Journal of Applied Physics REVIEW pubs.aip.org/aip/jap

Machine-learning-assisted searching for thermally conductive polymers: A mini review

Cite as: J. Appl. Phys. **135**, 120701 (2024); doi: 10.1063/5.0201613
Submitted: 31 January 2024 · Accepted: 6 March 2024 ·
Published Online: 25 March 2024

Yinglong Hu,[1] Qi Wang,[2] and Hao Ma[1,a)]

AFFILIATIONS

[1]CAS Key Laboratory of Mechanical Behavior and Design of Materials, Department of Thermal Science and Energy Engineering, University of Science and Technology of China, Hefei, Anhui 230027, China
[2]School of Biomedical Engineering and Suzhou Institute for Advanced Research, University of Science and Technology of China, Suzhou 215123, China

Note: This paper is part of the special topic, Machine Learning for Thermal Transport.
[a)]**Author to whom correspondence should be addressed:** mhao@ustc.edu.cn

ABSTRACT

Polymers, known for their lightweight, high strength, and ease of processing, serve as a key component in engineering materials. Polymers with high thermal conductivity (TC) present enormous potential applications in thermal management for high-performance electronic devices. However, the discovery of thermally conductive polymers is still in a time-consuming and labor-intensive trial-and-error process, which undoubtedly hinders the progress in related applications. Fortunately, machine learning (ML) enables to overcome this obstacle by building precise models to predict the TC of polymers through learning from a large volume of data and it can quickly identify polymers with high TC and provide significant insights to guide further design and innovation. In this mini review, we briefly describe the general process of using ML to predict polymers with high TC and then give guidance for the selection and utilization of three important components: database, descriptor, and algorithm. Furthermore, we summarize the predicted thermally conductive single polymer chains, amorphous polymers, and metal-organic frameworks via ML and identify the key factors that lead to high TC. Finally, we touch on the challenges faced when utilizing ML to predict the TC of polymer and provide a foresight into future research endeavors.

I. INTRODUCTION

Polymers, exhibiting lightweight, high strength, corrosion resistance, and ease of processing, have become integral to materials science. They have seen widespread use as specialty materials in devices such as batteries,[1,2] solar cells,[3–5] flexible wearable devices,[6–8] and thermal exchangers[9]. With modern electronic devices trending toward higher performance, lighter weight, and greater integration, the need for effective thermal management using compact, cost-effective materials is becoming increasingly critical. Polymers are now showing promising potential in this area. However, they have an inherent shortcoming: low thermal conductivity (TC), specifically in amorphous polymers whose TCs typically lie in the region of 0.1–0.3 W m^{-1} K^{-1}.[10] It has been indicated that the TC of polymers can be significantly enhanced via stretching to align polymer chains, increasing crystallinity, engineering inter-/intra-molecular interactions, and blending with thermally conductive fillers.[11–13] These endeavors carry exciting and promising implications for electronic thermal management. However, the traditional development of polymers with high TC resorts to the laborious and time-consuming trial-and-error process, which often lacks insightful guidance. To overcome this obstacle, a combination of machine learning (ML) and high-throughput molecular dynamics (MD) simulations has been introduced into the discovery of thermally conductive polymers in recent years. Facing the expansive chemical space of polymers, ML is capable of handling and analyzing volumes of data far beyond human capacity. By analyzing and learning from existing data, ML can quickly decipher hidden interconnections within the data, thereby leading to valuable insights. The high-precision predictive model aids in pre-selecting polymers likely to yield high TC, which not only enables more focused and efficient efforts in accelerating the discovery of thermally conductive polymers but also provides clear guidance for further design and innovation. In this mini review, we first show the

general process to discover polymers with high TC via the ML method, then discuss the importance of three key components (database, descriptor, and algorithm) to provide a guide for readers on how to use or select them, and finally summarize the current progress of discovering thermally conductive single polymer chains, amorphous polymer blends, as well as metalorganic frameworks (MOFs) via the ML method and illustrate key factors that determine their high TC. In the end, we provide our thoughts on current challenges and point out future opportunities for ML methods.

II. MAIN TEXT

A. The process of predicting thermally conductive polymers via ML

The simple process of ML predicting polymer with high TC is depicted in Fig. 1. First, one needs to select the appropriate database based on the specific polymer structures and target properties. For polymers, commonly used databases include PoLyInfo,[14] PI1M,[15] QM9,[16] Polymer Genome,[17] along with CoRE MOF (Computation-Ready, Experimental Metal-Organic Framework Database)[18] and QMOF[19] for the design of MOFs. These databases, developed for different purposes, require diligent selection or further processing to meet usage needs. For instance, a considerable amount of structures in PoLyInfo lack TC data as conducting experiment measurements is challenging. To rectify this situation and overcome the time-consuming trial-and-error experimental process, one could perform high-precision MD simulations to compute the TC of preselected structures, thereby expanding the dataset. Furthermore, due to different fabrication methods of polymers, the experimental TC values in the existing databases may be quite different and to mitigate this instability in data, further optimizations are needed. To select a reasonable experimental TC value, we provide several optimization approaches for consideration: First, the thermal conductivity of polymers often has known and widely accepted values, e.g., 0.1–0.5 $W\,m^{-1}\,K^{-1}$ for amorphous polymers. Based on these standard values, one can eliminate experimental TC that deviates significantly from the norm and average the remaining values with an error range to represent the final descriptor value. Alternatively, one can annotate and categorize the experimental TC values based on synthesis or measurement methods. The values obtained from the most commonly employed synthesis or widely used measurement methods can be chosen as the final descriptor value. Moreover, the variations in these synthesis parameters can also be added as "process descriptors" to training dataset for further research if interested.[20] After completing all the preparation procedures, various related descriptors of these vast volumes of structural or attribute datasets can be constructed in

