

Recent advances and outstanding challenges for machine learning interatomic potentials

Tsz Wai Ko & Shyue Ping Ong



Machine learning interatomic potentials (MLIPs) enable materials simulations at extended length and time scales with near-ab initio accuracy. They have broad applications in the study and design of materials. Here, we discuss recent advances, challenges, and the outlook for MLIPs.

Atomistic simulations are an important tool in the study of materials. The reliability of an atomistic simulation is critically dependent on the quality of the given potential energy surface (PES). Electronic structure methods provide ab initio descriptions of the PES across the periodic table, but at a high computational expense and poor scaling with system size. On the other hand, traditional interatomic potentials (IPs) parameterize the PES as a function of atomic degrees of freedom, gaining orders of magnitude speed-ups and linear scaling with respect to the number of atoms, but sacrificing significant accuracy and transferability in return.

In the past two decades, machine learning IPs (MLIPs) have emerged as a transformative alternative that bridges the gap between these two extremes. MLIPs are trained to a data set of reference structures with their corresponding energies, forces, and/or stresses, which

are typically obtained from accurate ab initio calculations. State-of-the-art MLIPs often achieve a prediction accuracy of <5 meV atom⁻¹ in energies and <0.1 eV Å⁻¹ in forces, almost an order of magnitude better than traditional IPs and sufficient to resolve the small energy differences between many polymorphs¹.

In this Comment, we will discuss and compare recent advances in MLIP architectures, as well as the outstanding challenges and opportunities.

Local environment versus graph descriptions

An MLIP comprises two interconnected components: the features used to describe the atomic character and positions, and the ML model used to map these features to the PES. MLIPs can be broadly categorized into two types (Fig. 1). The majority of MLIPs employ a set of descriptors to represent the local chemical environment of each atom, such as the atomic density or bond distances and angles, together with various ML algorithms ranging from simple linear regression to kernel methods to neural networks². More recently, a new class of MLIPs has emerged that utilizes a graph description, in combination with message-passing neural networks. Henceforth, we will refer to the former as local MLIPs (L-MLIPs) and the latter as graph MLIPs (G-MLIPs).

The key advantages of L-MLIPs are their relative compactness and computational efficiency. A major recent advance in L-MLIPs is the atomic cluster expansion³, which provides a hierarchical, complete, and efficient approach to incorporate many-body terms.

