Applied Physics Letters

Predicting lattice thermal conductivity of semiconductors from atomic-information-enhanced CGCNN combined with transfer learning

Cite as: Appl. Phys. Lett. **122**, 152106 (2023); https://doi.org/10.1063/5.0142150
Submitted: 12 January 2023 • Accepted: 31 March 2023 • Published Online: 12 April 2023

Zeyu Wang, Jinlong Ma, Run Hu, et al.

AIP Publishing

Applied Physics Letters ARTICLE scitation.org/journal/apl

Predicting lattice thermal conductivity of semiconductors from atomic-information-enhanced CGCNN combined with transfer learning

Cite as: Appl. Phys. Lett. **122**, 152106 (2023); doi: 10.1063/5.0142150
Submitted: 12 January 2023 · Accepted: 31 March 2023 ·
Published Online: 12 April 2023

Zeyu Wang, Jinlong Ma,[a] Run Hu, and Xiaobing Luo

AFFILIATIONS

School of Energy and Power Engineering, Huazhong University of Science and Technology, Wuhan 430074, China

[a]Author to whom correspondence should be addressed: majinlong@hust.edu.cn

ABSTRACT

Rapid identification of lattice thermal conductivity of semiconductors from their crystal structure is required in the discovery of functional materials. A promising strategy is using a machine learning method based on a first-principles dataset, which, however, suffers from the dilemma of too little data available. In this work, the crystal graph convolutional neural networks (CGCNN) model was improved by enhancing the information of atomic descriptors (for short CGCNN-D), and the transfer learning (TL) method was combined to overcome the problem of small datasets. It is found that the CGCNN-D has improved predicting performance for both electronic bandgap with large data volume and thermal conductivity with small data volume, with the mean absolute error reducing 7% and 10%, respectively, indicating the importance of the improved atomic description. Applying TL with electronic bandgap as a proxy into the CGCNN-D further upgrades the prediction accuracy for thermal conductivity that has only 95 pieces of data, yielding 19% decrease in the mean absolute error as compared to the original CGCNN. The trained CGCNN-D-TL model was used to quickly estimate the thermal conductivities of thousands of semiconductors, and the materials identified with potentially high thermal conductivity were further screened by the optimized Slack model. Finally, the most promising BC_2N was discovered and then confirmed by the first-principles calculations, which shows room-temperature thermal conductivities of 731, 594, and 500 W m^{-1} K^{-1} along the three principal axes of its lattice structure.

Semiconductors are the key components of electronic products and known as the cornerstone of the modern industry. The thermal transport property of semiconductors is one of the most crucial concerns that must be considered in devices due to its inherent nature. High thermal conductivity is expected for rapid heat dissipation to preserve the performance of electronic devices. Finding a material with desired functionalities while taking into account the thermal conductivity in a balanced manner is the basic criterion. Assisted by the computing methods, some materials are predicted to have potentially superior properties and are progressively synthesized, yielding a lot of exciting systems such as cubic BAs.[1–4] Among the state-of-the-art calculation methods for thermal conductivity, the first-principles calculation based on the density functional theory (DFT) is a parameter-free framework with predictive accuracy and has been well developed in recent decades.[5–7] Nonetheless, one disadvantage that needs to be acknowledged is its heavy computational burden. Up to now, the intrinsic thermal conductivities of only a few hundred systems are known in total of the tens of thousands semiconductors documented in the Materials Project database.[8] Therefore, an efficient method is desired for predicting the thermal conductivities of the large amount of semiconductors to provide the knowledge of their thermal transport potentials.

To accelerate the calculations of the thermal conductivities of materials while maintaining a certain degree of accuracy, the machine learning is used to combine with the first-principles calculations.[9–11] The applications mainly focus on training interatomic potentials for molecular dynamics simulations, predicting the properties of materials with the same structure but different elements, and designing the atomic architecture of nanostructured systems.[12–15] The process can be briefly introduced as that the first-principles calculations are performed on a small portion of the systems, and the obtained data are used for training of the machine learning model, which is then

employed to predict the properties of the rest. However, this strategy is strictly requiring for the dataset, which would not be suitable for screening of semiconductors with target thermal conductivities from thousands of candidates; therefore, more general machine learning strategies are needed.

