Benchmarking of Fast and Interpretable UF Machine Learning Potentials

Pawan Prakash
Department of Physics
University of Florida
Gainesville, FL 32611
pawanprakash@ufl.edu

Richard Hennig
Department of Material Science and Engineering
University of Florida
Gainesville, FL 32611
rhennig@ufl.edu

Abstract

Ab initio methods offer great promise for materials design, but they come with a hefty computational cost. Recent advances with machine learning interatomic potentials (MLIPs) have revolutionized molecular dynamic simulations by providing high accuracies similar to ab initio models but at much reduced computational cost. Our study evaluates the ultra-fast force fields (UF3) potential, employing linear regression with cubic B-spline basis for assessing effective two- and three-body potentials. On benchmarking, UF3 displays comparable precision to established models like GAP, MTP, NNP (Behler Parrinello), and qSNAP MLIPs, yet is significantly faster by two to three orders of magnitude. A distinct feature of UF3 is its capability to render visual representations of learned two- and three-body potentials, shedding light on potential gaps in the learning model. In refining UF3's performance, a comprehensive sweep of the hyperparameter space was undertaken. While our current optimizations are concentrated on energies and forces, we are primed to broaden UF3's evaluation spectrum, focusing on its applicability in critical areas of molecular dynamics simulations. The outcome of these investigations will not only enhance the predictability and usability of UF3 but also pave the way for its broader applications in advanced materials discovery and simulations.

1 Introduction

Molecular dynamics (MD) simulations play a pivotal role in understanding and predicting the behavior of materials at atomic and molecular scales [1]. Density functional theory (DFT) [2, 3], a quantum mechanical modeling method, offers detailed insights and precise predictions about these systems. However, the high computational costs and $O(N) - O(N^3)$ scaling associated with DFT make it challenging for use in extensive MD simulations spanning long durations [4]. Classical interatomic potentials [5, 6] offer computational efficiency but lack DFT's accuracy [7, 8]. The advent of machine learning interatomic potentials (MLIPs) has provided an effective alternative, combining the efficiency of classical potentials with near-DFT accuracy in energy, force, and property predictions [9, 10, 11, 12]. However, a challenge arises with many state-of-the-art MLIPs: while more efficient than DFT they can still be computationally too demanding. This heightened computational requirement complicates MD simulations on expansive systems over extended durations and also presents hurdles in model interpretation.

In the landscape of machine learning potentials, several have emerged as prominent contenders in accurately modeling atomic systems. These include the neural network potentials (NNP) [9], Gaussian approximation potential (GAP) [10], moment tensor potential (MTP) [11], spectral neighbor analysis potentials (SNAP) [12], and its variant, the quadratic SNAP (qSNAP) [13]. Central to the efficacy of these potentials is the concept of 'fingerprints' or 'descriptors'. These are intricate

Machine Learning and the Physical Sciences Workshop, NeurIPS 2023.

representations of atomic environments with flexible functional form, which are subsequently passed through an embedding function to yield the desired predictions. For readers desiring a comprehensive understanding of each method's details, the provided references offer an in-depth exploration.

This study builds on previous work comparing machine learning potentials in energy and force prediction accuracy using a consistent dataset of elements [14]. A primary focus of this study is the ultra-fast force field machine learning interatomic potential (UF3 - MLIP) [15]. We embark on a detailed juxtaposition of UF3 against prevailing state-of-the-art potentials, particularly in energy and force predictions.

2 Ultra-Fast Machine Learning Potential

The UF3-MLIP is formulated as a function of atomic positions R_i and species σ_i. It builds on the many-body expansion of the energy of atomic systems. By truncating its expansion at the three-body term, UF3 strikes an optimal compromise between computational efficiency and predictive accuracy [15]:

$$E = \sum_{i,j} V_2(r_{ij}) + \sum_{i,j,k} V_3(r_{ij}, r_{ik}, r_{jk}) \tag{1}$$

where indices i, j, k of the summation run over all atoms, $V_2(r_{ij})$ is the two-body contribution and $V_3(r_{ij}, r_{ik}, r_{jk})$ is the three-body contribution.

