doi:10.1017/S143192761700157X
Microsc. Microanal. 23 (Suppl 1), 2017

Leveraging First Principles Modeling and Machine Learning for Microscopy Data Inversion

Eric Schwenker,[1] Fatih Sen,[1] Spencer Hills,[1] Tadas Pualauskas,[2] Ce Sun,[3] Liang Li,[1] Alper Kinaci,[1] Kendra Letchworth-Weaver,[1] Moon Kim,[3] Robert Klie,[2] Jianguo Wen,[1] and Maria K. Y. Chan[1]

[1.] Center for Nanoscale Materials, Argonne National Laboratory, Lemont IL, USA.
[2.] Department of Physics, University of Illinois at Chicago, Chicago IL, USA.
[3.] Department of Materials Science and Engineering, University of Texas at Dallas, Dallas TX, USA.

In microscopy employing electron and x-ray beams, advances in instrumentation and techniques have substantially improved the ability to image, track, and characterize materials at ever-higher resolution and precision. However, determining atomistic arrangements (i.e. configurations) from microscopy data remains a substantial challenge. Whether due to projection of a three-dimensional structure onto one or two dimensions as in pair distribution functions (PDF) and (scanning) transmission electron microscopy (STEM/TEM), or the reduction of a large number of matrix elements into an overall energy-dependent amplitude as in x-ray absorption spectroscopy (XAS) and electron energy loss spectroscopy (EELS), the result is that inversion of this mapping is time consuming, imprecise, and sometimes even fruitless, despite a large number of excellent tools which produces characterization data from input atomistic configurations, or further perform fitting to produce configurations from characterization data. The reasons for such difficulties include the extremely high dimensional search space for atomistic configurations, especially for nanostructures and defected or disordered solids.

In order to tackle this challenge, we use first principles density functional theory (DFT) and DFT-based interatomic potential calculations to constrain the configurational space and guide microscopy data inversion. This approach is robust because (1) energetic information (and hence thermodynamic likelihood) is available for each atomistic configuration, (2) local structural optimization tools are widely available, and (3) configuration generation for equilibrium and non-equilibrium processes can be achieved through various deterministic and stochastic sampling approaches. A high-throughput framework is employed to enable the evaluation of a large number of configurations for energy and fit to microscopy data.

The final ingredient that enables an efficient and effective march towards the physical configurations consist of machine learning and optimization techniques including genetic algorithms, basin hopping, and computer vision. These methods either drive the sampling of configurations towards optimal solutions, or allow matching between measured and simulated microscopy data, e.g. for two-dimensional STEM/TEM images. Computer vision methods scale-invariant feature transform (SIFT) and histogram of oriented gradients (HOG) are used to select matches to the experimental image among hundreds or thousands of simulated images corresponding to different configurations. Figure 1 shows the selection of the correct match, among hundreds of simulated STEM images, of an experimentally-obtained image of a CdTe interface [3].

Taken together, we have used a combination of first principles-based modeling to evaluate a large number of different atomistic arrangements, and used the results in conjunction with machine learning to enable inversion of microscopy data for the determination of atomic-scale structural information. The

results enhance the power of microscopy and atomistic modeling, as well as accelerate materials understanding and design [4].

References:

[1] D. Lowe, Proceedings of the International Conference on Computer Vision **2** (1999), p. 1150.
[2] N. Dalal, Navneet and B. Triggs, IEEE Computer Society Conference on Computer Vision and Pattern Recognition **1** (2005), p. 886.
[3] T. Paulauskas *et al,* 43rd IEEE Photovoltaic Specialists Conference (PVSC) (2016).
[4] We acknowledge helpful discussions with Dane Morgan, Paul Voyles, Rebecca Willet, Simon Billinge, Amanda Petford-Long, and John Mitchell. Use of the Center of Nanoscale Materials, an Office of Science user facility, was supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under contract no. DE-AC02-06CH11357. This research used resources of the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. We gratefully acknowledge the computing resources provided on Blues, a high-performance computing cluster operated by the Laboratory Computing Resource Center at Argonne National Laboratory.

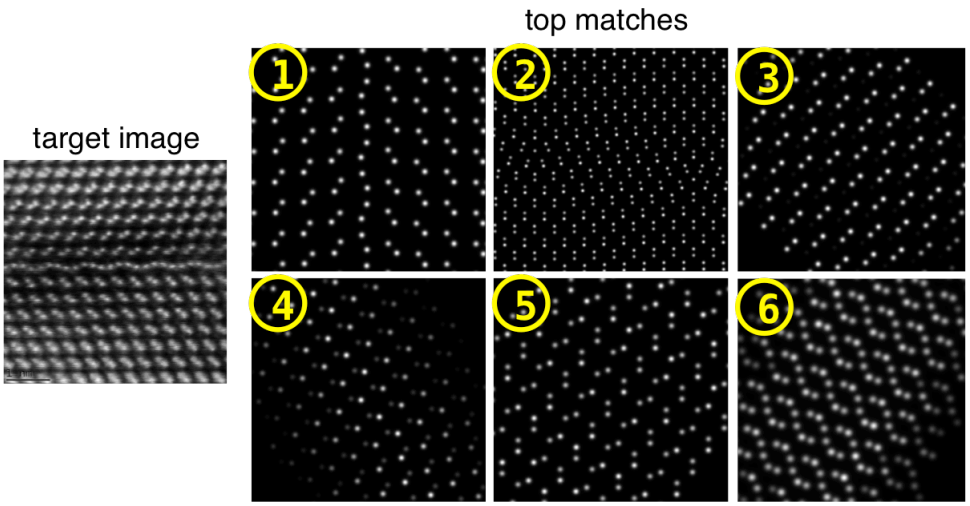

Figure 1. (Left) Target consists of STEM image measured from a grain boundary in CdTe. (Right) Using SIFT/HOG, a match among simulated images is obtained despite different overall orientation and scale, as well as noise in the experiment image.

https://doi.org/10.1017/S143192761700157X Published online by Cambridge University Press