Many machine learning models have demonstrated sound ability in predicting the properties of materials from their crystal structures such as formation energy[16–18] and electronic bandgap.[19–22] Among the various models, a neural network[23,24] shows great potential for its strong scalability. Some of its applications can construct the structure of materials and plainly mine the composition information in a deeper level. Considering the thermal conductivity is significantly affected by the crystal structure, the atomic component information and geometric structure information should be the basic inputs of the machine learning models. Recently, crystal graph convolutional neural networks (CGCNN)[17] have attracted great attention, because they provide a simple and general framework that directly learns material properties from the atoms and their connections in the crystal. However, as compared to the electronic properties, the investigation of modeling thermal conductivity is rare due to the lack of thermal conductivity data. The available study is found using extremely simplified models to approximately calculate the thermal conductivity, leading to biases in the data itself.[25] The data volume of thermal conductivity obtained in the first-principles level is much smaller than the electronic properties accounting for its much more time-consuming density functional perturbation (DFPT) calculations on dozens of Brillouin zone points or DFT calculations on hundreds of displaced supercells, which are at least several orders of magnitude heavier than the self-consistent field calculation for electronic properties. The small dataset issue is one of the major challenges in the machine learning prediction of thermal transport properties.[12] In addition, current improvement efforts of the CGCNN model emphasize a more complex description of the geometric structure, such as the iCGCNN and A-CGCNN models, but ignore the description of atomic features.[26,27]

Therefore, in this work, we improved the atomic descriptors of the CGCNN model to describe the crystal traits in more detail and call it as the CGCNN-D model in order to distinguish from the original model. We collected the thermal conductivity values calculated by the first-principles theory for 95 semiconductor single crystals. To use this small dataset for prediction, the transfer learning (TL) technique was adopted to overcome the problem of data dilemma, the main idea of which is to train the model for a proxy property on a large dataset and then transfer the learned knowledge to the machine learning model of the target property, which is limited in a much small dataset. The enhanced performance by TL has been verified in the materials' informatics studies.[18,28,29] The predicting performances of the CGCNN and CGCNN-D models were compared, and it is found that the CGCNN-D model has better performance than the CGCNN, whatever by using large samples or small samples datasets. In addition, the TL method using electronic bandgap (E_g) as proxy property is significantly benefited to the prediction of thermal conductivity (T_c). After obtaining the well-trained model for T_c, we used it to screen several potentially high T_c crystals from thousands of systems based on their ground state lattice structure. Considering the T_c is the result of lattice vibration, the optimized Slack model[30,31] was used for further identification. Finally, orthorhombic BC_2N was found to has high thermal conductivity and was confirmed by the first-principles calculations.

The machine learning prediction of T_c for crystalline materials needs a basic dataset, descriptors, and a machine learning model.[12–15] Herein, two sets of data are acquired. The first group is the room-temperature T_c of single-crystal semiconductors. A total of 95 pieces of data calculated by first principles are searched from the literature[1,5,9–11,32–51] as listed in Table S1 of the supplementary material. Significantly, the values are distributed over four orders of magnitude. The T_c of diamond with naturally occurring isotopes can be as high as 2290 W m^{-1} K^{-1};[1] on the contrary, it is as small as 0.368 W m^{-1} K^{-1} for Tl_3VSe_4.[32] Moreover, the distribution is extremely uneven, as shown in Fig. S1 of the supplementary material, with 80% of the data falling below 100 W m^{-1} K^{-1}, so the machine learning model may be prone to low T_c materials. Therefore, following the usual course,[25] the values of T_c are taken the logarithm to make the distribution of data more suitable for the regression algorithms. The other group of data aims to the crystal structure of material and its corresponding E_g. The former is used to identify the semiconductors throughout the machine learning process, and the latter is used as proxy property when the TL method is employed. The data of the crystal structure and E_g are extracted from the materials project database[8] by traversing all binary and ternary compounds and filtering them by the formation energy smaller than 0.2 eV for stable and metastable structures. Then, the materials with E_g between 0.2 and 5 eV are picked out for semiconductors. By additionally considering the difficulty of theoretical calculation and practical synthesization for complex crystals, the systems with a number of atoms in the primitive cell less than 10 are selected. Finally, there are 4160 pieces of data as the source domain. The selection process is shown in Fig. S2 of the supplementary material. The E_g is chosen as the proxy property for two reasons: One is the data accessibility, because E_g is a quantity available for all these semiconductors provided by the database. The other is the possible relation between E_g and T_c. In DFT, nearly all physical properties are related to the energy of the ground state or the energy change under perturbation. The E_g of electronic states is determined by the charge density of the ground state, while the T_c is related to phonon states that are determined by the ground state and the response to the perturbation of atomic displacement. It has been reported that the E_g acting as proxy property for deep neural networks improved the performance in predicting the phonon related group velocity and heat capacity.[52]

