Simulating lattice thermal conductivity in semiconducting materials using high-dimensional neural network potential

Emi Minamitani1,2,3*, Masayoshi Ogura2, and Satoshi Watanabe2,4 *

1Institute for Molecular Science, Japan
2Department of Materials Engineering, The University of Tokyo, Japan
3Japan Science and Technology Agency, PRESTO, Japan
4Center for Materials Research by Information Integration, Research and Services Division of Materials Data and Integrated System, National Institute for Materials Science, Japan

E-mail: eminamitani@ims.ac.jp

Received May 21, 2019; revised July 18, 2019; accepted July 30, 2019; published online August 12, 2019

We demonstrate that a high-dimensional neural network potential (HDNNP) can predict the lattice thermal conductivity of semiconducting materials with an accuracy that is comparable to that of density functional theory (DFT) calculation. After a training procedure based on force, the root mean square error between the forces predicted by HDNNP and DFT is less than 40 meV Å⁻¹. As typical examples, we present the results of Si and GaN bulk crystals. The deviation from the thermal conductivity calculated using DFT is within 1% at 200 to 500 K for Si and within 5.4% at 200 to 1000 K for GaN. © 2019 The Japan Society of Applied Physics

H eat generation in semiconducting materials has become a critical problem in modern nanoscale electronics. As the size of electric devices decreases, the power density and device temperature increase, which is one of the major factors contributing to the degradation of device performance and reliability.1) To design semiconductor materials with better thermal manageability, efficient methods for the theoretical simulation of thermal conductivity are required.

The main carrier of heat in semiconductors is the phonon, which is a quantum of lattice vibration. Current methods of simulating lattice thermal conductivity can be classified into three categories:2,3) (1) anharmonic lattice dynamics (ALD) in combination with phonon transport calculation using the Boltzmann transport equation (BTE) and Fourier’s law,4,5) (2) equilibrium molecular dynamics (EMD) using the Green–Kubo formula,6,7) and (3) direct evaluation of the heat flux by nonequilibrium molecular dynamics (NEMD).8,9) These theoretical frameworks require the accurate prediction of interatomic force in the solid. Density functional theory (DFT) calculation is one of the most well-established techniques for accurate force prediction, including that of the effect of changes in the electronic state with atomic displacement. However, the high computational cost limits the application of DFT calculation in thermal conductivity simulations. Although the combination of DFT calculations with ALD10) or EMD11) approaches has been successful in accurate prediction of the lattice thermal conductivities of semiconductor crystals, the application of this technique to systems with more complex structures such as defective or disordered ones is not realistic because of the rapid increase in computational cost with increasing system size. Direct combination of NEMD with DFT calculation is almost impossible because the system size used in NEMD simulations must be much larger to reproduce a reasonable temperature gradient.12) Computational time can be reduced by using the empirical potential, but the accuracy is insufficient compared to that of DFT-based calculations.8) An alternative simulation technique that can resolve the trade-off between the accuracy of force prediction and computational cost is urgently needed.

The application of machine learning techniques to solid-state physics has rapidly developed in recent years.15,16) These techniques are promising for thermal conductivity simulation.17) As shown in early works by Behler et al.18,19) and subsequent studies20–26) the high-dimensional neural network potential (HDNNP) can describe the relation between the total energy of a system and its atomic arrangement. It is naturally expected that the force acting on atoms can also be described by the HDNNP because the derivative of the total energy with respect to the atomic displacement gives the force. Force prediction in semiconducting materials using the HDNNP or other machine learning techniques has already been reported in several studies27–29) and the phonon dispersions have been well reproduced in these reports.28,29) However, the accuracy of the force prediction is limited to the order of 100 meV Å⁻¹, which is much higher than the accuracy of DFT calculations (a few tens of millielectron-volts per angstrom) required to simulate thermal conductivity. To the best of our knowledge, the HDNNP has never been applied to thermal conductivity simulation. Here, we show that much higher accuracy can be obtained by training HDNNP parameters with a focus on force fitting. We chose crystalline Si and GaN as representative semiconducting materials with one and two atom types, respectively. In both systems, we obtained an HDNNP that can predict the force with the accuracy of DFT. The obtained thermal conductivities are within 5.4% of those calculated by DFT.

