International Journal of Heat and Mass Transfer 162 (2020) 120381

Contents lists available at ScienceDirect

International Journal of Heat and Mass Transfer

journal homepage: www.elsevier.com/locate/hmt

Machine-learning-driven discovery of polymers molecular structures with high thermal conductivity

Ming-Xiao Zhu[a,*], Heng-Gao Song[a], Qiu-Cheng Yu[a], Ji-Ming Chen[a], Hong-Yu Zhang[b]

[a] College of New Energy, China University of Petroleum (East China), Qingdao 266580, PR China
[b] College of Chemical Engineering, China University of Petroleum (East China), Qingdao 266580, PR China

ARTICLE INFO

Article history:
Received 6 May 2020
Revised 5 July 2020
Accepted 24 August 2020
Available online 4 September 2020

Keywords:
Polymer chains
Thermal conductivity
Deep learning
Molecular dynamics
Molecular structure-property relationship

ABSTRACT

The ability to efficiently design new and advanced polymers with functional thermal properties is hampered by the high-cost and time-consuming experiments. Machine learning is an effective approach that can accelerate materials development by combining material science and big data techniques. Here, machine learning methods were used to predict the thermal conductivity of various single-chain polymers, and the relationship between molecular structures of polymer repeating units and thermal conductivity was also been investigated. The predict model starts from a benchmark dataset generated by large-scale molecular dynamics computations. In predict models, the polymers were 'fingerprinted' as simple, easily attainable numerical representations, which helps to develop an on-demand property prediction model. Further, potential quantitative relationship between molecular structures of polymer and thermal conductivity property was analyzed, and hypothetical polymers with ideal thermal conductivity were identified. The methods are shown to be general, and can hence guide the screening and systematic identification of high thermal conductivity.

1. Introduction

Due to the numerous advantages such as good processability, light weight, high voltage breakdown strength, polymeric materials are universally applied in modern electronics and electrical systems [1–3]. However, most polymers show low thermal conductivity (TC) in the range of 0.1–0.5 W/(mK), which is not sufficient for many applications that require high heat conduction. Molecular dynamics (MD) simulation and experimental studies have shown that polymer chains, as low-dimensional materials, can theoretically have high TC [4–6]. Therefore, it is expected to address the heat dissipation problems by developing novel polymer materials with enhanced TC [7,8].

The factors affecting the TC of polymers can be stated in terms of a subset of the following characteristics: chain structure [9–11], crystallinity and crystal form [12], orientation of polymers [13] and so on. Among these factors, chain structure including molecular composition and conformation is the first level of structure that determines TC of various polymers [14]. According to the Debye equation, i.e.$k = C_p \upsilon l/3$, the TC of the polymers depends on C_p (the specific heat capacity per unit volume), υ (the phonon velocity), and l (the phonon mean free path). In general, the C_p and υ of polymers are primarily determined by the characteristics of monomer repeat unit and characteristics/strength of backbone bonding in the individual chain [14]. Therefore, it can be expected the TC can be improved by adjusting the molecular structures of the polymer repeating units [15]. In fact, many MD simulations have been performed to understand how nanoscale molecular structures affect the TC. Ma and Tian [16] computed the TC of Kevlar (polyparaphenylene terephthalamide) and PBDT (poly(2,2′-disulfonyl-4,4′-benzidine terephthalamide)) single chains through MD, and the results indicated that the chain rotation is the key to reducing the TC. Wu et al [11] found that polymers with rigid backbones generally have high thermal stability, while the presence of heavier atoms and weak backbone bonds may reduce the TC. Using large-scale MD simulations. Liu and Yang [17] stated that the TC of polymer chain is closely related to the monomer type. The results showed that the TC of the polymers with aromatic rings is even 5 times higher than that of the polyethylene chain, while the TC is reduced due to the low bond-strength or mass disorder of the polymer chain.

Despite the polymers structure exhibits great influence on the thermal characteristics [17–20], the library of polymers is relatively large which is hardly to be comprehensively handled by the high-

* Corresponding author.
E-mail address: zhumx@upc.edu.cn (M.-X. Zhu).

https://doi.org/10.1016/j.ijheatmasstransfer.2020.120381
0017-9310/

cost and time-consuming experiments. In recent years, the machine learning (ML) technique trained on massive amounts of data matches or even outperforms humans in intellectually demanding tasks across various fields. With the remarkable innovation of ML methods, some efficient algorithms, such as kernel ridge reegression (KRR) [21], artificial neural network (ANN) [22] and convolutional neural network (CNN) [23] are developing rapidly. As such, there is growing interest in using ML to develop new materials with various properties, which is both cost and time saving compared to trial and error methods [24–27]. For example, *Carrete* et al [28] established a random forest classification model to predict the TC of half-Heusler materials, and found that the average atomic radius in structural positions is the key factor affecting the TC. Using several supervised ML and ANN models, *Yang* et al [29] predicted the interfacial thermal resistance between graphene and hexagonal boron–nitride, and concluded that ML algorithms have significant effects on the prediction accuracy. Due to the lack of TC data for polymers, Wu et al [30] used the QM9 database, which records thermal parameters related to TC such as glass transition temperature, to build a transferred TC model. Finally, the effectiveness of the method was verified by the chemical synthesis of the target material. Although ML methods have been applied to settle many research problems, they rarely been used in the field of thermal sciences in relation to polymer materials. In fact, it is urgent to look for polymer materials with high TC that can dissipate the massive heat.

In this work, a workflow based on ML was designed to discover polymers with high TC, and the relationship between molecular structures of polymer repeating units and the TC of polymer was analyzed. Fig. 1 summarizes the analytic workflow of this work. Firstly, 300 kinds of single-chain polymers were built, and nonequilibrium molecular dynamics (NEMD) [31,32] was used to calculate the TC of these polymers. These datas were used as the basic training dataset for the TC prediction model. Then, three advanced ML models including KRR, ANN and CNN were established to predict TC. The ML models could directly extract molecular features of polymer repeating units and establish structure-property linkages with high accuracy. Once the model is obtained, it can predict the effective properties of many new structures with negligible computational time. In addition, with ML predicting model, the thermal properties of polymers could be investigated in larger chemical spaces, which helps to screen promising candidates. Finally, the relationship between molecular structures and TC of polymer chains was analyzed. In this study, the polymer repeating unit was fingerprinted with Extended-Connectivity Fingerprints (ECFP) [33,34], which is well suited to the recognition of the presence or absence of particular structures. As a result, ECFP is used for building quantitative structure-property relationship model revealing key correlations between the TC and polymer structures.

2. Simulation methods

2.1. Molecular dynamic simulation

As many as 10^8 kinds of small organic molecules are known to exist in the chemical space [35], while such a large space contains many target molecules. By selecting effective chemical fragments in the chemical space and then randomly combining these fragments in a reasonable manner, polymers with excellent target properties could be discovered [36–39]. The findings above inspired this work to focus on several kinds of chemical fragments and their random recombination to find polymers with high TC properties. Due to the lack of thermal properties data and the practical synthesizing difficulties of polymers, the TC dataset were obtained by MD simulations using a Large-scale Atomic/Molecular Massively Parallel Simulator (LAMMPS) [40] in this work.

