

# Are Foundational Atomistic Models Reliable for Finite-Temperature Molecular Dynamics?

Denan Li,<sup>§</sup> Jiyuan Yang,<sup>§</sup> Xiangkai Chen, Lintao Yu, and Shi Liu\*



Cite This: <https://doi.org/10.1021/acs.jpcc.5c07541>



Read Online

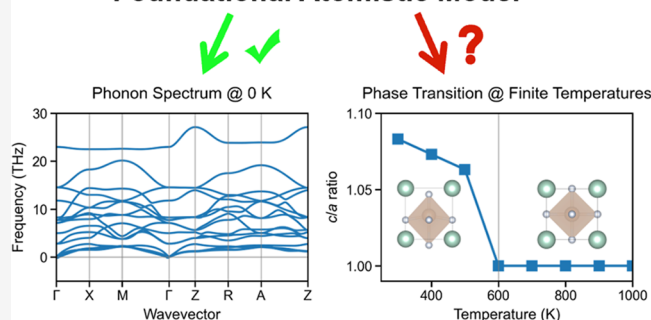
ACCESS |

Metrics & More

Article Recommendations

**ABSTRACT:** Machine learning force fields have emerged as promising tools for molecular dynamics (MD) simulations, potentially offering quantum-mechanical accuracy with the efficiency of classical MD. Inspired by foundational large language models, recent years have seen considerable progress in developing foundational atomistic models, sometimes referred to as universal force fields, designed to cover most elements in the periodic table. This perspective adopts a practitioner's viewpoint to ask a critical question: Are these foundational atomistic models reliable for one of their most compelling applications, in particular simulating finite-temperature dynamics? Instead of a broad benchmark, we use the canonical ferroelectric-paraelectric phase transition in  $\text{PbTiO}_3$  as a focused case study to evaluate prominent foundational atomistic models. Our findings suggest a potential disconnect between static accuracy and dynamic reliability. While 0 K properties are often well-reproduced, we observed that the models can struggle to consistently capture the correct phase transition, sometimes exhibiting simulation instabilities. We believe these challenges may stem from inherent biases in training data and a limited description of anharmonicity. These observed shortcomings, though demonstrated on a single system, appear to point to broader, systemic challenges that can be addressed with targeted fine-tuning. This perspective serves not to rank models, but to initiate a crucial discussion on the practical readiness of foundational atomistic models and to explore future directions for their improvement.

## Foundational Atomistic Model



## INTRODUCTION

Artificial intelligence (AI) is rapidly emerging as the fifth paradigm of scientific research, joining the established paradigms of experiments, theory, computation, and data. This transformative technology is fundamentally reshaping the nature of scientific inquiry and has the potential to significantly accelerate the pace of scientific discovery. The recognition of AI's contributions to science, highlighted by its acknowledgment in the 2024 Nobel Prizes in both Physics and Chemistry,<sup>1,2</sup> firmly establishes the era of "AI for Science." In materials science, AI holds immense potential to elucidate complex structure–property relationships, thereby enhancing and expediting the processes of materials discovery and design.

In this perspective, we primarily focus on machine learning force fields (MLFFs) for classical MD simulations. Classical MD simulations employ parametrized interatomic potentials, enabling computationally efficient exploration of dynamic processes across large temporal and spatial scales. These simulations not only reveal atomic-level mechanisms but also provide foundational data for coarse-grained models,<sup>3,4</sup> further extending their importance in multiscale simulations. Classical MD simulations have long been an indispensable tool for computer-aided drug design,<sup>5</sup> driven by classical force fields that accurately describe biomolecular interactions in proteins

and nucleic acids.<sup>6,7</sup> In contrast, the adoption of MD in computer-aided materials discovery lags behind due to the absence of force fields capable of handling diverse elements, especially transition-metal oxides and complex alloys. This challenge mainly stems from the high-dimensional potential energy surface inherent in multielement systems, where traditional analytical functionals struggle to balance accuracy and generality.

The emergence of MLFFs is transforming the field of MD simulations. By leveraging advanced techniques such as deep neural networks and graph neural networks, MLFFs could achieve quantum-mechanical accuracy while retaining the computational efficiency of classical MD. The standard protocol for developing an MLFF involves training the model on databases computed using density functional theory (DFT), which include energies, atomic forces, and virial

**Received:** November 4, 2025

**Revised:** November 13, 2025

**Accepted:** November 13, 2025

Table 1. Summary of Key Properties of Various MLFFs Used for PTO-Test<sup>a</sup>

model	version	training set (size)	parameters	energy MAE (meV/atom)	force MAE (meV/Å)
CHGNet <sup>9</sup>	0.3.0	MPtrj (146 K)	413 K	26	60
GPTFF <sup>11</sup>	v2	Atomly (37.8 M)	502 K	32	71
M3GNet <sup>8</sup>	MP-2021.2.8-PES	MPF-2021.2.8 (62.8 K)	228 K	18.7	63
MACE <sup>10</sup>	MP-0b-medium	MPtrj (146 K)	4.69 M	20	45
ORB <sup>12</sup>	orb-v2	MPtrj (146 K) + Alexandria (3.1 M)	25.2 M		
SevenNet <sup>42</sup>	7net-13i5	MPtrj (146 K)	1.17 M	8.3	29
UniPero <sup>33</sup>	v1	Customized (19 K)	≈500 K	1.75	54

<sup>a</sup>Notes: Some models have been updated since their initial publication. When available, the latest version is used, and the energy and force mean absolute errors (MAEs, if reported) are taken from the latest version if available; otherwise, they are taken from the original references. Abbreviations: MPtrj = Materials Project trajectories; MPF = Materials Project structure relaxations.

