ELECTRON DENSITY-BASED MACHINE LEARNING FOR ACCELERATING QUANTUM CALCULATIONS

Joshua Lansford, University of Delaware
2019–2020 Graduate Fellow

EXECUTIVE SUMMARY

This project aims to develop a machine learning (ML) algorithm that can be used to generate catalytic reaction mechanisms and kinetic models at reduced computational cost. The ML-based model will be trained to electron density to accelerate the three most computationally intensive components of the reaction energy profile: transition states, global minima, and entropy. The work will for the first time combine structural and density data to enhance ML convergence. The representation of the molecular systems will be invariant to rotations, translations, and reordering of atoms. Because it will use distances, partial charges, and bond orders, it should be generalizable to any size system. All generated surface data will be made available to the public upon publication of the work. On Blue Waters, the PI has computed electron density for local minima of carbon monoxide platinum nanoparticles; these data will also serve as the input for the data generation necessary for entropy predictions.

RESEARCH CHALLENGE

A vast array of materials properties can be computed using quantum chemical calculations that cannot be identified with even the best experiments available today. Machine learning (ML) algorithms can emulate these calculations at speeds that are orders of magnitude faster. For bulk materials properties, training ML algorithms to either structural or one-dimensional electron density data can result in very accurate predictions [1]. However, such techniques are currently lacking for much more demanding calculations involving catalytic reaction mechanisms and predictions of better materials.

Construction of a reaction mechanism and building a kinetic model require mapping of the reaction coordinate along the minimum energy path of the potential energy surface. Identifying global minima and their transition states on this path is expensive because of the large number of states that must be sampled and calculated. Vibrational contributions to entropy require separate energy calculations for every degree of freedom of every atom at a high convergence [2].

METHODS & CODES

The electronic density distribution completely specifies the energy of a chemical system's state and can be calculated using Density Functional Theory (DFT) based on the Kohn–Sham equation [3]. With ML, descriptors of input data are mapped to desired quantitative output, often using complex nonlinear fitting or distances between input data and training data [4]. ML algorithms based on atomic positions have already sped up energy and transition state calculations by an order of magnitude [5–7]. Without an infinitely large training set, however, relying only on atomic positions does not allow extrapolation. Combining geometric and electronic density information will help alleviate this issue and is an innovation of this project.

The method for accelerating quantum-based calculations used in this work includes electronic density as input for ML and should significantly accelerate transition state, energy minima, and entropic calculations. This representation will also enable the extrapolation of the algorithms to systems that include atom types not found in the original training set. This method will combine structural and three-dimensional electron density data to accelerate calculations of molecular reactions on catalyst surfaces. Partial charges, atomic dipoles, and possibly effective bond orders can represent the electron density, as partial charges represent the electron density localized to specific atoms, atomic dipoles represent the asymmetry of the charge, and bond orders represent the density in the space between atoms. This representation of electron density is easily combined with geometric information and can be made invariant to translations, rotations, and reindexing of atoms.

The PI uses the Vienna Ab initio Simulation Package (VASP) [8] for generating electron densities of the ground state and CHARGEMOL [9] for integrating those densities. In addition, he uses his own code for calculating frequencies and spectral intensities used in the physics- and data-based surrogate models.

RESULTS & IMPACT

This project has generated thousands of DFT calculations already and the PI has implemented parallel versions of VASP and CHARGEMOL on Blue Waters, which are generating data. These calculations serve as the basis of the modeling. Previous work, which the PI has submitted for publication, provides evidence that the calculated electronic densities are accurate and can be used in surrogate modeling. That work used the electron density and physics-based surrogate models to generate complex infrared spectra that were then used to learn the mapping from spectra to local structure via multinomial regression. The PI plans to implement this electron density-based machine learning algorithm to speed up the calculations of spectra for this model and to aid automatic structure generation. Further plans are to learn the forces and electron asymmetry directly to compute vibrational frequencies and spectral intensities. This approach should also drastically reduce the number of calculations needed to relax a structure to its local optimum. A training set of local optima will be used to identify global optima and identify transition states. Upon completion of this work, a database of all DFT calculations will be freely available to research groups studying surface science and catalysis.

WHY BLUE WATERS

Blue Waters provides the necessary computational power to generate massive data. In the first week of the fellowship, the PI had already used several thousand node hours. The highly parallel architecture is necessary to train neural networks on large amounts of data with many features.

Joshua Lansford is in the fourth year of a Ph.D. program in chemical engineering at the University of Delaware and expects to graduate in 2020. His advisor is Dionisios G. Vlachos.