Frontiers of Physics
https://doi.org/10.1007/s11467-020-1041-x

Front. Phys.
16(4), 43200 (2021)

TOPICAL REVIEW

Machine learning approach for the prediction and optimization of thermal transport properties

Yulou Ouyang[1,*], **Cuiqian Yu**[1,*], **Gang Yan**[1,2,†], **Jie Chen**[1,‡]

[1] *Center for Phononics and Thermal Energy Science, China–EU Joint Lab for Nanophononics, School of Physics Science and Engineering, Tongji University, Shanghai 200092, China*
[2] *Shanghai Institute of Intelligent Science and Technology, Tongji University, Shanghai 200092, China*
Corresponding authors. E-mail: [†]*gyan@tongji.edu.cn,* [‡]*jie@tongji.edu.cn*

Received October 7, 2020; accepted December 10, 2020

Traditional simulation methods have made prominent progress in aiding experiments for understanding thermal transport properties of materials, and in predicting thermal conductivity of novel materials. However, huge challenges are also encountered when exploring complex material systems, such as formidable computational costs. As a rising computational method, machine learning has a lot to offer in this regard, not only in speeding up the searching and optimization process, but also in providing novel perspectives. In this work, we review the state-of-the-art studies on material's thermal properties based on machine learning technique. First, the basic principles of machine learning method are introduced. We then review applications of machine learning technique in the prediction and optimization of material's thermal properties, including thermal conductivity and interfacial thermal resistance. Finally, an outlook is provided for the future studies.

Keywords machine learning, thermal transport, optimization, prediction

Contents

1 Introduction

A comprehensive and solid understanding of materials' thermal transport properties and mechanisms [1–5] is essential for a wide range of applications, from high-power density electronic devices [6, 7] to thermoelectrics [8–11], interfacial thermal management [12–15], energy storage [16, 17], thermal metamaterials [18], aerospace [19], and phonon engineering [20–28]. For example, materials with low thermal conductivity can be used for thermal barrier coatings [29, 30] and thermoelectric power generators [2, 31], while materials with high thermal conductivity can be used for heat dissipation [32, 33]. Therefore, accurate characterization of material's thermal conductivity is crucial.

Several advanced experimental techniques [34–40] have been developed to measure the intrinsic thermal conductivity of materials, yet challenges remain in experiments to provide improved control and characterization of complex structures, especially for low dimensional materials. On the other hand, theoretical and simulation tools [41–43], including first-principles calculations, Boltzmann transport equations (BTE), molecular dynamics (MD) simulations, lattice dynamics (LD) simulation, and non-equilibrium Green's function (NEGF) theory, can provide reliable predictions of thermal conductivity and physical insights to understand the underlying thermal transport mechanisms in materials.

The development of new materials synthesis technology and the fast-growing demand for the rapid and accurate prediction of physical properties require novel computational approaches. The machine learning (ML) method provides a promising solution to address such need. For

---

[*] Those two authors contributed equally to the work. Special Topic: Thermodynamics and Thermal Metamaterials (Editor: Ji-Ping Huang). This article can also be found at http://journal.hep.com.cn/fop/EN/10.1007/s11467-020-1041-x.

example, using a ML model based on support vector regression (SVR), Liu *et al.* [44] studied the thermodynamic properties of pure fluids and their mixtures, and found accurate prediction of the thermodynamics properties of fluids without prior knowledge of the underlying physical mechanisms. Wu *et al.* [45] used different ML models to predict the interface thermal resistance and found that the Bi/Si material system has high interfacial thermal resistance of $51.8\pm4.5 \times 10^{-9}$ $\mathrm{m^2{\cdot}K{\cdot}W^{-1}}$, which is in good agreement with the experimental measurement results. Zendehboudi *et al.* [46] demonstrated that a reliable ML model trained from 993 experimental data could exactly predict the effective thermal conductivity of 26 different classes of nanofluids. In general, the above-mentioned ML methods can predict the macroscopic thermal properties of materials based on data analysis, without the knowledge of underlying thermal transport mechanisms.

Moreover, previous studies have shown that ML potential combined with traditional transport calculation methods is an emerging avenue for predicting thermal conductivity of materials and even for probing underlying thermal transport mechanisms. For instance, by using spectral neighbor analysis potential, Gu *et al.* [47] studied the lattice dynamics and thermal transport of single-layer $\mathrm{MoS}_{2(1-x)}\,\mathrm{Se}_{2x}$ alloys, and further revealed the influence of mass disorder and force-field disorder on its thermal conductivity.

A recent burst of publications in applying ML methods to study thermal properties of materials suggest that associated techniques have become a powerful tool for predicting various physical properties [48–53]. Indeed, ML technique has been widely applied in many fields, such as structure prediction [52], materials design [49, 54], semiconductor manufacturing [55], and drug development [56].

In this review, we highlight the state-of-the-art advances of ML methods in prediction and optimization of the thermal transport properties. The rest of this article is organized as follows: In Section 2, we will introduce the basic principles of ML methods for thermal transport. In Sections 3, we will review the recent advances on the prediction of the intrinsic thermal conductivity of materials by using ML models and ML potentials. In Section 4, we will review the applications of ML models in predicting interfacial thermal resistance. The applications of ML methods in optimizing thermal related properties of materials will be discussed in Section 5. Finally, we present the conclusion and a brief outlook for future study in Section 6.

2 Machine learning method for thermal transport

Theoretical calculations and simulations are important tools to study the material's lattice thermal conductivity (κ_p), and serious computational challenges exist due to the enhancement of the complexity of materials and the demand for high-accuracy prediction. Approaches to improve the accuracy and reduce the computational cost are always a concern, and ML technique provides a novel avenue to address such need. In this section, we briefly introduce the framework of the ML methods used for the prediction of thermal conductivity.

2.1 The development of machine learning methods

In 1959, Samuel first defined Machine Learning as a "Field of study that gives computers the ability to learn without being explicitly programmed". Actually, the origin of ML can be traced back to the study of artificial neural networks (ANNs). In 1943, Mcculloch and Pitts [57] first proposed the hierarchical structure model of neural network and established the computational model theory of neural network. In 1957, Rosenblat [58] proposed the concept of perceptron, defined the mathematical models of self-organizing and self-learning neural networks with algorithms for the first time, and designed the original neural network algorithms. For a thorough introduction to the development of ML, interested readers can refer to Refs. [59–62]. So far, many ML algorithms have been developed that allow robots and computers to learn autonomously.

ML techniques can be roughly divided into three main categories: supervised learning, unsupervised learning and reinforcement learning [51, 63, 64]. In the supervised learning approach, the training data come from a labeled set of input (x_i)-output ($f(x_i)$) pairs, i from 1 to N, and N is the number of training samples. In supervised version, the goal of ML model is to produce a prediction y in response to a query x. However, for unsupervised learning approach, the training set is only composed of the unlabeled input (x_i) data, that is, there is no labeled sample available in the training set. This is sometimes called knowledge discovery [64]. The reinforcement learning trains an algorithm by interacting with its environment in order to maximize an objective of cumulative rewards, which can be used in inverse design of materials but is much less used to predict the thermal properties of materials. Most of the ML techniques used for the prediction of thermal conductivity utilize the supervised ML strategies.

2.2 Machine learning models

In this section, we will introduce the detailed process of using ML models to directly predict the thermal conductivity of complex or new materials based on experimental or theoretical calculation data. The main workflow of building excellent ML models comprises the following four steps: i) collection of training data, ii) feature generation, iii) training or learning process, iv) model optimization and validation. The scheme adopted to build ML models

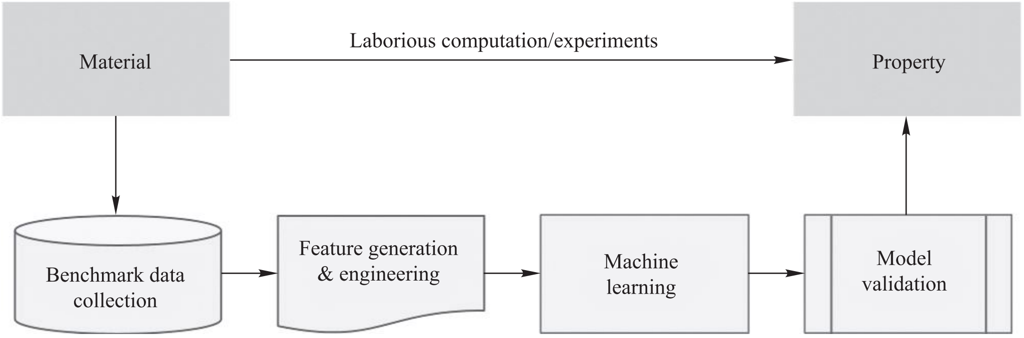

Fig. 1 Schematic workflow for the construction of machine learning models for the prediction of thermal conductivity. Reproduced with permission from Ref. [65].

of thermal conductivity is illustrated in Fig. 1. Next, we will introduce each step in detail to clearly illustrate the whole process.

First, one should select relevant and useful data as the input for training ML models. In the process of collecting training data, the main principle is to make the training data represent the complete diversity of reality as much as possible. Benefiting from the development of information storage technology, many research groups have built databases to collect information related to material properties, including both experimental and theoretical calculation data. Here, we briefly summarize some commonly used databases, such as the Cambridge Structural Database [66], Protein Data Bank [67], Crystallographic Open Database [68], Inorganic Crystal Structure Database (ICSD). [69], Atomwork [70], Materials Project Database [71] and aflowlib [72].

After selecting the training data, it is necessary to set up a feature (descriptor) dataset that not only represents the materials in training data, but also has the ability to describe the new materials. The quality of descriptor selection greatly affects the accuracy of the ML models. The descriptor usually includes atomic structures information, such as lattice constant, symmetry, and other physical and chemical properties. In previous studies, the most studied descriptor in these fields is related to the crystal structure fingerprints [73, 74], which can distinguish the local atomic environments in crystal materials. Besides, the elemental properties of materials are also important descriptors, such as the position of elements in the periodic table of elements, atomic mass, and atom number. Moreover, Wu *et al.* [45] found that heat capacity, density, melting point, unit cell volume, elastic modulus, bulk modulus and thermal expansion coefficient are also good descriptors for the prediction of the related thermal properties. In most situations, computational algorithms are often used to determine suitable descriptors [65, 75, 76], especially for structural features.