In this study, we adopted the HDNNP model developed by Behler et al.,16,17) in which the total energy of the system \( E_{\text{tot}} \) is expressed as the sum of the energy contributions from each atom \( E_i \), i.e., \( E_{\text{tot}} = \sum E_i \). Here, we neglect the effect of the long-range electrostatic potential, and \( E_i \) is determined by the local atomic environment, which is described by the cutoff and symmetry functions. The cutoff

Supplementary material for this article is available online.
function determines the sphere of the local environment, whereas the symmetry functions represent the radial and angular distributions of neighboring atoms. The following cutoff function is used:

\[
 f_c(R_{ij}) = \begin{cases} \tanh^3 \left( 1 - \frac{R_{ij}}{R_c} \right) & \text{for } R_{ij} \leq R_c, \\ 0 & \text{for } R_{ij} > R_c \end{cases}
\]  

(1)

where \( R_c \) is the cutoff distance, and \( R_{ij} \) is the interatomic distance. For the symmetry functions, we use the following three types of functional forms:

\[
 G_1^i = \sum_{j \neq i} f_c(R_{ij}),
\]

\[
 G_2^i = \sum_{j=1}^{N_{\text{atom}}} e^{-\eta(R_{ij} - R_c)^2} f_c(R_{ij}),
\]

\[
 G_3^i = 2^{1-\zeta} \sum_{j=1}^{N_{\text{atom}}} [1 + \lambda \cos \theta_{ijk}] e^{-\eta(R_{ij}^2 + R_{jk}^2)} f_c(R_{ij}) f_c(R_{jk})
\]

(4)

Here, \( N_{\text{atom}} \) represents the total number of atoms inside the cutoff sphere. The hyperparameters \( R_c, R_s, \lambda, \zeta, \) and \( \eta \), which should be set before the HDNNP is trained, must be tuned for better prediction performance. We manually tuned these hyperparameters. The setting used in this study was the one that returned the least root mean square error (RMSE) among the \( \sim 20 \) patterns of tests for GaN and the \( \sim 10 \) patterns for Si. The values of the hyperparameters used in this work are summarized in the supplementary material, available online at stacks.iop.org/APEX/12/095001/mmedia.

As a simple example, let us consider a neural network (NN) comprising a single hidden layer with \( N_h \) nodes and \( N_i \) input nodes associated with the symmetry functions. The atomic energy and force with respect to the atomic displacement and/or bond breaking are redundant for this purpose. Therefore, we selected temperatures between room temperature and the melting temperature for the classical MD simulation. For Si, we chose 300, 600, and 900 K, and for GaN, we chose 300, 700, 1100, 1500, 1900, 2300, and 2700 K.

Next, we randomly selected 100 snapshots from each MD trajectory and performed DFT calculations for the corresponding atomic configurations without structure relaxation by VASP with the projector augmented-wave potential. The LDA exchange–correlation functional was used for Si, and GGA-PBE was used for GaN. The cutoff energy of the plane wave basis was set to 550 eV. In all these MD and DFT calculations, we chose supercell sizes of 64 atoms for Si and 32 atoms for GaN (see the supplementary material). To increase the variation in the atomic configurations in the training data and to extend the flexibility of HDNNP, we also included data obtained using slightly different lattice constants. The total numbers of atomic configurations in the DFT data sets were 2100 and 3500 for Si and GaN, respectively.

In the training procedure using the above data set, we used an NN topology with two hidden layers. The number of nodes in each layer was set to 500, and the hyperbolic tangent was adopted as the activation function. Adaptive moment estimation was chosen as the optimization algorithm.

