https://doi.org/10.1007/s11467-022-1170-5

Enhancing thermal transport in multilayer structures: A molecular dynamics study on Lennard-Jones solids

Cuiqian Yu, Yulou Ouyang, Jie Chen†

Center for Phononics and Thermal Energy Science, China–EU Joint Lab for Nanophononics, MOE Key Laboratory of Advanced Micro-structured Materials, School of Physics Science and Engineering, Tongji University, Shanghai 200092, China

Corresponding author. E-mail: †jie@tongji.edu.cn

Received March 27, 2022; accepted April 29, 2022

Supporting Information

I. Optimal Parameter Selection for Structure Design

Guidelines for designing 2SL structures

Study of the wave-particle duality of phonons in multilayer structures is a research hotspot. Fig. S1 depicts a typical diagram of the P_0-κ relationship. The P_t in Fig. S1 can be seen as a signature of the transition from coherent dominant phonon transport to incoherent phonon dominant phonon transport [1].

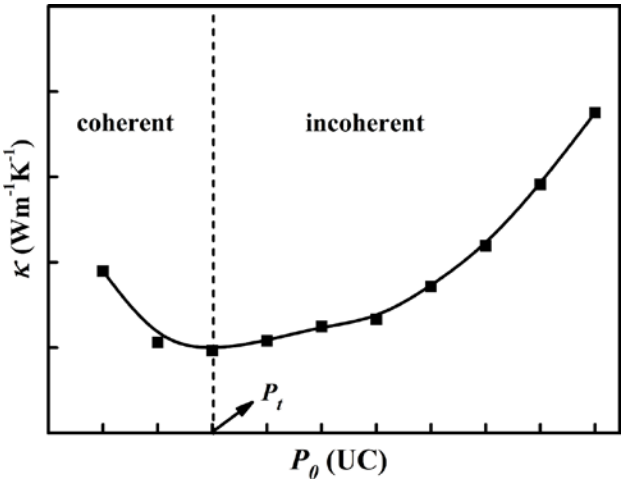

Figure S1 A typical diagram of SL's κ as a function of P_0.

We designed the 2SL with coherent and incoherent phonon synergy by modulating P_1 and P_2. We set L to be 1024 UC, P_0 to be 4, 8, 16, 32, 64, 128, 256, 512, 1024 UC. Through traversing the design space, Fig. S2 shows that the number of 2SL with the configuration that $P_1 < 8$ UC and 2SL with the configuration that $P_1 > 8$ UC are both dependent on P_t. In order to get enough 2SL structures for both of the situations, P_t should be large enough. As P_t increase to a certain extent, the number of structures for the two situations are both sufficient for ML research. RML is converted from 2SL via swapping the positions of several layers. So increasing P_t is the key to creating enough 2SL and RML structures. And then we use MD simulations to increase P_t.

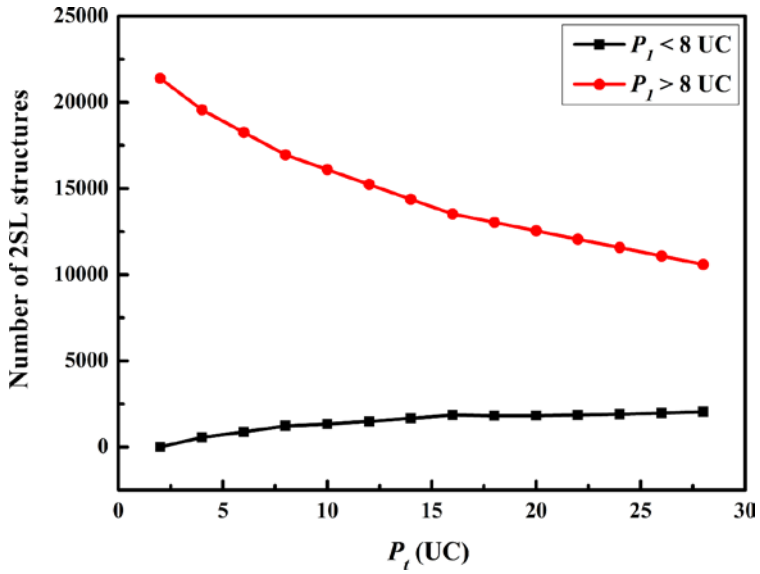

Figure S2 The relationship between P_t and the numbers of 2SL with two configurations: P_1 < 8 UC and P_1 > 8 UC.

Increase P_t through MD simulations

In order to improve P_t, we optimized the parameters of L, S, bond strength mismatch, mass mismatch and temperature. NEMD is performed to calculate κ_{eff} of these structures.

The κ calculated by NEMD based on Fourier law is always used to do convergence test for L of the system. Note that κ is calculated by $\kappa = -\frac{J}{S\nabla_T}$, where ∇_T is the temperature gradient. κ_{eff} is calculated by $\kappa_{eff} = \frac{JL}{S(T_h - T_c)}$. The reason that κ_{eff} cannot be used in the convergence test for L is that it treats the temperature gradient of each structure as the same regardless of L of the structure. Fig. S3(a) shows the κ of SL is a function of L. According to the linear extrapolation method [2,3], $\frac{1}{\kappa} = \frac{1}{\kappa_\infty}(1 + \frac{L_{MFP}}{L})$, where κ_∞ is the converged κ and L_{MFP} is the effective phonon MFP. The inset in Fig. S3(a) shows the relationship between $\frac{1}{\kappa}$ and $\frac{1}{L_{SL}}$. As we can see, for both curves in Fig. S3(a), a L of 1024 UC is long enough to obtain converged κ.

Fig. S3(b) shows P_0 dependence of κ_{eff} for different L. As we can see, P_t doesn't change as L changes, consistent with the result of previous work [4,5]. This phenomenon illustrates that the values of P_t for any mini-SLs in 2SL and RML are 8 UC. The κ of different structures should be compared at the same L because of the size effect. Since the κ converges at the L of 1024 UC, a 5% error tolerance of L is introduced in the design of three forms of multilayered structures, which will not greatly affect the κ_{eff}. Therefore, the length of the three forms of multilayered structures is fixed at 1024 UC with a 5% error tolerance.

