Benchmarking machine learning interatomic potentials via phonon anharmonicity

IOP Publishing Mach. Learn.: Sci. Technol. 5 (2024) 030502 https://doi.org/10.1088/2632-2153/ad674a

MACHINE LEARNING
Science and Technology

OPEN ACCESS

RECEIVED
4 April 2024

REVISED
16 June 2024

ACCEPTED FOR PUBLICATION
18 July 2024

PUBLISHED
14 August 2024

BENCHMARK

Benchmarking machine learning interatomic potentials via phonon anharmonicity

Sasaank Bandi[1,*], **Chao Jiang**[2] **and Chris A Marianetti**[1]

[1] Department of Applied Physics and Applied Mathematics, Columbia University, New York, NY 10027, United States of America
[2] Computational Mechanics and Materials Department, Idaho National Laboratory, Idaho Falls, ID 83415, United States of America
* Author to whom any correspondence should be addressed.

E-mail: sasaank.bandi@columbia.edu

Keywords: phonon interactions, thermal conductivity, machine learning interatomic potentials

Supplementary material for this article is available online

Abstract

Machine learning approaches have recently emerged as powerful tools to probe structure-property relationships in crystals and molecules. Specifically, machine learning interatomic potentials (MLIPs) can accurately reproduce first-principles data at a cost similar to that of conventional interatomic potential approaches. While MLIPs have been extensively tested across various classes of materials and molecules, a clear characterization of the anharmonic terms encoded in the MLIPs is lacking. Here, we benchmark popular MLIPs using the anharmonic vibrational Hamiltonian of ThO_2 in the fluorite crystal structure, which was constructed from density functional theory (DFT) using our highly accurate and efficient irreducible derivative methods. The anharmonic Hamiltonian was used to generate molecular dynamics (MD) trajectories, which were used to train three classes of MLIPs: Gaussian approximation potentials, artificial neural networks (ANN), and graph neural networks (GNN). The results were assessed by directly comparing phonons and their interactions, as well as phonon linewidths, phonon lineshifts, and thermal conductivity. The models were also trained on a DFT MD dataset, demonstrating good agreement up to fifth-order for the ANN and GNN. Our analysis demonstrates that MLIPs have great potential for accurately characterizing anharmonicity in materials systems at a fraction of the cost of conventional first principles-based approaches.

1. Introduction

Phonon anharmonicity plays an essential role in the thermodynamics of materials systems. Thus, constructing accurate vibrational Hamiltonians is critical for materials property prediction at finite temperatures. In particular, faithfully resolving phonon interactions up to fourth-order is essential for capturing leading order behavior of the phonon linewidths, phonon lineshifts, and thermal conductivity. In order to systematically assess the phonons and their interactions, it is critical to use the minimum set of derivatives allowed by group theory, which may be computed using irreducible derivative approaches [1–4]. Though these irreducible derivative methods are significantly more efficient than competing finite difference approaches, higher-order derivatives in materials with large supercells may still be prohibitively expensive. Recently, machine learning interatomic potentials (MLIPs) have emerged as powerful tools to represent the Born-Oppenheimer potential of materials. MLIP methods can achieve near quantum chemical accuracy relative to the *ab initio* method they were trained on while avoiding the polynomial scaling [5]. Here, we assess the fidelity of anharmonic derivatives computed from MLIPs.

MLIPs aim to represent the total energy of a given atomic configuration in terms of individual contributions from each atom. Since the complete space of atomic displacements is high dimensional for very large molecules and crystals, several descriptors have been developed to encode local atomic environments [6] such as coulomb matrices [7], atom centered symmetry functions [8], smooth overlap of atomic positions [9], and histograms of distances, angles, and dihedrals (HDAD) [10]. While understanding

how well certain descriptor sets can represent anharmonic energy surfaces is important, it will not be directly discussed in this work and will be the subject of further study. Our present goal is to assess the most prevalent MLIPs in the literature, using the commonly associated descriptors. Therefore, our benchmarks will be limited to three popular classes of potentials with publicly available software.