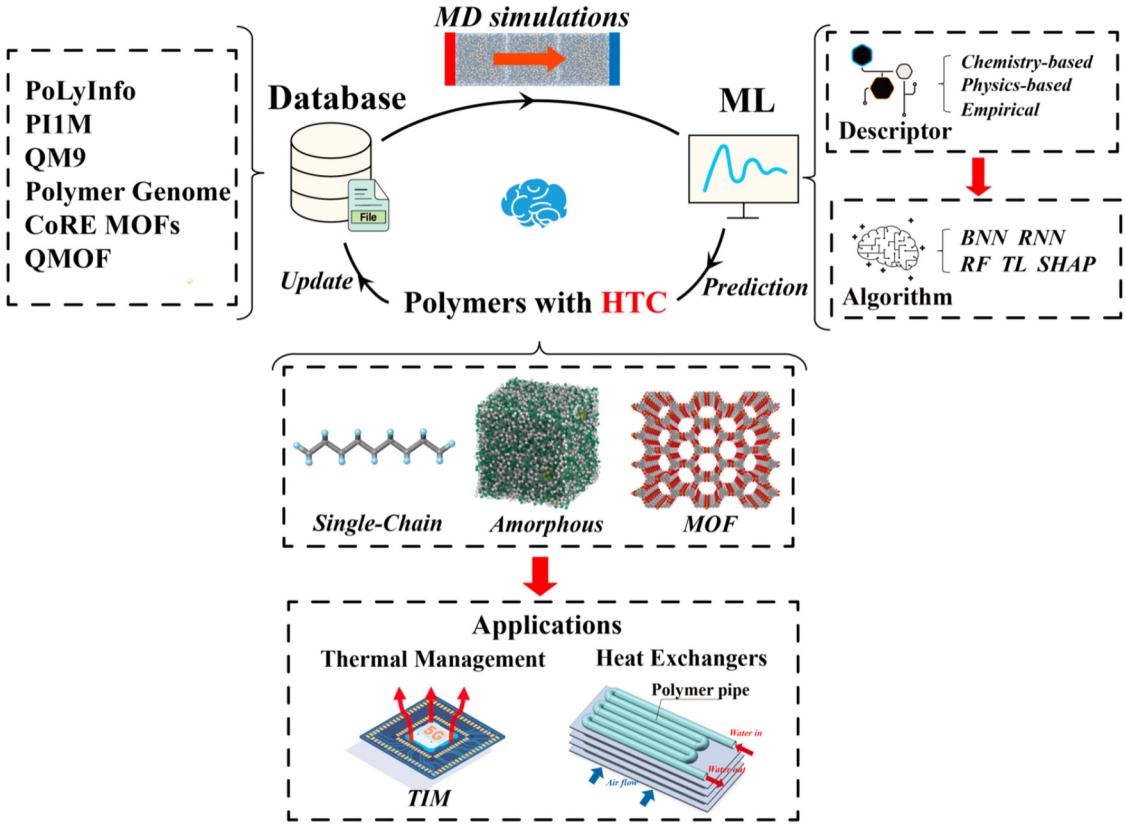

FIG. 1. Schematics of a general process of screening the polymers with high TC via machine learning combined with MD simulations.

the ML model, either directly or indirectly through software processing. Depending on their performances, one can select different algorithms to construct a relationship between the descriptors and properties, eventually establishing ML predictive models. These models will assist in quickly predicting and screening polymers with high TC. Finally, this information can serve to update the old database as a new dataset, which could undergo a new round of iterative analysis. Alternatively, these predicted polymers can become the target structures for experimental synthesis to verify whether they can meet the actual requirements for thermal management applications.

B. The guidance for selection and utilization of database, descriptor, and algorithm

To ensure the accuracy and efficiency of ML methods in predicting thermally conductive polymers, the selection and utilization of the database, descriptors, and algorithms are critical. We will take a close look at each of them in this section.

1. Database

One of the necessary conditions for machine learning is to have a large and high-quality database. The quality of the database significantly affects the accuracy and universality of the machine learning prediction results. PoLyInfo is a commonly used database for polymers, systematically providing various data needed for the design of polymeric materials, including structural and thermodynamic information such as glass transition temperature T_g, melting temperature T_m, density ρ, and TC κ. Similarly, the QM9 database consists of geometric, energetic, and thermodynamic properties of 134 thousand stable organic small molecules composed of elements: C, H, O, N, and F. The geometric data and thermodynamic properties (especially the volumetric heat capacity, C_v) from this database are critically important for training the TC prediction model. However, accessing these databases is challenging. For example, the PoLyInfo database is not easily accessible as it is embedded in web applications, causing significant inconveniences when undertaking large-scale machine learning. To solve this problem, an alternative database called PI1M was developed. After training manually collected data from 12 000 polymers in PoLyInfo, $\sim 1 \times 10^6$ new polymers were generated. PI1M has a similar chemical space as PoLyInfo and significantly populates areas where PoLyInfo data are sparse. This database, with its enormous data volume and easy accessibility, can be the prime choice for machine learning.

In terms of MOFs, owing to the excellent work of Bobbitt *et al.*,[21] the database MOFX-DB significantly facilitates the high-throughput machine learning. They have collected crystal structure files from prevalent databases such as CoRE MOF[18] and hMOF.[22] By screening the crystal structures according to the most commonly used MOF structural data such as void fraction, surface area, pore-limiting diameters (PLD), and largest cavity diameters (LCD), all needed structures can be downloaded all at once, undeniably accelerating the process of in-depth investigations of MOFs. Of course, due to the specificity of MOFs that are composed of nodes and linkers representing a vast chemical space, the existing databases might still be insufficient to further explore the influencing factors of TC for MOFs. In this case, a virtual database can be built based on the relevant knowledge. As demonstrated by Islamov *et al.*,[23] they built 10 194 MOFs with 1015 different topologies using inorganic/organic nodes with different coordination numbers and organic linkers with different topology structures as building blocks. This mode of building a virtual database enables us to explore beyond existing materials and offers a better understanding of the relationship between structures and target properties, serving as an important method of predicting thermal conductivity through machine learning in the future.

2. Descriptor

Descriptors are used to establish the relationship between the molecular structures and their properties. Accurate descriptors form a strong correlation with the TC of polymers, enhancing understanding of the link between the property and its structure. They have a high degree of universality and can provide insightful conclusions and guidance on discovering and designing thermally conductive polymers. Descriptors can be divided into three major categories: chemical descriptors, physical descriptors, and empirical descriptors.

