May 2022

Breakthrough accuracy in semi-empirical quantum chemistry via deep learning

Semi-empirical electronic Hamiltonians have nearly 30 years of successful history modeling chemistry processes such as photon absorption and chemical reactivity. However, they have always suffered from limited accuracy and generalizability due to the use of static parameters fit to limited datasets. New research from Los Alamos National Laboratory uses a Machine Learning model to dynamically reparametrize Semi-empirical Hamiltonians. Allowing physically interpretable model parameters to be adjusted according to the local chemical environment both significantly increases the model accuracy and enables new ways of interpreting results produced by ML models. This new paradigm of physics informed machine learning seeks to revolutionize the way machine learning models are applied to physics problems.

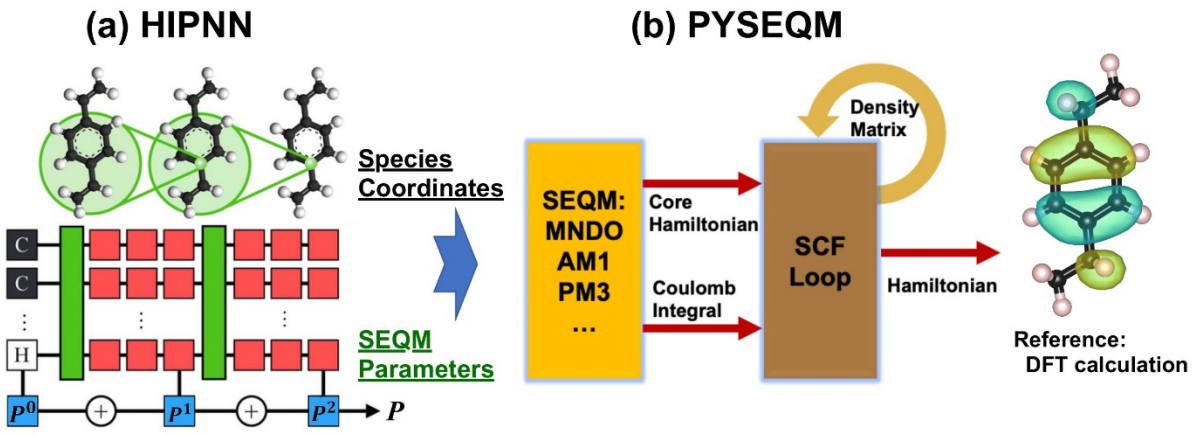

Figure 1: Diagram of the HIPNN+SEQM model. A neural network, HIPNN (a), generates Hamiltonian parameters based on a molecular configuration. These parameters are used in a semi-empirical Hamiltonian (b) to solve the Schrödinger equation and predict atomic properties.

The Science

The full workflow of HIPNN+SEQM is illustrated in Fig. 1. The HIPNN architecture is a message passing neural network for use on atomistic systems. HIPNN takes the molecular configuration as input, where each molecule is represented as a set of atom types and positions. The input features are passed through on-site layers (red blocks in Fig. 1), which are applied to the local features for each individual atom. Then, this information is shared through continuous message-passing layers (green blocks in Fig. 1) which pass information between nearby atoms and allow atoms to see their chemical environments. An inference layer is applied to the output from each of the last on-site layers to obtain zero- to higher order corrections of PM3 Hamiltonian parameters.

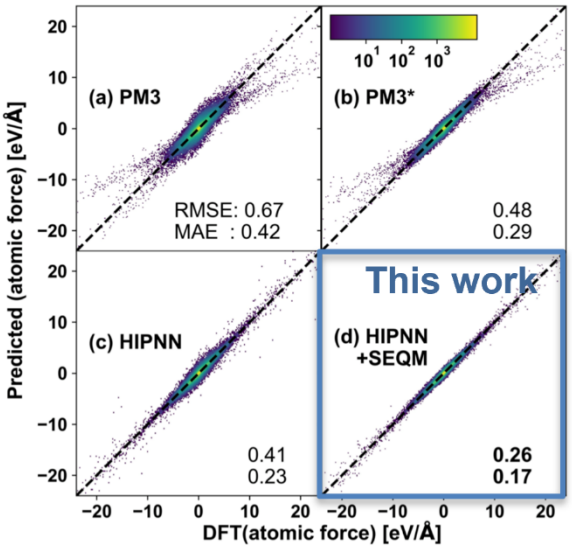

Figure 2: The combined HIPNN+SEQM model (d) makes significantly more accurate predictions than either the pure NN (c) or SEQM (a,b) models.

