Materials Today Physics xxx (xxxx) 101345

Contents lists available at ScienceDirect

Materials Today Physics

journal homepage: www.elsevier.com

Accelerating thermal conductivity prediction through machine-learning: Two orders of magnitude reduction in phonon-phonon scattering rates calculation

Yagyank Srivastava, Ankit Jain *

Mechanical Engineering Department, IIT, Bombay, India

ABSTRACT

The calculation of phonon-phonon scattering rates is a computational bottleneck in first-principles based phonon thermal conductivity prediction of materials. Here, we report a machine learning approach for phonon scattering rates prediction which is capable of predicting thermal conductivity by employing only 5 % of the total computational cost. We test this approach on more than 200 diverse materials and found that when this approach is combined with that of Guo et al. [npj Computational Materials **9**, 95 (2023)] for phonon-phonon linewidth prediction, the cumulative speed-up is more than two orders of magnitude while the accuracy of thermal conductivity prediction is preserved to within 10 %. This drastic speedup is translated into computational time reduction for phonon scattering rates calculation from more than 60,000 cpu-hours to less than 500 cpu-hours for considered 230 materials.

The fast and accurate prediction/determination of lattice thermal conductivity (κ) of materials is crucial in the discovery of new materials for applications such as thermoelectrics, thermal barrier coating, and heat dissipation [1–4]. The conventional search of materials with desired κ is via an experimental trial-and-error approach. Recently, with advances in computational resources, it has become possible to predict the κ of materials via the Boltzmann transport equation (BTE) approach with input from ab-initio calculations [5–9]. Such calculations have been reported in the literature for simple and complex materials and an excellent agreement with experiments is obtained, where possible [6,7,10]. The computation of κ via the BTE approach requires interatomic force constants and phonon-phonon scattering rates and the latter is a computational bottleneck in κ prediction of many material systems [11]. For instance, the calculation of phonon scattering rates for clathrates requires evaluation of $\sim 10^9$ three-phonon scattering processes with a total computational time of ~ 1400 cpu-hours. This high computational cost of three-phonon scattering rates calculation severely hinders the high-throughput discovery of materials and requires new methodology developments. Guo et al. [12] made significant advances in this regard and proposed a neural network machine learning-based model for prediction of scattering linewidths by training the model on a small fraction of processes for which linewidths are actually computed. The authors showed that while the linewidths of individual processes obtained from the model are less accurate [coefficient of determination, $R^2 < 0.5$] the total scattering rate of a given phonon mode (obtained by summing over all processes) are accurate with $R^2 \geq 0.9$ and the associated prediction error for $\kappa < 10$ %; thus facilitating accurate prediction of κ, along with its spectral dependence, while having a computational speed-up of 4x (70x for four-phonon processes). In their recent work [13], the authors further expanded on this and suggested a Maximum Likelihood Estimator (MLE) model based on the average linewidth of training datasets to replace the machine learning model and provide similar accuracy with additional speedup.

In this work, we develop further on these approaches and report an additional 20x reduction in computational time for κ prediction while maintaining the accuracy of predicted κ to within 10 % of the actual values. Our approach is complimentary to that by Guo et al. [12] and when combined together, the accuracy of κ-prediction and its spectral dependence is preserved to within 10 % while the computational time for phonon scattering rates calculation is reduced drastically by more than two orders of magnitude.

The phonon contribution to κ of material in the α-direction (κ_α) is obtainable using the BTE along with the Fourier's law as [14–16]:

$$\kappa_\alpha = \sum_\lambda c_\lambda v_{\lambda,\alpha}^2 \tau_{\lambda,\alpha}, \tag{1}$$

where the summation is over all phonon modes, c_λ is the phonon specific heat, $v_{\lambda,\alpha}$ is the α component of phonon group velocity vector $\boldsymbol{v}_\lambda$, and $\tau_{\lambda,\alpha}$ is the phonon scattering time. The phonon heat capacity and group velocity are obtained from phonon dispersion obtained by diago-

* Corresponding author.
E-mail address: a_jain@iitb.ac.in (A. Jain).

https://doi.org/10.1016/j.mtphys.2024.101345
Received 15 December 2023; Received in revised form 18 January 2024; Accepted 25 January 2024
2542-5293/

nalizing the phonon dynamical matrix [15,17]. The phonon scattering rates are obtained by considering the three-phonon as [15,16]:

$$\frac{1}{\tau_\lambda} = \sum_{\lambda'}\sum_{\lambda''}\Gamma(\lambda, \lambda', \lambda'') \quad (2)$$

The details regarding the calculation of phonon linewidth for three-phonon scattering processes, $\Gamma(\lambda, \lambda', \lambda'')$, are presented in Refs. [8,18].