Within chemical descriptors, the Simplified Molecular Input Line Entry System (SMILES)[24] is one of the most significant descriptors, which embodies molecular structures via a simple ASCII string. This format effectively conveys abundant structural information of polymers. Given its segmental and interpretable nature, using SMILES as a descriptor allows us to identify which molecular fragments or functional groups can give high TC[25] and simplifies the reverse-engineering of virtual structures with high TC for further verification of acquired conclusions.[26] Likewise, many other chemical descriptors such as relative molecular mass, the number of rings, and electronegativity can be rapidly generated using the RDKit[27] software package, which also facilitates data visualization, making descriptor distribution more intuitive and subsequent result analysis easier. Given the broader usage of physical descriptors, they have been integrated into many software for high-throughput computational purposes. The software Mordred[28] can swiftly calculate over 1800 descriptors that convey specific or abstract physical meanings. However, the vast number of descriptors escalates the computational costs and the learning time. Therefore, it is encouraged to make a prudent selection to balance the accuracy of the prediction model and the conservation of computational resources. It is noteworthy that MD simulation has increasingly become critical in high-throughput machine learning, and it simulates various properties of molecules by assigning force field to atoms. Not only can this extend the originally missing TC dataset, but the force field parameters including intermolecular forces, dihedral angles, and non-bonding interactions can serve as MD-inspired descriptors that participate in model training. As these parameters are directly related to the calculation of TC during MD simulations, they naturally bear a strong correlation with TC. For example, Huang *et al.*[29] conducted high-throughput computations using Mordred and MD-inspired descriptors, and in their optimized top eight important descriptors, the number of Mordred-based descriptors equals MD-inspired descriptors, thereby affirming the significance of MD descriptors. Additional parameters such as volumetric heat capacity, glass transition temperature, radius of gyration, chain length, and MOF structure's pore size, density, and surface area possess the most intuitive physicochemical meanings. Their

relationship with the TC of polymers has been studied or summarized. We categorize such professional knowledge provided in the literature as empirical descriptors. As these descriptors have already been repeatedly validated through numerous pieces of research, they carry the most direct guidance significance and should, therefore, be included in the training set.

3. Algorithm

The construction of a prediction model between the descriptors and TC requires the appropriate selection of algorithms. Different algorithms function distinctively and, therefore, vary in their accuracy. The primary algorithms used for predicting TC of polymers include random forest (RF),[30] Bayesian neural network (BNN),[31] recurrent neural network (RNN),[32] transfer learning (TL),[33] and SHapley Additive exPlanations (SHAP).[34] The RF algorithm is particularly popular because of its fast training speed, high accuracy, and strong generalization capacity. Most importantly, RF is often more robust and easier to train than neural networks when dealing with small datasets.[35–38] This is one of the biggest advantages for predicting TC in polymers, as the volume of existing data on this subject is quite limited, leading to the underperformance of direct predictions using neural networks. Given the inherent randomness of the training datasets, especially when combined with MD simulations, BNN provides reasonably accurate prediction uncertainty estimates without incurring much overhead.[39–41] It also demonstrates superior overall performance, and its prediction uncertainty range can cover the actual values computed by the MD simulations.[38] The RNN excels at handling sequence data of varying lengths, and coincidentally,[42] SMILES is a method of linear representation of chemical structures, also considered sequential data format. When using SMILES to depict molecular structures, certain segments within a molecule might have a substantial impact on TC. The RNN is proficient at capturing these key latent pieces of information, despite significant variances in the length of SMILES strings.[26]

When lacking data volume and diversity, the TL algorithm developed by Yamada *et al.*[33] proves to be a revelation. TL relies on the concept of physically interlinked property types such as physical, chemical, electronic, thermodynamic, and mechanical properties and identifies common features related to the target task through sufficient data trained in a proxy model. This algorithm can generate superior prediction performance even for very small datasets. Furthermore, a TC prediction model by applying TL with T_g and T_m as the proxy target properties was successfully developed.[43] Astonishingly, the predicted polymers with high TC by the model were synthesized, achieving a high degree of consistency between the experimental measurements and predicted values.[43] It fully exemplified the unlimited potential of TL in predicting high TC polymers. While machine learning models allow us to clearly see the inputs and outputs, their inner workings and decision-making processes are often difficult to understand, limiting our ability to probe deeper into the relationship between properties. The SHAP method aims to address the black-box challenge of uninterpretable machine learning algorithms by calculating the contribution of features to model output. This method facilitates analysis and selection of factors influencing the TC, guiding us to more efficient and precise discovery and design of targeted polymers. For example, Huang *et al.*[44] utilized the SHAP method to analyze the strong linear relationship between the effective cross-sectional area of the polymer chain and the average of the dihedral force constant with its TC [partly shown in Fig. 2(f)]. The usage of the SHAP method undoubtedly adds a more directive significance to the results.

C. The predicted thermally conductive single polymer chains, amorphous polymers, and MOFs via ML

1. Single polymer chains

Based on simple linear structures of one single chain, SMILES is often employed to describe their structures when conducting machine learning. SMILES allows to rapidly construct a variety of single chains by changing the order of functional groups or elements. Newly generated structures are typically subjected to MD simulations to obtain TC for the usage in prediction models. For example, Zhu *et al.*[25] crafted 300 types of single polymer chains using SMILES descriptors, and the TC of these structures were all calculated using MD and trained with three different ML models [Kernel Ridge Regression (KRR),[45] artificial neural network (ANN),[46] and convolutional neural network (CNN)[47]] to predict their properties. As shown in Fig. 2(a), six promising single polymer chains exhibiting higher TC than polyethylene chains were screened out, with the TC reaching as high as 103 $W m^{-1} K^{-1}$. Interestingly, polymers with a carbon–oxygen double bond (−C = O) have high TC, whereas the oxygen–hydrogen (−OH) functional group has a markedly negative impact on TC [Fig. 2(b)]. Because carbon–oxygen double bonds lead to more rigid backbones that inhibit the rotation of chain segments and, thus, the large phonon group velocities.[48] To probe the impact of the chain rotation on TC, the probability density estimate (PDE)[48] was also calculated [see Fig. 2(c)]. A higher $\overline{PDE}$ denotes a less rotated chain and, thus, larger TC. Huang *et al.*[29] employed similar strategies to conduct a more general study on the TC of single polymer chains using an extensive database. They predicted the single polymer chains with TC greater than 20 $W m^{-1} K^{-1}$ via the ML method, and MD simulations were used to calculate TC of these predicted structures. Eight polymers with high TC at a range of 33–1028 $W m^{-1} K^{-1}$ calculated by MD simulations are shown in Fig. 2(d). Twenty optimized descriptors (including Mordred and MD-inspired ones) were used to analyze their impact on TC. Through the analysis of SHAP values, two most representative important parameters were identified: the cross-sectional area and the average of the dihedral force constant (in the GAFF2 force field) [see Figs. 2(e) and 2(f)]. These two descriptors reiterate that the bond strength and chain rotation dominate the TC of single-chain polymers. In summary, to achieve high TC in single polymer chains, it is necessary to have a more rigid and straight backbone and enhance the bond strength through double bonds or π conjugation to achieve high phonon group velocities.