The first MLIP considered in this work is the Gaussian approximation potential (GAP) [11]. Here, each atomic energy is expanded in terms of basis functions of the descriptors, and the distribution of basis weights is assumed to be normally distributed about zero. Therefore, the energy of a given configuration is also normally distributed, where the mean provides a prediction from the model, and the variance can be a measure of uncertainity. The advantages of this type of potential are the characterization of model uncertainty and that the analytical form of the prediction is a sum of basis functions. In practice, the number of basis functions may be prohibitively large, and the 'kernel trick' is employed to reduce memory and cost requirements. It should be noted that the energy prediction provided by GAP is equivalent to the prediction of the kernel ridge regression on the same basis set. Recently, GAP has been used to predict second and third-order derivatives in two dimensional materials where good agreement with density functional theory (DFT) is achieved with integrated properties such as phonon lifetimes and thermal conductivity [12–16]. Online trained GAP models have also been implemented in the Vienna *ab initio* software package (VASP) [17, 18] and have also been used along with approaches such as the stochastic self-consistent harmonic approximation (SSCHA) [19] to predict temperature dependent phonon properties using Monte Carlo sampling [20, 21].

While not directly formulated as a MLIP, several potential energy surface fitting approaches, such as compressive sensing lattice dynamics (CSLD) [22, 23], have a similar mathematical foundation as GAP. In both models, a Bayesian polynomial regression framework is used to fit model parameters to sampled DFT data. While GAP generally fits descriptors of the local atomic environment, CSLD uses the complete Cartesian coordinate basis which ensures that fitting coefficients are more analogous to force constants. Due to the combinatorial explosion of high-order derivatives when using the Cartesian basis, least squares solutions tend to overfit parameters. CSLD overcomes this impediment by fitting force constants using the least absolute shrinkage and selection operator (LASSO), which has the same objective function as ordinary least squares but with an additional penalty on the L_1 norm of the model parameters [24]. The sparse solutions provided by LASSO have yielded accurate anharmonic properties in various systems [22, 23, 25].

The second class of machine learning approaches we consider is the high dimensional artificial neural network (ANN) first described by Behler and Parrinello (BPNN) [26]. In this model, each atomic species is represented by a convolutional neural network, and the total energy is computed as a sum of the output of each neural network. Historically, atom-centered symmetry functions [8] have been used as descriptors to encode rotational invariance and have been systematically improved over time. Recently, a set of descriptors constructed as irreducible representations of the Euclidean group has been shown to encode rotational equivariance, providing better predictions of tensor properties [27], such as the forces. BPNNs have been used to predict phonon dispersions and thermal conductivity in semiconductors [28–33], new thermoelectric candidate materials [34], and superlattices [35].

Finally, the most recent developments in machine learning interatomic potentials have focused on graph neural networks [5, 36–40]. In this architecture, graphs are constructed for all neighboring atoms where nodes encode atomic information, and bonding information is incorporated through edge connections. Of the three models described in this work, graph neural networks (GNN) have been shown to achieve the highest accuracy in reproducing forces and energies while requiring fewer training samples than the other two approaches. Recently, several developments have been aimed at developing a universal interatomic potential trained on large DFT databases, such as the materials project [41], and phonon dispersions computed using these universal models have shown good agreement with DFT [36, 39, 40]. While GNNs have been used to predict thermal conductivity directly by training on experimental and first principles data from materials property databases [42, 43], we are not aware of any other work using GNN interatomic potentials to compute anharmonic observables via phonon interactions.

In this work, we benchmark how well MLIPs reconstruct anharmonic potentials in ThO_2. We trained the GAP, BPNN, and GNN using two datasets: one containing configurations evaluated by an anharmonic Taylor series, which does not contain noise, and one generated using DFT calculations. In our analysis, we quantify errors at the level of the irreducible derivatives of the potential energy surface, making our benchmark more robust than benchmarks using integrated observables. In all three models, third-order derivatives are in good agreement with the reference. Up to fifth-order derivatives are mostly reproduced by the ANN and GNN, with the GNN demonstrating particularly promising accuracy. We also provide comparisons of observables such as phonon dispersions, phonon linewidths, phonon lineshifts, and thermal conductivity to quantify the effects of the errors in the irreducible derivatives.

2. Computational methods and details

DFT calculations were performed using the projector augmented wave (PAW) method implemented in the VASP using the local density approximation [17, 18]. The plane-wave basis cutoff energy was set to 500 eV. A Γ centered $10\times10\times10$ k-point mesh was used for primitive cell calculations, and measurements in other supercells were conducted with corresponding $k-$mesh densities. Gaussian smearing with $\sigma = 0.1$ was used for k-point integrations. DFT energies were converged to within 10^{-6} eV, and the unit cell was relaxed until all forces were within 0.02 eV Å^{-1}.