In Fig. 1, we compare the forces in the Si and GaN systems predicted by DFT and HDNNP. During training, 90% of the DFT data sets were used as training data, and the remaining 10% were used as validation data. We set the parameter \( \alpha = 0.99 \) in the loss function in Eq. (7) so that the RMSE of the force is dominant in the training procedure. The final RMSE of force prediction in the validation data was 25.5 meV Å\(^{-1}\) for Si and 37.8 meV Å\(^{-1}\) for GaN. Even at a very low weight of the loss function, the final RMSE of the
The phonon dispersion curves obtained using HDNNP agree well with the DFT calculation results and previous reports\(^\text{45,46}\) for both Si and GaN, as shown in Fig. 2. Note that we focus on the comparison between the HDNNP and DFT results, and we do not include the non-analytic correction for LO–TO splitting.\(^\text{47}\) Thus, the frequencies of the highest LO mode in GaN largely differ from the experimental values.

Next, we simulated the lattice thermal conductivity based on ALD by combining HDNNP and the phonopy package,\(^\text{4}\) using a procedure similar to that for the calculation of phonon dispersion. The irreducible displacements for evaluating the third-order potential for three phonon processes were obtained using phonopy, and the forces acting on atoms in each displacement pattern were predicted using HDNNP. The phonon–phonon interaction strength and the corresponding lifetime were extracted from these force data; then, the lattice thermal conductivity (\(\kappa\)) was calculated using the single-mode relaxation time approximation (RTA) of the linearized phonon BTE. Figure 3 shows a comparison of the temperature dependence of the thermal conductivity obtained from the force predictions of HDNNP and VASP calculations. We note that in both calculations, the thermal conductivities in Si and GaN are underestimated compared with previous reports (approximately 150 W m\(^{-1}\) K\(^{-1}\) and 400 W m\(^{-1}\) K\(^{-1}\) at 300 K, respectively\(^\text{8,48}\)). Three factors can be listed as plausible reasons for the underestimation: i) the small supercell size in the thermal conductivity simulation; ii) the DFT calculation setup, including the choice of pseudopotentials, exchange–correlation functional, and lattice constants; and iii) the RTA in the phonon BTE solution (iterative treatment is required to obtain the exact thermal conductivity\(^\text{8,48}\)). Here, we focus on the reproducibility of the DFT calculation by the HDNNP; thus, we do not discuss this point in detail. The calculation results from HDNNP and DFT under the same simulation conditions for both Si and GaN are in good agreement, indicating the strong potential of HDNNP for application in thermal conductivity simulations. The deviation from the DFT calculation results is within 1% at 200 to 500 K for Si and within 5.4% from 200 to 1000 K for GaN. For example, for Si, the \(\kappa\) value at 300 K is estimated to be 110.4 W m\(^{-1}\) K\(^{-1}\) from the HDNNP results and 112.1 W m\(^{-1}\) K\(^{-1}\) from the DFT results. For GaN, the thermal conductivities along the in-plane direction, \(\kappa_{ij}\), and along the out-of-plane direction, \(\kappa_{il}\), at 300 K are 275.5 and 309.7 W m\(^{-1}\) K\(^{-1}\), respectively. These values are in good agreement with the DFT calculation results, which are \(\kappa_{ij} = 274.2\) and \(\kappa_{il} = 325.5\) W m\(^{-1}\) K\(^{-1}\).

Finally, we comment on the computational time required to calculate the thermal conductivity using DFT and HDNNP. For Si, the DFT calculation for a total of 111 displacement patterns required a total of 24 cores of two CPUs (Intel\textsuperscript{®} Xeon\textsuperscript{®} E5-2680v3) and 4.5 h, which reduced to 1 core of a CPU and 10 s for the HDNNP calculation. For GaN, the DFT calculations for a total of 582 displacements required 24 cores of the CPUs and 21.0 h, which reduced to 24 cores of the CPUs and 1.5 min for the HDNNP calculation. The computational time for the thermal conductivity simulation itself became less than 1/800 on using HDNNP. On the other hand, to obtain reliable HDNNP, considerable time is required to prepare the training data sets and tune the hyperparameters. In this study, the generation of the data set required 24 cores of the CPUs and 140.4 h for Si and 126.8 h for GaN. Hyperparameter tuning took 24 cores of the CPUs and \(~\sim 40\) h for Si and \(~\sim 50\) h for GaN. Thus, if we compute the thermal conductivity in perfect bulk crystals, HDNNP is not an efficient method. The situation changes when we treat a system with defects and/or impurities. Currently, the thermal conductivity of systems with defects...
Fig. 2. (Color online) Comparison of phonon dispersions in (a) Si and (b) GaN obtained by HDNNP and DFT calculations. The supercell sizes for calculations of the second-order force constant were set to $2 \times 2 \times 2$ for the conventional unit cell in Si and $3 \times 3 \times 2$ for the primitive unit cell in GaN. The experimental data for comparison were obtained from Ref. 46 for Si and Ref. 45 for GaN.

