1-12-2015

Polymer Dielectrics Design Using First Principles Computations and Machine Learning

Chenchen Wang
University of Connecticut - Storrs, dvdcwangcc@gmail.com

Polymer Dielectrics Design Using First Principles Computations and Machine Learning

Chenchen Wang, Ph.D.

University of Connecticut, 2015

Polymers offer a nearly infinite variety of material systems with diverse properties. Until recently, the formulation of polymers for specific applications was based on trial and error, guided by intuition. In this work, first principles computations and machine learning approach are employed to guide the design of polymers, in the present case for dielectric applications. Specifically, we adopt two strategies, (1) functionalization of a well understood polymer dielectrics, such as PE and PP, to enhance its dielectric response, and (2) discovery of entirely new classes of polymer dielectrics, both organic and organometallic. Different polymer classes are explored, from C-based organic polymers to novel Si-, Ge-, and Sn-based polymers, and the search is based on two properties, band gap and dielectric constant. Newly developed high throughput DFT methods were used first to accurately determine the dielectric constant and band gap of different polymer systems for a set of limited compositions and configurations. Machine learning methods were then used to predict the properties of systems spanning a much larger part of the configurational and compositional space. Based on this strategy, we are able to provide a "map" of the achievable combination of properties within the chemical space explored.

Polymer Dielectrics Design Using First Principles Computations and Machine Learning

Chenchen Wang

B.S., Zhejiang University, China, 2008

A Dissertation

Submitted in Partial Fulfillment of the

Requirements for the Degree of

Doctor of Philosophy

at the

University of Connecticut

2015

APPROVAL PAGE

Doctor of Philosophy Dissertation

Polymer Dielectrics Design Using First Principles Computations and Machine Learning

Presented by

Chenchen Wang, B.S.

Major Advisor _____
Ramamurthy (Rampi) Ramprasad

Associate Advisor _____
Avinash M. Dongare

Associate Advisor _____
George Rossetti, Jr.

University of Connecticut

2015

to my husband Tony (Xun) Jiang

to my mother Yuying Chen and my father Li Wang

ACKNOWLEDGEMENTS

First and foremost, I would like to express my utmost gratitude to my major advisor Professor Rampi Ramprasad, for his insightful advice, his sincere support, and his generous time and commitment. His enthusiasm and attitude towards research encouraged and inspired me to get into this interesting yet challenging research area, and his guidance led me through the tough times in my Ph.D. pursuit.

I am really grateful that I can be part of this Multidisciplinary University Research initiative (MURI) project. It is a great team. The rigorous scientific attitude and insightful ideas of all the professors in the team influence me deeply. The useful discussions between students help me a lot in my research. Especially, I want to thank Prof. Boggs for his guidance and thoughtful academic suggestions. I also want to thank Prof. Sotzing and his students, Aaron Baldwin, Rui Ma, and Robert Lorenzini for their great help in my research.

I want to thank my associate advisors Professor Avinash M. Dongare and Professor George Rossetti for their kind help and valuable advice, which are important to me in the whole process of my graduate studies.

Many thanks are owed to my current and past group members for their help and priceless friendship, including Dr. Vinit Sharma, Dr. Clive Bealing, Dr. Huan Tran, Dr. Ghanshyam Pilania, Dr. Hong Zhu, Dr. Satyesh Kumar Yadav, Yenny P. Cardona Quintero, Hom Sharma, Venkatesh Botu, Lihua Chen, Arun Kumar Mannodi Kanakkithodi. I also would like to acknowledge the help from all my friends in UConn.

I want to dedicate this dissertation to my husband Tony (Xun) Jiang, my mother Yuying Chen, and my father Li Wang. I'll never forget the warm but firm support they give whenever I fall. Their everlasting love is the motivation for me to overcome and to achieve.

Finally, the financial support of this work from a Multi-University Research Initiative (MURI) grant of the Office of Naval Research, and computational support from an NSF sponsored Teragrid Resource Allocation are gratefully acknowledged.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

1.1 Polymer dielectrics

In the two decades after World War II, many polymers were invented, synthesized and commercialized for various applications. By the early 1960s, a number of polymers, which include polyethylene (PE), polypropylene (PP), polystyrene (PS), poly(vinylchloride), and other commodity polymers, was considered as dielectrics for electrical insulation. PE, PP, and PS have continued to be used for insulation, with high-density PE still used for HV extruded cables and PP dominating in many areas of utility and industrial capacitor applications.

Traditionally, inorganic materials are the most pervasive dielectric materials, e.g., silicon dioxide. But nowadays, polymer dielectrics have started to attract more and more attention, which is due to the fact that polymers are easier to process, cheaper, and able to be tailored for specific uses. Polymer dielectrics have many applications such as in capacitive energy storage [7–12], electronics [13–15], photovoltaics [16–18],

encapsulation [19–22] and insulation [23, 24]. The selection and design of a polymer for a given application depend on the requirements specific to that application. For instance, dielectrics for capacitive energy storage are required to meet stringent needs, including low dielectric loss, high dielectric breakdown strength and large dielectric constant [9, 25]. Those for electronics applications are required to have large band gap and large dielectric constant [25–27], while photovoltaic applications require polymers with band gap in a specific range to offer the best match with the solar spectrum [25, 28]. Encapsulation and insulation applications need polymers with low dielectric constant and high resistance to degradation under high electric fields [25, 29]. Needless to say, over and above these broad requirements, additional attributes, such as low defect and impurity concentrations, stability at elevated temperatures, etc., may be important in specific applications or sub-applications.

Required by different applications, dielectric constant (ϵ) and band gap (E_g) are the two key properties. Table 1.1 shows the E_g and ϵ values of several polymers compared with some inoganic materials. In general, polymer dielectrics which are commonly used, such as PE, PP, PMMA, *etc*, have low dielectric constant and relatively large band gap. In different applications, attempts have been made to surpass the current standard polymer dielectrics. For example, in capacitor and electronic applications, efforts are put into development of polymer dielectrics with larger dielectric constant and band gap, while designing polymers with smaller band gap is the focus in the field of photovoltaics.

Table 1.1: Band gap (E_g) and dielectric constant (ϵ) values of several polymers and inorganic materials [1–6].

Materials	E_g (eV)	ϵ	Application
TiO_2	3.5	80	pigments, photocatalysts
SiO_2	9	3.9	gate dielectrics in field effect transistors
HfO_2	5.8	25	gate dielectrics in field effect transistors
Fluorosilicate glass	–	3.2-4.0	circuit interconnect layers
PE	8.8	2.3-2.7	packaging, HV extruded cables
PP	7.0	2.4	packaging, capacitors
PS	4.4	2.5-2.9	packaging, gate dielectrics in organic thin-film transistors
Polyimide	–	2.8-3.2	electronics industry for flexible cables
PVDF	–	12	electromechanical sensors and actuators
Nylon	–	3.2-5	tires, ropes
Butyl rubber	–	2.35	tires, fuel and lubricant additive, sporting equipment
Teflon (PTFE)	–	2.0-2.1	insulators in cables and connectors
PMMA	5.8-6.1	2.5-2.9	gate dielectrics in organic thin-film transistors
PVP	–	3.9	gate dielectrics in organic thin-film transistors
cross linked PVP	–	4.0	gate dielectrics in organic thin-film transistors

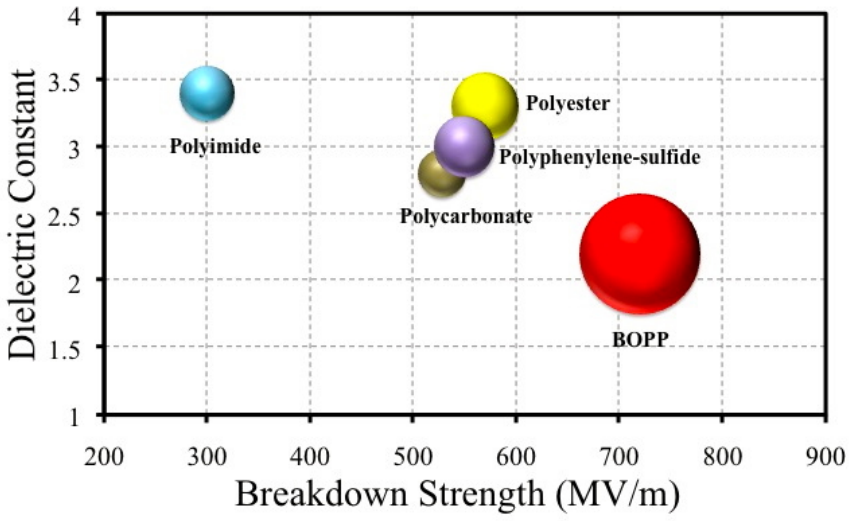


Figure 1.1: Dielectric constant and breakdown strength of common dielectric polymers. Each polymer is represented as a sphere, and the volume of the sphere corresponds to the energy density.

1.2 Polymer dielectrics for capacitive energy storage applications

Polymer dielectrics with high energy density are of great importance for a broad range of applications such as food preservation, nuclear test simulations, electric propulsion of ships, and hybrid electric vehicles [7,25,30,31]. The maximum possible stored energy density of a capacitor is given by $U = 0.5\epsilon_0\epsilon E_b^2$, where ϵ_0 and E_b are, respectively, the permittivity of free space and the dielectric breakdown field.

Compared with ceramic dielectrics, polymer-film capacitors can provide several attractive properties, including high breakdown field, low loss, and graceful failure [25]. Commercial capacitor polymers include biaxially oriented polypropylene (BOPP) [32], polycarbonate (PC) [6], polyester (PET) [6], polyphenylnene-sulfide

(PPS) [6], polyimide [33], *etc.* Figure 1.1 summarizes the dielectric and energy storage properties of such dielectric polymers. Among these polymers, BOPP is the current state-of-the-art polymer dielectric film due to its high breakdown strength and low loss. However, the low dielectric constant (∼2.2) of BOPP limits its energy density [6, 34, 35].

Continuous miniaturization in electrical power systems demands further increases in energy density of dielectric materials since these capacitors contribute significant volume and weight to these systems. Attempts to surpass the properties of BOPP has been mainly focused on poly(vinylidene fluoride) (PVDF) and its copolymers. However, the relatively high remanant polarization significantly limits the energy density in PVDF [10]. Later, it was demonstrated that ferroelectric PVDF polymers can be converted to relaxor ferroelectric polymers with very small remanant polarization by copolymerization. The modified PVDF polymers such as P(VDF-CTFE) and P(VDF-HFP) (CTFE: chlorotrifluoroethylene; HFP: hexafluoropropylene) can have very high energy densities at high electric fields [36, 37]. Unfortunately, these polar polymers with strongly coupled dipoles exhibit pronounced polarization hysteresis at high electric fields which leads to high loss. Another area of active investigation is in developing polymer nanocomposites to increase the dielectric constant and hence the energy density of capacitors [38–40]. The challenge for the nanocomposite approach is achieving a low dielectric loss while raising the dielectric

constant with inorganic nanofillers. Given the limited success of the above mentioned approaches, high energy density polymer dielectrics research has begun to shift towards the development of new classes of dielectric polymers.

1.3 Polymer dielectrics for electronics applications

For many years, inorganic materials have dominated the electronic materials research. In order to achieve comparable device performance at reduced cost, organic electronics have became the subject of intensive research. Thin-film field-effect transistor is one of the fundamental electronic devices and in the 1980s, organic materials began to be applied in such devices. Since then, organic thin-film transistors (OTFT) have gained more and more attention because they can provide several attractive properties, such as inexpensive large area processing, lightweight, and compatibility with flexible plastic substrates.

In OTFT, the organic semiconductor is not the only critical component, it is also very important to incorporate a suitable dielectric layer (gate dielectric). A promising gate dielectric should have some key properties to allow its application in transistors, *e.g.*, high dielectric constant, large band gap, large enough band offsets with respect to semiconductor (minimize carrier injection into its bands), high breakdown field, good immunity against moisture, *etc.* Polymer dielectrics have been considered as promising candidates for gate dielectrics due to the fact that they can exhibit a broad and complementary solubility, processability, and dielectric properties. Several polymers have been used as OTFT gate dielectrics, including

poly(methyl methacrylate) (PMMA) [41], polystyrene (PS) [42], poly(vinyl phenol) (PVP) [43, 44], and *etc.* The band gap and dielectric constant values of these polymers are listed in Table 1.1. Among various polymer gate dielectrics, cross linked PVP is an attractive gate dielectric because of its relatively high dielectric constant ($\sim$4). But it still has its own limitations, which are high curing temperature and high density of moisture-sensitive hydroxyl groups.

In general, polymer gate dielectrics exhibit low capacitances and the corresponding OTFTs operate at relatively high voltages. Thus, more and more efforts have been put into designing new polymer gate dielectrics which can achieve larger gate capacitance and higher breakdown strength.

1.4 Polymer dielectrics for photovoltaics applications

The need of global energy in the next 50 years is expected to double. But fossil fuels are running out and they are responsible for the increased concentration of carbon dioxide in the atmosphere. Photovoltaics (PV), which generates electricity directly from sunlight, is one of the renewable energy technologies and can be a promising solution to the energy crisis. Currently, the active materials used for PV based devices are mainly inorganic materials, such as silicon (Si), gallium arsenide (GaAs), cadmium telluride (CdTe), and *etc.* However, at the current stage, conventional PV technology only contributes to less than 0.1% of total US energy generation due to the fact that it is far too expensive. One of the major obstacles is the large production costs for silicon solar cells. In recent years, organic-based

(OPV) technology has attracted growing interest, which is based on solar cells made of entirely new materials, such as conjugated polymers.

Conjugated polymers based electron donor materials have the immense advantage of facile, chemical tailoring to adjust their properties, such as band gap. They show great promise owing to their low temperature processing, synthetic variability, and the possbility of producing lightweight, flexible, easily manufactured, and inexpensive solar cells. One of the best performing polymeric materials in organic solar cells is poly(3-hexylthiophene) (P3HT) [45–47], achieving an efficiency surpassing 5%. P3HT is also the most commonly used material owing to its high charge carrier mobility, good processability, easy synthesis, and *etc.* However, the band gap of P3HT is around 1.9 eV and its HOMO level is high, which limit the absorbance to below a wavelength of 650 nm and lead to low open circuit voltage. Attempts have been made to design and synthesize new polymeric materials with smaller band gap for PV applications, including the carbazole and benzothiadiazole based polymer (PCDTBT) (band gap is 1.88 eV) [48], poly(3-hexylselenophene) (P3HS), cyclopenta[2,1-b;3,4-b']dithiophene and dithieno[3,2-b:2',3'-d]silole based polymers (PCPDTBT and PSBTBT) (band gap is $\sim$1.5 eV) [50, 51], and *etc.* Although, in the last decade, there is amazing progress in this field, the search for new polymers with suitable band gap value is still quite active and challenging.

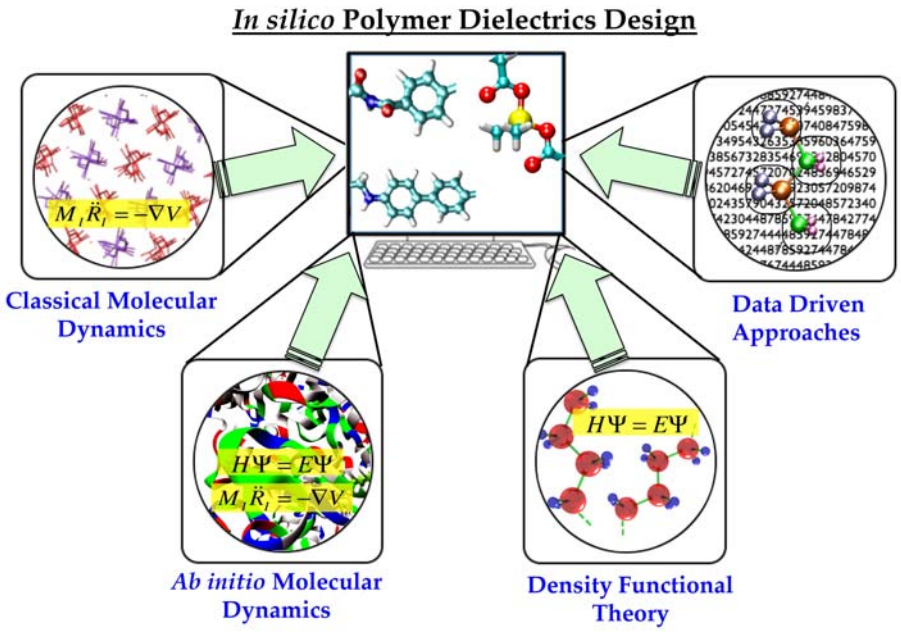


Figure 1.2: Computational strategies for polymer dielectrics design.

1.5 Computational strategies for polymer dielectrics design

As mentioned before, there is an increasing demand for new polymer dielectrics. However, until recently, the discovery of new materials has largely been guided by expensive, serendipitous and lengthy "trial and error" experimental cycles. The *in silico* design of new materials using computational approaches is increasingly being considered as an attractive alternate option [52–61, 87]. Several complementary classes of computational methods appropriate to model polymer dielectrics have reached a high level of sophistication, accuracy and utility (Figure 1.2).

At the most fundamental level, computational quantum mechanics, e.g., density functional theory (DFT), can be used to determine properties of dielectrics at the scale of a crystalline unit cell [62–66]. Such properties include structural and

thermodynamic details, reasonable estimates of the band gap, electronic dielectric constant, ionic dielectric constant, and intrinsic breakdown field [67–73]. In addition, impurity states in the band gap caused by common chemical impurities can be computed [74–77]. Realistic models can also be developed for metal-polymer interfaces in order to predict charge injection characteristics.

Larger scale morphological features of polymers can be accessed practically at the present time only using molecular dynamics (MD) based on empirical interatomic potentials or force fields [78–82]. Such simulations can predict crystal structure, semicrystalline morphology and provide rough estimates of glass transition temperature and dielectric loss, although the latter is presently limited to loss in the GHz range [83–86].

The above methods can be classified as "physics-based", as they are based on quantum mechanics, classical mechanics, and classical electromagnetism. An emerging class of methods, often referred to as "data-driven", use various forms of multivariate analysis on experimental or computational data, based on complex variables with a physical relationship to the properties being predicted [87–95]. Such systems are "trained" on available data and then used to predict properties of interest for polymers for which data are not available. An example of such data-driven approaches is quantitative structure property relationships (QSPR) [96–100], which can predict properties, such as glass transition temperature, melting temperature, *etc.*, for which no fundamental approach is presently available.

1.6 My thesis in a nutshell

The overarching goal of this work is to design new polymer dielectrics using first principles computations and the machine learning approach. As mentioned before, polymer dielectrics have many applications and the selection and design of a polymer for a given application depends on the requirements specific to that application. Dielectric constant and band gap have been identified to be the two common properties required by different applications, *e.g.*, capacitor and electronic applications, which require both high dielectric constant and large band gap. Thus, in this work, the search for new polymer dielectrics is based on these two properties.

Below we briefly describe the organization of the thesis.

In Chapter 2, we briefly touch upon the computational methods used in the present work, which include DFT, density functional perturbation theory (DFPT) and the machine learning approach.

Chapter 3 addresses functionalization of a well-understood polymeric dielectrics, such as polyethylene (PE) and polypropylene (PP), to enhance its dielectric response. While the discovery of completely new classes of polymers for capacitor dielectric applications is exciting in its own right and may prove to be transformative, the evolutionary strategy of enhancing the properties of existing standard materials may be a reasonable risk-mitigation strategy. In this part of the work, we study functionalized polyethylene (PE) and identify how the functional groups, *e.g.*, -OH, -NH_2, -SH, *etc*, will affect the dielectric properties of PE.

Chapter 4 and 5 discusses approaches to the discovery of entirely new classes of polymer dielectrics, both organic and organometallic. Strategies discussed include exploration of large chemical spaces and efficient computation of some relevant properties. The screening of systems considered in Chapter 4 and 5 is based on dielectric constant and band gap. Specifically, the high energy density capacitors applications require the polymer dielectrics to have high dielectric constant and large band gap. After the first-level screening for these properties, promising candidates can be subjected to more detailed analysis.

The role data-driven approaches, in particular, the emerging area of "machine learning", is surveyed in Chapter 6 and 7. The materials discovery process can be significantly expedited and simplified if we can learn effectively from past knowledge. This part of the work involve the development of a similarity-based machine learning approach that can help predict the band gap and dielectric constant of a new polymer in a minuscule fraction of the time necessitated by a typical DFT computation. The machine learning approach is validated and used to explore a larger scale polymer chemical space, both for organic and organometallic polymers. Promising candicates are then identified from a larger pool of polymer systems.

In the last chapter, we provide a summary of this work, and a broad outlook for future directions, which include detailed studies of promising systems and possible approaches to expand the polymer search space.

Chapter 2

Computational methods

2.1 Introduction

In the present work, density functional theory (DFT), density functional perturbation theory (DFPT), and the machine learning approach have been employed to computationally design new polymer dielectrics with both high dielectric constant and large band gap. The goal of the current chapter is to provide a brief overview of these computational methods. As conventional DFT is well established and documented in the literature, a comprehensive and detailed description is avoided here. A new method based on polymer single chain has been developed to rapidly estimate the dielectric constant of polymers. Aided with this single-chain approach, a high throughput method, which allows us to explore the polymer chemical space in a relatively short period of time, has been established. A detailed description of this method is provided. The use of machine learning techniques in materials science is

considered as a new rising field. A brief introduction to machine learning and its applications in materials science are also provided in this chapter.