After selecting the dataset and descriptors, one needs to train the ML model through ML algorithm to obtain the relation between descriptors (input data) and the thermal conductivity (output data) in dataset. The whole dataset should be divided into 3 sub-groups for training, validation and testing purposes, respectively. The training data, including input and output data, are used in the learning process which is realized mainly by the ML algorithms to obtain the ML models parameters. The popular algorithms used for the prediction of physical properties of materials include Support Vector Machines (SVM) [77], ANNs [78], Decision Trees [79], Naive Bayes [80], Bayesian optimization [81], etc. More up-to-date and practical algorithms suitable for materials informatics can be found in Refs. [48, 51, 82–86]. Afterwards the ML models parameters are obtained by using training dataset and the validation data is used to optimize the hyper-parameters in ML models for the best performance. When the obtained ML model cannot successfully predict the thermal conductivity in the testing dataset, the previous steps should be repeated by improving the number of training dataset and feature selection, or even changing the ML algorithm altogether until the ML model succeeds.

2.3 Machine learning potentials

Although the ML models mentioned above can predict thermal conductivity value of materials without the knowledge of associated transport physics, they fail to provide the insight for the underlying thermal transport mechanisms, particularly for nanomaterials where interface and surface effects are important. Recently, many research groups have developed the ML potentials, which can provide more detailed physical insights of phonon transport in realistic materials, in addition to the prediction of thermal conductivity.

The overall workflow for predicting thermal conductivity with ML potentials can be briefly summarized in Fig. 2. First, one needs to build the ML potentials. In the construction of ML potentials, the system total energy and atomic force calculated by density-functional theory (DFT) calculation are selected as the training data. Besides, the atomic positions need to be transformed into a suitable set of descriptors by using ML algorithms as the input. It is extremely important that the descriptors should be useful for similar atomic structures. The

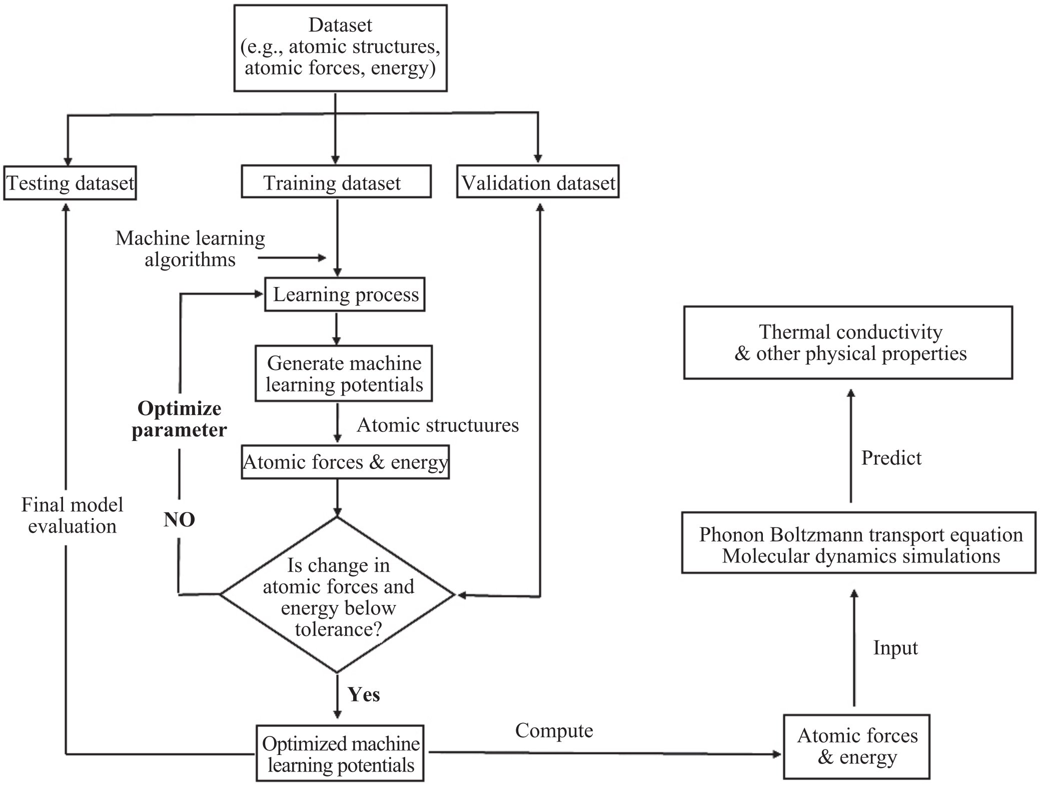

Fig. 2 Scheme for the generation of machine-learning potentials and its application in thermal transport calculations.

ML potentials can be obtained by using ML algorithms to map the input (descriptor) to output (total energy, atomic forces) dataset.

Then, the atomic force and system total energy can be computed directly from the obtained ML potentials. Finally, these data are used as input information for the thermal transport calculations, such as the phonon Boltzmann transport equations and molecular dynamics simulations, which can also provide physical understanding of the phonon transport mechanisms involved in the system.

It is worth noting that the computational cost of developing ML potentials is an important factor that should be considered, when using ML potentials combined with traditional calculation methods to predict the thermal properties of materials. These costs mainly come from the generation of training datasets that usually obtained by DFT calculations, and the process of tuning the hyper-parameters of algorithms.

For instance, Minamitani *et al.* [87] found that the generation of the dataset required 24 cores of the CPUs and 140.4 h for Si and 126.8 h for GaN, during the development of the high-dimensional neural network potential (HDNNP) for Si and GaN bulk crystals. Hyper-parameter tuning took 24 cores of the CPUs and ∼40 h for Si and ∼50 h for GaN. Their results show that HDNNP is not an efficient method for the prediction of thermal conductivity for perfect bulk crystals. However, it is a time-saving method for the systems with defects.

Therefore, when choosing an appropriate ML potential for each application, balancing the computational cost of data generation, the convenience of model training and the time of model evaluation is still an important consideration.

Generally speaking, in ML potentials, the algorithms for generating descriptors and training models can be different and be combined arbitrarily among each other. In short, ML potentials can establish the relationship between atomic structures and the potential energy surface (PES) which provides the potential energy of a system as a function of the atomic positions. Here, we review some ML potentials widely used in the study of thermal transport.

ANNs are traditional ML algorithms [88]. Behler and Parrinello [89] developed Neural-network potentials (NNPs) and applied it to MD simulation. Figure 3 shows the framework of Neural-networks used in NNPs to fit the PES. In the input layer, the two input nodes (G_i^1 and G_i^2) represent the generalized coordinates that determine the energy of configuration i. The output layer energy E_i is given by the expression [89]:

$$E_i = f_a^2\left[W_{01}^2+\sum_{j=1}^{3}W_{j1}^2 f_a^1\left(W_{0j}^1+\sum_{u=1}^{2}W_{uj}^1 G_i^u\right)\right], \tag{1}$$

where W_{ij}^k is the weight parameter connecting node j in layer k with node i in layer $k-1$, and W_{0j}^k is a bias weight

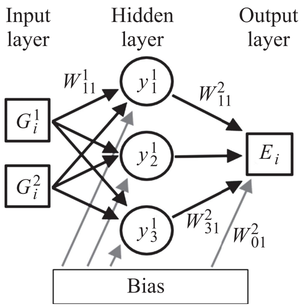

Fig. 3 Structure of a neural network. Reproduced with permission from Ref. [89].

used as an adjustable offset for the activation functions f^k_a.

Thompson *et al.* [90] presented the Spectral Neighbor Analysis Potential (SNAP) for solids and liquids. In the SNAP, the energy of the system is given as the sum of a reference energy E_{ref} and a local energy E_{local} [90]:

$$E_{tot} = E_{ref} + E_{local} = E_{ref} + \sum\nolimits_{n=1} E_n, \tag{2}$$

where E_n is the local energy of atom n. The reference energy includes long-range electrostatic interactions, which can be obtained by the well-established methods, such as Ewald summation [91] and Particle–Particle Particle–Mesh method [92]. The atomic energy of atom n is then given as a linear combination of the M projected bispectrum components by [90]

$$E_n = \beta_{0\alpha_n} + \sum_{m=1}^{M} \beta_{m\alpha_n} B_{mn}, \tag{3}$$

where α_n is the chemical identity of atom n and $\beta_{m\alpha_n}$ is the coefficients. B_{mn} is the bispectrum component m of atom n, and $\beta_{0\alpha_n}$ is a constant element-specific contribution.

The Gaussian Approximation Potential (GAP) was introduced by Csányi *et al.* [93, 94]. In GAP, the local energy E_n for atom n is given by a linear combination of the kernel functions:

$$E_n = \sum_l \alpha_l K(\boldsymbol{q}_n, \boldsymbol{q}_l), \tag{4}$$

where the summation over l includes all the atomic neighborhood environments in the training datasets. The α_l is a weighting factor obtained during the fitting process, and $K(\boldsymbol{q}_n, \boldsymbol{q}_l)$ is the kernel function that can quantify the similarity between any two atomic neighborhood environments.

3 Prediction of thermal conductivity

Recent advances in ML have shown that thermal conductivity can be well predicted by using ML methods with the accuracy comparable to experimental measurements or DFT calculations. Recently, Zhang *et al.* [54] reviewed the research progress in the screening of thermal conductivity with ML and high-throughput methods. Besides, Wan *et al.* [95] introduced recent progress in predicting thermal transport properties of materials by using materials informatics. Up to now, there have been increasing studies to investigate thermal conductivity based on ML methods. In this section, we summarize the latest works for the prediction of material's intrinsic thermal conductivity by using ML methods.

Table 1 Four materials with low thermal conductivity screened by machine learning. Reproduced from Ref. [96].

Formula	ICSD number	Space group	κ_L^{pre} (W·m^{-1}·K^{-1})	κ_L^{cal} (W·m^{-1}·K^{-1})
$BiTe_2Tl$	617200	166	0.76	0.2
$Br_2Cs_2F_2$	84022	139	0.49	0.55
Au_3CsSe_2	82542	164	0.59	0.32
Cl_2CsI	23982	166	0.58	0.22

3.1 Prediction by machine learning models

The ML methods are widely employed for the prediction of thermal conductivity in various systems. One of the primary uses of ML models in thermal properties research is to discover the materials with low thermal conductivity for the application of thermoelectrics and thermal insulation. Moreover, the ML methods can be used to predict the thermal conductivity of complex systems, such as nuclear fuels, nanofluids, composite materials, alloys, inorganic materials and porous media, which usually require a lot of experimental or computational costs.

