

MOLECULAR GRAPH STRUCTURE BASED MODELING TO PREDICT PROPERTIES OF POLYMERS UTILIZING LOCAL CLUSTERS

Technical Field

[0001] The present disclosure relates to molecular graph structure based modeling to predict properties of polymers utilizing local oligomeric structures. Such techniques can be particularly useful to predict the density and/or dielectric constant of designated polymers to alter the behavior of designated polymer formulations.

Background

[0002] Polymer materials can be utilized in a variety of products. One material utilized in flexible organic light emitting diodes (OLEDs) and other products is low dielectric constant (Dk) thin film encapsulation (TFE) material. Using TFE material with a low Dk value offers several benefits over other materials including reducing parasitic capacitance, improving response time, reducing electric consumption, and enhancing insulation performance when utilized with flexible OLEDs. Therefore, it is highly desirable to utilize TFE with low Dk values in OLED panels. However, the current TFE material used for OLED touch panels has a Dk value greater than 3.0, limiting the ability of customers to develop highly performing new flexible displays. There is a demand for TFE materials with even lower Dk values (e.g., less than 2.4) for the next generation of flexible OLED displays.

[0003] Currently, the investigation of polymer properties necessitates a multitude of experiments involving diverse polymer monomers and various polymerization conditions. These methodologies often entail significant manpower and material resources, leading to high costs and reduced efficiency in the development of new materials through trial and error. The cost of traditional high-precision theoretical calculations for polymer systems remains prohibitive, due to the exponential increase in computational complexity.

Summary of the Disclosure

[0004] The present disclosure is directed to using improvements in machine learning technology to predict a property of a polymer sample. The prediction can be based on a set of descriptors that have been categorized based local cluster properties of a plurality of polymers

and/or molecular representation from graph neural network. The descriptor can be calculated by DFT (density functional theory) on the local oligomeric structures, i.e., the local cluster properties (e.g., dielectric constant, etc.) of a plurality of polymers. The molecular representation can be generated by inputting a molecular graph to a pre-trained graph neural network. The combination of descriptors on oligomeric structures and representation on molecular structures can predict the polymer properties with significantly fewer computer resources compared to previous systems or methods, which can be advantageous over these systems and methods.

[0005] The above summary of the present disclosure is not intended to describe each disclosed embodiment or every implementation of the present disclosure. The description that follows more particularly exemplifies illustrative embodiments. In several places throughout the application, guidance is provided through lists of examples, which examples can be used in various combinations. In each instance, the recited list serves only as a representative group and should not be interpreted as an exclusive list.

Brief Description of the Drawings

[0006] Figure 1 illustrates one example flow diagram of a method for molecular graph structure based modeling to predict properties of polymers utilizing local structures.

[0007] Figure 2 illustrates one example flow diagram of a method for molecular graph structure based modeling to predict properties of polymers utilizing local structures.

[0008] Figure 3 illustrates one example flow diagram of a method for molecular graph structure based modeling to predict properties of polymers utilizing local structures.

[0009] Figure 4 illustrates an example of a machine readable medium for molecular graph structure based modeling to predict properties of polymers utilizing local structures.

[0010] Figure 5 illustrates an example of a device for molecular graph structure based modeling to predict properties of polymers utilizing local structures.

Detailed Description

[0011] The present disclosure relates to methods and devices for molecular graph structure based modeling to predict properties of polymers, which may utilize machine learning models to predict the dielectric constant of designated polymers. A machine learning model can

be a function or equation for identifying patterns in data. A machine learning module can be a plurality of machine learning models utilized together to identify patterns in data. In a specific example, a machine learning module can be organized as a neural network. A neural network can include a set of instructions that can be executed to recognize patterns in data. Some neural networks can be used to recognize underlying relationships in a set of data in a manner that mimics the way that a human brain operates. A neural network can adapt to varying or changing inputs such that the neural network can generate a best possible result in the absence of redesigning the output criteria.

[0012] A neural network can include multiple neurons, which can be represented by one or more equations or functions. In the context of neural networks, a neuron can receive a quantity of numbers or vectors as inputs and, based on properties of the neural network, produce an output. For example, a neuron can receive X_k inputs, with k corresponding to an index of inputs. For each input, the neuron can assign a weight vector, W_k , to the input. The weight vectors (e.g., weight value, etc.) can, in some embodiments, make the neurons in a neural network distinct from one or more different neurons in the network. In some neural networks, respective input vectors can be multiplied by respective weight vectors to yield a value, as shown by Equation 1, which shows an example of a linear combination of the input vectors and the weight vectors.

$$f(x_1, x_2) = w_1x_1 + w_2x_2$$

Equation 1

[0013] In some neural networks, a non-linear function (e.g., an activation function) can be applied to the value $f(x_1, x_2)$ that results from Equation 1. An example of a non-linear function that can be applied to the value that results from Equation 1 is a rectified linear unit function (ReLU). Application of the ReLU function, which is shown by Equation 2, yields the value input to the function if the value is greater than zero, or zero if the value input to the function is less than zero. The ReLU function is used here merely as an illustrative example of an activation function and is not intended to be limiting. Other non-limiting examples of activation functions that can be applied in the context of neural networks can include sigmoid functions, binary step functions, linear activation functions, hyperbolic functions, leaky ReLU functions, parametric ReLU functions, softmax functions, and/or swish functions, among others.

$$ReLU(x) = \max(x, 0)$$

Equation 2

[0014] During a process of training a neural network the weight vectors can be altered to “tune” the network. In at least one example, a neural network can be initialized with random weights. Over time, the weights can be adjusted to improve the accuracy of the neural network. This can, over time, yield a neural network with high accuracy. The present disclosure utilizes machine learning such as neural networks for predicting product properties through modeling of input data. In these embodiments, the weights can be tuned based on a number of factors. For example, the weights can be tuned utilizing a plurality of local clusters (e.g., two body structures, four body structures) and corresponding density values to determine a dielectric constant for a particular polymer. In this example, the machine learning model can utilize monomer data and an output value of a graph neural network to determine an unknown dielectric constant of a particular polymer based on identified local clusters of the particular polymer.

