# Interpretable Force Fields for On-the-Fly Machine Learning

Jonathan Vandermause,[1,2] Steven B. Torrisi,[1] Simon Batzner,[2,3] and Boris Kozinsky[2]

[1] *Department of Physics, Harvard University, Cambridge, MA 02138, USA*
[2] *John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA 02138, USA*
[3] *Center for Computational Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, USA*

(Dated: March 27, 2019)

Machine learning provides a path toward fast, accurate, and large-scale materials simulation, promising to combine the accuracy of *ab initio* methods with the computational efficiency of classical potentials. However, training current state-of-the-art models often requires databases of first principles calculations containing thousands of structures. We present an on-the-fly Bayesian inference scheme for automating and accelerating the construction of interatomic force fields in a single molecular dynamics simulation. Gaussian process regression is coupled to a first principles DFT code to learn two- and three-body force fields on-the-fly with minimal training data. The resulting force field is easily extended to structures outside the training set and compares favorably to state-of-the-art classical and machine learned potentials.

Recent machine learning (ML) approaches to modelling the Born-Oppenheimer potential energy surface have been shown to approach first principles accuracy in several molecular and extended systems. However, most of these methods return only point estimates of the quantities of interest (typically energies, forces, and stress) rather than a predictive distribution that reflects model uncertainty. Without knowledge of the highest uncertainty training points, a laborious fitting procedure is required, in which thousands of reference structures are selected *ad hoc* from a database of first principles calculations. At test time, lack of predictive uncertainty makes it difficult to determine when the fitted model is out-of-sample, leading to unreliable results and making the model difficult to update in the presence of new data.

Here, we show that on-the-fly Bayesian inference can be used to both accelerate the training of a high-quality machine learned force field and flexibly adapt the model to out-of-sample structures. By coupling Gaussian process regression and density functional theory in a single molecular dynamics engine, it is shown that a high quality potential can be obtained with only a few dozen first principle calculations. Such a reduction in the computational cost of training a high-quality potential promises to extend ML modelling to a wider class of materials than has been possible to date.

To reason about model uncertainty, we construct a Gaussian process model trained directly on *ab initio* forces. The training database of the Gaussian process is populated with individual atomic environments by expressing the total energy of the system as a sum over two- and three-body terms,

$$E = \sum_{ij} \varepsilon_{ij} + \sum_{ijk} \varepsilon_{ijk}, \tag{1}$$

where the sums range over all unique pairs and triplets of atoms containing at least one atom from the unfolded primary cell. In practice, the sums are truncated by considering local atom-centered environments surrounding each atom in the primary cell and neglecting contributions from atoms beyond a chosen cutoff distance from the central atom. The covariance between bond and triplet energies is defined by a kernel that directly compares interatomic distances, from which a fully covariant and energy conserving force kernel follows immediately [1, 2].

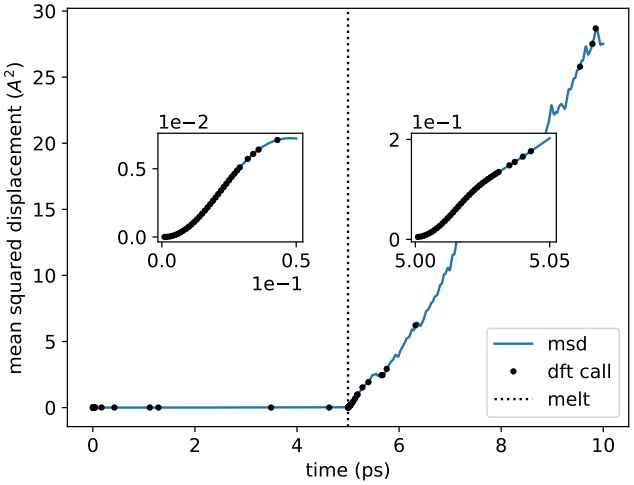


FIG. 1. On-the-fly learning of an aluminum force field at multiple temperatures.

[1] A. Glielmo, P. Sollich, and A. De Vita, Physical Review B **95**, 214302 (2017).
[2] A. Glielmo, C. Zeni, and A. De Vita, Physical Review B **97**, 184307 (2018).

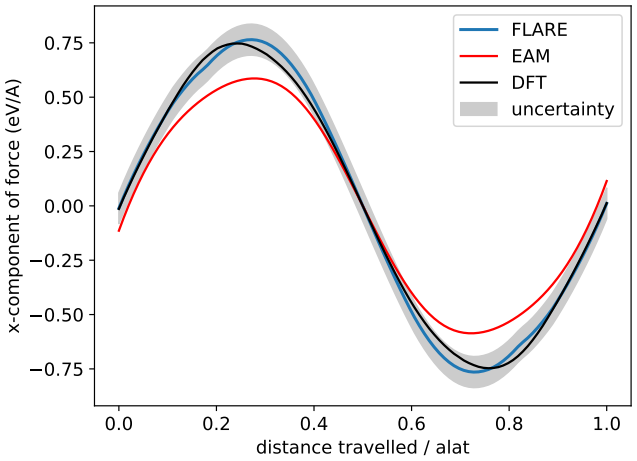


FIG. 2. On-the-fly learning of an aluminum force field at multiple temperatures.

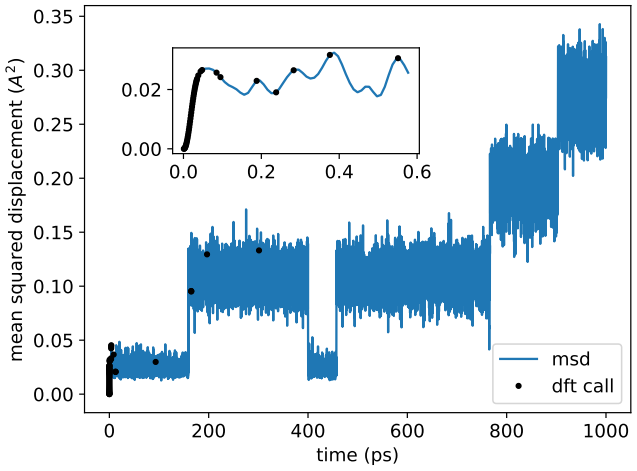


FIG. 3. On-the-fly learning of an aluminum force field at multiple temperatures.