Accelerated computation of lattice thermal conductivity using neural network interatomic potentials

Jeong Min Choi\textsuperscript{a,1}, Kyeongpung Lee\textsuperscript{a,1}, Sangtae Kim\textsuperscript{a}, Minseok Moon\textsuperscript{a}, Wonseok Jeong\textsuperscript{b,\ast}, Seungwu Han\textsuperscript{a,\ast}

\textsuperscript{a} Department of Materials Science and Engineering, Seoul National University, Seoul 08826, Republic of Korea
\textsuperscript{b} Lawrence Livermore National Laboratory, 7000 East Ave., Livermore, CA 94550, USA

\ast Corresponding authors.

E-mail addresses: jeong3@llnl.gov (W. Jeong), hansw@snu.ac.kr (S. Han).
\textsuperscript{1} Jeong Min Choi and Kyeongpung Lee contributed equally to this paper.

1. Introduction

The prediction of thermal conductivity is crucial in selecting materials for various applications [1]. For example, materials with low thermal conductivities can be used for thermoelectrics [2–4] or thermal insulations [5], while materials with high thermal conductivities are suitable for the thermal management of electronic devices [6]. Over the last decades, the development of \textit{ab initio} methods combined with increasing computational power enabled a reliable prediction of lattice thermal conductivity ($\kappa_l$) [4,7–12]. This is particularly impressive because the computed $\kappa_l$ spans a wide range of scales from $10^{-1}$ to $10^3$ Wm$^{-1}$K$^{-1}$. However, the \textit{ab initio} evaluation of $\kappa_l$ becomes expensive in multicomponent or low-symmetry materials due to high computational costs in obtaining anharmonic terms in the interatomic force constant. For instance, the monoclinic $\beta$-Ga$_2$O$_3$ requires thousands of single-point density functional theory (DFT) calculations to obtain $\kappa_l$. Several approaches that exploit regression techniques have been proposed to save the computational cost, resulting in a significant increase in the computational efficiency [13–19]. One example is the compressive sensing lattice dynamics method, which uses sparsity in the force constants [13,14]. The potential energy surface is Taylor-expanded with the high-order force constants (fourth-order and beyond) in this method, which is fitted to DFT atomic forces for reference structures [14]. The temperature-dependent effective-potential method (TDEP) also expands...
Machine-learning interatomic potentials (MLPs) have recently been used as surrogate models of DFT in calculating force constants and $\kappa_i$ [26–43]. To fit the potential energy surface produced by ab initio calculations, preferably DFT [44], MLPs use flexible regression models. For regression models, artificial neural networks [45], kernel-based methods [46], and linear fitting [47] are popular choices. Once trained, MLPs infer the energy, atomic force, and virial stress of the given structure with an accuracy comparable to DFT, but at a fraction of the cost. Until now, several crystals and alloys have been studied, and $\kappa_i$ values obtained by MLPs have been close to the reference DFT data [38,42,43]. Notably, the $\kappa_i$ of Ba was successfully predicted by MLP, which demands up to four-phonon scattering in the Boltzmann transport equation [38]. MLPs are also advantageous for investigating the $\kappa_i$ of disordered phases such as amorphous [40] and liquid systems [39]. For materials dominated by nonperturbative phonon scattering, extensive molecular dynamics (MD) simulations are necessary to get an accurate $\kappa_i$, which can be handled efficiently by MLP [41].