The computation of κ via Eqn. (1) requires phonon properties for all phonon modes (typically 10, 000 – 50, 000) in the Brillouin zone. However, all phonon modes are not independent and these calculations are carried out only for symmetry-unrelated (reduced/unique) phonon modes. For instance, for computation of κ in silicon with 32 × 32 × 32 phonon wavevector grid and two atoms in the primitive unitcell, only 5382 phonon modes (2.75 %) are symmetry-unrelated of the total of 196608 modes requiring 25 cpu-hours for phonon scattering rates computations, while for type-I filled clathrates with 8 × 8 × 8 phonon wavevector grid and 54 atoms in the unitcell, 14580 modes (17.5 %) are symmetry-unrelated of the total of 82944 modes requiring 1370 cpu-hours for phonon scattering rates computations (by considering only three-phonon scattering).

Our machine learning (ML) approach is driven by realizations that (a) some phonon properties, such as vibration frequency, heat capacity, mode localization, Grüneissen parameter, etc. are relatively cheaper (by 100x-10000x) to compute (referred to, hereafter, as fast-to-compute properties) than the phonon scattering rates and (b) phonon scattering rates computations are needed for ~ 5000-15,000 phonon modes for typical materials even after symmetry reductions. Accordingly, we propose to evaluate fast-to-compute phonon properties for all phonon modes and scattering rates only for a fraction of symmetry-unrelated phonon modes. Subsequently, these fast-to-compute phonon properties can be employed as the phonon mode descriptor to train ML model on computed phonon scattering rates. Finally, the trained model can be used to predict scattering rates for the remaining symmetry-unrelated modes.

We test our proposed approach for silicon and type-I filled clathrate ($Ba_8Ga_{16}Ge_{30}$) in Fig. 1. For this, we employ a simple random forest ML model with 100 trees [19]. We first evaluate fast-to-compute phonon properties for all phonon modes and sort symmetry-unrelated phonon modes on the basis of their vibration frequencies. We sample the desired fraction of modes uniformly from this sorted list and train the ML model by employing phonon vibration frequency, heat capacity, mode localization, and Grüneissen parameter as the mode descriptor. We tested training the model on $\log(\tau)$, $\tau^{0.4}$, and $\tau^{0.01}$ (by employing the same distribution of phonon modes) and found $\tau^{0.01}$ to result in the best performance. The phonon scattering rates/lifetimes obtained by employing 5 % and 25 % of the symmetry-unrelated phonon modes in the training data are reported in Fig. 1. We find that while $R^2[\log(\tau)]$ (for prediction) is lower than 0.95 when 5 % of symmetry-unrelated phonons are used in the training data, the $R^2[\log(\tau)]$ increases to more than 0.95 with the inclusion of 25 % of symmetry-unrelated phonons in the training dataset for both materials; thus, suggesting a 4x saving in computational cost for phonon scattering rates calculations with the use of ML model.

