

ABSTRACT

Title of Dissertation: MATERIALS MODELING BY AB INITIO
METHODS AND MACHINE LEARNING
INTERATOMIC POTENTIALS: A CRITICAL
ASSESSMENT

Yunsheng Liu, Doctor of Philosophy, 2023

Dissertation directed by: Professor Yifei Mo, Department of Materials
Science and Engineering

Atomistic modeling is a crucial research technique in materials science to simulate physical phenomena based on the interactions of atoms. Density functional theory (DFT) calculation has been the standard technique for evaluating atom interactions, but their applications are limited due to high computation cost. Classical interatomic potential, as another widely adopted technique, provides an alternative by enabling large-scale simulations that are computationally inexpensive. However, the employment of classical potentials for cheap atomistic simulations is achieved in exchange of accurately evaluating atom interactions and the transferability to different chemistries. Most recently, machine learning interatomic potential (MLP) emerges as a new computation technique to bridge the gap between first-principles computation and classical potentials.

I first utilized the atomistic modeling based on DFT calculations to find novel Li superionic conductor, a key component of the emerging all-solid-state Li-ion battery technology. I performed a systematic study on the Li-ion conduction of lithium chloride materials system and predicted a dozen potential Li superionic conductors. I revealed that the Li-ion migration in the materials is

greatly impacted by the Li content, the cation configuration, and the cation concentrations. I further demonstrated tuning these three factors in designing new chloride Li-ion conductors.

Then, I studied the atomistic dynamics predicted by MLPs in comparison to DFT calculations to answer the open question whether MLPs can accurately reproduce dynamical phenomena and related physical properties in molecular dynamics simulations. I examined the current state-of-the-art MLPs and uncovered a number of discrepancies related to atomistic dynamics, defects, and rare events compared to DFT methods. I found that testing averaged errors of MLPs are insufficient and developed evaluation metrics that better indicate the accurate prediction of related properties by MLPs in MD simulations. I further demonstrated that the MLPs optimized by the proposed evaluation metrics have improved prediction in multiple properties.

I also study the performance of MLPs in assessing the elemental orderings in a large variety of phases across composition range in alloy system. Using the Li-Al alloy system as a case study, I trained MLPs using only a few phases and the trained MLP demonstrated good performances over a number of existing and other hypothetical materials in- and out- of the training data across the Li-Al binary alloy system. We developed several new evaluation metrics on energy rankings to evaluate the elemental ordering, which is critical for studying the phase stabilities of materials. I tested MLP transferability to other phases and the limits of MLP applications on commonly performed simulations. I also studied the effect of diverse training data on MLP performances.

With these efforts evaluating MLP performances for a number of properties, metrics, prediction errors, and dynamical phenomena, I constructed a dataset with a large number of MLPs and their performances and performed an empirical analysis to identify the challenging properties to be predicted by the MLPs. Further, I identified pairs of properties that are challenging to predict.

This series of works demonstrated that atomistic modeling is an effective computation technique for studying atomistic dynamic mechanisms, evaluating materials thermodynamics, and guiding materials discovery. As an emerging computational technique, MLPs show good performance on many materials and have great potential to enhance the materials research, but the results show that critical assessments are needed to examine their capabilities to accurately reproduce physical phenomena and understand their limits of performing reliable simulations. My thesis evaluates MLP performances on a number of critical issues related to materials simulations and provides guidance to improve MLPs for future studies.

MATERIALS MODELING BY AB INITIO METHODS AND MACHINE LEARNING
INTERATOMIC POTENTIALS: A CRITICAL ASSESSMENT

by

Yunsheng Liu

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctoral of Philosophy
2023

Advisory Committee:

Professor Yifei Mo, Chair

Professor Ichiro Takeuchi

Professor Lourdes G. Salamanca-Riba

Professor Mark D. Fuge

Professor Pratyush Tiwary

© Copyright by
Yunsheng Liu
2023

Acknowledgements

First and foremost, I would like to thank my advisor, Professor Yifei Mo. The experience of working in his group is probably the most fulfilling time up to now. The greatest lesson I've learned from working with him is 'how to accomplish things'. Plain, simple, but most valuable. Accomplishment is all we can leave behind. I would also like to thank him for his support letting me explore different topics I feel interested, especially on machine learning and quantum computing. I enjoyed a lot learning all the knowledge, though these two fields are not directly related to battery research. Now I have a clear direction for my following career. I greatly appreciate such free environment for exploration in the group. This 6-year journey is long, much longer than my previous colleagues, but it's choice I don't regret.

I would also like to thank my dissertation committee, Professor Lourdes G. Salamanca-Riba, Professor Ichiro Takeuchi, Professor Mark D. Fugre, and Professor Pratyush Tiwary. Their suggestions are very valuable to this dissertation.

Special thanks to all my colleagues in the group, including Dr. Xingfeng He, Dr. Yizhou Zhu, Dr. Qiang Bai, Dr. Addie Mei-Chun Nolan, Dr. Shafiqul Islam, Dr. Shuo Wang, Dr. Menghao Yang, Alex Hall, and Alex Epstein. It's great and fun to learn, discuss, chat, and live the lab life with them. I would also hope these two workstations back online soon.

Most important thank to my parents for their love, support, and phone calls in the past years. They are always irreplaceable.

Table of Contents

Acknowledgements.....	iii
Table of Contents.....	iv
Chapter 1: Introduction.....	1
1.1 First-principles computation study of halide-based Li-ion conductors	2
1.2 Unravel discrepancies and develop evaluation metrics for MLPs.....	3
1.3 Assessing the accuracy of MLPs in predicting the elemental ordering: a case study on Li-Al alloys	5
1.4 High-dimensional assessment of MLP performances for multiple properties....	7
Chapter 2: First-principles computation study of halide-based Li-ion conductors.....	8
2.1 Introduction.....	8
2.2 High-throughput computation of chloride Li-ion conductors.....	11
2.3 Effect of cation sublattice on Li-ion migration.....	17
2.4 Tailoring cation sublattice for fast Li-ion conduction	21
2.5 Conclusion and discussion.....	23
Chapter 3: Unravel discrepancies and develop evaluation metrics for machine learning interatomic potentials.....	29
3.1 Introduction.....	29
3.2 Discrepancies of MLPs on atomic dynamics between MLP and DFT occur even when low average errors are reported	33
3.3 Discrepancies in atomic dynamics.....	35
3.3.1 Rare events are sources of discrepancies	36
3.3.2 Configurations with similar energies are discrepancy sources	38
3.3.3 Vibration near defects is a discrepancy source	40
3.4 Development of error evaluation metrics	42
3.4.1 Quantifying the force errors on RE migrating atoms.....	44
3.4.2 Force performance scores as effective metrics	47
3.4.3 Process of developing error evaluation metrics	50
3.5 Trade-offs in MLP performances: Pareto fronts of MLPs.....	52
3.6 Conclusion and discussion.....	54
Chapter 4: Assessing the accuracy of MLPs in predicting the elemental ordering: a case study on Li-Al alloys.....	60
4.1 Introduction.....	60
4.2 Evaluating the performance of the MLP across the alloy phase diagram.....	63
4.3 Energy rankings of different orderings	662
4.4 Effect of training data diversity	71
4.5 Other errors in the applications of MLPs.....	74
4.5.1 Errors in large supercells	74
4.5.2 Phase diagram and phase transitions by MLP	75
4.5.3 Errors of forces on migrating atoms in MD simulations	77
4.6 Conclusion and discussion.....	79
Chapter 5: High-dimensional assessment of MLP performances for multiple properties.....	81
5.1 Introduction.....	81

5.2 Constructing the dataset of many MLPs and performance metrics	83
5.3 Finding challenging properties for MLPs to predict.....	88
5.4 Finding pairs of properties that are challenging	90
5.5 Conclusion	91
Chapter 6: Conclusions and future work	93
6.1 Conclusions.....	93
6.2 Future work.....	95
Publications, Patents, and Presentations	96
Appendices.....	98
Appendix A	98
Appendix B	108
Appendix C	115
Appendix D	148
Appendix E	157
Appendix F.....	179
Appendix G.....	194
Appendix H.....	201
Bibliography	211

Chapter 1: Introduction¹

Atomistic modeling, which simulates physical phenomena based on the interactions of atoms, is a crucial research technique in a wide range of disciplines including physics, chemistry, biology, and materials science. First-principles computation, such as density functional theory (DFT) calculation have been the standard technique for evaluating atom interactions among a diverse range of configurations and chemistries. However, their applications are limited to small system sizes of a few hundred atoms (up to a few nm) due to high computation costs.^[1–3] By contrast, classical interatomic potentials, also known as force fields, have significantly lower computation costs and thus can be employed for atomistic simulations with much larger length-scale (up to 10 nm to μm) and longer timescale (up to ns – μs), but they lack the transferability to different atomistic configurations that are not considered in the potential fitting.^[1,4,5] As an emerging technique to bridge the gaps among different computational techniques,^[1–3,6–10] machine learning interatomic potentials (MLPs) utilize machine learning (ML) models to predict energies and forces of atomistic structures, which are mapped into the atomistic descriptors as input. Current state-of-the-art MLPs are trained using DFT-calculated energies and forces from a diverse range of atomistic configurations, typically encompassing bulk and defected structures, equilibrium and non-equilibrium structures, and solid and liquid phases. Current state-of-the-art MLPs are claimed to achieve accuracies similar to ab initio calculations,^[1,5,11–16] while maintaining low computation costs and linear size scaling akin to classical interatomic potentials.

¹ Some text of this chapter is adapted from **Y. Liu**, S. Wang, A. M. Nolan, C. Ling, Y. Mo, *Adv. Energy Mater.* **2020**, *10*, 2002356. **Y. Liu**, X. He, Y. Mo, **2023**. arXiv:2306.11639

All three computational approaches, the ab initio calculations,^[17,18] the classical interatomic potentials,^[19] and the emerging MLP techniques,^[20] have been widely employed in studying the mechanisms of the materials at the atomistic dynamics level and conducting computational materials discovery for an extensive range of materials research and applications.

1.1 First-principles computation study of halide-based Li-ion conductors

A key application of atomic modeling is to design and discover novel materials with desired properties. The all solid-state Li-ion battery (ASSLB) is regarded as a key emerging rechargeable-battery technology with potential advantages over current Li-ion batteries including improved safety, higher energy density, and longer cycle life.^[21–26] Solid electrolyte (SE) materials, which replace the liquid electrolyte in current Li-ion batteries, are critical for enabling ASSLBs. SE materials are lithium superionic conductors (SICs) with high Li-ion conductivities $> 1 \text{ mS cm}^{-1}$ at room temperature (RT), comparable to those of liquid electrolytes.^[21,27,28] However, only a limited number of materials systems have been demonstrated as Li SICs.

First-principles computation studies have previously led to significant progresses in understanding Li-ion diffusion mechanisms and in designing and predicting new SIC materials.^[29–33] Recently, first-principles computation studies^[17,34] confirmed that the chloride chemistry in general provide favorable properties, including fast Li-ion diffusion, wider electrochemical window, good interface compatibility with cathode, and good air/moisture stability compared to sulfide SEs. With a combination of multiple desired properties, lithium halides, particularly chlorides, are a promising class of Li SICs for SEs in ASSLBs.

In Chapter 2, I performed a systematic first-principles computational study on lithium metal chloride systems. By using the high-throughput screening based on DFT calculations, we identified 18 novel materials as potential Li-ion conductors in lithium metal chloride systems.

Further, we investigated the mechanisms of Li diffusions in lithium metal chloride systems and proposed several design principles for Li-containing halide materials systems. Following our design principles, sparse cation configuration, low cation concentration, and low Li content are needed to achieve high Li-ion conductivities in Li halides.

1.2 Unravel discrepancies and develop evaluation metrics for MLPs

Though first-principles computation is an effective approach for studying the atomistic level mechanisms of materials and for discovering new materials discovery with improved properties as demonstrated in our previous results (Chapter 2), the high computation cost of the DFT calculations limits their applications for larger materials systems (up to a few hundred atoms and a few nm) or longer time scales (up to ns – μ s). In Chapter 2, all DFT calculations and the ab initio molecular dynamics (AIMD) simulations cost $\sim 10^6$ CPU core hours on dual Intel Ivy Bridge E5-2680v2 processors (2019) for less than ~ 100 materials. Currently, AIMD simulations for studying ion diffusion are still limited to small supercell size of the materials system, high temperatures, and short simulation durations,^[4,35] thus there are urgent needs in further lowering the computation costs of such simulations.

As an emerging computation approach for atomistic modeling, the MLP technique is a potential solution to overcome the obstruction of high computational cost. In addition to its low computation cost, the MLP models trained by DFT-calculated energies and forces are claimed to achieve near-DFT level accuracy in reproducing the energies and forces for a range of given atomic configurations.

However, as an emerging technology in its early stage of development, the MLPs are black-box predictors not directly based on physical principles. While MLPs are well-known to reproduce

the calculated values of energies and atomic forces, an open question is whether MLPs can always accurately reproduce physical phenomena in atomistic simulations. Conventional ML error testing primarily quantifies MLP accuracies through average errors, such as root-mean-square error (RMSE) or mean-absolute error (MAE), of energies and atomic forces across a range of configurations known as testing dataset. These atomistic configurations in the testing dataset are randomly split from the entire datasets generated in the same manner as the training dataset, and thus are similar to the training dataset but may differ from the atomistic configurations that may encounter during MD simulations. Most MLPs are reported to achieve small, average errors of energies and atomic forces as low as 1 meV/atom and 0.05 eV/Å,^[1,5,8,11,36,37] respectively, in conventional ML testing. These MLPs with small average errors may not always accurately reproduce the physical phenomena in atomistic simulations, as shown by quite a few previous studies.^[10,38,39]

In Chapter 3, I conduct a systematic testing on MLP prediction of atomic dynamics and physical properties, compared with DFT calculations. We reveal several significant discrepancies between AIMD and MLP-MD simulations in a collection of MLPs, leading to considerable errors between true and predicted physical properties. To address the observed discrepancies, we develop a novel evaluation metric to assess the atomic dynamical performances of MLPs. We integrated the newly developed metric in the training process to improve the performance of MLPs. By developing improved evaluation metric, our results highlight general guidance to develop future reliable MLPs.

1.3 Assessing the accuracy of MLPs in predicting the elemental ordering: a case study on Li-Al alloys

Studying the thermodynamics of materials systems requires sampling a large number of complex atomistic configurations. The investigation of thermodynamics in binary alloy materials systems holds significant interest, particularly using atomistic modeling techniques. However, a key challenge is that the computations of a large number of elementary orderings are usually computationally too expensive to be carried out by DFT calculations. Furthermore, it is also a great challenge for interatomic potentials to fully capture a variety of phases with a wide composition range and complex elemental orderings even for a binary alloy materials system.

As an emerging technique to bridge the gap between these two computation approaches, MLP technique has great potential to explore the complex orderings of different phases in alloy systems. Since MLPs are trained by a limited number of configurations with pre-selected lattice structures and compositions, it is still an open question about the capability of MLP to make accurate predictions on lattice structures, compositions, and the ordering of atomic sites outside the training data, which can be considered as the transferability of MLPs.^[2,40–43] Thus, it is critical to assess the MLPs for their performances on different lattice structures and compositions in binary alloy materials systems, especially for alloys outside the training data.

In Chapter 4, I test how MLPs can reproduce complex orderings of different phases in alloy systems. I train a set of MLPs for Li-Al materials system and study the MLP performances on Li-Al binary alloys that are not considered in the training data. To conduct a thorough examination of such performances, we assess the MLPs using a range of evaluation metrics, including both conventional energy errors and several newly developed metrics. We find that MLPs trained by even a few materials may have good performances on materials outside the training data. In

addition, we also observe that MLPs, re-trained by dataset with more diversity but similar amount of data, show worse performance than original MLPs.

1.4 High-dimensional assessment of MLP performances for multiple properties

In previous chapters, we assess MLP performances on atomistic dynamics and elemental ordering in many phases. While these assessments provide critical and comprehensive insights on MLP predictions, they mostly focus on individual properties or perspectives separately. Given that we clearly observe trade-offs between different predicted properties in Chapter 3, as a follow-up our studies in Chapter 3 and 4, I conduct a systematic and comprehensive analysis the MLPs based on many evaluation metrics in Chapter 5.

Since previous studies clearly show that conventional metrics based on errors of energies and forces are inadequate to fully describe discrepancies in MLP performances, as shown in Chapter 3 and 4 and Fu et al.^[44]. We aim to find 1) evaluation metrics or physical properties that are challenging to predict separately, and 2) sets of metrics or properties that are challenging to achieve high prediction accuracies at the same time. As a result, we find that four major categories of physical properties, such as energy rankings, formation energies for defects, elastic constants, and force predictions on rare-event (RE) atoms, are challenging to predict in Chapter 5. We also highlight our work in Chapter 5 that, instead of analyzing a few MLPs, we conduct an empirical analysis of MLP performances systematically using a few thousand MLPs. This approach of statistical analysis provides additional insights of MLP predictions and opens probabilities to perform complicated analysis in high-dimensional space. We summarize our statistical analysis procedures and propose a standard process to conduct future empirical analysis.

Chapter 2: First-principles computation study for halide-based Li-ion conductors²

2.1 Introduction

The all-solid-state Li-ion battery (ASSLB) is regarded as the next-generation rechargeable-battery technologies. The solid electrolyte (SE) materials based on Li superionic conductor (SIC) is a critical component in enabling ASSLB.^[18,32] Currently, Li SICs, including Li₁₀GeP₂S₁₂ (LGPS),^[24] garnet Li₇La₃Zr₂O₁₂ (LLZO),^[45,46] and NASICON Li_{1.3}Al_{0.3}Ti_{1.7}(PO₄)₃ (LATP),^[47] are limited and they don't have all properties required for ASSLBs. Compared to oxide SICs, sulfide SICs have desired mechanical deformability for forming good interface physical contacts in ASSLB cell assembly, but have narrow electrochemical window,^[48,49] poor cathode interface compatibility,^[50–52] and poor air/moisture stability,^[25,34] which impede the large-scale commercialization of sulfide-based ASSLBs. Asano et al. discovered Li₃YCl₆ (LYC) and Li₃YBr₆ (LYB) as new Li SICs with high Li-ion conductivities on the order of 1 mS cm⁻¹ at RT and with good ASSLB cell performances^[53]. In addition to the desired deformable mechanical properties as sulfides, first-principles computation studies^[17,34] confirmed that the chloride chemistry in general gives a lower barrier for Li-ion migration, wider electrochemical window, good interface compatibility with cathode, and good air/moisture stability compared to sulfide SEs. With a combination of multiple desired properties, lithium halides, particularly chlorides, are a promising class of Li SICs for SEs in ASSLBs.

The chloride anion chemistry is advantageous for Li-ion migration, thanks to the relatively large anion radius, large anion polarizability, and weak interaction with Li-ion.^[17,30,53] First-

² Some text of this chapter is adapted from

Y. Liu, S. Wang, A. M. Nolan, C. Ling, Y. Mo, *Adv. Energy Mater.* **2020**, *10*, 2002356.

principles computation confirms that Li-ion migration exhibits a low energy barrier of 0.28 eV in face-centered cubic (fcc) and of 0.29 eV in hexagonal close-packed (hcp) Cl^- anion sublattices with no cation under typical lattice volume of lithium chlorides.^[17] Recent experimental studies reported a series of Li-containing chlorides Li_3MCl_6 ($\text{M} = \text{In}, \text{Er}, \text{Sc}$) and their doped variations that achieved Li-ion conductivities on the order of 1 mS cm^{-1} at RT.^[17,54–58] While the discovery of new SIC systems in oxides and sulfides is greatly limited by the unique crystal structures required for achieving fast Li-ion conduction^[29,32,59], Li-containing chlorides with common face-centered cubic (fcc) and hexagonal close packed (hcp) anion sublattices exhibit adequately low Li-ion migration barrier and are a promising chemical space for new Li-ion conductors.

A systematic fundamental understanding of chloride Li-ion conductors is essential for guiding the discovery and design new Li-containing chloride SICs. While Li-containing chlorides are common in close-packed fcc and hcp anion sublattices, they exhibit a range of compositions with different Li/cation configurations, which have significant effects on ion conduction. For example, some earlier studied Li-containing chlorides, such as Li_2CdCl_4 and Li_2MgCl_4 , were reported to have low Li-ion conductivities on the order of $10^{-6} \text{ S cm}^{-1}$ at RT ^[60–63], though these chlorides have the same close-packed fcc or hcp anion frameworks as Li_3MCl_6 ($\text{M} = \text{In}, \text{Er}, \text{Sc}$) SICs. Earlier reports of Li_3InCl_6 and Li_3YCl_6 also show much lower Li-ion conductivity on the order of $10^{-5} \text{ S cm}^{-1}$, than recently measurements on the order of $10^{-3} \text{ S cm}^{-1}$ with the same composition.^[53,54,64] It is not understood why different configurations and concentrations of Li/metal cations in Li_2MCl_4 ($\text{M} = \text{Cd}, \text{Mg}$) exhibit much lower ion conductivities than newly reported Li_3MCl_6 ($\text{M} = \text{In}, \text{Er}, \text{Sc}$) SIC systems with the same anion sublattice. In addition, doping and substitutions on these chloride SICs to tune Li concentration and cation concentration were demonstrated to further improve ionic conductivities.^[55–58] It is not clear why certain doping

strategies lead to improved Li-ion conductivity. To rationally guide future materials design of new halide Li-ion conductors, a scientific understanding about the effects of cation concentration, cation configuration, and Li concentration on Li-ion conduction is needed.

First-principles computation studies have led to significant progresses in understanding Li-ion diffusion mechanisms and in designing and predicting new SIC materials.^[29,30,32,33,65] In this study, we perform a systematic first-principles computational study on lithium metal chloride systems spanning a wide range of Li/cation concentrations and configurations to understand the key factors determining fast Li-ion diffusion in these chloride systems. Using first-principles calculations, we study Li-ion diffusion in over seventy Li-containing chlorides from the Inorganic Crystal Structure Database (ICSD) and identify 19 materials as potential Li-ion conductors. By analyzing a large set of computation data generated on a wide range of chloride materials with different levels of doping, our computation reveals that cation configurations and Li contents significantly affect the Li-ion conductivity of chloride materials. From this understanding, we propose design principles that sparse cation configuration, low cation concentration, and low Li content increase ionic conductivity in Li chlorides. Following these design principles, we artificially modify the cation lattice in a few poor Li-ion conductors and achieve increased Li-ion conductivities. Our study identifies multiple novel chloride systems as fast Li-ion conductors. Given the similarities among halide anion chemistries, our results provide the guidance to tailor cation sublattice for designing novel Li halide SICs.

2.2 High-throughput computation of chloride Li-ion conductors

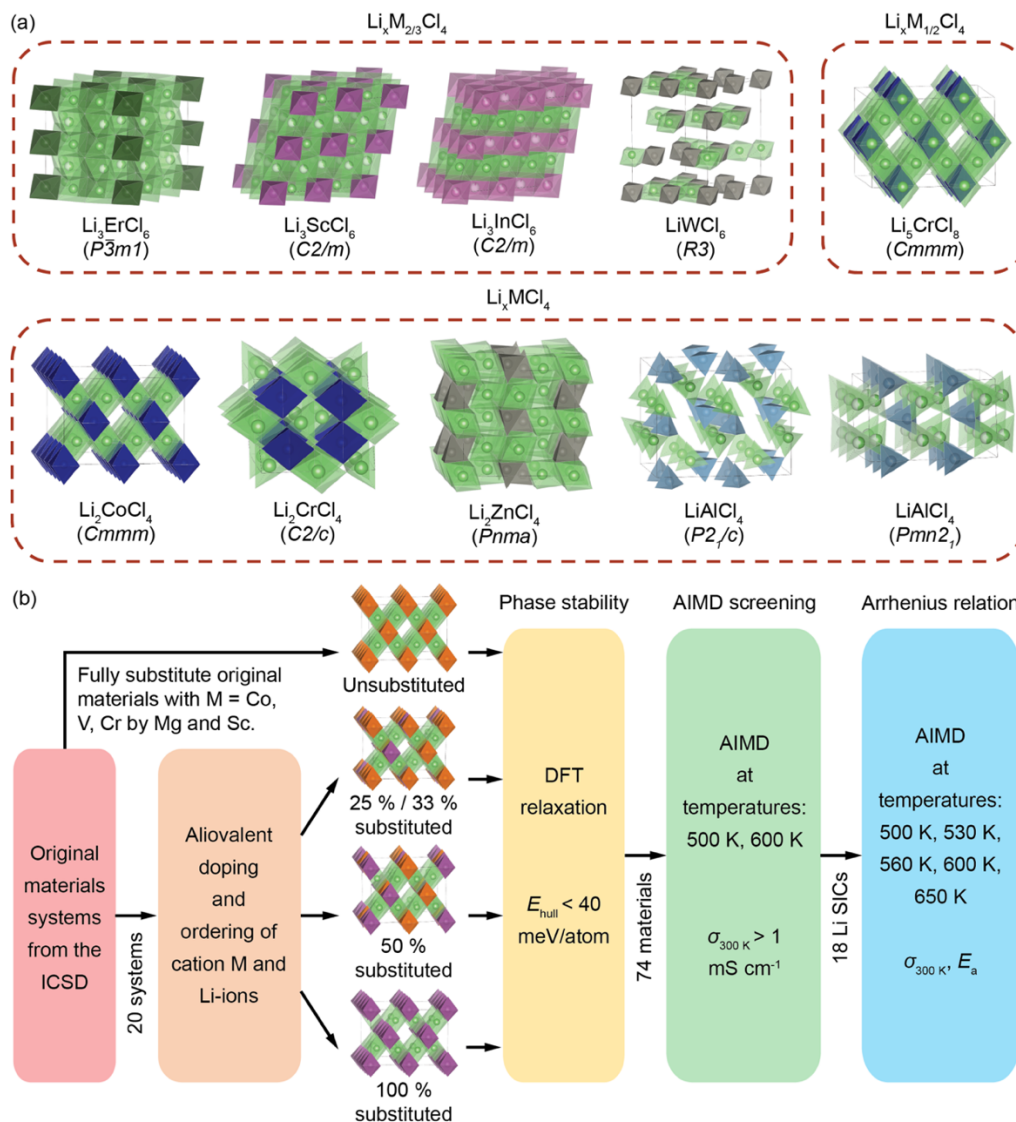


Figure 2.1. Crystal structures of chlorides and the computation scheme. (a) Example crystal structures $\text{Li}_x\text{M}_{2/3}\text{Cl}_4$, $\text{Li}_x\text{M}_{1/2}\text{Cl}_4$, and Li_xMCl_4 materials systems, with LiCl_6 octahedra and LiCl_4 tetrahedra colored green, and cation MCl_6 octahedra and MCl_4 tetrahedra in different colors. (b) The high-throughput computation workflow for studying halide Li-ion conductors.

From a total of 202 Li-containing chlorides in the ICSD, we selected 20 chloride systems that have fcc and hcp anion frameworks (14 fcc and 6 hcp) with unique cation sublattices (Figure A1 and Table A1 in Appendix A). The selection steps are described in the Appendix B. In these 20 candidate systems, there are six compositions, LiMCl_4 , Li_2MCl_4 , Li_3MCl_6 , LiMCl_6 , Li_5MCl_8 ,

and Li_6MCl_8 , where M is a non-Li metal cation. Examples of crystal structures are shown in **Figure 2.1a**. In addition, we generated a large number of compounds with different Li contents through aliovalent substitution, which is a common strategy to increase Li-ion conductivities.^[29,57] For each materials system, we substituted original M cations by aliovalent elements with different levels of substitutions (e.g. 25%/33%, 50% and 100%) to either create Li-ion vacancies or insert extra Li-ions into empty sites (if available). The doping elements were selected according to valence, ionic radius, and the substitution probability based on Ref.^[66] (Appendix B). The structures of substituted materials were generated following the ordering procedure as in Ref.^[29,35,67] (Appendix B). The energies of these substituted materials were evaluated in density functional theory (DFT) calculations, and the substituted materials with calculated energy above the hull $E_{\text{hull}} < 40$ meV/atom (**Figure 2.2** and Table A2) were further investigated for Li-ion diffusion.

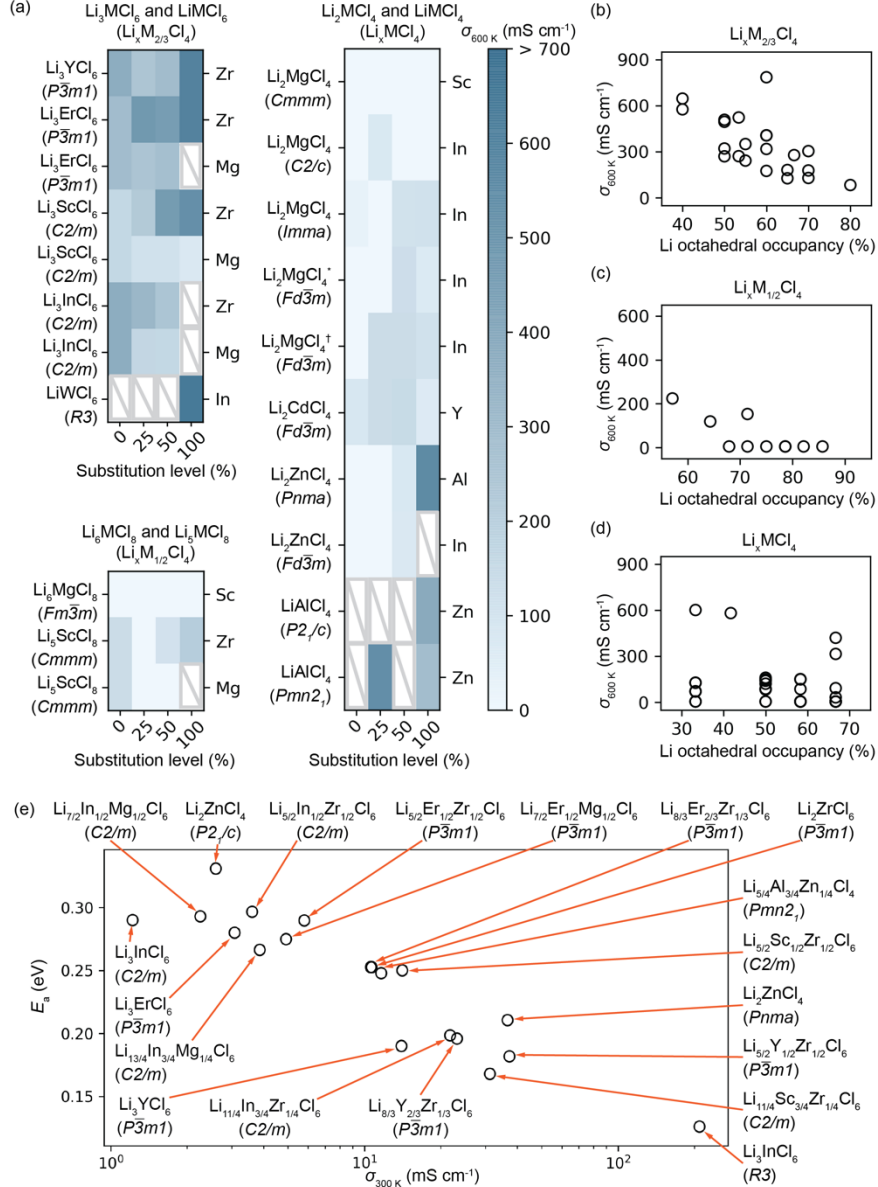


Figure 2.2. Li-ion conduction in chlorides by AIMD simulations. (a) Li-ion conductivities of Li-containing chlorides at 600 K. Each box in the heatmap corresponds to a compound with the original material (left) substituted by the aliovalent dopant (right) with different substitution levels (For the structures in $P\bar{3}m1$, a substitution level of 1/3 is used instead of 25%), and its Li-ion conductivity at 600 K from AIMD simulations are shown in color. Crossed grey boxes indicate predicted materials with poor phase stability ($E_{\text{hull}} > 40$ meV/atom) or significant structure change (e.g., melting) during AIMD simulations. $\text{Li}_2\text{MgCl}_4^*$ and $\text{Li}_2\text{MgCl}_4^\dagger$ correspond to Li_2MgCl_4 (ICSD-73229) and Li_2MgCl_4 (ICSD-74957), respectively. Materials are grouped by their renormalized compositions, Li_xMCl_4 , $\text{Li}_x\text{M}_{2/3}\text{Cl}_4$, and $\text{Li}_x\text{M}_{1/2}\text{Cl}_4$. Li-ion conductivities versus Li octahedral occupancy in renormalized composition (b) $\text{Li}_x\text{M}_{2/3}\text{Cl}_4$, (c) $\text{Li}_x\text{M}_{1/2}\text{Cl}_4$, or (d) Li_xMCl_4 . (e) The activation energies E_a and the extrapolated Li-ion conductivities at 300 K of potential Li chloride SICs predicted by the AIMD simulations.

For the candidate materials and their substituted materials with decent phase stability, we performed AIMD simulations to evaluate their Li-ion conductivities (Table A2). To identify Li-ion conductors, we employed a screening scheme as in Ref.^[29,68], in which we extrapolated the Li-ion conductivities at 300K using two higher temperatures (500 K and 600 K). For the candidates with high extrapolated conductivities at 300 K, we performed AIMD simulations at five temperatures (500 K, 530 K, 560 K, 600 K, and 650 K) to obtain more accurate Arrhenius relations of Li-ion conductivities on temperature and corresponding activation energies (**Figure 2.2e**, Figure A2, and Table A3). 19 materials were identified as potential Li SICs with Li-ion conductivities higher than 1 mS cm^{-1} at 300 K (Figure 2.2e) with wide electrochemical window (Figure A3) and large band gaps between 2 eV and 7 eV calculated by DFT hybrid functional calculations (Table A3). The hypothetical Li_3InCl_6 compound generated by a full In-to-W substitution in the LiWCl_6 (*R3*) structure with increased lithium content has an exceptional predicted Li-ion conductivity on the order of $10^{-1} \text{ S cm}^{-1}$ at 300 K. Li_3YCl_6 ($P\bar{3}m1$), Li_3ErCl_6 ($P\bar{3}m1$), Li_3InCl_6 (*C2/m*) and Li_3ScCl_6 (*C2/m*) are all confirmed as Li SICs, in agreement with previous experimental reports. Our calculated Li-ion conductivity of Li_3InCl_6 (*C2/m*) is 1.2 mS cm^{-1} , in good agreement with the experimental reported value of 1.49 mS cm^{-1} .^[53–56] A recent experiment study reports Li-ion conductivity of Li_3YCl_6 is reported to be 0.51 mS cm^{-1} , which is slightly lower than our calculated Li-ion conductivity of 14 mS cm^{-1} , with error bounds in the range of 5 to 47 mS cm^{-1} and the difference may be attributed to the existence of grain boundary^[53] and blocking defects.^[17] In addition, a wide range of Zr-doped Li_3MCl_6 were reported with greatly enhanced Li-ion conductivity, for example $\text{Li}_{5/2}\text{Y}_{1/2}\text{Zr}_{1/2}\text{Cl}_6$ ($P\bar{3}m1$), $\text{Li}_{11/4}\text{Sc}_{3/4}\text{Zr}_{1/4}\text{Cl}_6$ (*C2/m*), and $\text{Li}_{8/3}\text{Y}_{2/3}\text{Zr}_{1/3}\text{Cl}_6$ ($P\bar{3}m1$). In addition to confirming Li_3MCl_6 as a promising family of Li SICs, these computation

results suggest that M^{4+} substitution to create Li vacancies is an effective strategy to further increase the ionic conductivity.

Given that all studied chlorides have close-packed fcc or hcp Cl^- anion sublattices, the differences in Li-ion diffusional properties are mostly a result of different cation configurations and Li concentrations. We categorize all materials according to renormalized compositions into Li_xMCl_4 (corresponding to $LiMCl_4$ and Li_2MCl_4), $Li_xM_{2/3}Cl_4$ (corresponding to Li_3MCl_6 and $LiMCl_6$), and $Li_xM_{1/2}Cl_4$ (corresponding to Li_5MCl_8 and Li_6MCl_8), since most of currently known Li-containing chlorides have Li_xMCl_4 compositions. These renormalized compositions allow a straightforward comparison of the concentration of Li-ions and M cations, and the Li-ion diffusional properties vary among different renormalized composition groups. The $Li_xM_{2/3}Cl_4$ materials (i.e., unsubstituted and substituted Li_3MCl_6) in general exhibit the highest Li-ion conductivities (Figure 2.2). Among all compositions studied, 18 of 28 $Li_xM_{2/3}Cl_4$ materials have Li-ion conductivities higher than 200 mS cm^{-1} at 600 K, while only four of 47 Li_xMCl_4 materials and one of 11 $Li_xM_{1/2}Cl_4$ materials can achieve a similar level of Li-ion conductivity at 600 K. Among the final 19 identified Li SICs, 16 $Li_xM_{2/3}Cl_4$ materials are identified to have Li-ion conductivity of $> 1 \text{ mS cm}^{-1}$ at RT (Figure 2.2e). Our results further confirm recent computation and experimental studies that Li_3MCl_6 are promising Li SICs for a wide range of cation and substitutions.^[17,30,53–58] Most Li_xMCl_4 and $Li_xM_{1/2}Cl_4$ materials even with substantial levels of substitution have Li-ion conductivities much lower than 200 mS cm^{-1} at 600 K, and 24 of 51 materials exhibit a negligible amount of Li-ion hopping to obtain a statistically reliable ionic diffusivity. There are a few Li-ion conductors as the exceptions in the Li_xMCl_4 category, as a result of their distinct local cation coordination, which are further discussed in the Chapter 2.5.

Our results indicate that the Li content is an important factor for Li-ion conduction. In the fcc and hcp anion sublattices of these materials, the total number of octahedral sites for Li/M cations is the same as the number of anions N_{Cl} , and the percentile of Li-ions over the octahedral sites available for Li-ion hopping can be a good quantification for Li content. Here, we define the Li octahedral occupancy $f_{\text{Oct}}^{\text{Li}} = \frac{N_{\text{Li}}}{N_{\text{Cl}} - N_{\text{M}}}$ as a ratio of the number of Li-ions N_{Li} over the available octahedral sites excluding those N_{M} sites occupied by M cations. As the Li octahedral occupancy $f_{\text{Oct}}^{\text{Li}}$ decreases, there is an increased number of available octahedral sites for Li-ion hopping, resulting in increased Li-ion conductivity at lower Li octahedral occupancy (Figure 2.2b - d). In $\text{Li}_x\text{M}_{2/3}\text{Cl}_4$, Li-ion conductivities at 600 K increase with decreasing Li octahedral occupancy $f_{\text{Oct}}^{\text{Li}}$ until 40% (Figure 2.2d and A4a). Among the limited number of materials tested, the materials in $\text{Li}_x\text{M}_{1/2}\text{Cl}_4$ follow a similar trend at 600 K, as a few doped $\text{Li}_x\text{M}_{1/2}\text{Cl}_4$ compositions exhibit higher Li-ion conductivity at low Li octahedral occupancy, and most exhibit low conductivity at higher Li octahedral occupancy (Figure 2.2c and A4b). Li-ion conductivities of $\text{Li}_x\text{M}_{2/3}\text{Cl}_4$ at 500 K have a similar trend of increased Li-ion conduction as Li octahedral occupancy $f_{\text{Oct}}^{\text{Li}}$ decreases until 50%, and a slight decline of Li-ion conductivity at 500 K occurs when $f_{\text{Oct}}^{\text{Li}}$ decreases below 50% (Figure A4a), suggesting an increase in activation energy at low Li octahedral occupancy. In contrast to $\text{Li}_x\text{M}_{2/3}\text{Cl}_4$, Li-ion conductivities of Li_xMCl_4 materials are low across a wide range of Li octahedral occupancies even for highly doped materials (Figure 2.2a, d, and A4c), except for a few exceptions mentioned above. Given the same fcc/hcp anion sublattice in all these materials, the low Li-ion conductivity in Li_xMCl_4 compared to the high Li-ion conductivity in $\text{Li}_x\text{M}_{2/3}\text{Cl}_4$ (i.e., Li_3MCl_6) across a similar range of Li octahedral occupancies can be attributed to different cation concentrations and configurations. Understanding these cation effects on ion conduction is important for guiding the design of novel halide SICs.

2.3 Effect of cation sublattice on Li-ion migration

To understand the effect of the M cation sublattice on Li-ion diffusion, we performed NEB calculations to evaluate Li-ion migration barriers ΔE_b in 20 materials with different levels of cation substitutions from six representative material systems, Li_6MgCl_8 ($Fm\bar{3}m$), Li_2MgCl_4 ($Cmmm$), Li_2MgCl_4 ($Fd\bar{3}m$), Li_5ScCl_8 ($Cmmm$), Li_3ErCl_6 ($P\bar{3}m1$), and Li_3ScCl_6 ($C2/m$) (Table A5). In line with the Li-ion conductivities from AIMD simulations, Li-ion migration energies of substituted Li_xMCl_4 (between 0.3 eV and 0.5 eV) are higher than those in substituted $\text{Li}_x\text{M}_{2/3}\text{Cl}_4$ (between 0.1 eV and 0.3 eV). For all original unsubstituted materials, the Li-ion migration mediated by a vacancy has an energy barrier on the order of 0.2 eV, which is consistent with the low intrinsic barrier in close-packed Cl^- anion lattice found in previous computation study.^[17] However, in unsubstituted Li_xMCl_4 materials, empty octahedron sites for Li-ion migration are blocked by M cations (see *Blocking Effect* in Appendix A), leading to low ionic conduction in unsubstituted Li_xMCl_4 compounds. Even in the substituted Li_xMCl_4 materials with empty octahedron sites created for Li-ion migration, Li-ion migration barriers ΔE_b increase significantly, consistent with the low Li-ion conductivity in substituted Li_xMCl_4 materials from AIMD simulations (Figure 2.2). The high energy barriers of Li-ion migration ΔE_b in cation substituted Li_xMCl_4 are caused by increased energy differences between Li sites (**Figure 2.3d**), which include initial octahedral (oct), intermediate tetrahedral (tet) and final oct sites, along Li-ion migration pathways. In unsubstituted materials, the site energy differences between initial and final oct sites ΔE_f are almost negligible, leading to low ΔE_b at around 0.2 eV (Figure 2.3d), and the energy barrier mainly comes from site energy difference between initial oct and intermediate tet sites ΔE_m . In contrast, for substituted Li_xMCl_4 materials, site energy differences ΔE_f increase from 0 eV in unsubstituted materials by 0.05 to 0.4 eV (by 0.18 eV on average), and ΔE_m change by -0.1 to 0.3 eV (by 0.1 eV on average)

(Figure A7), lifting the overall migration barrier ΔE_b to above 0.3 eV. The statistical trends based on a large number of Li-ion migration pathways (Figure 2.3e and f) confirm that the migration barrier ΔE_b and site energy differences (ΔE_f and ΔE_m) are strongly correlated, and thus increased site energy differences cause increased migration barriers in substituted Li-containing chloride materials.

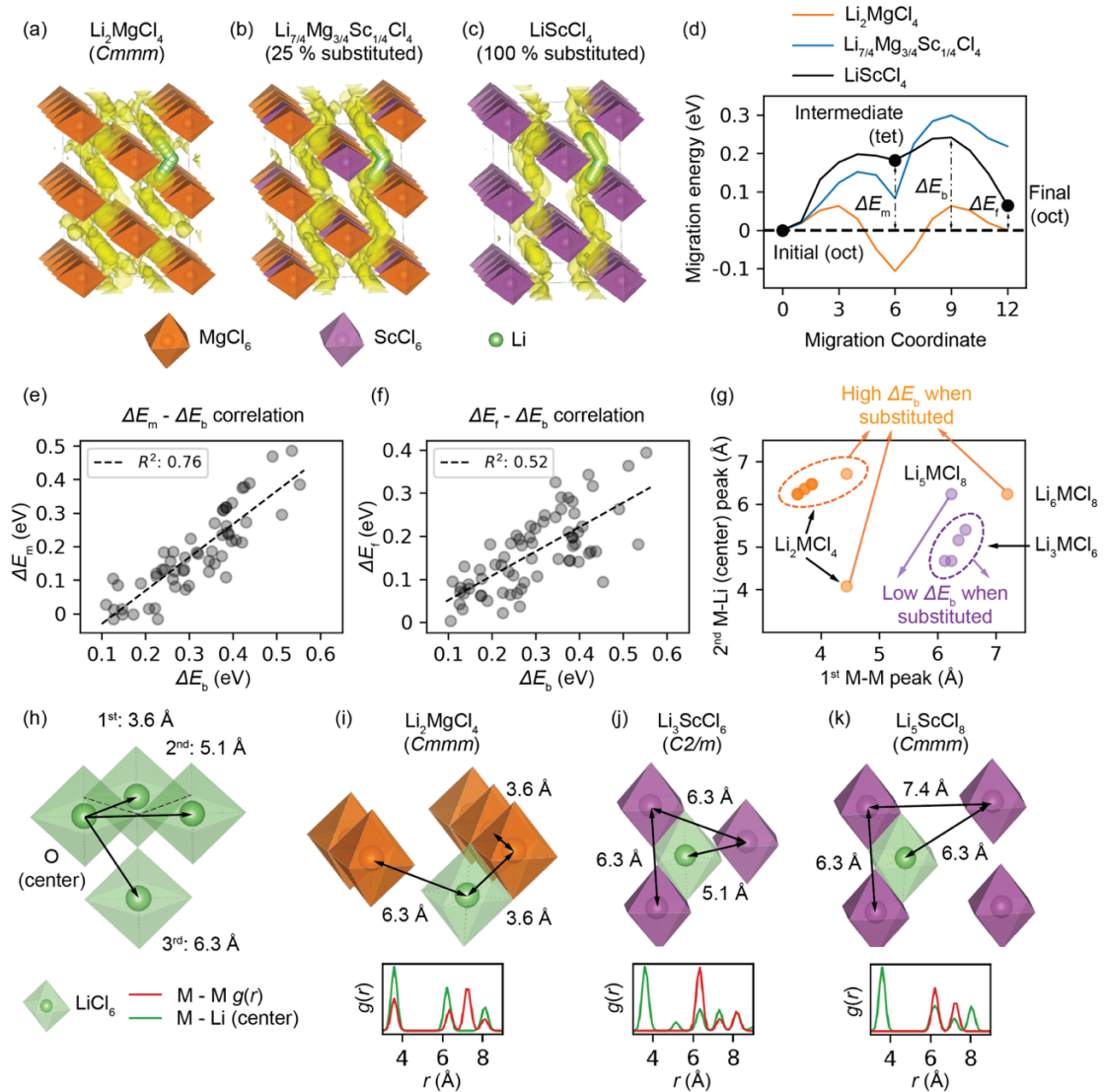


Figure 2.3. Effect of cation sublattice on Li-ion migration. Li-ion probability density (yellow) from AIMD simulations at 600 K in crystal structures of (a) original, (b) 25% Sc-substituted, and (c) 100% Sc-substituted Li_2MgCl_4 ($Cmmm$) and Li-ion migration pathways (green) from NEB

calculations. (d) Energy profile of corresponding migration pathways in a)–c) from NEB methods. Correlations between migration energy barrier ΔE_b with site energy differences (e) with intermediate site ΔE_m and (f) with final site ΔE_f for all migration pathways. (g) The 2nd M-Li peak and the 1st M-M peak of RDF $g(r)$ differentiate the materials with high barriers (orange) versus low barriers (purple). (h) The configurations of LiCl_6 or MCl_6 octahedral sites in fcc Cl^- sublattice in (i) Li_2MgCl_4 ($Cmmm$), (j) Li_3ScCl_6 ($C2/m$), and (k) Li_5ScCl_8 ($Cmmm$), with corresponding RDF $g(r)$ of M cations (red) and M cations surrounding Li centers (green).

According to our analyses of the crystal structures of unsubstituted and substituted materials, the increase of site energy differences is caused by the surrounding cation configurations. Unsubstituted materials have low ΔE_f because initial and final oct sites have identical cation environments (Figure 2.3a and d). However, in partially substituted materials, the cation environments of the initial and final oct sites are different due to the different valence of the substituting cation, causing an increase in the site energy differences (Figure 2.3b and d). Though fully substituted materials have only a single type of cation, the observed increase in ΔE_f is caused by the change of surrounding Li configurations (Figure 2.3c and d). In summary, the configurations of cations affect the energies of Li sites and hence the migration barrier.^[69,70]

In addition, we performed quantitative analyses of the cation distributions in the structures with high and low energy barriers, which are Li_xMCl_4 and $\text{Li}_x\text{M}_{2/3}\text{Cl}_4$, respectively, to understand the differences in cation sublattice. We analyzed the radial distribution function (RDF) $g(r)$ of all M cations (red curves in Figure. 2.3i - k) and of cations surrounding Li-ions (green curves, Figure 2.3i - k) on migration paths (Appendix B) and obtained a range of features (See *Quantitative Analysis of Surrounding Environments* in Appendix A) describing the cation configuration. We performed a systematic regression analysis of ΔE_f and ΔE_m from many migration pathways on these features of the cation configuration, and the key descriptors are the 2nd M-Li peak in the $g(r)$ of M cations surrounding Li-ions (green curves, Figure 2.3i - k) and the 1st peak in $g(r)$ of M cation

sublattice (red curves, Figure 2.3i - k). Therefore, compared to the Li_xMCl_4 structures with shorter cation distances, in general, lower cation concentration and sparse cation distribution, such as in $\text{Li}_x\text{M}_{2/3}\text{Cl}_4$, give lower migration barriers even with a substantial level of cation substitution (Figure 2.3g).

2.4 Tailoring cation sublattice for fast Li-ion conduction

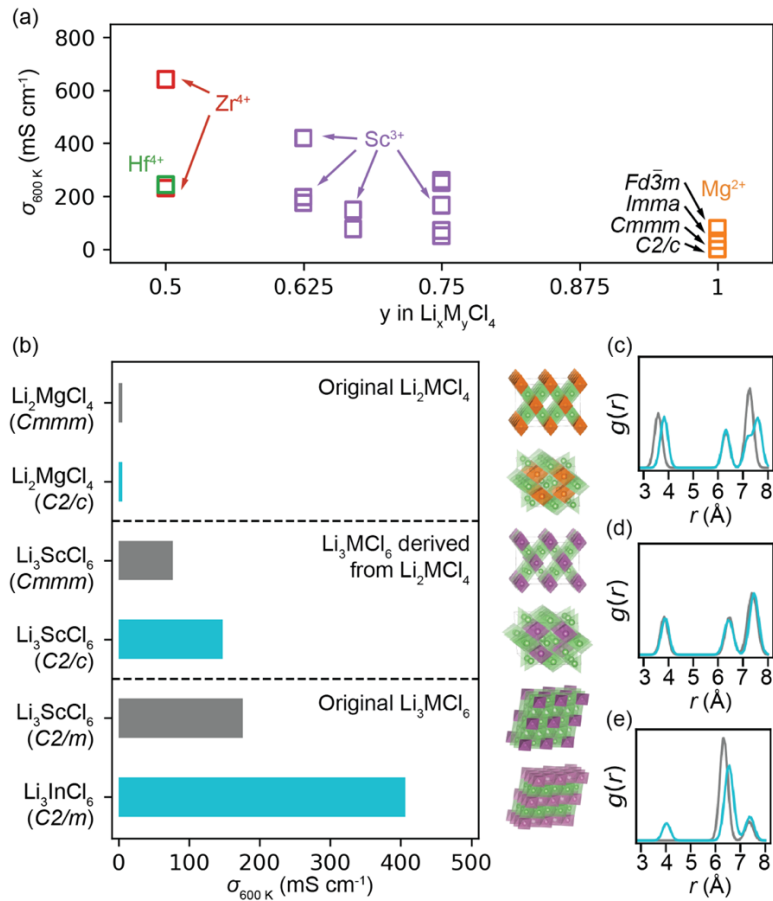


Figure 2.4. (a) Li-ion conductivities at 600 K as a function of M cation concentration y in artificial Li-ion conductors $\text{Li}_x\text{M}_y\text{Cl}_4$, $\text{M} = \text{Sc}^{3+}$ (purple), Zr^{4+} (red), and Hf^{4+} (green), substituted with reduced M cation concentration from original Li_2MgCl_4 materials (orange). (b) Li-ion conductivities at 600 K of (upper) unsubstituted Li_2MgCl_4 , (middle) artificial Li-ion conductors derived from Li_2MgCl_4 with reduced cation concentrations, and (lower) original Li_3MCl_6 materials, along with their crystal structures (right) and (d)–(g) RDF $g(r)$ of M cations.

We constructed a number of hypothetical chloride materials with reduced cation concentrations to further investigate the effect of cation sublattice on Li-ion conduction. From poor Li-ion conductors Li_2MCl_4 with fcc anion framework, such as Li_2MgCl_4 ($Fd\bar{3}m$), Li_2MgCl_4 ($Imma$), Li_2MgCl_4 ($C2/c$) and Li_2MgCl_4 ($Cmmm$), which have Li-ion conductivities at 600K lower than 80 mS cm^{-1} , we constructed hypothetical $\text{Li}_x\text{M}_y\text{Cl}_4$ with higher-valence $\text{M}^{3+}/\text{M}^{4+}$ cations at reduced concentrations $y = 0.5 - 0.75$, such as $\text{Li}_x\text{M}_{1/2}\text{Cl}_4$, $\text{Li}_x\text{M}_{5/8}\text{Cl}_4$ or $\text{Li}_x\text{M}_{2/3}\text{Cl}_4$, and $\text{Li}_x\text{M}_{3/4}\text{Cl}_4$, ($\text{M} = \text{Sc}^{3+}, \text{Zr}^{4+}, \text{Hf}^{4+}$), in which the M cation sublattice was the same as in Li_2MgCl_4 with a partial occupancy y and the Li content x was adjusted for charge balance following a similar approach used in Chapter 2.2 (Appendix B). By comparing the Li-ion conductivities between these hypothetical $\text{Li}_x\text{M}_y\text{Cl}_4$ materials and original Li_2MCl_4 materials, we can differentiate how cation concentration and cation configuration (as shown by the $g(r)$, **Figure 2.4c, d, and e**) affect Li-ion conduction. The AIMD simulations show that all hypothetical materials with the decreased cation concentration have significantly increased Li-ion conductivities than the original Li_2MCl_4 materials. Nine of 13 hypothetical Li-ion conductors showed higher Li-ion conductivities of $> 150 \text{ mS cm}^{-1}$ at 600 K compared to those of the original Li_xMCl_4 materials (Figure 2.4a and Table A6). All six materials in $\text{Li}_x\text{M}_{1/2}\text{Cl}_4$ and $\text{Li}_x\text{M}_{5/8}\text{Cl}_4$ compositions with most reduced cation concentrations have the Li-ion conductivities near or higher than 200 mS cm^{-1} at 600 K. In agreement with our computation, a recent experimental study of a halospinel material $\text{Li}_2\text{Sc}_{2/3}\text{Cl}_4$ derived from the structure of Li_2MgCl_4 ($Fd\bar{3}m$) achieves high Li-ion conductivity of 1.5 mS cm^{-1} at RT, significantly higher than the original Li_2MgCl_4 .^[58] The increased ionic conductivities of these derived Li-ion conductors with reduced cation concentration in Li_xMCl_4 confirm that low cation concentrations benefits Li-ion conduction in halides.

In addition, we compared materials with identical compositions but different cation configurations, in order to identify the effect of cation configuration on Li-ion conduction. We compared the above two hypothetical Li_3ScCl_6 materials derived from Li_2MgCl_4 with $C2/c$ and $Cmmm$ space groups, to the original Li_3ScCl_6 ($C2/m$) and Li_3InCl_6 ($C2/m$) (Figure 2.4b -e), which have highly distinct cation configurations, as indicated by $g(r)$ and crystal structures (Figure 2.4). The Li-ion conductivities of our hypothetical Li-ion conductors are around 100 mS cm^{-1} at 600 K, significantly lower than the Li-ion conductivities of existing Li_3MCl_6 materials at 176 mS cm^{-1} and 406 mS cm^{-1} at 600 K. Therefore, original Li_3MCl_6 materials with sparse cation distribution have higher Li-ion conductivities than hypothetical Li_3MCl_6 materials with denser cation distributions derived from poor Li-ion conducting Li_2MCl_4 . These results confirm that sparse cation distribution is beneficial for ion conduction in Li-containing chlorides.

2.5 Conclusion and discussion

Using first-principles computation, we studied Li-ion conduction in Li-containing chlorides from the ICSD with fcc and hcp anion frameworks. The Li-ion conduction mechanisms were revealed in these materials, with compositions Li_3MCl_6 , Li_2MCl_4 , and Li_5MCl_8 , and their substituted compounds with a wide range of lithium or cation concentrations. By analyzing the trend of Li-ion conductivity among these materials, we revealed that Li content and cation sublattices (including concentration and distribution) are crucial in determining Li-ion conduction in chloride materials, and the trends with respect to these factors are proposed and demonstrated as design principles for chloride fast Li-ion conductors. Given the similarity among halide anions, these understandings and design principles are likely to be applicable in other halide systems. Among these large number of Li-containing chlorides analyzed, Li-ion conductivities generally

increase as Li content decreases, and the optimal Li octahedral occupancy is around 50%, in which Li-ions occupy around half of the octahedral sites excluding M cations and the other half of the sites remain accessible for Li-ion hopping. Our computation confirm Li_3MCl_6 are generally fast Li-ion conductors and have higher Li-ion conductivities than Li_xMCl_4 , which tend to have high energy barriers for Li-ion migration in cation-substituted compounds. According to the analyses of a large number of materials, including hypothetically constructed materials, low cation concentration and sparse cation distributions are advantageous for fast Li-ion conduction. Based on the new understanding from our results and analyses, three principles for designing halide Li-ion conductors are 1) decrease Li octahedral occupancy to have a significant fraction of available sites (approximately 40% to 60%), 2) decrease cation concentrations, and 3) increase cation distances for sparse cation distribution.

Consistent with previous experimental and computational studies, most Li_3MCl_6 materials and their substituted variations are fast Li-ion conductors, as Li_3MCl_6 materials generally satisfy the above principles. According to the above principles, the materials design strategy to further increase the Li-ion conductivity in Li_3MCl_6 is to decrease Li content, which can be accomplished through aliovalent doping by higher valence cations, such as Zr^{4+} . From our computation, substituted Li_3MCl_6 materials such as $\text{Li}_{5/2}\text{Y}_{1/2}\text{Zr}_{1/2}\text{Cl}_6$ ($P\bar{3}m1$), $\text{Li}_{11/4}\text{Sc}_{3/4}\text{Zr}_{1/4}\text{Cl}_6$ ($C2/m$) and $\text{Li}_{11/4}\text{In}_{3/4}\text{Zr}_{1/4}\text{Cl}_6$ ($C2/m$), where M^{3+} are substituted by Zr^{4+} , exhibit high Li-ion conductivities on the level of $10^{-2} \text{ S cm}^{-1}$ at room temperature. These predictions are consistent with recent experimental studies that Li_3MCl_6 ($\text{M} = \text{Er}, \text{Y}$) substituted by Zr has increased Li-ion conductivity to over 1 mS cm^{-1} at RT.^[57] Another recent study on $\text{Li}_x\text{ScCl}_{3+x}$ showed increased cation concentration leads to lower Li-ion conductivity, which is also consistent with our design principle regarding cation concentration.^[56] As observed in experiments,^[56,57] cation redistribution or

reordering may happen in these Li_3MCl_6 systems after substitution, affecting the ion conduction. Given our computation method for materials substitution did not consider these cation redistributions during cation substitution, our computation-predicted Li-ion conductivities may differ from experimental values for some substituted materials, and the effect of Li concentration and cation distribution in substituted Li_3MCl_6 materials merits further study. In summary, our computation study identifies the effect of cation on Li-ion conduction and providing design strategies of tailoring cation distribution and concentrations for fast ion conduction in chlorides, and we expect our conclusions to be applicable to other halide systems such as bromides and iodides. Our computation indicates further opportunities in designing novel halide SICs with even higher conductivities.

In addition, our computation results also confirm generally low Li-ion conduction in Li_xMCl_4 materials, consistent with earlier experimental studies.^[60,61] In contrast to the earlier speculation of poor Li-ion diffusion in close-packed structures,^[71,72] first-principles calculations demonstrated that a single Li-ion vacancy in fact has a low barrier in this structure,^[17] and the poor Li-ion conductivity in the unsubstituted Li_2MCl_4 is caused by a lack of empty sites, i.e. a low concentration of vacancy sites, which would result in a high activation energy. Our calculations show that certain empty sites are high-energy sites and are unavailable for Li-ion migration in the unsubstituted Li_xMCl_4 materials (see *Blocking Effect* in Appendix A). While cation-substituted Li_xMCl_4 materials can create more empty sites for Li-ion hopping, they are still poor Li-ion conductors because of greatly increased site energy differences and migration energy barriers. Thus, the poor Li-ion conduction in Li_2MCl_4 is a result of the cation sublattice, rather than the anion sublattice as previously speculated.^[71,72] According to our design principles and computation results of hypothetical Li-ion conductors, Li-ion conductivities of Li_xMCl_4 materials can be

improved by reducing cation concentrations. Li_xMCl_4 materials with heavily reduced cation concentrations, such as $\text{Li}_2\text{Zr}_{1/2}\text{Cl}_4$ ($C2/c$), $\text{Li}_2\text{Zr}_{1/2}\text{Cl}_4$ ($Cmmm$) and $\text{Li}_{17/8}\text{Sc}_{5/8}\text{Cl}_4$ ($Imma$), have substantially increased Li-ion conductivities (Figure 2.2). Li_4ZrCl_8 ($Cmmm$), a fully substituted material originating from Li_5CrCl_8 ($Cmmm$), equivalent to $\text{Li}_x\text{M}_{1/2}\text{Cl}_4$ with half the cation concentration of Li_xMCl_4 , has considerable Li-ion conduction. A new experimental study on a halospinel $\text{Li}_2\text{Sc}_{2/3}\text{Cl}_4$ ($Fd\bar{3}m$) with reduced cation concentration from Li_2MgCl_4 ($Fd\bar{3}m$) reports greatly increased Li-ion conductivity to 1.5 mS cm^{-1} at RT,^[58] confirming our computation predictions and the design principle of reducing cation concentration for increasing Li-ion conduction in Li_xMCl_4 systems. In addition, Li_xMCl_4 systems with different cation site coordination can also lead to higher Li-ion diffusion. From our high-throughput computation, there are a few notable exceptions in substituted Li_xMCl_4 with decent Li-ion conductivity, such as LiAlCl_4 ($Pnma$), $\text{Li}_{5/4}\text{Al}_{3/4}\text{Zn}_{1/4}\text{Cl}_4$ (Zn-substituted LiAlCl_4 ($P2_1/c$)), and Li_2ZnCl_4 (Zn-substituted LiAlCl_4 ($Pmn2_1$)), which have cations located on tetrahedral sites instead of on octahedral sites as in other Li_xMCl_4 . These materials with cations on tetrahedral sites free octahedral sites for Li-ion migration and open up conduction pathways. A recent experimental study on LiAlCl_4 ($P2_1/c$) suggests a Li-ion conductivity of 0.02 mS cm^{-1} at RT, higher than previously reported Li_2MgCl_4 and Li_2CdCl_4 , and supports our computation results about potential good Li-ion diffusion in these Li_xMCl_4 systems.^[73] Therefore, the strategies for improving ionic conductivity in common Li_xMCl_4 systems are reducing cation concentrations or tailoring cation coordination, which deserve further study.

In conclusion, we systematically studied Li diffusion in a wide range of Li-containing chlorides including a total of 74 compounds derived from 20 known chlorides using first-principles calculations. The high-throughput computation identified and predicted 19 materials as potential

Li SICs with Li-ion conductivity of $> 10^{-3}$ S/cm at RT, wide electrochemical window, and large band gaps, as potential SE materials. Our computation on this wide range of chloride materials reveals that Li content and cation sublattice significantly affect Li-ion conduction in Li-containing chloride materials and that low cation concentration and sparse cation distribution are key factors for low Li-ion migration barriers and high Li-ion conductivity. With this understanding, we propose and demonstrate design principles for increasing Li-ion conductivities of poor Li-ion conductors by artificially altering their cation concentrations and configurations. Our results provide fundamental understanding, new materials systems, and guiding principles to design novel halides as fast Li-ion conductors.

Chapter 3: Unravel discrepancies and develop evaluation metrics for machine learning interatomic potentials³

3.1 Introduction

Our results in Chapter 2 demonstrated that first principles density functional theory (DFT) calculation is an effective approach to predict new Li superionic conductors (SICs) and understand fundamental ionic diffusion mechanisms in Li-chloride materials systems. However, the DFT calculations and ab initio molecular dynamics (AIMD) simulations in the previous analysis were computationally expensive, with a total cost on the order of 10^6 CPU core hours (dual Intel Ivy Bridge E5-2690v2, 2019) for around 150 AIMD simulations of 74 Li-containing chlorides. AIMD simulations in Li-chloride materials systems are more time-consuming because the crystal framework of these materials is subjected to melting at such high temperatures (of > 600 K). However, at low simulation temperatures below 600 K, the AIMD simulations should be performed over a long time up to 1 ns to obtain reliable sampling for Li-ion diffusivities.^[1,2]

To overcome the challenge of the high computation cost of AIMD simulations, applying machine learning interatomic potentials (MLPs) for molecular dynamics (MD) simulations is a promising direction for materials modeling. A variety of MLPs implemented by different software packages and workflows include Gaussian Approximation Potential (GAP) based on Smooth Overlap of Atomic Positions (SOAP) descriptors,^[11,74] Neural Network Potential (NNP),^[12,75] Spectral Neighbor Analysis Potential (SNAP),^[14] Moment Tensor Potential (MTP),^[15] and Deep Potential (DeePMD) models,^[5] and many MLP variances derived from different modifications and combinations of ML models and descriptors. Trained by DFT-calculated energies and atomic

³ Some text of this chapter is adapted from **Y. Liu**, X. He, Y. Mo, **2023**. arXiv:2306.11639

forces, state-of-the-art MLPs are claimed to achieve near-DFT accuracies with low errors on predicted energies and forces. These low averaged errors reported have created the impression that MLPs are as accurate as DFT calculations. However, these MLPs with small average errors may not always accurately reproduce the physical phenomena in atomistic simulations, as shown in the following examples.^[10,38,39] An MLP of Al by Botu *et al.* was reported a low MAE force error of 0.03 eV/Å, but its MD simulations predicted the activation energy of Al vacancy diffusion with an error of 0.1 eV compared to the DFT value of 0.59 eV, even though vacancy structures and vacancy diffusion were included in the training dataset.^[76] In Vandermause *et al.* ^[37], an Al MLP with a low RMSE force error of 0.05 eV/Å for solid phase and 0.12 eV/Å for liquid phase also exhibited discrepancies with DFT in in surface adatom migration, which were considered during the on-the-fly training. Zuo *et al.*^[1] reported a number of MLPs (such as GAP, NNP, SNAP, MTP) with small RMSEs of atomic forces at the level of 0.15 - 0.4 eV/Å and 10-20% errors in the vacancy formation energy and migration barrier for a number of materials (such as Li, Mo, Si, Ge), while vacancy structures were also included in the training. Additionally, in the aforementioned studies, there were large errors of the predicted migration energy barriers of defects, even though these defects and their migrations were considered in the training and testing datasets. Atomistic diffusion is determined by the dynamics of atoms (often in the form of point defects) and the PES beyond their equilibrium sites, but the direct testing of MLPs on these atomistic-level details in MD simulations still shows errors and failures.^[44] As reported by Fu *et al.*^[44], the MD simulations based on MLPs observe errors, such as radial density functions, and even the failure of the MD simulations after a certain duration. These results suggest there are errors in the MLPs in causing these errors and failures from actual MD simulations. It is crucial to examine the accuracy of MLPs in simulating

the atomic dynamics and reproducing physical properties, understand potential discrepancies, and develop appropriate testing metrics.