Although previous works support that the MLP is a powerful tool in calculating $\kappa_i$, several issues need to be addressed. For example, most materials in the previous studies retain high symmetries, three or fewer elements, relatively simple configurations, and a limited range of $k_i (>10$ Wm$^{-1}$K$^{-1}$). Consequently, it is unclear whether the MLP’s prediction accuracy can be maintained for materials in general. Furthermore, recipes for creating the training set differ significantly across the literature, making it difficult to establish a standard approach. Motivated by these observations, we herein investigate the effect of material complexity and different types of training sets on the accuracy and computational cost of $\kappa_i$ prediction by MLPs. We adopt Behler-Parrinello-type neural network potential (NNP) as an MLP model [45] and consider three types of training sets: i) snapshots of crystals with randomly displaced atoms [42], ii) AIMD trajectory [26–30,33–39], and iii) snapshots of crystals with atoms displaced along phonon eigenmodes [40]. Our main goal is to develop a recipe for building a training set that predicts the room-temperature (300 K) $\kappa_i$ of general bulk materials with high efficiency and reasonable accuracy. The following is how the remaining sections are built: the details of computational methods are described in Section 2. In Section 3.1, we conduct a preliminary test on the methods of constructing training sets. In Section 3.2, based on the method chosen in the previous section, we calculate $\kappa_i$ of 25 materials with diverse cell symmetry and a wide range of $k_i$ values and analyze relative errors and computational efficiencies. We also check the effect of reducing the size of the training set. Finally, we summarize and conclude in Section 4.

2. Methods

2.1. Density functional theory calculations

All the DFT calculations in the present work are performed using Vienna Ab initio Simulation Package (VASP) [48–51]. The generalized gradient approximation by Perdew–Burke–Ernzerhof (PBE) [52] is used for the exchange-correlation functional. The initial structure for each material is obtained from the inorganic crystal structure database (ICSD) [53], which is relaxed further within DFT. The plane-wave cutoff energy and k-points grids for the unit-cell optimization are selected such that the energy and atomic forces converge to within 1 meV/atom and 5 meV/Å, respectively. The selected parameters are summarized in the Supplementary Information. In addition, the PREC tag is set to “Accurate”, and the convergence criteria for the self-consistent cycle is set to $10^{-8}$ eV. Then, the unit cells are fully optimized including lattice vectors until remaining atomic forces become smaller than 1 meV/Å. The final structure becomes the reference one in generating the training set. The computational parameters used for constructing the training data will be discussed in Section 3.1.
calculated with a finite displacement method using the PHONOPY package [56]. The side lengths of supercells used in the computation are ~20 Å each. The third-order interatomic force constants ($\Phi_{\alpha\beta\gamma}^{ijk}$ in Eq. (1)), also calculated with a finite displacement method, and lattice thermal conductivities at 300 K are evaluated by the ShengBTE package [9]. Here the supercell dimension is typically ~10 Å with ~6 Å for the cutoff radii of the interatomic interaction. The number of single-point force calculations increases with the number of possible atomic pairs in the supercell, considering the cutoff radii of the interatomic interaction. For the $q$-point sampling, a uniform mesh grid with a density of ~19 points Å$^{-1}$ is used. The Dirac delta functions in Eqs. (5) and (6) are approximated by the adaptive Gaussian with a proportionality factor of 0.1 [9,57]. Under the relaxation time approximation, the Boltzmann transport equation is not solved iteratively. Because the current work focuses on comparing the results of DFT and NNP, we do not consider nonanalytic corrections for LO-TO splitting. The full details on the computational parameters used to evaluate force constants can be found in the Supplementary Information.

2.3. Neural network potential

For training NNPs, we employ the SIMPLE-NN package [58]. Atom-centered symmetry functions $G^2$ and $G^4$ [59] are adopted to describe radial and angular distributions of neighboring atoms, respectively, defined as follows.

Fig. 1. $\kappa_l$ of test materials calculated by density functional theory (DFT) and neural network potentials (NNPs) that are generated with various training sets. (a) Random displacements of atoms (RDA), (b) $ab$ initio molecular dynamics (AIMD), (c) superposition of phonon eigenmodes (SPE). (d) Comparison between DFT and experimental values. Multiple values for each symbol denote the diagonal components of $\kappa_l$ and gray lines denote the error with a factor of 2. The data for experimental $\kappa_l$ values are from Ref. [64] (BAs), Ref. [65] (CoSb$_3$), Ref. [66] ($\beta$-Ga$_2$O$_3$), Ref. [67] (GaP), Ref. [68] ($\alpha$-SiO$_2$), and Ref. [69] (Tl$_9$BiTe$_6$).
\[
G_i^2 = \sum_{j} e^{-\|\mathbf{R}_i - \mathbf{R}_j\|^2} f_j (\mathbf{R}_j),
\]
(8)