The CGCNN model regards the crystal structure as a crystal graph with nodes representing atoms and edges representing bonding connections and encodes the atomic information and the distance between atoms into crystal descriptors through a convolutional neural network.[17] The information for describing atom and bond feature vectors consists of binary digits. Suitable descriptors may not only improve the accuracy of prediction but also make the model to have better robustness. In the CGCNN model, discrete features in one-hot encoding are used to describe the element feature, which then represents each atom in the crystal by nine features. To have more sufficient description, in the CGCNN-D model, we ameliorate the descriptors of elements in consecutive value, and each atom vector is expressed with 58 features for all elements obtained from the XenonPy program.[28] The details of the implementation for atomic information enhancement as well as the following TL are provided in Sec. 2 of the supplementary material.

To assess the performance of CGCNN-D compared with CGCNN, these two models were trained against the E_g and T_c datasets.

The fivefold cross-validations were performed with mean absolute error (MAE) and coefficient of determination (R^2) as the metrics to evaluate the model performance. Figure 1 illustrates the MAE and R^2 of eight independent cases. The values do not change much in multiple cases, indicating strong robustness and no overfitting problem for both CGCNN and CGCNN-D models. The histograms clearly show that the CGCNN-D performs better than CGCNN whatever using the large E_g samples or small T_c samples. For E_g, the CGCNN-D obtains 7% decrease in the average MAE as compared to CGCNN, from 0.480 to 0.447 eV, and the R^2 also has an increase from 0.686 to 0.703. Training with the much smaller dataset of T_c, higher prediction accuracy is also gained by CGCNN-D. The average MAE reduces 10% from 0.228 to 0.207 W m^{-1} K^{-1}, and the average R^2 increases from 0.874 to 0.904. It can be seen that the enhancement in atomic information indeed yields performance gains for both large and small scales of datasets. This is because the message of crystal is based on the atomic features; therefore, when enhancing the description of the element, the information can pass in more detail.

To build the TL model, the CGCNN and CGCNN-D models are pretrained on the 4160 data of E_g. Then the hyperparameters from the embedding and convolutional layers are transferred to the target model as the initial parameters and retrained using the fine-tuning approach. Figure 2 shows the MAE and R^2 of CGCNN and CGCNN-D models with and without TL in eight independent fivefold cross-validations for T_c. It is obvious that the TL decreases the MAE and increases the R^2 for all cases, indicating the considerable improvement brought by TL in the prediction accuracy. Therefore, it can be inferred that the TL can act on the CGCNN and CGCNN-D models, and it can extract some information from proxy properties and then make sense for the target property although the relation between the proxy property and target property is not direct. When the TL is combined to CGCNN, the average MAE changes from 0.228 to 0.207, which is about 9% reduction. When the improved atomic descriptor and TL are both employed, the average MAE further decreases to 0.185, about 19% reduction as compared to the original CGCNN. Meanwhile, the average R^2 of CGCNN and CGCNN-D-TL for T_c are 0.874 and 0.921, respectively, giving an improvement of 5% by the enhanced atomic information and the TL method. For intuitive display of predictive power, Fig. 3 shows the T_c predicted by the CGCNN-D-TL model against the first-principles calculated values. The MAE is as small as 0.183 W m^{-1} K^{-1}, and the R^2 is as high as 0.928, suggesting a fairly good prediction.

