3078
doi:10.1017/S1431927622011473
Microsc. Microanal. 28 (Suppl 1), 2022

Finding Features from Microscopes to Simulations Via Ensemble Learning and Atomic Manipulation

Ayana Ghosh,[1,2], Kevin M. Roccapriore,[1] Bobby Sumpter,[1] Ondrej Dyck,[1] Maxim Ziatdinov[1,2] and Sergei V. Kalinin[3]

[1.] Center for Nanophase Materials Sciences, Oak Ridge National Laboratory, Oak Ridge, TN, USA
[2.] Computational Sciences and Engineering Division, Oak Ridge National Laboratory, Oak Ridge, TN, USA
[3.] Department of Materials Science and Engineering, University of Tennessee, Knoxville, TN, USA

Electron and scanning probe microscopies have become one of the primary techniques to investigate systems in the domain of physical and life sciences at the atomic to mesoscopic length scales [1-5]. From the perspective of computational simulations performed at different length scales, there has also been significant advancements with accessibility to faster, parallelizable CPU/GPU architectures. This means there is a heap of information from the simulations performed on a variety of materials, particularly investigating corresponding structures and functionalities. Hence, structural, and spectral data resulting from these experiments along with simulations create an enormous exploratory platform to utilize, design deep learning (DL), machine learning (ML) workflows for feature finding along with atomic manipulation. However, systematic studies combining all three avenues to go beyond comparing endpoint properties, performing DL/ML analyses on static data, assessing structure-property relationships are still in its infancy.

There are multiple challenges associated with designing such across-the-board frameworks. The time scales of STEM observations and intrinsic molecular dynamics are strongly different, with the DFT and MD models capable of simulating system sizes of Å to nanometer scales for up to microseconds, but taking multiple CPU hours, while STEM images are typically available at the fraction of a second. At the same time, disparity in the latencies of calculations is significant, with the DFT or MD simulations often taking many hours to days and weeks of time, well above the timescale of STEM measurements. The differences in length and timescales of such simulations, DL/ML approaches and retrieval of data from microscopes are shown by **Figure 1**. In addition, finding features such as atoms, defects (nearly identical objects) from a microscopic image using DL model and extending it to recognize features from images retrieved under different experimental conditions leading to out-of-distribution effects, is itself a challenge. Defining regions of interest such as parts of the image showing defects and determining the origins of such defects (could be electron-beam induced) is also not trivial. Importing coordinates of atoms directly predicted by either a DL model or experiments to simulations require quantifying uncertainties at all stages of the framework. Ways to close the loop where information from the theoretical simulations can guide future experiments and on-the-fly analyses, are also yet to be investigated.

In this work, we show how ensemble-based framework (**Figure 2**) can be constructed for robust feature finding, uncertainty quantification in predictions. This ensemble learning iterative training (ELIT) [6-8] workflow aims at rapidly adapting to changes to the imaging conditions and successfully locate features to make the analyses efficient. The EL section of the workflow allows for selection of artifact-free features and pixel-wise uncertainty maps by combining multiple networks. Switching from image-

specific to materials-specific descriptors and training networks on the first frame of images obtained at the initial stages of the experiment can help to account for out-of-distribution effects and improve model performance. The IT part retrains the networks with already realized features, focusing its attention on features present in the (heavily degenerate) data and thus increasing the detection limit of the network on the dataset(s) of interest. This framework has been implemented on-the-fly to analyze STEM images. The outputs of the network models are then used to construct supercells perform density functional theory simulations to find optimized geometry of the structures followed by studying temperature-dependent dynamics of system evolutions with ad-atoms and defects. The results along with associated uncertainties in predictions at various levels as obtained utilizing this framework may be used to evaluate and modify experimental conditions and regions of interest.

We aim to further expand this approach for atomic manipulation by incorporating edge-computing. This includes direct transfer of image-based data from microscopes, training suitable DL models using a GPU-based platform followed by performing simulations using CPU-based high-performance computing resources. A real-time-feedback loop can then be established to better guide experiments while learning from theoretical models.