Repeating units consist of at least one chemical fragments from the following pool of possibilities: -C_6H_5, -NH_2, -(=O), -OH, -Cl, -CH_2NH_2, -O-, -N-, etc., as depicted in Fig. 1. These chemical fragments were chosen because they are common in polymer backbones, for example, polyethylene, polyester, polyamide, polyether and polyurea. We constructed 300 kinds of single-chain polymers whose TC was calculated using the NEMD simulation. Since the relationship between polymer structure of repeating units and TC was mainly investigated in this work, we explored TC of single-

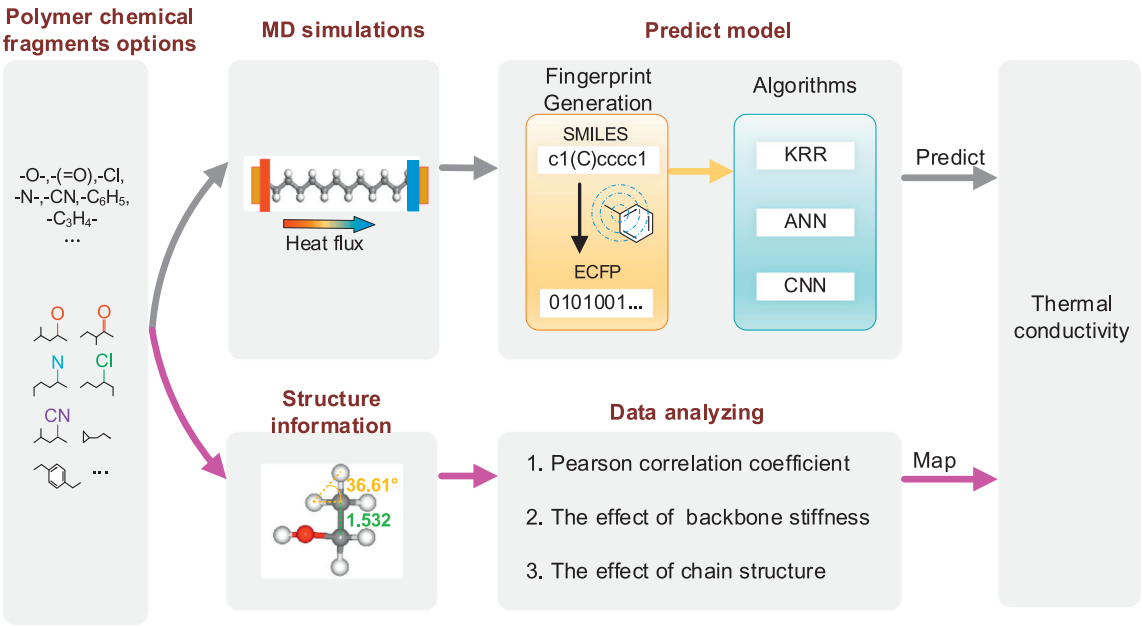

Fig. 1. Schematics of polymer chain TC prediction and molecular structure-property analysis.

chain polymers instead of precisely calculating the TC of polymers either in their amorphous or crystalline states.

Dimension of polymer chains is 600 × 50 × 50 ($x \times y \times z$) Å, in which the length in y and z directions is set as 50 Å to prevent interactions between neighboring polymer chains [16]. Materials Studio software was used to construct the initial configuration of the polymer chain. The interactions between atoms are described by the polymer consistent force field (PCFF) [41]. In this force field, each atom is assigned a corresponding coordinate position and mass, and the force cutoff distance was set to 10 Å [16]. For instance, after geometry optimization of the ethylene monomer (-CH_2-CH_2-) by adjusting the atomic coordinates iteratively to reach the minimum energy, a single polyethylene chain is obtained by polymer building module in Materials Studio. Here, the monomer is also called repeating unit.

After the initial structure was built using Materials Studio software, the simulation system was equally divided into 20-25 slabs along the x direction (the chain backbone direction), and the regions at both ends of the chain were set as a heat-insulating walls. The slab adjacent to the left heat-insulating wall was set as hot regions and the slab adjacent to the right wall was set as cold regions. Before the TC was calculated, the structure needs to be relaxed to obtain a stable conformation. The energy of the structure was firstly minimized, then NVT (constant number of particles, volume, and temperature) MD and NVE (constant number of particles, volume, and energy) MD were calculated at 300 K to release the thermal stress of the chain. After that, at regular intervals, an amount of heat was added into the hot region along the x direction, and the same heat was extracted from the cold region to form a heat flux (J) in the system. The applied heat value was obtained by multiple tests and generally ranges from 0.005 to 0.05(eV/ps), and the heat value varies slightly for the different polymer chain structures. When the heat flux was stable, the effective temperature of each slab of the chain was averaged over the following 2-4 ns. Finally, according to the Fourier's law of heat conduction, the TC of the polymer chain could be calculated using $\kappa = -(J/dT/dx)$, where κ is the TC, dT/dx is temperature gradient, which was obtained by fitting the linear temperature region using the least-square method. Figure S1 in Supporting Information depicts the schematic for calculating TC.

The TC dataset of polymers was obtained by high-throughput simulations combining python script, LAMMPS computations and MATLAB post-processing. Python scripting was used to complete the preparatory work. Then the job was submitted to the LAMMPS for calculating the temperature distribution of polymer chain. Once the job was finished, the collected data were post-processed with MATLAB scripts to obtain the TC.

2.2. Machine learning methods

Three ML models were built in this work to investigate the complex nonlinear relationship between the molecular structures and thermal properties of polymer chains, aiming at satisfying the need for quantitative predictions of TC. The polymers in the dataset were firstly fingerprinted, that is, the structures of polymers were represented in a form suitable for ML input. Each input was then reduced to a string of numbers. After that, the mapping relation between the fingerprints and the TC of polymers was established by using ML algorithms.

Encoding molecules into fixed-length strings or vectors is a core challenge for molecular ML. For the ML predict models considered in this work, we focused on following forms of representation: sequences of tokens via Simplified molecular input line entry system (SMILES) [42,43] and ECFP. SMILES is a line notation for entering and representing molecules and reactions, which uses a string of characters to describe a three-dimensional chemical structure.

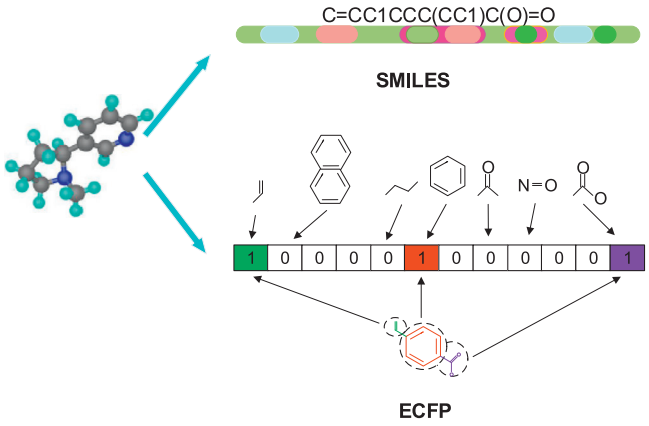

Fig. 2. The general form of molecular fingerprint. The ECFP generation process of a given molecule was summarized as follows. Firstly, the substructures in a circular range with a given diameter for each non-hydrogen atom were recorded. These substructures characteristics were then mapped to integer identifiers using the hash function. The resulting identifiers list formed after the hash process contains all the substructures of a given molecule.