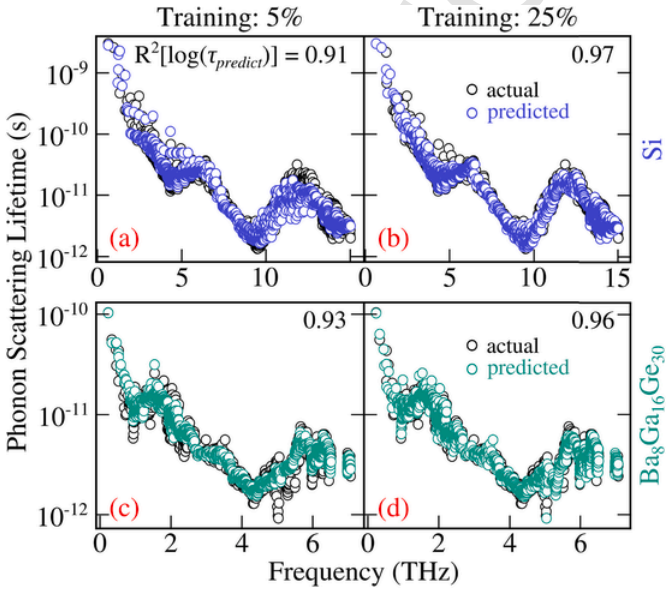

Fig. 1. The predicted phonon scattering lifetimes for (a), (b) silicon and (c), (d) $Ba_8Ga_{16}Ge_{30}$ using (a), (c) 5 % and (b), (d) 25 % of the symmetry-unrelated phonons in the ML training dataset. The predictions are made by training a random forest ML model with 100 trees. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Motivated by this, we next test the performance of ML model for κ-prediction by training the model on varying fractions of symmetry-unrelated phonon modes. After obtaining $\tau_{predict}$ on all symmetry-unrelated modes, we map them to the full list of phonon modes to obtain the $\kappa_{predict}$ of the underlying material. The obtained $\kappa_{predict}$ and $R^2[\log(\tau_{predict})]$ with varying amount of training data are reported in Fig. 2 for silicon and $Ba_8Ga_{16}Ge_{30}$.

We find that the obtained $\kappa_{predict}$ are within 10 % of κ_{actual} even with inclusion of only 5 % of the total symmetry-unrelated phonon modes in the ML training for both silicon and $Ba_8Ga_{16}Ge_{30}$. With an increase in the fraction of training data, the prediction error decreases and falls below 5 % with the inclusion of 25 % of phonon modes in the training of the ML model.

It is worthwhile to note that the κ obtained from the BTE approach typically have uncertainties of around 10 % owing to choices of various numerical parameters (phonon wavevector grid, broadening, cutoffs, etc) [20]. As such, the accelerated prediction of κ with 10 % accuracy with the inclusion of only 5 % of phonon modes in the training data is extremely promising and suggests 20x saving in the computational cost for phonon scattering rates. To test if such speed-up in κ prediction is obtainable for a wider variety of materials, we report $\kappa_{predict}$ of 230 different ternary materials in Fig. 3. The κ of these considered materials span around three orders of magnitude and was obtained using the high-throughput density functional theory calculations as detailed in Ref. [21].

We find that, as shown in Fig. 3, our ML model performance is amazingly good across all considered materials: as with silicon and $Ba_8Ga_{16}Ge_{30}$, the mean absolute percent error (MAPE) obtained across all materials is 10 % with inclusion of 5 % of the symmetry-unrelated phonons in ML training and falls below 5 % with inclusion of 25 % of data in the training; thus confirming a 4-20x speed-up in κ prediction across a wide range of materials. To pinpoint the beneficial role of the ML model, we also computed κ of considered materials by employing a coarser phonon wavevector grid consisting of a similar number of phonon modes as that in the training dataset of the ML model. We find that the κ obtained using these coarse grids have MAPE of 19 % with 5 % phonon modes compared to only 10 % from the ML model; thus suggesting that the improved performance from the ML approach is owing to learning of underlying phonon physics by the ML model.

In the ML approach recently reported by Guo et al. [12], the authors trained a ML model on phonon linewidths and found that even though phonon-phonon scattering linewidths are less accurate [$R^2(\Gamma(\lambda, \lambda', \lambda'')) < 0.5$], the total scattering rates are accurate with $R^2(\tau_\lambda) \geq 0.9$. The approach employed by Guo et al. [12] (ML-Γ) is fundamentally different from our approach (ML-τ), and the two approaches are complementary to each other and could be combined [ML-(τ, Γ)] to obtain cumulative computational speed-ups. To demonstrate this, we follow the following procedure [see Fig. 4(a)]:

