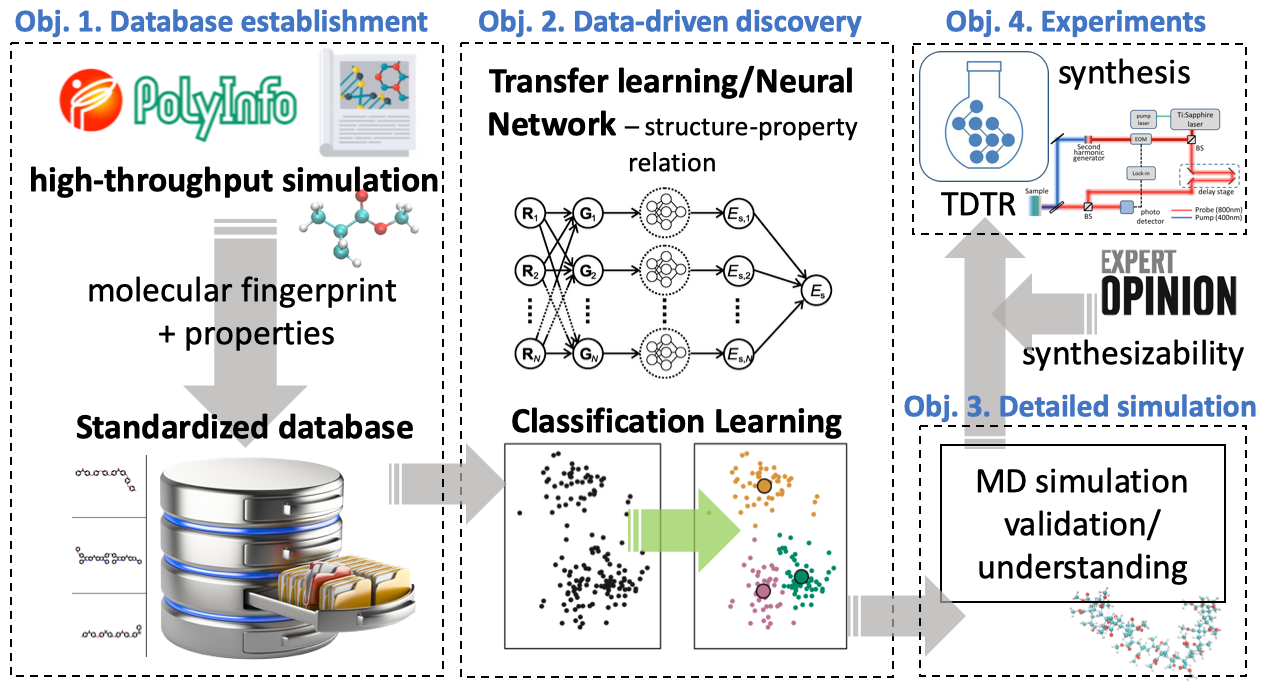
**Polymer Informatics for Developing Thermally Conductive Polymers**

# 1. Goals and Objectives

Bulk amorphous polymers are usually thermal insulators with thermal conductivity in the range of 0.1-0.5 W/mK. However, they are largely used in heat transfer applications such as plastic heat exchangers and electronics thermal packaging. Industry has resorted to compositing with thermally conductive fillers (*e.g.*, ceramics and carbon) to improve polymer thermal conductivity, but a bottleneck is still the low thermal conductivity of the polymer matrices. Theoretical calculation has pointed out that a composite thermal conductivity greater than 20 W/mK is more likely to be achieved if the polymer matrix has an intrinsic thermal conductivity larger than 1 W/mK.1 The ultimate **goal** of this project is to establish a Materials Informatics approach facilitated by physical understanding to predict and design amorphous polymers with high thermal conductivity (> 1 W/mK). To reach this goal, the **objectives** of this project are (**Fig. 1**): **(1)** establish a standardized polymer database by combining data from existing database (*e.g.*, PolyInfo), open literature (*e.g.*, papers, reports and websites) and high-throughput molecular dynamics (MD) simulations; **(2)** employ a unique Machine Learning (ML) technique called Transfer Learning to develop accurate surrogate models that map out structure-property relations for polymer chemistry and thermal conductivity, and use classification and clustering to understand molecular features impacting thermal conductivity; **(3)** use the established model to predict ~30 polymers with high thermal conductivity and perform detailed MD simulations to calculate the thermal conductivity to verify the prediction and understand the thermal transport physics; **(4)** down-select 10 polymers with the expert opinion from collaborating polymer chemists on the synthesizability of the predicted polymers, synthesize them and measure their thermal properties.

**Figure 1**. A schematic summary of the proposed research and methods.



# 2. Intellectual Merit and Broader Impacts

## 2.1. Intellectual Merit

This project represents the first systematic attempt to understand the structure-property relation for polymers and their thermal conductivity using Materials Informatics, which has been largely hindered by the lack of publicly available thermal conductivity data. It unites all four paradigms of materials science (empirical, theoretical, computational and data-driven) to tackle a grant challenge in thermal science. The expected outcomes are (1) the critical understanding of the key molecular features impacting polymer thermal conductivity; (2) a prediction and validation of a number (~5) of thermally conductive polymers; and (3) a Polymer Informatics framework which can be extended to design molecules with other properties of interest (*e.g.*, mechanical strength, electrical conductivity). The PIs will collaborate with the University of Tokyo to leverage the novel Transfer Learning technique to largely mitigate the risk related to the lack of thermal conductivity data at the beginning of the project. This new technique allows the transfer of ML models (*e.g.*, neural network (NN)) trained on highly populated targets (*e.g.*, heat capacity, glass transition temperature) to improve the accuracy of models for scarcely populated properties (*e.g.*, thermal conductivity). This project will also leverage PI Luo’s long-term experience in MD simulation and experimental study of polymer thermal transport and co-PI Jiang’s specialty on representation learning and Transfer Learning.

## 2.2. Broader Impact

Electronics, energy and healthcare are among the most important growth fields of today’s world. Heat conduction related to polymers largely exists in these areas as significant challenges but also unique opportunities. For example, polymers are critical components in the thermal management of electronics, but their low thermal conductivity is one of the limiting factors in device cooling. The data-driven approach to be established in this project will be the first of its kind in the field concerning thermally conductive polymers. The established protocol can be followed to design polymers with other desirable properties beyond thermal conductivity. This project will lead to a Polymer Informatics-based approach that can save the cost and time traditionally needed in new material development. This project will also help establish international collaborations between the U.S. and Japanese research institutions, which are leading in many aspects in the application of ML for thermal science. The data collected and generated from this project will be made available on open-source data servers. With a preliminary agreement, the PIs will also explore student internship opportunities at DuPont in the duration of this project to enrich the database and transferability of the research to industry. Besides advancing knowledge and technology, the proposed research will also enable the education and training of graduate students and undergraduate students from under-represented groups at ND. It will also promote international student exchange. The topics of this project and the expected academia-industry collaboration will cultivate future work force for the U.S. manufacturing industry. Besides education in the universities, ND PIs will outreach to the local Riley High School, which has a large portion (48%) of minority students, mostly African-American and Latino. Students involved in this research will also participate in the South Bend Science Alive event.

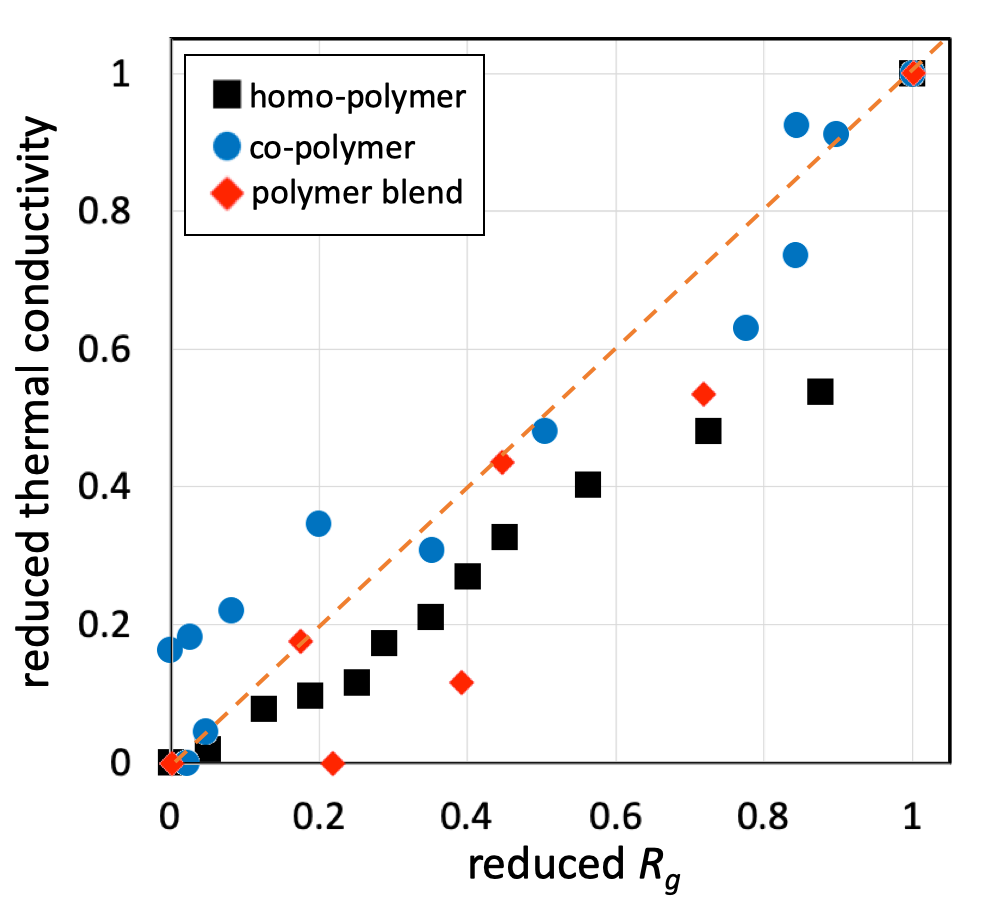
# 3. Background and Prior Results Motivating This Research

## 3.1. Heat Conduction in Disordered Materials

In non-metal crystals, heat conduction is dominated by phonons. Their transport has been extensively studied and well-understood.2-6 Heat conduction in polymer can be fundamentally different. The most dominant feature of bulk amorphous polymer is their disordered atomistic arrangements. Such lack of periodicity makes even defining wavevector for phonon modes difficult. For disordered materials, Kittel7 used a simple kinetic formula () to describe the thermal conductivity, where the mean free path (MFP), *l*,was set to the atomic spacing by assuming that the morphology scattering makes the phonon MFP similar to the interatomic bond length. A similar picture was proposed by Slack,8 Cahill and Pohl9, 10 based on Einstein’s thermal random walk model,11 in which they assumed that the heat carriers’ propagation length is limited to half of their wavelengths in highly disordered materials. Both models assume that heat carriers travel at the speed of sound. While using sound speed can be a good approximation for the group velocity of very low frequency modes, these modes are not necessarily the dominant heat carriers in amorphous materials.12 Allen and Feldman12 categorized the heat carriers in amorphous materials into three kinds depending on their nature: propagons(phonon-like propagating wave delocalized over large distances), diffusons(extended vibration modes delocalized over a short distances)and locons(highly localized modes)*.* According to their calculation, diffusons, rather than the long wavelength propagons, are the most dominant heat carriers at room temperatures in disordered materials.12

## 3.2. Conformation-Influenced Thermal Conductivity of Amorphous Polymers

The minimum thermal conductivity model derived by Cahill and Pohl,9, 10 which has shown success for inorganic amorphous materials, has been used to calculate the thermal conductivity of different polymers, but it consistently over-estimate the values compared to experiments13 or MD simulations.14 One important feature of polymers compared to inorganic amorphous materials is their local anisotropic bonding environment consisting of both covalent bonding and non-bonding interactions like van der Waals (vdW) and electrostatic interactions. By changing the global morphology from amorphous to highly aligned crystal structures through external strain, the thermal transport becomes dominated by the strong covalent polymer backbone and the thermal conductivity can be increased by as much as three orders of magnitude.15-23 However, in the cross-chain direction, the thermal conductivity, which is dominated by the weak non-bonding interactions, is still low.24 Such global morphology-influenced thermal conductivity has been studied recently by a number of groups, including ours.20, 25-28



**Figure 2.** The relationship between reduced thermal conductivity and reduced radius of gyration of different types of polymers in amorphous states. Reduced thermal conductivity , where and are respectively the minimal and maxium thermal conductivity in each type of polymer. Similary, reduced radius of gyration .