As mentioned in Section 2, the success of ML models in predicting thermal conductivity requires the selection of an appropriate algorithm as well as descriptors. Very recently, based on the materials in ICSD, Wang *et al.* [96] trained four different ML models (SVR, kernel ridge regression, eXtreme Gradient Boosting (XGBoost), fully connected neural networks) to predict the thermal conductivity of crystal materials for the search of materials with low thermal conductivity. They found that the XGBoost model has the best performance among the four models with lowest mean absolute error, 0.975 and 0.902 for train set and test set, respectively. Using the XGBoost model, Wang *et al.* [96] found that some compounds ($BiTe_2Tl$, $Br_2Cs_2F_2$, Au_3CsSe_2, Cl_2CsI) from 54 790 compounds show low thermal conductivity less than 0.8 W·m^{-1}·K^{-1} which was also validated by the first-principles method, as shown in Table 1. These results show that the ML method is useful in the rapid search for materials with low thermal conductivity in a large number of materials.

In the study of thermal conductivity of aluminum alloys, by using maximal information coefficient (MIC) and conventional Pearson's correlation coefficient (PCC) data analysis techniques, Wang *et al.* [76] analyzed the relationship between precipitate features and effective thermal conductivity (κ_{eff}) of aluminum alloys. They found that the precipitate morphology parameters, such as surface area, average diameter, and volume fraction of precipitates, can serve as excellent descriptors to evaluate the κ_{eff} of the aluminum alloy. Moreover, their result shows that the precipitate features can be used to quickly and precisely predict the κ_{eff} of alloys through ML models, without the need for separate theoretical simulation or experimental measurement. Therefore, identifying the key features that affect the physical properties of interest is very important when training ML models.

Features used in ML models should be a common property of each material that is easy to obtain. Chen *et al.* [65] found that bulk modulus (B), mean average bond length, mean atomic mass, and $\sqrt{B/\rho}$ ($\rho:$ density) are key features controlling the κ_p in inorganic materials, showing a strong correlation with logκ_p. Based on these features, they trained the ML models with the κ_p of 100 inorganic materials as input. Their results show that the accuracy of the developed ML model is comparable to that of past semi-empirical models in the prediction of lattice thermal conductivity. Moreover, Juneja *et al.* [97] adopted maximum phonon frequency, integrated Grüeisen parameter up to 3 THz, average atomic mass, and volume of the unit cell as the descriptors to develop a ML model for lattice thermal conductivity. Their prediction results of log-scaled κ_p show excellent accuracy with a very small root mean square error of ~0.21, which is one order of magnitude improvement in accuracy compared to the widely used Slack model.

In addition to the prediction of the thermal conductivity according to the microstructure and physical characteristics of materials, ML models also have the capability to establish the relationship between experimental operating conditions and desired thermal conductivity, which can improve the probability of success for the experimental test to achieve the target properties. For instance, a ML model developed by Kautz *et al.* [98] was used to predict the thermal conductivity of irradiated nuclear fuel at various irradiation test parameters, including major alloying element concentration, depletion, fission density, fission power, surface heat flux, neutron flux, and temperature. Therefore, in this way, one can further establish the structure-property-processing performance relationships in the system of interest.

More interestingly, Fujii *et al.* [99] have quantified the relationship between thermal conductivity of MgO grain boundary structure and local atomic distortions by using ML model with a multi-dimensional dataset. The ML model developed by Fujii *et al.* [99] reveals that small structural distortions to the lattice are sufficient to reduce thermal conductivity by grain boundary. Their results demonstrate that ML method is a good choice to determine quantitatively the correlation between the microstructure and material's property at the atomic level.

However, a large number of quantified thermal conductivity values are often required as the training data to train a ML model for the prediction of thermal conductivity, which needs huge experimental or computational resources and thus is a serious challenge. Interestingly, Ju *et al.* [100] developed a feature-based transfer learning method to search high thermal conductivity materials from 60 000 crystal compounds, in which only a small amount of thermal conductivity training data is required. They first chose the scattering phase space (P_3) as a key feature for characterizing thermal conductivity and established a pre-trained ML model connecting the crystal structures and P_3. Using the subnetwork of the pre-trained ML model, they then trained a transferred ML model that can obtain the relation between crystal structures and the thermal conductivity based on 45 thermal conductivity data.

Moreover, high-throughput screening based on ML is an effective way to overcome such problem and accelerate the prediction of thermal conductivity. For example, using a set of descriptors and random-forest regression, Carrete *et al.* [79] built an effective classification model to prescreen the material with low thermal conductivity from 79 000 half-Heusler compounds. Then, they performed full DFT calculations over the reduced candidate compounds and found that ordered half-Heusler compounds with a low thermal conductivity $<3\ \mathrm{W\cdot m^{-1}\cdot K^{-1}}$ very likely exist. Base on the thermal conductivity of 101 compounds, by using Bayesian optimization, Seko *et al.* [101] virtual screened 54779 compounds and found 221 materials with very low thermal conductivity. In addition, through an iterative ML scheme using principal component analysis and regression, Roekeghem *et al.* [102] extrapolated interatomic force constants (IFCs) at finite temperatures from a few IFCs values to predict the thermal conductivity at finite temperatures. These works demonstrate the great potential of ML method in accelerating the prediction of thermal conductivity.

3.2 Prediction by machine learning potentials

Different from the prediction of the thermal conductivity value by using ML models, ML potentials combined with the BTE or MD simulations can provide detailed phonon information to better understand the thermal transport mechanisms. In sharp contrast to the empirical potential, ML potentials do not make any prior assumptions on the form of the interatomic potentials, and have no adjustable empirical parameter either. Once ML potential has been fitted, it can accurately predict energies and atomic forces for large systems according to the atomic positions. Therefore, using ML potentials can significantly reduce the expense of quantum–mechanical calculations in generating interatomic force constants for large and com-

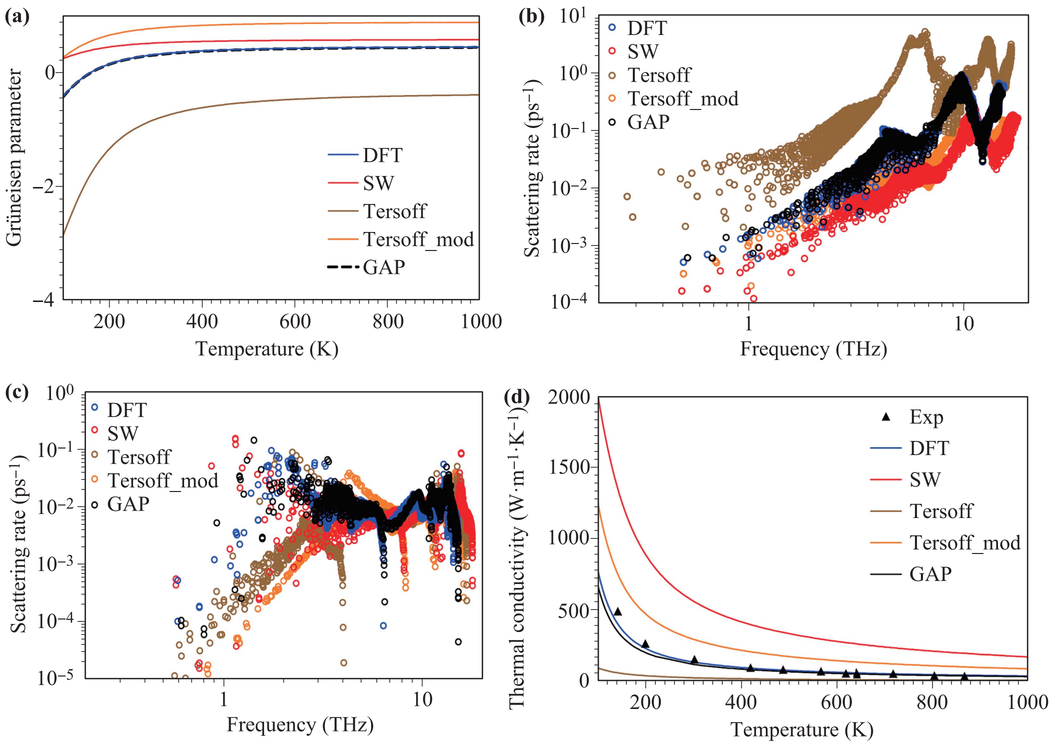

Fig. 4 **(a)** Grüneisen parameter of crystalline silicon at different temperature from empirical potentials, ML potentials and DFT. **(b)** Three-phonon scattering rates of crystalline silicon as a function of frequency at 300 K. **(c)** Phonon-vacancy scattering rates at 300 K assuming a volumetric vacancy concentration of 0.01%. **(d)** Thermal conductivity of crystalline silicon as a function of temperature. Reproduced with permission from Ref. [104].

plex structures.

Recently, ML potentials have received increasing attention. By using MD simulations with the GAP potentials, Zhang *et al.* [103] calculated the thermal conductivity of silicone, and found good agreement with results from DFT calculations. Moreover, they found that the effect of high-order anharmonic scatterings on the thermal conductivity of silicene is negligible. By learning the forces generated by DFT calculations, Minamitani *et al.* [87] developed the HDNNP for Si and GaN bulk crystals. They found that HDNNP has the ability to predict the lattice thermal conductivity of semiconducting materials with the same accuracy as DFT calculations.