Typical approaches to improve the MLPs include adjusting the fraction or weights of certain structures in the training dataset, modifying the cost/loss functions, and tuning hyperparameters.^[77] The average errors in energies and forces or a few easily computable properties, such as elastic constants, energy vs. volume curves for different crystal structures, and formation energies of point defects,^[1,78] are often used to optimize and select the MLP models. However, to test and quantify those errors that can only be directly observed in actual MD simulations, such as the errors in diffusional properties, one would need to conduct numerous tests in MD simulations for an extended duration before selecting the final MLP models.^[44] This approach requires a large computational cost of running MD simulations to select MLPs, which may be impractical for optimizing MLPs with many combinations of training variables and hyperparameters. Therefore, appropriate testing metrics should be developed to thoroughly gauge the ability of an MLP in reproducing atomic dynamics and physical properties in a range of typically encountered physical situations, and such quantitative metrics are crucial for the further improvement of MLPs.

In this chapter, by systematically comparing atomic dynamics from MLP-MD simulations and AIMD simulations, we reveal that state-of-the-art MLPs, even with carefully selected training datasets and small average errors evaluated by conventional testing, may not fully reproduce atomic dynamics or related properties (Chapter 3.2). The tested MLPs show discrepancies in diffusions or rare events (Chapter 3.3.1), defect configurations (Chapter 3.3.2), and atomic vibrations (Chapter 3.3.3). We then develop the error evaluation metrics based on the atomic forces of RE atoms (Chapter 3.4.1) and demonstrate them for indicating the performance of MLPs on

atomistic dynamics in MD simulations (Chapter 3.4.2). The MLPs trained with enhanced RE data and selected by newly developed metrics show improved predictions of atom dynamics and diffusional properties. In the end, we summarize our process of developing the evaluation metrics for the observed simulation-based discrepancies. The identified discrepancies, our novel evaluation metrics, and their development process are general to all MLPs, and can serve as valuable guidance for future development and improvements of accurate, robust, and reliable MLPs for atomistic modeling.

3.2 Discrepancies of MLPs on atomic dynamics between MLP and DFT occur even when low average errors are reported

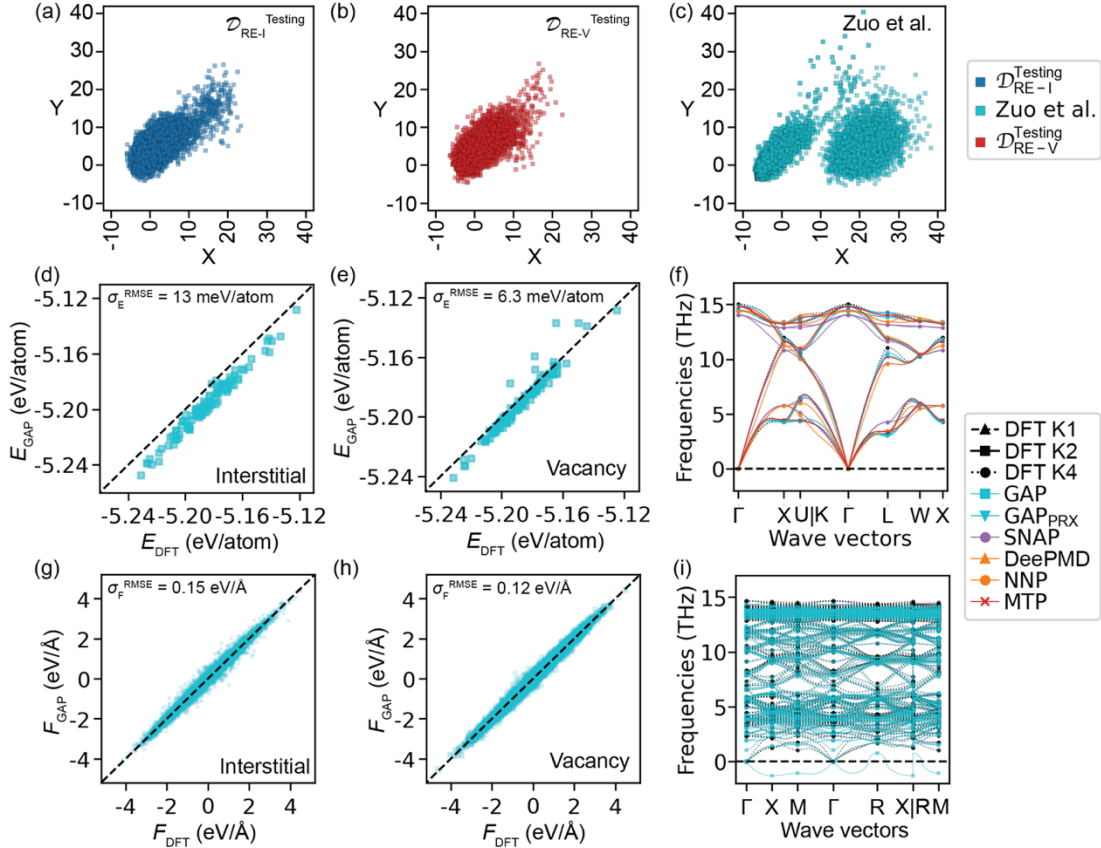


Figure 3.1. Testing of MLPs. The comparison of the atomic environments from (a) the interstitial-RE testing set $\mathcal{D}_{\text{RE-I}}^{\text{Testing}}$ (blue), (b) the vacancy-RE testing set $\mathcal{D}_{\text{RE-V}}^{\text{Testing}}$ (red), and (c) the original training dataset from Ref.^[1] (cyan), using 6500, 6300, and 13553 atomic environments respectively. All atomic environments here are quantified by the 1st and the 2nd principal components from principal component analysis (PCA) of the SOAP descriptors. Comparison of the (g, h) atomic forces and (d, e) energies predicted from MLP (GAP) versus the benchmark (DFT K4) on (e, h) the vacancy-RE testing set $\mathcal{D}_{\text{RE-V}}^{\text{Testing}}$ and (d, g) the interstitial-RE testing set $\mathcal{D}_{\text{RE-I}}^{\text{Testing}}$. Phonon dispersion of (f) the bulk Si and (i) the Si supercell with a single vacancy calculated by DFT K4 and GAP.

We conduct our study on a number of MLPs to show the observed phenomena are general among MLPs. In order to perform a consistent comparison, we directly retrieve the MLP (GAP, GAP_{PRX}, NNP, SNAP, and MTP) models of Si from previous studies^[1,78], besides the DeePMD model^[79] trained using the same training dataset from Ref. ^[1]. This training dataset of Ref. ^[1]

includes a diverse range of carefully selected atomistic structures of solid Si, melted liquid Si, strained Si, Si surfaces, and Si-vacancy from AIMD simulations at a wide range of temperatures (**Figure 3.1c**, Appendix D and *Analysis of errors of MLPs* in Appendix C). In order to quantify the errors on predicted energies and atomic forces of these MLPs, we construct interstitial-RE and vacancy-RE testing sets, $\mathcal{D}_{\text{RE-I}}^{\text{Testing}}$ and $\mathcal{D}_{\text{RE-V}}^{\text{Testing}}$, respectively, each consisting of 100 snapshots of atomic configurations with a single migrating vacancy or interstitial, respectively, from AIMD simulations at 1230 K (Figure 3.1a and b, Append D), with the true values of energies and atomic forces evaluated by DFT calculations with a k -point mesh of $4 \times 4 \times 4$ (DFT K4). The MLPs accurately predict the energies and atomic forces in the training dataset and the testing dataset in consistency with the original studies,^[1,78] showing low RMSEs below 10 meV/atom and 0.3 eV/Å (with the median force magnitude of 1.67 eV/Å) for most MLPs on the vacancy-RE testing set $\mathcal{D}_{\text{RE-V}}^{\text{Testing}}$ (Figure 3.1e, h, C1, and Table C1). Given similar structures with vacancies are covered in the training dataset, the good performance of MLPs on vacancy structures should be expected.

We also test the MLPs on the structures not included in the training data (Figure 3.1c), specifically the interstitial-RE testing dataset $\mathcal{D}_{\text{RE-I}}^{\text{Testing}}$, which comprises the snapshots from the AIMD simulations of Si supercell with an interstitial. Most MLPs (except for GAP_{PRX}) do not include the configurations with an interstitial in the training dataset. While the MLPs prediction of atomistic energies show a bias with an average offset of 10 – 13 meV/atom lower than DFT values, the overall RMSEs are below 15 meV/atom and 0.3 eV/Å (with the median force magnitude of 1.69 eV/Å) for these interstitial structures for most MLPs (Figure 3.1d, g, C2, and Table C1). These small error values in energy and force on training and testing datasets are often interpreted as good performance of MLPs.

In addition to the errors of energies and forces, we evaluate and compare the phonon dispersion relations predicted by MLPs with DFT (Figure 3.1f, i, C3, and C4). While most MLPs have good agreement on phonon dispersions of bulk Si (Figure 3.1f)^[78,80], a number of MLPs exhibit noticeable differences on phonon dispersions in the Si supercell with a vacancy (Figure 3.1i and C4, Appendix D). For example, the phonon dispersion by GAP has imaginary frequencies (Figure 3.1i), and the phonon dispersion by SNAP has additional bands with lower frequencies compared to DFT K4 in high frequency sections (Figure C4). The training data includes the atom vibration from AIMD simulations with vacancies, which should help the training of phonon dispersion relations. Therefore, these discrepancies require further investigation.

3.3 Discrepancies in atomic dynamics

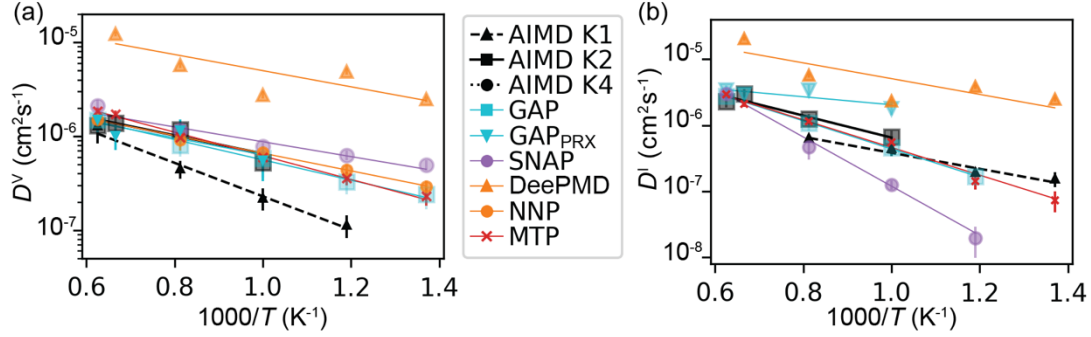


Figure 3.2. Diffusions of point defects in Si. Arrhenius plots of the diffusivity of (a) vacancy and (b) interstitial in Si from AIMD and MLP-MD simulations. The missing data points indicate the failure of the MD simulations due to either the melting of the crystal structure or an insufficient number of atomic hops to quantify diffusivities (See Table C4, and C5). The error bars of diffusivities are estimated based on the scheme in Ref. ^[4] (Appendix D).

3.3.1 Rare events are sources of discrepancies

Here the MD simulations using these MLPs are performed to simulate atom dynamics and related physical properties in a Si supercell with a single vacancy or a single interstitial, to identify potential discrepancies between MLP-MD and AIMD simulations. The Si diffusivities are evaluated at a range of temperatures 730 – 1600 K from the mean-squared-displacements of Si over time (Appendix D) (**Figure 3.2a** and **b**). Given the MLPs are trained on the DFT calculations based on a fine k -point mesh of $4 \times 4 \times 4$ (K4), the MLP results should be compared to DFT K4 benchmark as in many of our tests. For the MD simulations, the MLP results are compared to the AIMD simulations based on coarser accuracy setting of a k -point mesh of $2 \times 2 \times 2$ (K2) and a single Γ -point (K1), because AIMD simulations based on K4 is takes prohibitively long to obtain enough number of atom hops. In addition, these lower accuracy settings are commonly utilized for AIMD simulations in previous studies.^[4,81,82] The errors with lower-accuracy DFT calculations also can be used as error ranges for comparison.

For vacancy diffusion, which is covered in the training data, most MLPs predict diffusivities within a reasonable error range of the diffusivities predicted by K2 AIMD simulations. Most MLPs perform better than DFT-K1. DeePMD gives higher diffusivities than AIMD simulations, but an agreement on the fitted activation energy (0.17 eV compared with 0.19 eV given by AIMD K2). Some MLPs show discrepancies and deviations among each other in the diffusivities at lower temperatures, leading to discrepancies in the fitted activation energies and extrapolated RT diffusivities. Nonetheless, the comparison of the MLPs and AIMD simulations should take into consideration the stochastic nature of estimating diffusivities from MD simulations and the limited number of data points, which would lead to large uncertainty of the fitted activation energy and extrapolated diffusivity.^[4] As shown in the previous study^[4], even for a total mean square displacement of ion diffusion over a few thousand Å², the standard deviation of obtained diffusivity can be as large as 20-30%. The error bars and range of all obtained diffusivity values are in Table C4, C5 and C6, Appendix C.

While the Si interstitials are not covered in the training of MLPs (except for GAP_{PRX}), all MLPs show low RMSEs in the testing dataset (Figure 3.1d, and g) with interstitial configurations. Here the MD simulations of interstitial diffusion is performed to test whether the atomic dynamics can be correctly reproduced. Some MLPs, such as GAP_{PRX}, DeePMD, and SNAP, show significant deviations in the interstitial diffusional properties predicted by MLP-MD simulations (Figure 3.2a, and Table C2). The diffusivities predicted by GAP and MTP agree reasonably with AIMD simulations over the temperature range 1000 – 1600 K, and the fitted activation energies of interstitial diffusion, E_a^I , show minor differences. GAP_{PRX}, which is the only MLP considered interstitial in the training, also shows discrepancies in interstitial diffusivities with a low E_a^I of 0.12 eV. SNAP shows a E_a^I of 0.74 eV, much higher than AIMD K2 (0.30 eV). For NNP, the

crystalline Si structure melted during the MD simulations at the temperature of 730 – 1600 K (see *Diffusional properties* in Appendix C). According to the tests, many MLPs show some discrepancies in predicting diffusivity, activation energy, or both, and thus the ability to fully reproduce the atom dynamics of Si interstitial diffusion is limited. The discrepancies in diffusional properties indicate that having low average errors in energies and forces are insufficient error evaluation metrics to judge whether the atomic dynamics of diffusions in MD simulations are accurate.

3.3.2 Configurations with similar energies are discrepancy sources

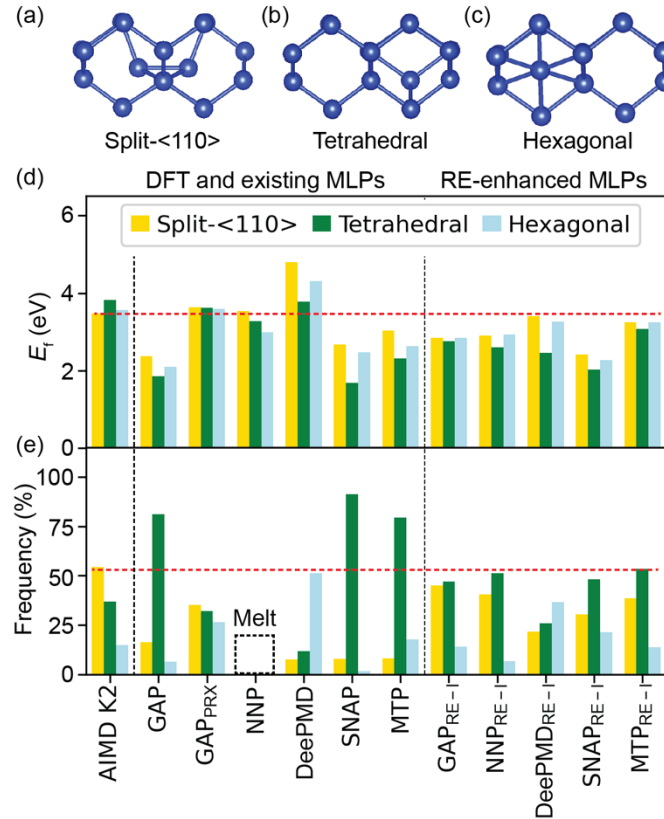


Figure 3.3. Si interstitials by MLPs. Atomistic configurations of (a) split-<110>, (b) tetrahedral, and (c) hexagonal Si interstitials. Comparison of DFT (red dashed lines) and MLPs on (d) the

formation energies E_f of the Si interstitials and (e) the occurrence frequencies in AIMD (K2) and MLP-MD simulations at 1230K (Appendix D).

In order to reveal the discrepancies of MLPs in predicting interstitial diffusions, we analyze the snapshots from MLP-MD simulations of interstitials in comparison with AIMD simulations. DFT studies^[83–85] reported three types of Si interstitials, such as the ground-state split- $\langle 110 \rangle$ (**Figure 3.3a**), tetrahedral (Figure 3.3b), and hexagonal (Figure 3.3c) interstitials, with the formation energies E_f of 3.56, 3.72, and 3.59 eV, respectively, from DFT (K4) calculations (Figure 3.3). Consistent with the trend of the formation energies, the split- $\langle 110 \rangle$ interstitial has higher occurrence frequency in the AIMD simulations than tetrahedral or hexagonal (Appendix D) (Figure 3.3d and e).^[83–85] By contrast, MLP-MD simulations show higher occurrence frequencies of either tetrahedral interstitial (by GAP, SNAP, and MTP) or hexagonal interstitial (by DeePMD) (Figure 3.3e and C7). Consistent with the occurrence frequencies in MLP-MD simulations, most MLPs (except for GAP_{PRX}) give lower E_f for the tetrahedral interstitial than the ground-state split- $\langle 110 \rangle$ interstitial (Figure 3.3d). The trend of increasing occurrence frequencies for different interstitials can be largely explained by the decreasing formation energies, while the entropy of these defects may also play an effect (see *Analysis of Si interstitials* and *Convergence of interstitial occurrence frequencies* in Appendix C). Therefore, the errors in interstitial diffusion during MLP-MD simulations are caused by the discrepancies of MLPs in the predictions of different interstitial configurations. These discrepancies should be expected given interstitial configurations are not considered in the training (except for GAP_{PRX}). However, it is worth noting that small errors are reported for the testing of energies and forces for interstitial configurations (in the interstitial-RE test set $\mathcal{D}_{\text{RE-I}}^{\text{Testing}}$). Thus, having small average errors in energies and forces in conventional error

testing on the defect configurations may be insufficient in determining whether the MLPs would correctly reproduce these defect configurations in MD simulations.

3.3.3 Vibration near defects is a discrepancy source

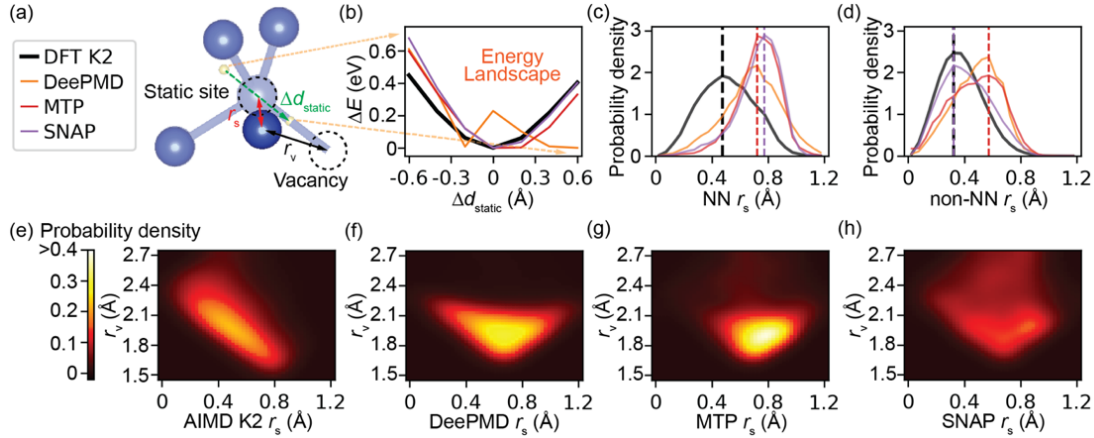


Figure 3.4. Errors in atom vibrations. (a) Illustration of a vibrating Si atom by the distance r_s to its static site and the distance r_v to the nearest static vacancy site. (b) The potential energy surface (PES) of the vibrating Si atom moving along the direction (green arrow) towards the vacancy. The probability density (c) the nearest neighbor (NN) atoms and (d) non-NN atoms of the vacancy plotted as a function of r_s . (e-h) The probability density for the NN atoms of the vacancy plotted as a function of r_v and r_s . All distributions are obtained from MD simulations at 1230K, using an averaging scheme described in the Appendix D.

The good performance of MLPs on Si vacancies is expected given the dynamical snapshots of Si vacancies in a wide range of conditions are well covered in the training dataset. However, the phonon dispersion of the vacancy structure predicted by the MLPs (Figure 3.1i) suggests potential discrepancies of potential energy surface (PES) and atomic vibrations. Additionally, for the vacancy diffusion (Figure 3.2a, and Table C3), the discrepancies of diffusional properties are observed in a few cases (see *Diffusional properties* in Appendix C). The diffusional properties, such as activation energies and the pre-exponential factors of the Arrhenius relation of diffusivity, are critically dependent on the PES of a migrating atom next to a defect such as a vacancy.

We further analyze the vibrations of Si atoms neighboring the vacancy, which is directly related to vacancy diffusion, in order to gain insights into the observed discrepancies of the PES. We visualize the vibrations of Si atoms near the vacancy by plotting the distribution of the distance r_s and r_v of the atoms from the nearest static Si site to its nearest static vacancy site, respectively (**Figure 3.4a**) (Appendix D). For the atoms that are not the nearest neighbor (NN) atoms of the vacancy (Figure 3.4d), the atom vibrations are similar to those in crystalline bulk. The agreement in the vibration of non-NN atoms is consistent with the good agreement in the phonon dispersion relations (Figure 3.1f) and elastic constants of bulk crystalline Si.^[1,78,80]

The vibrations of the vacancy nearest neighbor (vacancy-NN) atoms in MLP-MD simulations show major discrepancies with AIMD simulations (Figure 3.4e – h, and C8). In the MLP-MD simulations, the vacancy-NN atoms vibrate much further away from their static sites, as indicated by the distributions of r_s (Figure 3.4c and C8). The discrepancies in vacancy-NN vibrations reveal that MLPs do not accurately reproduce atom dynamics around the defect.

The observed discrepancies in the vibrations of vacancy-NN atoms can be explained by the errors of the PES of a vacancy-NN Si atom moving along the direction toward the vacancy (Figure 3.4b and C10). While some pathways of the migration for a single atom in a static configuration with fully relaxed static sites (Figure C10) may be poorly predicted by the MLPs, some different pathways can be accurately predicted in very similar configurations with slightly adjusted atom positions (near-equilibrium snapshots from AIMD simulations) (see *Analysis of potential energy landscapes* in Appendix C, Figure C11). These observed discrepancies indicate that the PES predicted by the MLPs may not be always accurate and reliable under similar atomic configurations with minor variances in positions.

These errors in the atom vibrations and PESs nearby a defect are surprising, given that a range of dynamical configurations of vacancies from AIMD simulations is considered in the training data and that small errors in energies and forces are reported for these configurations. These results show the challenges of MLPs in accurately reproducing the PES of atoms related to defects, even if these defects and related dynamics are included in the training. In addition, the conventional error testing of MLPs based on average errors of energies and forces are insufficient in indicating these errors in atom vibration and PES near defects. Our results reveal previously neglected errors on the PES of MLPs in causing inaccurate atom dynamics.

In summary, we have identified discrepancies based on observations of atomic dynamics in MD simulations, such as diffusions, configurations of defects, and atomic vibrations. In order to train MLPs that can more accurately reproduce these dynamical phenomena, we need to quantify these discrepancies by developing corresponding error evaluation metrics, which can be further used to train and select the MLPs with the highest metric scores.

3.4 Development of error evaluation metrics

Using the discrepancies on atom diffusions as an example, we here develop corresponding error evaluation metrics, and improve the performances of MLPs on diffusional properties. The process is as follows. We first develop a number of metrics for quantifying the aforementioned sources of discrepancies (3.4.1). The evaluation metrics are then statistically verified to effectively indicate diffusional properties derived from atom dynamics in MD simulations (3.4.2). This process can be generalized to improve the training and testing of MLPs, as summarized in 3.4.3.

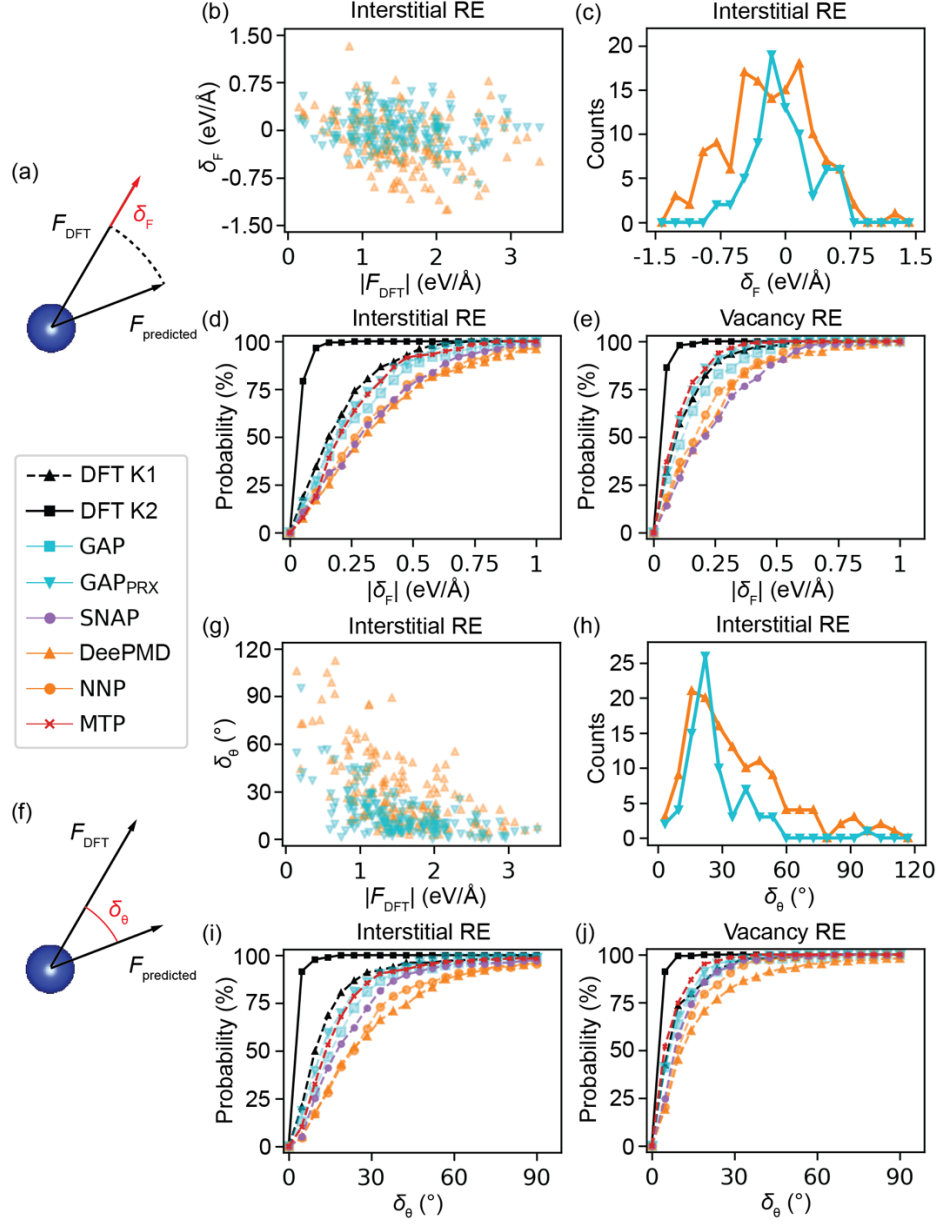


Figure 3.5. Errors of atomic forces. (a) Illustration of force magnitude error δ_F and force direction δ_θ of atomic force $F_{\text{predicted}}$ predicted by an MLP in comparison to F_{DFT} by DFT. For the RE atoms in the $\mathcal{D}_{\text{RE-I}}^{\text{Testing}}$, the errors (b) δ_F and (g) δ_θ of DeePMD (orange) and GAP_{PRX} (cyan), the corresponding distribution of (c) δ_F and (h) δ_θ , and the cumulative distribution function (CDF) of (d) δ_F and (i) δ_θ . The CDF of (e) δ_F and (j) δ_θ on vacancy RE atoms in the $\mathcal{D}_{\text{RE-V}}^{\text{Testing}}$ set. Interstitial- and vacancy-RE atoms are atoms in the middle of migration identified in the $\mathcal{D}_{\text{RE-I}}^{\text{Testing}}$ and $\mathcal{D}_{\text{RE-V}}^{\text{Testing}}$ datasets, respectively.

3.4.1 Quantifying the force errors on RE migrating atoms

Given the aforementioned discrepancies in diffusional properties, we identify the sources of errors on migrating atoms, which are the atoms of interstitial or vacancy in the middle of the hopping from the current equilibrium site to the neighboring site (Appendix D). These atom hops are known as rare events (REs) in MD simulations. To quantify the errors of MLP predictions for RE atoms, we compare the predicted atomic forces with DFT results on these RE atoms. We evaluate the error of atomic force in the magnitude δ_F and the direction δ_θ of as

$$\delta_F = |\vec{F}_{\text{predicted}}| - |\vec{F}_{\text{DFT}}| \quad (3.1)$$

$$\delta_\theta = \arccos\left(\frac{\vec{F}_{\text{predicted}} \cdot \vec{F}_{\text{DFT}}}{|\vec{F}_{\text{predicted}}| |\vec{F}_{\text{DFT}}|}\right) \quad (3.2)$$

where $\vec{F}_{\text{DFT}}$ is the benchmark true force values calculated by DFT K4, and $\vec{F}_{\text{predicted}}$ is the force predicted by the MLPs (or DFT K1, or K2). For the RE atoms in the $\mathcal{D}_{\text{RE-I}}^{\text{Testing}}$ and $\mathcal{D}_{\text{RE-V}}^{\text{Testing}}$ datasets, significant errors were found in the force magnitude δ_F (**Figure 3.5a**) and force direction δ_θ (Figure 3.5f) of the atomic forces predicted by MLPs. For the RE atoms and the nearby atoms within a distance $r < 3 \text{ \AA}$, 50% - 80% and >40% exhibit large errors of $\delta_F > 0.5 \text{ eV/\AA}$ or $\delta_\theta > 15^\circ$, respectively, whereas around 20% of the other atoms that are $> 3 \text{ \AA}$ away from the migrating atoms show similar levels of errors (see *Analysis of forces with large errors* in Appendix C, Table C7). While other factors in addition to the force errors on RE atoms may contribute to the discrepancies, the analyses nonetheless confirm that there are large errors on the atoms near defects and RE atoms are a major source responsible for discrepancies in diffusions.

The cumulative distribution functions (CDFs) of δ_F and δ_θ on RE atoms show the distributions of force errors of MLPs (Figure 3.5d, e, i, j, and C12). Most MLPs (except for GAP_{PRX}) exhibit significant errors in force prediction on RE atoms (8-25% for $\delta_F > 0.5 \text{ eV/\AA}$ and

40-70% for $\delta_\theta > 15^\circ$), whereas lower accuracy DFT K2 ($< 0.5\%$ for $\delta_F > 0.5 \text{ eV/\AA}$ and 1% for $\delta_\theta > 15^\circ$) and DFT K1 calculations (5% for $\delta_F > 0.5 \text{ eV/\AA}$ and 30% for $\delta_\theta > 15^\circ$) have much smaller errors. These results clearly show MLP predictions have large errors on RE atoms. The force magnitude error δ_F of larger than $0.5 \text{ eV/\AA}$ is significant, as the median force magnitude of all atoms is around $1.7 \text{ eV/\AA}$ (Figure 3.5b). Among these atoms with $\delta_F > 0.5 \text{ eV/\AA}$, 10 – 35% interstitials and 3 – 15% vacancies also exhibit significant force direction errors of $\delta_\theta > 30^\circ$ (Figure 3.5g), which would lead to major errors in predicting atom dynamics in MD simulations. Therefore, these large errors in MLP-predicted forces on RE atoms (Figure 3.5d, e, i, and j) and their nearby atoms (Table C7 and Figure C12) are major sources of the observed discrepancies in the atom vibration and diffusion between MLPs and DFT.

By identifying what atoms tend to give large errors, it can be understood why conventional error testing of MLPs shows very small errors. In conventional error testing of MLPs, most of the atomistic configurations evaluated are for atoms near equilibrium positions, which are accurately predicted by MLPs with very small errors. The RE atoms only consist of a very small fraction in most typical testing datasets, e.g., $< 1\%$ in the dataset of Zuo et al.^[1] and Bartok et al.^[78] ($2 - 3\%$ in our RE-testing datasets). Therefore, the large errors on RE atoms are averaged out by the majority of near-equilibrium atom configurations. The total RMSEs of the forces on all atoms in conventional error testing mostly reflect the accurate MLP prediction of near-equilibrium atoms, which often dominate the typical testing dataset (see *Analysis of forces with large errors* in Appendix C). However, it should be noted that a few percent of RE-related atoms are critical for the correct prediction of atom dynamics and physical phenomena in MD simulations.

In order to overcome the limitations of conventional error testing, we propose to quantify the CDFs of force errors δ_F or δ_θ on RE atoms, using the normalized area of curve (NAC) under

the CDF curves (see details in Appendix D) as new metrics. The NAC of CDF equals to 1 for an MLP that completely agree with the benchmark true values (DFT K4), and a lower value of NAC suggests larger errors. To combine the metrics of force magnitude $NAC(\delta_F, \mathcal{D})$ and force direction $NAC(\delta_\theta, \mathcal{D})$, we define the force performance score $P(\mathcal{D})$, as the product of $NAC(\delta_F, \mathcal{D})$ and $NAC(\delta_\theta, \mathcal{D})$ for errors of atomic forces in a given dataset $\mathcal{D}$ (Appendix D),

$$P(\mathcal{D}) = NAC(\delta_F, \mathcal{D}) \times NAC(\delta_\theta, \mathcal{D}). \quad (3.3)$$

These quantitative metrics can be effectively used in training, validation, and testing to improve MLPs as demonstrated in the next chapter.

3.4.2 Force performance scores as effective metrics

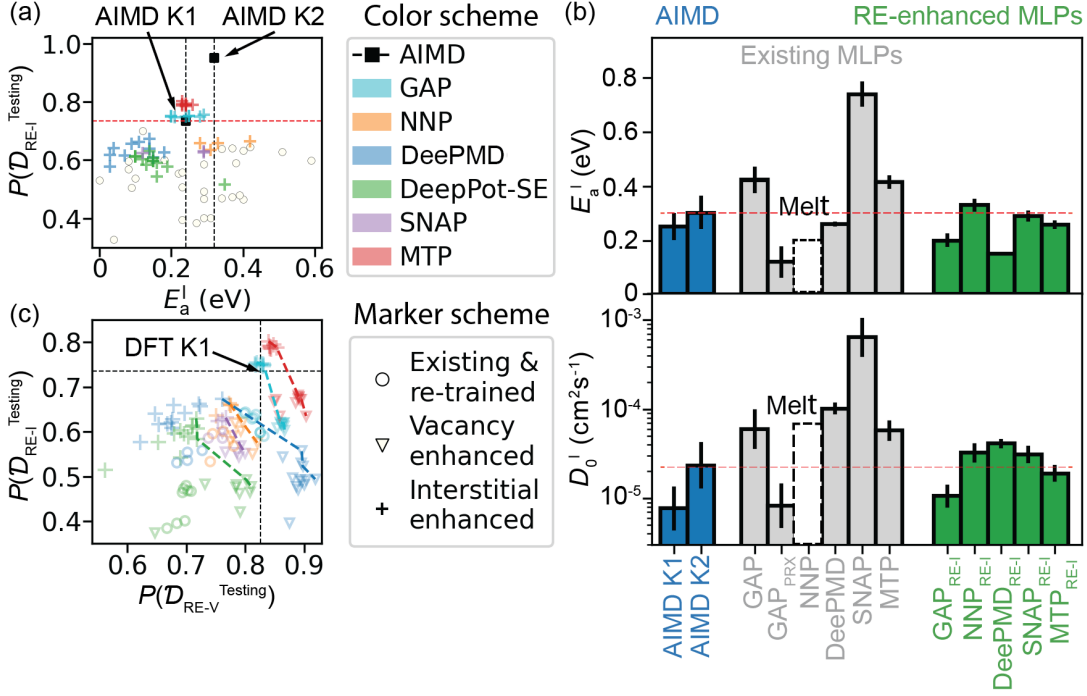


Figure 3.6. The performance of RE-enhanced MLPs. (a) Comparison of MLPs for their force performance score $P(\mathcal{D}_{\text{RE-I}}^{\text{Testing}})$ on interstitial RE atoms versus the fitted activation energy E_a^{I} of interstitial diffusion in MLP-MD simulations. Black dashed lines are E_a^{I} from AIMD simulations, and the red dashed line indicates $P(\mathcal{D}_{\text{RE-I}}^{\text{Testing}})$ calculated by DFT K1. (b) The calculated E_a^{I} and D_0^{I} of interstitial diffusion from AIMD simulations (blue bar and red dash line), original MLPs (grey), and interstitial-enhanced MLPs (green). (c) Comparison of MLPs for their force accuracies $P(\mathcal{D}_{\text{RE-I}}^{\text{Testing}})$ on interstitial RE dataset versus $P(\mathcal{D}_{\text{RE-V}}^{\text{Testing}})$ on vacancy RE dataset, showing the Pareto fronts (dash lines). These MLPs are comprised of newly trained MLPs by RE-enhanced training data with interstitials RE-I (cross) or vacancies RE-V (triangles), original MLPs (circles), or MLPs re-trained by the original dataset in Ref.^[1] and selected by the new RE-validation process (Appendix D). Thin black dashed lines are $P(\mathcal{D})$ calculated by DFT K1.

To evaluate the effectiveness of $P(\mathcal{D})$, we here obtain MLPs with high $P(\mathcal{D})$ scores and compare their atomic dynamics in diffusions with previous MLPs. A RE-enhanced training dataset is generated to train the MLPs that can achieve higher $P(\mathcal{D})$ scores. For the RE-enhanced training dataset, we replace a fraction (54%) of the structures in the original dataset in Ref.^[1] by those containing identified migrating interstitials. In this way, the size of the training dataset is kept the same to eliminate the impact of training data size in the comparison with original MLPs (Appendix

D). In the RE-enhanced validation, the force errors on RE atoms are evaluated for trained MLPs using the enhanced validation set (EVS) (Appendix D), which includes those structures with RE atoms (migrating interstitials and vacancies) identified in addition to a fraction of the original dataset in Ref.^[1]. We use the evaluation metrics on errors of energies, overall forces, forces of RE atoms (migrating interstitials or vacancies) to fine-tune the hyperparameters of MLPs and select the MLPs with good performances on all evaluation metrics in the validation process (Appendix D). Following this process, we train and obtain 135 MLPs to study the statistical effectiveness of our force performance score and six representative RE-enhanced MLPs (denoted by the subscript RE-I), GAP_{RE-I} , NNP_{RE-I} , $DeePMD_{RE-I}$, $SNAP_{RE-I}$, and MTP_{RE-I} and compare their atomic dynamics in MD simulations with previous MLPs (Appendix D).

To evaluate the effectiveness of our force performance scores, here we calculate and compare the force performance scores $P(\mathcal{D})$ and the diffusional properties from MD simulations, for a total of 81 MLPs with a range of models and different hyperparameters, such as 35 original MLPs and 46 interstitial-enhanced MLPs selected by the metrics of force performance scores $P(\mathcal{D})$ in the validation step (Appendix D). Eight criteria measuring the energy errors and force errors of both force magnitudes and directions (Appendix D) on RE atoms are evaluated. The diffusional properties, such as activation energies, diffusivities, and diffusion pre-factors, from MD simulations of RE-enhanced MLPs with high force performance scores, show improved agreement with AIMD simulations (**Figure 3.6b**). Among these 81 MLPs (Figure 3.6a), the MLPs with higher force accuracies $P(\mathcal{D}_{RE-I}^{Testing})$ on interstitial RE atoms than DFT K1 predict the activation energy E_a^I of interstitial diffusion in good agreement with the AIMD K2 value (Figure 3.6a and C14) (see *Force accuracy as an indicator for MLP performance* in Appendix C). The other MLPs with lower $P(\mathcal{D}_{RE-I}^{Testing})$ lower scores than DFT K1 (below the red dashed line, Figure 3.6a) show much large

variations and errors in their predictions of diffusion. Among the RE-enhanced MLPs with high performance scores, $\text{GAP}_{\text{RE-I}}$, $\text{SNAP}_{\text{RE-I}}$, $\text{NNP}_{\text{RE-I}}$, and $\text{MTP}_{\text{RE-I}}$, give E_a^{I} of 0.2, 0.33, 0.29, and 0.26 eV, respectively, in better agreement with 0.25 and 0.30 eV from K1 and K2 AIMD simulations, compared to 0.42, 0.74, and 0.42 eV by the original MLPs, (Figure 3.6b, and Table C2). The predicted diffusion pre-factors D_0^{I} of enhanced MLPs improve from $> 5.7 \times 10^{-5} \text{ cm}^2\text{s}^{-1}$ (except for GAP_{PRX}) in original MLPs to $1.1 \times 10^{-5} - 4.1 \times 10^{-5} \text{ cm}^2\text{s}^{-1}$, in better agreement with $7.7 \times 10^{-6} - 2.3 \times 10^{-5} \text{ cm}^2\text{s}^{-1}$ from AIMD simulations (Figure 3.6b). In addition, the interstitial configurations observed in MLP-MD simulations also agree better with AIMD simulations. The occurrence frequencies of split- $\langle 110 \rangle$ interstitials increase to around 35% in RE-enhanced $\text{GAP}_{\text{RE-I}}$, $\text{NNP}_{\text{RE-I}}$, $\text{SNAP}_{\text{RE-I}}$, and $\text{MTP}_{\text{RE-I}}$ from less than 10% by the original MLPs (Figure 3.3e). In addition to these improvements, the occurrence frequencies of ground-state split- $\langle 110 \rangle$ interstitial are still lower than those of ground-state tetrahedral, even the configurations of all three types of interstitials are included in the RE-enhanced training data. Given the total energy difference of 0.16 eV between the two interstitial configurations is merely 3 meV/atom in the 65-atom supercell used, these results indicate the difficulties of MLPs in accurately reproducing the atomic configurations with similar energies. In summary, the force performance scores on RE atoms are effective error evaluation metrics to indicate related diffusional properties predicted by the MD simulations of the MLPs. Using improved error evaluation metrics for optimization and validation is demonstrated as an effective step in training MLPs with improved performance.

3.4.3 Process of developing error evaluation metrics

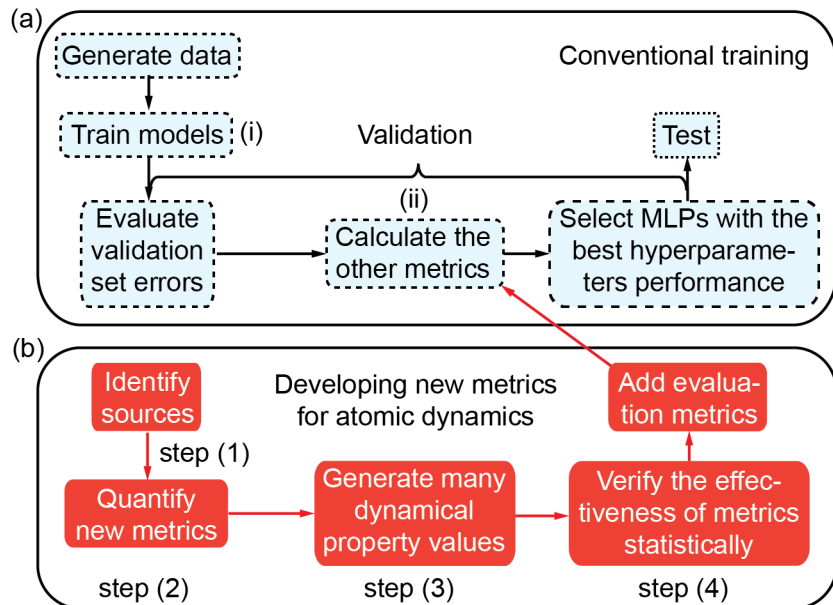


Figure 3.7. Process of MLP training and developing new metrics. (a) Conventional process of MLP training including the validation process of fine-tuning hyperparameters. (b) The process of developing evaluation metrics.

Developing error evaluation metrics that are indicative of the predictions of atomic dynamics is essential for the development of MLPs. In the conventional training process, optimizing MLPs on properties other than the errors of predicted energies and forces can be either done by 1) adding additional terms and weights into the loss functions as training targets (**Figure 3.7a-i**),^[1,77] or 2) adding additional metrics or material properties, such as elastic tensors in Ref. ^[1], when deciding the optimal hyperparameters for MLPs (**Figure 3.7a-ii**). However, this conventional approach would be computationally expensive for evaluating atomic dynamics (**Figure 3.7a-ii**), because it requires running multiple MD simulations for each trained model for the fine-tuning of hyperparameters.^[44] Thus, effective error evaluation metrics that do not require extensive MD simulations and are indicative of atomic dynamics are essential, so one can train

MLPs that can accurately predict physical phenomena in MD simulations with lower computational costs in the training and validation process.

Here we summarize the process of developing error evaluation metrics as follows, which can be generalized for future development to improve simulation-based performance of MLPs (Figure 3.2b): (1) Identify the sources of discrepancies in the MD simulations, (2) propose related but easy-to-calculate metrics that are related to the physical properties based on the simulation (e.g. forces on RE atoms to diffusional properties in our case), (3) quantify the proposed error evaluation metrics and dynamical properties based on simulations for a range of MLPs, and (4) verify the effectiveness of proposed metrics statistically among these MLPs with different models, hyperparameters, or training data. If the metrics prove to have a statistically significant effect on improving the atomic dynamics for MLPs, we then can add the proposed and verified evaluation metrics into the training process (Figure 3.7a-ii).

There are important considerations in generalizing this process of developing new error evaluation metrics for simulation-derived properties. It's important to use a large amount of data, on metrics and dynamical properties, predicted by diverse MLPs, and conduct statistical verification instead of using a single or a few MLPs because the developed metrics should be general enough to be applied to most MLPs, which may be in the future trained by other descriptors, models, or training data. In some exception cases, MLPs (e.g., NNPs (orange crosses), SNAPs (purple crosses), and the other models (circles), Figure 3.6a) may give poor scores in evaluation metrics but may have small errors on dynamical properties, thus statistically verifying the evaluation metrics is important. There are many factors affecting the outcome of the training of ML models, and many are random in nature and yet poorly understood.

3.5 Trade-offs in MLP performances: Pareto fronts of MLPs

Additionally, we compare the performance and force accuracies of MLPs based on a variety of ML algorithms and atomic descriptors and observe the trade-off the accuracies on different properties (Figure 3.6c). We compare MLPs including the 46 interstitial-enhanced MLPs which used training data with both interstitials and vacancies, and 54 vacancy-enhanced MLPs trained using vacancy REs data (see *Vacancy-enhanced MLPs* in Appendix C), 11 original MLPs, and 24 MLPs retrained from the original dataset.^[1] We compare the force accuracies $P(\mathcal{D}_{\text{RE-I}}^{\text{Testing}})$ and $P(\mathcal{D}_{\text{RE-V}}^{\text{Testing}})$ on interstitial- and vacancy-RE atoms, respectively. For all MLPs, there is a clear trend of decreasing $P(\mathcal{D}_{\text{RE-I}}^{\text{Testing}})$ with increasing $P(\mathcal{D}_{\text{RE-V}}^{\text{Testing}})$ as shown in Fig. 6c, indicating trade-offs on force accuracies of different defects (interstitials versus vacancy). The Pareto front lines for each type of MLPs (GAP, NNP, SNAP, MTP, DeePMD, and DeepPot-SE) can be shown (Figure 3.6c) for the force accuracies $P(\mathcal{D}_{\text{RE-V}}^{\text{Testing}})$ and $P(\mathcal{D}_{\text{RE-I}}^{\text{Testing}})$ on different defects, such as interstitial-RE and vacancy-RE atoms.

These Pareto fronts indicate the difficulties of MLPs in achieving accurate predictions for all properties and serve as a good way to compare the performance of different MLPs. Some MLPs have lower Pareto fronts, indicating low force accuracies in interstitials at given force accuracies in vacancies. Notably, a few MLPs, such as GAP and MTP, are able to achieve higher force accuracies than DFT K1 in both interstitial and vacancy. These MLP models and descriptors may be more effective in accurate prediction for different defect configurations at the given training dataset, though the trade-offs between different predicted properties generally exist for all MLPs.

There are more examples showing the trade-offs of MLPs on other properties. MTP gives accurate predictions for atomic forces and activation energies but shows discrepancies in the predicted atomic vibrations near vacancy (Figure 3.3f, 3.6a, and c). GAP has good predictions on

diffusivities in MD simulations but shows discrepancies on phonon spectra (Figure 3.1f, g, and h). DeePMD reproduces activation energies but shows errors for diffusivities, atomic forces, and interstitial configurations (Table C2, Table C3, Figure 3.1, 3.2, 3.4, and C12). Therefore, the performance of MLPs should not be judged by a single property or a few properties. Even if an MLP may give good performance for a number of properties, good performance for other properties should not be assumed. Furthermore, overcoming the trade-off of MLPs' performance on different properties is required to further improve MLPs for a wide variety of physical simulations. While a careful selection and balance for different types of defect data are known to be essential, it is also important to have a systematic process and quantitative metrics to train MLPs with balanced accuracies in a range of structures and properties.

3.6 Conclusion and discussion

This chapter presents a systematic testing on the current state-of-the-art MLPs, regarding the discrepancies in the prediction of atom dynamics and physical properties, in comparison with DFT calculations. Even with the carefully selected training dataset and very low average errors in energies and atomic forces from conventional ML testing, MLPs may exhibit significant discrepancies in their calculated properties and atom dynamics in MD simulations. Examples of these discrepancies include different defect types, atom vibration near defects, phonon dispersion, and diffusional properties. These discrepancies of the MLPs in actual MD simulations should not be neglected. While the discrepancies for cases that are not included in the training data may be expected, significant discrepancies for presumed cases that are fully covered in the training dataset (e.g., atomic vibration around vacancy) are also significant in some cases and should be carefully

tested. These sources of potential discrepancies are summarized as follows and should be carefully considered for improved training and rigorous testing in the future development of MLPs.

The properties related to rare events. Atomistic diffusion is a typical example of physical property derived from MD simulations. However, the diffusion is largely determined by the REs, which are often less sampled in the training and testing data of the MLPs.^[86,87] The MLP predictions on the atomic forces (both in force magnitude and force direction) of RE atoms deviate significantly from DFT values, causing errors in the predicted atomic dynamics and diffusional properties from the MD simulations. On the other hand, these errors associated with RE atoms serve as effective metrics for quantifying errors of MLPs. Similar discrepancies are expected for diffusions in other materials systems or other types of rare events, such as reactions and state transitions. In addition, while the RE atoms analyzed in this study are identified using a hand-designed algorithm based on local atom distances, unsupervised machine learning, e.g., the k-means clustering method, as we demonstrated in the SI, can also be employed to identify these RE atoms.

The defects with similar energies. As shown in the examples of Si interstitials, the formation energies and the occurrence in MD simulations of different defect configurations (e.g., split- $\langle 110 \rangle$ vs tetrahedral interstitials) may be incorrectly predicted by MLPs. Even if these various defects are considered in the training, as shown in our results generated by RE-enhanced MLPs (Figure 3.6b), the MLPs' predictions still do not completely agree with the DFT benchmark. These discrepancies are not revealed by low average errors on energies and forces in conventional ML testing for MLPs, even if the testing dataset includes these defect configurations.

Atomic vibration. The dynamics of atoms near defects are poorly reproduced by the MLPs, even though these atom dynamics are considered in the training data. These errors are caused by

the poor prediction of PES around defects. Since the PES is a high dimensional function of multi-atom configurations, the capability of MLPs covering these configuration spaces may still have limitations, which should be further studied. It's worthwhile to note that the atomic vibration and the aforementioned formation energy of defect are persistent challenges for MLPs in our study and are not fully resolved.

In general, these errors of MLPs in defect energies, atom vibrations, and REs are related to defects and their nearby atoms. These aforementioned errors given by MLPs, though yield significant errors in the prediction of atom dynamics and properties in MLP-MD simulations, are not reflected in the conventional testing of MLPs. In conventional testing, a majority of the testing data are near-equilibrium atomic configurations, which are well described by the MLPs, leading to low average error values (see *Analysis of errors of MLPs* in Appendix C). Those aforementioned error-prone atomic configurations, such as RE atoms and those around defects, account for a very small fraction of atomic configurations in the testing dataset and their errors are averaged out in the RMSE/MAE error values.

This understanding provides important guidance for the training, testing, and development of MLPs. Careful considerations are required for complex defects, their surrounding atoms, and non-equilibrium structures in the training of MLPs.^[1,78] In addition to the evaluation of average error values, the accuracies and robustness of MLPs should be carefully considered and tested on a number of key scenarios with a high likelihood of errors, including defects, atom vibrations near defects, and forces (both direction and magnitude) on RE atoms. As demonstrated in our study, the error evaluation metrics quantifying these errors can effectively improve the training, validation, and testing of MLPs. For example, force accuracies on RE atoms are demonstrated examples of such error metrics to be used in the validation (Figure 3.6a) and testing of MLPs

(Figure 3.3 and 3.6c). Furthermore, our demonstrated process of developing error evaluation metrics, which were verified over a large number of MLPs with different models and training datasets and were applied to selecting MLP models, can be generally applied and extended to alleviate or overcome these MLP errors.

It is very important to develop these error evaluation metrics that are indicative of simulation-based errors. These error evaluation metrics serve as an applicable and effective method in the verification, testing, and selection of better MLPs, because it is computationally expensive to run a large number of MD simulations for all MLPs over an extended time duration to obtain related physical properties (such as diffusion). For these error evaluation metrics to be effective and generally applicable, it is important to statistically verify these metrics on a large number of MLPs with different models and even different training datasets and to compare many metrics and their values with the targeted dynamical properties. Since evaluation metrics are effective in a statistical sense, applying them to different MLP models, descriptors, and hyperparameters may show various levels of improvement. Nonetheless, developing and applying error evaluation metrics is critical to developing MLPs with improved performance on atomic dynamics and other properties derived from MD simulations. The process we demonstrate to develop such evaluation metrics can be generalized and extended for future studies.

Applying our RE-based evaluation metrics provides significant improvements in MLPs, and the aforementioned errors are significantly reduced but not fully eliminated, including atom vibration near defects and defect configuration occurrences since they're not directly related to RE atoms. Beyond the aforementioned issues of averaged errors and limited PES sampling, our results reveal a number of other potential challenges of MLPs for improving performances on these defect-related errors. The defect-related errors, such as on different types of interstitials or the PES

around defects, can be understood as atomistic configurations with small perturbations and small energy differences ($\sim 10^1$ meV/atom). Given the defect has a low concentration in the supercell model and the total energies of the entire supercell model are used for training, it can be challenging for the MLPs to accurately reproduce the PES around all these atomistic configurations with small differences and similar energies. Small differences in the PES would lead to significant change in the probability density, atomic forces, occurrence frequency, and dynamics of atoms. While more systematic testing and more training data may be helpful, it is possible that the descriptors and ML models can be further improved to capture the PES of these varying atomistic configurations. Given the PES is a high-dimension function of the atomic descriptor, it is possible the interpolation among the given configurations from the training dataset may be inadequate for accurately covering the entire relevant portion of PES encountered during MD simulations. Overall, to accurately reproduce all aforementioned error cases with similar energies, more studies are needed to further test and improve the atomistic descriptors, ML models, or the schemes of training and testing.

In conclusion, we study a number of MLPs and identified a number of potential discrepancy sources in their applications. Leveraging these, we develop novel evaluation metrics into a process that identifies sources of discrepancies in atomic dynamics, quantifies the discrepancies, statistically verifies the effectiveness of the developed metrics, and optimizes MLPs using enhanced quantitative metrics. By proposing and demonstrating improved evaluation metrics and the general process to develop such metrics, we show the improvement of the MLPs in the prediction of physical properties. Overall, our results highlight general guidance and potential challenges in the future development of accurate, robust, and reliable MLPs for atomistic modeling.

Chapter 4: Assessing the accuracy of MLPs in predicting the elemental ordering: A case study on Li-Al alloys

4.1 Introduction

In Chapter 3, we revealed a number of discrepancies in the machine learning interatomic potentials (MLPs) in predicting atom dynamics and the properties.^[41–43] Given that the training of MLPs cannot include all possible atomic configurations that can be encountered during their atomistic modeling, such as molecular dynamics (MD) / Monte Carlo (MC) simulations, it is worth testing how MLPs would perform on these atom configurations outside of the training dataset. In this chapter, we test how MLPs can reproduce complex orderings of different phases in alloy systems. We conduct the MLP tests in binary alloy materials systems, which provide a variety of phases with a wide composition range of complex elemental ordering in each phase. Studying the thermodynamics of alloy materials systems requires sampling a large number of complex atomistic configurations that are usually beyond the computation power of using density functional theory (DFT) calculations. Generally, MLPs are trained on a limited number of configurations with pre-selected lattice structures and compositions. It is still an open question about the capability of MLP to make accurate predictions on lattice structures, compositions, and the ordering of atomic sites outside the training data, which can be considered as the transferability of MLPs.^[2,40–43]

Multiple previous studies employ MLPs to study the thermodynamics of alloys systems and their complex orderings in a range of lattice structures and compositions.^[88–91] To correctly predict the thermodynamics of different phases in an alloy system, the energies of atomistic configurations, in particular their lowest energy configurations (LEC), for all relevant phases with different structures and compositions must be accurately reproduced. In addition to the LECs, the

relative energies of different atomistic configurations or element orderings are also important for the corresponding thermodynamics, including determining the anti-site defects, diffusion, element dissolution, solid solution, and disordering of the system. A common application is in the (MC) simulations of alloys, which evaluate and compare the energies of a wide range of configurations for the alloy system. However, during MD/MC simulations, it is possible encountering atomistic configurations, site orderings, lattice structures, or compositions that were not included in the training data of MLP. Thus, the testing of the MLPs on reproducing these properties, which can be considered the transferability to these other phases and orderings, is essential before conducting MD/MC simulations.^[16] Therefore, testing the MLP performance is required and MLPs should be tested on a variety of lattice structures and compositions, both in- and out- of the training data. In addition, the testing scheme should be conducted in a manner that are commonly employed in MD/MC simulations, as a true testing of the MLPs. Such testing is essential to correctly employ MLPs on a wide range of relevant phenomena of alloys, to identify potential errors, and to improve MLPs for optimal performance on atomistic modeling of alloys.

A number of studies have investigated the effect of training data selection and have developed various strategies to train MLPs with improved performance. Zapiain et al. suggest that increasing the diversity of training data and the training data size are important by comparing the performance of MLPs trained by different datasets.^[41] They find the MLPs trained by the dataset that is larger and has more diversity have better transferability, and they develop a protocol to select training data and maximize the complexity of the data. Similarly, Zhang et al. develop an active-learning scheme, DP-GEN, to train accurate and transferable MLPs.^[43] In Chapter 4.3, we evaluate the performance of MLPs affected by the selection of training dataset, such as the different alloy phases included.

In this chapter, we train moment tensor potential (MTP) models on Li-Al binary alloys and systematically test the performance and transferability of these models on various lattice structures and compositions in- and out- of the training data and compare the results with DFT calculations. We find that MLPs trained by the training data with only a few phases may achieve high accuracy and good transferability to the orderings in different phases with low energy errors (Chapter 4.2). In addition to conventional testing errors of energies, we develop new evaluation metrics to evaluate the accuracy of energy rankings of elemental orderings and show that the MLPs well reproduce the energy ranking for other phases not included in the training data (Chapter 4.3). We find that increased diversity without significantly increasing training data size may lead to worse performance (Chapter 4.4). Finally, we perform three sets of MC and MD simulations with different supercell sizes and find mixed results as discrepancies caused by the supercell sizes in some of these applications (Chapter 4.5).

4.2 Evaluating the performance of the MLP across the alloy phase diagram

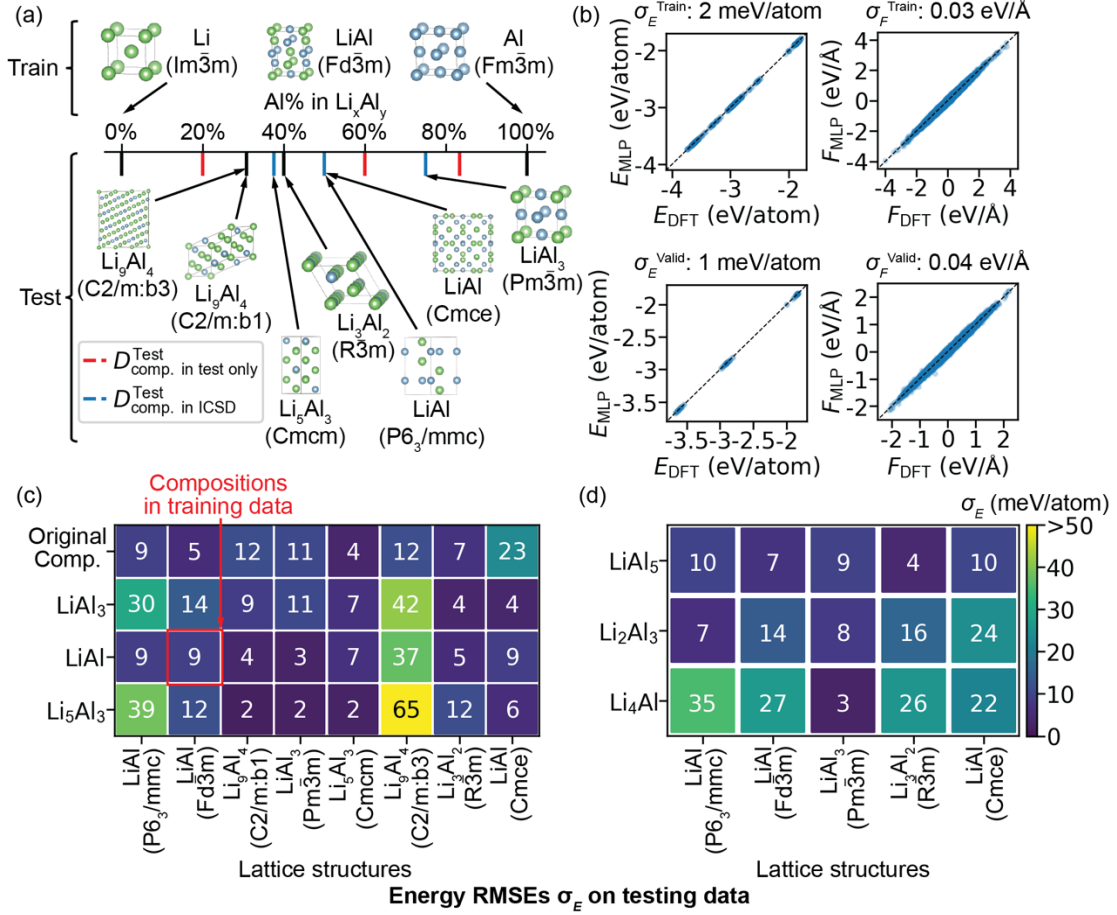


Figure 4.1. The training and testing of MLP in Li-Al alloy system. (a) Among 10 known phases of Li_xAl_y , BCC Li ($Im\bar{3}m$), BCC LiAl ($Fd\bar{3}m$), and FCC Al ($Fm\bar{3}m$) are used in training data, and the other seven intermediate phases are used in testing data $\mathcal{D}^{\text{test}}$. (b) Comparison of the energies and atomic forces predicted by MTP versus the benchmark DFT K4 results on training and validation datasets, $\mathcal{D}^{\text{train}}$ and $\mathcal{D}^{\text{validation}}$. (c, d) The RMSEs of predicted energies of up to 30 ordered configurations with different compositions (y axis) in a given lattice structure of known phases (x axis).

We perform our study using the Li-Al alloy system, which includes body-centered cubic (BCC) Li ($Im\bar{3}m$), face-centered cubic (FCC) Al ($Fm\bar{3}m$), and eight known Li_xAl_y intermediate phases with different compositions from the Inorganic Crystalline Structure Database (ICSD). The intermediate phases include BCC structure, such as Li_9Al_4 ($C2/m:b1$), Li_9Al_4 ($C2/m:b3$), Li_5Al_3

($Cmcm$), Li_3Al_2 ($R\bar{3}m$), and $LiAl$ ($Fd\bar{3}m$), FCC structure $LiAl_3$ ($Pm\bar{3}m$), HCP structure $LiAl$ ($P6_3/mmc$), and others structures $LiAl$ ($Cmce$), as shown (**Figure 4.1a**). To train the MTP model for Li-Al binary alloy system, the training dataset includes the end phases BCC Li ($Im\bar{3}m$) and FCC Al ($Fm\bar{3}m$) (Appendix F), and only one intermediate phase BCC LiAl ($Fd\bar{3}m$). The training dataset $\mathcal{D}^{\text{train}}$ is generated from these three phases, including a variety of atomic configurations from crystalline bulk structures, strained and distorted bulk structures, liquid structures, defected structures with multiple vacancies or interstitials, interfaces between each pair of the three selected phases with the surfaces with Miller indices of $\{001\}$, $\{011\}$ and $\{111\}$, and the snapshots obtained from AIMD simulations (Appendix F) of bulk crystalline phases, bulks with point defects, and interface supercells. The trained MLP accurately predicts atomic forces and energies on $\mathcal{D}^{\text{train}}$ and the validation dataset $\mathcal{D}^{\text{validation}}$ (consisting of 66 snapshots from AIMD simulations, Appendix F), compared with DFT calculations using a k -point mesh of $4 \times 4 \times 4$ (DFT K4). The RMSEs of energies are as low as 3 meV/atom for $\mathcal{D}^{\text{train}}$ and 2 meV/atom for $\mathcal{D}^{\text{validation}}$, and the RMSEs of atomic forces are 0.03 eV/Å in both $\mathcal{D}^{\text{train}}$ and $\mathcal{D}^{\text{validation}}$ (Figure 4.1b).

Here, we test whether the MLP trained on only three phases (i.e., two end phases and one intermediate phase), referred as MLP_{Train} , can properly predict a number of key properties for other intermediate phases over the entire composition range, with different orderings in different crystal structures and lattices over the entire Li-Al alloy system. The testing dataset, $\mathcal{D}^{\text{test}}$, uses the all intermediate phases, including other intermediate phases, i.e., four BCCs for Li_9Al_4 ($C2/m:b1$), Li_9Al_4 ($C2/m:b3$), Li_5Al_3 ($Cmcm$), and Li_3Al_2 ($R\bar{3}m$), FCC $LiAl_3$ ($Pm\bar{3}m$), HCP $LiAl$ ($P6_3/mmc$), and the $LiAl$ ($Cmce$) (Figure 4.1c). $\mathcal{D}^{\text{test}}$ also includes hypothetical phases that are generated by a number of different Li-Al compositions with the lattice structures of known intermediate phases (Figure 4.1c and d) (Appendix F). Given atomic environments differ significantly in different

phases (Figure E19), it is important to test MLPs' capabilities in correctly modeling the hypothetical phases that are generated by with other compositions, which may be the same as known existing intermediate phases (blue lines, Figure 4.1a, and lattice structures, Figure 4.1c) or compositions that deviate significantly from those of known intermediate phases (red lines, Figure 4.1a, and lattice structures, Figure 4.1d). Up to 30 configurations are generated for each of the intermediate and the hypothetical phases (Figure 4.1c and d) by swapping the positions of a number of randomly selected atoms in known intermediate phases and by random sampling the ordering the partially substituted sites in hypothetical phases (Appendix F). The $\mathcal{D}^{\text{test}}$ covers a wide range of lattice structures, compositions, and orderings, allowing us to test the MLP for a diverse range of cases that are distinct from the training data.