Using a spline basis with compact support, UF3 represents effective two- and three-body interactions. It employs cubic B-splines and tensor product splines for its smooth and efficient representation:

$$V_2(r_{ij}) = \sum_{n=0}^{K} c_n B_n(r_{ij}) \tag{2}$$

$$V_3(r_{ij}, r_{ik}, r_{jk}) = \sum_{l=0}^{K_l} \sum_{m=0}^{K_m} \sum_{n=0}^{K_n} c_{lmn} B_l(r_{ij}) B_m(r_{ik}) B_n(r_{jk}) \tag{3}$$

UF3's training uses regularized linear regression to optimizes the coefficients c_n and c_{lmn} of the two-and three-body such that energies and forces are fit to data from quantum mechanical data. The loss function incorporates Tikhonov regularization. The ridge regularizer λ_1 imposes a penalty on large coefficients to prevent overfitting and mitigates coefficient fluctuation in response to change in data. Meanwhile, the curvature regularizer λ_2 ensures the smoothness of the learned potential. The loss function for two-body is

$$\begin{aligned} L = &\frac{\kappa}{\sigma_{\mathcal{E}}^2 |\mathcal{E}|} \sum_{s \in S} (E(s) - \mathcal{E}_s)^2 + \frac{1-\kappa}{\sigma_{\mathcal{F}}^2 |\mathcal{F}|} \sum_{s \in S} (-\nabla E(s) - \mathcal{F}_s)^2 \\ &+ \lambda_1 \sum_{n}^{K} c_n^2 + \lambda_2 \sum_{n}^{K} (c_n - 2c_{n+1} + c_{n+2})^2 \end{aligned} \tag{4}$$

where $\kappa \in [0, 1]$ is the effective weight of energy/force residual, $E(s)$ is the predicted energy, and $|\mathcal{F}|$ and $|\mathcal{E}|$ denote the number of forces and energy observation in training set, respectively.

A standout attribute of the UF3 framework is its interpretability. Given its reliance on tangible physical parameters like inter-atomic distances, it paves the way for straightforward visualization of its model. This facilitates easy identification of any anomalies or unphysical behaviors in the learned patterns.

3 Methodology

The objective of this study is to compare the performance of UF3 on energy and force prediction by juxtaposing it with other state-of-the-art MLIPs.

While recent advancements in MLIPs are commendable, a meticulous and systematic evaluation, especially on a standardized dataset, remains largely elusive. It is in this context that the efforts of Ong et al. [14] stand out; they put forth a comprehensive dataset for MLIP comparison, which

encompasses a diverse range of atomic local environments associated with six elements: Ni, Cu, Li, Mo, Si, and Ge. Selected for their diverse crystal structures and chemistries, the dataset covers ground-states, strained structures, ab initio MD simulations of bulk supercells at elevated temperatures, and even structures with single vacancies. It's worth noting that this data, drawn from first-principle calculations executed via VASP, come pre-partitioned into training and testing sets.

The UF3 potential adopts a two-tiered hyperparameter system. The 'outer' hyperparameters in featurization guide both initialization and the learning of atomic environments, while the 'inner' hyperparameters of the loss function direct the least square optimization. Specifically, the outer hyperparameters set the maximum and minimum cutoff distances between atoms and determine the number of basis splines for each spline. In contrast, the inner hyperparameters address aspects like the effective weight of the energy/force residual in the loss function and the regularization parameters, namely the two- and three-body ridge and curvature regularizers. It's important to underline that adjustments to the outer hyperparameters necessitate the generation of new representation vectors, a process that is both time-intensive and computationally demanding, making it the most resource-intensive step in the UF3 framework. This distinction of hyperparameters into tiers is based on their computational complexity.

4 Results and Optimization

Ong et al. performed a detailed comparison of state-of-the-art MLIPs in energy and force prediction for elemental metals [14]. To ensure a consistent comparison, we optimized our outer hyperparameter in the vicinity of the parameters specified in Ong et al. Following the inner-hyperparameter optimization, the resultant energy and force errors are depicted in Fig. 1. These figures underscore that UF3 offers accuracy on par with GAP, MTP, and qSNAP MLIPs for face-centered cubic (FCC) crystals, all while significantly reducing computational cost [15]. Notably, in more open structures like diamond and body-centered cubic (BCC), all MLIPs, including UF3, showed diminished accuracy, with UF3 ranking either third or fourth among sophisticated models. Additionally, UF3 and other MLIPs outperformed spline-based empirical potentials like s-MEAN [16].