Second (i.e. phonons), third, fourth, fifth, and sixth-order derivatives of the Born–Oppenheimer potential of ThO_2 were computed from DFT and the MLIPs via irreducible derivative approaches [1]. The finite difference calculations were extrapolated to a discretization size of zero using quadratic errortails with ten discretizations (for more information see [1, 3]). Phonons and third-order derivatives commensurate with the $2 \times 2 \times 2$ supercell, which contains 24 atoms, were computed using the lone irreducible derivative approach with forces (LID_1) and energy derivatives (LID_0) for DFT and the MLIPs, respectively. Fourth-order phonon interactions commensurate with the conventional cubic supercell, which contains 12 atoms, were computed using the bundled irreducible derivative (BID) approach for DFT due to the large computational cost. While BID generally evaluates derivatives using the smallest set of displacements allowed by group theory, in this case the number of displacements was tripled to ensure robust derivatives. The fourth-order derivatives were evaluated with LID_1 for all three machine learning models. Fifth-order derivatives associated with the following Q points were selectively computed using LID_1 for both DFT and the MLIPs: $Q = \left(\left(\frac{1}{2},0,0\right),\left(\frac{1}{2},0,0\right),(0,0,0),(0,0,0),(0,0,0)\right)$ and $Q = ((0,0,0),(0,0,0),(0,0,0),(0,0,0),(0,0,0))$. Finally, only sixth-order interactions commensurate with the primitive cell were measured with LID_1.

Two datasets were prepared to train the machine learning models. The first dataset derives from a vibrational Hamiltonian containing second, third, and fourth-order irreducible derivatives computed from DFT. Datapoints are generated by performing molecular dynamics (MD) on the irreducible derivative Hamiltonian using the microcanonical ensemble, which we refer to as irreducible derivative molecular dynamics (IDMD) [4]. As a result, all energies and forces in the dataset are in perfect agreement with the irreducible derivatives, yielding a noiseless dataset, which we refer to as the *IDMD* dataset. The IDMD was conducted in the $4 \times 4 \times 4$ supercell, containing 192 atoms, for 4000 steps with a timestep of 4 fs. MD trajectory velocities were initialized by sampling from the Maxwell–Boltzmann distribution at 2000 K and a temperature range of 1522–2238 K was observed throughout the simulation. The second dataset, referred to as *DFTMD*, was generated using *ab initio* MD in the conventional cubic cell tripled in all three dimensions ($3S_C$), which contains 324 atoms. MD was conducted using the canonical ensemble at 2000 K and 3000 K, and 1346 snapshots were taken along the two trajectories. The *DFTMD* dataset was previously used to study defect properties in ThO_2 [44]. While this *DFTMD* dataset contains noise inherent to the numerics of the DFT calculation, it probes all derivatives commensurate with the $3S_C$ supercell up to infinite-order. Each dataset was constructed to provide a unique test of the MLIPs. Since *IDMD* is a noiseless dataset, this benchmark purely probes each MLIP's ability to learn anharmonic interactions while avoiding any influence of how well the model accuracy scales with noise. On the other hand, *DFTMD* is a dataset that has been constructed in accordance with the current state-of-the-art and represents a typical dataset used for MLIP training. Since *DFTMD* contains infinite-order interactions, it provides the opportunity to test the limits of each MLIP and determine the order at which the MLIP fails to replicate DFT results.

All three MLIPs were trained on *IDMD*, and only the BPNN and GNN were trained on *DFTMD*. The GAP model was generated using the quantum mechanics and interatomic potentials (QUIP) code [11, 45, 46]. Descriptors were constructed using two-body terms with a cutoff of 5 Å, three-body terms with a cutoff of 4 Å, and with the smooth overlap of positions (SOAP) descriptor with a cutoff of 6 Å. The neural network potential was trained using the n2p2 code [26, 47]. Atom-centered radial and angular symmetry functions with a cutoff radius of 10 Å and 6 Å were used, respectively. For each atom, a four-layer neural network was constructed which contained two hidden layers with 21 neurons each. The size of the input layer for thorium atoms was 41, while the size was 46 for oxygen atoms. Finally, the graph neural network was trained using nequIP [5]. Graph edges were constructed with a cutoff of 6 Å with four interaction blocks. The E(3) equivariant spherical harmonics were used with radial basis functions to form descriptor sets [27]. More detailed model information for all three MLIPs is provided in supplementary information [48].