This MLP_{Train} trained on only three phases shows low RMSEs of energies at 2 – 14 meV/atom for most of the other intermediate Li-Al phases and their hypothetical variations not included in the training process (35 out of 47, Figure 4.1c and d). The HCP LiAl (*P6₃/mmc*) and the LiAl (*Cmce*), which are neither BCC nor FCC, show energy RMSEs greater than 20 meV/atom (the 1st row, Figure 4.1c). The MLP shows a larger error of energy RMSEs higher than 30 meV/atom for five materials, HCP Li₅Al₃ (*P6₃/mmc*), HCP LiAl₃ (*P6₃/mmc*), HCP Li₄Al (*P6₃/mmc*), BCC Li₅Al₃ (*C2/m:b3*), BCC LiAl (*C2/m:b3*), and BCC LiAl₃ (*C2/m:b3*) (Figure 4.1c and d). The notably larger errors in HCP structures may be caused by the identical nearest-neighbor atomic environments between HCP and FCC in the atomistic descriptor. Similarly, the larger errors on *C2/m:b3* BCC may also be explained by the confusion with another BCC *C2/m:b1* (Figure 4.1c). Additionally, we test our MLP in the structures with single Li or Al vacancy and the energy RMSEs are as low as 19 meV/atom, which confirms the good performance of the MLP to vacancy structures outside the training data (Appendix F). In summary, our tests of the MLP in general

show good results for predicted energies over a wide range of the lattice structures, compositions, and orderings in Li-Al alloy systems. Given that the MLP is trained using only two end phases and one intermediate phase, this implies the good performance and transferability of MLPs for the energies of atomistic configurations in the lattice model for alloy systems.

4.3 Energy rankings of different orderings

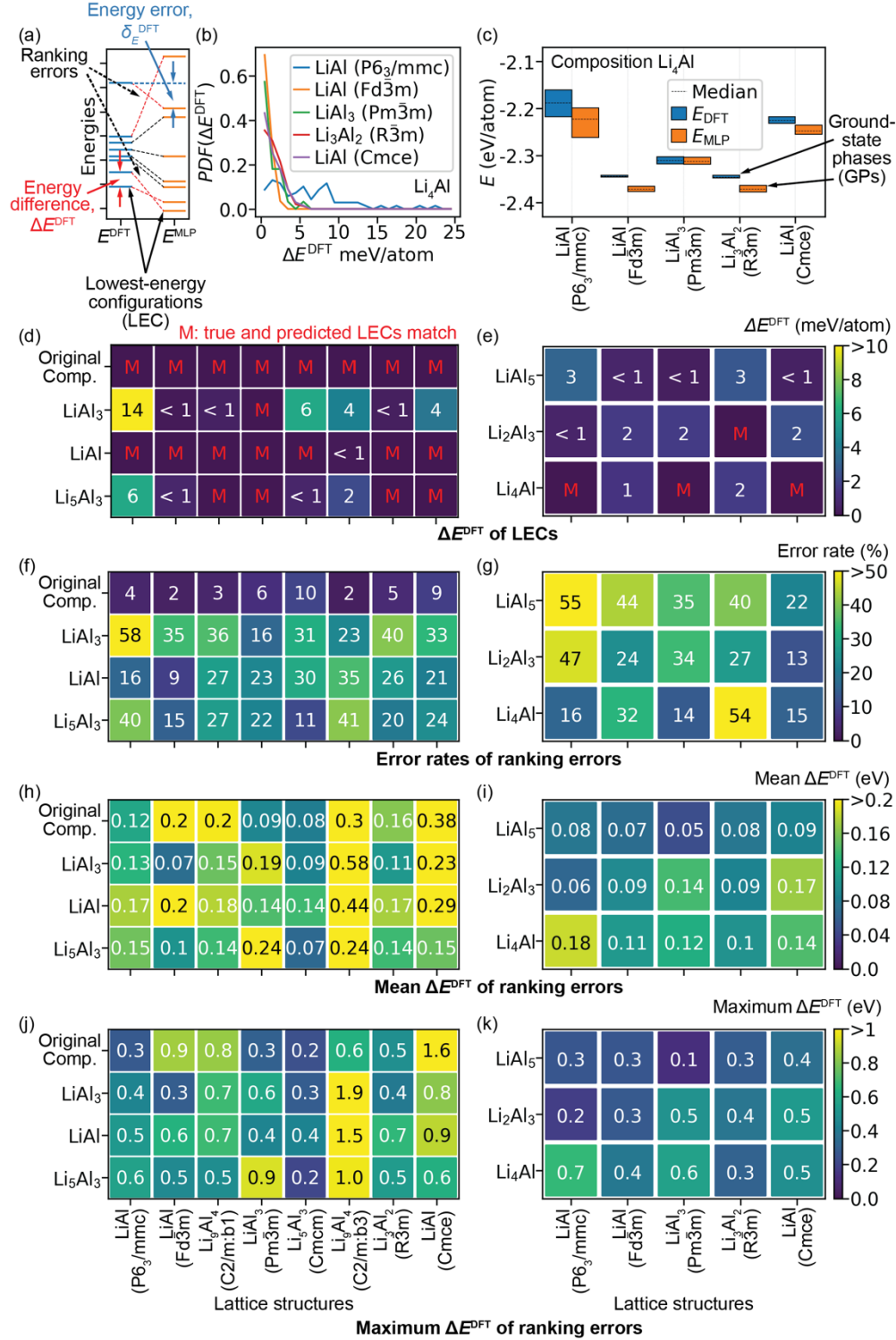


Figure 4.2. The predicted energy rankings of configurations with different atomistic orderings. (a) The illustration of the energy rankings of different orderings by DFT and MLP. The crossings of red dashed lines indicate ranking errors. δ_E^{DFT} is the energy error between the true DFT and the predicted MLP energies of the same configuration (blue gap). ΔE^{DFT} is the DFT energy difference

of two mismatched configurations in a ranking error (red gap). The lowest energy lines correspond to lowest-energy configurations (LECs). (b) The distributions of ΔE^{DFT} for all ranking errors and (c) the box plot of the energy distributions (the maximum, minimum, and median shown in dashed lines) of up to 30 configurations from Li_4Al in five lattice structures ($R\bar{3}m$ as the true phase, and others are hypothetical) from the $\mathcal{D}^{\text{test}}$. (d, e) The differences of DFT energies, ΔE^{DFT} , between the LECs calculated by DFT and predicted by MLP, (f, g) ranking error rates, (h, i) the mean ΔE^{DFT} and (j, k) the maximum ΔE^{DFT} for all ranking errors in $\mathcal{D}^{\text{test}}$. A match of LEC predicted by DFT and MLP is marked as M.

Here we examine the MLP prediction on the energies, the relative stabilities of different elemental orderings for both the known existing and the hypothetical phases. The MLP's capability to correctly reproduce the energetics of elemental ordering in alloys is important for studying the thermodynamical stabilities, constructing the phase diagram, understanding defects and disordering, and conducting MC/MD simulations.

We first test the lowest-energy configuration (LEC) for each of all phases from $\mathcal{D}^{\text{test}}$. The LEC is the configuration with the lowest calculated or predicted energy, i.e., the most favorable atomistic ordering of sites in the phase, among all competing configurations of a given phase. The LECs of all eight known Li-Al intermediate phases are correctly reproduced by the MLP (the 1st row of Figure 4.2d), for 50% of the 24 predictions of hypothetical materials for **Figure 4.2d** and four of the 15 predictions for Figure 4.2e. The MLP correctly predicts most of the LECs of hypothetical LiAl (except for LiAl $C2/m:b3$) in all lattice structures, possibly because the composition LiAl is included in the training data. For a given composition among selected lattice structures (Figure 4.2c, E2, and E3), the MLP correctly predicts the ground-state phase (GP), which is the most favorable lattice structure of a given composition, ($R\bar{3}m$) of LiAl_4 among the hypothetically generated phases with lattice structures of $P6_3/mmc$, $Fd\bar{3}m$, $Pm\bar{3}m$, $R\bar{3}m$, and $Cmce$ (Figure 4.2c). In addition to LiAl_4 , GPs are also correctly predicted for LiAl ($Fd\bar{3}m$), LiAl_3 ($Pm\bar{3}m$), and LiAl_5 ($Pm\bar{3}m$) (Appendix E). The GPs were incorrectly predicted to be $Fd\bar{3}m$ for

Li_5Al_3 and Li_2Al_3 by MLP, whereas correct GPs predicted by DFT are $Cmcm$ and $Pm\bar{3}m$, respectively. Even for the GPs incorrectly predicted by the MLP, the energy differences, ΔE^{DFT} , between the LECs of the true and the MLP-predicted GPs are as low as 10 and 7 meV/atom for Li_5Al_3 and Li_2Al_3 respectively (Appendix E). For most phases, ΔE^{DFT} are generally small, as low as 2 meV/atom for 13 out of the 23 MLP incorrectly predicted GPs. Generally, the MLP shows good performance in predicting LECs and GPs for a wide variety of phases in the materials system.

We further examine whether the MLP can correctly predict the relative rankings of the energies for each of the original and the hypothetical phases. The relative energy ranking of a sequence of atomistic configurations with different orderings is important for studying the relative stabilities, such as in anti-site defects and order-disorder transitions, of alloys and for conducting MC simulations. To quantify how the true and the predicted rankings match, we develop several evaluation metrics based on ranking errors, i.e., a pair of configurations that their true and the predicted energy rankings are opposite (crossings with red dashed lines, Figure 4.2a). While the energy error, δ_E^{DFT} between the true DFT and the MLP predicted energies of the same configuration (blue gap, Figure 4.2a), we also define the difference of DFT energies between these two mismatched configurations, ΔE^{DFT} (red gap, Figure 4.2a) as another measure of the ranking error. We calculate the error rates, the mean ΔE^{DFT} , and the maximum ΔE^{DFT} of all ranking errors as three novel evaluation metrics to evaluate the predicted energy ranking (Appendix F). The error rates for all intermediate phases in the ICSD are low ($< 10\%$ as shown in the 1st row of Figure 4.2f). In $\mathcal{D}^{\text{test}}$, a majority of (24 out of 39) hypothetical phases have mismatch probabilities in the range of 11 - 33% (Figure 4.2f and g), while only three are higher than 50%. We also quantify the mean and maximum values of (Figure 4.2b as an example) ΔE^{DFT} , mean ΔE^{DFT} and maximum ΔE^{DFT} , respectively, between DFT energies of all mismatched configuration pairs. Three of the

eight intermediate phases have mean ΔE^{DFT} lower than 0.15 eV, so as 26 of the total 39 hypothetical materials, and four maximum ΔE^{DFT} of existing intermediate phases and 21 of the total 39 hypothetical phases are lower than 0.5 eV (Figure 4.2h - k). Notably, the ΔE^{DFT} distributions of HCP Li-Al alloys with lattice structure $P6_3/mmc$, in general, have larger spread (possible ranking errors even though configurations are clearly more favorable in DFT) than the other Li-Al alloys. This may be caused by the fact that the specific HCP atomic arrangement is not considered in the training data but may share similar descriptors for the nearest-neighbor environments as the FCC configurations (blue line, Figure 4.2b and E4). Overall, the MLP only trained on three phases shows decent performance in predicting the relative energy rankings in most intermediate and hypothetical phases, except for the HCP phases. This implies decent performance in MD/MC simulations for different concentrations and different lattice structures. Relatively large errors of energy rankings for some hypothetical phases suggest potential directions for further improvements.

4.4 Effect of training data diversity

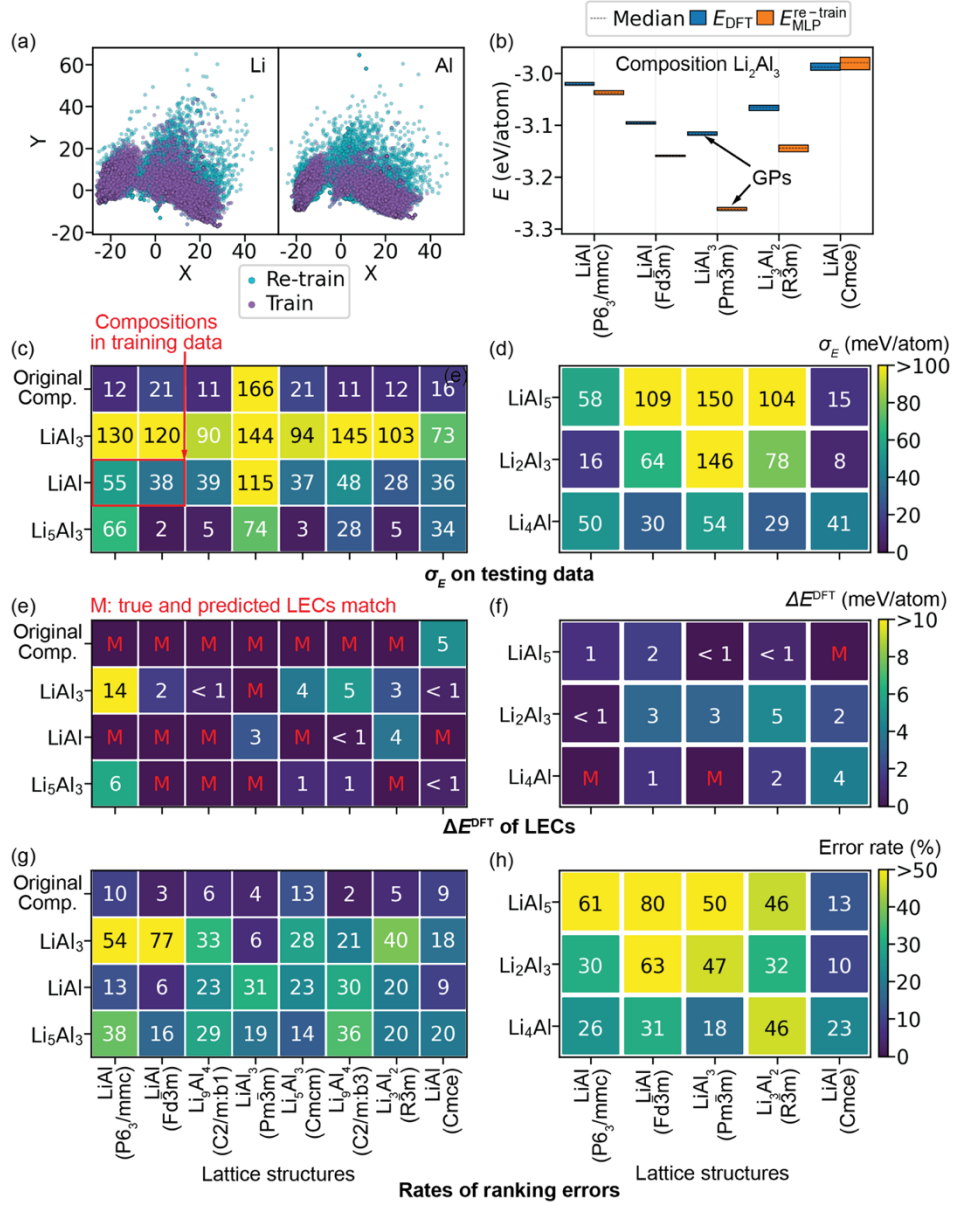


Figure 4.3. Performances of $MLP_{\text{Re-train}}$. The atomic environments of (a) Li and Al atoms in $\mathcal{D}^{\text{train}}$ (blue) and $\mathcal{D}^{\text{re-train}}$ (orange), generated by principal component analysis (PCA) 1st and 2nd principal components from Smooth Overlap of Atomic Positions (SOAP) descriptors (Appendix F). (b) The box plot of the energy distributions (the maximum, minimum, and median shown in dashed lines), as Figure 4.2c, of up to 30 configurations generated in five lattice structures with the composition Li_2Al_3 from $\mathcal{D}^{\text{test}}$ (the $Pm\bar{3}m$ as the true phase, and others are hypothetical). (c, d) The RMSEs of predicted energies of up to 30 ordered configurations with different compositions (y axis) in a given lattice structure of known intermediate phases (x axis). (e, f) The differences of DFT energies, ΔE^{DFT} between the LECs calculated by DFT and predicted by MLP for $\mathcal{D}^{\text{test}}$ as Figure 4.2d and e. (g, h) The ranking error rates. A match of LEC predicted by DFT and MLP is marked as M.

While the $\text{MLP}_{\text{Train}}$ tested in 4.1-4.3 was trained on only three phases, here we explore the MLP by increasing the diversity of training data, which is a common practice to improve ML models including MLPs.^[40,41,92,93] The new training dataset, $\mathcal{D}^{\text{re-train}}$, included one more material, HCP LiAl ($P6_3/mmc$), whereas only BCC and FCC lattices were in the original training dataset $\mathcal{D}^{\text{train}}$ for $\text{MLP}_{\text{Train}}$ (4.1-4.3). 144 atomistic configurations with the HCP lattice including crystalline bulk with a range of orderings, liquid, point defects, and AIMD snapshots of bulk and defect bulks, were generated in the same manner as other phases, and replaced 24% of configurations in $\mathcal{D}^{\text{train}}$. With the added HCP data, the size of $\mathcal{D}^{\text{re-train}}$ has increased 16% compared to $\mathcal{D}^{\text{train}}$.

Here, we perform a similar test for the re-trained MLP, $\text{MLP}_{\text{Re-train}}$, on $\mathcal{D}^{\text{test}}$ as in 4.3. For hypothetical phases outside the training data, $\text{MLP}_{\text{Re-train}}$ performs significantly poorer than $\text{MLP}_{\text{Train}}$, with only six hypothetical phases having energy RMSEs lower than 15 meV/atom, compared to 28 from $\text{MLP}_{\text{Train}}$ (**Figure 4.3c** and d). For example, $\text{MLP}_{\text{Re-train}}$ gives a higher energy RMSE of 103 meV/atom for hypothetical LiAl₃ in the BCC ($R\bar{3}m$) lattice structure, whereas $\text{MLP}_{\text{Train}}$ gives only 4 meV/atom for the same hypothetical phase. $\text{MLP}_{\text{Re-train}}$ also gives poorer predictions in energy rankings, especially for hypothetical phases with high error rates. There are six materials exhibiting large error rates ($> 50\%$) for our $\text{MLP}_{\text{Re-train}}$, with an error rate as high as 80% on LiAl₅ $Fd\bar{3}m$ (**Figure 4.3g** and h). The $\text{MLP}_{\text{Re-train}}$ also performs poorer than $\text{MLP}_{\text{Train}}$ on other metrics for energy rankings. The mean ΔE^{DFT} of ranking errors increases to 0.07 – 0.11 eV for hypothetically generated LiAl $Pm\bar{3}m$, Li₂Al₃ $Fd\bar{3}m$, Li₄Al $Pm\bar{3}m$, and Li₄Al $Cmce$. There are 22 hypothetical phases from $\mathcal{D}^{\text{test}}$ with the mean ΔE^{DFT} of ranking errors lower than 0.15 eV, downing from 26 of $\text{MLP}_{\text{Train}}$ (**Appendix E**). The number of correctly predicted LECs also decline to 13 from 16 by $\text{MLP}_{\text{Train}}$ (**Figure 4.3e** and f). We examine three other re-trained MLPs, $\text{MLP}_{\text{Re-}}$

$\text{MLP}_{\text{train}}^{\text{extra-1}}$, $\text{MLP}_{\text{Re-train}}^{\text{extra-2}}$, and $\text{MLP}_{\text{Re-train}}^{\text{extra-3}}$, that are selected by the approach described in Appendix F, and all of them have low RMSE of energies as low as at 8 meV/atom on $\mathcal{D}^{\text{train}}$ (24 meV/atom on $\mathcal{D}^{\text{re-train}}$), but have similarly poor performances in $\mathcal{D}^{\text{test}}$ and in the same tests as above (Figure E14, E16, and E18 in Appendix E). This indicates that the poorer performance of $\text{MLP}_{\text{Re-train}}$ is a result of different training datasets not model selection (Figure E14, E16, and E18 in Appendix E). In summary, the MLP trained with more diverse data does not show improved performance to other materials out of the training data without significantly increasing the training data size. While only two training datasets are tested here, further in-depth studies would provide insights about how to select the training data for further MLP improvement.

4.5 Other errors in the applications of MLPs

4.5.1 Errors in large supercells

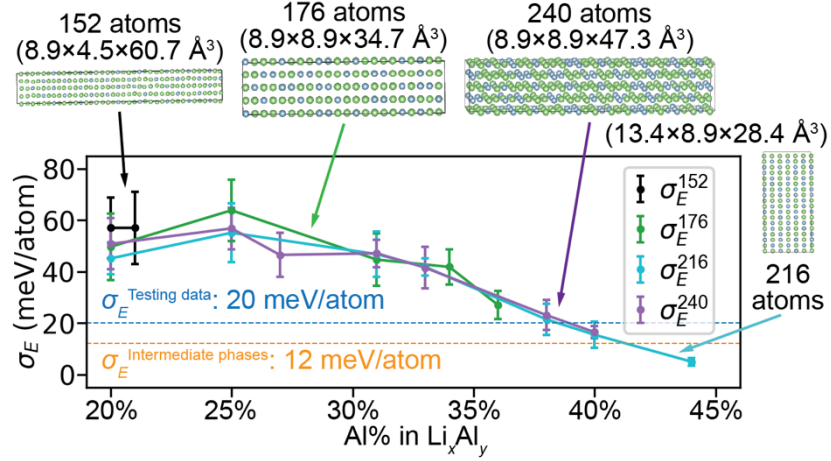


Figure 4.4. The energy RMSEs of Li-Al supercells with different compositions and sizes. Dashed lines are referred to energy RMSEs of all existing intermediate phases (orange) and hypothetical phases from $\mathcal{D}^{\text{test}}$ (blue).

Here we further test MLPs in three application cases, in which MLPs are directly utilized to study physical phenomena. A key advantage of MLP technique is to conduct MD/MC simulations in the atomistic models with large cell sizes, which are not easily accessible by DFT, for studying elemental orderings, surfaces, and grain boundaries in materials.^[94–97] We here test the MLP in large supercells with four different sizes, such as $8.9 \times 4.5 \times 60.7 \text{ \AA}^3$, $8.9 \times 8.9 \times 34.7 \text{ \AA}^3$, $8.9 \times 8.9 \times 47.3 \text{ \AA}^3$, and $13.4 \times 8.9 \times 28.4 \text{ \AA}^3$ (Appendix F), which consist of 152, 176, 216, and 240 atoms with the aspect ratios of 1: 13.5, 1: 3.9, 1: 3.2, and 1: 5.3, respectively. For each supercell, we generate multiple compositions ranging from Li_4Al to Li_5Al_4 (**Figure 4.4**) (Appendix F) for testing, while specific compositions vary due to the number of atoms in the supercell. While our $\text{MLP}_{\text{Train}}$ gives low energy RMSEs as low as 12 meV/atom for all existing Li-Al intermediate phases and 20 meV/atom for all hypothetical phases, here the same MLP tested on such large modeling supercells exhibit significantly higher the energy RMSEs well above 30 meV/atom at

Al percentage $< 35\%$ (Figure 4.4). The energy RMSEs of all supercells show large errors as high as 50 – 60 meV/atom at around Li_3Al (roughly 25% Al) (Figure 4.4). Given the low errors on different ordering as tested for $\text{MLP}_{\text{Train}}$ in $\mathcal{D}^{\text{test}}$, this may be caused by the different elemental ordering or configurations encountered in these large supercells, as shown by the PCA of quantified atomic environments using SOAP in these large supercells (Figure E19) (Appendix E). This observed error may be a potential issue in applying MLPs in larger simulation cells.

4.5.2 Phase diagrams and phase transitions by MLP

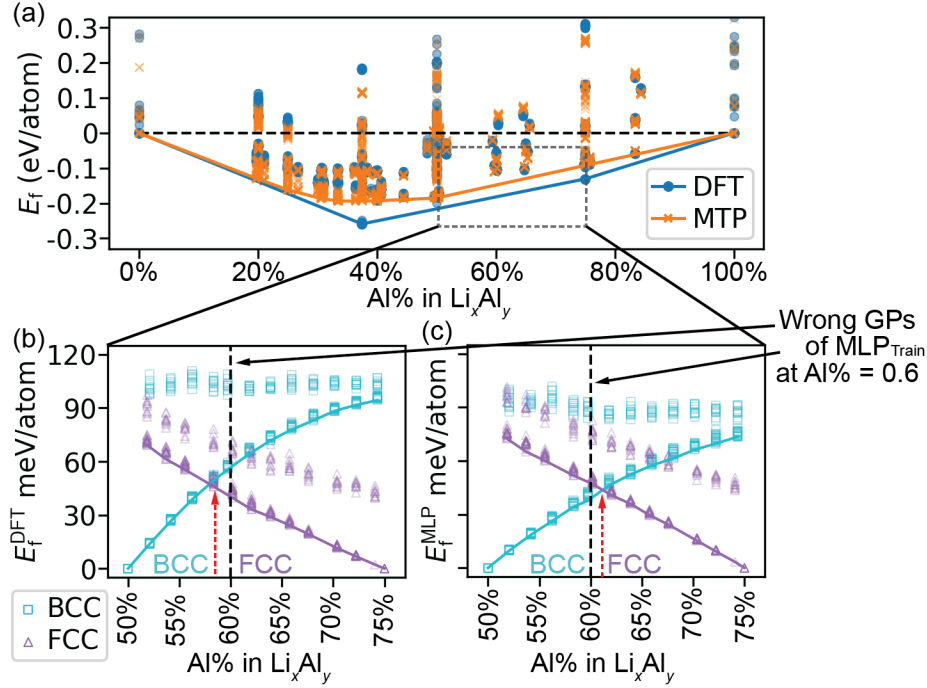


Figure 4.5. Convex hull and phase stability of Li-Al alloy system. (a) The formation energies E_f and the convex hulls of Li-Al alloy system were constructed using all training and testing data calculated by DFT (blue line and circles) and MLP (orange line and crosses). (b, c) The phase transition of Li-Al between BCC ($\text{LiAl } Fd\bar{3}m$, cyan line and squares) and FCC ($\text{LiAl}_3 Pm\bar{3}m$, purple line and triangles) by (b) DFT and (c) MLP. Red dash arrow indicates the transition point by E_f . The black dashed line at Al% = 60% highlights the Li_2Al_3 composition, which has GP incorrectly predicted by $\text{MLP}_{\text{Train}}$ (Figure 4.2c), within the area that $\text{MLP}_{\text{Train}}$ predicts wrong GPs between BCC and FCC (Appendix E).

Here, we construct and compare the energy convex hull of Li-Al alloy system by using MLP and DFT. The convex hull is constructed using calculated formation energies E_f of all configurations from $\mathcal{D}^{\text{train}}$ and $\mathcal{D}^{\text{test}}$ (Figure 4.1c and d), the relaxed structures from the ICSD (Figure 4.1a), the long slabs $\mathcal{D}^{\text{slabs}}$ in Figure. 4.4a, and the structures with single vacancies $\mathcal{D}^{\text{Vacancy}}$ (**Figure 4.5**) (Appendix E). Compared with DFT, the MLP largely reproduces the overall trend of the convex hull, with some discrepancies in equilibrium phases and the corresponding energies (Figure 4.5). Both DFT and MLP give the lowest E_f configuration at Li_5Al_3 , though the

minimum E_f differ by 65 meV/atom. However, while there are only two stable intermediated phases given DFT, MLP predicted a total of eight stable intermediate phases with most of the additional phases in the composition range of 19% – 50% of Al.

We further examine the phase transition between BCC LiAl and FCC LiAl₃ intermediate phases. By comparing the energies of different configurations across the compositions range from 50% to 75% Al based on the BCC ($Fd\bar{3}m$) and FCC ($Pm\bar{3}m$) lattices (Figure 4.5b and c), the transition between BCC and FCC is at 57% Al predicted by MLP in good agreement with that at 62% Al by DFT. The deviations of the energies of other compositions based on these phases may cause the deviation (Figure 4.5b and c).

4.5.3 Errors of forces on migrating atoms in MD simulations

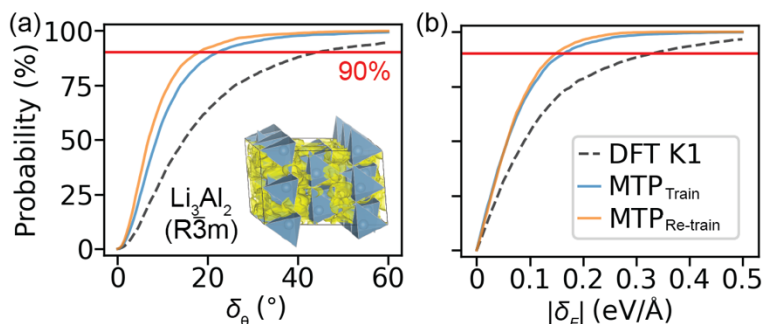


Figure 4.6. Errors of forces predicted by MLP on migrating atoms in MD simulations. The accuracies of force predictions, including force (a) directions and (b) magnitudes, on migrating atoms in snapshots taken from AIMD simulations for Li₃Al₂ ($R\bar{3}m$). The cumulative distribution functions (CDFs) are plotted for forces by DFT K1 (dashed black lines), MLP_{Train} (blue), and MLP_{Re-train} (orange). Red lines are visual guidance for 90%.

Here we test the MLP on studying atom diffusion in MD simulations, another key application of MLPs. In the earlier study (Chapter 3), we reveal that atomic forces on migrating atoms are a key error source for MLPs. Here, we quantify the errors of atomic forces on migrating

atoms, which are extracted from AIMD simulations (200 snapshots) for $\text{LiAl } Fd\bar{3}m$, $\text{Li}_3\text{Al}_2 R\bar{3}m$, $\text{Li}_5\text{Al}_3 Cmc m$, and $\text{Li}_9\text{Al}_4 C2/m:b1$ with single Al vacancies (Appendix F, **Figure 4.6a, b**, and E20). The atomic forces predicted by the MLP is compared to the low-accuracy setting of the DFT with a single Γ -centered k -point (DFT K1) same as in Chapter 3, since K1 is commonly used in AIMD simulations for lower computation costs. The MLP achieves good accuracy in predicting atomic forces on migrating atoms for all four materials, even though the three intermediate materials Li_3Al_2 , Li_5Al_3 , and Li_9Al_4 are outside the training data. As shown in the CDF curves of force errors on RE atoms obtained from Li_3Al_2 AIMD diffusion snapshots, over 90% of migrating atoms predicted by MLP has less than 22 degrees for force directional error and less than 0.16 eV/Å for magnitude error, while the CDF curve of DFT K1 shows at the same 90% cut-off larger errors of force directional error of 44 degrees and magnitude error of 0.33 eV/Å (Figure 4.6a, b, and E20). The force performance score $P(\mathcal{D}^{\text{Li}_3\text{Al}_2})$, which is defined as $P(\mathcal{D}^{\text{Li}_3\text{Al}_2}) = \text{NAC}(|\delta_F|, \mathcal{D}^{\text{Li}_3\text{Al}_2}) \times \text{NAC}(\delta_\theta, \mathcal{D}^{\text{Li}_3\text{Al}_2})$ as proposed in Chapter 3 as an error evaluation metrics, is 0.69 for the MLP compared to 0.50 from DFT K1, indicating better performance of MLP in predicting forces on migrating atoms.

The re-trained MLP $\text{MLP}_{\text{Re-train}}$, which was trained with an additional HCP phases, also gives better performance on force errors with its $P(\mathcal{D}^{\text{Li}_3\text{Al}_2})$ at 0.73. Notably, though $\text{Li}_3\text{Al}_2 R\bar{3}m$ and $\text{Li}_5\text{Al}_3 Cmc m$ are not in $\mathcal{D}^{\text{train}}$ or $\mathcal{D}^{\text{re-train}}$, both CDF curves of force errors, predicted by either $\text{MLP}_{\text{Train}}$ or $\text{MLP}_{\text{Re-train}}$, exceed those of DFT K1 calculations, indicating even our $\text{MLP}_{\text{Re-train}}$ may outperform DFT K1 on predicting forces of RE atoms in unseen structures. It should be noted that DFT calculations for metal often require denser k -point density, and K1 is coarse k -point sampling even for the metal supercell models used. Since force predictions on migrating atoms are critical for diffusional properties, this good performance of MLP on the migrating atoms

not included in the training data is very encouraging, but careful testing as we have shown is still needed before application.

4.6 Conclusion and discussion

In this chapter, we train MLPs on Li-Al binary alloy materials system and test the model predictions on a variety of lattice structures and compositions in- and out- of the training data. We evaluate our MLP predictions on hypothetical phases that are completely outside the training process. The large number of the tested hypothetical phases allows us to closely examine the transferability of MLPs on materials with unknown lattices, compositions, and structures across the phase diagram. Our results show low energy errors for most of the hypothetical phases, including some materials having lattice structures far different from common BCC/FCC phases included in the training data. These results imply that our MLPs can largely capture the physics of interatomic bonds and may lead to good transferability to other alloy intermediate phases for a range of atomistic modeling applications. Furthermore, our practice of testing MLPs on a large number of hypothetical phases, including lattices, structures, and compositions completely outside the training process, is demonstrated as a feasible process to assess the errors, performance, and transferability of MLPs, and should be further evaluated and adopted in the testing of MLPs.

Additionally, in this chapter, we develop multiple new evaluation metrics and thoroughly examine the MLP predictions. Beyond the conventional error evaluation metric of energy RMSE, we propose evaluation metrics, such as the ranking error rates, the mean ΔE^{DFT} of ranking errors, the maximum ΔE^{DFT} of ranking errors, the GPs, and the LECs, test the predicted energy ranking of energies of different elemental ordering in different phases as a key component of the MLP performances. Our results show that good predictions on energy rankings though further

improvement can be made for hypothetical phases. Our novel evaluation metrics related to energy rankings should be further considered in the testing process of MLPs in the future.

In this chapter, our test shows promising results that good transferability to other intermediate alloy phases may be achieved by using a small number of training data, only three phases and 356 configurations, to predict 47 original and hypothetical phases with mostly low energy errors. In comparison, the re-trained MLPs with more phases and more data diversity in the training process have worse performances. This phenomenon suggests that data diversity alone, without substantially increasing the amount of training data, may not always lead to improved MLPs. The tests mimicking real applications schemes show previously not well-known pitfalls in predicting configurations with large supercell models that are often used in MD/MC simulations, mixed results in predicting the convex hull of the phase diagram, and promising performances in predicting forces on migrating atoms based on MD simulations. Both good results and discrepancies are observed in these three application tests, suggesting that further studies are needed to verify the accuracy in different application cases of MLPs.

To sum up, we systematically test our MLPs on materials in- and out- of the training data, assess the transferability of MLPs to a variety of hypothetical phases, develop evaluation metrics related to energy ranking, and examine MLPs in schemes close to applications. The newly developed metrics and the process of evaluating the transferability of MLPs as well as the effect on selecting appropriate training data can be further studied and adopted. Our study provides a comprehensive case study using a binary alloy system on the application and on MLP performances, which may offer some guidance for future studies in the field.

Chapter 5: High-dimensional assessment of MLP performances for multiple properties

5.1 Introduction

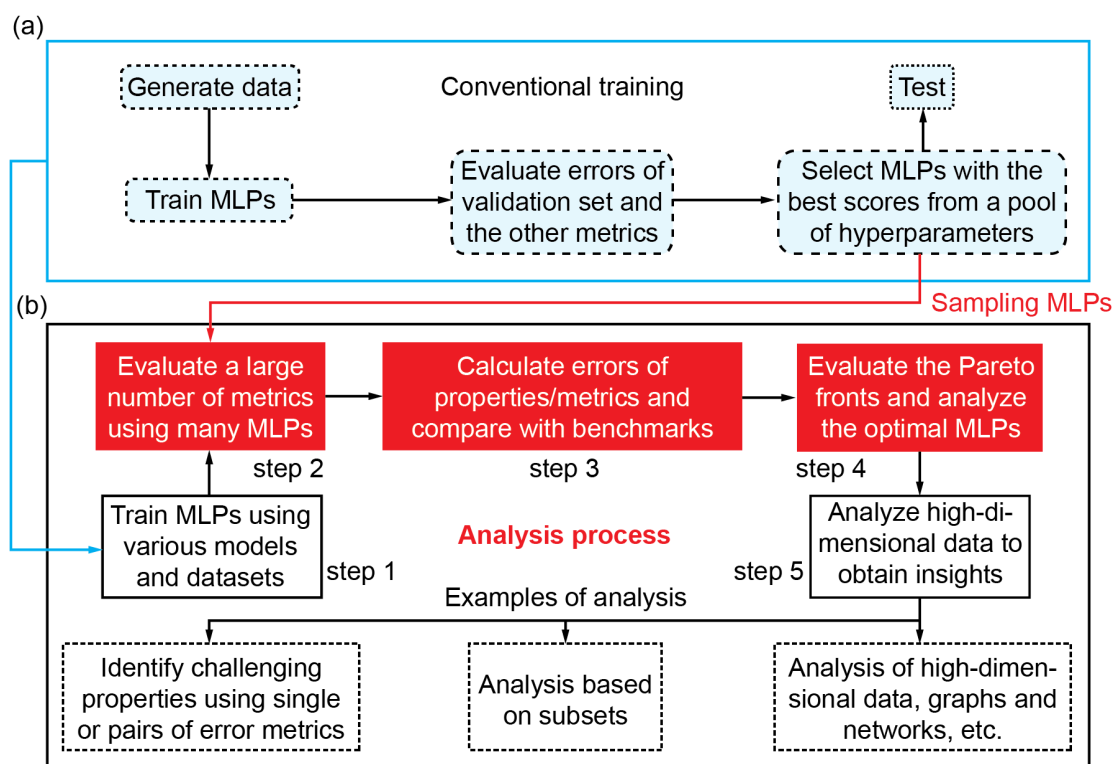


Figure 5.1. Schematic plot of our process to conduct an empirical analysis of MLP performances for multiple properties. (a) The scheme of typical MLP training process. A model pool consisting of many MLPs with different hyperparameters is usually generated in the validation step to fine tune the hyperparameter set for the final MLP. (b) The procedure of our analysis process.

In our systematic studies of machine learning interatomic potential (MLP) performances in Chapter 3 and 4, there are significant trade-off between two performance scores of force accuracies on migrating atoms in interstitial RE testing data versus those on migrating atoms in vacancy rare-event (RE) testing data as shown (Figure 3.6c). However, little is understood about how MLPs perform for multiple properties or performance metrics and how they correlate. In this

chapter, we introduce and employ an analysis process to collect and analyze a large number of MLPs and their performances on a diverse range of physical properties (**Figure 5.1**). This process samples MLPs from the pool generated in the validation step in MLP training and conducts empirical analysis of MLP performances using multiple evaluation metrics (Figure 5.1). This process consists five steps: 1) train MLPs using various models (such as GAP, NNP, MTP, etc.) and different training datasets, 2) sample a large number of MLPs from each of the model pools, 3) calculate a large number of physical properties of materials using each MLP, compute the errors of these properties as evaluation metrics, and adopt reasonable error ranges as benchmarks to assess these calculated errors, 4) construct Pareto fronts using multiple metrics, evaluate these Pareto fronts, and analyze the optimal MLPs on the Pareto fronts, and 5) assess MLP performances using multiple metrics, including analysis of subsets, clusters of MLPs, and graphs or networks of metrics.

In this chapter, we follow this process to perform an empirical analysis of MLP performances, using Si as a case study and utilizing the larged number of MLPs we trained in Chapter 3. We include many of our new evaluation metrics, developed in Chapter 3 and 4, into our dataset. With the rich number of metrics, we fully assess MLP performances for multiple properties and successfully identify multiple metrics that are challenging for MLPs to accurately predict. These results on challenging metrics provide valuable insights of MLP performances and may guide future training of improved MLPs.

5.2 Constructing the dataset of many MLPs and performance metrics

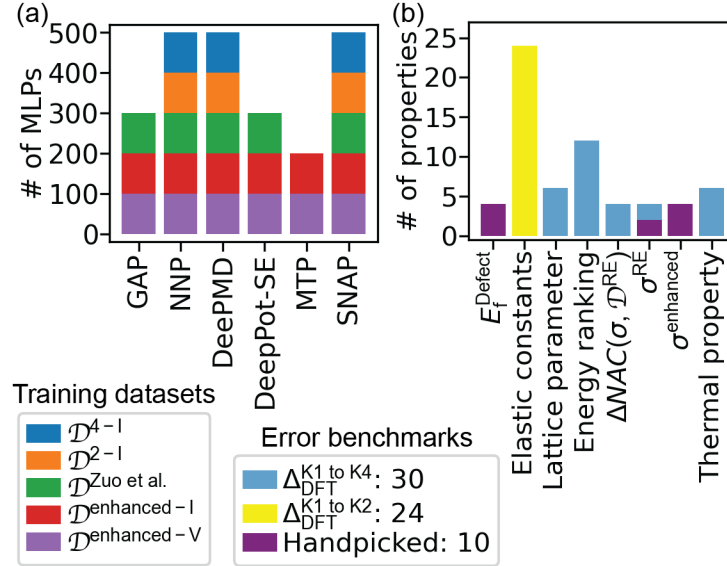


Figure 5.2. (a) The distribution of 2300 MLPs using six models and trained by five different datasets, $\mathcal{D}^{\text{Zuo et al.}}$ (green), $\mathcal{D}^{\text{enhanced-I}}$ (red), $\mathcal{D}^{\text{enhanced-V}}$ (purple), $\mathcal{D}^{2-I}$ (orange), and $\mathcal{D}^{4-I}$ (blue). (b) The 64 calculated properties, performance scores, and evaluation metrics over eight categories and the benchmarks applied on each metric (Table 5.1), including 30 metrics using the differences between DFT K4 and K1, $\Delta_{\text{DFT}}^{K1 \text{ to } K4}$, as benchmarks (blue), 24 metrics using the differences between DFT K2 and K1, $\Delta_{\text{DFT}}^{K1 \text{ to } K2}$, as benchmarks in elastic constants (yellow), and 10 handpicked values (purple).

We conduct our empirical analysis on 2300 MLPs with different models and training datasets to analyze the accuracy and errors of MLPs in predicting a large array of different properties in Si as a model materials system (step 1 to 3, Figure 5.1). In order to obtain a general conclusion of MLPs, the MLPs are trained using six current state-of-the-art MLP models, Gaussian Approximation Potential (GAP), Neural Network Potential (NNP), Spectral Neighbor Analysis Potential (SNAP), Moment Tensor Potential (MTP), Deep Potential (DeePMD), and Deep Potential-Smooth Edition (DeepPot-SE), the same with those used in Chapter 3, and trained on five different training dataset, $\mathcal{D}^{\text{Zuo et al.}}$,^[1] $\mathcal{D}^{\text{enhanced-I}}$ (the enhanced training dataset in Chapter 3.4.2 and Appendix D), $\mathcal{D}^{\text{enhanced-V}}$ (the vacancy-enhanced training dataset in Appendix D),

$\mathcal{D}^{2-I}$, and $\mathcal{D}^{4-I}$, (Appendix H) (**Figure 5.2a**). Every training dataset contains 219 configurations over a wide range of selected structures of solid Si, liquid Si, strained or distorted Si, Si surfaces, and bulk Si with single vacancy from AIMD simulations (Appendix H). To increase the data diversity among training datasets, $\mathcal{D}^{\text{enhanced}-I}$, $\mathcal{D}^{2-I}$, and $\mathcal{D}^{4-I}$ were generated by replacing 54.7% (120 configurations) of training data in $\mathcal{D}^{\text{Zuo et al.}}$ by the configurations of single interstitial, di-interstitials, and four interstitials, respectively, per supercell of Si bulks from AIMD simulations. The values of energies and forces are calculated by DFT with a k -point mesh of $4 \times 4 \times 4$ (DFT K4).

We select fully trained MLPs with different hyperparameters (Appendix H) in two ways. For each MLP model and each training dataset, we train a pool of MLPs with a wide range of hyperparameters to choose from, and the total number of MLPs selected from all the pools is 2300 MLPs (Appendix G). The first half of the MLPs in the entire pool are selected as the optimized MLPs that have the best (lower validation scores refer to better MLPs in our approach) validation scores from our validation dataset (Appendix H). The other half of the MLPs in the pool are randomly selected from the rest of the pool excluding those already selected in the first half. Through this selection, the pool of 2300 MLPs balances the optimized MLPs following the conventional training process and the MLPs that may not exhibit the lowest validation scores in conventional testing but may show particularly good performances for certain properties. This selection gives us a pool of MLPs to understand their limits on predicting certain properties (further discussed in Appendix G).

For this pool of 2300 MLPs, we evaluate every MLP by many properties and performance metrics (e.g., RMSEs in the predictions of physical properties) in the Si (**Figure 5.2b** and **Table 5.1**). As summarized (Table 5.1), these physical properties include the formation energies $E_{\text{f}}^{\text{Defect}}$ of four different types of point defects, such as vacancy, split- $\langle 110 \rangle$ interstitial, tetrahedral

interstitial, and hexagonal interstitial, the elastic constants of bulk crystalline Si and the Si structure supercell with a single point defect, the lattice parameters of bulk crystalline Si and the Si structure supercell with single point defect, the energy rankings (as defined in Chapter 4) for four sets of Si configurations with multiple vacancies, and the other properties, such as the free energy E_{free} , the entropy S , and the heat capacity K , for Si bulk crystalline phase and Si supercell with a single vacancy. Additionally, the performance metrics for the MLPs include the magnitude and directional errors of forces on rare event (RE) atoms in interstitial or vacancy diffusions, the normalized area of curve (NAC) $NAC(\delta, \mathcal{D}^{\text{RE}})$, as defined in Chapter 3.4.1. Other metrics include the root-mean-square errors (RMSEs) of energies and forces for configurations in $\mathcal{D}^{\text{RE-I}}$ (as defined in Chapter 3), and RMSEs of energies and forces on RE atoms σ^{RE} , as defined in Chapter 3, and the on RMSEs of energy and forces for enhanced training datasets $\mathcal{D}^{\text{enhanced-I}}$ and $\mathcal{D}^{\text{enhanced-V}}$ in Chapter 3.4.2, σ^{enhanced} . Here, to simplify our description, we collectively refer to all 64 properties and their performance metrics as ‘properties’ in this chapter, summarized as Table 5.1. In order to have a fair comparison, we calculate and analyze the log error of properties, and the error for each property are normalized as follows:

$$\delta M = \log_{10}|P_{\text{predicted}} - P_{\text{ideal}}| \quad \text{Eq. (5.1)}$$

Where δM is referred as ‘error metric’ with higher value indicating better MLP performance, $P_{\text{predicted}}$ is the property calculated by MLPs, and P_{ideal} is the ideal value, which for most cases is given by DFT K4 calculations. For 24 elastic constants, DFT calculations, using a k -point mesh of $2 \times 2 \times 2$ (DFT K2), are used as ideal values due to the high computation cost of DFT K4 calculations for the large supercell of interstitial structures (Figure 5.2b). The ideal value of the performance score of force directional errors $NAC(\delta_{\theta}, \mathcal{D})$ is 1 and the ideal value of energy RMSE is 0 (Appendix H).

In addition, we also compare the errors of the MLPs to the errors of prediction given by DFT in lower accuracy settings such as a single Γ -point (K1), $\Delta_{\text{DFT}}^{\text{K1 to K4}}$ ($\Delta_{\text{DFT}}^{\text{K1 to K2}}$ for elastic constants), as reasonable benchmarks of our metrics for comparison. For σ^{enhanced} and σ_E^{RE} metrics, the benchmarks were handset to 15 meV/atom for all energy RMSEs σ_E and to 0.12 eV/Å for all force RMSEs σ_F . The benchmarks of E_f^{Defect} were set to 0.1 eV, because even DFT K1 results produce errors as large as 1 eV in E_f^{Defect} (Figure 5.2b) (Appendix H).

Table 5.1. Summary of 64 properties and error metrics, their corresponding categories of physical properties, the configurations/datasets they are calculated from, and their benchmarks (shown as error values before logarithm).

Category	Configuration/Dataset	Error metric	Benchmark error
Defect formation energy, E_f^{Defect}	Vacancy	$\delta E_f^{\text{vacancy}}$	0.1 eV (handpicked)
	Split-<110> interstitial	$\delta E_f^{\text{split-<110>}}$	0.1 eV (handpicked)
	Tetrahedral interstitial	$\delta E_f^{\text{tetrahedral}}$	0.1 eV (handpicked)
	Hexagonal interstitial	$\delta E_f^{\text{hexagonal}}$	0.1 eV (handpicked)
Elastic constants	Vacancy	$\delta C_{11}^{\text{vacancy}}$	492 GPa ($\Delta_{\text{DFT}}^{\text{K1 to K2}}$)
		$\delta C_{12}^{\text{vacancy}}$	251 GPa ($\Delta_{\text{DFT}}^{\text{K1 to K2}}$)
		$\delta C_{44}^{\text{vacancy}}$	84 GPa ($\Delta_{\text{DFT}}^{\text{K1 to K2}}$)
		Bulk modulus, δK^{vac}	3.8 GPa ($\Delta_{\text{DFT}}^{\text{K1 to K2}}$)
	Split-<110> interstitial	$\delta C_{11}^{\text{split-<110>}}$	3.2 GPa ($\Delta_{\text{DFT}}^{\text{K1 to K2}}$)
		$\delta C_{12}^{\text{split-<110>}}$	4.9 GPa ($\Delta_{\text{DFT}}^{\text{K1 to K2}}$)
		$\delta C_{44}^{\text{split-<110>}}$	0.7 GPa ($\Delta_{\text{DFT}}^{\text{K1 to K2}}$)
		$\delta K^{\text{split-<110>}}$	2.1 GPa ($\Delta_{\text{DFT}}^{\text{K1 to K2}}$)
	Tetrahedral interstitial	$\delta C_{11}^{\text{tetrahedral}}$	21 GPa ($\Delta_{\text{DFT}}^{\text{K1 to K2}}$)
		$\delta C_{12}^{\text{tetrahedral}}$	20 GPa ($\Delta_{\text{DFT}}^{\text{K1 to K2}}$)
		$\delta C_{44}^{\text{tetrahedral}}$	26 GPa ($\Delta_{\text{DFT}}^{\text{K1 to K2}}$)
		$\delta K^{\text{tetrahedral}}$	5.8 GPa ($\Delta_{\text{DFT}}^{\text{K1 to K2}}$)
	Hexagonal interstitial	$\delta C_{11}^{\text{hexagonal}}$	0.5 GPa ($\Delta_{\text{DFT}}^{\text{K1 to K2}}$)
		$\delta C_{12}^{\text{hexagonal}}$	4.4 GPa ($\Delta_{\text{DFT}}^{\text{K1 to K2}}$)
		$\delta C_{44}^{\text{hexagonal}}$	3.0 GPa ($\Delta_{\text{DFT}}^{\text{K1 to K2}}$)
		$\delta K^{\text{hexagonal}}$	2.8 GPa ($\Delta_{\text{DFT}}^{\text{K1 to K2}}$)
	Unit cell with 8 atoms	$\delta C_{11}^{\text{unit cell}}$	57 GPa ($\Delta_{\text{DFT}}^{\text{K1 to K2}}$)
		$\delta C_{12}^{\text{unit cell}}$	5.5 GPa ($\Delta_{\text{DFT}}^{\text{K1 to K2}}$)
		$\delta C_{44}^{\text{unit cell}}$	17 GPa ($\Delta_{\text{DFT}}^{\text{K1 to K2}}$)
		$\delta K^{\text{unit cell}}$	15 GPa ($\Delta_{\text{DFT}}^{\text{K1 to K2}}$)
	Supercell with 64 atoms	$\delta C_{11}^{\text{supercell}}$	1.4 GPa ($\Delta_{\text{DFT}}^{\text{K1 to K2}}$)
		$\delta C_{12}^{\text{supercell}}$	4.0 GPa ($\Delta_{\text{DFT}}^{\text{K1 to K2}}$)
		$\delta C_{44}^{\text{supercell}}$	3.4 GPa ($\Delta_{\text{DFT}}^{\text{K1 to K2}}$)
		$\delta K^{\text{supercell}}$	3.2 GPa ($\Delta_{\text{DFT}}^{\text{K1 to K2}}$)
Lattice parameters	Vacancy	δa^{vac}	0.05 Å ($\Delta_{\text{DFT}}^{\text{K1 to K4}}$)
	Split-<110> interstitial	$\delta a^{\text{split-<110>}}$	0.04 Å ($\Delta_{\text{DFT}}^{\text{K1 to K4}}$)
	Tetrahedral interstitial	$\delta a^{\text{tetrahedral}}$	0.02 Å ($\Delta_{\text{DFT}}^{\text{K1 to K4}}$)
	Hexagonal interstitial	$\delta a^{\text{hexagonal}}$	0.03 Å ($\Delta_{\text{DFT}}^{\text{K1 to K4}}$)
	Unit cell with 8 atoms	$\delta a^{\text{unit cell}}$	0.16 Å ($\Delta_{\text{DFT}}^{\text{K1 to K4}}$)
	Supercell with 64 atoms	$\delta a^{\text{supercell}}$	0.03 Å ($\Delta_{\text{DFT}}^{\text{K1 to K4}}$)
Rate of ranking errors, $p_{\text{ranking error}}^1$			7.6% ($\Delta_{\text{DFT}}^{\text{K1 to K4}}$)

Energy rankings	Ranking dataset 1: 2-vacancies in supercells with 214 atoms	Mean ΔE_{DFT} of ranking errors, Mean ΔE_{DFT}^1	16 meV ($\Delta_{\text{DFT}}^{\text{K1 to K4}}$)
		Max ΔE_{DFT} of ranking errors, Maximum ΔE_{DFT}^1	54 meV ($\Delta_{\text{DFT}}^{\text{K1 to K4}}$)
		$P^2_{\text{ranking error}}$	7.4% ($\Delta_{\text{DFT}}^{\text{K1 to K4}}$)
	Ranking dataset 2: 3-vacancies in supercells with 141 atoms	Mean ΔE_{DFT}^2	26 meV ($\Delta_{\text{DFT}}^{\text{K1 to K4}}$)
		Maximum ΔE_{DFT}^2	67 meV ($\Delta_{\text{DFT}}^{\text{K1 to K4}}$)
		$P^3_{\text{ranking error}}$	9.9% ($\Delta_{\text{DFT}}^{\text{K1 to K4}}$)
	Ranking dataset 3: 3-vacancies in supercells with 93 atoms	Mean ΔE_{DFT}^3	78 meV ($\Delta_{\text{DFT}}^{\text{K1 to K4}}$)
		Maximum ΔE_{DFT}^3	263 meV ($\Delta_{\text{DFT}}^{\text{K1 to K4}}$)
		$P^4_{\text{ranking error}}$	5.5% ($\Delta_{\text{DFT}}^{\text{K1 to K4}}$)
	Ranking dataset 4: 3-vacancies in supercells with 63 atoms	Mean ΔE_{DFT}^4	69 meV ($\Delta_{\text{DFT}}^{\text{K1 to K4}}$)
		Maximum ΔE_{DFT}^4	240 meV ($\Delta_{\text{DFT}}^{\text{K1 to K4}}$)
Forces on RE atoms, $NAC(\delta, \mathcal{D}^{\text{RE}})$	$\mathcal{D}^{\text{RE-I}}$	Directional error of forces, $\Delta NAC(\delta_{\theta}, \mathcal{D}^{\text{RE-I}})$	0.22 ($\Delta_{\text{DFT}}^{\text{K1 to K4}}$)
		Magnitude error of forces, $\Delta NAC(\delta_F , \mathcal{D}^{\text{RE-I}})$	0.20 ($\Delta_{\text{DFT}}^{\text{K1 to K4}}$)
	$\mathcal{D}^{\text{RE-V}}$	$\Delta NAC(\delta_{\theta}, \mathcal{D}^{\text{RE-V}})$	0.15 ($\Delta_{\text{DFT}}^{\text{K1 to K4}}$)
		$\Delta NAC(\delta_F , \mathcal{D}^{\text{RE-V}})$	0.13 ($\Delta_{\text{DFT}}^{\text{K1 to K4}}$)
δ^{RE}	$\mathcal{D}^{\text{RE-I}}$	$\sigma_F^{\text{RE-I}}$	15 meV/atom (handpicked)
		$\sigma_F^{\text{RE-I}}$	0.25 eV/Å ($\Delta_{\text{DFT}}^{\text{K1 to K4}}$)
	$\mathcal{D}^{\text{RE-V}}$	$\sigma_F^{\text{RE-V}}$	15 meV/atom (handpicked)
		$\sigma_F^{\text{RE-V}}$	0.17 eV/Å ($\Delta_{\text{DFT}}^{\text{K1 to K4}}$)
δ^{enhanced}	$\mathcal{D}^{\text{enhanced-I}}$	$\sigma_F^{\text{enhanced-I}}$	15 meV/atom (handpicked)
		$\sigma_F^{\text{enhanced-I}}$	0.12 eV/Å (handpicked)
	$\mathcal{D}^{\text{enhanced-V}}$	$\sigma_F^{\text{enhanced-V}}$	15 meV/atom (handpicked)
		$\sigma_F^{\text{enhanced-V}}$	0.12 eV/Å (handpicked)
Thermal properties	Bulk: unit cell with 8 atoms	Free energy, $\delta E_{\text{free}}^{\text{bulk}}(T)$	0.25 ($\Delta_{\text{DFT}}^{\text{K1 to K4}}$)
		Entropy, $\delta S^{\text{bulk}}(T)$	0.33 ($\Delta_{\text{DFT}}^{\text{K1 to K4}}$)
		Heat capacity, $\delta c^{\text{bulk}}(T)$	0.12 ($\Delta_{\text{DFT}}^{\text{K1 to K4}}$)
	Vacancy	$\delta E_{\text{free}}^{\text{vacancy}}(T)$	39 ($\Delta_{\text{DFT}}^{\text{K1 to K4}}$)
		$\delta S^{\text{vacancy}}(T)$	92 ($\Delta_{\text{DFT}}^{\text{K1 to K4}}$)
		$\delta c^{\text{vacancy}}(T)$	41 ($\Delta_{\text{DFT}}^{\text{K1 to K4}}$)

5.3 Finding challenging properties for MLPs to predict

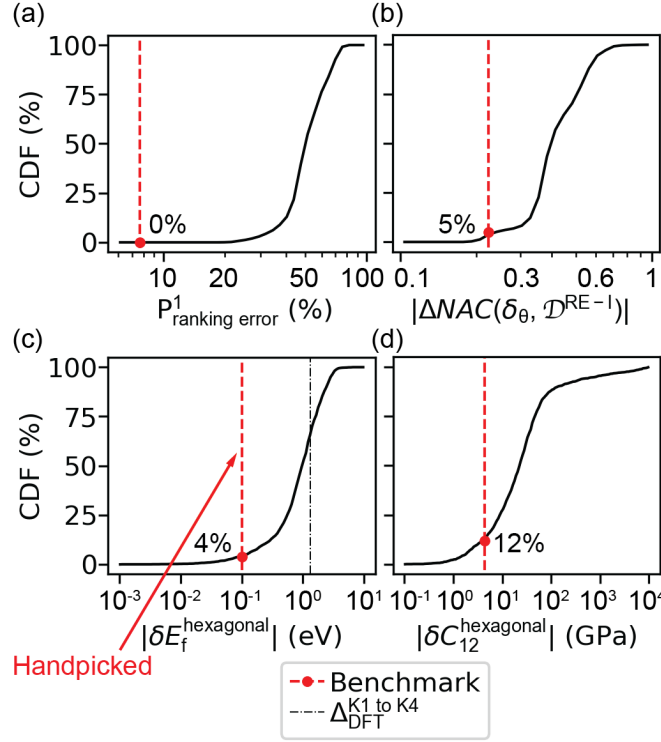


Figure 5.3. Distribution of the error metrics for identified ‘challenging’ properties. The cumulative distribution function (CDF) curves of (a) the error rates of ranking errors in ranking dataset of two vacancies in supercells with 214 atoms (Table 5.1), (b) the NAC of errors on force directions in $\mathcal{D}^{RE-I}$, $|\Delta NAC(\delta_\theta, \mathcal{D}^{RE-I})|$, (c) the error of the formation energy of hexagonal interstitial, $|\delta E_f^{\text{hexagonal}}|$, and (d) the error of the elastic constant C_{12} for the supercell with hexagonal interstitial, $|\delta C_{12}^{\text{hexagonal}}|$. The red dashed lines in (a-d) are benchmarks for each metric and the fractions below benchmarks are shown beside the benchmark (red dot). For handpicked benchmark, black dashed line indicates the $\Delta_{\text{DFT}}^{K1 \text{ to } K4}$ value for the metric.

Among the large array of properties for the large pool of MLPs, we first identify the properties that are difficult to be calculated accurately by the MLPs (step 3 from the process, Figure 5.1). If less than 15% of MLPs can achieve the errors lower than the benchmark, we call it a ‘challenging’ property to be calculated by MLP. Our results show there are 36 challenging properties (Appendix G), mostly fall into four groups (except for the entropy of $\text{Si } S^{\text{bulk}}(T)$), listed as follows: 1) the defect formation energies E_f^{Defect} (only 2.8% – 5% of MLP predictions can

achieve the proposed 0.1 eV benchmark), 2) 14 metrics in the elastic constants (as only 1% – 12% are lower than the proposed $\Delta_{\text{DFT}}^{\text{K1 to K2}}$ benchmarks), 3) all 12 metrics for energy rankings (less than four MLPs, or 0.2%, for each of the metrics can achieve the $\Delta_{\text{DFT}}^{\text{K1 to K4}}$ benchmarks), and 4) five metrics for force predictions on RE atoms (**Figure 5.3a, b, c, d, and Table G2**). The energy rankings are particularly challenging for MLPs as shown by the much higher error than the $\Delta_{\text{DFT}}^{\text{K1 to K4}}$ (Figure 5.3a) and almost no MLP can achieve the same level of prediction errors as the proposed $\Delta_{\text{DFT}}^{\text{K1 to K4}}$ benchmark for any of the property in this category. For example, in the error rate in ranking dataset 1, $P^1_{\text{ranking error, top 10\% MLPs}}$ (except for two outliers) are between 18% – 38%, compared to the benchmark 7.6% for $\Delta_{\text{DFT}}^{\text{K1 to K4}}$ (Figure 5.3a). The error rates in ranking in Si materials system in this chapter is much higher than those in Li-Al materials system in Chapter 4, in which $\text{MLP}_{\text{Train}}$ on known intermediate phases are all below 10% (Figure 4.2f). To sum up, we analyze the MLP performances on the errors of predicting a single property and identify that energy ranking, E_f of defects, elastic constants, and force predictions on RE atoms are challenging for MLPs to predict.

5.4 Finding pairs of properties that are challenging

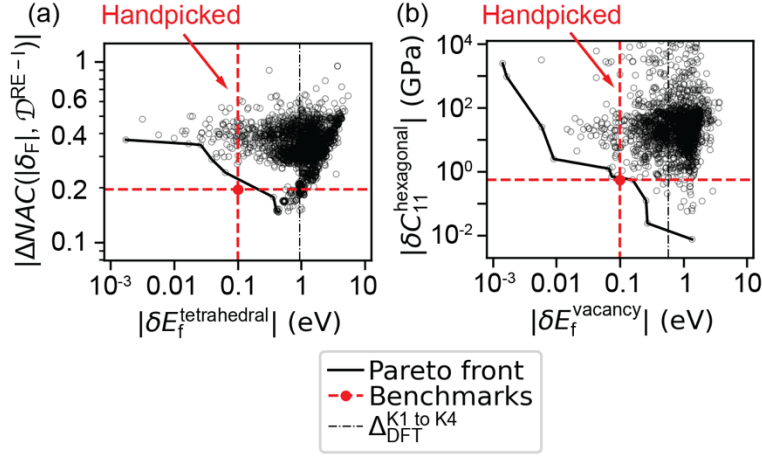


Figure 5.4. Examples of error metric distributions for challenging property pairs. The scattering plots of two properties for each MLP: (a) $|\Delta NAC(|\delta_F|, \mathcal{D}^{RE-1})|$ versus $|\delta E_f^{\text{tetrahedral}}|$, and (b) $|\delta C_{11}^{\text{hexagonal}}|$ versus $|\delta E_f^{\text{vacancy}}|$. The black solid lines are Pareto fronts.

We further analyze the performances of the MLPs for accurately predicting two properties and to identify certain pairs of properties that are challenging to simultaneously achieve the proposed benchmarks (step 4 from process, Figure 5.1). Given that all 12 metrics for energy rankings exhibit high prediction errors and are regarded as challenging, we exclude them from the analyses below (Chapter 5.4 and 5.5), and only examine the pairs of properties from the remaining 52 properties. As shown in the Pareto fronts for each pair of properties, there are clear trade-offs for the optimal MLPs (which are the vertices on Pareto fronts), i.e., decreasing prediction errors on one property will increase the errors on the other, as shown (Figure 5.4a, b, and Figure G2, G3, and G4, and Appendix G). Due to this trade-off, training an MLP that exceeds the benchmark (e.g., DFT K1) on both properties are challenging. From our analyses of all property pairs, there are 20 pairs of properties that no optimal MLP can simultaneously meet or exceed the benchmarks in both properties (Table G3). All properties in these 20 pairs are from the following five categories of properties, such as defect formation energy E_f^{Defect} , elastic constants, force performance score

on RE atoms $NAC(\delta, \mathcal{D}^{\text{RE}})$, RMSE force errors on RE atoms σ_F^{RE} , and $\sigma_F^{\text{enhanced}}$, which are also among the challenging properties. These 20 pairs are seven pairs in ' E_f^{Defect} – elastic constants', six pairs in ' E_f^{Defect} – $NAC(\delta, \mathcal{D}^{\text{RE}})$ ', four pairs in ' E_f^{Defect} – σ_F^{RE} ', and three pairs in ' E_f^{Defect} – $\sigma_F^{\text{enhanced}}$ ', (Table G3, details of Pareto fronts are further discussed Appendix G). Given predicting one of these properties can be challenging, it is expected that no MLP can meet both benchmarks for two properties in two distinct categories. This highlights the difficulty of having MLPs that can accurately predict multiple properties from challenging categories.

5.5 Conclusion

In this chapter, we conduct a systematic analysis on performance for a wide range of MLPs and evaluation metrics, through our analysis process (Figure 5.1). Following our previous work in Chapter 3 on the MLP performances in Si materials system, we construct our performance dataset by sampling a large number of MLPs from a pool of trained MLPs, and by generating many physical properties and evaluation metrics. After setting benchmarks, an empirical analysis identifies challenging properties/metrics on MLP and investigate the correlations between these metrics. We identify 36 challenging properties, a majority of which belonging to four categories: the energy rankings, the defect formation energies, the elastic constants, and the force performances on migrating atoms. While the defect formation energies and the force performances on migrating atoms are discussed in Chapter 3, the energy ranking metrics, which are newly developed in Chapter 4, are difficult to predict for MLPs. Additionally, we investigate pairs of properties that are challenging to predict spontaneously. By excluding the energy ranking metrics that are specifically difficult, we identify 20 challenging pairs, for which no optimal MLP can achieve both proposed benchmarks. Among these 20 pairs, a majority of them are defect formation

energies combined with elastic constants, force performance NACs on migrating atoms, or force RMSEs on migrating atoms. We emphasize that the combination of properties from distinct categories makes property pairs difficult to predict. In summary, our proposed process is effective to sample a large number of MLP performance data, to identify challenging properties and their combinations, providing valuable guidance to improve MLP performances in future studies.