The Hamiltonian parameters are fed into the SEQM module which uses the Self-Consistent Field procedure to solve the motion of electrons in the current molecular configuration. The solution of this problem provides full quantum information including wavefunctions, charge distributions, molecular energy and atomic forces. The properties predicted by the combined HIPNN+SEQM model generally outperform both traditional SEQM and NN results (Fig. 2). This new type of machine learning model will facilitate the accurate solution of complex electronic structure problems, such as excited state and open shell systems.

Impact

This work represents a significant development in both machine learning models and semi-empirical quantum mechanics. While most machine learning models are 'black box' methods, the presented HIPNN+SEQM model provides additional physical insight through the parameters predicted by the NN (Fig. 3). Interestingly, the NN assigns different values to atomic Hamiltonian parameters based on traditional views of atomic bonding. Fig. 3 shows that if an atom forms a triple bond with a neighbor, it has a significantly different p-orbital energy parameter than if it forms a double or only single bonds.

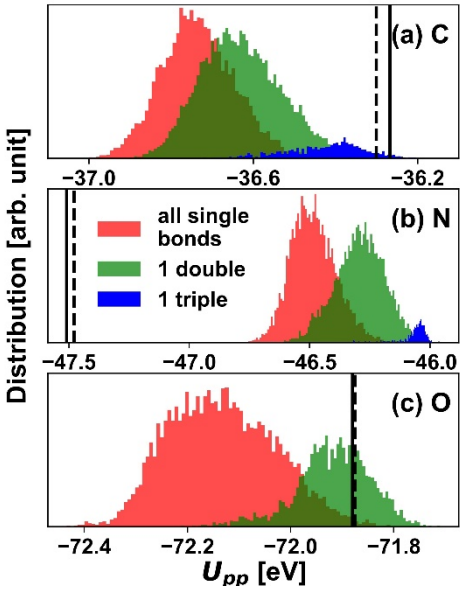

Figure 3: The parameters predicted by HIPNN obey different distributions according to traditional views of chemical bonding.

This new class of machine learning enhanced semi-empirical quantum mechanics models will enable increasingly accurate theoretical studies of chemical reactivity, photo-absorption, excited state dynamics, polaritonics and other related phenomena.

These applications are critical to the LANL mission, either through the simulation of reactive molecules or simulating future organic photovoltaic systems critical to national energy security.

Research Details

Both the HIPNN and SEQM codes used to construct these models have been developed at LANL and openly released on github (see related links). HIPNN is a message passing neural network built capable of simulating energies, forces, atomic charges, bonding patterns, and now Hamiltonian parameters. PYSEQM is a semi-empirical quantum chemistry package built in Pytorch. Using back propagation, it is possible to optimize a neural network which parameterizes the Hamiltonian. Further, PYSEQM is capable of performing calculations on GPU devices and operating in batches. These are critical features for the machine learning tasks performed in this paper, but also efficient, large scale molecular excited state simulations.

Contact

Benjamin Nebgen bnebgen@lanl.gov

Kipton Barros kbarros@lanl.gov

Sergei Tretiak serg@lanl.gov

Funding

This work was funded by the Los Alamos National Laboratory (LANL) Directed Research and Development project 20210087DR the U.S. DOE, Office of Science, Basic Energy Sciences, Chemical Sciences, Geosciences, and Biosciences Division under Triad National Security, LLC ("Triad") contract grant # 89233218CNA00000. LA-UR-21-30735

Publications

Deep Learning of Dynamically Responsive Chemical Hamiltonians with Semi-Empirical Quantum Mechanics, *PNAS*, 2022 *In Press*.
Guoqing Zhou, Nicholas Lubbers, Kipton Barros, Sergei Tretiak, and Benjamin Nebgen

Hierarchical modeling of molecular energies using a deep neural network, *J. Chem. Phys.* **148**, 421715 (2018).
Nicholas Lubbers, Justin S. Smith and Kipton Barros.

Graphics Processing Unit-Accelerated Semiempirical Born Oppenheimer Molecular Dynamics Using PyTorch, *J. Chem. Theo. Comp.* **16**, 4951-4962 (2020).
Guoqing Zhou, Ben Nebgen, Nicholas Lubbers, Walter Malone, Anders M. N. Niklasson, and Sergei Tretiak

Related Links

HIPNN: www.github.com/lanl/HIPPYNN
PYSEM: www.github.com/lanl/PYSEQM