[0015] The present disclosure relates to a neural network combining property computation based on microstructural properties (e.g., local cluster properties, etc.) with molecular representations from graph neural networks. In some embodiments, the monomer structure file of the polymer of interest (e.g., designated polymer, etc.) can be obtained and converted into an oligomeric structure or local clusters. In some embodiments, the oligomeric structures can refer to a complex formed from a few monomer units. In some embodiments, the oligomeric structures can be synthesized through controlled chemical reactions.

[0016] To resemble the structure in the actual polymer, a methyl group or other auxiliary group is added to the terminal linkage of the repeating unit. As used herein, the terminal linkage (e.g., terminal connection, etc.) of the repeating unit refers to the chemical bonds or groups that exist at the ends of a polymer chain, connecting the last repeating unit of the polymer to the rest of the molecule or to some other entity. These terminal linkages can affect properties of the polymer, including, but not limited to: stability, reactivity, and/or how the polymer interacts with other substances.

[0017] For example, the polymer microstructure can be approximated by local clusters (e.g., two-body oligomers, four-body oligomers, etc.) and supplementary groups, such as methyl groups, can be added to the terminal linkages. For the two-body and four-body molecules within the local clusters, quantum chemical methods such as Parametric Method 6 (PM6) or density functional methods like Becke's three-parameter exchange functional combined with the Lee,

Yang, and Parr correlation function (B3LYP) can be utilized to simulate and calculate the free energy of the local clusters of the polymer (e.g., monomers, two-body, and four-body molecules), as well as properties to be predicted.

[0018] The machine learning model (e.g., trained polymer property prediction neural network, etc.) can utilize the calculated information of density and/or dielectric constant along with the previously obtained graph structure information of monomer molecules from the graph neural network as a combined input. This combined input serves to predict particular properties of the target polymer. For example, the combined input can be utilized to predict a dielectric constant of a designated polymer that includes a designated quantity of particular local clusters.

[0019] As used herein, the singular forms “a”, “an”, and “the” include singular and plural referents unless the content clearly dictates otherwise. Furthermore, the word “may” is used throughout this application in a permissive sense (e.g., having the potential to, being able to), not in a mandatory sense (e.g., must). The term “include,” and derivations thereof, mean “including, but not limited to.”

[0020] As will be appreciated, elements shown in the various embodiments herein can be added, exchanged, and/or eliminated so as to provide a number of additional embodiments of the present disclosure. In addition, as will be appreciated, the proportion and the relative scale of the elements provided in the figures are intended to illustrate certain embodiments of the present invention and should not be taken in a limiting sense.

[0021] Figure 1 illustrates one example flow diagram of a method 100 for molecular graph structure based modeling to predict properties of polymers utilizing local structures (e.g., local oligomeric structures, etc.). The method 100 can be executed by a computing device as described herein. The method 100 can be utilized to train and/or utilize a neural network or other type of machine learning model to determine properties of a polymer utilizing possible local structures.

[0022] In some embodiments, the method 100 can be utilized to determine a density and/or dielectric constant (Dk) of a polymer. As described herein, the dielectric constant is a measure of a material's ability to store electrical energy in an electric field. It's a dimensionless quantity that represents a ratio of the permittivity of a material to the permittivity of a vacuum. The higher the dielectric constant of a material, the more charge it can store. As described herein, a polymer with a relatively low dielectric constant can be beneficial for particular devices by

reducing parasitic capacitance, improving response time, reducing electric consumption, and enhancing insulation performance. For example, polymers with relatively low dielectric constants can have properties that are useful when utilized with flexible OLEDs.

[0023] The method 100 can include generating a molecular graph for a plurality of polymers utilizing a Simplified Molecular Input Line Entry System (SMILES) representation 102 of the plurality of polymers. SMILES representations refer to a notation system for representing chemical structures in a concise and human-readable format. SMILES is designed to be both machine-readable and relatively easy to write and understand. SMILES uses simple American Standard Code for Information Interchange (ASCII) characters to represent atoms, bonds, and the connectivity of atoms within a molecule.

[0024] The method 100 can utilize the SMILES representations to identify complementary monomers 104 of the plurality of polymers. As used herein, complementary monomers refers to pairs of monomers that have chemical structures or properties that enable them to react together in a specific manner to form a polymer. These monomer pairs are often chosen strategically to create polymer chains with specific characteristics or functionalities. For example, complementary monomers can have characteristics or functionalities such as: condensation polymerization, addition polymerization, copolymerization, and/or cross-linking.

[0025] In some embodiments, the method 100 can categorize the complementary monomers 104 a plurality of categories. For example, the complementary monomers 104 can be categorized into two-body local clusters 106-1 and/or four-body local clusters 106-2. In other embodiments, a greater number of categories can exist. For example, the method 100 can include additional categories such as monomer clusters. As used herein, a cluster or structure of polymer molecules can refer to groups of molecules of the polymer that are interacting or associating in a particular way. These clusters can vary in size and structure, and they play a role in determining the physical properties of polymer materials.