Chapter 6: Conclusions and future work

6.1 Conclusions

The key conclusions of this dissertation are summarized as follows:

- 1) We systematically study the Li-ion conduction of known Li-containing chlorides and a wide range of doped compounds. We identify 19 potential Li superionic conductors in the materials system and reveal that the Li-ion conduction is greatly affected by Li content and cation sublattice. By analyzing a number of materials, we summarize that low Li content, low cation concentration, and sparse cation distribution are mainly responsible for high Li-ion conduction in the Li-containing chlorides system. Our work provides guidance for developing future halide Li-ion conductors.
- 2) We examine the MLP performance on reproducing atomistic dynamics in Si system. By performing MD simulations using several state-of-the-art MLPs, we reveal major discrepancies between MLP predictions and DFT calculations, such as properties related to rare events, defects with similar energies, and atomic vibrations. By investigating the MLP predictions on RE atoms, we develop novel evaluation metrics to quantify the errors of forces predicted by MLPs and improve the MLPs by integrating the newly developed evaluation metrics into the training process. Our summarized process of developing evaluation metrics provides guidance to improve MLP performance for future work.
- 3) We study the performance and transferability of MLPs for elemental orderings on a wide range of phases and lattice structures in- and out- of the training data in Li-Al alloy materials system. By using only three materials in Li-Al binary alloy system, the trained MLP can well predict energies, energy rankings, ground-state phases, and lowest energy configurations over all known phases and many hypothetical phases across the phase diagram. Our developed

evaluation metrics for the energy ranking of elemental orderings can be further used for the testing and development of MLPs. We further showed that the MLP performance may not be improved by solely increasing the diversity of training data without significantly increasing the data size.

- 4) We thoroughly assess MLP performances in the Si system using thousands of MLPs and tens of calculated properties, testing errors, evaluation metrics, and performance scores. By setting benchmarks on each of the tested properties or performance metrics, we identify 36 properties and 20 pairs of properties that are challenging to predict. Challenging properties mainly belong to four categories such as energy rankings, defect formation energies, elastic constants, and force predictions on migrating atoms.

6.2 Future work

This dissertation demonstrates that systematically assessing MLP performance on atomistic dynamics is critical to enable wide applications of MLP techniques in atomistic modeling studies. According to our results on the MLP errors, several future directions are listed as follows:

- 1) The further development of MLPs to address the errors and discrepancies observed require further study on the effect of descriptors, ML models, and training data selection. For example, the performances on energy rankings of MLPs are rarely studied previously. Efforts such as modifying the training data to improve MLP with lower ranking errors will be valuable. The use of our evaluation metrics will greatly aid this process. Investigating how MLPs with high/low scores on different evaluation metrics may perform in actual atomistic simulations may be another interesting research topic.
- 2) Examining more phenomena related to atom dynamics and develop corresponding evaluation metrics will expand the understanding of MLPs' performance and their limits. A diverse set of assessments on properties and atom dynamics will likely uncover many more discrepancies in MLP applications, which can be further employed as demonstrated in our study to improve MLPs.
- 3) Since surfaces and interfaces are of great interest for industrial applications and are widely studied in atomistic modeling. A systematic testing of MLP performances on surfaces and interfaces will be valuable for studying surface/interface phenomena and related applications.

Publications, Patents, and Presentations

Publications

- 1) **Liu, Y.**, Wang, S., Nolan, A. M., Ling, C., & Mo, Y. (2020). Tailoring the cation lattice for chloride lithium-ion conductors. *Advanced Energy Materials*, 10(40), 2002356.
- 2) **Liu, Y.**, He, X., & Mo, Y. (2023). Discrepancies and the Error Evaluation Metrics for Machine Learning Interatomic Potentials. *arXiv preprint arXiv:2306.11639*.
- 3) Yang, M., **Liu, Y.**, & Mo, Y. (2023). Lithium crystallization at solid interfaces. *Nature Communications*, 14(1), 2986.
- 4) Wang, S., **Liu, Y.**, & Mo, Y. (2023). Frustration in Super-Ionic Conductors Unraveled by the Density of Atomistic States. *Angewandte Chemie*, 135(15), e202215544.
- 5) Kwak, H., Wang, S., Park, J., **Liu, Y.**, Kim, K. T., Choi, Y., ... & Jung, Y. S. (2022). Emerging halide superionic conductors for all-solid-state batteries: Design, synthesis, and practical applications. *ACS Energy Letters*, 7(5), 1776-1805.
- 6) Yang, M., **Liu, Y.**, Nolan, A. M., & Mo, Y. (2021). Interfacial atomistic mechanisms of lithium metal stripping and plating in solid-state batteries. *Advanced Materials*, 33(11), 2008081.
- 7) Wang, R., Ping, W., Wang, C., **Liu, Y.**, Gao, J., Dong, Q., ... & Hu, L. (2020). Computation-Guided Synthesis of New Garnet-Type Solid-State Electrolytes via an Ultrafast Sintering Technique. *Advanced Materials*, 32(46), 2005059.
- 8) He, X., Bai, Q., **Liu, Y.**, Nolan, A. M., Ling, C., & Mo, Y. (2019). Crystal structural framework of lithium super-ionic conductors. *Advanced Energy Materials*, 9(43), 1902078.

- 9) Nolan, A. M., **Liu, Y.**, & Mo, Y. (2019). Solid-state chemistries stable with high-energy cathodes for lithium-ion batteries. *ACS Energy Letters*, 4(10), 2444-2451.
- 10) Wang, S., Bai, Q., Nolan, A. M., **Liu, Y.**, Gong, S., Sun, Q., & Mo, Y. (2019). Lithium chlorides and bromides as promising solid-state chemistries for fast ion conductors with good electrochemical stability. *Angewandte Chemie International Edition*, 58(24), 8039-8043.

Patents

- 1) Mo, Y., **Liu, Y.**, & Ling, C. (2021). *U.S. Patent Application No. 16/909,645*.
- 2) Mo, Y., **Liu, Y.**, & Ling, C. (2023). *U.S. Patent No. 11,641,029*. Washington, DC: U.S. Patent and Trademark Office.
- 3) Mo, Y., **Liu, Y.**, Xingfeng, H. E., & Ling, C. (2022). *U.S. Patent No. 11,462,766*. Washington, DC: U.S. Patent and Trademark Office.

Presentation

Liu, Y., & Mo, Y. (2023). Exploring the discrepancies, the extrapolability, and the interpolability of machine learning interatomic potentials. *Bulletin of the American Physical Society*.

Liu, Y., & Mo, Y. (2023). Exploring the discrepancies, the extrapolability, and the interpolability of machine learning interatomic potentials. *MRS Spring meeting*, San Fransico, 2023, April 14th.

Appendices

Appendix A

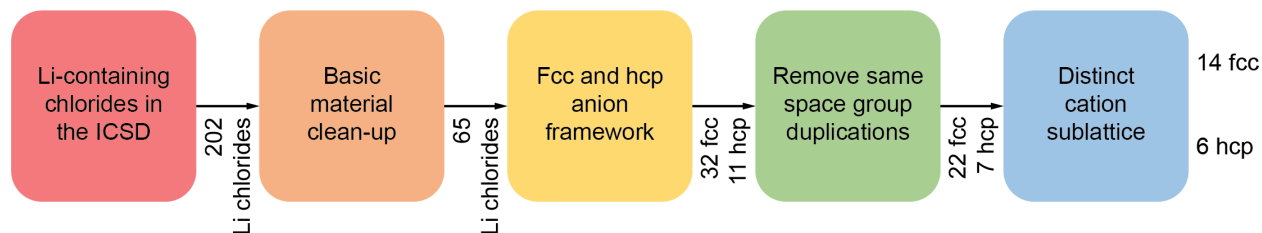


Figure A1. Steps and Criteria of selecting candidate chloride systems with face-centered cubic (fcc) and hexagonal close packed (hcp) anion sublattice from the Inorganic Crystal Structure Database (ICSD).

Table A1. Selected candidate Li-containing chloride systems.

Composition	E_{hull} (meV/atom)	Anion framework	Space Group	ICSD IDs
Li_6VCl_8	13	fcc	$Fm\bar{3}m$	100834
Li_2ZnCl_4	17	fcc	$Fd\bar{3}m$	202743
Li_5CrCl_8	11	fcc	$Cmmm$	67742
Li_2CrCl_4	38	fcc	$C2/c$	202627
Li_2CoCl_4	20	fcc	$Cmmm$	73227
Li_2MgCl_4	5	fcc	$Fd\bar{3}m$	74957
Li_2CdCl_4	10	fcc	$Fd\bar{3}m$	74958
LiTiCl_3	148	fcc	$Fd\bar{3}m$	56481
Li_2CrCl_4	61	fcc	$Fd\bar{3}m$	67259
Li_2CoCl_4	24	fcc	$Imma$	202402
Li_2CoCl_4	36	fcc	$Fd\bar{3}m$	73229
LiTiCl_3	150	fcc	$Fm\bar{3}m$	62248
Li_3ScCl_6	18	fcc	$C2/m$	50153
Li_3InCl_6	20	fcc	$C2/m$	89617
Li_2ZnCl_4	13	hcp	$Pnma$	73222
LiAlCl_4	0	hcp	$P2_1/c$	1040
LiWCl_6	13	hcp	$R3$	409938
Li_3ErCl_6	0	hcp	$P\bar{3}m1$	50151
Li_3YbCl_6	75	hcp	$Pnma$	50152
LiAlCl_4	0	hcp	$Pmn2_1$	244127

Table A2. AIMD screening of unsubstituted and substituted Li-containing chlorides in Figure 2.2.

ICSD ID	Composition	E_{hull} (meV/atom)	σ at 500K (mS cm ⁻¹)	Error bound [σ_{min} , σ_{max}] (mS cm ⁻¹)	σ at 600K (mS cm ⁻¹)	Error bound [σ_{min} , σ_{max}] (mS cm ⁻¹)
100834	Li ₆ MgCl ₈	15	/	/	Very low	/
	Li _{23/4} Mg _{3/4} Sc _{1/4} Cl ₈	28	/	/	Very low	/
	Li _{11/2} Mg _{1/2} Sc _{1/2} Cl ₈	19	/	/	Very low	/
	Li ₅ ScCl ₈	22	/	/	Very low	/
202743	Li ₂ ZnCl ₄	17	/	/	Very low	/
	Li _{7/4} Zn _{3/4} In _{1/4} Cl ₄	25	/	/	Very low	/
	Li _{3/2} Zn _{1/2} In _{1/2} Cl ₄	25	/	/	83	[43, 124]
	LiInCl ₄	16	184	[115, 253]	Melt	/
67742	Li ₅ InCl ₈	0	/	/	154	[93, 215]
	Li _{19/4} Sc _{3/4} Zr _{1/4} Cl ₈	21	/	/	Very low	/
	Li _{9/2} Sc _{1/2} Zr _{1/2} Cl ₈	24	/	/	119	[66, 172]
	Li ₄ ZrCl ₈	26	53	[28, 79]	224	[167, 281]
	Li _{21/4} Sc _{3/4} Mg _{1/4} Cl ₈	22	/	/	Very low	/
	Li _{11/2} Sc _{1/2} Mg _{1/2} Cl ₈	27	/	/	Very low	/
	Li ₆ MgCl ₈	30	/	/	Very low	/
73227	Li ₂ MgCl ₄	12	/	/	Very low	/
	Li _{7/4} Mg _{3/4} Sc _{1/4} Cl ₄	15	/	/	Very low	/
	Li _{3/2} Mg _{1/2} Sc _{1/2} Cl ₄	19	/	/	Very low	/
	LiScCl ₄	34	/	/	Very low	/
73222	Li ₂ ZnCl ₄	13	/	/	Very low	/
	Li _{7/4} Zn _{3/4} Al _{1/4} Cl ₄	20	/	/	Very low	/
	Li _{3/2} Zn _{1/2} Al _{1/2} Cl ₄	21	/	/	88	[31, 147]
	LiAlCl ₄	17	269	[201, 338]	601	[448, 753]
202627	Li ₂ MgCl ₄	7	/	/	Very low	/
	Li _{7/4} Mg _{3/4} In _{1/4} Cl ₄	8	/	/	86	[53, 120]
	Li _{3/2} Mg _{1/2} In _{1/2} Cl ₄	6	/	/	Very low	/
	LiInCl ₄	0	/	/	Very low	/
202402	Li ₂ MgCl ₄	7	/	/	32	[18, 47]
	Li _{7/4} Mg _{3/4} In _{1/4} Cl ₄	3	/	/	Very low	/
	Li _{3/2} Mg _{1/2} In _{1/2} Cl ₄	4	/	/	121	[70, 173]
	LiInCl ₄	0	/	/	128	[44, 212]
73229	Li ₂ MgCl ₄	26	/	/	Very low	/
	Li _{7/4} Mg _{3/4} Sc _{1/4} Cl ₄	28	/	/	Very low	/
	Li _{3/2} Mg _{1/2} In _{1/2} Cl ₄	8	/	/	140	[91, 190]
	LiInCl ₄	0	/	/	73	[30, 116]
50153	Li ₃ ScCl ₆	18	/	/	175	[130, 221]
	Li _{11/4} Sc _{3/4} Zr _{1/4} Cl ₆	15	167	[126, 210]	242	[192, 294]
	Li _{5/2} Sc _{1/2} Zr _{1/2} Cl ₆	14	179	[133, 227]	508	[409, 608]
	Li ₂ ZrCl ₆	8	122	[92, 154]	576	[431, 722]
	Li _{13/4} Sc _{3/4} Mg _{1/4} Cl ₆	22	/	/	127	[60, 194]
	Li _{7/2} Sc _{1/2} Mg _{1/2} Cl ₆	26	/	/	129	[44, 215]
	Li ₄ MgCl ₆	30	/	/	83	[41, 125]
89617	Li ₃ InCl ₆	20	120	[88, 153]	406	[304, 509]
	Li _{11/4} In _{3/4} Zr _{1/4} Cl ₆	19	152	[113, 191]	350	[262, 440]
	Li _{5/2} In _{1/2} Zr _{1/2} Cl ₆	30	172	[129, 216]	270	[219, 321]
	Li _{13/4} In _{3/4} Mg _{1/4} Cl ₆	17	97	[71, 124]	181	[145, 218]
	Li _{7/2} In _{1/2} Mg _{1/2} Cl ₆	30	71	[48, 95]	178	[133, 223]
1040	LiAlCl ₄	0	235	[114, 356]	Melt	/
	Li _{5/4} Al _{3/4} Zn _{1/4} Cl ₄	11	226	[168, 285]	Melt	/
	Li _{3/2} Al _{1/2} Zn _{1/2} Cl ₄	26	259	[194, 324]	Melt	/
	Li ₂ ZnCl ₄	19	135	[105, 174]	420	[315, 525]
409938	Li ₂ ZrCl ₆	13	70	[45, 95]	Melt	/
	Li ₃ InCl ₆	19	543	[416, 670]	785	[611, 961]
244127	LiAlCl ₄	0	/	/	Very low	/
	Li _{5/4} Al _{3/4} Zn _{1/4} Cl ₄	6	198	[149, 249]	581	[434, 728]
	Li _{3/2} Al _{1/2} Zn _{1/2} Cl ₄	13	280	[210, 350]	Melt	/
	Li ₂ ZnCl ₄	12	39	[8, 72]	315	[236, 394]
74958	Li ₂ CdCl ₄	10	/	/	92	[43, 141]
	Li _{7/4} Cd _{3/4} Y _{1/4} Cl ₄	20	/	/	151	[93, 211]
	Li _{3/2} Cd _{1/2} Y _{1/2} Cl ₄	29	/	/	160	[104, 217]
	LiInCl ₄	0	/	/	70	[20, 121]

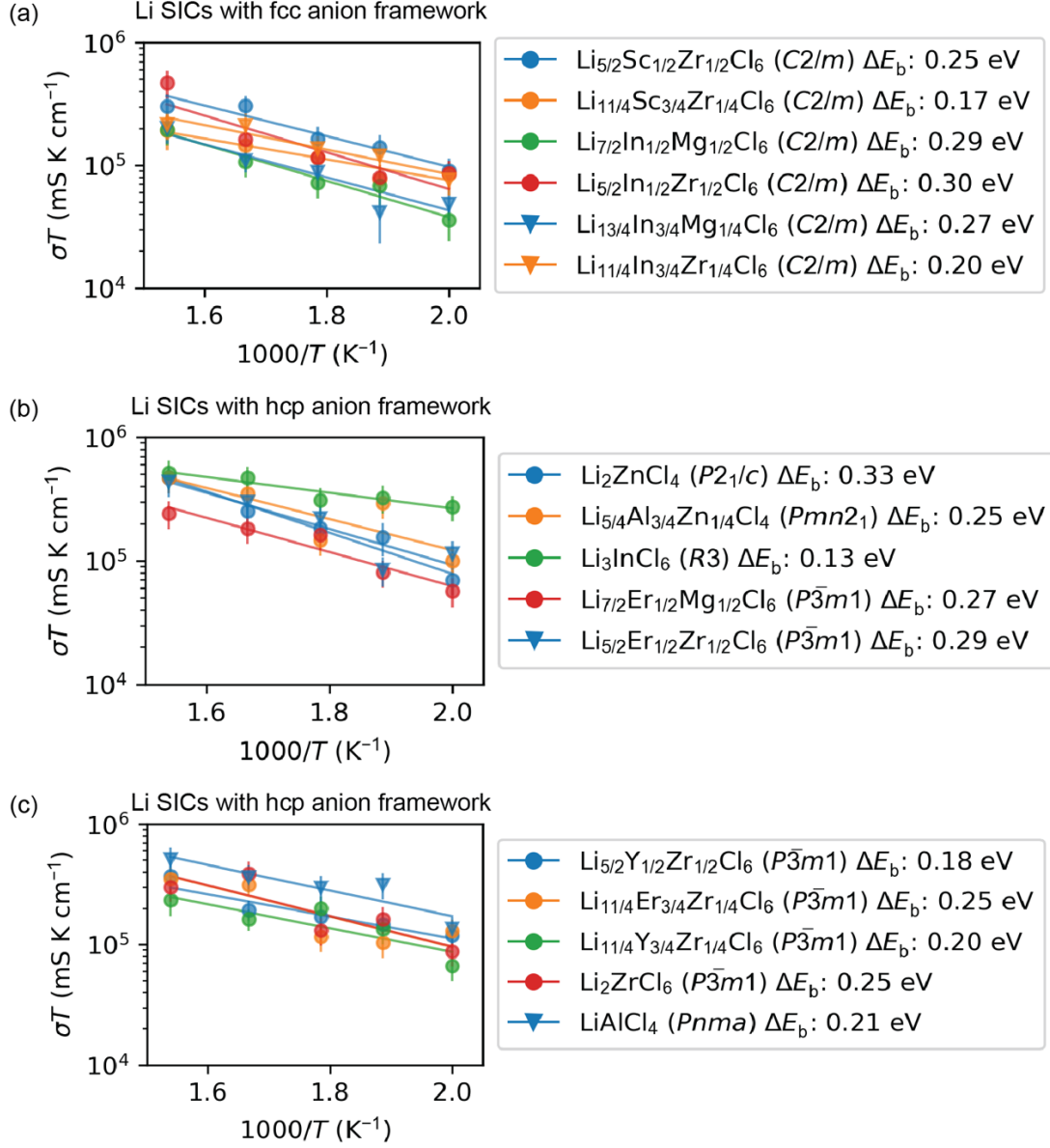


Figure A2. Arrhenius plots of potential lithium superionic conductors (SICs) predicted by *ab initio* molecular dynamics (AIMD) simulations.

Table A3. Energies above the hull (E_{hull}), Li-ion conductivities at 300 K (σ) with error bounds, activation energies (E_a), and the band gaps (E_g) of potential Li SICs predicted by AIMD simulations.

ICSD ID of original compound	Composition	E_{hull} (meV/atom)	σ at 300K (mS cm ⁻¹)	Error bound [σ_{min} , σ_{max}] (mS cm ⁻¹)	E_a (eV)	E_g (eV)
89617	Li ₃ InCl ₆	20	1.2	[0.1, 11]	0.29 ± 0.05	2.5
89617	Li _{7/2} In _{1/2} Mg _{1/2} Cl ₆	30	2	[0.1, 44]	0.29 ± 0.07	4.0
1040	Li ₂ ZnCl ₄	19	3	[0.2, 38]	0.33 ± 0.06	5.0
50151	Li ₃ ErCl ₆	0	3	[0.1, 118]	0.28 ± 0.02	6.1
89617	Li _{7/2} In _{1/2} Zr _{1/2} Cl ₆	30	4	[0.3, 48]	0.30 ± 0.06	3.4
89617	Li _{13/4} In _{3/4} Mg _{1/4} Cl ₆	17	4	[0.2, 62]	0.27 ± 0.06	4.0
50151	Li _{7/2} Er _{1/2} Mg _{1/2} Cl ₆	18	5	[0.3, 71]	0.27 ± 0.06	5.9
50151	Li _{5/2} Er _{1/2} Zr _{1/2} Cl ₆	8	6	[0.4, 81]	0.29 ± 0.06	4.3
50151	Li _{8/3} Er _{2/3} Zr _{1/3} Cl ₆	5	11	[0.8, 133]	0.25 ± 0.06	4.1
50151	Li ₂ ZrCl ₆	13	11	[0.8, 149]	0.25 ± 0.06	4.1
50153	Li _{5/2} Sc _{1/2} Zr _{1/2} Cl ₆	14	12	[0.9, 158]	0.25 ± 0.06	4.1
50151	Li ₃ YCl ₆	22	14	[5, 47]	0.28 ± 0.02	6.0
244127	Li _{5/4} Al _{3/4} Zn _{1/4} Cl ₆	6	14	[1, 193]	0.25 ± 0.06	6.0
89617	Li _{11/4} In _{3/4} Zr _{1/4} Cl ₆	19	22	[2, 316]	0.20 ± 0.06	3.6
50151	Li _{8/3} Y _{2/3} Zr _{1/3} Cl ₆	25	23	[2, 308]	0.20 ± 0.06	4.3
50153	Li _{11/4} Sc _{3/4} Zr _{1/4} Cl ₆	15	31	[2, 525]	0.17 ± 0.06	4.3
73222	LiAlCl ₄	17	37	[3, 527]	0.21 ± 0.06	5.8
50151	Li _{5/2} Y _{1/2} Zr _{1/2} Cl ₆	24	37	[3, 423]	0.18 ± 0.06	4.1
409938	Li ₃ InCl ₆	19	209	[16, 2724]	0.13 ± 0.06	4.0

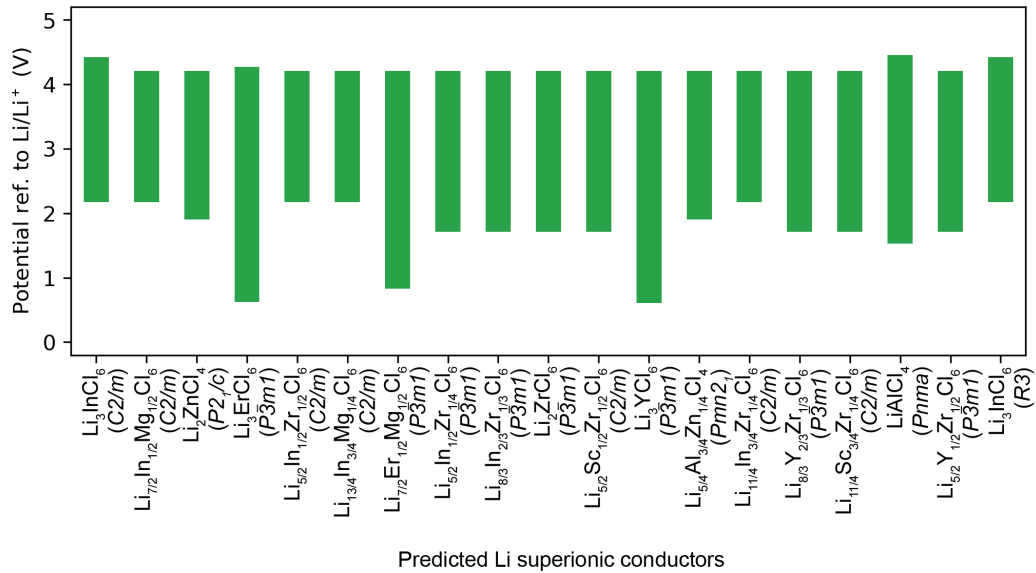


Figure A3. Thermodynamic intrinsic electrochemical windows of potential Li SICs evaluated using first principles computation scheme in Ref. [48].

Table A4. Thermodynamic intrinsic electrochemical window and phase equilibria of predicted Li SICs evaluated using first principles computation scheme in Ref. [48].

ICSD ID of original compound	Composition	Reduction onset /V ref. to Li/Li ⁺	Phase equilibria at reduction onset	Oxidation onset /V ref. to Li/Li ⁺	Phase equilibria at oxidation onset
89617	Li ₃ InCl ₆	2.17	LiCl, InCl	4.42	Cl ₂ , InCl ₃
89617	Li _{7/2} In _{1/2} Mg _{1/2} Cl ₆	2.17	LiCl, InCl, Li ₂ MgCl ₄	4.21	Cl ₂ , Li ₃ InCl ₆ , Li ₂ MgCl ₄
1040	Li ₂ ZnCl ₄	1.90	LiCl, Zn	4.21	Cl ₂ , ZnCl ₂
50151	Li ₃ ErCl ₆	0.62	LiCl, Er	4.27	Cl ₂ , ErCl ₃
89617	Li _{7/2} In _{1/2} Zr _{1/2} Cl ₆	2.17	LiCl, InCl, ZrCl ₄	4.21	Cl ₂ , Li ₃ InCl ₆ , ZrCl ₄
89617	Li _{13/4} In _{3/4} Mg _{1/4} Cl ₆	2.17	LiCl, InCl, Li ₂ MgCl ₄	4.21	Cl ₂ , Li ₃ InCl ₆ , Li ₂ MgCl ₄
50151	Li _{7/2} Er _{1/2} Mg _{1/2} Cl ₆	0.82	LiCl, Li ₃ ErCl ₆ , Mg	4.21	Cl ₂ , Li ₃ ErCl ₆ , Li ₂ MgCl ₄
50151	Li _{5/2} Er _{1/2} Zr _{1/2} Cl ₆	1.71	LiCl, Li ₃ ErCl ₆ , ZrCl ₃	4.21	Cl ₂ , Li ₃ ErCl ₆ , ZrCl ₄
50151	Li _{8/3} Er _{2/3} Zr _{1/3} Cl ₆	1.71	LiCl, Li ₃ ErCl ₆ , ZrCl ₃	4.21	Cl ₂ , Li ₃ ErCl ₆ , ZrCl ₄
50151	Li ₂ ZrCl ₆	1.71	LiCl, ZrCl ₃	4.21	Cl ₂ , ZrCl ₄
50153	Li _{5/2} Sc _{1/2} Zr _{1/2} Cl ₆	1.71	LiCl, ScCl ₃ , ZrCl ₃	4.21	Cl ₂ , ScCl ₃ , ZrCl ₄
50151	Li ₃ YCl ₆	0.61	LiCl, Y ₂ Cl ₃	4.21	Cl ₂ , YCl ₃
244127	Li _{5/4} Al _{3/4} Zn _{1/4} Cl ₆	1.90	LiCl, LiAlCl ₄ , Zn	4.21	Cl ₂ , LiAlCl ₄ , ZnCl ₂
89617	Li _{11/4} In _{3/4} Zr _{1/4} Cl ₆	2.17	LiCl, InCl, ZrCl ₄	4.21	Cl ₂ , Li ₃ InCl ₆ , ZrCl ₄
50151	Li _{8/3} Y _{2/3} Zr _{1/3} Cl ₆	1.71	LiCl, YCl ₃ , ZrCl ₃	4.21	Cl ₂ , YCl ₃ , ZrCl ₄
50153	Li _{11/4} Sc _{3/4} Zr _{1/4} Cl ₆	1.71	LiCl, ScCl ₃ , ZrCl ₃	4.21	Cl ₂ , ScCl ₃ , ZrCl ₄
73222	LiAlCl ₄	1.53	LiCl, Al	4.45	Cl ₂ , AlCl ₃
50151	Li _{5/2} Y _{1/2} Zr _{1/2} Cl ₆	1.71	LiCl, YCl ₃ , ZrCl ₃	4.21	Cl ₂ , YCl ₃ , ZrCl ₄
409938	Li ₃ InCl ₆	2.17	LiCl, InCl	4.42	Cl ₂ , InCl ₃

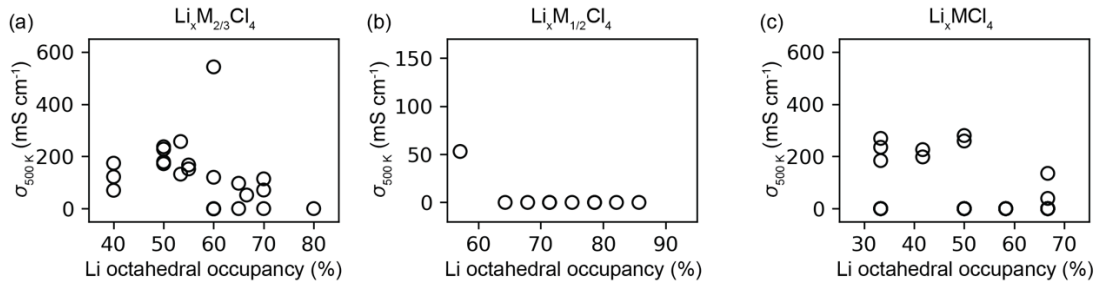


Figure A4. Li-ion conductivities of original and substituted Li-containing chlorides at 500 K versus Li octahedral occupancy for a) $\text{Li}_x\text{M}_{2/3}\text{Cl}_4$, b) $\text{Li}_x\text{M}_{1/2}\text{Cl}_4$ and c) Li_xMCl_4 .

Table A5. Summary of nudged elastic band (NEB) calculations for Li-ion migration in different materials (Figure 2.3).

ICSD ID	Composition	Substitution level (%)	Percolated barrier (eV)	All barriers (eV)
100834	Li_6MgCl_8	0	0.21	0.21
	$\text{Li}_{23/4}\text{Mg}_{3/4}\text{Sc}_{1/4}\text{Cl}_8$	25	0.30	0.30
	Li_5ScCl_8	100	Depends on Li	0.38, 0.38
73227	Li_2MgCl_4	0	0.17	0.17, 0.38
	$\text{Li}_{7/4}\text{Mg}_{3/4}\text{Sc}_{1/4}\text{Cl}_4$	25	0.30	0.30, 0.22
	$\text{Li}_{3/2}\text{Mg}_{1/2}\text{Sc}_{1/2}\text{Cl}_4$	50	0.42	0.42, 0.36, 0.29
	LiScCl_4	100	Depends on Li	0.24, 0.40, 0.36
74957	Li_2MgCl_4	0	0.25, 0.16	0.25, 0.16
	$\text{Li}_{7/4}\text{Mg}_{3/4}\text{In}_{1/4}\text{Cl}_4$	25	0.38	0.38, 0.38
	$\text{Li}_{3/2}\text{Mg}_{1/2}\text{In}_{1/2}\text{Cl}_4$	50	0.44	0.42, 0.43, 0.44, 0.40, 0.53
50151	Li_3YCl_6	0	0.20	0.20, 0.24, 0.35
	$\text{Li}_{8/3}\text{Y}_{2/3}\text{Zr}_{1/3}\text{Cl}_6$	33	0.18, 0.27, 0.35	0.18, 0.26, 0.17, 0.11, 0.27, 0.27, 0.29, 0.41, 0.35, 0.39, 0.35
	$\text{Li}_{5/2}\text{Y}_{1/2}\text{Zr}_{1/2}\text{Cl}_6$	50	0.18, 0.25	0.18, 0.27, 0.13, 0.39, 0.25, 0.13, 0.14, 0.38, 0.23, 0.40
50153	Li_3ScCl_6	0	0.24	0.24, 0.24, 0.32
	$\text{Li}_{11/4}\text{Sc}_{3/4}\text{Zr}_{1/4}\text{Cl}_6$	25	0.24, 0.36	0.24, 0.43, 0.36, 0.55, 0.51, 0.35, 0.38, 0.24, 0.21, 0.23, 0.29
	$\text{Li}_{5/2}\text{Sc}_{1/2}\text{Zr}_{1/2}\text{Cl}_6$	50	0.20, 0.26	0.26, 0.20, 0.37, 0.26, 0.32, 0.19, 0.49, 0.40, 0.45, 0.22, 0.15
67742	Li_5ScCl_8	0	0.16	0.16, 0.37, 0.55
	$\text{Li}_{19/4}\text{Sc}_{3/4}\text{Zr}_{1/4}\text{Cl}_8$	25	0.29	0.29
	$\text{Li}_{9/2}\text{Sc}_{1/2}\text{Zr}_{1/2}\text{Cl}_8$	50	0.15	0.15, 0.17, 0.13, 0.26, 0.11, 0.13
	Li_4ZrCl_8	100	Depends on Li	0.14, 0.32

1. Cation blocking effect in Li_2MgCl_4

Using NEB calculations, we found that, in unsubstituted Li_2MgCl_4 (*Cmmm*), some empty octahedral sites have high energy barriers for Li-ions to access (green octahedra, Figure A5a). The energy barrier of Li-ion migration from original octahedral site to these high-energy sites is as high as 0.51 eV (Figure A5b and c). This result suggests the blocking effect of cations for Li-ion migration to certain empty octahedral sites. In addition, the Li^+ probability density from AIMD simulations in substituted materials, such as $\text{Li}_{3/2}\text{Mg}_{1/2}\text{Sc}_{1/2}\text{Cl}_4$ (*Cmmm*), also confirms that these sites were difficult to access during Li-ion diffusion.

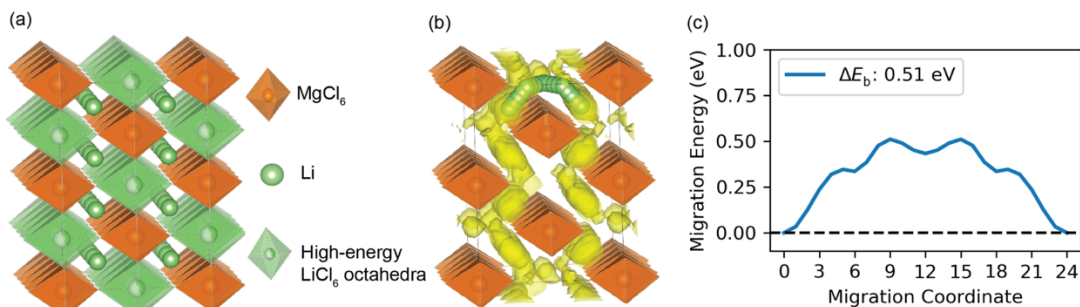


Figure A5. a) Crystal structure of unsubstituted Li_2MgCl_4 ($Cmmm$) with high-energy Li sites (shown in octahedra). b) Li-ion migration pathway (green spheres) through the high-energy sites and Li^+ probability density (yellow) from AIMD simulation. c) The energy landscape of Li-ion migration pathway in b) from NEB calculation.

2. Quantitative analysis of surrounding environments

To generate quantitative features of surrounding cations M and Li-ions for Li-ion diffusion pathways, we obtained radial distribution functions (RDFs) $g(r)$ of other non-mobile Li-ions in migration path surrounding Li-ion centers (Li-Li $g(r)$) and M surrounding Li-ion centers (Li-M $g(r)$), and took the radial distance (r) and coordination number (CN) of the first three peaks of each $g(r)$ (Figure A6a and b). Standard deviation (STD) of r and CN are quantified to describe the distortion of crystal structures, which have multiple small peaks. A total of 54 quantitative features were generated for each migration pathway.

To identify key features for Li-ion migration, we applied Lasso model, which was linear regression model with L1 regularization, on ΔE_f , ΔE_m and ΔE_b of migration pathways. The model parameter a of L1 regularization was generated by using 70% of dataset as training data, 20% as validation data, while leaving 10% of dataset out. The variable coefficients of Lasso models were descriptors of feature importance. We used a bootstrap strategy and performed 200-fold cross-validation process which took 80% of total dataset for each fold, and the average coefficients were obtained.

We further investigated correlations of site energy differences and energy barriers on the features suggested by averaged coefficients of Lasso models. Our analysis found that the 2nd nearest M location, i.e., the 2nd peak position of Li-M distances in $g(r)$ of M with mobile Li-ions at the center, is an important feature, compared to the closest M (1st nearest M). The radial distances of the 2nd peak in Li-M $g(r)$ are either around 5.1 Å or 6.3 Å (Figure 2.3). Migration paths with 2nd nearest M locating at 5.1 Å are more likely to have low site energy differences (Figure A6c and d). Migration pathways with high site energy differences are more likely to have farther 2nd nearest M at 6.3 Å (Figure A6c and d). Different 2nd nearest M locations correspond to distinct cation configurations, where the 2nd Li-M peaks of materials with sparse cation configurations are around 5.1 Å (Figure 2.3j). Other important features from Li-M $g(r)$ include the CN of the 1st nearest M (the corresponding CN of the 1st Li-M peak) on initial sites, i.e., 1st CN_{initial} M, and the change of CN of the 1st nearest M between initial and final sites, i.e., 1st CN_{Δ} M (Figure A6e and f). Detailed analysis and crystal structures on individual migration pathways show that, the high correlations between site energy differences and these 1st nearest M CN features are caused by several migration pathways with very large site energy differences (around 0.5 eV). Features from Li-Li $g(r)$ may have high correlation with either ΔE_f or ΔE_m , but not with both of them, implying no universal Li-Li feature for site energy differences.

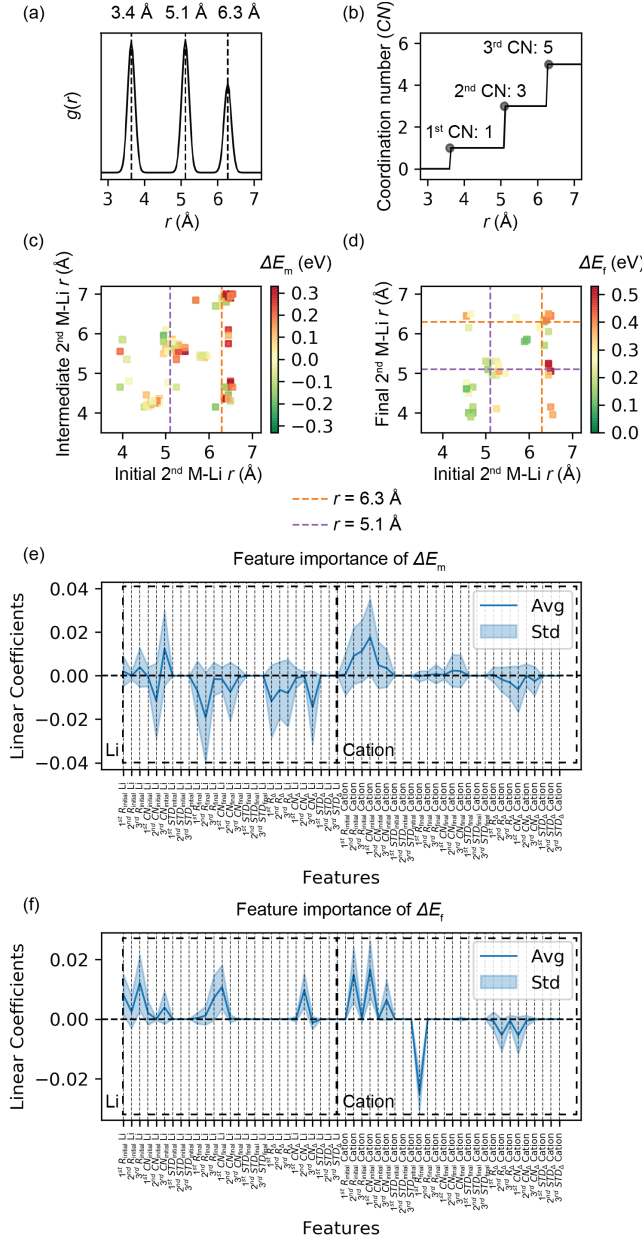


Figure A6. a) The RDF $g(r)$ of octahedron sites in an fcc anion framework (Figure 2.3h). The radial distance r of the first three peaks correspond to three nearest octahedron sites, with distances 3.6 Å, 5.1 Å and 6.3 Å. b) the integrated CN of $g(r)$ in a). c,d) The site energy differences ΔE_f and ΔE_m (colored squares) plotted with r of the 2nd Li-M peaks for initial and c) intermediate, d) final sites. e,f) the variable coefficients generated from Lasso models, showing the correlation of different features with site energy difference e) ΔE_f and f) ΔE_m .

Table A6. Summary of hypothetical Li-containing chloride materials with reduced cation concentrations.

ICSD ID of original compounds	Composition	E_{hull} (meV/atom)	σ at 600K (mS cm ⁻¹)	Error bound [σ_{min} , σ_{max}] (mS cm ⁻¹)
202402	Li _{17/8} Sc _{5/8} Cl ₄	39	420	[315, 525]
	Li _{7/4} Sc _{3/4} Cl ₄	42	252	[188, 316]
202627	Li ₂ Sc _{2/3} Cl ₄	30	148	[88, 209]
	Li _{7/4} Sc _{3/4} Cl ₄	45	71	[31, 111]
	Li ₂ Hf _{1/2} Cl ₄	46	244	[183, 305]
	Li ₂ Zr _{1/2} Cl ₄	41	641	[481, 801]
73227	Li ₂ Sc _{2/3} Cl ₄	21	77	[47, 108]
	Li _{7/4} Sc _{3/4} Cl ₄	37	52	[30, 74]
	Li ₂ Zr _{1/2} Cl ₄	31	231	[173, 289]
74957	Li _{17/8} Sc _{5/8} Cl ₄	44	198	[147, 249]
	Li _{7/4} Sc _{3/4} Cl ₄	37	262	[195, 329]

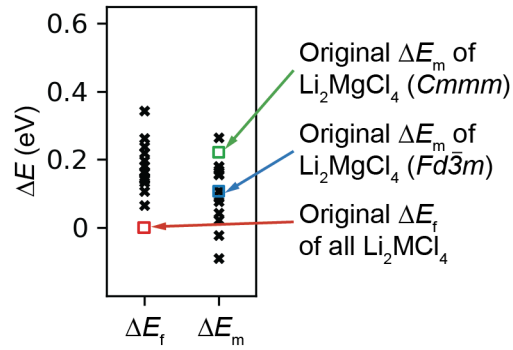


Figure A7. The site energy difference ΔE_f and ΔE_m in substituted (black) and in unsubstituted (color) Li_2MgCl_4 ($Cmmm$) and Li_2MgCl_4 ($Fd\bar{3}m$).

Appendix B

B1. First-principles computation

We performed all density functional theory (DFT) calculations by using the Vienna ab initio simulation package^[98] (VASP) within the projector augmented-wave^[99] (PAW) approach. We adopted Perdew-Burke-Ernzerhof^[100] (PBE) functionals by generalized-gradient approximation (GGA) to calculate total energies in most of our calculations. In addition, we used the Heyd-Scuseria-Ernzerhof (HSE) functional^[101] to calculate the band gap and the density of states (DOS) for potential Li SIC materials. The convergence parameters of all static DFT calculations were consistent with the *Materials Project*.^[102,103]

B2. Ab Initio molecular dynamics simulation

AIMD simulations were performed following the same scheme adopted in Ref.^[29]. The supercell models for all materials were generated with lattice parameters at around 10 Å or larger. AIMD simulations used non-spin-polarized DFT calculations with a Γ -centered $1 \times 1 \times 1$ k -point grid. The initial temperature of simulations was set to 100K after static relaxation of all initial structures. With the time step setting to 2 fs, the structures were heated to the final temperatures at a constant rate by velocity scaling during a period of 2 ps. All simulations adopted the NVT ensemble with Nosé-Hoover thermostat, where an Nosé-mass corresponding to period of 40 time steps was chosen (i.e., SMASS = 0 setting in VASP).^[104] The evaluations of ionic diffusivity, conductivity and corresponding errors followed the method in Ref.^[4] using the Einstein relation. During the AIMD simulations, the trajectories of all lithium ions were tracked, and the total mean squared displacement (TMSD) of lithium ions is evaluated as a function of time interval Δt as

$$\text{TMSD}(\Delta t) = \sum_{i=1}^N \frac{1}{N\Delta t} \sum_{t=0}^{t_{\text{tot}}-\Delta t} |\mathbf{r}_i(t + \Delta t) - \mathbf{r}_i(t)|^2 \quad \text{Eq. (B1)}$$

where $\mathbf{r}_i$ is the trajectory of the lithium ion i and $N_{\Delta t}$ is the total number of time intervals Δt during the entire time duration t_{tot} of AIMD simulation. TMSD(Δt) is an averaged value of TMSD over all possible time intervals of Δt during the AIMD simulation and represents the total diffusion of all lithium ions over a period of time.

The diffusivity of a specific specie of mobile carrier can be evaluated according to Einstein relation:

$$D = \frac{1}{N} \frac{\text{TMSD}(\Delta t)}{2d\Delta t} \quad \text{Eq. (B2)}$$

where N is the number of the mobile carriers and d is set to 3 for 3D ion conductors.

The diffusivity or mobility is defined for a specific species of mobile carrier, for example, Li-ion versus vacancy, and has different values for different mobile carriers, so as the N is different.

The lithium ionic conductivity is calculated following the Nernst-Einstein relation as

$$\sigma = \frac{N}{V} \frac{q^2}{k_B T} D = \frac{q^2}{V k_B T} \frac{\text{TMSD}(\Delta t)}{2d\Delta t} \quad \text{Eq. (B3)}$$

where V is the volume of the model, q is the charge of the carrier, k_B is the Boltzmann constant, and T is the temperature. This equation assumes independent hopping of ions, which are confirmed by the AIMD simulations.^[17] The ion conductivity is independent from the choice of specific species of mobile carrier, as the TMSD represents the total diffusion of all lithium ions in the material.

Given that the ion hopping is a stochastic process, the statistical deviations of the diffusivities were evaluated according to the values of total mean squared displacement using the scheme in Ref.^[4] The total time duration of AIMD simulations were within the range of 100 to 1000 ps until the ionic diffusivity converged with relative standard deviation lower than 25%, which correspond to total mean squared displacements in an approximate range of 2500 to 6000

Å². The Li-ion probability density was calculated as the time fraction of Li-ions at each spatial location during the AIMD simulations.^[32]

B3. AIMD screening of Li-ion conductors

As an initial screening for fast Li-ion conductors, AIMD simulations were performed at 500 K and 600 K in unsubstituted and substituted compounds. The materials with $\sigma_{600\text{ K}} < 200\text{ mS cm}^{-1}$ or $\sigma_{500\text{ K}} < 50\text{ mS cm}^{-1}$ were first excluded because they were unlikely to be Li SICs at RT. Then we selected materials that satisfy $\log(\sigma T)_{600\text{ K}} < 1.25 \log(\sigma T)_{500\text{ K}} - 0.62$, for which extrapolated Li-ion conductivities are higher than 1 mS cm^{-1} at 300 K. The materials that passed the initial AIMD screening were further evaluated in AIMD simulations at a total of five temperatures (conducting additional simulations at 530 K, 560 K and 650 K) to obtain the Arrhenius relation:

$$\sigma T = \sigma_0 \exp\left(-\frac{E_a}{k_B T}\right) \quad \text{Eq. (B4)}$$

where T is temperature, k_B is the Boltzmann constant, E_a is activation energy, σ_0 is the pre-exponential factor and σ is conductivity.^[29] Using Arrhenius relationship, the activation energy and the evaluated Li-ion conductivity at 300 K were reported.

B4. Nudged elastic band calculations

NEB calculations were performed in the same supercell models. Static relaxation of initial and final structures used an energy convergence criterion at 10^{-7} eV and a force convergence criterion within 0.01 eV/Å. For NEB calculations, five images were linearly interpolated between the fully relaxed initial and final structures. The energy convergence criterion of NEB calculations was adjusted to 10^{-6} eV and the force convergence criterion was set to 0.05 eV/Å.

We constructed NEB migration pathways from Li-ion probability density generated by AIMD simulations^[32] to best represent the Li-ion hopping with realistic lithium and cation configurations. For unsubstituted compounds with no vacant Li-ion sites, we removed a Li ion from the original structure to create a Li-ion vacancy with a background electron for charge compensation. Two NEB calculations were performed for the first part of migration pathways between initial oct and intermediate tet sites and for the second part between intermediate tet and final oct sites. The energy barrier ΔE_b of the complete migration pathway was calculated by the difference between the maximum and minimum energies along the entire oct-tet-oct migration pathway. To have a consistent comparison among different pathways, the initial site was set to the site with the lowest energy, so that ΔE_f was equal to or greater than 0 eV.

B5. Selection of Li-containing chlorides systems

The candidate Li-containing chlorides were selected according to the following criteria. We first excluded compounds containing H, C, O, S, N, F, Br, I, or radioactive elements or having unbalanced charge. From the remaining structures, we selected the anion sublattices with fcc or hcp framework by applying the structural matching algorithm using pymatgen^[105], following a similar scheme in Ref.^[59]. The tolerance parameters of supercell lattice angle and the supercell lattice length was set to 5° and 20%, respectively, and the site root-mean-square tolerance was set to $0.3(V/n)^{1/3}$, where V/n was the volume V normalized by the number of atoms n . We kept only one representative compound for each unique structure and distinct cation configuration as a candidate material system. The numbers of compounds that passed each screening step are summarized (Figure A1).

For candidate systems with redox-active transition metal cations such as Co, Cr, and V, we first substituted the original cations by isovalent non-redox-active elements, such as Mg and Sc, as suggested by the ionic substitution probability.^[66] For candidate systems with disordered and partially occupied sites, we performed site ordering for these materials following the procedure described below. Then, static DFT relaxations were performed to select the material with the lowest E_{hull} .

B6. Aliovalent substitution of materials

The aliovalent substitution elements were identified from the main group elements and Sc, Y, Zr, Zn, Cd, that have high ionic substitution probability suggested by Ref.^[66] and have ionic radii that satisfy Pauling's first rule^[106].

To obtain substituted materials, we substituted 25%/33%, 50% and 100% of original cations with a partial occupancy. The Li content of the material was changed accordingly to balance the valence after aliovalent doping. For decreasing Li content, the partial occupancies of original Li-ion sites were correspondingly lowered. For increasing Li content, extra Li-ions were added to empty oct sites with corresponding partial occupancy. The site ordering process was performed to substituted material to obtain the ordered structure with lowest energy. Only materials with good phase stability, i.e., the energies above the hull $E_{\text{hull}} < 40$ meV per atom, were kept for further study. For artificially constructed materials with reduced cation concentration in Chapter 2.3, an identical procedure was followed.

B7. Ordering of structures with partial occupancy

To obtain the ordered structures, we generated 10,000 structures by randomly occupying the disordered sites with probabilities as the corresponding partial occupancies. Among all these structures, 20 symmetrically distinct structures with minimal electrostatic energies were selected for DFT calculations. The structure with the lowest DFT energy was selected as the representative ordered structure for further studies.

B8. Quantitative analysis of local environment surrounding cation and Li-ion

Quantitative analyses of local environments surrounding cation and Li-ion were performed on a large number of individual Li-ion migration pathways. For each individual Li-ion migration pathway, we generated RDFs $g(r)$ of initial, intermediate and final sites of Li-ion. We generated two RDFs of surrounding environments for each site, one for surrounding cations and the other one for surrounding Li-ions. The radial distances and coordination numbers obtained from $g(r)$ were used as quantitative features for linear regression models of migration energy barrier and site energy differences (details in Appendix A).

Appendix C

1. Errors of MLPs

Here, as shown in **Figure C1 and C2**, we perform additional testing for the energy and force errors of MLPs and DFT calculations using vacancy- and interstitial- RE testing sets, where the construction of these testing sets is described in the *Error Evaluation of Atomic Forces* in Appendix D. Here we analyze a total of 11 known MLPs, including six MLPs from previous studies as shown in main text (GAP, GAP_{PRX}, NNP, DeePMD, SNAP, and MTP), and GAP_{RPA} and GAP_{SCAN},^[40] DeePMD (original) and DeepPot-SE (original) directly retrieved from Ref.^[79], from Zuo et al.,^[1] Bartók et al.,^[78] Deringer et al.,^[40] and Zhang et al.,^[79] and the reproduced DeepPot-SE retrained from the same training data and optimization scheme, as described in Appendix D.

Table C1 shows the energy and force performances of 12 RE-enhanced MLPs on vacancy- and interstitial- RE testing sets. The six interstitial-enhanced MLPs are the representative MLPs obtained through our RE-enhanced work flow in *Optimizing Enhanced MLPs* in Appendix D and in main text. We obtain the root-mean-squared errors (RMSEs) of the energies and forces, σ_E^{Test} and σ_F^{Test} , predicted by MLPs (Table C1).^[1,78,79] RMSEs of the energies and forces on vacancy- RE testing set are labelled $\sigma_E^{\text{Vac Test}}$ and $\sigma_F^{\text{Vac Test}}$, while those on interstitial-RE testing set are labelled $\sigma_E^{\text{Int Test}}$ and $\sigma_F^{\text{Int Test}}$, respectively. Most MLPs show small averaged errors for atomic forces of all atoms in the dataset where σ_F^{Test} are mostly within the range between 0.1 to 0.4 eV/Å, consistent with the force RMSEs range 0.1 – 0.35 eV/Å in original publications^[1,40,78,79], except for DeepPot-SE (original), which may be because its original training data lacks of similar configurations in the testing dataset. All RMSEs of energies and forces on training data, σ_E^{Train} and σ_F^{Train} , are comparable with the values reported in previous publications.^[1,40,78,79]

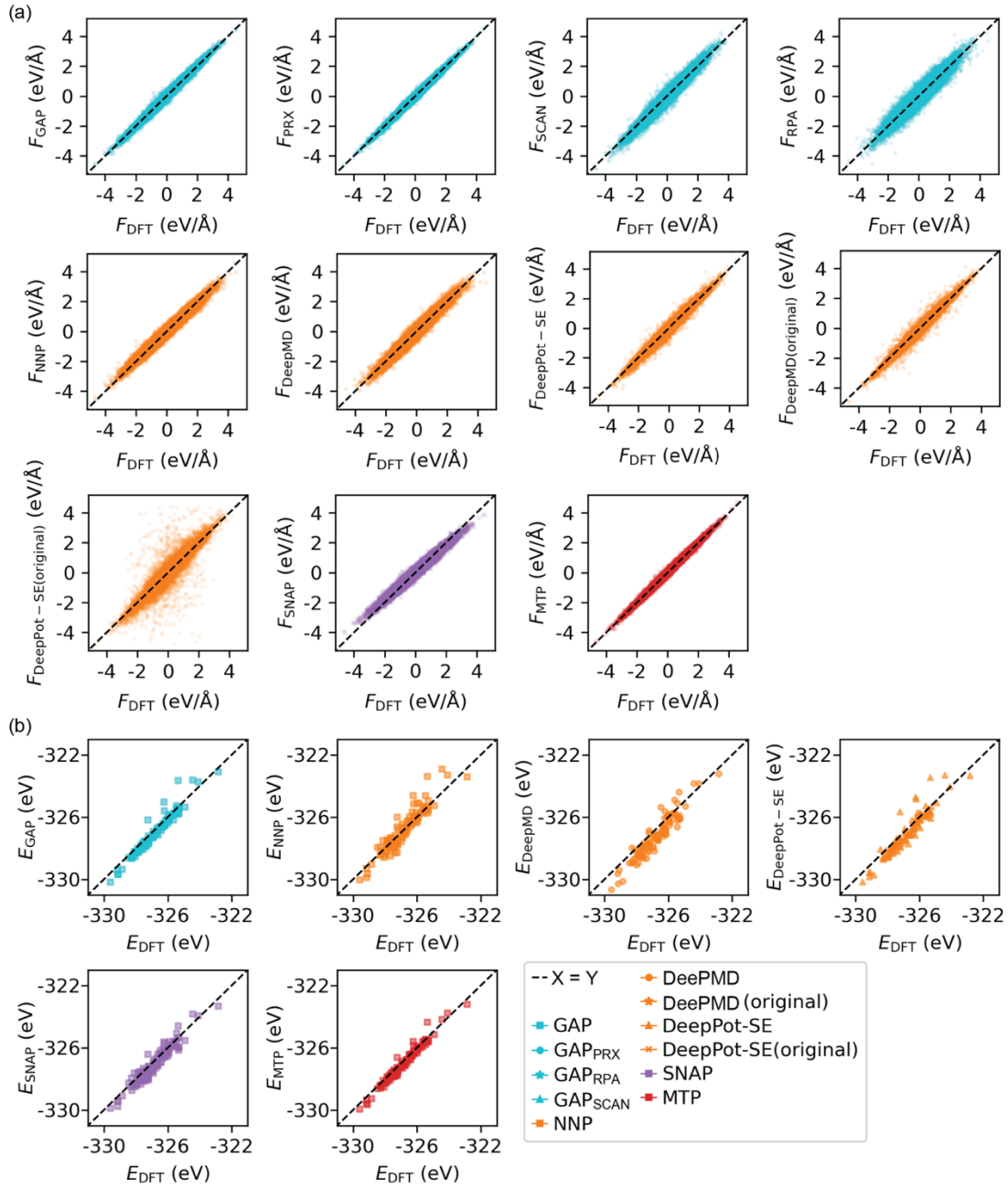


Figure C1. Comparison of the (a) atomic forces and (b) energies predicted by the known MLPs to DFT calculations for the vacancy-RE testing set.

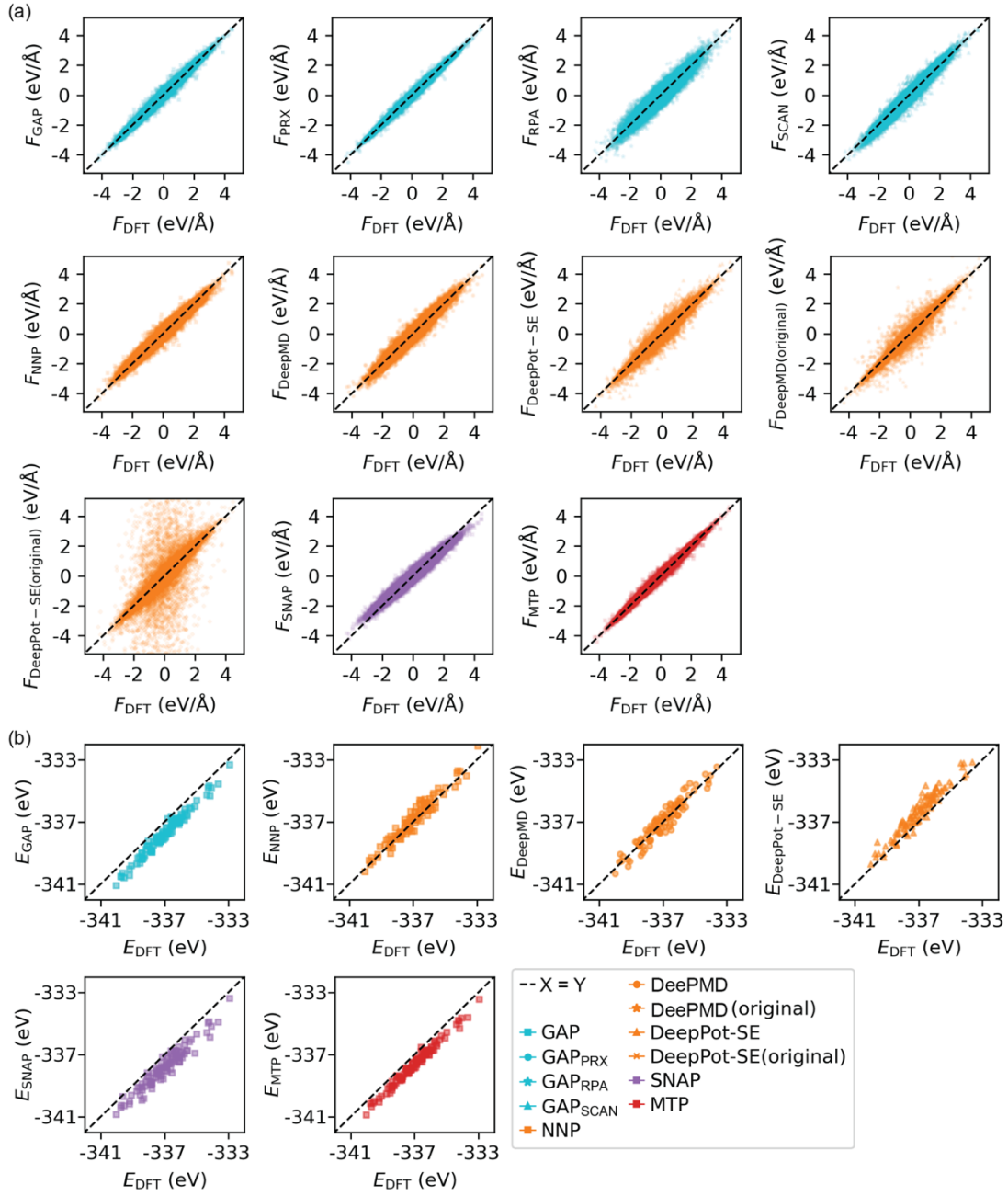


Figure C2. Comparison of the (a) atomic forces and (b) energies predicted by known MLPs to DFT calculations for the interstitial-RE testing set.

Table C1. Summary of MLP errors. RMSEs of energies on the training data, on the testing data, and of atomic forces on the testing data, using the vacancy-RE testing set and interstitial-RE testing set. The training energy RMSEs, σ_E^{Train} , are based on the original training dataset in previous studies.^[1,40,78,79]

Models		RMSE σ_E^{Train} (meV/atom)	RMSE $\sigma_E^{\text{Vac Test}}$ (meV/atom)	RMSE $\sigma_F^{\text{Vac Test}}$ (eV/Å)	RMSE $\sigma_E^{\text{Int Test}}$ (meV/atom)	RMSE $\sigma_F^{\text{Int Test}}$ (eV/Å)	Ref.
Known MLPs Trained in previous studies	GAP	5	6	0.12	13	0.15	[1]
	GAP _{PRX}	1	N/A	0.10	N/A	0.12	[78]
	GAP _{RPA}	46	N/A	0.38	N/A	0.38	[40]
	GAP _{SCAN}	3	N/A	0.28	N/A	0.27	[40]
	NNP	11	8	0.20	6	0.22	[1]
	DeePMD	5	10	0.24	7	0.26	
	DeepPot-SE	5	8	0.19	12	0.26	
	DeePMD (original)	1	N/A	0.21	N/A	0.29	[79]
	DeepPot-SE (original)	0.3	N/A	0.51	N/A	1.43	[79]
	MTP	3	5	0.10	10	0.12	[1]
Vacancy- enhanced MLPs	SNAP	8	7	0.25	14	0.26	[1]
	GAP _{RE-V}	3	3	0.09	8	0.13	
	NNP _{RE-V}	10	6	0.16	12	0.21	
	DeePMD _{RE-V}	11	6	0.14	12	0.24	
	DeepPot-SE _{RE-V}	7	5	0.14	14	0.19	
	SNAP _{RE-V}	12	5	0.15	5	0.18	
Interstitial- enhanced MLPs	MTP _{RE-V}	1	1	0.07	12	0.12	
	GAP _{RE-I}	3	4	0.10	3	0.10	
	NNP _{RE-I}	10	6	0.17	5	0.17	
	DeePMD _{RE-I}	14	6	0.22	6	0.19	
	DeepPot-SE _{RE-I}	9	7	0.16	4	0.16	
	SNAP _{RE-I}	13	6	0.15	5	0.16	
Reported MLP errors from previous studies	MTP _{RE-I}	2	4	0.09	3	0.09	
	GAP	5	4	0.12			[1]
	GAP _{PRX}	/	/	0.1 ^a			[78]
	NNP	11	10	0.17			[1]
	MTP	3	3	0.09			[1]
	SNAP	8	8	0.34			[1]
	DeePMD	0.5	/	0.03			[79]
	DeepPot-SE	0.2	/	0.04			[79]

^a We used the reported estimation of force RMSE for GAP_{PRX} from Bartók et. al.^[78] here.

** Energy errors of testing data on GAP_{PRX}, GAP_{RPA}, GAP_{SCAN}, DeePMD (original), and DeepPot-SE (original) are not evaluated (N/A) since they were trained using different DFT functionals or k -point meshes in previous publications, which cannot be directly compared with our DFT calculations.

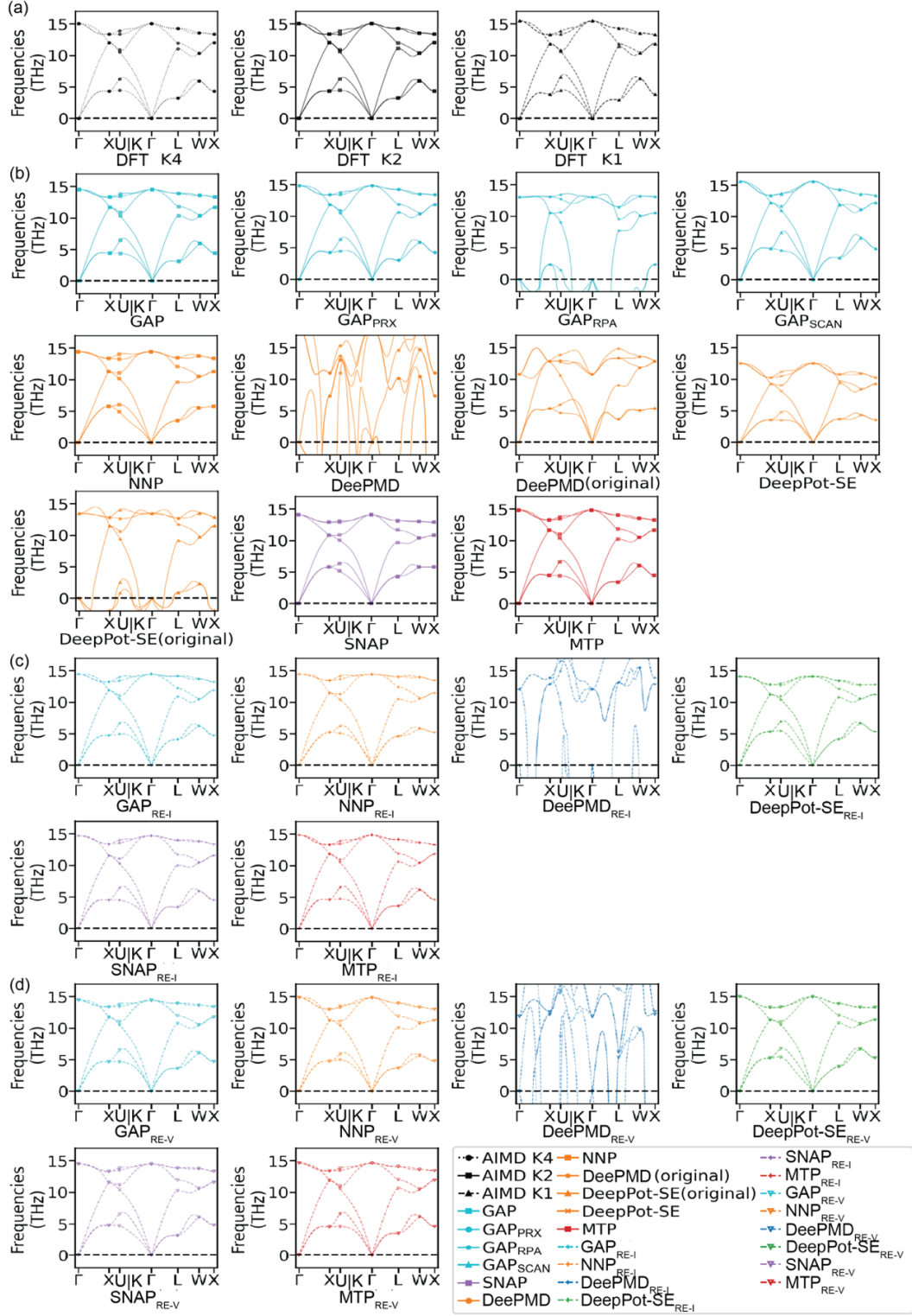


Figure C3. Comparison of the phonon dispersion spectra of the bulk Si calculated by (a) DFT calculations, (b) original MLPs, (c) interstitial-enhanced MLPs, and (d) vacancy-enhanced MLPs.