In addition to the crystalline materials, ML potentials can also be employed to predict the thermal conductivity of disordered crystalline phases and amorphous materials, which is in general a formidable task for DFT calculations due to the disordered structures. By using MD simulation with GAP potentials, Babaei *et al.* [104] found GAP potential can predict various thermal transport properties of both perfect crystalline Si and crystalline Si with vacancies in excellent agreement with the results from DFT calculations, including the Grüneisen parameter, three-phonon scattering rate, phonon-vacancy scattering rate and thermal conductivity, as shown in Fig. 4. More significantly, for the calculation of atomic forces, the computational cost of the GAP is five orders of magnitude lower than that of the DFT calculations [104]. Moreover, Qian *et al.* [105] developed the GAP for predicting thermal conductivity of both crystalline and amorphous silicon, as shown in Figs. 5(a, b). The thermal conductivity of amorphous silicon obtained from equilibrium molecular dynamics with GAP is ~1.4 W·m^{-1}·K^{-1} at room temperature [Fig.5(b)], within the range of measurement values [106, 107] around 1–2 W·m^{-1}·K^{-1}.

In general, it is difficult to calculate the thermal conductivity of complex or disordered structure by using DFT calculations, due to the large number of atoms in primitive cell and the expensive quantum-mechanical calculation of high-order interatomic force constants. In addition, due to the low accuracy of empirical potentials, the MD simulation with empirical potentials cannot accurately reproduce phonon scattering and ensure long-time simulation of complex structures. These obstacles seriously limit the application of both DFT calculation and MD simulation in studying thermal transport in complex materials.

Combining MP potentials with the BTE or MD simulations for the prediction of thermal conductivity can well solve the above-mentioned difficulties. For instance, Korotaev *et al.* [108] developed a ML potential named moment tensor potential (MTP) for $CoSb_3$ skutterudite. Based on MTP, they used MD simulation and BTE method to

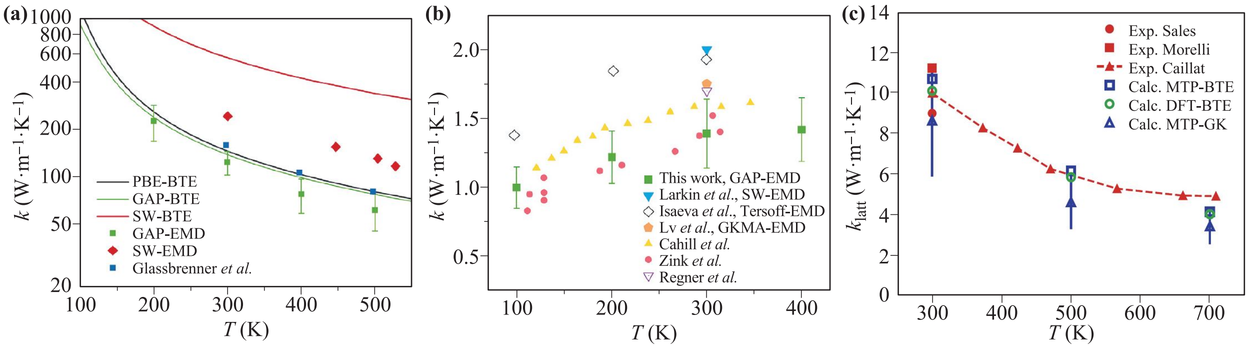

Fig. 5 Thermal conductivity of **(a)** crystalline silicon **(b)** amorphous silicon derived from the GAP compared with DFT calculations and experimental measurements. **(c)** The results of thermal conductivity of $CoSb_3$, open symbols represent the calculations data, filled symbols represent the experimental data. (a, b) Reproduced with permission from Ref. [105], (c) Reproduced with permission from Ref. [108].

calculate the thermal conductivity of $CoSb_3$, as shown in Fig. 5(c). Their results demonstrate that the thermal conductivity of skutterudite predicted with ML potentials has the accuracy comparable to that of DFT calculations. In other words, ML potentials are among the best options to study the intrinsic thermal transport properties of materials with high accuracy.

4 Prediction of interfacial thermal resistance

Interface thermal transport between two different materials is a key issue in nanoelectronic devices. Due to various factors affecting the interfacial thermal resistance (ITR), it is important to accurately predict ITR for the design of the thermal management in nanoelectronic devices. With the development of materials informatics [109], the ML based prediction of the interfacial thermal properties can help to greatly accelerate the process of designing high-performance thermal interface materials. To identify interfaces with ultralow and ultrahigh ITR, Zhan *et al.* [110] used ML approach to scan over 368 interfaces formed by 45 different materials. Four kinds of ML algorithms were used in their study, including generalized linear regression (GLR), least absolute shrinkage and selection operator regularization (LASSOR), Gaussian process regression (GPR), and SVR. They found that among four algorithms, the GPR and the SVR have better accuracy. Moreover, the choice of descriptors also affects the prediction accuracy. In their study, the film thickness is found to be an important descriptor in the prediction of ITR by comparing the prediction results among different descriptor sets. Moreover, their results demonstrate that the ML results can match the experimental values better than that of the traditional acoustic mismatch model and the diffuse mismatch model.

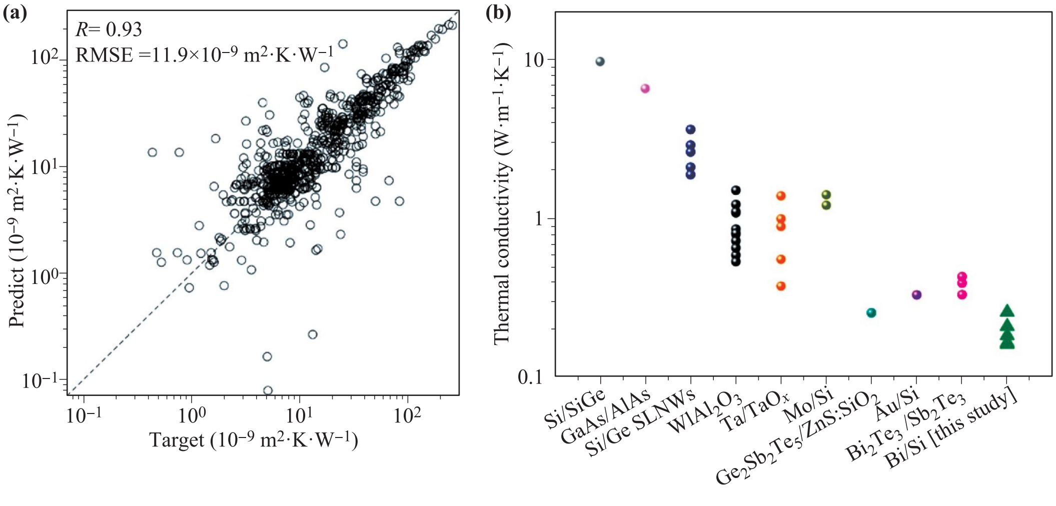

Fig. 6 **(a)** The mismatch between target result from the literature data and the prediction by the ML model. **(b)** A comparison of the thermal conductivity between Bi/Si nanocomposites and other inorganic material systems. Reproduced with permission from Ref. [111].

Besides, the properties predicted by various ML methods can be verified experimentally. By introducing ML methods, the Bi/Si nanocomposites with ultralow thermal conductivity have been discovered and experimentally verified by Wu *et al.* [111]. They used ML model based on the algorithms of regression tree ensembles of Least-squares boosting (simplified as LSBoost) to screen out potential candidates from hundreds of interface material systems. The mismatch between the literature data and the prediction is shown in Fig. 6(a), demonstrating that the mismatch is small and ML method is suitable for the prediction of ITR. Their results showed that the Bi/Si interface selected from 2025 kinds of interfaces systems exhibits a high ITR and can be used to design nanostructures with low thermal conductivity. In order to verify the accuracy of ML methods in predicting the ITR of Bi/Si interface, Wu *et al.* [111] prepared the Bi/Si nanocomposites and measured their thermal conductivity. As shown in Fig. 6(b), Bi/Si has the lowest thermal conductivity compared with other materials systems. These results show that ML models have great potential in predicting the ITR and can promote the discovery of potential systems with low thermal conductivity.

Selecting suitable descriptor is crucial for improving the accuracy of ML prediction. For instance, Wu *et al.* [45] chose the property descriptors, compound descriptors, and process descriptors for the prediction of ITR, which could represent the physical, chemical, and material properties of ITR, respectively. As listed in Table 2, for various ML models considered in their study (LSBoost, SVM, and GPR), the predictive performance with all three descriptors sets is better than that with only the thickness and property descriptors. However, the physical origin for the relationship between these physical quantities and ITR was not discussed, which deserves further investigation to explore the underlying physics.

Table 2 The predictive performance evaluation of the correlation coefficient (R), the square of the correlation coefficient (R^2) and the root-mean-squared error (RMSE) by various algorithms. Reproduced from Ref. [45].

Algorithms	R	R^2	RMSE
All descriptors			
LSBoost	0.958	0.919	8.944
SVM	0.938	0.879	10.897
GPR	0.957	0.916	9.073
Property descriptors & thickness			
LSBoost	0.952	0.907	9.575
SVM	0.815	0.664	18.171
GPR	0.934	0.871	11.271

After achieving high-accuracy prediction, Wu *et al.* [45] screened out the top-100 high-ITR material systems over 80 282 material systems (excluding metal/metal interface) by using LSBoost, GPR, and SVM models, respectively. They found that 25 of the top-100 material systems were repeatedly predicted by at least two ML models, as shown in Fig. 7(a). The ITR values of the above 25 material systems predicted by the three algorithms are shown in Fig. 7(b). The predicted ITR value of Bi/Si interface by ML models is in good agreement with the experimental measurements.