3. Results

We begin by analyzing the results of the machine learning models trained on *IDMD*. In figure 1, we show the results of all three models where GAP, BPNN, and GNN results are shown in green, blue, and red,

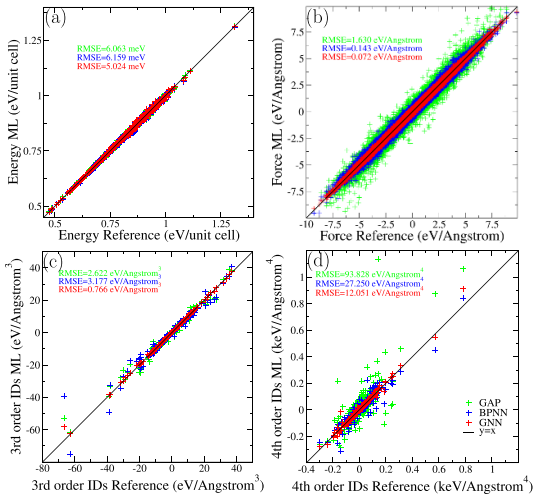

Figure 1. Results of the MLIPs trained on *IDMD* compared to the reference, where green, blue, and red dots indicate GAP, BPNN, and GNN results, respectively. Panels (a) and (b) show the energies and forces for a distinct testing set, respectively. Panels (c) and (d) compare the third and fourth-order irreducible derivatives computed from the MLIPs to the reference irreducible derivatives, respectively.

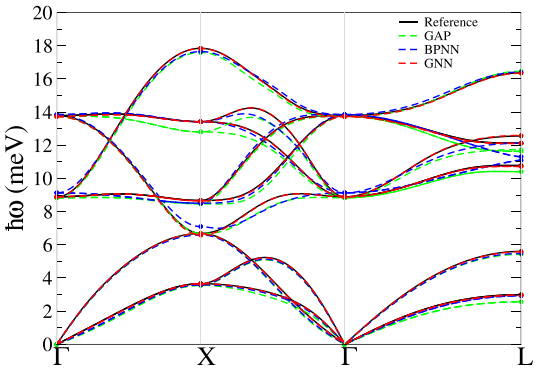

Figure 2. Computed phonon dispersion, where green, blue, red, and black points represent explicit measurements of the irreducible derivatives using GAP, BPNN, GNN, and the reference, respectively. Solid lines are a Fourier interpolation.

respectively. In figure 1(a), we compare energies computed on a distinct testing set generated using IDMD, showing excellent agreement. Figure 1(b) compares the forces from all three models on the testing set; the forces from GAP are in reasonable agreement with the reference, while the BPNN and GNN are in much better agreement. In panels 1(c), (d) and table 1, the third and fourth-order irreducible derivatives are compared to the reference. In all three models, the third-order derivatives are in good agreement with the reference, with the GNN yielding the best fidelity of the three. At fourth-order, the BPNN and GNN are in reasonable and good agreement with the reference, respectively, while the GAP model fails to provide accurate irreducible derivatives.

To demonstrate the practical impact of the errors in the computed irreducible derivatives, we compare several observables computed using irreducible derivatives obtained from each model. We begin by presenting the phonon dispersion (see figure 2). Solid dots represent grid points where irreducible derivatives are computed, and the solid lines are a Fourier interpolation. We demonstrate that for most phonon branches, all three models are in good agreement with the reference, though the BPNN and the GAP both have non-trivial discrepancies at the X and L points. We proceed by presenting the phonon linewidths from the bubble diagram [4], which uses the third-order irreducible derivatives, in a typical acoustic and optical branch (see figure 3). All other branches are included in supplementary material [48]. In the acoustic branch (figure 3(a)), all three models are in relatively good agreement with the reference, though GAP overestimates the linewidths near the X point by about 16%. The optical branch linewidths (figure 3(b)) are more sensitive to errors in the irreducible derivatives, and thus, both GAP and BPNN yield discrepancies up to 30% between Γ and X. We note that the GNN yields near perfect agreement with the reference for both branches. Next, we present the phonon lineshifts computed by evaluating the loop diagram [4] using the

