Modelling and
Simulation in Materials
Science and
Engineering

ROADMAP • **OPEN ACCESS**

Roadmap for the development of machine learning-based interatomic potentials

To cite this article: Yong-Wei Zhang *et al* 2025 *Modelling Simul. Mater. Sci. Eng.* **33** 023301

OPEN ACCESS
IOP Publishing
Modelling and Simulation in Materials Science and Engineering

Modelling Simul. Mater. Sci. Eng. **33** (2025) 023301 (54pp)
https://doi.org/10.1088/1361-651X/ad9d63

Roadmap

Roadmap for the development of machine learning-based interatomic potentials

Yong-Wei Zhang[1,16,*], **Viacheslav Sorkin**[1], **Zachary H Aitken**[1], **Antonio Politano**[2,16,*], **Jörg Behler**[3,4], **Aidan P Thompson**[5], **Tsz Wai Ko**[6], **Shyue Ping Ong**[6], **Olga Chalykh**[7], **Dmitry Korogod**[8], **Evgeny Podryabinkin**[7], **Alexander Shapeev**[7], **Ju Li**[9], **Yuri Mishin**[10], **Zongrui Pei**[11], **Xianglin Liu**[12], **Jaesun Kim**[13], **Yutack Park**[13], **Seungwoo Hwang**[13], **Seungwu Han**[13,14], **Killian Sheriff**[15], **Yifan Cao**[15] **and Rodrigo Freitas**[15]

[1] Institute of High-Performance Computing (IHPC), Agency for Science, Technology and Research (A*STAR), 1 Fusionopolis Way, #16-16 Connexis, Singapore 138632, Singapore
[2] Department of Physical and Chemical Sciences, University of L'Aquila, via Vetoio, 67100 L'Aquila, Italy
[3] Lehrstuhl für Theoretische Chemie II, Ruhr-Universität Bochum, 44780 Bochum, Germany
[4] Research Center Chemical Sciences and Sustainability, Research Alliance Ruhr, 44780 Bochum, Germany
[5] Center for Computing Research, Sandia National Laboratories, Albuquerque, New Mexico, United States of America
[6] Aiiso Yufeng Li Family Department of Chemical and Nano Engineering, University of California, San Diego, 9500 Gilman Dr Mail Code #0448, La Jolla, CA 92093-0448, United States of America
[7] Skolkovo Institute of Science and Technology, Russia, Bolshoy Boulevard 30, bld. 1., Skolkovo 121205, Russia
[8] Moscow Institute of Physics and Technology, Russia, Institutsky lane 9, Dolgoprudny, Moscow region 141700, Russia
[9] Department of Nuclear Science and Engineering and Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, United States of America

---

[16] Guest Editor of the roadmap.
[*] Authors to whom any correspondence should be addressed.

[10] Department of Physics and Astronomy, George Mason University, 4400 University Drive, MSN 3F3, Fairfax, VA 22030, United States of America
[11] New York University, New York, NY 10012, United States of America
[12] Peng Cheng Laboratory, Shenzhen 518066, People's Republic of China
[13] Department of Materials Science and Engineering, Seoul National University, Seoul 08826, Republic of Korea
[14] Korea Institute for Advanced Study, Seoul 02455, Republic of Korea
[15] Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, United States of America

E-mail: zhangyw@ihpc.a-star.edu.sg and antonio.politano@univaq.it

Received 2 August 2024; revised 13 November 2024
Accepted for publication 11 December 2024
Published 28 January 2025

Abstract

An interatomic potential, traditionally regarded as a mathematical function, serves to depict atomic interactions within molecules or solids by expressing potential energy concerning atom positions. These potentials are pivotal in materials science and engineering, facilitating atomic-scale simulations, predictive material behavior, accelerated discovery, and property optimization. Notably, the landscape is evolving with machine learning transcending conventional mathematical models. Various machine learning-based interatomic potentials, such as artificial neural networks, kernel-based methods, deep learning, and physics-informed models, have emerged, each wielding unique strengths and limitations. These methods decode the intricate connection between atomic configurations and potential energies, offering advantages like precision, adaptability, insights, and seamless integration. The transformative potential of machine learning-based interatomic potentials looms large in materials science and engineering. They promise tailor-made materials discovery and optimized properties for specific applications. Yet, formidable challenges persist, encompassing data quality, computational demands, transferability, interpretability, and robustness. Tackling these hurdles is imperative for nurturing accurate, efficient, and dependable machine learning-based interatomic potentials primed for widespread adoption in materials science and engineering. This roadmap offers an appraisal of the current machine learning-based interatomic potential landscape, delineates the associated challenges, and envisages how progress in this domain can empower atomic-scale modeling of the composition-processing-microstructure-property relationship, underscoring its significance in materials science and engineering.

Keywords: machine learning, interatomic potentials, neural networks, atomic simulations

Contents

1. Introduction 4
2. Databases for machine learning-based interatomic potentials 7
3. Machine learning-based empirical interatomic potentials 11
4. Artificial neural networks-based interatomic potentials 15
5. Large-scale atomistic simulation using machine learning-based interatomic potentials 18
6. Graph deep learning potentials for atomistic simulations 22
7. MTPs 25
8. UNIP 28
9. Physically informed machine-learning potential for atomistic simulations 33
10. Machine learning-based potentials for medium and HEAs 37
11. Pretrained UMLP 41
12. Capturing chemical complexity in high-entropy materials 45

Data availability statement 48

References 49

1. Introduction

Yong-Wei Zhang
Institute of High-Performance Computing (IHPC), Agency for Science, Technology and Research (A*STAR), 1 Fusionopolis Way, #16-16 Connexis, Singapore 138632, Republic of Singapore

Atomistic simulations have become an indispensable tool in materials research, providing detailed insights into the fundamental behaviors and properties of materials at the atomic level. These simulations enable researchers to predict material properties, understand mechanisms of material transformations, and design new materials with tailored properties. However, the accuracy and efficiency of these simulations heavily depend on the quality of the interatomic potentials used. Traditional interatomic potentials often face significant challenges in balancing computational efficiency with the accuracy required to faithfully represent complex atomic interactions. This bottleneck has limited the predictive power and applicability of atomistic simulations across various materials systems. Addressing these challenges is crucial for advancing the field and unlocking the full potential of atomistic simulations in materials research.

The current research landscape in interatomic potentials is marked by the widespread use of empirical interatomic potentials, such as Lennard–Jones, embedded atom method (EAM), Stillinger–Weber, ReaxFF, Tersoff, and reactive empirical bond order (REBO). These empirical interatomic potentials are traditionally fitted by tuning parameters to align with experimental data or first-principles calculations. This process starts by selecting target properties, such as lattice constants, elastic moduli, and energy landscapes. A suitable potential form (e.g. EAM, REBO) is then chosen based on the bonding characteristics of the material. Parameters are optimized using techniques like least-squares or genetic algorithms to minimize deviations from target data. The potential is validated against additional properties to check for transferability and refined through iterative adjustments. Final testing across different configurations confirms its reliability, yielding a practical approximation of atomic interactions. Their importance lies in their simplicity and computational efficiency, enabling simulations of millions of atoms over long timescales. However, empirical interatomic potentials come with significant limitations, including limited transferability and lack of accuracy in capturing subtle atomic interactions, leading to potential inaccuracies in the predicted properties and behaviors of materials.

This is where the development of machine learning-based interatomic potentials (MLIPs) comes into play. By leveraging large datasets and advanced algorithms, machine learning methods have the potential to learn complex atomic interactions directly from data, offering a promising pathway to overcome the limitations of traditional empirical potentials. The integration of machine learning in this domain aims to achieve a balance between computational efficiency and predictive accuracy, paving the way for more reliable and versatile atomistic simulations in materials research.

MLIPs have become a transformative tool in atomistic simulations by combining computational efficiency with quantum-level accuracy. Foundational work by Behler [1] offers a perspective on neural network potentials (NNPs), introducing key methodologies and challenges in atomistic MLIPs, such as the need for accurate, transferable models across material types. Mishin [2] provides an overview of MLIPs in materials science, discussing various ML approaches, such as neural networks and Gaussian process models, and how they overcome the limitations of classical models in terms of flexibility and transferability. Expanding on these ideas, Mueller *et al* [3] explore diverse ML frameworks, highlighting how ML refines interatomic potentials and enables complex, multiscale simulations. Unke *et al* [4] introduce SpookyNet, a model incorporating nonlocal electronic effects, which opens new possibilities in electronic-structure-informed MLIPs. Last but not least, Friederich *et al* [5] and Fedik *et al* [6] discuss next-generation MLIPs with applications in molecular property prediction and how these methods extend beyond force-field approximations, respectively. These perspectives and reviews provide a broad spectrum of valuable references for understanding the evolution, methodologies, and future potential of MLIPs.

Currently, several types of MLIPs are being actively researched and developed, including NNPs, Gaussian approximation potentials (GAPs), and moment tensor potentials (MTPs). These models leverage vast amounts of data and sophisticated algorithms to capture complex atomic interactions with high accuracy. For instance, NNPs can model intricate energy landscapes by learning from quantum mechanical calculations, while GAPs provide a probabilistic framework to account for uncertainties in predictions. Despite their potential, MLIPs come with their own set of challenges. One significant advantage is their ability to achieve unprecedented accuracy and transferability across different materials and conditions, surpassing traditional empirical potentials. However, this accuracy often comes at the cost of increased computational demands, making them less efficient for large-scale simulations compared to empirical methods. Additionally, the training process for these models requires extensive datasets and computational resources, and there is a need for rigorous validation to ensure their reliability.

This roadmap article addresses these challenges by providing a comprehensive guide for the development and application of MLIPs in materials research. It aims to highlight the current progress, identify existing gaps, and propose future directions to enhance the performance and applicability of these potentials. By doing so, the article serves as a crucial resource for researchers, guiding them through the complexities of MLIPs and fostering advancements that can revolutionize atomistic simulations and materials design.

The roadmap article is structured to provide a comprehensive overview of the various facets involved in the development of MLIPs. The first section, authored by Viacheslav Sorkin and Yong-Wei Zhang, delves into the importance of databases for MLIPs. This section discusses the critical role of high-quality, extensive datasets in training accurate and reliable interatomic potentials, setting the foundation for subsequent advancements in the field. Following this, Zachary H Aitken, Antonio Politano, and Yong-Wei Zhang explore machine learning-based empirical interatomic potentials, examining how machine learning can enhance traditional empirical potentials, improving their accuracy and transferability while maintaining computational efficiency. Jörg Behler's section on artificial neural networks-based interatomic potentials introduces neural network models tailored for atomistic simulations, highlighting their capabilities to model complex atomic interactions with high accuracy. Aidan Thompson discusses large-scale atomistic simulations using MLIPs, focusing on the scalability of these methods and addressing the challenges and solutions for applying MLIPs to simulate extensive systems, which is crucial for studying materials with significant atomic complexity. The section by Tsz Wai Ko and Shyue Ping Ong covers graph deep learning potentials for atomistic simulations, leveraging graph theory and deep learning to capture atomic interactions more effectively. MTPs are examined by Olga Chalykh, Dmitry Korogod, Evgeny Podryabinkin, and Alexander Shapeev, exploring their balance between accuracy and computational efficiency, making them suitable for a wide range of materials and conditions. Ju Li introduces the universal neural interatomic potential (UNIP), focusing on creating versatile models applicable across diverse material systems, emphasizing their potential to streamline the development process and enhance the predictive power of MLIPs. Yuri Mishin's section on physically informed machine-learning potentials integrates physical principles with machine learning techniques, aiming to improve the reliability and interpretability of MLIPs. Zongrui Pei and Xianglin Liu discuss machine learning-based potentials for medium and high entropy alloys (HEAs), addressing the unique challenges posed by these complex materials and demonstrating how MLIPs can provide accurate predictions for their properties and behaviors. The section by Jaesun Kim, Yutack Park, Seungwoo Hwang, and Seungwu Han presents the concept of pretrained universal machine learning potentials (UMLPs), leveraging pretrained models to accelerate the development process and improve the efficiency of MLIPs for various applications. Finally, Killian Sheriff, Yifan Cao, and Rodrigo Freitas cover capturing chemical complexity in high-entropy materials, highlighting the role of MLIPs in accurately modeling intricate interactions in high-entropy materials, facilitating their design and optimization.

By providing detailed insights into each of these areas, this roadmap article aims to guide researchers through the complexities and opportunities in the development of machine learning-based interatomic potentials, ultimately advancing the field of atomistic simulations in materials research.

In conclusion, the development of MLIPs marks a significant milestone in the advancement of atomistic simulations. This roadmap article provides a thorough exploration of the current state of MLIPs, the challenges faced, and the innovative solutions being proposed across various domains. By integrating contributions from leading experts in the field, this article aims to offer a comprehensive guide for researchers, facilitating the development and application of more accurate, efficient, and versatile interatomic potentials. As we continue to harness the power of machine learning, the future of materials research looks promising, with the potential to achieve unprecedented insights and breakthroughs in material design and discovery.

Acknowledgments

The author acknowledges the support from the Italy-Singapore Science and Technology Cooperation Project with Award Number R22I0IR121, Singapore A*STAR AME Programmatic Project (Grant No. A1898b0043), and the support from Singapore A*STAR SERC CRF Award.

2. Databases for machine learning-based interatomic potentials

Viacheslav Sorkin and Yong-Wei Zhang
Institute of High-Performance Computing (IHPC), Agency for Science, Technology and Research (A*STAR), 1 Fusionopolis Way, #16-16 Connexis, Singapore 138632, Republic of Singapore

Status

Databases containing atomistic structures with DFT-calculated energies and forces are crucial for designing machine learning interatomic potentials (MLIPs) [2]. To accurately represent the potential energy surface (PES), these databases encompass a diverse set of atomistic structures [7].

Early MLIP databases faced limitations. They relied on hand-picked data, chosen by researchers based on their expertise and specific problems (see figure 1), which could introduce biases into the MLIP [2]. Additionally, *ab-initio* molecular dynamics (MD) simulations, while used to generate data at various temperatures and pressures, favored regions around minima on the PES [8]. This bias limited the ability of these databases to capture the full complexity of the PES, including crucial features like saddle points and transition paths [9]. Consequently, these early datasets are now considered a starting point for more advanced approaches like active learning (AL) [9].

AL is a very powerful approach for building databases: It starts with an initial dataset used to train a preliminary MLIP. This model is then used in MD simulations to explore the PES region. During this exploration, AL identifies configurations that significantly differ from those in the training set, according to a predefined selection strategy like query-by-committee [10] and extrapolation grade [11]. The most informative configurations are chosen, their geometries are optimized using DFT. This data is then fed back into the MLIP for retraining, improving its accuracy for the specific PES domain. These three steps—structure selection, DFT optimization, and MLIP retraining—are repeated iteratively until convergence is reached (see figure 2). Convergence signifies that **AL no longer identifies new configurations of low-confidence predictions**, indicating sufficient exploration of the relevant PES domain [7]. Moving beyond MD simulations, evolutionary algorithms [12] and random structure searching tools [13] offer alternative ways to generate initial datasets. They can then be complemented by AL for fine-tuning the database.

Metadynamics is a powerful tool for exploring complex free energy landscapes by adding a bias to the system that helps escape local minima and steer MD simulations toward previously unvisited atomic environments such that each atom experiences diverse chemical environments without redundancy. This is especially useful in cases where traditional MD simulations may struggle with barriers between configurations, making it a valuable technique for enhancing sampling efficiency in challenging systems. If dataset diversity is crucial, metadynamics can indeed be an excellent method for generating varied configurations for machine learning in material science applications. Currently, the application of metadynamics to create databases for MLIPs is gaining traction, as it facilitates the collection of diverse atomic configurations critical for training accurate MLIPs [14–17].

Advancements in AL database generation hold immense potential for unlocking the full richness of the PES by efficiently finding key features like barriers and transition paths, even

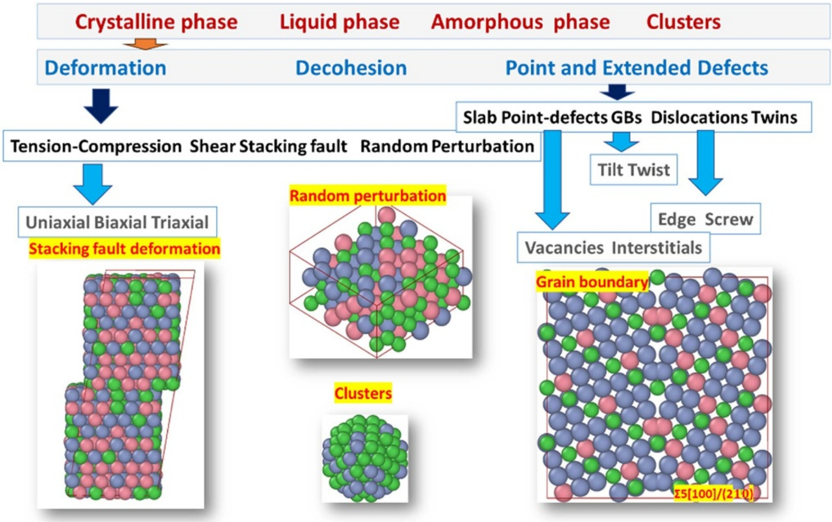

Figure 1. Constructing a database for a ternary alloy: the database incorporates ideal crystal structures, deformed structures, and defected structures (point and extended defects) across various alloy compositions. It also includes liquid, amorphous phases, and small clusters for comprehensive MLIP training. (inset) Examples of deformed, defected structures, and clusters.