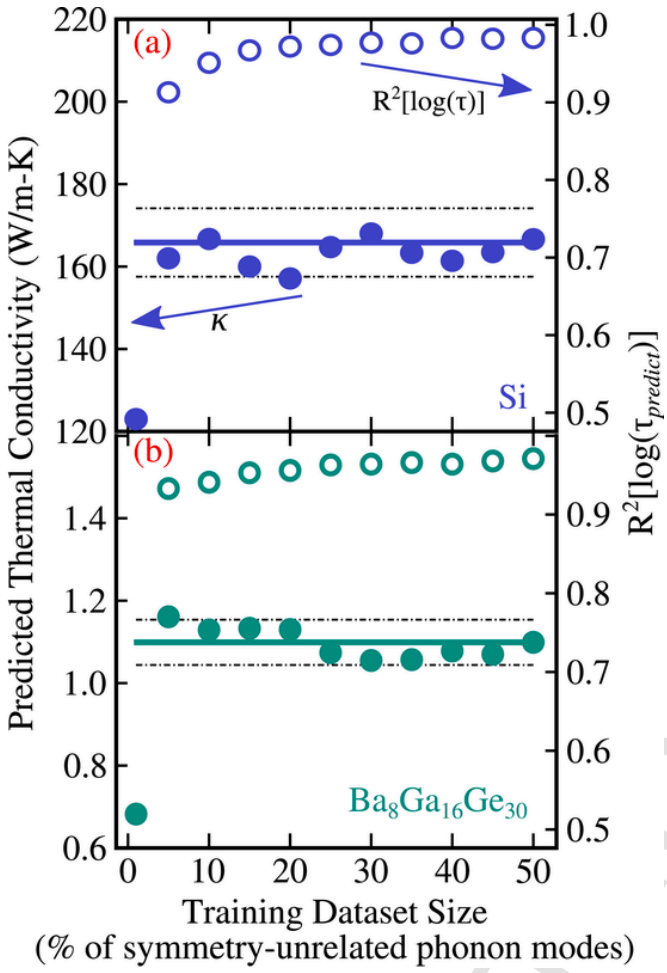

Fig. 2. The variation of ML predicted κ (filled markers) of (a) silicon and (b) $Ba_8Ga_{16}Ge_{30}$ with training dataset size. The corresponding prediction errors on phonon scattering rates ($R^2[\log(\tau_{predict})]$) are reported on the secondary y-axis (open markers). The actual κ values and 5 % variations around actual values are indicated using solid and dashed lines. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