2.2 Density functional theory (DFT)

Among all modern electronic structure methods, density functional theory (DFT) offers the best tradeoff between computational cost and accuracy. DFT [62,63] based quantum mechanical electronic structure methods have reached a state that allows atomic level interactions in diverse chemical environments to be studied accurately. The only required inputs for such quantum mechanical calculation are the electronic and ionic charges and masses, which make it parameter-free and therefore do not rely on any experimental input. Such method is also referred to as "first-principles" or "*ab initio*" techniques.

The pioneering work of Hohenberg, Kohn and Sham [62, 63] provides a practical, computationally-tractable, and enormously successful strategy to solve many-electron system in which the problem of a real system with interacting electrons is solved by mapping it onto an auxiliary system of noninteracting electrons moving a an external effective potential. Within Kohn-Sham DFT, the following eigenvalue equation (in atomic units) is solved:

$$[-\bigtriangledown^2 + V_{eff}(r)]\Psi_i^{KS}(r) = \epsilon_i^{KS}\Psi_i^{KS}(r) \tag{2.1}$$

where $-\bigtriangledown^2$ represents the electronic kinetic energy, and $V_{eff}(r)$ represents the effective potential energy seen by an electron. $V_{eff}(r)$ contains all the electron-electron and electron-nuclear interactions, as well as the potential caused by an external

electric field. In practice, the quantum mechanical part of the electron-electron interaction is approximated using local functionals within the local density approximation (LDA), semilocal functionals within the generalized gradient approximation (GGA) [101], or nonlocal hybrid functionals. ϵ_i^{KS} and $\Psi_i^{KS}(r)$ are the energy eigenvalues and wavefunctions of the Kohn-Sham orbitals. For a given set of atomic positions, the above equation is solved self-consistently to result in converged charge densities (obtained from the wave functions of the occupied states), total energies (obtained from the wave functions and eigenenergies of the occupied states) and atomic forces (obtained from the first derivatives of the total energy with respect to the position of any given atom). The atomic coordinates are optimized by requiring that the total energy of the system is a minimum, and the forces on each atom are close to zero.

Once the geometry is optimized, several other properties can be predicted. DFT is capable of predicting structural properties, vibrational or phonon frequencies, elastic constants, work functions, surface energies, and *etc.*, with high accuracy as summarized elsewhere [102]. In this thesis, band gap and dielectric constant are the two key properties that need to be computed. Conventional DFT is plagued by the well-known band gap underestimation problem due to the spurious electron self-interactions within the GGA and LDA functionals. DFT band gaps are usually underestimated relative to experimental values by up to 50%. Techniques to handle such deficiencies are currently available and include the use of hybrid functionals

that are rising in popularity [103]. The calculation of the other important property, dielectric constant, will be provided in the following section.

2.3 Density functional perturbation theory (DFPT)

Many physical properties, such as polarisabilities, phonons, Raman intensities, *etc.*, depend on a system response to some form of perturbation. DFPT is a particularly powerful and flexible theoretical technique which can be used to calculate the electronic linear response to an external perturbation and thus allows calculation of such properties. Although calculating system responses to external perturbations is also possible within DFT, by taking into account of the perturbing potential in the Hamiltonian, such methods are somewhat crude and asethetically unappealing, since they involve obtaining the system response through a series of single-point energy calculations carried out at varying strengths of the external perturbation. More fundamentally, the applications of such techniques are sometimes restricted. For example, they cannot readily be used to calculate the response of a three-dimensional periodic systems to electric field perturbations. The DFPT approach, on the other hand, provides an elegant, efficient and mathematically more appealing alternative to calculate linear response of a system.

There are two main formalisms of the modern DFPT which are due to Baroni [104] and Gonze [105, 106]. Although the two approaches are essentially equivalent, there are differences in the implementations which may lead to one method being preferable to the other. While the former formalism is based on Green's function

methods, the latter relies on a perturbative expansion of the Kohn-Sham energy functional.

DFPT is now ripe enough to allow a systematic application to systems and materials of increasing complexity. Most promising application fields of DFPT include prediction of thermal dependence of different materials properties using the quasi-harmonic approximation, study of structural stability of materials at extreme pressure conditions, prediction of dielectric response and vibrational frequencies, prediction of superconductive properties via the calculation of electron-phonon coupling coefficients [65]. In the present work, DFPT has been used to calculate the dielectric constant of various polymer systems.

The total dielectric constant derives its contribution from electronic and ionic part, both of which can be calculated using DFPT. The behavior of the electronic contribution to the dielectric constant may be understood in terms of the electronic structures of systems. Indeed, the electronic contribution to the polarizability can be written in terms of a sum over electronic transitions from the valence band to the conduction band manifolds. Thus, the electronic dielectric constant is correlated with the band gap of a system.

The ionic contribution to the dielectric constant, on the other hand, is not directly correlated with the band gap, as this contribution is purely controlled by the infrared (IR) active zone center phonon modes (*i.e.*, the modes that display a time-varying dipole moment) [129,130]. The extent to which each IR-active phonon mode contributes to the dielectric tensor is determined by the frequency of the mode

(smaller the frequency larger the dielectric constant) and the IR intensity (I_i^{IR}) of the corresponding zone center mode (larger the IR intensity, larger the dielectric constant), which is given by [131, 132]

$$I_i^{IR} \propto \sum_{\alpha} \left| \sum_{k,\beta} Z^*_{k,\alpha\beta} X_i(k,\beta) \right|^2$$

where i labels the modes, k labels atoms, and α and β are the cartesian coordinates. ${Z^*}_{k,\alpha\beta}$ and $X_i(k,\beta)$ represent appropriate components of the Born effective charge tensor and the phonon mode eigenvector, respectively. For IR-inactive phonon modes, this intensity vanishes.

The ionic dielectric constant can then be defined as:

$$\epsilon_i = \frac{4\pi e^2}{V} \sum_i \frac{I_i}{{\omega_i}^2} \tag{2.2}$$

where ω_i is the frequency of the ith IR-active phonon normal mode, and V is the volume of the system. The contribution of each mode to the ionic dielectric constant can be calculated using the expression $\frac{I_i}{{\omega_i}^2}$.

DFPT has been used to compute the optical (the high frequency, or the electronic) and static (the low frequency, or the ionic plus electronic, total) dielectric constant of different materials. As a validation, Figure 2.1 shows the correlation between DFPT predicted optical and static dielectric constant and experimental values for several classes of systems. It can be seen that the calculated results are in good agreement with experiments.

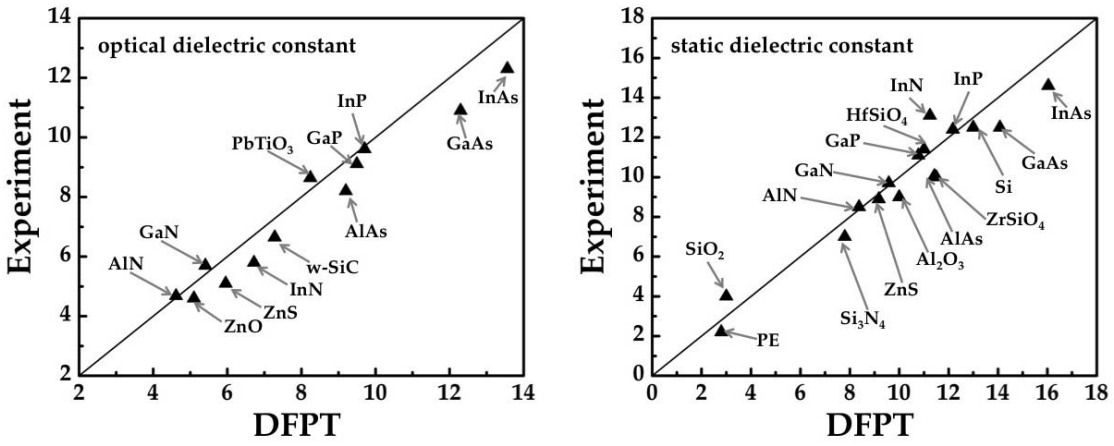


Figure 2.1: Comparison between DFPT predictions and experimental values for optical and static dielectric constant. Data are from various sources including refs. [1, 107, 108, 154]

2.4 High-throughput method for polymer dielectric constant prediction

The investigation of polymers is confounded by the necessity of a knowledge of the appropriate crystal structure to be used in DFT computations. Moreover, even if such information is available (or can be "guessed"), each computation may be time consuming. In order to reduce the time involved in each such computation, and to obtain rapid estimates of the dielectric constant and band gap values in the absence of reliable crystallographic information, we have developed a method that is based on purely single chain computations.

As shown in Figure 2.2(a), we consider an isolated infinite chain of a polymer placed in a supercell volume V_{tot}, and use density functional perturbation theory (DFPT) to compute its dielectric constant. We note that the dielectric constant

calculated from DFPT for such a supercell includes the contributions from the polymer as well as from the vacuum region of the supercell. Treating the supercell as a vacuum-polymer composite, effective medium theory may then be used to estimate the dielectric constant of just the polymer [68].

According to the Maxwell-Garnett equation, the principal components of dielectric constant of a vacuum-filler composite (ϵ_{ii}) containing a volume fraction δ of polarizable fillers (polymer chain in this case) with dielectric constant ($\epsilon_{ii}^{polymer}$) can be written as [109]:

$$\frac{\epsilon_{ii} - 1}{1 + (\epsilon_{ii} - 1)P_i} = \delta \frac{\epsilon_{ii}^{polymer} - 1}{1 + (\epsilon_{ii}^{polymer} - 1)P_i} \tag{2.3}$$

Here i represents the cartesian axes x, y or z, and P_i is a geometry-dependent depolarizing factor [110–112]. In our case, assuming that the polymer chain is along the z direction, $P_z = 0$, and $P_x = P_y = 0.5$. This leads to the following formula for the axial and off-axis components of the dielectric constant [113, 114]:

$$\epsilon_{zz} - 1 = \delta(\epsilon_{zz}^{polymer} - 1) \tag{2.4}$$

$$\frac{\epsilon_{xx} - 1}{\epsilon_{xx} + 1} = \delta \frac{\epsilon_{xx}^{polymer} - 1}{\epsilon_{xx}^{polymer} + 1} \tag{2.5}$$

In these equations, δ ($= V_{polymer}/V_{tot}$) is the volume fraction of the polymer in the supercell as shown in Figure 2.2(a). In order to use this method to estimate the dielectric constant of just the polymer, *i.e.*, $\epsilon_{ii}^{polymer}$, the volume occupied by the polymer chain in the supercell is needed. Here, a procedure based on charge

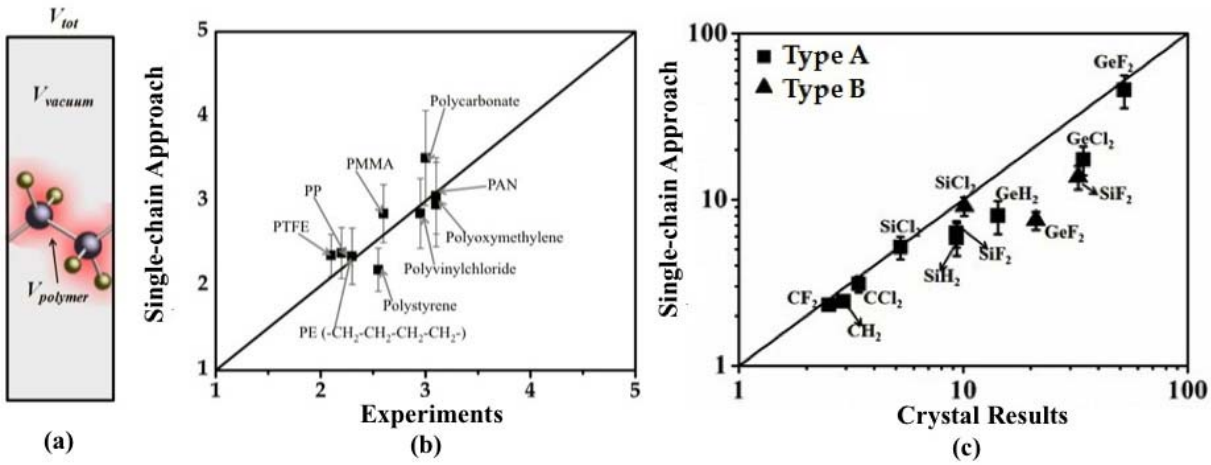


Figure 2.2: (a) Single-chain model represented as a polymer-vacumm composite. (b) Comparison of dielectric constant results from single-chain approach and experiments for common polymers. (c) Comparison of the single-chain based results vs the full crystal results for several XY_2 polymers in different types of crystal structure. Polymers SiF_2, $SiCl_2$, and GeF_2 can be stablized in two different types of crystal structures which are represented as Type A and Type B.

density cutoffs is used to estimate the volume. If the electronic charge density is larger than a cutoff value at a particular location, then this location is deemed occupied by the polymer. In order to determine the charge density cutoff value, we have considered several polymers for which experimental volumes (or densities) are available to determine the charge density cutoff value that would result in the experimental density. The polymers used for this analysis include polyethylene, polypropylene, polyacetylene, polythiophene, polypyrole, polydimethylsiloxane, *etc.* In general, we find that the charge density cutoff needs to be in the 0.003 electron/Å^3 to 0.007 electron/Å^3 range in order to reproduce the experimental densities. We note that this range of charge density cutoffs translates to a range of volumes, $V_{polymer}$, and hence to error bars in the calculated dielectric constants using Eq. 2.4 and 3.1.

To validate the single-chain approach, we choose a few common polymers with known dielectric constant and compare the results from this method with those from the experiments [115]. As shown in Figure 2.2(b), the estimated dielectric constants from our single-chain approach are in relatively good agreement with the experimental values if considering the associated error bars in our method. But the dielectric constants of common polymers are only in the range of around 2 to 4. In order to validate that the single-chain approach can also make a good prediction for the systems with high dielectric constant, we considered homopolymers with a XY_2 repeat unit, where X = C, Si or Ge and Y = H, F, Cl [67]. The dielectric constants of these polymers span a large range from around 2 to 45. We compare the dielectric constant results of our single-chain approach with those from the three-dimensional crystal based calculations as shown in Figure 2.2(c). Overall, the agreement between the two methods is good, given that interchain interactions are completely neglected in the single-chain computations. The complete neglect of interchain interactions in the single-chain approach may account for the consistent underestimation of the dielectric constant with respect to the crystal results (as interchain interaction tend to "soften" the intrachain phonon modes, and, consequently, increase the dielectric constant values). Although there are some discrepancies, the correct trends are mostly captured by the single-chain approach. Hence, we believe that, this approach can offer a practical strategy for rapidly screening candidate polymer chain structures within high throughput chemical space exploration efforts.

2.5 Accelerating materials discovery using machine learning approach

Owing to the staggering compositional and configurational degrees of freedom possible in materials, it is fair to assume that the chemical space of even a restricted subclass of materials is far from being exhausted, and an enormous number of new materials with useful properties are yet to be discovered. Given this formidable chemical landscape, a fundamental bottleneck to an efficient materials discovery process is the lack of suitable methods to rapidly and accurately predict the properties of a vast array of new yet-to-be-synthesized materials. The standard approaches adopted thus far involve either expensive and lengthy synthesis-testing experimental cycles, or laborious and time-intensive computations, performed on a case-by-case manner. In the present work, a radically different paradigm, namely, machine learning, has been used for materials property predictions.

2.5.1 A brief introduction to machine learning

Machine learning, also known as data mining, or statistical learning, deals with making predictions from data. It is a field at the intersection of computer science and modern mathematical analysis. Machine learning is a topic central to cognitive/decision/game theory [116], pattern recognition [117–119], and event forecasting [120, 121]. The overall goal of machine learning is extracting information from a large dataset and transforming it into understandable structure, thus leading to knowledge discovery. There are different learning paradigms in machine learning, mainly *supervised learning* and *unsupervised learning* (Figure 2.3). In supervised

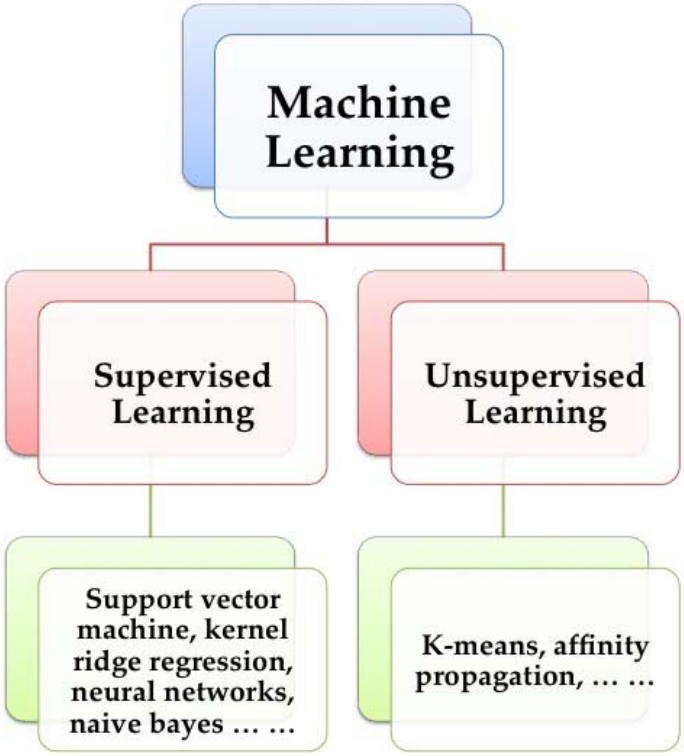


Figure 2.3: Machine learning hierarchy.

learning, the aim is to learn a mapping from the input to an output whose correct values are provided by a supervisor. While in unsupervised learning, there is no such supervisor and the input data is the only thing one has. The aim is to find the regularities in the input.

Both regression and classification are supervised learning problems where there is an input, X, an output, Y, and the task is to learn the mapping from the input to the output. The approach is that one can assume a model defined up to a set of parameters:

$$Y = f(X \mid \theta) \tag{2.6}$$

Where f(·) is the model and θ are its parameters. Y is a number in regression problem, while in the case of classification, is a class code, *e.g.*, 0 or 1. f(·) is the regression function, or in classification, it is the discriminant function that separates the instances of different classes. The parameters, θ, are optimized so that the approximation error is minimized, which means that the estimates are as close as possible to the correct values given in the training data. The training data consists of a set of training examples. Certain supervised learning algorithms analyze the training data and produce g(·), which can be used for mapping new examples, which are called test data. There are different supervised algorithms available, such as support vector machines (SVM), kernel ridge regression (KRR), artificial neural networks (ANN), naive bayes, random forests, *etc.*. SVM, which constructs a hyperplane or set of hyperplanes in a high dimensional space, can be used for classification and regression analysis. KRR is an nonlinear regression algorithm, which is applicable in many different fields ranging from optical character recognition to business forecasting. KRR has proven to be better than many other predictors and it is the choice of regression algorithm in this work. The details of KRR will be provided in Chapter 6. Inspired by animals' central nervous systems, ANN are usually presented as systems of interconnected "neurons" that can compute values from inputs by feeding information through the network. Naive bayes is a probabilistic classifier based on applying Bayes theorem with strong independence assumptions, and the random forests method can be used for both classification and regression.

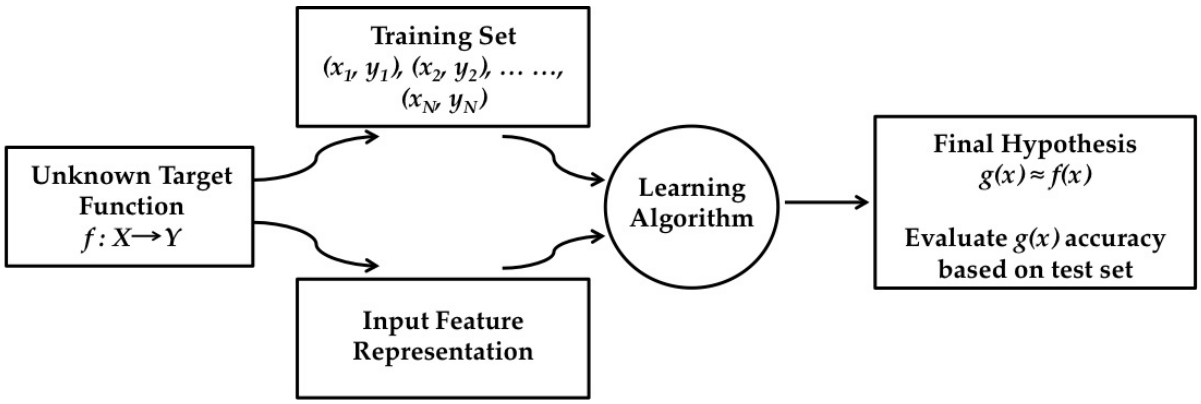


Figure 2.4: Basic supervised learning setup.