However, such highly aligned crystalline polymers are not ideal for a wide range of applications, where amorphous polymers dominate. Our group has also obtained critical insights into the thermal transport of amorphous polymers. In recent MD simulations, we have shown that the thermal conductivity of amorphous polymer is closely related to the chain confirmation.29-31 We have observed that thermal transport along the covalent chain backbone dominates the thermal conductivity of amorphous polymers, which is found to be related to the chain conformation, especially its spatial extension as characterized by the radius of gyration (*Rg*).29 We have observed positive correlations between *Rg* and thermal conductivity for homopolymers,29 polymer blends30 and bi-/tri-block copolymers31 (**Fig. 2**). Through a parameteric study, we have further shown that by tuning the chain backbone stiffness, the *Rg* can be increased and the thermal conductivity increases accordingly.29 These results suggest that finding stiff polymer chains may help identify polymers with high thermal conductivity. Our preivous simulation has predicted that π-conjugated polymers can have stiffer backbone, leading to higher thermal conductivity.27 Other research groups have also contributed to the understanding of thermal transport in polymers using MD simulations.32

An experimental study by Singh *et al*.33 shows that polythiophene (PT), a π-conjugated polymer, can have a large thermal conductivity (~4 W/mK) in the amorphous state. This was attributed to the stiff backbone of PT, which keeps the chain straight over a long distance. In another study by Xu *et al.*,34 it is shown that the strong backbone of poly(3-hexylthiophene) leads to a high thermal conductivity of 2.2 W/mK in a oxidative-CVD synthesized amorphous film. In addition to the strong backbone, the authors also argued that the strong π-π stacking facilitated inter-chain thermal transport. In a recent simulation (not yet published), we observe that the π-π stacking does not directly enhance thermal transport across chains, but instead, it helps straighten chains and lead to enhanced intra-chain thermal transport along the covalent backbone.

There are also other factors that might impact the thermal conductivity of amorphous polymers. In a recent study by Shanker *et al.*,35 it is found that the thermal conductivity of polyeletrolyte increases with the ionic strength indicated by the pH, and a thermal conductivity of ionized polyacrylic acid (PAA) can reach ~1.2 W/mK. The authors attribute this observation to that the electrostatic repulsion between neighboring ionized side groups on the same backbone pushes the chain straighter and thus lead to larger spatial extension of the backbone (*Rg*), resulting in higher thermal conductivity. However, our recent simulations show that the *Rg* is not obviously impacted by ionization since the electrostatic repulsion is largely screened by the electrostatic attraction with the condensated ions. Instead, we find that these couterions work as bridges to enhance thermal transport between the side chain and the ions.14, 36

As discussed above, different factors that can influence polymer thermal conductivity, and they are rooted in the molecular chemistry. Mapping out the relation between polymer chemistry and thermal conductivity is thus of critical significance to understanding thermal transport in polymers.

## 3.3. Polymer Informatics

Materials Informatics using ML techniques is an important component towards the eventual realization of material by design.37 A common goal of different Materials Informatics studies is to identify structure-property relations to singificantly reduce both the time and cost of developing new materials with desirable properties. Actually, Materials Informatics has been recently applied to thermal conductivity,38 especially for designing inorganic semiconductors with low thermal conductivity, mainly motivated by the demand of high performance thermoelectrics.39-42 However, Materials Informatics concerning polymers (*i.e.*, Polymer Informatics) has been hindered by the lack of data in a unified format.43 Unlike inorganic crystals which have well-defined lattice structures, it is challenging to produce large datasets for amorphous polymers using high-throughput simulations due to the uncertainty and difficulty in generating reasonable amoprhous structures.

In a recent study, ML techniques was used to contruct a surrogate model between polymer chemistry represented by SMILES (Simplified molecular-input line-entry system) strings and thermal conductivity.44 Interestingly, they were able to obtain a deep neural network (DNN) model with reasonable accuracy (*R2*=0.73) using merely 28 training data points in thermal conductivity available from the PolyInfo database. This was realized by implementing a novel technique called Transfer Learning, which “transfers” knowledge learned from training against highly populated proxy labels (in their case, ~6000 data points on glass transition temperature, *Tg*) into constructing DNN for sparsely populated labels (*i.e.*, thermal conductivity). By directly training a DNN against the 28 thermal conducitivty data, *R2* was -0.4 (*i.e.*, no prediction power), but Transfer Learning helped improve the model accuracy to *R2*=0.73. The essence of Transfer Learning is that different properties are different expressions of the inherent chemistry of materials. However, the transferred DNN is still much inferior compared to the model for the highly populated label, *Tg* (*R2*=0.92). It is our belief that Transfer Learning can further improve the model accuracy if more thermal conductivity labels can be leveraged in the training.

In addition, *Tg* might not be the only proxy label that can be used for Transfer Learning. Other intermediate labels, such as *Rg*, modulus, effective charges, and even force field parameters,45 which can be calculated more quickly than thermal conductivity, may also provide a bridge to construct more accurate thermal conductivity surrogate models *via* Transfer Learning. Besides ML models, our recent findings showed that properly representing polymer molecules can be important for constructing accurate surrogate models, since representations describe the chemistry of molecules.46 We found that a deep learning representation scheme (*i.e.*, Mol2Vec) based on the Natural Language Process (NLP) algorithm was better than the conventionally used Morgan Fingerprint, a one-hot encoding scheme based on the simple chemical connections. These previous studies from our collabortors and ourselves motivate us to look at Polymer Informatics holistically and explore different aspects of the framework (*e.g.*, database, molecular representation, ML models, Transfer Learning, proxy labels) for designing thermally conductive polymers.

## 3.4. Fundamental Research Needs

From the above discussions, the scientific questions remain to be answered are: (1) What are the molecular level features that impact the thermal transport properties? and (2) How to develop highly accurate structure-property relation models that can be used to guide the design of thermally conductive polymers?

# 4. Project Plan

The research in this project is designed to answer the above-stated fundamental questions through a combination of ML techniques, detailed MD simulations and experimental validations. The project is guided by four objectives, and the corresponding tasks are discussed below.

## 4.1 Objective 1: Database Development

As previously stated, the largest hurdle for the implementation of Polymer Informatics is the lack of a standardized database that includes the target properties of interest43 – in our case – thermal conductivity. In this objective, we plan to use three parallel thrusts to enrich polymer data. These include data collection from existing database, open literature and high-throughput MD simulations.

### Task 1.1. Data Collection from Existing Database (Luo)

Currently, the lack of polymer data is especially true for thermal conductivity. There are a few open-source polymer databases, such as PolyInfo,47 Polymer Genome,48 and PolymerDatabase.49 While they do contain many properties, very few thermal conductivity data are available. The only database that has thermal conductivity data is PolyInfo, but it has merely 28 of them. However, these databases are still of value since (1) they provide some, although limited, thermal conductivity values; (2) they include chemical formula of more than 10,000 realistic polymers; (3) they have other highly populated labels, like glass transition temperature (*Tg*) that can be leveraged as proxy labels to perform Transfer Learning for thermal conductivity targets. We have developed a semi-automatic Application Programming Interface (API) to download and convert the chemical formula from PolyInfo into ML-compatible representation using a unsupervised deep learning scheme, Mol2vec.50 The multi-dimensional tensor generated from Mol2vec is unique to each polymer, which has considered chemical composition, atomic connectivity and other information that is needed to determine a specific polymer. Such tensors can be used as direct inputs for ML. In this process, we are also able to convert the chemical formula into atomic coordinate files which can be used in MD simulations to generate property data in a high-throughput manner in Task 1.3. Using the above-mentioned API, we have established a standardized database of ~2,000 homo-polymers, including SMILES strings, molecular representations, potential proxy properties (*e.g.*, *Tg* and *Tm*) and a limited number of thermal conductivity data.46 In this task, we will extend the API to be compatible with other existing database and continue to enrich the data in our standardized database.

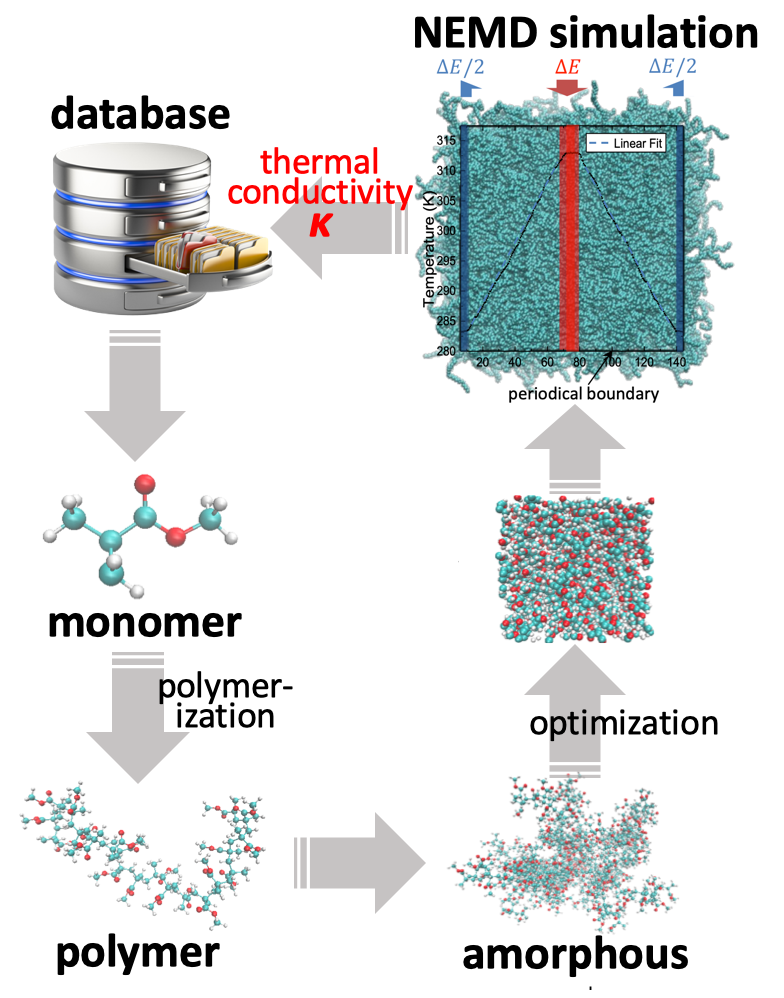
### Task 1.2. Data Collection from Open Literature (Jiang and Luo)

