NSF DIBBs Award 1640899: The Local Spectroscopy Data Infrastructure

PI: Kristin A. Persson, Assistant Professor
UC Berkeley

The Local Spectroscopy Data Infrastructure (LSDI) Project embraces the concepts of the Materials Genome Initiative (MGI) and will establish a unique computational infrastructure for the production, curation, analysis, dissemination and sharing of X-Ray Absorption (XAS) and Nuclear Magnetic Resonance (NMR) spectroscopy data, and correlating such spectra with specific electronic structure, bonding and local atomic environments. The proposed effort will leverage the software infrastructure of the Materials Project to develop a completely integrated, one-stop platform for first principles XAS/NMR data, including experimental benchmarks and comparison tools.

The LSDI was launched October 1, 2016 and as such, is in its start-up phase. Nevertheless, significant progress has been made such that fully automated workflows for both NMR and XAS spectra have been designed including database import such that all computed information is retrievable via the database. The XAS spectra calculations are performed using the real (and reciprocal) space green function formalism as implemented in the widely used software FEFF. Each FEFF calculation depends on a considerable number of parameters that need to optimized to obtain the spectra that reproduces experimentally observed features. However, computing the spectra in high-throughput requires the parameters to be chosen and refined automatically. Hence, rigorous benchmarking on over 100 experimental spectra has provided us with a conservative set of FEFF parameters. Currently, we are in production on K-edge spectra, which so far indicates the lowest sensitivity to the parameter choice, and over 10,000 spectra for 1,500 compounds have been successfully computed. These are imported in the Materials Project database and will be displayed online in the coming year. For higher edge spectra, we are currently performing systematic convergence studies of various FEFF parameters for reliability. We are also in the process of purchasing a new cluster dedicated to this project, which will accelerate productivity.

The NMR workflow poses different challenges such that it currently suffers from low success rate; about 50%. However, the failures are hardware and highly platform dependent. Hence, we are developing i) a multi-platform strategy, ii) different parameter sets for specific failures and iii) correlations between high accuracy settings and specific failure modes.

One of the major future components of the project is the automated matching of experimental spectra with that of the computed ones in order to accelerate materials discovery. For XAS, a major hurdle in that process is the issue of merging the near-edge and the extended edge spectra obtained from FEFF. Near edge and extended spectra quantify different aspects of the structure and the FEFF software employed for the spectra calculation uses different levels of the theory to compute the spectra. This poses a major challenge and we are looking into various spectra filtering techniques to seamlessly merge the two spectra. Once we can produce reliable spectra the next step would be to use statistical learning techniques to match the experimental spectra with the spectra in the database and provide rapid characterization-on-the-fly for the extended materials science community.

Finally, the platform MPContribs will allow us to work with experimental partners to import experimental spectra into the Materials Project for reference, comparison and future datamining efforts. MPContribs has been soft-launched and XAS spectra from the Advanced Light Source at the Lawrence Berkeley National Laboratory are in the progress of being imported.