tensors across a diverse set of atomic configurations. A significant recent trend in materials modeling, paralleling the rise of foundational models in machine learning, is the development of what are often called “universal force fields”.<sup>8–16</sup> In this perspective, we refer to these as **foundational atomistic models**, a term we find useful to highlight their intended role as a general-purpose base for a wide range of downstream applications. These models are characterized by their training on vast and chemically diverse data sets, often encompassing a large portion of the periodic table, with the ultimate aim of enabling efficient simulations of complex materials systems at an accuracy approaching that of first-principles methods. Notable examples include M3GNet,<sup>8</sup> CHGNet,<sup>9</sup> and MACE,<sup>10</sup> which are based on graph neural network architectures and are trained on extensive materials science databases.<sup>17–19</sup> Proprietary advancements include GNoME,<sup>20</sup> built upon E(3)-equivariant graph neural networks, and PFP,<sup>21</sup> which leverages the TeaNet architecture<sup>22</sup> to combine attention mechanisms with graph-based atomic representations. The GPTFF model<sup>11</sup> integrates graph neural network and transformer architectures with attention mechanisms and is trained on the proprietary Atomly database. The DPA-2 model<sup>13</sup> positions itself as a pretrained model covering more than 90 elements. It is designed to significantly reduce downstream data requirements by leveraging transfer learning, enabling efficient on-demand fine-tuning to create tailored models for specific materials of user interest. These developments mark a paradigm shift toward general-purpose force fields capable of simulating complex multielement systems, from battery electrolytes to high-entropy alloys. Recently, Riebesell et al. developed Matbench Discovery, an evaluation framework for MLFFs, applied as prefilters for high-throughput searches of stable inorganic crystals.<sup>23</sup>

Many excellent reviews have focused on comprehensively benchmarking these foundational atomistic models, often ranking them based on their accuracy for static properties across vast chemical spaces.<sup>24–29</sup> In this perspective, we adopt a deliberately narrower, more practitioner-focused viewpoint. The distinctive advantage of MD over static calculations is its ability to reveal time-dependent atomic behaviors and emergent properties. From our perspective, this suggests that while static accuracy is a necessary foundation, a critical test of a foundational atomistic model's practical utility is its performance under realistic, dynamic conditions. This leads us to pose a question that we believe precedes simple model selection: Can a practitioner trust an “out-of-the-box” foundational atomistic model to reliably capture the complex physics of their specific material system, particularly during finite-temperature MD simulations? Therefore, our focus here is not

on static accuracy alone, but on the ability of these models to reproduce the dynamic behavior of materials over the time scales, particularly tens to hundreds of picoseconds, where emergent phenomena like phase transitions occur.

## METHODS

All calculations are performed using the ASE,<sup>30</sup> with each MLFF integrated as an ASE calculator to compute energy, forces, and stress. To simulate the temperature-driven phase transition of PbTiO<sub>3</sub>, a  $5 \times 5 \times 5$  supercell containing 625 atoms is constructed from the ground-state structure. MD simulations are carried out in the NPT ensemble using a Parrinello–Rahman barostat coupled with a Nosé–Hoover thermostat. At each temperature, the simulation runs with a 2 fs time step for 50,000 steps, totaling 100 ps. The last 50 ps of the trajectory is used to compute the averaged lattice constants. Our tests confirm that the cumulative average of the  $c/a$  ratio converges after a 10 ps production trajectory. For the NVT ensemble simulation, the same  $5 \times 5 \times 5$  supercell is used, with lattice parameters fixed at their experimental values measured at room temperature. Langevin dynamics are employed with a 2 fs time step for 50,000 steps. For performance benchmarking, all computations are performed on a single V100-SXM2-16G GPU, with each data point averaged over three independent runs.

## BENCHMARK DESIGN

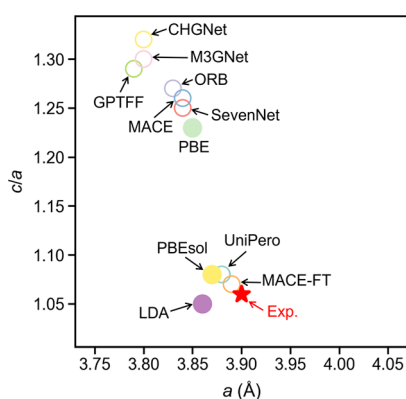
In the following (admittedly not comprehensive) performance assessment, we use the temperature-driven ferroelectric-paraelectric phase transition of PbTiO<sub>3</sub> as a test case, referred to as the PTO-test. As a prototypical ferroelectric material, PbTiO<sub>3</sub> is one of the most extensively studied perovskite oxides. Its ground state adopts a tetragonal phase (space group  $P4mm$ ) characterized by spontaneous electrical polarization, which transitions to a nonpolar cubic phase (space group  $Pm3m$ ) at temperatures above 760 K, as observed in experiments.<sup>31</sup> The tetragonal phase features a short axis,  $a$ , and a long axis,  $c$ , with the tetragonality defined by the ratio  $c/a$ , which correlates with the magnitude of the polarization. The energy difference between the ferroelectric and paraelectric phases, determined by DFT calculations at zero Kelvin, is 16 meV/atom. This moderate energy difference falls within the accuracy range of typical force fields. Furthermore, the sub-800 K transition temperature allows for direct MD validation without requiring extrapolation to extreme thermal regimes ( $>1000$  K), where anharmonic effects could introduce significant complexities. For these reasons, we consider the PTO-test an ideal benchmark: it is sufficiently complex to reveal potential limitations of foundational atomistic models in

simulating structural dynamics, yet tractable enough to enable systematic error analysis.