Besides the available open-source databases, we will also collect thermal conductivity data from the open literature, including journal papers, reports, catalogs and *etc.* The scientific literature including chemical properties from simulations and experiments contain many relevant polymer property data, but with widely varying coverage, accuracy and precision. We will apply methods encapsulated in available open source tools51, 52 to locate and extract relevant data from the literature using a mix of fully automated and hybrid human-computer methods. The data extraction tools will leverage a range of existing algorithms (*e.g.*, named entity recognition, NER, and semantic role labeling, SRL) to enable rapid extraction of data from a variety of textual sources. The most challenging data source to work with is from unstructured sources, mainly from journal articles in PDF format, such as from the *Journal of Physical and Chemical Reference Data* and *ACS Macromolecules*. For data contained in such sources, we anticipate employing a mix of simple text searches to find references to known molecules and NER methods based on Word2vec53 representations to find references to other molecules used in similar contexts, and then NLP methods to extract desired properties. We will use a hybrid Information Extraction pipeline that combines automation and crowdsourcing in ways that leverage the complementary strengths of computational modules and humans. This pipeline first extracts candidate properties automatically and subsequently assigns tasks to humans to accept or reject candidate properties, with the goal of maximizing throughput and accuracy while minimizing human curators. We will use Word2Vec to generate context-aware word vectors for every word in a corpus of polymer publications. We will then use those customized word vectors to build a classifier capable of differentiating between polymers and other words using the word vector coordinates as features. The resulting classifier will then be integrated with other elements of an NLP pipeline to extract references to desired molecular properties. A graduate student will be charged to construct this framework, and human curation can be performed by undergraduate researchers to be funded by supplemental REU requests to NSF.

### Task 1.3 Data Generation from High-Throughput MD Simulations (Luo)

At the beginning of this project, it is expected that the thermal conductivity data available for this project is limited. To largely mitigate the limitation related to the volume of data, we will use MD simulations to calculate the thermal conductivity. The molecular structures for these high-throughput calculations will mainly be those collected and converted from existing polymer databases (*e.g.*, PolyInfo). We note that we already have ~2,000 such structures for homopolymers that are ready for the high-throughput data generation. These structures will be fed into LAMMPS (Large-scale Atomic/Molecular Massively Parallel Simulator) with a python wrapper automating the calculations. This pipeline has already been established in our lab and has successfully produced ~50 thermal conductivity values (see later in **Fig. 6**) with a rate of ~5-10 values/week. The data production rate is to be further increased by leveraging high-performance computers at XSEDE.

In this pipeline (**Fig. 3**), monomer information of a polymer is first drawn from the standardized database, which include atom types and connectivity. This information is then ported into pysimm54 for polymerization and in the meantime generating LAMMPS-compatible data files containing atom positions and forcefield parameters. A number of such polymer chains will be packed into a simulation box using a custom-built Python code. Another Python code has been written to create control input files (*e.g.*, time step, ensemble, target temperature, *etc.*) for the LAMMPS MD simulations. The model will first be run to optimize the polymer conformation *via* repeated annealing processes to ensure reaching reasonable morphology. Conformation features, such as density and *Rg*, will be monitored automatically to check convergence. Once the morphology is optimized, thermal conductivity will be calculated using non-equilibrium MD (NEMD) simulations. Since local crystallization can happen, as we have already observed in some polymers, we will calculate the thermal conductivity in all three spatial directions and also record the average value. It is expected that in a globally amorphous polymer, the local crystal domains are randomly orientated and the polymer should still show isotropic thermal conductivity in real samples.

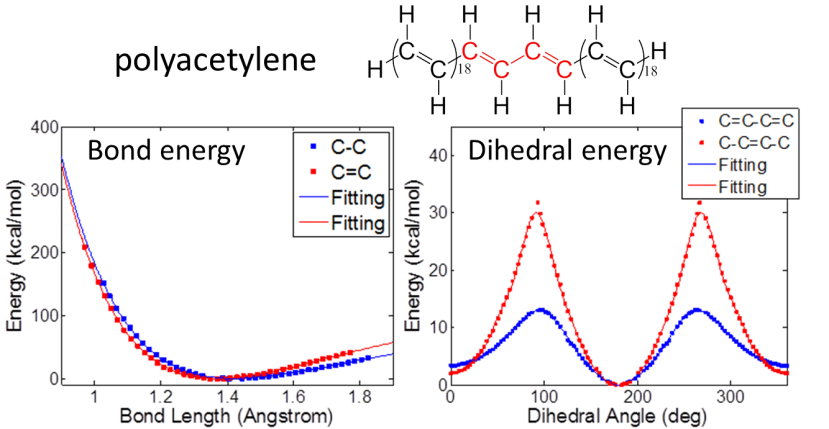


**Figure 3.** High-throughput MD simulation pipeline for thermal conductivity calculation.

The forcefield we will use is the General Amber forcefield (GaFF),55 but we will also test others like Consistent Valence forcefield (CVFF) and Optimized Potentials for Liquid Simulations (OPLS). One **challenge** we have faced is the lack of forcefield parameters for some bonding interactions (*e.g.*, Cl-H). To mitigate this problem, the PI has extensive experience in characterizing molecular interactions using first-principles calculations.56 In such a process, first-principles calculations using programs like Gaussian that do not need any empirical inputs are used to obtain benchmark data on energy hypersurfaces (energy-coordinate relation) for molecules of interest. Empirical potential functions will then be used to model the same system, and the potential parameters will be optimized until the functions reproduce the first-principles data. **Figure 4** shows our parameterization of a polyacetylene molecule that correctly captures the feature of the alternating single and double bonds as well as the induced dihedral angle energy variation. In the same calculation, the electrostatic potential (ESP) charges for every atom on the molecule can be calculated. We plan to use high fidelity hybrid functional like B3LYP to improve the accuracy of these calculations. Using this procedure, we have recently calculated charges for a PAA molecule and its ionized form with different counterions.14

The **expected** **outcomes** from this objective will be a sizable and easily extensible standardized database of polymer properties, accessible *via* a simple API, plus a set of tools for obtaining additional data as required by our research. Relevant metadata (*e.g.*, source, method, uncertainty) for each data value will be included. Another outcome is the automated pipeline for calculating properties of polymers based on high-fidelity MD simulations.

**Figure 4**. Fitted empirical potential for polyacetylene. Dots correspond to first-principles calculations. Solid curves show fitted results.

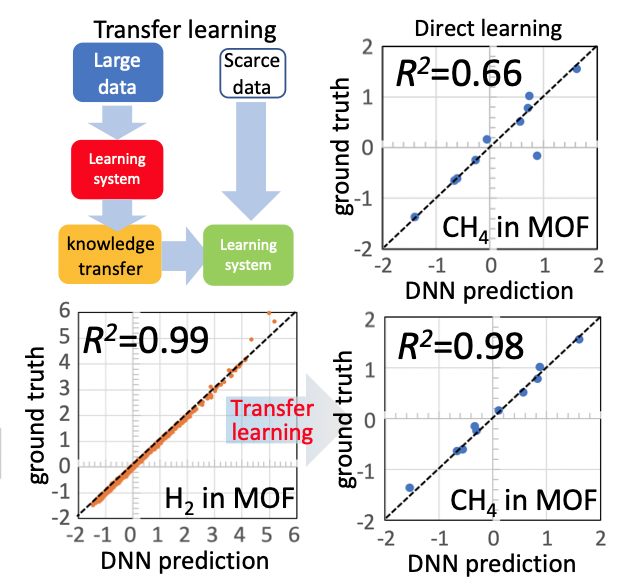


## 4.2 Objective 2: Machine Learning

Based on the database constructed in Objective 1, we will focus on using Transfer Learning to establish ML models for the structure-property relation. In addition, classification and clustering will be performed to identify common features of polymers with thermal conductivity at different ranges.

### Task 2.1. Transfer Learning (Jiang and Luo)

Given the anticipated scarcity of the thermal conductivity data, which is our ultimate target, we will leverage Transfer Learning approaches to “transfer” the “knowledge” learned from highly populated proxy properties (*e.g.*, *Tg*) in the database or easily computable intermediate labels (*e.g.*, dihedral angle energy, charges and *Rg*) to the training of ML models against the less populated thermal conductivity data. In such a technique, models like DNN is first trained on the highly populated data. Then, the DNN model with all corresponding weights, except those of the very last layer, is transferred to the training process of the DNN against thermal conductivity data, during which only the last layer weights are updated (**Fig. 5**). This technique, which is from the computer science field, has recently been applied to materials informatics and preliminary success has been achieved.57 The philosophy behind Transfer Learning is that the inherent structure-property relations share certain physical principles that are captured by the weights and biases of the DNN, although the final expression can be different. PI Luo has been working with researchers from University of Tokyo and the Japanese Institute of Statistical Mechanics to use this new technique for polymer exploration. Recently, the PIs have successfully applied Transfer Learning to study gas adsorption in different metal-organic frameworks (MOFs). In this study, a DNN was first trained against the 13,506 H2 adsorption data, and it was then transferred to training using the 100 CH4 adsorption data. Compared to direct training, the Transfer Learning improved the model accuracy from *R2*=0.66 to *R2*=0.98 (**Fig. 5**). Shown in **Fig. 5** are the testing data that were not included in the training or validation sets. The results also highlight how Transfer Learning effectively eliminates outliers, which makes the model more generalizable.



**Figure 5**. Schematic of transfer learning and preliminary results showing the that Transfer learning can improve DNN accuracy.