With standard SMILES, the name of a molecule is synonymous with its structure. In the SMILES method, atoms are symbolized by chemical elements. The double bond is denoted by "="; The triple bond is denoted by "#". The C, O, S, and N atoms in an aromatic ring are denoted by the lowercase letters c, o, s, and n, respectively. When chemical structures are converted to SMILES, hydrogen atoms are often omitted, and if there is a ring in the structure, it is opened. The two atoms at the breaking point are marked with the same number to indicate that there is a bond between the atoms. The branch chain is written in parentheses. ECFP is a novel class of topological fingerprints for molecular characterization. In this work, the polymer repeating unit was first transformed into SMILES, and then ECFP was generated by RDKit software [44] as input to ML algorithms. ECFP uses a list of integer identifiers to naturally and accurately represent the molecular structures of polymer repeating units. Each integer represents the substructure contained in a particular atomic neighborhood of the molecule, as shown in Fig. 2.

Based on the MD simulation results and the molecular representation in ML, we trained three kinds of ML models including KRR, ANN, and CNN to predict the TC of polymers. The typical ML models are shown in Fig. 3. In the present work, 90% of the samples in the dataset were used as the training set, and the remaining 10% were used as the test set. The parameters in ML algorithms were optimized to improve the prediction accuracy. For KRR, GridSearchCV function in scikit-learn [45] was used to determine the optimal parameter δ of the gaussian kernel, with the value of 0.0032. In ANN and CNN, the adaptive moment estimation optimization method was used for gradient descent computation. With a large learning rate, the training loss does not improve and probably even grows worse. With an extremely small (near-zero) learning rate, the training results will not update at all. As a result, the learning rate was set as 0.005 here. In the constructed neural network structures, the size of the input layer was determined by the length of the ECFP-based molecular fingerprint, which is a one-dimensional sequence of length 250 in this work. Neurons in all fully-connected layers were activated by Rectified Linear Unit. The size of convolution kernel in CNN is 2, and stride of convolutional kernel is 1. Pooling size and stride of max pooling layers are 2 and 1, respectively. Both the convolution layers and the pooling layers used "same" mode for padding operation. (Table S1 of the supporting information summarizes the convolutional neural network model used in this paper)

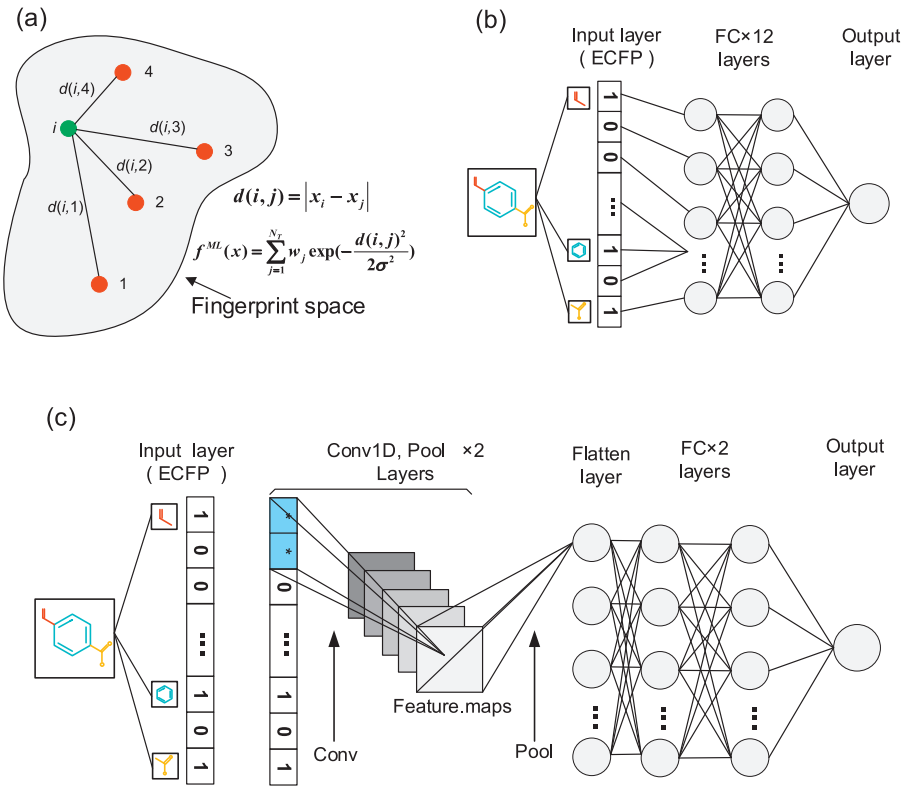

Fig. 3. Typical ML architectures used in this work. First, the molecular structures of polymer repeating units were encoded as ECFP. This sequence was taken as the input of the ML. (a) Schematic of the KRR scheme showing the example cases in fingerprint(x) space. The distance, $d(i, j)$, between the point (in fingerprint space) corresponding to a new case, j, and each of the training example cases, i, is used to predict the property. (b) The ANN model has 12 fully connected layers, each with 250 neurons. (c) CNN: where conv1D represents convolutional layer, pool represents max pooling layer and FC represents fully connected layer.

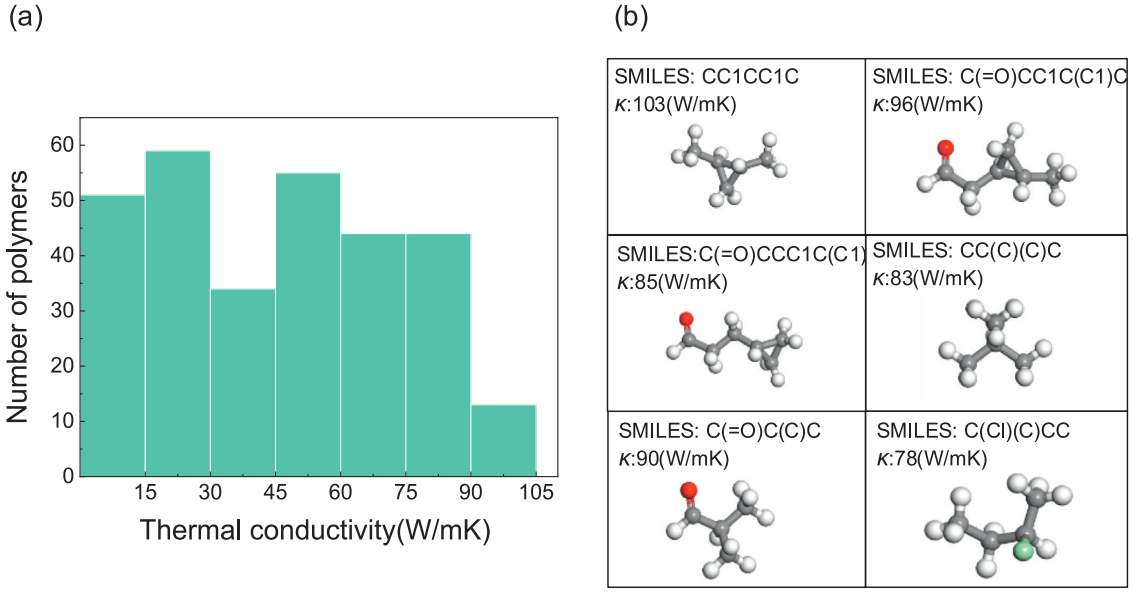

Fig. 4. The calculated TC of various polymers. (a) Summary of molecular dynamics results. (b) 6 polymer repeating units were screened.