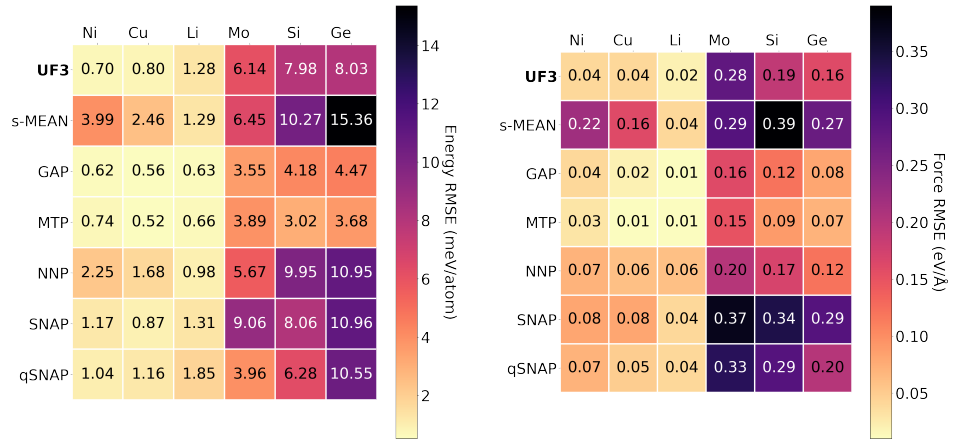


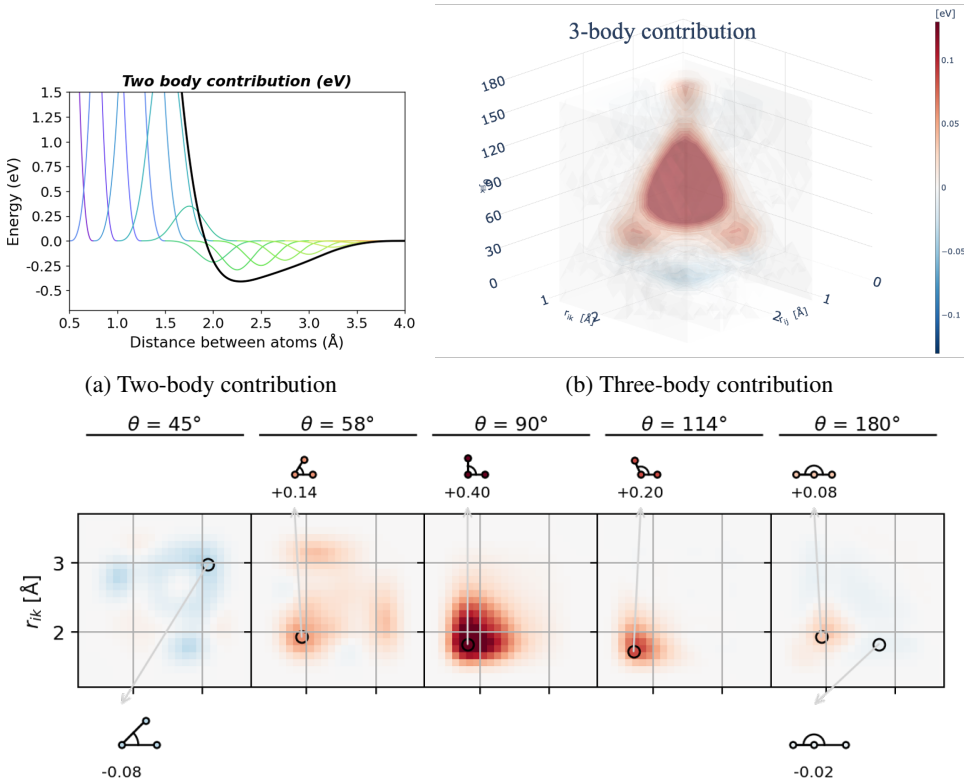
Figure 1: Energy errors (left) and Force errors (right) for UF3 compared to other MLIPs. UF3 exhibits comparable accuracy to GAP, MTP, and qSNAP for FCC crystals, highlighting its efficiency. However, its precision is slightly lower for more open structures like diamond and BCC as well as others, but it still ranks third or fourth in terms of accuracy.

We further demonstrate the utility of UF3 potentials through visual analysis, as exemplified in Fig. 2 for nickel. These visual representations, by dissecting contributions to the interatomic interactions, lay bare the intricacies of the model. Critical features such as minima, both repulsive and attractive contributions, and the presence of inflection points are discernible, granting a deeper understanding of nickel's chemical bonding characteristics. This visualization makes the intricacies of UF3 directly interpretable.