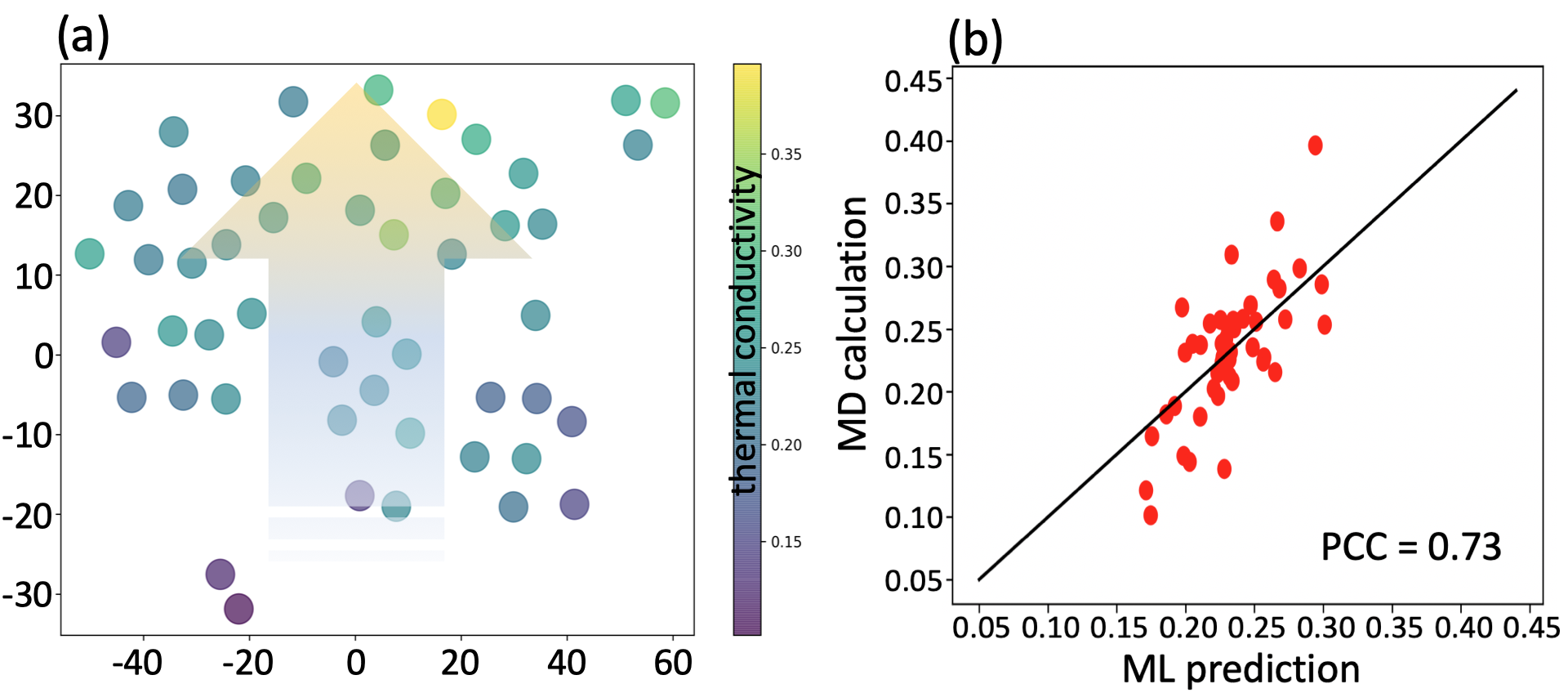
In training the DNN, the molecular embedding from Mol2vec46, 50 will be used as descriptors. The whole dataset will be divided into a training set and a validation set. The training set will be varied to have 70-90% of the total data selected in a random manner. After training, the DNN will take the molecular descriptors from a test set to predict the thermal conductivity, which will then be compared with reference values in the database. The results will then be validated with detailed MD simulations and experiments in Objectives 3 and 4.

The **expected outcome** from this task is the establishment of DNN surrogate models that can accurately and efficiently predict thermal conductivity of polymers based on their chemistry.

### Task 2.2. Exploring Influential Molecular Features (Jiang and Luo)

It would be ideal to understand what chemistry will likely induce high thermal conductivity and what would lead to low thermal conductivity. This will help rationalize the structure-property relations obtained from Task 2.1, and provide guidance to more detailed MD simulations in Task 3. To achieve this understanding, we will leverage co-PI Jiang’s expertise in classification (supervised learning) and clustering (unsupervised learning). In classification, we will use the thermal conductivity as the label. After the classification task, we will have groups of polymers with high, intermediate and low thermal conductivity. The specific borderlines for dividing the thermal conductivity regions will be rationally chosen after examining the values in the standardized database.

Since our molecular representations from Mol2vec are continuous vectors, they are especially suited for similarity analysis. We expect that polymers with similar thermal conductivity values will tend to have similarities in their chemistry. As did in our recent molecular representation study,46 we will quantify the distances among the polymers within each group (*via* vector subtraction) and visualize their positions using the t-SNE (T-distributed Stochastic Neighbor Embedding) scheme58 which map high-dimensional vectors into two-dimensional vectors. We will measure the similarity of whole molecules and do the same for sub-structures of each molecules. It is our belief that polymers with similar thermal conductivity values should share certain similarity in chemistry, and this can be reflected by the possession of common or similar substructures. For example, our previous study have shown that polymer fibers with high thermal conductivity values are mostly from the conjugated polymer structure.27 We have performed a preliminary study using the thermal conductivity calculated from MD simulations for 50 different polymers (**Fig. 6**). The results show rough clustering of low and high thermal conductivity polymers (**Fig. 6a**), suggesting that unique chemistry-thermal conductivity relationships exist. This is confirmed by a direct training of a DNN using the 50 data, which yields acceptable model accuracy despite the scarcity of the data (**Fig. 6b**).



**Figure 6**. (a) Projected molecular embeddings (300D to 2D by t-SNE) colored by corresponding thermal conductivity. There is a general increasing trend of thermal conductivity from bottom to top in the t-SNE plot. This trend is verified by the parity plot of DNN-predicted and MD-calculated thermal conductivity in (b), where PCC (Pearson correlation coefficient) = 0.73 suggest a good prediction.

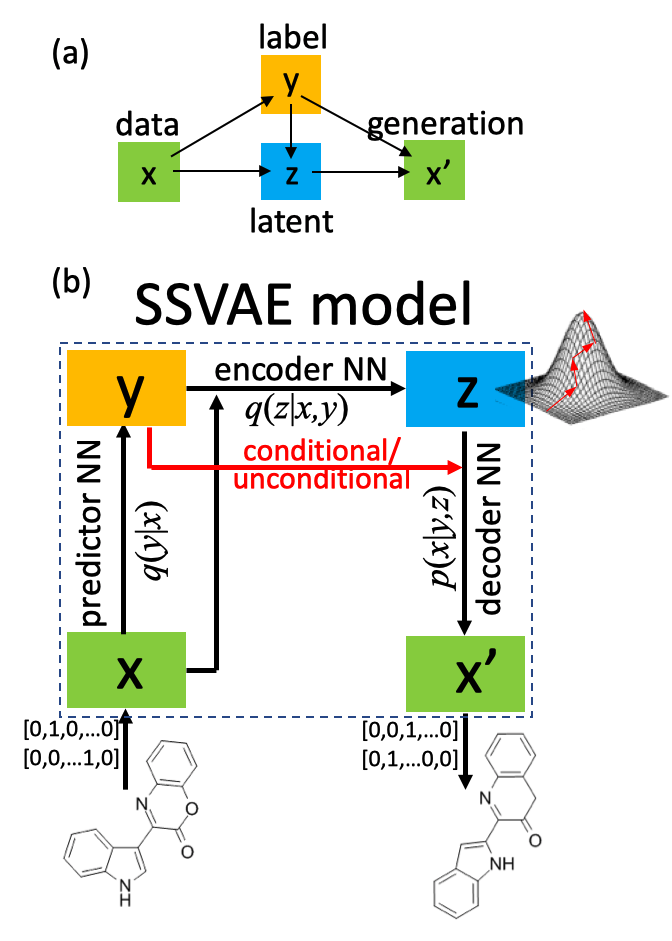
In addition, we will also perform clustering to categorize polymers without referring to their labels (*i.e.*, unsupervised learning). This will help us select diverse polymers for labeling their thermal conductivity using MD simulations. The diversity of descriptors as well as in labels will enhance the transferability of the ML models to be established. Clustering, which will divide polymers into different groups based on their chemistry similarity, will also help establish local models that would capture finer-grained chemistry features among similar polymers and may increase the accuracy of the model when trained using data categorized into the same group.

The **expected outcome** from this task is the understanding of the chemistry origin of thermal conductivity for polymers. The clustering results will also help rationally select polymers for labeling to maximize data diversity.

## 4.3 Objective 3: Validation and Detailed Study Using MD Simulations

### Task 3.1. Inverse Design of Polymer with Target Thermal Conductivity (Luo)

**Figure 7.** (a) Relation of different components of SSVAE; (b) Predictor, encoder and decoder NNs in SSVAE. Note the latent space, *z*, is a continuous distribution function.



Using the established DNN, we will predict ~30 promising polymers with high thermal conductivity (> 1 W/mK). This will involve the inverse design of molecules based on given objectives of a label. In order to predict a polymer chemistry with a given thermal conductivity value, a generative model is needed. The generative model59 to be used will be based on the Semi-Supervised Variational Autoencoder (SSVAE) method,60, 61 which combines probabilistic modeling and DNNs. The model is capable of both predicting properties and generating molecules. Given specific target properties, polymers can be generated by directly sampling new molecules from a conditional generative distribution. The semi-supervised model, which means it can effectively leverage unlabeled data, will be especially beneficial for this project, where thermal conductivity data is expected to be small.

The generative model (**Fig. 7a**) aims to establish the probability, *p(x|y)p*(*x*|*y*), for a given molecular descriptor *x* with label (property) *y*. In the generative process, a new molecule is generated from a generative distribution *p*(*x*|*y*,*z*), which is conditioned on the label *y* and latent variable *z*. The latent variable is a continuous distribution with a much smaller dimension than *x*. The prior distributions of *y* and *z* can be assumed to be normal, and variational inference approximations will be used to address the intractability of the exact posterior inference in the model. The posterior distributions of *y* and *z* are *q*(*y*|*x*) and *q*(*z|x,y*). When some data are not labeled, *y* are treated as latent variables and the missing values are predicted by *q*(*y*|*x*). The different distribution functions are non-linear functions, and in our case will be represented as DNNs. There will be three DNNs in the model, including the predictor DNN, *q*(*y*|*x*), the encoder DNN *q*(*z*|*x,y*) and the decoder DNN *p*(*x*|*y,z*). **Figure 7b** shows their structures and relations. These DNNs are trained simultaneously, and since our data will consist of both labeled and unlabeled data, a combined loss function for the entire data set is minimized in the training.

Once the model is trained, the predictor DNN, *q*(*y*|*x*), can be used to predict the property for unlabeled data. The decoder DNN, *p*(*x*|*y,z*), can be used to generate molecules both conditionally and unconditionally (**Fig. 7b**). To generate arbitrary molecules unconditionally, *y* and *z* are sampled from their prior distributions *p*(*y*) and *p*(*z*), respectively. For conditional molecular generation given a target property, *z* is sampled from *p*(*z*), while *y* is set to the target value. This framework is already established in PI’s group, and preliminary success has been achieved using data from the ZINC database.