The foundational atomistic models selected for this benchmark include CHGNet, GPTFF, MACE, M3GNet, ORB, and SevenNet. Unfortunately, our request for access to EquiformerV2-OMat<sup>19,32</sup> was denied. Table 1 provides an overview of these MLFFs, detailing the model versions used in our tests, their training data sets, the number of trainable parameters, and the mean absolute error for energy and forces during training. Additionally, we evaluate UniPero, a “professional model” designed as a force field for perovskite oxides, covering 14 metal elements.<sup>33</sup> It mainly follows the architecture of DPA-1,<sup>34</sup> an earlier version of DPA-2.

## RESULTS

**Static Properties.** We begin by determining the ground-state structure of the tetragonal phase of PTO through structural optimizations employing various models. Figure 1



**Figure 1.** Lattice parameter  $a$  and tetragonality ( $c/a$ ) of ground-state  $\text{PbTiO}_3$  predicted by various MLFFs and exchange-correlation functionals.

compares the lattice parameter  $a$  and the tetragonality ( $c/a$ ) predicted by these models with results from standard exchange-correlation functionals, including LDA, PBE, and PBEsol. The values are also summarized in Table 2. It is well-known that the PBE functional significantly overestimates the

**Table 2.** Lattice Parameters ( $a$  and  $c$ ), Tetragonality ( $c/a$ ) of the Tetragonal  $\text{PbTiO}_3$  Phase, and the Energy Difference ( $\Delta E$ ) between the Tetragonal and Cubic Phases Predicted by Different MLFFs<sup>a</sup>

model	$a$ (Å)	$c$ (Å)	$c/a$	$\Delta E$ (eV/f.u.)
CHGNet	3.80	5.01	1.32	0.24
GPTFF	3.79	4.88	1.29	0.25
M3GNet	3.80	4.92	1.30	0.16
MACE	3.84	4.83	1.26	0.19
ORB	3.83	4.87	1.27	0.21
SevenNet	3.84	4.79	1.25	0.21
MACE-FT	3.89	4.16	1.07	0.08
UniPero	3.88	4.21	1.08	0.09
LDA <sup>46</sup>	3.86	4.04	1.05	
PBE	3.85	4.73	1.23	0.20
PBEsol	3.87	4.20	1.08	0.08
Exp. <sup>47</sup>	3.90	4.15	1.06	

<sup>a</sup>Experimental and DFT results are also included for comparison.

$c/a$  ratio (experimental value: 1.06), yielding a value of 1.23, whereas PBEsol gives a closer estimate of 1.10. This discrepancy explains why models trained on PBE-based databases, such as CHGNet, M3GNet, and MACE, inherit this bias, predicting  $c/a$  ratios even larger than those from PBE itself. The exception is UniPero, which aligns with PBEsol due to its training on PBEsol-derived data. This reveals an expected limitation in foundational atomic models: their accuracy is inherently tied to the exchange-correlation functional used in their training database. For systems like PTO, where even conventional exchange-correlation functionals struggle to reproduce key properties like tetragonality, selecting an appropriate functional a priori becomes essential for developing reliable MLFFs.

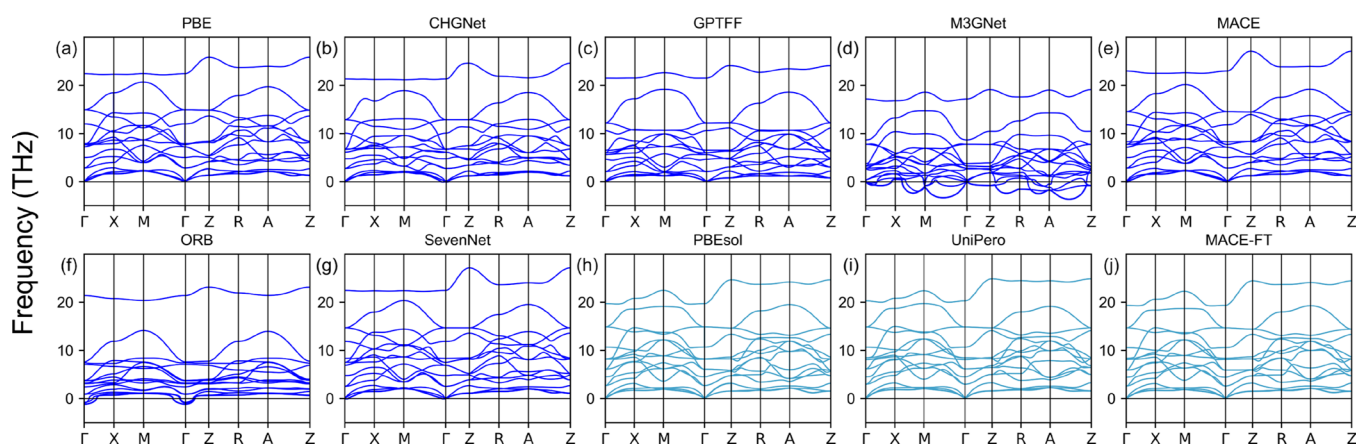
To show how such limitations might be addressed, we investigate the effect of fine-tuning. Specifically, we fine-tuned the MACE model on the small PBEsol-based data set from the UniPero study to create a new model, MACE-FT. This relatively simple procedure appears to be highly effective; as shown in Figure 1, the MACE-FT model yields a ground-state structure in strong agreement with the PBEsol reference, a result the original model did not achieve.

We further calculate the phonon spectrum of the optimized tetragonal PTO for each model using the finite-displacement method implemented in Phonopy package,<sup>35</sup> with atomic forces evaluated directly by the respective model. Despite their overestimated tetragonality, most models, including CHGNet, MACE, and SevenNet, generate phonon spectra free of imaginary frequencies (Figure 2), confirming dynamical stability. As shown in Figure 2, the phonon spectra of CHGNet, GPTFF, MACE, and SevenNet closely align with the PBE reference. In contrast, the phonon spectrum of M3GNet exhibits instability across the Brillouin zone; ORB displays localized instabilities near the  $\Gamma$  point, characterized by weak imaginary frequencies (below  $20 \text{ cm}^{-1}$ ), and also predicts notably flat bands for low-frequency phonons. Both UniPero and MACE-FT accurately reproduce the PBEsol phonon spectrum. Since phonon spectra are highly sensitive to the second derivatives of the potential energy surface near equilibrium, this benchmark highlights that most models effectively capture the local curvature of the energy landscape corresponding to their parent exchange-correlation functional.