\[
G_i^4 = 2^{-3} \sum_{j \neq k} \left(1 + \lambda \cos \theta_{jk}\right)^2 e^{-\|\mathbf{R}_i - \mathbf{R}_j - \mathbf{R}_k\|^2} f_j (\mathbf{R}_j) f_k (\mathbf{R}_k) f_j (\mathbf{R}_j) f_k (\mathbf{R}_k),
\]
(9)

where \(i, j, \) and \(k\) are atomic indices, \(\mathbf{R}_j\) is the distance between atom \(i\) and \(j\), and \(\eta, \zeta, \) and \(\lambda\) are hyperparameters. \(f_j\) is the cutoff function such that atoms outside the cutoff radius \(R_c\) do not contribute to the atomic energy.

\[
f_j (\mathbf{R}_j) = \begin{cases} 
0.5 \left[ \cos \left( \frac{\eta \mathbf{R}_j}{R_c} \right) + 1 \right] & (\mathbf{R}_j \leq R_c) \\
0 & (\mathbf{R}_j > R_c). 
\end{cases}
\]
(10)

The cutoff radius is set to 6.5 Å, and 26, 70, and 132 symmetry functions are employed for unary, binary, and ternary systems, respectively. The selected hyperparameters [60] are summarized in the Supplementary Information. The network architecture comprises two hidden layers with 60 hidden nodes each and one output layer that provides atomic energy. The input vector is decorrelated by principal component analysis and then whitened to increase the learning speed [60]. The NNP is trained by minimizing the loss function (\(\Gamma\)) defined as follows.

\[
\Gamma = \frac{1}{M} \frac{\sum_{i=1}^{M} (E^\text{DFT(NNP)}_i - E^\text{DFT}}{N_i} \leq \frac{\text{MP}}{N_i} \leq \frac{\text{NPP}}{N_i} \leq \frac{\text{MP}}{N_i}}{2} + \frac{\mu_1}{2} \sum_{i=1}^{M} \sum_{j=1}^{N_i} \left( E^\text{DFT}_i - E^\text{DFT}_j \right)^2 + \frac{\mu_2}{6M} \sum_{i=1}^{M} \sum_{j=1}^{N_i} \left( S^\text{DFT}_i - S^\text{DFT}_j \right)^2,
\]
(11)

where \(M\) is the total number of structures in the training set, \(N_i\) is the number of atoms in the \(i\)th structure, and \(E^\text{DFT(NNP)}_i, E^\text{DFT(NNP)}_j, \) and \(S^\text{DFT(NNP)}_i, S^\text{DFT(NNP)}_j\) are its total energy, atomic force of the \(i\)th atom, and the \(k\)th component of the virial stress tensor, respectively. \(\mu_1\) and \(\mu_2\) are the parameters that scale the relative importance of atomic force and stress with respect to the total energy when minimizing the loss function. We set the parameters \(\mu_1\) and \(\mu_2\) as \(10^2\) and \(10^{-6}\), respectively. To avoid overfitting and obtain more regularized NNPs, we apply a dropout technique in which half of the nodes are randomly selected and fixed at each training iteration [61]. The learning rate starts at 0.01 and scales overfitting and obtain more regularized NNPs, we apply a dropout