In addition to the choice of proper descriptors, another key issue for the ML based predictions is the completeness of the dataset. In this regard, Wu *et al.* [112] proposed a dataset which can be used to predict ITR of metal/nonmetal interfaces by using ML technique. As shown in Fig. 8, they divided the whole dataset into two parts: one part is the data about the material properties of interface thermal resistance in the literature and the other part was the physical, chemical, and process descriptors of 298 different materials. They made this database available to the public, and elaborated the data collection methods and

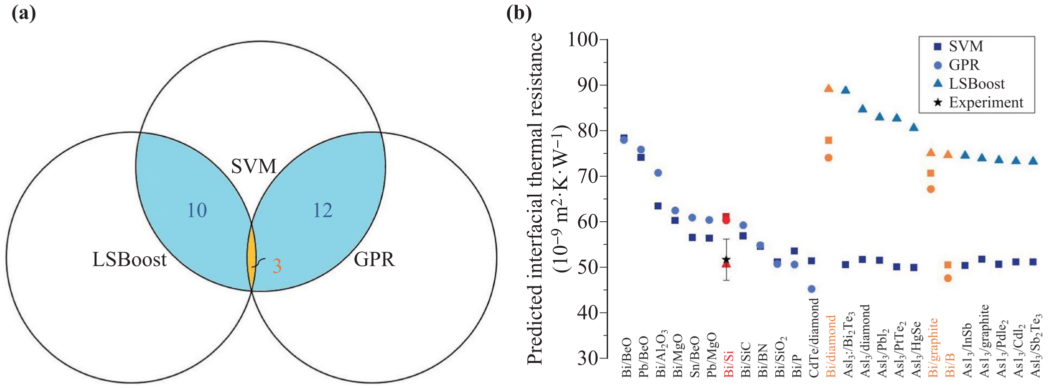

Fig. 7 **(a)** Each circle represents the top-100 high-ITR material systems predicted by LSBoost, SVM, GPR respectively. The blue and yellow regions represent repeated predictions by two and three models, respectively. **(b)** The ITR predicted by the three models of SVM (square), GPR (circle), and LSBoost (triangle). The orange points represent the material systems repeated predicted by the three algorithms. The red points show the predicted values of Bi/Si system. The black point means that the experimental value of Bi/Si system in Ref. [111]. Reproduced with permission from Ref. [45].

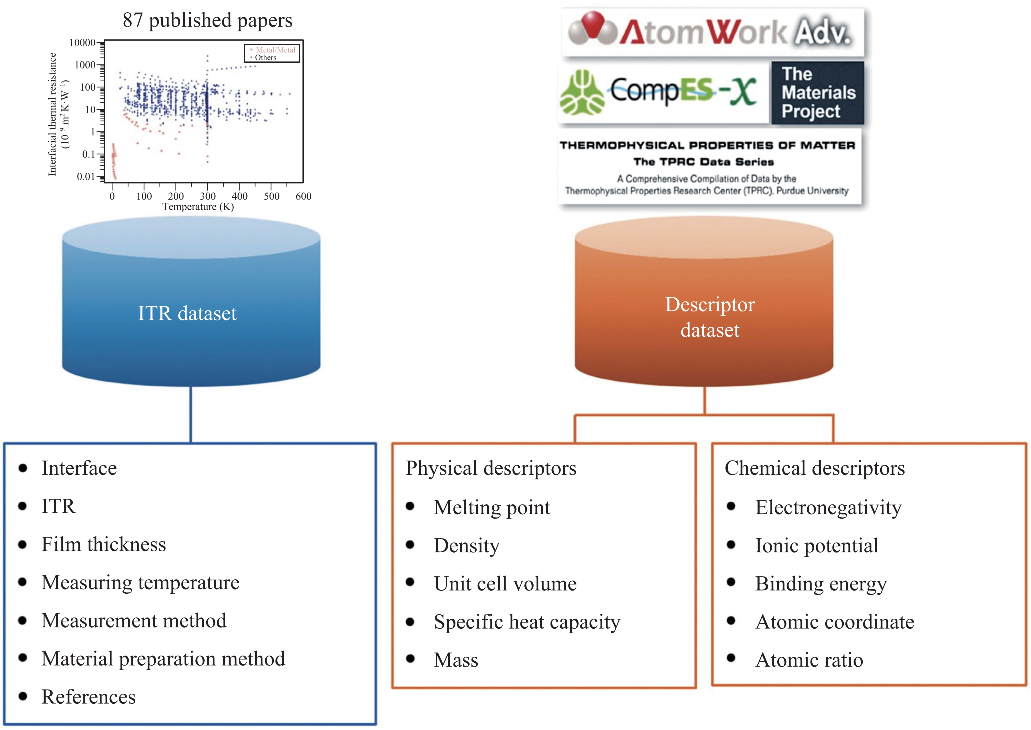

Fig. 8 The contents contained in the dataset. The ITR dataset includes literature data. The descriptor dataset is comprised of the data about the physical and chemical properties of different materials. Reproduced with permission from Ref. [112].

sources of various descriptors' data. The availability of this data set can provide a complete set of descriptors for future work, thus accelerating the development of thermal management materials.

In addition to the collection from literature studies, the training data of ML models can also be obtained by high-throughput computations. Yang *et al.* [113] collected thousands of training data points by classical molecular dynamics simulations. Taking the system temperature, coupling strength between two layers, and in-plane tensile strains as the input, they predicted ITR between graphene and hexagonal boron nitride by training several ML models. They found that the predictions by ANN model were the best. Therefore, ML technique can help to break through the limitation of traditional methods which can only study the properties of materials in a few dimensions, and allows for the exploration of more exotic properties in complex material systems.

5 Optimization of thermal related properties of materials

In addition to the prediction of physical properties, the optimization of the physical properties of materials is also very important for the design of materials with targeted properties. The ML approach is an effective tool to design functional materials in a large structure space. For instance, Ju *et al.* [114] introduced in detail how to use ML combined with materials informatics and thermal transport calculations to design materials with target properties for various applications, such as thermoelectrics and thermal radiation. Here we review the recent advances on the optimization of thermal related properties based on ML.

To get the maximum or minimum interfacial thermal conductance sandwiched between Si-Si segments and Si-Ge segments, Ju *et al.* [81] developed a method combining atomistic Green's function and Bayesian optimization. The sandwiched interface region can be constructed in the arbitrary concentration of Si or Ge in the diamond lattice structure. This method can greatly save computational resources to identify the optimal structure, compared to traditional simulation method. As a result, by optimizing the structure, the influence of interference was maximized so as to minimize the thermal conductance of the interface. By introducing the optimal hole distribution into the porous graphene, the thermal conductivity of porous graphene can be tuned to a maximum or minimum value. Furthermore, Wan *et al.* [115] used a reverse design method based on convolutional neural network model to determine the optimal porous graphene structure with the lowest thermal conductivity. The searching algorithm

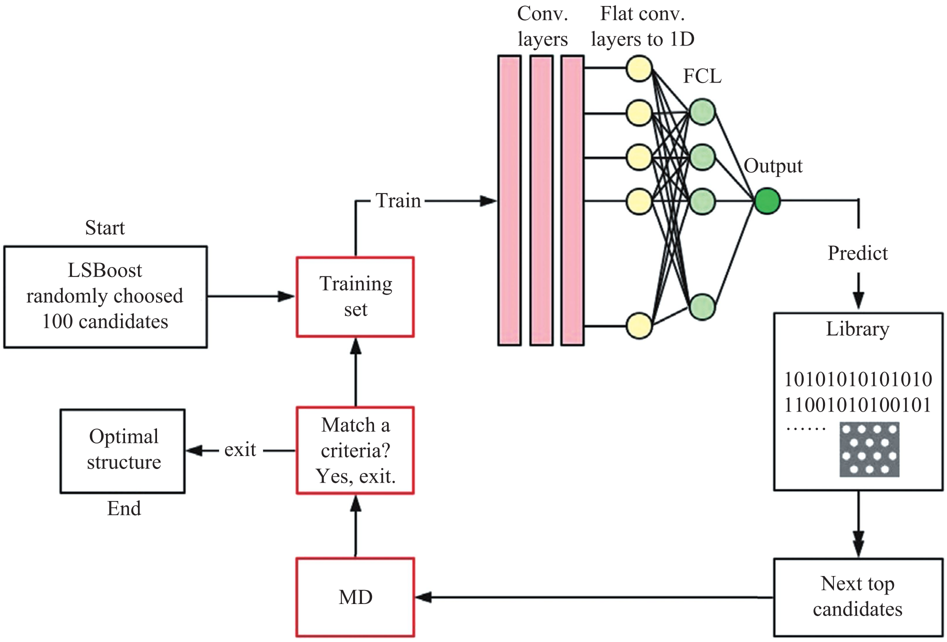

Fig. 9 Schematic diagram of the searching algorithm based on the convolutional neural network model. Reproduced with permission from Ref. [115].

used in their work is shown in Fig. 9. The training set expands as the number of iterations increases. When the targeted criterion is matched, the search process is completed and the optimized porous graphene structure can be obtained. The advantage of ML over traditional methods is that it can traverse the whole design space in a very efficient way to obtain the optimal material structure.

The aperiodic superlattice structure has a lower thermal conductivity compared to the superlattice counterpart, because aperiodic superlattice structure can induce the localization of coherent phonons [27, 116, 117]. But the lowest thermal conductivity remains elusive because of the large design space. Chakraborty *et al.* [118] used ML to accelerate the design process of the multi-layer structure with ultra-low thermal conductivity by reasonably defining the disorder of aperiodic superlattice structure. They used the thickness based index of randomization and period based index of randomization as two key parameters to quantify the degree of disorder in aperiodic superlattice structure. By introducing these randomness parameters into a neural network based ML model, they identified the optimized multilayer structures with the lowest thermal conductivity and strong phonon Anderson localization.

Similarly, in order to search for the minimum value of thermal conductivity in Si/Ge random multilayer structure, Chowdhury *et al.* [119] compared the genetic algorithm based approach with an intuition-guided manual search. They found that the traditional intuition-guided manual search could only identify the local minimum thermal conductivity, but the genetic algorithm could efficiently find the global minimum thermal conductivity. They have shown that the global minimum thermal conductivity appears in Si/Ge aperiodic superlattice structures with an average superlattice period which is smaller than the period length at which the minimum thermal conductivity occurs in the superlattice structure. Their work demonstrates that by studying the thermal properties of random multilayer structures through ML technique, the barrier of large design space can be resolved, and the optimal structure can be identified efficiently.

Using Bayesian optimization technique, Yamawaki *et al.* [120] optimized the thermoelectric properties of graphene nanoribbons (GNRs). They found that the thermoelectric figure of merit for optimized GNRs structure can reach 11 times that of the pristine GNRs. In addition to the optimization of the lattice structure, the variation of chemical composition is another way to tune the thermoelectric properties of materials. Hou *et al.* [121] employed the GPR model to enhance the thermoelectric power factor by tuning the composition of Al and Si in the $Al_2Fe_3Si_3$ compound. The optimal composition of Al and Si for high power factor over the intermediate temperature range was found efficiently. The power factor for the optimal composition at ~510 K was about 670 $\mu W\cdot m^{-1}\cdot K^{-2}$. Their results demonstrate that ML method can provide effective guidance for optimizing thermoelectric properties of materials.