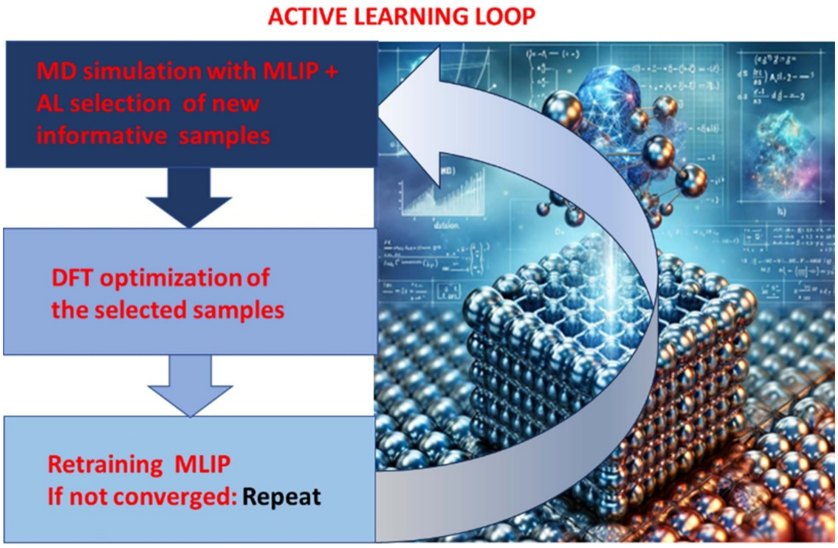

Figure 2. Active learning for MLIPs: an iterative process of selecting informative structures, performing DFT optimization, and retraining the MLIP until convergence is achieved.

without prior knowledge. AL iteratively uses the MLIP to predict features, verifies them with high-fidelity calculations, and refines the model for better accuracy—all without constant user input [11].

Current and future challenges

A critical challenge in MLIP development is ensuring its transferability, its ability to predict properties of atomistic structures not included in the training database) [2]. To achieve this, databases must encompass diverse structures, chemical compositions, and system sizes, pushing the boundaries of MLIP applicability.

A key question in MLIP design is choosing between a universal PES model and a domain-specific one. For exploration, a universal MLIP might be appealing, but capturing the whole PES requires a large training database [2]. Additionally, a universal MLIP might suffer from overfitting or underfitting in specific regions which limits its transferability. Training a universal MLIP becomes particularly daunting problem for multi-component system due to their vast compositional space. Capturing all PES regions in such systems with a single reference dataset is impossible, making domain-specific MLIPs an only solution. Domain-specific MLIPs offer higher accuracy with less data, reducing computational costs, which ideal for applications with limited structural variations [2, 18].

A key challenge in using AL to create a database for MLIPs arises from the vast difference in scale between MD and DFT simulations. DFT typically handles hundreds of atoms, whereas MD simulations, particularly those involving dislocations, cracks, inclusions, and grain boundaries can involve from thousands to billions of atoms [19]. This mismatch makes it impractical to optimize the entire MD configurations with DFT during AL cycles.

AL relies on a well-designed inquiry scheme to efficiently select informative configurations for training MLIPs. Two common approaches are query-by-committee and extrapolation grade. Query-by-committee involves creating multiple MLIP versions with differently fitted parameters. If these models disagree significantly on the energy and forces of a new structure, that structure is included in the training database. Extrapolation grade, favored for MTPs MLIPs, prioritizes structures that significantly deviate from the existing training data by using D-optimality criterion [11].

However, a key challenge for all these inquiry schemes is the lack of a rigorous mathematical foundation. Additionally, AL can introduce bias by focusing on highly uncertain configurations, potentially neglecting subtle errors relevant for specific applications. This can lead the model to prioritize irrelevant aspects of the PES and miss important features.

Advances in science and technology to meet challenges

Overcoming these challenges in MLIP database development requires advancements in methodology. A promising approach is the continued development of AL-based methods for database construction. By strategically selecting new configurations for training based on uncertainty estimates [20–23], AL can guide the MLIP towards a more comprehensive representation of the PES, even beyond the data it was explicitly trained on.

A hybrid approach can strike a balance between the comprehensiveness of database for a universal MLIP developed for the entire PES and the accuracy of domain-specific MLIP. This strategy may combine multiple MLIPs, each tailored to a specific domain. For less critical regions, coarser models can be employed to improve computational efficiency without sacrificing essential accuracy.

The challenge of mismatched scales between DFT and MD can be mitigated by a subsampling approach. This method may focus on small DFT scale PES regions where the current MLIP, guided by uncertainty inquiry, performs poorly in predicting energy and forces. These subregions can be extracted and surrounded by a padding region, similar to multiscale methods, to satisfy periodic boundary conditions with minimally possible disruption to the internal structure. Once optimized with DFT, these smaller regions can be used within the standard AL framework for retraining and improving the MLIP. The selection criteria for these subregions would be similar to the standard approach [19].

To overcome current limitations of AL-based database development for MLIPs, novel error detection algorithms are crucial: quantifying uncertainty and error associated with the data is essential for reliable MLIP predictions. Research is ongoing to develop methodologies for estimating uncertainty and error propagation in MLIPs.

The growing need for MLIP training data has fuelled the creation of open-access databases like The Materials Project, Open-KIM, and Open Quantum Materials Database containing structures which can be used to create a starting training dataset. To ensure the accuracy of DFT based database, high-throughput computational methods were used. Efforts are underway to standardize data formats to enhance data interoperability and accessibility. These resources, along with popular software (VASP, ESPRESSO, LAMMPS), enhance usability for researchers. Cross-disciplinary collaboration is essential to maximize the potential of these databases and accelerate materials discovery.

Concluding remarks

High-quality databases with diverse and representative atomistic structures are the cornerstone of accurate MLIP predictions for material properties and behavior. These databases are crucial for achieving the goals of materials informatics and accelerating materials discovery and design. Database development for MLIPs is a rapidly evolving field, and continued collaboration, data sharing, and innovative curation methods are essential to maximize their utility and reliability for materials science and beyond. Future advancements in algorithms, theoretical frameworks, and computational efficiency hold promise for overcoming current limitations and delivering the high-quality databases needed to unlock the full potential of MLIPs. This will allow MLIPs to become even more powerful tools for materials modeling and simulation.

Acknowledgments

This contribution is supported by Singapore A*STAR AME Programmatic Project (Grant No. A1898b0043). Y-W Zhang acknowledges the support from Singapore A*STAR SERC CRF Award.

3. Machine learning-based empirical interatomic potentials

Zachary H Aitken[1], *Antonio Politano*[2] *and Yong-Wei Zhang*[1]
[1] Institute of High-Performance Computing (IHPC), Agency for Science, Technology and Research (A*STAR), 1 Fusionopolis Way, #16-16 Connexis, Singapore 138632, Republic of Singapore
[2] University of L'Aquila, Department of Physical and Chemical Sciences, via Vetoio, 67100 L'Aquila, Italy

Status

A blend of empirical and machine learning (ML) approaches highlights the evolving nature of computational materials science, where empirical assumptions informed and refined by ML are paving the way for new materials discoveries and innovations.

Traditional empirical potentials such as Lennard–Jones potentials, bond-order potentials, and EAM potentials, and their modifications (MEAM) have played a pivotal role in the investigation of various material properties and behaviors, yet they frequently encounter limitations in accuracy and their applicability across diverse material systems. These limitations arise partly due to the simplistic nature of the empirical functions or assumptions they employ to model interatomic interactions, which may not capture the complexity of real-world materials.

On the other hand, ML-based interatomic potentials, utilizing sophisticated algorithms such as neural networks, Gaussian process regression, and various deep learning models, aim to overcome these challenges. They achieve this by learning the relationships between atomic configurations and their corresponding energies directly from quantum mechanical data, predominantly derived from density functional theory (DFT) calculations. Areas where ML-based potentials have been able to advance beyond classical potentials include capturing the complex energy landscape of amorphous phases in Si [24] and reproducing the mechanical behavior of 2D materials [25].

An integrated approach, ML-based empirical interatomic potentials, has its formalism rooted in empirical functions or assumptions and refined by ML insights. Combining the strengths of these two interatomic potential formalisms, enables the investigation of material properties and phenomena that classical potentials could not explore due to computational inaccuracy, or that were beyond the reach of DFT methods because of their extensive computational demands for larger systems or prolonged simulation durations.

The integration of ML-based empirical interatomic potentials into established MD and computational chemistry platforms will undoubtedly enhance their accessibility and utility. Their development has already spurred a wave of research, aiming to achieve high computational efficiency, near-quantum mechanical accuracy and offering deep insights into material behavior under various conditions, thus aiding in the discovery and design of novel materials.

Current and future challenges

As an amalgam of ML and empirical potentials, a prominent challenge common to both these features is the computationally expensive creation and utilization of large, high-quality datasets for training, which in turn necessitates exhaustive quantum mechanical simulations. This is especially challenging when attempting to refine and deploy new empirical potentials, as the empirical basis of these models necessitates careful calibration and validation against quantum mechanical data.

ML-based empirical potentials allow the possibility to capitalize on the strengths of ML methods to modify empirical formalisms. The resulting effectiveness and broad applicability of ML-based empirical potentials significantly depends on the chosen functions and the diversity and representativeness of the training data. This highlights the need for judicious selection of the empirical functions, ML models, and the comprehensive datasets that span a wide array of material behaviors and atomic configurations. This will further be integral to address the resource-intensive training phase, where scalability issues can increase dramatically for multicomponent systems [26]. The scalability of size and time can be overcome with prudent design of ML and empirical functions and datasets. These empirical functions and training datasets are vital for ensuring that the empirical models are both versatile and universally applicable, effectively capturing the intricate atomic interactions as dictated by the underlying empirical functions or assumptions while maintaining the efficiency of classical potentials.

Challenges also arise from the empirical foundations of these models, particularly concerning interpretability and transferability. The ‘black box’ nature of ML models and the empirical nature of classic interatomic potentials may hinder the ability to extract meaningful scientific insights and impacting their trustworthiness and reliability. Furthermore, the issue of transferability—where a model trained on specific materials or conditions fails to accurately predict the properties of dissimilar materials—presents a significant obstacle. This challenge is exacerbated in the context of empirical interatomic potentials, as the specific empirical functions or assumptions may not be universally valid across different material systems. Overcoming this hurdle is crucial for achieving widespread acceptance and application of ML-based empirical interatomic potentials.

Advancing science and technology to meet these challenges

To fully harness the capabilities of ML-based empirical interatomic potentials and address the inherent challenges, a multi-faceted approach is essential, focusing on the empirical nature of these models. The empirical-machine learning hybrid approach balances between computational efficiency and predictive accuracy by combining the strengths of empirical potentials with ML-tuned parameters. Traditional empirical potentials, such as EAM and REBO, are computationally inexpensive but often lack the flexibility to adapt to diverse chemical environments and complex interactions. Integrating ML into this framework allows for more adaptive parameter tuning, enhancing the accuracy of empirical potentials without imposing the high computational demands typically associated with fully ML-based models. This hybrid method is particularly advantageous in large-scale MD simulations, where computational cost is a key constraint. By maintaining the simplicity of empirical potentials, hybrid models show significant speed advantages, enabling simulations that are orders of magnitude faster than those of ML potentials, such as neural network or Gaussian process models. Enhancing data efficiency and the scalability of models through innovative algorithmic developments is paramount.

In empirical interatomic potentials, there can be complex relations between the analytic framework and high-order material properties and processes that must be reproduced. Where empirical formulations lack flexibility, ML models are particularly apt at capturing trends in complex systems [27] This has been used where ML approaches were applied to optimize fitting parameters for EAM and ADP potentials [28]. For a given empirical formalism, many important lattice properties, including dislocation core structure, and surface energies as examples in metallic systems, often show correlation that can complicate the empirical fitting procedure. Application of ML approaches to uncover these property correlations [29] can reduce the order complexity of the fitting.

Techniques such as adaptive learning (AL), which allows the model to identify and prioritize data points that significantly enhance its accuracy, can drastically reduce the amount of data required and ensures that the models focus on the most informative atomic interactions as defined by their empirical functions.

Similarly, transfer learning strategies, enabling a model trained on one dataset to adapt to new, related data with minimal additional training, can mitigate issues related to the generalizability and computational efficiency of empirical models. Here, the hybrid approach helps to overcome this critical limitation of transferability across different materials and chemical environments. ML-tuned parameters can be easily adjusted, allowing the hybrid model to better capture complex atomic behaviors in diverse, complex chemical environments. By tuning empirical parameters through ML, the hybrid model can be a versatile tool that is both computationally feasible and scientifically robust. This combination of efficiency and enhanced predictive capacity makes the empirical-ML hybrid approach a valuable tool for material modeling and simulations.

Improving the interpretability of ML models is a critical endeavor in this context. Developing models that incorporate empirical functions or assumptions alongside physical insights or constraints can bridge the gap between data-driven approaches and simulations informed by physical laws. This not only enhances the scientific utility and acceptance of ML-based empirical potentials but also aligns with the need for empirical models to be grounded in physical reality. Approaches that implement empirical features of MEAM models have shown initial promise at engaging the quantum-accuracy of ML models while maintaining the interpretability of the interatomic potential [30, 31]. Figure 3 provides an example of such a ML/empirical architecture that provides interpretability. Tools and methodologies that facilitate the examination and visualization of the models' decision-making processes are crucial for fostering a deeper understanding and trust in these technologies, particularly when they are used to extrapolate empirical relationships in materials science.

Furthermore, advancing the field of ML-based empirical interatomic potentials necessitates a multidisciplinary approach that leverages expertise from materials science, computational physics, and machine learning. Collaborative initiatives, including the development of open-source software and data-sharing platforms, are vital for accelerating the development, validation, and dissemination of these potentials. Such collaborations can promote the creation of standardized datasets and benchmarks tailored to the empirical nature of these models, facilitating model comparison and iterative improvement.

Concluding remarks

ML-based empirical interatomic potentials represent an evolving shift in computational materials science, merging the speed of classical empirical methods with the precision of machine learning. By learning from quantum mechanical data, these potentials overcome traditional limitations, enabling highly accurate simulations across diverse materials systems. Their integration into computational platforms will undoubtedly broaden their application, gain new insights into material behavior and accelerate the discovery of novel materials. The synergy

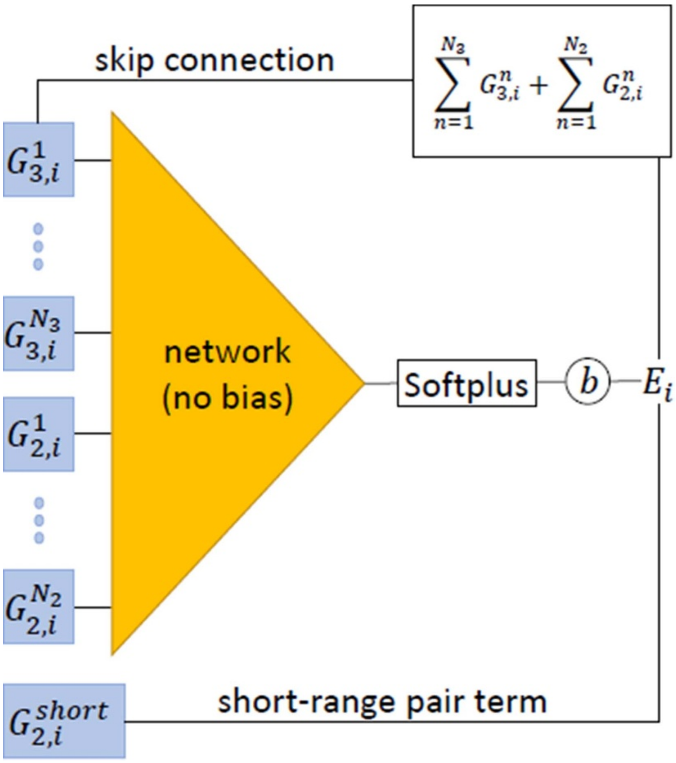

Figure 3. Diagram of a spline-based MEAM potential that employs a neural network. The G functions compose 2- and 3-body spline filters that are directly used as well as fed into a neural network to compute energy. Reprinted from [30], Copyright 2023, with permission from Elsevier.

between empirical assumptions and machine learning not only enhances predictive power but also pushes the boundaries of material science research. Clearly, refining these models for greater efficiency, interpretability, and applicability remains crucial. It is expected that ML-based empirical interatomic potentials will stand as a transformative tool for integrating physical, computational and data sciences, accelerating materials discovery and design.

Acknowledgments

The authors acknowledge the MAECI-A*STAR Science and Technology Italy-Singapore Cooperation Project with Award Number R22I0IR121 for A*STAR and ‘Progetti Grande Rilevanza-PGR’ SG23GR07 for MAECI. AP acknowledges funding from European Community-Next Generation EU, Mission 4, Component 1 under the Ministry of University and Research (MUR) of Italy PRIN 2022 (CUP: E53D23001750006, Grant No. 2022LFWJBR, acronym PLANET) and PRIN PNRR (CUP: E53D23018280001, Grant No. P20223LXTA, acronym ENTANGLE) projects.

4. Artificial neural networks-based interatomic potentials

Jörg Behler[1,2]
[1] Lehrstuhl für Theoretische Chemie II, Ruhr-Universität Bochum, 44780 Bochum, Germany
[2] Research Center Chemical Sciences and Sustainability, Research Alliance Ruhr, 44780 Bochum, Germany

Status

The advent of modern machine learning potentials (MLP) as a tool to compute highly accurate energies and forces of even complex systems has revolutionized the field of atomistic simulations. Due to their very flexible functional form, MLPs are able to learn the shape of the multidimensional PES from reference electronic structure data without the need for a cumbersome construction of approximate and often highly empirical functional forms. As a prominent family of machine learning algorithms, in particular various forms of artificial neural networks have found many applications in the construction of NNPs. The first generation of NNPs applicable to low-dimensional systems has been introduced as early as 1995 and made use of single feed-forward neural networks to derive a functional relation between the atomic positions and the potential energy of the system [32]. Larger systems containing thousands of atoms became accessible with the introduction of high-dimensional NNPs (HDNNP) in 2007 [33], which represent the first MLP of the second generation. Here, the total energy is defined as a sum of local atomic energies that are functions of the geometric environments up to a cutoff radius. The atomic energies are obtained as outputs of a set of individual atomic neural networks, while the input feature vectors vary for different types of second-generation NNPs. Long-range electrostatic interactions based on environment-dependent charges represented, e.g. by a second set to atomic neural networks, can be taken into account in third-generation MLPs [34]. Still, as these potentials build on local energies and charges only, they are unable to take physical phenomena like long-range charge transfer into account. Fourth-generation MLPs overcome this limitation by employing a charge equilibration step, which provides the charge distribution of the systems taking the global structure and total charge of the system into account [35]. In summary, the four generations of MLPs currently available are summarized in figure 4 [36].