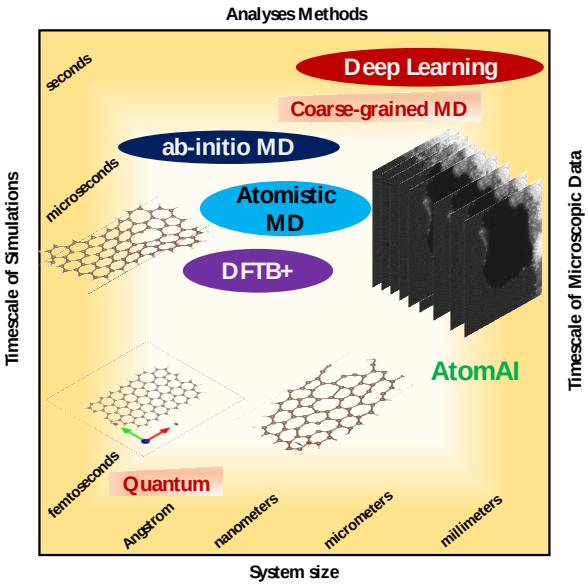

Figure 1. Primary disparities of time-length scales associated with dynamic integration of data from STEM, SPM measurements with simulations and deep learning models.

https://doi.org/10.1017/S1431927622011473 Published online by Cambridge University Press

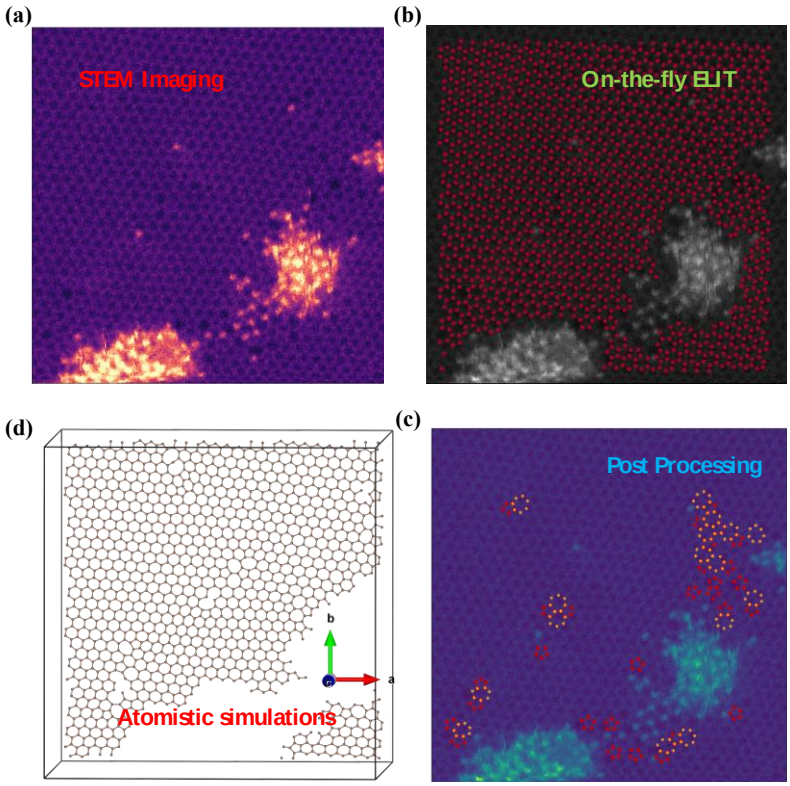

Figure 2. Workflow showing four primary steps of imaging, DL framework, post-processing of outputs and piping those to simulations environment.

References:

[1] S. J. Pennycook, Ultramicroscopy, 2017 **180**, 22-33.
[2] S. J. Pennycook, Ultramicroscopy, 2012 **123**, 28-37.
[3] C. Gerber and H.P. Lang, Nat. Nanotechnol. 2006 **1**, 3-5.
[4] C. Barth, C. et al., Adv. Mater. 2011 **23**, 477-501.
[5] D. A. Bonnell, and J. Garra, Rep. Prog. Phys. 2008 71 **4**, 044501.
[6] A. Ghosh, B.G. Sumpter., O. Dyck, S. V. Kalinin, M. Ziatdinov, npj Comput. Mater., 2021 7, 1.
[7] A. Ghosh, M. Ziatdinov, O. Dyck, B. G. Sumpter, S. V. Kalinin, arXiv preprint, 2021. arXiv:2109.04541.
[8] M. Ziatdinov, A. Ghosh, T. Wong, S.V. Kalinin, arXiv preprint, 2021. arXiv:2105.07485.
[9] This effort (machine learning) is based upon work supported by the U.S. Department of Energy (DOE), Office of Science, Office of Basic Energy Sciences Data, Artificial Intelligence Nanoscale Science Research (NSRC AI) Centers program (A.G., BGS, S.V.K.). The STEM experiments were supported by the DOE, Office of Science, Basic Energy Sciences (BES), Materials Sciences and Engineering Division (O.D.) and by the DOE, Office of Science, Basic Energy Sciences (BES), Materials Sciences and Engineering Division (K.M.R., S.V.K.). Work was performed and partially supported (M.Z) at Oak Ridge National Laboratory's Center for Nanophase Materials Sciences (CNMS), a DOE Office of Science User Facility. This manuscript has been authored by UT-Battelle, LLC, under contract DE-AC05-00OR22725 with the US Department of Energy (DOE).