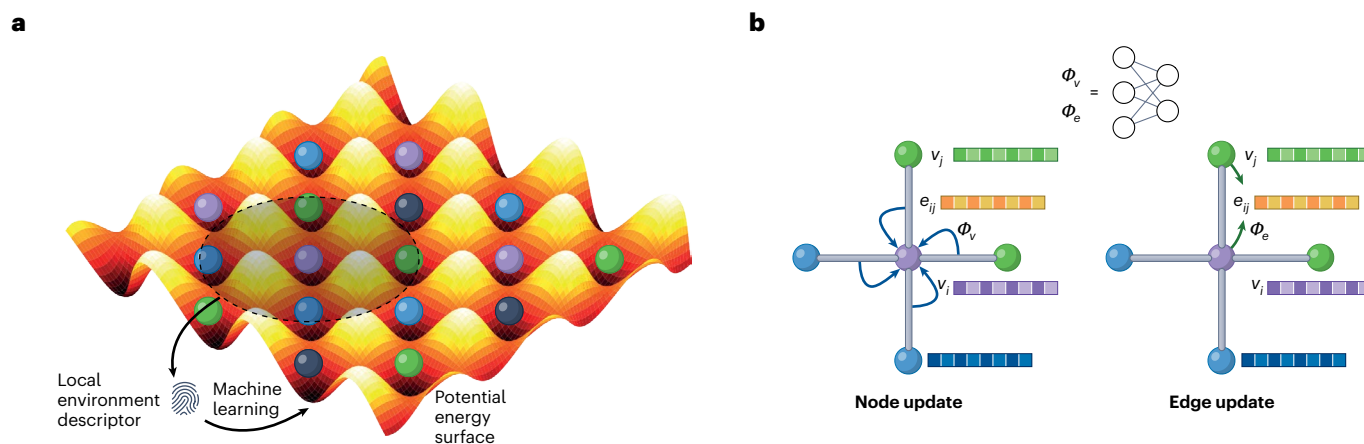


Fig. 1 | Types of machine learning interatomic potentials. **a**, Local MLIPs describe the local environment around each atom using descriptors based on the atomic density, bond distances and angles, and so on. An ML model is then trained to model the relationship between the local environment descriptors and the PES. The corrugated surface represents the PES of a system, which depends on the local chemical environment of each atom. **b**, Graph MLIPs utilize a graph-based representation where atoms are the nodes (v_i, v_j) and bonds are the

edges (e_{ij}). Each node is a vector uniquely defined by elemental character. Information flows through the graph via message passing (arrows). The message passing functions ϕ_e and ϕ_v are usually feed-forward neural networks that are fitted as part of model training. In equivariant MLIPs, the message-passing functions are chosen such that they preserve the transformation of tensors under coordinate transformations.

This effectively addresses the incomplete representation between distinct atomic environments with only two-body and three-body descriptors⁴. Despite their major successes, a critical limitation of L-MLIPs is the combinatorial explosion in the number of features with the number of distinct elements, which has a corresponding impact on the amount of training data required, as well as computational efficiency. Thus, L-MLIPs have mostly been applied to relatively chemically ‘simple’ systems that contain up to four elements.

In contrast, the primary advantage of G-MLIPs is their ability to handle systems of arbitrary complexity. In G-MLIPs, each species is represented as a unique vector. As such, G-MLIPs do not suffer from the same combinatorial explosion with the number of distinct elements⁵. Furthermore, the range of atomic interactions can be extended beyond the cutoff radius for graph construction via the number of message-passing steps. However, the message-passing framework is inefficient in terms of modeling many-body interactions. Most modern G-MLIPs such as Materials 3-body Graph Network (M3GNet)⁵ and Message-passing Atomic Cluster Expansion (MACE)⁶ therefore combine message-passing graphs with many-body terms based on either traditional IP or L-MLIPs formalism to achieve the best of both worlds. Also, nonlocal message-passing G-MLIPs are inherently more difficult to parallelize than local L-MLIPs. Hence, G-MLIPs are still about an order of magnitude more computationally intensive than L-MLIPs, and developing efficient implementations that can make the best use of large high-performance computing (CPU and GPU) resources remains an open challenge.

Invariance versus equivariance

All IPs are symmetry aware. At the minimum, all IPs must respect the invariance of the PES under translation, rotation, and permutation of identical atoms, which can be accomplished using scalar features. In recent years, there has been great interest in applying the concept of equivariance to MLIPs, particularly G-MLIPs. Equivariant G-MLIPs preserve how tensorial quantities change under coordinate transformations through the design of the input attributes and message-passing functions. For example, the neural equivariant interatomic potential (Nequip), a popular equivariant G-MLIP, uses the normalized relative position vectors projected onto a spherical harmonic basis as edge attributes and a geometric tensor product to combine the edge features with neighboring atomic features.

The primary advantage of equivariant G-MLIPs is that the relative position vectors between atoms can be used, which are more expressive than the scalar bond distances and angles used in invariant G-MLIPs. A series of benchmarks conducted on small molecular systems⁶ and a few bulk systems⁷ have found that equivariant G-MLIPs can achieve accuracy comparable to that of invariant G-MLIPs with substantially fewer training points. It should be stressed that these benchmarks have been carried out on a limited number of systems, and more recent studies do not seem to support that there is a substantial advantage to equivariant over invariant MLIPs, especially on bulk crystals⁸. An open challenge is a conclusive, preferably mathematically rigorous, demonstration of the advantages (or lack thereof) of equivariance over invariance in terms of performance on systems across diverse structures and chemistries.