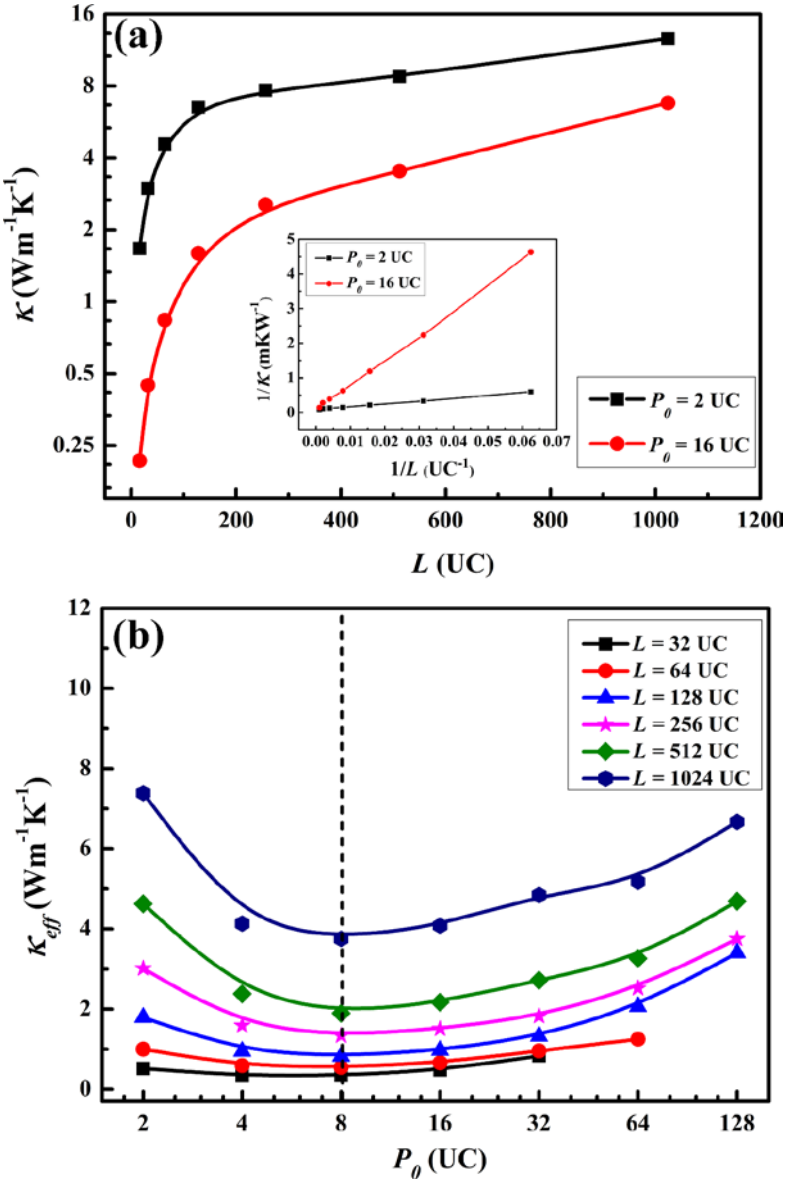

Figure S3 (a) κ of SL as a function of L. The two curves represent SL with P_0 = 2 UC and 16 UC, respectively. Inset: the relationship between $\frac{1}{\kappa}$ and $\frac{1}{L}$. **(b)** P_0 dependence of κ_{eff} for different L. The dotted line is used to indicate P_t for different L. S is 3 UC × 3 UC, ε of materials A and B are 4 and 16 times of ε_{Ar}, masses of materials A and B are equal to that of argon and temperature is 30K.

The other aspect of the size effect is S. When periodic boundary condition is used and S of the simulated structure is too small, it can't accurately simulate the phonon interaction in the bulk material. So it's necessary to do convergence tests for S. Fig. S4 shows the P_0 dependence of κ_{eff} for different S. As we can see, a S of 3 UC × 3 UC is large enough to generate converged κ_{eff}. Therefore, to save computing resources, the S of three forms of multilayered structures is fixed at 3 UC × 3 UC. Also, P_t doesn't change as the S changes. So far, it can be seen that the size of the structure does not affect the P_t. Next, we change bond strength mismatch, mass mismatch and temperature.

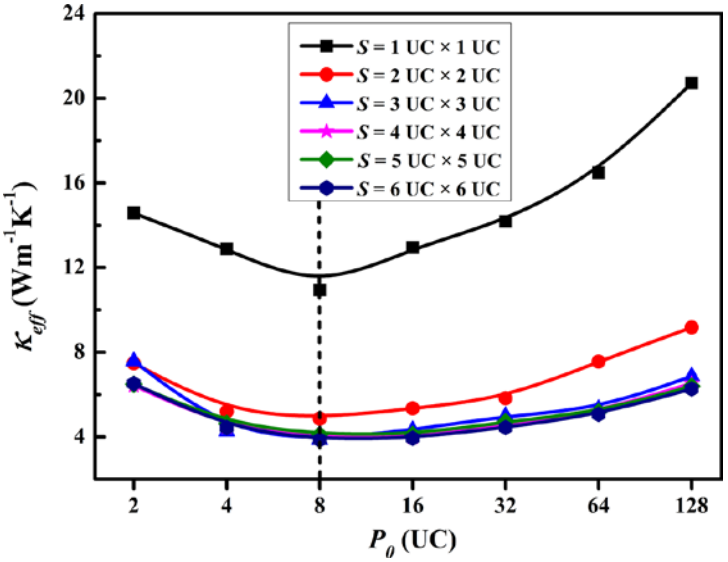

Figure S4 P_0 dependence of κ_{eff} for different S. The dotted line is used to indicate P_t for different S. L is 1024 UC, ε of materials A and B are 4 and 16 times of ε_{Ar}, masses of materials A and B are equal to that of argon and temperature is 30 K.

The parameter ε of materials A and B in LJ potential determine the bond strength of materials A and B, respectively. As shown in Fig. S5, for the comparison between curve $1\varepsilon_{Ar}$-$16\varepsilon_{Ar}$ and $4\varepsilon_{Ar}$-$16\varepsilon_{Ar}$, κ_{eff} of SL with larger bond strength mismatch is smaller, which depends on the high thermal resistance of the interfaces, consistent with the result of previous work [6]. P_t and MFP are positively correlated. When MFP increases, P_t increases[1]. Larger ε corresponds to longer MFP [7] because that the enhancement of ε will weaken the anharmonic phonon-phonon scattering [8,9]. The overall average ε of the case $4\varepsilon_{Ar}$-$16\varepsilon_{Ar}$ is largest in the three cases, the phonon MFP in this case is the largest, so the P_t is largest. For the purpose of improving P_t, we fix the ε of materials A and B at 4 and 16 times of ε_{Ar}, respectively.