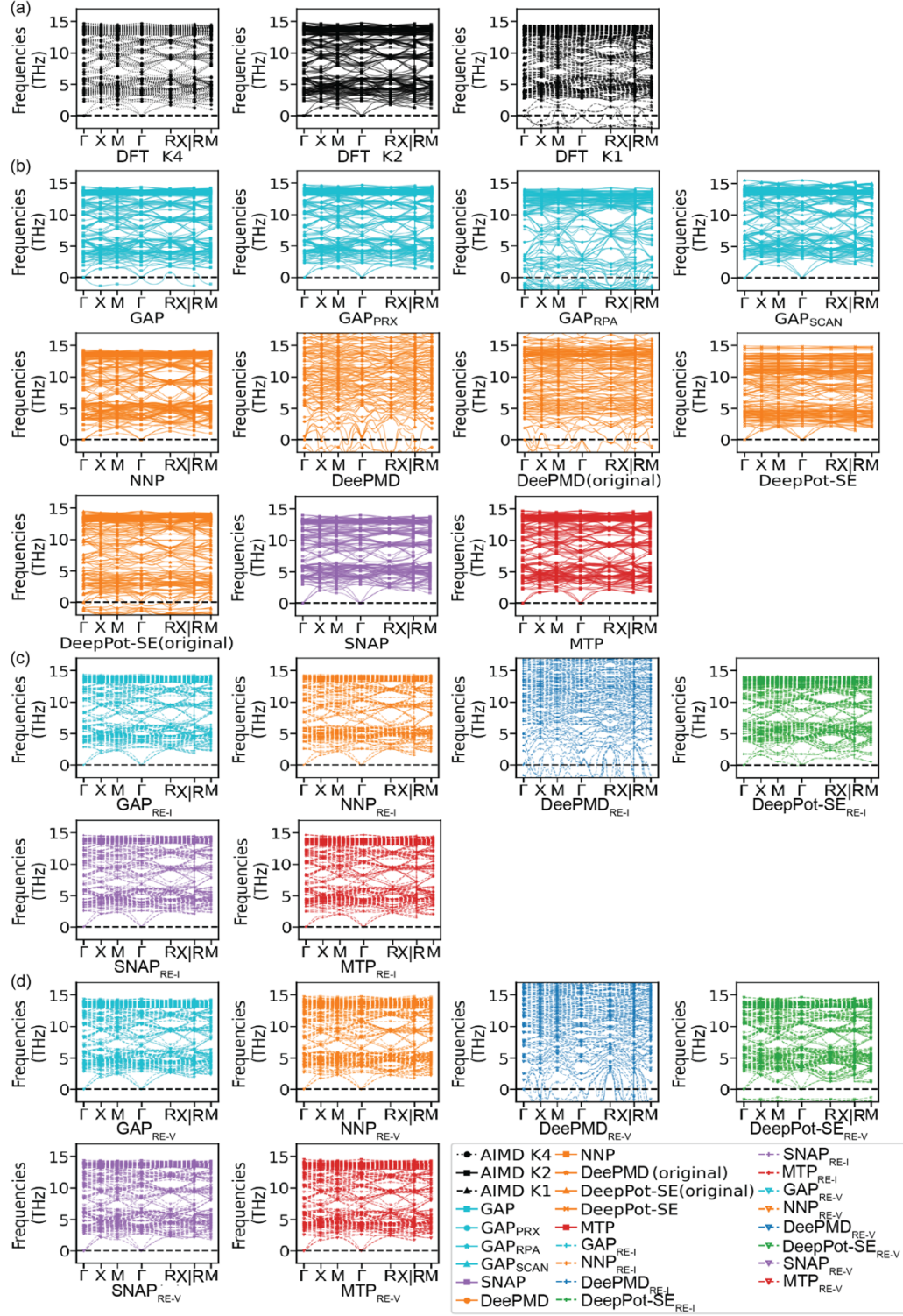


Figure C4. Comparison of the phonon dispersion spectra of the Si supercell with a single vacancy calculated from (a) DFT calculations, (b) the known MLPs, (c) the interstitial-enhanced MLPs, and (d) the vacancy-enhanced MLPs.

2. Comparing atomic environments of training datasets

Here we quantify and compare the atomic environments in different testing datasets and training datasets from previous studies and this study. The Smooth Overlap of Atomic Positions (SOAP) descriptor is employed using the same hyperparameter set and following the same approach as described in the Appendix D and Ref. [1], and has a feature vector of 325 elements for each atomic environment. To visualize the atomic environments and keep the computation process consistent for different datasets, we use all 657451 SOAP descriptors (17785 for GAP_{SCAN} training data in Deringer et al. (SCAN),^[40] 9498 for GAP_{RPA} training data in Deringer et al. (RPA),^[40] 171815 for GAP_{PRX} training data in Bartok et al.,^[78] 13553 for dataset in Zuo et al.,^[1] 432000 for DeePMD (original) training data in Zhang et al.,^[79] 6500 for $\mathcal{D}_{\text{RE-I}}^{\text{Testing}}$, and 6300 for $\mathcal{D}_{\text{RE-V}}^{\text{Testing}}$) here and employ the following steps for dimension reduction and visualization. Firstly, we standardize all SOAP vectors by subtracting the mean and then scaling them by the standard deviation. Then, to visualize the distributions of atomic environments, we perform principal component analysis (PCA) on all standardized descriptors and plot the data distributions using the 1st and the 2nd principal components. We use the two principal components generated here to plot the atomic environments (Figure 3.1a, b, and c in the main text, and Figure C5).

To numerically quantify the differences of all atomic environments across the datasets, we first calculate a central average SOAP vector for each dataset by averaging the SOAP vector of all atom environments in the corresponding dataset. We then calculate the Euclidean distances d between the atomic environments in the given dataset (y axis) and the central SOAP vectors of the target datasets (x axis) as is shown (**Figure C6**). Thus, the differences of the atomic environments within the same dataset are quantified by the average and the standard deviation of the distances d to its central SOAP vector (the diagonal elements, Figure C6). The differences between two

different datasets are calculated using the same method shown as the non-diagonal elements (Figure C6).

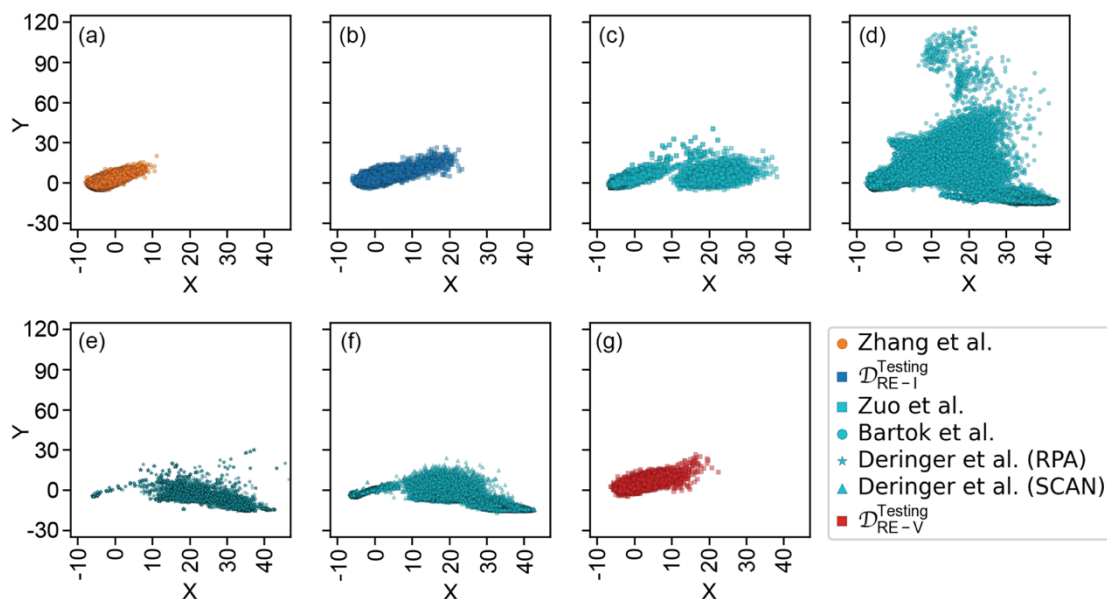


Figure C5. Comparison of atomic environments using the first two principal components of PCA based on the SOAP descriptors for the training datasets adopted in (a) Zhang et al.^[79] (orange dots), (c) Zuo et al.^[1] (cyan squares), (d) Bartok et al.^[78] (cyan dots), (e) RPA in Deringer et al.^[40] (cyan stars), (f) SCAN in Deringer et al.^[40] (cyan triangles), and the testing datasets (b) $\mathcal{D}_{\text{RE-I}}^{\text{Testing}}$ (blue squares) and (g) $\mathcal{D}_{\text{RE-V}}^{\text{Testing}}$ (red squares).

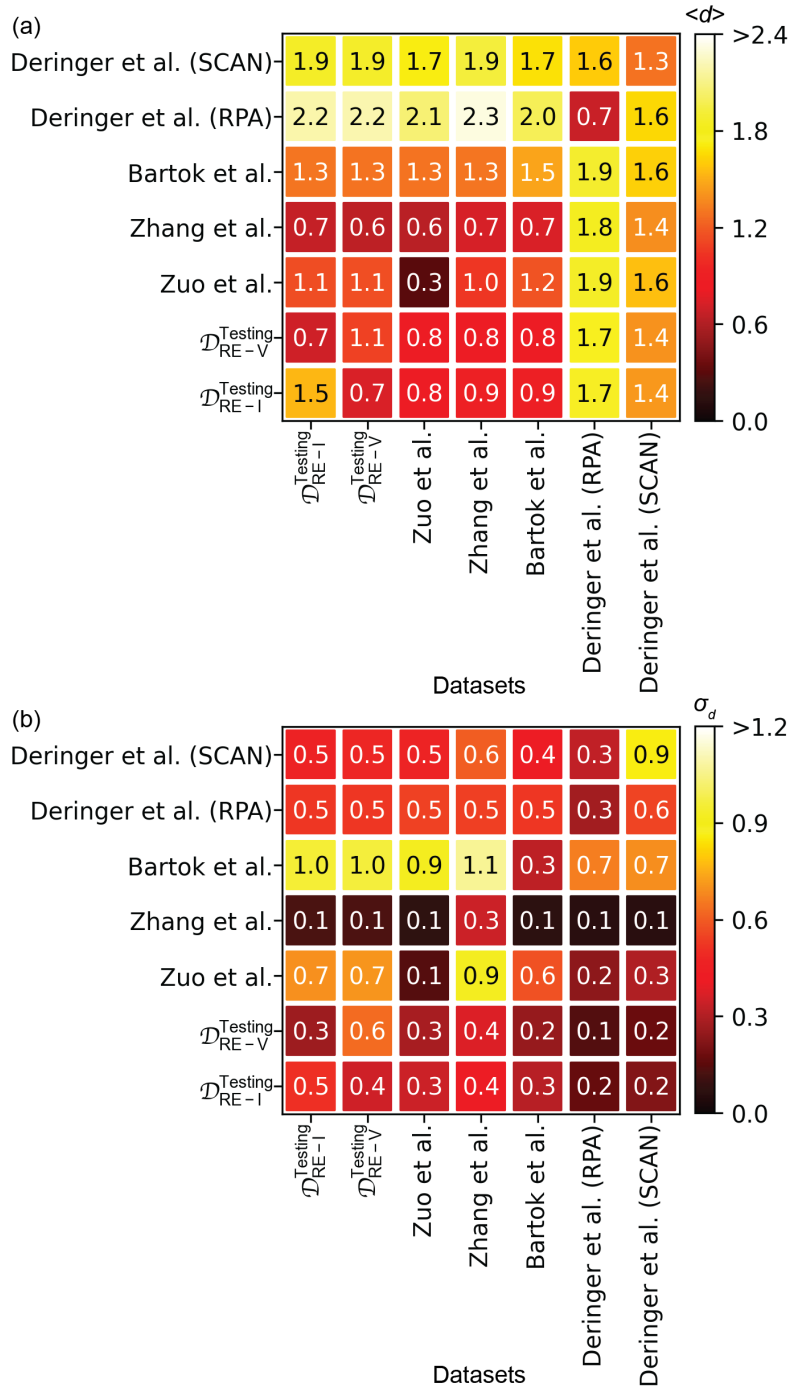


Figure C6. The differences of atomic environments within and between the datasets. (a) The average Euclidean distances, $\langle d \rangle$, and (b) the standard deviation of Euclidean distances, σ_d , for each pair of datasets (non-diagonal elements) and within the datasets (diagonal elements). The target datasets (x -axis) use their central SOAP vectors to calculate d with all the atomic environments in the given datasets (y -axis).

3. Vacancy-enhanced MLPs

In addition to interstitial-enhanced MLPs shown in main text, we perform the RE-enhanced workflow (Appendix D) for vacancies, and we construct vacancy-enhanced training dataset and select six representative vacancy-enhanced MLPs. Following the same procedure of constructing interstitial-enhanced dataset in *Generating RE-Enhanced Training Dataset*, Appendix D, vacancy-enhanced training dataset is constructed by replacing 120 structures in the original dataset in Ref.^[1] by the snapshots containing vacancy REs that are selected by the vacancy k-means clustering model described in Chapter 3 and Appendix D. All processes and protocols, including the replaced structures and the sampling approach for vacancy snapshots, remain the same as described in Appendix D.

Trained by the vacancy-enhanced training data, we select 54 good vacancy-enhanced MLPs, using the enhanced validation set, the optimization processes, and the 8-selection criterion same as in the *Optimizing Enhanced MLPs* in Appendix D. Finally, we calculate the joint force error score $P(\mathcal{D}_{\text{RE-I}}^{\text{EVS}})^2 + P(\mathcal{D}_{\text{RE-V}}^{\text{EVS}})^2$ of these 54 selected vacancy-enhanced MLPs, and pick six of them with the highest scores (*Optimizing Enhanced MLPs* in the Appendix D) for each model as the representative vacancy-enhanced MLPs.

4. Diffusional properties

The diffusional properties evaluated from AIMD and MLP-MD simulations for all MLPs (including known and RE-enhanced MLPs) are summarized (**Table C2 and C3**). For NNP (with interstitial) and DeepPot-SE (with either vacancy or interstitial), the crystal structures of Si melt at all temperatures tested (730 – 1600 K). For GAP_{SCAN}, MD simulations with a vacancy show no diffusion at all temperatures tested, and for GAP_{RPA}, the Si structures melt at temperatures 1230 K to 1600 K and show no diffusion below 1230 K. For NNP_{RE-I}, DeePMD_{RE-I}, and SNAP_{RE-I}, MD simulations with a single vacancy show no diffusion at 1000 K.

Table C2. The diffusivities at 1000 K, the pre-exponential factors D_0 and activation energies E_a from fitted Arrhenius relations for single-interstitial (superscript I) from AIMD and MLP-MD simulations.

Models		D_{1000K}^I (cm ² /s)	Error bound [D_{min} , D_{max}] (cm ² /s)	D_0^I (cm ² /s)	Error bound [D_{min} , D_{max}] (cm ² /s)	E_a^I (eV)
AIMD	K2	6.6*10 ⁻⁷	[8.9*10 ⁻⁷ , 4.4*10 ⁻⁷]	2.3*10 ⁻⁵	[4.1*10 ⁻⁵ , 1.3*10 ⁻⁵]	0.30 ± 0.06
	K1	4.5*10 ⁻⁷	[5.7*10 ⁻⁷ , 3.4*10 ⁻⁷]	7.7*10 ⁻⁶	[1.3*10 ⁻⁵ , 4.5*10 ⁻⁶]	0.25 ± 0.05
Known MLPs trained from previous studies	GAP	4.6*10 ⁻⁷	[5.8*10 ⁻⁷ , 3.4*10 ⁻⁷]	5.9*10 ⁻⁵	[9.4*10 ⁻⁵ , 3.7*10 ⁻⁵]	0.42 ± 0.05
	GAP _{PRX}	1.8*10 ⁻⁶	[2.2*10 ⁻⁶ , 1.4*10 ⁻⁶]	8.3*10 ⁻⁶	[1.4*10 ⁻⁵ , 4.9*10 ⁻⁶]	0.12 ± 0.06
	GAP _{RPA}	1.8*10 ⁻⁶	[2.1*10 ⁻⁶ , 1.5*10 ⁻⁶]	6.1*10 ⁻⁵	[1.4*10 ⁻⁴ , 2.6*10 ⁻⁵]	0.29 ± 0.06
	GAP _{SCAN}	6.7*10 ⁻⁷	[8.7*10 ⁻⁷ , 4.7*10 ⁻⁷]	1.6*10 ⁻⁵	[2.2*10 ⁻⁵ , 1.1*10 ⁻⁵]	0.23 ± 0.04
	NNP	/	/	/	/	/
	DeePMD	2.3*10 ⁻⁶	[2.5*10 ⁻⁶ , 2.2*10 ⁻⁶]	1.0*10 ⁻⁴	[1.1*10 ⁻⁴ , 9.2*10 ⁻⁵]	0.24 ± 0.01
	DeepPot-SE	7.7*10 ⁻⁷	[8.7*10 ⁻⁷ , 6.8*10 ⁻⁷]	2.6*10 ⁻⁵	[3.1*10 ⁻⁵ , 2.2*10 ⁻⁵]	0.31 ± 0.02
	DeePMD (original)	3.1*10 ⁻⁶	[3.4*10 ⁻⁶ , 2.9*10 ⁻⁶]	5.5*10 ⁻⁶	[6.6*10 ⁻⁶ , 4.5*10 ⁻⁶]	0.04 ± 0.01
	DeepPot-SE (original)	/	/	/	/	/
	MTP	5.5*10 ⁻⁷	[6.3*10 ⁻⁷ , 4.7*10 ⁻⁷]	5.7*10 ⁻⁵	[7.1*10 ⁻⁵ , 4.6*10 ⁻⁵]	0.42 ± 0.02
Vacancy- enhanced MLPs	SNAP	1.3*10 ⁻⁷	[1.5*10 ⁻⁷ , 1.1*10 ⁻⁷]	6.4*10 ⁻⁴	[1.0*10 ⁻³ , 4.1*10 ⁻⁴]	0.74 ± 0.05
	GAP _{RE-V}	3.7*10 ⁻⁷	[4.5*10 ⁻⁷ , 2.9*10 ⁻⁷]	5.9*10 ⁻⁵	[7.7*10 ⁻⁵ , 4.5*10 ⁻⁵]	0.42 ± 0.03
	NNP _{RE-V}	1.4*10 ⁻⁶	[1.6*10 ⁻⁶ , 1.2*10 ⁻⁶]	3.6*10 ⁻⁶	[4.1*10 ⁻⁶ , 3.2*10 ⁻⁶]	0.08 ± 0.01
	DeePMD _{RE-V}	5.1*10 ⁻⁶	[5.4*10 ⁻⁶ , 4.8*10 ⁻⁶]	2.2*10 ⁻⁵	[2.4*10 ⁻⁵ , 2.1*10 ⁻⁵]	0.13 ± 0.01
	DeepPot-SE _{RE-V}	1.0*10 ⁻⁶	[1.1*10 ⁻⁶ , 9.1*10 ⁻⁷]	1.5*10 ⁻⁵	[1.7*10 ⁻⁵ , 1.3*10 ⁻⁵]	0.22 ± 0.01
	SNAP _{RE-V}	2.3*10 ⁻⁶	[2.5*10 ⁻⁶ , 2.0*10 ⁻⁶]	1.4*10 ⁻⁵	[1.7*10 ⁻⁵ , 1.2*10 ⁻⁵]	0.17 ± 0.02
	MTP _{RE-V}	4.3*10 ⁻⁸	[6.1*10 ⁻⁸ , 2.6*10 ⁻⁸]	4.1*10 ⁻⁴	[7.1*10 ⁻⁴ , 2.3*10 ⁻⁴]	0.77 ± 0.07
Interstitial- enhanced MLPs	GAP _{RE-I}	1.0*10 ⁻⁶	[1.2*10 ⁻⁶ , 8.3*10 ⁻⁷]	1.1*10 ⁻⁵	[1.4*10 ⁻⁵ , 8.2*10 ⁻⁶]	0.20 ± 0.02
	NNP _{RE-I}	7.8*10 ⁻⁷	[9.2*10 ⁻⁷ , 6.4*10 ⁻⁷]	3.2*10 ⁻⁵	[3.9*10 ⁻⁵ , 2.7*10 ⁻⁵]	0.33 ± 0.02
	DeePMD _{RE-I}	1.1*10 ⁻⁵	[1.1*10 ⁻⁵ , 1.0*10 ⁻⁵]	4.1*10 ⁻⁵	[4.4*10 ⁻⁵ , 3.8*10 ⁻⁵]	0.15 ± 0.01
	DeepPot-SE _{RE-I}	1.9*10 ⁻⁶	[2.0*10 ⁻⁶ , 1.7*10 ⁻⁶]	8.5*10 ⁻⁶	[9.5*10 ⁻⁶ , 7.6*10 ⁻⁶]	0.14 ± 0.01
	SNAP _{RE-I}	1.2*10 ⁻⁶	[1.3*10 ⁻⁶ , 9.7*10 ⁻⁷]	3.1*10 ⁻⁵	[3.7*10 ⁻⁵ , 2.6*10 ⁻⁵]	0.29 ± 0.02
	MTP _{RE-I}	1.1*10 ⁻⁶	[1.2*10 ⁻⁶ , 9.5*10 ⁻⁷]	1.9*10 ⁻⁵	[2.2*10 ⁻⁵ , 1.6*10 ⁻⁵]	0.26 ± 0.02

Table C3. The diffusivities at 1000 K, the pre-exponential factors D_0 and activation energies E_a from fitted Arrhenius relations for single-vacancy (superscript V) from AIMD and MLP-MD simulations.

Models		D_{1000K}^V (cm ² /s)	Error bound [D_{min} , D_{max}] (cm ² /s)	D_0^V (cm ² /s)	Error bound [D_{min} , D_{max}] (cm ² /s)	E_a^V (eV)
AIMD	K2	5.2*10 ⁻⁷	[6.9*10 ⁻⁷ , 3.6*10 ⁻⁷]	6.0*10 ⁻⁶	[1.1*10 ⁻⁵ , 3.1*10 ⁻⁶]	0.20 ± 0.07
	K1	2.2*10 ⁻⁷	[2.8*10 ⁻⁷ , 1.6*10 ⁻⁷]	7.4*10 ⁻⁶	[1.5*10 ⁻⁵ , 3.7*10 ⁻⁶]	0.31 ± 0.07
Known MLPs trained from previous studies	GAP	5.8*10 ⁻⁷	[7.1*10 ⁻⁷ , 4.5*10 ⁻⁷]	6.8*10 ⁻⁶	[1.0*10 ⁻⁵ , 4.6*10 ⁻⁶]	0.21 ± 0.04
	GAP _{PRX}	5.6*10 ⁻⁷	[7.8*10 ⁻⁷ , 3.3*10 ⁻⁷]	5.1*10 ⁻⁶	[1.2*10 ⁻⁵ , 2.1*10 ⁻⁶]	0.18 ± 0.10
	GAP _{RPA}	/	/	/	/	/
	GAP _{SCAN}	/	/	/	/	/
	NNP	6.7*10 ⁻⁷	[7.3*10 ⁻⁷ , 6.1*10 ⁻⁷]	5.8*10 ⁻⁶	[8.0*10 ⁻⁶ , 4.2*10 ⁻⁶]	0.19 ± 0.03
	DeePMD	2.8*10 ⁻⁶	[3.0*10 ⁻⁶ , 2.5*10 ⁻⁶]	3.6*10 ⁻⁵	[4.0*10 ⁻⁵ , 3.3*10 ⁻⁵]	0.17 ± 0.01
	DeepPot-SE	9.1*10 ⁻⁷	[1.0*10 ⁻⁶ , 8.0*10 ⁻⁷]	6.3*10 ⁻⁶	[7.4*10 ⁻⁶ , 5.4*10 ⁻⁶]	0.17 ± 0.02
	DeePMD (original)	1.1*10 ⁻⁵	[1.1*10 ⁻⁵ , 9.9*10 ⁻⁶]	4.6*10 ⁻⁵	[5.1*10 ⁻⁵ , 4.2*10 ⁻⁵]	0.16 ± 0.01
	DeepPot-SE (original)	/	/	/	/	/
	MTP	6.0*10 ⁻⁷	[6.8*10 ⁻⁷ , 5.2*10 ⁻⁷]	1.1*10 ⁻⁵	[1.4*10 ⁻⁵ , 9.3*10 ⁻⁶]	0.25 ± 0.02
	SNAP	7.8*10 ⁻⁷	[8.5*10 ⁻⁷ , 7.1*10 ⁻⁷]	5.4*10 ⁻⁶	[7.0*10 ⁻⁶ , 4.1*10 ⁻⁶]	0.16 ± 0.02
Vacancy- enhanced MLPs	GAP _{RE-V}	3.9*10 ⁻⁷	[4.7*10 ⁻⁷ , 3.0*10 ⁻⁷]	8.7*10 ⁻⁶	[1.1*10 ⁻⁵ , 6.7*10 ⁻⁶]	0.24 ± 0.03
	NNP _{RE-V}	9.4*10 ⁻⁷	[1.1*10 ⁻⁶ , 7.7*10 ⁻⁷]	9.4*10 ⁻⁶	[1.2*10 ⁻⁵ , 7.3*10 ⁻⁶]	0.21 ± 0.03
	DeePMD _{RE-V}	3.5*10 ⁻⁶	[3.7*10 ⁻⁶ , 3.2*10 ⁻⁶]	1.2*10 ⁻⁵	[1.3*10 ⁻⁵ , 1.1*10 ⁻⁵]	0.12 ± 0.01
	DeepPot-SE _{RE-V}	7.5*10 ⁻⁷	[8.3*10 ⁻⁷ , 6.7*10 ⁻⁷]	3.6*10 ⁻⁶	[4.0*10 ⁻⁶ , 3.2*10 ⁻⁶]	0.13 ± 0.01
	SNAP _{RE-V}	5.7*10 ⁻⁷	[7.0*10 ⁻⁷ , 4.4*10 ⁻⁷]	8.5*10 ⁻⁶	[1.0*10 ⁻⁵ , 6.9*10 ⁻⁶]	0.23 ± 0.02
	MTP _{RE-V}	8.0*10 ⁻⁷	[9.0*10 ⁻⁷ , 7.0*10 ⁻⁷]	6.8*10 ⁻⁶	[8.1*10 ⁻⁶ , 5.7*10 ⁻⁶]	0.19 ± 0.02
Interstitial- enhanced MLPs	GAP _{RE-I}	9.7*10 ⁻⁷	[1.2*10 ⁻⁶ , 7.9*10 ⁻⁷]	6.4*10 ⁻⁶	[8.2*10 ⁻⁶ , 4.9*10 ⁻⁶]	0.17 ± 0.02
	NNP _{RE-I}	/	/	1.8*10 ⁻⁶	[2.3*10 ⁻⁶ , 1.4*10 ⁻⁶]	0.08 ± 0.03
	DeePMD _{RE-I}	/	/	2.7*10 ⁻⁵	[2.9*10 ⁻⁵ , 2.5*10 ⁻⁵]	0.12 ± 0.01
	DeepPot-SE _{RE-I}	6.2*10 ⁻⁷	[6.9*10 ⁻⁷ , 5.4*10 ⁻⁷]	7.1*10 ⁻⁶	[8.1*10 ⁻⁶ , 6.3*10 ⁻⁶]	0.21 ± 0.01
	SNAP _{RE-I}	/	/	2.6*10 ⁻⁶	[3.2*10 ⁻⁶ , 2.1*10 ⁻⁶]	0.14 ± 0.03
	MTP _{RE-I}	1.2*10 ⁻⁶	[1.3*10 ⁻⁶ , 1.1*10 ⁻⁶]	4.6*10 ⁻⁶	[5.4*10 ⁻⁶ , 4.0*10 ⁻⁶]	0.11 ± 0.01

Table C4. The diffusivities at 1600 K, 1500K, 1230 K, 1000 K, 840 K, and 730 K for interstitial (superscript I) from AIMD and MLP-MD simulations. The MD simulations with melting structures (noted as melt) and not enough hopping events (noted as N/A) are marked.

Models		D_{1600K} (cm ² /s)	D_{1500K} (cm ² /s)	D_{1230K} (cm ² /s)	D_{1000K} (cm ² /s)	D_{840K} (cm ² /s)	D_{730K} (cm ² /s)
AIMD	K2	2.3×10^{-6}	3.0×10^{-6}	1.3×10^{-6}	6.6×10^{-7}	N/A	N/A
	Error bound	$[2.7 \times 10^{-6},$	$[3.6 \times 10^{-6},$	$[1.6 \times 10^{-6},$	$[8.9 \times 10^{-7},$		
	$[D_{min}, D_{max}]$ (cm ² /s)	$1.9 \times 10^{-6}]$	$2.4 \times 10^{-6}]$	$9.5 \times 10^{-7}]$	$4.4 \times 10^{-7}]$		
	K1	Melt	Melt	6.4×10^{-7}	4.5×10^{-7}	1.9×10^{-7}	1.6×10^{-7}
Known MLPs trained from previous studies	Error bound			$[8.0 \times 10^{-7},$	$[5.7 \times 10^{-7},$	$[2.6 \times 10^{-7},$	$[2.2 \times 10^{-7},$
	$[D_{min}, D_{max}]$ (cm ² /s)			$4.8 \times 10^{-7}]$	$3.4 \times 10^{-7}]$	$1.2 \times 10^{-7}]$	$8.7 \times 10^{-8}]$
	GAP	2.8×10^{-6}	Melt	1.1×10^{-6}	4.6×10^{-7}	1.7×10^{-7}	N/A
	Error bound	$[3.2 \times 10^{-6},$		$[1.3 \times 10^{-6},$	$[5.8 \times 10^{-7},$	$[2.3 \times 10^{-7},$	
	$[D_{min}, D_{max}]$ (cm ² /s)	$2.4 \times 10^{-6}]$		$8.2 \times 10^{-7}]$	$3.4 \times 10^{-7}]$	$1.1 \times 10^{-7}]$	
	GAP _{PRX}	3.4×10^{-6}	2.9×10^{-6}	3.4×10^{-6}	1.8×10^{-6}	N/A	N/A
	Error bound	$[4.0 \times 10^{-6},$	$[3.4 \times 10^{-6},$	$[4.1 \times 10^{-6},$	$[2.2 \times 10^{-6},$		
	$[D_{min}, D_{max}]$ (cm ² /s)	$2.8 \times 10^{-6}]$	$2.4 \times 10^{-6}]$	$2.8 \times 10^{-6}]$	$1.4 \times 10^{-6}]$		
	GAP _{RPA}	Melt	Melt	Melt	1.8×10^{-6}	1.5×10^{-6}	4.5×10^{-7}
	Error bound				$[2.1 \times 10^{-6},$	$[1.8 \times 10^{-6},$	$[5.5 \times 10^{-7},$
	$[D_{min}, D_{max}]$ (cm ² /s)				$1.5 \times 10^{-6}]$	$1.3 \times 10^{-6}]$	$3.5 \times 10^{-7}]$
	GAP _{SCAN}	2.9×10^{-6}	2.8×10^{-6}	1.6×10^{-6}	6.7×10^{-7}	7.7×10^{-7}	4.7×10^{-7}
	Error bound	$[3.5 \times 10^{-6},$	$[3.3 \times 10^{-6},$	$[2.0 \times 10^{-6},$	$[8.7 \times 10^{-7},$	$[1.0 \times 10^{-6},$	$[6.4 \times 10^{-7},$
	$[D_{min}, D_{max}]$ (cm ² /s)	$2.4 \times 10^{-6}]$	$2.3 \times 10^{-6}]$	$1.3 \times 10^{-6}]$	$4.7 \times 10^{-7}]$	$5.3 \times 10^{-7}]$	$2.9 \times 10^{-7}]$
	NNP	Melt	Melt	Melt	Melt	Melt	Melt
	Error bound						
	$[D_{min}, D_{max}]$ (cm ² /s)						
	DeepPMD	Melt	2.1×10^{-5}	5.7×10^{-6}	2.3×10^{-6}	3.8×10^{-6}	2.5×10^{-6}
	Error bound		$[2.2 \times 10^{-5},$	$[6.1 \times 10^{-6},$	$[2.5 \times 10^{-6},$	$[4.1 \times 10^{-6},$	$[2.6 \times 10^{-6},$
	$[D_{min}, D_{max}]$ (cm ² /s)		$2.0 \times 10^{-5}]$	$5.3 \times 10^{-6}]$	$2.2 \times 10^{-6}]$	$3.5 \times 10^{-6}]$	$2.3 \times 10^{-6}]$
	DeepPot-SE	2.9×10^{-6}	2.1×10^{-6}	1.7×10^{-6}	7.7×10^{-7}	3.2×10^{-7}	2.0×10^{-7}
	Error bound	$[3.1 \times 10^{-6},$	$[2.3 \times 10^{-6},$	$[1.8 \times 10^{-6},$	$[8.7 \times 10^{-7},$	$[3.7 \times 10^{-7},$	$[2.4 \times 10^{-7},$
	$[D_{min}, D_{max}]$ (cm ² /s)	$2.7 \times 10^{-6}]$	$1.9 \times 10^{-6}]$	$1.5 \times 10^{-6}]$	$6.8 \times 10^{-7}]$	$2.7 \times 10^{-7}]$	$1.6 \times 10^{-7}]$
	DeepPMD (original)	Melt	Melt	4.0×10^{-6}	3.1×10^{-6}	3.1×10^{-6}	3.1×10^{-6}
	Error bound			$[4.3 \times 10^{-6},$	$[3.4 \times 10^{-6},$	$[3.3 \times 10^{-6},$	$[3.3 \times 10^{-6},$
	$[D_{min}, D_{max}]$ (cm ² /s)			$3.8 \times 10^{-6}]$	$2.9 \times 10^{-6}]$	$2.9 \times 10^{-6}]$	$2.8 \times 10^{-6}]$
	DeepPot-SE (original)	Melt	Melt	Melt	Melt	Melt	Melt
	Error bound						
	$[D_{min}, D_{max}]$ (cm ² /s)						
	MTP	2.9×10^{-6}	2.1×10^{-6}	1.1×10^{-6}	5.5×10^{-7}	1.4×10^{-7}	7.4×10^{-8}
	Error bound	$[3.2 \times 10^{-6},$	$[2.3 \times 10^{-6},$	$[1.3 \times 10^{-6},$	$[6.3 \times 10^{-7},$	$[1.8 \times 10^{-7},$	$[1.0 \times 10^{-7},$
	$[D_{min}, D_{max}]$ (cm ² /s)	$2.7 \times 10^{-6}]$	$1.9 \times 10^{-6}]$	$1.0 \times 10^{-6}]$	$4.7 \times 10^{-7}]$	$1.1 \times 10^{-7}]$	$4.8 \times 10^{-8}]$
	SNAP	3.0×10^{-6}	Melt	4.7×10^{-7}	1.3×10^{-7}	2.0×10^{-8}	N/A
	Error bound	$[3.5 \times 10^{-6},$		$[6.4 \times 10^{-7},$	$[1.5 \times 10^{-7},$	$[2.9 \times 10^{-8},$	
	$[D_{min}, D_{max}]$ (cm ² /s)	$2.6 \times 10^{-6}]$		$3.0 \times 10^{-7}]$	$1.1 \times 10^{-7}]$	$9.9 \times 10^{-9}]$	
	Vacancy-enhanced MLPs						
	GAP _{RE-V}	2.9×10^{-6}	2.2×10^{-6}	1.1×10^{-6}	3.7×10^{-7}	1.6×10^{-7}	1.1×10^{-7}
	Error bound	$[3.3 \times 10^{-6},$	$[2.5 \times 10^{-6},$	$[1.2 \times 10^{-6},$	$[4.5 \times 10^{-7},$	$[2.1 \times 10^{-7},$	$[1.5 \times 10^{-7},$
	$[D_{min}, D_{max}]$ (cm ² /s)	$2.6 \times 10^{-6}]$	$2.0 \times 10^{-6}]$	$9.1 \times 10^{-7}]$	$2.9 \times 10^{-7}]$	$1.1 \times 10^{-7}]$	$7.1 \times 10^{-8}]$
	NNP _{RE-V}	Melt	Melt	Melt	1.4×10^{-6}	1.5×10^{-6}	9.4×10^{-7}
	Error bound				$[1.6 \times 10^{-6},$	$[1.6 \times 10^{-6},$	$[1.1 \times 10^{-6},$
	$[D_{min}, D_{max}]$ (cm ² /s)				$1.2 \times 10^{-6}]$	$1.3 \times 10^{-6}]$	$7.8 \times 10^{-7}]$
	DeepPMD _{RE-V}	Melt	Melt	7.3×10^{-6}	5.1×10^{-6}	3.3×10^{-6}	3.4×10^{-6}
	Error bound			$[7.8 \times 10^{-6},$	$[5.4 \times 10^{-6},$	$[3.6 \times 10^{-6},$	$[3.6 \times 10^{-6},$
	$[D_{min}, D_{max}]$ (cm ² /s)			$6.8 \times 10^{-6}]$	$4.8 \times 10^{-6}]$	$3.1 \times 10^{-6}]$	$3.2 \times 10^{-6}]$
	DeepPot-SE _{RE-V}	3.1×10^{-6}	2.7×10^{-6}	1.6×10^{-6}	1.0×10^{-6}	6.2×10^{-7}	4.9×10^{-7}
	Error bound	$[3.3 \times 10^{-6},$	$[2.9 \times 10^{-6},$	$[1.7 \times 10^{-6},$	$[1.1 \times 10^{-6},$	$[7.0 \times 10^{-7},$	$[5.5 \times 10^{-7},$
	$[D_{min}, D_{max}]$ (cm ² /s)	$2.8 \times 10^{-6}]$	$2.5 \times 10^{-6}]$	$1.4 \times 10^{-6}]$	$9.1 \times 10^{-7}]$	$5.5 \times 10^{-7}]$	$4.3 \times 10^{-7}]$
	SNAP _{RE-V}	4.0×10^{-6}	3.7×10^{-6}	3.1×10^{-6}	2.3×10^{-6}	1.1×10^{-6}	1.0×10^{-6}
	Error bound	$[4.4 \times 10^{-6},$	$[4.1 \times 10^{-6},$	$[3.5 \times 10^{-6},$	$[2.5 \times 10^{-6},$	$[1.3 \times 10^{-6},$	$[1.2 \times 10^{-6},$
	$[D_{min}, D_{max}]$ (cm ² /s)	$3.6 \times 10^{-6}]$	$3.4 \times 10^{-6}]$	$2.7 \times 10^{-6}]$	$2.0 \times 10^{-6}]$	$9.6 \times 10^{-7}]$	$8.4 \times 10^{-7}]$
	MTP _{RE-V}	1.6×10^{-6}	9.6×10^{-7}	3.1×10^{-7}	4.3×10^{-8}	N/A	N/A
	Error bound	$[1.7 \times 10^{-6},$	$[1.1 \times 10^{-6},$	$[3.6 \times 10^{-7},$	$[6.1 \times 10^{-8},$		
	$[D_{min}, D_{max}]$ (cm ² /s)	$1.4 \times 10^{-6}]$	$8.6 \times 10^{-7}]$	$2.6 \times 10^{-7}]$	$2.6 \times 10^{-8}]$		
	Interstitial-						
	GAP _{RE-I}	2.4×10^{-6}	2.3×10^{-6}	1.6×10^{-6}	1.0×10^{-6}	6.5×10^{-7}	4.6×10^{-7}

enhanced MLPs	Error bound	[2.7*10 ⁻⁶ ,	[2.7*10 ⁻⁶ ,	[1.9*10 ⁻⁶ ,	[1.2*10 ⁻⁶ ,	[7.9*10 ⁻⁷ ,	[5.8*10 ⁻⁷ ,
	[D _{min} , D _{max}] (cm ² /s)	2.1*10 ⁻⁶]	2.0*10 ⁻⁶]	1.4*10 ⁻⁶]	8.3*10 ⁻⁷]	5.1*10 ⁻⁷]	3.5*10 ⁻⁷]
	NNP _{RE-I}	Melt	Melt	1.3*10 ⁻⁶	7.8*10 ⁻⁷	2.7*10 ⁻⁷	1.7*10 ⁻⁷
	Error bound			[1.6*10 ⁻⁶ ,	[9.2*10 ⁻⁷ ,	[3.5*10 ⁻⁷ ,	[2.3*10 ⁻⁷ ,
	[D _{min} , D _{max}] (cm ² /s)			1.1*10 ⁻⁶]	6.4*10 ⁻⁷]	2.0*10 ⁻⁷]	1.1*10 ⁻⁷]
	DeePMD _{RE-I}	Melt	Melt	7.4*10 ⁻⁶	1.1*10 ⁻⁵	4.3*10 ⁻⁶	3.5*10 ⁻⁶
	Error bound			[7.9*10 ⁻⁶ ,	[1.1*10 ⁻⁵ ,	[4.5*10 ⁻⁶ ,	[3.7*10 ⁻⁶ ,
	[D _{min} , D _{max}] (cm ² /s)			6.9*10 ⁻⁶]	1.0*10 ⁻⁵]	4.0*10 ⁻⁶]	3.2*10 ⁻⁶]
	DeepPot-SE _{RE-I}	6.9*10 ⁻⁶	2.8*10 ⁻⁶	2.2*10 ⁻⁶	1.8*10 ⁻⁶	1.0*10 ⁻⁶	9.7*10 ⁻⁷
	Error bound	[7.3*10 ⁻⁶ ,	[3.1*10 ⁻⁶ ,	[2.4*10 ⁻⁶ ,	[2.0*10 ⁻⁶ ,	[1.1*10 ⁻⁶ ,	[1.1*10 ⁻⁶ ,
	[D _{min} , D _{max}] (cm ² /s)	6.5*10 ⁻⁶]	2.6*10 ⁻⁶]	2.1*10 ⁻⁶]	1.7*10 ⁻⁶]	9.3*10 ⁻⁷]	8.8*10 ⁻⁷]
	SNAP _{RE-I}	Melt	Melt	2.1*10 ⁻⁶	1.2*10 ⁻⁶	3.6*10 ⁻⁷	4.3*10 ⁻⁷
	Error bound			[2.4*10 ⁻⁶ ,	[1.3*10 ⁻⁶ ,	[4.5*10 ⁻⁷ ,	[5.4*10 ⁻⁷ ,
	[D _{min} , D _{max}] (cm ² /s)			1.8*10 ⁻⁶]	9.7*10 ⁻⁷]	2.7*10 ⁻⁷]	3.2*10 ⁻⁷]
	MTP _{RE-I}	3.1*10 ⁻⁶	2.4*10 ⁻⁶	1.6*10 ⁻⁶	1.1*10 ⁻⁶	4.9*10 ⁻⁷	3.2*10 ⁻⁷
	Error bound	[3.4*10 ⁻⁶ ,	[2.6*10 ⁻⁶ ,	[1.7*10 ⁻⁶ ,	[1.2*10 ⁻⁶ ,	[5.6*10 ⁻⁷ ,	[3.8*10 ⁻⁷ ,
	[D _{min} , D _{max}] (cm ² /s)	2.9*10 ⁻⁶]	2.2*10 ⁻⁶]	1.4*10 ⁻⁶]	9.5*10 ⁻⁷]	4.2*10 ⁻⁷]	2.7*10 ⁻⁷]

Table C5. The diffusivities at 1600 K, 1500K, 1230 K, 1000 K, 840 K, and 730 K for vacancy (superscript V) from AIMD and MLP-MD simulations. The MD simulations with melting structures (noted as melt) and not enough hopping events (noted as N/A) are marked.

Models		D_{1600K}^V (cm ² /s)	D_{1500K}^V (cm ² /s)	D_{1230K}^V (cm ² /s)	D_{1000K}^V (cm ² /s)	D_{840K}^V (cm ² /s)	D_{730K}^V (cm ² /s)
AIMD	K2	1.3×10^{-6}	1.4×10^{-6}	1.2×10^{-6}	5.2×10^{-7}	N/A	N/A
	Error bound	$[1.6 \times 10^{-6}, 9.9 \times 10^{-7}]$	$[1.7 \times 10^{-6}, 1.0 \times 10^{-6}]$	$[1.5 \times 10^{-6}, 8.2 \times 10^{-7}]$	$[6.9 \times 10^{-7}, 3.6 \times 10^{-7}]$		
	K1	1.3×10^{-6}	Melt	4.5×10^{-7}	2.2×10^{-7}	1.1×10^{-7}	N/A
	Error bound	$[1.8 \times 10^{-6}, 8.4 \times 10^{-7}]$		$[5.7 \times 10^{-7}, 3.3 \times 10^{-7}]$	$[2.8 \times 10^{-7}, 1.6 \times 10^{-7}]$	$[1.6 \times 10^{-7}, 6.4 \times 10^{-8}]$	
Known MLPs trained from previous studies	GAP	1.5×10^{-6}	Melt	8.1×10^{-7}	5.8×10^{-7}	3.3×10^{-7}	2.4×10^{-7}
	Error bound	$[1.8 \times 10^{-6}, 1.2 \times 10^{-6}]$		$[1.0 \times 10^{-6}, 6.1 \times 10^{-7}]$	$[7.1 \times 10^{-7}, 4.5 \times 10^{-7}]$	$[4.1 \times 10^{-7}, 2.4 \times 10^{-7}]$	$[3.2 \times 10^{-7}, 1.7 \times 10^{-7}]$
	GAP _{PRX}	1.5×10^{-6}	1.0×10^{-6}	1.2×10^{-6}	5.6×10^{-7}	N/A	N/A
	Error bound	$[1.9 \times 10^{-6}, 1.1 \times 10^{-6}]$	$[1.3 \times 10^{-6}, 7.2 \times 10^{-7}]$	$[1.5 \times 10^{-6}, 8.0 \times 10^{-7}]$	$[7.8 \times 10^{-7}, 3.3 \times 10^{-7}]$		
	GAP _{RPA}	Melt	Melt	Melt	N/A	N/A	N/A
	Error bound						
	GAP _{SCAN}	N/A	N/A	N/A	N/A	N/A	N/A
	Error bound						
	[D _{min} , D _{max}] (cm ² /s)						
	NNP	1.5×10^{-6}	Melt	9.2×10^{-7}	6.7×10^{-7}	4.4×10^{-7}	2.9×10^{-7}
	Error bound	$[1.8 \times 10^{-6}, 1.2 \times 10^{-6}]$		$[1.1 \times 10^{-6}, 7.0 \times 10^{-7}]$	$[7.3 \times 10^{-7}, 6.1 \times 10^{-7}]$	$[5.0 \times 10^{-7}, 3.8 \times 10^{-7}]$	$[3.4 \times 10^{-7}, 2.5 \times 10^{-7}]$
	DeePMD	Melt	1.2×10^{-5}	5.7×10^{-6}	2.8×10^{-6}	4.9×10^{-6}	2.5×10^{-6}
	Error bound		$[1.3 \times 10^{-5}, 1.2 \times 10^{-5}]$	$[6.1 \times 10^{-6}, 5.4 \times 10^{-6}]$	$[3.0 \times 10^{-6}, 2.5 \times 10^{-6}]$	$[5.3 \times 10^{-6}, 4.5 \times 10^{-6}]$	$[2.6 \times 10^{-6}, 2.3 \times 10^{-6}]$
	DeepPot-SE	1.8×10^{-6}	1.8×10^{-6}	1.4×10^{-6}	9.1×10^{-7}	5.4×10^{-7}	5.0×10^{-7}
	Error bound	$[2.0 \times 10^{-6}, 1.7 \times 10^{-6}]$	$[1.9 \times 10^{-6}, 1.6 \times 10^{-6}]$	$[1.6 \times 10^{-6}, 1.3 \times 10^{-6}]$	$[1.0 \times 10^{-6}, 8.0 \times 10^{-7}]$	$[6.1 \times 10^{-7}, 4.7 \times 10^{-7}]$	$[5.7 \times 10^{-7}, 4.3 \times 10^{-7}]$
	DeePMD (original)	Melt	Melt	9.5×10^{-6}	1.1×10^{-5}	3.8×10^{-6}	4.3×10^{-6}
	Error bound			$[1.0 \times 10^{-5}, 8.9 \times 10^{-6}]$	$[1.1 \times 10^{-5}, 9.9 \times 10^{-6}]$	$[4.0 \times 10^{-6}, 3.5 \times 10^{-6}]$	$[4.5 \times 10^{-6}, 4.0 \times 10^{-6}]$
	DeepPot-SE (original)	Melt	Melt	Melt	Melt	Melt	Melt
	Error bound						
	[D _{min} , D _{max}] (cm ² /s)						
	MTP	1.8×10^{-6}	1.7×10^{-6}	9.7×10^{-7}	6.0×10^{-7}	3.6×10^{-7}	2.3×10^{-7}
	Error bound	$[2.0 \times 10^{-6}, 1.7 \times 10^{-6}]$	$[1.9 \times 10^{-6}, 1.6 \times 10^{-6}]$	$[1.1 \times 10^{-6}, 8.6 \times 10^{-7}]$	$[6.8 \times 10^{-7}, 5.2 \times 10^{-7}]$	$[4.2 \times 10^{-7}, 3.0 \times 10^{-7}]$	$[2.8 \times 10^{-7}, 1.8 \times 10^{-7}]$
Vacancy-enhanced MLPs	SNAP	2.1×10^{-6}	Melt	1.2×10^{-6}	7.8×10^{-7}	6.3×10^{-7}	5.0×10^{-7}
	Error bound	$[2.5 \times 10^{-6}, 1.8 \times 10^{-6}]$		$[1.4 \times 10^{-6}, 9.2 \times 10^{-7}]$	$[8.5 \times 10^{-7}, 7.1 \times 10^{-7}]$	$[7.0 \times 10^{-7}, 5.6 \times 10^{-7}]$	$[5.5 \times 10^{-7}, 4.3 \times 10^{-7}]$
	GAP _{RE-V}	1.6×10^{-6}	1.4×10^{-6}	9.1×10^{-7}	3.9×10^{-7}	3.0×10^{-7}	2.4×10^{-7}
	Error bound	$[1.8 \times 10^{-6}, 1.4 \times 10^{-6}]$	$[1.6 \times 10^{-6}, 1.2 \times 10^{-6}]$	$[1.1 \times 10^{-6}, 7.6 \times 10^{-7}]$	$[4.7 \times 10^{-7}, 3.0 \times 10^{-7}]$	$[3.7 \times 10^{-7}, 2.3 \times 10^{-7}]$	$[3.0 \times 10^{-7}, 1.7 \times 10^{-7}]$
	NNP _{RE-V}	2.4×10^{-6}	1.6×10^{-6}	1.2×10^{-6}	9.4×10^{-7}	5.3×10^{-7}	3.4×10^{-7}
	Error bound	$[2.7 \times 10^{-6}, 2.1 \times 10^{-6}]$	$[1.8 \times 10^{-6}, 1.3 \times 10^{-6}]$	$[1.3 \times 10^{-6}, 9.8 \times 10^{-7}]$	$[1.1 \times 10^{-6}, 7.7 \times 10^{-7}]$	$[6.5 \times 10^{-7}, 4.1 \times 10^{-7}]$	$[4.3 \times 10^{-7}, 2.5 \times 10^{-7}]$
	DeePMD _{RE-V}	Melt	Melt	4.8×10^{-6}	3.5×10^{-6}	1.3×10^{-6}	5.7×10^{-6}
	Error bound			$[5.1 \times 10^{-6}, 4.5 \times 10^{-6}]$	$[3.7 \times 10^{-6}, 3.2 \times 10^{-6}]$	$[1.4 \times 10^{-6}, 1.2 \times 10^{-6}]$	$[6.1 \times 10^{-6}, 5.3 \times 10^{-6}]$
	DeepPot-SE _{RE-V}	Melt	1.5×10^{-6}	1.0×10^{-6}	7.5×10^{-7}	6.7×10^{-7}	5.1×10^{-7}
	Error bound		$[1.6 \times 10^{-6}, 1.3 \times 10^{-6}]$	$[1.1 \times 10^{-6}, 9.4 \times 10^{-7}]$	$[8.3 \times 10^{-7}, 6.7 \times 10^{-7}]$	$[7.4 \times 10^{-7}, 6.0 \times 10^{-7}]$	$[5.7 \times 10^{-7}, 4.5 \times 10^{-7}]$
	SNAP _{RE-V}	Melt	1.4×10^{-6}	1.1×10^{-6}	5.7×10^{-7}	3.1×10^{-7}	2.7×10^{-7}
	Error bound		$[1.6 \times 10^{-6}, 1.2 \times 10^{-6}]$	$[1.2 \times 10^{-6}, 9.1 \times 10^{-7}]$	$[7.0 \times 10^{-7}, 4.4 \times 10^{-7}]$	$[3.9 \times 10^{-7}, 2.3 \times 10^{-7}]$	$[3.5 \times 10^{-7}, 1.9 \times 10^{-7}]$
	MTP _{RE-V}	1.7×10^{-6}	1.7×10^{-6}	1.1×10^{-6}	8.0×10^{-7}	4.7×10^{-7}	3.5×10^{-7}
	Error bound	$[1.9 \times 10^{-6}, 1.5 \times 10^{-6}]$	$[1.8 \times 10^{-6}, 1.5 \times 10^{-6}]$	$[1.2 \times 10^{-6}, 9.4 \times 10^{-7}]$	$[9.0 \times 10^{-7}, 7.0 \times 10^{-7}]$	$[5.4 \times 10^{-7}, 4.0 \times 10^{-7}]$	$[4.1 \times 10^{-7}, 2.9 \times 10^{-7}]$
	GAP _{RE-I}	1.9×10^{-6}	1.8×10^{-6}	1.3×10^{-6}	9.7×10^{-7}	5.5×10^{-7}	5.0×10^{-7}
Interstitial-							

enhanced MLPs	Error bound	$[2.2*10^{-6}, 1.6*10^{-6}]$	$[2.1*10^{-6}, 1.5*10^{-6}]$	$[1.6*10^{-6}, 1.1*10^{-6}]$	$[1.2*10^{-6}, 7.9*10^{-7}]$	$[6.7*10^{-7}, 4.2*10^{-7}]$	$[6.2*10^{-7}, 3.9*10^{-7}]$
	$[D_{\min}, D_{\max}]$ (cm ² /s)						
	NNP _{RE-I}	Melt	Melt	$8.4*10^{-7}$	N/A	$5.9*10^{-7}$	$5.0*10^{-7}$
	Error bound			$[9.9*10^{-7}, 6.8*10^{-7}]$		$[7.2*10^{-7}, 4.5*10^{-7}]$	$[6.2*10^{-7}, 3.9*10^{-7}]$
	$[D_{\min}, D_{\max}]$ (cm ² /s)						
	DeePMD _{RE-I}	Melt	Melt	$7.0*10^{-6}$	Melt	$6.4*10^{-6}$	$2.9*10^{-6}$
	Error bound			$[7.4*10^{-6}, 6.6*10^{-6}]$		$[6.9*10^{-6}, 6.0*10^{-6}]$	$[3.1*10^{-6}, 2.7*10^{-6}]$
	$[D_{\min}, D_{\max}]$ (cm ² /s)						
	DeepPot-SE _{RE-I}	Melt	$1.6*10^{-6}$	$8.3*10^{-7}$	$6.2*10^{-7}$	$4.5*10^{-7}$	$2.8*10^{-7}$
	Error bound		$[1.8*10^{-6}, 1.5*10^{-6}]$	$[9.2*10^{-7}, 7.5*10^{-7}]$	$[6.9*10^{-7}, 5.4*10^{-7}]$	$[5.1*10^{-7}, 4.0*10^{-7}]$	$[3.2*10^{-7}, 2.3*10^{-7}]$
	$[D_{\min}, D_{\max}]$ (cm ² /s)						
	SNAP _{RE-I}	Melt	Melt	$6.7*10^{-7}$	Melt	$4.0*10^{-7}$	$2.5*10^{-7}$
	Error bound			$[8.0*10^{-7}, 5.3*10^{-7}]$		$[5.0*10^{-7}, 3.0*10^{-7}]$	$[3.3*10^{-7}, 1.7*10^{-7}]$
	$[D_{\min}, D_{\max}]$ (cm ² /s)						
	MTP _{RE-I}	$2.2*10^{-6}$	$1.8*10^{-6}$	$1.5*10^{-6}$	$1.2*10^{-6}$	$1.0*10^{-6}$	$7.4*10^{-7}$
	Error bound	$[2.4*10^{-6}, 2.0*10^{-6}]$	$[2.0*10^{-6}, 1.6*10^{-6}]$	$[1.6*10^{-6}, 1.3*10^{-6}]$	$[1.3*10^{-6}, 1.1*10^{-6}]$	$[1.1*10^{-6}, 9.0*10^{-7}]$	$[8.3*10^{-7}, 6.4*10^{-7}]$
	$[D_{\min}, D_{\max}]$ (cm ² /s)						

Table C6. The additional diffusivities at 1333 K and 1080K for single-interstitials and single-vacancies from AIMD simulations.

	AIMD	K2	K1
Interstitial	D_{1333K}^I (cm ² /s)	$1.5*10^{-6}$	$8.7*10^{-7}$
	Error bound	$[1.7*10^{-6}, 1.2*10^{-6}]$	$[1.0*10^{-6}, 7.4*10^{-7}]$
	$[D_{\min}, D_{\max}]$ (cm ² /s)		
	D_{1080K}^I (cm ² /s)	$1.0*10^{-6}$	$5.4*10^{-7}$
	Error bound	$[1.2*10^{-6}, 8.0*10^{-7}]$	$[6.4*10^{-7}, 4.4*10^{-7}]$
	$[D_{\min}, D_{\max}]$ (cm ² /s)		
Vacancy	D_{1333K}^V (cm ² /s)	$1.2*10^{-6}$	$3.7*10^{-7}$
	Error bound	$[1.4*10^{-6}, 9.4*10^{-7}]$	$[4.5*10^{-7}, 2.9*10^{-7}]$
	$[D_{\min}, D_{\max}]$ (cm ² /s)		
	D_{1080K}^V (cm ² /s)	$7.9*10^{-7}$	$2.5*10^{-7}$
	Error bound	$[9.7*10^{-7}, 6.1*10^{-7}]$	$[3.2*10^{-7}, 1.8*10^{-7}]$
	$[D_{\min}, D_{\max}]$ (cm ² /s)		

5. Analysis of Si interstitials

The comparison of E_f predicted by all MLPs and DFT calculations are summarized in **Figure C7a**. As shown in the main text, existing MLPs poorly predict the formation energies E_f (and relative trends) of different interstitial configurations and their occurrence frequencies in MD simulations (Figure 3.3d). The interstitial-enhanced MLPs show significantly improved prediction of the occurrence frequencies of different interstitials as shown (Figure C7b). However, RE-enhanced MLPs still cannot fully predict of E_f in enhanced MLPs. Most RE-enhanced MLPs still give higher E_f of DFT-ground-state split- $\langle 110 \rangle$ than those of tetrahedral, except for DeepPot-SE_{RE-I}. In addition, the original and RE-enhanced MLPs generally underestimate E_f by around 1 eV. GAP, GAP_{RPA}, SNAP, MTP, GAP_{RE-V}, SNAP_{RE-V}, MTP_{RE-V}, GAP_{RE-I}, NNP_{RE-I}, DeepPot-SE_{RE-I}, and SNAP_{RE-I} predict E_f of all three interstitial configurations to be less than 3 eV, compared to the E_f range of 3.5 – 3.8 eV calculated by DFT K4. As a comparison to MLPs, DFT K1 gives lower E_f range of DFT K1 between 2.3 – 2.8 eV but reproduces the correct trend of E_f among three configurations, with the split- $\langle 110 \rangle$ as the lowest and the tetrahedral as the highest. The result suggests low accuracy DFT still performs better than MLPs on predicting complex defect configurations.

In most of our MD simulations, we also note that the lower formation energies don't always lead to higher occurrence frequencies. For example, the occurrence frequency of split- $\langle 110 \rangle$ in the MD simulation by GAP is higher than that of hexagonal (Figure C7b), while the predicted E_f of hexagonal (2.1 eV) is lower than that of split- $\langle 110 \rangle$ (2.3 eV) (Figure C7a). Similarly, the predicted E_f of split- $\langle 110 \rangle$ and hexagonal are the same (2.85 eV) for GAP_{RE-I}, while the occurrence frequency of split- $\langle 110 \rangle$ is higher than that of hexagonal in GAP_{RE-I}-MD simulations. This may be caused by one or multiple of the following reasons. First, in addition to the formation

energy E_f , the occurrence frequency may also be affected by the entropy of different interstitials. Second, given the different interstitials are formed by the transition among them during MD simulations, the occurrence frequencies are influenced by their energy barriers, which is not identical to the formation energy E_f . Nonetheless, lower formation energy E_f is still highly correlated with higher occurrence frequencies. We compare the E_f and occurrence frequencies for each pair of interstitial configurations and count the number of pairs that have a consistent trend of increasing E_f and decreasing occurrence frequencies, and 73% of the interstitial pairs have the correct trend for all MLPs (Figure 3.3), and 61% for existing MLPs and 87% for RE-enhanced MLPs. Thus, the formation energy E_f may largely account for the occurrence frequencies of different interstitials in MD simulations.

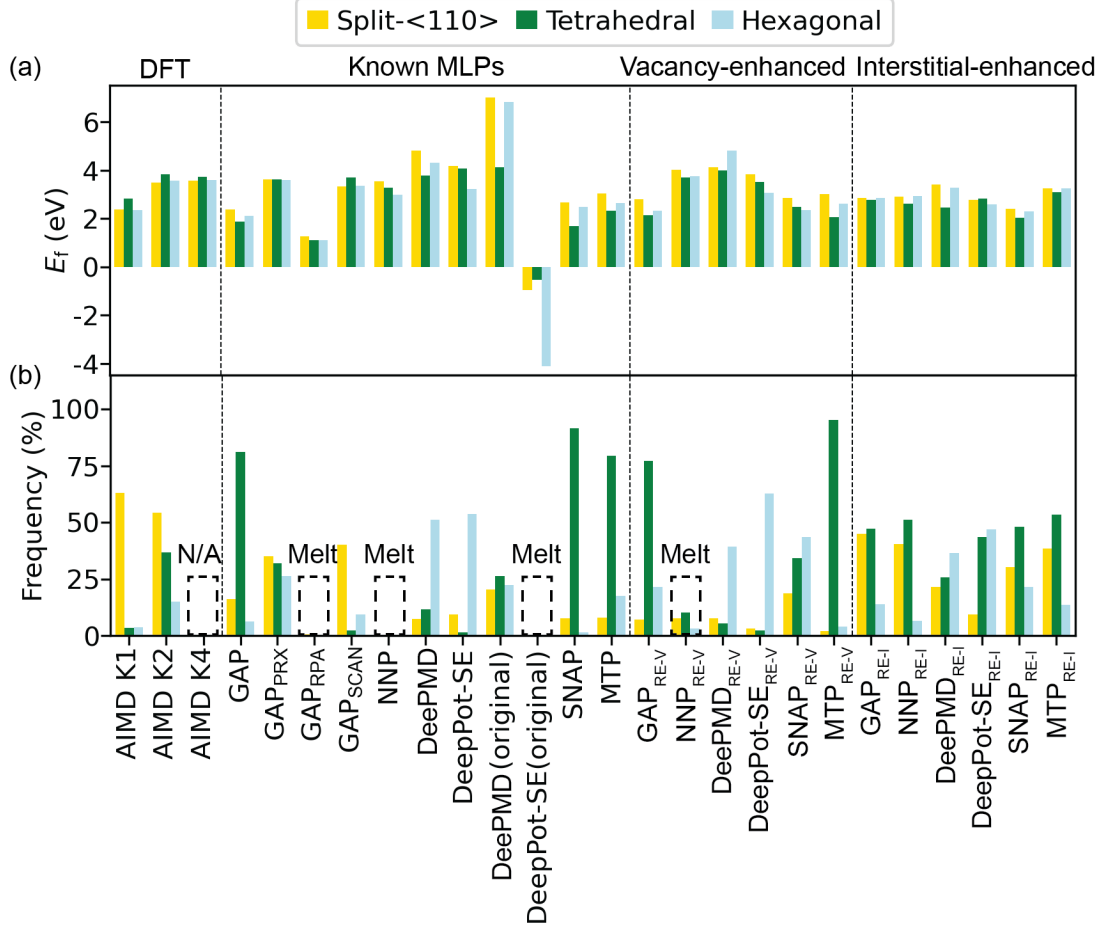


Figure C7. Comparison of DFT, known MLPs, and RE-enhanced MLPs, for (a) the formation energies E_f and (b) occurrence frequencies in MD simulations of different interstitial configurations. The occurrence frequency for AIMD K4 is unavailable (N/A) due to the high computation cost to run AIMD simulations with $4 \times 4 \times 4$ k -point mesh.

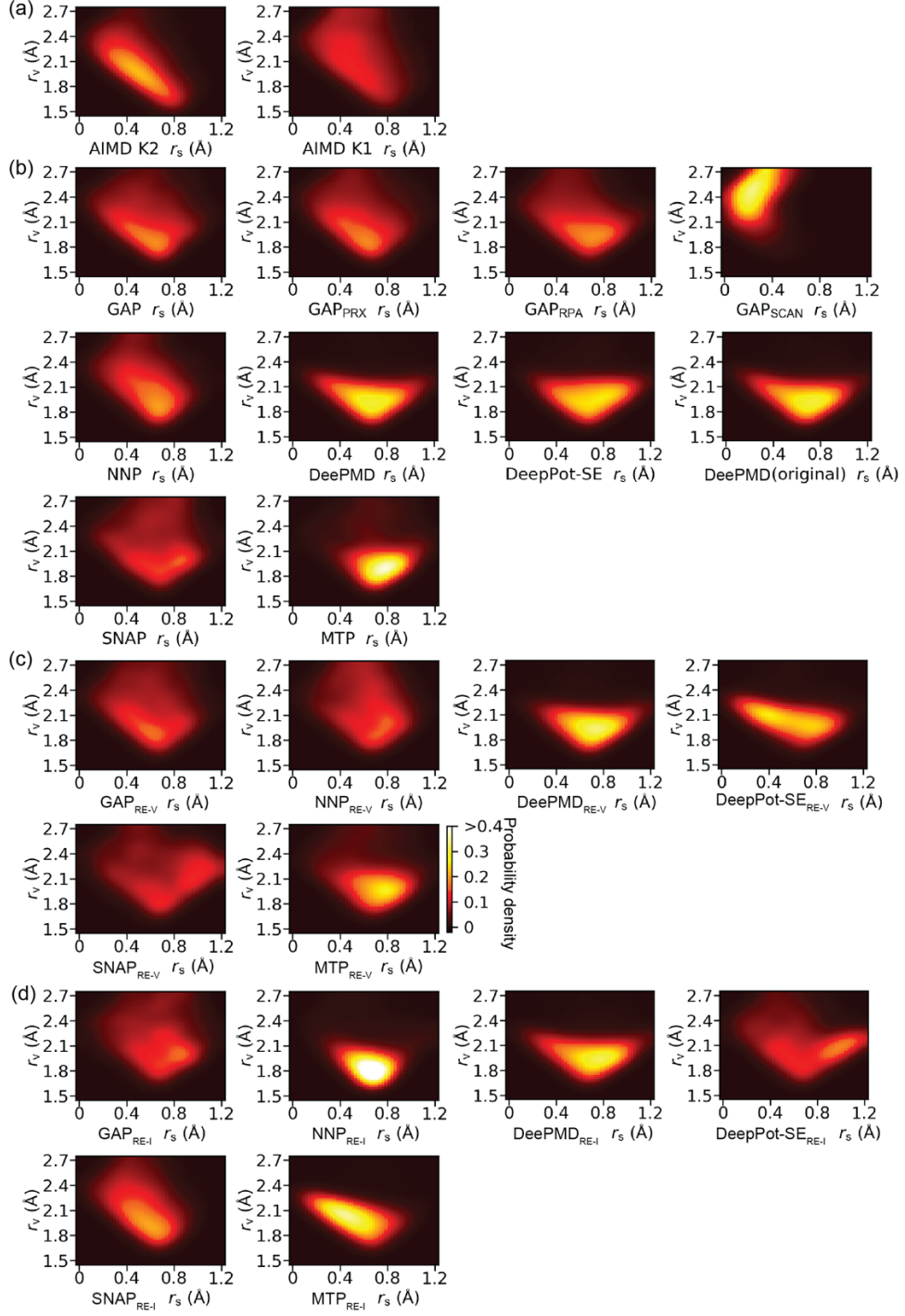


Figure C8. The distributions of r_v versus r_s of the vacancy nearest neighbor (vacancy-NN) atoms from the MD simulations with single vacancy by (a) AIMD simulations, (b) original MLPs, (c) vacancy-enhanced MLPs, and (d) interstitial-enhanced MLPs, in addition to Figure 3.3.