3. Results and discussion

3.1. Molecular dynamics for thermal conductivity

Fig. 4(a) summarizes the calculated TC of various polymers. The computation accuracy was verified by comparing the TC values of several polymers with the results in existing references [17,46,47]. It can be observed that the TC values obtained by MD simulations are in the range of 1-120 W/mK, indicating that the TC of polymer chains is significantly dependent on the molecular structure, even small changes in chemical structure have a profound impact on the TC. In addition, some polymer chains present TC values greater than 80W/mK, implying that there are numerous polymer materials with high TC which potentially hidden in the huge chemical space. Finally, six kinds of typically promising targets were screened. Examples of the screened chemical structures

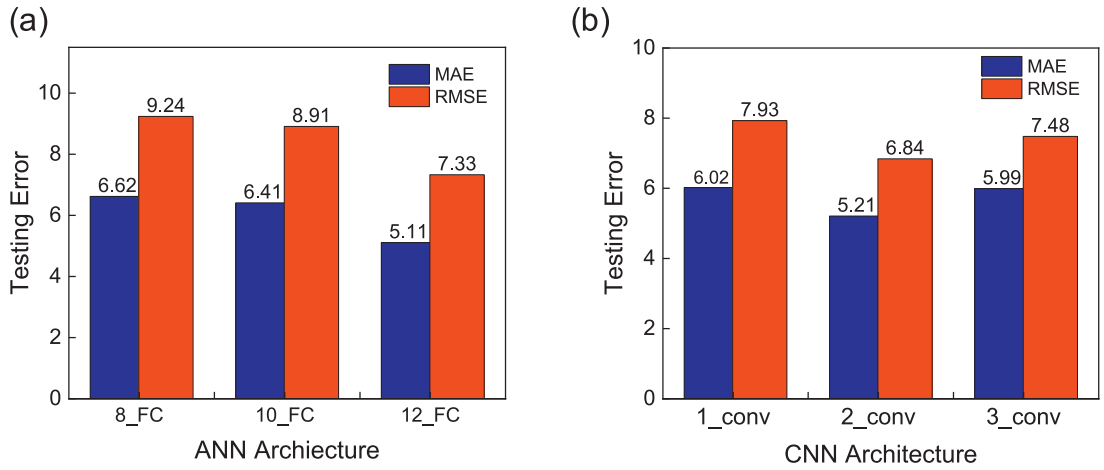

Fig. 5. Testing errors of ANNs and CNNs. (a) Dependence on different ANN architectures. ANNs with 8, 10 and 12 fully connected layers are denoted by 8_FC, 10_FC and 12_FC, respectively. (b) Dependence on different CNN architectures. CNNs with 1, 2 and 3 convolutional layers are denoted by 1_conv, 2_conv and 3_conv, respectively.

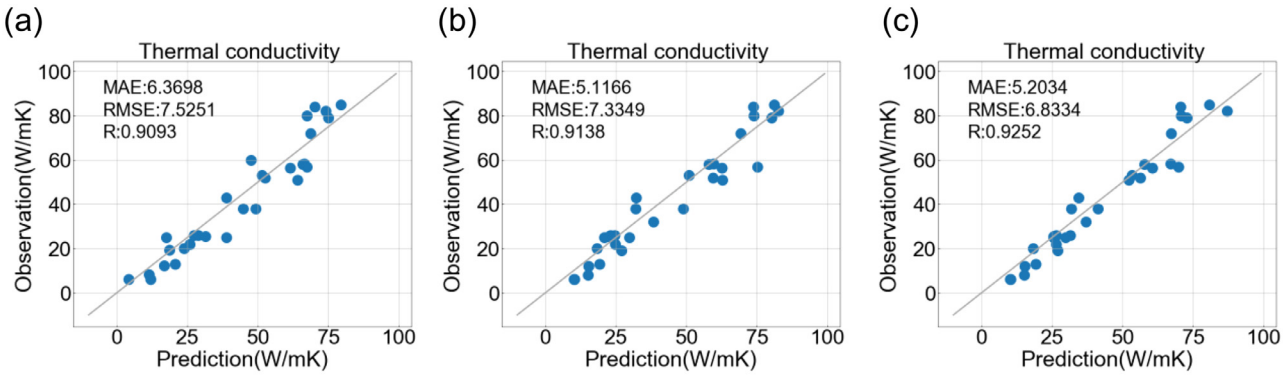

Fig. 6. ML results of (a) KRR, (b) ANN, (c) CNN. The blue dots represent ML predicted values. The MAE, RMSE and R are shown in each plot. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

are depicted in Fig. 4(b). The TC values of selected polymer chains are all higher than that of polyethylene, i.e., 20W/mK.

3.2. Prediction of thermal conductivity

In order to improve the prediction accuracy of ANN and CNN, one possible way is to increase the number of layers [48]. The mean absolute errors (MAE) and root mean square error (RMSE) of three different depths were tested, i.e., ANNs with 8, 10 and 12 fully connected layers, CNNs with 1, 2 and 3 convolutional layers. As shown in Fig. 5(a), the performance of ANN with 12 fully connected layers is better than the network with 8 and 10 fully connected layers. This is because deeper networks can capture more abstract features of the input fingerprints. As shown in Fig. 5(b), CNN with 2 convolutional layers shows the best accuracy. The reason for that is fewer convolutional layers is not enough for capturing abundant features, while too many convolutional layers may lead to overfitting, which is detrimental to network prediction performance.

By selecting appropriate molecular fingerprint and optimizing the training model, an excellent prediction accuracy of the thermal property of the polymers may be obtained. The training results of three kinds of models are shown in Fig. 6, in which the x-axis and y-axis represent the model predicted and target values, respectively. As can be seen, three kinds of ML algorithms could properly describe the relationship between TC and the features (fingerprint), with R-squared (R) values all above 0.9. In other words, these ML algorithms could recognize quantitative structure-property relationships with respect to TC. The RMSE for KRR and ANN are 7.53 and 7.33 W/mK, respectively, indicating that the test errors of them are close. While the CNN model outperforms both KRR and ANN, with the MAE of 5.20 W/mK, RMSE of 6.83 W/mK and R value of 0.93, which implies that the generated predictive values are closer to the target values. Compared to ANN architecture, CNN has multiple convolutional layers, which gives rise to the better predictive performance. It shows that convolution layers can capture more abstract features of the input sequence.