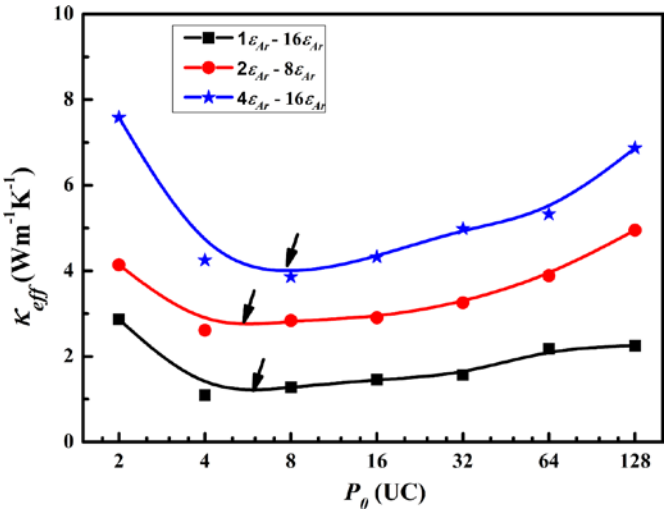

Figure S5 P_0 dependence of κ_{eff} for different bond strength mismatch of materials A and B. The arrows are used to indicate P_t for different bond strength mismatch. L is 1024 UC, S is 3 UC × 3 UC, masses of materials A and B are equal to that of argon and temperature is 30 K.

To explore the relationship between mass mismatch and P_t, we study the P_0 dependence of κ_{eff} for different mass mismatch of materials A and B. In Fig. S6, 1m-2m means that masses of materials A and B are 1 and 2 times of that of Ar. With the increase of mass

mismatch, the nonmonotonic phenomenon of κ_{eff} gradually disappears. To ensure that the synergistic effect of coherent phonons and incoherent phonons can be greatly studied, although P_t increases when the mass mismatch initially increases, we still end up with a fixed mass mismatch of 1m-1m to avoid its unnecessary effects.

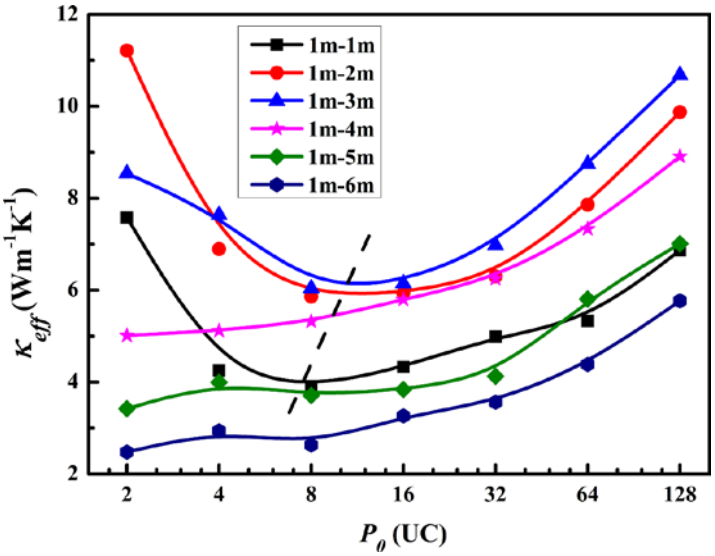

Figure S6 P_0 dependence of κ_{eff} for different mass mismatch of materials A and B. The dotted line is used to indicate P_t for mass mismatch of 1m-1m,1m-2m and 1m-3m. L is 1024 UC, S is 3 UC × 3 UC, ε of materials A and B are 4 and 16 times of ε_{Ar} and temperature is 30 K.

Previous works have demonstrated that the relationship between P_0 and temperature [5,10,11]. Higher temperature increases anharmonic phonon-phonon scattering, which will reduce the phonon MFP, leading to the suppression of phonon wave interference and the decreasing of κ_{eff}. So as shown in Fig. S7, P_t of shifts to a smaller value with increasing temperature. Meanwhile, high temperature in SL will lead to unstable heat flow in MD simulation, which will affect the accuracy of the results. But, even lower temperatures won't work because MD simulations are not suitable for extremely low temperatures. Therefore, the temperature is fixed at 30 K.

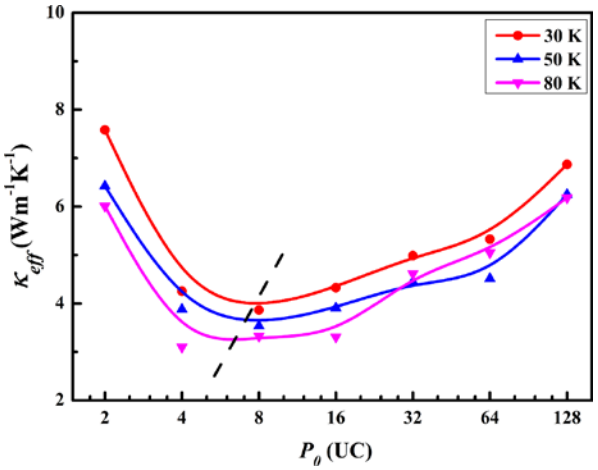

Figure S7 P_0 dependence of κ_{eff} for different temperatures. The dotted line is used to indicate P_t for different temperatures. L is 1024 UC, S is 3 UC × 3 UC, ε of materials A and B are 4 and 16 times of ε_{Ar} and masses of materials A and B are equal to that of argon.

In order to improve P_t, we fix the parameters as follows: L is 1024 UC, S is 3 UC × 3 UC, ε of materials A and B are 4 and 16 times of ε_{Ar}, masses of materials A and B are equal to that of argon and temperature is 30 K. In this framework, P_t is 8 UC. The final relationship between P_0 and κ_{eff} of SL in this work is shown in Fig. S8.