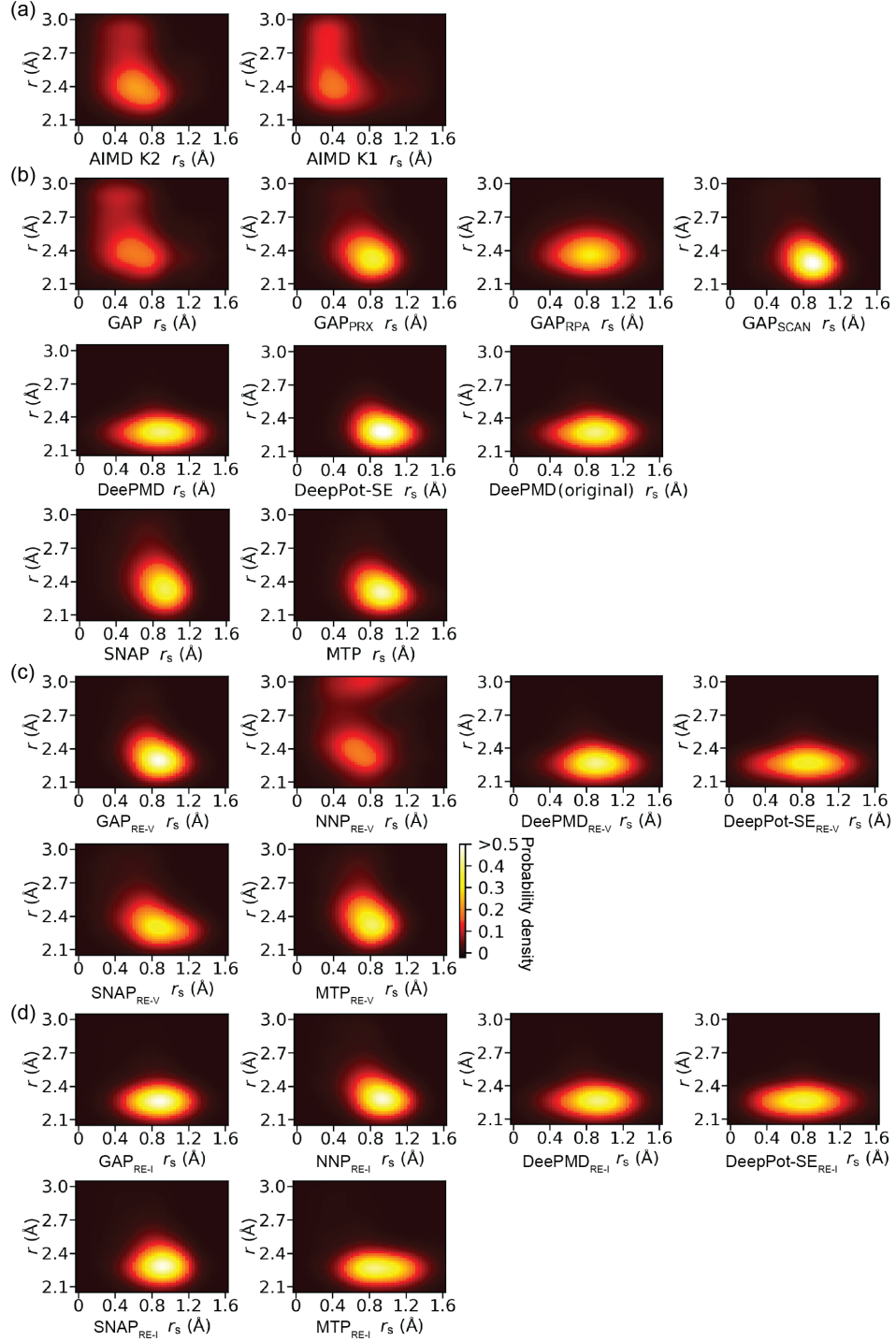


Figure C9. The distributions of r versus r_s of the interstitial nearest neighbor (interstitial-NN) atoms from the MD simulations with single interstitial defect by (a) AIMD simulations, (b) original MLPs, (c) vacancy-enhanced MLPs, and (d) interstitial-enhanced MLPs, following the same plotting protocols (Appendix D) (Figure 3.3).

6. Analysis of potential energy landscapes

Figure C10 shows the MLP-predicted potential energy surfaces (PESs) of a Si atom moving toward the nearest-neighbor vacancy same as Figure 3.4a and f. The predicted PESs of GAP_{RPA} , GAP_{SCAN} , DeePMD, DeepPot-SE, DeePMD (original), and DeepPot-SE (original) deviate significantly from DFT calculations (Figure C10b, c, and e), suggesting poorly reproduced PESs by MLPs. In addition, we also test the PES of a Si atom moving from its initial position in the snapshot towards the static site position in a relaxed cell, around a vacancy in a snapshot randomly taken from AIMD simulation with single vacancy defect (**Figure C11**). Interestingly, these PESs agree well with DFT calculations (Figure C11b, c, and e). The results of different PESs suggests that MLPs may potentially show large discrepancies when predicting PESs with similar but slightly distorted atomic configurations.

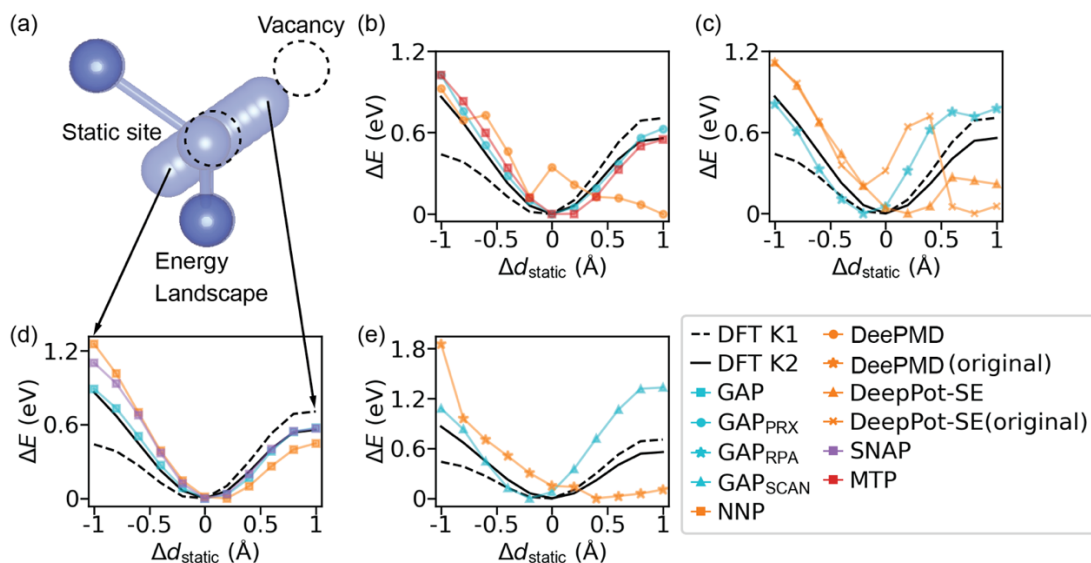


Figure C10. Comparison of the PESs predicted by known MLPs and DFT calculations on (a) a single Si atom around a static site moving towards the vacancy in a static relaxed configuration, the same as the main text Figure 3.3a.

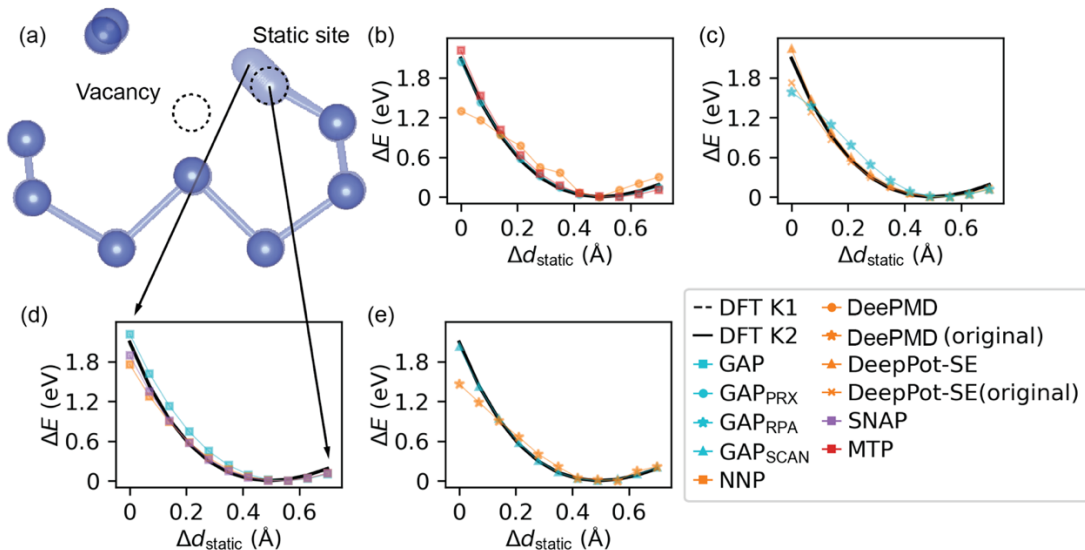


Figure C11. Comparison of the PESs predicted by known MLPs and DFT calculations on (a) a vacancy NN atom moving towards the static relaxed position, in a snapshot taken from AIMD simulations with single vacancy.

7. Analysis of forces with large errors

Following the same methods illustrated in main text Chapter 3.4.1, here we analyze the errors of atomic forces on all atoms in interstitial-RE and vacancy-RE testing sets and compare the force errors among three categories of atoms, such as RE migrating atoms ($r = 0 \text{ \AA}$), NN atoms of RE migrating atoms ($r = 0 - 3 \text{ \AA}$), and other non-NN atoms ($r > 3 \text{ \AA}$) (**Table C7**). We count the fraction of those atoms having large force errors with either the magnitude errors of atomic forces $\delta_F > 0.5 \text{ eV/\AA}$ or directional errors of atomic forces $\delta_\theta > 15^\circ$. The fraction of atoms with large force errors are substantially higher in known MLPs than DFT calculations (below 2% for DFT K2) in all three categories. For known MLPs, 42% - 97% of RE migrating atoms have large force errors in interstitial-RE testing set. For vacancy-RE testing set, 31% to 64% of RE migrating atoms have large force errors by most MLPs, except for GAP, GAP_{PRX}, and MTP, which are similar to or better than DFT K1. Therefore, atomic force predictions of most MLPs are prone to show large errors than even low-accuracy DFT calculations.

We compare known MLPs for the cumulative density functions (CDFs) of force errors on RE migrating atoms (**Figure C12a and b**) using the same approach as in the main text (Figure 3.5d, e, i, and j), in interstitial-RE and vacancy-RE testing sets. We also perform the same analyses on NN atoms around RE (Figure C12c and d) and on all atoms (Figure C12e and f). Known MLPs from previous studies clearly underperform DFT calculations for force predictions on RE migrating atoms. The CDFs of most MLPs on RE-NN atoms are below the CDF curves of DFT K1, except for GAP_{PRX} and MTP (Figure C12c and d).

In contrast to RE or RE-NN atoms, the CDFs of force errors on all atoms show better agreement with DFT calculations (Figure C12e and f). Particularly, the CDFs of GAP, GAP_{PRX}, and MTP on all atoms are similar to or higher than the DFT K1 CDF curve in interstitial-RE testing set. The similar CDFs of force errors for all atoms are in consistent with the low RMSEs of force errors given by conventional error testing, because averaged total errors are evaluated on the dataset with a large fraction of near equilibrium atomistic configurations with small errors.

Table C7. Fraction of atoms with large errors on atomic forces for three categories of atoms (RE migrating atoms, NN atoms of RE migrating atoms, and other non-NN atoms) in interstitial-RE testing set and in vacancy-RE testing set. Those atoms having large force errors with force magnitude errors $\delta_F > 0.5 \text{ eV/\AA}$ or force direction errors $\delta_\theta > 15^\circ$ are counted as large errors.

Models	Interstitial-RE Testing Set			Vacancy-RE Testing Set		
	RE $r = 0 \text{ \AA}$ (%)	RE-NN $0 - 3 \text{ \AA}$ (%)	Non-NN $r > 3 \text{ \AA}$ (%)	RE $r = 0 \text{ \AA}$ (%)	RE-NN $0 - 3 \text{ \AA}$ (%)	Non-NN $r > 3 \text{ \AA}$ (%)
DFT K1	33	17	5	21	8	5
DFT K2	1	0.4	0	0	0.6	0
GAP	58	23	4	21	12	4
GAP _{PRX}	42	16	2.5	13	6	3
GAP _{RPA}	97	59	47	64	60	47
GAP _{SCAN}	74	40	22	48	36	25
NNP	73	36	14	31	25	15
DeePMD	86	53	21	45	35	23
DeepPot-SE	93	59	14	39	27	11
DeePMD (original)	86	58	7	43	23	11
DeepPot-SE (original)	93	89	30	67	59	29
MTP	49	15	2	11	6	3
SNAP	68	44	19	34	26	19

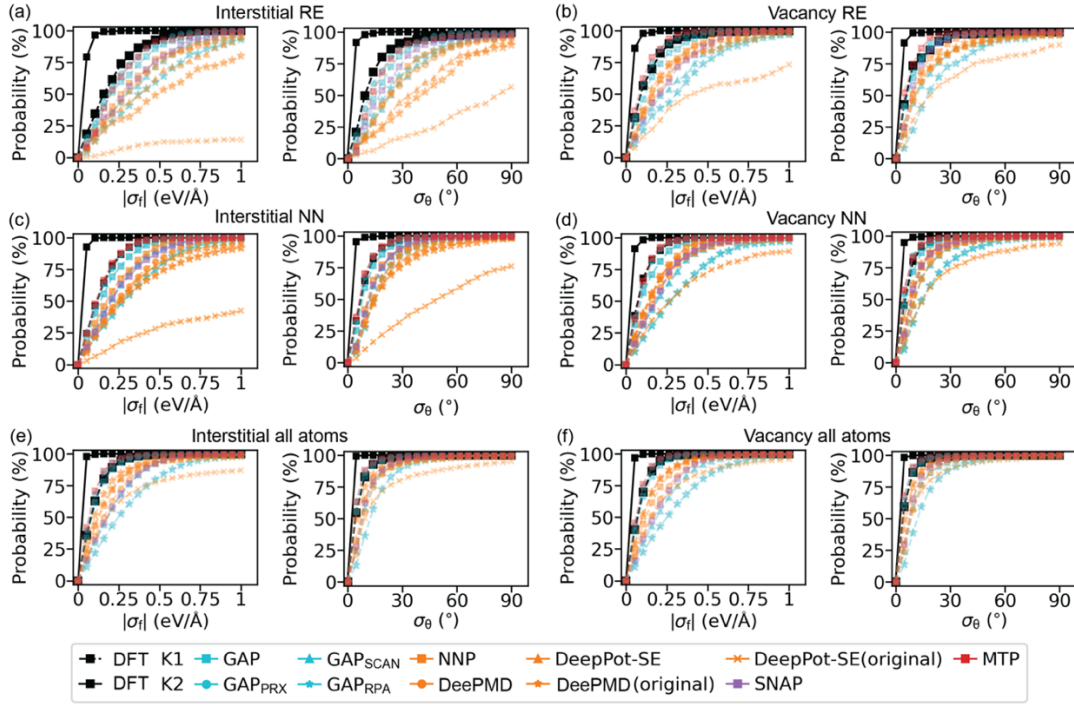


Figure C12. The CDFs of force errors δ_F and δ_θ on (a) interstitial RE, (b) vacancy RE, (c) interstitial RE-NN, (d) vacancy RE-NN, (e) all atoms in interstitial-RE testing set, and (f) all atoms in vacancy-RE testing set of DFT calculations and known MLPs.

8. K-means clustering analysis

We here employ the clustering algorithms to automate the process of identifying RE atoms from the dataset encompassing snapshots from MD simulations. We apply the k-means clustering technique using the SOAP descriptor to identify RE atoms (Appendix D). In the $\mathcal{D}_{\text{RE-I}}^{\text{Testing}}$ and $\mathcal{D}_{\text{RE-V}}^{\text{Testing}}$ datasets, the projections of SOAP show three clusters (**Figure C13**), clearly distinguishing RE atoms (migrating vacancies and interstitials) from non-RE atoms (**Table C8**). In addition, for a range of hyperparameters tested, this clustering approach is effective in identifying RE atoms. We trained three k-means clustering models, an all-RE model, an interstitial-RE model, and a vacancy-RE model, using different datasets to identify RE atoms in simulations with single vacancy or interstitial, described as follows.

The all-RE k-means model is employed to distinguish RE atoms from non-RE atoms in a dataset of 12800 atomic environments from snapshots in the $\mathcal{D}_{\text{RE-I}}^{\text{Testing}}$ and $\mathcal{D}_{\text{RE-V}}^{\text{Testing}}$ dataset. The SOAP descriptors of atomic environments, which are calculated using QUIP package,^[107] had 325 features and the same hyperparameters as Ref. ^[1]. The k-means clustering is performed using *scikit-learn* package, with the maximal iteration of 1000, a convergence tolerance of 10^{-4} , and a number of clusters between three to eight (Table C8). The results of the k-means models as shown in Figure C13 are visualized by projecting onto a 2-dimensional plane using the t-distributed stochastic neighbor embedding (TSNE) model in *scikit-learn* package^[108]. The hyperparameter perplexity in TSNE model was set to be 50. The visualization in Figure C13 is for a subset of data points consisting of 520 atomistic environments from four categories (130 each), such as RE atoms in $\mathcal{D}_{\text{RE-V}}^{\text{Testing}}$, RE atoms in $\mathcal{D}_{\text{RE-I}}^{\text{Testing}}$, non-RE atoms in $\mathcal{D}_{\text{RE-I}}^{\text{Testing}}$ and $\mathcal{D}_{\text{RE-V}}^{\text{Testing}}$.

The interstitial-RE k-means clustering model to distinguish interstitials is trained by using 6500 atomic environments from $\mathcal{D}_{\text{RE-I}}^{\text{Testing}}$ set, following the same protocols described above (the

cluster number was set to 3). The 3rd k-means clustering model (vacancy-RE k-means clustering model) is trained from 6300 atomic environments taken from $\mathcal{D}_{\text{RE-V}}^{\text{Testing}}$ following the same training procedure of interstitial k-means model.

Here we show the precisions and recalls of the k-means clustering models, including the general all-RE k-means models trained by the $\mathcal{D}_{\text{RE-I}}^{\text{Testing}}$ and $\mathcal{D}_{\text{RE-V}}^{\text{Testing}}$ datasets (Table C8), the interstitial-RE k-means models (**Table C9**), and the vacancy-RE k-means models (**Table C10**) described above. Our results show that the efficiencies of most k-means models deviate less than 5% for a range of cluster number parameters. In addition, the interstitial-RE k-means models, trained by atomistic environments in interstitial-RE testing set, is highly effective for identifying interstitial REs, with around 90% of real REs identified and 70% of identified REs being real REs.

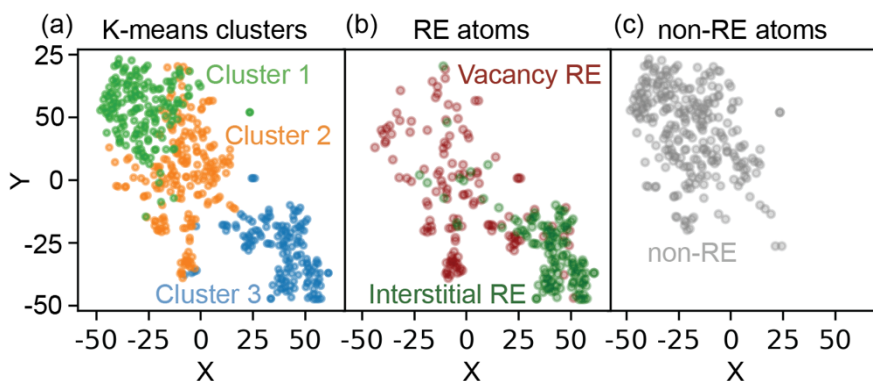


Figure C13. The 2-dimensional t-SNE projections of the SOAP descriptor of 520 atomic environments, showing (left) the clusters by all-RE k-means model compared with (middle) migrating vacancies (red) or interstitials (green) and (right) non-migrating atoms (grey), from $\mathcal{D}_{\text{RE-V}}^{\text{Testing}}$ and $\mathcal{D}_{\text{RE-I}}^{\text{Testing}}$.

Table C8. The precisions and recalls of the all-RE k-means models, with different cluster number parameters, for identifying both interstitial and vacancy RE migrating atoms (Appendix D).

All-REs		
Cluster number	Precision (%)	Recall (%)
3	62	61
4	62	59
5	64	59
6	65	57
7	66	57
8	66	57

Table C9. The precisions and recalls of interstitial-RE k-means models with different cluster numbers, which is trained using interstitial-RE testing set and for identifying interstitial RE migrating atoms for RE-enhanced workflow (*Generating RE Data and Constructing the Enhanced Training Dataset* in Appendix D).

Interstitial-RE K-means Models		
Cluster number	Precision (%)	Recall (%)
3	65	90
4	67	90
5	69	89
6	68	89
7	69	89
8	70	88

Table C10. The precisions and recalls of vacancy-RE k-means models with different cluster numbers, which is trained using vacancy-RE testing set and for identifying vacancy RE migrating atoms (*Optimizing Enhanced MLPs* in Appendix D).

Cluster number	Vacancy-RE K-means Models	
	Precision (%)	Recall (%)
3	10	75
4	8	80
5	10	74
6	20	60
7	23	51
8	24	47

9. Force accuracy as an indicator for MLP performance

Here, we compare the 11 known MLPs and a large number of selected enhanced MLPs for the force performance scores, $P(\mathcal{D}_{\text{RE-I}}^{\text{Testing}})$ and $P(\mathcal{D}_{\text{RE-V}}^{\text{Testing}})$, (see Appendix D) and their activation energy, E_a^{I} and E_a^{V} respectively, from the fitted Arrhenius relation of the diffusivities from their MLP-MD simulations (**Figure C14**). The studied MLPs include 11 known MLPs, 46 interstitial-enhanced MLPs (*Optimizing Enhanced MLPs* in Appendix D), 54 selected vacancy-enhanced MLPs (*Analysis of Errors of MLPs* above), and 24 additional MLPs produced using the original training dataset in Ref.^[1] but with different hyperparameters, which were selected by the optimization processes using the eight selection criterion (*Optimizing Enhanced MLPs* in Appendix D). The force performance scores, $P(\mathcal{D}_{\text{RE-I}}^{\text{Testing}})$ and $P(\mathcal{D}_{\text{RE-V}}^{\text{Testing}})$, were evaluated on RE migrating atoms in interstitial-RE testing and vacancy-RE testing datasets, respectively (*Optimizing Enhanced MLPs* in Appendix D). The interstitial-enhanced MLPs generally outperform other MLPs on $P(\mathcal{D}_{\text{RE-I}}^{\text{Testing}})$ as shown in main text (Figure 3.6a and C14a), and similarly MLPs trained by vacancy-enhanced dataset have higher $P(\mathcal{D}_{\text{RE-V}}^{\text{Testing}})$ than other MLPs (Figure C14b). By comparing a range of MLPs with different force performance scores versus the activation energy from MLP-MD simulations with corresponding defect, the MLPs generally give better predicted values of E_a if their force performance scores are higher than DFT K1 calculations.

There are a few exceptions, which may deserve future study. In general, force performance scores are good metrics to evaluate MLPs for their performance in MD simulations.

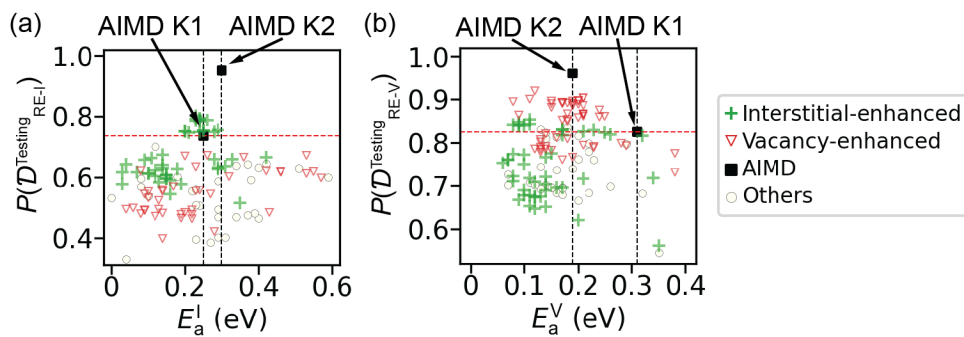


Figure C14. The force performance scores vs. E_a for known, interstitial-enhanced, vacancy-enhanced, and additional MLPs on (a) interstitial RE migrating atoms and (b) vacancy RE migrating atoms.

10. Convergence of interstitial occurrence frequencies

Here we examine whether the occurrence frequencies of different interstitials have converged in the MD simulations. We calculate the average and standard deviation of occurrence frequencies for simulations conducted over 0.25, 0.3, 0.4, 0.6, 0.8, and 1 ns shown in **Figure C15**. The average values and standard deviations of occurrence frequencies are calculated for 10 configuration sets obtained from the MD simulations with different durations. Every set of configurations consists of 2000 configurations that are randomly selected during MD simulations, with a minimum separation time of 100 fs, following the same scheme in Appendix D. The occurrence frequencies of three interstitials for all MLP-MD simulations (Figure. 3.3) are well converged over different durations of the MD simulations, with standard deviations of occurrence frequencies lower than 1.3% for all three interstitials for all MLPs. Therefore, the MD simulations have a long duration and have sampled a large number of configurations, and thus the obtained occurrence frequencies are well converged.

We also note that the occurrence frequencies of three interstitials do not always sum up to 1, because some interstitial configurations, e.g., migrating interstitials or concerted migrations of multiple interstitials, happened in MD simulations cannot be classified as either of three interstitials.

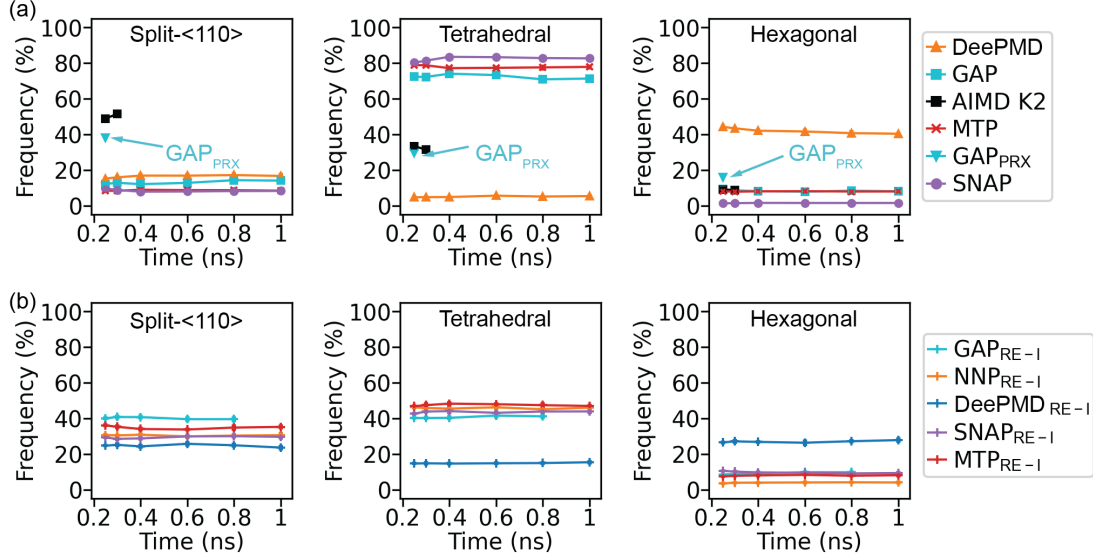


Figure C15. The average occurrence frequencies and their standard deviations for MD simulations with different durations. The average and standard deviation of occurrence frequencies for split-<110> (left), tetrahedral (middle), and hexagonal (right) interstitials, in (a) AIMD K2 (black squares) simulations and MD simulations performed by known MLPs: GAP (cyan squares), GAP_{PRX} (cyan triangles), DeePMD (orange triangles), SNAP (purple circles), and MTP (red crosses), and (b) MD simulation performed by RE-enhanced MLPs: GAP_{RE-I} (cyan crosses), NNP_{RE-I} (orange crosses), DeePMD_{RE-I} (blue crosses), SNAP_{RE-I} (purple crosses), and MTP_{RE-I} (red crosses). The error bars of occurrence frequencies are smaller than 1.3% in all MD simulations.

Appendix D

D1. Obtaining and training MLPs

To obtain existing MLPs in consistent with previous studies, we retrieved MLPs (GAP, NNP, SNAP, and MTP) directly from the corresponding *mlearn* repository of Zuo et al.^[1], or trained MLPs (DeePMD) by using the training data the same as in Zuo et al.^[1] The *mlearn* package and the corresponding MLP models, including QUIP for GAP,^[11] N2P2 for NNP,^[109] MLIP for MTP,^[15,110] and SNAP coded in LAMMPS, were used for the energy and force evaluation. The GAP_{PRX} model was retrieved from Ward et al.^[111] Since different training data was used for the existing DeePMD model by Zhang et al.^[5], the DeePMD model was trained here using the training data from Zuo et al.^[1] in order to have a fair comparison with other MLP models. Here we refer to the early version of Deep Potential Molecular Dynamics as DeePMD^[5] and the later version of Deep Potential–Smooth Edition as DeepPot-SE in Supporting Information.^[79] The training of DeePMD models was performed using the DeePMD-kit package.^[79,112] The hyperparameters of the DeePMD model, including the size of the neural network, cut-off radius, and the number of iteration steps, were optimized using a grid search algorithm (two to six values were considered for each hyperparameter) by the RMSEs of energies and forces in a separate set of 100 snapshots from AIMD simulations of bulk Si with single vacancy defect (vacancy validation dataset). The length of the neighbor list in DeePMD model was estimated based on cutoff radius. The final set of hyperparameters was selected to have the lowest RMSEs of energies and forces in both training dataset and the vacancy validation dataset.

D2. First-principles computation

DFT calculations were performed to generate additional data of energies, forces, PES and atomic configurations for training and testing. All DFT calculations were performed by Vienna ab initio simulation package^[98] (VASP) with the projector augmented-wave approach on these snapshots. The Perdew-Burke-Ernzerhof^[100] (PBE) functionals by generalized-gradient approximation (GGA) were adopted to calculate the total energies of snapshots. Static relaxation of atomic configurations were performed using spin-polarized DFT calculations with an energy cutoff of 520 eV, an electronic relaxation convergence cut-off of 10^{-5} eV, and other parameters set similar to those used in Materials Project.^[102,105] Our benchmark true values of energies and forces were calculated using $4\times4\times4$ k-point mesh (K4), while Γ -centered $1\times1\times1$ (K1) and $2\times2\times2$ (K2) were also evaluated for comparison.

D3. Ab initio molecular dynamics simulation

Ab initio molecular dynamics (AIMD) simulations were performed using non-spin-polarized, an electronic energy convergence cut-off of 10^{-4} eV, a Γ -centered $1\times1\times1$ (K1) or $2\times2\times2$ (K2) *k*-point mesh, and a time step of 2 fs. The PBE functionals by GGA were adopted as in the *First-Principles Computation* section (Appendix D). The supercell model consists of Si bulk-phase with $2\times2\times2$ conventional unit cells (64 atoms) with a single vacancy (63 atoms) or a single interstitial (65 atoms). The initial temperature of the simulations was set to 100 K after a static relaxation of the initial structures, and the structures were heated to the final temperatures during a period of 2 ps with a constant rate by velocity scaling, and afterwards an NVT ensemble with Nosé-Hoover thermostat was adopted. To obtain diffusional properties, AIMD simulations were performed at different temperatures 730, 840, 1000, 1230, 1500, and 1600 K following the same

scheme in Ref.^[4,32]. Missing diffusivities at certain temperatures (Figure 3.2a and b) are either due to the melting of the crystal structures or inadequate numbers of hopping events (specified, Table C4, C5, and C6). Given the stochastic nature of ion hopping in estimating diffusivities, the diffusivity calculations were converged using our developed scheme.^[4,32] and the error bars are estimated correspondingly based on the total number of ion hops. The diffusivities and their error bars of AIMD simulations at all temperatures are available (Table C4, C5, and C6).

The diffusivities and their error bars were evaluated according to the total mean-squared-displacement (TMSD) of Si atoms as in Ref.^[4]. The total time durations of AIMD simulations were in the range of 100 and 1000 ps, so the values of TMSD were in the range of 1200 to 3000 Å² and were 600 Å² for AIMD K2 simulations at temperatures of 1000 K and 1230 K due to high computation cost.

MLP-MD simulations. All classical MD simulations based on MLPs were carried out using LAMMPS. The MLP-MD simulations for vacancy and interstitial diffusion were performed on the same supercells with single interstitial or vacancy defect as in the AIMD simulations with NVT ensemble. The time step of all MD simulations was set to 1 fs. MLP-MD simulations of all MLPs are performed at six different temperatures 730, 840, 1000, 1230, 1500, and 1600 K. The total time duration of MD simulations was in the range of 500 to 5000 ps, similar to those used in AIMD simulations. The TMSDs of MD simulations were between 6000 - 320000 Å² for NNP, SNAP, MTP, DeePMD, and DeepPot-SE, and > 1500 Å² for GAP models. The convergence of diffusivity and the error bars are estimated correspondingly based on the total number of ion hops in the same scheme as in the AIMD simulations.^[4,32] The diffusivities and their error bars from MLP-MD simulations are summarized in Table C4 and C5, following the same scheme described in *Ab initio molecular dynamics simulation* Section (Appendix D).

D4. Analyzing Si interstitial in MD simulations

The occurrence frequency of each interstitial configuration (split- $\langle 110 \rangle$, tetrahedral, and hexagonal) was counted among 2000 snapshots (taken every 100 fs) from the single-interstitial supercell model from MD simulations at 1230K. To determine the interstitial type, a structural matching algorithm was performed for each snapshot following the scheme adopted in Ref. [59] using *pymatgen*^[105]. The interstitial configurations used as matching templates were fully relaxed using the fixed lattices of a crystalline Si bulk and a Γ -centered $4 \times 4 \times 4$ k-point mesh, with energy and the force convergence criteria at 10^{-7} eV and 0.01 eV/Å, respectively. The matching algorithm used the tolerance parameters of the lattice angle of 5° , the lattice length of 20%, and the site root-mean-square tolerance of $0.275(V/n)^{1/3}$, where V/n was the volume V normalized by the number of atoms n . Some interstitial configurations, e.g., migrating interstitials or concerted migrations of multiple interstitials, happened in MD simulations, cannot be classified as any of the three interstitials, and thus the occurrence frequencies of three interstitials do not always sum up to 1.

D5. Identifying static sites, vacancies, and migrating atoms

The algorithms for identifying static sites, vacancies, and migrating atoms during MD simulations were as follows. The static sites of Si were set to the sites of the perfect crystalline Si bulk. A static site that has no Si atoms within 1.1 Å was identified as a vacancy. A migrating Si atom (RE atom) was a Si atom between two nearest-neighbor static sites, which was selected if the distances to their 1st nearest and 2nd nearest static sites has a difference below 0.75 Å (approximately 31% of distances between two static sites). This criterion can effectively distinguish migrating atoms from atoms vibrating around static sites.

The values of the distance to the nearest static site, r_s , the distances to its nearest vacancy, r_v , and the distances to its nearest RE atom, r , were quantified correspondingly from these static sites. All distributions (Figure 3.3c – h) was generated with 50×50 grid size by collecting r_v and r_s distances of vacancy NN or non-NN atoms over 2000 snapshots taken randomly (> every 100 fs) from each simulation of MLPs or AIMD at 1230 K. The Gaussian filter with smoothing parameter σ set to 3 was applied in plotting the distributions using *SciPy* library^[113].

D6. Testing dataset with REs

Two testing sets, $\mathcal{D}_{\text{RE-I}}^{\text{Testing}}$ and $\mathcal{D}_{\text{RE-V}}^{\text{Testing}}$ sets, are constructed from 100 snapshots taken from AIMD simulations at 1230 K with single-vacancy and single-interstitial, respectively, with every snapshot having a RE. The energies and atomic forces of each snapshot were further converged by single-step self-consistent DFT calculations to a higher accuracy (Γ -centered 4×4×4 k-point mesh) with fixed atom positions and lattices, and these converged values (DFT K4) were used as true values for testing.

D7. Error evaluation of atomic forces

Using the $\mathcal{D}_{\text{RE-I}}^{\text{Testing}}$ and $\mathcal{D}_{\text{RE-V}}^{\text{Testing}}$ sets, the error evaluation of atomic forces was performed for all MLPs. The errors of atomic forces by DFT calculations with lower accuracy setting (K1 and K2) commonly used in AIMD simulations were also evaluated in comparison to the MLPs. The magnitude error δ_f and the directional error δ_θ of forces were evaluated according to Eq. (1) and (2) in Chapter 3, respectively. From the error evaluation results on the RE atoms (i.e., migrating atoms as defined above) in the $\mathcal{D}_{\text{RE-I}}^{\text{Testing}}$ or $\mathcal{D}_{\text{RE-V}}^{\text{Testing}}$ dataset, the CDFs of δ_f and δ_θ were generated. The normalized area under the curves $NAC(\delta, \mathcal{D})$ of the CDFs of δ_f and δ_θ were quantified in corresponding dataset $\mathcal{D}$ for δ_f over 0 – 1 eV/Å (20 bins) or for δ_θ over 0 – 90 degrees (20 bins), respectively, for each MLP or DFT K1/K2 as shown (Figure 3.4), and were normalized by the total area of the evaluation range of 0 – 1 eV/Å for x-axis and 0 – 100% for y-axis, giving to a value between 0 and 1. Higher NAC values (closer to 1) correspond to smaller errors in the MLP prediction compared to DFT K4 benchmark.

D8. Generating RE-enhanced training dataset

To train MLPs with low errors of forces on RE atoms, 120 snapshots with identified RE atoms were randomly selected to replace the structures in the original training dataset^[1] from the following categories, such as liquid Si, AIMD simulations of Si bulk, and the strained Si bulk, from the original training dataset, in order to maintain a balance of structures for each category. The RE-enhanced training dataset has the same size and similar diversity of different types of atomic configurations as the original dataset in Ref.^[1] so we can make a fair comparison between existing and RE-enhanced MLPs.

D9. Optimizing enhanced MLPs

The validation dataset to optimize the MLPs trained by RE-enhanced training data was constructed as follows. The enhanced validation set (EVS) contains 50 total structures, consisting of 20 structures randomly selected from the 120 replaced structures of the original training dataset, 11 snapshots with vacancy RE from AIMD simulations, and 19 snapshots with interstitial RE from AIMD simulations. The optimization of the hyperparameters of MLPs considered four to ten values for each parameter including the band limit of spherical harmonic basis functions and the number of radial basis functions for GAP, the size of neural network for NNP, the number of radial basis functions number for MTP. For each MLP model, grid search algorithms were performed among 300 – 2000 sets of hyperparameters to identify the selected MLPs as explained below.

For the selection of optimal MLP models, which were used to study the force performances of MLPs (Figure 3.6c), we evaluated the following eight metrics, such as the RMSEs of energies and forces of 31 structures (excluding the 19 interstitial structures), the RMSEs of energies and forces of RE atoms in 11 vacancy structures, the RMSEs of the energies and forces of RE atoms in 19 interstitial structures, the RMSEs of the energies and forces of RE atoms in 30 vacancy and interstitial structures in the EVS. We used the force performance score $P(\mathcal{D})$ for the dataset $\mathcal{D}$. The evaluations were performed for different data such as all atoms in EVS, the RE atoms of the interstitial structures in EVS, the RE atoms of the vacancy structures in EVS, or all RE atoms in EVS.

Using these evaluation scores, we first selected those MLPs with the lowest RMSE or the highest NAC in any one of the eight criteria, giving a total of 46 MLPs as interstitial-enhanced MLPs. The same optimization process using eight criterion was also applied on selecting additional MLPs trained by the original dataset in Ref.^[1] obtaining 24 additional MLPs. All these

46 interstitial-enhanced MLPs and 24 additional MLPs were used (Figure 3.6a) for testing of their force performances on interstitial RE atoms. These RE-enhanced MLPs were further down selected to have the optimal one as the representative MLPs, using the joint force error score of $P(\mathcal{D}_{\text{RE-I}}^{\text{EVS}})^2 + P(\mathcal{D}_{\text{RE-V}}^{\text{EVS}})^2$. All NACs of δ_F and δ_θ were computed using the RE atoms in the EVS. Six optimal MLPs, one for each model, were selected. Both MLP-MD simulations with single vacancy and MD simulations with single interstitial were further performed at six different temperatures to analyze their diffusional properties.

D10. Testing interstitial-enhanced MLPs

Our testing of the MLPs from RE-enhanced training dataset (all 46 interstitial-enhanced MLPs) included the evaluation of joint force error matrices calculated from $\mathcal{D}_{\text{RE-I}}^{\text{Testing}}$ and $\mathcal{D}_{\text{RE-V}}^{\text{Testing}}$ sets, $P(\mathcal{D}_{\text{RE-I}}^{\text{Testing}})$ and $P(\mathcal{D}_{\text{RE-V}}^{\text{Testing}})$ respectively, and the evaluation of their diffusivities, pre-exponential factors, interstitial occurrence frequencies and activation energies of vacancy and interstitial in MD simulations. The joint force error metrics and their calculated diffusional properties are compared as shown (Figure 3.6a). The joint force error metrics, $P(\mathcal{D}_{\text{RE-I}}^{\text{Testing}})$ and $P(\mathcal{D}_{\text{RE-V}}^{\text{Testing}})$, of the existing MLPs and the additional MLPs were also calculated using the $\mathcal{D}_{\text{RE-I}}^{\text{Testing}}$ and $\mathcal{D}_{\text{RE-V}}^{\text{Testing}}$ sets.

Appendix E

1. Errors of MLPs on dataset with single vacancies

We test the energies of intermediate Li-Al phases with single vacancies for the machine-learning interatomic potential (MLP) trained by only three materials, $\text{MLP}_{\text{Train}}$, used in Chapter 4.2. The dataset for intermediate Li-Al phases with single vacancies, $\mathcal{D}^{\text{Vacancy}}$, was constructed using all eight known Li-Al intermediate phases (Figure 4.1). The 30 symmetrically distinct structures were generated by swapping a pair of randomly selected atom, and by randomly removing either a Li atom or an Al atom to create a single Li vacancy (V^{Li}) or Al vacancy (V^{Al}).

The same tests as Chapter 4.2 were performed for these vacancy structures, $\mathcal{D}^{\text{Vacancy}}$. The overall energy root-mean-squared-error (RMSE) of all configurations in $\mathcal{D}^{\text{Vacancy}}$ is as low as 19 meV/atom (18 and 19 meV/atom for V^{Li} and V^{Al} configurations, respectively). Only one intermediate phase, LiAl (*Cmce*), which has a non-BCC or non-FCC lattice structure, has higher energy RMSEs of 37 and 36 meV/atom for V^{Li} and V^{Al} , respectively (Figure E1a). 11 out of the 16 lowest energy configurations (LECs) are correctly predicted. For those LECs that are not correctly predicted by MLP, the errors with the DFT energies, ΔE^{DFT} , are very small at < 2 meV/atom (Figure E1b). The error rates of ranking errors are low for most intermediate phases, and only LiAl (*Fd $\bar{3}m$*), LiAl₃ (*Pm $\bar{3}m$*), and LiAl (*Cmce*) have the error rates of ranking errors of $> 33\%$ (**Figure E1c**). These results show that the $\text{MLP}_{\text{Train}}$ has decent transferability to the configurations with vacancies of other intermediate phases, even the original $\mathcal{D}^{\text{train}}$ doesn't include these phases.

In consistency with the findings of Chapter 4.4, our re-trained MLP ($\text{MLP}_{\text{Re-train}}$) on more intermediate phases as in Chapter 4.4 exhibits worse performances compared to the original $\text{MLP}_{\text{Train}}$. The $\text{MLP}_{\text{Re-train}}$ shows higher overall energy RMSE of 42 meV/atom (41 and 42 meV/atom for configurations with V^{Li} and V^{Al} respectively) and higher error rates of ranking errors

(4 defected intermediate phases have error rates above 33%) (Figure E1b and f). The $\text{MLP}_{\text{Re-train}}$ also predicts one less correct LEC compared with $\text{MLP}_{\text{Train}}$ (Figure E1d).

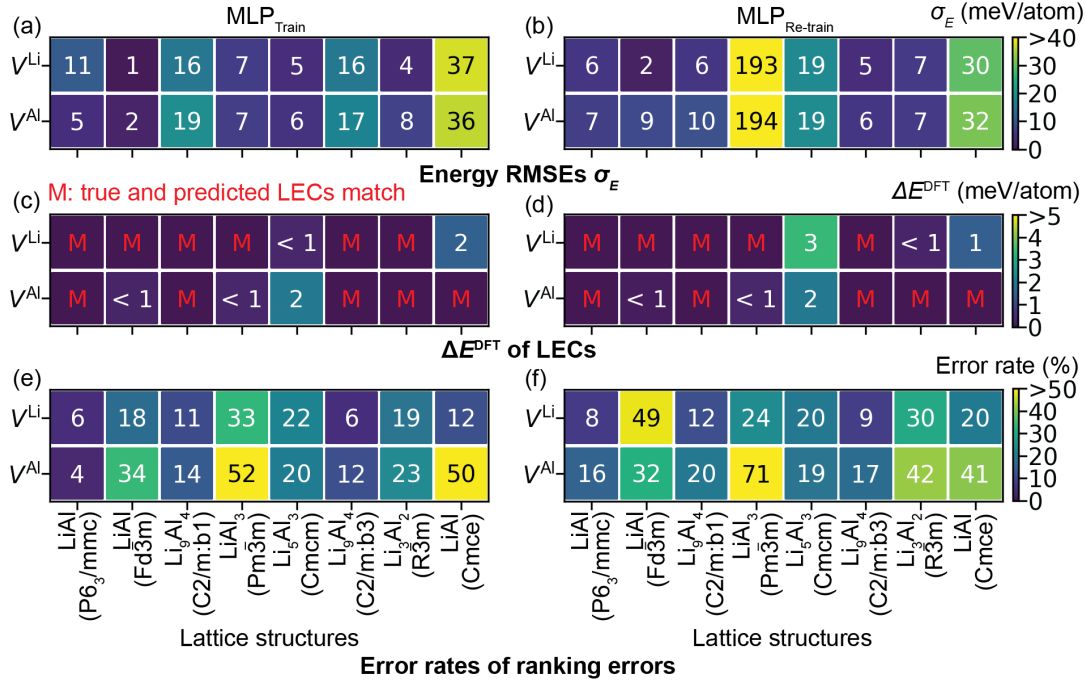


Figure E1. The performance of MLPs on Li-Al intermediate phases with single vacancies. The RMSEs of predicted energies of up to 30 ordered configurations on eight intermediate phases (x axis) with either V^{Li} or V^{Al} (y axis), for (a) the $\text{MLP}_{\text{Train}}$ in Chapter 4.2 and (b) the $\text{MLP}_{\text{Re-train}}$ in Chapter 4.4. The differences of DFT energies, ΔE^{DFT} between the LECs calculated by DFT and predicted by (c) $\text{MLP}_{\text{Train}}$, or (d) $\text{MLP}_{\text{Re-train}}$. A match of LECs predicted by DFT and MLP is marked as M. The error rates of ranking errors for (e) $\text{MLP}_{\text{Train}}$, and (f) $\text{MLP}_{\text{Re-train}}$.

2. Prediction of ground-state phases (GPs) for different compositions

Following the same scheme in Chapter 4.3 in main text (Figure 4.2c and Figure 4.3b), here we compare the MLP-predicted GPs of Li_5Al_3 , Li_4Al , LiAl , Li_2Al_3 , LiAl_3 , and LiAl_5 . Those phases in $\mathcal{D}^{\text{test}}$ were hypothetically generated using lattice structures of intermediate phases (all eight lattice structures for Li_5Al_3 , LiAl , and LiAl_3 , the five selected lattice structures for Li_4Al , Li_2Al_3 , and LiAl_5 , **Figure E2**). The $\text{MLP}_{\text{Train}}$ incorrectly predicts the GPs of hypothetical Li_5Al_3 and Li_2Al_3 to be $Fd\bar{3}m$, in comparison to $Cmcm$ and $Pm\bar{3}m$, respectively, predicted by DFT (Figure E2a and d). The energy difference, ΔE^{DFT} , between the true and the predicted LECs of two phases is as small as 4 meV/atom (Figure E2a). The incorrect prediction of hypothetical Li_5Al_3 may be attributed to the difficulties of the MLP in accurately capturing similar BCC lattice structures of $Fd\bar{3}m$ and $Cmcm$ with different site orderings. For Li_2Al_3 , the ΔE^{DFT} is relatively high (21 meV/atom). This may also explain the discrepancy of the DFT and MLP predicted at 62% Al and 57% Al, respectively, for the phase transition between the BCC (the MLP-predicted GP on Li_2Al_3) lattice and the FCC (the DFT GP of Li_2Al_3) lattice shown in Chapter 4.5.2.

Additionally, the $\text{MLP}_{\text{Re-train}}$ correctly predicts the GPs of all six compositions from $\mathcal{D}^{\text{test}}$ (Figure E3), even though the predicted energies of the LECs of predicted GPs (such as the LECs for hypothetical Li_2Al_3 , LiAl_3 , and LiAl_5) differ more than 0.1 eV/atom from DFT calculations (Figure E3d, e, and f). It is notable that the $\text{MLP}_{\text{Re-Train}}$, even having large errors on predicted energies, can still correctly predict the GPs. It is notable that low energy RMSEs are not always equivalent to the correct prediction of GPs, which correspond to the prediction of energy rankings. For example, the energy RMSE of the wrongly predicted GP of Li_5Al_3 $Cmcm$ (Figure E2a) is as low as 2 meV/atom, while the energy RMSE of the correctly predicted GP of Li_2Al_3 $Fd\bar{3}m$ is 12

meV/atom (Figure 4.1). Thus, the errors on energy values and relative energy rankings should be tested about the correct prediction of GPs.

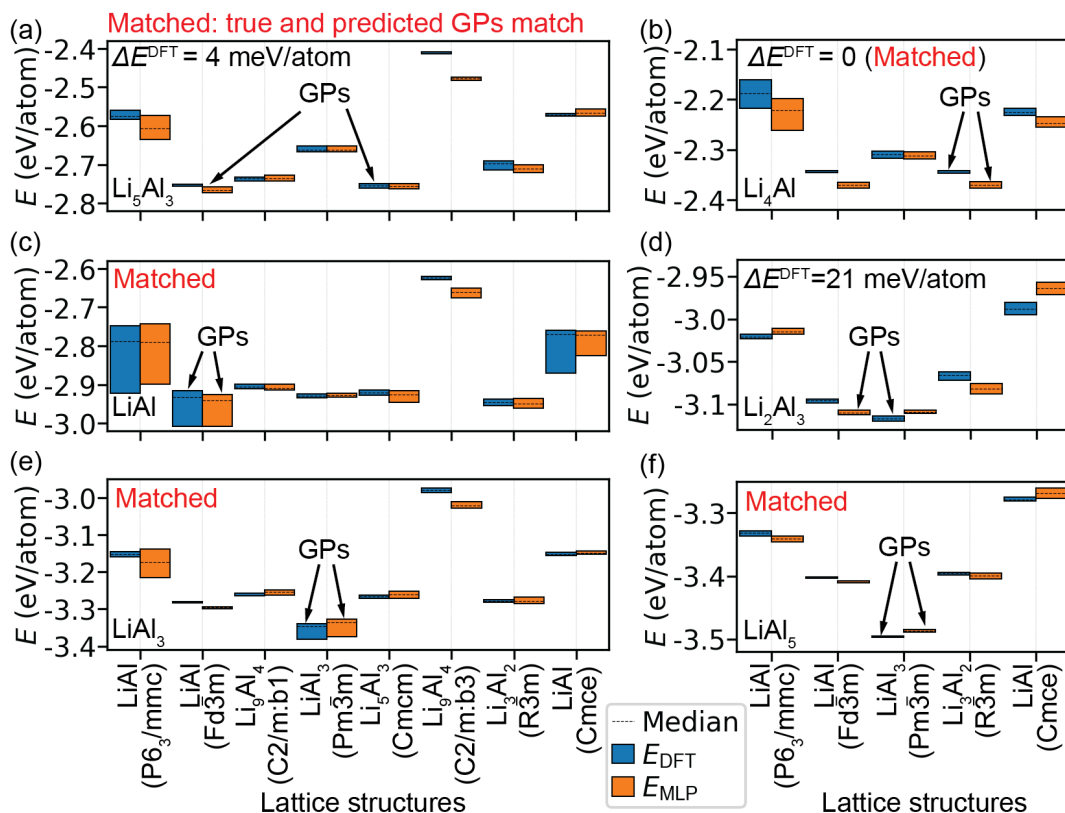


Figure E2. Distributions of the predicted and the true energies of MLP_{Train} and DFT for up to 30 configurations in each phase as Figure 4.2c in main text Chapter 4. The box plots show the maximum, the minimum, and the median (dashed lines) of energies in all eight lattice structures (x axis) for the compositions (a) Li_5Al_3 , (c) LiAl , and (e) LiAl_3 , and five selected lattice structures for (b) Li_4Al , (d) Li_2Al_3 , and (f) LiAl_5 . A match of the true and the predicted GPs produced by DFT and MLP is marked as red ‘matched’.

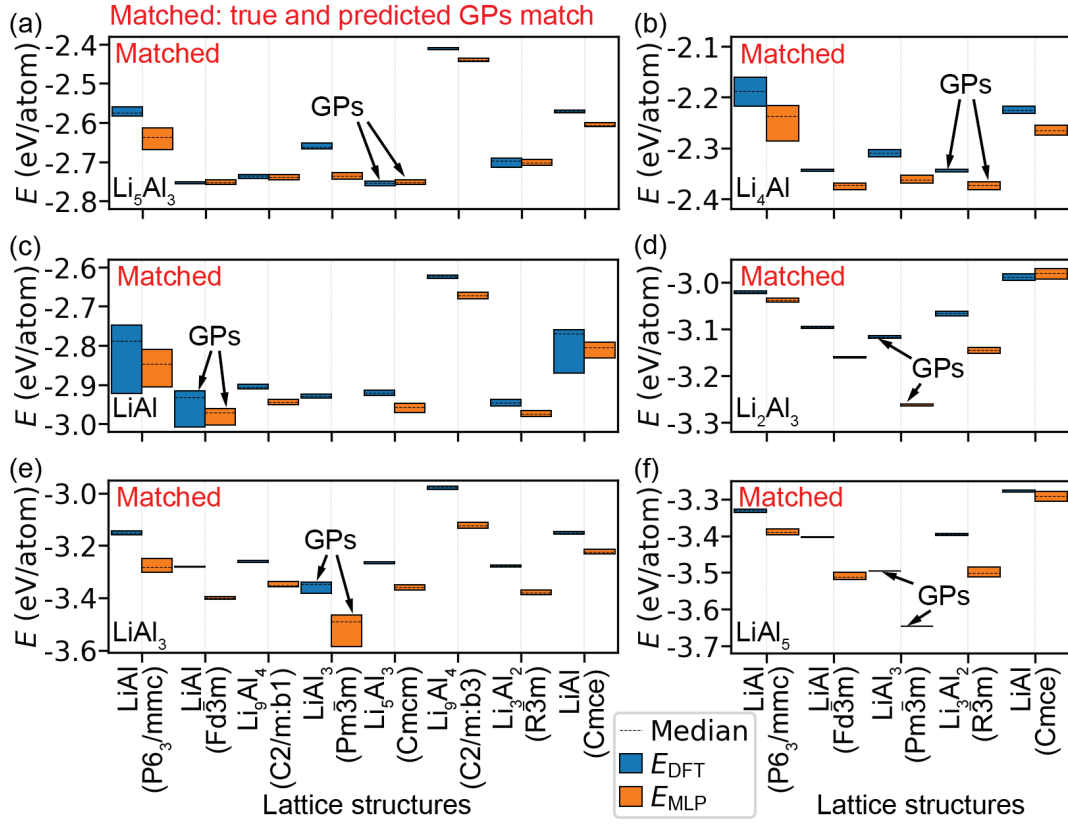


Figure E3. Distributions of the predicted and the true energies of GPs produced by MLP_{Re-train} and DFT. The box plots show the maximum, the minimum, and the median (dashed lines) of energies in all eight lattice structures (x axis) for compositions (a) Li_5Al_3 , (c) LiAl , and (e) LiAl_3 , and five selected lattice structures for compositions (b) Li_4Al , (d) Li_2Al_3 , and (f) LiAl_5 . A match of the true and the predicted GPs produced by DFT and MLP is marked as red ‘matched’.

3. The ranking errors of hypothetical phases in $\mathcal{D}^{\text{test}}$

As discussed in Chapter 4.3, the ranking error refers to the errors in predicting the relative stability of different elemental orderings. In Chapter 4.3, we analyzed the energy ranking of the 30 configurations randomly generated in Li_4Al (Figure 4.2b). Here we show the distributions of the energy differences, ΔE^{DFT} , of ranking errors for all hypothetical phases in $\mathcal{D}^{\text{test}}$, as predicted by $\text{MLP}_{\text{Train}}$ (Figure E4) and $\text{MLP}_{\text{Re-train}}$ (Figure E5). For most materials, except for LiAl ($P6_3/mmc$), the probability density function (PDF) curves show low probability at ΔE^{DFT} higher than 15 meV/atom, suggesting that ranking errors mostly occur between those configurations with similar DFT energies in a close range (Figure E4, and E5). For the HCP LiAl ($P6_3/mmc$) lattice structure (blue lines, Figure E4a, b, c, e, Figure E5a, b, c, and e) with compositions Li_5Al_3 , Li_4Al , LiAl , and LiAl_3 , the large spreads of PDFs in both $\text{MLP}_{\text{Train}}$ and $\text{MLP}_{\text{Re-train}}$ imply that the atomic environments in this lattice structure are poorly captured, even though the configurations of the original HCP LiAl ($P6_3/mmc$) intermediate phase were included in the $\mathcal{D}^{\text{re-train}}$ for the training of $\text{MLP}_{\text{Re-train}}$.

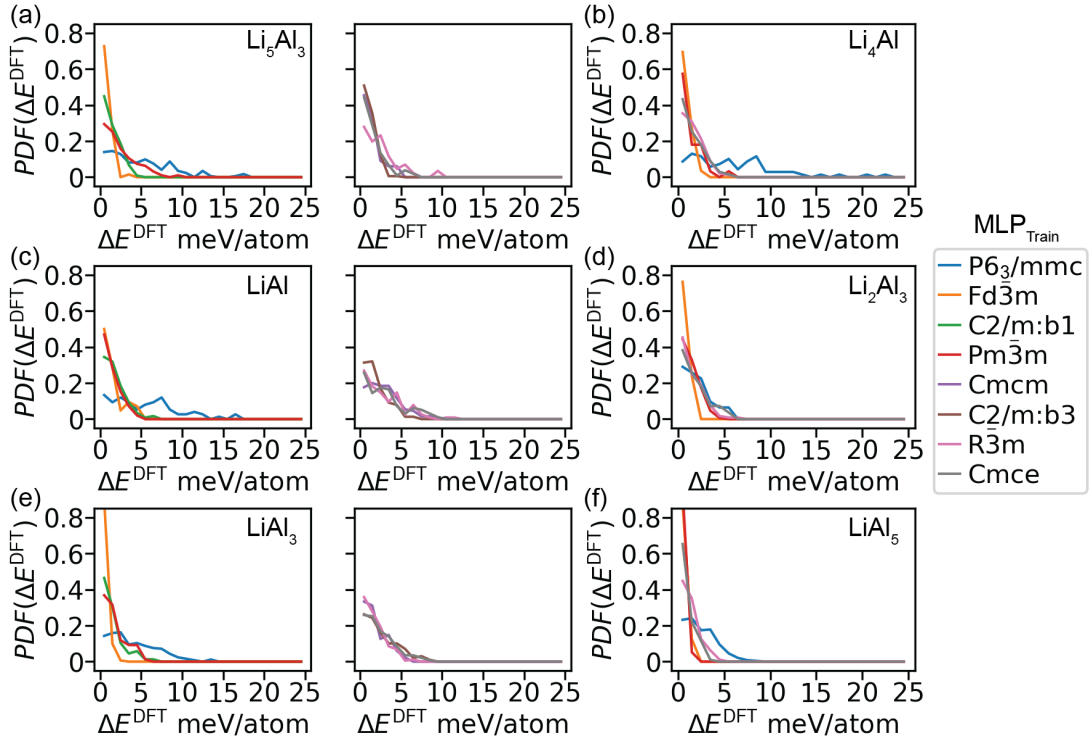


Figure E4. The probability density functions (PDFs) of ΔE^{DFT} produced by $\text{MLP}_{\text{Train}}$ for all hypothetical phases with compositions (a) Li_5Al_3 , (b) Li_4Al , (c) LiAl , (d) Li_2Al_3 , (e) LiAl_3 , and (f) LiAl_5 derived from the lattice structures (legend, color) in $\mathcal{D}^{\text{test}}$.

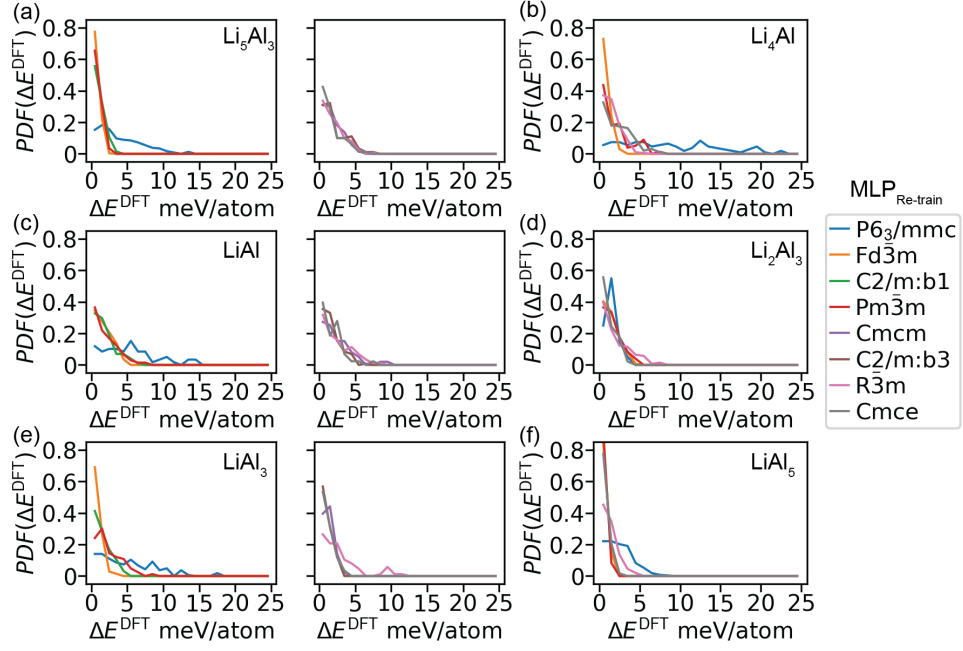


Figure E5. The PDFs of ΔE^{DFT} produced by $\text{MLP}_{\text{Re-train}}$, for all hypothetical phases with compositions (a) Li_5Al_3 , (b) Li_4Al , (c) LiAl , (d) Li_2Al_3 , (e) LiAl_3 , and (f) LiAl_5 derived from the lattice structures (legend, color) in $\mathcal{D}^{\text{test}}$.

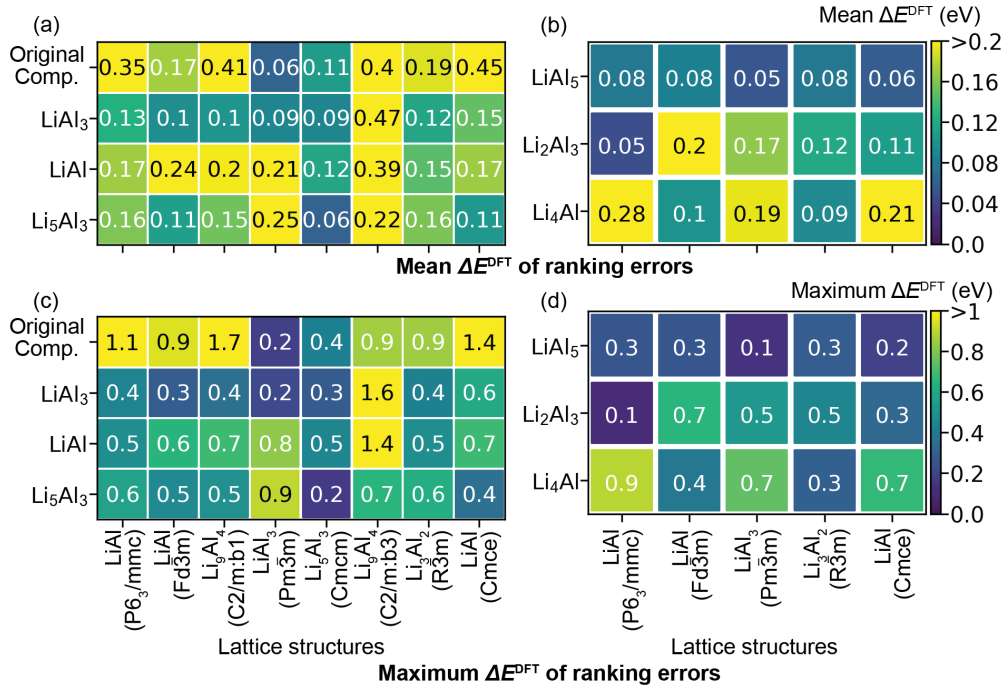


Figure E6. The predictions of energy rankings, (a, b) the mean ΔE^{DFT} and (c, d) the maximum ΔE^{DFT} , by $\text{MLP}_{\text{Re-train}}$ for all hypothetical phases in $\mathcal{D}^{\text{test}}$.