[0026] The two-body local clusters 106-1 can refer to a polymer dimer. In this case, two polymer chains associate with each other through various types of interactions, such as van der Waals forces, hydrogen bonding, or hydrophobic interactions. This dimeric association can occur between two identical polymer chains or between two different polymer chains. Dimeric associations can affect the properties of the polymer, including its solubility, viscosity, and mechanical behavior. Four-body local clusters 106-2 can refer to a tetrameric association of

polymer molecules. In this arrangement, four polymer chains come together through various interactions.

[0027] Tetrameric associations can occur in more complex polymer systems and may involve multiple intermolecular forces and entanglements. These clusters can be important for understanding the behavior of polymers in solutions, melts, or solid-state configurations. In the case of associating polymers like certain types of polyelectrolytes, four polymer chains might form a tetrameric association through a combination of electrostatic interactions, van der Waals forces, and hydrogen bonding. Such associations can influence the polymer's behavior in solution and its response to changes in pH and/or ionic strength.

[0028] In some embodiments, conformational sampling by Molecular Dynamics (MD) simulation can be utilized to model the two-body local clusters 106-1 and/or the four-body local clusters 106-2. As used herein, conformational sampling by a MD simulation refers to a computational method used to study the physical movements and conformational changes of molecules, particularly large biological macromolecules like proteins, nucleic acids, and lipids. In some embodiments, the MD simulation can include a dynamic simulation to calculate time-dependent behavior of a molecular system (e.g., molecules interactions that evolve over a period of time under an influence of physical laws, etc.). In addition, the MD simulation can include force field interactions within the molecules and between the molecules and the environment. In these simulations, the interactions are governed by mathematical functions that are referred to as a force field. The force field can refer to bonds, angles, dihedrals, van der Waals forces, and/or electrostatic forces.

[0029] In some embodiments, the MD simulation can include predicting movements of molecules through Newton's laws of motion. In addition, the conformational sampling refers to the process of exploring the various spatial arrangements that a particular molecule can adapt. In this way, the conformational sampling can utilize the MD simulation to explore the various spatial relationships of the two-body local clusters 106-1 and/or four-body local clusters 106-2.

[0030] The two-body local clusters 106-1 can be analyzed to determine two-body properties 108-1 (e.g., density, dielectric constant, electronic structures, etc.) and the four-body local clusters 106-2 can be analyzed to determine four-body properties 108-2 (e.g., density, dielectric constant, electronic structures, etc.). In some embodiments, the two-body properties 108-1 and four-body properties 108-2 can be based on particular properties to be predicted by the

method 100. For example, the two-body properties 108-1 and the four-body properties 108-2 can be density values and/or dielectric constant values associated with the two-body local clusters 106-1 and four-body local clusters 106-2 respectively.

[0031] In some embodiments, the two-body properties 108-1 can include calculated two-body density values 110-1. In addition, the four-body properties 108-2 can include calculated four-body density values 110-2. In some embodiments, two-body density values 110-1 and four-body density values 110-2 can be provided as input values to a first machine learning model 114. In addition, graph neural network data 112 can be provided as an input to the first machine learning model 114. As described herein, the graph neural network data 112 can be data associated with a graph neural network of the monomeric molecules to form the designated polymers. As used herein, a graph neural network (GNN) is a type of neural network architecture designed to process and analyze data represented as graphs. A GNN can be used to learn and analyze the structural and chemical properties of the molecules, where the molecule's atoms and their connections form a graph-like structure. The output of a GNN applied to a monomeric molecule can provide different data associated with properties and/or structures of the monomeric molecules. For example, the output of GNN data 112 can include, but is not limited to: node embeddings, graph embeddings, property predictions, chemical activity, visualization, quantitative descriptors, graph-based analysis, and/or interactions of the polymer.

[0032] As described herein, the input data can be provided to the GNN. In these embodiments, the input data provided to the GNN includes the monomeric molecules that are identified after substituting the functional groups. As described herein, the functional groups can be positioned at terminal junctions to produce different stable monomeric molecules.

[0033] As described herein, the GNN data 112 can include node embeddings, which can refer to vector representations of each atom in the monomeric molecule. These embeddings encode information about the local environment of each atom, including its connectivity, neighboring atoms, and chemical properties. Node embeddings can be useful for various downstream tasks, such as property prediction, classification, or clustering of monomeric molecules. As described herein, the GNN data 112 can include graph embedding, which can refer to a graph-level embedding that summarizes the entire monomeric molecule. This embedding captures the overall structural and chemical characteristics of the monomeric

molecule. It can be used for tasks like polymer classification, similarity comparison, or property prediction at the molecule level.

[0034] As described herein, the GNN data 112 can include property predictions of such properties as mechanical properties (e.g., tensile strength, elasticity), thermal properties (e.g., melting point, glass transition temperature), and/or chemical properties (e.g., reactivity, solubility). The predicted properties can be scalar values or multi-dimensional vectors. As described herein, the GNN data 112 can include the chemical reactivity or activity of specific functional groups or bonds within the polymer. This information can be valuable for understanding how the polymer might react in different chemical environments or for designing polymers with specific reactivity profiles. As described herein, the GNN data 112 can include visualization, which can refer to visualizing the polymer's structure and properties. The GNN data 112 can generate two-dimensional (2D) or three-dimensional (3D) representations of the polymer molecule, highlighting key structural features or regions of interest. Visualization can aid researchers in understanding the polymer's behavior.

[0035] As described herein, the GNN data 112 can include quantitative descriptors such as molecular fingerprints or topological indices. These descriptors can be used for similarity searching, virtual screening, or other chemoinformatics tasks. The GNN data 112 can analyze the graph structure of the polymer, identifying important substructures (e.g., functional groups, repeating units) or detecting anomalies or defects in the polymer chain. As described herein, the GNN data 112 can predict interactions between polymer molecules and other molecules, such as solvents, additives, or nanoparticles. This information is crucial for studying the behavior of polymers in various applications.