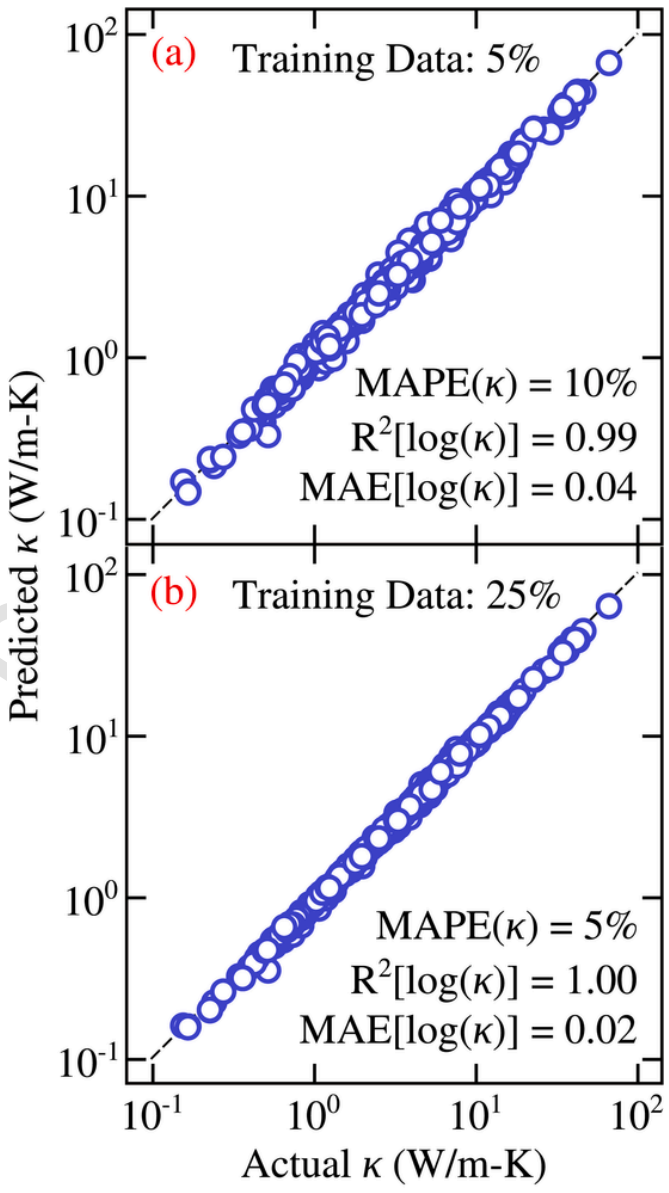

Fig. 3. The predicted κ of 230 diverse ternary materials using (a) 5 % and (b) 25 % of the symmetry-unrelated phonons in the ML training dataset. The prediction mean absolute percentage error (MAPE) on predicted κ is 10 % with the use of only 5 % phonon modes in the training dataset; indicating a computation speed-up of 20x in phonon scattering rate calculations with the ML approach.

1. Randomly select f_τ fraction of total phonons for which $\tau_\lambda^{pseudo\text{-}actual}$ is to be actually evaluated to train an ML-τ model.
2. Randomly select f_Γ fraction of three-phonon phonon scattering processes for each of the selected phonons from step (i) and perform $\Gamma^{actual}(\lambda, \lambda', \lambda'')$ calculations for these selected scattering processes.
3. Use evaluated $\Gamma^{actual}(\lambda, \lambda', \lambda'')$ from step-(ii) to predict/estimate $\Gamma^{predict}(\lambda, \lambda', \lambda'')$ of remaining $1 - f_\Gamma$ scattering processes by either training a ML model or using MLE estimator (ML-Γ).
4. Use $\Gamma^{predict}(\lambda, \lambda', \lambda'')$ to obtain $\tau_\lambda^{pseudo\text{-}actual}$ [using Eqn. (2)] of selected phonon modes from step (i).
5. Train ML model (ML-τ) on $\tau_\lambda^{pseudo\text{-}actual}$ to predict $\tau_\lambda^{predict}$ of remaining $1 - f_\tau$ phonon modes.

The $\kappa_{predict}$ obtained from this combined approach using f_Γ of 0.1 and f_τ of 0.05 are reported in Fig. 4(b). For $\Gamma^{predict}(\lambda, \lambda', \lambda'')$, we employed the MLE estimator based on the average linewidth of 10 % of the total three-phonon scattering processes of a given phonon mode.

Noticeably, we find that the κ prediction accuracy is maintained with this combined approach and the obtained MAPE on the considered diverse dataset of 230 materials is 10 % by employing only $f_\Gamma f_\tau = 0.005$ fraction of the total three-phonon scattering calculations, thus, resulting in a computational speed-up of 200x as shown in Fig. 4(c), thereby allowing for high-throughput discovery of materials. We note that the employed random forest ML model adds only a few seconds to computational cost compared to several orders of magnitude reduction in cpu-hours.

It is worth emphasizing that while data-driven approaches, focusing on the direct end-to-end prediction of κ from atomic structures, are currently being explored for accelerated discovery of materials [22,23], the accuracy of such approaches is limited to 55 % due to the non-availability of a large amount of high-quality κ data [21]. In the absence of such data, the κ calculation is currently possible through the BTE approach requiring interatomic force constants and phonon-phonon scattering rates. This BTE-based approach is, however, computationally expensive and requires 10-10^4 cpu-hours for each material and calls for method developments to accelerate such calculations.