3. Results

3.1. Construction of training sets

To select an optimal approach to generate training sets, preliminary tests are conducted on six materials: BaS, CoSb, β-Ga2O3, GaP, α-SiO2, and Tl2BiTe6 that span three orders of \(k_f\) from \(10^{-3}\) to \(10^{2}\) W m\(^{-1}\) K\(^{-1}\). We compare the three methods of constructing the training set: (i) random displacements of atoms (RDA), (ii) AIMD, and (iii) superposition of phonon eigenmodes (SPE). Each atom is displaced with random directions and amplitudes in the RDA method. The amplitudes are chosen from a normal distribution with a standard deviation of 0.1 Å. The AIMD method comprises two steps. First, the AIMD using a canonical ensemble is conducted over 1 ps with the time step of 1 fs, at temperatures of 50, 300, 500, and 700 K, following the choice of the simulation time and temperatures in Ref. [30]. (We also tested the AIMD method with NPT ensemble, but the resulting \(k_f\) showed no significant difference.) The simulation cell, which contains \(\sim 100\) atoms, is identical to the one used to calculate the third-order interatomic force constants for the most cases of the test materials (see Section 2.2, and Supplementary Information). During AIMD, the computational parameters are slightly loosened such that the self-consistency criteria is set to \(10^{-4}\) eV with the default plane-wave energy cutoff and the \(\Gamma\)-point sampling. Next, we sample the 4-ps AIMD trajectories in 10- or 80-fs intervals depending on the target size of the training set, and more accurate DFT calculations are performed on the sampled structures. This is required because the interatomic force constants are sensitive to computational precision, so the training data must maintain high precision. In these calculations, the plane-wave energy cutoff, k-point grids, and convergence criteria of self-consistent calculation are set to the same as those of the unit cell optimization (See Section 2.1). Finally, the SPE method requires the second-order interatomic force constants that determine the phonon eigenmodes. To roughly calculate the second-order interatomic force constants with minimal computational costs, we use the \(\Gamma\) point and a small supercell size adopted in calculating third-order interatomic force constants (see Section 2.2). Each atom is displaced along the superposed phonon modes with random amplitudes and phase factors [19,63].

In each of the three approaches described above, the training set comprises 400 structures. To calculate third-order interatomic force constants, interactions up to \(\sim 6\) Å or 6th, 7th, 19th, 5th, 11th, and 4th nearest neighbors are considered for BaS, CoSb, α-Ga2O3, GaP, α-SiO2, and Tl2BiTe6, respectively. For the phonon dispersion, every NNP trained by the three methods reproduces phonon dispersions for all six materials (see Fig. S1 in the Supplementary Information). Fig. 1 (a)–(c) present the comparisons of the diagonal components of the \(k_f\) tensor between NNP and DFT. We also compare the DFT results with experimental data in Fig. 1(d) to validate computational settings. It is seen that NNP-AIMD shows higher accuracy than other methods, with the error for all test materials being less than 30%. Conversely, NNP-RDA and NNP-SPE show error levels larger than a factor of 2 for GaP and β-Ga2O3, respectively. This means that RDA and SPE method may produce atomic displacements that have no bearing on third-order interatomic force constants. Since the RDA method does not make any assumptions about atomic correlations, it is more likely to produce unphysical atomic configurations with small interatomic distances and large repulsive atomic forces. This analysis is supported by the phonon scattering rates of GaP shown in Fig. 2, while NNP-AIMD shows good accuracy in predicting frequency-scattering rates relations, the other two NNPs show large discrepancies, especially NNP-RDA.

The computational cost of the SPE method varies greatly depending on the cell symmetry due to the part on obtaining phonon dispersion. The costs of the AIMD and RDA methods, however, are less affected by crystal symmetry. Consequently, we conclude that the AIMD method is the best for generating training data. In passing, we note that for BaS, the present DFT result of \(k_f\) (~1200) in Fig. 1 is at variance with the previous literature (>2000) [30,70]. This is because we did not solve the Boltzmann transport equation iteratively (i.e., relaxation time approximation), which is known to affect the prediction accuracy for materials.
with high $\kappa_l$ [9]. When the Boltzmann transport equation is solved iteratively, we find that $\kappa_l$ increases to $\sim 1900$.

