Title: Development of a Fourth Generation Machine Learned Potential with Long-Range Flexible Charges for Two-Dimensional CsPbI Perovskites

Authors: Jos Thomas, Anders M. N. Niklasson, Robert Stanton

Affiliations: Texas A&M Department of Materials Science and Engineering, Los Alamos National Laboratory

Acknowledgements: This work was supported and funded by the Texas A&M O-REU Program. We acknowledge Los Alamos National Laboratory (LANL) as a whole for providing research facilities and computational resources. In particular, I thank Robert Stanton and Anders Niklasson for mentorship and guidance.

Abstract

Two-dimensional perovskites (2DPKs) have emerged as a compelling path forward for perovskites, demonstrating immense potential in applications across catalysis, photovoltaics, nanotechnology, superconductivity, and beyond. However, nearly all methods for their large-scale dynamic simulation, including traditional force fields and neural-network potentials, have been parameterized exclusively for three-dimensional perovskites. This leaves 2DPKs without an efficient and computationally inexpensive simulation framework to assess their diverse physicochemical properties. Moreover, conventional machine-learning interatomic potentials (MLIPs) are generally short-ranged, truncating atomic interactions beyond a fixed cutoff. This neglects critical long-range Coulomb forces that dominate in ionic, defect-rich systems such as those often found in Pb-based 2DPKs. In these materials, strongly localized charge regions develop around vacancies, defects, or under strain, creating highly inhomogeneous electrostatic environments. Ignoring such interactions significantly reduces accuracy in predicting defect formation energies and charge-transport properties. Therefore, developing a computationally inexpensive, fourth-generation MLIP framework that explicitly incorporates long-range electrostatics is essential to accurately simulate 2DPKs and accelerate high-quality materials discovery.

This project pursues three interrelated goals. First, it will generate a flexible machine-learning model that determines each atom's electronegativity from its local chemical environment, enabling the real-time determination of atomic partial charges. Second, it will produce a dataset of 2D-CsPbI perovskite configurations that adequately sample the mid- to low-temperature regime. Third, using the forces and energies obtained in step two, it will create a charge-neutral MLIP for two-dimensional CsPbI perovskite systems by combining learned electronegativities with the short-range energy and force model.

To learn per-atom electronegativities, we will employ HIP-NN (Hierarchically Interacting Particle Neural Network), an in-house deep neural network architecture developed at LANL. Predicted electronegativities will feed into the charge-equilibration (QEQ) model that obtains partial charges for every atom in a given configuration. Simultaneously, a second HIP-NN network will

be trained on total energies and atomic forces from semi-empirical (SE) quantum mechanical (QM) reference data. To generate representative atomic configurations, atomistic simulation toolkits such as the atomic simulation environment (ASE) and Pymatgen will be employed, in combination with the extended tight-binding SEQM framework. From these xTB-sampled energies and forces, we will train the short-ranged, charge-neutral portion of the MLIP. A pre-existing dataset of atomic partial charges from density functional theory (DFT) calculations will be used for training the electronegativity-based neural network, with the option of generating additional partial charges as reference data if required.

All data preprocessing, feature extraction, and model training will be conducted in Python using NumPy and pandas, while plotting and visualization of results, will employ Matplotlib and Seaborn. Final validation will consist of comparing the MLIP+QEQ predictions of energies, forces, and partial charges against xTB and DFT reference data to confirm that the combined short-range/long-range framework quantitatively reproduces key properties obtained from using methods at a higher level of theory.

References

[1] J.-C. Blancon, J. Even, C. C. Stoumpos, M. G. Kanatzidis, and A. D. Mohite, "Semiconductor physics of organic–inorganic 2D halide perovskites," *Nat. Nanotechnol.*, 15:969–985 (2020).

[2] R. Stanton and D. J. Trivedi, "Pyrovskite: A software package for the high-throughput construction, analysis, and featurization of two- and three-dimensional perovskite systems," *J. Chem. Phys.*, 159:064803 (2023).

[3] R. Stanton, W. Nie, S. Tretiak, and D. J. Trivedi, "Data mining and computational screening of Rashba–Dresselhaus splitting and optoelectronic properties in two-dimensional perovskite materials," *J. Chem. Phys.* (2025).