2. Amorphous polymer

Amorphous polymers are constructed by repeatedly polymerizing monomers; hence, it is clear that the amorphous polymers with thermally conductive monomers tend to have relatively high

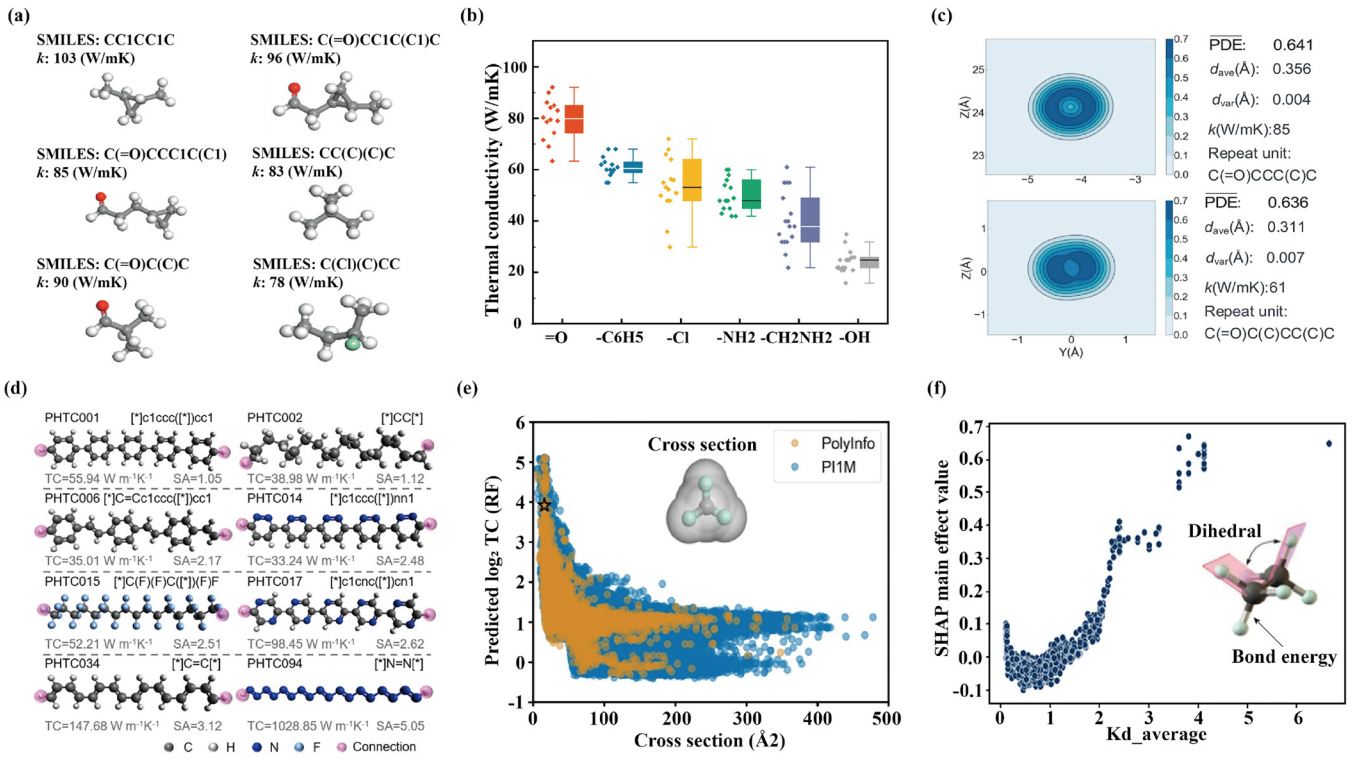

FIG. 2. (a) Six single polymer chains with high TC screened by Zhu *et al.*[25] (b) The distribution of TC of polymers with different molecular fragments. (c) Comparison of the thermal conductivity κ of structures that have different $\overline{PDE}$, d_{ave} and d_{var}. d_{ave} and d_{var} represent the concentration degree and the discrete level of the region of interest, respectively. d_{ave} is the average of the distances, and d_{var} is the variances of the distances. (d) Considering factors such as TC and synthetic accessibility (SA) score, eight polymers with high TC calculated via MD simulations and low SA are selected by Huang *et al.*[29] (e) The distinct relationship between the TC and cross-sectional area in the RF model. (f) SHAP values for the Kd_average of the training data set polymers as a function of descriptor value. (a)–(c) are reproduced with permission from Zhu *et al.*, Int. J. Heat Mass Transfer **162**, 120381 (2020). Copyright 2020 Elsevier.[25] (d)–(f) are reproduced with permission from Huang *et al.*, npj Comput. Mater. **9**(1), 191 (2023), licensed under a Creative Commons Attribution (CC BY) license.[29]

TC as well. Therefore, the key factors that dominate the TC of single polymer chains are also applicable to the screening of polymer monomers. However, the differences lie in the fact that monomers in amorphous polymers become more disordered and entangled, and they are connected via weak interactions such as van der Waals force (vdW) interactions. As a result, macroproperties of amorphous polymers such as density, molecular weight, and radius of gyration become critical.[44, 49] For example, Huang *et al.*[44] found that two MD descriptors (i.e., MW_ratio and K_bond_ave) are important to the TC of amorphous polymers. As shown in Fig. 3(a), the MW_ratio is the ratio of the main chain molecular weight to the total monomer molecular weight, and K_bond_ave represents the average of different bond force constants. The closer the MW_ratio value is to 1, the fewer side chains the polymer has, reducing the probability of heat flux dissipation along the side chains and tending to yield a larger TC. K_bond_ave reflects the overall level of bond strength in polymers. Polymers with strong bond strength are conducive to heat flux transmission along the chain, leading to higher TC. The degree of twist in the polymer chain can be represented by the radius of gyration R_g. A larger radius of gyration (indicating less chain twisting) is more conducive to high TC. In addition, when the density of amorphous polymers increases, the intermolecular interaction is enhanced, promoting heat transfer. The relationship between R_g and ρ with TC is displayed in Fig. 3(b).[50] Moreover, electrostatic interactions can enhance the TC of amorphous polymer as well.[51] In summary, fewer side chains, strong bonds, small R_g, large ρ, and strong electrostatic interactions are critical to achieving high TC in amorphous polymers. A large number of polymers with a TC of above 0.4 $W m^{-1} K^{-1}$, even up to 1.22 $W m^{-1} K^{-1}$, have been predicted via ML.[34] After considering the synthesis feasibility, Wu *et al.*[43] selected three polymers with a predicted TC above 0.2 $W m^{-1} K^{-1}$ for synthesis. The experimental results were highly consistent with the predicted TC, and after annealing treatment, the TC could even reach about 0.4 $W m^{-1} K^{-1}$. It spurs the vibrant development of predicting TC using ML.