In order to solve a given problem of supervised learning, the following steps need to be performed (as shown in Figure 2.4):

- Choose the appropriate training set. The training set needs to be representative of the given problem. Thus, a set of input objects and the corresponding outputs are gathered.

- Determine the input feature representation. The accuracy of the learned function depends strongly on how the input objects are represented. Usually, the input object is represented by a feature vector, which includes a number of features. The number of features should not be too large and contain enough information to accurately predict the output. But for most cases, it is not clear which features are important and should be included in the feature vector. It is also conceivable that some of the feature vector components may be hurtful rather than helpful in predictions. Thus, dimensionality or feature reduction algorithms can be applied to choose the appropriate number of features used

in learning, *e.g.*, principal component analysis, random forests, and *etc.*. Random forests can be used to rank the importance of features in a regression or classification problem.

- Choose learning algorithm and apply it on the gathered training set.
- Evaluate the accuracy of the learned function based on test set.

In unsupervised learning, the goal is to find a structure to the input space such that certain patterns occur more often than others. One method of unsupervised learning is clustering where the aim is to find clusters or groupings of input. Different algorithms are available for clustering, such as k-means, affinity propagation, and *etc.*. The k-means algorithm clusters data by trying to separate samples in n groups of equal variance, minimizing a criteria known as the 'inertia' of the groups. While the affinity propagation creates clusters by sending messages between pairs of samples until convergence.

2.5.2 Machine learning applications

Machine learning has started to penetrate into different areas, from our daily life, *e.g.*, emails, online shopping, banking, weather forecasting, to scientific world, such as robot control and accelerating science development. In the case of retail, for example, a Walmart supermarket chain, one application of machine learning is basket analysis. It aims to find associations between products bought by customers. If people who buy X typically also buy Y, then a custom can be a potential Y buyer if he or she buys X and does not buy Y. Once the supermarket finds such

customs, they can be targeted for cross-selling. Such association rule is also applied in Amazon and Netflix recommendation system.

In credit scoring, the bank calculates the risk given the amount of credit the information about the customer. The information about the customer includes data that are relevant to the customers' financial capacity, such as income, savings, profession, past financial history, and so forth. From such data of particular applications, the machine learning is used to fit a model to the past data to be able to calculate the risk for a new application and then decides to accept or refuse it accordingly.

2.5.3 Machine learning in Materials Science

In Materials Science, when confronted with a new material, it would be advantageous if its properties may be predicted using past knowledge pertaining to other similar known materials, rather than by resorting to explicit new experiments or laborious calculations. This 'training of our intuition' requires a critical amount of prior information or results. More and more data are generated experimentally and computationally in Materials Science. Applying machine learning to learn from these data becomes a new area of interest. In the present effort, quantum mechanical first principles computational methods based on density functional theory (DFT) are used to generate the initial data set (i.e., the 'training set'). Machine learning techniques can then be used to establish a mapping between a suitable representation of a material (i.e., its fingerprint or its feature) and any or all of its properties using known historic, or intentionally generated data. Once the profile⇔property

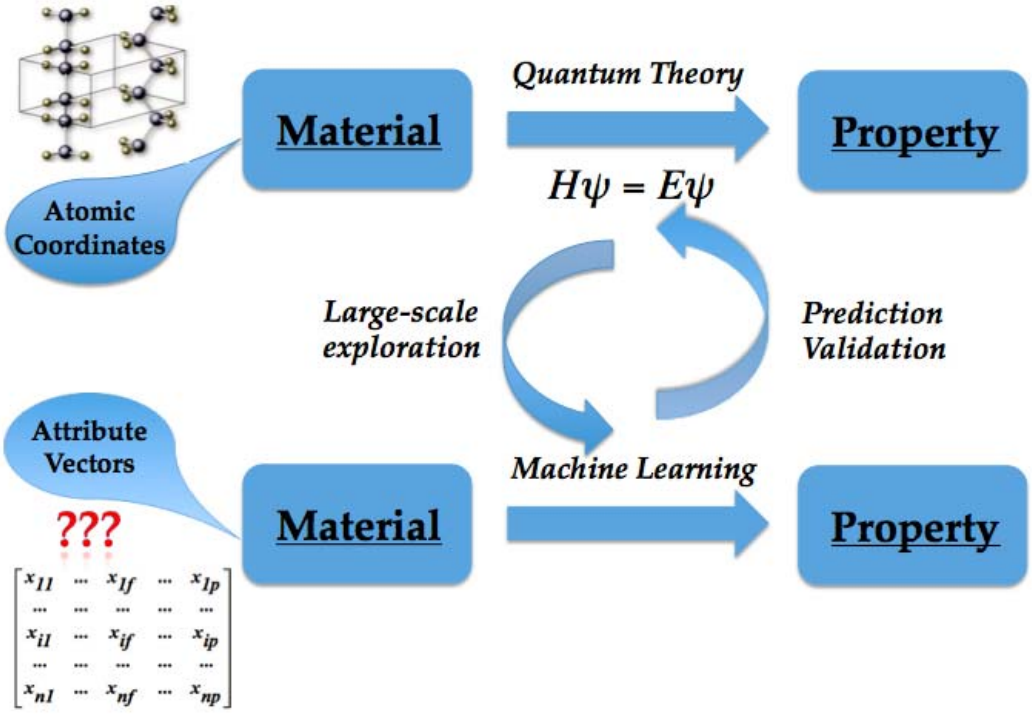


Figure 2.5: A perspective on the role machine learning methods can play in accelerating materials discovery.

mapping has been established, the properties of a vast number of new materials within the same subclass may then be directly predicted (and correlations between properties may be unearthed) at negligible computational cost, thereby completely by-passing the conventional laborious approaches towards material property determination and understanding alluded to above.

Applying machine learning in Materials Science has started to attract more and more attentions. Machine learning methods were used in the work of Bucholz et al. [122] to generate a predictive model which allows efficient high throughput screening of ceramic materials. They developed a decision tree based model to estimate the friction coefficients of a wide range of materials through the combination of principal

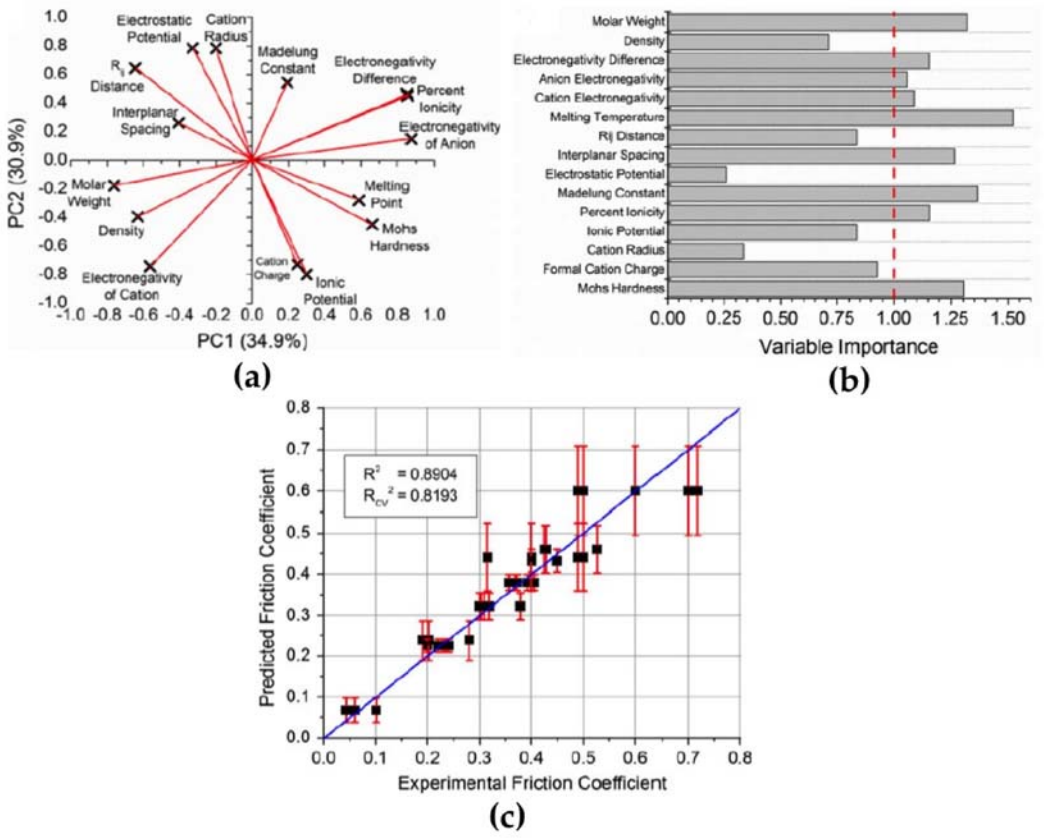


Figure 2.6: (a) Principal component analysis loadings plot for PC1 versus PC2. The plot indicates which materials properties contained in the feature vector are influential and how the properties are related to each other within the plane of the first two PCs. (b) Feature importance identified by variable importance in projection (VIP) method. (c) Predicted versus experimental friction coefficient from recursive partitioning. Reprint with permission from ref. [122].

component analysis and recursive partitioning. The materials were represented by a feature vector containing parameters ranging from crystal structure to electronic structure. The relative influence of these parameters were assessed quantitatively. This data driven model was shown to have a high degree of accuracy (Figure 2.6).

Based on nuclear charges and atomic positions, Rupp et al. [93] introduced a nonlinear statistical regression model to predict atomization energies of a diverse

set of organic molecules. Over more than seven thousand organic molecules were used in cross validation, which yielded a mean absolute error of ∼10 kcal/mol. Zakutayev et al. [124] presented an inverse design approach to accelerate discovery of missing materials. There are 29 missing out of the 45 possible V-IX-IV compounds. Eight new stable compounds were revealed by theoretical screening of their thermodynamica stability.

Hautier et al. [123] presented a probabilistic model to estimate the likelihood for ionic species to substitute for each other while still retaining the crystal structure. This model can be used to suggest novel compounds and their crystal structures, and the predictive accuracy is demonstrated using cross validation on quaternary ionic compounds. Figure 2.7 shows the procedure to predict new compounds formed by the a, b, c, and d species using the substitutional probabilistic model adopted in their work .

Several machine learning techniques, including ordinary learst squares, sparse partial least squares, and elastic net/least absolute shrinkage and selection operator regression methods coupled to rough set and principal component analysis methods, were used to predict the band gaps of over 200 new chalcopyrite compounds in the work of Dey et al. [125]. The regression was based on a model using the features that are most related to band gap. Feature ranking algorithms were then used to identify the most relevant features for band gap prediction.

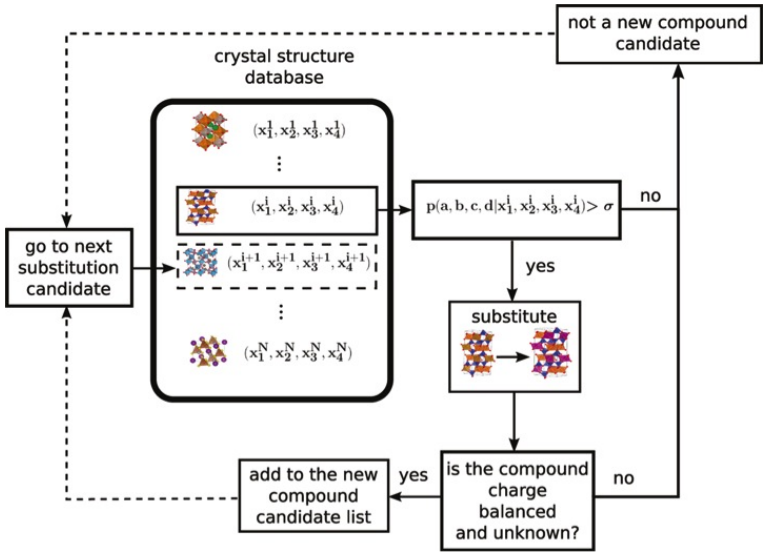


Figure 2.7: Procedure to predict new compounds formed by the a, b, c, and d species using the substitutional probabilistic model. Reprint with permission from ref. [123].

Although applying machine learning in Materials Science is a blooming field, it is still new and there are questions and problems that need to be answered and solved.

- How to choose the features or fingerprints of a materials that may govern specific properties? This is probably the first and the most important problem in materials informatics. The features should reflect the analysis task. In many cases, one may have to extract more features than actually is necessary. A reduction in the number of features is then required to find the key features for predicting certain property.

- How to choose the appropriate machine learning techniques for specific applications? There are different regression and classification algorithms available.

Once the features are determined, the next important step is to choose the appropriate algorithm for good predictive power.

- How to predict for outliers? In materials discovery, outliers are usually the interesting and important cases. But in most cases, machine learning techniques are for extrapolation and cannot be used for outlier prediction.

Chapter 3

Modification of existing polymer dielectrics

3.1 Introduction

A good place to start in applying computational methods to design new materials is to understand or identify the factors that lead to surprising properties when a well-studied material is modified creatively. Polypropylene (PP), in its biaxially oriented form, is the most common polymer dielectric for high energy density capacitors as a result of its high dielectric breakdown strength, low dielectric loss, and good clearing characteristics; however, its dielectric constant is only 2.2 [6, 34, 35]. Increasing the dielectric constant while maintaining or improving other properties, such as operating temperature range, is highly desirable. One possible way to increase the dielectric constant of PP is by adding polar functional groups to the side chain. In this chapter, a comprehensive study of functional groups modified polyethylene (PE), which is chemically similar to PP, is provided. In particular,

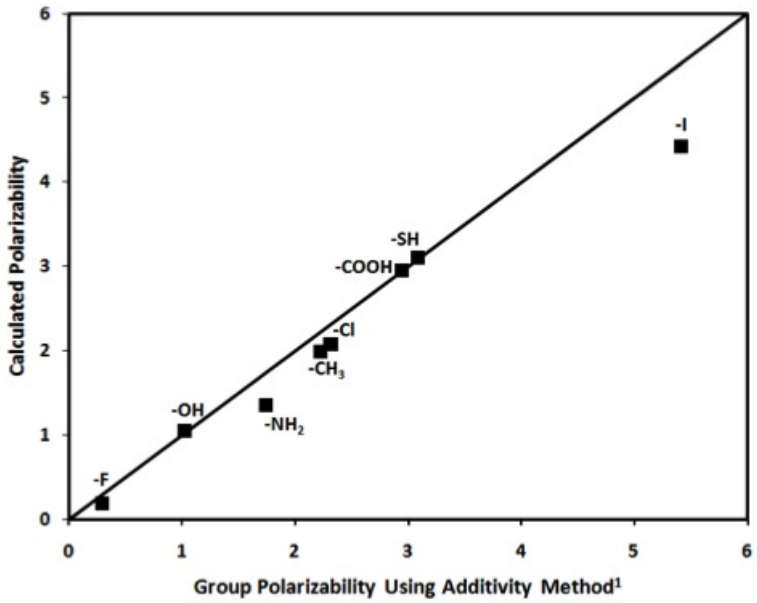


Figure 3.1: Polarizabilities of -F, -Cl, -I, -OH, -NH_2, -CH_3, -COOH, -SH calculated from DFT and additivity method.

hydroxyl groups modified PE is shown to be promising and have been studied in detail. Our results reveal that the trapped water molecules will inevitably accompany -OH incorporation (due to hydrogen bonding) and that both the -OH groups and water molecules will contribute cooperatively to increase the dielectric constant.

3.2 Functional groups modification

In order to examine how different functional groups will affect the dielectric constant of PP, we have developed an approach. Let's consider a polyethylene (PE) chain which is chemically similar to PP, and attach one functional group to the PE backbone. In Clausius-Mossetti equation:

$$\frac{\epsilon - 1}{\epsilon + 2} = \frac{4\pi}{3}\frac{\alpha}{V} \tag{3.1}$$

α is the polarizability of the functional groups and can be determined accurately and quickly using DFT calculations. V is the volume of each functional group and in our approach, it is estimated using either tabulated molar volumes or experimental densities.

The quantity polarizability per volume can be used to qualify how much each functional group can contribute to the increase of PE dielectric constant. In order to validate our approach, we have compared our results for several functional groups with their group polarizability computed using additivity method. Figure 3.1 shows that, overall, the DFT caculated polarizabilities are in good agreement with the values computed from additivity method.

Once the approach has been validated, it was used to calculate the polarizability per volume values for a number of different functional groups. Figure 3.2 shows the structures of different functional groups we have considered, and the polarizability per volume values of them are presented in Figure 3.3. The plot like Figure 3.3 can provide us with an idea on how much each functional group can contribute to the increase of dielectric constant of PE qualitatively. The nonpolar functional groups, e.g., -CH_3, -C_6H_5, seem to have little impact on the dielectric constant increase. While functional groups -OH, -NCS, -ONO, *etc.*, appear to be promising candidates.

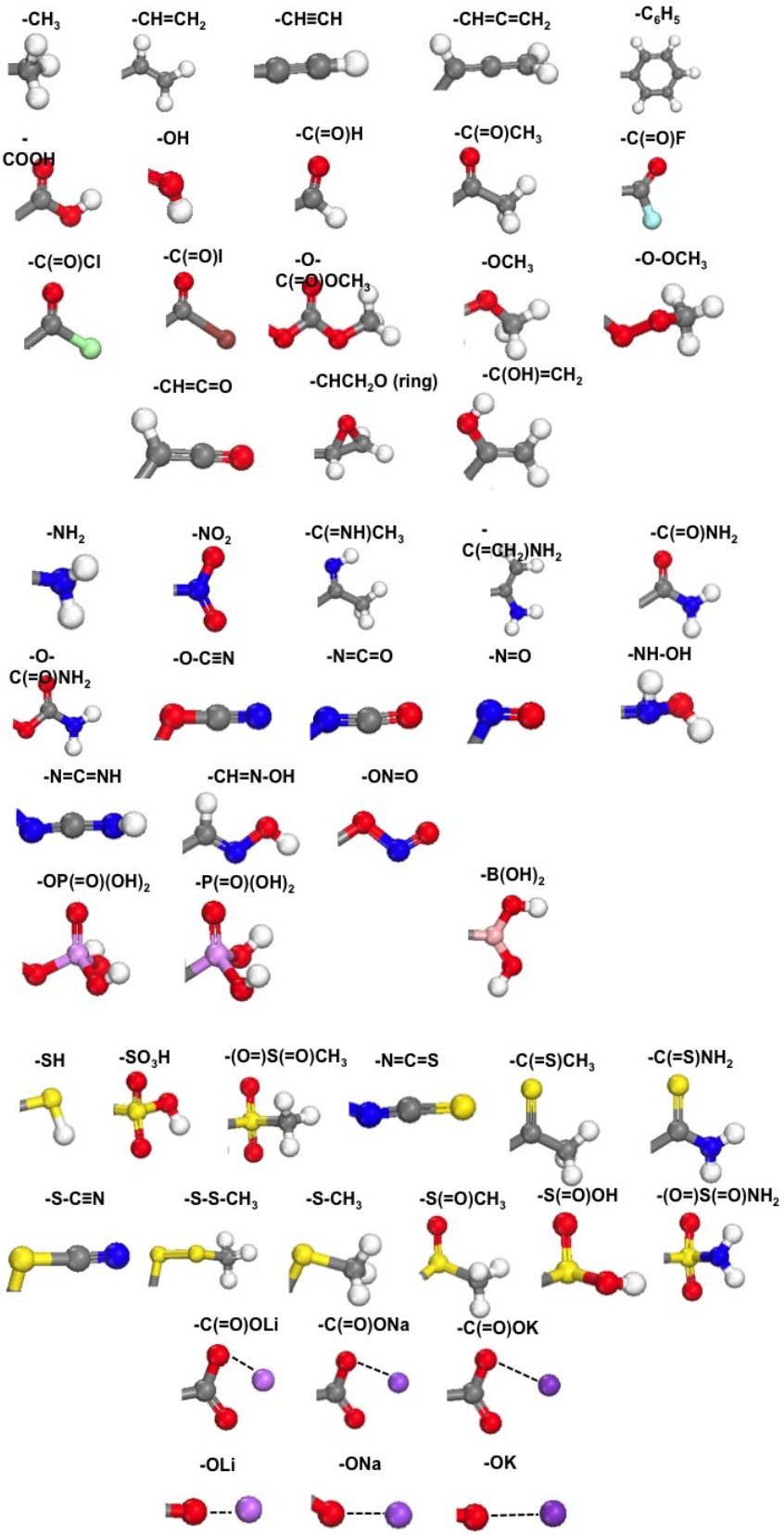


Figure 3.2: Structures of different functional groups.