4. Test and compare multiple MLPs from $\mathcal{D}^{\text{train}}$ and $\mathcal{D}^{\text{re-train}}$

In order to show that our findings for the MLPs shown in Chapter 4 are representative, we select, test, and compare three additional MLPs for each of the $\mathcal{D}^{\text{train}}$ and $\mathcal{D}^{\text{re-train}}$. The additional MLPs were selected as follows. The $\text{MLP}_{\text{Train}}^{\text{extra-1}}$ and $\text{MLP}_{\text{Re-train}}^{\text{extra-1}}$ are obtained the same as $\text{MLP}_{\text{Train}}$ and $\text{MLP}_{\text{Re-train}}$ used in the Chapter 4 but with the second lowest total validation score V (methods described in the Appendix F). Additionally, $\text{MLP}_{\text{Train}}^{\text{extra-2}}$, $\text{MLP}_{\text{Train}}^{\text{extra-3}}$, $\text{MLP}_{\text{Re-train}}^{\text{extra-2}}$, and $\text{MLP}_{\text{Re-train}}^{\text{extra-3}}$, were selected as follows. We first calculate the selection score S , using the energy (σ_E) and force (σ_F) RMSEs, σ_E^{Train} , σ_F^{Train} , $\sigma_E^{\text{Validation}}$, and $\sigma_F^{\text{Validation}}$, based on $\mathcal{D}^{\text{train}}$ (or $\mathcal{D}^{\text{re-train}}$ respectively) and $\mathcal{D}^{\text{validation}}$ as:

$$S = \sum_i \sum_j (v_i^j)^2, \quad \text{Eq. (E1)}$$

$$v_i^j = [\sigma_i^j - \min(\sigma_i^j)] / \text{std}(\sigma_i^j), I \in \{E, F\}, \text{ and } j \in \{\text{Train, Validation}\}, \quad \text{Eq. (E2)}$$

where σ_i^j is one of the four RMSEs, and $\min(\sigma_i^j)$ and $\text{std}(\sigma_i^j)$ are the minimum and the standard deviation of σ_i^j , respectively. $\text{MLP}^{\text{extra-2}}$ and $\text{MLP}^{\text{extra-3}}$ are the MLP models with the lowest and second lowest S scores.

Our results show that all MLPs within the same training dataset of either $\mathcal{D}^{\text{train}}$ or $\mathcal{D}^{\text{re-train}}$ have similar performances (Table E1, and Figure E7 – 12). The energy RMSEs of $\text{MLP}_{\text{Train}}$, $\text{MLP}_{\text{Train}}^{\text{extra-1}}$, $\text{MLP}_{\text{Train}}^{\text{extra-2}}$, $\text{MLP}_{\text{Train}}^{\text{extra-3}}$, $\text{MLP}_{\text{Re-train}}$, $\text{MLP}_{\text{Re-train}}^{\text{extra-1}}$, $\text{MLP}_{\text{Re-train}}^{\text{extra-2}}$, and $\text{MLP}_{\text{Re-train}}^{\text{extra-3}}$ are low. For example, the energy RMSEs of $\text{MLP}_{\text{Train}}$, $\text{MLP}_{\text{Train}}^{\text{extra-1}}$, $\text{MLP}_{\text{Train}}^{\text{extra-2}}$, and $\text{MLP}_{\text{Train}}^{\text{extra-3}}$ on $\mathcal{D}^{\text{validation}}$ are 1, 2, 2, and 2 meV/atom, respectively (Table E1). The force RMSEs of $\text{MLP}_{\text{Train}}$, $\text{MLP}_{\text{Train}}^{\text{extra-1}}$, $\text{MLP}_{\text{Train}}^{\text{extra-2}}$, $\text{MLP}_{\text{Train}}^{\text{extra-3}}$, $\text{MLP}_{\text{Re-train}}$, $\text{MLP}_{\text{Re-train}}^{\text{extra-1}}$, $\text{MLP}_{\text{Re-train}}^{\text{extra-2}}$, and $\text{MLP}_{\text{Re-train}}^{\text{extra-3}}$ (excluding $\text{MLP}_{\text{Train}}$ on $\mathcal{D}^{\text{re-train}}$) are within a close range of

0.02 - 0.05 eV/Å. These results confirm that the performances and conclusions about $\text{MLP}_{\text{Train}}$ and $\text{MLP}_{\text{Re-train}}$ shown in the main text (Chapter 4.2, 4.3, and 4.4) are representative.

These results further confirm that $\text{MLP}_{\text{Re-train}}$, $\text{MLP}_{\text{Re-train}}^{\text{extra-1}}$, $\text{MLP}_{\text{Re-train}}^{\text{extra-2}}$, and $\text{MLP}_{\text{Re-train}}^{\text{extra-3}}$ trained on the $\mathcal{D}^{\text{retrain}}$ with four phases (two intermediate phases) show worse performance than the $\text{MLP}_{\text{Train}}$, $\text{MLP}_{\text{Train}}^{\text{extra-1}}$, $\text{MLP}_{\text{Train}}^{\text{extra-2}}$, and $\text{MLP}_{\text{Train}}^{\text{extra-3}}$ trained on the $\mathcal{D}^{\text{train}}$ with only three phases (one intermediate phase), as shown by the energy RMSEs of $\text{MLP}_{\text{Train}}$ and $\text{MLP}_{\text{Re-train}}$ on $\mathcal{D}^{\text{test}}$ and the ranking error rates (Figure E13, E14, E17, and E18). For the energy RMSEs, $\text{MLP}_{\text{Train}}$, $\text{MLP}_{\text{Train}}^{\text{extra-1}}$, $\text{MLP}_{\text{Train}}^{\text{extra-2}}$, and $\text{MLP}_{\text{Train}}^{\text{extra-3}}$ only have no more than two hypothetical phases (2, 2, 2, and 0 respectively) in $\mathcal{D}^{\text{test}}$ (Figure E13a, b, c, and d) with the RMSEs higher than 40 meV/atom for, whereas $\text{MLP}_{\text{Re-train}}$, $\text{MLP}_{\text{Re-train}}^{\text{extra-1}}$, $\text{MLP}_{\text{Re-train}}^{\text{extra-2}}$, and $\text{MLP}_{\text{Re-train}}^{\text{extra-3}}$ have 24, 26, 24, and 23 hypothetical phases with RMSEs higher than 40 meV/atom (Figure E14a, b, c, and d). For the ranking error rates, $\text{MLP}_{\text{Train}}$, $\text{MLP}_{\text{Train}}^{\text{extra-1}}$, $\text{MLP}_{\text{Train}}^{\text{extra-2}}$, and $\text{MLP}_{\text{Train}}^{\text{extra-3}}$ have only 3, 2, 2, and 2 hypothetical phases with error rates higher than 50% respectively (Figure E17a, b, c, and d), whereas $\text{MLP}_{\text{Re-train}}$, $\text{MLP}_{\text{Re-train}}^{\text{extra-1}}$, $\text{MLP}_{\text{Re-train}}^{\text{extra-2}}$, and $\text{MLP}_{\text{Re-train}}^{\text{extra-3}}$ have 6, 8, 4, and 4 hypothetical phases with error rates higher than 50% in $\mathcal{D}^{\text{test}}$ (Figure E18a, b, c, and d). For the worst performing materials, $\text{MLP}_{\text{Train}}$, $\text{MLP}_{\text{Train}}^{\text{extra-1}}$, $\text{MLP}_{\text{Train}}^{\text{extra-2}}$, and $\text{MLP}_{\text{Train}}^{\text{extra-3}}$ exhibit error rates as high as 58%, 54%, 57%, and 59%, respectively, while $\text{MLP}_{\text{Re-train}}$, $\text{MLP}_{\text{Re-train}}^{\text{extra-1}}$, $\text{MLP}_{\text{Re-train}}^{\text{extra-2}}$, and $\text{MLP}_{\text{Re-train}}^{\text{extra-3}}$ have higher error rates of 80%, 82%, 63%, and 63%.

Table E1. Energy and force RMSEs of $\text{MLP}_{\text{Train}}$, $\text{MLP}_{\text{Train}}^{\text{extra-1}}$, $\text{MLP}_{\text{Train}}^{\text{extra-2}}$, $\text{MLP}_{\text{Train}}^{\text{extra-3}}$, $\text{MLP}_{\text{Re-train}}$, $\text{MLP}_{\text{Re-train}}^{\text{extra-1}}$, $\text{MLP}_{\text{Re-train}}^{\text{extra-2}}$, and $\text{MLP}_{\text{Re-train}}^{\text{extra-3}}$ on $\mathcal{D}^{\text{train}}$, $\mathcal{D}^{\text{validation}}$, and $\mathcal{D}^{\text{re-train}}$.

Models	RMSE $\sigma_{E^{\text{Train}}}$ (meV/atom)	RMSE $\sigma_{F^{\text{Train}}}$ (eV/Å)	RMSE $\sigma_{E^{\text{Validation}}}$ (meV/atom)	RMSE $\sigma_{F^{\text{Validation}}}$ (eV/Å)	RMSE $\sigma_{E^{\text{Re-train}}}$ (meV/atom)	RMSE $\sigma_{F^{\text{Re-train}}}$ (eV/Å)
$\text{MLP}_{\text{Train}}$	2	0.03	1	0.03	103	0.13
$\text{MLP}_{\text{Train}}^{\text{extra-1}}$	3	0.02	2	0.03	55	0.07
$\text{MLP}_{\text{Train}}^{\text{extra-2}}$	3	0.02	2	0.03	41	0.06
$\text{MLP}_{\text{Train}}^{\text{extra-3}}$	3	0.02	2	0.03	81	0.22
$\text{MLP}_{\text{Re-train}}$	9	0.04	6	0.04	26	0.05
$\text{MLP}_{\text{Re-train}}^{\text{extra-1}}$	9	0.04	6	0.04	25	0.05
$\text{MLP}_{\text{Re-train}}^{\text{extra-2}}$	8	0.04	5	0.04	24	0.05
$\text{MLP}_{\text{Re-train}}^{\text{extra-3}}$	8	0.04	5	0.04	24	0.05

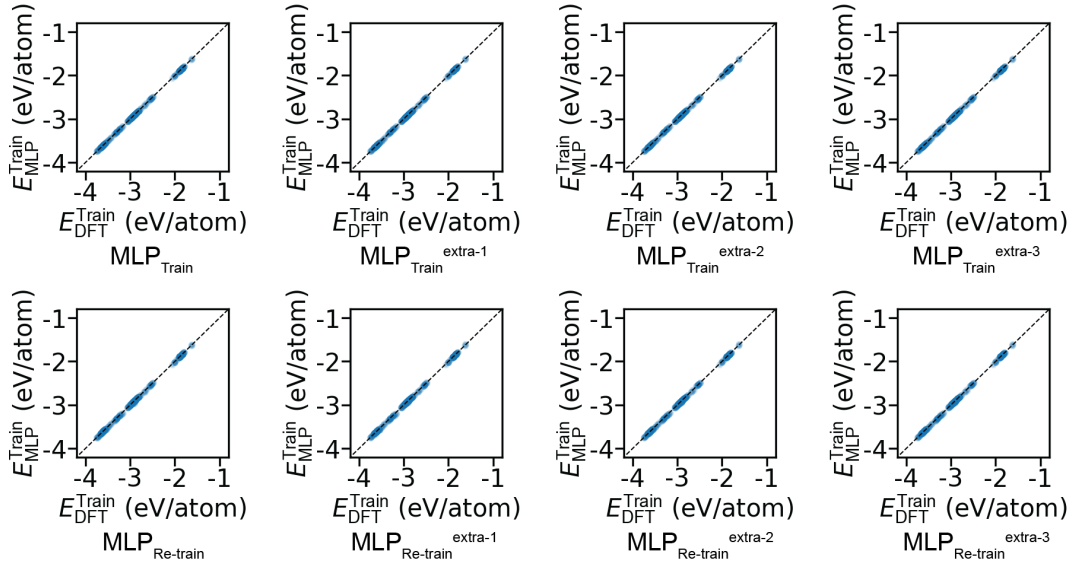


Figure E7. The energies predicted by $\text{MLP}_{\text{Train}}$, $\text{MLP}_{\text{Train}}^{\text{extra-1}}$, $\text{MLP}_{\text{Train}}^{\text{extra-2}}$, $\text{MLP}_{\text{Train}}^{\text{extra-3}}$, $\text{MLP}_{\text{Re-train}}$, $\text{MLP}_{\text{Re-train}}^{\text{extra-1}}$, $\text{MLP}_{\text{Re-train}}^{\text{extra-2}}$, and $\text{MLP}_{\text{Re-train}}^{\text{extra-3}}$ for $\mathcal{D}^{\text{train}}$.

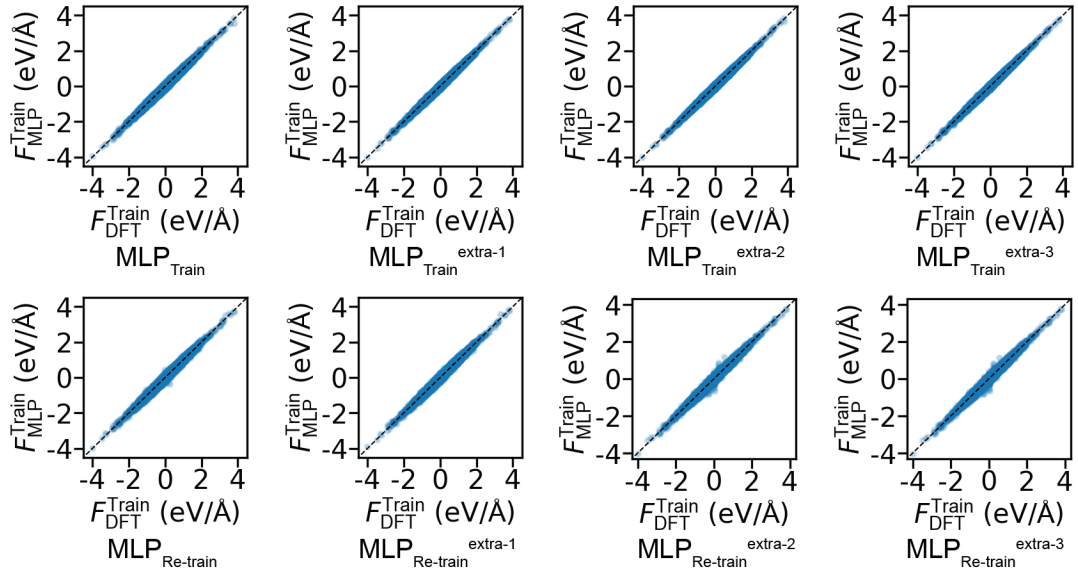


Figure E8. The atomic forces predicted by $\text{MLP}_{\text{Train}}$, $\text{MLP}_{\text{Train}}^{\text{extra-1}}$, $\text{MLP}_{\text{Train}}^{\text{extra-2}}$, $\text{MLP}_{\text{Train}}^{\text{extra-3}}$, $\text{MLP}_{\text{Re-train}}$, $\text{MLP}_{\text{Re-train}}^{\text{extra-1}}$, $\text{MLP}_{\text{Re-train}}^{\text{extra-2}}$, and $\text{MLP}_{\text{Re-train}}^{\text{extra-3}}$ for $\mathcal{D}^{\text{train}}$.

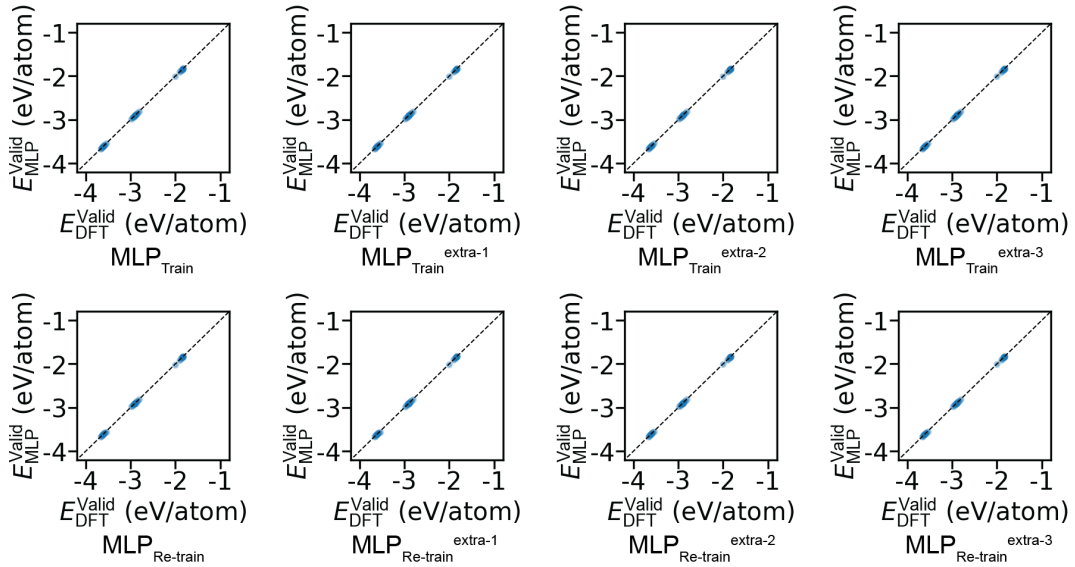


Figure E9. The energies predicted by $\text{MLP}_{\text{Train}}$, $\text{MLP}_{\text{Train}}^{\text{extra-1}}$, $\text{MLP}_{\text{Train}}^{\text{extra-2}}$, $\text{MLP}_{\text{Train}}^{\text{extra-3}}$, $\text{MLP}_{\text{Re-train}}$, $\text{MLP}_{\text{Re-train}}^{\text{extra-1}}$, $\text{MLP}_{\text{Re-train}}^{\text{extra-2}}$, and $\text{MLP}_{\text{Re-train}}^{\text{extra-3}}$ for $\mathcal{D}^{\text{validation}}$.

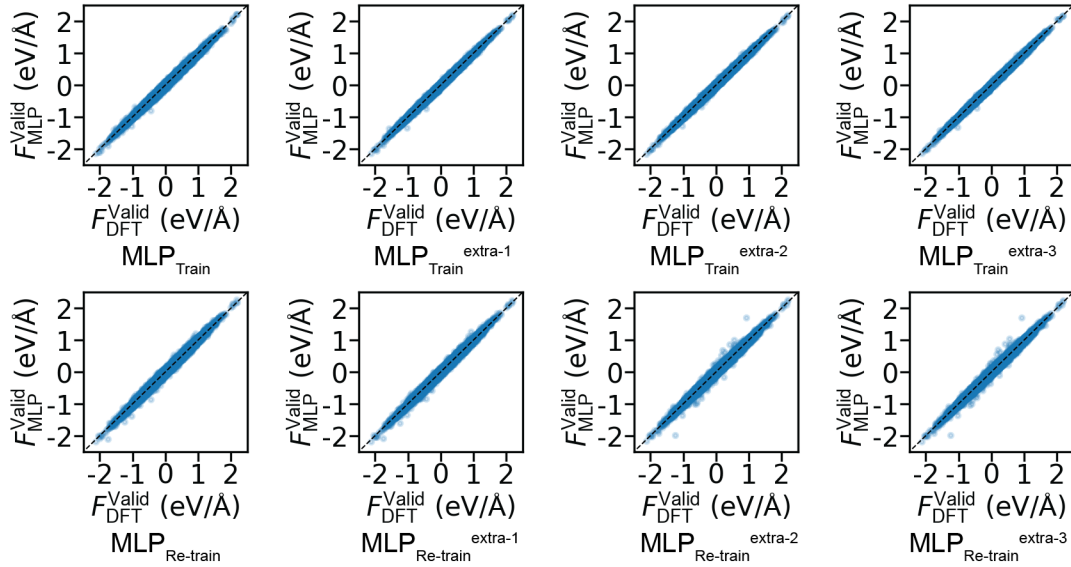


Figure E10. The atomic forces predicted by $\text{MLP}_{\text{Train}}$, $\text{MLP}_{\text{Train}}^{\text{extra-1}}$, $\text{MLP}_{\text{Train}}^{\text{extra-2}}$, $\text{MLP}_{\text{Train}}^{\text{extra-3}}$, $\text{MLP}_{\text{Re-train}}$, $\text{MLP}_{\text{Re-train}}^{\text{extra-1}}$, $\text{MLP}_{\text{Re-train}}^{\text{extra-2}}$, and $\text{MLP}_{\text{Re-train}}^{\text{extra-3}}$ for $\mathcal{D}^{\text{validation}}$.

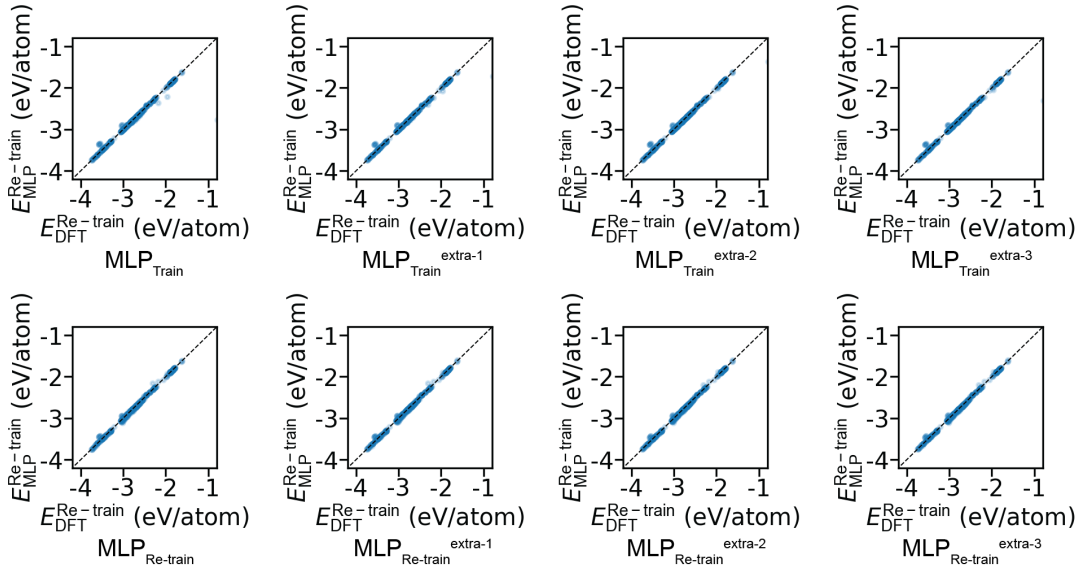


Figure E11. The energies predicted by $\text{MLP}_{\text{Train}}$, $\text{MLP}_{\text{Train}}^{\text{extra-1}}$, $\text{MLP}_{\text{Train}}^{\text{extra-2}}$, $\text{MLP}_{\text{Train}}^{\text{extra-3}}$, $\text{MLP}_{\text{Re-train}}$, $\text{MLP}_{\text{Re-train}}^{\text{extra-1}}$, $\text{MLP}_{\text{Re-train}}^{\text{extra-2}}$, and $\text{MLP}_{\text{Re-train}}^{\text{extra-3}}$ for $\mathcal{D}^{\text{re-train}}$.

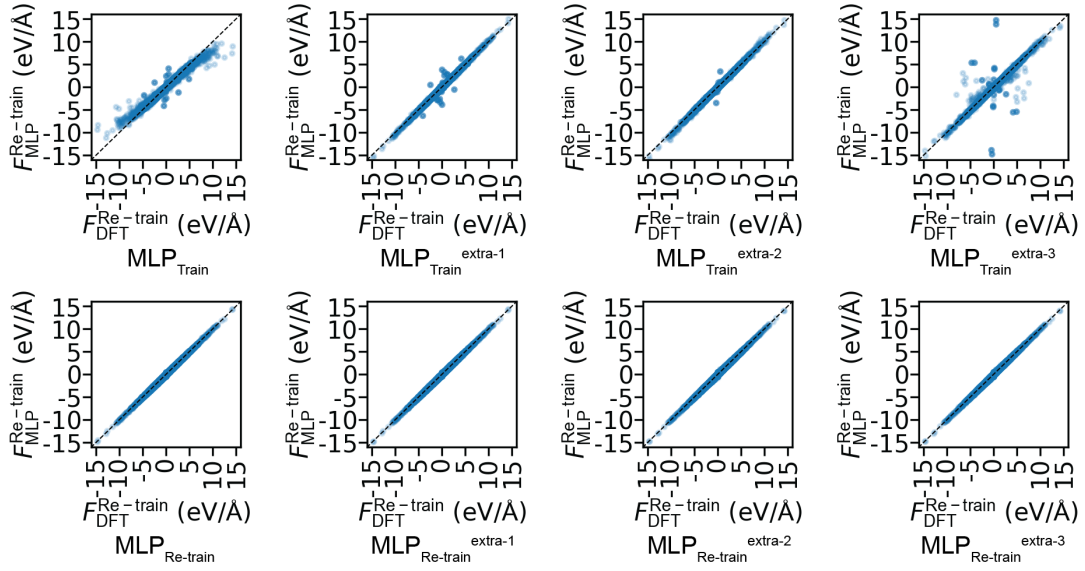


Figure E12. The atomic forces predicted by $\text{MLP}_{\text{Train}}$, $\text{MLP}_{\text{Train}}^{\text{extra-1}}$, $\text{MLP}_{\text{Train}}^{\text{extra-2}}$, $\text{MLP}_{\text{Train}}^{\text{extra-3}}$, $\text{MLP}_{\text{Re-train}}$, $\text{MLP}_{\text{Re-train}}^{\text{extra-1}}$, $\text{MLP}_{\text{Re-train}}^{\text{extra-2}}$, and $\text{MLP}_{\text{Re-train}}^{\text{extra-3}}$ for $\mathcal{D}^{\text{re-train}}$.

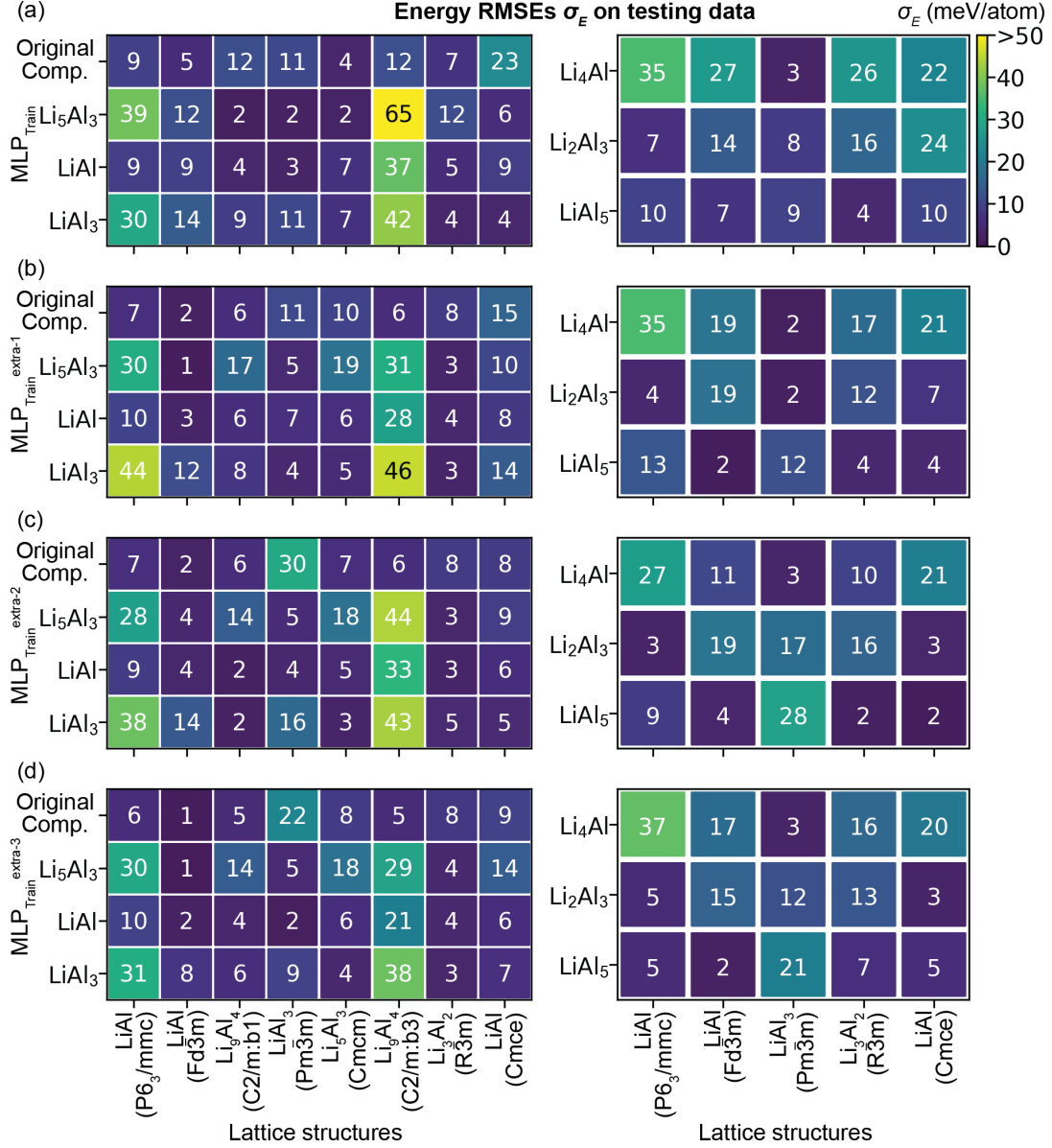


Figure E13. The energy RMSEs predicted by the (a) MLP_{Train}, additional (b) MLP_{Train}^{extra-1}, (c) MLP_{Train}^{extra-2}, and (d) MLP_{Train}^{extra-3} on $\mathcal{D}^{\text{test}}$.

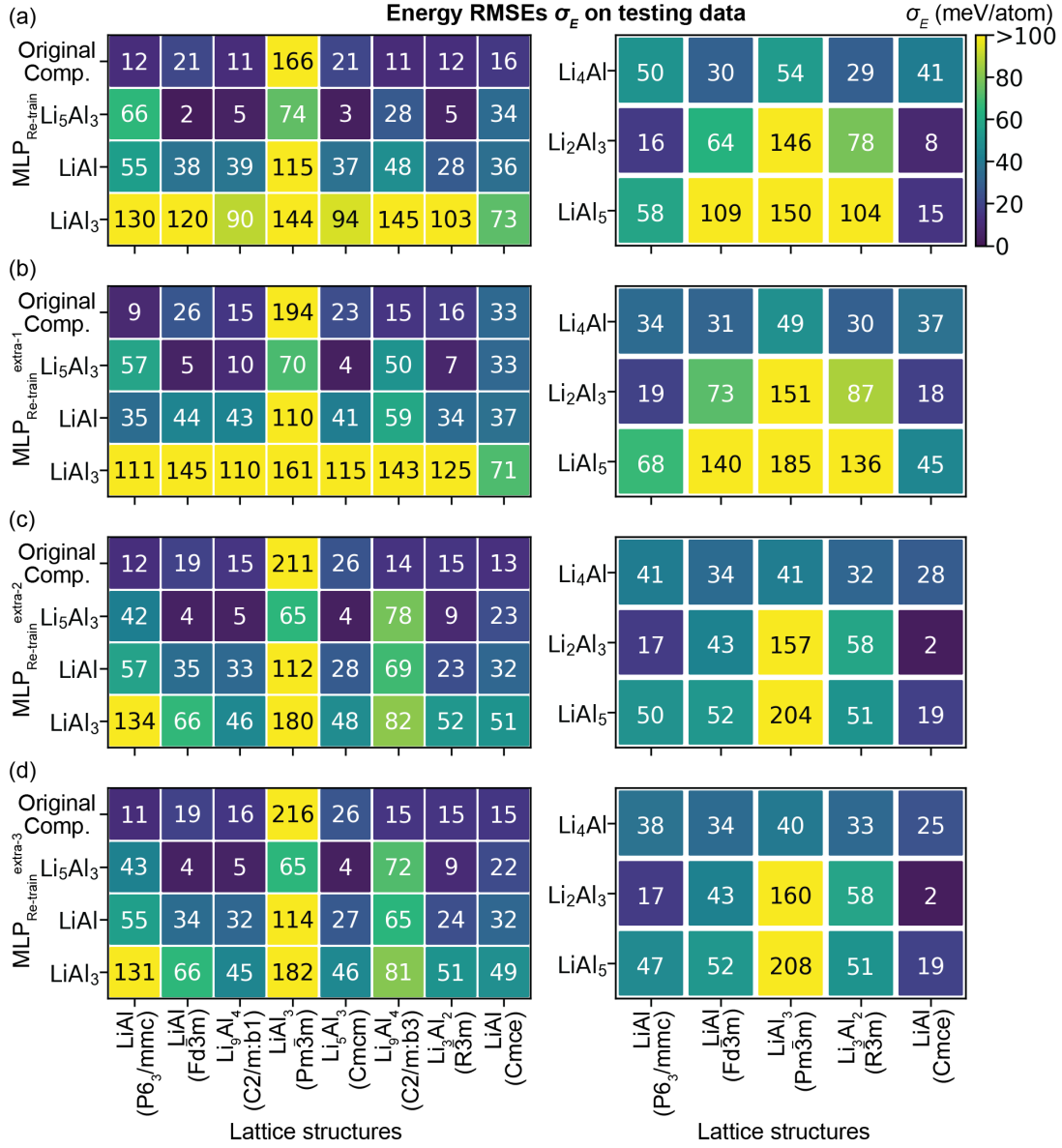


Figure E14. The energy RMSEs predicted by the (a) $\text{MLP}_{\text{Re-train}}$, additional (b) $\text{MLP}_{\text{Re-train}}^{\text{extra-1}}$, (c) $\text{MLP}_{\text{Re-train}}^{\text{extra-2}}$, and (d) $\text{MLP}_{\text{Re-train}}^{\text{extra-3}}$ on $\mathcal{D}^{\text{test}}$.

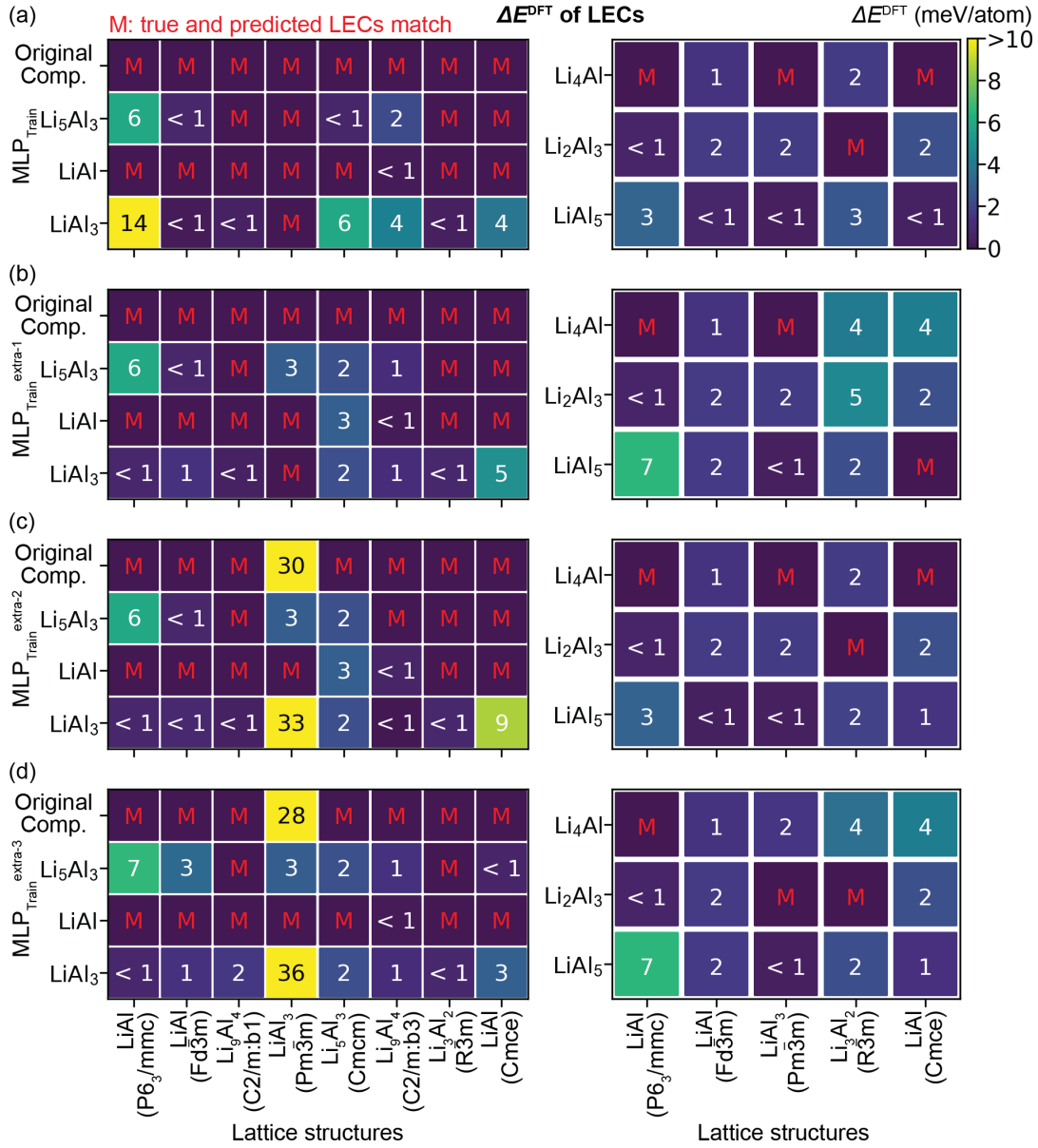


Figure E15. The ΔE^{DFT} of LECs predicted by the (a) MLP_{Train}, additional (b) MLP_{Train}^{extra-1}, (c) MLP_{Train}^{extra-2}, and (d) MLP_{Train}^{extra-3}. A match of LEC predicted by DFT and MLP is marked as red 'M'.

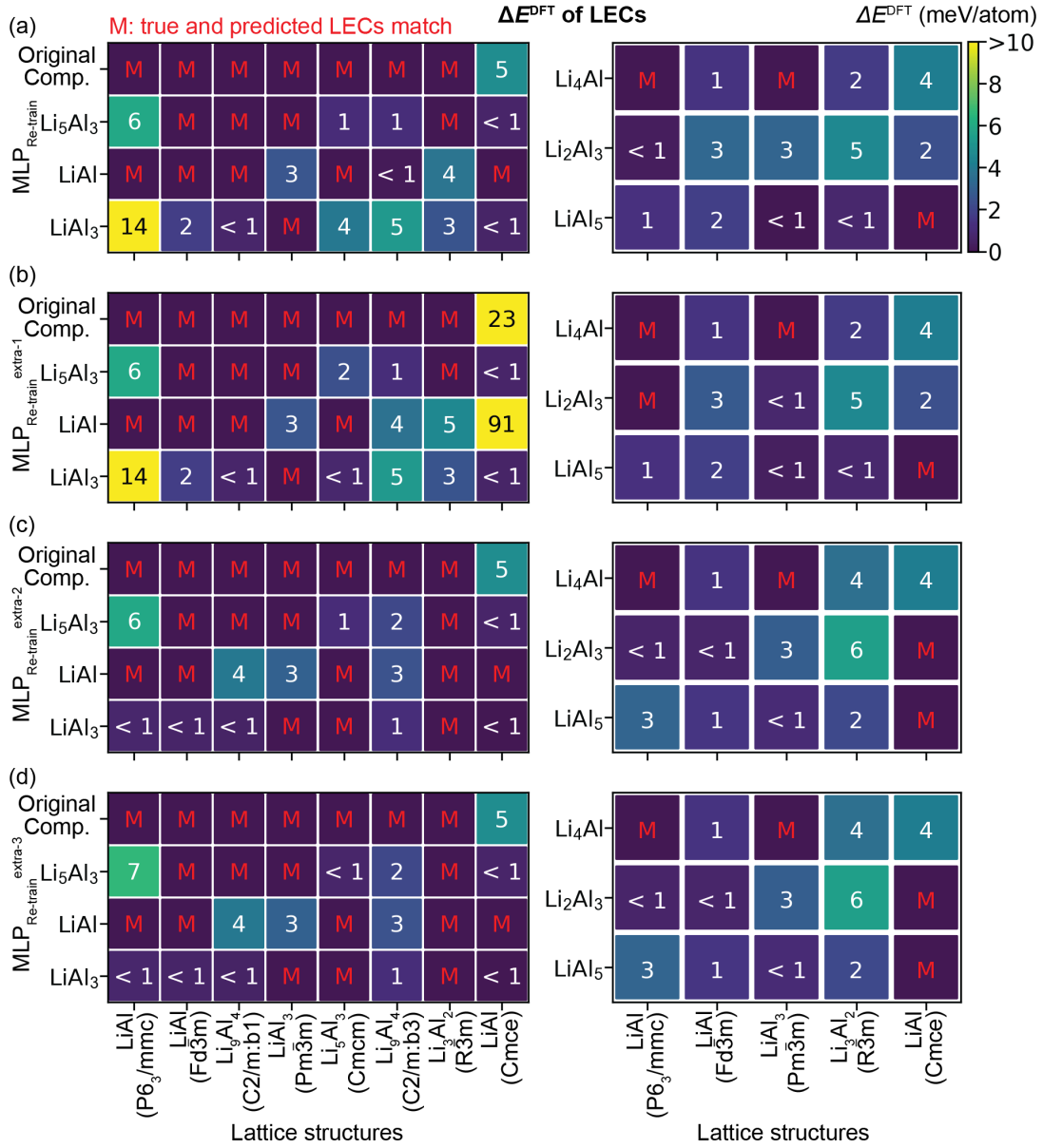


Figure E16. The ΔE^{DFT} of LECs predicted by the (a) MLP_{Re-train}, additional (b) MLP_{Re-train^{extra-1}}, (c) MLP_{Re-train^{extra-2}}, and (d) MLP_{Re-train^{extra-3}}. A match of LEC predicted by DFT and MLP is marked as red 'M'.

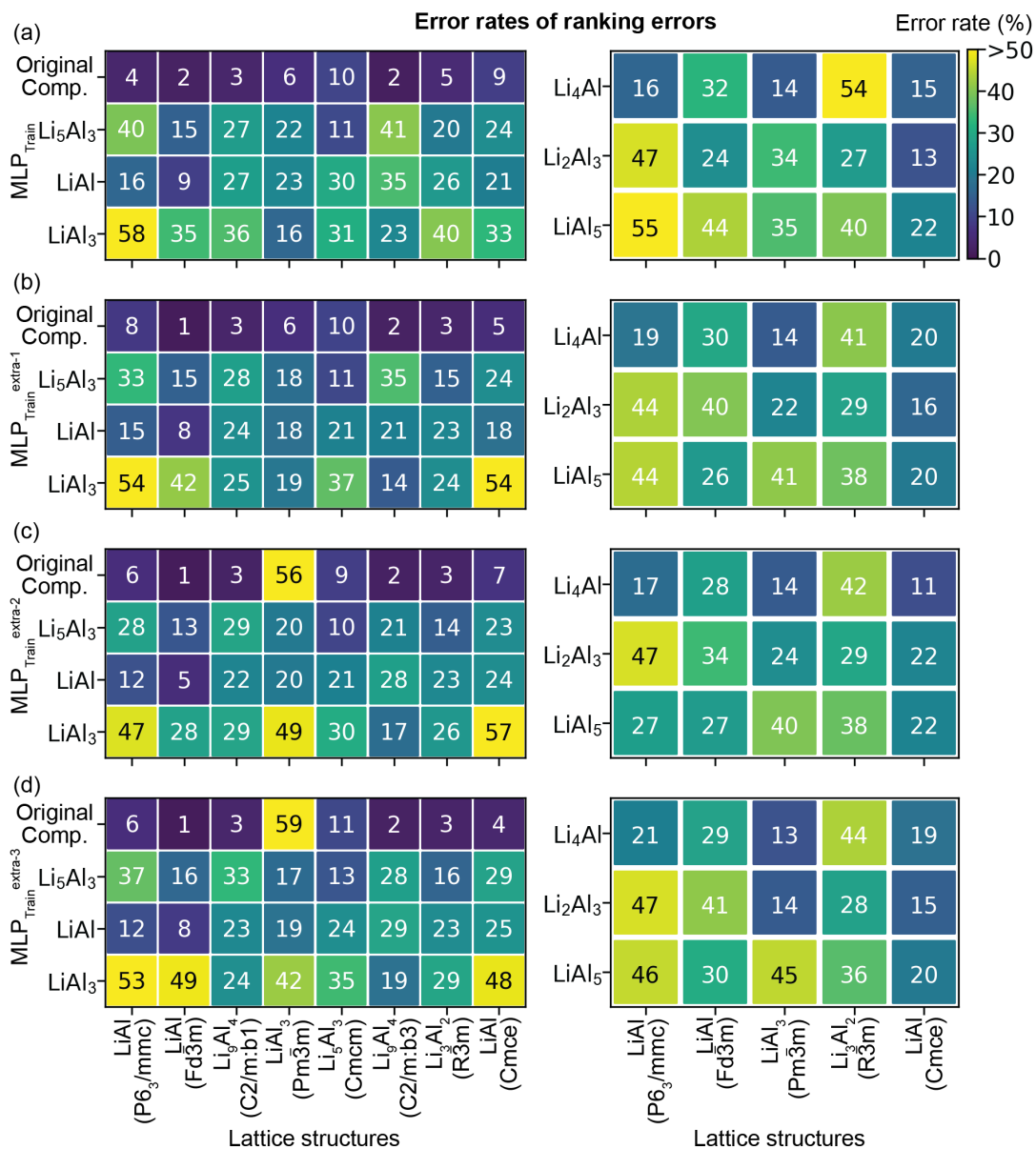


Figure E17. The ranking error rates predicted by the (a) MLP_{Train}, additional (b) MLP_{Train}^{extra-1}, (c) MLP_{Train}^{extra-2}, and (d) MLP_{Train}^{extra-3}.

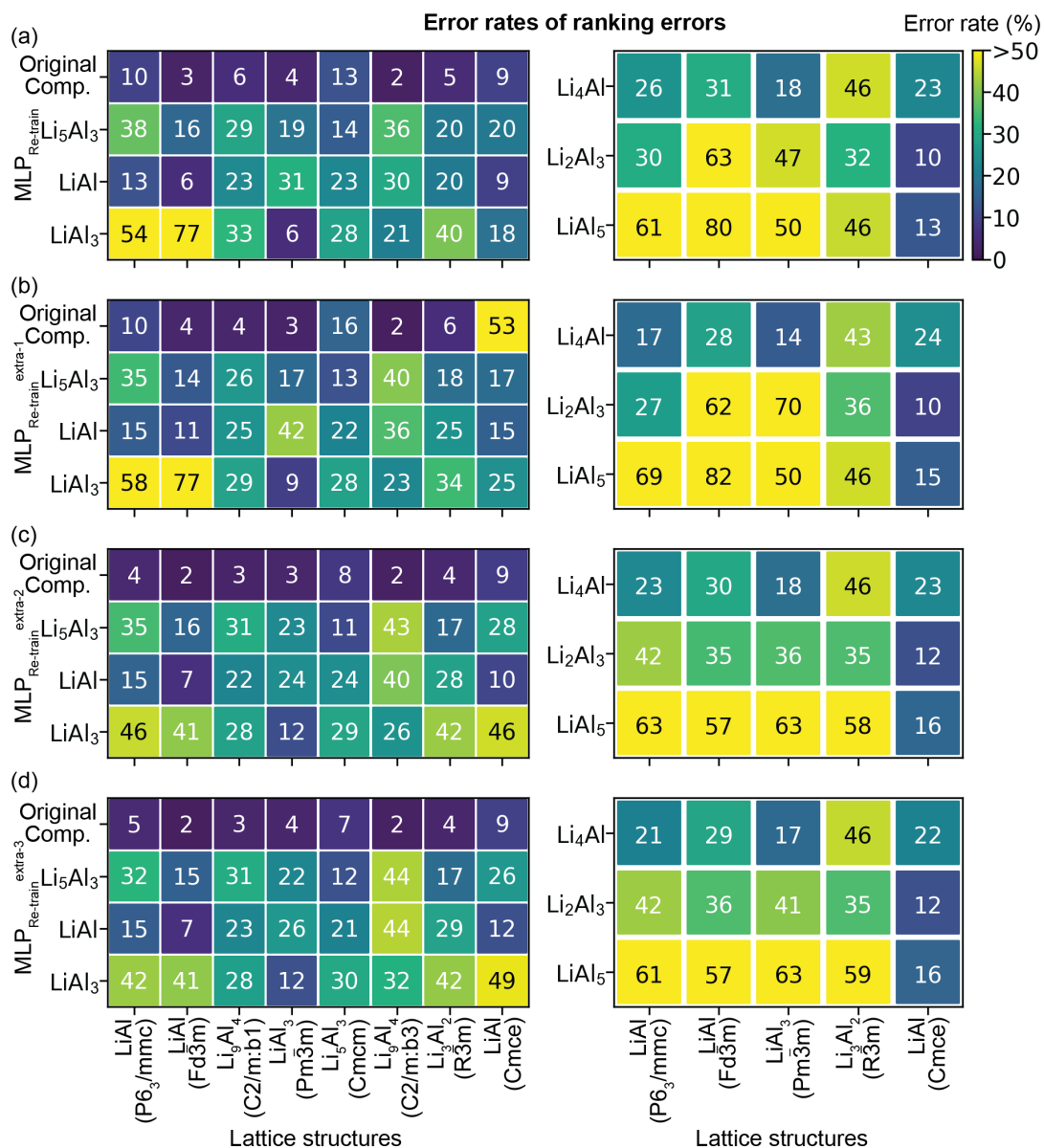


Figure E18. The ranking error rates predicted by the (a) $\text{MLP}_{\text{Re-train}}$, additional (b) $\text{MLP}_{\text{Re-train}}^{\text{extra-1}}$, (c) $\text{MLP}_{\text{Re-train}}^{\text{extra-2}}$, and (d) $\text{MLP}_{\text{Re-train}}^{\text{extra-3}}$.

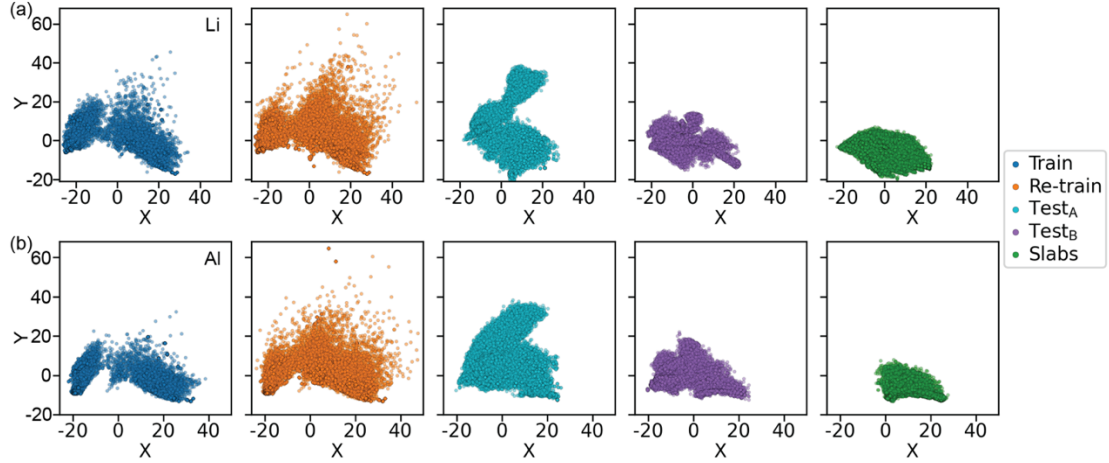


Figure E19. The atomic environments of (a) Li and (b) Al atoms in (blue) $\mathcal{D}^{\text{train}}$, (orange) $\mathcal{D}^{\text{re-train}}$, (cyan) $\mathcal{D}^{\text{test}}$ as in Chapter 4 (Figure 4.1c) (labeled as Test_A), (purple) $\mathcal{D}^{\text{test}}$ as in Chapter 4 (Figure 4.1d) (labeled as Test_B), and (green) $\mathcal{D}^{\text{slabs}}$ as in Chapter 4 Figure. 4.4. The x and y coordinates are the 1st and the 2nd principal components of Principal Component Analysis (PCA) based on Smooth Overlap of Atomic Positions (SOAP) descriptors, the same as in Chapter 4 (Figure 4.3a).

Table E2. The force performance scores, errors of force directions, and errors of force magnitudes for cumulative density function (CDF) curves at 90% probability on predicted forces of migrating atoms. All scores are calculated from AIMD snapshots for LiAl ($Fd3m$), Li₉Al₄ ($C2/m$: $b1$), and Li₅Al₃ ($Cmcm$) using MLP_{Train}, MLP_{Re-train}, and DFT calculations using single Γ -centered k -point (DFT K1).

Models	DFT K1	MLP _{Train}	MLP _{Re-train}
$P(\mathcal{D}^{\text{LiAl}})$	0.79	0.83	0.81
$ \delta_F $ at 90% of CDF on LiAl (eV/Å)	0.075	0.065	0.090
δ_θ at 90% of CDF on LiAl (°)	16.5	12	13
$P(\mathcal{D}^{\text{Li}_9\text{Al}_4})$	0.73	0.77	0.77
$ \delta_F $ at 90% of CDF on Li ₉ Al ₄ (eV/Å)	0.130	0.105	0.105
δ_θ at 90% of CDF on Li ₉ Al ₄ (°)	22.5	16.5	16.5
$P(\mathcal{D}^{\text{Li}_5\text{Al}_3})$	0.61	0.70	0.76
$ \delta_F $ at 90% of CDF on Li ₅ Al ₃ (eV/Å)	0.220	0.145	0.115
δ_θ at 90% of CDF on Li ₅ Al ₃ (°)	36	22	16.5

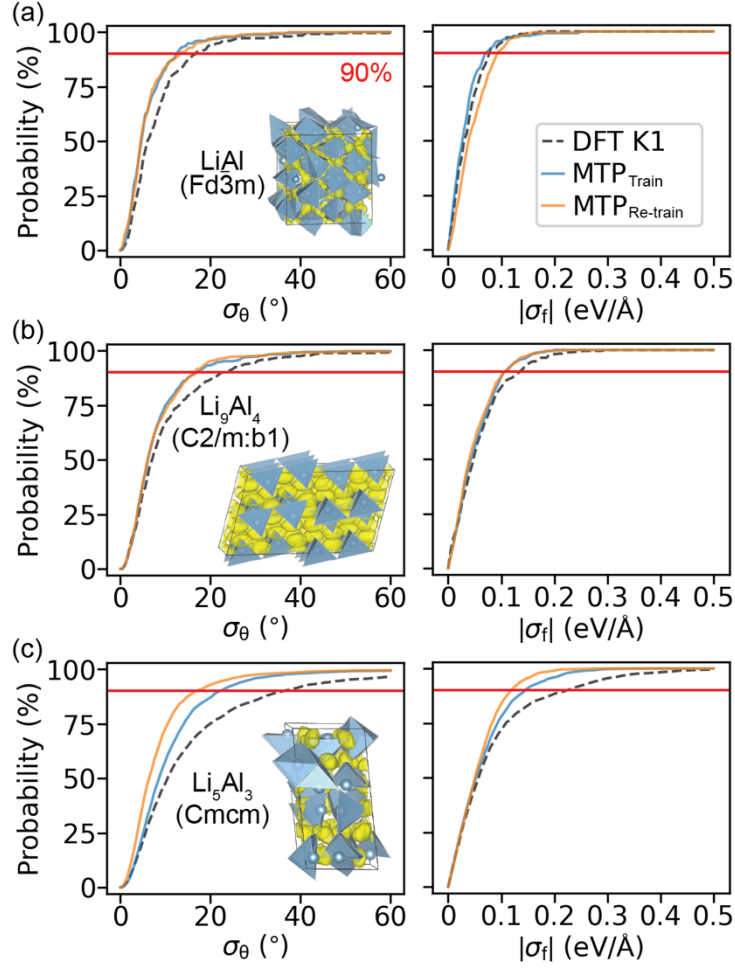


Figure E20. Errors of forces predicted by MLPs on the migrating atoms in AIMD simulations. The directional and magnitude accuracies of forces on migrating atoms in snapshots taken from AIMD simulations for intermediate Li-Al phases (a) LiAl ($Fd3m$), (b) Li_9Al_4 ($C2/m: b1$), and (c) Li_5Al_3 ($Cmcm$) with single Al vacancies. The CDF curves are plotted for forces (dashed black lines) calculated by DFT K1, (blue) predicted by $\text{MLP}_{\text{Train}}$, and (orange) by $\text{MLP}_{\text{Re-train}}$. Red lines are visual guidance for 90% cumulative probability, following the same scheme in Chapter 4 (Figure 4.6).

Appendix F

F1. First-principles computation

Vienna ab initio simulation package^[98] (VASP) with Perdew-Burke-Ernzerhof^[100] (PBE) functionals by generalized-gradient approximation (GGA) was used to perform all DFT calculations. The projector augmented-wave approach was adopted in DFT calculations to relax crystal structures and get energies and forces. All true values of energies and forces for training, validating, and testing MLPs were calculated using $4\times 4\times 4$ k -point mesh (K4). In addition, forces using Γ -centered $1\times 1\times 1$ k -point mesh (K1) were also calculated on migrating rare-event (RE) atoms in diffusions as benchmarks to compare with MLPs in Chapter 4.5.3. The Gaussian smearing method was adopted to calculate partial occupancies for each orbital (ISMEAR = 0) in combination with smearing width set to 0.05 eV (SIGMA = 0.05). Furthermore, all DFT calculations were spin-polarized with an energy cutoff set to 520 eV, electronic relaxation cutoff set to 10^{-7} eV, and other parameters set to be compatible with the Materials Project.^[102,105]

F2. Ab initio molecular dynamics simulation

Ab initio molecular dynamics (AIMD) simulations were performed as the same scheme described in Chapter 2, Chapter 3, Appendix B, and Appendix D. All simulations adopted GGA for PBE functionals, as described in *First-Principles Computation* section. In addition, the AIMD simulations were non-spin-polarized, with the setting of electronic energy convergence cutoff to 10^{-4} eV, a time step of 2 fs, and a Γ -centered $1\times 1\times 1$ k -point mesh. All simulations adopted the NVT ensemble with Nose-Hoover thermostat (SMASS = 0). The initial structures were heated up from 100 K to the target temperatures with a constant heating rate using velocity scaling during a period of 2 ps. All supercells for AIMD simulations have lattice parameters at around 10 Å or

larger. The temperatures of 600 K, 2000 K, and 400 K were used to generate the near-equilibrium snapshots of relaxed intermediate phases, liquids of intermediate phases, and interfaces, respectively.

The diffusion simulations for the selected Li_5Al_3 ($Cmcm$) and Li_3Al_2 ($R\bar{3}m$) intermediate phases in Chapter 4.5.3 were carried out following the established process in Chapter 2, Chapter 3, Appendix B, and Appendix D. Since counting atom hopping is stochastic in nature, all diffusion simulations were converged following the scheme in Ref.^[4] To obtain adequate migrating atoms, the total mean-squared-displacements of Li atoms were in the range between 150 to 517 $\text{\AA}^2$ within 400 ps for Li_5Al_3 ($Cmcm$), and between 180 and 4300 $\text{\AA}^2$ during 600 to 1000 ps for Li_3Al_2 ($R\bar{3}m$). The diffusional properties of Li_5Al_3 ($Cmcm$) were evaluated at the temperatures of 400, 450, 500, 550, and 600 K, while the diffusional properties of Li_3Al_2 ($R\bar{3}m$) were evaluated at the temperatures of 400, 450, 500, 550, 600, 650, and 700 K.

F3. Datasets for training, validation, and testing

A total of eight datasets were generated for training, validating, and testing MLPs, such as two training datasets $\mathcal{D}^{\text{train}}$ and $\mathcal{D}^{\text{re-train}}$, two validation datasets $\mathcal{D}^{\text{validation}}$ and $\mathcal{D}^{\text{irregular}}$, and four testing datasets $\mathcal{D}^{\text{test}}$, $\mathcal{D}^{\text{slabs}}$, $\mathcal{D}^{\text{Vacancy}}$, and $\mathcal{D}^{\text{bcc/fcc}}$. The energies and the forces of all configurations in these datasets (except for $\mathcal{D}^{\text{irregular}}$) were converged using K4 by single-step self-consistent DFT calculations (unless indicated otherwise).

The $\mathcal{D}^{\text{train}}$ dataset contains 356 configurations based on BCC Li metal ($Im\bar{3}m$, ICSD 77370), LiAl ($Fd\bar{3}m$, ICSD 240122), and FCC Al metal ($Fm\bar{3}m$, ICSD 606000). All supercells in $\mathcal{D}^{\text{train}}$ have the lattice parameters around or larger than 10 Å. For each of these three phases, a variety of configurations were generated as follows:

- (1) The ground-state configuration that was fully relaxed by DFT K4.
- (2) Five near-equilibrium configurations of the ground-state generated by NVT AIMD simulations at 600 K. Each of the near-equilibrium configurations was generated by a separate AIMD simulation of 2 ps.
- (3) Two liquids configurations generated by NVT AIMD simulations at 2000 K, following the same process as 2).
- (4) Strained supercells with a) -15% or +15% strain on each of the three lattice parameters, and b) -15% or +15% strain on a pair of distinct lattice parameters with four combinations (+15%, +15%), (-15%, -15%), (+15%, -15%), and (-15%, +15%) strains.
- (5) Distorted supercells generated by distorting each of the three lattice angles by -13.5° or 13.5° for cubic lattices (-15% or +15% of corresponding lattice angles if lattice angles are not 90°).
- (6) Supercells with vacancies. We removed 5%, 10%, or 15% of atoms in the supercell of Li and Al. For LiAl ($Fd\bar{3}m$), the defected supercells with either Li or Al vacancies were generated,

but not both Li and Al vacancies at the same time. All supercells with vacancies were then fully relaxed by DFT K4.

- (7) Five near-equilibrium configurations for each supercell with vacancies. All configurations were generated by NVT AIMD simulations at 600 K, following the same process in 2).
- (8) Supercells with interstitials. We added one, two, or four atoms of either Li or Al in the supercells of Li, Al, and LiAl. All bulk supercells were then fully relaxed by DFT K4.
- (9) Five near-equilibrium configurations for each of the supercells with interstitials. All configurations were generated by NVT AIMD simulations at 600 K, following the same process in 2).
- (10) Additionally, 20 interfaces with every pair of two phases in the dataset. All 20 interfaces were generated by directly joining the conventional unit cell of each material by the Zur and McGill algorithm^[114] implemented in Pymatgen^[105] Python package. The maximum matching area of Zur and McGill algorithm was set to $400 \text{ \AA}^2$ and the maximum angle tolerance was set to 0.01. Slabs of both materials on either side of the interface had four layers of unit cells. We generated interfaces using the Miller indices of $\{100\}$, $\{110\}$, and $\{111\}$. The gap distance and the vacuum space at the interface core were both set to $2.5 \text{ \AA}$. Due to the high computational cost of evaluating energies and forces for these large interface supercells, we excluded the supercells that had more than 250 atoms or had lattice parameters larger than $25 \text{ \AA}$. For each interface supercell, four configurations were generated as follows,
 - i. A supercell of the interface configuration without relaxing lattice or atom positions.
 - ii. A supercell of the interface configuration relaxed by DFT K1. The energy and atomic forces of the relaxed supercell were then calculated by DFT K4 self-consistent run with fixed lattice and atom positions.

- iii. Two near-equilibrium configurations for the relaxed supercell generated by NVT AIMD simulations at 400 K.

The $\mathcal{D}^{\text{re-train}}$ dataset contains 414 configurations based on four phases: BCC Li metal ($Im\bar{3}m$, ICSD 77370), LiAl ($Fd\bar{3}m$, ICSD 240122), LiAl ($P6_3/mmc$, ICSD 262069), and FCC Al metal ($Fm\bar{3}m$, ICSD 606000). All configurations were generated following the same process as $\mathcal{D}^{\text{train}}$, except for near-equilibrium configurations generated in 2), 7), and 9). In $\mathcal{D}^{\text{re-train}}$, near-equilibrium configurations were generated by displacing each atom to a random direction and the displacement of each atom was randomly selected using a uniform distribution between 0 and 0.5 Å, and only three near-equilibrium configurations were generated for 2), 7), and 9) in $\mathcal{D}^{\text{re-train}}$. A total of 42 interface supercells were generated in the $\mathcal{D}^{\text{re-train}}$ following the same process of $\mathcal{D}^{\text{train}}$. Three configurations of each interface supercell were generated, including a supercell configuration without relaxation, a supercell configuration relaxed by DFT K1, and a near-equilibrium configuration generated by the random displacement.

The $\mathcal{D}^{\text{validation}}$ dataset contains 66 configurations generated from NVT AIMD simulations of relaxed bulk, vacancy, and interstitial supercells at 600 K, following the same process used in 2) for $\mathcal{D}^{\text{train}}$. The bulk, vacancy, and interstitial supercells were the same as 1), 6), and 8) in $\mathcal{D}^{\text{train}}$ (a total of 33 supercells), and two near-equilibrium configurations of each supercell were taken following the same as 2), 7), and 9) in $\mathcal{D}^{\text{train}}$.

The $\mathcal{D}^{\text{test}}$ dataset contains 1391 configurations based on eight intermediate phases from the ICSD and 39 hypothetical phases (Figure 3.1c and d). Hypothetical phases were generated using the lattice structures of eight known intermediate phases and changing the site occupancies to given compositions, consisting of 24 in three compositions LiAl₃, LiAl, and Li₅Al₃ in eight

known intermediate phases (Figure 3.1c) and 15 phases with handpicked compositions, LiAl_5 , Li_2Al_3 , and Li_4Al (Figure 3.1d). The configurations for the dataset were generated by swapping random numbers of Li and Al atoms in the supercell as follows. First, we generated a random number n between 1 and 10% of the total number of atoms in the supercell. Then, we randomly picked n Li and n Al atoms, and swapped their positions. We generated up to 30 configurations for each intermediate phase. All configurations were symmetrically distinct to each other, according to the established scheme in Ref.^[29,115] The hypothetical phases were generated by adjusting their compositions from the original intermediate phases to the target compositions in the relaxed supercells. The compositions were adjusted by correspondingly changing the occupancies of Li and Al sites in the relaxed supercells of the original intermediate phases, and the partial occupancies may be adjusted to include vacancies to guarantee the numbers of Li and Al atoms are integers (such as Li_5Al_3 with lattice structures $Pm\bar{3}m$, $C2/m:b3$, and $R\bar{3}m$, and Li_4Al with lattice structures $P6_3/mmc$, $Pm\bar{3}m$, and $Cmce$). Then, the structure with modified site partial occupancies was ordered following the scheme in Ref.^[29] We generated up to 30 configurations for each hypothetical phase. All configurations were symmetrically distinct from each other.

The $\mathcal{D}^{\text{slabs}}$ dataset contains 420 configurations of interface supercells with a range of compositions with different orderings generated by Monte Carlo (MC) methods. Interface supercells in $\mathcal{D}^{\text{slabs}}$ were between BCC metal Li and BCC LiAl ($Fd\bar{3}m$) based on the relaxed conventional unit cell, following the same process of generating interfaces in $\mathcal{D}^{\text{train}}$. The maximum matching area of the Zur and McGill algorithm was set to $100 \text{ \AA}^2$ and the maximum angle tolerance was set to 0.01. The Miller indices of both the substrate and the film were set to $\{100\}$. four types of supercells were used, $8.9 \times 4.5 \times 60.7 \text{ \AA}^3$, $8.9 \times 8.9 \times 34.7 \text{ \AA}^3$, $8.9 \times 8.9 \times 47.3 \text{ \AA}^3$, and $13.4 \times 8.9 \times 28.4 \text{ \AA}^3$, corresponding to $2 \times 2 \times 5.5$, $2 \times 2 \times 7.5$, $2 \times 1 \times 9.5$, and $3 \times 2 \times 4.5$ of the

conventional LiAl ($Fd\bar{3}m$) unit cell (Figure 4.4). Restricted by the total number of atoms in the supercell, the compositions of supercells were handpicked and rounded to the closest Al concentration at 20%, 21%, 25%, 33%, 34%, 36%, 38%, 40%, and 45%. Compositions were changed by adjusting the number of layers of unit cells for Li and LiAl. A total of 21 different compositions in four supercells were generated by MC simulations using MTP_{Train} (Figure 4.4). A MC simulation at 300 K for each interface supercell was performed to find the lowest energy configurations by swapping atom positions of a random Li and a random Al for each attempted random move. The MC simulations were terminated when the relative standard deviations of the total energies of the last 1000 attempts were lower than 10^{-5} . During the MC simulations, supercells were relaxed with energy convergence set to 10^{-5} . The MC simulations were conducted using LAMMPS by our house-customized scripts. The final 20 configurations that were accepted in each simulation were included in $\mathcal{D}^{\text{slabs}}$, with their energies were calculated by DFT K4 without relaxing lattices or atom positions.