To ensure the reliability, 25 CGCNN-D-TL models were trained and used to predict the T_c of the remaining thousands of semiconductors and then the obtained T_c were averaged. The materials with estimated T_c larger than 500 W m^{-1} K^{-1} are listed in Table S4 of the supplementary material. It should be pointed out that the prediction of the CGCNN-D-TL model is based on the ground state crystal structure. Unlike electronic properties that are mainly determined by the

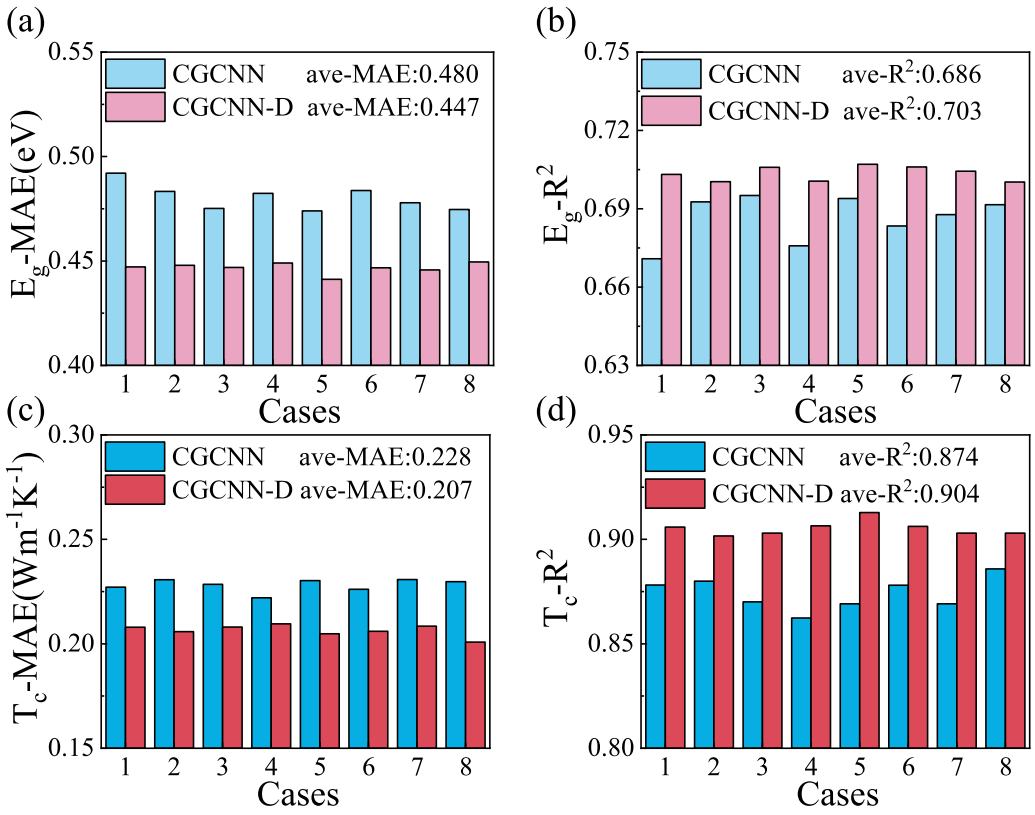

FIG. 1. MAE and R^2 of eight independent predictions for (a) and (b) electronic bandgap and (c) and (d) thermal conductivity using CGCNN and CGCNN-D models, respectively.