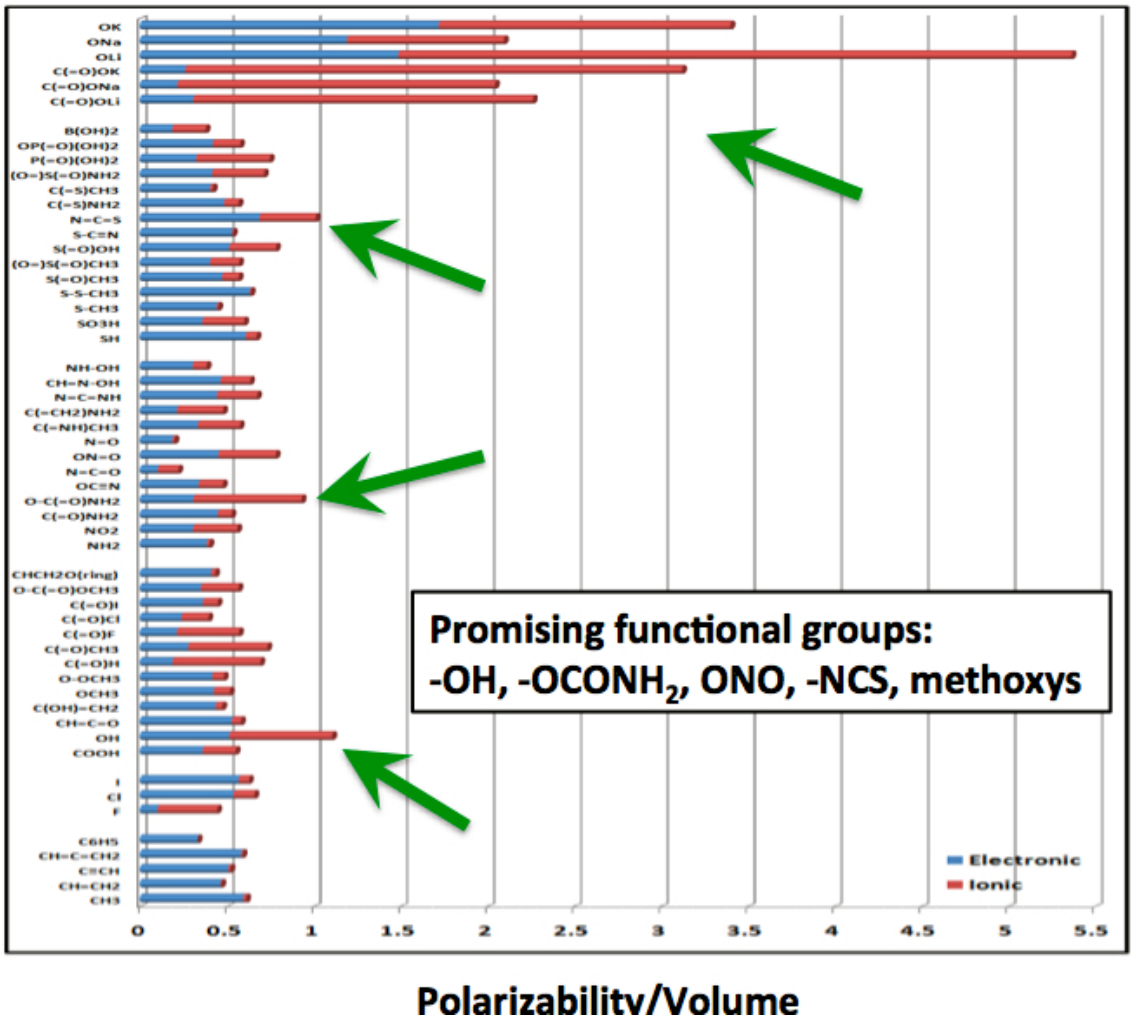


Figure 3.3: polarizability per volume of different functional groups.

3.3 Hydroxyl functionalized PE

Consistent with what we have predicted, Chung has developed an -OH functionalized PP which doubles the dielectric constant and maintains relatively low loss with only 4.2 mole% -OH (Figure 3.4) [126–128]. The experimental data also imply that the PP-OH contains roughly 0.5 water molecule per -OH moiety on the chain and that the -OH groups are hydrogen-bonded in pairs.

DFT computations can investigate the role played by -OH functional groups and the trapped moisture. Since modeling PP or any polyolefin, inclusive of morphological details, is cumbersome using first principles methods, a short $C_{11}H_{24}$ hydrocarbon was used to model the unfunctionalized polyolefin, and $C_{11}H_{22}(OH)_2$ (with one -OH unit in each of its two ends) to model the -OH functionalized polyolefin. The tendency for hydrogen bond formation was probed using two such hydrocarbon chains within our unit cell, and varying numbers of water molecules were introduced close to the chain ends to interrogate the tendency for water trapping. Similar computations with an even smaller hydrocarbon molecule, namely, ethane (*i.e.*, C_2H_6) were also performed to more exhaustively explore various configurations. These latter computations yielded hydrogen bonding and water bonding results that were similar to those obtained using the longer chains. The optimized geometries obtained using the longer 11 C atom hydrocarbon systems were used to determine their dielectric constant using density functional perturbation theory (DFPT) followed by extrapolation of these results to correspond to the volume actually occupied in real polyolefin systems using the new single-chain approach.

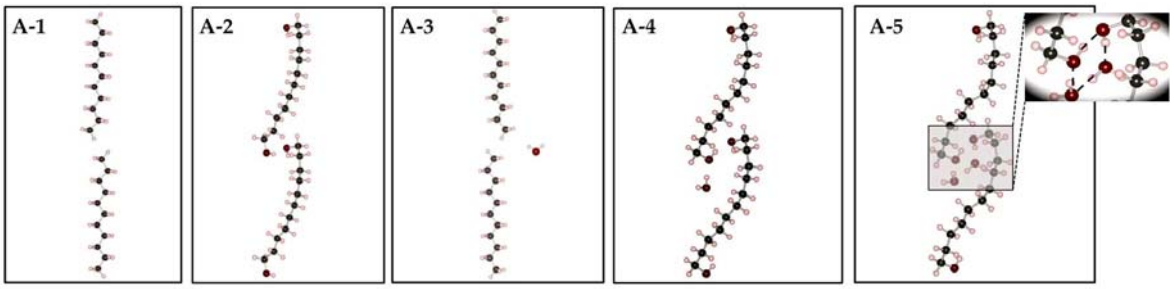


Figure 3.4: (Color online) Optimized structures of complexes A-1, A-2, A-3, A-4, and A-5. Black, white and red spheres represent C, H and O atoms, respectively. The inset shows a typical hydrogen bonded ring.

The optimized structures of complexes involving $C_{11}H_{24}$, $C_{11}H_{22}(OH)_2$, and water molecules are shown in Figure 3.4. The complex A-1 is the system with two $C_{11}H_{24}$ chains in a periodic supercell attempting to interact with each other at the chain ends, and, likewise, the complex A-2 contains two $C_{11}H_{22}(OH)_2$ chains. In complex A-3, where -OH species are absent, complex A-1 interacts with one water molecule, while in complexes A-4 and A-5, one and two water molecules, respectively, interact with the complex A-2, in the vicinity of the -OH species. The interaction between chains, in the presence and absence of -OH species, and the interaction between the chains and water molecules, are quantified by calculating the energy of interaction (E_{int}), as defined in Table 3.1. When there is no -OH species, the chain-chain interaction energy is 0.01 eV per CH_3 group, consistent with the weak van der Waals interaction between chains. With the incorporation of -OH species, the chain-chain interaction energy increases to 0.28 eV per -OH group. As can be seen in Figure 3.4, the relative position and geometry of the two chains get altered dramatically (with respect to the case when there is no -OH species) to accommodate the interchain hydrogen bonding interaction mediated by the -OH groups. These

results show that introduction of hydroxyl groups will increase the interchain interactions significantly due to hydrogen bonding. In order to probe the tendency for water molecules to bind to the chains (in the absence and presence of -OH species), we next explore the larger A-3, A-4, and A-5 complexes with one and two H_2O molecules. When there is one water molecule in the complex, the energy of interaction between the chain and water is 0.09 eV when there is no -OH species present (complex A-3). Presence of the -OH species enhances the chain-water interaction to 0.41 eV per water molecule when one water molecule is present (complex A-4), and to 0.59 eV per water molecule when two water molecules are present (complex A-5). The optimized structure of complex A-5 shows a typical hydrogen bonding ring involving the two water molecules and the -OH species. The strong interaction between H_2O and -OH functionalized polyolefin chains indicates that the trapping of water molecules due to the -OH groups is inevitable.

To further validate the results obtained using our long chain model, a parallel study with a small hydrocarbon system, consisting of C_2H_6, C_2H_5OH, and water molecules was also performed. A large number of configurations and complexes involving these species were considered, and the optimized structures of the most stable (and the most relevant) of these complexes are shown in Figure 3.5. Complexes B-1, B-2 and B-3 are analogous to complexes A-1, A-2 and A-3. Complex B-1 involves the interaction of two hydrocarbons (*i.e.*, C_2H_6), and complex B-2 involves two C_2H_5OH molecules. Likewise, complex B-3 represents the interaction between water and the hydrocarbons (*i.e.*, in the absence of -OH). Finally, complexes B-4,

Table 3.1: The energy of interaction, E_{int}, between chains and water molecules under various situations. E_X indicates the total energy of either the isolated molecule X or the complex X. Snapshots of the stable complexes are shown in Figures 3.4 and 3.5. In the E_{int} formula, the denominator represents the number of interacting groups, such as -CH_3, -OH, or water molecules depending on the case.

Interaction	-OH species	E_{int} Formula	E_{int} (eV)
chain-chain interaction	absent	$\frac{2E_{C_{11}H_{24}} - E_{A-1}}{4}$	0.01
chain-chain interaction	present	$\frac{2E_{C_{11}H_{22}(OH)_2} - E_{A-2}}{4}$	0.28
chain-water interaction	absent	$E_{A-1} + E_{H_2O} - E_{A-3}$	0.09
chain-water interaction	present	$E_{A-2} + E_{H_2O} - E_{A-4}$	0.41
chain-water interaction	present	$\frac{E_{A-2} + 2E_{H_2O} - E_{A-5}}{2}$	0.59
molecule-molecule interaction	absent	$\frac{2E_{C_2H_6} - E_{B-1}}{2}$	0.03
molecule-molecule interaction	present	$\frac{2E_{C_2H_5OH} - E_{B-2}}{2}$	0.20
molecule-water interaction	absent	$E_{B-1} + E_{H_2O} - E_{B-3}$	0.06
molecule-water interaction	present	$E_{B-2} + E_{H_2O} - E_{B-4}$	0.66
molecule-water interaction	present	$E_{B-2} + E_{H_2O} - E_{B-5}$	0.56
molecule-water interaction	present	$E_{B-2} + E_{H_2O} - E_{B-6}$	0.51

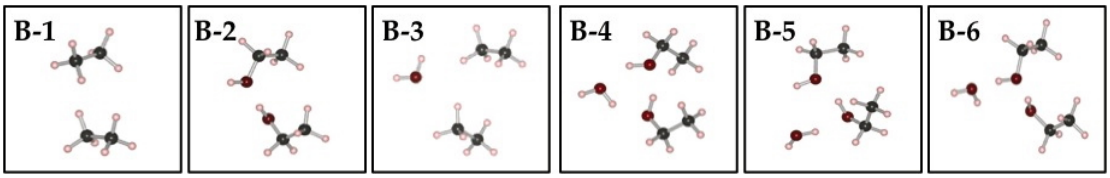


Figure 3.5: (Color online) Optimized structures of complexes B-1, B-2, B-3, B-4, B-5, and B-6. Black, white and red spheres represent C, H and O atoms, respectively.

B-5 and B-6 represent three different configurations involving two C_2H_5OH and one water molecule (and hence, analogous with complex A-4). The molecule-molecule interaction energy with -OH species being absent is 0.03 eV per CH_3 group (c.f. Table 3.1). When there are -OH species, the molecule-molecule interaction energy increases to 0.20 eV per -OH group due the formation of hydrogen bonding between the two hydroxyl groups. The interaction of water with the hydrocarbons (*i.e.*, when -OH species are absent) is rather weak, and is estimated to be 0.06 eV. However, the presence of -OH species dramatically alters this picture. In complexes B-4, B-5, and B-6, we placed the water molecule either near one of two -OH groups (complexes B-5 and B-6) or close to both of the -OH groups (complex B-4). The binding energy of water to the -OH containing molecules is in the 0.51-0.66 eV range depending on the configuration. These findings are consistent with our longer chain model (c.f. Table 3.1). Our DFT results clearly demonstrate that when there are hydroxyl groups present in the system, water molecules have a strong tendency to form hydrogen bonding with the hydroxyl groups and will be trapped in the -OH containing environment.

The dielectric constants pertaining to the A-1, A-2, A-4, and A-5 complexes were then calculated. The trace of the dielectric constant tensor corresponding to both

Table 3.2: The trace of the dielectric constant tensor of complexes A-1, A-2, A-4, and A-5. ϵ^{elec} represents the electronic part of the dielectric constant, while ϵ^{tot} is the total dielectric constant (i.e., including electronic and ionic contributions). The error bars indicate the uncertainty inherent in the procedure used to extrapolate the computed value to the correct volume.

Complex	ϵ^{elec}	ϵ^{tot}
A-1	2.4±0.4	2.4±0.4
A-2	2.4±0.4	3.3±0.2
A-4	2.4±0.4	3.7±0.7
A-5	2.4±0.4	4.3±1.5

the electronic (the high frequency, or optical part) and total (electronic plus ionic, or the static) values for these systems is presented in Table 3.2. For complex A-1, the dielectric constant is almost entirely due to electronic contributions, and the predicted value is close to the experimental value of polyethylene [115]. Successive increases in the dielectric constant values can be seen due to the introduction of the -OH species, and subsequently, due to the introduction of one and two water molecules. The increase in the dielectric constant due to the introduction of -OH groups (without and with water molecules) is entirely due to the ionic contributions.

In order to understand the increased ionic contribution to the dielectric constant with substitution of a H atom by a hydroxyl group and the presence of bound water molecule(s), we present the following analysis. The ionic contribution to the dielectric constant (total minus electronic) is purely controlled by the infrared (IR) active zone center phonon modes (*i.e.*, the modes that display a time-varying dipole moment) [129, 130]. The extent to which each IR-active phonon mode contributes

to the dielectric tensor is determined by the frequency of the mode (smaller the frequency larger the dielectric constant) and the IR intensity (I_i^{IR}) of the corresponding zone center mode (larger the IR intensity, larger the dielectric constant), which is given by [131, 132]

$$I_i^{IR} \propto \sum_{\alpha} \left| \sum_{k,\beta} Z^*_{k,\alpha\beta} X_i(k,\beta) \right|^2$$

where i labels the modes, k labels atoms, and α and β are the cartesian coordinates. ${Z^*}_{k,\alpha\beta}$ and $X_i(k,\beta)$ represent appropriate components of the Born effective charge tensor and the phonon mode eigenvector, respectively. For IR-inactive phonon modes, this intensity vanishes. The zone center phonon modes and frequencies were computed for the complexes A-1, A-2, A-4, and A-5, and the IR intensities determined. Results of these computations are summarized in Figure 3.6 in the form of histograms that capture the dominant IR-active modes. From Figure 3.6 we can see that compared with complex A-1, introducing OH groups results in new modes in complex A-2, and the dominant ones are the C-O stretching mode and O-H stretching mode. For the complexes with H_2O (*i.e.*, complexes A-4 and A-5), a new mode at around 1600 cm^{-1} corresponding to the H_2O O-H bending mode appears. A few new modes appear at around 3200 cm^{-1} to 3300 cm^{-1}, which are also O-H stretching modes arising from both the hydroxyl group and the H_2O. We are thus persuaded to conclude that -OH functionalizations and the inevitable trapped water act cooperatively to increase the ionic part of the dielectric constant of polyolefins.

While the DFT computations provide some understanding of the role played by -OH functional groups and the accompanying water molecules, these data are

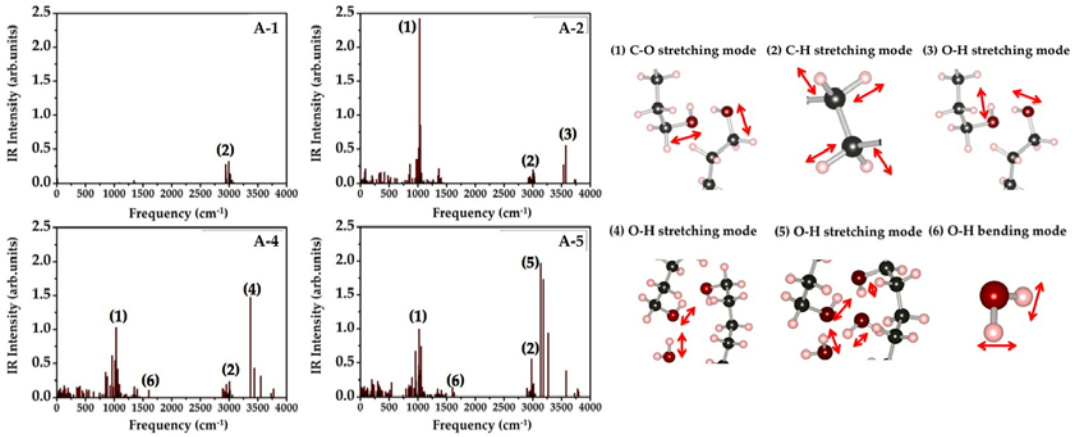


Figure 3.6: (Color online) IR intensity of the dominant IR-active zone center phonon modes of complexes A-1, A-2, A-4, and A-5. The character of IR-active modes are illustrated using schematics.

qualitative. The DFT analysis ignores complexities associated with morphological variations of PE when functionalized with -OH, and owing to computational considerations, the concentration of -OH modeled is much greater than in experiments. Large-scale MD simulations based on force fields are required to address morphological aspects of the system at realistic -OH concentrations. Recent force-field based MD simulations have extended the DFT work discussed above [133]. In the MD simulations, PE-OH with 4.2 mol% -OH groups was considered, and varying amounts of water were added into PE-OH system. The morphology of -OH groups, trapped water and H-bonding was well captured, insofar as can be concluded by comparisons with measured infrared spectra. In addition, the MD simulations indicate that the -OH groups tend to collect at amorphous-crystalline interfaces, likely as a result of -OH groups "expelled" from the crystalline regions during formation. These simulations also imply that hydrogen bonding of the -OH groups into pairs is essential to maintaining low dielectric loss. Figure 3.7 summarizes data for the

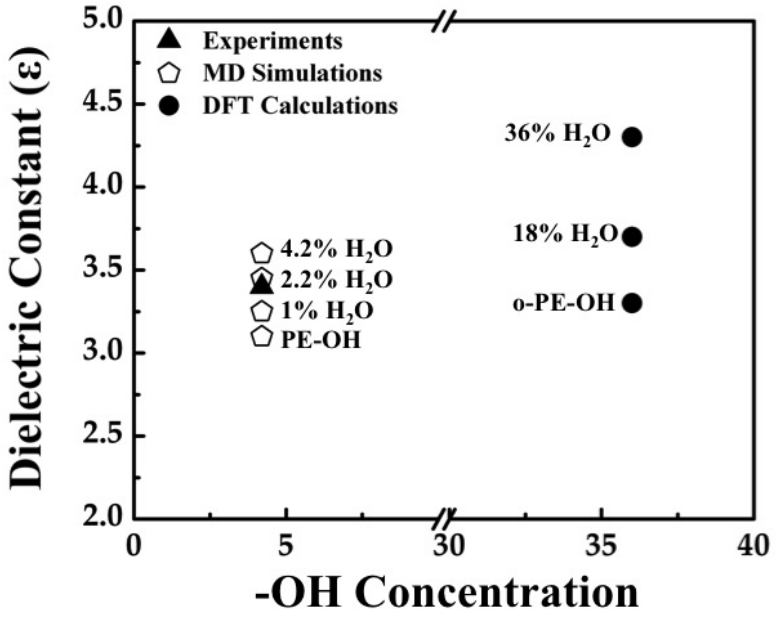


Figure 3.7: Comparison of the dielectric constant of hydroxyl functionalized polyethylene (PE-OH), obtained from experiment, MD simulations and DFT calculations.

dielectric constant derived from the MD simulations, in addition to the results from DFT computations and experiment. The addition of water results in a significant increase in dielectric constant. For ∼0.5 water per -OH group, the MD data for dielectric constant match closely the experimental value of ∼3.4, and larger amounts of water, so long as they are "trapped" by the -OH groups much like in Figure 3.4 (A-5), increase the dielectric constant further without increasing loss.

The DFT, MD and experimental data, collectively, have allowed us to understand the factors that lead to a significant increase in the dielectric constant of a saturated hydrocarbon due to the incorporation of a small amount of -OH groups. These findings are significant, as they imply a path toward tunable control of dielectric

properties of polyolefins through functionalization. While the analysis presented underlines the complexities that can result from incorporation of a functional group, it also demonstrates the utility of computational strategies for systematic studies of other functional groups.

3.4 Conclusion

In summary, the DFT calculations indicate that introduction of -OH functional groups will enhance the interaction between PE chains due to interchain hydrogen bonding. Moreover, when -OH groups are present, water molecules tend to be trapped due to the strong interaction between water and the -OH functionalized PE chains. The existence of -OH groups and water molecules in the system contribute cooperatively to the increase of the dielectric constant of PE. These findings are significant, as they point towards a rational pathway for the tunable control of materials properties of polyolefins as well as other related materials systems via creative utilization of molecular functionalization (and moisture).

Chapter 4

Pathways to the discovery of new organic polymer dielectrics

4.1 Introduction

While modifying an existing polymer to enhance its properties is a promising approach and offers a "risk mitigation" strategy, identifying or discovering entirely new classes of polymers can be transformative but requires systematic schemes to explore the polymer chemical space. Figure 4.1 portrays one such scheme in which single polymeric chains consist of four distinct building blocks drawn from a pool of possibilities in a combinatorial manner [68, 69]. Depending on the pool of blocks, various polymer classes can be studied.