Current and future challenges

The emerging field of MLPs has led to a paradigm change in the construction of atomistic potentials, and apart from significant methodical advances also an increasing number of applications to challenging problems—from chemistry to materials science—reported in the literature demonstrate the capabilities of MLPs. Still, there are a large number of remaining challenges that need to be overcome to turn MLPs, and hence also NNPs as an important subgroup of MLPs, into a routinely applicable tool. First, while the flexibility of neural networks is the reason for the high numerical accuracy that can be achieved with respect to the underlying electronic structure method, it is also the origin of the limited transferability of NNPs beyond the structural motifs present in the training data. As the physical shape of MLPs depends on the information provided for these structures, a careful selection is required, while at the same time the computational costs of the parameterization of MLPs is determined by the effort of the electronic structure calculations. Modern NNPs combining environment descriptors learned by message passing [37] and equivariant features [38] promise to strongly reduce the amount

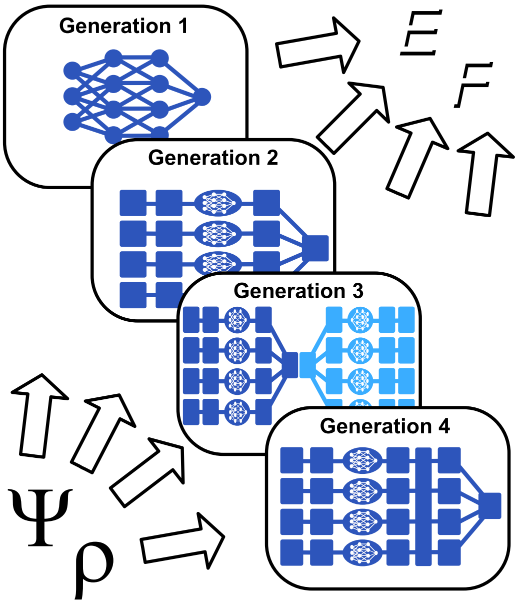

Figure 4. Four generations of neural network potentials.

of data needed to construct reliable potentials but come along with the disadvantage of an increased number of parameters and higher complexity of the method slowing down the evaluation of energies and forces. Moreover, the predictive power of NNPs depends on the quality of the underlying reference electronic structure method, which forms an inevitable limit for the accuracy of the method. In addition, as MLPs enable simulations of very large systems, which are inaccessible by direct electronic structure calculations, the validation of simulation results becomes increasingly challenging. Finally, while in the early years of MLP development a strong focus has been on further and further reducing the errors in learning the energies and forces of the reference data, it has now been realized that low errors for given data sets are not a sufficient criterion for high-quality simulation results or even stable simulations at all. Therefore, the validation of MLPs is increasingly shifting to predicting physical properties and even comparisons with experimental data.

Advances in science and technology to meet challenges

An obvious need for the more efficient use of NNPs in atomistic simulations is the further development of user-friendly software and high-performance hardware. A lot of progress has been made for both, e.g. by the introduction of machine learning libraries and powerful GPUs. Still, they key to a successful application of NNPs does not only require advances on the computational side, which are also beneficial for generating higher-quality reference data sets. In addition, it is and will remain of crucial importance to take the physics of the systems to be studied into account. If phenomena like spin ordering [39] or long-range charge transfer [40] are important but not explicitly considered in the MLP, the obtained results will be at best unreliable. Physical insights gained in NNP-based atomistic simulations thus crucially depend on the choice of a suitable type of MLP, and the extended combination of the numerical flexibility and accuracy of machine learning with physical concepts will remain an important research topic for the years to come. As a guiding principle, whenever possible established physical concepts like long-range electrostatic or dispersion interactions should be included explicitly, while machine learning should focus on those interactions, for which no exact analytic relations are readily available, to improve the transferability and reliability of potentials.

Concluding remarks

Advances in machine learning have opened the door for a novel approach to generate atomistic potentials with essentially first-principles, i.e. quantum mechanical, accuracy that are applicable to a wide range of systems. Much progress has been made in up to now almost three decades of research on neural network-based PESs. Several generations of potentials applicable to various types of systems and taking different physical phenomena explicitly into account are now available, with many successful applications reported in the literature to date. Remaining challenges like reducing the required data set size, increasing the transferability, the development of new validation strategies, and the combination with physical concepts offer many opportunities for future advances and can be expected to substantially push the boundaries of atomistic simulations with predictive accuracy in many fields, from chemistry via biomolecular systems to materials science.

Acknowledgments

The author is grateful for support by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) in the framework of the DFG priority program SPP 2363 (Project 495842446) and under Germany's Excellence Strategy—EXC 2033 RESOLV (Project 390677874).

5. Large-scale atomistic simulation using machine learning-based interatomic potentials

Aidan P Thompson
Center for Computing Research, Sandia National Laboratories, Albuquerque, New Mexico, United States of America

Status

Atomistic simulation provides insight that is complementary to experimental and theoretical studies. By making very few approximations, other than the choice of interatomic potential, it can be used to both test the assumptions of theoretical models and interpret experimental observations. It is particularly valuable for emergent phenomena that are not directly manifested at the atomic scale and are also not captured by continuum models. A good example of this is dynamic heterogeneity in glasses, which was first observed in molecular dynamics simulations [41]. The origins of atomistic simulation of materials can be traced back to the origins of computing itself. Each new advancement in computing hardware has been used to access new levels of accuracy, length, and time. In the last decade, the emergence of machine learning-based potentials (MLPs) has coincided with the proliferation of massively parallel computing resources. Together, these developments have transformed the practice of atomistic simulation, shifting it from qualitative studies of general behaviors to quantitative predictions for specific materials.

LAMMPS [42] is an open source particle simulation code that has played an important role in this transformation. It uses a spatial decomposition parallelization approach to distribute the force calculation over many processors in a manner that is largely independent of the details of the potential. This has allowed each new innovation in MLPs to quickly be made available for use on large parallel computing platforms around the world, thus accelerating the widespread adoption of the most effective approaches. The rapid growth in the use of MLPs in LAMMPS can be quantitatively measured by examining historical citation statistics. In each calendar year, citation lists in papers citing LAMMPS were extracted from the ISI Web of Knowledge (4513 papers in 2024). The 200 most frequently cited papers in 2024 were manually categorized, including 36 categories of interatomic potential. Figure 5 shows how the fraction of papers citing the four most popular potentials (EAM; REBO; Tersoff; ReaxFF) [43] has declined over the last decade, with a corresponding rapid growth in the machine learning category (DeepPot; BP-NNP; GAP; MTP; SNAP; ACE; NEquIP; SchNet; MACE; Allegro) [44]. It is important to note that this type of citation analysis is a lagging indicator. The fraction of LAMMPS simulations using MLPs today is likely even higher than indicated here.

Current and future challenges

The transformation from qualitatively descriptive to quantitatively predictive atomistic simulations, as well as the unprecedented ability to capture emergent behavior on larger scales is particularly well exemplified by the recent multi-billion atom simulations of Oleynik and coworkers [45]. They used the FitSNAP [46] package to train a quadratic SNAP [47] potential that reliably captured structural transformations in compressed diamond, including accurately reproducing the melting curves for diamond, BC8, and simple cubic carbon polymorphs from DFT. They then ran MD simulations of dynamic compression of micron-scale (1.8 billion atoms) samples of diamond, observing the emergence of stress relaxation nucleation sites in

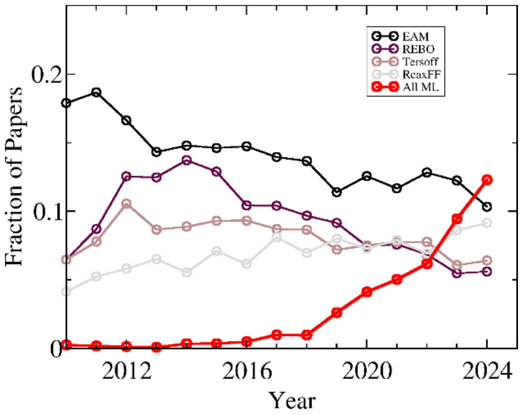

Figure 5. Growth in the use of machine learning-based potentials (red) and slow decline in the use of prominent hand-fitted potentials (EAM black, REBO maroon, Tersoff brown, ReaxFF grey), as measured by citation analysis on papers citing LAMMPS.

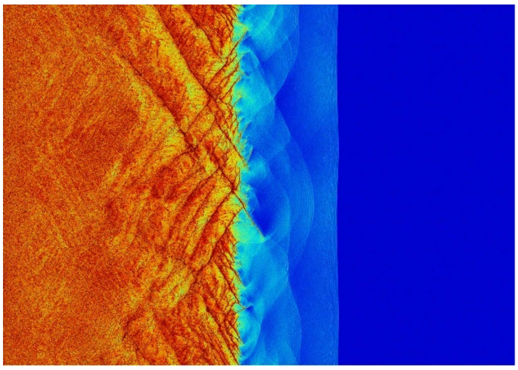

Figure 6. Simulation of split elastic–inelastic shock wave propagating in single crystal diamond (dark blue). The elastic precursor (light blue) is followed by an inelastic wave (red), which exhibits an unexpected stress relaxation mechanism. The SNAP potential in LAMMPS was used to simulate 1.8 billion SNAP carbon atoms, requiring several hours on the entire summit supercomputer (27900 GPUs). Reproduced with permission from [45c]. Copyright © 2021 ACM.

the elastically compressed material due to constructive interference of sound waves emanating from earlier cracks, as shown in figure 6. These large scale production simulations of shocked carbon could not have been performed without a robust MLP, an efficient implementation in LAMMPS, and access to extreme scale computing resources. While most science applications require just a few million atoms or less, they do require large timescales. Using massively parallel implementations of MLPs in LAMMPS, it is possible to achieve simulation rates of 10–20 ns/day, similar to the maximum rate achievable with less accurate empirical potentials like EAM, albeit using fewer nodes. With increasing proliferation of parallel computing resources, researchers can absorb the increased computational cost of MLPs by running on more nodes.

As is indicated in figure 5, the development of MLPs has been rapid. It has also been very heterogeneous, with many new ideas emerging each year, and a lot of hybridization between different approaches. The first wave of relatively simple descriptor-based kernel approaches (BP, GAP, MTP, SNAP, ACE) [44] have been followed by more complex graph-based message-passing neural network models (SchNet, NEquIP, MACE, Allegro, GrACE) [44]. Increased complexity provides better absolute training accuracy and may also capture more detailed physical and chemical interactions between atoms. For large-scale simulations, this increased complexity presents two distinct challenges: stability and performance. Stability is important because in production simulations (millions of atoms, millions of timesteps) even a single bad force prediction can propagate forward in time and space, causing inaccurate results or even a catastrophic failure (blackholing, lost atoms). The complexity of the potentials also presents obstacles to achieving good computational performance on large parallel platforms. The use of sophisticated software stacks, diverse libraries, multiple programming languages, hardware accelerators, etc., create difficulties with portability, accessibility, and performance. For example, if the force calculation requires redundant data movement between different hardware or software components, memory latency may severely limit the overall simulation rate. New models may also present new algorithmic challenges. For example, message-passing models require access to information from a hierarchy of neighbor shells, reducing the efficacy of spatial decomposition parallelization [48]. When implementing new MLPs in LAMMPS, previous implementations of similar MLPs should be carefully studied and the programmer guide should be reviewed [49].

Advances in science and technology to meet challenges

As the complexity of MLPs continues to grow, and their use becomes more widespread, models need to be intrinsically stable by construction. Alternatively, they need to detect and remedy instability without user intervention. Variations on existing descriptor-based kernel approaches may emerge in the first category. For example, certain subsets of atomic cluster expansion (ACE) [44] descriptors may provide good balance between stability and accuracy. In the second category, uncertainty quantification and active learning (AL) methods may lead to robust bootstrapping workflows. AL has been demonstrated to work well in particular applications, but much work is still required to make it sufficiently general for widespread use by non-experts.

Community codes like LAMMPS must continue to evolve to handle the high rate of innovation in both MLPs and hardware. New interatomic potential architectures and new hardware architectures should be accommodated without a lot of additional software development. The Kokkos [50] performance portability library has proven to be an effective abstraction layer that isolates LAMMPS from the complexity of the hardware. Similarly, the MLIAP package in LAMMPS has provided an abstraction layer that isolates LAMMPS from the complexity of the potential and vice versa. Currently, only one MLP (HIP-NN) [51] has successfully used both abstractions together. In the future, this will become more common and may require extensions or replacement of the MLIAP package, as well as extensions to the Kokkos library.

Concluding remarks

Over the last decade, the construction and use of machine learning-based interatomic potentials has gone from being a rather esoteric activity to an important skill for new practitioners of atomistic simulation. The challenge for the next decade is to develop robust methods that will allow non-experts to achieve large-scale high-accuracy simulations of specific materials under specific conditions without much human intervention. Historically, LAMMPS has served as a marketplace where users can quickly try out different potentials. It serves a similar role for MLPs. However, the software complexity of these models is far greater, stretching the limited time resources of both MLP developers and LAMMPS developers. Despite these challenges, it is essential that the barriers to adding new MLPs be kept low, so that the best ideas can be rapidly made available to users, thus continuing the high rate of innovation for the entire field.

Acknowledgments

This work was supported by the U.S. Department of Energy, Office of Fusion Energy Sciences (OFES) under Field Work Proposal Number 20-023149 and by the Exascale Computing Project (17-SC-20-SC), a collaborative effort of the U.S. Department of Energy Office of Science and the National Nuclear Security Administration.

This article has been authored by an employee of National Technology and Engineering Solutions of Sandia, LLC under Contract No. DE-NA0003525 with the U.S. Department of Energy (DOE). The employee owns all right, title and interest in and to the article and is solely responsible for its contents. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this article or allow others to do so, for United States Government purposes. The DOE will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan www.energy.gov/downloads/doe-public-access-plan.

6. Graph deep learning potentials for atomistic simulations

Tsz Wai Ko and Shyue Ping Ong
Aiiso Yufeng Li Family Department of Chemical and Nano Engineering, University of California, San Diego, 9500 Gilman Dr Mail Code #0448, La Jolla, CA 92093–0448, United States of America

Status

Graphs are a natural universal representation for a collection of atoms. As depicted in figure 7, atoms are represented as nodes, and the bonds between them are represented as edges. In some representations, there may also be additional global state features [52]. For machine learning (ML) purposes, each unique node/edge/global state is represented by a fixed-length vector. Information flows through the graph through sequential message passing, typically modeled using artificial neural networks. Such graph networks have become increasingly popular for the construction of machine learning interatomic potentials (MLPs) [38, 53–55].

Graph learning interatomic potentials (GLPs) have several advantages over MLP architectures based on local environment descriptors (simply referred to as MLPs). In MLPs, the number of basis functions scales combinatorially with the number of elements in the system. For systems with more than a few elements, this can result in a massive increase in training data and computational cost for MLPs. In contrast, GLPs represent each element using a unique, learned embedding vector, obviating the need for multielement interaction terms. This unique benefit of GLPs has enabled the recent development of 'universal' GLPs with coverage of the entire periodic table of the elements, such as the Materials 3-body graph network (M3GNet) [54], Crystal Hamiltonian Graph Neural Network (CHGNet) [55] and message-passing ACE (MACE) [53]. In addition, GLPs can model longer-range interactions beyond the graph construction cutoff radius through the number of message-passing steps. The main disadvantage of GLPs is that they are less compact in representing many-body interactions compared to MLPs.

GLPs can be broadly divided into two classes based on their symmetry awareness [56]. All GLPs respect invariance to translational, permutational, and rotational symmetry. Invariant GLPs (iGLPs) accomplish this through the use of scalar features such as atomic numbers and bond distances. However, purely distance-based graphs have been shown to be incomplete in their ability to represent different atomic environments [57]. More modern iGLPs typically incorporate higher-body terms [54, 55]. Equivariant GLPs (eGLPs), on the other hand, preserve the transformation of tensorial quantities under coordinate transformations through an appropriate choice of the message-passing functions. eGLPs directly use the relative positional vectors between atoms as input features, which is a more informative and efficient geometric representation. Extensive benchmarks such as 3BPA and MD17 have shown that eGLPs [38, 58, 59] generally outperform iGLPs in accuracy, data efficiency and transferability to unseen configurations in molecular systems, though the evidence on bulk materials is more mixed. The relative position vectors in eGLPs can be represented using spherical tensors [38, 58, 59], and more recently, using Cartesian tensors [60]. Generally, the Cartesian tensor representation offers greater computational efficiency, but at the expense of lower accuracy, though the recently developed Cartesian ACE (CACE) [60] has demonstrated an efficient construction of many-body message functions for constructing accurate eGLPs in Cartesian space.