Fig. 3. (Color online) Comparison of thermal conductivities (a) in Si, (b) along the in-plane (100) direction in GaN, and (c) along the out-of-plane (001) direction in GaN obtained by HDNNP and DFT calculations. The supercell sizes for calculations of the third-order force constant were set to $2 \times 2 \times 2$ for the conventional unit cell in Si and $2 \times 2 \times 2$ for the primitive unit cell in GaN. In the calculation of the linearized Boltzmann equation with the RTA, the Brillouin zone was sampled on an $11 \times 11 \times 11$ mesh in all cases.
and impurities is evaluated based on the first-order perturbation\(^49\) or ab initio Green’s function technique\(^50\)–\(^52\) which can include the defect/impurity effect on harmonic interatomic force constants (IFCs). In contrast, the direct extension of DFT-based ALD for such systems is cumbersome because of the high computational cost of anharmonic IFCs. For example, the DFT calculation of anharmonic IFCs for a system with a single defect/impurity in a 64-atom supercell of Si requires calculations for ∼6418 displacement patterns, which will take 24 cores of CPUs and ∼430 h. If we can construct the data set including defects with reasonable computational time, and train an HDNNP that accurately reproduces the forces in such systems, the HDNNP will become beneficial for both Green’s function based techniques and the evaluation of anharmonic IFCs. This would provide an alternative efficient method to investigate the thermal properties in complex systems with high accuracy. The flexibility and extensibility of the HDNNP in this direction remain to be elucidated.

Acknowledgments The following financial support is acknowledged: IST PRESTO JPMIPR1717 (E. M.), the IST “Materials Research by Information Integration” Initiative (MIZI) of the Support Program for Starting Up Innovation Hub (S. W.), and MEXT KAKENHI 17H05330 (S. W. and E. M.). We thank Prof. Togo for fruitful discussions on the connection between HDNNP and phonopy/phono3py. The calculations were performed using the computer facilities of the Institute for Solid State Physics, Information Technology Center, the University of Tokyo, and RIKEN (HOKUSAI GreatWave).

ORCID iDs
Emi Minamitani https://orcid.org/0000-0002-8003-6526 Satoshi Watanabe https://orcid.org/0000-0002-8069-6938