Very recently, the work of Sakurai *et al.* [122] demonstrated that Bayesian optimization is an effective way to design thermal radiation materials. Guo *et al.* [123] used the Bayesian optimization to optimize the radiative cooling properties of thermal photonic structures. With the help of machine learning, they found that only less than 1% of the total candidates needed to be accurately evaluated in order to find the structure with optimal thermal radiative property. Besides, many other properties of materials can also be optimized by using ML methods, such as decomposition energies [124], Li-ion conductivity [125], afterglow lifetime in phosphorescent material [126], phase design for high entropy alloys [127], stress-strain curves [128] and hardness [129]. Therefore, ML has great potential in the studies related to the structure search and material property optimization.

6 Conclusion

As the complexity of computing tasks increases, machine learning, which is an accurate and cost-effective approach, is becoming increasingly popular. Machine learning method plays an important role in guiding chemical synthesis, assisting multi-dimensional material characterization, and acquiring new material design methods. In this work, we review the basic principles of machine learning methods, and the recent advances in applying machine learning approach to the prediction and optimization of thermal transport properties, such as thermal conductivity and interfacial thermal resistance. These advances in this field demonstrate that machine learning has become an important tool for the study of thermal transport in complex materials, and can provide valuable insights to the material informatics.

Despite the inspiring successes, there still lacks a comprehensive database for the thermo-physical properties of materials. Meanwhile, there is a gap in most studies between theoretical calculations and experimental measurements for mutual verifications. In addition, in most ML algorithms, there will be a problem of underfitting, if the size of material descriptor dataset is small. Moreover, when dealing with problems with multiple degrees of freedom, it is necessary to develop novel and effective descriptors. The selection criteria of suitable descriptors are also important aspects of machine learning approach. As a reasonable set of descriptors is the key to the accurate predictions, a more rigorous method for descriptor analysis is needed. These challenges need to be addressed in the future development of machine learning, which will be the driving force behind machine learning's work on thermal transport for decades to come.

Acknowledgements This project was supported in part by the grants from the National Natural Science Foundation of China (Grant Nos. 12075168 and 11890703), the National Key Research and Development Program of China (Grant No. 2017YFB0406000), Science and Technology Commission of Shanghai Municipality (Grant Nos. 19ZR1478600, 18JC1410900, and 18ZR1442000), the Fundamental Research Funds for the Central Universities (Grant No. 22120200069), and Open Fund of Hunan Provincial Key Laboratory of Advanced Materials for New Energy Storage and Conversion (Grant No. 2018TP1037_201901).

References

1. Z. Zhang, Y. Ouyang, Y. Cheng, J. Chen, N. Li, and G. Zhang, Size-dependent phononic thermal transport in low-dimensional nanomaterials, *Phys. Rep.* 860, 1 (2020)
2. X. Xu, J. Zhou, and J. Chen, Thermal transport in conductive polymer-based materials, *Adv. Funct. Mater.* 30(8), 1904704 (2020)
3. Z. Zhang and J. Chen, Thermal conductivity of nanowires, *Chin. Phys. B* 27(3), 035101 (2018)
4. X. Xu, J. Chen, J. Zhou, and B. Li, Thermal conductivity of polymers and their nanocomposites, *Adv. Mater.* 30(17), 1705544 (2018)
5. Z. Zhang, Y. Ouyang, Y. Guo, T. Nakayama, M. Nomura, S. Volz, and J. Chen, Hydrodynamic phonon transport in bulk crystalline polymers, *Phys. Rev. B* 102(19), 195302 (2020)
6. R. J. McGlen, R. Jachuck, and S. Lin, Integrated thermal management techniques for high power electronic devices, *Appl. Therm. Eng.* 24(8–9), 1143 (2004)
7. S. G. Kandlikar and C. N. II Hayner, Liquid Cooled Cold Plates for Industrial High-Power Electronic Devices—Thermal Design and Manufacturing Considerations, *Heat Transf. Eng.* 30(12), 918 (2009)
8. J. P. Rojas, D. Singh, S. B. Inayat, G. A. T. Sevilla, H. M. Fahad, and M. M. Hussain, Review — Micro and nano-engineering enabled new generation of thermoelectric generator devices and applications, *ECS J. Solid State Sci. Technol.* 6(3), N3036 (2017)
9. Y. Ouyang, Z. Zhang, D. Li, J. Chen, and G. Zhang, Emerging theory, materials, and screening methods: New opportunities for promoting thermoelectric performance, *Ann. Phys. (Berlin)* 531(4), 1800437 (2019)
10. L. Shi, J. Chen, G. Zhang, and B. Li, Thermoelectric figure of merit in Ga-doped [0001] ZnO nanowires, *Phys. Lett. A* 376(8–9), 978 (2012)
11. X. L. Zhu, P. F. Liu, J. Zhang, P. Zhang, W. X. Zhou, G. Xie, and B. T. Wang, Monolayer SnP_3: An excellent p-type thermoelectric material, *Nanoscale* 11(42), 19923 (2019)
12. D. Alexeev, J. Chen, J. H. Walther, K. P. Giapis, P. Angelikopoulos, and P. Koumoutsakos, Kapitza resistance between few-layer graphene and water: Liquid layering effects, *Nano Lett.* 15(9), 5744 (2015)
13. Y. Ma, Z. Zhang, J. Chen, K. Sääskilahti, S. Volz, and J. Chen, Ordered water layers by interfacial charge decoration leading to an ultra-low Kapitza resistance between graphene and water, *Carbon* 135, 263 (2018)
14. X. Liu, J. Gao, G. Zhang, and Y. W. Zhang, Design of phosphorene/graphene heterojunctions for high and tunable interfacial thermal conductance, *Nanoscale* 10(42), 19854 (2018)

15. J. Wang, Z. Zhang, R. Shi, B. N. Chandrashekar, N. Shen, H. Song, N. Wang, J. Chen, and C. Cheng, Impact of nanoscale roughness on heat transport across the solid–solid interface, *Adv. Mater. Interfaces* 7(4), 1901582 (2020)
16. M. Zhou, H. Bi, T. Lin, X. Lü, D. Wan, F. Huang, and J. Lin, Heat transport enhancement of thermal energy storage material using graphene/ceramic composites, *Carbon* 75, 314 (2014)
17. F. Gong, Z. Ding, Y. Fang, C.J. Tong, D. Xia, Y. Lv, B. Wang, D. V. Papavassiliou, J. Liao, and M. Wu, Enhanced electrochemical and thermal transport properties of graphene/MoS_2 heterostructures for energy storage: Insights from multiscale modeling, *ACS Appl. Mater. Interfaces* 10(17), 14614 2018)
18. C. Z. Fan, Y. Gao, and J. P. Huang, Shaped graded materials with an apparent negative thermal conductivity, *Appl. Phys. Lett.* 92(25), 251907 (2008)
19. J. Doty, K. Yerkes, L. Byrd, J. Murthy, A. Alleyne, M. Wolff, S. Heister, and T. S. Fisher, Dynamic thermal management for aerospace technology: Review and outlook, *J. Thermophys. Heat Transfer* 31(1), 86 (2017)
20. P. Jiang, S. Hu, Y. Ouyang, W. Ren, C. Yu, Z. Zhang, and J. Chen, Remarkable thermal rectification in pristine and symmetric monolayer graphene enabled by asymmetric thermal contact, *J. Appl. Phys.* 127(23), 235101 (2020)
21. A. L. Moore and L. Shi, Emerging challenges and materials for thermal management of electronics, *Mater. Today* 17(4), 163 (2014)
22. Z. Zhang, J. Chen, and B. Li, Negative Gaussian curvature induces significant suppression of thermal conduction in carbon crystals, *Nanoscale* 9(37), 14208 (2017)
23. M. Soltanimehr and M. Afrand, Thermal conductivity enhancement of COOH-functionalized MWCNTs/ethylene glycol–water nanofluid for application in heating and cooling systems, *Appl. Therm. Eng.* 105, 716 (2016)
24. Z. Zhang, S. Hu, T. Nakayama, J. Chen, and B. Li, Reducing lattice thermal conductivity in schwarzites via engineering the hybridized phonon modes, *Carbon* 139, 289 (2018)
25. G. Xie, Z. Ju, K. Zhou, X. Wei, Z. Guo, Y. Cai, and G. Zhang, Ultra-low thermal conductivity of two-dimensional phononic crystals in the incoherent regime, *npj Comput. Mater.* 4, 1 (2018)
26. Z. Zhang, S. Hu, Q. Xi, T. Nakayama, S. Volz, J. Chen, and B. Li, Tunable phonon nanocapacitor built by carbon schwarzite based host–guest system, *Phys. Rev. B* 101(8), 081402 (2020)
27. S. Hu, Z. Zhang, P. Jiang, W. Ren, C. Yu, J. Shiomi, and J. Chen, Disorder limits the coherent phonon transport in two-dimensional phononic crystal structures, *Nanoscale* 11(24), 11839 (2019)
28. Y. Ouyang, Z. Zhang, Q. Xi, P. Jiang, W. Ren, N. Li, J. Zhou, and J. Chen, Effect of boundary chain folding on thermal conductivity of lamellar amorphous polyethylene, *RSC Adv.* 9(57), 33549 (2019)
29. M. Yang, Y. Zhu, X. Wang, Q. Wang, L. Ai, L. Zhao, and Y. Chu, A novel low thermal conductivity thermal barrier coating at super high temperature, *Appl. Surf. Sci.* 497, 143774 (2019)
30. M. Arai, H. Ochiai, and T. Suidzuc, A novel low-thermal-conductivity plasma-sprayed thermal barrier coating controlled by large pores, *Surf. Coat. Tech.* 285, 120 (2016)
31. M. Bharti, A. Singh, S. Samanta, and D. K. Aswal, Conductive polymers for thermoelectric power generation, *Prog. Mater. Sci.* 93, 270 (2018)
32. D. Ghosh, I. Calizo, D. Teweldebrhan, E. P. Pokatilov, D. L. Nika, A. A. Balandin, W. Bao, F. Miao, and C. N. Lau, Extremely high thermal conductivity of graphene: Prospects for thermal management applications in nanoelectronic circuits, *Appl. Phys. Lett.* 92(15), 151911 (2008)
33. Z. Zhang, S. Hu, J. Chen, and B. Li, Hexagonal boron nitride: A promising substrate for graphene with high heat dissipation, *Nanotechnology* 28(22), 225704 (2017)
34. A. A. Balandin, S. Ghosh, W. Bao, I. Calizo, D. Teweldebrhan, F. Miao, and C. N. Lau, Superior thermal conductivity of single-layer graphene, *Nano Lett.* 8(3), 902 (2008)
35. P. Kim, L. Shi, A. Majumdar, and P. L. McEuen, Thermal transport measurements of individual multiwalled nanotubes, *Phys. Rev. Lett.* 87(21), 215502 (2001)
36. L. Shi, D. Li, C. Yu, W. Jang, D. Kim, Z. Yao, P. Kim, and A. Majumdar, Measuring thermal and thermoelectric properties of one-dimensional nanostructures using a microfabricated device, *J. Heat Transfer* 125(5), 881 (2003)
37. D. G. Cahill, M. Katiyar, and J. R. Abelson, Thermal conductivity of a-Si: H thin films, *Phys. Rev. B* 50(9), 6077 (1994)
38. D. G. Cahill, Analysis of heat flow in layered structures for time-domain thermoreflectance, *Rev. Sci. Instrum.* 75(12), 5119 (2004)
39. C. C. Williams and H. K. Wickramasinghe, Scanning thermal profiler, *Appl. Phys. Lett.* 49(23), 1587 (1986)
40. X. Xu, J. Chen, and B. Li, Phonon thermal conduction in novel 2D materials, *J. Phys.: Condens. Matter* 28(48), 483001 (2016)
41. A. J. H. McGaughey, A. Jain, and H. Y. Kim, Phonon properties and thermal conductivity from first principles, lattice dynamics, and the Boltzmann transport equation, *J. Appl. Phys.* 125(1), 011101 (2019)
42. H. Bao, J. Chen, X. Gu, and B. Cao, A review of simulation methods in micro/nanoscale heat conduction, *ES Energy Environ.* 1, 16 (2018)
43. J. S. Wang, J. Wang, and J. Lü, Quantum thermal transport in nanostructures, *Eur. Phys. J. B* 62(4), 381 (2008)
44. Y. Liu, W. Hong, and B. Cao, Machine learning for predicting thermodynamic properties of pure fluids and their mixtures, *Energy* 188, 116091 (2019)
45. Y. J. Wu, L. Fang, and Y. Xu, Predicting interfacial thermal resistance by machine learning, *npj Comput. Mater.* 5, 56 (2019)