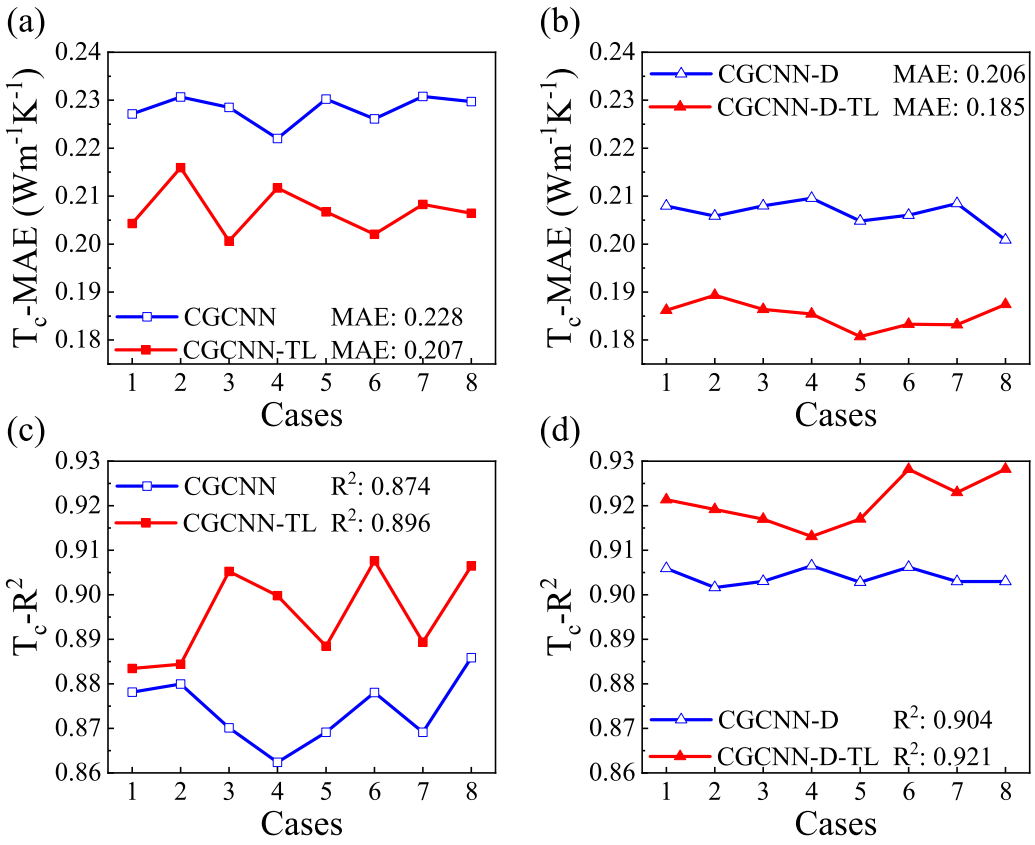

FIG. 2. (a) and (b) MAE and (c) and (d) R^2 of eight independent calculations for thermal conductivity using CGCNN and CGCNN-D models with (red lines) and without (blue lines) transfer learning, respectively.

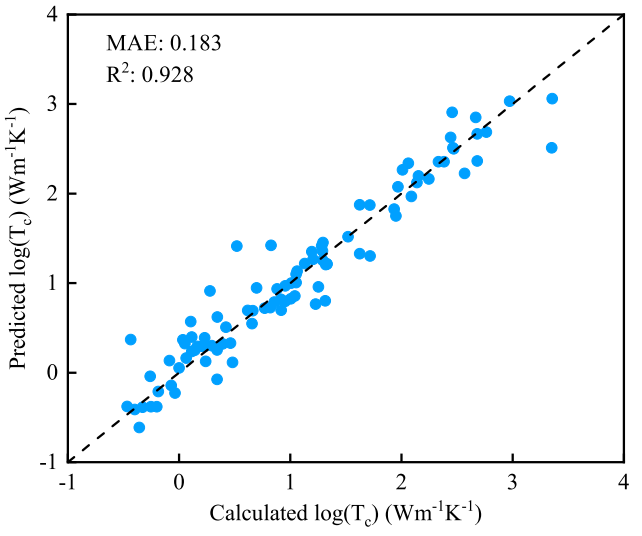

FIG. 3. Predicted thermal conductivities by the CGCNN-D-TL model vs the first-principles calculated values.