3. Metal-organic frameworks

When it comes to MOFs, the strategy to discover high TC differs from single polymer chains and amorphous polymers due to their unique structures. MOFs consist of nodes and linkers. The

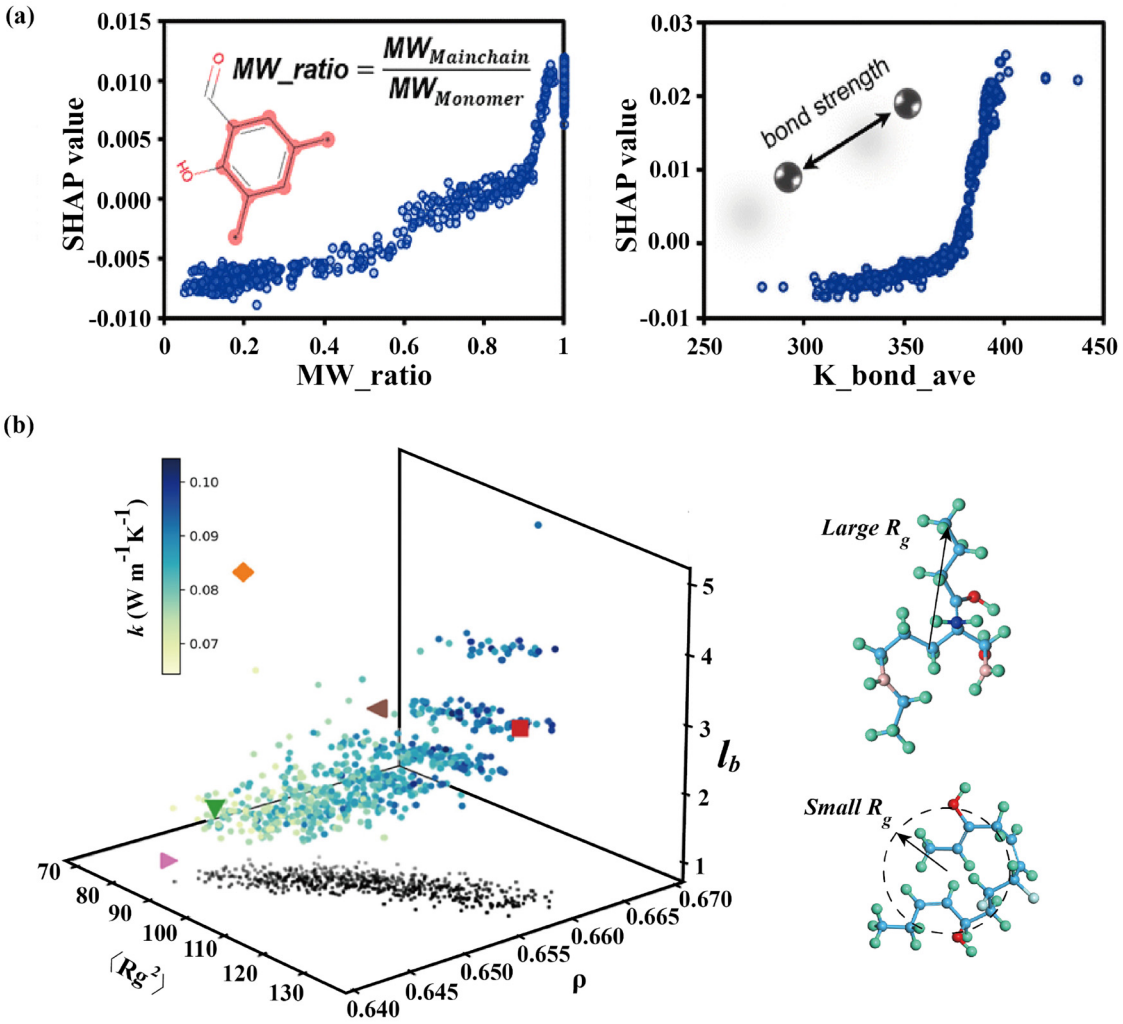

FIG. 3. (a) SHAP values for MW_ratio and K_bond_ave of the training set polymer as functions of descriptor value. The larger the SHAP value, the greater the contribution to the TC. (b) TC of copolymers as a function of $\langle R_g^2 \rangle$, ρ, and the mean block length l_b. The color bar denotes the TC of polymers. (a) is reproduced with permission from Huang *et al.*, J. Mater. Chem. A **11**(38), 20539–20548 (2023). Copyright 2023 Royal Society of Chemistry.[44] (b) is adapted with permission from Zhou *et al.* J. Chem. Theory Comput. **17**(6), 3772–3782 (2021). Copyright 2021 American Chemical Society.[50]

same strategy for selecting thermally conductive single-chain polymers can be utilized for the selection of linkers. In addition. the focus is more on the combination of nodes and linkers or the construction of the entire framework. Islamov *et al.*[23] constructed a large number of MOFs by pairing organic and inorganic nodes with different coordination numbers with organic linkers of different topological structures and some of the important inorganic nodes are shown in Fig. 4(a). High-throughput computational analysis was conducted, and more than 60 MOFs were found to possess TC greater than 3 $W m^{-1} K^{-1}$. Remarkably, the TC of two MOFs even exceeded 30 $W m^{-1} K^{-1}$. In Fig. 4(c), MOFs with high density and low largest pore diameter (LPD) (Inset III) have higher TC, while MOFs with extremely low density and high porosity (Inset I), as well as high density and high porosity ones(Inset IV), exhibit a drastic reduction in TC. In terms of geometric/compositional characteristics, the relationship between TC and mass mismatch presents a mountain shape as shown in Fig. 4(d). A relatively small mass mismatch causes a large overlap in the phonon density of state (PDOS), thereby reducing phonon scattering at the node–linker interface. Interestingly, structures with four-connected node showed very high TC (>30 $W m^{-1} K^{-1}$) [Fig. 4(d)]. In summary, MOFs with high density, low pore size, four-connected node topology, and low mass mismatch show high TC.