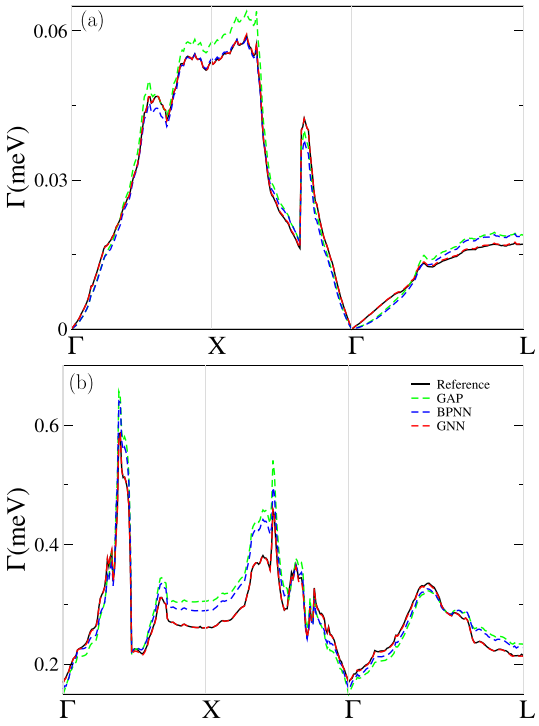

Figure 3. Phonon linewidth contribution from the bubble diagram, which uses the third-order irreducible derivatives. Panels (a) and (b) show the linewidth of an acoustic and optical branch along a high symmetry path, respectively. The GAP, BPNN, GNN, and the reference are shown in green, blue, red, and black, respectively.

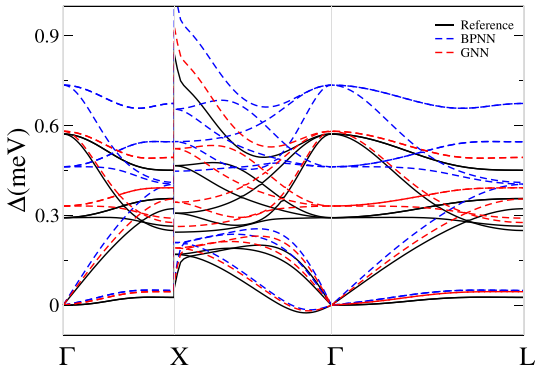

Figure 4. Phonon lineshift contribution from the loop diagram, which uses the fourth-order irreducible derivatives. The BPNN, GNN, and the reference results are shown in blue, red, and black, respectively.

fourth-order irreducible derivatives (see figure 4). Only the BPNN and GNN lineshifts have been plotted due to the large discrepancies in the GAP results (see supplementary material [48]). While the BPNN and GNN are in good agreement with the reference for the acoustic branches, the BPNN significantly overestimates the shifts of the optical branches, while the GNN is still within 18% of the reference. Finally, we compare the thermal conductivity computed using the phonon lifetimes in the Boltzmann transport equation within the relaxation time approximation. In panel 5(a), the phonon lifetimes have been evaluated using the bubble diagram, which only uses the third-order derivatives, while in panel 5(b) phonon lifetimes have been evaluated using only the sunset diagram [4], which uses the fourth-order irreducible derivatives. In both cases, it is evident that the thermal conductivity is less sensitive to errors in the irreducible derivatives, and all three models are in good agreement with the reference.

We proceed by discussing the results of the BPNN and GNN models trained on *DFTMD*. Since we have already demonstrated that the GAP model fails to capture interactions beyond third-order, it was not used for this benchmark aimed at determining the highest order of derivatives that can be faithfully reproduced using MLIPs. We present a comparison of the third, fourth, fifth, and sixth-order irreducible derivatives computed from the MLIPs and DFT (see figure 6 and table 1). For third, fourth, and fifth-order derivatives,

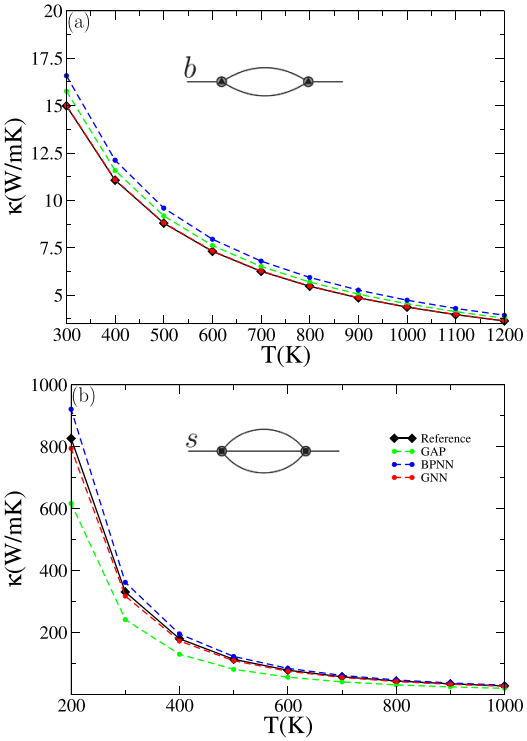

Figure 5. Thermal conductivity evaluated using lifetimes exclusively from the bubble diagram in panel (a) and lifetimes exclusively from the sunset diagram in panel (b). GAP, BPNN, and GNN results are shown with green, blue, and red solid dots and the reference results are shown with black diamonds. Solid lines are drawn to direct the eyes.