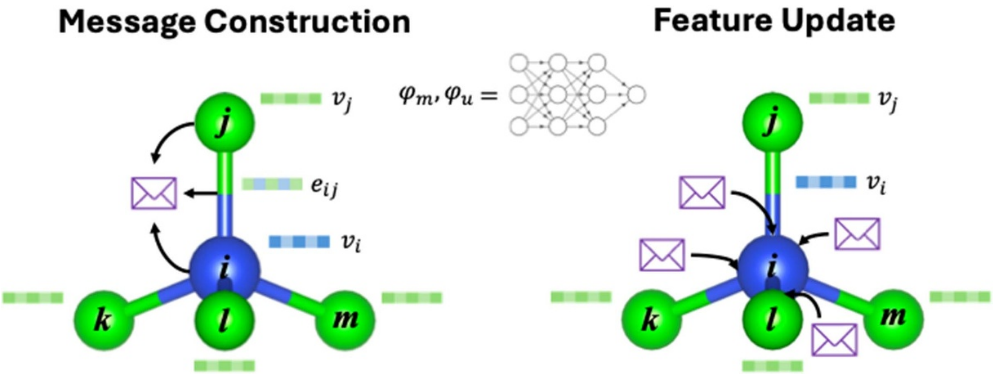

Figure 7. Message passing in graph interatomic potentials.

Current and future challenges

Similar MLPs, the primary challenge in the development of GLPs is the availability and quality of training data. Here, we will discuss this challenge in the context of *universal* GLPs, which is a unique capability of GLPs over MLPs. There are two main considerations in constructing a robust training dataset—the accuracy and cost of the method used to compute the PES and capturing a diversity of structures/local environments for the target system of interest. Both considerations are significantly compounded in the context of universal GLPs. As implicit in the name, universal GLPs typically cover nearly all elements of the periodic table and a large fraction of known crystal structures. Currently, all universal GLPs start from a training dataset extracted from the DFT relaxation trajectories from large open materials databases such as the Materials Project [61]. However, the workhorse semi-local generalized gradient approximation (GGA) functionals used in such databases have well-known limitations in their treatment of strongly correlated systems, hydrogen and van der Waals (vdW) bonding. While higher fidelity methods such as meta-GGAs [62], range-separated and screened hybrids [63] and CCSD(T) [64] can provide a more accurate description of the PES, they are prohibitively expensive for larger systems. To capture a diversity of local environments, the most common approach is to use some form of AL, whereby structures that extrapolate beyond the training domain are iteratively added to the training dataset. However, the AL procedure can be highly inefficient, especially for complex systems with large chemical and structural diversity [65].

A second significant shortcoming of GLPs is that they are approximately an order of magnitude less efficient than that of MLPs due to their more complex model architectures. Although graphical processing units (GPUs) can provide a significant acceleration, parallelization of GLPs is not trivial with the conventional domain decomposition method since the broader region for communications is dynamic due to message passing.

Finally, the design of GLPs to describe long-range interactions such as electrostatics, dispersion and magnetic interactions remains an open challenge.

Advances in science and technology to meet challenges

The above challenges can be addressed through several coordinated efforts. First, efforts are already underway to develop robust datasets for universal GLPs using high-throughput DFT computations. Unlike existing datasets, these datasets should contain well-converged PES information (energies and forces). Second, to address the computational expense associated with higher fidelity methods, one promising approach is to leverage large, lower fidelity datasets based on inexpensive *ab initio* methods to reduce the amount of high-fidelity computations needed, for instance, through delta learning and multi-fidelity models [66]. Finally, improved sampling techniques can be developed to reduce the need for inefficient AL [65]. Effective sampling techniques to ensure coverage of rare events such as transition states would also need to be developed. Interestingly, some of these efforts can be significantly accelerated using existing universal GLPs. For instance, existing universal GLPs can be used to perform multi-temperature MD simulations to rapidly generate a comprehensive configuration space [65].

To handle long-range interactions, new GLP architectures can take inspiration from recent MLP developments such as the 4G-HDNNPs [40] which incorporate long-range electrostatics based on globally distributed charges governed by charge equilibration scheme without requiring additional message-passing steps, and spin-DimeNet++ and spin-Allegro [67] which incorporate magnetism.

Finally, a few graph libraries such as Pytorch-Geometric [68] and Deep Graph Library [69] have been implemented to optimize the efficiency and scalability of graph constructions and operations. Additionally, several packages have developed new parallelization strategies to enable large-scale GLP-driven MD simulations with multi-GPU support.

Concluding remarks

There has been tremendous progress in the development of GLPs in recent years. With their unique ability to handle complex chemical and structural spaces, GLPs have broad applications in materials discovery and dynamic property predictions. Indeed, sufficiently accurate universal GLPs can serve as foundational models, upon which more accurate customized MLPs can be constructed much more efficiently. While long-range interactions and computational efficiency remain significant challenges, we are optimistic that continued advances in architectures, data generation and algorithmic improvements can mitigate many of the current shortcomings of GLPs.

Acknowledgments

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.

7. MTPs

Olga Chalykh[1]*, Dmitry Korogod*[2]*, Evgeny Podryabinkin*[1] *and Alexander Shapeev*[1]
[1] Skolkovo Institute of Science and Technology, Russia; Bolshoy Boulevard 30, bld. 1. Skolkovo, 121205, Russia
[2] Moscow Institute of Physics and Technology, Russia; Institutsky lane 9, Dolgoprudny, Moscow region, 141700, Russia

Status

MTPs [18] have been developed as the first example of machine-learned potentials whose functional form is sufficiently flexible to approximate any local functional of atomic positions. Together with the D-optimality-based AL algorithm [11, 70], an interatomic potential can be constructed automatically [71], leading to seamless acceleration of quantum-mechanical calculations as a part of solving a problem at hand, e.g. finding the convex hull of stable structures [70]. In this context, it is worth noting that, for many problems, machine-learned potentials (MLPs) approximate DFT so closely that the overall accuracy, when compared to experimental results, is ultimately limited by the accuracy of the density functional itself [71].

In a 2020 study [72], MTPs showed a favorable balance between the error and the CPU computing time compared to other potentials existing at the time of the study. Belonging to the class of polynomial-based potentials, which approximate the energy as a polynomial-like function of atomic positions (adjusted to have a finite cutoff sphere), MTPs have the advantage of performing fast on CPUs and requiring a relatively small training set compared to Neural-Network-based potentials.

Current and future challenges

The challenges overcoming which would advance the capabilities of MTPs are hence:

GPU implementation

While neural networks typically involve learning parameters of large matrices, the multiplication of which is inefficient on CPUs, readily available GPU implementations of NNs make them computationally competitive on modern architectures. To make polynomial-based models effective on GPUs, one needs to code the existing models (or perhaps even develop new models) in a way that many computing threads can process one atomic neighborhood.

Long-range interactions

While all interatomic potentials are generally effective at capturing short-range interactions (such as covalent/metallic or even well-screened ionic bonds), learning long-range interactions poses challenges because the phase space grows rapidly as the interaction cutoff sphere increases. Alternative approaches, potentially more explicit than merely learning a general form of interaction, should be explored to incorporate long-range interactions.

AL on large-scale simulations

AL approaches assume that an atomistic configuration occurring in a simulation can be ‘plugged’ into a quantum-mechanical code to produce training data. However, in a large simulation of surfaces or extended defects, this is clearly not feasible due to the large size of the required supercell and the superlinear increase in complexity of typical quantum-mechanical calculations. Therefore, approaches to learn from fragments of simulations should be further developed. This would enable modeling processes such as atomic layer deposition, catalysis, and the behavior of dislocations and grain boundaries.

Incorporating experimental data

Since MLPs have been shown to achieve higher accuracy than DFT itself, it raises the question of incorporating experimental data and training MLPs to surpass DFT. For instance, it has long been known that the melting points of metals can deviate by hundreds of Kelvin when computed using DFT [73]. A common workaround is to calculate melting points at high pressures (several GPa) to predict a more accurate ambient-condition melting point. This approach compensates for the underbinding in metals inherent to the GGA approximation. The question arises: can we utilize experimentally known parameters such as density to refine MLPs and enhance their accuracy?

Advances in science and technology to meet challenges

GPU implementation

Tools like PyTorch or TensorFlow, which accelerate expensive linear algebra operations on GPUs, were developed thanks to the large community working on neural networks. In order for MTPs, as well as similar potentials like SNAP and ACE, to be efficient on GPUs, similar libraries need to be created in this field. As of now, we are not aware of any efforts in this regard.

Long-range interactions

Several approaches to overcome the inherent short-sightedness of MLPs have been proposed in the literature, including the incorporation of explicit Coulomb and vdW) terms [74]. Within the latter approach, one can include terms corresponding to the electrostatic interaction of either point or Gaussian charges. The charges can be either fixed, environment-dependent, or obtained using a charge equilibration (QEq) method, allowing for global charge redistribution, with the latter being the most accurate. As for the vdW interaction, the most common strategy is to model it as the sum of an existing vdW interaction model (e.g. the D3 dispersion correction), and a general-purpose MTP that is fitted to reproduce the short-range (e.g. covalent) interactions.

AL on large-scale simulations

AL on fragments of large-scale simulations requires two components: (1) the detection of extrapolative fragments, and (2) the reconstruction of training configurations from these fragments. The first problem is solved by the neighborhood version of the D-optimality criterion [75]. A solution for the second problem is not yet at the ‘automatic seamless DFT acceleration’ level, however, the first steps are made. The solution consists of intelligently completing the extrapolative fragment to a periodic configuration [19] or a semi-periodic structure or a non-periodic cluster [75] depending on the structure of the large-scale simulation and the fragment’s location within it.

Incorporating experimental data

Recently, the concept of differentiable simulations has been developed, allowing for the incorporation of measures of simulation proximity (e.g. the radial distribution function) to experimental data and thus facilitating potential fitting. An example of fitting simple two-body potentials has recently been demonstrated [76]. This approach could evolve into a method for fine-tuning potentials fitted to large amounts of *ab initio* data to match existing (typically scarce) experimental data.

Concluding remarks

MTPs have provided important advancements in materials science by offering an automated and accurate approach to accelerate quantum-mechanical calculations. Despite the growing utility of machine-learned potentials (MLPs), challenges remain in optimizing their performance. Key areas for improvement include efficient GPU implementation, addressing long-range interactions, and refining AL methodologies for large-scale simulations. To overcome these challenges, advancements in technology and methodology are essential. This includes the creation of specialized libraries for GPU acceleration, incorporating explicit Coulomb and vdW terms to enhance MLP accuracy, and refining methods for AL on large-scale simulations. Furthermore, the integration of experimental data into MLP training processes, facilitated by concepts such as differentiable simulations, holds promise for further enhancing the accuracy and applicability of machine-learned interatomic potentials.

Acknowledgments

This work was supported by the Russian Science Foundation (Grant No. 23-13-00332, https://rscf.ru/project/23-13-00332/).

8. UNIP

Ju Li
Department of Nuclear Science and Engineering and Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA 02139, United States of America

Status

Given the rapid progress in machine-learning interatomic potentials, tackling the full chemical complexities of materials in atomistic modeling becomes the next imperative. The historical success of DFT can be largely attributed to its chemical transferability, that is, once the pseudopotentials are chosen (say for V, O, C, Pt), most atomic combinations of these elements from zero-valent metallic alloys to oxides can be modeled in DFT with a decent expected error, typically within $\sim$100 meV atom^{-1} in the formation enthalpy compared to the experimental value. However, because on-the-fly iterative electronic relaxation is required, DFT computation tends to have poor size scaling. Even for so-called order-N DFT methods, the prefactor is quite large, or the error is quite a bit larger than basis-converged DFT calculations.

Recently, the so-called UNIPs that aim to cover the entire periodic table without restraint [77] (figure 8) have been developed. The tensor embedded atom network (TeaNet) was able to cover the first 18 elements (from H to Ar) of the periodic table in 2019. TeaNet architecture is based on passing Cartesian tensors (scalars, vectors, rank-2 tensors and beyond) on graph neural networks (GNNs) from atoms to bonds, and bonds to atoms, where simply using the Einstein notation for tensors (contraction or direct product) guarantees rotational equivariance. A key feature of training TeaNet is that there were no pre-conceived chemical or structural knowledge, as if the materials world were just starting primordial gases. See table 1, where all types of chemical bonds were invoked in generating the DFT training set, including between Ar–Cl, Ar–H, Ar–Na, etc, which hardly exist in crystal structure databases. In 2022, Preferred Networks Inc. and ENEOS Corp. adapted the TeaNet architecture [78] into the commercial PreFerred Potential (PFP), which covers 45 elements on the periodic table [79]. In the summer of 2024, PFP officially supports all 96 long-lived elements on the periodic table, from Hydrogen to Curium. This marks the completion of a true UNIP at production scale that covers arbitrary materials chemistries. Open-source UNIPs, e.g. M3GNet [54], CHGNet [55] and MACE [80], have been developed and are also becoming widely used.

The popularization of UNIPs is poised to be the biggest revolution in materials modeling in the coming years, because it offers a great compromise between computational efficiency and accuracy for a very wide range of materials problems. While the current generation of UNIPs still has much to be desired, they represent a step advance in coverage, accuracy, and reliability compared to the empirical potentials we had previously.

Current and future challenges

The training of UNIP is very costly. The development of PFP v6.0 has utilized 42 million DFT configurations as the training dataset, generated by the active-learning algorithm [81], with a combined 2000 GPU years (2 GPU millennia) of computations. Thanks to the extraordinary expressive power of deep neural networks, UNIP can indeed handle the challenge of the

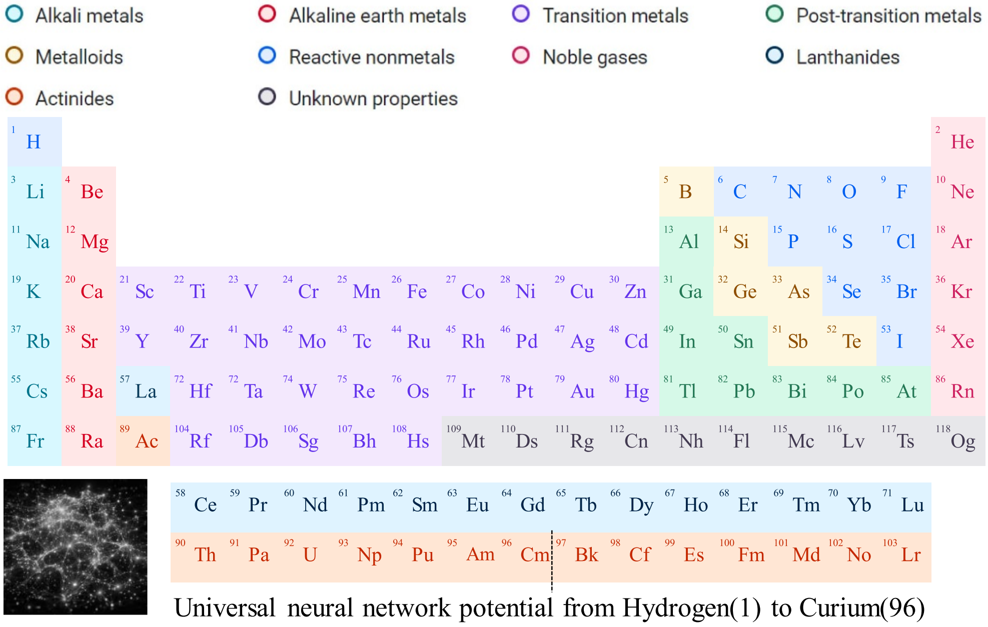

Figure 8. Universal neural interatomic potentials (UNIP) such as PFP [77] aims to cover the entire periodic table of elements, with no limitations on what types of chemistries or configurations (metals, oxides, molecules etc) can be simulated, including chemical reaction barriers.

periodic table, as long as sufficient *ab initio* training data are provided. The prediction errors of UNIP formation enthalpy compared to DFT 'ground truth' are typically on the order of $\sim$20 meV atom^{-1} for stable crystals and only few meV/atom for organic molecules, even for chemically complex configurations, which is an order of magnitude smaller than the DFT-to-experiment error of $\sim$100 meV atom^{-1}. In other words, UNIP indeed accomplishes its primary goal of representing the DFT energy landscape *accurately enough*, while reducing the computational cost.

Because *ab initio* calculations have been used to generate the UNIP training data, getting more accurate (compared to experiment) *ab initio* results is the main avenue to more accurate and transferrable UNIP. Since error cancelation is key to the success of DFT models, how to maintain systematically uniform DFT errors (planewave energy cutoff, **k**-point mesh, etc) for a wide variety of configurations such as molecules, metals, clusters, etc is important. In terms of absolute energies, DFT can have significant errors, especially when the coordination number changes drastically or when strong electron localization and correlation occurs, for example, in generating the metal–oxide formation energies in reference to the O_2 gaseous state [82]. Methods such as DFT + U have been developed, but U is an empirical adjustable parameter that introduces uncertainty in the systematic errors. Even with better density functionals such as SCAN for which the magnitude of U can be bounded smaller than for LDA or GGA [82], this can still introduce significant uncertainty to the training dataset. Also, for problems involving magnetism, radicals, or multi reference character [83], DFT can have large uncontrolled errors,

Table 1. The frequency of occurrences of different nearest-neighbor chemical bond types in the DFT training data set of TeaNet [78].

	H	He	Li	Be	B	C	N	O	F	Ne	Na	Mg	Al	Si	P	S	Cl	Ar
H	981	403	622	300	277	916	684	1002	564	535	438	327	537	341	741	401	469	372
He		455	582	554	518	594	601	527	401	637	384	598	586	477	500	780	539	501
Li			2091	522	705	543	697	1313	662	676	509	902	741	802	502	432	779	460
Be				1208	671	766	929	590	401	606	414	411	621	1391	283	491	385	442
B					3324	505	739	3514	659	886	459	474	625	824	572	751	448	551
C						1333	681	538	393	554	477	599	841	460	396	868	676	490
N							1068	759	536	927	825	586	525	678	576	433	622	416
O								2745	387	758	404	520	625	462	505	414	506	455
F									815	499	520	369	402	415	488	519	503	372
Ne										1201	717	588	668	621	778	350	498	553
Na											741	451	374	393	295	346	527	321
Mg												917	468	489	302	482	630	383
Al													1515	516	475	599	1106	510
Si														1098	297	412	889	451
P															1155	716	408	342
S																1600	738	464
Cl																	1413	469
Ar																		421

which would lead to UNIPs giving unreasonable results. In these cases, it might be helpful to develop machine-learning 'flags of unreliability' [83].