III. CHALLENGES AND OUTLOOKS

Given the complexity of polymers, there is a significant lack of data regarding their TC in the current database. This dearth of

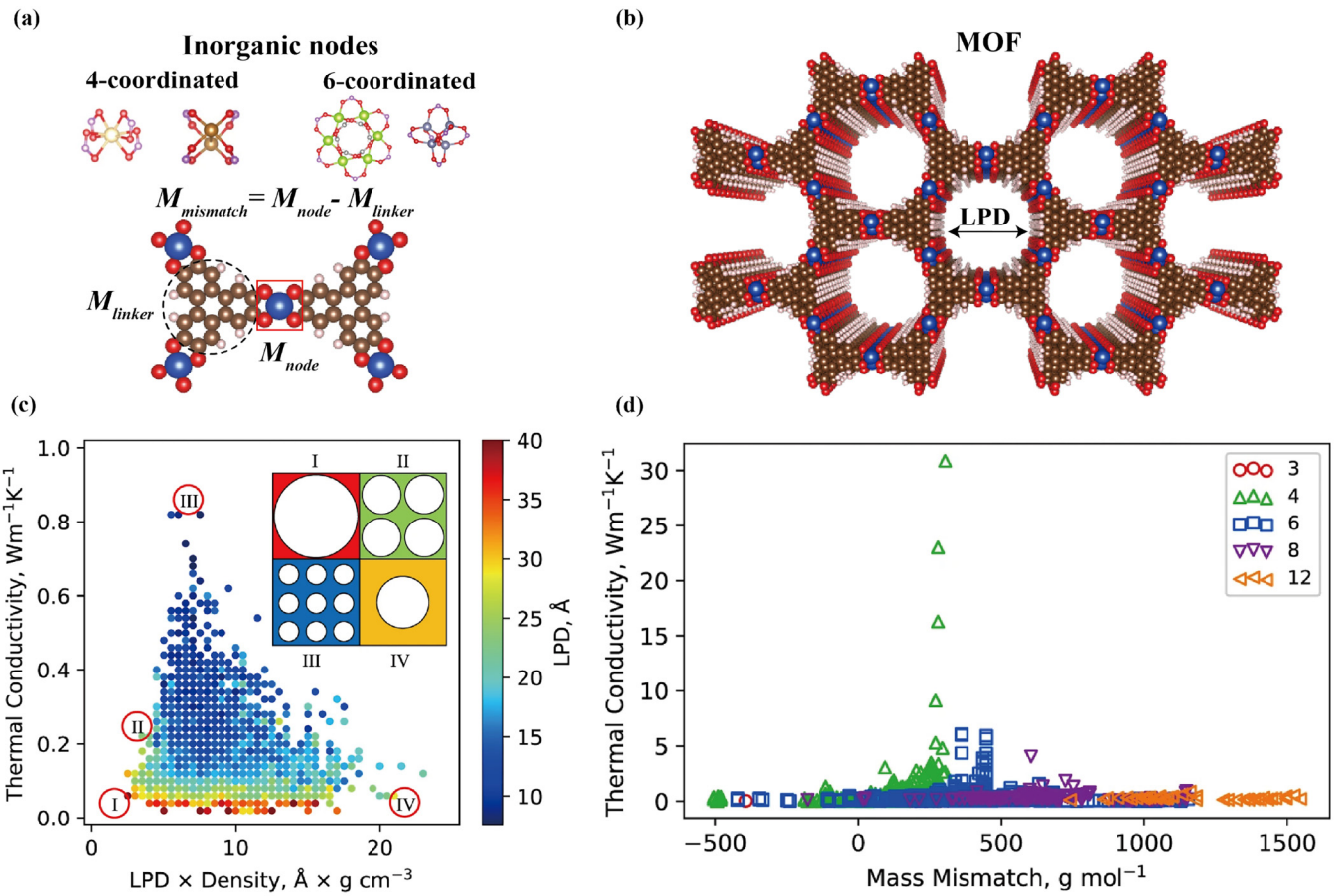

FIG. 4. (a) Four-coordinated and six-coordinated inorganic nodes. The graph at the bottom shows the meaning of mass mismatch. (b) Schematic diagram of a MOF and the location of its LPD. (c) TC as a function of product of density and LPD with four different types of pores (inset). Insets I–IV mean the structure with ultra-low density and very high LPD, low density and high LPD, high density and low LPD, and high density and high LPD, respectively. (d) The relationship between TC and mass mismatch, and the legend indicates the coordination number of the node. (a), (c), and (d) are reproduced with permission from Islamov *et al.* npj Comput. Mater. **9**(1), 11 (2023), licensed under a Creative Commons Attribution (CC BY) license.[23]

information poses a considerable challenge to the direct development of prediction models of TC. Additionally, the loss of structural information can reduce the diversity of datasets, leading to possible less extensive prediction power of ML models. Therefore, an urgent issue that needs to be addressed is the refinement and standardization of the databases. Furthermore, the credibility of the database or prediction needs to be evaluated carefully. For example, the choice of force field in MD simulations can lead to significant variations in the MD-inspired descriptors, directly impacting the credibility of calculated TC. Therefore, a future research direction could focus on developing more accurate force fields such as machine learning interatomic potential (MLIP). Developing MLIP for common polymers is highly desired. Most importantly, the credibility of the ML predictions needs to be verified by experiments. The majority of existing studies utilizing ML for the prediction of polymers with high TC are in early phases, with uncertainty surrounding the viability of synthesizing the identified polymers. Several studies[26,43] incorporated "Synthetic Accessibility (SA)" into ML predictions to evaluate the synthesizability of predicted polymers. For example, Wu *et al.*[43] fabricated the predicted polymers based on the consideration of SA score and the experimentally measured properties of the polymers are highly consistent with the predicted values from ML. Therefore, incorporating SA score within ML predictions to evaluate the feasibility of synthesis and further experimental validations are encouraged in the future ML research. Notably, predefined physical descriptors applied in ML hinder us to think outside the box. Developing ML algorithms that do not rely on predefined physical descriptors but can identify the descriptors implicitly by themselves may help to discover the unexpected or novel results. It is imperative to acknowledge that machine learning techniques lead to substantial computational resource consumption. The development of more efficient algorithms for streamlining computational processes can enhance the acceptability of such research methods.

Although there are many existing challenges, ML is still a powerful method for the discovery of polymers with high TC. We hope that this mini review could inspire more research into the discovery of thermally conductive polymers via the ML method and accelerate their applications for thermal management.

ACKNOWLEDGMENTS

Hao Ma acknowledges the funding support from USTC Startup Program (No. KY2090000117) and the Excellent Young Scholars Program of the National Natural Science Foundation of China (NNSFC) (Overseas) (No. GG2090007013).

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Yinglong Hu: Writing – original draft (equal). **Qi Wang:** Writing – review & editing (supporting). **Hao Ma:** Funding acquisition (lead); Supervision (lead); Writing – review & editing (lead).