[0036] The combined input of the two-body density values 110-1, four-body density values 110-2, and GNN data 112 can be provided to the first machine learning model 114. The first machine learning model 114 is trained to determine a predicted density 116 of the designated polymer based on those inputs. In some embodiments, two-body dielectric constant data 118-1 can be calculated from the two-body properties 108-1. In a similar way, the four-body dielectric constant data 118-2 can be calculated from the four-body properties 108-2. The predicted density 116 of the designated polymer can be utilized as a combined input with two-body dielectric constant data 118-1 and four-body dielectric constant data 118-2 to a second machine learning model 122. that the second machine learning model 122 is trained to determine

an output dielectric constant 124 of the designated polymer based on those inputs. In this way, the method 100 can be utilized to calculate a dielectric constant of a large polymer molecule with relatively less computing resources compared to previous systems and methods.

[0037] Figure 2 illustrates one example flow diagram of a method 220 for molecular graph structure based modeling to predict properties of polymers utilizing local structures (e.g., local oligomeric structures, etc.). In some examples, the method 220 can be executed by a computing device as described herein. The method 220 can be utilized to train a neural network or other type of machine learning model to determine a polymer properties of a designated or desired polymer based on local structure data and GNN output data. In some embodiments, the trained neural network can be utilized to reverse engineer polymers that are capable of reducing parasitic capacitance, improving response time, reducing electric consumption, and enhancing insulation performance when utilized with flexible OLEDs.

[0038] At step 244, the method 220 can include generating 3D molecular structures from SMILES structures of the polymer monomers. In some embodiments, the polymer monomers can be monomer structures for a plurality of different polymers or possible polymers. In this way, the plurality of SMILES representations can reflect properties that can be parts of different polymers. The SMILES representations can be one-dimensional notations that focus on chemical connectivity while the 3D structures provide a detailed and three-dimensional depiction of the polymer's spatial arrangement, enabling a more comprehensive understanding of its structure, conformation, and properties. In some embodiments, multiple 3D structures can be provided for each of the plurality of SMILES representations. For example, a particular SMILES representation can be utilized to generate a “cis” 3D structure orientation and a “trans” 3D structure orientation.

[0039] At step 246, the method 220 can include substituting methyl groups or other functional groups to the terminal junctions of the repeating units to form stable monomer structures. In some embodiments, substituting the methyl groups or other functional groups can be utilized to determine a plurality of stable monomer structures that can be utilized to generate different polymer structures. For example, different combinations methyl groups or other functional groups can generate stable monomer structures while other combinations can generate non-stable monomer structures. In this way, the non-stable monomer structures can be filtered out or not utilized. In some embodiments, the orientation of the 3D structure can be altered when

different functional groups are added to the terminal junctions, which can either create a more stable molecule or less stable molecule.

[0040] At step 248, the method 220 can include connecting the monomers to generate a plurality of possible two-body local clusters and four-body local clusters to form stable structures. The different combinations of methyl groups or other functional groups can generate a plurality of two-body local structures and four-body local structures that are stable or substantially stable to be utilized by particular polymers. In this way, unstable two-body clusters and/or unstable four-body clusters can be disregarded or removed from future calculations.

[0041] At step 252, the method 220 can include simulating a plurality of monomer, two-body, and four body local clusters (e.g., stable structures, etc.) to determine different conformations and corresponding partition functions. In some embodiments, a plurality of simulated monomer, two-body local clusters, and/or four-body local clusters can be simulated by determining the stable structures of step 248. At step 254, the method 220 can include utilizing quantum chemical methods for property calculations. In some examples, the quantum chemical methods can include, but are not limited to: PM6, extended tight binding (XTB), B3LYP and/or M06-2X to determine properties of the two-body and four-body local structures.

[0042] As used herein, the XTB method can include a semi-empirical (uses both theoretical approximations and empirical data) quantum chemical method used to approximate the electronic structure of molecules. As used herein, the M06-2X method refers to a method for density functional theory (DFT) calculations. The M06-2X is a hybrid meta-GGA (generalized gradient approximation) functional. It includes a higher amount of Hartree-Fock exchange (about 54%) compared to many other functionals, which helps in accurately modeling noncovalent interactions and transition states. In some embodiments, the M06-2X method can accurately describe noncovalent interactions and the kinetics of chemical reactions.

[0043] As used herein, the DFT calculations can refer to a quantum mechanical modeling method used to investigate the electronic structure of many-body systems, especially atoms, molecules, and the condensed phases. For example, the DFT calculations can include a computational quantum mechanical modeling method, focusing on the electron density rather than the wave function to calculate properties of matter. In some embodiments, the DFT calculations can be based on the electron density as the primary quantity. It can utilize theorems that allow properties of a system to be determined by the spatial electron density distribution. In

some embodiments, the DFT calculations can utilize Hohenberg-Kohn theorems and/or Kohn-Sham equations. The DFT calculations can utilize energy calculations, material properties, and/or functional approximations.

[0044] For example, the DFT calculations can calculate a total energy of a system (e.g., two-body local cluster, four-body local cluster, etc.), which can include contributions from kinetic, potential, and/or electron-electron interaction energies. In addition, the DFT calculations can predict physical and/or chemical properties of materials and molecules, such as molecular structure, reaction energies, electronic properties, and/or other properties. In some embodiments, the exact functional form may not be known, thus approximations can be utilized. Approximations such as local density approximations (LDA) and/or generalized gradient approximations (GGA) can be utilized.