### Task 3.2. Detailed MD Simulation of Generated Polymers (Luo)

We will then perform detailed MD simulations to verify the thermal conductivity of the generated polymers following the procedure outlined previously in **Fig. 3**. To form the amorphous structures, we will use the modified Markov Monte Carlo method62 to pack molecule chains in a randomly disordered pattern. The structures will then be annealed to fully relax them. We have extensive experience in using such a strategy to obtain amorphous polymers.63-65 We will then use NEMD to calculate the thermal conductivity of the amorphous polymers and correlate it to the molecular level conformation to identify its influence on thermal transport. In NEMD, a temperature difference is established across the simulation domain by applying two thermostats at the ends of the simulation domain (**Fig. 8**). After steady state is reached, the heat flux (*J*) can be calculated from the energy input and output at the thermostated heat source and sink regions. The thermal conductivity of the liquid can be calculated as , where is the linear fit of the temperature gradient of the liquid. Size effect and temperature gradient effect will be studied to ensure the results are converged. We will compare the MD calculated thermal conductivity with the ML prediction to evaluate the accuracy of the surrogate model.



**Figure 8**. (a) NEMD setup for thermal conductivity calculation. (b) A typical temperature profile at steady state.

We will also decompose the thermal conductivity into contributions from different interactions, including bonding interaction, vdW and electrostatic interactions, to understand the heat transfer along the chain backbone and across-chains, like we did in our previous studies.29-31 We will also correlate the decomposed thermal conductivity with the chain confirmation features like *Rg*, π-π stacking, and radial distribution function concerning inter-chain atoms.

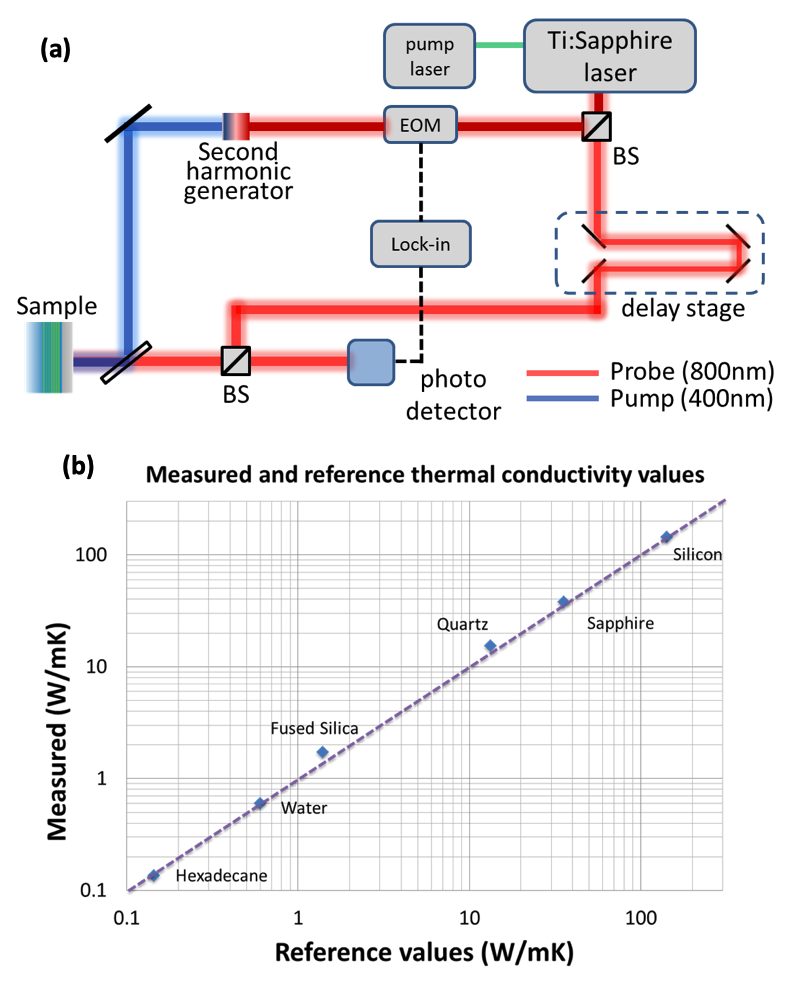
Through the detailed simulations of the 30 polymers generated from Task 3.2, we **expect** to significantly advance the understanding of the polymer chemistry-morphology-thermal conductivity relationships. We also **expect** to confirm 10 high thermal conductivity polymers using the MD simulation from this objective.

## 4.4 Objective 4: Experimental Validation (Luo and Collaborators)

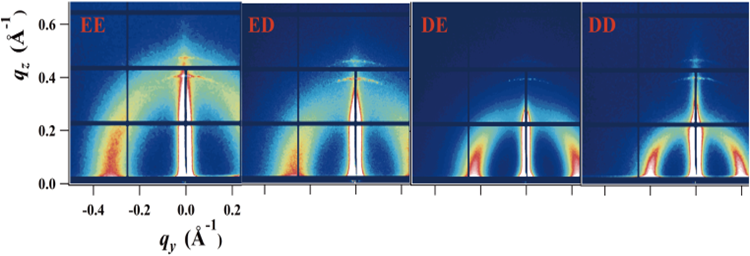
For the simulation-down-selected 10 polymers, we will work with our Japanese collaborator, Prof. Morikawa (see letter), local chemist, Prof. Gao (see letter) and potentially DuPont chemists (preliminary agreement reached), to identify synthesizable polymers, synthesize them, and then the PI will measure their thermal conductivity. To prepare amorphous samples for characterization and measurements, the polymers will be dissolved in organic solvents and then drop-cast on a glass substrate. If the polymer is thermoelastic, melt casting can also be used to fabricate the film. We will make sure that the polymer layer is thick so that there is no thickness-dependent morphology effect recently implied in Ref. [66].

Thermal conductivity will be measured using a high fidelity pump-probe technique, called time-domain thermoreflectance.67, 68The two-color TDTR system in the PI’s lab (**Fig. 9**) has been extensively tested against standard samples with thermal conductivities spanning four orders of magnitude (O(0.1-100) W/mK). In brief, the system shines pump laser pulses on a photothermal transducer that is usually a metal layer (*e.g.*, Al) deposited on the sample. The electrons in the metal absorb the photon energy and convert it into thermal energy through electron-phonon coupling. The thermal energy is transferred from the heated metal layer to the sample. The time history of the metal surface temperature is probed by a temporally delayed probe laser by detecting the evolution of the optical reflectance, which is proportional to the metal surface temperature. Using heat transfer models to fit the decay curve, the thermal conductivity can be obtained. The TDTR experiments will be cross-checked by collaborator, Morikawa’s lab (see letter), which has expertise in using MEMS devices for thermal conductivity measurements.57

**Figure 9**. (a) Simplified TDTR optical diagram. (b) thermal conductivity of different materials measured using our TDTR system and comparison with reference data.



To better understand the thermal conductivity measurements, SAXS, GIXS and tunneling electron microscopy (TEM) will be used to characterize the morphology, especially the local morphology. The polymers to be studied are supposed to be globally amorphous with possible ordered domains scattered in a disordered matrix. These materials will be only weakly scattering to x-ray sources. Such a feature requires the use of an x-ray source with high brilliance. The GIXS in the APS facility at Argonne National Lab will be used to offer detailed and accurate morphology data. We will apply for user time allocation in the APS beamlines, which we used previously to characterize detailed morphology of amorphous PBDTTT polymers (**Fig. 10**).69 The combination of the above-mentioned characterizations and their complementary incident angles will yield a wide range of morphology data on difference length scales, such as crystallinity and local characteristic lengths, which will be leveraged to understand the measured thermal conductivity.



**Figure 10.** GIXS patterns of PBDTTT polymers.

The **expected outcome** of this objective is the experimental validation of 3-5 polymers with thermal conductivity values greater than 1 W/mK.

A **potential challenge** for the experimental validation is the synthesizability of the generated polymer from Objective 3. We will take two routes to mitigate such a challenge if encountered. In one route, we will generate additional polymers from Task 3.1 to increase the possibility of finding synthesizable molecules. In another route, we may add the synthesizability score70 as a label and use it as an additional target in the conditional polymer generation process in Task 3.1.

# 5. PI Qualification

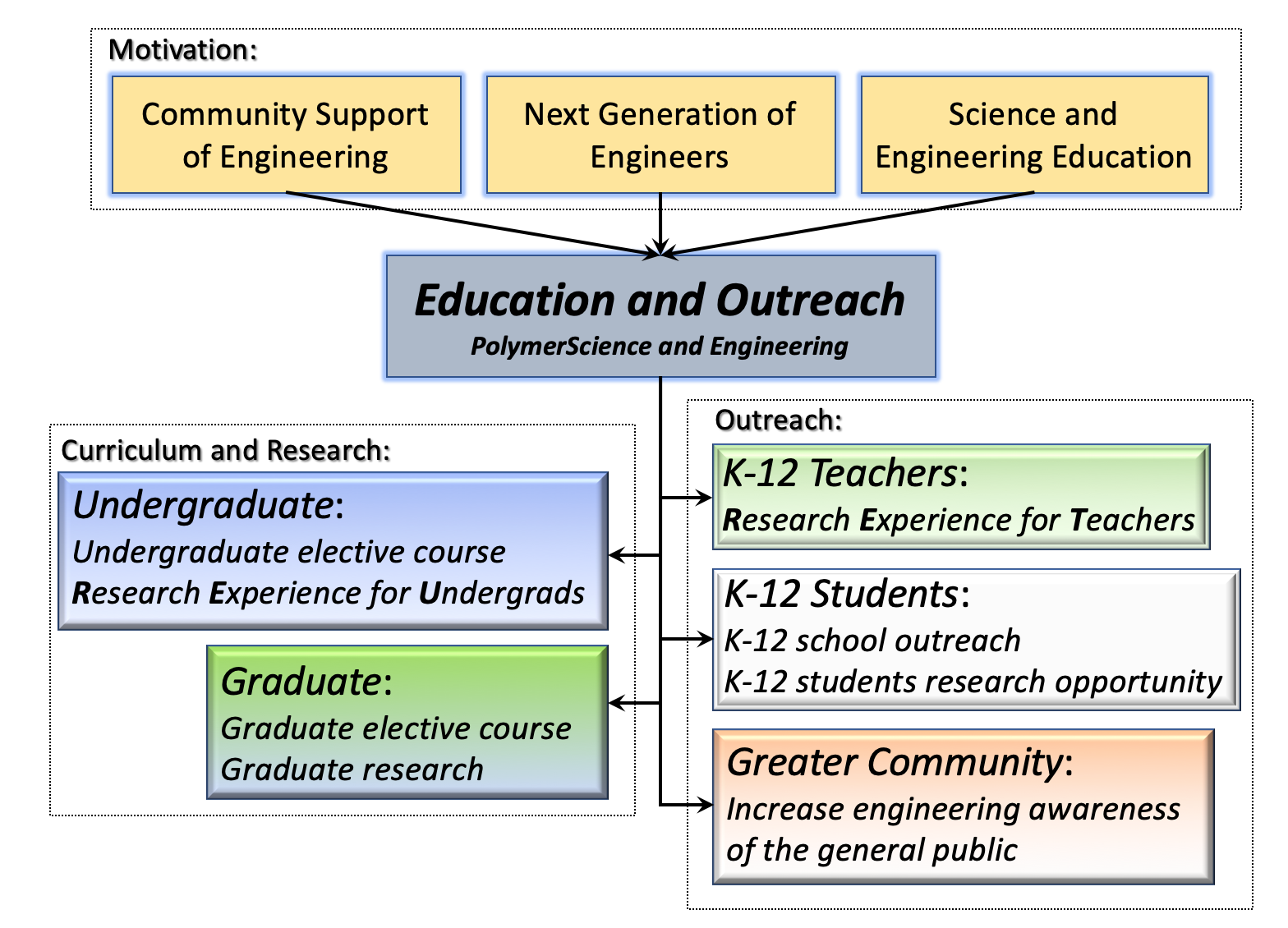
The PI, **Tengfei Luo**, is the Dorini Family Associate Professor at the Aerospace and Mechanical Engineering Department with a concurrent appointment in the Chemical and Biomolecular Engineering Department of Notre Dame (ND). He has strong expertise in polymer thermal transport physics. Over the years, his group has used MD simulation to reveal new physics related to the structure-property relations of polymer thermal conductivity.27, 29-31, 69, 71, 72 It is this experience that motivated him to combine the knowledge learned from these physics-based simulations with the ML techniques to accelerate the material discovery of thermally conductive polymers. His group has also developed expertise in TDTR measurement of polymer thermal conductivity.69, 72 In addition, his group has also obtained preliminary results in Polymer Informatics and has developed collaboration with Japanese institutions to leverage Transfer Learning (see letter from Shiomi). He was the 2016 DuPont Young Professor Awardee and has since maintained a strong collaboration with DuPont scientists. These collaborations will significantly contribute to the coordination of detailed simulation, ML, synthesis and characterization of polymers in this project.