ground state, the thermal conductivity additionally involves lattice vibration. Therefore, the optimized Slack model[30,31] was used for further screening. The details were provided in Sec. 3 of the supplementary material. As expected, some of the high T_c materials predicted by the CGCNN-D-TL model actually have low T_c, but there are still some materials that do have high T_c potential. Among these materials, orthorhombic BC_2N with a space group of *Pmm*2 is the most promising candidate. The anisotropy of T_c is not considered in the CGCNN-D-TL model and the Slack model, and thus, the corresponding predicted T_c are 1130 and 483 W m^{-1} K^{-1} at room temperature, respectively. The first-principles calculations were used to examine the T_c of BC_2N, which is an anisotropic system with lattice constants of 2.538, 2.566, and 3.645 Å along the three principal axes as shown in the inset of Fig. 4. The calculations were performed using the Vienna *Ab initio* simulation package (VASP) with Perdew–Burke–Ernzerhof (PBE) pseudopotentials. The second and third order interatomic force constants (IFCs) were calculated using $6 \times 6 \times 4$ supercell, and a truncation of ninth nearest neighbors was adopted for third order IFCs. The ShengBTE package[6] was used to calculate T_c by iteratively solving the phonon Boltzmann transport equation (BTE) with $30 \times 30 \times 20$ **q** grids. The related convergence test was provided in Sec. 4 of the

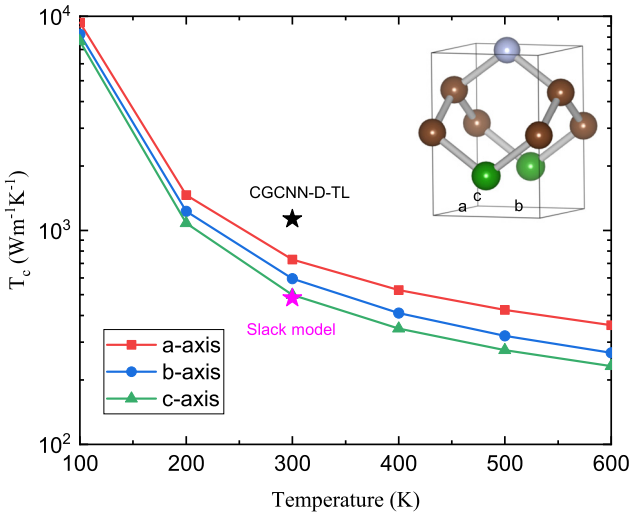

FIG. 4. Thermal conductivity of BC_2N along three principal axes calculated by first-principles, where the black and pink points are values predicted by the CGCNN-D-TL model and the optimized Slack model, respectively. The inset shows the primitive cell of BC_2N.

supplementary material. At room temperature, the first-principles calculated T_c of BC_2N are 731, 594, and 500 W m^{-1} K^{-1} along a, b, and c axes, respectively. It can be seen that BC_2N indeed has high T_c, indicating the feasibility of quick screening of materials with high T_c from huge dataset by CGCNN-D-TL combined with the Slack model.

In conclusion, this work provided a quick screening scheme for the thermal conductivities of semiconductors using improved CGCNN combined with the TL method and the Slack model. The atomic descriptor of CGCNN was modified with more precise element descriptions, increasing from nine features to 58 features. This improvement brings considerable boost in the prediction performance of semiconductors both for the large dataset of electronic bandgap and for the extremely small dataset of thermal conductivity. Moreover, the TL method was applied to the atomic-information-enhanced CGCNN in the prediction of thermal conductivity, which yields a further increase in the prediction accuracy. This suggests that the TL method is effective in solving the dilemma of too little data that generally exists in machine learning for predicting physical properties of materials. The improved CGCNN combined with TL has robust prediction performance with the coefficient of determination of predicted thermal conductivities against first-principles calculated values as high as 0.928. The well-trained models are then used to predict the thermal conductivities of thousands of semiconductors based on their crystal structures. The semiconductors estimated to have high thermal conductivities were further screened by the optimized Slack model, giving the most promising candidate of BC_2N. The potential of BC_2N was validated by first-principles calculations, showing a thermal conductivity as high as 731 W m^{-1} K^{-1} at room temperature.

See the supplementary material for the data description of thermal conductivity and electronic bandgap as well as the crystal structure of semiconductors, the implementation of atomic information enhancement and transfer learning on CGCNN, the optimized Slack model and estimated thermal conductivities of potential materials, and the convergence check for the first-principles calculations of BC_2N.