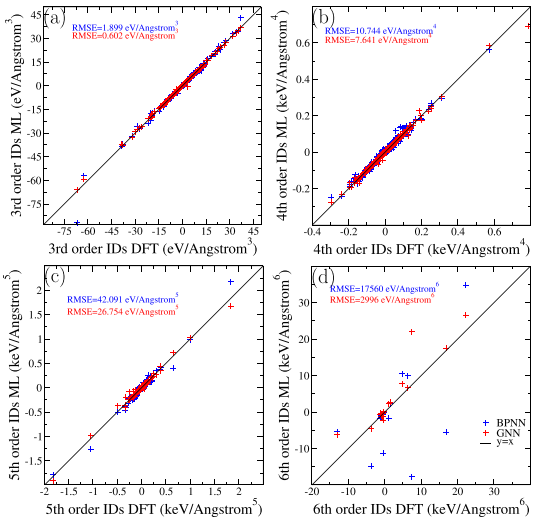

Figure 6. Anharmonic derivatives of the MLIPs trained on *DFTMD* compared to the DFT reference where BPNN and GNN results are shown in blue and red, respectively.

both models are in excellent agreement with DFT, with the GNN outperforming the BPNN by a factor of 1.5 to 3 depending on the derivative order. At sixth-order, both models struggle to reproduce the irreducible derivatives computed from DFT, though the GNN generates reasonable results, especially on low-magnitude derivatives.

In the preceding, we benchmarked the ability of MLIPs to learn the anharmonic potential energy surface in thoria, which is a band insulator. To understand the transferability of this benchmark to metallic systems, we constructed a similar dataset to *IDMD* using the irreducible derivatives of face-centered cubic aluminum and used it to train the GAP, BPNN, and GNN, where we found good agreement with the reference for phonons, third-order, and fourth-order derivatives. The details and results of this benchmark can be found in supplementary materials [48].

Table 1. Table of root mean squared error divided by the average magnitude of the irreducible derivatives for each model trained on the *IDMD* and *DFTMD* datasets.

	IDMD		*DFTMD*			
	3rd derivative	4th derivative	3rd derivative	4th derivative	5th derivative	6th derivative
GAP	0.2207	1.652	—	—	—	—
BPNN	0.3100	0.480	0.1853	0.1890	0.4707	0.7965
GNN	0.0747	0.212	0.0587	0.1346	0.2991	0.1359

4. Conclusions

In summary, we have benchmarked three popular machine learning interatomic potentials on the anharmonic irreducible derivatives of ThO_2. We have developed two training datasets: one that is computed from the anharmonic vibrational Hamiltonian of ThO_2, containing up to quartic terms, and one that is representative of training conducted via *ab initio* methods in the literature. The Behler Parrinello ANN and the GNN yield robust irreducible derivatives up to fifth-order, while the Gaussian approximation potential is only able to accurately capture anharmonic interactions up to third-order. Our work demonstrates the promising potential of machine learning methods in characterizing anharmonicity in materials systems. While we obtained accurate results using MLIPs in the present study, it would be informative to verify that our findings hold in systems with sensitivities, such as soft phonon modes. Future work will include extending this analysis to strongly correlated electronic materials, where the generation of accurate training data comes at a significant premium. Additionally, to study materials with strong anharmonicity, further work will be conducted to develop training algorithms to more accurately capture sixth-order and beyond interactions.

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: 10.5281/zenodo.10 928 194 [49].

Acknowledgments

The development of MLIPs for ThO_2 was supported by the INL Laboratory Directed Research & Development (LDRD) Program under DOE Idaho Operations Office Contract DE-AC07-05ID14517. The construction of anharmonic Hamiltonians was supported by the Grant DE-SC0016507 funded by the U.S. Department of Energy, Office of Science. This research used resources of the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

ORCID iDs

Sasaank Bandi https://orcid.org/0009-0002-8854-4215
Chris A Marianetti https://orcid.org/0000-0003-3812-2751