46. A. Zendehboudi and R. Saidur, A reliable model to estimate the effective thermal conductivity of nanofluids, *Heat Mass Transf.* 55(2), 397 (2019)
47. X. Gu and C. Y. Zhao, Thermal conductivity of single-layer $MoS_{2(1-x)}Se_{2x}$ alloys from molecular dynamics simulations with a machine-learning-based interatomic potential, *Comput. Mater. Sci.* 165, 74 (2019)
48. G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto, and L. Zdeborová, Machine learning and the physical sciences, *Rev. Mod. Phys.* 91(4), 045002 (2019)
49. A. O. Oliynyk and J. M. Buriak, Virtual issue on machine-learning discoveries in materials science, *Chem. Mater.* 31(20), 8243 (2019)
50. G. R. Schleder, A. C. M. Padilha, C. M. Acosta, M. Costa, and A. Fazzio, From DFT to machine learning: recent approaches to materials science — a review, *J. Phys.: Mater* 2(3), 032001 (2019)
51. J. Schmidt, M. R. G. Marques, S. Botti, and M. A. L. Marques, Recent advances and applications of machine learning in solid-state materials science, *npj Comput. Mater.* 5, 83 (2019)
52. J. Graser, S. K. Kauwe, and T. D. Sparks, Machine learning and energy minimization approaches for crystal structure predictions: A review and new horizons, *Chem. Mater.* 30(11), 3601 (2018)
53. S. K. Kauwe, J. Graser, R. Murdock, and T. D. Sparks, Can machine learning find extraordinary materials? *Comput. Mater. Sci.* 174, 109498 (2020)
54. H. Zhang, K. Hippalgaonkar, T. Buonassisi, O. M. Løvvik, E. Sagvolden, and D. Ding, Machine learning for novel thermal-materials discovery: Early successes, opportunities, and challenges, *ES Energy Environ.* 2, 1 (2018)
55. D. H. Kim, T. J. Y. Kim, X. Wang, M. Kim, Y.-J. Quan, J. W. Oh, S.-H. Min, H. Kim, B. Bhandari, I. Yang, and S.-H. Ahn, Smart machining process using machine learning: A review and perspective on machining industry, *Int. J. Precis. Eng. Manuf. Green Tech.* 5, 555 (2018)
56. S. Ekins, A. C. Puhl, K. M. Zorn, T. R. Lane, D. P. Russo, J. J. Klein, A. J. Hickey, and A. M. Clark, Exploiting machine learning for end-to-end drug discovery and development, *Nat. Mater.* 18(5), 435 (2019)
57. W. S. McCulloch and W. Pitts, A logical calculus of the ideas immanent in nervous activity, *Bull. Math. Biophys.* 5(4), 115 (1943)
58. F. Rosenblatt, The Perceptron, A Perceiving and Recognizing Automaton Project Para, Cornell Aeronautical Laboratory, Buffalo, NY, 1957
59. S. B. Kotsiantis, Supervised machine learning — a review of classification techniques, *Informatica (Vilnius)* 31, 249 (2007)
60. S. B. Kotsiantis, I. D. Zaharakis, and P. E. Pintelas, Machine learning: A review of classification and combining techniques, *Artif. Intell. Rev.* 26(3), 159 (2006)
61. M. Nickel, K. Murphy, V. Tresp, and E. Gabrilovich, A review of relational machine learning for knowledge graphs, *Proc. IEEE* 104(1), 11 (2016)
62. C. Voyant, G. Notton, S. Kalogirou, M. L. Nivet, C. Paoli, F. Motte, and A. Fouilloy, Machine learning methods for solar radiation forecasting: A review, *Renew. Energy* 105, 569 (2017)
63. M. I. Jordan and T. M. Mitchell, Machine learning: Trends, perspectives, and prospects, *Science* 349(6245), 255 (2015)
64. K. P. Murphy, Machine Learning: A Probabilistic Perspective, MIT Press, 2012
65. L. Chen, H. Tran, R. Batra, C. Kim, and R. Ramprasad, Machine learning models for the lattice thermal conductivity prediction of inorganic materials, *Comput. Mater. Sci.* 170, 109155 (2019)
66. F. H. Allen, The Cambridge Structural Database: A quarter of a million crystal structures and rising, *Acta Crystallogr. B* 58(3), 380 (2002)
67. F. C. Bernstein, T. F. Koetzle, G. J. B. Williams, E. F. Meyer, M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi, and M. Tasumi, The Protein Data Bank: A computer-based archival file for macromolecular structures, *Eur. J. Biochem.* 80(2), 319 (1977)
68. S. Gražulis, D. Chateigner, R. T. Downs, A. F. T. Yokochi, M. Quirós, L. Lutterotti, E. Manakova, J. Butkus, P. Moeck, and A. Le Bail, Crystallography Open Database — an open-access collection of crystal structures, *J. Appl. Cryst.* 42(4), 726 (2009)
69. A. Belsky, M. Hellenbrandt, V. L. Karen, and P. Luksch, New developments in the Inorganic Crystal Structure Database (ICSD): Accessibility in support of materials research and design, *Acta Crystallogr. B* 58(3), 364 (2002)
70. Y. Xu, M. Yamazaki, and P. Villars, Inorganic materials database for exploring the nature of material, *Jpn. J. Appl. Phys.* 50(11S), 11RH02 (2011)
71. 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, Commentary: The Materials Project: A materials genome approach to accelerating materials innovation, *APL Mater.* 1(1), 011002 (2013)
72. S. Curtarolo, W. Setyawan, S. Wang, J. Xue, K. Yang, R. H. Taylor, L. J. Nelson, G. L. W. Hart, S. Sanvito, M. Buongiorno-Nardelli, N. Mingo, and O. Levy, AFLOWLIB.ORG: A distributed materials properties repository from high-throughput *ab initio* calculations, *Comput. Mater. Sci.* 58, 227 (2012)
73. N. E. R. Zimmermann and A. Jain, Local structure order parameters and site fingerprints for quantification of coordination environment and crystal structure similarity, *RSC Adv.* 10(10), 6063 (2020)
74. L. Zhu, M. Amsler, T. Fuhrer, B. Schaefer, S. Faraji, S. Rostami, S. A. Ghasemi, A. Sadeghi, M. Grauzinyte, C. Wolverton, and S. Goedecker, A fingerprint based metric for measuring similarities of crystalline structures, *J. Chem. Phys.* 144(3), 034203 (2016)
75. A. Seko, H. Hayashi, K. Nakayama, A. Takahashi, and I. Tanaka, Representation of compounds for machine-learning prediction of physical properties, *Phys. Rev. B* 95(14), 144110 (2017)

