

Accelerating *Ab Initio* Molecular Dynamics with On-the-Fly Machine Learning

Jonathan Vandermause,^{1,2} Steven B. Torrisi,² Simon Batzner,³ and Boris Kozinsky¹

¹*John A. Paulson School of Engineering and Applied Sciences,
Harvard University, Cambridge, Massachusetts 02138, USA*

²*Department of Physics, Harvard University, Cambridge, Massachusetts 02138, USA*

³*Center for Computational Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*

Ab initio molecular dynamics is a powerful tool for accurately probing the dynamics of molecules and solids, but it is limited to system sizes on the order of 1000 atoms and time scales on the order of 10 ps. We present a scheme for rapidly training a machine learning model of the interatomic force field that approaches the accuracy of *ab initio* force calculations but can be applied to larger systems over longer time scales. Gaussian Process models are trained on-the-fly, with density-functional theory calculations of the atomic forces performed whenever the model encounters chemical configurations outside of the training set. We demonstrate the flexibility of our approach by testing it on a range of single- and multi-component systems.

I. INTRODUCTION

Key literature:

1. On-the-fly force fields with GPs [1–3]
2. Gaussian Approximation Potentials (GAP) [4–6]
3. The SOAP kernel [7]
4. Other covariant kernels (two- and three-body) [5, 6, 8, 9]
5. Vector-valued GPs and ICM [10]
6. DFT and Quantum Espresso [11, 12]

II. COVARIANT KERNELS

This section will discuss our approach to modelling multi-species systems.

1. Review our two- and three-body single-component kernels, which are similar to the ones used in [6, 9]. Our three-body kernel is slightly different from both: [9] does not distinguish between cross-bonds and central-bonds, and [6] combines the central-bond lengths in a way that doesn’t easily generalize to multi-component systems.

We should also highlight our approach to selecting the GP hyperparameters, which hasn’t been discussed at length in the GAP literature. Our approach maximizes the marginal likelihood, and is therefore data-driven rather than empirical. Optimizing the hyperparameters in this way can reveal interesting physics. For example, in the two-plus-three body kernel,

$$k(\rho_1, \rho_2) = \delta_2 k_2(\rho_1, \rho_2) + \delta_3 k_3(\rho_1, \rho_2), \quad (1)$$

the size of the hyperparameters δ_2 and δ_3 indicates the relative importance of two- and three-body interactions in the system.

There is one “special” hyperparameter, the cutoff radius, which can’t be easily optimized using gradient descent. We should justify our cutoff choice using the “freeze-and-perturb” method of [6]. This will also give a lower bound on the error of any local ML model, which is a useful benchmark. It might be interesting to compare this lower bound with the noise variance of the model, which is an internal measure of how well the model is able to capture the data.

It would be nice to benchmark the performance of the two-body, three-body, and two-plus-three body single-component kernels by showing the model performance as a function of training set size for a representative system (e.g. solid-state C at 1000 K). Since the focus of this section is on characterizing and benchmarking the kernel, we can train on a constant-T AIMD trajectory.

2. Discuss our generalization to multi-component systems. There are three plausible ways of going about this: comparing every bond to every other bond (no distinction between different species, which might make sense for similar atoms); comparing bonds of the same type only (which wastes information by neglecting similarities between different bond types); and weighting each bond pair with a hyperparameter. The third method, which is equivalent to an ICM model, provides a general way to capture covariance between different outputs that should be more accurate than the first approach and more efficient than the second.

We can plot model performance versus training set size on a SiC system for each of the three methods mentioned above and for each of the three kernels (two-body, three-body, and two-plus-three body). The two-plus-three body kernel will have a lot of hyperparameters (on the order of 100), but it should be feasible with our gradient descent scheme.

III. ON-THE-FLY LEARNING

We’ve so far come up with three interesting applications that highlight the utility of our on-the-fly program: large systems outside the reach of AIMD, bond formation and bond breaking in molecules, and diffusion in solids. In all examples that we try, it will be important to use the converged force field to predict a physical property (e.g. phonon dispersion, bond length, activation energy) and compare the prediction against the “exact” value computed with DFT.

1. Large systems. As a test case, we can consider a simple solid-state example, e.g. a 5000-atom bulk SiC sample, and use the “punchout” method to train an initial force field. In this approach, an atom is first selected at random, a small cubic cell containing the atom and 20-30 of its neighbors is sent to Quantum Espresso, and the resulting force data is used to build a GP model. The model then loops through the remaining atoms and makes predictions, augmenting the database whenever the uncertainty threshold is reached. Once the model

converges, it can be tested by predicting the phonon spectrum of SiC and comparing against DFPT.

2. Bond formation and bond breaking. The simplest example is on-the-fly modelling of the bond energy of H_2 . We might also try more complicated examples:
 - Starting a simulation with separated C and O atoms and allowing bonds to form, comparing the model’s predicted bond lengths against bond lengths computed with DFT.
 - Taking a perturbed benzene molecule and seeing if the model correctly predicts the equilibrium configuration.
3. Diffusion in solids. Gold diffusing on a silver slab or self-diffusion in bulk aluminum are worth trying. Following [2, 3], once the model has converged, a good test is to predict the potential energy along the migration trajectory by integrating the force field. This energy activation profile can be compared against the corresponding DFT prediction.

-
- | | |
|--|---|
| <p>[1] Z. Li, J. R. Kermode, and A. De Vita, Physical review letters 114, 096405 (2015).</p> <p>[2] V. Botu and R. Ramprasad, International Journal of Quantum Chemistry 115, 1074 (2015).</p> <p>[3] V. Botu and R. Ramprasad, Physical Review B 92, 094306 (2015).</p> <p>[4] A. P. Bartók, M. C. Payne, R. Kondor, and G. Csányi, Physical review letters 104, 136403 (2010).</p> <p>[5] A. P. Bartók and G. Csányi, International Journal of Quantum Chemistry 115, 1051 (2015).</p> <p>[6] V. L. Deringer and G. Csányi, Physical Review B 95, 094203 (2017).</p> | <p>[7] A. P. Bartók, R. Kondor, and G. Csányi, Physical Review B 87, 184115 (2013).</p> <p>[8] A. Glielmo, P. Sollich, and A. De Vita, Physical Review B 95, 214302 (2017).</p> <p>[9] A. Glielmo, C. Zeni, and A. De Vita, Physical Review B 97, 184307 (2018).</p> <p>[10] M. A. Alvarez, L. Rosasco, N. D. Lawrence, <i>et al.</i>, Foundations and Trends® in Machine Learning 4, 195 (2012).</p> <p>[11] W. Kohn, Reviews of Modern Physics 71, 1253 (1999).</p> <p>[12] P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. L. Chiarotti, M. Cococcioni, I. Dabo, <i>et al.</i>, Journal of physics: Condensed matter 21, 395502 (2009).</p> |
|--|---|