References

[1] Fu L, Kornbluth M, Cheng Z and Marianetti C A 2019 Group theoretical approach to computing phonons and their interactions *Phys. Rev.* B **100** 014303
[2] Mathis M A *et al* 2022 Generalized quasiharmonic approximation via space group irreducible derivatives *Phys. Rev.* B **106** 014314
[3] Bandi S and Marianetti C A 2023 Precisely computing phonons via irreducible derivatives *Phys. Rev.* B **107** 174302
[4] Xiao E and Marianetti C A 2023 Anharmonic phonon behavior via irreducible derivatives: self-consistent perturbation theory and molecular dynamics *Phys. Rev.* B **107** 094303
[5] Batzner S, Musaelian A, Sun L, Geiger M, Mailoa J P, Kornbluth M, Molinari N, Smidt T E and Kozinsky B 2022 E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials *Nat. Commun.* **13**
[6] Langer M F, Goeβmann A and Rupp M 2022 Representations of molecules and materials for interpolation of quantum-mechanical simulations via machine learning *npj Comput. Mater.* **8** 41
[7] Rupp M, Tkatchenko A, Müller K-R and von Lilienfeld O A 2012 Fast and accurate modeling of molecular atomization energies with machine learning *Phys. Rev. Lett.* **108** 058301
[8] Behler J 2011 Atom-centered symmetry functions for constructing high-dimensional neural network potentials *J. Chem. Phys.* **134** 074106
[9] Bartók A P, Kondor R and Csányi G 2013 On representing chemical environments *Phys. Rev.* B **87** 184115