3.3. Molecular structure and thermal conductivity relations

ECFP can represent a complete molecular structure as a series of substructures. To better understand the correlations between each substructure and the TC, the Pearson correlation coefficients (r) between various substructures and TC were calculated, as presented in Fig. 7(a). The calculated r ranges from -1 to 1 with -1 representing strong negative correlations and 1 representing strong positive correlations. Fig. 7(a) shows that various substructures of polymer repeating unit exhibit different effects on TC. For instance, -(=O) group has positive r value of 0.45, indicating that polymers containing -(=O) structure usually have high TC, while -OH group has significant negative effect on the TC. The calculated TC of various polymers with different groups are shown in Fig. 7(c),

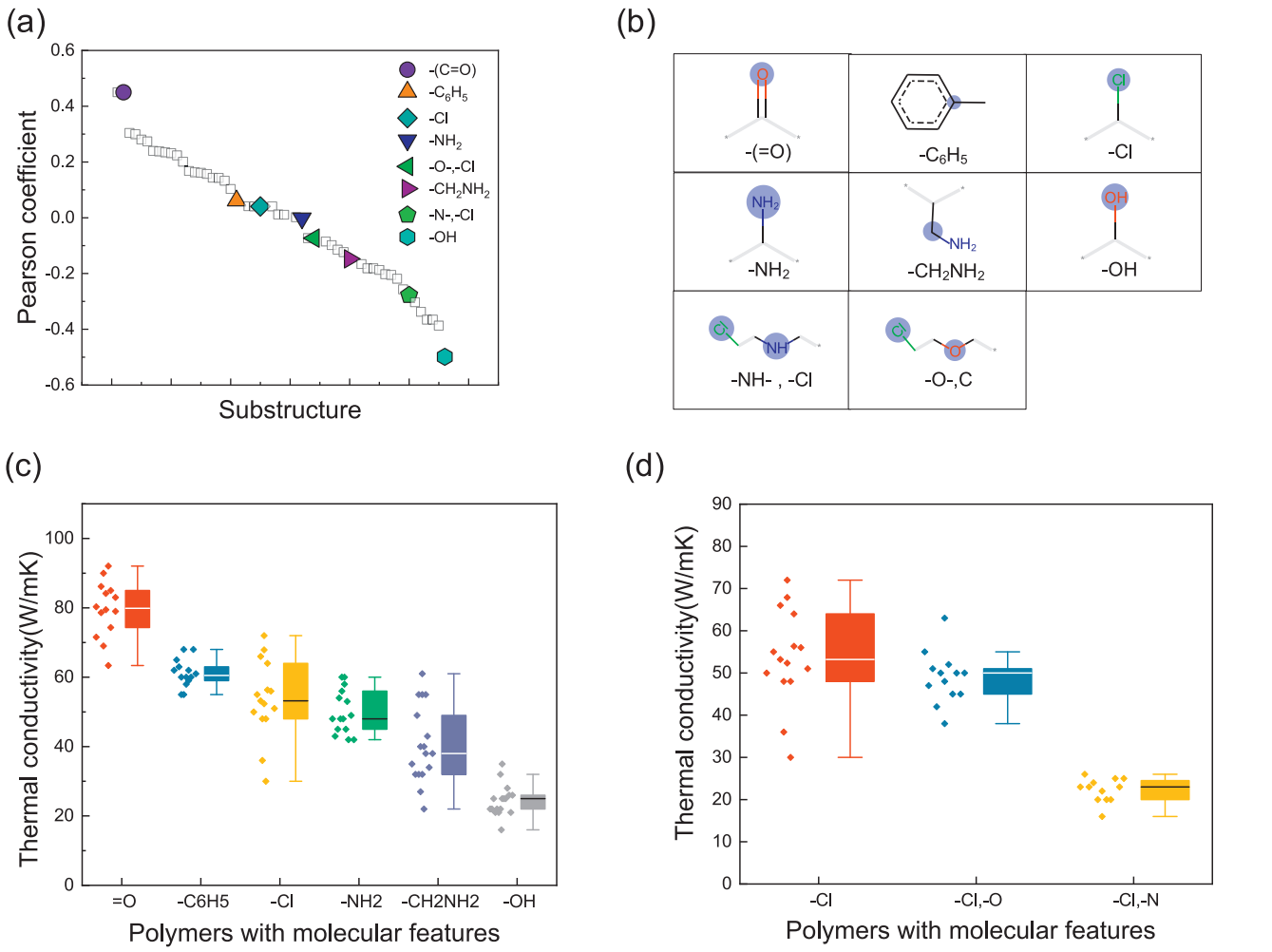

Fig. 7. The relationship between chemical structures in polymers with TC. (a) Calculated Person correlation coefficients between molecular structures of polymer repeating units and TC. (b) Molecular substructures diagram. (c), (d) Polymers with different molecular structures and their TC.

(d). Fig. 7(c) shows that polymers with rigid skeleton could obtain higher TC, such as these containing carbon-oxygen double bond and π-bond. The reason for that is the rigid main chain has a higher bond strength to inhibit rotation of the segments. At the molecular level, strong bonds usually enable high phonon group velocities, and stiffer dihedral angle tends to suppress the segmental rotation, can enable better thermal stability and higher TC. In the PCFF, bond stretching, angle bending strengths and the dihedral angle torsion are defined by the energy constants in the following formula

$$E_{\text{bond}} = \sum_{n=2}^{4} K_{bn}(b - b_o)^n \tag{1}$$

$$E_{\text{angel}} = \sum_{n=2}^{4} K_{an}(\theta - \theta_o)^n \tag{2}$$

$$E_{\text{dihedral}} = \sum_{n=1}^{3} K_{dn}[1 - \cos(n\Phi)] \tag{3}$$

where b is the distance between the two bonded atoms; b_o is equilibrium bond distance; and K_{bn} are the energy constants defining the strength of the bond. θ is the angle value; θ_o is equilibrium angle; and K_{an} are the energy constants defining the strength of the angle. Φ is the dihedral angle value; K_{dn} are the dihedral angle energy constant.

Functional groups affect the TC by influencing the morphological structure of the polymer backbone. Table 1 shows the PCFF energy constants and the bond energy of various chemical groups. The carbon-oxygen double bond (C=O) provides the highest bond energy of 728 kJ/mol and energy constants, hence polymers containing this group shows the maximum TC. The energy constant distribution and bond energy of C_p-C_p and C-C_p in π-conjugated polymers are second only to the carbon-oxygen double bond, whose Pearson correlation coefficient is lower than polymers containing carbon-oxygen double bond and higher than others. Carbon-nitrogen bond and carbon-chlorine bond present similar bond energies. Moreover, carbon-nitrogen bond has higher bond energy constants and dihedral constants, while the angular constant is slightly lower compared to carbon-chlorine bond. As a result, polymers with -Cl, -NH_2 groups show similar TC and Pearson correlation coefficient. The TC of polymer contains -CH_2NH_2 group is usually lower, which is attributed to the complex group structure with a larger atomic mass. The above results indicate that small atomic masses, higher stiffness of the backbone (bond stretching and angle bending strengths) and the dihedral angle torsion usually give rise to high TC. At the molecular level, strong bonds usually enable high phonon group velocities, and stiffer dihedral angle tends to suppress the segmental rotation, which leads to better

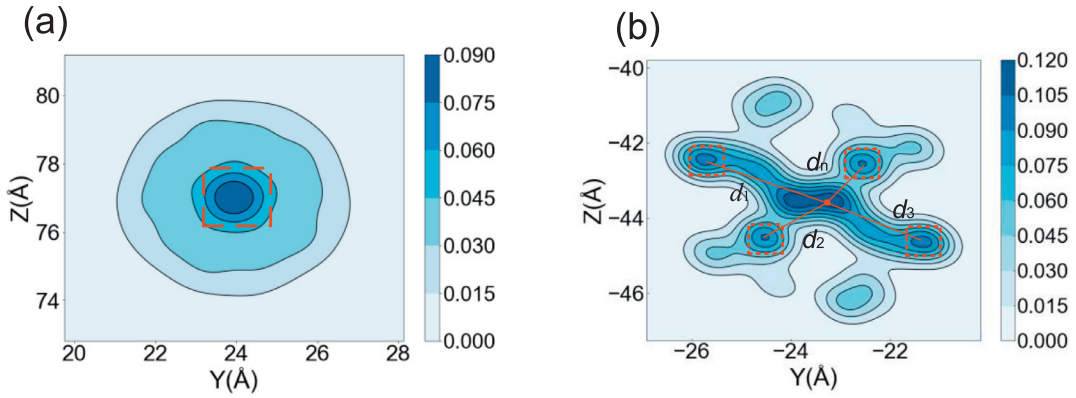

Fig. 8. The definition of PDE(y, z) and related parameters. (a) Contour plot of PDE(y, z) (where (y, z) is the y- and z-Cartesian coordinates of atoms in a given single polymer chain along x-axis) of cartesian coordinate points (y, z) of all atoms in a polymer chain. Larger PDE(y, z) in the color bar represents a larger overlap level of atoms when projecting onto the y–z plane. The red dashed square contains mostly backbone atoms.(b) The distances (d) between the region of interest and the center of the coordinate system. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