The authors acknowledge support from the Hubei Provincial Natural Science Foundation of China (Grant No. 2022CFB104).

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Zeyu Wang: Investigation (equal); Methodology (equal); Writing – original draft (equal). **Jinlong Ma:** Conceptualization (equal); Investigation (equal); Supervision (equal); Writing – review & editing (equal). **Run Hu:** Conceptualization (equal). **Xiaobing Luo:** Resources (equal); Supervision (equal).

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

[1]L. Lindsay, D. A. Broido, and T. L. Reinecke, Phys. Rev. Lett. **111**, 025901 (2013).
[2]J. S. Kang, M. Li, H. Wu, H. Nguyen, and Y. Hu, Science **361**, 575 (2018).
[3]S. Li, Q. Zheng, Y. Lv, X. Liu, X. Wang, P. Y. Huang, D. G. Cahill, and B. Lv, Science **361**, 579 (2018).
[4]F. Tian, B. Song, X. Chen, N. K. Ravichandran, Y. Lv, K. Chen, S. Sullivan, J. Kim, Y. Zhou, T.-H. Liu, M. Goni, Z. Ding, J. Sun, G. A. G. U. Gamage, H. Sun, H. Ziyaee, S. Huyan, L. Deng, J. Zhou, A. J. Schmidt, S. Chen, C.-W. Chu, P. Y. Huang, D. Broido, L. Shi, G. Chen, and Z. Ren, Science **361**, 582 (2018).
[5]D. A. Broido, M. Malorny, G. Birner, N. Mingo, and D. A. Stewart, Appl. Phys. Lett. **91**, 231922 (2007).
[6]W. Li, J. Carrete, N. A. Katcho, and N. Mingo, Comput. Phys. Commun. **185**, 1747 (2014).
[7]Z. Han, X. Yang, W. Li, T. Feng, and X. Ruan, Comput. Phys. Commun. **270**, 108179 (2022).
[8]A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, and K. A. Persson, APL Mater. **1**, 011002 (2013).
[9]J. Carrete, W. Li, N. Mingo, S. Wang, and S. Curtarolo, Phys. Rev. X **4**, 011019 (2014).
[10]A. van Roekeghem, J. Carrete, C. Oses, S. Curtarolo, and N. Mingo, Phys. Rev. X **6**, 041061 (2016).
[11]Y. Xia, V. I. Hegde, K. Pal, X. Hua, D. Gaines, S. Patel, J. He, M. Aykol, and C. Wolverton, Phys. Rev. X **10**, 041029 (2020).
[12]H. Wei, H. Bao, and X. Ruan, Energy AI **8**, 100153 (2022).
[13]S. Ju and J. Shiomi, Nanosc. Microsc. Therm. **23**, 157 (2019).
[14]X. Wan, W. Feng, Y. Wang, H. Wang, X. Zhang, C. Deng, and N. Yang, Nano Lett. **19**, 3387 (2019).
[15]X. Qian and R. Yang, Mat. Sci. Eng. R **146**, 100642 (2021).
[16]W. Ye, C. Chen, Z. Wang, I.-H. Chu, and S. P. Ong, Nat. Commun. **9**, 3800 (2018).
[17]T. Xie and J. C. Grossman, Phys. Rev. Lett. **120**, 145301 (2018).
[18]D. Jha, K. Choudhary, F. Tavazza, W.-K. Liao, A. Choudhary, C. Campbell, and A. Agrawal, Nat. Commun. **10**, 5316 (2019).
[19]J. Lee, A. Seko, K. Shitara, K. Nakayama, and I. Tanaka, Phys. Rev. B **93**, 115104 (2016).
[20]G. Pilania, J. Gubernatis, and T. Lookman, Comput. Mater. Sci. **129**, 156 (2017).