Universal versus custom MLIPs

Most IPs (ML or traditional) are custom-fitted to a limited set of elements and structures of interest. This can be attributed to the unfavorable combinatorial scaling of traditional and L-MLIP descriptors with

the number of elements as well as the absence of a large standardized dataset covering a wide range of chemistries. Both constraints have been significantly mitigated with the development of G-MLIPs and large open databases of DFT materials data such as the [Materials Project](#). This has led to the emergence of ‘universal’ G-MLIPs^{5,6,9} trained on datasets that cover the entire periodic table.

At the present stage of their development, universal and custom MLIPs serve distinct and complementary purposes. While G-MLIPs perform comparably with custom MLIPs given the same training data^{5,9}, the current generation of universal G-MLIPs generally makes compromises in accuracy for specific chemistries in exchange for reasonably good accuracy across diverse chemistries. This is likely due to the inter-related issues of the relatively less stringent DFT convergence criteria of existing whole-periodic-table PES datasets as well as insufficient model complexity. In fact, recent work suggests that addressing both these issues can improve the accuracy and extrapolability of universal G-MLIPs¹⁰. Despite these limitations, existing universal G-MLIPs have broad applications in the discovery and simulations of materials across diverse chemical spaces. They effectively replace *ab initio* methods where speed and scalability are prioritized over accuracy, such as conducting preliminary screening of millions of materials. Universal G-MLIPs can also be thought of as ‘foundation models’ for materials simulations, that is, pre-trained models with reliable performance across a wide range of tasks that can be further retrained quickly as custom MLIPs for specific chemistries or applications.

Data and learning

Regardless of architecture, MLIPs are only as accurate as the data on which they are trained. The predominant approach to generate such datasets is DFT calculations using semi-local generalized gradient approximation (GGA) functionals. However, it is well established that GGA functionals have major shortcomings that can be addressed with modern meta-GGA functionals¹¹. The main impediment to using these improved functionals is their higher computational cost. Nevertheless, the generation of open, high-quality PES datasets using state-of-the-art methods is a critical immediate priority for future MLIP development.

Another major priority is the development of sampling and training approaches with better data efficiency. Current MLIP development best practices often incorporate some form of active learning (AL) to ensure comprehensive coverage of the configuration space of interest. New approaches such as entropy maximization and uncertainty-driven dynamics have recently been applied to bias the sampling to unexplored spaces, which accelerates AL. Despite these advances, AL can be highly inefficient for complex configuration spaces. Qi et al.¹⁰ recently proposed an approach that utilizes universal G-MLIPs to generate a large configuration space without the need for DFT calculations, followed by dimensionality reduction and stratified sampling to ensure comprehensive coverage of the entire configuration space. This approach has been shown to reduce and, in some cases, eliminate the need for iterative AL. Finally, transfer learning and multi-fidelity approaches can also significantly reduce the number of high-fidelity training data points required¹².

Charge and magnetism

With a few notable exceptions^{9,13,14}, most MLIP architectures do not explicitly incorporate atomic charges and magnetism. Accurate descriptions of these long-ranged interactions are necessary for a wide range of materials science problems, including reactive systems and interfaces, electronically driven phase transitions, magnetic materials, and so on. In our view,

the main challenge in incorporating charges and magnetism into MLIPs lies not in the architectural developments but in the generation of training data and computational efficiency. Several effective approaches, such as charge equilibration schemes and spin-lattice G-MLIPs, have already been developed. However, training such MLIPs requires data sets that include charge and magnetic degrees of freedom, in addition to the coupled structural and chemical degrees of freedom, which massively increases the computational cost and complexity of data generation. Furthermore, the added complexity of incorporating electrostatic and magnetic interactions also tends to have a negative impact on the computational efficiency and scaling of the MLIPs, reducing the length and time scales to which they can be applied¹³. Practical algorithmic improvements are necessary to optimize the efficiency of such models, and MLIPs in general, to broaden the scope of their applicability.