### 3.2. Computation of $\kappa_l$ for diverse materials.

This section increases the benchmark cases to 25 materials with diverse cell symmetries and a wide range of $\kappa_l$ values. The training set for each NNP comprises 400 structures generated by the AIMD method. Fig. 3(a) compares $\kappa_l$ between NNP-AIMD and DFT. The root-mean-squared relative error (RMSRE) of test materials is 18.6%. Fig. 3(b) shows the results when the training data is reduced, as discussed below. For comparison, we gather the $\kappa_l$ values computed by MLP and DFT from the literature and plot them in Fig. 3(c) (For a fair comparison, only those obtained by solving the Boltzmann transport equation are presented). Except for a few materials, the error level in Fig. 3(c) is similar. Fig. 3(d) explicitly compares the error values for common materials in Fig. 3(a) and Fig. 3(c). It can be seen that the errors from the present approach are comparable to those of other references even though we use a consistent choice of the training set. However, BAs shows a much larger error than Ref. [38]. It is known that $\kappa_l$ of BAs is
mainly determined by the scattering of acoustic phonon modes. Ref. [38] employed NPT simulations and adopted DFT results for the harmonic part, which may have contributed to more accurate scatterings by acoustic phonons.

We note that materials with low symmetries often exhibit large errors in Fig. 3 (a). To be specific, we classify the test materials into four groups based on the crystal system: (1) cubic, (2) tetragonal and hexagonal, (3) orthorhombic, and (4) monoclinic and triclinic. For each material, we first select the component of $\kappa_l$ showing the largest error. For the selected components, we calculate the average error within each group. The results are 7.6%, 13.5%, 15.7%, and 24.2% for the group (1)-(4), respectively, which shows a trend of increasing errors with low crystal symmetries. The low symmetry may require longer AIMD simulations for sufficient sampling. To test this, we extend AIMD of $K_2Bi_8Se_{13}$, which has $P\overline{1}$ space group and thus has the largest error, up to 2 ps, increasing the training set by two folds. However, the resulting NNP for $K_2Bi_8Se_{13}$ produces almost identical $\kappa_l$. Therefore, the origin of the error and a systematic solution need further investigation in future.

The high-precision DFT calculations used to construct the training set account for about 80% of the computational costs in the present work. Consequently, it would be worthwhile to see if the training set could be shrunk further to save on computational costs. To this end, we investigate the effect of dataset size on $\kappa_l$ using $\alpha$-SiO$_2$ as an example (see Fig. 4(a)). First, we choose 12, 28, 52, 100, 200, and 400 structures from 108-atom AIMD data, which train six different NNPs. The simulation time and temperatures are the same as those in the previous section. In Fig. 4(b)-(d), we calculate phonon dispersions, scattering rates, and cumulative lattice thermal conductivities as a function of the phonon mean free path. Fig. 4(b) shows that harmonic properties are well described by every NNP, but a slight deviation of the transverse acoustic mode is observed along the $\Gamma$–$A$ line with the data size of 12. The scattering rates and cumulative $\kappa_l$ in Fig. 4(c) and 4(d), respectively, are well reproduced for data sizes exceeding 28. With data sizes of 12 or 28, significant deviations in $\kappa_l$ are observed mostly at mean free paths of 1–10 nm.

Following the above analysis, we reduce the training set to 50 structures for 18 materials selected from Fig. 3 (a) by sampling more sparsely over the same MD trajectories. The $\kappa_l$ values compared between the resulting NNPs and DFT are shown in Fig. 3 (b). The RMSRE of the $\kappa_l$ is 31%, increasing from 18% in Fig. 3(a), which is reasonable considering the drastic decrease in computational cost. Again, $K_2Bi_8Se_{13}$ shows the largest error as in Fig. 3(a), which can be understood by the low
symmetry in the crystal structure.