[0045] At step 256, the method 220 can include calculating a statistical average of corresponding properties calculated by the partition function. Calculating the statistical average of corresponding properties can be based on the quantity of two-body local structure properties and four-body local structure properties within a designated polymer. In some embodiments, the statistical average can refer to a mathematical calculation such as a mean, mode, or other mathematical averaging. In this way, the quantity of the two-body and four-body local structures with particular properties can be utilized to determine the statistical averages for each property. In some embodiments, the statistical average can be a measure of the central tendency of a set of values for that property. The statistical average can summarize a set of data with a single number representing the center point of the data distribution.

[0046] At step 258, the method 220 can include utilizing the computed properties to predict the polymer properties. As described herein, the combined properties of the two-body local structures and four-body local structures can be provided to generate the predicted properties of the polymer.

[0047] Figure 3 illustrates one example flow diagram of a method 330 for molecular graph structure based modeling to predict properties of polymers utilizing local structures (e.g., local oligomeric structures, etc.). In some examples, the method 330 can be executed by a computing device as described herein. The method 330 can be utilized to train a neural network or other type of machine learning model to determine a polymer properties of a designated or desired polymer based on local structure data and GNN output data. In some embodiments, the

trained neural network can be utilized to reverse engineer polymers that are capable of reducing parasitic capacitance, improving response time, reducing electric consumption, and enhancing insulation performance when utilized with flexible OLEDs.

[0048] At step 361, the method 330 can be executed to generate a data set that includes a respective structure and properties for each of a plurality of monomers corresponding to a designated polymer. In some embodiments, the respective structure and properties corresponding to a designated polymer can be a particular structure and/or properties that are desired for a particular function. For example, the respective structure and properties can be for a particular purpose, such as, but not limited to OLED display materials. In some embodiments, the generated data set of the plurality of monomers can be selected based on a stability of the plurality of monomers. In some embodiments, the data set can include a plurality of monomers that are simulated based on known monomers. That is, the plurality of monomers can be virtual representations of monomers that include particular structure features and/or particular properties that correspond to desired properties of monomers that have desired properties such as a particular dielectric constant or a dielectric constant within a particular range.

[0049] In some embodiments, the method 330 can be executed to generate the structure of the plurality of polymers utilizing a SMILES representation of the plurality of polymers. As described herein, the structure of the plurality of polymers can be SMILES representations that can be utilized to represent the chemical connections of the plurality of polymers. In some embodiments, the plurality of polymers can be converted from the SMILES representations to 3D representations to represent the orientations of the plurality of polymers.

[0050] In some embodiments, the method 330 can be executed to simulate the plurality of polymers for the data set by substituting functional groups at terminal junctions of repeating monomer units of the plurality of polymers. In some embodiments, substituting functional groups can be utilized to determine a plurality of stable monomer structures that can be utilized to generate different polymer structures. For example, different combinations of methyl groups or other functional groups can generate stable monomer structures while other combinations can generate non-stable monomer structures. In this way, the non-stable monomer structures can be filtered out or not utilized. In some embodiments, the orientation of the 3D structure can be altered when different functional groups are added to the terminal junctions, which can either create a more stable molecule or less stable molecule.

[0051] At step 362, the method 330 can be executed to convert the respective structures to graph connectivity structures. In some embodiments, the plurality of local clusters and/or polymers that are determined to be stable structures can be converted or utilized to generate graph connectivity structures. The graph connectivity structures can be GNN representations of the plurality of local clusters and/or GNN representations of a polymer.

[0052] At step 363, the method 330 can be executed to determine a plurality of local clusters based on the plurality of monomers. Determining the plurality of local clusters can include identifying two-body clusters and/or three-body clusters from the plurality of monomers that are determined to be stable. In some embodiments, determining the plurality of local clusters can include identifying a plurality of local clusters for a designated polymer. Identifying the plurality of local clusters can include identifying the designated polymer as a combination of numerous local clusters, and the polymer properties can be approximately expressed by the statistical average of the properties of the local clusters. In some embodiments, the topological connection structure and interaction between local clusters (e.g., dipole-dipole interaction) can remain basically unchanged during the optimization process.

[0053] In principle, the overall polymer properties remain unchanged as long as the conformation distribution of local clusters and the interaction energy therein are identified. The local cluster model focuses on the local microstructures of polymers, expecting that the properties corresponding to these structures can become the basis for the prediction of overall polymer properties. In principle, if a size of the local clusters increases, the investigated scope can continue to increase until it approaches the whole polymer. Unfortunately, with the increase of the local cluster size and the exponential increase of conformation numbers, the computational cost increases exponentially, and eventually may exceed computational power.

[0054] Considering the balance between calculation accuracy and efficiency, a threshold size (e.g., molecular weight, quantity of particularly sized atoms, etc.) of the monomers can be selected. For example, a threshold of 20 heavy atoms can be selected where each of the plurality of local clusters does not exceed 20 heavy atoms. For some cases with cross-connections, a single chain may need to meet a basic structure of the T-cross or cross-cross. As used herein, cross-connections or cross-links refer to bonds that link one polymer chain to another polymer chain. For example, the cross-connections can be covalent cross-connections, ionic cross-connections, and/or physical cross-connections. As used herein, T-cross and cross-cross

connections can refer to specific types of cross-connections or cross-linking. For example, T-cross connections can occur when one polymer chain forms a 'T' shape with another polymer chain. For example, a single polymer chain acting as a branch can connect at a point along the length of another polymer chain, forming a 'T' shape. The cross-cross connection can occur when multiple cross-links occur between the same two polymer chains or where one polymer chain is linked to many others at several points. Some special cross structures may cause all local clusters to be connected into a whole. At the same time, the representative molecular conformation can affect characteristics such as charge mobility, molecular shape and cavity size.