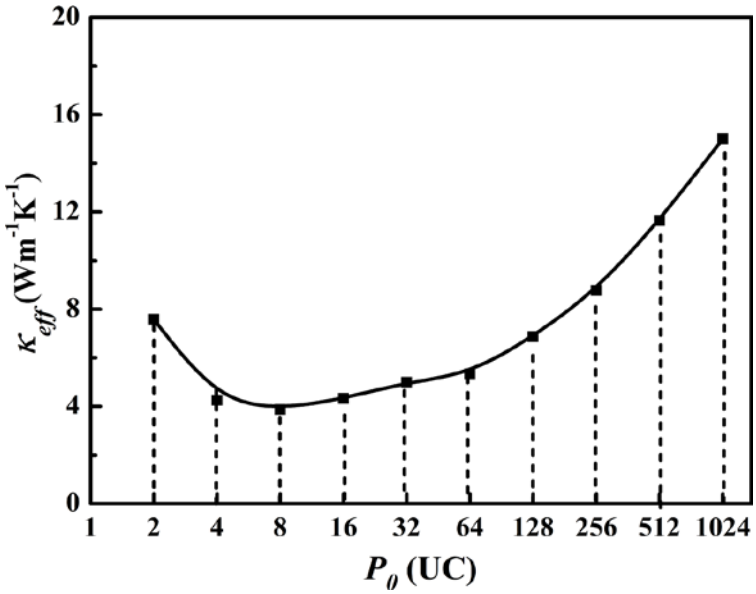

Figure S8 The relationship between P_0 and κ_{eff} of SL in this work.

II. Convergence test for κ_{eff} of RML with changes in *S_num*

To increase the randomness of structures, we converted 2SL to RML by swapping the positions of layers for several times. *S_num* is used as a criterion of randomness. We start from three representative 2SL structures in the structure library. The values of N_1 and N_2 in these three 2SL structure are close to each other. The relationship between κ_{eff} and *S_num* is shown in Fig. S9. A decreasing trend of κ_{eff} is observed as *S_num* increases because more and more coherent phonons are localized. However, κ_{eff} of RML doesn't decrease 100% monotonically with increasing randomness, which is due to the formation of new mini-SLs with small P_0 and regenerate coherent phonons. The illustration shows a visual representation of new mini-SLs. The essence of this phenomenon is that *S_num* could not represent the randomness, so Chakraborty *et al*. [12] took more parameters into account to quantify randomness. In Fig. S9, we can see that when *S_num* reaches 5000, κ_{eff} converges. In this work, the role of RML is to serve as a reference for 2SL. Therefore, for each 2SL, we adopt the structure which swaps the positions of several layers for 5000 times as its corresponding RML in our final prediction structure library.

However, when N_1 and N_2 in the initial 2SL are very different from each other, for example, the configuration of the initial 2SL is as follows: P_1 = 2 UC, N_1 = 248, P_2 = 194 UC, N_2 = 8. No matter how many times the structure swaps, there are always plenty of mini-SLs in RML, which can promote the increment of κ_{eff}.

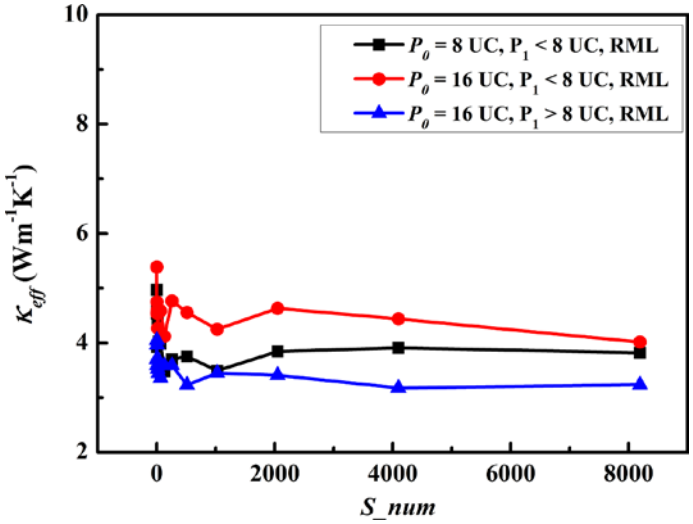

Figure S9 The relationship between *S_num* and κ_{eff} of RML. The illustration represents typical structures with different *S_num*. (n and m are positive integers)

III. Structure details of 2SL with high κ_{eff}

Table S1 Configuration of 2SL in Fig. 2(d) with κ_{eff} higher than its SL counterpart. All the 2SL with high κ_{eff} have the feature that the difference between P_1 and P_2 is large. The κ_{eff} of SL with P_0 of 16 UC is 4.32 Wm^{-1}K^{-1}.

Structure	P_0	P_1	P_2	$\lvert P_2-P_1 \rvert$	κ_{eff} (Wm^{-1}K^{-1})
2SL	16	10	778	768	5.99
2SL	16	10	394	384	5.78
2SL	16	394	10	384	5.50
2SL	16	266	10	256	5.44
2SL	16	10	202	192	5.43
2SL	16	12	524	512	5.32
2SL	16	778	10	768	5.26
2SL	16	202	10	192	5.24
2SL	16	524	12	512	5.16
2SL	16	138	10	128	5.09
2SL	16	268	12	256	5.08
2SL	16	12	268	256	5.06
2SL	16	106	10	96	5.04
2SL	16	140	12	128	4.92
2SL	16	10	266	256	4.86
2SL	16	142	14	128	4.83

2SL	16	76	12	64	4.75
2SL	16	10	138	128	4.59
2SL	16	12	140	128	4.54
2SL	16	270	14	256	4.50
2SL	16	46	14	32	4.47
2SL	16	74	10	64	4.45
2SL	16	14	270	256	4.43
2SL	16	12	76	64	4.42
2SL	16	14	142	128	4.37
2SL	16	10	74	64	4.34

IV. κ_{eff} of multilayer structures with larger P_0

Compared Fig. 2(d) and Fig. S10, which are for P_0 of 16 UC and 256 UC, respectively, we can see that the κ_{eff}'s difference between RML and corresponding 2SL is greatly smaller in Fig. S10 than that in Fig. 2(d). At low interface density, the effect of randomness is small, which is a result of weak phonon coherence. The existence of phonon coherence is a prerequisite for Anderson localization, which has been experimentally [13,14] and theoretically [15,16] observed.