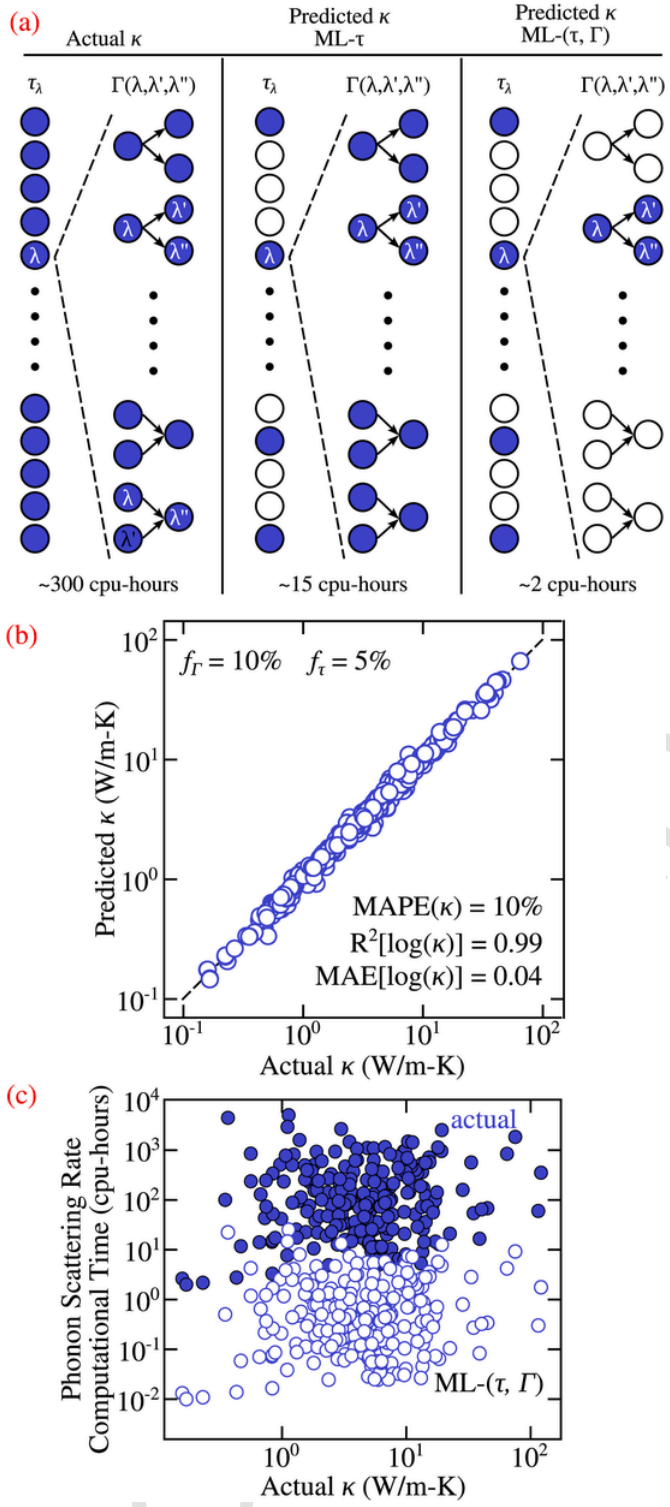

Fig. 4. (a) The calculation of phonon scattering rates from actual (left), ML-τ (middle), and ML-(τ, Γ) (right) approaches. The filled and open circles represent actual calculations and ML-based predictions. (b) The predicted κ of 230 ternary materials using the combined ML-(τ, Γ) approach. (c) The computational time required for phonon scattering rates calculation with actual (filled circles) and ML-(τ, Γ) (open circles) approaches. The computational times reported in (a) are obtained by averaging over 230 ternary materials.

Various research efforts are underway to train ML forcefield on first-principles data to accelerate the calculation of interatomic interactions [24]. While these efforts are not particularly focused on κ prediction, the developed forcefields are generic and are directly employable for interatomic force constant evaluation without requiring any additional effort. As such, with the possibility of availability of such high accuracy forcefields, the phonon scattering rates evaluation will be the remaining computational bottleneck for high-throughput BTE-based κ calculations. In this work, we have demonstrated that this computational bottleneck in the phonon scattering rate calculations can be overcome with ML-based approaches without compromising prediction accuracy. When employed in conjunction with suitably accurate ML forcefields, our approach will allow for the computation of κ of thousands of materials with first-principles accuracy.