In this chapter, we have explored the organic polymer chemical space using the chain scheme (Figure 4.1). To be specific, we allow the blocks to be -CH_2-, -NH-, -C(=O)-, -C_6H_4- (benzene), -C_4H_2S- (thiophene), -C(=S)-, or -O- units in a

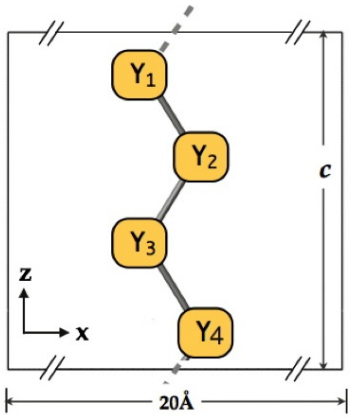


Figure 4.1: Schematic representation of the polymeric chain model. c is the lattice constant along the polymer chain direction that was allowed to relaxed along with the internal coordinates. Four blocks Y_1, Y_2, Y_3, and Y_4 can be filled up by different motifs.

systematic, progressive, and exhaustive manner. All of these units are commonly seen in polymer backbones. High throughput methods were used first to accurately determine the dielectric constant and band gap of all the systems. Based on the DFT results, we have identified several polymer sub-classes which offer a better tradeoff between the dielectric constant and the band gap.

4.2 High throughput DFT calculations

In theory, there is a huge number of different organic polymers, but only a limited amount has been synthesized. Out of the synthesized polymers, only a few are considered as dielectric materials. We first consider a set of organic building blocks, for example, -CH_2-, -NH-, -C(=O)-, -C_6H_4- (benzene), -C_4H_2S- (thiophene), -C(=S)-, and -O- [134, 144], which are common in polymer backbones and various combinations of which form traditional polymers, including polyesters, polyamides,

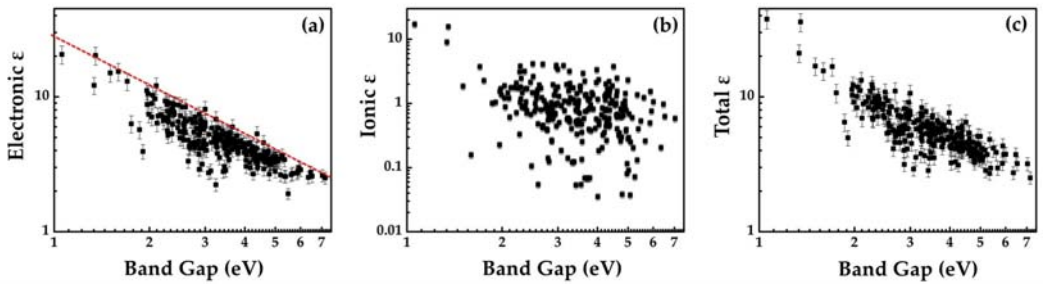


Figure 4.2: (a) Electronic, (b) ionic and (c) total dielectric constant (ϵ) as a function of the band gap, for the class of organic polymers considered, computed using DFT within the single-chain approach. The building blocks were drawn from the following pool of possibilities: -CH_2-, -NH-, -C(=O)-, -C_6H_4- (benzene), -C_4H_2S- (thiophene), -C(=S)-, and -O-. The axes are in logarithmic scale.

polyethers, polyureas, etc. Accounting for translational and inversion symmetries, and removing unstable systems, leaves 267 symmetry-unique systems.

The two most important properties of dielectrics for most applications are dielectric constant and band gap. Hence, we focus on these two properties in this section. With the aid of the single chain method discussed in Chapter 2.4 [68], high throughput DFT computations can be used to obtain the band gap and dielectric constant of the 267 polymer systems. Figure 4.2(a), (b) and (c), respectively, show the relationship between the electronic, ionic and total dielectric constant and band gap for the 267 polymer systems. A near perfect inverse Pareto optimal front relationship between the band gap and the electronic dielectric constant can be seen from Figure 4.2(a), which imposes a theoretical limit on the electronic part of the dielectric constant as a function of band gap, a limit that can be understood by regarding the electronic part of the dielectric response as a sum over electronic transitions from occupied to unoccupied states. On the other hand, the ionic dielectric

constant is not correlated with the band gap, as seen from Figure 4.2(b). The ionic contribution is determined by the infrared active zone center phonon modes (i.e., the modes that display a time-varying dipole moment).

The ionic dielectric constant can thus be exploited to increase the total dielectric constant without compromising the band gap. Figure 4.2(c) shows the variation of the total dielectric constant with the band gap. Plots such as Figure 4.2(c) provide us with a "map" of the achievable combination of properties within the chemical space explored. Capacitive energy storage and some electronics applications, e.g., gate insulations, could draw from the large dielectric constant and moderate band gap region of this plot. Desirable material properties are likely to be associated with certain building block structures. In order to identify the building blocks that correlate with certain properties, the following analysis has been conducted. Let's consider a certain property of organic polymers, e.g., band gap. The band gap values vary from 0 to 7.2. This range can then be binned with a certain step size. Each bin will correspond to a number of polymers. The number of singles (CH_2, NH, C(=O), C_6H_4, C_4H_2S, C(=S), and O), doubles (CH_2-CH_2, CH_2-NH, NH-CO, NH-C_6H_4, *etc.*), and triples (CH_2-CH_2-CH_2, NH-CO-NH, NH-CO-O, C_6H_4-C_6H_4-C_6H_4, *etc.*) are counted for each of the polymers in each bin. Then add up the total number of each singles, each doubles, and each triples for all the polymers in the bin, and identify the top three singles, doubles and triples. All three propeties, band gap, electronic dielectric constant and ionic dielectric constant, have undergone this kind of analysis. The results are shown in Figure 4.3 with the counts in each bin

normalized according to the number of polymers. From the plot, we can see that in the low band gap (high electronic dielectric costant) region, C_4H_2S-containing doubles and triples are dominant. While in the high band gap (low electronic dielectric constant) part, CH_2-containing doubles and triples are the important blocks. In addition, CS shows up in the high ionic dielectric constant regime, and in the low band gap regime as well. But CO appears in both the high ionic dielectric cosntant (not as high as CS) and in the high band gap area, making it more attractive than CS. Such observations could accelerate the design process by identifying correlations between material properties and specific building blocks structures.

Once a set of promising polymers has been identified, their crystal structures and morphologies must be investigated. If inter-atomic potentials are available to handle the systems, MD simulations can be used to determine their crystal structure and morphology, for example, using a melt and quench approach [83–86]. An additional option, especially if force fields are not available for the new identified systems, is to use 3-dimensional structure searching schemes [135–140] to determine the ground state structures based on DFT.

4.3 Guidance to synthesis efforts

Among all 267 polymers we have calculated, besides PE (CH_2-CH_2-CH_2-CH_2), we identified 13 polymers which have already been synthesized before. The list of these polymers is provided in Table 4.1 along with their calculated band gap, electronic and total dielectric constant values (The polymers are listed in the decreasing

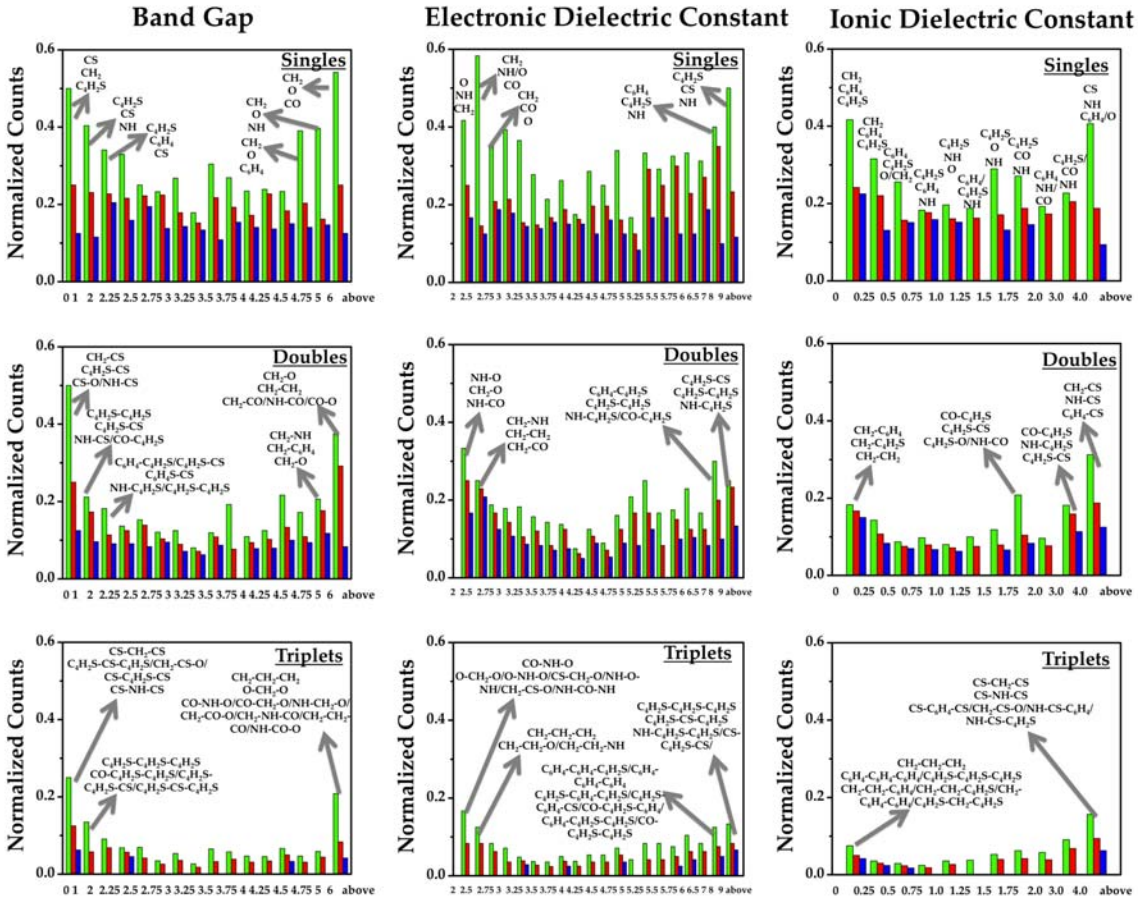


Figure 4.3: Normalized counts of top three singles, doubles and triples in each bin for band gap, electronic dielectric constant and ionic dielectric constant.

order of total dielectric constant). Most of these polymers have not been considered as dielectric materials in the past. As shown in Table 4.1, all 13 polymers have larger dielectric constant than PE and PP, and relatively large band gap. This study may lead a new direction for them to be considered for dielectric applications. In fact, the top three of these 13 polymers have been synthesized by our experimental collaborators (Prof. Sotzing, UCONN), which are NH-CO-C_6H_4-CO, NH-CO-NH-C_6H_4, and NH-CS-NH-C_6H_4. The values in the parenthesis are experimental dielectric constant at 1k Hz. Considering that the calculated dielectric constant is from single-chain approach, the agreement between computed and experimental dielectric constant is reasonable.

Since we only consider four block systems, there is only a small fraction of the 267 polymers that have been synthesized. In terms of synthesis efforts, the repeat units of our systems can be modified to make them synthesizable. Indeed, guided by the DFT results, a few modified polyureas and polyurethanes containing -NH-, -C(=O)-, -C_6H_4-, and -O- building blocks have been synthesized using step polymerization [144]. The structures of these modified polyureas and polyurethanes are shown in Figure 4.4. The details of the syntheis effort are provided in Appendices A and B.

4.4 Conclusion

In this chapter, we have explored the organic polymer chemical space using the chain scheme, in which the building blocks were drawn from the following pool

Table 4.1: Already synthesized polymers. The band gap (E_g), electronic (ϵ_e) and total (ϵ_t) dielectric constant are DFT calculated results. The values in the parenthesis are experimental dielectric constant at 1k Hz.

System repeat unit	Polymer class	E_g (eV)	ϵ_e	ϵ_t
NH-CO-C_6H_4-CO	Polyimide	4.5	3.7±0.4	7.4±1.1 (4.3)
NH-CS-NH-C_6H_4 [141]	Polyurea	3.0	5.7±0.8	6.7±0.9 (5.9)
NH-CO-NH-C_6H_4 [141]	Polyurea	3.8	5.0±0.6	6.5±0.9 (5.5)
NH-C_6H_4-C_6H_4-C_6H_4	Polyamine	3.2	5.8±0.8	6.3±0.8
CO-C_6H_4-CO-O	Polyester, polyanhydride	4.6	3.8±0.4	6.2±0.8
CH_2-C_4H_2S-C_4H_2S-C_4H_2S [142]	Polyother	2.8	5.8±0.8	5.9±0.8
C_6H_4-C_6H_4-C_6H_4-O	Polyether	3.6	5.1±0.6	5.4±0.7
CH_2-C_6H_4-C_6H_4-O	Polyether	4.3	4.2±0.5	4.6±0.6
CH_2-CO-C_6H_4-O [143]	Polyether, polyketone	4.2	3.8±0.4	4.6±0.6
CH_2-C_6H_4-CO-O	Polyester	4.9	3.6±0.4	4.4±0.5
CH_2-NH-CO-NH [141]	Polyurea	5.9	2.8±0.3	4.3±0.6
CH_2-NH-CS-NH [141]	Polyurea	4.1	3.3±0.4	4.1±0.5
CH_2-C_6H_4-CH_2-O	Polyether	5.0	3.4±0.4	3.7±0.5

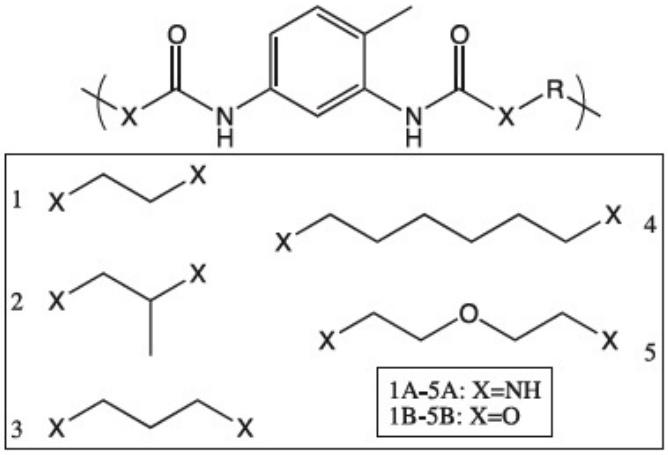


Figure 4.4: Structures of synthesized polyureas and polyurethanes.

of possibilities: -CH_2-, -NH-, -C(=O)-, -C_6H_4- (benzene), -C_4H_2S- (thiophene), -C(=S)-, and -O-. The dielectric constant and band gap were computed using DFT for a manageable combination of polymer chains (totaling 267 cases). There is an apparent inverse relationship between the electronic part of the dielectric constant and the band gap, which set restrictions in the degree to which the electronic dielectric constant can be enhanced. On the other hand, the ionic dielectric constant is not correlated with the band gap, and can thus be exploited to increase the total dielectric constant without compromising the band gap. Based on the DFT results, promising polymer systems with both high dielectric constant and large band gap have been identified. The systems can provide guidance to synthesis efforts. Indeed, a few modified polyureas and polyurethanes containing -NH-, -C(=O)-, -C_6H_4-, and -O- building blocks have been synthesized by our experimental collaborators.

Chapter 5

Pathways to the discovery of Group 14 element-based hybrid polymer dielectrics

5.1 Introduction

The strategy described in the previous chapter can be extended by considering building units based on non-carbon elements from Group 14 of the periodic table, i.e., Si, Ge and Sn. Replacing C with Si, Ge or Sn offers the opportunity to manipulate the band gap and the electronic part of the dielectric constant through control of σ conjugation along the chain, and to manipulate the ionic part of the dielectric constant through control of dipole moments. Substituting C with Si, Ge and/or Sn ensures chemical compatibility by preserving the local chemical environment and bonding. The dipole moments of each building block can be enhanced by introducing small atoms with high electronegativity such as F and Cl to the side chain.

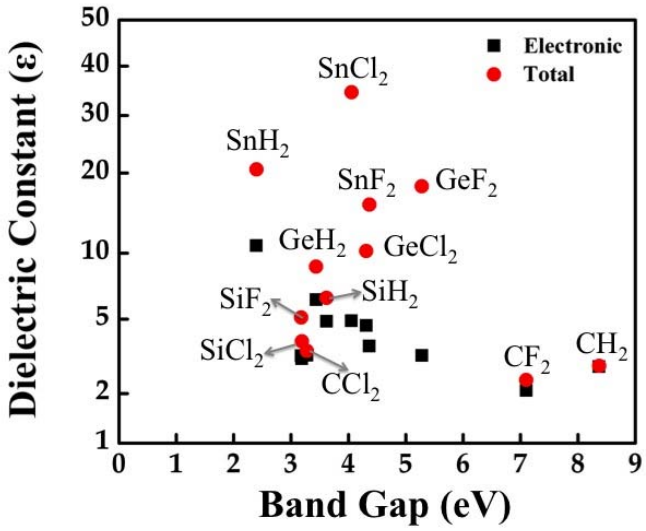


Figure 5.1: Dielectric constant and band gap results for -XY_2- homopolymer crystals at their respective ground state structure (where X = C, Si, Ge, Sn and Y = H, F, Cl).

In this chapter, the building blocks in Figure 4.1 are drawn from the following pool of possibilities: -CH_2-, -SiF_2-, -$SiCl_2$-, -GeF_2-, -$GeCl_2$-, -SnF_2- and -$SnCl_2$-. We then use the high throughput method that employs DFT calculations of isolated single polymer chains to accurately estimate the dielectric constant and band gap of all possible polymer systems. A set of most promising polymers (with simultaneously large dielectric constant and band gap), which contains Ge and Sn in the polymer backbone, is identified.

5.2 High throughput DFT calculations

Before exploring the single chain systems based on the combinatorial exercise, we investigate -XY_2- (X = C, Si, Ge, Sn and Y = H, F, Cl) "homopolymer" crystals to see if any pattern emerges. Group 14 systems display rich chemistry and crystallize

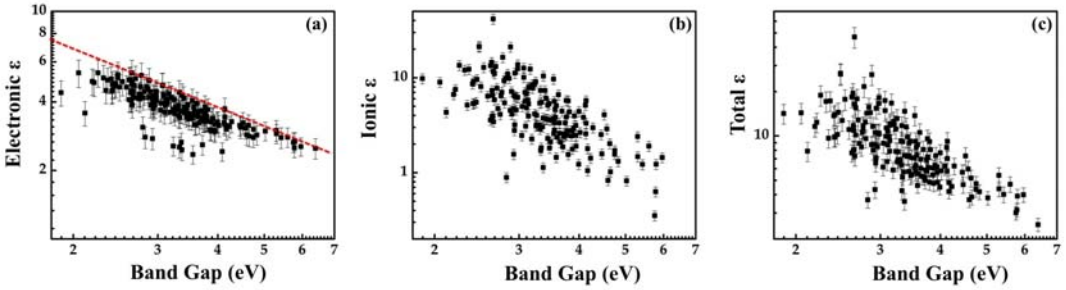


Figure 5.2: (a) Electronic, (b) ionic and (c) total dielectric constants (ϵ) as a function of the band gap, for the class of polymers containing Group 14 elements, computed using DFT within the single-chain approach. The building blocks were drawn from the following pool of possibilities: -CH_2-, -SiF_2-, -$SiCl_2$-, -GeF_2-, -$GeCl_2$-, -SnF_2- and -$SnCl_2$-. The axes are in logarithmic scale.

in forms with differing coordination geometries. While C always prefers a 4-fold coordination environment, Ge favors a 5-fold environment, and Sn can occur in 6 or 7 fold coordination geometries [67, 68]. Figure 5.1 presents dielectric constant and band gap data for homopolymer ground state geometries. The electronic part of the dielectric constant tends to vary inversely with the band gap, as expected, while the ionic contribution to the dielectric constant is negligible for the C based systems, low for the Si based systems, but quite high for the Ge and Sn containing polymers. Figure 5.1 also reveals that -XY_2- homopolymers with Ge and Sn backbone have large dielectric constants but smaller band gaps than C and Si-based polymers. Pure PE has a large band gap but relatively small dielectric constant.

A natural next step is to "mix" PE with the -XY_2- homopolymers to identify compositions which span a large range of band gap and dielectric constant. Such a combinatorial exploration based on the scheme presented in Figure 4.1 involving 7 building block possibilities (namely, -CH_2-, -SiF_2-, -$SiCl_2$-, -GeF_2-, -$GeCl_2$-, -SnF_2-

and -$SnCl_2$-) was recently undertaken [69, 87]. Results for dielectric constant vs. band gap for the 175 such single chain polymers is shown in Figure 5.2. While PE has the greatest calculated band gap of the systems explored, addition of Group 14 elements leads to progressive decrease in the band gap and increase in the electronic dielectric constant. The total dielectric constant of the Group 14 element-based polymers spans over a large range between 2.5 to 47, with the smallest and largest values corresponding to -$(CH_2)_4$- and -CH_2-$(SnF_2)_3$-, respectively. As the backbone atoms vary from C to Sn with all other units in the chain are held fixed, both the electronic and total dielectric constant increase, as the band gap decreases, probably because -SnF_2- has the largest dipole moment and the Sn-Sn bond rotation has the lowest barrier among all the X-X' (with X or X' = C, Si, Ge, and Sn). Compared with organic polymers, the Group 14 element-based hybrid polymers, especially those involving Sn, can achieve greater dielectric constants without large reductions of the band gap, which makes them attractive candidates for high dielectric constant polymeric dielectrics [69, 87].