Another issue is the computational complexity in runtime. PFP v6.0 is much cheaper than planewave DFT calculations, especially when the number of atoms is large, as it is an order-N method. Still UNIPs are much more expensive to run than standard EAM potential calculations, which can easily handle tens of millions of atoms on a workstation. The UNIPs of today (year 2024) can typically handle tens of thousands of atoms, which, while much bigger than typical DFT calculation, is still small for some atomistic problems.

Advances in science and technology to meet challenges

Because of the well-known issues with DFT, researchers are also generating *ab initio* energetics based on quantum chemical coupled-cluster theory with single, double, and perturbative triple excitations (CCSD(T)), that may be used to train UNIPs in the future. CCSD(T) is widely regarded as the gold standard of computational chemistry, therefore, to the extent it can converge and be accessible, the results should be definitive for UNIP training. Recently, there has been effort [84] to use equivariant neural networks to accurately represent CCSD(T) energetics with DFT-level computational cost, which may be leveraged to generate the beyond-DFT training dataset.

We note that the original motivation for TeaNet architecture [78] was to mimic electronic-structure relaxations in a self-consistent tight-binding calculation. In self-consistent tight-binding calculations, the total energy is not the only output; the quantum wavefunctions, charge density distribution, and electronic transport properties can also be derived. Thus, PFP (and probably all future UNIPs) can also reproduce charge distribution, electrostatic potential, electronic transport properties, etc, along with the total energy, which were simply not available with traditional empirical interatomic potential.

To mitigate the runtime computational complexity problem for large-scale atomistic simulations (1 million atoms or above), the so-called LightPFP is developed. LightPFP takes the form of MTP developed by Shapeev, but is trained rapidly with PFP-generated energetics. This approach presents a huge practical advantage because generating chemistry-specific training data for the fitting of LightPFP takes only a few hours, in contrast to developing DFT-based MTP which usually takes weeks or months. Such 'overnight' training of machine-learning interatomic potential based on UNIP-generated data allows a great number of chemistries to be explored, and in runtime, the use of MTP would easily allow millions of atoms to be simulated with near-DFT accuracy.

Currently, PFP (Matlantis$^{\mathrm{TM}}$) has already been applied in a wide range of materials, including semiconductor materials, high-entropy alloys, battery electrolytes and electrodes, catalysts, lubricants, metal–organic frameworks, nuclear materials, geological minerals, etc, and is finding new applications rapidly.

Concluding remarks

The development of Universal Interatomic Potentials signifies a great step advance in the atomistic modeling field. No longer is one constrained by the limited chemistries with the empirical potential approach before the 2010s. Now, researchers can simulate complex chemistries with DFT-level accuracy, or even beyond, at a small fraction of the cost in runtime. This allows one to simulate curved extended defects and their reactions, heterogeneous nucleation, electrocatalysis at complex interfaces, etc with no constraint on the material chemistries.

Acknowledgments

J L acknowledges support by DTRA (Award No. HDTRA1-20-2-0002) Interaction of Ionizing Radiation with Matter (IIRM) University Research Alliance (URA), as well as extremely helpful discussions with Dr So Takamoto and Dr Daisuke Okanohara.

9. Physically informed machine-learning potential for atomistic simulations

Yuri Mishin
Department of Physics and Astronomy, George Mason University, 4400 University Drive, MSN 3F3, Fairfax, VA 22 030, United States of America

Status

Modern atomistic simulations of materials utilize two different approaches to energy and force calculations: traditional interatomic potentials and the relatively new class of machine learning (ML) potentials [2]. These approaches can be classified as model-based and data-based, respectively. Traditional potentials are based on physical models of interatomic bonding. They are computationally fast but have limited accuracy due to the small ($\sim$10) number of fitting parameters and the approximate character of the underlying physical models. Developing a reliable general-purpose potential is a long process heavily relying on human expertise. However, transferability to unknown atomic configurations is often reasonable (even if not very accurate) as long as the underlying model remains physically meaningful (figure 9(a)). By contrast, ML potentials do not rely on any physics specific to the system other than smoothness, locality, and energy invariance under translations and rotations. Instead, the energy and forces are predicted by purely numerical interpolation between reference data points (energy, forces, and often stresses) generated by quantum-mechanical (usually DFT) calculations. The interpolation is implemented by a high-dimensional regression containing a large number (usually thousands) of adjustable parameters, which are fitted to a large (often 10^6 atoms) DFT database (figures 9(b) and (d)). DFT-level accuracy of training and testing can be achieved, and MD simulations can be run orders of magnitude faster than *ab initio* MD. On the downside, the ML regression is a black box with uninterpretable parameters. More importantly, the regression only interpolates but is unable to provide physically meaningful extrapolation. Predictions of physical properties outside the interpolation domain are based on a purely mathematical algorithm and can give unpredictable and often physically meaningless results (figure 9(b)).

The strengths and weaknesses of the traditional and ML potentials are complementary. A new direction has recently emerged, aiming to take the best from both worlds. This can be achieved by choosing a physics-based interatomic potential and letting an ML regression predict its parameters according to each atom's local environment. Because the regression's output is 'piped' through a physics-based model, the extrapolation is guided by the physics underlying the potential and can be more meaningful than the purely mathematical extrapolation (figures 9(c) and (e)). For example, the physically informed neural network (PINN) potentials connect a neural network (NN) to an analytical bond order potential (BOP) [85]. The recently developed general-purpose PINN potentials for Al [86] and Ta [87] accurately describe a broad spectrum of properties of these metals, including the mechanical and thermal properties most relevant to materials science applications.

Current and future challenges

Despite the recent successes in developing and applying ML, their broader applications are hampered by the lack of transferability outside the training domain of atomic configurations. Any physically meaningful extrapolation requires a physical model. A database, no matter

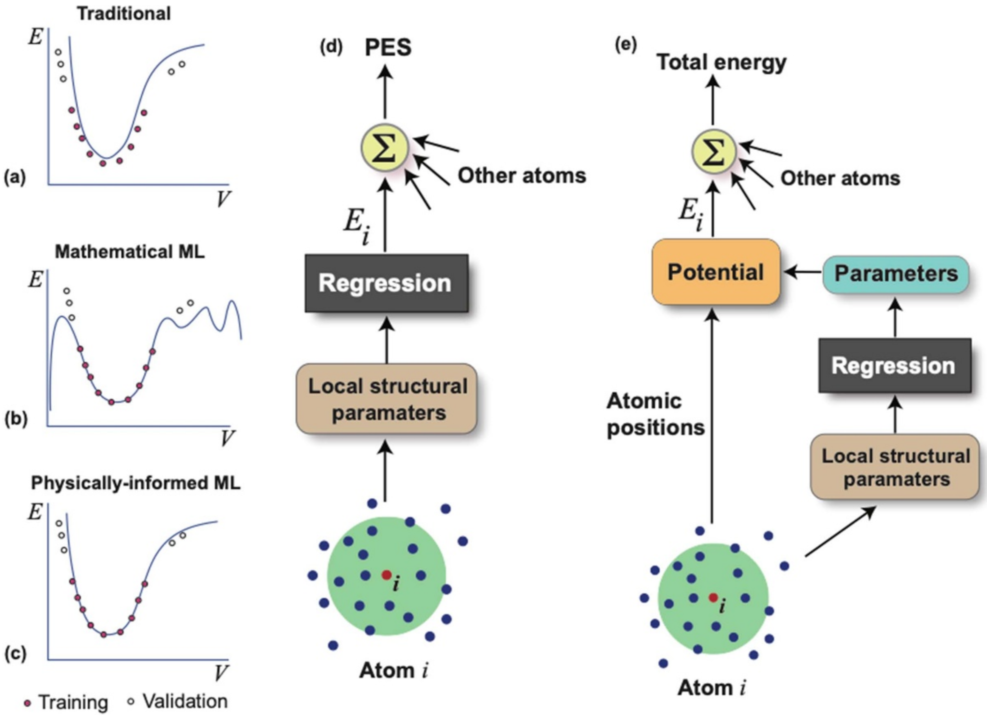

Figure 9. (a)–(c) Schematic illustration of accuracy and transferability of (a) traditional, (b) mathematical ML, and (c) physically informed ML interatomic potentials. (d), (e) Flowcharts of total energy calculations with (d) mathematical ML potentials and (e) physically informed ML potentials. Reprinted from [2], Copyright 2021, with permission from Elsevier.

how big, cannot replace a model. An ML regression, no matter how sophisticated, only knows the long string of numbers fed into it from the reference database. The only way to ensure that extrapolation to unknown configurations makes physical sense is to ‘teach’ the ML potential some basic laws of physics.

Incorporating physics into ML potentials has been explored by several authors [2, 24, 85–89]. For example, recent versions of GAP potentials include analytical functions representing short-range pairwise repulsion or vdW interactions [90]. NN potentials inspired by the EAM have been proposed [91]. Other explorations include the hybrid potentials [31] and spline-based NN potentials [30]. In the PINN model, the BOP parameters are locally adjusted during the simulation in response to the changing local atomic environments. The existing PINN potentials [2, 85–87] demonstrate reasonable transferability to atomic environments unseen during the training and testing. Benchmarking against straight NN potentials shows that incorporating the BOP only increases the computational overhead by 25%.

The BOP potential embedded into the PINN formalism has the advantage of being applicable to both metallic and covalent materials. Despite the small number of fitting parameters, it captures some of the key features of interatomic bonding, such as the many-body character of interactions, the bond-order effect, and the bond screening by neighbors. However, in the future, the BOP can be replaced by a simpler but faster potential, such as an EAM potential for metallic systems. Another regression algorithm can replace the NN, and the local structural descriptors can be modified.

Future research should address the following problem common to both the mathematical and physically informed ML potentials. Traditional potentials developed for chemical elements can be included in a binary (or multi-component) potential by only fitting the cross-interaction functions. In some cases, both self-interactions and cross-interactions are fitted simultaneously, but inheriting the elemental potentials is the most typical and practical approach. By contrast, multicomponent ML potentials are incapable of inheriting elemental potentials. This leads to a proliferation of different potentials for the same element, depending on whether it was constructed as a stand-alone potential or part of binaries or ternaries. The inheritance of elemental potentials is a highly desirable feature. It is worth exploring if the architectures of ML potentials could be modified to enable the inheritance feature. This would help avoid duplication of potentials and facilitate their standardization and organization in repositories.

Advances in science and technology to meet challenges

Atomic-scale computer simulations are an effective tool for gaining fundamental insights into the microscopic mechanisms of many processes and providing quantitative input to mesoscale and continuum models. Such simulations heavily rely on the quality of classical interatomic potentials, which constitute the critical ingredient that often determines the validity of the simulation results. The ‘quality’ of the potential has three main components: accuracy, reliability, and computational efficiency. ML potentials easily outperform the traditional potentials in accuracy but are significantly slower. Generating a DFT database also requires substantial computational efforts. Both issues become less of a problem as computers become faster and new and more efficient algorithms for training, feature engineering, and reference database assembly are developed. The reliability is a more challenging problem. It largely reduces the potential’s ability to predict accurate, or at least physically meaningful, results when the simulation trajectory wanders away from the domain of configurations represented in the reference database. The purely numerical extrapolation to such configurations implemented by today’s ML potentials produces uncontrollable and often nonsensical results. The critical challenge for future research is to improve the transferability of ML potentials. This goal can be achieved by integrating the ML engines with physics-based models. Work in this direction is already going on. The PINN potentials give one example. A multicomponent PINN format has been developed [2] and can be used to generate alloy potentials in the future. The method could be generalized to include long-range forces and magnetism. More importantly, the PINN example can motivate the development of other and possibly more advanced physics-based atomic interaction models in the future.

Concluding remarks

The strengths and weaknesses of the traditional and ML potentials complement each other. Combining them into physically informed ML models can improve their transferability to unknown environments while maintaining the DFT-level accuracy during the training and testing. This, in turn, can significantly improve the reliability of potential-based atomistic simulations. Every model eventually fails. However, the extrapolation guided by a physics-based model can expand the reliability domain of ML potentials in the configuration space of the material.

Acknowledgments

This research was supported by the National Science Foundation, Division of Materials Research, under Award No. DMR-2103431.

10. Machine learning-based potentials for medium and HEAs

Zongrui Pei[1] *and Xianglin Liu*[2]
[1] New York University, New York, NY10012, United States of America
[2] Peng Cheng Laboratory, Shenzhen 518066, People's Republic of China

Status

Interatomic potentials describing the interactions between atoms are the cornerstone of materials simulation methods such as MD or Monte Carlo (MC) simulations. Accurate interatomic potentials can be constructed by solving the quantum-mechanical equations governing the motion of electrons. Still, such a method quickly becomes impractical due to the rapid increase of computational cost with respect to the number of atoms, even after applying approximations with DFT methods. The computational cost would be even higher if atomistic thermodynamics needed to be incorporated to predict physical properties at finite temperatures. On the other hand, a large number of atoms are necessary for the study of a range of essential materials science problems, such as dislocations, nano-precipitates, and order-disorder transitions, all of which are crucial for understanding the origin of the exceptional mechanical properties in medium and HEAs. Various empirical interatomic potentials were proposed to increase atomistic simulation efficiency, such as Lennard–Jones potential and EAM potential. These traditional empirical potentials are highly efficient for computation and, therefore, can easily simulate millions or billions of atoms. Nevertheless, compared to first-principles methods, empirical potentials typically require extra tuning efforts for unknown or complex materials, and their accuracy can also be compromised due to their limited functional form.

The rapid development of machine learning interatomic potentials (MLIP) sheds light on the problem of simulating medium and HEAs with high accuracy and efficiency [92]. Compared to the traditional empirical potentials, MLIPs take much more flexible functional forms and, therefore, can describe more complex interatomic interactions. Moreover, the large number of parameters in MLIP can be automatically determined by fitting datasets with machine learning methods. In general, MLIP is computationally very efficient compared to DFT since there is no need to solve the quantum-mechanical equations of the electrons in the simulation stage. Overall, MLIPs can reach an accuracy close to the DFT training datasets but with orders of magnitudes higher efficiency. Due to their complex functional form, most ML potentials still show slower computational speed than conventional empirical potentials for energy/force evaluation, typically by one or two orders of magnitudes. It should also be noted that the exact speed-up ratio in actual simulation depends on many factors, such as the ML model architecture, the size of the simulation system, computing hardware, as well as the computing characteristics (e.g. computing or memory-access boundedness) of the atomistic simulation code.

Various MLIPs are proposed based on different machine-learning methods and physics-embedding schemes. For example, PhysNet adopts a neural-network structure using features based on the local environment of each atom [93]. Similarly, Zhang *et al* developed deep neural-network potentials, or, in short, deep potentials [44]. They are based on the local environment, like distance, angles, etc. Also, some ML potentials are not based on neural networks, such as GAP, which adopts the Gaussian process [94]. Other empirical potentials with machine-learning flavors include ACE [95], MTP [18], and equivariant GNNs [38]. Recently, with the emergence of foundation language models, there have been foundation models for potentials [96]. Assisted by such models, there is no need to find potentials for each chemical element since all commonly used chemical elements are included in a single model.

Since 2004, studying medium and HEAs has gradually become one of the central topics in the materials science community [97]. Given the multicomponent features of these materials, modeling and simulations usually require the average over multiple configurations to calculate their physical properties. This feature indicates computationally more expensive calculations with large system sizes than ordered and few-component materials, which are more challenging using even empirical potentials. The main challenge lies in constructing an accurate potential for the multiple components. ML-based potentials for medium to HEAs include the CoCrFeNiMn alloy and NbMoTaW [98]. For example, Choi *et al* [99] obtained the melting point, generalized stacking fault energy, and stress-strain curve of MoNbTaW refractory HEA through hybrid MC /MD simulations with a system of 36 000 atoms using the SNAP potential. They also investigated the influence of short-range ordering on dislocation migration rates using the MTP method in a larger supercell containing 573 672 atoms [100].

Current and future challenges

Constructing accurate ML potentials for medium and HEAs is still challenging. The challenges can be roughly grouped into (i) the requirement of relatively large datasets, (ii) the transferability of potentials, and (iii) the development of massively parallel simulation software for MLIPs. For example, for a five-component alloy, there are ten pairs of chemical elements or ten binary systems whose phase diagrams must be taken care of. The enhanced chemical complexity demands a large amount of training data to reach satisfactory accuracy. The larger dataset size not only increases the technical difficulty of fitting the MLIPs but, more importantly, also increases the computational cost of constructing the training data, which is typically obtained by evaluating the energies and force tensors with high-accuracy first-principles methods such as plane-wave DFT. In other words, the dataset should be large enough for the MLIPs to accurately describe medium and HEAs but not too large to compromise the efficiency advantage of MLIPs. Therefore, accuracy and computational efficiency must be compromised for all empirical potentials based on traditional methods (e.g. EAM) and machine-learning potentials. One strategy to overcome or balance this tradeoff is AL, in which the simulations are monitored with a query policy to determine if the MLIPs are accurate enough at the sampled phase space and whether direct evaluation with DFT is necessary.