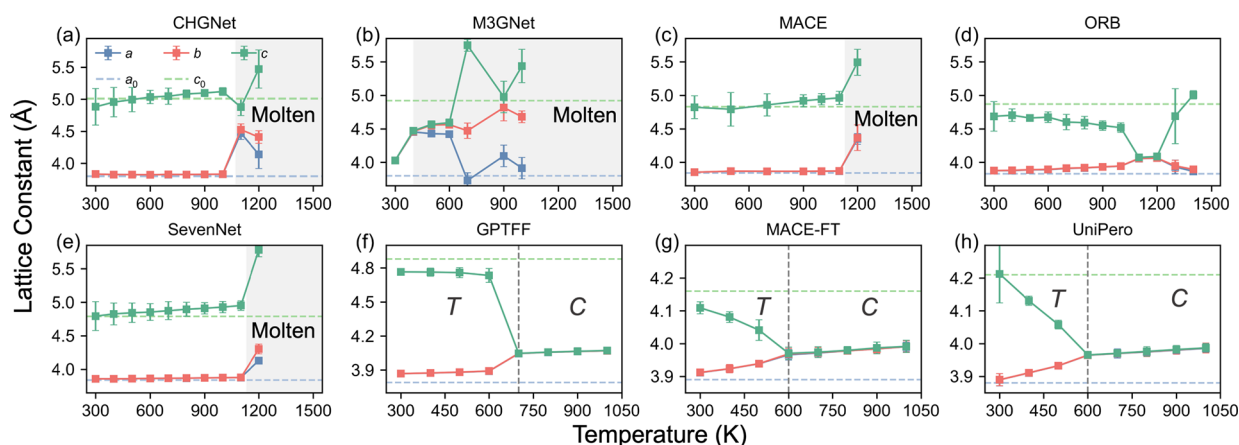
**Finite-Temperature Properties.** One might expect that an accurate representation of the local potential energy surface near the ground-state structure would ensure at least qualitative reliability for finite-temperature lattice dynamics. However, our findings reveal a significant limitation: most foundational atomistic models struggle to capture dynamic behavior accurately. As shown in Figure 3, the majority of tested models fail to reproduce the expected temperature-driven tetragonal-to-cubic phase transition during constant-pressure, constant-temperature (NPT) MD simulations.

For example, MD simulations using CHGNet, M3GNet, MACE, and SevenNet show abrupt instabilities above a critical temperature, with the system collapsing into a disordered, molten state. Before melting, CHGNet, MACE, and SevenNet stabilize an unphysical, persistent supertetragonal phase. ORB correctly captures the tetragonal-to-cubic transition near 1100 K but incorrectly predicts a reverse cubic-to-tetragonal transition at higher temperatures. Among the models tested, only GPTFF predicts a temperature-driven (super)tetragonal-to-cubic phase transition. These results indicate a key limitation: accurate modeling of local curvature near the





**Figure 2.** Phonon spectra of PbTiO<sub>3</sub> calculated using various MLFFs, each based on the optimized ground-state tetragonal structure. The panels show results for: (a) PBE, (b) CHGNet, (c) GPTFF, (d) M3GNet, (e) MACE, (f) ORB, (g) SevenNet, (h) PBEsol, (i) UniPero, and (j) MACE-FT. The spectra obtained from (a) PBE and (h) PBEsol are also included for comparison. (i) UniPero and (j) MACE-FT are trained on a PBEsol-derived database.



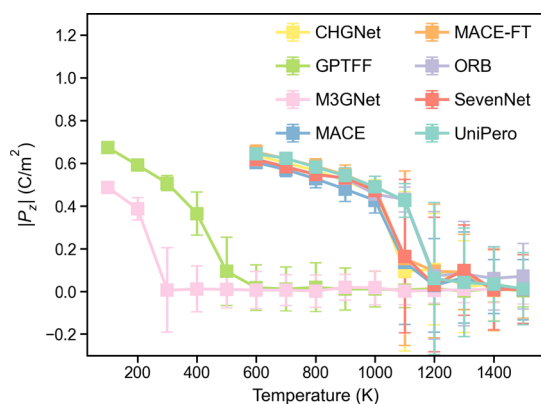
**Figure 3.** Temperature-dependent lattice constants ( $a$  and  $c$ ) obtained from  $NPT$  MD simulations using various machine learning force fields. The panels show results for: (a) CHGNet, (b) M3GNet, (c) MACE, (d) ORB, (e) SevenNet, (f) GPTFF, (g) MACE-FT, and (h) UniPero. The dashed lines indicate the ground-state lattice parameters ( $a_0$  and  $c_0$ ) of tetragonal PbTiO<sub>3</sub> for each force field. The error bars represent the standard deviation over the 50 ps production trajectory, reflecting the extent of thermal fluctuations.

ground state does not guarantee correct treatment of anharmonic interactions or free-energy landscapes governing temperature-dependent structural transitions. In contrast, both UniPero and MACE-FT successfully reproduce the expected ferroelectric–paraelectric transition, though with an underestimated Curie temperature by approximately 160 K compared to the experiment.