76. J. Wang, A. Yousefzadi Nobakht, J. D. Blanks, D. Shin, S. Lee, A. Shyam, H. Rezayat, and S. Shin, Machine learning for thermal transport analysis of aluminum alloys with precipitate morphology, *Adv. Theory Simul.* 2(4), 1800196 (2019)
77. R. M. Balabin and E. I. Lomakina, Support vector machine regression (LS-SVM) — an alternative to artificial neural networks (ANNs) for the analysis of quantum chemistry data? *Phys. Chem. Chem. Phys.* 13(24), 11710 (2011)
78. H. Kurt and M. Kayfeci, Prediction of thermal conductivity of ethylene glycol–water solutions by using artificial neural networks, *Appl. Energy* 86(10), 2244 (2009)
79. J. Carrete, W. Li, N. Mingo, S. Wang, and S. Curtarolo, Finding unprecedentedly low-thermal-conductivity half-Heusler semiconductors via high-throughput materials modeling, *Phys. Rev. X* 4(1), 011019 (2014)
80. M. J. Peet, H. S. Hasan, and H. K. D. H. Bhadeshia, Prediction of thermal conductivity of steel, *Int. J. Heat Mass Transf.* 54(11–12), 2602 (2011)
81. S. Ju, T. Shiga, L. Feng, Z. Hou, K. Tsuda, and J. Shiomi, Designing nanostructures for phonon transport via bayesian optimization, *Phys. Rev. X* 7(2), 021024 (2017)
82. T. Zhou, Z. Song, and K. Sundmacher, Big data creates new opportunities for materials research: A review on methods and applications of machine learning for materials design, *Engineering* 5(6), 1017 (2019)
83. Y. Liu, T. Zhao, W. Ju, and S. Shi, Materials discovery and design using machine learning, *Journal of Materiomics* 3(3), 159 (2017)
84. S. Ju and J. Shiomi, Materials informatics for heat transfer: Recent progresses and perspectives, *Nanoscale Microscale Thermophys. Eng.* 23(2), 157 (2019)
85. T. Ishikawa, T. Miyake, and K. Shimizu, Materials informatics based on evolutionary algorithms: Application to search for superconducting hydrogen compounds, *Phys. Rev. B* 100(17), 174506 (2019)
86. C. Chen, Y. Zuo, W. Ye, X. Li, Z. Deng, and S. P. Ong, A critical review of machine learning of energy materials, *Adv. Energy Mater.* 10(8), 1903242 (2020)
87. E. Minamitani, M. Ogura, and S. Watanabe, Simulating lattice thermal conductivity in semiconducting materials using high-dimensional neural network potential, *Appl. Phys. Express* 12(9), 095001 (2019)
88. H. D. Landahl, W. S. McCulloch, and W. Pitts, A statistical consequence of the logical calculus of nervous nets, *Bull. Math. Biophys.* 5(4), 135 (1943)
89. J. Behler and M. Parrinello, Generalized neural-network representation of high-dimensional potential-energy surfaces, *Phys. Rev. Lett.* 98(14), 146401 (2007)
90. A. P. Thompson, L. P. Swiler, C. R. Trott, S. M. Foiles, and G. J. Tucker, Spectral neighbor analysis method for automated generation of quantum-accurate interatomic potentials, *J. Comput. Phys.* 285, 316 (2015)
91. M. P. Marder, Condensed Matter Physics, John Wiley & Sons, INC, New Jersey, 2010
92. R. Hockney, and J. Eastwood, Computer Simulation Using Particles, Taylor & Francis Group, New York, 1988
93. P. Rowe, G. Csányi, D. Alfè, and A. Michaelides, Development of a machine learning potential for graphene, *Phys. Rev. B* 97(5), 054303 (2018)
94. A. P. Bartók, J. Kermode, N. Bernstein, and G. Csányi, Machine learning a general-purpose interatomic potential for silicon, *Phys. Rev. X* 8(4), 041048 (2018)
95. X. Wan, W. Feng, Y. Wang, H. Wang, X. Zhang, C. Deng, and N. Yang, Materials discovery and properties prediction in thermal transport via materials informatics: A mini review, *Nano Lett.* 19(6), 3387 (2019)
96. X. Wang, S. Zeng, Z. Wang, and J. Ni, Identification of crystalline materials with ultra-low thermal conductivity based on machine learning study, *J. Phys. Chem. C* 124(16), 8488 (2020)
97. R. Juneja, G. Yumnam, S. Satsangi, and A. K. Singh, Coupling the high-throughput property map to machine learning for predicting lattice thermal conductivity, *Chem. Mater.* 31(14), 5145 (2019)
98. E. J. Kautz, A. R. Hagen, J. M. Johns, and D. E. Burkes, A machine learning approach to thermal conductivity modeling: A case study on irradiated uranium-molybdenum nuclear fuels, *Comput. Mater. Sci.* 161, 107 (2019)
99. S. Fujii, T. Yokoi, C. A. J. Fisher, H. Moriwake, and M. Yoshiya, Quantitative prediction of grain boundary thermal conductivities from local atomic environments, *Nat. Commun.* 11(1), 1854 (2020)
100. S. Ju, R. Yoshida, C. Liu, K. Hongo, T. Tadano, and J. Shiomi, Exploring diamond-like lattice thermal conductivity crystals via feature-based transfer learning, arXiv: 1909.11234 (2019)
101. A. Seko, A. Togo, H. Hayashi, K. Tsuda, L. Chaput, and I. Tanaka, Prediction of low-thermal-conductivity compounds with first-principles anharmonic lattice-dynamics calculations and bayesian optimization, *Phys. Rev. Lett.* 115(20), 205901 (2015)
102. A. van Roekeghem, J. Carrete, C. Oses, S. Curtarolo, and N. Mingo, High-throughput computation of thermal conductivity of high-temperature solid phases: The case of oxide and fluoride perovskites, *Phys. Rev. X* 6(4), 041061 (2016)
103. C. Zhang and Q. Sun, Gaussian approximation potential for studying the thermal conductivity of silicene, *J. Appl. Phys.* 126(10), 105103 (2019)
104. H. Babaei, R. Guo, A. Hashemi, and S. Lee, Machine-learning-based interatomic potential for phonon transport in perfect crystalline Si and crystalline Si with vacancies, *Phys. Rev. Mater.* 3(7), 074603 (2019)
105. X. Qian, S. Peng, X. Li, Y. Wei, and R. Yang, Thermal conductivity modeling using machine learning potentials: application to crystalline and amorphous silicon, *Mater. Today Phys.* 10, 100140 (2019)
106. B. L. Zink, R. Pietri, and F. Hellman, Thermal conductivity and specific heat of thin-film amorphous silicon, *Phys. Rev. Lett.* 96(5), 055902 (2006)

107. K. T. Regner, D. P. Sellan, Z. Su, C. H. Amon, A. J. H. McGaughey, and J. A. Malen, Broadband phonon mean free path contributions to thermal conductivity measured using frequency domain thermoreflectance, *Nat. Commun.* 4(1), 1640 (2013)

108. P. Korotaev, I. Novoselov, A. Yanilkin, and A. Shapeev, Accessing thermal conductivity of complex compounds by machine learning interatomic potentials, *Phys. Rev. B* 100(14), 144308 (2019)

109. A. Agrawal and A. Choudhary, Perspective: Materials informatics and big data: Realization of the "fourth paradigm" of science in materials science ok, *APL Mater.* 4(5), 053208 (2016)

110. T. Zhan, L. Fang, and Y. Xu, Prediction of thermal boundary resistance by the machine learning method, *Sci. Rep.* 7(1), 7109 (2017)

111. Y. J. Wu, M. Sasaki, M. Goto, L. Fang, and Y. Xu, Electrically conductive thermally insulating Bi Si nanocomposites by interface design for thermal management, *ACS Appl. Nano Mater.* 1(7), 3355 (2018)

112. Y. J. Wu, T. Zhan, Z. Hou, L. Fang, and Y. Xu, Physical and chemical descriptors for predicting interfacial thermal resistance, *Sci. Data* 7(1), 36 (2020)

113. H. Yang, Z. Zhang, J. Zhang, and X. C. Zeng, Machine learning and artificial neural network prediction of interfacial thermal resistance between graphene and hexagonal boron nitride, *Nanoscale* 10(40), 19092 (2018)

114. S. Ju, S. Shimizu, and J. Shiomi, Designing thermal functional materials by coupling thermal transport calculations and machine learning, *J. Appl. Phys.* 128(16), 161102 (2020)

115. J. Wan, J. W. Jiang, and H. S. Park, Machine learning-based design of porous graphene with low thermal conductivity, *Carbon* 157, 262 (2020)

116. 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)

117. T. Juntunen, O. Vänskä, and I. Tittonen, Anderson localization quenches thermal transport in aperiodic superlattices, *Phys. Rev. Lett.* 122(10), 105901 (2019)

118. 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)

119. P. Roy 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)

120. M. Yamawaki, M. Ohnishi, S. Ju, and J. Shiomi, Multifunctional structural design of graphene thermoelectrics by Bayesian optimization, *Sci. Adv.* 4(6), eaar4192 (2018)

121. Z. Hou, Y. Takagiwa, Y. Shinohara, Y. Xu, and K. Tsuda, Machine-learning-assisted development and theoretical consideration for the $Al_2Fe_3Si_3$ thermoelectric material, *ACS Appl. Mater. Interfaces* 11(12), 11545 (2019)

122. A. Sakurai, K. Yada, T. Simomura, S. Ju, M. Kashiwagi, H. Okada, T. Nagao, K. Tsuda, and J. Shiomi, Ultranarrow-band wavelength-selective thermal emission with aperiodic multilayered metamaterials designed by bayesian optimization, *ACS Cent. Sci.* 5(2), 319 (2019)

123. J. Guo, S. Ju, and J. Shiomi, Design of a highly selective radiative cooling structure accelerated by materials informatics, *Opt. Lett.* 45(2), 343 (2020)

124. Z. Li, Q. Xu, Q. Sun, Z. Hou, and W. J. Yin, Thermodynamic stability landscape of halide double perovskites via high-throughput computing and machine learning, *Adv. Funct. Mater.* 29(9), 1807280 (2019)

125. K. Homma, Y. Liu, M. Sumita, R. Tamura, N. Fushimi, J. Iwata, K. Tsuda, and C. Kaneta, Optimization of heterogeneous ternary Li_3PO_4–Li_3BO_3–Li_2SO_4 mixture for Li-ion conductivity by machine learning, *J. Phys. Chem. C* 124(24), 12865 (2020)

126. H. Hazama, S. Sobue, S. Tajima, and R. Asahi, Phosphorescent material search using a combination of high-throughput evaluation and machine learning, *Inorg. Chem.* 58(16), 10936 (2019)

127. Z. Zhou, Y. Zhou, Q. He, Z. Ding, F. Li, and Y. Yang, Machine learning guided appraisal and exploration of phase design for high entropy alloys, *npj Comput. Mater.* 5, 128 (2019)

128. H. Salmenjoki, M. J. Alava, and L. Laurson, Machine learning plastic deformation of crystals, *Nat. Commun.* 9(1), 5307 (2018)

129. K. Xia, H. Gao, C. Liu, J. Yuan, J. Sun, H. T. Wang, and D. Xing, A novel superhard tungsten nitride predicted by machine-learning accelerated crystal structure search, *Sci. Bull. (Beijing)* 63(13), 817 (2018)