[0055] In some embodiments, the method 330 can be executed to simulate monomer, two-body, and four body stable structures from the plurality of polymers to determine the plurality of local clusters. In these embodiments, the method 330 can be executed to calculate a density of the plurality of local clusters utilizing a density functional method and calculating the dielectric constant of the plurality of local clusters utilizing a quantum chemical method. As described herein, each of the monomer, two-body, and four-body structures can be utilized to determine a corresponding density, dielectric constant, or other property. In this way, the statistical average of the monomer, two-body, and/or four-body structures within the designated polymer can be utilized to determine the corresponding property of the entire polymer.

[0056] At step 364, the method 330 can be executed to calculate a respective polymeric property of each of the plurality of local clusters based on a density functional theory (DFT). In some embodiments, the density of the plurality of local clusters can be utilized to determine a dielectric constant of the plurality of local clusters. In a similar way as the density, a statistical average of the determined dielectric constant for the plurality of local clusters can be utilized to determine the dielectric constant for the whole polymer or the polymer as a whole.

[0057] At step 365, the method 330 can be executed to input the respective polymeric properties and an output value of a graph neural network to a machine learning model trained to determine an unknown polymeric property of a designated polymer. In some embodiments, monomer data or polymeric properties of the monomers can be provided as inputs to a machine learning model trained to determine an unknown polymeric property of a designated polymer. The input monomer data and/or local cluster data can include the respective polymeric properties of the local clusters. The local cluster data can be provided to the machine learning model as

described herein. In some examples, the input monomer data and/or local cluster data can include the density data associated with the monomer data and/or local cluster data.

[0058] As described herein, the density data of the two-body clusters and four-body clusters can be provided along with the output of the GNN to determine a density of an entire designated polymer. In these examples, the density data of the designated polymer can be utilized as an input along with calculated two-body and four-body dielectric constant data to determine a dielectric constant for the entire designated polymer.

[0059] In some embodiments, the method 330 can be executed to provide the GNN with inputs including an atomic nuclear charge number, an atomic net charge of each of a plurality of monomer molecules, a heavy atom connection number, and a hydrogen atom connection number of the designated polymer. With this type of data, the GNN can be used to learn and analyze the structural and chemical properties of the designated polymer. As described herein, the output of a GNN applied to a polymer molecule can provide different data associated with properties and/or structures of the polymer. For example, the output of GNN data can include, but is not limited to: node embeddings, graph embeddings, property predictions, chemical activity, visualization, quantitative descriptors, graph-based analysis, and/or interactions of the polymer.

[0060] At step 366, the method 330 can be executed to receive a prediction of the unknown polymeric property of the designated polymer from the machine learning model. As described herein, the polymeric property can be a predicted unknown dielectric constant. The predicted unknown dielectric constant can be based on a statistical average of the dielectric constant of the plurality of local clusters associated with the designated polymer. As described herein, the machine learning model or plurality of machine learning models can be trained to predict the unknown dielectric constant of the designated polymer based on the monomer data and/or local cluster data. In some embodiments, the method 330 can be executed to determine the dielectric constant of the designated polymer utilizing a statistical average of corresponding properties for the plurality of local clusters.

[0061] Figure 4 illustrates an example of a machine readable medium 440 for molecular graph structure based modeling to predict properties of polymers utilizing local structures (e.g., local oligomeric structures, etc.). The machine readable medium 440 can be communicatively connected to a processor resource 471 by a communication path 472. In some examples, a communication path 472 can include a wired or wireless connection that can allow

communication between devices and/or components within a single device. As used herein, the processor resource 471 can include, but is not limited to: a central processing unit (CPU), an application specific integrated circuit (ASIC), a field programmable gate array (FPGA), a metal-programmable cell array (MPCA), a semiconductor-based microprocessor, or other combination of circuitry and/or logic to orchestrate execution of instructions 473, 474, 475, 476, 477, 478. In a specific example, the processor resource 471 utilizes a non-transitory computer-readable medium storing instructions 473, 474, 475, 476, 477, 578, that, when executed, cause the processor resource 471 to perform corresponding functions.

[0062] The machine readable medium 440 may be electronic, magnetic, optical, or other physical storage device that stores executable instructions. Thus, a non-transitory machine-readable medium (MRM) (e.g., machine readable medium 440) may be, for example, a non-transitory MRM comprising Random-Access Memory (RAM), read-only memory (ROM), an Electrically Erasable Programmable ROM (EEPROM), a storage drive, an optical disc, and the like. The machine readable medium 440 may be disposed within a controller and/or computing device. In this example, the executable instructions 473, 475, 476, 477, 478, can be “installed” on the device. Additionally, and/or alternatively, the machine readable medium 440 can be a portable, external, or remote storage medium, for example, which allows a computing system to download the instructions 473, 475, 476, 477, 478, from the portable/external/remote storage medium. In this situation, the executable instructions may be part of an “installation package”.

[0063] The machine readable medium 440 includes instructions 473 to generate a data set that includes a structure and properties for a plurality of known polymers. In some embodiments, the data set includes structure and properties that are desired structures and properties of a designated polymer. In this way, the data set can include information regarding the structure and properties of known polymers that include the properties of the designated polymer.

[0064] In some embodiments, machine readable medium 440 can include instructions to predict a plurality of polymer structures that are within a defined dielectric constant range and a defined density range utilizing the machine learning model. As described herein, the plurality of polymer structures can be simulated representations of real polymer structures that include desired properties. In this way, the defined dielectric constant range and/or defined density range can be desired properties of a designated polymer.