Another critical issue is transferability. As a data-driven approach, the MLIPs designed for specific purposes may not be transferrable for different tasks. There is a trade-off or competition between the accuracies of different properties. This is mainly due to the approximations of the ML ansatzes, which may not capture the essential complexity of the real potentials. This argument is not complex to understand when we compare the small number of input variables in empirical potentials and those in the Schrödinger equation [i.e. $3N$ variables with N as the number of electrons in the system]. Usually, it is easier to fit energies than forces. Due to this feature, empirical potentials are more favorable to studying thermodynamic or kinetic properties than mechanical properties, such as modeling dislocation core structure and the associated mechanical properties (e.g. dislocation mobility and critical resolved shear stress). In these applications, the stress field and mechanical response are inherently long-range and larger systems are needed to describe them. Additionally, server local distortion can seriously affect the accuracy, making it challenging for MLIPs with relatively short radial cutoffs. The key to enhancing the transferability of MLIPs is the potential to capture the underlying physical information of atomic species. Since machine learning is better at interpolation than extrapolation, ensuring the training data are diverse enough is essential. For example, potentials intended to study the glass transition of alloys may need to be more suitable for studying mechanical properties. Poor transferability is a common feature of empirical potentials. It is always recommended to check if potentials are ideal for specific tasks by comparing their output with first-principles calculations or available experimental data.

The third challenge is developing high-performance simulation software using MLIPs. For MLIPs to shine in real applications, they must be able to simulate at a large scale. This requires porting MLIPs into highly parallel MC/MD simulation codes and optimizing them to fully harness the power of current computing hardware, such as GPUs. Considering the rapid development of MLIPs, these packages need to constantly incorporate new MLIP implementations, which can be written with different programming languages. Moreover, compared to the deep learning models for images or natural languages, MLIPs are small in terms of the number of parameters and typically involve a large number of irregular memory accesses to construct the input features, making optimizing the computing performance challenging.

Advances in science and technology to meet challenges

The advance of computing hardware plays a critical role in meeting the challenges confronted by ML empirical potentials. Theoretically, we do not need empirical potentials for medium and HEAs if we have infinitely powerful computers to solve the Schrödinger equation. However, our supercomputers are still insufficient to solve the Schrödinger equation of multicomponent alloys with actual microstructure (dislocations, twins, grain boundaries, etc). This reality is the fundamental reason for the existence of ML and other empirical potentials. Nonetheless, high-quality ML potentials are still computationally expensive in simulations. Supercomputers have become increasingly powerful, and computational resources have become inexpensive, allowing us to compute more complicated systems more accurately. The supercomputer advances can significantly mitigate most challenges, including lowering computational costs and data storage. In particular, the advance of the graphics processing unit (GPU) and other accelerator technology substantially reduces the time to perform the matrix operations, a bottleneck in many algorithms when calculating atomic interactions, including NNPs.

In addition to the hardware, new algorithms/methods of ML techniques that can fully harvest the hardware power are also crucial. The development of new machine-learning algorithms is very dynamic and fast-evolving. New algorithms can construct more accurate potentials with comparable computational costs. The emergence of GAP, PhysNet, ACE [95], GNNs, and other ML potentials already address substantial challenges confronted by classical empirical potentials for multicomponent alloys, with their enhancement in accuracy and/or efficiency.

When designing general-purpose potentials for multicomponent alloys, the challenges become even more serious than potentials for specific applications (e.g. for dislocations in CoCrNi). In the future, we expect general foundation models to appear that include all chemical elements and generally apply to multiple situations (e.g. mechanical, chemical, and other problems). Like the successful foundation models for natural languages, the foundation models for empirical potentials may be based on the transformer framework. Considering the development in quantum computing, it would also be interesting to construct MLIPs based on parameterized quantum circuits and utilize quantum computing to efficiently capture the complex latent space of chemical interactions in HEAs.

Machine learning potentials are data-driven. The most important thing to increase the data efficiency is to encode physical constraints into MLIPs. These physical constraints can remove unnecessary degrees of freedom in the model, substantially reducing the training dataset size. One well-known example is E(3) equivariant GNNs [38], which demonstrate high data efficiency and use 1000x less data than DeepMD. Another strategy is creating a standardized dataset for researchers to reuse for different tasks. A well-known example is the Open Catalyst dataset. High-quality data is also essential to constructing high-quality potentials, and new experimental data can be used to design high-quality ML potentials. Other integrated computational materials engineering methods, like the Calculation of Phase Diagram (Calphad), are also helpful in validating ML potential quality, and data generated from them can help improve ML potentials, too. Finally, for some simplified applications, interatomic potentials, such as on-lattice models, can be beneficial in harnessing high symmetry to design models focusing on chemical complexity. This reduced model complexity can significantly reduce the demand for training datasets.

To increase the transferability, one possible solution is to develop a general-purpose neural network. For instance, Song *et al* developed general-purpose MLIPs for 16 elemental metals and their diverse alloys. They used them to study the plasticity and primary radiation damage in the refractory high-entropy alloy MoTaVW [101]. While it generally requires an even more significant amount of training data, these general-purpose models promise to serve as foundation models for downstream tasks, probably with only a small number of additional data needed to fine-tune for various tasks.

A few strategies can be used for high-performance simulations. For instance, we can port MLIPs into high-performance simulation packages such as LAMMPS, which is widely used MD software well-tested on various HPC platforms. On the other hand, we can utilize optimized packages such as PyTorch geometric to construct MLIP. These packages use highly efficient math libraries, which can harness the computation power of various accelerators.

Concluding remarks

Empirical potentials have played a critical role in studying the material properties of medium and HEAs. We are in the AI era. AI and ML potentials will show more power in helping design high-quality ML potentials, greatly facilitating the simulations and modeling of multicomponent alloys. ML potentials allow us to study problems in medium- and high-entropy alloys that were impossible before regarding accuracy, efficiency, and system sizes, such as studying nano-precipitates, dislocation, and interaction between significant defects. With the advance of computing hardware technology and new ML algorithms, ML potentials will become even more critical, which may allow us to design materials automatically based on AI and ML potential without much human involvement. Encouraged by the success of large language foundation models, one attractive direction is constructing transferrable foundation models for different applications, overcoming the non-transferrable drawbacks shared by all empirical potentials for complex, concentrated alloys and other materials nowadays.

11. Pretrained UMLP

Jaesun Kim[1], *Yutack Park*[1], *Seungwoo Hwang*[1] *and Seungwu Han*[1,2]
[1] Department of Materials Science and Engineering, Seoul National University, Seoul 08826, Republic of Korea
[2] Korea Institute for Advanced Study, Seoul 02455, Republic of Korea

Status

Recently, pretrained UMLPs [48, 54, 55, 79, 96, 102, 103] have garnered attention as they can overcome the long development period of traditional MLPs, which results from a lengthy process of iterative procedures and trial-and-error selection of training sets [104]. These models are pretrained on large material databases and can be applied to diverse applications with decent accuracy without additional training. The most popular method for preparing large datasets is to utilize open material databases. For instance, M3GNet [54], CHGNet [55], SevenNet-0 [48], and MACE-MP-0 [96] are trained using crystal databases from the Materials Project [61]. Beyond near-equilibrium crystal relaxation configurations from the Materials Project, EquiformerV2 uses the Open Materials 2024 (OMat24) database [105], which includes far-from-equilibrium structures generated by perturbing atomic positions and performing *ab initio* MD starting from crystal structures. DPA-1 [102] is trained using the OC20-2M [106] database, which consists of small molecule adsorbates on various metal surfaces. MACE-OFF-23 [103] is trained with MD trajectories of various organic molecules, including PubChem data and amino acids. In contrast, PFP [79] uses a self-generated database covering molecules, crystals, clusters, and disordered structures to improve generalizability to both unstable structures and stable phases. The most popular architecture for UMLPs is GNNs, which represent structures using graph representations and employ message-passing networks to model interactions between neighboring atoms.

One significant strength of these UMLPs lies in their ability to generalize effectively to out-of-distribution (OOD) data. For instance, PFP accurately predicts the cell volume of metal–organic frameworks, which are absent from its training set [79]. Similarly, MACE-MP-0 accurately predicts the formation energy of crystals with elemental combinations not present in the training database [96]. As an example, we show in figure 10(a) the equilibrium energy of SevenNet-0 [48] for theoretical Ac-Tc crystals from OQMD [107]. Although Ac and Tc have never been present together in any crystal structure and very rarely sampled in the training set, SevenNet-0 reasonably reproduces the DFT energy ordering of the Ac-Tc system. This notable generalization across chemical space may be attributed to element-type embedding. In figure 10(b), UMAP analysis illustrates the embedding vector of each element in SevenNet-0, which was learned during training. This spiral pattern reflects the three-dimensional periodic table, where periods expand outward along an, and families are periodically positioned along a circular path. This indicates that UMLPs can comprehend the chemical nature of each element, facilitating knowledge transfer to unseen interactions.

Current and future challenges

There are several challenges in the broad application of UMLPs across diverse fields. First, there is no standardized method for assessing the generalizability of UMLPs. As a result, their

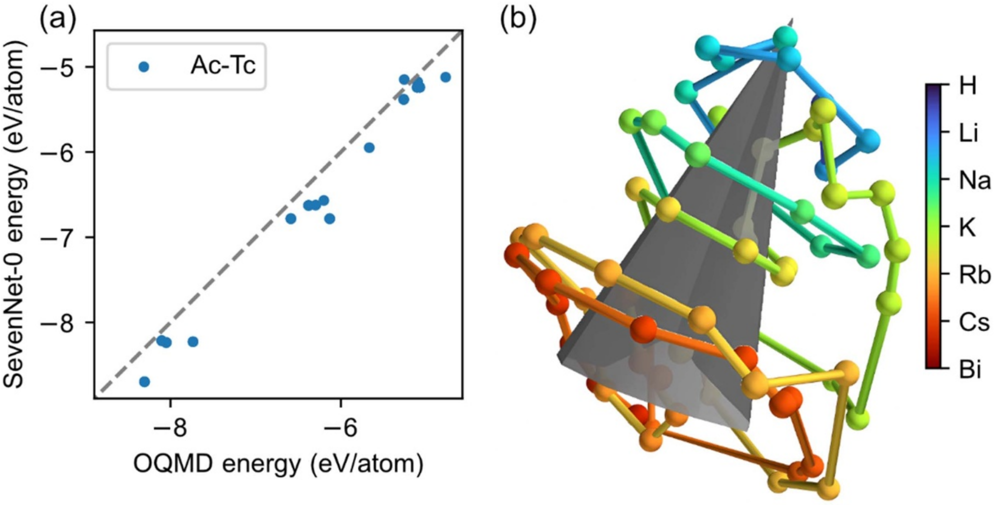

Figure 10. Generalization capability of UMLPs toward extrapolative chemical space. (a) Equilibrium energy of SevenNet-0 with respect to the OQMD reference in the hypothetical Ac-Tc crystal system with various polymorphs. (b) 3D UMAP plot of element-embedding vectors of SevenNet-0. From H to Bi, elements are connected sequentially in the atomic number and colored based on the period.

accuracy in downstream tasks is uncertain, especially when the applications involve systems that appear to deviate from the training domain. Since most UMLPs are primarily trained on inorganic crystal structures [48, 54, 55, 96], their accuracy may not be sufficient for simulating configurations such as defects, surfaces, amorphous and liquid phases, as well as organic and biomolecules. Therefore, robust metrics to evaluate the extrapolative power of UMLPs are in high demand, not only for applying UMLPs to downstream tasks but also for improving their generalizability.

Secondly, the accuracy of *ab initio* calculations for the training data for UMLPs may not be enough for the target application. Most of the extensive *ab initio* databases typically employ GGA for the exchange-correlation functionals, which allows for high-throughput computations. However, atomistic simulations often require higher-level theories, such as meta-GGA functionals or hybrid functionals, to accurately describe lattice parameters or the strong correlation effects among localized electrons.

Finally, the simulation scales of UMLPs are limited by their demanding computational costs. The UMLPs, which typically employ GNNs [48, 54, 55, 79, 96, 103], achieve high accuracy and generalizability at the expense of complex model architectures and increased numbers of parameters ($>10^6$). As a result, these MLPs are orders of magnitude slower than traditional descriptor-based MLPs. Moreover, parallelizing GNNs is challenging because multiple message-passing layers complicate their integration with spatial decomposition, the parallelization strategy commonly used in many MD simulators. These limitations present significant obstacles when applying UMLPs to simulations that require long timescales, such as ionic diffusivity, or to systems involving large length scales, such as those with extended defects.

Advances in science and technology to meet challenges

Similar to other machine learning applications, UMLPs reduce errors in OOD as datasets grow larger, following a power law [108]. To achieve higher generalizability across different chemical systems, UMLPs require extensive training data covering different scenarios. One direction to augment the training data is to consider that the present material database is biased towards popular materials such as oxides. Balancing the chemistry in the database will significantly improve the generalizability of UMLPs. Another strategy to expand training data is to simultaneously train multiple datasets with distinct characteristics. The differences in DFT settings (such as pseudo-potentials, Hubbard U parameters, etc) for each database can be managed by introducing database labels into the model architecture (e.g. embedding features). When training crystal and molecule databases together, the knowledge from crystal data can assist in predicting molecules with new element combination [79]. Also, training datasets with different levels of theory can improve results, especially in high-fidelity calculations, by incorporating the diversity of low-level DFT datasets [66]. Recently, a multi-fidelity training approach has been applied to develop UMLP that achieve meta-GGA level energetics for crystal systems [109]. By combining data from both GGA and meta-GGA datasets from the Materials Project, UMLP can achieve high accuracy with minimal use of meta-GGA data. This approach holds potential for further exploration, aiming to elevate the fidelity of UMLPs toward beyond-DFT methods, such as coupled cluster theory.

When conducting specific simulations, it is often preferable to select a UMLP with moderate complexity as a pretrained model and fine-tune it with a small dataset [96, 102]. Fine-tuning may offer advantages over training MLPs from scratch, including reduced computational costs and enhanced performance across a broad chemical space. However, during fine-tuning, preserving knowledge from pretraining databases and avoiding issues like catastrophic forgetting is essential. Techniques such as layer freezing and continual learning approaches like elastic weight consolidation [110] and experience replay [111] should be explored in fine-tuning MLPs.

The computational performance of GNN-based UMLP can be improved by optimizing algorithms for GPU architecture. In contrast to CPU programming, GPU codes often require low-level optimization for optimal GPU usage. For instance, current tensor product libraries could be significantly accelerated by implementing GPU-friendly algorithms [112].

To compare various UMLPs, standardizing OOD tests that encompass various aspects of material simulation would be beneficial. For example, Matbench Discovery [113] provides a unified framework and benchmark dataset for assessing UMLPs, offering insights into their performance in predicting thermodynamic stability and thermal conductivity [114] in OOD solid systems. Similarly, efforts to actively expand these benchmarks across various applications can assist in selecting appropriate UMLPs based on benchmark results resembling the target system.

Concluding remarks

Trained on a large dataset with deep neural network models, the pretrained UMLP is one of the few machine learning models in materials science that exhibit generalization capabilities in inference, extending well beyond the trained domain. The accuracy of current UMLPs is not yet sufficient to fully replace DFT calculations. However, expanding datasets and improving model architecture may eventually lead to the development of UMLPs with accuracies surpassing conventional DFT. Such advanced UMLPs will open a new avenue in materials science and engineering.

Acknowledgments

This work was supported by the National Research Foundation of Korea (NRF) Grant funded by the Korea government (MSIT) (RS-2023-00247245).

12. Capturing chemical complexity in high-entropy materials

Killian Sheriff[1], *Yifan Cao*[1] *and Rodrigo Freitas*
Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, MA, United States of America
[1] These authors contributed equally to this work.

Status

High-entropy materials are metallic or ceramic systems composed of three or more chemical elements mixed in nearly equiatomic concentrations [97, 115]. These design choices lead to substantially greater chemical complexity than dilute systems, where small quantities of secondary elements are dispersed into a primary-element matrix (figure 11). Chemical complexity functions as the background against which microstructural evolution occurs, thereby affecting various material properties through chemistry–microstructure relationships [116, 117]. Thus, manipulating chemical complexity is a direct route to control microstructure with the purpose of designing properties and performance.

Machine learning potentials (MLPs) are particularly well-suited for investigating the role of chemical complexity on chemistry–microstructure relationships. Experimental and computational evidence shows that variations in chemical complexity can extend for as long as 3 nm [118–120], which is far above what can be properly sampled with electronic-structure calculations. Traditional interatomic potentials can sample such length scales but seem to not be flexible enough to capture chemical complexity from DFT calculations [118]. Meanwhile, MLPs' computational efficiency enables them to reach the appropriate length scale while being sufficiently expressive to capture the large variety of atomic environments present due to chemical complexity, as shown by the authors in [118] and [121].

The accuracy of MLPs for high-entropy materials depends significantly on the construction of training datasets that appropriately sample the chemical complexity. Current designs of such training sets often incorporate a variety of single-element structures, with the chemical complexity in multicomponent systems being sampled with special quasi-random structures of varying concentrations [122]. Alternatively, active-learning strategies have been employed to augment training sets with high-uncertainty structures identified through MC simulations [123]. Yet, there is a persistent lack of validation strategies to verify whether such training-set designs adequately capture chemical complexity and ensure accurate predictions across compositions.

Current and future challenges

Navigating the complexities of the chemical space of high-entropy materials introduces several challenges for MLPs, as listed below.

How can the enormous chemical complexity in these systems be captured with size-limited training datasets?
Combining numerous chemical elements in similarly large concentrations results in a total number of possible chemical configurations that far surpasses the limited size of state-of-the-art MLP datasets. For example, a ternary equiatomic face-centered cubic alloy has more than 6.8×10^{18} distinct chemical motifs when considering a typical radial cutoff distance between the third and fourth nearest-neighbor shell [118, 121].