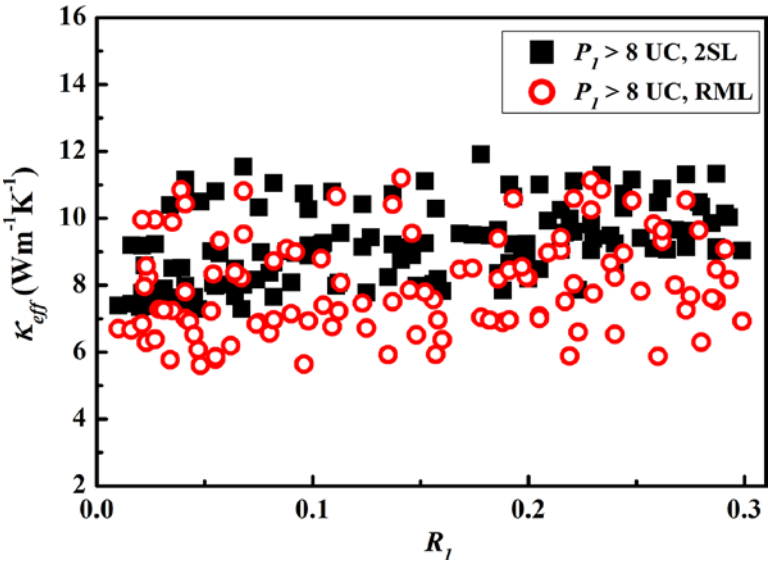

Figure S10 The relationship between R_1 and κ_{eff} of multilayer structures with that the P_0 of the corresponding SL is 256 UC. The P_1 of 2SL and RML is greater than 8 UC, i.e. the value of P_t.

V. Evaluation of algorithms and descriptors

For the case that the prediction target (e.g. κ in this work) is continuous values, the regression algorithms are used to fit the initial input data, and then the trained models are used to predict κ of a larger design space.

The seven regression algorithms used in this work.

Linear regression (LR)

Linear regression is an approach describing the linear relationship between target response and regressors [17]. This method can easily lead to overfitting due to the problem of descriptor multicollinearity, so it needs some modifications.

Least absolute shrinkage and selection operator - linear regression (LASSO-LR)

In order to prevent overfitting of LR, the L_1-norm term was introduced as a penalty term by the method LASSO to modify LR [18]. The method LASSO can also be used to modify other algorithms. The essence of this method is to remove the interference of redundant descriptors so that it can reflect the important descriptors more accurately. Here we need to adjust one hyperparameter to get the best modification performance.

Ridge - linear regression (Ridge-LR)

The same as LASSO, the method Ridge modifies LR by introducing the L_2-norm term [19]. The difference between LASSO and Ridge is the form of the penalty item. Here we also need one hyperparameter to get the best performance.

Elastic Net - linear regression (Elastic Net-LR)

The method Elastic Net can be described as a combination of LASSO and Ridge [20]. In this method, both the L_1-norm term and the L_2-norm term are introduced and another hyperparameter is introduced to adjust the ratio of the two terms.

Decision tree regression (DTR)

Decision tree regression is a ML method based on the tree structure. Decision tree learning adopts a top-down recursive method and its basic idea is to directly simulate the process of multistage selection and decision making. The algorithms based on the decision tree is superior to other algorithms because the models have interpretable rules and logic statements [21]. But noisy data has a great impact on the prediction performance of the decision tree regression [22], resulting in that it lacks good generalization ability and is easy to overfit. To solve this, we need to prune the decision tree or generate strong learners based on multiple decision trees. Here we need to adjust serveral hyperparameters in the tree structure to get the best modification performance.

Random forest regression (RFR)

Random forest is based on the bagging integration of decision trees [23]. As a strong learner, RFR has better generalization ability. The random forest needs to perform the convergence test for the number of the base learner, i.e. the decision tree, so as to obtain a

lower generalization error. Here we need to adjust serveral hyperparameters in the tree structure and the number of the tree structures to get the best modification performance.

Gradient boosting decision tree (GBDT)

The decision trees in the random forest are independent of each other and have no correlation. The generalization ability of the algorithm is further improved by establishing the correlation between decision trees. Based on the Boosting integration learning method, GBDT [24], which also treats the decision tree as the base learner, approaches the ideal prediction model by introducing the concept of base learner weighting. Here we need to adjust serveral hyperparameters in the tree structure, the number of the tree structures and the learning rate to get the best modification performance.

Algorithm performance comparison and descriptor importance judgment

The selection of algorithms allows us to get the appropriate algorithm for our work, thus improving the accuracy of prediction. Some algorithms need to adjust the hyperparameters which have no analytic solutions, we can use the module GridSearchCV in Python to get the optimal hyperparameters [25]. Meanwhile, effective descriptors can free the ML model from the interference of invalid input data and make the prediction more reliable.

We take R^2, *MSE*, *RMSE* of the test dataset as criteria to evaluate the models generated by different algorithms. R^2 is the direct evaluation index of the regression model. A larger value of R^2 means that the prediction performance of the algorithm is better, and its optimal value is 1. The smaller the value of *MSE* and *RMSE*, the better the prediction performance is. In those algorithms which take the decision tree as the base learner, i.e. DTR, RFR and GBDT, the function 'feature_importances' in Python can help us obtain the importance of each descriptor.

Figure S11 and Fig. S12 show the correlation between κ_{eff} calculated by NEMD and that predicted by linear regression-based and decision tree-based algorithms for the training dataset and test dataset, respectively. The prediction performance of algorithms based on the decision tree in Fig. S12 is better than that of algorithms based on linear regression in Fig. S11 in all three aspects. In Fig. S12, R^2 of DTR, RFR and GBDT trained with all the descriptors are 0.90, 0.91 and 0.92, whereas *MSE* are 0.52, 0.46, 0.39 $Wm^{-1}K^{-1}$, respectively. This shows that the algorithms based on the decision tree, especially GBDT, can reasonably describe the relationship between the descriptors and κ_{eff}.