We evaluated property predictions using UF3 models optimized for energy and force, comparing them with DFT data and other advanced MLIPs. These comparisons, illustrated in spider plots (Fig. 3), show UF3's high predictive accuracy for properties like cubic elastic constants, bulk modulus, and lattice constants, particularly for Ni and Cu. Discrepancies notably arise in predicting properties for Li, where UF3 and other MLIPs diverge from DFT results, partly due to the small magnitude of Li's properties. This variation in performance across elements highlights the critical role of having diverse and representative training data. Although the dataset used was comprehensive, the effectiveness of UF3 and similar models is limited when training data lacks representation for specific properties. Thus, ensuring diverse and representative training data is paramount for consistent and optimal model performance.

5 Future Works

Our study revealed that optimizing the UF3-MLIP for energy and forces results in impressive performance, despite its straightforward physically-inspired structure. Notably, our optimization focused primarily on minimizing energy and force errors, which led to the employment of minimal regularization. However, we propose that a slight increase in curvature regularization may yield a smoother potential. This adjustment could marginally reduce the model's accuracy on energy and forces on the testing data, yet potentially enhance its application in property prediction and dynamic simulations.



(a) Two-body contribution

(b) Three-body contribution

(c) Sliced view of three-body contributions at certain angles.

Figure 2: Visualization of UF3 applied to Nickel, showcasing the mapping of energy and force data onto effective two- and three-body terms. Through this representation, various features — from minima to inflection points — elucidate Nickel's chemical bonding characteristics. This clear, direct depiction highlights the interpretability advantages of UF3 over many other machine learning potentials.

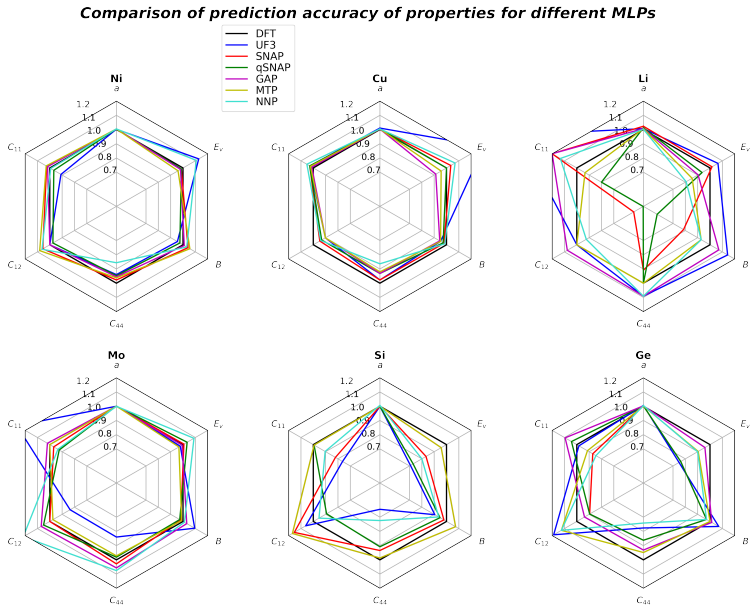


Figure 3: Spider plots comparing the predictive capabilities of various MLIPs with DFT benchmarks for different properties - cubic elastic constants C_{11}, C_{12}, C_{44}, bulk modulus (B), vacancy formation energy (E_v) and lattice constant (a).

It is also crucial to acknowledge that the UF3 potentials, as utilized in this study, were specifically tailored for energy and force predictions. Their effectiveness in predicting other properties was not considered during the optimization process. Moving forward, we aim to incorporate these relevant properties directly into the optimization loop. This approach is intended not only to enhance accuracy but also to yield more functionally optimal potentials. Emphasizing the practicality of MLIPs in dynamic simulations is essential, as the most accurate potential is not always the most applicable or physically relevant one [17, 18]. Our subsequent steps include evaluating UF3's performance in MD simulations, especially for targeted applications. Additionally, we emphasize the need for tailored human oversight in MLIP training. For specific uses like elasticity calculations, training data should predominantly consist of bulk structures, which is feasible given UF3's reduced computational needs and effective training with less data.

In conclusion, we aim to refine UF3 optimizations to balance accuracy with practical usability in dynamic simulations, ensuring the developed potentials are both precise and widely applicable.