In summary, we presented a machine-learning approach for accelerating phonon scattering rate calculations for the thermal conductivity prediction of materials. We tested our approach on more than 200 diverse materials and showed that the computational cost of phonon scattering rate calculations can be reduced by more than two orders of magnitude while preserving the thermal conductivity prediction accuracy to 10 %. Our approach alleviates the computational bottleneck in phonon scattering rate calculations and will pave the way forward for high-throughput BTE-based computational discovery of materials.

CRediT authorship contribution statement

Yagyank Srivastava: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Ankit Jain:** Writing – review & editing, Writing – original draft, Supervision, Software, Resources, Project administration, Investigation, Funding acquisition, Conceptualization.

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.

Data availability

Data will be made available on request.

Acknowledgement

The authors acknowledge the financial support from the Core Research Grant, Science & Engineering Research Board, India (Grant Number: CRG/2021/000010) and Nano Mission, Government of India (Grant No.: DST/NM/NS/2020/340). The calculations are carried out on the SpaceTime-II supercomputing facility of IIT Bombay.

References

[1] G.A. Slack, in: D.M. Rowe (Ed.), In CRC Handbook of Thermoelectrics, CRC, Boca Raton, 1995, pp. 407–440.
[2] D.R. Clarke, Surf. Coating. Technol. 163 (2003) 67.
[3] A.J. Minnich, M.S. Dresselhaus, F. Ren, G. Chen, Energy Environ. Sci. 2 (2009) 466.
[4] L. Lindsay, C.A. Polanco, Handbook of Materials Modeling, 2018, pp. 1–31.
[5] K. Esfarjani, H.T. Stokes, Phys. Rev. B 77 (2008) 144112.
[6] K. Esfarjani, G. Chen, H.T. Stokes, Phys. Rev. B 84 (2011) 085204.
[7] L. Lindsay, C. Hua, X. Ruan, S. Lee, Mater. Today Phys. 7 (2018) 106.
[8] A.J. McGaughey, A. Jain, H.Y. Kim, B. Fu, J. Appl. Phys. 125 (2019) 11101.
[9] H.P. Veeravenkata, A. Jain, Carbon 183 (2021) 893.
[10] L. Lindsay, D.A. Broido, T.L. Reinecke, Phys. Rev. B 87 (2013) 165201.
[11] T. Feng, X. Ruan, Phys. Rev. B 93 (2016) 045202.
[12] Z. Guo, P. Roy Chowdhury, Z. Han, Y. Sun, D. Feng, G. Lin, X. Ruan, npj Comput. Mater. 9 (2023) 95.
[13] Z. Guo, Z. Han, D. Feng, G. Lin, X. Ruan, Sampling-accelerated First-Principles Prediction of Phonon Scattering Rates for Converged Thermal Conductivity and

Radiative Properties, 2023 arXiv:2311.12935 [cond-mat.mtrl-sci].
[14] J.M. Ziman, Electrons and Phonons, Oxford University Press, Clarendon, Oxford, 1960.
[15] J.A. Reissland, The Physics of Phonons, John Wiley and Sons Ltd, 1973.
[16] G.P. Srivastava, The Physics of Phonons, Adam Hilger, Bristol, 1990.
[17] M.T. Dove, Introduction to Lattice Dynamics, 1993 Cambridge, Cambridge.
[18] A. Jain, Phys. Rev. B 102 (2020), https://doi.org/10.1103/physrevb.102.201201.
[19] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, J. Mach. Learn. Res. 12 (2011) 2825.
[20] A. Jain, A.J. McGaughey, Comput. Mater. Sci. 110 (2015) 115.
[21] Y. Srivastava, A. Jain, J. Appl. Phys. 134 (2023).
[22] T. Zhu, R. He, S. Gong, T. Xie, P. Gorai, K. Nielsch, J.C. Grossman, Energy Environ. Sci. 14 (2021) 3559.
[23] Z. Liu, M. Jiang, T. Luo, Mater. Today Phys. 28 (2022) 100868.
[24] A. Rodriguez, C. Lin, C. Shen, K. Yuan, M. Al-Fahdi, X. Zhang, H. Zhang, M. Hu, Commun. Mater. 4 (2023) 61.