5.3 Guidance to synthesis efforts

DFT results indicate that, in general, in order to enhance the dielectric constant without sacrificing the good insulating properties of polyethylene, it is desirable to incorporate SiF_2, GeF_2, and SnF_2 motifs in the polyethylene backbone. Cholrides of Si, Ge, and Sn usually result in relatively lower band gaps. The order of relative importance of fluorides in improving the dielectric constant is predicted to be SiF_2

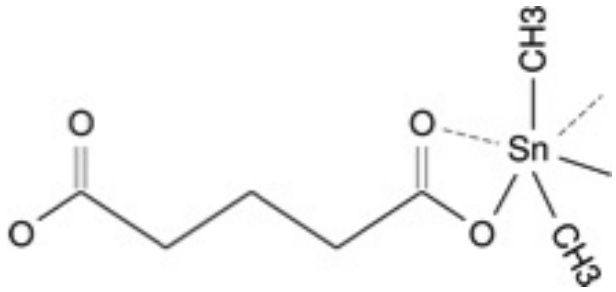


Figure 5.3: The structure of poly(dimethyltin glutarate).

$< GeF_2 < SnF_2$. Since Ge is more expansive, Sn containing polymers have greater potential, in that a lower mole fraction of Sn within the backbone in comparison, will result in a higher overall dielectric constant.

Considering that the Sn containing polymers we have calculated are difficult to synthesize, our experimental collaborators have successfully made modified Sn containing polymer, which is named as tin ester. The structure of this organometallic polymer, poly(dimethyltin glutarate), is shown in Figure 5.3. DFT is then used to calculate this polymer's dielectric constant ($\epsilon = 6.04$), band gap ($E_g = 6.14$ eV), stable crystal structure and IR spectrum. Comparison of the computed values and spectra to the experimental measurements of poly(dimethyltin glutarate) shows a direct correlation. The measured dielectric constant is 7.39 with an band gap of 4.88 eV, while the IR and XRD spectra confirm the presence of the four most stable crystal structures in varying proportion dependent upon processing conditions.

5.4 Conclusion

In this chapter, we expand our search space to a new class of polymer systems, the Group 14 element-based hybrid polymers. The building blocks in the chain

scheme were drawn from the following pool of possibilities: -CH_2-, -SiF_2-, -$SiCl_2$-, -GeF_2-, -$GeCl_2$-, -SnF_2- and -$SnCl_2$-. Improvement in the dielectric constant by purely increasing electronic contribution tends to degrade the insulating properties, since the electronic dielectric constant correlates inversely with the band gap. The ionic dielectric constant, on the other hand, is not limited by the band gap, and thus can provide more latitude. SiF_2, GeF_2, and SnF_2 have been identified to be promising in terms of increasing the dielectric constant of polyethylene while still maintaining a relatively large band gap, especially when SnF_2 or GeF_2 motifs are arranged in a consecutive manner.

Chapter 6

Advanced search for Group 14 element-based hybrid polymer dielectrics

6.1 Introduction

DFT-based strategies are limited to polymers based on short, periodic chains. For longer chains required to screen larger regions of compositional space, the computational cost associated with DFT rises rapidly. Furthermore, as the system size increases, the number of candidates within the system grows exponentially, which leads to combinatorial explosion. Take the Group 14 element-based hybrid polymers for example. Doubling the supercell size along the chain direction to include 8 distinct building units in a periodic repeating cell results in a total of 29,365 symmetry unique systems. Clearly, exploration of such a vast chemical space using present first principles based approaches is impractical. A new approach is needed for this large class of systems.

The materials discovery process can be significantly expedited and simplified if we can learn effectively from past knowledge. This part of the work involves the development of a similarity-based machine learning approach that can help predict the band gap and dielectric constant of a new polymer in a minuscule fraction of the time necessitated by a typical DFT computation. The machine learning approach is validated and used to explore a larger scale polymer chemical space for Group 14 element-based hybrid polymers. The strategy developed in this chapter allows us to systematically explore and mine vast chemical spaces, and can significantly accelerate the discovery of new application specific materials.

6.2 Machine learning Scheme

To effect such large scale explorations, a machine learning approach has been developed and applied to the Group 14 element-based hybrid polymers. The essential ingredients of our machine learning approach are captured schematically in Figure 6.1. The first step in the approach prescribed in the panels of Figure 6.1 is to reduce each material system under inquiry to a string of numbers—we refer to this string as the feature (or fingerprint) vector. For the polymer chains composed of seven possible building blocks, the following coarse-level chemo-structural feature vector was considered first: $|f_1, \ldots, f_6, g_1, \ldots, g_7, h_1, \ldots, h_7\rangle$, where f_i, g_i and h_i are, respectively, the number of building blocks of type i, number of $i-i$ pairs, and number of $i-i-i$ triplets, normalized to total number of units (note that f_7 is

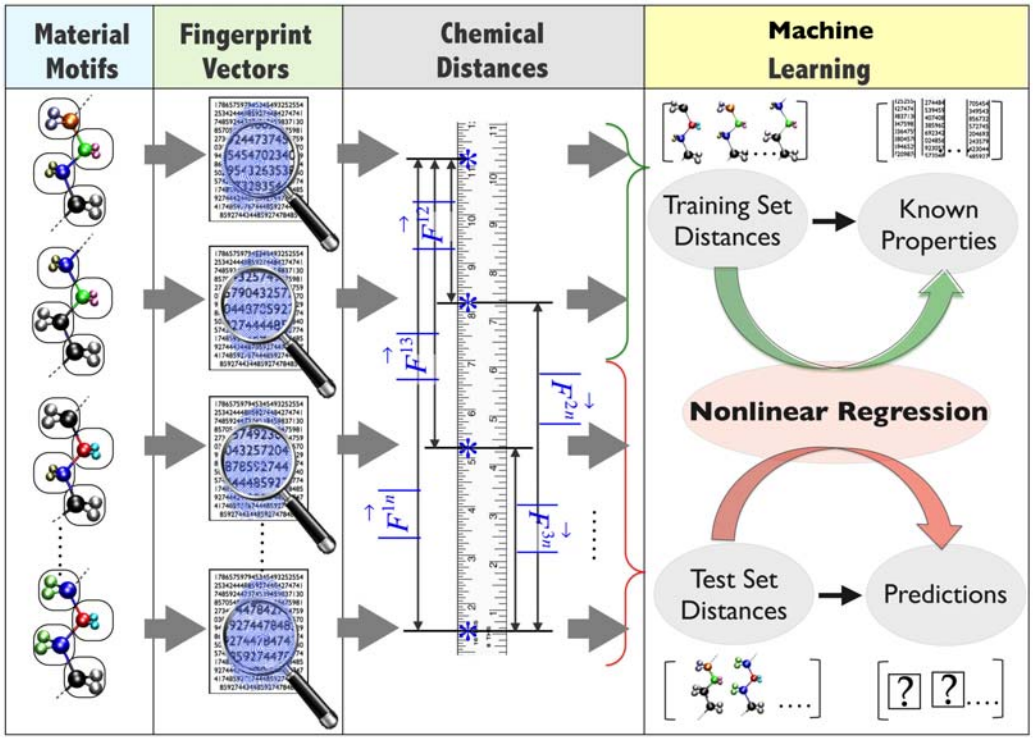


Figure 6.1: Schematic of the machine (or statistical) learning methodology employed. First, material motifs within a class are reduced to numerical feature vectors. Next, a suitable measure of chemical (dis)similarity, or chemical distance, is used within a learning scheme—in this case, kernel ridge regression—to map the distances to properties.

missing in above vector as it is not an independent quantity owing to the relation: $f_7 = 1 - \Sigma_{i=1}^{6} f_i$).

Next, a suitable measure of *chemical distance* is defined to allow for a quantification of the degree of (dis)similarity between any two feature vectors. Consider two systems a and b with feature vectors $\vec{F}^a$ and $\vec{F}^b$. The similarity of the two vectors may be measured in many ways, e.g., using the Euclidean norm of the difference between the two vectors, $|\vec{F}^a - \vec{F}^b|$, or the dot product of the two vectors $\vec{F}^a \cdot \vec{F}^b$. In the present work, we use the former, which we refer to as $|\vec{F}^{ab}|$ (Figure 9). Clearly, if $|\vec{F}^{ab}| = 0$, materials a and b are equivalent (insofar as we can conclude based

on the feature vectors), and their property values P^a and P^b are the same. When $|\vec{F}^{ab}| \neq 0$, materials a and b are not equivalent, and $P^a - P^b$ is not necessarily zero, and depends on $|\vec{F}^{ab}|$. This observation may be formally quantified when we have a prior materials-property dataset, in which case we can determine the parametric dependence of the property values on $|\vec{F}^{ab}|$. In the present work, we use the *statistical learning* algorithm referred to as kernel ridge regression (KRR). Using this scheme, the property of a new system b is given by a sum of weighted Gaussians,

$$P^b = \sum_{a=1}^{N} \alpha_a \exp\left(-\frac{1}{2\sigma^2}|F^{ab}|^2\right). \tag{6.1}$$

where a runs over the systems in the previously known dataset. The coefficients α_as and the parameter σ are obtained by 'training' the above form on the systems a in the previously known dataset, while adhering to the best practices of statistical and cross-validation methods [145, 146].

Within the present similarity-based learning model, the learning process is built on minimizing the expression $\sum_{a=1}^{N}\left(P^a_{Est} - P^a_{DFT}\right)^2 + \lambda \sum_{a=1}^{N} \alpha_a^2$, with P^a_{Est} being the estimated property value, P^a_{DFT} the DFT value, and λ a regularization parameter. The explicit solution to this minimization problem is $\boldsymbol{\alpha = (K + \lambda I)^{-1} P_{DFT}}$, where $\boldsymbol{I}$ is the identity matrix, and $K_{ab} = \exp\left(-\frac{1}{2\sigma^2}|F^{ab}|^2\right)$ is the kernel matrix elements of all polymers in the training set. The parameters λ, σ and α_as are determined in an inner loop of fivefold cross validation using a logarithmically scaling fine grid.

The above scheme has been applied to the Group IV element-based hybrid polymers. As mentioned above, the initial dataset was generated using DFT for systems

with repeat units containing 4 distinct building blocks. Of the total 175 such systems, 130 were classified to be in the 'training' set (used in the training of the KRR model), and the remainder in the 'test' set. To validate the machine learning approach, besides the band gap and dielectric constant, five more properties (atomization energy, formation energy, c lattice parameter, spring constant, and electron affinity) were considered. Figure 6.2 shows the agreement between the predictions of the learning model and the DFT results for the training and the test sets. Furthermore, we considered several chains composed of 8-block repeat units (in addition to the 175 4-block systems), performed DFT computations on these, and compared the DFT predictions of the 8-block systems with those predicted using our learning scheme. As can be seen, the level of agreement between the DFT and the learning schemes is uniformly good for all eight properties across the 4-block training and test set, as well as the somewhat out-of-sample 8-block test set (regardless of the variance in the property values). Overall, the high fidelity nature of the learning predictions is particularly impressive, given that these calculations take a minuscule fraction of the time necessitated by a typical DFT computation.

The entire discussion thus far has focused on feature vectors defined in terms of coarse-level chemo-structural descriptors. This brings up a question as to whether other more fundamental quantities may be used as a fingerprint to profile a material. The first Hohenberg-Kohn theorem of DFT [62] proves that the electronic charge density of a system is a universal descriptor containing the sum total of the information about the system, including all its properties. The shape [147] and

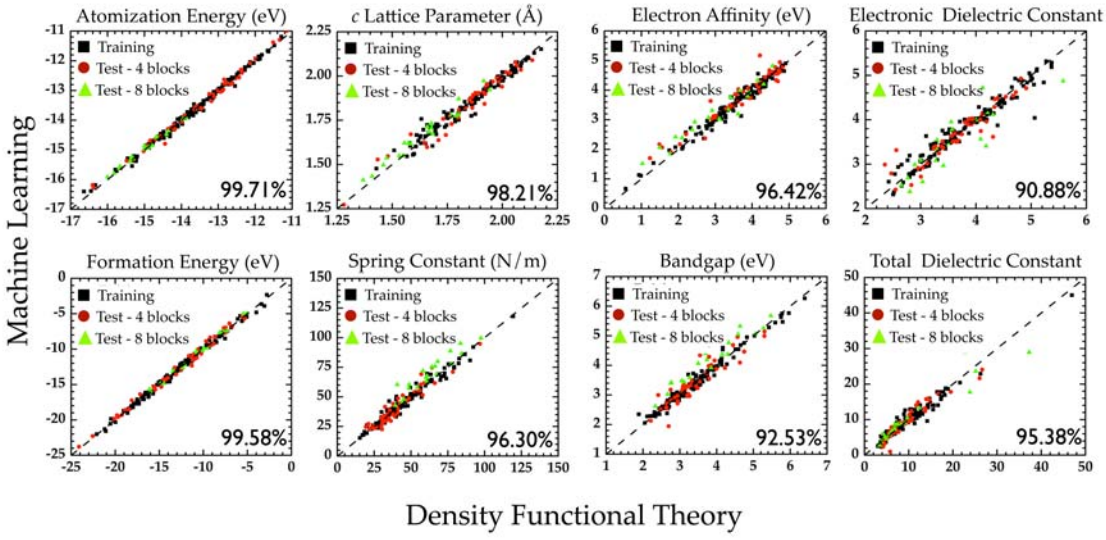


Figure 6.2: Parity plots comparing property values computed using DFT against predictions made using learning algorithms trained using chemo-structural feature vectors. Pearson's correlation index is indicated in each of the panels to quantify the agreement between the two schemes.

the holographic [148] electron density theorems constitute further extensions of the original Hohenberg-Kohn theorem. Inspired by these theorems, we propose that machine learning methods may be used to establish a mapping between the electronic charge density and various properties.

A fundamental issue related to this perspective deals with defining a (dis)similarity criterion that can enable a fair comparison between the charge density of two different systems. Note that any such measure has to be invariant with respect to relative translations and/or rotations of the systems. In the present work, we have employed Fourier coefficients of the 1-d charge density of our systems (averaged along the plane normal to the chain axis). The Fourier coefficients are invariant to translations of the systems along the chain axis, and consideration of the 1-d planar

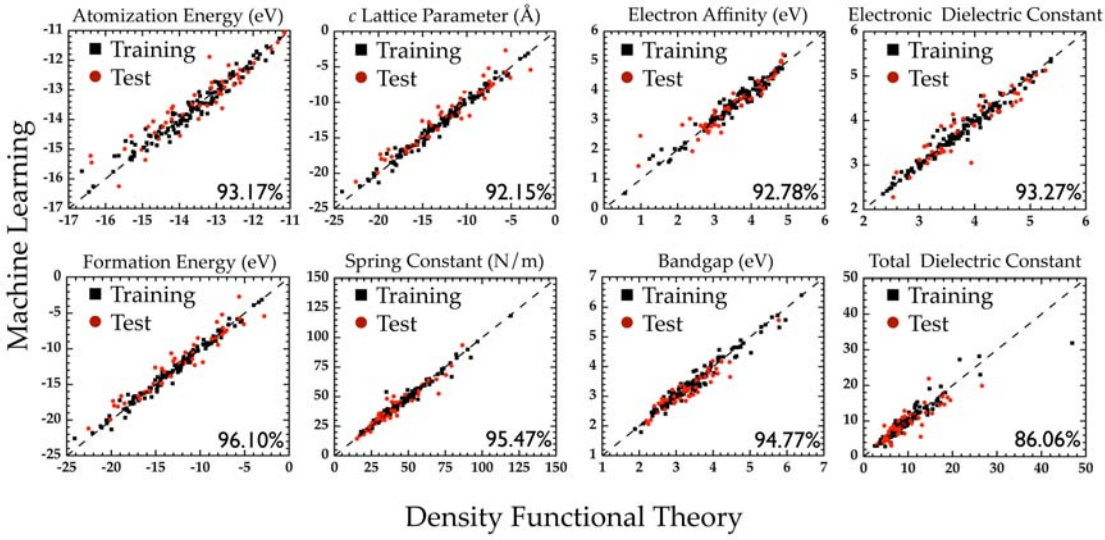


Figure 6.3: Parity plots comparing property values computed using DFT against predictions made using learning algorithms trained using electron density-based feature vectors. The Fourier coefficients of the planar-averaged Kohn-Sham charge density are used to construct the feature vector. Pearson's correlation index is indicated in each of the panels to quantify the agreement between the two schemes.

averaged charge density makes the rotational degrees of freedom irrelevant. Figure 6.3 shows a comparison of the predictions of the learning model based on charge density with the corresponding DFT results. While the agreement between the learning scheme and DFT is not as remarkable as with the chemo-structural fingerprint approach adopted earlier, this can most likely be addressed by the utilization of the actual 3-d charge density. Nevertheless, we believe that the performance of the learning scheme is satisfactory, and heralds the possibility of arriving at a 'universal' approach for property predictions solely using the electronic charge density.

A second issue with the charge density based materials profiling relates to determining the charge density in the first place. If indeed a mapping between charge density and the properties can be made for the training set, how do we obtain the

charge density of a new system without explicitly performing a DFT computation? We suggest that the 'atoms in molecules' concept may be exploited to create a patched-up charge density distribution [149]. Needless to say, barring some studies in the area of atoms and molecules [150], these concepts are in a state of infancy, and there is much room available for both fundamental developments and innovative applications.

6.3 Larger scale chemical space exploration of Group 14 element-based hybrid polymers

While the favorable agreement between the machine learning and the DFT results for a variety of properties is exciting, in and of itself, the real power of this prediction paradigm lies in the possibility of exploring a *much* larger chemical-configurational space than is practically possible using DFT computations (or laborious experimentation). For instance, merely expanding into a family of 1-d systems with 8-block repeat units leads to 29,365 symmetry unique cases (an extremely small fraction of this class was scrutinized above for validation purposes). Not only can the learning approach make the study of this staggeringly large number of cases possible, it also allows for a search for correlations between properties in a systematic manner. In order to unearth such correlations, we first determined the properties of the 29,365 systems using our machine learning methodology, followed by the estimation of Pearson's correlation coefficient for each pair of properties. Figure 6.4a

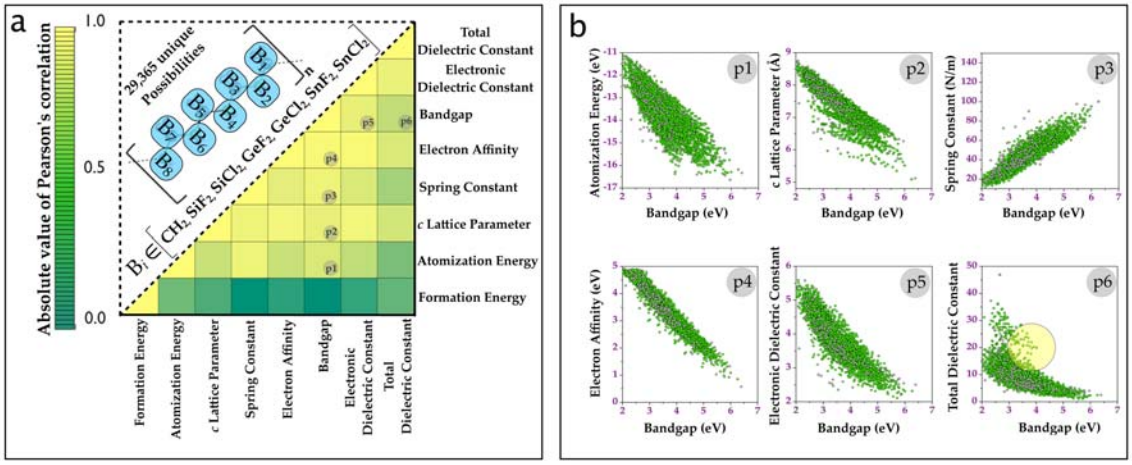


Figure 6.4: **(a)** The upper triangle presents a schematic of the atomistic model composed of repeat units with 8 building blocks. Populating each of he 8 blocks with one of the seven units leads to 29,365 systems. The matrix in the lower triangle depicts the Pearson's correlation index for each pair of the eight properties of the 8-block systems predicted using machine learning. **(b)** Panels p1 to p6 show the correlations between the band gap and six properties. The panel labels are also appropriately indexed in (a). The circle in panel p6 indicates systems with a simultaneously large dielectric constant and band gap.

shows a matrix of the correlation coefficients, color-coded to allow for immediate identification of pairs of properties that are most correlated.