Additionally, we analyze the computational efficiency of the NNP-based calculation of $\kappa_l$ with respect to the DFT-based calculation. Fig. 5 shows how the computational cost changes as the material complexity increases. The number of structures required to calculate the third-order force constants, which encompasses the number of atomic elements, the number of atoms in the unit cell, and the crystal symmetry, is used to rank the material complexity. The computational cost is normalized with respect to that of Si computed using the full DFT approach (left scale) or the NNP trained with 50 structures. (right scale). Here, the computational cost of the DFT-based $\kappa_l$ calculation is the summation of those obtaining second- and third-order force constants. The majority of the computational costs in NNP-based calculations come from building the training set (AIMD and single-point calculations) and training NNPs, while the cost of calculating force constants using the finite displacement method is negligible. As the material complexity increases, the number of structures needed to generate the third-order force constants increases rapidly, which is confirmed in Fig. 5. For instance, the DFT computational cost for KCuS is more than 20 times higher than that of Si. Conversely, NNP-based $\kappa_l$ calculations show nearly constant computational costs regardless of the material complexity. This is because the cost of AIMD and high-precision DFT calculations are similar among the materials and the atomic forces can be calculated with almost no cost once the NNP has been trained, even for a very large supercell.

To check the applicability for the metallic materials, we additionally calculate the $\kappa_l$ of Al with the same procedure as in the above. The results calculated by DFT and NNP trained with 400 and 50 training sets are 10.2, 10.6, and 9.9 Wm$^{-1}$K$^{-1}$, respectively, and the corresponding errors are 3.9%, and 2.9%. In addition, the $\kappa_l$ result of a narrow-gap material InAs, which PBE functional misclassifies as a metal, is given with the error of 2.3% in Fig. 3(a). Both results of metallic, and narrow-gap materials show that the present method can be applied for a wide range of materials with reasonable errors. However, for the metallic materials, electron-phonon scatterings may have significant effects on the $\kappa_l$ [72], while our method only considers phonon-phonon scatterings. Lastly, special care is needed when calculating the $\kappa_l$ of materials with high anharmonicity or $\kappa_l$ at high temperatures, because we did not consider the phonon renormalization and higher-order phonon processes than third order.

4. Conclusion

In conclusion, we propose a standard protocol for building the training set of NNPs targeted for computing $\kappa_l$ efficiently without fine-tuning for each material. The protocol requires 1-ps AIMD simulations at various temperatures and accurate single-point calculations for 400 structures sampled along the MD trajectory. Testing over 25 materials with diverse symmetries and wide range of $\kappa_l$, it is confirmed that NNP-AIMD provides consistent accuracies comparable to reported values in the literature. The uniform cost across material types for the proposed method makes it especially efficient for complex materials whose $\kappa_l$ prediction would be costly in a full-DFT approach. Furthermore, the NNPs showed reasonable accuracy even when the training set was reduced to 50 structures. The current work will broaden the application scope of machine learning potentials by establishing a robust framework for accurately computing lattice thermal conductivity with machine learning potentials.

Data Availability

The entire data set of the lattice thermal conductivity calculated by DFT and NNP are available in Table S1 of the Supplementary Information.

CRediT authorship contribution statement

Jeong Min Choi: Data curation, Formal analysis, Writing – original draft. Kyeongpung Lee: Data curation, Formal analysis, Writing – original draft. Sangtae Kim: Investigation. Minseok Moon: Validation. Wonseok Jeong: Supervision, Writing – review & editing. Seungwu Han: Supervision, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported Korea Institute of Ceramic Engineering and Technology (KICET) (N0002599), and Creative Materials Discovery Program (RIAM) through the National Research Foundation of Korea (NRF) funded by Ministry of Science and ICT (2017M3D1A1040689). The computations were carried out at Korea Institute of Science and Technology Information (KISTI) National Supercomputing Center (KSC-2020-CRE-0296). Part of the work is performed under the auspices of the U.S. Department of Energy by Lawrence Livermore National Laboratory under Contract Number DE-AC52-07NA27344.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.commatsci.2022.111472.

References