Acknowledgments and Disclosure of Funding

We extend our sincere thanks to Stephen R. Xie, Ajinkya Hire, and Jason Gibson for their invaluable discussions and insights that significantly contributed to this research. I am also grateful to the University of Florida for providing computational resources through the HiPerGator computing facility. This work was supported in part by the National Science Foundation, Division of Materials Research under Award No. NSF-DMR-2118718 and by the Center for Molecular Magnetic Quantum Materials, an Energy Frontier Research Center funded by the U.S. Department of Energy, Office of Science, Basic Energy Sciences under Award No. DE-SC0019330.

References

[1] Aidan P. Thompson, H. Metin Aktulga, Richard Berger, Dan S. Bolintineanu, W. Michael Brown, Paul S. Crozier, Pieter J. in 't Veld, Axel Kohlmeyer, Stan G. Moore, Trung Dac Nguyen, Ray Shan, Mark J. Stevens, Julien Tranchida, Christian Trott, and Steven J. Plimpton. Lammps - a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales. *Computer Physics Communications*, 271:108171, 2022.

[2] W. Kohn and P. Hohenberg. Inhomogeneous electron gas. *Physical Review*, 136:B864–B871, 1964.

[3] W. Kohn and L.J. Sham. Self-consistent equations including exchange and correlation effects. *Physical Review*, 140:A1133–A1138, 1965.

[4] A. E. Mattsson, P. A. Schultz, M. P. Desjarlais, T. R. Mattsson, and K. Leung. Designing meaningful density functional theory calculations in material science–a primer. *Modelling Simul. Mater. Sci. Eng.*, 13:R1–R31, 2005.

[5] J. E. Jones. On the determination of molecular fields. —i. from the variation of the viscosity of a gas with temperature. *Proc. R. Soc. Lond. A*, 106:441–462, 1924.

[6] J E Lennard-Jones. Cohesion. *Proceedings of the Physical Society*, 43(5):461, sep 1931.

[7] A. K. Rappe, C. J. Casewit, K. S. Colwell, W. A. III Goddard, and W. M. Skiff. Uff, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *Journal of the American Chemical Society*, 114(25):10024–10035, 1992.

[8] J Godet, L Pizzagalli, S Brochard, and P Beauchamp. Comparison between classical potentials and ab initio methods for silicon under large shear. *Journal of Physics: Condensed Matter*, 15(41):6943, oct 2003.

[9] Jörg Behler and Michele Parrinello. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys. Rev. Lett.*, 98:146401, Apr 2007.

[10] Albert P. Bartók, Mike C. Payne, Risi Kondor, and Gábor Csányi. Gaussian approximation potentials: The accuracy of quantum mechanics, without the electrons. *Phys. Rev. Lett.*, 104:136403, Apr 2010.

[11] Alexander V. Shapeev. Moment tensor potentials: A class of systematically improvable interatomic potentials. *Multiscale Modeling & Simulation*, 14(3):1153–1173, 2016.

[12] A.P. Thompson, L.P. Swiler, C.R. Trott, S.M. Foiles, and G.J. Tucker. Spectral neighbor analysis method for automated generation of quantum-accurate interatomic potentials. *Journal of Computational Physics*, 285:316–330, 2015.

[13] Mitchell A. Wood and Aidan P. Thompson. Extending the accuracy of the SNAP interatomic potential form. *The Journal of Chemical Physics*, 148(24):241721, 03 2018.

[14] Yunxing Zuo, Chi Chen, Xiangguo Li, Zhi Deng, Yiming Chen, Jörg Behler, Gábor Csányi, Alexander V. Shapeev, Aidan P. Thompson, Mitchell A. Wood, and Shyue Ping Ong. Performance and cost assessment of machine learning interatomic potentials. *The Journal of Physical Chemistry A*, 124(4):731–745, 2020. PMID: 31916773.

[15] Stephen R Xie, Matthias Rupp, and Richard G Hennig. Ultra-fast interpretable machine-learning potentials. *npj Computational Materials*, 9(1):162, 2023.

[16] Joshua A. Vita and Dallas R. Trinkle. Exploring the necessary complexity of interatomic potentials. *Computational Materials Science*, 200:110752, 2021.

[17] Xiang Fu, Zhenghao Wu, Wujie Wang, Tian Xie, Sinan Keten, Rafael Gomez-Bombarelli, and Tommi Jaakkola. Forces are not enough: Benchmark and critical evaluation for machine learning force fields with molecular simulations, 2022.

[18] Joe D. Morrow, John L. A. Gardner, and Volker L. Deringer. How to validate machine-learned interatomic potentials. *The Journal of Chemical Physics*, 158(12):121501, 03 2023.