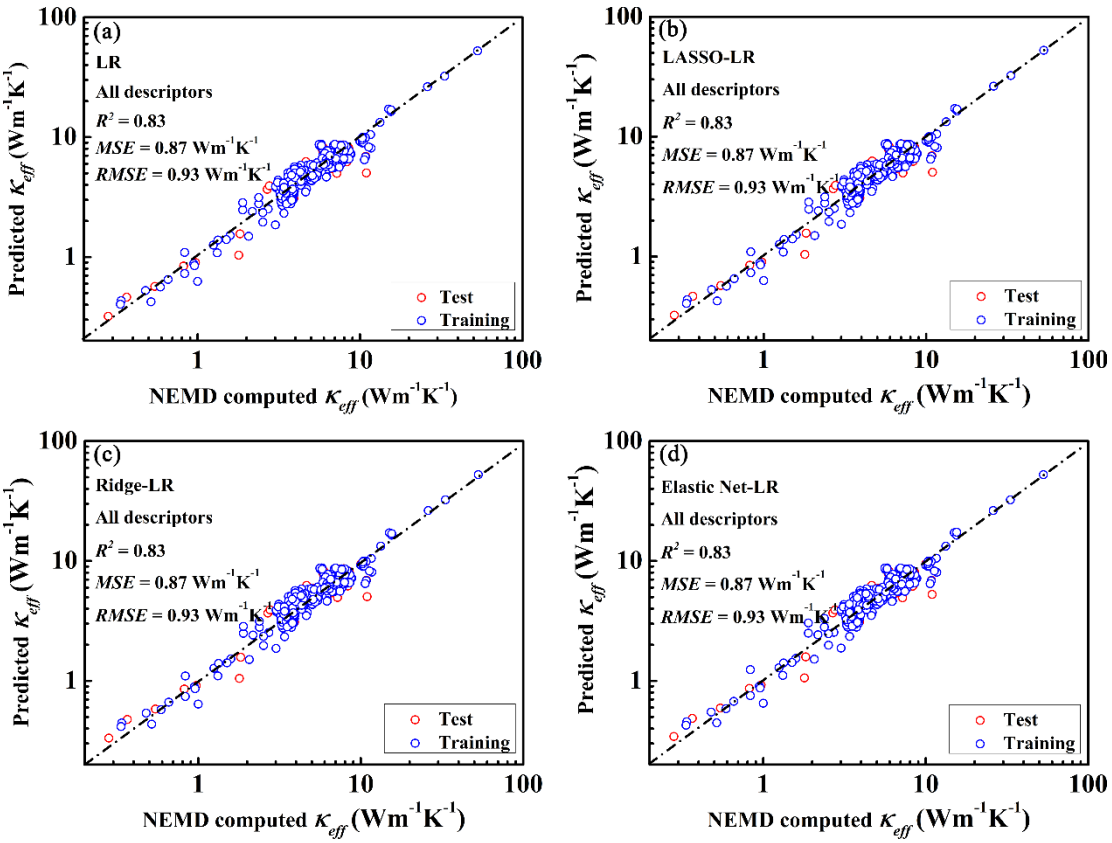

Figure S11 The comparison of κ_{eff} calculated by NEMD and that predicted by algorithms which take linear regression as the base learner, i.e., **(a)** LR, **(b)** LASSO-LR, **(c)** Ridge-LR and (d) Elastic Net-LR.

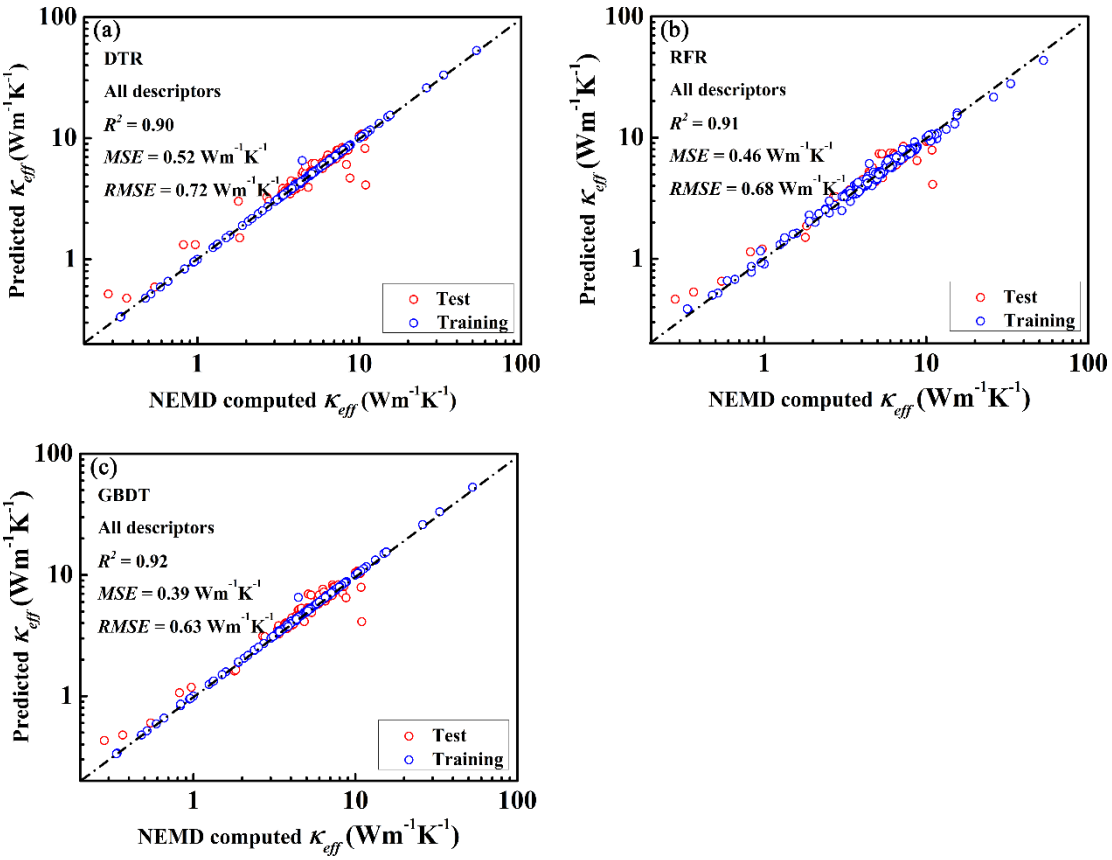

Figure S12 The comparison of κ_{eff} calculated by NEMD and that predicted by algorithms which take the decision tree as the base learner, i.e., **(a)** DTR, **(b)** RFR and **(c)** GBDT.

For most problems, it is not the case that more descriptors make better performance because invalid descriptors will cause unnecessary interference on the results. The presence of less important descriptors only wastes computational resources and has little effect on the prediction target. The purpose of the LR modification, i.e. LASSO-LR, Ridge-LR and Elastic Net-LR, is to remove the interference of redundant descriptors. In Fig. S11, we can see that the prediction performance of the modified models is unchanged, which means that the descriptors here are all reliable.