1) P. Ball, Nature 492, 174 (2012).
2) D. G. Cahill et al., Appl. Phys. Rev. 1, 011305 (2014).
3) P. K. Schelling, S. R. Phillpot, and P. Keblinski, Phys. Rev. B 65, 144306 (2002).
4) A. Togo, L. Chaput, and I. Tanaka, Phys. Rev. B 91, 094306 (2015).
5) K. Mizokami, A. Togo, and I. Tanaka, Phys. Rev. B 97, 224306 (2018).
6) J. Zhou, B. Liao, and G. Chen, Semicond. Sci. Technol. 31, 043001 (2016).
7) L. Lindsay, D. A. Broideo, and T. L. Reinecke, Phys. Rev. B 87, 165201 (2013).
8) D. A. Broideo et al., Appl. Phys. Lett. 91, 11305 (2007).
9) R. Kubo, M. Yoskota, and S. Nakajima, J. Phys. Soc. Japan 12, 1203 (1957).
10) S. G. Volz and G. Chen, Phys. Rev. B 61, 2651 (2000).
11) A. J. C. Ladd, B. Moran, and W. G. Hoover, Phys. Rev. B 34, 5058 (1986).
12) F. Müller-Plathe, J. Chem. Phys. 106, 6082 (1997).
13) M. S. El-Genk, K. Talaat, and B. J. Cowen, J. Appl. Phys. 123, 205104 (2018).
14) C. Carbogni, R. Ramprasad, and M. Scheffler, Phys. Rev. Lett. 118, 175901 (2017).
15) L. Zeleborová, Nat. Phys. 13, 420 (2017).
16) J. Behler, J. Chem. Phys. 145, 170901 (2016).
17) F. Eriksson, E. Fransson, and P. Erhart, Adv. Theory Simul. 2, 1800184 (2019).
18) J. Behler and M. Parrinello, Phys. Rev. Lett. 98, 146401 (2007).
19) J. Behler, Angew. Chem. Int. Ed. 56, 12828 (2017).
20) W. Li et al., J. Chem. Phys. 147, 214106 (2017).
21) W. Li, Y. Ando, and S. Watanabe, J. Phys. Soc. Japan 86, 104004 (2017).
22) N. Arrirth, U. Urban, and G. Ceder, Phys. Rev. B 96, 14112 (2017).
23) N. Arrirth and U. Urban, Comput. Mater. Sci. 114, 135 (2016).
24) B. Onat et al., Phys. Rev. B 97, 94106 (2018).
25) R. Kobayashi et al., Phys. Rev. Mater. 1, 53604 (2017).
26) W. Jeong et al., J. Phys. Chem. C 122, 22790 (2018).
27) Y. Huang et al., Phys. Rev. B 99, 064103 (2019).
28) M. R. G. Marques et al., Phys. Chem. Chem. Phys. 21, 6506 (2019).
29) A. P. Bartók et al., Phys. Rev. Lett. 104, 136403 (2010).
30) Our HDNNP code can be downloaded at [https://github.com/ogura-edu/HDNNP].
31) S. Tokui et al., [https://learningsys.org/papers/LearningSys_2015_paper_33.pdf].
32) T. Akiba, F. Fukuda, and S. Suzuki, [https://learningsys.org/papers/LearningSys_2015_paper_33.pdf].
33) S. Plimpton, J. Comput. Phys. 117, 1 (1995).
34) LAMMPS is provided at [https://lammps.sandia.gov/].
35) G. Kresse and D. Joubert, Phys. Rev. B 54, 11169 (1996).
36) G. Kresse and J. Furthmüller, Comput. Mater. Sci. 6, 15 (1996).
37) G. Kresse and J. Hafner, Phys. Rev. B 47, 558 (1993).
38) G. Kresse and J. Furthmüller, Phys. Rev. B 54, 11169 (1996).
39) F. H. Stillinger and T. A. Weber, Phys. Rev. B 31, 5262 (1985).
40) P. E. Blöchl, Phys. Rev. B 50, 17953 (1994).
41) J. P. Perdew, K. Burke, and M. Ernzerhof, Phys. Rev. Lett. 77, 3865 (1996).
42) D. Yoo et al., arXiv:1903.04366 (2019).
43) D. P. Kingma and J. Ba, arXiv:1412.6980 (2014).
44) A. Togo and I. Tanaka, Scr. Mater. 108, 1 (2015).
45) T. Ruf et al., Phys. Rev. Lett. 86, 906 (2001).
46) A. Ward and D. A. Broideo, Phys. Rev. B 81, 085205 (2010).
47) S. Baroni et al., Rev. Mod. Phys. 73, 515 (2001).
48) L. Lindsay, D. A. Broideo, and T. L. Reinecke, Phys. Rev. Lett. 109, 095901 (2012).
49) S. Tamura, Phys. Rev. B 27, 858 (1983).
50) N. A. Katcho et al., Phys. Rev. B 90, 094117 (2014).
51) A. Katre et al., Phys. Rev. Mater. 2, 050602 (2018).
52) C. A. Polanco and L. Lindsay, Phys. Rev. B 98, 014306 (2018).