[21]Y. Zhuo, A. Mansouri Tehrani, and J. Brgoch, J. Phys. Chem. Lett. **9**, 1668 (2018).
[22]A. C. Rajan, A. Mishra, S. Satsangi, R. Vaish, H. Mizuseki, K.-R. Lee, and A. K. Singh, Chem. Mater. **30**, 4031 (2018).
[23]G. E. Hinton and R. R. Salakhutdinov, Science **313**, 504 (2006).
[24]Y. LeCun, Y. Bengio, and G. Hinton, Nature **521**, 436 (2015).
[25]X. Wang, S. Zeng, Z. Wang, and J. Ni, J. Phys. Chem. C **124**, 8488 (2020).
[26]C. W. Park and C. Wolverton, Phys. Rev. Mater. **4**, 063801 (2020).
[27]B. Wang, Q. Fan, and Y. Yue, J. Phys.: Condens. Matter **34**, 195901 (2022).
[28]H. Yamada, C. Liu, S. Wu, Y. Koyama, S. Ju, J. Shiomi, J. Morikawa, and R. Yoshida, ACS Cent. Sci. **5**, 1717 (2019).
[29]R. Ma, Y. J. Colon, and T. Luo, ACS Appl. Mater. Interfaces **12**, 34041 (2020).
[30]G. Qin, A. Huang, Y. Liu, H. Wang, Z. Qin, X. Jiang, J. Zhao, J. Hu, and M. Hu, Mater. Adv. **3**, 6826 (2022).
[31]T. Jia, G. Chen, and Y. Zhang, Phys. Rev. B **95**, 155206 (2017).
[32]S. Mukhopadhyay, D. S. Parker, B. C. Sales, A. A. Puretzky, M. A. McGuire, and L. Lindsay, Science **360**, 1455 (2018).
[33]W. Li, L. Lindsay, D. A. Broido, D. A. Stewart, and N. Mingo, Phys. Rev. B **86**, 174307 (2012).
[34]L. Lindsay, D. A. Broido, and T. L. Reinecke, Phys. Rev. Lett. **109**, 095901 (2012).
[35]Z. Tian, J. Garg, K. Esfarjani, T. Shiga, J. Shiomi, and G. Chen, Phys. Rev. B **85**, 184303 (2012).
[36]H. Dekura, T. Tsuchiya, and J. Tsuchiya, Phys. Rev. Lett. **110**, 025904 (2013).
[37]W. Li and N. Mingo, J. Appl. Phys. **114**, 183505 (2013).
[38]L. Lindsay, D. A. Broido, and T. L. Reinecke, Phys. Rev. B **87**, 165201 (2013).
[39]L. Lindsay, D. A. Broido, and T. L. Reinecke, Phys. Rev. B **88**, 144306 (2013).
[40]W. Li and N. Mingo, Phys. Rev. B **89**, 184304 (2014).
[41]W. Li and N. Mingo, Phys. Rev. B **90**, 094302 (2014).
[42]L. Feng, T. Shiga, and J. Shiomi, Appl. Phys. Express **8**, 071501 (2015).
[43]A. van Roekeghem, J. Carrete, and N. Mingo, Phys. Rev. B **94**, 020303 (2016).
[44]M. Zeeshan, T. Nautiyal, J. van den Brink, and H. C. Kandpal, Phys. Rev. Mater. **2**, 065407 (2018).
[45]I. Pallecchi, M. Pani, F. Ricci, S. Lemal, D. I. Bilc, P. Ghosez, C. Bernini, N. Ardoino, G. Lamura, and D. Marré, Phys. Rev. Mater. **2**, 075403 (2018).
[46]J. Park, Y. Xia, A. M. Ganose, A. Jain, and V. Ozoliņš, Phys. Rev. Appl. **14**, 024064 (2020).
[47]J. Ma, A. S. Nissimagoudar, S. Wang, and W. Li, Phys. Status Solidi-R **14**, 2000084 (2020).
[48]J. Ma, W. Li, and X. Luo, Appl. Phys. Lett. **105**, 082103 (2014).
[49]M. D. Santia, N. Tandon, and J. D. Albrecht, Appl. Phys. Lett. **107**, 041907 (2015).
[50]D. A. Broido, L. Lindsay, and A. Ward, Phys. Rev. B **86**, 115203 (2012).
[51]J. Dai and Z. Tian, Appl. Phys. Lett. **118**, 041901 (2021).
[52]Z. Liu, M. Jiang, and T. Luo, Sci. Adv. **6**, eabd1356 (2020).