The $\mathcal{D}^{\text{Vacancy}}$ dataset contains 350 configurations based on eight intermediate phases with a single vacancy of either Li or Al atom (see Appendix E). In the relaxed supercells of intermediate phases, a random Li or Al atom was then chosen, swapped its position with any other atom (either Li or Al), and then removed. By using this method, we generated up to 30 configurations that were symmetrically distinct (including an anti-site in some cases) for each phase containing one vacancy.

The $\mathcal{D}^{\text{bcc/fcc}}$ dataset contains 701 configurations based on the structures with BCC and FCC lattices with 12 compositions with Al % of 52%, 54%, 56%, 57%, 60%, 62%, 64%, 66%, 68%, 70%, 72%, and 74% for each lattice. The structure lattice of metal Li ($Im\bar{3}m$) and LiAl ($Fd\bar{3}m$) for BCC, and metal Al ($Fm\bar{3}m$) and LiAl₃ ($Pm\bar{3}m$) for FCC, were adopted to generate

configurations with different elemental orderings. This led to a total of 48 hypothetical phases were generated. Compositions of these phases were adjusted following the same steps in $\mathcal{D}^{\text{test}}$. For each phase, following the same process of ordering of hypothetical phases in $\mathcal{D}^{\text{test}}$, up to 10 symmetrically distinct configurations were generated for metal Li ($Im\bar{3}m$) and Al ($Fm\bar{3}m$), and up to 20 configurations were generated for LiAl ($Fd\bar{3}m$) and LiAl₃ ($Pm\bar{3}m$).

The $\mathcal{D}^{\text{irregular}}$ dataset was a validation dataset only used to select the optimized MLPs as described in the F5. This dataset contains 141 configurations, which were intentionally generated as far-from-equilibrium configurations to give either exceptionally low energies predicted by MLP_{Train}^{extra-2} (Appendix E) (< -10 eV/atom, compared with -1.9 eV/atom for bulk Li and -3.7 eV/atom for bulk Al) or exceptionally high predicted forces (> 15 eV/Å on any atom). Since the configurations in this dataset were far-from-equilibrium, predicted energies were expected to be high (> -4 eV/atom for typical Li_xAl_y). Thus, the exceptionally low predicted energies contradict to the physical nature of far-from-equilibrium configurations. This validation dataset was used to exclude those MLPs that have exceptionally low predicted energies, which can be viewed as a failure in capturing the correct physical interactions of Li-Al materials system. These 141 configurations were selected from the snapshots produced by performing MLP-MD simulations using grand canonical ensemble (allowing the change of supercell size and the number of Li atoms) at 600 K. The supercell of LiAl ($Fd\bar{3}m$) has a size of $25.3 \times 25.3 \times 24.0$ Å³ and contains 576 Al atoms with varying Li atoms (289 to 578), and another supercell is the interface between bulk Li and Al with the size of $27.9 \times 19.7 \times 22.6$ Å³ and 384 Al atoms with varying Li atoms (157 to 260). A random amount of Li atoms was removed (varying between 1 to 287 Li atoms removed for supercells with 576 Al atoms, and between 124 to 227 Li atoms removed for supercells with 384 Al atoms) or inserted (< 2 Li atoms inserted for supercells with 576 Al atoms) from the

supercells in the dataset. During the MD simulations using grand canonical ensemble, an attempt to either remove or insert (50% probability for either action) a Li atom is tried every 0.1 ps. Then, by fixing the number of Al atoms, the acceptance of each attempt is determined by the final energies predicted by MLP with a given constant chemical potential of Li only,^[116,117] following the Metropolis algorithm in Monte Carlo simulations. The probability of accepting a removal/insertion attempt using the final predicted energies and Li chemical potential is calculated following the equations developed by Ref^[116,117]. Many of these generated configurations were intentionally generated to be less reasonable, far-from-equilibrium configurations, as some atoms get as close as within 1 Å, which were given by the erroneous prediction (very low energies) of MLP_{Train}^{extra-2}. These configurations produced by MLP_{Train}^{extra-2} (Appendix E) during MLP-MD simulations using grand canonical ensemble were re-used in the validation process (Appendix F5) to select MLPs having the minimum amount of erroneously predicted energies (< -10 eV/atom). For some configurations with atoms very close to each other, large forces were predicted by the MLP_{Train}^{extra-2} even on atoms that were far away (> 5 Å) from the erroneous pairs. Given The $\mathcal{D}^{\text{irregular}}$ dataset was a validation dataset only used to select the optimized MLPs as described in the F5, no DFT values are needed.

F4. Training MLPs

The MTP models were trained using the same scheme implemented in MAML^[118] Python package interfaced with MLIP for MTP^[15,110] as in previous studies (Zuo et al.^[1] and Chapter 3). A grid search approach was employed to identify the combination of hyperparameters, including the cutoff radius, the choice of the radial basis function sets, and the maximum number of iterations, as in Zuo et al.^[1] and Chapter 3. Two to ten values were tested for each hyperparameter, and there were a total of 2488 MTPs trained by $\mathcal{D}^{\text{train}}$ and 1537 MTPs trained by $\mathcal{D}^{\text{re-train}}$.

F5. Optimizing MLPs and fine tuning the hyperparameters

The processes of evaluating the validation scores, fine tuning the hyperparameters, and selecting the optimal MLPs with the best scores were as follows. To calculate the validation scores, 10 criteria, consisting of the RMSE energies and forces and the NAC of force errors, were used:

- (1) The RMSE of energies for training data $\mathcal{D}^{\text{train}}/\mathcal{D}^{\text{re-train}}$ (with respect to the training process of MLP_{Train} or MLP_{Re-train}), σ_E^{train} ;
- (2) The RMSE of energies for validation data $\mathcal{D}^{\text{validation}}$, $\sigma_E^{\text{validation}}$;
- (3) The RMSE of forces for $\mathcal{D}^{\text{train}}/\mathcal{D}^{\text{re-train}}$, σ_F^{train} ;
- (4) The RMSE of forces for $\mathcal{D}^{\text{validation}}$, $\sigma_F^{\text{validation}}$;
- (5) The normalized area of the cumulative distribution function curve (NAC) of errors on force magnitudes for $\mathcal{D}^{\text{train}}/\mathcal{D}^{\text{re-train}}$, $\Delta NAC(|\delta_F|, \mathcal{D}^{\text{train}}) = 1 - NAC(|\delta_F|, \mathcal{D}^{\text{train}})$;
- (6) The NAC of errors on force directions for $\mathcal{D}^{\text{train}}/\mathcal{D}^{\text{re-train}}$, $\Delta NAC(\delta_\theta, \mathcal{D}^{\text{train}}) = 1 - NAC(\delta_\theta, \mathcal{D}^{\text{train}})$;
- (7) The NAC of errors on force magnitudes for $\mathcal{D}^{\text{validation}}$, $\Delta NAC(|\delta_F|, \mathcal{D}^{\text{validation}}) = 1 - NAC(|\delta_F|, \mathcal{D}^{\text{validation}})$;

- (8) The NAC of errors on force directions for $\mathcal{D}^{\text{valid}}$, $\Delta NAC(\delta_\theta, \mathcal{D}^{\text{validation}}) = 1 - NAC(\delta_\theta, \mathcal{D}^{\text{validation}})$;
- (9) The number of large forces predicted for all atoms in $\mathcal{D}^{\text{irregular}}$, N_F ;
- (10) The number of configurations that the energies predicted were < -10 eV/atom in $\mathcal{D}^{\text{irregular}}$, N_E ;

The NAC s were the normalized area of the cumulative distribution function (CDF) curves as we proposed in Chapter 3 and were calculated following the same steps in Chapter 3.4.1. The CDF curves of force errors were generated to show the distributions of errors over a specific error range, and the NAC was then calculated as dividing the area under the CDF curve by the total area (between 0 to 100% on y -axis and the specified error range on x -axis). Here, the NAC of errors on force magnitudes, $NAC(|\delta_F|, \mathcal{D})$, was calculated by using the CDF curves over an error range of 0 to 1.5 eV/Å, and the NAC of errors on force directions, $NAC(\delta_\theta, \mathcal{D})$, was calculated by using the CDF curves over an error range of 0 to 180 degrees.

The value of each criterion, c , was then standardized by subtracting its minimum value $\min(c)$ and dividing by its standard deviation $\text{std}(c)$,

$$v = [c - \min(c)] / \text{std}(c).$$

The total validation score was calculated as

$$V = \sum_v (w_v v)^2,$$

where the weights w_v of N_F and N_E were set to 1.5 and w_v of all the other criteria were set to 1.

The MLPs with the lowest total validation score V were selected as $\text{MLP}_{\text{Train}}$ and $\text{MLP}_{\text{Re-train}}$.

F6. Principal component analysis of atomic environments

We adopted the Smooth Overlap of Atomic Positions (SOAP) descriptors to quantify the atomic environments in all datasets. The SOAP descriptors were calculated using the QUIP^[11] package and the parameters set as follows (or default values if not specified),

- (1) The band limit of spherical harmonics basis function (l_{max}) was set to 12.
- (2) The number of radial basis function (n_{max}) was set to 6.
- (3) The covariance function type was set to ‘dot_product’ and its zeta was set to 4.
- (4) The cutoff radius was set to 6 Å.
- (5) The number of sparse points was set to 4000.

We visualize the quantified atomic environments by performing the principal component analysis (PCA) function implemented in *sci-kit learn* package. We first standardized all elements of the SOAP descriptors by subtracting the means and dividing by the standard deviations. Then, we conducted PCA and plotted the 1st and the 2nd principal component (Figure 4.3 and Figure E19).

F7. Identifying migrating atoms and evaluating force performance

The process of identifying migrating atoms to evaluate force performance of MLPs (Chapter 4.5.3) follows the same approach in Chapter 3.4.1, and was explained in Appendix D. Since most of the diffusion events are by Li atoms in Li-Al materials systems, migrating Li atoms were identified to evaluate MLP performances on migrating atoms. Li migrating atoms were selected following a similar approach as identifying migrating atoms in Chapter 3 (Appendix D). If the distances of the atom between their 1st and 2nd nearest static sites have a difference below 0.9 Å (approximately 33% of distances between two static sites), the atom is identified as migrating

atom. The errors of force magnitudes and force directions between the true and the predicted atomic forces were then calculated for these migrating atoms, and the error distribution were plotted as the CDF curves in Chapter 4.5.3 (Figure 4.6).

Appendix G

1. Pareto fronts of MLPs and “outlier” MLPs

Our analyses of the optimal MLPs on Pareto fronts of the MLP pool may be beneficial for selecting and comparing MLP models with different descriptors, training data, etc. These analyses test a set of representative MLPs instead of an optimized single MLP model, as a single optimized model may be susceptible to many factors, such as hyperparameters, training data and process, validation, and many more.

In a typical MLP training process, the selection of optimized MLPs is based on a few metrics, most often the energy and force RMSEs of training datasets and of validation datasets. These metrics can be combined to calculate a validation score, e.g.,

$$Score = \sum_i (w_i M_i^{\text{validation}})^2, \quad \text{Eq. (G1)}$$

where $M^{\text{validation}}$ is the validation metric, and w is the weight of $M^{\text{validation}}$. The MLP having the best validation score is often selected as the optimized MLP for further testing. One may choose

different weights and the formulas of validation scores, e.g., $\sqrt{\frac{\sum_i^N (w_i M_i^{\text{validation}})^2}{N}}$ or $\frac{\sum_i^N w_i |M_i^{\text{validation}}|}{N}$,

which would produce different ‘optimized’ MLPs with different performances. It is not obvious what are the best choices of the weights and the formulas of validation scores to achieve the optimal MLPs. Having different validation scores may greatly affect the choice of the optimal models and their performance, thus making the analyses and comparison of MLPs difficult as there are many other factors involved. Using optimal MLPs on the Pareto fronts (the vertices on the Pareto fronts) based on a collection of MLPs may help circumvent the issue of choosing the weights or formulas of validation scores.

Given the importance of the Pareto fronts of MLPs, sampling MLPs to construct Pareto fronts that clearly show the boundary of MLP performance on each error metrics is essential. Our

study uses both optimized MLPs and MLPs that are randomly sampled from the model pool of the MLPs. Though many MLPs randomly sampled from the model pool have bad performances, a few of them, which are referred as “outliers”, may exhibit good performance on only a very limited number of properties but not others. These outlier MLPs are usually excluded from the optimization step in typical MLP training processes, because they perform poorly on the many other metrics, but our random sampling scheme can still include these MLPs for our analysis purposes. For example, the outlier MLP (Figure G1d) has an error of 3.7 eV on E_f^{vacancy} . Similar outlier MLPs can be observed for many properties (as shown in Figure G2, G3, and G4).

These outliers provide interesting insights into the upper limits of MLP performance on certain properties or metrics in forming complete Pareto fronts, such as the ones highlighted by green boxes (**Figure G1**). For example, the outlier MLPs (Figure G1c) and d can achieve an accuracy lower than 3% error rates on energy rankings, while the error rates of other MLPs are higher than 20%, compared with the $\Delta_{\text{DFT}}^{\text{K1 to K4}}$ benchmark of 7%. These outlier MLPs suggest that having good predictions on energy rankings can be achieved using current MLP models and descriptors.

Including randomly sampled MLPs, which have been generated in the training process, into the MLP model pool, incurs no additional computation costs. Analyzing both the optimal MLPs along with these outlier MLPs provides a more comprehensive understanding of the MLP behavior.

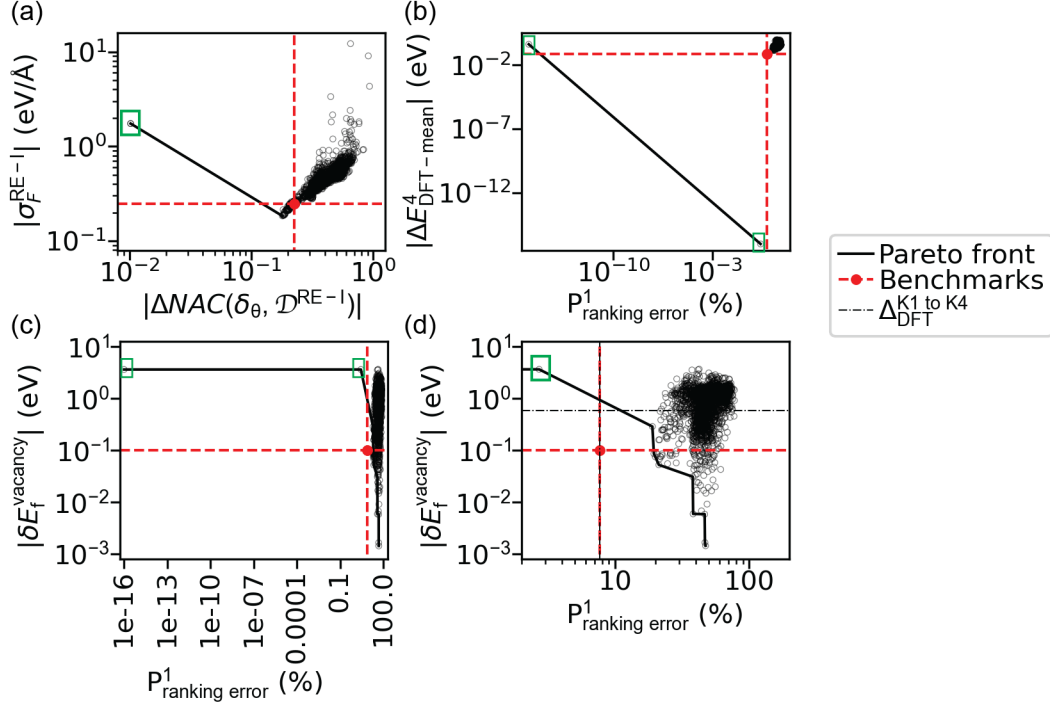


Figure G1. Examples of error metric distributions of property pairs containing outlier MLPs. The scattering plots of (a) the RMSEs, $\sigma_F^{\text{RE-I}}$, versus the directional errors $\Delta \text{NAC}(\delta_\theta, \mathcal{D}^{\text{RE-I}})$ of forces on RE atoms for Si interstitials, (b) the mean energy differences of ranking errors, mean ΔE_{DFT}^4 , versus the ranking error rate, $P_{\text{ranking error}}^1$, and (c) the formation energies of Si vacancy, $\delta E_f^{\text{vacancy}}$, versus $P_{\text{ranking error}}^1$. (d) The zoom-in plot of (c) with $P_{\text{ranking error}}^1$ range from 2% to 100%. Green boxes highlight the outlier MLPs in each plot.

Table G1. Summary of MLP pools categorized by their models and training datasets.

MLP models	Training data	Pool size of hyperparameters
GAP	$\mathcal{D}^{\text{Zuo et al.}}$	2304
	$\mathcal{D}^{\text{enhanced-I}}$	2304
	$\mathcal{D}^{\text{enhanced-V}}$	2304
NNP	$\mathcal{D}^{\text{Zuo et al.}}$	576
	$\mathcal{D}^{\text{enhanced-I}}$	576
	$\mathcal{D}^{\text{enhanced-V}}$	576
	$\mathcal{D}^{\text{Di-I}}$	720
	$\mathcal{D}^{4\text{-I}}$	720
	$\mathcal{D}^{\text{Zuo et al.}}$	300
DeePMD	$\mathcal{D}^{\text{enhanced-I}}$	300
	$\mathcal{D}^{\text{enhanced-V}}$	300
	$\mathcal{D}^{\text{Di-I}}$	294
	$\mathcal{D}^{4\text{-I}}$	294
DeepPot-SE	$\mathcal{D}^{\text{Zuo et al.}}$	200
	$\mathcal{D}^{\text{enhanced-I}}$	197
	$\mathcal{D}^{\text{enhanced-V}}$	194
MTP	$\mathcal{D}^{\text{enhanced-I}}$	1760
	$\mathcal{D}^{\text{enhanced-V}}$	1526
SNAP	$\mathcal{D}^{\text{Zuo et al.}}$	1965
	$\mathcal{D}^{\text{enhanced-I}}$	1980
	$\mathcal{D}^{\text{enhanced-V}}$	1980
	$\mathcal{D}^{\text{Di-I}}$	1977
	$\mathcal{D}^{4\text{-I}}$	1980

Table G2. Summary of 36 challenging properties and the fraction of MLPs that meet the benchmark of each challenging property.

Category	Error metric	MLPs meeting the benchmark (%)
E_f^{Defect}	$\delta E_f^{\text{split}-\langle 110 \rangle}$	5.0
	$\delta E_f^{\text{hexagonal}}$	4.4
	$\delta E_f^{\text{vacancy}}$	3.7
	$\delta E_f^{\text{tetrahedral}}$	2.8
Elastic constants	$\delta C_{12}^{\text{split}-\langle 110 \rangle}$	12.7
	$\delta C_{11}^{\text{split}-\langle 110 \rangle}$	7.0
	$\delta K^{\text{split}-\langle 110 \rangle}$	7.0
	$\delta C_{44}^{\text{split}-\langle 110 \rangle}$	3.2
	$\delta C_{12}^{\text{hexagonal}}$	12.0
	$\delta C_{44}^{\text{hexagonal}}$	9.1
	$\delta K^{\text{hexagonal}}$	8.2
	$\delta C_{11}^{\text{hexagonal}}$	1.0
	$\delta C_{12}^{\text{supercell}}$	8.5
	$\delta C_{44}^{\text{supercell}}$	8.3
	$\delta K^{\text{supercell}}$	7.3
	$\delta C_{11}^{\text{supercell}}$	2.0
	$\delta C_{12}^{\text{unit cell}}$	11.5
	$\delta K^{\text{vacancy}}$	9.3
Energy rankings	$P_{\text{ranking error}}^1$	0.1
	Maximum ΔE_{DFT}^1	0.1
	Mean ΔE_{DFT}^1	0.0
	$P_{\text{ranking error}}^2$	0.2
	Maximum ΔE_{DFT}^2	0.0
	Mean ΔE_{DFT}^2	0.0
	$P_{\text{ranking error}}^3$	0.0
	Maximum ΔE_{DFT}^3	0.0
	Mean ΔE_{DFT}^3	0.0
	$P_{\text{ranking error}}^4$	0.1
	Maximum ΔE_{DFT}^4	0.0
	Mean ΔE_{DFT}^4	0.0
Forces on RE atoms, $NAC(\delta, \mathcal{D}^{\text{RE}})$	$\Delta NAC(\delta_F , \mathcal{D}^{\text{RE}-V})$	12.7
	$\Delta NAC(\delta_\theta, \mathcal{D}^{\text{RE}-I})$	4.8
	$\Delta NAC(\delta_F , \mathcal{D}^{\text{RE}-I})$	3.8
σ^{RE}	$\sigma_F^{\text{RE}-V}$	11.7
	$\sigma_F^{\text{RE}-I}$	3.7
Thermal properties	$\delta S^{\text{bulk}}(T)$	14.5

Table G3. Summary of 20 challenging property pairs that no MLP can meet both benchmarks at the same time.

Category A	Error metric A	Error metric B	Category B
E_f^{Defect}	$\delta E_f^{\text{hexagonal}}$	$\delta C_{11}^{\text{hexagonal}}$	Elastic constants
		$\delta C_{11}^{\text{supercell}}$	
		$\delta C_{44}^{\text{supercell}}$	
		$\Delta NAC(\delta_{\theta}, \mathcal{D}^{\text{RE-I}})$	Forces on RE atoms, $NAC(\delta, \mathcal{D}^{\text{RE}})$
		$\Delta NAC(\delta_E , \mathcal{D}^{\text{RE-I}})$	
		$\sigma_F^{\text{RE-I}}$	σ^{RE}
		$\sigma_F^{\text{RE-V}}$	
		$\sigma_F^{\text{enhanced-V}}$	σ^{enhanced}
	$\delta E_f^{\text{tetrahedral}}$	$\delta C_{11}^{\text{split-}\langle 110 \rangle}$	Elastic constants
		$\delta C_{44}^{\text{split-}\langle 110 \rangle}$	
		$\delta C_{11}^{\text{hexagonal}}$	
		$\Delta NAC(\delta_{\theta}, \mathcal{D}^{\text{RE-I}})$	Forces on RE atoms, $NAC(\delta, \mathcal{D}^{\text{RE}})$
		$\Delta NAC(\delta_E , \mathcal{D}^{\text{RE-I}})$	
		$\sigma_F^{\text{RE-I}}$	σ_{EF}^{RE}
		$\sigma_F^{\text{enhanced-V}}$	
	$\delta E_f^{\text{split-}\langle 110 \rangle}$	$\Delta NAC(\delta_{\theta}, \mathcal{D}^{\text{RE-I}})$	Forces on RE atoms, $NAC(\delta, \mathcal{D}^{\text{RE}})$
		$\Delta NAC(\delta_E , \mathcal{D}^{\text{RE-I}})$	
		$\sigma_F^{\text{RE-I}}$	σ^{RE}
		$\sigma_F^{\text{enhanced-I}}$	σ^{enhanced}
	$\delta E_f^{\text{vacancy}}$	$\delta C_{11}^{\text{hexagonal}}$	Elastic constants

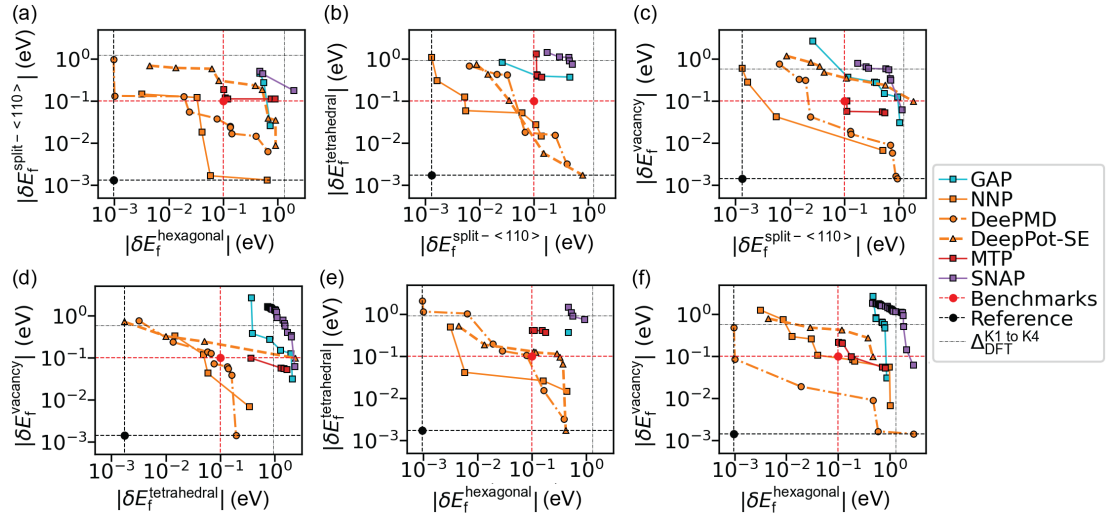


Figure G2. Pareto fronts for six property pairs in E_f^{Defect} . Scattering plots of optimal MLPs on Pareto fronts for each of the six MLP models, GAP (cyan lines with squares), NNP (orange lines with squares), DeePMD (orange dashed lines with circles), DeepPot-SE (orange dash-dotted lines with triangles), MTP (red lines with squares), and SNAP (purple lines with squares). Red dashed lines with red dots are the handpicked benchmarks. Black dashed lines with black dots are reference points to evaluate hypervolumes or inverted generational distances of Pareto fronts. Thin black dashed lines are benchmarks of differences between DFT K4 and K1 calculations, $\Delta_{\text{DFT}}^{\text{K1 to K4}}$.

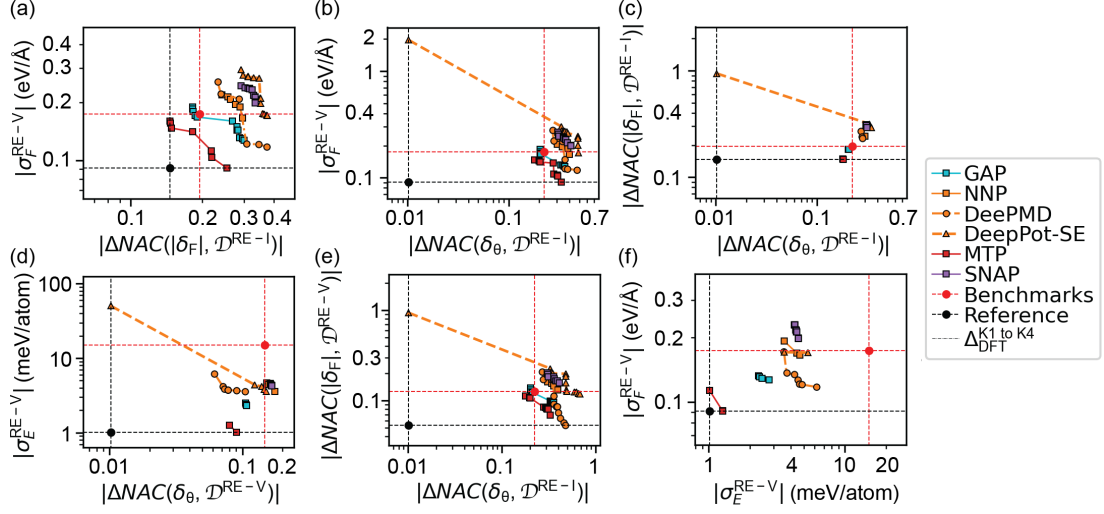


Figure G3. Pareto fronts for six selected property pairs in either the forces on RE atoms $NAC(\delta, D^{RE})$ category or σ^{RE} category, following the same scheme of Figure G2. The benchmark of σ_E^{RE-V} is handicapped as 15 meV/atom in (d) and (f), while all the other error metrics in the figure use $\Delta_{DFT}^{K1 \text{ to } K4}$ as benchmarks.

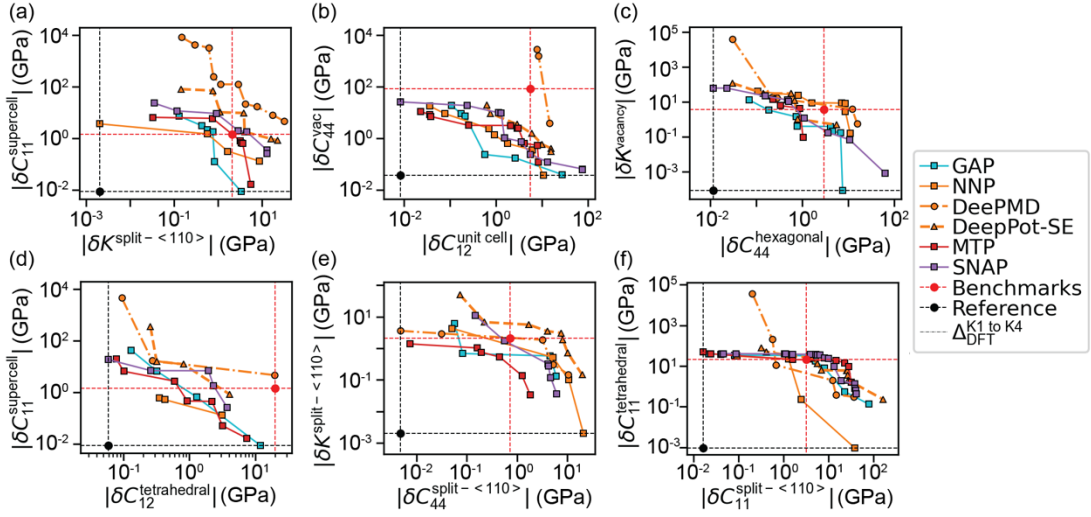


Figure G4. Pareto fronts for six selected property pairs in the elastic constant category. All error metrics use $\Delta_{DFT}^{K1 \text{ to } K2}$ as benchmarks.

Appendix H

H1. First-Principles Computation

Density functional theory (DFT) calculations were performed to compute energies and forces of configurations and to relax structures. Since a majority of MLPs were retrieved from Chapter 3, all DFT calculations were performed as described in Appendix D, using Vienna ab initio simulation package^[98] (VASP) with the projector augmented-wave approach. Generalized-gradient approximation (GGA) with Perdew-Burke-Ernzerhof^[100] (PBE) functionals were used. All true values of energies, forces, physical properties, and evaluation metrics were calculated with a $4\times 4\times 4$ k -point mesh (K4), except for elastic constants, which were calculated with $2\times 2\times 2$ k -point mesh (K2). DFT calculations using a Γ -centered single k -point $1\times 1\times 1$ (K1) were also obtained as benchmarks of error metrics of properties. All DFT calculations were spin-polarized, with electronic relaxation cutoff of 10^{-5} eV, energy cutoff of 520 eV, and other parameters set as in the Materials Project.^[102,105]

H2. Ab initio molecular dynamics simulation

Ab initio molecular dynamics (AIMD) simulations were performed to generate atomistic configurations in training, validation, and testing datasets (Appendix H3). The supercell models of bulk or defected Si for AIMD simulations have lattice parameters larger than 10 Å. AIMD simulations were non-spin-polarized, with electronic energy convergence cutoff to 10^{-4} eV and a time step of 2 fs. A Γ -centered single k -point of $1\times 1\times 1$ (K1) was used for AIMD simulations. All AIMD simulations to obtain migrating (rare event, RE) atoms were performed at 1000 K or 1230 K, following the same scheme described in Chapter 3 and Appendix D. For each AIMD simulation, an initial period of heating up with static relaxed supercells from 100 K to the final temperatures

using velocity scaling at a constant rate over 2 ps. Then, AIMD simulations were conducted using NVT ensemble with Nosé-Hoover thermostat.

H3. Constructing training, validation, and testing datasets

This study used a total of five training datasets, $\mathcal{D}^{\text{Zuo et al.}}$, $\mathcal{D}^{\text{enhanced-I}}$, $\mathcal{D}^{\text{enhanced-V}}$, $\mathcal{D}^{2-I}$, and $\mathcal{D}^{4-I}$, a validation dataset, $\mathcal{D}^{\text{validation}}$, and two testing datasets, $\mathcal{D}^{\text{RE-I}}$ and $\mathcal{D}^{\text{RE-V}}$, each consisting of a diverse range of bulk, defected, and distorted Si configurations. All datasets, except for $\mathcal{D}^{2-I}$ and $\mathcal{D}^{4-I}$, were identical to those in Chapter 3, and $\mathcal{D}^{2-I}$ and $\mathcal{D}^{4-I}$ were generated as the same methods as $\mathcal{D}^{\text{enhanced-I}}$ but with more interstitials as described below. All true values of energies and forces used calculation results of DFT K4, as indicated in the *First-Principles Computation* section. The five training datasets are generated by the same approach following Chapter 3.4.2. First, we retrieved the training dataset $\mathcal{D}^{\text{Zuo et al.}}$ from Zuo et al.,^[1] which consists of 219 configurations for a wide range of Si structures, including solid Si, melted Si, distorted Si, thin slabs of surface Si, and Si bulks with single vacancy from AIMD simulations at different temperatures as described in Ref. ^[1]. Then, we randomly selected and removed 120 Si configurations from liquid Si, AIMD simulations of Si bulk, and the strained Si bulk in $\mathcal{D}^{\text{Zuo et al.}}$. The removed 120 configurations were replaced by 120 snapshots generated by AIMD simulations (K1), with different defects: $\mathcal{D}^{\text{enhanced-V}}$ uses Si bulk with single vacancy; $\mathcal{D}^{\text{enhanced-I}}$ uses Si bulk with single interstitial; $\mathcal{D}^{2-I}$ uses Si bulk with two interstitials; and $\mathcal{D}^{4-I}$ uses Si bulk with four interstitials. Thus, all five training datasets have totally 219 configurations. All defected structures are fully relaxed to perform AIMD simulations, with lattice parameters of supercells changing -0.4%, 0.3%, 0.3%, 0.3%, 0.5%, and 1.6% (with respect to single vacancy, split- $\langle 110 \rangle$ interstitial, tetrahedral interstitial, hexagonal interstitial, 2-interstitial, and 4-interstitial supercells)

compared with the supercell of perfect Si bulk. All energies and atomic forces of the AIMD snapshots were converged using single-step self-consistent DFT K4 without relaxing atom positions and lattices. Specifically, all snapshots contain at least one identified rare-event (RE) atom and the identification of RE atoms follows the method in Chapter 3 and Appendix D. If the difference between the distances to the 1st and the 2nd nearest neighbor of a Si atom is less than 0.75 Å (approximately 31% of the distance between two nearest static sites), that Si atom will be identified as a migrating RE atom.

The validation dataset $\mathcal{D}^{\text{validation}}$ is the enhanced validation set used in Chapter 3.4.2, consisting of 50 Si structures. 20 configurations are randomly selected from the 120 removed structures in the original $\mathcal{D}^{\text{Zuo et al.}}$, 11 configurations are AIMD snapshots with single vacancy RE, and 19 configurations are AIMD snapshots with single interstitial RE.

The two testing datasets, $\mathcal{D}^{\text{RE-I}}$ and $\mathcal{D}^{\text{RE-V}}$, are the interstitial RE testing set and the vacancy RE testing set used in Chapter 3 and Appendix D. Each testing dataset consists of 100 AIMD snapshots with single interstitial (for $\mathcal{D}^{\text{RE-I}}$) or single vacancy (for $\mathcal{D}^{\text{RE-V}}$) respectively and each snapshot contains an identified RE atom as previously described in Appendix D.

H4. Training MLPs

We adopted the same approach of training MLPs as in Chapter 3.2 and 3.4.2, and in Zuo et al.^[1] The *mlearn* Python package and the corresponding MLP source codes, such as QUIP for GAP,^[11] N2P2 for NNP,^[109] MLIP for MTP,^[15,110] and SNAP embedded in LAMMPS were used to train these MLPs respectively. We also used the *DeePMD-kit* package to train MLPs for DeePMD and DeepPot-SE models, following Ref. ^[111] and Ref. ^[5]. DeePMD refers to the early version of Deep Potential Molecular Dynamics and the later version of Deep Potential-Smooth

Edition is referred as DeepPot-SE. A pool of MLPs with different hyperparameter values were trained for each MLP model and by each training dataset, following the same approach in Chapter 3.2. A grid search approach was adopted using two to ten values for each hyperparameter, including the number of radial basis functions and the band limit of spherical harmonic basis functions for GAP, the cutoff radius, and the size of neural network for NNP, the choice of radial basis function sets for MTP, and the cutoff radius and the number of iteration steps for DeepPMD and DeepPot-SE. Each model pool has 194 to 2304 MLPs with various hyperparameters for different models and training datasets. There were 25327 MLPs generated for our pool.

H5. Optimizing MLPs and selecting the hyperparameter sets

The validation scores of trained MLP models were calculated as follows. We first calculated the energy and the force RMSEs of the corresponding training data, σ_E^{train} and σ_F^{train} , and those RMSEs of the validation data $\mathcal{D}^{\text{validation}}$, $\sigma_E^{\text{validation}}$ and $\sigma_F^{\text{validation}}$, totally four criteria. These four RMSEs criteria were normalized as,

$$\sigma_k^{\mathcal{D}'} = \frac{\sigma_k^{\mathcal{D}} - (\sigma_k^{\mathcal{D}})_{\min}}{(\sigma_k^{\mathcal{D}})_{\text{median}} - (\sigma_k^{\mathcal{D}})_{\min}}, k \in \{E, F\} \text{ and } \mathcal{D} \in \{\text{train, validation}\} \quad \text{Eq. (H1)}$$

where $\sigma_k^{\mathcal{D}}$ is one of the four RMSEs on dataset $\mathcal{D}$, and $(\sigma_k^{\mathcal{D}})_{\min}$ and $(\sigma_k^{\mathcal{D}})_{\text{median}}$ is the minimum and the median RMSE, in order to have normalized distributions and to have lower values correspond to better performances. Then, the validation score was calculated as:

$$\text{Score} = \sqrt{\sum_{\mathcal{D}} [\sigma_k^{\mathcal{D}'}]^2} \quad \text{Eq. (H2)}$$

Lower validation scores correspond to better MLP performances.

H6. Construction of MLP model pools

The MLP model pool consisted of optimized MLPs and MLPs randomly selected. If the pool size is larger than 1000, we picked the top 10% MLPs from each pool with the lowest validation scores, and among them chose 50 MLPs with even spacing according to the ranking by the validation scores. If the pool size is smaller than 1000, we picked the top 100 MLPs with the lowest validation scores and chose every other MLP (a total of 50) according to the ranking by the validation scores. After picking the optimized MLPs, we randomly selected 50 MLPs from the remaining pool.

H7. Calculating different physical properties

The methods for the calculations of the physical properties by DFT (either K4 or K2) as true values were described as follows. The supercells of perfect Si bulk, Si with single vacancy, and Si interstitials, such as split- $\langle 110 \rangle$, tetrahedral, and hexagonal interstitials, were relaxed by DFT K4. MLP predictions on all properties used either fixed configurations relaxed by DFT K4 or fixed snapshots from AIMD simulations without further relaxations, except for lattice parameters. The true values of all properties, such as energies, forces, energy RMSEs, force RMSEs, lattice parameters, energy rankings, the normalized area of curve (NAC), and the force constants, were calculated by DFT K4 except for elastic constants of bulk and defected Si bulks.

Defect formation energy E_f^{Defect} : The supercells with defects were constructed using perfect Si bulk with $2 \times 2 \times 2$ conventional unit cells (64 atoms) and then statically relaxed by DFT K4. The defects considered were vacancy, and interstitial with split- $\langle 110 \rangle$, tetrahedral, or

hexagonal interstitial configurations. The energies of DFT K4 calculations were obtained to get the defect formation energy E_f^{Defect} :

$$E_f^{\text{Defect}} = E^{\text{defect}} - \frac{N^{\text{defect}}}{N^{\text{bulk}}} E^{\text{bulk}}, \quad \text{Eq. (H3)}$$

where E^{defect} is the energy of defected configuration, E^{bulk} is the energy of perfect bulk supercell, N^{bulk} is the number of atoms in perfect bulk supercell, and N^{defect} is the number of atoms of defected supercell.

Elastic constants: The true values of elastic constants were calculated by finite differences approach via a stress-strain relationship implemented in VASP (IBRION = 6 and ISIF = 3). The width of the displacement of each ion was set to 0.015 Å (POTIM = 0.015) and the number of the ionic displacements was set to 2 (NFREE = 2). Since the defect supercells require higher computation cost due to lower symmetry, the true values were calculated by DFT K2. For MLPs, such as GAP, NNP, MTP, and SNAP, the functions to calculate elastic constants were implemented in the *maml* Python package.^[118] For DeePMD and DeepPot-SE, we utilized the LAMMPS scripts from the *maml* package to compute the elastic constants.

Lattice parameters: While other property evaluations used the identical supercells relaxed by DFT K4, here the supercells were relaxed by MLPs (See H4 for the methods) to determine the lattice parameters given by the MLPs.

Energy rankings: The MLP performances on energy rankings of the different ordering of multiple defects in Si using the following four datasets: 1) di-vacancies in Si bulks in the supercell with 3×3×3 conventional unit cells (214 atoms); 2) tri-vacancies in Si bulks in the supercell with 3×3×2 conventional unit cells (141 atoms); 3) tri-vacancies in Si bulks in the supercell with 3×2×2 conventional unit cells (93 atoms); 4) tri-vacancies in Si bulks in the supercell with 2×2×2 conventional unit cells (63 atoms). For each set, we randomly removed two to three Si atoms and

generated up to 30 symmetrically distinct configurations (using Pymatgen package ^[105]). All configurations were not relaxed and had fixed lattices and atom positions.

The energy rankings for each dataset were determined by the true energies converged by single-step self-consistent calculations of DFT K4. There were three performance metrics for energy ranking, the error rate of ranking errors, $P_{\text{ranking error}}^{\mathcal{D}}$ for dataset $\mathcal{D}$, the mean, and the maximum values of the energy differences ΔE^{DFT} of ranking errors, the mean $\Delta E_{\text{DFT}}^{\mathcal{D}}$ and the maximum $\Delta E_{\text{DFT}}^{\mathcal{D}}$. Following Chapter 4.3, we used the error rates of ranking errors in all ranking pairs to get $P_{\text{ranking error}}^{\mathcal{D}}$. We calculated ΔE^{DFT} of all ranking errors in dataset $\mathcal{D}$ and took the mean and the maximum $\Delta E_{\text{DFT}}^{\mathcal{D}}$.

Force performances on RE atoms: NAC as performance metric, as proposed in Chapter 3.4.1, was used to evaluate the magnitude or the directional errors of forces on rare-event (RE) atoms. We took the identified RE atoms from $\mathcal{D}^{\text{RE-I}}$ and $\mathcal{D}^{\text{RE-V}}$, and compared the true and the predicted forces on these RE atoms to get force errors. We constructed cumulative distribution functions (CDFs) curves of force errors to evaluate *NACs* of CDF curves of magnitude errors, δ_F , and directional errors, δ_θ . *NACs* of the CDF curves were then calculated within the error ranges of $0 < |\delta_F| < 1 \text{ eV/\AA}$ for magnitude errors $\text{NAC}(|\delta_F|, \mathcal{D})$, and $0^\circ < \delta_\theta < 60^\circ$ for directional errors $\text{NAC}(\delta_\theta, \mathcal{D})$.

Energy and force RMSEs on RE atoms: The energy RMSEs of configurations from $\mathcal{D}^{\text{RE-I}}$ and $\mathcal{D}^{\text{RE-V}}$ $\sigma_E^{\mathcal{D}}$, were calculated as evaluation metrics. For forces, we only took the forces on identified RE atoms from $\mathcal{D}^{\text{RE-I}}$ and $\mathcal{D}^{\text{RE-V}}$ and computed the RMSEs, $\sigma_F^{\mathcal{D}}$.

Energy and force RMSEs on enhanced training data: We calculated the energy and force RMSEs of all configurations and all atomistic forces in datasets $\mathcal{D}^{\text{enhanced-I}}$ and $\mathcal{D}^{\text{enhanced-V}}$.

Thermal properties: Thermal properties, such as the equation of state (EOS) curves of entropy versus temperature $S(T)$, free energy versus temperature $E_{\text{free}}(T)$, and heat capacity versus temperature $c(T)$, were calculated using *Phonopy* Python package.^[119–121] We calculated the thermal properties for both Si bulk and Si bulk with a single vacancy in the supercell with $2 \times 2 \times 2$ conventional unit cells with a single vacancy (63 atoms). *Phonopy* generated Si configurations with displaced atoms according to the symmetry of the given cell. Then, we calculated the atomistic forces of these configurations with small displacements, by either DFT or MLPs, to generate force constants. Given the calculated forces and force constants, we generated the EOS curves using *Phonopy* with a $40 \times 40 \times 40$ q -point mesh. The temperatures for EOS curves covered a range from 0 to 1000 K with a consistent spacing of 10 K.

H8. Calculating errors of physical properties and evaluation metrics

The errors were evaluated between the MLP predicted values and the true values (by DFT K4) of physical properties. The predicted physical properties of $E_{\text{f}}^{\text{Defect}}$, elastic constants, and lattice parameters were directly compared with their true values calculated by DFT K4, and then the logarithm of these errors were taken to construct and analyze their Pareto fronts.

For EOS curves of thermal properties, we used the following equation to calculate the error metrics of curves predicted by MLPs and the true EOS curves, Δ_{EOS} :

$$\Delta_{\text{EOS}} = \sqrt{\frac{\sum_{i=1}^N [X^a(T_i) - X^{\text{K4}}(T_i)]^2}{N}} \quad \text{Eq. (H4)}$$

where $X^a(T_i)$ denotes property X (either entropy S , free energy E_{free} , or heat capacity c) computed using method a (either MLP or DFT) at temperature T_i , X^{K4} is the property X calculated by DFT K4, and N is the total number of temperatures on the curve. We then took the logarithm of Δ_{EOS} .

For the properties, such as energy rankings, force performances on RE atoms, energy and force RMSEs on RE atoms, and energy and force RMSEs on enhanced training datasets, we used the differences between these properties and their ‘ideal’ values (1 for force performances on RE atoms and 0 for all the others) and took the logarithm of the differences as evaluation metrics.

The benchmarks of error metrics on all properties were calculated by DFT K1, except for $E_{\text{f}}^{\text{Defect}}$, energy RMSEs on RE atoms, and energy and force RMSEs on enhanced training datasets. The differences between DFT K4 and K1 on $E_{\text{f}}^{\text{Defect}}$ were large and the benchmarks were all handpicked to be 0.1 eV. The benchmarks of energy RMSEs for configurations on RE atoms ($\mathcal{D}^{\text{RE-I}}$ and $\mathcal{D}^{\text{RE-V}}$) were handpicked to the values of 15 meV/atom. Benchmarks of 15 meV/atom for energy RMSEs and 0.12 eV/Å for force RMSEs were used on enhanced training datasets ($\mathcal{D}^{\text{enhanced-I}}$ and $\mathcal{D}^{\text{enhanced-V}}$).

Bibliography

- [1] Y. Zuo, C. Chen, X. Li, Z. Deng, Y. Chen, J. Behler, G. Csányi, A. V. Shapeev, A. P. Thompson, M. A. Wood, S. P. Ong, *J. Phys. Chem. A* **2020**, *124*, 731.
- [2] R. Batra, S. Sankaranarayanan, *J. Phys. Mater.* **2020**, *3*, 031002.
- [3] H. Chan, B. Narayanan, M. J. Cherukara, F. G. Sen, K. Sasikumar, S. K. Gray, M. K. Y. Chan, S. K. R. S. Sankaranarayanan, *J. Phys. Chem. C* **2019**, *123*, 6941.
- [4] X. He, Y. Zhu, A. Epstein, Y. Mo, *npj Comput. Mater.* **2018**, *4*, 18.
- [5] L. Zhang, J. Han, H. Wang, R. Car, W. E, *Phys. Rev. Lett.* **2018**, *120*, 143001.
- [6] J. Schmidt, M. R. G. Marques, S. Botti, M. A. L. Marques, *npj Comput. Mater.* **2019**, *5*, 83.
- [7] P. Friederich, F. Häse, J. Proppe, A. Aspuru-Guzik, *Nat. Mater.* **2021**, *20*, 750.
- [8] J. Behler, *Chem. Rev.* **2021**, *121*, 10037.
- [9] O. T. Unke, S. Chmiela, H. E. Sauceda, M. Gastegger, I. Poltavsky, K. T. Schütt, A. Tkatchenko, K.-R. Müller, *Chem. Rev.* **2021**, *121*, 10142.
- [10] V. L. Deringer, M. A. Caro, G. Csányi, *Adv. Mater.* **2019**, *31*, 1.
- [11] A. P. Bartók, M. C. Payne, R. Kondor, G. Csányi, *Phys. Rev. Lett.* **2010**, *104*, 136403.
- [12] J. Behler, M. Parrinello, *Phys. Rev. Lett.* **2007**, *98*, 146401.
- [13] T. W. Ko, J. A. Finkler, S. Goedecker, J. Behler, *Nat. Commun.* **2021**, *12*, 398.
- [14] A. P. Thompson, L. P. Swiler, C. R. Trott, S. M. Foiles, G. J. Tucker, *J. Comput. Phys.* **2015**, *285*, 316.
- [15] A. V Shapeev, *Multiscale Model. Simul.* **2016**, *14*, 1153.
- [16] T. Zubatiuk, O. Isayev, *Acc. Chem. Res.* **2021**, *54*, 1575.
- [17] S. Wang, Q. Bai, A. M. Nolan, Y. Liu, S. Gong, Q. Sun, Y. Mo, *Angew. Chemie - Int. Ed.* **2019**, *58*, 8039.

- [18] X. He, Q. Bai, Y. Liu, A. M. Nolan, C. Ling, Y. Mo, *Adv. Energy Mater.* **2019**, *9*, 1902078.
- [19] M. Yang, Y. Liu, A. M. Nolan, Y. Mo, *Adv. Mater.* **2021**, *33*, 2008081.
- [20] J. Qi, S. Banerjee, Y. Zuo, C. Chen, Z. Zhu, M. L. Holekevi Chandrappa, X. Li, S. P. Ong, *Mater. Today Phys.* **2021**, *21*, 100463.
- [21] Z. Zhang, Y. Shao, B. Lotsch, Y. S. Hu, H. Li, J. Janek, L. F. Nazar, C. W. Nan, J. Maier, M. Armand, L. Chen, *Energy Environ. Sci.* **2018**, *11*, 1945.
- [22] J. Janek, W. G. Zeier, *Nat. Energy* **2016**, *1*, 16141.
- [23] J. Lau, R. H. DeBlock, D. M. Butts, D. S. Ashby, C. S. Choi, B. S. Dunn, *Adv. Energy Mater.* **2018**, *8*, 1.
- [24] N. Kamaya, K. Homma, Y. Yamakawa, M. Hirayama, R. Kanno, M. Yonemura, T. Kamiyama, Y. Kato, S. Hama, K. Kawamoto, A. Mitsui, *Nat. Mater.* **2011**, *10*, 682.
- [25] K. H. Park, Q. Bai, D. H. Kim, D. Y. Oh, Y. Zhu, Y. Mo, Y. S. Jung, *Adv. Energy Mater.* **2018**, *8*, 1.
- [26] Y. Kato, S. Hori, T. Saito, K. Suzuki, M. Hirayama, A. Mitsui, M. Yonemura, H. Iba, R. Kanno, *Nat. Energy* **2016**, *1*, 16030.
- [27] J. C. Bachman, S. Muy, A. Grimaud, H. H. Chang, N. Pour, S. F. Lux, O. Paschos, F. Maglia, S. Lupart, P. Lamp, L. Giordano, Y. Shao-Horn, *Chem. Rev.* **2016**, *116*, 140.
- [28] T. Famprikis, P. Canepa, J. A. Dawson, M. S. Islam, C. Masquelier, *Nat. Mater.* **2019**, *18*, 1278.
- [29] X. He, Q. Bai, Y. Liu, A. M. Nolan, C. Ling, Y. Mo, *Adv. Energy Mater.* **2019**, *9*, 1.
- [30] S. Muy, J. Voss, R. Schlem, R. Koerver, S. J. Sedlmaier, F. Maglia, P. Lamp, W. G. Zeier, Y. Shao-Horn, *iScience* **2019**, *16*, 270.

- [31] L. Kahle, A. Marcolongo, N. Marzari, *Energy Environ. Sci.* **2020**, *13*, 928.
- [32] X. He, Y. Zhu, Y. Mo, *Nat. Commun.* **2017**, *8*, 15893.
- [33] A. M. Nolan, Y. Zhu, X. He, Q. Bai, Y. Mo, *Joule* **2018**, *2*, 2016.
- [34] Y. Zhu, Y. Mo, *Angew. Chemie Int. Ed.* **2020**, anie. 202007621.
- [35] X. He, Y. Mo, *Phys. Chem. Chem. Phys.* **2015**, *17*, 18035.
- [36] H. Zong, G. Pilania, X. Ding, G. J. Ackland, T. Lookman, *npj Comput. Mater.* **2018**, *4*, 48.
- [37] J. Vandermause, S. B. Torrisi, S. Batzner, Y. Xie, L. Sun, A. M. Kolpak, B. Kozinsky, *npj Comput. Mater.* **2020**, *6*, 20.
- [38] B. Kailkhura, B. Gallagher, S. Kim, A. Hiszpanski, T. Y.-J. Han, *npj Comput. Mater.* **2019**, *5*, 108.
- [39] M. Wen, E. B. Tadmor, *npj Comput. Mater.* **2020**, *6*, 124.
- [40] V. L. Deringer, N. Bernstein, G. Csányi, C. Ben Mahmoud, M. Ceriotti, M. Wilson, D. A. Drabold, S. R. Elliott, *Nature* **2021**, *589*, 59.
- [41] D. Montes de Oca Zapiain, M. A. Wood, N. Lubbers, C. Z. Pereyra, A. P. Thompson, D. Perez, *npj Comput. Mater.* **2022**, *8*, 189.
- [42] P. Rowe, V. L. Deringer, P. Gasparotto, G. Csányi, A. Michaelides, *J. Chem. Phys.* **2020**, *153*, 034702.
- [43] L. Zhang, D.-Y. Lin, H. Wang, R. Car, W. E, *Phys. Rev. Mater.* **2019**, *3*, 023804.
- [44] X. Fu, Z. Wu, W. Wang, T. Xie, S. Keten, R. Gomez-Bombarelli, T. Jaakkola, **2022**, *1*.
- [45] V. Thangadurai, H. Kaack, W. J. F. Weppner, *J. Am. Ceram. Soc.* **2003**, *86*, 437.
- [46] R. Murugan, V. Thangadurai, W. Weppner, *Angew. Chemie - Int. Ed.* **2007**, *46*, 7778.
- [47] H. Aono, E. Sugimoto, Y. Sadaoka, N. Imanaka, G. ya Adachi, *Solid State Ionics* **1990**,

40–41, 38.

- [48] Y. Zhu, X. He, Y. Mo, *ACS Appl. Mater. Interfaces* **2015**, 7, 23685.
- [49] F. Han, T. Gao, Y. Zhu, K. J. Gaskell, C. Wang, *Adv. Mater.* **2015**, 27, 3473.
- [50] Y. Zhu, X. He, Y. Mo, *J. Mater. Chem. A* **2016**, 4, 3253.
- [51] A. Sakuda, A. Hayashi, M. Tatsumisago, *Chem. Mater.* **2010**, 22, 949.
- [52] J. H. Woo, J. E. Trevey, A. S. Cavanagh, Y. S. Choi, S. C. Kim, S. M. George, K. H. Oh, S.-H. Lee, *J. Electrochem. Soc.* **2012**, 159, A1120.
- [53] T. Asano, A. Sakai, S. Ouchi, M. Sakaida, A. Miyazaki, S. Hasegawa, *Adv. Mater.* **2018**, 30, 1.
- [54] X. Li, J. Liang, J. Luo, M. Norouzi Banis, C. Wang, W. Li, S. Deng, C. Yu, F. Zhao, Y. Hu, T. K. Sham, L. Zhang, S. Zhao, S. Lu, H. Huang, R. Li, K. R. Adair, X. Sun, *Energy Environ. Sci.* **2019**, 12, 2665.
- [55] R. Schlem, S. Muy, N. Prinz, A. Banik, Y. Shao-Horn, M. Zobel, W. G. Zeier, *Adv. Energy Mater.* **2020**, 10, 1.
- [56] J. Liang, X. Li, S. Wang, K. R. Adair, W. Li, Y. Zhao, C. Wang, Y. Hu, L. Zhang, S. Zhao, S. Lu, H. Huang, R. Li, Y. Mo, X. Sun, *J. Am. Chem. Soc.* **2020**, 142, 7012.
- [57] K. H. Park, K. Kaup, A. Assoud, Q. Zhang, X. Wu, L. F. Nazar, *ACS Energy Lett.* **2020**, 533.
- [58] L. Zhou, C. Y. Kwok, A. Shyamsunder, Q. Zhang, X. Wu, L. F. Nazar, *Energy Environ. Sci.* **2020**.
- [59] Y. Wang, W. D. Richards, S. P. Ong, L. J. Miara, J. C. Kim, Y. Mo, G. Ceder, *Nat. Mater.* **2015**, 14, 1026.
- [60] H. D. Lutz, W. Schmidt, H. Haeuseler, *J. Phys. Chem. Solids* **1981**, 42, 287.

- [61] H. LUTZ, Z. ZHANG, A. PFITZNER, *Solid State Ionics* **1993**, 62, 1.
- [62] R. Kanno, Y. Takeda, O. Yamamoto, *Mater. Res. Bull.* **1981**, 16, 999.
- [63] C. Catlow, M. . Wolf, *Proc. R. Soc. London. A. Math. Phys. Sci.* **1987**, 413, 201.
- [64] H.-J. Steiner, H. D. Lutz, *Zeitschrift für Anorg. und Allg. Chemie* **1992**, 613, 26.
- [65] L. Kahle, A. Marcolongo, N. Marzari, *Energy Environ. Sci.* **2020**, 3, 1.
- [66] G. Hautier, C. Fischer, V. Ehrlacher, A. Jain, G. Ceder, *Inorg. Chem.* **2011**, 50, 656.
- [67] Q. Bai, X. He, Y. Zhu, Y. Mo, *ACS Appl. Energy Mater.* **2018**, 1, 1626.
- [68] Z. Zhu, I. H. Chu, S. P. Ong, *Chem. Mater.* **2017**, 29, 2474.
- [69] A. Urban, J. Lee, G. Ceder, *Adv. Energy Mater.* **2014**, 4, 1.
- [70] J. Lee, A. Urban, X. Li, D. Su, G. Hautier, G. Ceder, *Science (80-.).* **2014**, 343, 519.
- [71] S. Hull, *Reports Prog. Phys.* **2004**, 67, 1233.
- [72] A. Manthiram, X. Yu, S. Wang, *Nat. Rev. Mater.* **2017**, 2, 16103.
- [73] N. Tanibata, S. Takimoto, K. Nakano, H. Takeda, M. Nakayama, H. Sumi, *ACS Mater. Lett.* **2020**, acsmaterialslett.0c00127.
- [74] A. P. Bartók, R. Kondor, G. Csányi, *Phys. Rev. B* **2013**, 87, 184115.
- [75] J. Behler, *J. Chem. Phys.* **2011**, 134, 074106.
- [76] V. Botu, R. Ramprasad, *Phys. Rev. B* **2015**, 92, 094306.
- [77] M. A. Cusentino, M. A. Wood, A. P. Thompson, *J. Phys. Chem. A* **2020**, 124, 5456.
- [78] A. P. Bartók, J. Kermode, N. Bernstein, G. Csányi, *Phys. Rev. X* **2018**, 8, 041048.
- [79] L. Zhang, J. Han, H. Wang, W. A. Saidi, R. Car, W. E, *Adv. Neural Inf. Process. Syst.* **2018**, 32, 4441.
- [80] H. Babaei, R. Guo, A. Hashemi, S. Lee, *Phys. Rev. Mater.* **2019**, 3, 074603.
- [81] J. Qi, S. Banerjee, Y. Zuo, C. Chen, Z. Zhu, M. L. Holekevi Chandrappa, X. Li, S. P. Ong,

- Mater. Today Phys.* **2021**, *21*, 100463.
- [82] E. A. Wu, S. Banerjee, H. Tang, P. M. Richardson, J.-M. Doux, J. Qi, Z. Zhu, A. Grenier, Y. Li, E. Zhao, G. Deysher, E. Sebt, H. Nguyen, R. Stephens, G. Verbist, K. W. Chapman, R. J. Clément, A. Banerjee, Y. S. Meng, S. P. Ong, *Nat. Commun.* **2021**, *12*, 1256.
- [83] W.-K. Leung, R. J. Needs, G. Rajagopal, S. Itoh, S. Ihara, *Phys. Rev. Lett.* **1999**, *83*, 2351.
- [84] D. Maroudas, R. A. Brown, *Appl. Phys. Lett.* **1993**, *62*, 172.
- [85] M. Tang, L. Colombo, J. Zhu, T. Diaz de la Rubia, *Phys. Rev. B* **1997**, *55*, 14279.
- [86] M. Yang, L. Bonati, D. Polino, M. Parrinello, *Catal. Today* **2021**, *18*.
- [87] S. Pant, Z. Smith, Y. Wang, E. Tajkhorshid, P. Tiwary, *J. Chem. Phys.* **2020**, *153*, 234118.
- [88] G. L. W. Hart, T. Mueller, C. Toher, S. Curtarolo, *Nat. Rev. Mater.* **2021**, *6*, 730.
- [89] S. Kharabadze, A. Thorn, E. A. Koulakova, A. N. Kolmogorov, *npj Comput. Mater.* **2022**, *8*, 136.
- [90] L. Tang, Z. J. Yang, T. Q. Wen, K. M. Ho, M. J. Kramer, C. Z. Wang, *Phys. Chem. Chem. Phys.* **2020**, *22*, 18467.
- [91] L. Tang, Z. J. Yang, T. Q. Wen, K. M. Ho, M. J. Kramer, C. Z. Wang, *Acta Mater.* **2021**, *204*, 116513.
- [92] V. L. Deringer, A. P. Bartók, N. Bernstein, D. M. Wilkins, M. Ceriotti, G. Csányi, *Chem. Rev.* **2021**, *121*, 10073.
- [93] N. Bernstein, G. Csányi, V. L. Deringer, *npj Comput. Mater.* **2019**, *5*, 1.
- [94] C. Sutton, S. V. Levchenko, *Front. Chem.* **2020**, *8*.
- [95] N. Artrith, A. M. Kolpak, *Nano Lett.* **2014**, *14*, 2670.
- [96] X. Chen, H. Zhou, Y. Li, *Mater. Des.* **2019**, *183*, 108085.

- [97] M. Shen, Y. Li, C. Hu, S. Xue, C. Xiang, J. Luo, Z. Yu, *Scr. Mater.* **2021**, 204, 114149.
- [98] G. Kresse, J. Furthmüller, *Phys. Rev. B - Condens. Matter Mater. Phys.* **1996**, 54, 11169.
- [99] P. E. Blöchl, *Phys. Rev. B* **1994**, 50, 17953.
- [100] J. P. Perdew, M. Ernzerhof, K. Burke, *J. Chem. Phys.* **1996**, 105, 9982.
- [101] J. Heyd, G. E. Scuseria, *J. Chem. Phys.* **2004**, 121, 1187.
- [102] A. Jain, G. Hautier, C. J. Moore, S. Ping Ong, C. C. Fischer, T. Mueller, K. A. Persson, G. Ceder, *Comput. Mater. Sci.* **2011**, 50, 2295.
- [103] A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, K. A. Persson, *APL Mater.* **2013**, 1.
- [104] S. Nosé, *Prog. Theor. Phys. Suppl.* **1991**, 103, 1.
- [105] S. P. Ong, W. D. Richards, A. Jain, G. Hautier, M. Kocher, S. Cholia, D. Gunter, V. L. Chevrier, K. A. Persson, G. Ceder, *Comput. Mater. Sci.* **2013**, 68, 314.
- [106] L. Pauling, *J. Am. Chem. Soc.* **1929**, 51, 1010.
- [107] QUIP and quippy documentation.
- [108] G. Varoquaux, L. Buitinck, G. Louppe, O. Grisel, F. Pedregosa, A. Mueller, *GetMobile Mob. Comput. Commun.* **2015**, 19, 29.
- [109] S. Desai, S. T. Reeve, J. F. Belak, **2020**.
- [110] E. V. Podryabinkin, A. V. Shapeev, *Comput. Mater. Sci.* **2017**, 140, 171.
- [111] L. Ward, A. Agrawal, A. Choudhary, C. Wolverton, *npj Comput. Mater.* **2016**, 2, 16028.
- [112] H. Wang, L. Zhang, J. Han, W. E, *Comput. Phys. Commun.* **2018**, 228, 178.
- [113] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, Í.

- Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, A. Vijaykumar, A. Pietro Bardelli, A. Rothberg, A. Hilboll, A. Kloeckner, A. Scopatz, A. Lee, A. Rokem, C. N. Woods, C. Fulton, C. Masson, C. Häggström, C. Fitzgerald, D. A. Nicholson, D. R. Hagen, D. V. Pasechnik, E. Olivetti, E. Martin, E. Wieser, F. Silva, F. Lenders, F. Wilhelm, G. Young, G. A. Price, G.-L. Ingold, G. E. Allen, G. R. Lee, H. Audren, I. Probst, J. P. Dietrich, J. Silterra, J. T. Webber, J. Slavič, J. Nothman, J. Buchner, J. Kulick, J. L. Schönberger, J. V. de Miranda Cardoso, J. Reimer, J. Harrington, J. L. C. Rodríguez, J. Nunez-Iglesias, J. Kuczynski, K. Tritz, M. Thoma, M. Newville, M. Kümmerer, M. Bolingbroke, M. Tartre, M. Pak, N. J. Smith, N. Nowaczyk, N. Shebanov, O. Pavlyk, P. A. Brodtkorb, P. Lee, R. T. McGibbon, R. Feldbauer, S. Lewis, S. Tygier, S. Sievert, S. Vigna, S. Peterson, S. More, T. Pudlik, T. Oshima, T. J. Pingel, T. P. Robitaille, T. Spura, T. R. Jones, T. Cera, T. Leslie, T. Zito, T. Krauss, U. Upadhyay, Y. O. Halchenko, Y. Vázquez-Baeza, *Nat. Methods* **2020**, *17*, 261.
- [114] A. Zur, T. C. McGill, *J. Appl. Phys.* **1984**, *55*, 378.
- [115] Y. Liu, S. Wang, A. M. Nolan, C. Ling, Y. Mo, *Adv. Energy Mater.* **2020**, *10*, 2002356.
- [116] T. P. Senftle, M. J. Janik, A. C. T. van Duin, *J. Phys. Chem. C* **2014**, *118*, 4967.
- [117] T. P. Senftle, R. J. Meyer, M. J. Janik, A. C. T. van Duin, *J. Chem. Phys.* **2013**, *139*, 044109.
- [118] C. Chen, Y. Zuo, W. Ye, Q. Ji, S. P. Ong, *GitHub Repos.* **2020**.
- [119] A. Togo, *J. Phys. Soc. Japan* **2023**, *92*.
- [120] A. Togo, I. Tanaka, *Scr. Mater.* **2015**, *108*, 1.
- [121] A. Togo, L. Chaput, I. Tanaka, *Phys. Rev. B* **2015**, *91*, 094306.