Outlook

To conclude, MLIPs have had a transformative impact on materials science by providing a robust, automatable approach to parameterize the PES. MLIPs enable the simulation of materials at length and time scales beyond that accessible by *ab initio* methods, while reproducing near-*ab initio* accuracy. The recent advent of universal G-MLIPs has further expanded the scope of their application beyond studies of select chemistries to broad-based materials discovery. We are confident that continued advances in featurization, architectures and best practices will lead to MLIPs with improved accuracy, efficiency, and generalizability. However, an area that requires more attention from the community is the optimization of MLIP software implementations. MLIPs, in particular, G-MLIPs, are still orders of magnitude more computationally expensive than traditional IPs, which can potentially limit the length and time scales to which they can be applied.

In our view, the main bottleneck for future MLIP development is ultimately the scarcity of PES data. Generating high-quality PES data, especially using state-of-the-art functionals, is the most expensive step in creating a MLIP. Existing materials databases were developed with a primary focus on property information for materials discovery and design; relatively little attention has been paid to the development, curation, and dissemination of comprehensive PES data. The development of a findable, accessible, interoperable and reusable

(FAIR)¹⁵ repository of PES data, as a complement to databases of IPs such as [OpenKim](#), is therefore a critical priority for the community. A PES database should include not only typical PES information, but also auxiliary information such as charge densities, magnetic moments, and so on, derived from electronic structure calculations that may be utilized in future model development. It can serve as the foundation for developing common benchmarks for MLIPs, providing a standardized dataset for training, testing and evaluating the computational efficiency of different MLIP architectures.

Tsz Wai Ko  & **Shyue Ping Ong**  

Department of NanoEngineering, University of California San Diego, La Jolla, CA, USA.

✉ e-mail: ongsp@ucsd.edu

Published online: 13 December 2023

References

1. Zuo, Y. et al. *J. Phys. Chem. A* **124**, 731–745 (2020).
2. Behler, J. & Csányi, G. *Eur. Phys. J. B* **94**, 142 (2021).
3. Drautz, R. *Phys. Rev. B* **99**, 014104 (2019).
4. Musil, F. et al. *Chem. Rev.* **121**, 9759–9815 (2021).
5. Chen, C. & Ong, S. P. A. *Nat. Comput. Sci.* **2**, 718–728 (2022).
6. Batatia, I., Kovacs, D. P., Simm, G., Ortner, C. & Csányi, G. *Adv. Neural Inf. Process. Syst.* **35**, 11423–11436 (2022).
7. Batzner, S. et al. *Nat. Commun.* **13**, 2453 (2022).
8. Riebesell, J. et al. Preprint at <https://arxiv.org/pdf/2308.14920.pdf> (2023).
9. Deng, B. et al. *Nat. Mach. Intell.* **5**, 1031–1041 (2023).
10. Qi, J. et al. Preprint at <https://arxiv.org/abs/2307.13710> (2023).
11. Sun, J. et al. *Nat. Chem.* **8**, 831–836 (2016).
12. Smith, J. S. et al. *Nat. Commun.* **10**, 2903 (2019).
13. Ko, T. W., Finkler, J. A., Goedecker, S. & Behler, J. A. *Nat. Commun.* **12**, 398 (2021).
14. Yu, H. et al. Preprint at <https://arxiv.org/abs/2203.02853> (2023).
15. Wilkinson, M. D. et al. *Sci. Data* **3**, 160018 (2016).

Acknowledgements

This work was primarily supported by the Materials Project, funded by the US Department of Energy, Office of Science, Office of Basic Energy Sciences, Materials Sciences and Engineering Division under contract no. DE-AC02-05-CH11231; Materials Project program KC23MP. T.W.K also acknowledges support from the Schmidt AI in Science Postdoctoral Fellowship.

Competing interests

The authors declare no competing interests.