It can be seen from Figure 6.4a that the band gap is most strongly correlated with many of the properties. Panels p1-p6 of Figure 6.4b explicitly show the correlation between the band gap and six of the remaining seven properties. Most notably, the band gap is inversely correlated with the atomization energy (p1), size (p2), electron affinity (p4), and the dielectric constants (p5 and p6), and directly correlated with the spring constant (p3). The relationships captured in panels p1-p3 follow from stability and bond strength arguments. The interesting inverse relationship between the band gap and the electron affinity is a consequence of the uniform shift of the conduction band minimum (due to changes in the band gap) with respect to the

vacuum level. The inverse correlation of the band gap with the electronic part of the dielectric constant follows from the quantum mechanical picture of electronic polarization being due to electronic excitations. As no such requirement is expected for the ionic part of the dielectric constant, it is rather surprising that a rough inverse correlation is seen between the total dielectric constant and the band gap, although clear deviations from this inverse behavior can be seen. Finally, we note that the formation energy is uncorrelated with all the other seven properties, including the band gap. This is particularly notable as it is a common tendency to assume that the formation energy (indicative of thermodynamic stability) is inversely correlated with the band gap (indicative of electronic stability).

Correlation diagrams such as the ones in Figure 6.4b offer a pathway to 'design' systems that meet a given set of property requirements. For the interest of dielectric applications, panel p5 and p6 of Figure 6.4b provide the relationship between the band gap and electronic (total) dielectric constant. The inverse correlation of the band gap with the electronic dielectric constant is confirmed once again from p5. The rough inverse correlation between total dielectric constant and band gap seen in p6 is surprising, although clear deviations from this inverse behavior are seen. This extensive search facilitates identification of candidate polymer dielectrics for various applications. For capacitor applications, a search for polymers with high dielectric constant and large band gap leads to systems in the top part of p6, corresponding to the 'deviations' from the inverse correlation and indicated by a circle in p6. These are systems that contain two or more contiguous SnF_2 units, but with an overall

CH_2 mole fraction of at least 25%. Such organo-tin systems may be appropriate for polymer dielectrics. By learning effectively from the available DFT data, the machine learning approach facilitates searching a much larger scale polymer chemical space efficiently. With the aid of this approach, the discovery of new polymer dielectrics can be accelerated significantly.

6.4 Conclusion

In summary, it has been shown that the efficient and accurate prediction of dielectric constant and band gap of polymer systems is possible by combining the notions of chemical (dis)similarity and machine learning approach. Using Group 14 element-based hybrid polymers, we have presented a general formalism that allow us to discover decision rules that establish a mapping between easily accessible attributes of a system and its properties. We have shown that simple feature vector based on compositional and configurational information can be used to profile a polymer system and make property predictions at an enormously small cost compared with quantum mechanical calculations. The methodology presented in this chapter is of direct relevance in identifying (or screening) undiscovered polymer dielectrics in an efficient manner with high fidelity.

Chapter 7

Advanced search for new organic polymer dielectrics

7.1 Introduction

For the chemo-structural feature vector used in our machine learning approach, actually one may generalize the above vector to include all possible $i-j$ pairs, $i-j-k$ triplets, $i-j-k-l$ quadruplets, etc. For the Group 14 element-based hybrid polymers, we intuitively chose the $i{-}i$ pairs, $i{-}i{-}i$ triplets as the components of the feature vector in the previous chapter. But for the organic polymers, it is not clear which ones are important for a certain propery and should be included in the feature vector. One question remaining in our machine learning approach is related to the proper choice of the components of the feature vector. If we can identify a subset of the feature vector components that are most important in determining a property, this knowledge can be used to perform targeted searches. It is also

conceivable that some of the feature vector components may be hurtful rather than helpful in predictions. For this reason, we have developed a scheme to reduce our feature vector in this chapter.

Genetic algorithm combined with kernel ridge regression is used to determine the proper components of the feature vector that most control a certain property. Trained machine learning models were then used to predict the dielectric constant and band gap of a large number of six block organic polymer systems. This extensive search allows us to identify polymers with suitable dielectric constant and band gap values

7.2 Genetic algorithm for feature selection

Genetic algorithm (GA) [90,151–153] is a powerful tool that can be used to solve computational tasks with high complexity. It is premised on the evolutionary ideas of natural selection and genetic. The basic concept of genetic algorithm is designed to simulate processes in natural system necessary for evolution, specifically those that follow the principles first laid down by Charles Darwin of survival of the fittest. GA can be used to solve optimization/simulation/modelling/design problems, and it also has a high predictive power.

Unlike the optimization problems, here we are facing 'yes or no' decisions regarding whether each component of the feature vector should be included. The flowchart of our GA approach is shown in Figure 7.1. The basic idea is that, the initial feature vector is represented by a array of n ones and zeros, indicating whether a particular

component is, or is not, included. Based on this array, we can convert our initial feature vector to a reduced one which we will use to run the machine learning prediction. If the prediction accuracy is the same or gets increased, it means that the eliminated components are not essential for a particular property; if the accuracy decreases, we probably lose some important components in the feature vector. To search for the best reduced feature vector, we need to go through the GA loop. As shown in Figure 7.1, first we need to create a random initial generation, and each one of the arrays of ones and zeros is used to evaluate the predction accuracy. Then the new generation is created by selecting and reproducing, namely, taking the initial generation as the parents and reproducing children by mutation and crossover. If the highest accuracy in the new generation is equal or higher than the set accuracy, the GA stops; if not, the next generation will be created and the loop goes on.

For the organic polymers, we construct the feature vectors using all possible i units, $i-j$ pairs, and $i-j-k$ triples, which lead to 162 features in the feature vector. Of the total 267 organic systems, 200 were classified to be in the 'training' set (used in the training of the KRR model), and the remainder in the 'test' set. As shown in Figure 7.2, if considering all 162 features in the machine learning scheme, the prediction of band gap is good compared with the DFT results. But for electronic and total dielectric constant, the machine learning predicted values have a big discrepancy with those computed from DFT, which probably indicate that some of the feature vector components are hurtful in predictions. After applying the GA scheme, the feature vector has been reduced. For band gap, 162 features are reduced

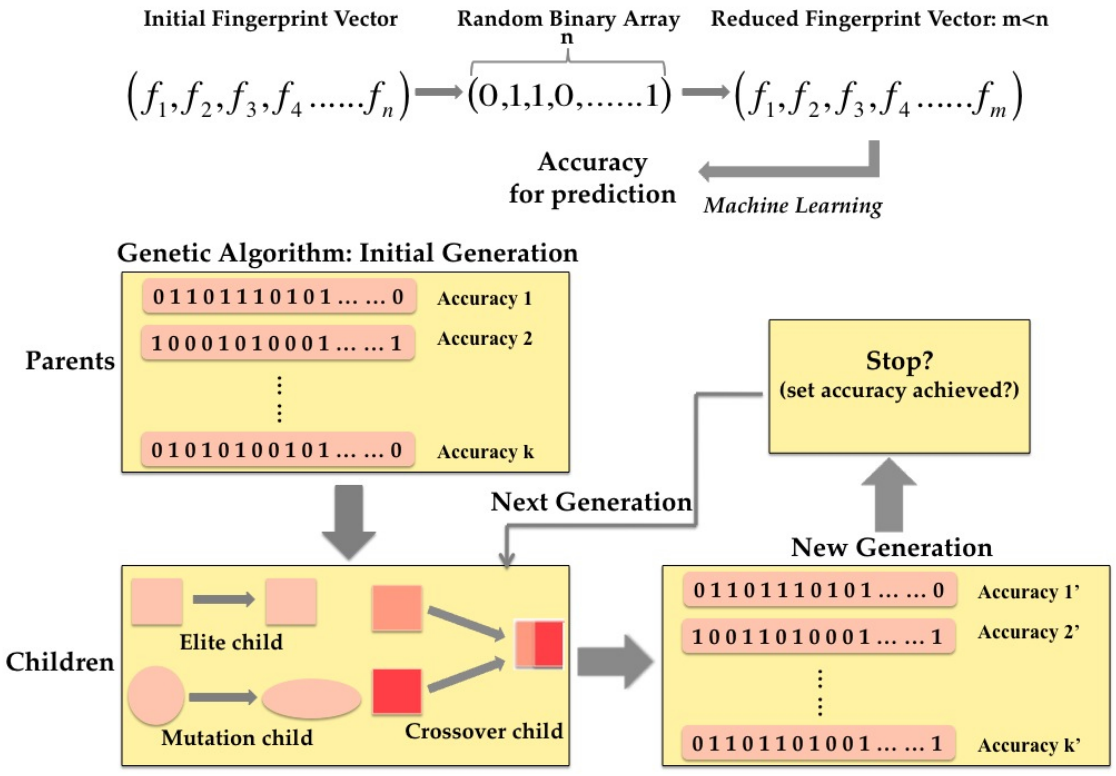


Figure 7.1: Flowchart of the genetic algorithm scheme for feature vector reduction.

to 78 and at the same time, the prediction accuracy increases to 94.25%. In the case of electronic dielectric constant, 82 features remain after applying the GA scheme and the machine learning predicted values are now in a much better agreement with those calculated from DFT with an accuracy of 90.98%. 162 features are reduced to 85 for total dielectric constant and the prediction accuracy increases from 64.00% to 90.55%. Overall, with the GA scheme, the features are reduced and the accuracy increases.

7.3 Larger scale chemical space exploration of new organic polymers

With the reduced features, the machine learning scheme is able to provide good predictions for band gap, electronic and total dielectric constant. Thus, it can be

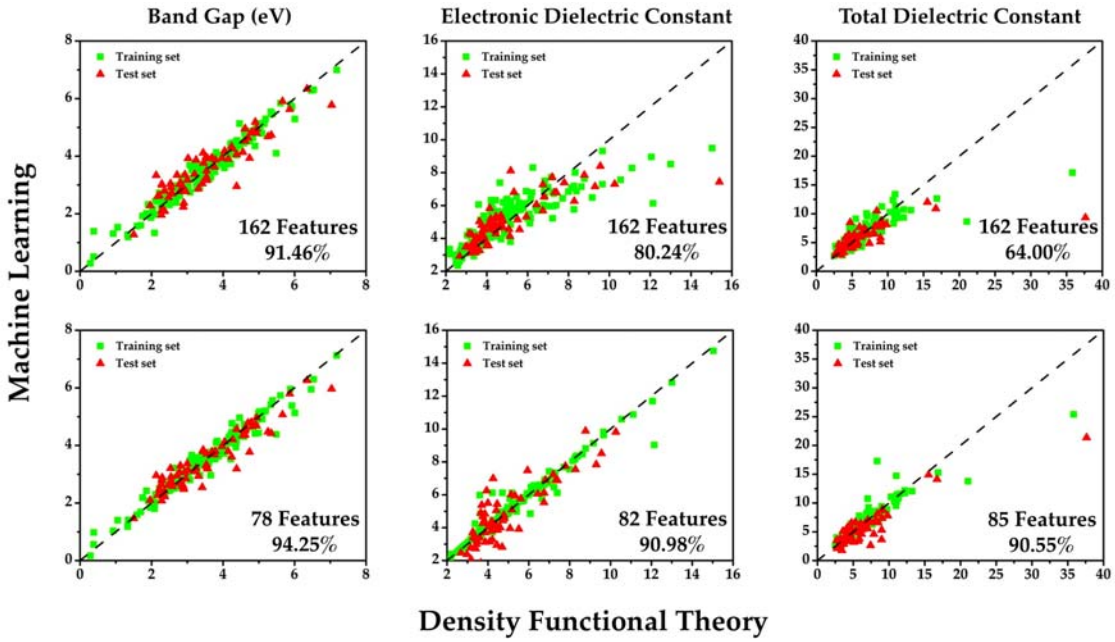


Figure 7.2: Parity plots comparing property values computed using DFT against predictions made using learning algorithms trained using chemo-structural feature vectors. Pearson's correlation index is indicated in each of the panels to quantify the agreement between the two schemes.

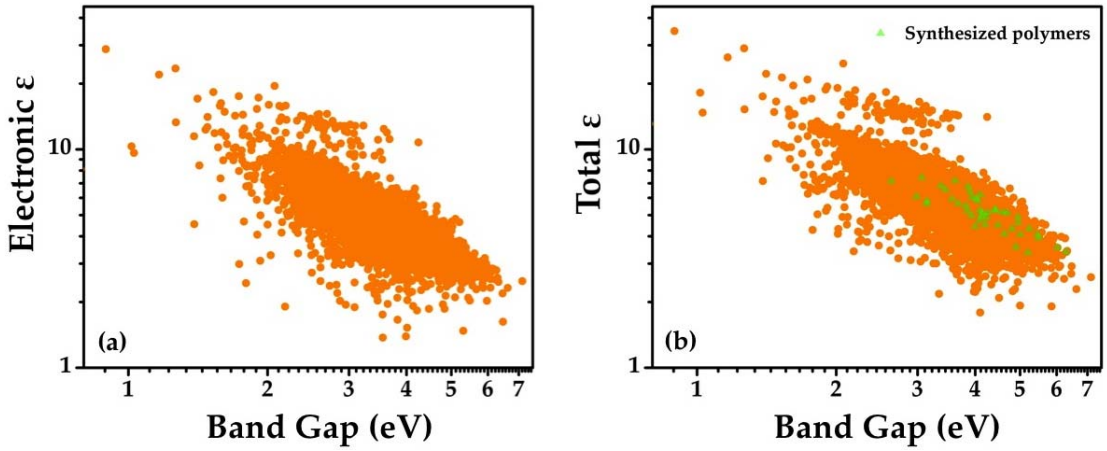


Figure 7.3: Machine learning predicted (a) electronic and (b) total dielectric constant as a function of the band gap. The green triangle represents the systems that are identified to be already synthesized experimentally.

used to study a much larger number of organic systems. In this study, organic polymers with 6 independent blocks are considered. Accounting for translational and inversion symmetries, and removing unstable systems, leaves 5,507 symmetry-unique systems. The machine learning scheme is then used to predict the band gap, electronic and total dielectric constant of these 5,507 organic systems, and the results are shown in Figure 7.3. From Figure 7.3(a), the approximate inverse relationship between the band gap and the electronic dielectric constant can still be observed. It is also noted that an inverse correlation is seen between the total dielectric constant and the band gap due to the fact that for organic systems, the total dielectric constant is dominated by the electronic contribution. This extensive search then allows us to identify polymers with suitable dielectric constant and band gap values, e.g., in the case of capacitive energy storage applications, high dielectric constant and large band gap.

Before exploring completely new systems, we should first examine those organic polymers that have already been synthesized. So far, 44 systems have been identified to be experimentally synthesized, and are represented in Figure 7.3(b). The complete list of these 44 systems is shown in Table 7.1 in decreasing order of total dielectric constant. The top few polymers are predicted to have moderate band gap and large dielectric constant values. These polymers may be considered as promising systems for capacitor and electronic applications and thus should undergo more detailed computional and experimental studies. Within these polymers, kevlar (NH-C_6H_4-NH-CO-C_6H_4-CO) is an commercial polymer which has many applications

from bicycle tires to racing sails. It has a dielectric constant of 3.5-4.5 (machine learning predicted value is 4.9). Another interesting polymer is NH-CS-NH-C_6H_4-CH_2-C_6H_4, polythiourea [12]. It is an recently synthesized polymer, which is shown to have high breakdown strength, low dielectric loss, and high energy density. The experimental dielectric constant of polythiourea is around 4.5.

Table 7.1: Already synthesized polymers. The band gap (E_g), electronic (ϵ_e) and total (ϵ_t) dielectric constant are machine learning predicted values.

System repeat unit	Polymer class	E_g (eV)	ϵ_e	ϵ_t
NH-C_6H_4-C_4H_2S-C_6H_4-NH-CS	Polyurea	3.1	5.8	7.4
NH-CO-NH-C_6H_4-CO-C_6H_4	Polyketone, polyurea	3.6	4.9	7.2
NH-C_6H_4-NH-CS-C_6H_4-CS	Polyamide	2.6	5.8	7.1
NH-C_6H_4-NH-C_6H_4-CO-C_6H_4	Polyketone, polyamine	3.4	6.2	6.8
NH-CO-C_4H_2S-CO-NH-C_6H_4	Polyamide	3.9	6.3	6.7
NH-C_6H_4-CO-C_6H_4-CO-C_6H_4	Polyamine, polyketone	3.4	6.5	6.5
NH-CO-NH-C_6H_4-O-C_6H_4	Polyether, polyurea	3.9	5.2	6.4
CH_2-C_4H_2S-CO-O-C_6H_4-O	Polyester	4.1	4.2	6.2

NH-CO-C_6H_4-O-CO-C_6H_4	Polyester, polyamide	3.9	4.2	6.0
C_6H_4-C_6H_4-C_6H_4-C_6H_4-C_6H_4-O	Polyether	3.0	5.6	6.0
CO-C_6H_4-O-C_6H_4-CO-O	Polyanhydride, polyester, polyether	4.0	4.6	5.9
NH-CO-C_6H_4-O-C_6H_4-C_6H_4	Polyamide, polyether	3.6	5.9	5.9
CO-C_6H_4-O-CO-O-C_6H_4	Polyketone, polycarbonate	4.0	4.1	5.8
CO-C_6H_4-C_6H_4-C_6H_4-O-C_6H_4	Polyketone, polyether	3.1	5.5	5.8
CO-C_6H_4-CO-C_6H_4-O-C_6H_4	Polyketone, polyether	3.7	5.6	5.7
CO-C_6H_4-C_6H_4-O-C_6H_4-C_6H_4	Polyketone, polyether	3.1	5.6	5.7
CO-C_6H_4-O-C_6H_4-O-C_6H_4	Polyether, polyketone	3.8	5.5	5.6
CH_2-C_6H_4-O-CS-O-C_6H_4	Polycarbonate	4.4	4.1	5.3
CH_2-C_6H_4-NH-CO-NH-C_6H_4	Polyurea	4.4	4.6	5.3

CH_2-NH-C_6H_4-O-C_6H_4-CO	Polyamine, polyether, polyketone	3.8	5.0	5.2
NH-CS-NH-C_6H_4-CH_2-C_6H_4	Polyurea	4.1	4.7	5.2
CH_2-NH-CO-C_6H_4-NH-CO	Polyamide	4.6	3.7	5.1
CH_2-C_6H_4-O-CO-O-C_6H_4	Polycarbonate	4.7	3.6	5.1
CH_2-C_6H_4-CO-C_4H_2S-CO-C_6H_4	Polyketone	4.1	4.6	5.1
CH_2-O-C_6H_4-CO-C_6H_4-O	Polyketone, polyether	4.3	3.8	5.1
CH_2-CO-C_6H_4-O-C_6H_4-CO	Polyether, polyketone	3.9	4.8	5.0
CH_2-C_6H_4-NH-C_6H_4-NH-C_6H_4	Polyamine	4.2	4.8	4.9
NH-C_6H_4-NH-CO-C_6H_4-CO (kevlar)	Polyketone, polyamide	4.1	4.6	4.9
CH_2-C_6H_4-O-C_6H_4-O-C_6H_4	Polyether	4.9	4.5	4.9
CO-C_6H_4-CO-C_6H_4-CO-C_6H_4	Polyketone	4.1	4.7	4.8
CH_2-C_6H_4-CH_2-C_6H_4-O-C_6H_4	Polyether	4.9	2.9	4.6
CH_2-CH_2-NH-CS-O-C_6H_4	Polyurethane	4.5	3.8	4.5
CH_2-C_6H_4-CO-C_6H_4-CO-C_6H_4	Polyketone	4.2	4.5	4.5
CH_2-C_6H_4-CH_2-C_6H_4-C_6H_4-C_6H_4	Polyother	4.0	3.7	4.4
CH_2-C_6H_4-CH_2-O-CO-O	Polycarbonate	5.2	3.6	4.3

CH_2-C_6H_4-CO-C_6H_4-CH_2-O	Polyketone, polyether	4.8	4.0	4.3
CH_2-C_6H_4-CH_2-O-C_6H_4-O	Polyether	5.0	3.2	4.1
CH_2-CH_2-CH_2-NH-CS-NH	Polyurea	4.6	3.2	4.1
CH_2-CO-CH_2-O-CO-O	Polycarbonate, polyketone	5.5	3.0	4.1
CH_2-NH-CO-NH-CO-NH	Polyurea	5.5	3.0	3.9
CH_2-CH_2-CH_2-NH-CO-O	Polyurethane	6.0	2.8	3.6
CH_2-CO-O-C_6H_4-O-CO	Polyester	4.9	3.1	3.6
CH_2-CH_2-O-CH_2-C_6H_4-O	Polyether	5.2	3.4	3.5
CH_2-O-CH_2-O-CH_2-O	Polyether	6.3	2.6	3.4

7.4 Conclusion

In summary, this part of the work involves the development of a feature selection approach based on genetic algorithm and kernel ridge regression. For organic polymers, we first used the feature selection approach to determine the proper components of the feature vector that most control a certain property. Trained machine learning models then predict the band gap and dielectric constant of a large number of new organic polymers with six blocks in the repeat unit at negligible cost with

high fidelity. The larger scale polymer chemical space exploration allows us to identify promising polymer systems with both high dielectric constant and large band gap.