[0065] The machine readable medium 440 includes instructions 474 to generate a molecular graph for each of the plurality of known polymers from a SMILES representation of the plurality of known polymers. The machine readable medium 440 includes instructions 475 to determine a plurality of local clusters for each of the plurality of known polymers, wherein the plurality of local clusters include simulated monomer, two-body, and four body stable structures from the plurality of known polymers.

[0066] The machine readable medium 440 includes instructions 476 to calculate a density and a dielectric constant for each of the plurality of local clusters. The machine readable medium 440 includes instructions 477 to calculate a statistical average of the density and the dielectric constant for the plurality of local clusters for each of the plurality of known polymers. In some embodiments, the statistical average is based on a quantity of each of the plurality of local clusters.

[0067] The machine readable medium 440 includes instructions 478 to train a machine learning model to determine an unknown dielectric constant of a designated polymer from an output of a GNN of the designated polymer and density data determined from a molecular graph structure of the designated polymer. Training the machine learning model can include altering and/or updating the weights or weight vectors of the machine learning model to improve the accuracy of the machine learning model's ability to accurately predict a property of the polymer based on the inputs. As described herein, the inputs can include the output of the GNN and the density data and/or dielectric constant data for two-body and four body local clusters.

[0068] In some embodiments, the machine readable medium 440 includes instructions to provide an atomic nuclear charge number, an atomic net charge of each of the monomer molecules, a heavy atom connection number, and a hydrogen atom connection number of the designated polymer as an input to the GNN. In these embodiments, the machine readable medium 440 can include instructions to provide an output of the GNN as an input for the machine learning model.

[0069] In some embodiments, the machine readable medium 440 can include instructions to train the machine learning model to utilize the density data of two-body and four body local clusters as descriptors. The two-body and four-body local clusters can be utilized as inputs or descriptors of the machine learning model. As used herein, descriptors refer to features or attributes used to represent and understand the data within the machine learning model. The

descriptors can serve as the foundational input that the machine learning model can utilize to make predictions or classifications.

[0070] Figure 5 illustrates an example of a device 550 for molecular graph structure based modeling to predict properties of polymers utilizing local structures (e.g., local oligomeric structures, etc.). In some examples, the device 550 is a computing device that includes a processor resource 571 and a machine readable medium 540 to store instructions 582, 583, 584, 585, 586, 587, 588 that are executed by the processor resource 571 to perform particular functions. Figure 5 illustrates how a computing device can execute instructions to perform functions described herein.

[0071] The device 550 includes instructions 582 stored by the machine readable medium 650 that is executed by the processor resource 571 to generate a data set that includes a structure, a density, and a dielectric constant for a plurality of polymers. As described herein, the data set can include information related to simulated polymers that can include the same or similar properties as desired properties of a designated polymer.

[0072] The device 550 includes instructions 583 stored by the machine readable medium 650 that is executed by the processor resource 571 to generate a molecular graph for each of the plurality of polymers utilizing a SMILES representation of the plurality of polymers.

[0073] The device 550 includes instructions 584 stored by the machine readable medium 650 that is executed by the processor resource 571 to determine a plurality of local clusters for each of the plurality of polymers. In some embodiments, the device 550 can include instructions to simulate a free energy and corresponding property of the plurality of local clusters. In some embodiments, the plurality of local clusters include monomer, two-body, and four-body interactions of the plurality of polymers.

[0074] The device 550 includes instructions 585 stored by the machine readable medium 540 that is executed by the processor resource 571 to calculate a local density and a local dielectric constant for each of the plurality of local clusters. In some embodiments, the device 550 can include instructions to calculate the density of the plurality of local clusters utilizing a Becke's three-parameter exchange functional method combined with a Lee, Yang, and Parr correlation functional method (B3LYP) and the unknown dielectric constant utilizing a Parametric Method (PM6).

[0075] The device 550 includes instructions 586 stored by the machine readable medium 540 that is executed by the processor resource 571 to calculate a statistical average of the density and the dielectric constant for the plurality of local clusters for each of the plurality of polymers.

[0076] In some embodiments, the device 550 can include instructions to calculate an aggregated mean and maximum values for elements of the plurality of polymers and calculate an aggregated mean and maximum values for each of the plurality of polymers as whole molecules.

[0077] The device 550 includes instructions 587 stored by the machine readable medium 650 that is executed by the processor resource 571 to input monomer data to a machine learning model trained to determine an unknown dielectric constant of a designated polymer from an output value of a GNN and density values for a plurality of identified local clusters from the designated polymer. In some embodiments, the device 550 can include instructions to provide polymer property data as an input to the GNN.

[0078] In some embodiments, the device 550 can include instructions to determine a topological connectivity pattern of the unknown polymer. As used herein, the topological connectivity pattern includes a representation of how atoms or segments of the polymer are connected to each other, including the arrangement of the polymer chains and any branches or cross-links that may be present. That is, the topological connectivity pattern is a spatial map of the polymer's molecular structure and how each of its constituent parts are linked.

[0079] The device 550 includes instructions 588 stored by the machine readable medium 650 that is executed by the processor resource 571 to receive a prediction of the unknown dielectric constant of the designated polymer from the machine learning model. As described herein, a first machine learning model can be utilized to determine the density of the plurality of local clusters and a second machine learning model can be utilized to determine the dielectric constant based on the determined density. In this way, the desired dielectric constant of the designated polymer can be achieved.

[0080] Although specific embodiments have been described above, these embodiments are not intended to limit the scope of the present disclosure, even where only a single embodiment is described with respect to a particular feature. Examples of features provided in the disclosure are intended to be illustrative rather than restrictive unless stated otherwise. The above description is intended to cover such alternatives, modifications, and equivalents as would be apparent to a person skilled in the art having the benefit of this disclosure.