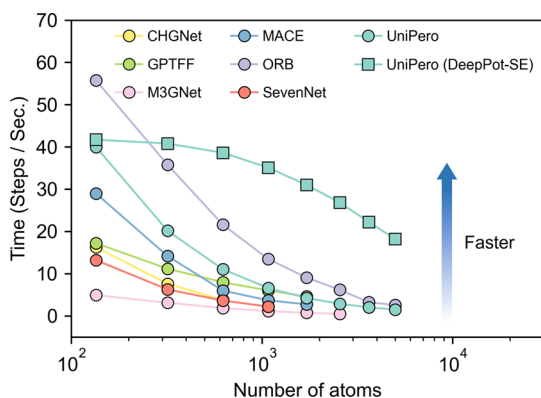
Running  $NPT$  simulations imposes stringent accuracy requirements on force fields, demanding precise parametrization to capture pressure–density relationships and accurate computation of virial contributions essential for pressure control. However, if a foundational atomistic model is primarily trained on equilibrium configurations (ground-state structures), it may lack the generalizability to handle the dynamic volume fluctuations inherent to  $NPT$  ensembles. To mitigate this challenge, we conduct a controlled validation test using constant-volume, constant-temperature ( $NVT$ ) MD simulations, fixing the lattice constants of PTO to experimental values. This approach eliminates volume relaxation, simplifying the system while still allowing us to probe temperature-driven phase transitions. Notably, most MLFFs, including CHGNet, MACE, MACE-FT, ORB, SevenNet, and UniPero, successfully

predict the ferroelectric-to-paraelectric transition, with the spontaneous polarization along the long axis ( $P_z$ ) dropping to near zero at  $\approx 1100$  K (Figure 4). In contrast, M3GNet and GPTFF exhibit significant deviations, predicting Curie temperatures far below expectations. These results indicate that under constrained  $NVT$  conditions, most models capture finite-temperature lattice behavior, as the reduced degrees of freedom simplify the energy landscape.

**Computational Efficiency.** While previous reviews have primarily focused on the accuracy of foundational atomistic models, a practitioner must also consider computational efficiency, an aspect that is often equally important, if not more so. In real-world R&D environments, where computational resources and physical time are limited, the ability to perform large-scale MD simulations efficiently becomes a critical factor in model selection. We also briefly discuss the computational efficiency of the tested models, as shown in Figure 5. It is noted that only SevenNet and the DPA-based UniPero are explicitly designed for multi-GPU parallelism, a crucial feature for large-scale MD simulations. Since some models do not support multi-GPU parallelism or specialized MD packages like LAMMPS,<sup>36</sup> the speed test is conducted on



**Figure 4.** Temperature-dependent spontaneous polarization along the  $c$ -axis ( $P_z$ ) obtained from NVT MD simulations, with lattice parameters fixed to the experimental room-temperature values ( $a = 3.90$  Å,  $c = 4.15$  Å). At high temperatures, the polarization does not fully converge to zero due to the imposed tetragonality constraint ( $c/a = 1.06$ ). The error bars represent the extent of polarization fluctuations arising from thermal effects over the 50 ps production trajectory.



**Figure 5.** Computational efficiency benchmark. The reported speed data are for reference only, as a model's optimal performance can be further improved through careful tuning and the implementation of multi-GPU parallelism.

a single GPU using the Atomic Simulation Environment (ASE),<sup>30</sup> which integrates each MLFF as a calculator. Therefore, the reported speed data are for reference only, as the optimal performance of a model could be improved with careful tuning.

Our benchmark reveals that most models have yet to fully optimize their performance for GPU acceleration. For instance, M3GNet's slower computational performance arises from unresolved GPU compatibility issues in our cluster, which defaults to CPU execution rather than leveraging GPU acceleration. While this issue might be resolved with proper settings, it highlights a potential engineering burden for users adopting a foundational atomistic model at this stage. Notably, ORB outperforms UniPero despite having a larger parameter count, leveraging the TensorFlow32 data format for enhanced efficiency; its nonconservative architecture, which directly predicts forces rather than deriving them via energy gradients, further accelerates computation. UniPero, originally based on the DPA-1 architecture with the self-attention mechanism, demonstrated significantly improved speed when simplified to a smooth edition of deep potential (DeepPot-SE)<sup>37</sup> and further compressed using techniques including tabulated

inference, operator merging, and precise neighbor indexing, making it the fastest model in our benchmarks. Furthermore, by integrating this optimized model into LAMMPS and fully harnessing multi-GPU parallelism, we successfully conducted an MD simulation of 240,000 atoms across 48 GPUs, achieving a computational speed of approximately 42 steps per second.

## DISCUSSION

While our evaluation centers on a single system,  $\text{PbTiO}_3$ , it was chosen randomly and without tuning to highlight model strengths or weaknesses. The fact that several state-of-the-art foundational atomistic models fail to reproduce well-established finite-temperature behavior in this prototypical material suggests that the issues we reveal are unlikely to be isolated. Instead, they point to broader, systemic challenges that may affect the practical deployment of foundational atomistic models across a wide range of materials and applications. In the following, we highlight three critical areas of concern from a practitioner's perspective: training data quality, computational efficiency and scalability, and the user experience within the current software ecosystem.

**Training Data Quality.** One of the most critical factors limiting model reliability is the quality of the training data. For MLFFs, this quality hinges on two key aspects: the fidelity of the reference calculations and the diversity of the sampled configurations. The choice of exchange-correlation functional plays a central role in determining reference accuracy. Recent efforts, such as MatPES and others,<sup>19,38,39</sup> have improved upon earlier data sets by adopting more advanced functionals like  $r^2\text{SCAN}$  and sampling configurations from finite-temperature MD simulations. However, our  $\text{PbTiO}_3$  results underscore that simply switching to a more sophisticated functional does not guarantee better outcomes. For example, a model trained on SCAN still overestimated the  $c/a$  ratio and Curie temperature, performing worse than one trained on PBEsol.<sup>40</sup> This highlights that the “best” exchange-correlation functional can vary by system and often requires prior domain knowledge.