The co-PI, Meng Jiang, is an Assistant Professor at the Computer Science and Engineering Department of ND. He has strong expertise in data mining and ML, especially Transfer Learning. Over the years, his group has developed computational methods to model human behaviors in dynamic, social, and spatiotemporal environments. The ML techniques include Transfer Learning,73-78 semi-supervised and distantly-supervised learning,79-81 and unsupervised representation learning.82-85 His group has established trustworthiness in successful dissemination of user-friendly software implementing the ML techniques.

The two PIs are currently working together on applying Transfer Learning to Materials Informatics related to thermal transport and other energy-related problems.

# 6. Educational and Outreach Program

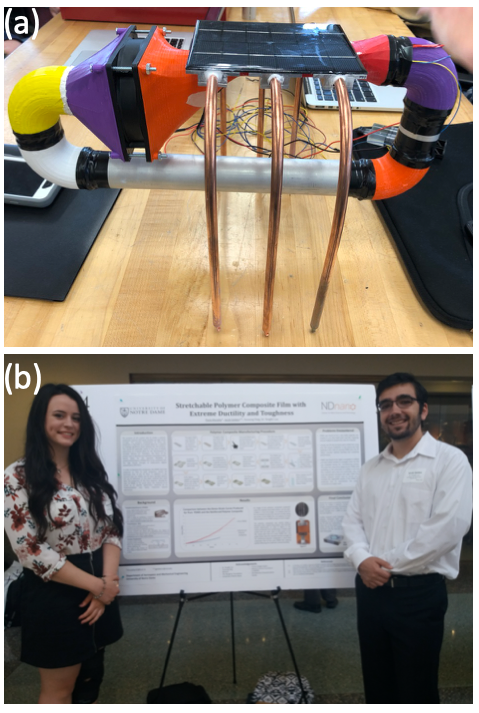
The improvement of science and engineering education at different levels is deemed as imperative in preparing the next generation of U.S. engineers to guarantee the country’s global competitiveness in engineering fields.86-95 Polymer and ML, being directly related to the most challenging engineering problems, is an ideal topic for such multi-tiered education. In this objective, we will develop an integrated educational and outreach program, which integrates university, K-12 and greater community education under the umbrella of science and engineering of polymer in the context of grand challenges. The activities in this education program largely leverage our fundamental research on polymers and ML and their importance in the world’s most critical applications including electronics, alternative energy and biotechnology. To ensure the success of this program, it is divided into subprograms targeting different groups. The contents of the subprograms are inter-connected and can be leveraged in a way that the whole program is efficient and complete (**Fig. 11**).



**Figure 11**. Overview of the educational and outreach program.

## Task 6.1. Undergraduate Course Enhancement and Research

The Heat Transfer course the PI is going to teach is an infamously difficult subject to students due to its abstract nature. While the contents traditionally taught in this course are essential to building a firm foundation for future engineers, exposing them to the state-of-the-art and emerging technologies has not been systematically incorporated, which is essential in preparing them to work on nontraditional engineering problems. In this project, the PI will further improve the effectiveness of teaching by adding an assignment related to analyzing the thermal transport in a CPU-heat sink junction mated by different polymers. Three different heat flux values will be given to mimic the CPU heat dissipation. The students will be required to look up different polymer based thermal interface materials (TIM) and calculate the temperature of CPU given a fixed heat sink temperature. This will allow the students to have a real understanding of the importance of polymers in the thermal management of modern electronics.



**Figure 12**. (a) A senior design project protocol of cooling system attached to a solar panel; (b) Undergraduate students funded by the ND Noughton fellowship and the NURF fellowship presenting a polymer project in the campus-wide undergraduate research conference.

The PI also teaches Senior Design. The projects he teaches are usually related to thermal control of a modern system. In 2017, he has instructed students to develop a closed loop temperature control system for a 3D printer. In 2018, he has mentored students to develop a closed loop cooling system for solar panels. Both of these projects are related to the forefront of the modern technology. In both projects, students will need to design thermal interfaces with low thermal resistances so that the cooling systems can effectively modulate the temperature of the target. The PI emphasizes the importance of such interfaces and encourage the students to do trade studies on polymer TIM. Towards the end of the projects, the students compare the system performance with and without the polymer TIM to highlight the importance of polymer in thermal design. Similar projects will be continued in the following years where polymers will continue to play important roles (**Fig. 12a**).

PI Luo has had 34 undergraduates doing research in his lab for the past 7 years. Such activities have not only resulted in journal publications96 and presentations, but also given him opportunities to teach and educate the students on a more personal level outside the classroom. Over the past two years, the Co-PI Jiang has had two high school students (the Trinity School at Greenlawn) and 14 REUs including 5 females. 6 REUs have published papers in international conferences as the first author. Three of the five females have applied the PhD program of the CSE Department at ND. This project will enable the PIs to provide positions for undergraduate students to participate in different aspects of the research in his lab. For example, the students can be involved in polymer data collection and ML. The positions will be funded through REU supplements to be requested. Assessment will be carried out *via* regular meetings, reports and a presentation in events at ND such as the *undergraduate energy research expo* and the campus-wide *undergraduate research conference*. **Figure 12b** shows a prior polymer project presented by undergraduate students.

## TASK 6.2. Graduate Course Development and Research

Atomistic simulations have long been effective tools in biology, physics and material science. These tools have recently become more widely used to study engineering problems like the topics of the present proposal. In line with this trend, the PI has developed a graduate course on “Engineering at the Molecular Level” with the support of previous NSF supports. He teaches specific molecular modeling techniques, including Density Functional Theory, Monte Carlo and MD, and how they are used to study contemporary engineering science like nanoscale heat and mass transfer. It also provides unique views from the molecular level of micro/nano systems such as nanocomposites, nanoparticles, nanofibers, thin films and their applications in the frontiers of technology. The PI will include specific polymer simulation cases to this course as homework assignment and term projects. In addition, the PI will also invite co-PI to give guest lectures on ML techniques and allow students to propose ML-related term projects. Assessment will be based on the students’ class participation, course projects and end-of-semester presentations.

The PIs have and will continue to mentor graduate students on research strategy, publication and presentation skills, and ethics. Both PIs’ groups currently have a total of 13 PhD students. The interdisciplinary groups are working on different topics concerning molecular level heat transfer and ML. The proposed project will support two graduate students. Students will be assessed by the quality of their research, publications and their capability to perform independent, high-quality research. In addition, DuPont will also provide student internship opportunities in the duration of this project from year 1. In this way, the student will be able to put the academic research in an industrial context. This internship will also help DuPont understand the merits and limitations of the ML model so as to better plan further research to improve the model.

### TASK 6.3. K-12 Outreach

The objective of the K-12 outreach program is to excite and engage students at an early age about the benefits and rewards of pursuing engineering as a career choice. This will be realized by three parallel activities through programs like RET and K-12 research opportunities.

**RET:** The PI has been actively participating in the ND Engineering a More Sustainable Energy Future RET program. Ms. Louann Kensinger, from Riley High School South Bend, worked on water desalination in his lab in 2012. Mr. Thomas Adams, from South Bend New Tech, worked on measuring heat transfer of polymers in 2013. In 2015, Ms. Pauline Alokolaro from Seattle Preparatory School is working on polymer nanofiber fabrication. The PI meets with the teachers every week to not only discuss the project progress but also to develop strategies for transforming their experience into classroom lectures to inspire interest in science and engineering among high school students. He is committed to having more teachers in the future to participate in polymer-related research with supplemental support from RET programs. Specifically, he will introduce the senior design project and seek opportunities to demonstrate the designed cooling system to high school classrooms. The teachers will participate in different aspects of the development of such a system. Assessment will be based on the end-of-summer presentation to the RET program and the students’ feedback on the specific contents brought back to the classrooms.

**K-12 Student Summer Research:** A part of this program is designed to make direct contact with high school students through the connection made with the Riley High School, which has a large portion (48%) of minority students, mostly African-American and Latino. The PI’s groups will recruit and mentor two to three, ideally underrepresented, high school students from Riley each summer to experience different parts of the proposed research such as data collection, MD trajectory movie generation and ML. These students will be encouraged to share their experience with their classmates and families, enabling the widespread awareness of the importance of science and engineering of soft materials. Assessment will be based on exiting presentations.

### TASK 6.4. Outreach to the Greater Community

The PIs also propose to demonstrate the senior design project models at local public events to showcase emerging engineering technologies. Through such prototype demonstrations they hope to excite the greater community on engineering innovations. For these activities, they will also design special brochures to be taken by the audience so that they can further spread the information to reach an even larger community. To assess these activities, attendees will be invited to participate in two quizzes on polymer engineering, with one being taken before and the other one being taken after each event. One of the events we will participate in is *Science Alive*, which is an annual event at the St. Joseph County Public Library that features a day of science activities and exhibits targeting K-8 youth, drawing in *more than 3,500 residents per year*. The library is transformed into a hands-on science museum for the day. Researchers, clubs, and organizations set up booths with hands-on, interactive demonstrations.

# 6. Proposed Timeline

| **Objectives and Subtasks** | **Year** | | | | |
| --- | --- | --- | --- | --- | --- |
| 1 | | | 2 | 3 |
| **Obj 1. Database Development** |  | | |  |  |
| 1.1. Data collection from existing database |  | | |  |  |
| 1.2. Data collection from literature |  | |  |  |  |
| 1.3. Data generation from high-throughput MD simulation |  | | |  |  |
| **Obj 2. Machine Learning** |  | | |  |  |
| 2.1. Surrogate model for structure-property relationships |  | | |  |  |
| 2.2. Classification and clustering to identify key molecular features |  | | |  |  |
| **Obj 3. MD Simulations** |  | | |  |  |
| 3.1. Molecular generation |  |  | |  |  |
| 3.2. MD simulation of the polymer chemistry-morphology-property relation |  |  | |  |  |
| **Obj 4. Experimental Validation** |  |  | |  |  |

# 7. Results of Prior NSF Support

**EFRI-2DARE-1433490**; 11/01/2014–10/31/2018; $1,999,997; Monolayer Heterostructures: Epitaxy to Beyond-CMOS Devices, PI: H. Xing; Co-PIs: T. Luo, D. Jena, M. Eskildsen, L. Huang. **Intellectual Merits:** The MBE growth of 2D layered materials and heterostructures will result in a paradigm shift in what is possible with layered materials and their heterostructures. **Broader Impacts:** The proposed research will be complemented by significant international and industrial interactions, and with a meaningful and vigorous educational/outreach program. Products:97-99.

