14] which allows to alter theoretically predicted classes based on the observed data. The underlying idea of this approach is that the theoretical and experimental datasets are similar yet different in such a way that traditional neural networks may not capture correct features just from the labeled data.

REFERENCES

(1) Kalinin, S. V.; Sumpter, B. G.; Archibald, R. K. Big-deep-smart data in imaging for guiding materials design. *Nat Mater* **2015**, *14*, 973-980.

(2) Kalinin, S. V.; Pennycook, S. J. Microscopy: Hasten high resolution. *Nature* **2014**, *515*.

(3) de Oteyza, D. G.; Gorman, P.; Chen, Y.-C.; Wickenburg, S.; Riss, A.; Mowbray, D. J.; Etkin, G.; Pedramrazi, Z.; Tsai, H.-Z.; Rubio, A.; Crommie, M. F.; Fischer, F. R. Direct Imaging of Covalent Bond Structure in Single-Molecule Chemical Reactions. *Science* **2013**.

(4) Wang, Y.; Wong, D.; Shytov, A. V.; Brar, V. W.; Choi, S.; Wu, Q.; Tsai, H.-Z.; Regan, W.; Zettl, A.; Kawakami, R. K.; Louie, S. G.; Levitov, L. S.; Crommie, M. F. Observing Atomic Collapse Resonances in Artificial Nuclei on Graphene. *Science* **2013**.

(5) Jia, C.-L.; Mi, S.-B.; Urban, K.; Vrejoiu, I.; Alexe, M.; Hesse, D. Atomic-scale study of electric dipoles near charged and uncharged domain walls in ferroelectric films. *Nat Mater* **2008**, *7*, 57-61.

(6) Borisevich, A.; Ovchinnikov, O. S.; Chang, H. J.; Oxley, M. P.; Yu, P.; Seidel, J.; Eliseev, E. A.; Morozovska, A. N.; Ramesh, R.; Pennycook, S. J.; Kalinin, S. V. Mapping Octahedral Tilts and Polarization Across a Domain Wall in BiFeO3 from Z-Contrast Scanning Transmission Electron Microscopy Image Atomic Column Shape Analysis. *ACS Nano* **2010**, *4*, 6071-6079.

(7) Kim, Y.-M.; He, J.; Biegalski, M. D.; Ambaye, H.; Lauter, V.; Christen, H. M.; Pantelides, S. T.; Pennycook, S. J.; Kalinin, S. V.; Borisevich, A. Y. Probing oxygen vacancy concentration and homogeneity in solid-oxide fuel-cell cathode materials on the subunit-cell level. *Nat Mater* **2012**, *11*, 888-894.

(8) Le, T.; Epa, V. C.; Burden, F. R.; Winkler, D. A. Quantitative Structure–Property Relationship Modeling of Diverse Materials Properties. *Chemical Reviews* **2012**, *112*, 2889-2919.

(9) Isayev, O.; Fourches, D.; Muratov, E. N.; Oses, C.; Rasch, K.; Tropsha, A.; Curtarolo, S. Materials Cartography: Representing and Mining Materials Space Using Structural and Electronic Fingerprints. *Chemistry of Materials* **2015**, *27*, 735-743.

(10) Ziatdinov, M.; Fujii, S.; Kusakabe, K.; Kiguchi, M.; Mori, T.; Enoki, T. Direct imaging of monovacancy-hydrogen complexes in a single graphitic layer. *Physical Review B* **2014**, *89*, 155405.

(11) Ziatdinov, M.; Maksov, A.; Li, L.; Sefat, A. S.; Maksymovych, P.; Kalinin, S. V. Deep data mining in a real space: separation of intertwined electronic responses in a lightly doped $BaFe_2As_2$. *Nanotechnology* **2016**, *27*, 475706.

(12) Ziatdinov, M.; Fujii, S.; Kiguchi, M.; Enoki, T.; Jesse, S.; Kalinin, S. V. Data mining graphene: correlative analysis of structure and electronic degrees of freedom in graphenic monolayers with defects. *Nanotechnology* **2016**, *27*, 495703.

(13) Zeljkovic, I.; Hoffman, J. E. Interplay of chemical disorder and electronic inhomogeneity in unconventional superconductors. *Physical Chemistry Chemical Physics* **2013**, *15*, 13462-13478.

(14) Ganin, Y.; Ustinova, E.; Ajakan, H.; Germain, P.; Larochelle, H.; Laviolette, F. o.; Marchand, M.; Lempitsky, V. Domain-Adversarial Training of Neural Networks. *ArXiv e-prints* 2015. http://adsabs.harvard.edu/abs/2015arXiv150507818G.