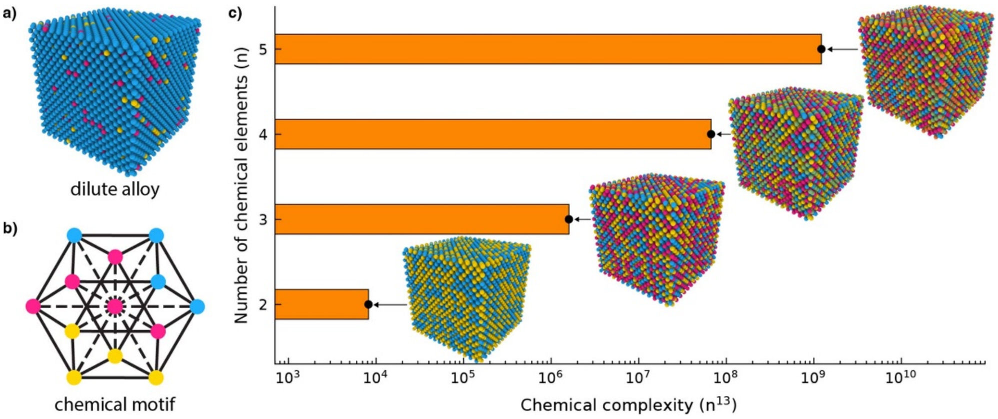

Figure 11. Chemical complexity of high-entropy materials. (a) Dilute alloys have considerably lower chemical complexity than high-entropy materials. (b) Nearest neighbors of an atom in the face-centered cubic (fcc) crystal structure, which is a common lattice for high-entropy alloys. The degree of chemical complexity in a system can be gauged by the number of distinct chemical motifs such as this one [118, 124]. There are n^{13} possible chemical motifs in a fcc system with n chemical elements. (c) The chemical complexity increases by orders of magnitude with each chemical element added to the material.

How can it be ensured that MLPs reproduce the relative phase stability of high-entropy materials?

Properly capturing the energetics of all relevant phases across chemical concentrations and temperatures is important for a variety of simulations of materials processing [125, 126]. For example, solidification simulations require an accurate description of the relative phase stability among the solid phase and the liquid phase at different concentrations. In high-entropy materials this task is complicated by the subtle influence of chemical complexity on the free-energy landscape of different phases, which includes chemical short-range ordering.

How can a MLP be properly validated to capture the relationship between chemical complexity and materials properties?

It has been recently demonstrated that energy accuracy on test sets often does not correlate with accuracy in reproducing material properties. Notably, ensembles of identically-trained MLPs for a high-entropy alloy presented significant variations in performance (figure 12), suggesting that chemical complexity leads to a highly-degenerate energy landscape for MLP fitting [121]. This observation calls attention to a growing need for more comprehensive MLP evaluation metrics that account for the performance in reproducing material properties.

How to find reduced representations of the chemical space that are also complete?

The chemical complexity of high-entropy materials increases by orders of magnitude with each chemical element added to the system (figure 11(c)). This requires MLP representations to increase in size accordingly to maintain the same accuracy for different numbers of chemical elements, which quickly becomes computationally unsustainable. Approaches to identify and optimize reduced (but complete) chemical representations are required to mitigate this immense growth in complexity.

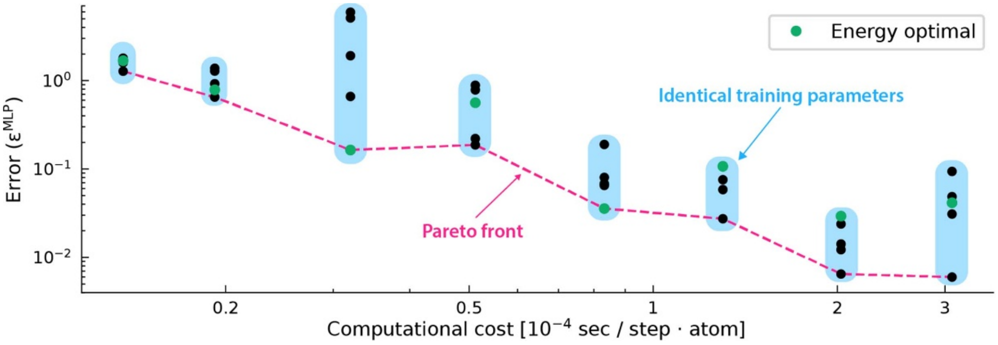

Figure 12. MLP accuracy on test sets often does not correlate with accuracy in reproducing material properties. The MLP error $\varepsilon^{\mathrm{MLP}}$ evaluates the potential's performance in predicting chemical short-range order in the solid and liquid phases, as well as the root-mean-square energy error on the test set [121]. Blue boxes indicate ensembles of identically trained potentials (i.e., same training parameters), and green circles mark the potential with the lowest energy error within each ensemble.

Advances in science and technology to meet challenges

We now turn to possible directions for overcoming the challenges outlined above:

Sampling chemical complexity

Relying on random sampling of chemical complexity for MLP training results in inadequate exploration of the chemical space and limited representation of thermal effects, such as short-range order (SRO) [121]. The field stands to benefit from innovative sampling strategies tailored to capture the chemical complexity inherent in these materials. For example, biasing the sampling of training data towards chemical configurations relevant to thermal equilibrium proves to be an effective strategy to manage the chemical complexity of high-entropy materials and mitigate the influence of the nearly-degenerate energy landscape on MLP fitting [121, 123]. Further efforts should consider how to more efficiently sample the chemical space of high-entropy materials while maintaining a reasonable training dataset size. Emerging approaches [116–118, 121, 124, 127] that quantify chemical complexity in atomistic data would be a powerful tool to guide the development of compact MLP datasets.

Validation

An initial effort to better validate MLPs is depicted in figure 12, where a comprehensive error metric evaluates their performance in capturing chemical complexity and material properties [119]. Further efforts should prioritize refining such MLP evaluation metrics to incorporate other relevant material properties affected by chemical complexity, such as defect energies and relative phase stability. Additionally, the field would greatly benefit from open-source frameworks designed to conduct uncertainty quantification regarding the predictive capabilities of MLPs for specific properties of high-entropy materials.

Representation

A plausible solution to the poor scaling of ML models with chemical complexity has emerged through the development of compressed lower-dimensional descriptors of chemical information. This approach aims to mitigate the immense growth in computational resources required to represent a growing array of chemical elements [53, 128, 129]. Additionally, equivariant GNNs have shown potential in providing another reduced yet complete representation of the chemical space [118, 124]. Evaluation of the fidelity and efficiency of such representations could lead to further insights into the description of atomic environments for high-entropy materials.

Concluding remarks

Not long ago, the prospect of performing computational investigations of chemistry–microstructure relationships with predictive accuracy was inconceivable. Now, this possibility is within our grasp due to the introduction of MLPs. This remarkable achievement has only been made possible due to the relentless development of increasingly accurate and efficient ML methodology, which has been promptly adopted and adapted for MLPs. Yet, the potential of atomistic simulations as a predictive tool for high-entropy materials will not be realized until we tame chemical complexity by demonstrating that it can be properly captured with MLPs.

Acknowledgments

This work was supported by the MathWorks Ignition Fund, MathWorks Engineering Fellowship Fund, and the Portuguese Foundation for International Cooperation in Science, Technology and Higher Education in the MIT-Portugal Program. We were also supported by the Research Support Committee Funds from the School of Engineering at the Massachusetts Institute of Technology.

Data availability statement

No new data were created or analysed in this study.

ORCID iDs

Yong-Wei Zhang https://orcid.org/0000-0001-7255-1678
Viacheslav Sorkin https://orcid.org/0000-0001-8284-5354
Antonio Politano https://orcid.org/0000-0001-6806-8130
Jörg Behler https://orcid.org/0000-0002-1220-1542
Aidan P Thompson https://orcid.org/0000-0002-0324-9114
Tsz Wai Ko https://orcid.org/0000-0002-0802-9559
Olga Chalykh https://orcid.org/0009-0008-2071-3725
Dmitry Korogod https://orcid.org/0009-0007-9378-6653
Evgeny Podryabinkin https://orcid.org/0000-0001-9057-7249
Ju Li https://orcid.org/0000-0002-7841-8058
Yuri Mishin https://orcid.org/0000-0001-7299-433X
Zongrui Pei https://orcid.org/0000-0003-0748-4629

Xianglin Liu https://orcid.org/0000-0001-7116-3649
Jaesun Kim https://orcid.org/0009-0000-6646-1318
Yutack Park https://orcid.org/0009-0008-8690-935X
Seungwoo Hwang https://orcid.org/0000-0002-1523-8340
Rodrigo Freitas https://orcid.org/0000-0001-8523-2916

References

[1] Behler J 2016 Perspective: machine learning potentials for atomistic simulations *J. Chem. Phys.* **145** 170901
[2] Mishin Y 2021 Machine-learning interatomic potentials for materials science *Acta Mater.* **214** 116980
[3] Mueller T, Hernandez A and Wang C 2020 Machine learning for interatomic potential model *J. Chem. Phys.* **152** 050902
[4] Unke O T, Chmiela S, Gastegger M, Schütt K T, Sauceda H E and Müller K R 2021 SpookyNet: learning force fields with electronic degrees of freedom and nonlocal effects *Nat. Commun.* **12** 7273
[5] Friederich P, Häse F, Proppe J and Aspuru-Guzik A 2021 Machine-learned potentials for next-generation matter simulations *Nat. Mater.* **20** 750–61
[6] Fedik N *et al* 2022 Extending machine learning beyond interatomic potentials for predicting molecular properties *Nat. Rev. Chem.* **6** 653–72
[7] Miksch A M, Morawietz T, Kästner J, Urban A and Artrith N 2021 Strategies for the construction of machine-learning potentials for accurate and efficient atomic-scale simulations *Mach. Learn.: Sci. Technol.* **2** 031001
[8] Peterson A A 2016 Acceleration of saddle-point searches with machine learning *J. Chem. Phys.* **145** 074106
[9] Yakutovich A V *et al* 2021 AiiDAlab—an ecosystem for developing, executing, and sharing scientific workflows *Comput. Mater. Sci.* **188** 110165
[10] Smith J S, Nebgen B, Lubbers N, Isayev O and Roitberg A E 2018 Less is more: sampling chemical space with active learning *J. Chem. Phys.* **148** 241733
[11] Podryabinkin E V and Shapeev A V 2017 Active learning of linearly parametrized interatomic potentials *Comput. Mater. Sci.* **140** 171–80
[12] Artrith N, Urban A and Ceder G 2018 Constructing first-principles phase diagrams of amorphous LixSi using machine-learning-assisted sampling with an evolutionary algorithm *J. Chem. Phys.* **148** 1–10
[13] Bernstein N, Csányi G and Deringer V L 2019 De novo exploration and self-guided learning of potential-energy surfaces *npj Comput. Mater.* **5** 1–9
[14] Yoo D, Jung J, Jeong W and Han S 2021 Metadynamics sampling in atomic environment space for collecting training data for machine learning potentials *npj Comput. Mater.* **7** 131
[15] Zaverkin V, Holzmüller D, Christiansen H, Errica F, Alesiani F, Takamoto M, Niepert M and Kästner J 2024 Uncertainty-biased molecular dynamics for learning uniformly accurate interatomic potentials *npj Comput. Mater.* **10** 83
[16] Salas-Estrada L, Fiorillo B and Filizola M 2022 Metadynamics simulations leveraged by statistical analyses and artificial intelligence-based tools to inform the discovery of G protein-coupled receptor ligands *Front. Endocrinol.* **13** 1099715
[17] Yang M, Bonati L, Polino D and Parrinello M 2022 Using metadynamics to build neural network potentials for reactive events: the case of urea decomposition in water *Catal. Today* **387** 143–9
[18] Shapeev A V 2016 Moment tensor potentials: a class of systematically improvable interatomic potentials *Multiscale Model. Simul.* **14** 1153–73
[19] Hodapp M and Shapeev A 2020 In operando active learning of interatomic interaction during large-scale simulations *Mach. Learn.: Sci. Technol.* **1** 045005
[20] Tan A R, Urata S, Goldman S, Dietschreit J C B and Gómez-Bombarelli R 2023 Single-model uncertainty quantification in neural network potentials does not consistently outperform model ensembles *npj Comput. Mater.* **9** 01754
[21] Christiansen M-P V, Rønne N and Hammer B 2024 Efficient ensemble uncertainty estimation in Gaussian Processes Regression *Mach. Learn.: Sci. Technol.* **1** 8894

[22] Heid E, Schörghuber J, Wanzenböck R and Madsen G K H 2024 Spatially Resolved Uncertainties for Machine Learning Potentials *J. Chem. Inf. Model.* **64** 6377–87

[23] Zhu A, Batzner S, Musaelian A and Kozinsky B 2023 Fast uncertainty estimates in deep learning interatomic potentials *J. Chem. Phys.* **158** 164111

[24] Erhard L C, Rohrer J, Albe K and Deringer V L 2022 A machine-learned interatomic potential for silica and its relation to empirical models *npj Comput. Mater.* **8** 90

[25] Mortazavi B, Zhuang X, Rabczuk T and Shapeev A V 2023 Atomistic modeling of the mechanical properties: the rise of machine learning interatomic potentials *Mater. Horiz.* **10** 1956–68

[26] Marchand D and Curtin W A 2022 Machine learning for metallurgy IV: a neural network potential for Al-Cu-Mg and Al-Cu-Mg-Zn *Phys. Rev. Mater.* **6** 053803

[27] Morrow J D, Gardner J L A and Deringer V L 2023 How to validate machine-learned interatomic potentials *J. Chem. Phys.* **158** 121501

[28] Chen X, Wang L-F, Gao X-Y, Zhao Y-F, Lin D-Y, Chu W-D and Song H-F 2021 Machine learning enhanced empirical potentials for metals and alloys *Comput. Phys. Commun.* **269** 108132

[29] Aitken Z H, Sorkin V, Yu Z G, Chen S, Tan T L, Wu Z and Zhang Y-W 2024 Controlling screw dislocation core structure and Peierls barrier in BCC interatomic potentials *Int. J. Solids Struct.* **303** 113004

[30] Vita J A and Trinkle D R 2024 Spline-based neural network interatomic potentials: blending classical and machine learning models *Comput. Mater. Sci.* **232** 112655

[31] Nitol M S, Dang K, Fensin S J, Baskes M I, Dickel D E and Barrett C D 2023 Hybrid interatomic potential for Sn *Phys. Rev. Mater.* **7** 043601

[32] Blank T B, Brown S D, Calhoun A W and Doren D J 1995 Neural network models of potential energy surfaces *J. Chem. Phys.* **103** 4129–37

[33] Behler J and Parrinello M 2007 Generalized neural-network representation of high-dimensional potential-energy surfaces *Phys. Rev. Lett.* **98** 146401

[34] Artrith N, Morawietz T and Behler J 2011 High-dimensional neural-network potentials for multicomponent systems: applications to zinc oxide *Phys. Rev.* B **83** 153101

[35] Ghasemi S A, Hofstetter A, Saha S and Goedecker S 2015 Interatomic potentials for ionic systems with density functional accuracy based on charge densities obtained by a neural network *Phys. Rev.* B **92** 045131

[36] Behler J 2021 Four generations of high-dimensional neural network potentials *Chem. Rev.* **121** 10037–72

[37] Gilmer J, Schoenholz S S, Riley P F, Vinyals O and Dahl G E 2017 Neural message passing for quantum chemistry *Proc. 34th Int. Conf. on Machine Learning* vol 70 pp 1263–72

[38] Batzner S, Musaelian A, Sun L, Geiger M, Mailoa J P, Kornbluth M, Molinari N, Smidt T E and Kozinsky B 2022 SE(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials *Nat. Commun.* **13** 2453

[39] Eckhoff M and Behler J 2021 High-dimensional neural network potentials for magnetic systems using spin-dependent atom-centered symmetry functions *npj Comput. Mater.* **7** 170

[40] Ko T W, Finkler J A, Goedecker S and Behler J 2021 A fourth-generation high-dimensional neural network potential with accurate electrostatics including non-local charge transfer *Nat. Commun.* **12** 398

[41] Donati C, Douglas J F, Kob W, Plimpton S J, Poole P H and Glotzer S C 1998 Stringlike cooperative motion in a supercooled liquid *Phys. Rev. Lett.* **80** 2338–41

[42] Thompson A P *et al* 2022 LAMMPS—a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales *Comput. Phys. Commun.* **271** 108171

[43] (a) EAM: Daw M S and Baskes M I 1984 Embedded-atom method: derivation and application to impurities, surfaces, and other defects in metals *Phys. Rev.* B **29** 6443–53
(b) ReaxFF: Chenoweth K, van Duin A C T, and Goddard W A 2008 ReaxFF reactive force field for molecular dynamics simulations of hydrocarbon oxidation *J. Phys. Chem.* A **112** 1040–53
(c) Tersoff: Tersoff J 1989 Modeling solid-state chemistry: interatomic potentials for multicomponent systems *Phys. Rev.* B **39** 5566–8
(d) REBO: Brenner D W *et al* 2002 A second-generation reactive empirical bond order (REBO) potential energy expression for hydrocarbons *J. Phys.: Condens. Matter* **14** 783

[44] (a) DeepPot: Zhang L, Han J, Wang H, Car R and Weinan E 2018 Deep potential molecular dynamics: a scalable model with the accuracy of quantum mechanics *Phys. Rev. Lett.* **120** 143001
(b) BP-NNP: Behler J and Parrinello M 2007 Generalized neural-network representation of high-dimensional potential-energy surfaces *Phys. Rev. Lett.* **98** 146401