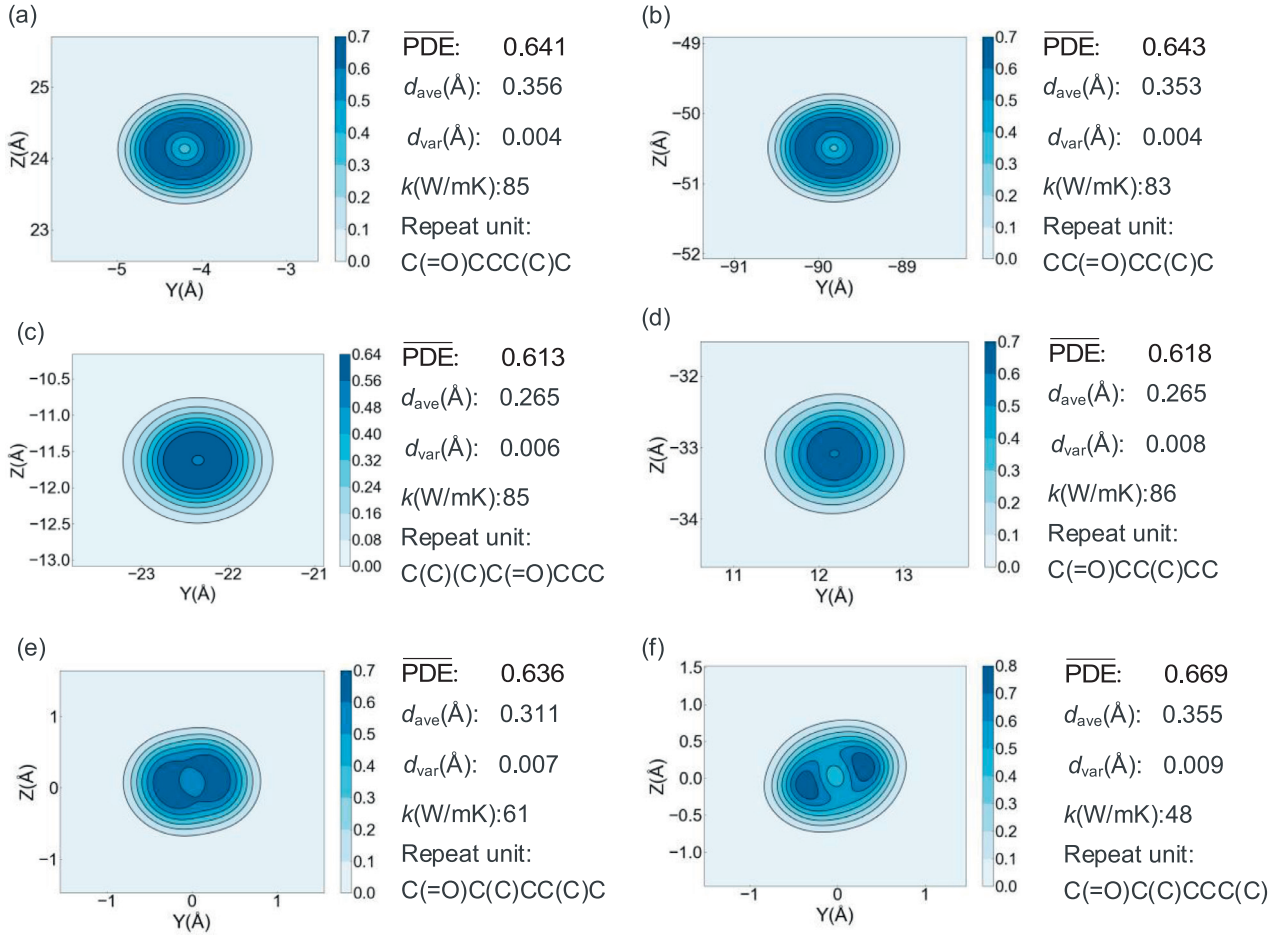

Fig. 9. $\overline{PDE(y,z)}$, d_{ave}, d_{var}, and thermal conductivity for the polymer chains.

thermal stability and higher TC. Fig. 7(d) shows the influence of different atoms in the molecular chain on the TC. Compared with nitrogen atoms, the existence of oxygen atoms results in higher TC. The reason for that is the bond energy of carbon-oxygen bond (~326kJ/mol) is higher than carbon-nitrogen bond (~305kJ/mol), which can be observed in Table 1.

The spatial conformation of polymer chain plays an important role in phonon scattering process, thereby affecting the TC. To further explore the influence of the backbone structure on the TC, probability density estimate (PDE) of atoms were defined to describe the spatial distribution of atoms in the polymer chains, as shown in Fig. 8. The PDE(y, z) in Fig. 8 denotes the probability density estimate for each atom (y, z) based on Kernel density estimator [49], which is a nonparametric way to estimate the probability of having an atom at (y, z). Then three indices including $\overline{PDE(y,z)}$, d_{ave}, and d_{var} were defined to represent the distribution of atoms. $\overline{PDE(y,z)}$ was calculated by averaging over the density estimate within the region where contains almost all backbone atoms, such

Table 1
Typical PCFF energy constants and the bond energy parameters.

Chemical bond	Leading bond energy constant K_{b2} (kcal/mol)	Leading angle energy constant K_{a2} (kcal/mol)	Dihedral angle energy constant K_{d1}/ K_{d2}/ K_{d3} (kcal/mol)	Bond energy (kJ/mol)
C=O	823.79	65.10	0.70/ 1.21/ -0.04	~728
C(=O)−C	312.37	43.96	0.29/1.02/ -0.11	~332
C_p−C_p	470.84	61.02	8.37/ 1.19/0.00	>332
C−C_p	321.90	43.96	0.00/0.00/0.00	~332
C−Cl	194.32	61.17	0.00/ 0.00/ 0.15	~328
C−N	365.81	60.71	0.18/ 0.18/ -0.52	~305
C−O	400.40	54.54	0.71/0.27/-0.25	~326
C−C	299.67	39.52	0.00/0.05/-0.14	~332

as the area contained in the red dashed square shown in Fig. 8(a). Thermal energy is defined by the microscopic vibrations of particles [50], therefore, the region of interest containing almost all the backbone atoms is the main heat carrier region. Higher $\overline{\mathrm{PDE}(y,z)}$ indicates that atoms in the polymer chain skeleton are more concentrated, and the atoms vibrate more violent when heated, leading to the improvement of heat transfer efficiency. In addition to focusing on the averaged overlap level of (y, z) coordinates in the region of interest represented by $\overline{\mathrm{PDE}(y,z)}$, it is also necessary to consider the distribution of the region of interest on the y-z plane. As shown in Fig. 8(b), the distance d between the region of interest and the center of the coordinate system was calculated by the Eq. (4). Then the indices d_{ave} and d_{var} were used to represent the concentration degree and the discrete level of the region of interest, respectively. d_{ave} is the average of the distances, d_{var} is the variances of the distances. It is well understanding that the more concentrated and uniform the spatial distribution of single-chain polymers, the smaller the corresponding indices d_{ave} and d_{var} are.

$$d = \frac{1}{m}\sum_{j}^{m}\sqrt{(y_j - y_0)^2 + (z_j - z_0)^2} \tag{4}$$

where m is the number of coordinate points on the region of interest; (y_j, z_j) is the coordinate of the point; (y_0, z_0)is the central coordinates of the PDE contour plot.