Fine-tuning offers a practical solution. When we fine-tuned a modern architecture like MACE using a small amount of system-specific data, its performance improved significantly, capturing the correct phase transition. However, this introduces a trade-off: if foundational atomistic models require new DFT data and system-specific tuning to work reliably, their benefits over traditional, from-scratch approaches become less compelling. Therefore, we urge the community to initiate a more rigorous discussion around the “scaling law” for foundational atomistic models. Specifically, we need quantitative metrics to compare the true cost and speed-up of fine-tuning a large pretrained model versus training a smaller model from scratch for a given system.

**Efficiency and Scalability.** Computational efficiency and scalability present another major challenge. Real-world MD simulations often demand long time scales and large system sizes, scenarios where the performance of many foundational atomistic models degrades significantly. This is partially due to their large model sizes, which increase inference cost. Promising strategies are emerging. For example, pretraining-distillation frameworks use a large model to label system-specific data for a smaller, faster model, offering a practical balance between accuracy and speed.<sup>13</sup> Interestingly, the NEP89 achieves empirical-potential-like speed while maintaining competitive accuracy across 89 elements by combining an efficient architecture with an innovative data strategy.<sup>41</sup>

In addition, software engineering plays a critical role. Many graph-neural-network-based MLFFs lack robust multi-GPU parallelism, limiting their scalability on modern hardware. Models like SevenNet, which implement spatial decomposition for multi-GPU execution, represent important progress.<sup>42</sup> We also found that model performance can vary significantly depending on the MD engine used; switching from ASE to LAMMPS yielded substantial speedups in some cases. These observations highlight that realizing the full potential of MLFFs often requires careful optimization across both model architecture and software infrastructure.

**User Experience.** Finally, the user experience remains a considerable barrier to adoption. Practitioners, especially those outside the ML community, often face steep learning curves when deploying new foundational atomistic models. The current ecosystem is fragmented, with each model packaged in its own framework, requiring users to learn different APIs, data formats, and dependencies. Moreover, integration with professional MD engines like LAMMPS is frequently incomplete or inefficient, making it difficult to run production-level simulations or interface with established analysis tools. Encouragingly, several recent efforts aim to unify and simplify this landscape. The DeePMD-GNN plugin supports external graph neural network models like NequIP and MACE within the DeePMD-kit framework, while DeePMD-kit v3 supports multiple machine learning backends and optimized multi-GPU inference.<sup>43,44</sup> Similarly, the Materials Graph Library consolidates various GNN-based MLFFs into a single, extensible library.<sup>45</sup> These developments represent meaningful progress, but much work remains in creating an ecosystem where foundational atomistic models are as easy to use and as robust as traditional force fields.

## CONCLUSIONS

There is little doubt that we are entering an era where AI is transforming scientific inquiry, particularly in computational materials science. While this perspective does not attempt to chart the entire frontier, it addresses a focused, practical question: Can we trust foundational atomistic models for finite-temperature MD simulations? Based on our targeted benchmark, the answer is a cautious “yes”, provided their limitations are carefully considered. Many foundational atomistic models demonstrate remarkable accuracy in predicting phonon spectra and equilibrium properties. However, their behavior under realistic MD conditions can be inconsistent, especially in capturing dynamic phase transitions. While our study focuses on a single material, the observed issues likely reflect broader challenges rooted in training data quality, functional choices, and model generalizability. These problems are not intrinsic flaws of the models themselves, but rather symptoms of a still-maturing ecosystem. Moving forward, we see value in hybrid strategies that combine pretraining with targeted fine-tuning, alongside greater emphasis on benchmarking dynamic performance, improving software integration, and fostering community-driven standards. Ultimately, the goal is not to replace specialized models, but to develop robust, adaptable tools that extend the reach of atomistic simulations, unlocking AI’s full potential in computational materials discovery.

## ASSOCIATED CONTENT

### Data Availability Statement

All implementation scripts and some model weights are publicly available at <https://github.com/MoseyQAQ/PTO-test>.

## AUTHOR INFORMATION

### Corresponding Author

**Shi Liu** – Department of Physics, School of Science, Westlake University, Hangzhou, Zhejiang 310030, China; Institute of Natural Sciences, Westlake Institute for Advanced Study, Hangzhou, Zhejiang 310024, China; [orcid.org/0000-0002-8488-4848](https://orcid.org/0000-0002-8488-4848); Email: [liushi@westlake.edu.cn](mailto:liushi@westlake.edu.cn)

### Authors

**Denan Li** – Department of Physics, School of Science, Westlake University, Hangzhou, Zhejiang 310030, China

**Jiyuan Yang** – Department of Physics, School of Science, Westlake University, Hangzhou, Zhejiang 310030, China

**Xiangkai Chen** – Department of Physics, School of Science, Westlake University, Hangzhou, Zhejiang 310030, China; Institute of Natural Sciences, Westlake Institute for Advanced Study, Hangzhou, Zhejiang 310024, China

**Lintao Yu** – Department of Physics, School of Science, Westlake University, Hangzhou, Zhejiang 310030, China; [orcid.org/0000-0002-9852-5993](https://orcid.org/0000-0002-9852-5993)

Complete contact information is available at: <https://pubs.acs.org/10.1021/acs.jpcc.5c07541>

### Author Contributions

<sup>§</sup>D.L. and J.Y. contributed equally.

### Notes

The authors declare no competing financial interest.

### Biographies

**Denan Li** Denan Li earned his B.S. in Materials Science and Engineering from Ningbo University in 2024. He is currently a Ph.D. candidate in Physics at Westlake University, working in the Multiscale Materials Modeling Laboratory under the supervision of Prof. Shi Liu. His research focuses on the development and application of machine-learning-assisted multiscale methods to understand the dynamic properties of ferroelectrics, with a particular emphasis on organic–inorganic hybrid systems.