[10] Faber F A, Hutchison L, Huang B, Gilmer J, Schoenholz S S, Dahl G E, Vinyals O, Kearnes S, Riley P F and von Lilienfeld O A 2017 Prediction errors of molecular machine learning models lower than hybrid dft error *J. Chem. Theory Comput.* **13** 5255
[11] Bartók A P, Payne M C, Kondor R and Csányi G 2010 Gaussian approximation potentials: the accuracy of quantum mechanics, without the electrons *Phys. Rev. Lett.* **104** 136403
[12] Kocabas T, Keeli M, Vázquez-Mayagoitia A and Sevik C 2023 Gaussian approximation potentials for accurate thermal properties of two-dimensional materials *Nanoscale* **15** 8772
[13] Ouyang Y, Zhang Z, Yu C, He J, Yan G and Chen J 2020 Accuracy of machine learning potential for predictions of multiple-target physical properties *Chin. Phys. Lett.* **37** 126301
[14] Zhang Y, Shen C, Long T and Zhang H 2020 Thermal conductivity of h-bn monolayers using machine learning interatomic potential *J. Phys.: Condens. Matter* **33** 105903
[15] Wei Z, Zhang C, Kan Y, Zhang Y and Chen Y 2022 Developing machine learning potential for classical molecular dynamics simulation with superior phonon properties *Comput. Mater. Sci.* **202** 111012
[16] Arabha S, Aghbolagh Z S, Ghorbani K, Hatam-Lee S M and Rajabpour A 2021 Recent advances in lattice thermal conductivity calculation using machine-learning interatomic potentials *J. Appl. Phys.* **130** 210903
[17] Kresse G and Furthmüller J 1996 Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set *Comput. Mater. Sci.* **6** 15
[18] Kresse G and Hafner J 1993 Ab initiomolecular dynamics for liquid metals *Phys. Rev.* B **47** 558
[19] Monacelli L, Bianco R, Cherubini M, Calandra M, Errea I and Mauri F 2021 The stochastic self-consistent harmonic approximation: calculating vibrational properties of materials with full quantum and anharmonic effects *J. Phys.: Condens. Matter* **33** 363001
[20] Ranalli L, Verdi C, Monacelli L, Kresse G, Calandra M and Franchini C 2023 Temperature dependent anharmonic phonons in quantum paraelectric ktao3 by first principles and machine learned force fields *Adv. Quantum Technol.* **6** 2200131
[21] Verdi C, Ranalli L, Franchini C and Kresse G 2023 Quantum paraelectricity and structural phase transitions in strontium titanate beyond density functional theory *Phys. Rev. Mater.* **7** l030801
[22] Zhou F, Nielson W, Xia Y and Ozolins V 2014 Lattice anharmonicity and thermal conductivity from compressive sensing of first-principles calculations *Phys. Rev. Lett.* **113** 185501
[23] Zhou F, Nielson W, Xia Y and Ozolins V 2019 Compressive sensing lattice dynamics. I. general formalism *Phys. Rev.* B **100** 184308
[24] Tibshirani R 1996 Regression shrinkage and selection via the lasso: a retrospective *J. R. Stat. Soc.* B **73** 273
[25] Tadano T and Tsuneyuki S 2015 Self-consistent phonon calculations of lattice dynamical properties in cubic $SrTio_3$ with first-principles anharmonic force constants *Phys. Rev.* B **92** 054301
[26] Behler J and Parrinello M 2007 Generalized neural-network representation of high-dimensional potential-energy surfaces *Phys. Rev. Lett.* **98** 146401
[27] Geiger M and Smidt T 2022 e3nn: euclidean neural networks
[28] Li R, Lee E and Luo T 2020a A unified deep neural network potential capable of predicting thermal conductivity of silicon in different phases *Mater. Today Phys.* **12** 100181
[29] Li R, Liu Z, Rohskopf A, Gordiz K, Henry A, Lee E and Luo T 2020 A deep neural network interatomic potential for studying thermal conductivity of β -Ga_2O_3 *Appl. Phys. Lett.* **117** 152102
[30] Mangold C, Chen S, Barbalinardo G, Behler J, Pochet P, Termentzidis K, Han Y, Chaput L, Lacroix D and Donadio D 2020 Transferability of neural network potentials for varying stoichiometry: phonons and thermal conductivity of MnxGey compounds *J. Appl. Phys.* **127** 244901
[31] Minamitani E, Ogura M and Watanabe S 2019 Simulating lattice thermal conductivity in semiconducting materials using high-dimensional neural network potential *Appl. Phys. Express* **12** 095001
[32] Babaei H, Guo R, Hashemi A and Lee S 2019 Machine-learning-based interatomic potential for phonon transport in perfect crystalline si and crystalline si with vacancies *Phys. Rev. Mater.* **3** 074603
[33] Qian X, Peng S, Li X, Wei Y and Yang R 2019 Thermal conductivity modeling using machine learning potentials: application to crystalline and amorphous silicon *Mater. Today Phys.* **10** 100140
[34] Choi J M, Lee K, Kim S, Moon M, Jeong W and Han S 2022 Accelerated computation of lattice thermal conductivity using neural network interatomic potentials *Comput. Mater. Sci.* **211** 111472
[35] Qu R, Lv Y and Lu Z 2023 A deep neural network potential to study the thermal conductivity of $MnBi_2Te_4$ and $Bi_2Te_3/MnBi_2Te_4$ superlattice *J. Electron. Mater.* **52** 4475
[36] Takamoto S, Izumi S and Li J 2022 Teanet: universal neural network interatomic potential inspired by iterative electronic relaxations *Comput. Mater. Sci.* **207** 111280
[37] Batatia I, Kovács D P, Simm G N C, Ortner C and Csányi G 2022 Mace: higher order equivariant message passing neural networks for fast and accurate force fields
[38] Gasteiger J, Groß J and Günnemann S 2020 Directional message passing for molecular graphs
[39] Chen C and Ong S P 2022 A universal graph deep learning interatomic potential for the periodic table *Nat. Comput. Sci.* **2** 718
[40] Deng B, Zhong P, Jun K, Riebesell J, Han K, Bartel C J and Ceder G 2023 Chgnet as a pretrained universal neural network potential for charge-informed atomistic modelling *Nat. Mach. Intell.* **5** 1031
[41] Jain A *et al* 2013 Commentary: the materials project: a materials genome approach to accelerating materials innovation *APL Mater.* **1** 011002
[42] Ojih J, Rodriguez A, Hu J and Hu M 2023 Screening outstanding mechanical properties and low lattice thermal conductivity using global attention graph neural network *Energy AI* **14** 100286
[43] Sun Y and Hu W 2021 Novel machine learning framework for thermal conductivity prediction by crystal graph convolution embedded ensemble *SmartMat* **3** 474
[44] Jiang C, Marianetti C A, Khafizov M and Hurley D H 2024 Machine learning potential assisted exploration of complex defect potential energy surfaces *npj Comput. Mater.* **10** 21
[45] Csányi G, Winfield S, Kermode J R, De Vita A, Comisso A, Bernstein N and Payne M C 2007 Expressive programming for computational physics in fortran 95+, IoP Comput. *Phys. Newsletter*

[46] Kermode J R 2020 f90wrap: an automated tool for constructing deep python interfaces to modern fortran codes *J. Phys.: Condens. Matter* **32** 305901
[47] Singraber A, Behler J and Dellago C 2019 Library-based lammps implementation of high-dimensional neural network potentials *J. Chem. Theory Comput.* **15** 1827
[48] See Supplemental Material at [URL will be inserted by publisher] for machine learning model details, more computational details for the calculation of observables, and additional linewidth, lineshift, and thermal conductivity plots.
[49] Bandi S, Jiang C and Marianetti C 2024 Supplemental data for benchmarking machine learning interatomic potentials via phonon anharmonicity