The normalized distribution of importances for different descriptors in DTR, RFR and GBDT can be found in Fig. S13. The importances of different descriptors for predicting κ_{eff} are shown here, which also illustrates the descriptors in this work are all useful. In Fig. S13(c), we can observe that in addition to *L*, P_1 and P_2 are extremely critical descriptors. The reason for adding the descriptor *L* is that there is a 5% error tolerance error in length when modeling. As shown in Fig. S3(a), a fixed *L* of 1024UC with a 5% tolerance error in this work is long enough to obtain converged κ, so actually *L* is not the most critical factor that determines κ_{eff}.

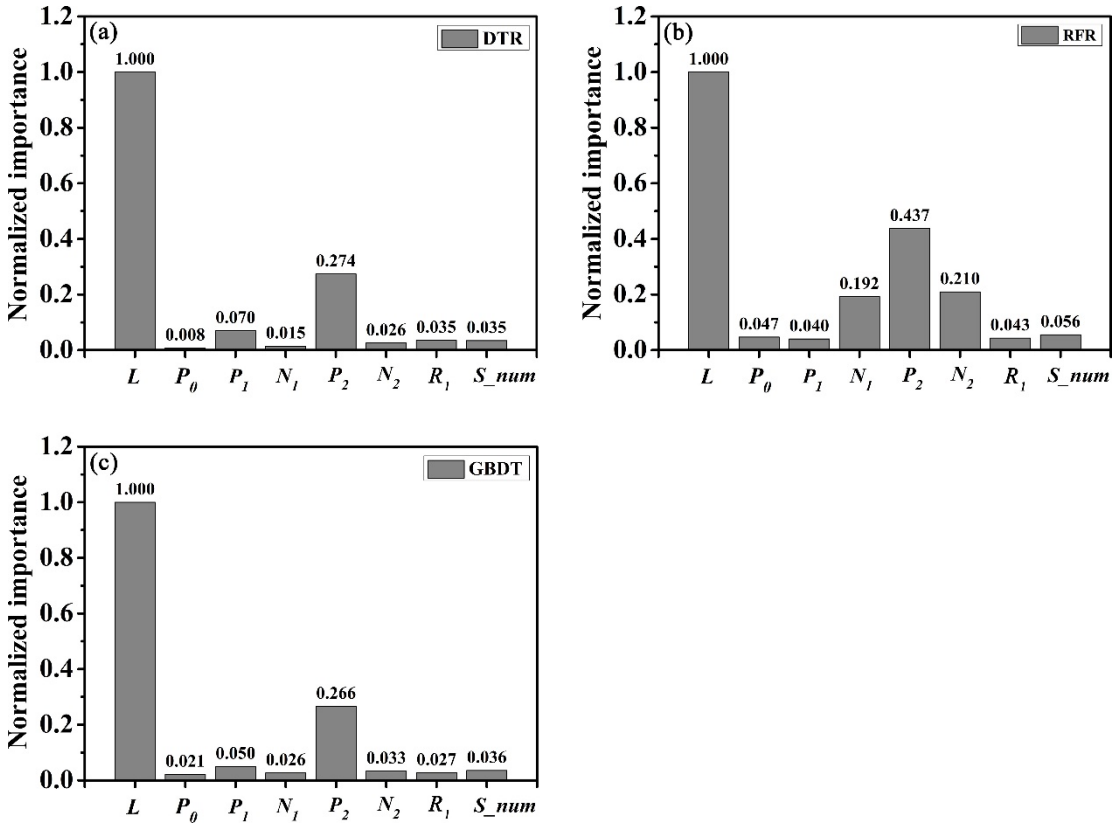

Figure S13 The normalized distribution of importances for different descriptors in algorithms which take the decision tree as base learner, i.e., **(a)** DTR, **(b)** RFR and **(c)** GBDT.

In order to achieve the best prediction performance, GBDT is selected as the final prediction algorithm.

Detailed method of ML

Figure S14 shows the schematic routine in this work. To begin with, the initial input data for ML is the data entries of 401 different structural systems, which is large enough to ensure the accuracy of prediction and avoid the occurrence of overfitting. The κ data of these 401 structures is calculated by NEMD. The initial input data is randomly separated into five folds: one fold used for testing and four folds used for training. We repeat this process five times to ensure that the testing data is orthogonal to each other. The final result is the average of the five predictions above. By this 5-fold cross validation, we can get the optimized state of the current model which eliminates the impact of fluctuations between each data fold.