(c) GAP: Bartók A P, Payne M C, Kondor R and Csányi G 2010 Gaussian approximation potentials: the accuracy of quantum mechanics,without the electrons *Phys. Rev. Lett.* **104** 136403
(d) MTP: Shapeev A V 2016 Moment tensor potentials: a class of systematically improvable interatomic potentials *Multiscale Model. Simul.* **14** 1153–73
(e) SNAP: Thompson A P, Swiler L P, Trott C R, Foiles S M and Tucker G J 2015 Spectral neighbor analysis method for automated generation of quantum-accurate interatomic potentials *J. Comput. Phys.* **285** 316–30
(f) ACE: Drautz R 2019 Atomic cluster expansion for accurate and transferable interatomic potentials *Phys. Rev.* B **99** 014104
(g) NEquIP: Batzner S, Musaelian A, Sun L, Geiger M, Mailoa J P, Kornbluth M, Molinari N, Smidt T E and Kozinsky B E 2022 E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials *Nat. Commun.* **13** 2453
(h) SchNet: Schütt K T, Sauceda H E, Kindermans P-J, Tkatchenko A and K-R Müller 2018 SchNet—a deep learning architecture for molecules and materials *J. Chem. Phys.* **148** 241722
(i) MACE: Batatia I, Kovacs D P, Simm G N C, Ortner C and Csanyi G 2022 MACE: higher order equivariant message passing neural networks for fast and accurate force fields *Advances in Neural Information Processing Systems*
(j) Allegro: Musaelian A, Batzner S, Johansson A, Sun L, Owen C J, Kornbluth M and Kozinsky B 2023 Learning local equivariant representations for large-scale atomistic dynamics *Nat. Commun.* **14** 579
(k) GrACE: Bochkarev A, Lysogorskiy Y and Drautz R 2024 Graph atomic cluster expansion for semilocal interactions beyond equivariant message passing *Phys. Rev.* X **14** 021036

[45] (a) Nguyen-Cong K *et al* 2024 Extreme metastability of diamond and its transformation to the BC8 post-diamond phase of carbon *J. Phys. Chem. Lett.* **15** 1152–60
(b) Willman J T *et al* 2022 Machine learning interatomic potential for simulations of carbon at extreme conditions *Phys. Rev.* B **106** L180101
(c) Nguyen-Cong K *et al* 2021 Billion atom molecular dynamics simulations of carbon at extreme conditions and experimental time and length scales *Proc. Int. Conf. for High Performance Computing, Networking, Storage and Analysis* pp 1–12

[46] Rohskopf A *et al* 2023 FitSNAP: atomistic machine learning with LAMMPS *J. Open Source Softw.* **8** 5118

[47] Wood M A and Thompson A P 2018 Extending the accuracy of the SNAP interatomic potential form *J. Chem. Phys.* **148** 241721

[48] Park Y, Kim J, Hwang S and Han S 2024 Scalable parallel algorithm for graph neural network interatomic potentials in molecular dynamics simulations *J. Chem. Theory Comput.* **20** 4857

[49] LAMMPS Programmer Guide (available at: https://docs.lammps.org/Manual.html#programmer-documentation)

[50] Trott C R *et al* 2022 Kokkos 3: programming model extensions for the exascale era *IEEE Trans. Parallel Distrib. Syst.* **33** 805–17

[51] Lubbers N, Smith J S and Barros K 2018 Hierarchical modeling of molecular energies using a deep neural network *J. Chem. Phys.* **148** 241715

[52] Chen C, Ye W, Zuo Y, Zheng C and Ong S P 2019 Graph networks as a universal machine learning framework for molecules and crystals *Chem. Mater.* **31** 3564

[53] Batatia I, Kovács D P, Simm G N C, Ortner C and Csányi G 2022 MACE: Higher order equivariant message passing neural networks for fast and accurate force fields *Advances in Neural Information Processing Systems* vol 35 pp 11423–36

[54] Chen C and Ong S P 2022 A universal graph deep learning interatomic potential for the periodic tabl *Nat. Comput. Sci.* **2** 718

[55] Deng B, Zhong P, Jun K, Riebesell J, Han K, Bartel C J and Ceder G 2023 CHGNet as a pretrained universal neural network potential for charge-informed atomistic modelling *Nat. Mach. Intell.* **5** 9

[56] Ko T W and Ong S P 2023 Recent advances and outstanding challenges for machine learning interatomic potentials *Nat. Comput. Sci.* **3** 998–1000

[57] Pozdnyakov S N and Ceriotti M 2022 Incompleteness of graph neural networks for points clouds in three dimensions *Mach. Learn.: Sci. Technol.* **3** 045020

[58] Musaelian A, Batzner S, Johansson A, Sun L, Owen C J, Kornbluth M and Kozinsky B 2023 Learning local equivariant representations for large-scale atomistic dynamics *Nat. Commun.* **14** 579

[59] Schütt K T, Kessel P, Gastegger M, Nicoli K A, Tkatchenko A and Müller K-R 2019 SchNetPack: a deep learning toolbox for atomistic systems *J. Chem. Theory Comput.* **15** 448

[60] Cheng B 2024 Cartesian atomic cluster expansion for machine learning interatomic potentials (arXiv:2402.07472)

[61] Jain A *et al* 2013 Commentary: the materials project: a materials genome approach to accelerating materials innovation *APL Mater.* **1** 011002

[62] Sun J, Ruzsinszky A and Perdew J P 2015 Strongly constrained and appropriately normed semi-local density functional *Phys. Rev. Lett.* **115** 036402

[63] Heyd J, Scuseria G E and Ernzerhof M 2003 Hybrid functionals based on a screened coulomb potential *J. Chem. Phys.* **118** 8207

[64] Bartlett R J, Watts J D, Kucharski S A and Noga J 1990 Non-iterative fifth-order triple and quadruple excitation energy corrections in correlated methods *Chem. Phys. Lett.* **165** 513

[65] Qi J, Ko T W, Wood B C, Pham T A and Ong S P 2024 Robust training of machine learning interatomic potentials with dimensionality reduction and stratified sampling *npj Comput. Mater.* **10** 1

[66] Chen C, Zuo Y, Ye W, Li X and Ong S P 2021 Learning properties of ordered and disordered materials from multi-fidelity data *Nat. Comput. Sci.* **1** 46

[67] Yu H, Zhong Y, Hong L, Xu C, Ren W, Gong X and Xiang H 2024 Spin-dependent graph neural network potential for magnetic materials *Phys. Rev.* B **109** 144426

[68] Fey M and Lenssen J E 2019 Fast graph representation learning with PyTorch geometric (arXiv:1903.02428)

[69] Wang M *et al* 2019 Deep graph library: a graph-centric, highly-performant package for graph neural networks (arXiv:1909.01315)

[70] Gubaev K, Podryabinkin E V, Hart G L and Shapeev A V 2019 Accelerating high-throughput searches for new alloys with active learning of interatomic potentials *Comput. Mater. Sci.* **156** 148–56

[71] Novikov I S, Gubaev K, Podryabinkin E V and Shapeev A V 2020 The MLIP package: moment tensor potentials with mpi and active learning *Mach. Learn.: Sci. Technol.* **2** 025002

[72] Zuo Y *et al* 2020 Performance and cost assessment of machine learning interatomic potentials *J. Phys. Chem.* A **124** 731–45

[73] Zhu L-F, Körmann F, Ruban A V, Neugebauer J and Grabowski B 2020 Performance of the standard exchange-correlation functionals in predicting melting properties fully from first principles: application to al and magnetic ni *Phys. Rev.* B **101** 144108

[74] Behler J and Csányi G 2021 Machine learning potentials for extended systems: a perspective *Eur. Phys. J.* B **94** 1–11

[75] Podryabinkin E V, Kvashnin A G, Asgarpour M, Maslenikov I I, Ovsyannikov D A, Sorokin P B, Popov M Y and Shapeev A V 2022 Nanohardness from first principles with active learning on atomic environments *J. Chem. Theory Comput.* **18** 1109–21

[76] Wang W, Wu Z, Dietschreit J C and Gómez-Bombarelli R 2023 Learning pair potentials using differentiable simulations *J. Chem. Phys.* **158** 044113

[77] Takamoto S, Okanohara D, Li Q-J and Li J 2023 Towards universal neural network interatomic potential *J. Materiom.* **9** 447–54

[78] Takamoto S, Izumi S and Li J 2022 TeaNet: universal neural network interatomic potential inspired by iterative electronic relaxations *Comput. Mater. Sci.* **207** 111280

[79] Takamoto S *et al* 2022 Towards universal neural network potential for material discovery applicable to arbitrary combination of 45 elements *Nat. Commun.* **13** 2991

[80] Kovács D P, Batatia I, Arany E S and Csányi G 2023 Evaluation of the MACE force field architecture: from medicinal chemistry to materials science *J. Chem. Phys.* **159** 044118

[81] Morgan D, Pilania G, Couet A, Uberuaga B P, Sun C and Li J 2022 Machine learning in nuclear materials research *Curr. Opin. Solid State Mater. Sci.* **26** 100975

[82] Artrith N, Garrido Torres J A, Urban A and Hybertsen M S 2022 Data-driven approach to parameterize $\mathrm{SCAN}+U$ for an accurate description of $3d$ transition metal oxide thermochemistry *Phys. Rev. Mater.* **6** 035003

[83] Liu F, Duan C and Kulik H J 2020 Rapid detection of strong correlation with machine learning for transition-metal complex high-throughput screening *J. Phys. Chem. Lett.* **11** 8067–76

[84] Tang H *et al* 2024 Multi-task learning for molecular electronic structure approaching coupled-cluster accuracy (arXiv:2405.12229)

[85] Pun G P P, Batra R, Ramprasad R and Mishin Y 2019 Physically informed artificial neural networks for atomistic modeling of materials *Nat. Commun.* **10** 2339

[86] Pun G P P, Yamakov V, Hickman J, Glaessgen E H and Mishin Y 2020 Development of a general-purpose machine-learning interatomic potential for aluminum by the physically informed neural network method *Phys. Rev. Mater.* **4** 113807

[87] Lin Y-S, Pun G P P and Mishin Y 2022 Development of a physically-informed neural network interatomic potential for tantalum *Comput. Mater. Sci.* **205** 111180

[88] Hodapp M 2024 Machine learning is funny but physics makes the money: how machine- learning potentials can advance computer-aided materials design in metallurgy *Comput. Mater. Sci.* **233** 112715

[89] Raabe D, Mianroodi J R and Neugebauer J 2023 Accelerating the design of compositionally complex materials via physics-informed artificial intelligence *Nat. Comput. Sci.* **3** 198–209

[90] Rowe P, Deringer V L, Gasparotto P, Cs ́anyi G and Michaelides A 2020 An accurate and transferable machine learning potential for carbon *J. Chem. Phys.* **153** 034702

[91] Zhang Y, Hu C and Jiang B 2019 Embedded atom neural network potentials: efficient and accurate machine learning with a physically inspired representation *J. Phys. Chem. Lett.* **10** 4962–7

[92] Liu X, Zhang J and Pei Z 2023 Machine learning for high-entropy alloys: progress, challenges and opportunities *Proc. Mater. Sci.* **131** 101018

[93] Unke O T and Meuwly M 2019 PhysNet: a neural network for predicting energies, forces, dipole moments, and partial charges *J. Chem. Theory Comput.* **15** 3678–93

[94] Bartók A P, Payne M C, Kondor R and Csányi G 2010 Gaussian approximation potentials: the accuracy of quantum mechanics, without the electrons *Phys. Rev. Lett.* **104** 136403

[95] Drautz R 2019 Atomic cluster expansion for accurate and transferable interatomic potentials *Phys. Rev.* B **99** 014104

[96] Batatia I *et al* 2023 A foundation model for atomistic materials chemistry (arXiv:2401.00096)

[97] George E P, Raabe D and Ritchie R O 2019 High-entropy alloys *Nat. Rev. Mater.* **4** 515–34

[98] Kostiuchenko T, Körmann F, Neugebauer J and Shapeev A 2019 Impact of lattice relaxations on phase transitions in a high-entropy alloy studied by machine-learning potentials *npj Comput. Mater.* **5** 55

[99] Choi W-M, Jo Y H, Sohn S S, Lee S and Lee B-J 2018 Understanding the physical metallurgy of the CoCrFeMnNi high-entropy alloy: an atomistic simulation study *npj Comput. Mater.* **4** 1

[100] Yin S, Zuo Y, Abu-Odeh A, Zheng H, Li X-G, Ding J, Ong S P, Asta M and Ritchie R O 2021 Atomistic simulations of dislocation mobility in refractory high-entropy alloys and the effect of chemical short-range order *Nat. Commun.* **12** 4873

[101] Song K *et al* 2023 General-purpose machine-learned potential for 16 elemental metals and their alloys (arXiv:2311.04732)

[102] Zhang D, Bi H, Dai F-Z, Jiang W, Zhang L and Wang H 2023 DPA-1: pretraining of attention-based deep potential model for molecular simulation (arXiv:2208.08236v4)

[103] Kovács D P, Moore J H, Browning N J, Batatia I, Horton J T, Kapil V, Witt W C, Magdău I-B, Cole D J and Csányi G 2023 MACE-OFF23: transferable machine learning force fields for organic molecules (arXiv:2312.15211v2)

[104] Hong C, Kim J, Kim J, Jung J, Ju S, Choi J M and Han S 2023 Applications and training sets of machine learning potentials *Sci. Technol. Adv. Mater.* **3** 2269948

[105] Barroso-Luque L, Shuaibi M, Fu X, Wood B M, Dzamba M, Gao M, Rizvi A, Zitnick C L and Ulissi Z W 2024 Open materials 2024 (OMat24) inorganic materials dataset and models (arXiv:2410.12771v1)

[106] Chanussot L *et al* 2021 Open catalyst 2020 (OC20) dataset and community challenges *ACS Catal.* **11** 6059–72

[107] Kirklin S, Saal J E, Meredig B, Thompson A, Doak J W, Aykol M, Rühl S and Wolverton C 2015 The open quantum materials database (OQMD): assessing the accuracy of DFT formation energies *npj Comput. Mater.* **1** 15010

[108] Merchant A, Batzner S, Schoenholz S S, Aykol M, Cheon G and Cubuk E D 2023 Scaling deep learning for materials discovery *Nature* **624** 80–85

[109] Kim J, Kim J, Kim J, Lee J, Park Y, Kang Y and Han S 2025 Data-efficient multifidelity training for high-fidelity machine learning interatomic potentials *J. Am. Chem. Soc.* **147** 1042–54

[110] Kirkpatrick J *et al* 2017 Overcoming catastrophic forgetting in neural networks *Proc. Natl Acad. Sci. USA* **114** 3521–6

[111] Rolnick D, Ahuja A, Schwarz J, Lillicrap T P and Wayne G 2019 Experience replay for continual learning (arXiv:1811.11682v2)

[112] Luo S, Chen T and Krishnapriyan A S 2024 Enabling efficient equivariant operations in the fourier basis via gaunt tensor products (arXiv:2401.10216v1)

[113] Riebesell J, Goodall R E A, Benner P, Chiang Y, Deng B, Lee A A, Jain A and Persson K A 2024 Matbench discovery A framework to evaluate machine learning crystal stability predictions (arXiv:2308.14920v2)

[114] Póta B, Ahlawat P, Csányi G and Simoncelli M 2024 Thermal conductivity predictions with foundation atomistic models (arXiv:2408.00755v4)

[115] Oses C, Toher C and Curtarolo S 2020 High-entropy ceramics *Nat. Rev. Mater.* **5** 295–309

[116] Ferrari A, Körmann F, Asta M and Neugebauer J 2023 Simulating short-range order in compositionally complex materials *Nat. Comput. Sci.* **3** 221–9

[117] McCarthy M J, Startt J, Dingreville R, Thompson A P and Wood M A 2023 Atomic representations of local and global chemistry in complex alloys (arXiv:2303.04311)

[118] Sheriff K, Cao Y, Smidt T and Freitas R 2024 Quantifying chemical short-range order in metallic alloys *Proc. Natl Acad. Sci. USA* **121** e2322962121

[119] Zhang R, Zhao S, Ding J, Chong Y, Jia T, Ophus C, Asta M, Ritchie R O and Minor A M 2020 Short-range order and its impact on the CrCoNi medium-entropy alloy *Nature* **581** 283–7

[120] He M, Davids W J, Breen A J and Ringer S P 2024 Quantifying short-range order using atom probe tomography *Nat. Mater.* **23** 1200–7

[121] Cao Y, Sheriff K and Freitas R 2024 Capturing short-range order in high-entropy alloys with machine learning potentials (arXiv:2401.06622)

[122] Li X-G, Chen C, Zheng H, Zuo Y and Ong S P 2020 Complex strengthening mechanisms in the NbMoTaW multi-principal element alloy *npj Comput. Mater.* **6** 1–10

[123] Ghosh S, Sotskov V, Shapeev A V, Neugebauer J and Körmann F 2022 Short-range order and phase stability of CrCoNi explored with machine learning potentials *Phys. Rev. Mater.* **6** 113804

[124] Sheriff K, Cao Y and Freitas R 2024 Chemical-motif characterization of short-range order with E(3)-equivariant graph neural networks *npj Comput. Mater.* **10** 215

[125] Islam M, Sheriff K, Cao Y and Freitas R 2024 Nonequilibrium chemical short-range order in metallic alloys (arXiv:2409.15474)

[126] Li Z, Pradeep K, Deng Y, Raabe D and Tasan C C 2016 Metastable high-entropy dual-phase alloys overcome the strength–ductility trade-off *Nature* **534** 227–30

[127] Li Y *et al* 2024 Machine learning-enabled tomographic imaging of chemical short-range atomic ordering *Adv. Mater.* **36** 2407564

[128] Lopanitsyna N, Fraux G, Springer M A, De S and Ceriotti M 2023 Modeling high-entropy transition metal alloys with alchemical compression *Phys. Rev. Mater.* **7** 045802

[129] Shapeev A 2017 Accurate representation of formation energies of crystalline alloys with many components *Comput. Mater. Sci.* **139** 26–30