**Jiyuan Yang** Jiyuan Yang earned his B.S. in Physics from Nanjing Normal University in 2017. He received his Ph.D. in Physics from Zhejiang University in 2025 through a joint program with Westlake University. He is currently a postdoctoral fellow at Westlake University, working in Prof. Shi Liu’s lab. His research focuses on the application of deep potential molecular dynamics to study ferroelectric domain dynamics.

**Xiangkai Chen** Xiangkai Chen received his Ph.D. in Materials Science and Engineering from the University of Science and Technology Beijing in 2023. He is currently a postdoctoral fellow at Westlake University, working in Prof. Shi Liu’s lab. His research focuses on the study of oxide plasticity using the deep potential molecular dynamics method.

**Lintao Yu** Lintao Yu earned his B.S. in Chemistry from Northeast Petroleum University in 2023. He is currently a Ph.D. candidate in Physics at Westlake University in Prof. Shi Liu’s lab. His research focuses on the design and construction of high-performance computing infrastructure for large-scale simulations.



**Shi Liu** Shi Liu received his B.S. from the University of Science and Technology of China. He obtained his Ph.D. from the University of Pennsylvania in 2015. He continued his postdoctoral research at the Carnegie Institution for Science in Washington, D.C., and later at the Army Research Laboratory. In June 2019, he joined the Department of Physics at Westlake University, where he is now a tenured Associate Professor. His research interests include novel ferroelectrics, emergent topological phases in condensed matter physics, and deep-learning-based large-scale modeling of complex systems.

## ■ ACKNOWLEDGMENTS

We acknowledge the support from the National Natural Science Foundation of China (92370104) and Westlake Education Foundation. The computational resource is provided by the Westlake HPC Center.