Taking the polymers with -(=O) group with -C groups as an example, we carried out correlation analysis between the above parameters and TC of polymer chains, and some results are shown in Fig. 9. The PDEs of more polymers can be found in Table S2, Supporting Information. It can be seen that the $\overline{\mathrm{PDE}(y,z)}$ and d_{ave} values of the polymers in Fig. 9 a, b are similar to Fig. 9 e, f, while Fig. 9 e, f exhibit lower TC accompanied by the higher d_{var} values. This indicates that higher d_{var} values have a negative correlation with TC. That is, the dispersion of the main heat carrier region leads to the decrease of TC. Compared to the polymers in Fig. 9 e, f, the TC of first four polymers (Fig. 9 a-d) are similar and higher, in which the excellent TC of Fig. 9 a, b benefit from the higher $\overline{\mathrm{PDE}(y,z)}$ and lower d_{var} values. Polymers in Fig. 9 c, d show higher d_{var} values, which is detrimental to the TC of the polymer, but the lower d_{ave} values result in excellent TC performance. This suggests that concentrated chain skeleton atoms are favorable for the heat transfer.

4. Conclusions

A ML framework was presented to predict the TC of polymer chains, in which the training data come from the high-throughput MD simulations. The ML methods could effectively model the relationships between material structure and properties, and the trained models present high prediction accuracy for TC, among which CNN model shows the best performance with R-squared value of 0.93. Both the chemical groups and spatial conformation of molecules plays an important role in phonon scattering process, thereby affecting the TC. Polymers containing functional groups with strong bond strength such as -(=O) structure generally give rise to high TC, owing to the inhibited rotation of the segments. And polymer chains with well-ordered spatial structures usually present higher TC.

Supporting information

The following files are available free of charge.

Detailed information on basic schematic of molecular dynamics, ML training process and the $\overline{\mathrm{PDE}(y,z)}$, d_{ave}, d_{var} and TC results of the polymers with -(=O) group. (PDF)

Data availability

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

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Ming-Xiao Zhu: Conceptualization, Methodology, Funding acquisition, Writing - original draft. **Heng-Gao Song:** Data curation, Writing - original draft, Software. **Qiu-Cheng Yu:** Writing - review & editing. **Ji-Ming Chen:** Supervision. **Hong-Yu Zhang:** Software.

Acknowledgements

This work was supported by the Natural Science Foundation of Shandong Province of China[ZR2019QEE014]; and the Fundamental Research Funds for the Central Universities [19CX02015A].

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.ijheatmasstransfer.2020.120381.

References

[1] T.J. Zhu, Y.T. Liu, C.G. Fu, J.P. Heremans, J.G. Snyder, X.B. Zhao, Compromise and synergy in high-efficiency thermoelectric materials, Adv. Mater. 29 (2017) 1605884 201729 (30), 1.
[2] D. Zhang, Y.H. Wang, Y. Yang, Design, performance, and application of thermoelectric nanogenerators, Small 15 (32) (2019) 13.
[3] L. Yang, Z.G. Chen, M.S. Dargusch, J. Zou, High performance thermoelectric materials: progress and their applications, Adv. Energy Mater. 8 (6) (2018) 28.