DATA AVAILABILITY

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

REFERENCES

[1]Z. Li, J. Fu, X. Zhou, S. Gui, L. Wei, H. Yang, H. Li, and X. Guo, "Ionic conduction in polymer-based solid electrolytes," Adv. Sci. **10**(10), 2201718 (2023).
[2]Y. Zhao, Y. Zhang, D. Gosselink, T. N. L. Doan, M. Sadhu, H.-J. Cheang, and P. Chen, "Polymer electrolytes for lithium/sulfur batteries," Membranes **2**(3), 553–564 (2012).
[3]H. Yu, Y. Wang, H. K. Kim, X. Wu, Y. Li, Z. Yao, M. Pan, X. Zou, J. Zhang, S. Chen, D. Zhao, F. Huang, X. Lu, Z. Zhu, and H. Yan, "A vinylene-linker-based polymer acceptor featuring a coplanar and rigid molecular conformation enables high-performance all-polymer solar cells with over 17% efficiency," Adv. Mater. **34**(27), 2200361 (2022).
[4]L. Gao, Z. Zhang, L. Xue, J. Min, J. Zhang, Z. Wei, and Y. Li, "All-polymer solar cells based on absorption-complementary polymer donor and acceptor with high power conversion efficiency of 8.27%," Adv. Mater. **28**(9), 1884–1890 (2016).
[5]T. Kim, J.-H. Kim, T. E. Kang, C. Lee, H. Kang, M. Shin, C. Wang, B. Ma, U. Jeong, T.-S. Kim, and B. J. Kim, "Flexible, highly efficient all-polymer solar cells," Nat. Commun. **6**, 8547 (2015).
[6]R. Ma, Z. Zhang, K. Tong, D. Huber, R. Kornbluh, Y. S. Ju, and Q. Pei, "Highly efficient electrocaloric cooling with electrostatic actuation," Science **357**(6356), 1130–1134 (2017).
[7]T.-H. Han, Y. Lee, M.-R. Choi, S.-H. Woo, S.-H. Bae, B. H. Hong, J.-H. Ahn, and T.-W. Lee, "Extremely efficient flexible organic light-emitting diodes with modified graphene anode," Nat. Photonics **6**(2), 105–110 (2012).
[8]Z. Wang, Y. Bo, P. Bai, S. Zhang, G. Li, X. Wan, Y. Liu, R. Ma, and Y. Chen, "Self-sustaining personal all-day thermoregulatory clothing using only sunlight," Science **382**(6676), 1291–1296 (2023).
[9]C. S. Roper, R. C. Schubert, K. J. Maloney, D. Page, C. J. Ro, S. S. Yang, and A. J. Jacobsen, "Scalable 3D bicontinuous fluid networks: Polymer heat exchangers toward artificial organs," Adv. Mater. **27**(15), 2479–2484 (2015).
[10]X. Xu, J. Chen, J. Zhou, and B. Li, "Thermal conductivity of polymers and their nanocomposites," Adv. Mater. **30**(17), 1705544 (2018).
[11]S. Shen, A. Henry, J. Tong, R. Zheng, and G. Chen, "Polyethylene nanofibres with very high thermal conductivities," Nat. Nanotechnol. **5**(4), 251–255 (2010).
[12]S. Ronca, T. Igarashi, G. Forte, and S. Rastogi, "Metallic-like thermal conductivity in a lightweight insulator: Solid-state processed ultra high molecular weight polyethylene tapes and films," Polymer **123**, 203–210 (2017).
[13]C. L. Choy, "Thermal conductivity of polymers," Polymer **18**(10), 984–1004 (1977).
[14]S. Otsuka, I. Kuwajima, J. Hosoya, Y. Xu, and M. Yamazaki, "Polyinfo: Polymer database for polymeric materials design," in *2011 International Conference on Emerging Intelligent Data and Web Technologies* (IEEE, 2011).
[15]R. Ma and T. Luo, "PI1M: A benchmark database for polymer informatics," J. Chem. Inf. Model. **60**(10), 4684–4690 (2020).
[16]R. Ramakrishnan, P. O. Dral, M. Rupp, and O. A. von Lilienfeld, "Quantum chemistry structures and properties of 134 kilo molecules," Sci. Data **1**(1), 140022 (2014).
[17]C. Kim, A. Chandrasekaran, T. D. Huan, D. Das, and R. Ramprasad, "Polymer genome: A data-powered polymer informatics platform for property predictions," J. Phys. Chem. C **122**(31), 17575–17585 (2018).
[18]Y. G. Chung, E. Haldoupis, B. J. Bucior, M. Haranczyk, S. Lee, H. Zhang, K. D. Vogiatzis, M. Milisavljevic, S. Ling, J. S. Camp, B. Slater, J. I. Siepmann, D. S. Sholl, and R. Q. Snurr, "Advances, updates, and analytics for the computation-ready, experimental metal–organic framework database: CoRE MOF 2019," J. Chem. Eng. Data **64**(12), 5985–5998 (2019).
[19]A. S. Rosen, S. M. Iyer, D. Ray, Z. Yao, A. Aspuru-Guzik, L. Gagliardi, J. M. Notestein, and R. Q. Snurr, "Machine learning the quantum-chemical properties of metal-organic frameworks for accelerated materials discovery," Matter **4**(5), 1578–1597 (2021).
[20]Y.-J. Wu, L. Fang, and Y. Xu, "Predicting interfacial thermal resistance by machine learning," npj Comput. Mater. **5**(1), 56 (2019).
[21]N. S. Bobbitt, K. Shi, B. J. Bucior, H. Chen, N. Tracy-Amoroso, Z. Li, Y. Sun, J. H. Merlin, J. I. Siepmann, D. W. Siderius, and R. Q. Snurr, "MOFX-DB: An online database of computational adsorption data for nanoporous materials," J. Chem. Eng. Data **68**(2), 483–498 (2023).
[22]C. E. Wilmer, M. Leaf, C. Y. Lee, O. K. Farha, B. G. Hauser, J. T. Hupp, and R. Q. Snurr, "Large-scale screening of hypothetical metal–organic frameworks," Nat. Chem. **4**(2), 83–89 (2012).
[23]M. Islamov, H. Babaei, R. Anderson, K. B. Sezginel, J. R. Long, A. J. H. McGaughey, D. A. Gomez-Gualdron, and C. E. Wilmer, "High-throughput screening of hypothetical metal-organic frameworks for thermal conductivity," npj Comput. Mater. **9**(1), 11 (2023).
[24]D. Weininger, "SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules," J. Chem. Inf. Comput. Sci. **28**(1), 31–36 (1988).
[25]M. X. Zhu, H.-G. Song, Q.-C. Yu, J.-M. Chen, and H.-Y. Zhang, "Machine-learning-driven discovery of polymers molecular structures with high thermal conductivity," Int. J. Heat Mass Transfer **162**, 120381 (2020).
[26]R. M. Ma, H. F. Zhang, and T. F. Luo, "Exploring high thermal conductivity amorphous polymers using reinforcement learning," ACS Appl. Mater. Interfaces **14**(13), 15587–15598 (2022).
[27]L. Greg, see http://www.rdkit.org. for "*RDKit: Open-Source Cheminformatics*" (2006).
[28]H. Moriwaki, Y.-S. Tian, N. Kawashita, and T. Takagi, "Mordred: A molecular descriptor calculator," J. Cheminf. **10**(1), 4 (2018).
[29]X. Huang, S. Ma, C. Y. Zhao, H. Wang, and S. Ju, "Exploring high thermal conductivity polymers via interpretable machine learning with physical descriptors," npj Comput. Mater. **9**(1), 191 (2023).
[30]J. L. Speiser, M. E. Miller, J. Tooze, and E. Ip, "A comparison of random forest variable selection methods for classification prediction modeling," Expert Syst Appl. **134**, 93–101 (2019).
[31]L. V. Jospin, H. Laga, F. Boussaid, W. Buntine, and M. Bennamoun, "Hands-on Bayesian neural networks—A tutorial for deep learning users," IEEE Comput. Intell. Mag. **17**(2), 29–48 (2022).