**CBET- 1706039**; 9/1/2017-8/31/2020; $356,000; Thermal Evaporation around Optically-Excited Functionalized Nanoparticles, PI: T. Luo; Co-PI: H. C. Chang, A. Raigoza. **Intellectual Merits:** This project will combine molecular simulation, mesoscale modeling and experiments understand how surface functionalization impact thermal transport and local viscosity to influence the bubble generation. **Broader Impacts:** Optical absorber-assisted thermal evaporation can enable a variety of innovative renewable energy applications, ranging from water purification to chemical fractionation. Many of these applications can drive global economic growth. Products: 100-104.

**IIS-1849816**; 10/1/2019-9/30/2021; $174,910; Complementarity Learning for Contextual Behavior Modeling, PI: M. Jiang. **Intellectual Merits:** The project will develop novel behavior modeling methods that learn representations of human behavior by preserving the structure of complementarity among the behavior’s components. **Broader Impacts:** The project will advance our understanding of human behaviors in dynamic, social, and spatiotemporal environments. A new course in data science has been offered to undergraduate/graduate students. Products: 73, 82-85.

**CCF-1901059**; 8/15/2019-8/14/2023; $1,019,030; Semantically-Enhanced Software Traceability for Supporting Human-Centric Tasks, PI: J. Huang; Co-PIs: M. Jiang, R. Metoyer. **Intellectual Merits:** This project will develop a holistic, interactive tracing environment, which incorporates diverse algorithmic solutions for dynamically generating trace links, visualizing the results, and guiding the user through the interactive process of using the results. **Broader Impacts:** The cross-disciplinary nature of the team will introduce new opportunities for software engineers and data scientists to collaborate in addressing open Software Engineering challenge. Products: 80, 81, 105-107.

**References:**

1. Henry, A. Thermal transport in polymers. *Annu.Rev.Heat Transfer* **17**, 485-520 (2013).

2. Ziman, J. M. in *Electrons and Phonons: The Theory of Transport Phenomena in Solids* (OUP Oxford, 2001).

3. Srivastava, G. P. in *The Physics of Phonons* (A. Hilger, 1990).

4. Chen, G. in *Nanoscale energy transport and conversion: a parallel treatment of electrons, molecules, phonons, and photons* (Oxford University Press, 2005).

5. Tien, C. L. in *Microscale Energy Transfer* (Taylor \& Francis, 1997).

6. Zhang, Z. in *Nano/Microscale Heat Transfer* (McGraw-Hill Education, 2007).

7. Kittel, C. Interpretation of the Thermal Conductivity of Glasses. *Phys.Rev.* **75**, 972-974 (1949).

8. Slack, G. A. in *Solid State Physics* (eds Henry Ehrenreich, F. S. & David, T.) 1-71 (Academic Press, 1979).

9. Cahill, D. G. & Pohl, R. O. Lattice Vibrations and Heat Transport in Crystals and Glasses. *Annu. Rev. Phys. Chem.* **39**, 93-121 (1988).

10. Cahill, D. G. & Pohl, R. O. Thermal conductivity of amorphous solids above the plateau. *Phys.Rev.B* **35**, 4067-4073 (1987).

11. Einstein, A. Elementare Betrachtungen über die thermische Molekularbewegung in festen Körpern. *Annalen der Physik* **340**, 679-694 (1911).

12. Allen, P. B., Feldman, J. L., Fabian, J. & Wooten, F. Diffusons, locons and propagons: Character of atomie yibrations in amorphous Si. *Philosophical Magazine Part B* **79**, 1715-1731 (1999).

13. Xie, X.High and low thermal conductivity of amorphous macromolecules. *Phys.Rev.B* **95**, 035406 (2017).

14. wei, x., MA, R. & Luo, T. Thermal Conductivity of Polyelectrolytes with Different Counterions. (2020).

15. Subramanyan, H.Role of angular bending freedom in regulating thermal transport in polymers. *J. Appl. Phys.* **125**, 095104 (2019).

16. Fujishiro, H., Ikebe, M., Kashima, T. & Yamanaka, A. Thermal Conductivity and Diffusivity of High-Strength Polymer Fibers. *Japanese Journal of Applied Physics* **36**, 5633-5637 (1997).

17. Shen, S., Henry, A., Tong, J., Zheng, R. T. & Chen, G. Polyethylene nanofibres with very high thermal conductivities. *Nat Nanotechnol* **5**, 251-255 (2010).

18. Wang, X., Ho, V., Segalman, R. A. & Cahill, D. G. Thermal Conductivity of High-Modulus Polymer Fibers. *Macromolecules* **46**, 4937-4943 (2013).

19. Zhang, T. & Luo, T. Morphology-Influenced Thermal Conductivity of Polyethylene Single Chains and Crystalline Fibers. *J Appl Phys* **112**, 094304 (2012).

20. Liu, J. & Yang, R. Length-dependent thermal conductivity of single extended polymer chains. *Phys.Rev.B* **86**, 104307 (2012).

21. Liu, J. & Yang, R. Tuning the thermal conductivity of polymers with mechanical strains. *Physical Review B* **81**, 174122 (2010).

22. Ma, H. & Tian, Z. Chain rotation significantly reduces thermal conductivity of single-chain polymers. *J. Mater. Res.* **34**, 126-133 (2019).

23. Ma, H., Ma, Y. & Tian, Z. Simple Theoretical Model for Thermal Conductivity of Crystalline Polymers. *ACS Appl. Polym. Mater.* **1**, 2566-2570 (2019).

24. He, J., Kim, K., Wang, Y. & Liu, J. Strain effects on the anisotropic thermal transport in crystalline polyethylene. *Appl. Phys. Lett.* **112**, 051907 (2018).

25. Henry, A. & Chen, G. Normal Mode Analysis of a Single Polyethylene Chain. *Proceedings of the Asme International Mechanical Engineering Congress and Exposition, Vol 13, Pts a and B*, 1157-1160 (2009).

26. Henry, A. & Chen, G. Anomalous heat conduction in polyethylene chains: Theory and molecular dynamics simulations. *Phys Rev B* **79** (2009).

27. Zhang, T., Wu, X. & Luo, T. Polymer Fibers with Outstanding Thermal Conductivity and Thermal Stability: Fundamental Linkage between Molecular Characteristics and Macroscopic Thermal Properties. *J. Phys. Chem. C* **118**, 21148 (2014).

28. Shenogin, S., Bodapati, A., Keblinski, P. & McGaughey, A. J. H. Predicting the thermal conductivity of inorganic and polymeric glasses: The role of anharmonicity. *J. Appl. Phys.* **105** (2009).

29. Zhang, T. & Luo, T. Role of Chain Morphology and Stiffness in Thermal Conductivity of Amorphous Polymers. *J Phys Chem B* **120**, 803-812 (2016).

30. Wei, X., Zhang, T. & Luo, T. Morphology-Dependent Thermal Conductivity of Amorphous Polymer Blends: The Impact of Inter- and Intra-chain Interactions. *Phys. Chem. Chem. Phys.* **in review** (2016).

31. Wei, X. & Luo, T. The effect of the block ratio on the thermal conductivity of amorphous polyethylene-polypropylene (PE-PP) diblock copolymers. *Phys.Chem.Chem.Phys.* **20**, 20534-20539 (2018).

32. Lu, T.Thermal transport in semicrystalline polyethylene by molecular dynamics simulation. *J. Appl. Phys.* **123**, 015107 (2018).

33. Singh, V.High thermal conductivity of chain-oriented amorphous polythiophene. *Nat Nano* **9**, 384-390 (2014).

34. Xu, Y.Molecular engineered conjugated polymer with high thermal conductivity. *Science Advances* **4** (2018).

35. Shanker, A.High thermal conductivity in electrostatically engineered amorphous polymers. *Science Advances* **3** (2017).

36. Wei, X. & Luo, T. Role of Ionization in Thermal Transport of Solid Polyelectrolytes. *J. Phys. Chem. C* **123**, 12659-12665 (2019).

37. Alberi, K.The 2019 materials by design roadmap. *J. Phys. D* **52**, 013001 (2018).

38. Gorai, P., Stevanović, V. & Toberer, E. S. Computationally guided discovery of thermoelectric materials. *Nature Reviews Materials* **2**, 17053 (2017).

39. Carrete, J., Li, W., Mingo, N., Wang, S. & Curtarolo, S. Finding Unprecedentedly Low-Thermal-Conductivity Half-Heusler Semiconductors via High-Throughput Materials Modeling. *Phys.Rev.X* **4**, 011019 (2014).

40. Oliynyk, A. O.High-Throughput Machine-Learning-Driven Synthesis of Full-Heusler Compounds. *Chemistry of Materials* **28**, 7324-7331 (2016).

41. Gaultois, M. W.Perspective: Web-based machine learning models for real-time screening of thermoelectric materials properties. *APL Materials* **4**, 053213 (2016).

42. Carrete, J., Mingo, N., Wang, S. & Curtarolo, S. Nanograined Half-Heusler Semiconductors as Advanced Thermoelectrics: An Ab Initio High-Throughput Statistical Study. *Advanced Functional Materials* **24**, 7427-7432.

43. Audus, D. J. & de Pablo, J. J. Polymer Informatics: Opportunities and Challenges. *ACS Macro Letters* **6**, 1078-1082 (2017).

44. Wu, S.Machine-learning-assisted discovery of polymers with high thermal conductivity using a molecular design algorithm. *npj Computational Materials* **5**, 66 (2019).

45. Ma, R., Huang, D., Zhang, T. & Luo, T. Determining influential descriptors for polymer chain conformation based on empirical force-fields and molecular dynamics simulations. *Chemical Physics Letters* **704**, 49-54 (2018).

46. Ma, R., Liu, Z., Zhang, Q., Liu, Z. & Luo, T. Evaluating Polymer Representations via Quantifying Structure–Property Relationships. *J. Chem. Inf. Model.* **59**, 3110-3119 (2019).

47. Otsuka, S., Kuwajima, I., Hosoya, J., Xu, Y. & Yamazaki, M. *PoLyInfo: Polymer Database for Polymeric Materials Design* (2011 International Conference on Emerging Intelligent Data and Web Technologies, 2011).

48. Kim, C., Chandrasekaran, A., Huan, T. D., Das, D. & Ramprasad, R. Polymer Genome: A Data-Powered Polymer Informatics Platform for Property Predictions. *J. Phys. Chem. C* **122**, 17575-17585 (2018).