[4] T.F. Luo, K. Esfarjani, J. Shiomi, A. Henry, G. Chen, Molecular dynamics simulation of thermal energy transport in polydimethylsiloxane (PDMS), J. Appl. Phys. 109 (7) (2011) 6.
[5] R.Y. Wang, R.A. Segalman, A. Majumdar, Room temperature thermal conductance of alkanedithiol self-assembled monolayers, Appl. Phys. Lett. 89 (17) (2006) 3.
[6] Z.H. Wang, J.A. Carter, A. Lagutchev, Y.K. Koh, N.H. Seong, D.G. Cahill, D.D. Dlott, Ultrafast flash thermal conductance of molecular chains, Science 317 (5839) (2007) 787–790.
[7] C. Han, Z. Li, S. Dou, Recent progress in thermoelectric materials, Chin. Sci. Bull. 59 (18) (2014) 2073–2091.
[8] X.F. Xu, J. Chen, J. Zhou, B.W. Li, Thermal conductivity of polymers and their nanocomposites, Adv. Mater. 30 (17) (2018) 10.
[9] H. Ma, Z.T. Tian, Effects of polymer topology and morphology on thermal transport: a molecular dynamics study of bottlebrush polymers, Appl. Phys. Lett. 110 (9) (2017) 5.
[10] Y. Xu, X. Wang, J. Zhou, B. Song, Z. Jiang, E.M.Y. Lee, S. Huberman, K.K. Gleason, G. Chen, Molecular engineered conjugated polymer with high thermal conductivity, Sci. Adv. 4 (3) (2018) eaar3031.
[11] T. Zhang, X. Wu, T. Luo, Polymer nanofibers with outstanding thermal conductivity and thermal stability: fundamental linkage between molecular characteristics and macroscopic thermal properties, J. Phys. Chem. C 118 (36) (2014) 21148–21159.
[12] J. Yu, B. Sundqvist, B. Tonpheng, O. Andersson, Thermal conductivity of highly crystallized polyethylene, Polymer 55 (1) (2014) 195–200.
[13] J. Zhao, A.C. Tan, P.F. Green, Thermally induced chain orientation for improved thermal conductivity of P(VDF-TrFE) thin films, J. Mater. Chem. C 5 (41) (2017) 10834–10838.
[14] H. Chen, V.V. Ginzburg, J. Yang, Y. Yang, W. Liu, Y. Huang, L. Du, B. Chen, Thermal conductivity of polymer-based composites: fundamentals and applications, Progr. Polym. Sci. 59 (2016) 41–85.
[15] C.L. Huang, X. Qian, R.G. Yang, Thermal conductivity of polymers and polymer nanocomposites, Mater. Sci. Eng. R-Rep. 132 (2018) 1–22.
[16] H. Ma, Z. Tian, Chain rotation significantly reduces thermal conductivity of single-chain polymers, J. Mater. Res. 34 (1) (2019) 126–133.
[17] J. Liu, R. Yang, Length-dependent thermal conductivity of single extended polymer chains, Phys. Rev. B 86 (10) (2012) 104307.
[18] T. Zhang, T. Luo, Morphology-influenced thermal conductivity of polyethylene single chains and crystalline fibers, J. Appl. Phys. 112 (9) (2012) 094304.
[19] J. Liu, S. Ju, Y. Ding, R. Yang, Size effect on the thermal conductivity of ultrathin polystyrene films, Appl. Phys. Lett. 104 (15) (2014) 153110.
[20] X. Wang, V. Ho, R.A. Segalman, D.G. Cahill, Thermal conductivity of high-modulus polymer fibers, Macromolecules 46 (12) (2013) 4937–4943.
[21] K. Vu, J.C. Snyder, L. Li, M. Rupp, B.F. Chen, T. Khelif, K.-R. Mueller, K. Burke, Understanding kernel ridge regression: common behaviors from simple functions to density functionals, Int. J. Quant. Chem. 115 (16) (2015) 1115–1128.
[22] J. Schmidhuber, Deep learning in neural networks: an overview, Neural Netw. 61 (2015) 85–117.
[23] A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet classification with deep convolutional neural networks, Commun. ACM 60 (6) (2017) 84–90.
[24] D.P. Tabor, L.M. Roch, S.K. Saikin, C. Kreisbeck, D. Sheberla, J.H. Montoya, S. Dwaraknath, M. Aykol, C. Ortiz, H. Tribukait, C. Amador-Bedolla, C.J. Brabec, B. Maruyama, K.A. Persson, A. Aspuru-Guzik, Accelerating the discovery of materials for clean energy in the era of smart automation, Nat. Rev. Mater. 3 (5) (2018) 5–20.
[25] R. Gomez-Bombarelli, J. Aguilera-Iparraguirre, T.D. Hirzel, D. Duvenaud, D. Maclaurin, M.A. Blood-Forsythe, H.S. Chae, M. Einzinger, D.G. Ha, T. Wu, G. Markopoulos, S. Jeon, H. Kang, H. Miyazaki, M. Numata, S. Kim, W.L. Huang, S.I. Hong, M. Baldo, R.P. Adams, A. Aspuru-Guzik, Design of efficient molecular organic light-emitting diodes by a high-throughput virtual screening and experimental approach, Nat. Mater. 15 (10) (2016) 1120-+.
[26] M.L. Green, C.L. Choi, J.R. Hattrick-Simpers, A.M. Joshi, I. Takeuchi, S.C. Barron, E. Campo, T. Chiang, S. Empedocles, J.M. Gregoire, A.G. Kusne, J. Martin, A. Mehta, K. Persson, Z. Trautt, J. Van Duren, A. Zakutayev, Fulfilling the promise of the materials genome initiative with high-throughput experimental methodologies, Appl. Phys. Rev. 4 (1) (2017) 011105.
[27] A. Mannodi-Kanakkithodi, A. Chandrasekaran, C. Kim, H. Tran Doan, G. Pilania, V. Botu, R. Ramprasad, Scoping the polymer genome: a roadmap for rational polymer dielectrics design and beyond, Mater. Today 21 (7) (2018) 785–796.
[28] J. Carrete, W. Li, N. Mingo, S. Wang, S. Curtarolo, Finding unprecedentedly low-thermal-conductivity Half-Heusler semiconductors via high-throughput materials modeling, Phys. Rev. X 4 (1) (2014) 011019.
[29] H. Yang, Z. Zhang, J. Zhang, X.C. Zeng, Machine learning and artificial neural network prediction of interfacial thermal resistance between graphene and hexagonal boron nitride, Nanoscale 10 (40) (2018) 19092–19099.
[30] 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, R. Yoshida, Machine-learning-assisted discovery of polymers with high thermal conductivity using a molecular design algorithm, Npj Comput. Mater. 5 (2019) 5.
[31] P.K. Schelling, S.R. Phillpot, P. Keblinski, Comparison of atomic-level simulation methods for computing thermal conductivity, Phys. Rev. B 65 (14) (2002) 144306.
[32] C. Oligschleger, J.C. Schön, Simulation of thermal conductivity and heat transport in solids, Phys. Rev. B 59 (6) (1999) 4125–4133.
[33] D. Rogers, M. Hahn, Extended-connectivity fingerprints, J. Chem. Inf. Model. 50 (5) (2010) 742–754.
[34] Y. Hu, E. Lounkine, J. Bajorath, Improving the search performance of extended connectivity fingerprints through activity-oriented feature filtering and application of a bit-density-dependent similarity function, ChemMedChem 4 (4) (2009) 540–548.
[35] S. Kim, P.A. Thiessen, E.E. Bolton, J. Chen, G. Fu, A. Gindulyte, L. Han, J. He, S. He, B.A. Shoemaker, J. Wang, B. Yu, J. Zhang, S.H. Bryant, PubChem substance and compound databases, Nucl. Acids Res. 44 (D1) (2016) D1202–D1213.
[36] H. Tran Doan, S. Boggs, G. Teyssedre, C. Laurent, M. Cakmak, S. Kumar, R. Ramprasad, Advanced polymeric dielectrics for high energy density applications, Progr. Mater. Sci. 83 (2016) 236–269.
[37] A. Mannodi-Kanakkithodi, G. Pilania, H. Tran Doan, T. Lookman, R. Ramprasad, Machine learning strategy for accelerated design of polymer dielectrics, Sci. Rep. 6 (2016) 20952.
[38] H. Tran Doan, A. Mannodi-Kanakkithodi, C. Kim, V. Sharma, G. Pilania, R. Ramprasad, A polymer dataset for accelerated property prediction and design, Sci. Data 3 (2016) UNSP 160012.
[39] V. Sharma, C. Wang, R.G. Lorenzini, R. Ma, Q. Zhu, D.W. Sinkovits, G. Pilania, A.R. Oganov, S. Kumar, G.A. Sotzing, S.A. Boggs, R. Ramprasad, Rational design of all organic polymer dielectrics, Nat. Commun. 5 (2014) 4845.
[40] S. Plimpton, Fast parallel algorithms for short-range molecular dynamics, J. Comput. Phys. 117 (1) (1995) 1–19.
[41] H. Sun, S.J. Mumby, J.R. Maple, A.T. Hagler, An ab Initio CFF93 all-atom force field for polycarbonates, J. Am. Chem. Soc. 116 (7) (1994) 2978–2987.
[42] D. Weininger, SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules, J. Chem. Inf. Comput. Sci. 28 (1) (1988) 31–36.
[43] D. Weininger, A. Weininger, J.L. Weininger, SMILES. 2. Algorithm for generation of unique SMILES notation, J. Chem. Inf. Comput. Sci. 29 (2) (1989) 97–101.
[44] RDKit: Open-source cheminformatics. http://www.rdkit.org.
[45] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: machine learning in python, J. Mach. Learn. Res. 12 (2011) 2825–2830.
[46] Y.-p. Lin, M.-y. Zhang, Y.-f. Gao, L.-y. Mei, Y.-z. Fu, Y.-q. Liu, Effect of stretching on the thermal conductivity of single polyethylene chains by molecular dynamics simulations, Acta Polymerica Sinica (6) (2014) 789–793.
[47] G.-J. Hu, B.-Y. Cao, Y.-W. Li, Thermal conduction in a single polyethylene chain using molecular dynamics simulations, Chin. Phys. Lett. 31 (8) (2014) 086501.
[48] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, IEEE, going deeper with convolutions, in: IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1–9. 2015.
[49] P.D. Hill, Kernel estimation of a distribution function, Commun. Stat.-Theory Methods 14 (3) (1985) 605–620.
[50] N. Burger, A. Laachachi, M. Ferriol, M. Lutz, V. Toniazzo, D. Ruch, Review of thermal conductivity in composites: mechanisms, parameters and theory, Progrm Polym. Sci. 61 (2016) 1–28.