Merging Deep Learning, Chemistry, and Diffraction for High-Throughput Material Structure Prediction

Matthew L. Gong[1,2], Brandon D. Miller[1], Ray R. Unocic[3], Khallid Hattar[4], Bryan Reed[5], Dan Masiel[5], Tolga Tasdizen[2] and Jeffery A. Aguiar[1]

[1.] Idaho National Laboratory, Nuclear Materials Department, Idaho Falls, Idaho USA.
[2.] University of Utah, Scientific Computing Imaging Institute, Department of Electrical and Computer Engineering, Salt Lake City, Utah USA.
[3.] Oak Ridge National Laboratory, Center for Nanophase Materials Science, Oak Ridge, Tennessee USA.
[4.] Sandia National Laboratories, Center for Integrated Nanotechnologies Albuquerque, New Mexico USA.
[5.] Integrated Dynamic Electron Solutions, Pleasanton, California USA.

Deep learning and augmented analysis have begun to disrupt the microscopy and microanalysis community with advancements made in material-specific models to solve narrow tasks. Developments in compressive imaging, where the concept of less is more in compressive modalities has begun to set the stage for high throughput and dose fractionation on our latest microscopes and characterization platforms. Crystallographic determination is crucial to many workflows within microscopy and materials research as a whole. At the core, crystallography is deeply rooted in pattern recognition and experts train for years to distinguish minute variations within the data.[1] Determining a crystal's space group often involves a lengthy process requiring fitting series of non-linear equations and intimate knowledge of a sample to be performed properly. This heavy dependence in complex pattern matching and time intensive process makes it an ideal case for automation with deep learning.

Building on recent work to classify material structure from diffraction data alone, the goal of this research is to boost these models' predictive capabilities and provide further disambiguation between high-level classifications by incorporating chemistry. Combining diffraction and EELS/EDS data sets that are collected simultaneously or separately, it is possible to improve the accuracy of these models trained solely on crystallographic data.[2,3] The additional information provided by chemistry data augments the model's understanding of higher-level structural classifications by drawing on the Open Crystallography Database, Materials Project Database, and experimental data, assembled in a robust training set that combines diffraction and chemistry.

In this presentation, we will discuss the additional benefits and challenges associated with learning on multi-modal datasets. In addition to extracting chemical composition from EELS and EDS data, it is necessary to create complex neural networks that utilize the multiple data streams through normalization and data-specific sub-networks. Sub-networks learn meaningful data-specific features before being concatenated and normalized with the other modules output before classification. Diffraction, chemistry ranked by composition and presence of elements were chosen to be learned on. To prevent overfitting and account for variations within experimental data and noise, additional functions were employed at training time in conjunction with randomly dropping out different channels. In this presentation, the basic approach to the workflow in **Figure 1** and analysis strategies creating neural networks that incorporate both chemistry and diffraction data will be presented. Results utilizing sub networks to better classify materials with minimal a priori knowledge in structure or chemistry will be presented and discussed. The potential

insights gained by increasing acquisition speed and/or decreasing the electron dose for future research projects will also be discussed [4].

References:

[1] R.K. Vasudevan et al., ACS Nano. **8** (2014). doi:10.1021/nn504730n.
[2] R.K. Vasudevan et al., Appl Phys Lett. **106** (2015). doi:10.1063/1.4914016.
[3] J Dongarra et al., Int J High Perform Comput Appl. **25** (2011). doi:10.1177/1094342010391989.
[4] Work supported through the INL Laboratory Directed Research& Development (LDRD) Program under DOE Idaho Operations Office Contract DE-AC07-05ID14517. This work was performed, in part, at the Center for Integrated Nanotechnologies, an Office of Science User Facility operated for the U.S. Department of Energy (DOE) Office of Science. Sandia National Laboratories is a multi-mission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. DOE's National Nuclear Security Administration under contract DE-NA-0003525. The views expressed in the article do not necessarily represent the views of the U.S. DOE or the United States Government. In part, this research was conducted at the Center for Nanophase Materials Sciences, which is a DOE Office of Science User Facility.
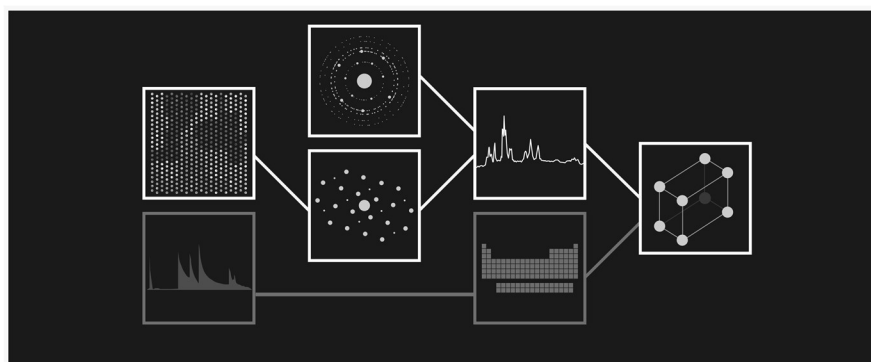
Figure 1. Robust Data Workflow for High Throughput Analysis. Illustrated above is a workflow for merging diffraction and chemistry data gathered from various modalities. Leveraging neural networks and machine learning, we have developed toolsets, and workflow models to benefit the community aiding their ability to perform high throughput analysis for several imaging modalities.