[32]H. G. Zhang, Z. S. Wang, and D. R. Liu, "A comprehensive review of stability analysis of continuous-time recurrent neural networks," IEEE Trans. Neural Networks Learn. Syst. **25**(7), 1229–1262 (2014).
[33]H. Yamada, C. Liu, S. Wu, Y. Koyama, S. Ju, J. Shiomi, J. Morikawa, and R. Yoshida, "Predicting materials properties with little data using shotgun transfer learning," ACS Cent. Sci. **5**(10), 1717–1730 (2019).
[34]S. M. Lundberg and S. I. Lee, "A unified approach to interpreting model predictions," in *31st Annual Conference on Neural Information Processing Systems (NIPS)*, Long Beach, CA (Curran Associates Inc., 2017).
[35]Y. Qi, Random Forest for Bioinformatics in *Ensemble Machine Learning*, edited by C. Zhang and Y. Ma (Springer, New York, 2012).
[36]K. Fawagreh, M. M. Gaber, and E. Elyan, "Random forests: From early developments to recent advancements," Syst. Sci. Control Eng. **2**(1), 602–609 (2014).
[37]S. Han, B. D. Williamson, and Y. Fong, "Improving random forest predictions in small datasets from two-phase sampling designs," BMC Med. Inf. Decis. Mak. **21**(1), 322 (2021).
[38]R. M. Ma, H. Zhang, J. Xu, L. Sun, Y. Hayashi, R. Yoshida, J. Shiomi, J.-x. Wang, and T. Luo, "Machine learning-assisted exploration of thermally conductive polymers based on high-throughput molecular dynamics simulations," Mater. Today Phys. **28**, 100850 (2022).
[39]Y. Kwon, J.-H. Won, B. J. Kim, and M. C. Paik, "Uncertainty quantification using Bayesian neural networks in classification: Application to biomedical image segmentation," Comput. Stat. Data Anal. **142**, 106816 (2020).
[40]P. Westermann and R. Evins, "Using Bayesian deep learning approaches for uncertainty-aware building energy surrogate models," Energy AI **3**, 100039 (2021).
[41]J. Gawlikowski, C. R. N. Tassi, M. Ali, J. Lee, M. Humt, J. Feng, A. Kruspe, R. Triebel, P. Jung, R. Roscher, M. Shahzad, W. Yang, R. Bamler, and X. X. Zhu, "A survey of uncertainty in deep neural networks," Artif. Intell. Rev. **56**(1), 1513–1589 (2023).
[42]Z. C. Lipton, "A critical review of recurrent neural networks for sequence learning," arXiv:abs/1506.00019 (2015).
[43]S. Wu, Y. Kondo, M.-a. Kakimoto, B. Yang, H. Yamada, I. Kuwajima, G. Lambard, K. Hongo, Y. Xu, J. Shiomi, C. Schick, J. Morikawa, and R. Yoshida, "Machine-learning-assisted discovery of polymers with high thermal conductivity using a molecular design algorithm," npj Comput. Mater. **5**, 66 (2019).
[44]X. Huang, S. Ma, Y. Wu, C. Wan, C. Y. Zhao, H. Wang, and S. Ju, "High-throughput screening of amorphous polymers with high intrinsic thermal conductivity via automated physical feature engineering," J. Mater. Chem. A **11**(38), 20539–20548 (2023).
[45]K. Vu, J. C. Snyder, L. Li, M. Rupp, B. F. Chen, T. Khelif, K. Müller, and K. Burke, "Understanding kernel ridge regression: Common behaviors from simple functions to density functionals," Int. J. Quantum Chem. **115**(16), 1115–1128 (2015).
[46]J. Schmidhuber, "Deep learning in neural networks: An overview," Neural Networks **61**, 85–117 (2015).
[47]E. A. Smirnov, D. M. Timoshenko, and S. N. Andrianov, "Comparison of regularization methods for ImageNet classification with deep convolutional neural networks," in *2nd AASRI Conference on Computational Intelligence and Bioinformatics (CIB)*, South Korea (Elsevier, 2013).
[48]H. Ma and Z. Tian, "Chain rotation significantly reduces thermal conductivity of single-chain polymers," J. Mater. Res. **34**(1), 126–133 (2019).
[49]A. Nagoya, N. Kikkawa, N. Ohba, T. Baba, S. Kajita, K. Yanai, and T. Takeno, "Autonomous search for polymers with high thermal conductivity using a rapid Green-Kubo estimation," Macromolecules **55**(9), 3384–3395 (2022).
[50]T. H. Zhou, Z. Wu, H. K. Chilukoti, and F. Müller-Plathe, "Sequence-engineering polyethylene-polypropylene copolymers with high thermal conductivity using a molecular-dynamics-based genetic algorithm," J. Chem. Theory Comput. **17**(6), 3772–3782 (2021).
[51]T. L. Feng, J. He, A. Rai, D. Hun, J. Liu, and S. S. Shrestha, "Size effects in the thermal conductivity of amorphous polymers," Phys. Rev. Appl. **14**(4), 044023 (2020).