[0081] The scope of the present disclosure includes any feature or combination of features disclosed herein (either explicitly or implicitly), or any generalization thereof, whether or not it mitigates any or all of the problems addressed herein. Various advantages of the present disclosure have been described herein, but embodiments may provide some, all, or none of such advantages, or may provide other advantages.

[0082] In the foregoing Detailed Description, some features are grouped together in a single embodiment for the purpose of streamlining the disclosure. This method of disclosure is not to be interpreted as reflecting an intention that the disclosed embodiments of the present disclosure have to use more features than are expressly recited in each claim. Rather, as the following claims reflect, inventive subject matter lies in less than all features of a single disclosed embodiment. Thus, the following claims are hereby incorporated into the Detailed Description, with each claim standing on its own as a separate embodiment.

Claims

What is claimed is:

1. A method, comprising:
generating a data set that includes a respective structure and properties for each of a plurality of monomers corresponding to a designated polymer;
converting the respective structures to graph connectivity structures;
determining a plurality of local clusters based on the plurality of monomers;
calculating a respective polymeric property of each of the plurality of local clusters based on density functional theory (DFT);
inputting the respective polymeric properties and an output value of a graph neural network to a machine learning model trained to determine an unknown polymeric property of a designated polymer; and
receiving a prediction of the unknown polymeric property of the designated polymer from the machine learning model.
2. The method of claim 1, further comprising providing the graph neural network with inputs including an atomic nuclear charge number, an atomic net charge of each of a plurality of monomer molecules, a heavy atom connection number, and a hydrogen atom connection number of the designated polymer.
3. The method of claim 1, further comprising generating the structure of the plurality of polymers utilizing a Simplified Molecular Input Line Entry System.
4. The method of claim 1, further comprising simulating the plurality of polymers for the data set by substituting functional groups at terminal junctions of repeating monomer units of the plurality of polymers.
5. The method of claim 1, further comprising simulating monomer, two-body, and four body stable structures from the plurality of polymers to determine the plurality of local clusters.

6. The method of claim 5, further comprising calculating a density of the plurality of local clusters utilizing the DFT and calculating the dielectric constant of the plurality of local clusters utilizing a quantum chemical method.

7. The method of claim 1, further comprising determining a dielectric constant of the designated polymer utilizing a statistical average of corresponding properties for the plurality of local clusters.

8. A machine-readable medium, storing machine-readable instructions which, when executed by a processor resource of a device, cause the processor to:

generate a data set that includes a structure and properties for a plurality of known polymers;

generate a molecular graph for each of the plurality of known polymers from a Simplified Molecular Input Line Entry System (SMILES) representation of the plurality of known polymers;

determine a plurality of local clusters for each of the plurality of known polymers, wherein the plurality of local clusters include simulated monomer, two-body, and four body stable structures from the plurality of known polymers;

calculate a density and a dielectric constant for each of the plurality of local clusters;

calculate a statistical average of the density and the dielectric constant for the plurality of local clusters for each of the plurality of known polymers; and

train a machine learning model to determine an unknown dielectric constant of a designated polymer from an output of a graph neural network of the designated polymer and density data determined from a molecular graph structure of the designated polymer.

9. The machine-readable medium of claim 8, comprising instructions to identify a plurality of polymer structures that are within a defined dielectric constant range and a defined density range utilizing the machine learning model.

10. The machine-readable medium of claim 8, comprising instructions to provide an atomic nuclear charge number, an atomic net charge of each of a plurality of monomer molecules, a heavy atom connection number, and a hydrogen atom connection number of the designated polymer as an input to the graph neural network.

11. The machine-readable medium of claim 10, comprising instructions to provide an output of the graph neural network as an input for the machine learning model.

12. The machine-readable medium of claim 8, comprising instructions to train the machine learning model to utilize the density data of two-body and four body local clusters as descriptors.

13. The machine-readable medium of claim 8, wherein the statistical average is based on a quantity of each of the plurality of local clusters.

14. A device, comprising:
a processor; and
a non-transitory memory resource storing machine-readable instructions stored thereon that, when executed, cause the processor resource to:
generate a data set that includes a structure, a density, and a dielectric constant for a plurality of polymers;
generate a molecular graph for each of the plurality of polymers utilizing a Simplified Molecular Input Line Entry System (SMILES) representation of the plurality of polymers;
determine a plurality of local clusters for each of the plurality of polymers;
calculate a local density and a local dielectric constant for each of the plurality of local clusters;
calculate a statistical average of the density and the dielectric constant for the plurality of local clusters for each of the plurality of polymers;
input monomer data to a machine learning model trained to determine an unknown dielectric constant of a designated polymer from an output value of a graph neural

network and density values for a plurality of identified local clusters from the designated polymer; and

receive a prediction of the unknown dielectric constant of the designated polymer from the machine learning model.

15. The device of claim 14, wherein the processor is to calculate the density of the plurality of local clusters utilizing a Becke's three-parameter exchange functional method combined with a Lee, Yang, and Parr correlation functional method (B3LYP) and the unknown dielectric constant utilizing a Parametric Method (PM6).

16. The device of claim 14, wherein the processor is to determine a topological connectivity pattern of the unknown polymer.

17. The device of claim 14, wherein the processor is to provide polymer property data as an input to the graph neural network.

18. The device of claim 14, wherein the processor is to simulate a free energy and corresponding property of the plurality of local clusters.

19. The device of claim 18, wherein the plurality of local clusters include monomer, two-body, and four-body interactions of the plurality of polymers.

20. The device of claim 14, wherein the processor is to calculate an aggregated mean and maximum values for elements of the plurality of polymers and calculate an aggregated mean and maximum values for each of the plurality of polymers as whole molecules.