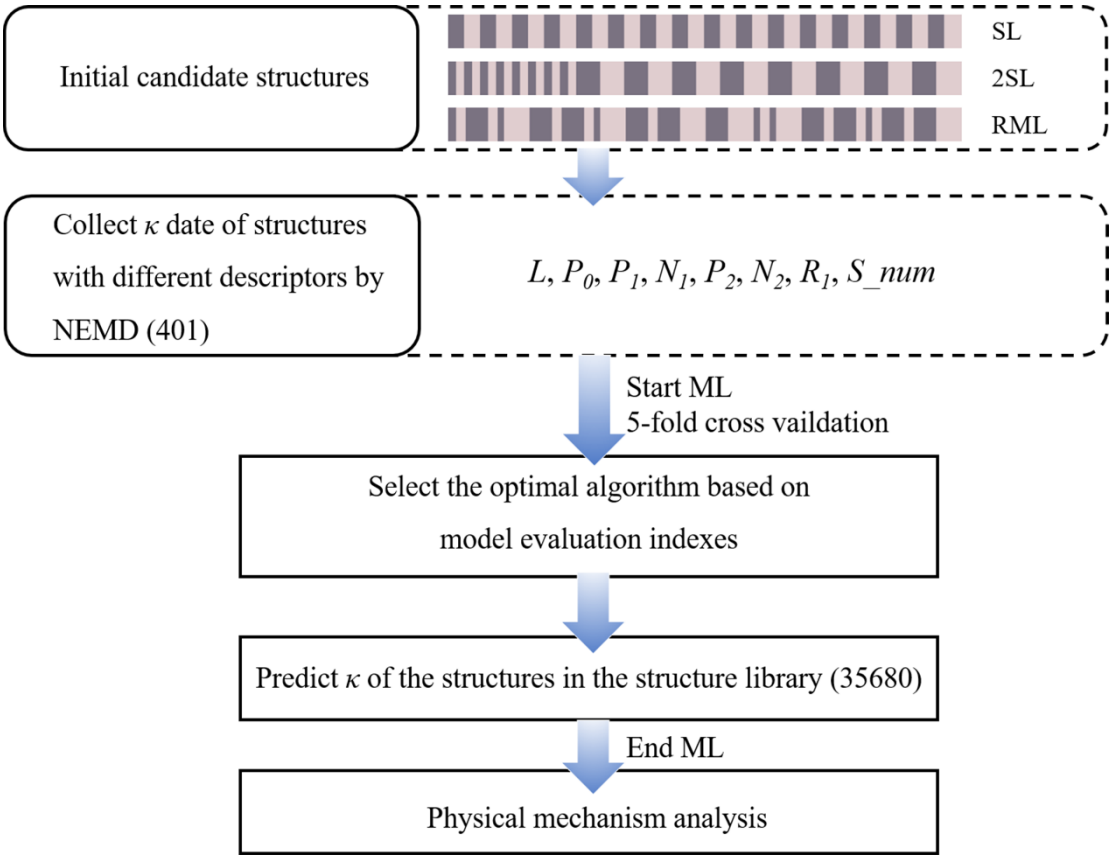

Figure S14 A framework for predicting κ_{eff} based on traditional ML regression algorithms.

References

1. M. Simkin and G. Mahan, Minimum thermal conductivity of superlattices, *Phys. Rev. Lett.* 84(5), 927 (2000)
2. P. K. Schelling, S. R. Phillpot, and P. Keblinski, Comparison of atomic-level simulation methods for computing thermal conductivity, *Phys. Rev. B* 65(14), 144306 (2002)
3. D. P. Sellan, E. S. Landry, J. Turney, A. J. McGaughey, and C. H. Amon, Size effects in molecular dynamics thermal conductivity predictions, *Phys. Rev. B* 81(21), 214305 (2010)
4. M. Hu and D. Poulikakos, Si/Ge superlattice nanowires with ultralow thermal conductivity, *Nano Lett.* 12(11), 5487 (2012)
5. X. Wang, M. Wang, Y. Hong, Z. Wang, and J. Zhang, Coherent and incoherent phonon transport in a graphene and nitrogenated holey graphene superlattice, *Phys. Chem. Chem. Phys.* 19(35), 24240 (2017)
6. Y. Wang, C. Gu, and X. Ruan, Optimization of the random multilayer structure to break the random-alloy limit of thermal conductivity, *Appl. Phys. Lett.* 106(7), 073104 (2015)
7. Y. Chen, D. Li, J. R. Lukes, Z. Ni, and M. Chen, Minimum superlattice thermal conductivity from molecular dynamics, *Phys. Rev. B* 72(17), 174302 (2005)
8. P. Hummel, A. M. Lechner, K. Herrmann, P. Biehl, C. Rössel, L. Wiedenhöft, F. H. Schacher, and M. Retsch, Thermal transport in ampholytic polymers: the role of hydrogen bonding and water uptake, *Macromol.* 53(13), 5528 (2020)

9. T. Ouyang and M. Hu, Competing mechanism driving diverse pressure dependence of thermal conductivity of X Te (X= Hg, Cd, and Zn), *Phys. Rev. B* 92(23), 235204 (2015)
10. X.-K. Chen, Z.-X. Xie, W.-X. Zhou, L.-M. Tang, and K.-Q. Chen, Phonon wave interference in graphene and boron nitride superlattice, *Appl. Phys. Lett.* 109(2), 023101 (2016)
11. S. Hu, Z. Zhang, P. Jiang, J. Chen, S. Volz, M. Nomura, and B. Li, Randomness-induced phonon localization in graphene heat conduction, *J. Phys. Chem. Lett.* 9(14), 3959 (2018)
12. P. Chakraborty, Y. Liu, T. Ma, X. Guo, L. Cao, R. Hu, and Y. Wang, Quenching thermal transport in aperiodic superlattices: A molecular dynamics and machine learning study, *ACS Appl. Mater. Interfaces* 12(7), 8795 (2020)
13. J. Maire, R. Anufriev, R. Yanagisawa, A. Ramiere, S. Volz, and M. Nomura, Heat conduction tuning by wave nature of phonons, *Sci. Adv.* 3(8), e1700027 (2017)
14. M. R. Wagner, B. Graczykowski, J. S. Reparaz, A. El Sachat, M. Sledzinska, F. Alzina, and C. M. Sotomayor Torres, Two-dimensional phononic crystals: Disorder matters, *Nano Lett.* 16(9), 5661 (2016)
15. B. Qiu, G. Chen, and Z. Tian, Effects of aperiodicity and roughness on coherent heat conduction in superlattices, *Nanoscale Microscale Thermophys. Eng.* 19(4), 272 (2015)
16. P. R. Chowdhury, C. Reynolds, A. Garrett, T. Feng, S. P. Adiga, and X. Ruan, Machine learning maximized Anderson localization of phonons in aperiodic superlattices, *Nano Energy* 69, 104428 (2020)
17. D. C. Montgomery, E. A. Peck, and G. G. Vining, Introduction to Linear Regression Analysis, John Wiley & Sons, 2012
18. R. Tibshirani, Regression shrinkage and selection via the lasso, *J. R. Statist. Soc. B* 58(1), 267 (1996)
19. A. E. Hoerl and R. W. Kennard, Ridge regression: Biased estimation for nonorthogonal problems, *Technometrics* 12(1), 55 (1970)
20. H. Zou and T. Hastie, Regularization and variable selection via the elastic net, *J. R. Statist. Soc. B* 67(2), 301 (2005)
21. G. K. Tso and K. K. Yau, Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks, *Energy* 32(9), 1761 (2007)
22. S. P. Curram and J. Mingers, Neural networks, decision tree induction and discriminant analysis: An empirical comparison, *J. Oper. Res. Soc. Am.* 45(4), 440 (1994)
23. A. Liaw and M. Wiener, Classification and regression by randomForest, *R news* 2(3), 18 (2002)
24. J. H. Friedman, Greedy function approximation: A gradient boosting machine, *Ann. Stat.* 29, 1189 (2001)
25. A. A. Abdullah, S. Rijal, and S. R. Dash, Evaluation on machine learning algorithms for classification of Autism Spectrum Disorder (ASD), *J. Phys.: Conf. Ser.* 1372(1), 012052 (2019)