49. <https://polymerdatabase.com>.

50. Jaeger, S., Fulle, S. & Turk, S. Mol2vec: Unsupervised machine learning approach with chemical intuition. *Journal of chemical information and modeling* **58**, 27-35 (2018).

51. Tchoua, R. B. *Towards a hybrid human-computer scientific information extraction pipeline* (e-Science (e-Science), 2017 IEEE 13th International Conference on, IEEE, 2017).

52. Swain, M. C. & Cole, J. M. Chemdataextractor: a toolkit for automated extraction of chemical information from the scientific literature. *Journal of chemical information and modeling* **56**, 1894-1904 (2016).

53. Goldberg, Y. & Levy, O. word2vec Explained: deriving Mikolov et al.'s negative-sampling word-embedding method. *arXiv preprint arXiv:1402.3722* (2014).

54. Fortunato, M. E. & Colina, C. M. pysimm: A python package for simulation of molecular systems. *SoftwareX* **6**, 7-12 (2017).

55. Vassetti, D., Pagliai, M. & Procacci, P. Assessment of GAFF2 and OPLS-AA general force fields in combination with the water models TIP3P, SPCE, and OPC3 for the solvation free energy of druglike organic molecules. *Journal of chemical theory and computation* **15**, 1983-1995 (2019).

56. Luo, T. Molecular dynamics study of thermal transport in GaAs-self-assembly monolayer-GaAs junctions with ab initio characterization of thiol-GaAs bonds. *J. Appl. Phys.* **109**, 034301 (2011).

57. Wu, S.Machine-learning-assisted discovery of polymers with high thermal conductivity using a molecular design algorithm. *npj Computational Materials* **5**, 66 (2019).

58. Maaten, L. v. d. & Hinton, G. Visualizing data using t-SNE. *Journal of machine learning research* **9**, 2579-2605 (2008).

59. Kang, S. & Cho, K. Conditional molecular design with deep generative models. *arXiv preprint arXiv:1805.00108* (2018).

60. Kingma, D. P., Mohamed, S., Rezende, D. J. & Welling, M. *Semi-supervised learning with deep generative models* (Advances in neural information processing systems, 2014).

61. Rezende, D. J., Mohamed, S. & Wierstra, D. Stochastic backpropagation and approximate inference in deep generative models. *arXiv preprint arXiv:1401.4082* (2014).

62. Allen, P. & Tildesley, D. J. in *Computer Simulation of Liquids* (Clarendon Press, 1989).

63. Luo, T. F. & Lloyd, J. R. Enhancement of Thermal Energy Transport across Graphene/Graphite and Polymer Interfaces - A Molecular Dynamics Study. *Advanced Functional Materials* **22**, 2495 (2012).

64. Luo, T. F., Esfarjani, K., Shiomi, J., Henry, A. & Chen, G. Molecular dynamics simulation of thermal energy transport in polydimethylsiloxane (PDMS). *J Appl Phys* **109**, 074321 (2011).

65. Zhang, T. & Luo, T. High-Contrast, Reversible Thermal Conductivity Regulation Utilizing the Phase Transition of Polyethylene Nanofibers. *ACS Nano* **7**, 7592 (2013).

66. Liu, J., Ju, S., Ding, Y. & Yang, R. Size effect on the thermal conductivity of ultrathin polystyrene films. *Appl. Phys. Lett.* **104** (2014).

67. Norris, P. M.Femtosecond pump-probe nondestructive examination of materials (invited). *Rev Sci Instrum* **74**, 400-406 (2003).

68. Cahill, D. G. Analysis of heat flow in layered structures for time-domain thermoreflectance. *Rev Sci Instrum* **75**, 5119-5122 (2004).

69. Guo, Z.Tuning the thermal conductivity of solar cell polymers through side chain engineering. *Phys.Chem.Chem.Phys.* **16**, 7764-7771 (2014).

70. Ertl, P. & Schuffenhauer, A. Estimation of synthetic accessibility score of drug-like molecules based on molecular complexity and fragment contributions. *Journal of Cheminformatics* **1**, 8 (2009).

71. Zhang, T. & Luo, T. Thermal Diodes: Giant Thermal Rectification from Polyethylene Nanofiber Thermal Diodes (Small 36/2015). *Small* **11**, 4656-4656 (2015).

72. Guo, Z.Thermal conductivity of organic bulk heterojunction solar cells: an unusual binary mixing effect. *Physical Chemistry Chemical Physics* **16**, 26359-26364 (2014).

73. Yao, H.Graph Few-shot Learning via Knowledge Transfer. *arXiv preprint arXiv:1910.03053* (2019).

74. Jiang, M. *Social contextual recommendation* (Proceedings of the 21st ACM international conference on Information and knowledge management, 2012).

75. Jiang, M. *Social recommendation across multiple relational domains* (Proceedings of the 21st ACM international conference on Information and knowledge management, 2012).

76. Jiang, M., Cui, P., Yuan, N. J., Xie, X. & Yang, S. *Little is much: Bridging cross-platform behaviors through overlapped crowds* (Thirtieth AAAI Conference on Artificial Intelligence, 2016).

77. Jiang, M., Cui, P., Wang, F., Zhu, W. & Yang, S. Scalable recommendation with social contextual information. *IEEE Trans. Knowled. Data Eng.* **26**, 2789-2802 (2014).

78. Jiang, M.Social recommendation with cross-domain transferable knowledge. *IEEE Trans. Knowled. Data Eng.* **27**, 3084-3097 (2015).

79. Zeng, Q.Valproic Acid Stimulates Hippocampal Neurogenesis via Activating the Wnt/beta-Catenin Signaling Pathway in the APP/PS1/Nestin-GFP Triple Transgenic Mouse Model of Alzheimer's Disease. *Front. Aging Neurosci.* **11**, 62 (2019).

80. Jiang, T. *Multi-Input Multi-Output Sequence Labeling for Joint Extraction of Fact and Condition Tuples from Scientific Text* (Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), 2019).

81. Jiang, T. *The Role of" Condition" A Novel Scientific Knowledge Graph Representation and Construction Model* (Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2019).

82. Yu, W., Yu, M., Zhao, T. & Jiang, M. Identifying Referential Intention with Heterogeneous Contexts.

83. Tang, P. *Multi-label Patent Categorization with Non-local Attention-based Graph Convolutional Network* (Proceedings of AAAI Conference on Artificial Intelligence, 2020).

84. Zhang, C.Few-Shot Knowledge Graph Completion. *arXiv preprint arXiv:1911.11298* (2019).

85. Wang, D., Jiang, T., Chawla, N. V. & Jiang, M. *TUBE: Embedding Behavior Outcomes for Predicting Success* (Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2019).

86. National Science Board. Science and Engineering Indicator 2012 -- A broad base of quantitative information on the U.S. and international science and engineering enterprise. (2012).

87. National Academy of Engineering. in *Educating the Engineer Of 2020: Adapting Engineering Education to the New Century* (National Academies Press, 2005).

88. President’s Council of Advisors on Science and Technology. PREPARE AND INSPIRE: K-12 EDUCATION IN SCIENCE, TECHNOLOGY, ENGINEERING, AND MATH (STEM) FOR AMERICA’S FUTURE. (2010).

89. National Science Foundation, Directorate for Engineering. Making the Case for Engineering -- Study and Recommendations. **1-40** (2004).

90. Committee on Science,Engineering, and Public Policy. in *Reshaping the Graduate Education of Scientists and Engineers* (National Academy Press, 1995).

91. Nerad, M. Improving Doctoral Education: Recommendations from the PhDs - Ten Years Later Study. *The Council Graduate Schools Communicator* **33**, 6 (2000).

92. Nyquist, J. D. & Woodford, B. J. Re-envisioning the Ph.D.: What Concerns Do We Have? .

93. Walker, G. E. & Carnegie Foundation for the Advancement of Teaching. in *The Formation of Scholars: Rethinking Doctoral Education for the Twenty-First Century* (Jossey-Bass, 2008).

94. Patrick, Helen, Mantzicopoulos, P. & Samarapungavan, A. Motivation for learning science in kindergarten: Is there a gender gap and does integrated inquiry and literacy instruction make a difference. *Journal of Research in Science Teaching* **46**, 166-191 (2009).

95. Rogers, C. & Portsmore, M. Bringing Engineering to Elementary School. *Journal of STEM Education* **5**, 17 (2004).

96. Rish, D., Luo, S., Kurtz, B. & Luo, T. Exceptional Ion Rejection Ability of Directional Solvent for Non-Membrane Desalination. *Appl. Phys. Lett.* **104**, 024102 (2014).

97. Wu, X., Yang, N. & Luo, T. Unusual isotope effect on thermal transport of single layer molybdenum disulphide. *Appl. Phys. Lett.* **107**, 191907 (2015).

98. Jin, Z.A Revisit to High Thermoelectric Performance of Single-layer MoS2. *arXiv preprint arXiv:1504.03852* (2015).

99. Wu, X.Hydrogenation of Penta-Graphene Leads to Unexpected Large Improvement in Thermal Conductivity. *Nano letters* (2016).

100. Wei, X. & Luo, T. Effects of Electrostatic Interaction and Chirality on the Friction Coefficient of Water Flow Inside Single-Walled Carbon Nanotubes and Boron Nitride Nanotubes. *The Journal of Physical Chemistry C* **122**, 5131-5140 (2018).

101. Schiffbauer, J. & Luo, T. Liquid phase stabilization versus bubble formation at a nanoscale curved interface. *Physical Review E* **97**, 033106 (2018).

102. Yang, Junlong and Pang, Yunsong and Huang, Weixin and Shaw, Scott K and Schiffbauer, Jarrod and Pillers, Michelle Anne and Mu, Xin and Luo, Shirui and Zhang, Teng and Huang, Yajiang and others. Functionalized Graphene Enables Highly Efficient Solar Thermal Steam Generation. *ACS Nano* **11**, 5510 (2017).

103. Huang, D., Ma, R., Zhang, T. & Luo, T. Origin of Hydrophilic Surface Functionalization-Induced Thermal Conductance Enhancement across Solid–Water Interfaces. *ACS applied materials & interfaces* **10**, 28159-28165 (2018).

104. Wei, X., Zhang, T. & Luo, T. Molecular Fin Effect from Heterogeneous Self-Assembled Monolayer Enhances Thermal Conductance across Hard–Soft Interfaces. *ACS Appl. Mater. Interfaces* **9**, 33740-33748 (2017).

105. Yu, W., Peng, W., Shu, Y., Zeng, Q. & Jiang, M. Experimental Evidence Extraction System in Data Science with Hybrid Table Features and Ensemble Learning.

106. Zeng, Q.Valproic Acid Stimulates Hippocampal Neurogenesis via Activating the Wnt/beta-Catenin Signaling Pathway in the APP/PS1/Nestin-GFP Triple Transgenic Mouse Model of Alzheimer's Disease. *Front. Aging Neurosci.* **11**, 62 (2019).

107. Zeng, Q. *Faceted hierarchy: A new graph type to organize scientific concepts and a construction method* (Proceedings of the Thirteenth Workshop on Graph-Based Methods for Natural Language Processing (TextGraphs-13), 2019).