## ■ REFERENCES

- (1) The Royal Swedish Academy of Sciences. *The Nobel Prize in Physics* 2024, 2024 (accessed 2025-03-10).
- (2) The Royal Swedish Academy of Sciences. *The Nobel Prize in Chemistry* 2024, 2024 (accessed 2025-03-10).
- (3) Schilling, T. *Phys. Rep.* **2022**, 972, 1–45.
- (4) Sun, T.; Minhas, V.; Korolev, N.; Mirzoev, A.; Lyubartsev, A. P.; Nordenskiöld, L. *Front. Mol. Biosci.* **2021**, 8, No. 645527.
- (5) Macalino, S. J. Y.; Gosu, V.; Hong, S.; Choi, S. *Arch. Pharm. Res.* **2015**, 38, 1686–1701.
- (6) Liebl, K.; Zacharias, M. *Biophys. J.* **2023**, 122, 2841–2851.
- (7) Samuel Russell, P. P.; Alaeen, S.; Pogorelov, T. V. *J. Phys. Chem. B* **2023**, 127, 9863–9872.
- (8) Chen, C.; Ong, S. P. *Nat. Comput. Sci.* **2022**, 2, 718–728.
- (9) Deng, B.; Zhong, P.; Jun, K.; Riebesell, J.; Han, K.; Bartel, C. J.; Ceder, G. *Nat. Mach. Intell.* **2023**, 5, 1031–1041.
- (10) Batatia, I.; Benner, P.; Chiang, Y.; Elena, A. M.; Kovács, D. P.; Riebesell, J.; Advincula, X. R.; Asta, M.; Avaylon, M.; Baldwin, W. J. et al. *arXiv preprint*, **2023**, arXiv:2401.00096.
- (11) Xie, F.; Lu, T.; Meng, S.; Liu, M. *Sci. Bull.* **2024**, 69, 3525–3532.
- (12) Neumann, M.; Gin, J.; Rhodes, B.; Bennett, S.; Li, Z.; Choubisa, H.; Hussey, A.; Godwin, J. *arXiv preprint*, **2024**, arXiv:2410.22570.
- (13) Zhang, D.; Liu, X.; Zhang, X.; Zhang, C.; Cai, C.; Bi, H.; Du, Y.; Qin, X.; Peng, A.; Huang, J.; et al. *npj Comput. Mater.* **2024**, 10, 293.
- (14) Kim, J.; Kim, J.; Kim, J.; Lee, J.; Park, Y.; Kang, Y.; Han, S. *J. Am. Chem. Soc.* **2025**, 147, 1042–1054.
- (15) Yin, B.; Wang, J.; Du, W.; Wang, P.; Ying, P.; Jia, H.; Zhang, Z.; Du, Y.; Gomes, C. P.; Duan, C.; Xiao, H.; Henkelman, G. *arXiv preprint*, **2025**, arXiv:2501.07155.
- (16) Yang, H.; Hu, C.; Zhou, Y.; Liu, X.; Shi, Y.; Li, J.; Li, G.; Chen, Z.; Chen, S.; Zeni, C. et al. *arXiv preprint*, **2024**, arXiv:2405.04967.
- (17) Jain, A.; Ong, S. P.; Hautier, G.; Chen, W.; Richards, W. D.; Dacek, S.; Cholia, S.; Gunter, D.; Skinner, D.; Ceder, G.; Persson, K. A. *APL Mater.* **2013**, 1, No. 011002.
- (18) Schmidt, J.; Wang, H.-C.; Cerqueira, T. F.; Botti, S.; Marques, M. A. *Sci. Data* **2022**, 9, 64.
- (19) Barroso-Luque, L.; Shuaibi, M.; Fu, X.; Wood, B. M.; Dzamba, M.; Gao, M.; Rizvi, A.; Zitnick, C. L.; Ulissi, Z. W. *arXiv preprint*, **2024**, arXiv:2410.12771.
- (20) Merchant, A.; Batzner, S.; Schoenholz, S. S.; Aykol, M.; Cheon, G.; Cubuk, E. D. *Nature* **2023**, 624, 80–85.
- (21) Takamoto, S.; Shinagawa, C.; Motoki, D.; Nakago, K.; Li, W.; Kurata, I.; Watanabe, T.; Yayama, Y.; Iriguchi, H.; Asano, Y.; et al. *Nat. Commun.* **2022**, 13, 2991.
- (22) Takamoto, S.; Izumi, S.; Li, J. *Comput. Mater. Sci.* **2022**, 207, No. 111280.
- (23) Riebesell, J.; Goodall, R. E.; Benner, P.; Chiang, Y.; Deng, B.; Ceder, G.; Asta, M.; Lee, A. A.; Jain, A.; Persson, K. A. *Nat. Mach. Intell.* **2025**, 7, 836.
- (24) Wines, D.; Choudhary, K. *ACS Mater. Lett.* **2025**, 7, 2105.
- (25) Yu, H.; Giantomassi, M.; Materzanini, G.; Wang, J.; Rignanese, G.-M. *Mater. Genome Eng. Adv.* **2024**, 2, No. e58.
- (26) Kim, G.; Na, B.; Kim, G.; Cho, H.; Kang, S.; Lee, H. S.; Choi, S.; Kim, H.; Lee, S.; Kim, Y. *Adv. Neural Inf. Process. Syst.* **2023**, 36, 51434.
- (27) Mortazavi, B. *Adv. Energy Mater.* **2025**, 15, No. 2403876.
- (28) Deng, B.; Choi, Y.; Zhong, P.; Riebesell, J.; Anand, S.; Li, Z.; Jun, K.; Persson, K. A.; Ceder, G. *npj Comput. Mater.* **2025**, 11, 9.
- (29) Poltavsky, I.; Charkin-Gorbulin, A.; Puleva, M.; Fonseca, G.; Batatia, I.; Browning, N. J.; Chmiela, S.; Cui, M.; Frank, J. T.; Heinen, S.; et al. *Chem. Sci.* **2025**, 16, 3738.
- (30) Hjorth Larsen, A.; Jørgen Mortensen, J.; Blomqvist, J.; Castelli, I. E.; Christensen, R.; Dulak, M.; Friis, J.; Groves, M. N.; Hammer, B.; Hargus, C.; et al. *J. Phys.: Condens. Matter* **2017**, 29, 273002.
- (31) Shirane, G.; Hoshino, S. *J. Phys. Soc. Jpn.* **1951**, 6, 265–270.
- (32) Liao, Y.-L.; Wood, B.; Das, A.; Smidt, T. *arXiv preprint*, **2024**, arXiv:2306.12059.
- (33) Wu, J.; Yang, J.; Liu, Y.-J.; Zhang, D.; Yang, Y.; Zhang, Y.; Zhang, L.; Liu, S. *Phys. Rev. B* **2023**, 108, L180104.
- (34) Zhang, D.; Bi, H.; Dai, F.-Z.; Jiang, W.; Liu, X.; Zhang, L.; Wang, H. *npj Comput. Mater.* **2024**, 10, 94.
- (35) Togo, A.; Chaput, L.; Tadano, T.; Tanaka, I. *J. Phys.: Condens. Matter* **2023**, 35, 353001.
- (36) Plimpton, S. J. *Comput. Phys.* **1995**, 117, 1–19.
- (37) Zhang, L.; Han, J.; Wang, H.; Saidi, W.; Car, R.; E, W. *Adv. Neural Inf. Process. Syst.* **2018**, 31, 4441–4451.
- (38) Kaplan, A. D.; Liu, R.; Qi, J.; Ko, T. W.; Deng, B.; Riebesell, J.; Ceder, G.; Persson, K. A.; Ong, S. P. *arXiv preprint*, **2025**, arXiv:2503.04070.
- (39) Levine, D. S.; Shuaibi, M.; Spotte-Smith, E. W. C.; Taylor, M. G.; Hasyim, M. R.; Michel, K.; Batatia, I.; Csányi, G.; Dzamba, M.; Eastman, P. et al. *arXiv preprint*, **2025**, arXiv:2505.08762.
- (40) Xie, P.; Chen, Y.; E, W.; Car, R. *Phys. Rev. B* **2025**, 111, No. 094113.
- (41) Liang, T.; Xu, K.; Lindgren, E.; Chen, Z.; Zhao, R.; Liu, J.; Berger, E.; Tang, B.; Zhang, B.; Wang, Y. et al. *arXiv preprint*, **2025**, arXiv:2504.21286.
- (42) Park, Y.; Kim, J.; Hwang, S.; Han, S. *J. Chem. Theory Comput.* **2024**, 20, 4857–4868.
- (43) Zeng, J.; Giese, T. J.; Zhang, D.; Wang, H.; York, D. M. *J. Chem. Inf. Model.* **2025**, 65, 3154.
- (44) Zeng, J.; Zhang, D.; Peng, A.; Zhang, X.; He, S.; Wang, Y.; Liu, X.; Bi, H.; Li, Y.; Cai, C.; et al. *J. Chem. Theory Comput.* **2025**, 21, 4375.
- (45) Ko, T. W.; Deng, B.; Nassar, M.; Barroso-Luque, L.; Liu, R.; Qi, J.; Thakur, A. C.; Mishra, A. R.; Liu, E.; Ceder, G.; Miret, S.; Ong, S. P. *npj Comput. Mater.* **2025**, 11, 253.
- (46) Meyer, B.; Vanderbilt, D. *Phys. Rev. B* **2002**, 65, No. 104111.
- (47) Mabud, S.; Glazer, A. J. *Appl. Crystallogr.* **1979**, 12, 49–53.