Chapter 8

Summary and future work

8.1 Summary

In this thesis, first principles computations and machine learning approach are employed to the design of polymer dielectrics with both high dielectric constant and large band gap. Specifically, we adopt two strategies, (1) functionalization of a well understood polymer dielectrics, such as PE and PP, to enhance its dielectric response, and (2) discovery of entirely new classes of polymer dielectrics, both organic and organometallic. Approaches include exploration of large chemical spaces and efficient computation of some relevant properties.

Although accurate dielectric properties of materials are readily accessible today through DFT, such a conventional approach requires an knowledge of the appropriate crystal structure for the materials to be investigated. For polymer systems, determination of a global minimum on the potential energy surface for each of the new system may not be a practically feasible solution. Towards this end, we develop

a new method, the single-chain approach, to rapidly estimate the dielectric constant of polymers, which has paved the way for all other work in this thesis.

In the first strategy, our first principles calculations indicate that introduction of -OH functional groups will enhance the interaction between PE chains due to inter-chain hydrogen bonding. Moreover, when -OH groups are present, water molecules tend to be trapped due to the strong interaction between water and the -OH groups functionalized PE chains. The existence of the -OH groups and water molecules in the system contribute cooperatively to the increase of the dielectric constant of PE. Our DFT results are also consistent with MD simulations and experimental results qualitatively. These are interesting findings, as they point towards a rational pathway for the tunable control of dielectric properties of polymers via creative utilization of molecular functionalization (and moisture).

In order to discover new polymer dielectrics with both high dielectric constant and large band gap, we develop a high throughput DFT method (with the aid of single-chain approach) to explore the polymer chemical space based on two properties, band gap and dielectric constant. Two classes of polymers are explored, new organic polymers and Group 14 element-based hybrid polymers. The results provide us with a "map" of the achievable combination of properties within the chemical space explored. For example, in the case of capacitive energy storage applications, the polymer systems with large dielectric constant and moderate band gap can lead to useful structure information for synthesis efforts. Indeed, our computed results have guided some interesting experimental results.

DFT-based approach are limited to polymers based on short, periodic chains. For longer chains required to screen larger regions of compositional space, the computational cost associated with DFT rises rapidly. Furthermore, as the system size increases, the number of candidates within the system grows exponentially, which leads to combinatorial explosion. One of the important findings of this work is the development of machine learning approaches to explore a much larger polymer chemical space. These machine learning approaches help us predict the band gap and dielectric constant of a new polymer in a minuscule fraction of the time necessitated by a typical DFT computation. With the aid of these approaches, the discovery of new polymer dielectrics can be accelerated significantly.

8.2 Future work

8.2.1 Detailed studies of promising systems

This thesis can be considered as a paving stone for the future work, as it provides methodologies to explore a large amount of polymer systems, from which possible candidates can be chosen. Only a small portion of the promising systems have been studied in this thesis. There are still a lot more that need to be carefully examined. The crystal structures and morphologies of these promising systems must be investigated. If inter-atomic potentials are available to handle the systems, MD simulations can be used to determine their crystal structure and morphology, for example, using a melt and quench approach. An additional option, especially if force fields are not available for the new identified systems, is to use 3-dimensional structure searching

schemes to determine the ground state structures based on DFT, *e.g.*, evolutionary algorithm [135–137], minima hopping [138, 139], simulated annealing [140], *etc.*

Once the crystal structure and morphology are determined for the promising polymers, relevant properties for dielectric applications should be calculated. The work in this thesis deals largely with dielectric constant and band gap. While these two properties are important, as mentioned in the Introduction, many other factors are relevant to practical applications. Then down-selection of promising candidates and real design of new polymer dielectrics require estimating a range of other parameters. In the following, some possible properties, which can be studied in details in the future, are listed.

1. *Electron/hole injection barriers at metal-polymer interfaces* Electronic and capacitor applications require the polymer dielectric to contact a metal electrode. Such an interface sometimes dominates the performance of polymer dielectrics, as charge injection from the metal electrode to the polymer is the primary source of charge during conduction. The property that determines charge injection at interfaces is the Schottky barrier heights for electron and hole injection, which can be computed using modern DFT. Such issues have been encountered in past studies of metal-dielectric or metal-semiconductor interfaces within the context of (inorganic) semiconductor devices [1, 2, 154, 155]. While methods developed earlier can be used within the context of metal-polymer interfaces, the primary impediment to doing so is the complex nature of the metal-polymer interface at the atomic-level, which is far from ideal or

"abrupt", unlike interfaces typically encountered by the semiconductor community. Future efforts to compute barrier heights must break new ground in terms of realistic models of interfaces, perhaps through a combination of DFT and force-field based computations.

2. *Dielectric breakdown* The dielectric breakdown field of a capacitor dielectric is correlated strongly with its energy density (i.e., its energy storage capacity). Recently an approach to treat intrinsic breakdown using DFT in the average electron model of Von Hippel has been developed [70, 71]. But extrinsic factors, e.g., chemical impurities, nano-sized cavities, etc., are important in determining breakdown field for polymer-based dielectrics, and need to be taken into account.

3. *Glass transition temperature* Many properties change dramatically when an amorphous polymer undergoes glass transition. For instance, the dielectric loss of a polar material normally increases dramatically as the polymer goes through its T_g. At temperatures below T_g, polymers are hard and glassy because motion of polymer chains is restricted to local vibration. When the temperature goes to above T_g, chains have greater mobility and the polymer softens. Various factors affect the T_g of polymers, e.g., molecular weight and the chemical structure.

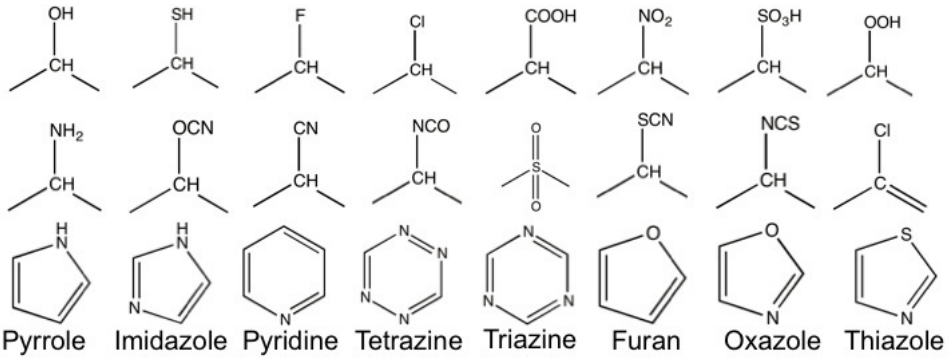


Figure 8.1: Possible motifs for expanding the chemical space exploration.

8.2.2 Expand search space

The chemical space of two different classes of polymers, namely, organic polymers and hybrid polymers, has been explored. Seven units, CH_2, NH, C(=O), C_6H_4 (benzene), C_4H_2S (thiophene), C(=S), and O, were used for organic polymers exploration, and to search the Group IV element-based hybrid polymers, units CH_2, SiF_2, $SiCl_2$, GeF_2, $GeCl_2$, SnF_2 and $SnCl_2$ were considered. More units can be then identified to expand the chemical space exploration. Figure 8.1 lists a few units that can be considered. For organic polymers, it can be found that the total dielectric constants of these polymers (*c.f.* Chapter 4) are mostly dominated by the electronic contribution which is inversely related to the band gap. It means that the dielectric constant values that one can achieve within this set of polymer are limited. Thus in order to realize higher dielectric constant values, we need to explore polymers which result in combinations of improved ionic and electronic dielectric constant. The top two rows in Figure 8.1 are units containing dipoles. The stretching or rotation of these dipoles may lead to increased ionic contribution to dielectric constant. The

units in the last row are different aromatic groups. How these heterocycles with localized conjugation in the presence of other functionalities (e.g., NH, CO) will affect the dielectric constant can then be investigated.

As more units are going to be considered, there will be a combinatorial explosion of the number of systems that one needs to study. Thus, the search strategy adopted in this thesis can be extremely time consuming. After defining the desired properties and selecting the polymer classes, we use high throughput DFT calculations for the initial screening, where all the possible polymer systems in the single chain model are subjected to band gap and dielectric constant calculations. With more and more units being considered, this step can be a curse. To relieve this curse, one way is to use a so-called guided search approach.

With the single chain model, what we are trying to do is fixing the configuration of a polymer system and varying the composition. The goal is to find compositions which correspond to high dielectric constant and large band gap. This can actually be generalized into a combinatorial optimization problem (combinatorial optimization is a topic that consists of finding an optimal object from a finite set of objects) [156], maximize both the dielectric constant and band gap values by changing the composition. By applying appropriate combinatorial optimization algorithms, such as genetic algorithm, greedy randomized adaptive search procedure [157], *etc.*, the number of calculations one need to do can be significantly reduced. The challenge of this guided search approach is the appropriate choice of combinatorial optimization algorithms.

Bibliography

[1] Robertson J, *Rep Prog Phys* 2006; 69:327-396.

[2] Zhu H, Tang C, Fonseca RC, Ramprasad R, *J Mater Sci* 2012; 47:7399-7416.

[3] Soltis FG, *Fortress Rochester: The Inside Story of the IBM ISeries* 2001; 29th Street Press.

[4] Less KJ, Wilson EG, *J Phys C: Solid State Phys* 1973; 6:3110.

[5] Ohki Y, Fuse N, Arai T, *In Proceedings of 2010 Annual Report on the Electrical Insulation and Dielectric Phenomena Conference* West Lafayette, LA, USA, October 2010; pp. 1-4.

[6] Rabuffi M, Picci G, *IEEE Trans Plasma Sci* 2002; 30:1939.

[7] Sarjeant WJ, Zirnheld J, MacDougall FW, *IEEE Trans Plasma Sci* 1998; 108:1368-1392.

[8] Ho J, Jow R. Characterization of high temperature polymer thin films for power conditioning capacitors. Army Research Laboratories; Jul. 2009. ARL-TR-4880.

[9] Zhu L, Wang Q, *Macromolecules* 2012; 45:2937-2954.

[10] Chu B, Zhou X, Ren K, Neese B, Lin M, Wang Q, Bauer F, Zhang QM, *Science* 2006; 313:334-336.

[11] Wang Y, Zhou X, Chen Q, Chu BJ, Zhang QM, *IEEE Trans Dielectr Electr Insul* 2010; 17:1036-1042.

[12] Wu S, Li WP, Li MR, Burlingame Q, Chen Q, Payzant A, Xiao K, Zhang QM, *Adv Mater* 2013; 25:1734-1738.

[13] Yoon M-H, Yan H, Facchetti A, Marks TJ, *J Am Chem Soc.* 2005; 127:10388-10395.

[14] Bertolazzi S, W*ü*nsche J, Cicoira F, Santato C, *Appl Phys Lett.* 2011; 99:013301-3.

[15] Roberts ME, Queralt*ó* N, Mannsfeld SCB, Reinecke BN, Knoll W, Bao Z, *Chem Mater* 2009; 21:2292-2299.

[16] Facchetti A, *Chem Mater* 2011; 23:733-758.

[17] Dang MT, Hirsch L, Wantz G, *Adv Mater* 2011; 23:3597-3602.

[18] Huo LJ, Zhang S, Guo X, Xu F, Li YF, Hou JH, *Angew Chem* 2011; 123:9871-9876.

[19] Manzione LT. Plastic packaging of microelectronic devices. New York: Van Nostrand, 1990.

[20] Tummala RR, Rymaszewski EJ. Microelectronic packaging handbook. New York: Van Nostrand Reinhold, 1997 (chap. 8).

[21] Rimdusit S, Ishida H, *Polymer* 2000; 14:7941-7949.

[22] Töpper M, Fischer T, Baumgartner T, Reichl H, *Electronic Components and Technology Conference* Las Vegas, NV, USA, June 2010.

[23] Hanley TL, Burford RP, Fleming RJ, Barber KW, *IEEE Electr Insul M* 2003; 19:13-24.

[24] Dadbin S, Frounchi M, Saeid MH, Gangi F, *J Appl Polym Sci* 2002; 86:1959-1969.

[25] Nalwa HS, Ed., *Handbook of Low and High Dielectric Constant Materials and Their Applications* Academic Press: San Diego, CA, 1999.

[26] Facchetti A, Yoon M-H, Marks TJ, *Adv Mater* 2005; 17:1705-1725.

[27] Ortiz RP, Facchetti A, Marks TJ, *Chem Rev* 2010; 110(1):205-239.

[28] Sun S-S, Sariciftci NS, *Organic Photovoltaics, Mechanisms, Materials, and Devices* CRC Press/Taylor & Francis 2005.

[29] Zhao XY, Liu HJ, *Polym Int* 2010; 59:597-606.

[30] Ennis J, MacDougall FW, Yang XH, Cooper RA, Seal K, Naruo C, et al., "Recent Advances in High Voltage, High Energy Capacitor Technology," 16th IEEE International Pulsed Power Conference, Albuquerque, NM, June 2007.

[31] Flynn PF, Meeting the Energy Needs of Future Warriors (National Academics, Washington, D. C., 2004).

[32] Ho J, Ramprasad R, Boggs S, *IEEE Trans. Dielectr. Electr. Insul.* 2007; 14:1295.

[33] Tu N, Kao K, *J Appl Phys* 1997; 85:7267.

[34] Barshaw EJ, White J, Chait MJ, Cornette JB, Bustamante J, Folli F, Biltchick D, Borelli G, Picci G, Rabuffi M, *IEEE Trans Magn* 2007; 43:223.

[35] Tortai JH, Bonifaci N, Denat A, Trassy CJ, *Appl Phys* 2005; 97:053304.

[36] Zhou X, Chu BJ, Neese B, Lin MR, Zhang QM, *IEEE Trans Dielectr Electr Insul* 2007; 14:1133.

[37] Zhou X, Zhao X, Suo Z, Zou C, Runt J, Liu S, Zhang S, Zhang QM, *Appl Phys Lett* 2009; 94:162901.

[38] Kim P, Doss NM, Tillotson JP, Hotchkiss PJ, Pan MJ, Marder SR, Li J, Calame JP, Perry JW, *ACS Nano* 2009; 3:2581.

[39] Dang ZM, Yuan JK, Zha JW, Zhou T, Li ST, Hu GH, *Prog Mater Sci* 2012; 57:660.

[40] Li J, Seok SI, Chu B, Dogan F, Zhang Q, Wang Q, *Adv Mater* 2009; 21:217.

[41] Narayanan Unni KN, Pandey AK, Nunzi JM, *Chem Phys Lett* 2005; 407:95-99.

[42] Kim C, Facchetti A, Marks TJ, *Science* 2007; 318:76-80.

[43] Klauk H, Halik M, Zschieschang U, Schmid G, Radlik W, Weber W, *J Appl Phys* 2002; 92:5259.

[44] Parashkov R, Becker E, Ginev G, Riedl T, Johannes HH, Kowalsky W, *J Appl Phys* 2004; 95:1594.

[45] McCullough RD, Lowe RDJ, *Chem Soc Chem Commun* 1992; 70.

[46] Chen TA, Rieke RD, *J Am Chem Soc* 1992; 114:10087.

[47] Loewe RS, Khersonsky SM, McCullough RD, *Adv Mater* 1999; 3:250.

[48] Blouin N, Michaud A, Leclerc M, *Adv Mater* 2007; 19:2295.

[49] Ballantyne AM, Chen L, Nelson J, Bradley DDC, Astuti Y, Maurano A, Shuttle CG, Durrant JR, Heeney M, Duffy W, McCulloch I, *Adv Mater* 2007; 19:4544-4547.

[50] Zhu Z, Waller D, Gaudiana R, Morana M, Muhlbacher D, Scharber M, Brabec C, *Macromolecules* 2007; 40:1981.

[51] Hou JH, Chen HY, Zhang SQ, Li G, Yang Y, *J Am Chem Soc* 2008; 130:16144.

[52] Hautier G, Jain A, Ong SP, *J Mater Sci* 2012; 47:7317-7340.

[53] Castelli IE, Olsen T, Datta S, Landis DD, Dahl S, Thygesen KS, Jacobsen KW, *Energy Environ Sci* 2012; 5:5814-5819.

[54] Yang K, Setyawan W, Wang S, Buongiorno Nardelli M, Curtarolo S, *Nat Mater* 2012; 11:614-619.

[55] Curtarolo S, Hart GLW, Buongiorno Nardelli M, Mingo N, Sanvito S, Levy O, *Nat Mater* 2013; 12:191-201.

[56] Olivares-Amaya R, Amador-Bedolla C, Hachmann J, Atahan-Evrenk S, Sanchez-Carrera RS, Vogt L, Aspuru-Guzik A, *Energy Environ Sci* 2011; 4:4849-4861.

[57] Hachmann J, Olivares-Amaya R, Jinich A, Appleton AL, Blood-Forsythe MA, Seress LR, Román-Salgado C, Trepte K, Atahan-Evrenk S, Er S, Shrestha S, Mondal R, Sokolov A, Bao ZA, Aspuru-Guzik A, *Energy Environ Sci* 2013; Advance Article.

[58] Ortiz C, Eriksson O, Klintenberg M, *Comput Mater Sci* 2009; 113 (12):4932-4939.

[59] Andersson MP, Bligaard T, Kustov A, Larsen KE, Greeley J, Johannessen T, Christensen CH, Nørskov JK, *J Catal* 2006; 239:501-506.

[60] Besenbacher F, Chorkendorff I, Clausen BS, Hammer B, Molenbroek AM, Nørskov JK, Stensgaard I, *Science* 1998; 279:1913-1915.

[61] Martin, RL, Haranczyk M, *J Chem Theory Comput* 2013; 9 (6):2816-2825.

[62] Hohenberg P, Kohn W, *Phys Rev* 1964; 136:B864-B871.

[63] Kohn W, Sham L, *Phys Rev* 1965; 140:A1133-A1138.

[64] Jones RO, Gunnarsson O, *Rev Mod Phys* 1989; 61:689-746.

[65] Baroni S, de Gironcoli S, Dal Corso A, Giannozzi P, *Rev Mod Phys* 2001; 73:515-562.

[66] Sharma V, Pilania G, Rossetti GA, Slenes K, Ramprasad R, *Phys Rev B* 2013; 87:134109.

[67] Wang CC, Ramprasad R, *J Mater Sci* 2011; 46:90-93.

[68] Wang CC, Pilania G, Ramprasad R, *Phys Rev B* 2013; 87:035103.

[69] Pilania G, Wang CC, Wu K, Sukumar N, Breneman C, Sotzing G, Ramprasad R, *J Chem Inf Model* 2013; 53(4):879-886.

[70] Sun Y, Boggs SA, Ramprasad R, *Appl Phys Lett* 2012; 101:132906.

[71] Sun Y, Bealing C, Boggs SA, Ramprasad R, *IEEE Electr Insul M* 2013; 29:8-15.

[72] Ranjan V, Buongiorno-Nardelli M, Bernholc J, *Phys Rev Lett* 2012; 108:087802.

[73] Ranjan V, Yu L, Buongiorno-Nardelli M, Bernholc J, *Phys Rev Lett* 2007; 99:047801.

[74] Huzayyin A, Boggs SA, Ramprasad R, *IEEE Electr Insul M* 2012; 28:23-29.

[75] Huzayyin A, Boggs SA, Ramprasad R, *IEEE Trans Dielectr Electr Insul* 2010; 17:920-925.

[76] Huzayyin A, Boggs SA, Ramprasad R, *IEEE Trans Dielectr Electr Insul* 2010; 17:926-930.

[77] Huzayyin A, Boggs SA, Ramprasad R, *Annual Report of the Conference on Electrical Insulation and Dielectric Phenomena* 2010; pp. 1-4.

[78] Jorgensen WL, Ulmschneider JP, Tirado-Rives J, *J Phys Chem B* 2004; 108:16264-16270.

[79] Wang JM, Wolf RM, Caldwell JW, Kollman PA, Case DA, *J Comp Chem* 2004; 25:1157-1174.

[80] Vanommeslaeghe K, MacKerell AD, Jr., *J Chem Inf Model* 2012; 52:3144-3154.

[81] Vanommeslaeghe K, Prabhu Raman E, MacKerell AD, Jr., *J Chem Inf Model* 2012; 52:3155-3168.

[82] Sun H, *J Phys Chem B* 1998; 102:7338-7364.

[83] Rigby D, Roe R, *J Chem Phys* 1987; 87:7285-7292.

[84] Bitsanis I, Hadziioannou G, *J Chem Phys* 1990; 92:3827-3847.

[85] Bennemann C, Paul W, Binder K, *Phys Rev E* 1998; 57:843-851.

[86] Moreno M, Casalegno M, Raos G, Meille SV, Po R, *J Phys Chem B* 2010; 114(4):1591-1602.

[87] Pilania G, Wang CC, Jiang X, Rajasekaran S, Ramprasad R, *Sci Rep* 2013; 3:2810.

[88] Sukumar N, Krein M, Luo Q, Breneman C, *J Mater Sci* 2012; 47:7703-7715.

[89] Hughes LD, Palmer DS, Nigsch F, Mitchell JBO, *J Chem Inf Model* 2008; 48(1):220-232.

[90] Hart GLW, Blum V, Walorski MJ, Zunger A, *Nature Mater* 2005; 4:391-394.

[91] Fischer CC, Tibbetts KJ, Morgan D, Ceder G, *Nature Mater* 2006; 5:641-646.

[92] Saad Y, Gao D, Ngo T, Bobbitt S, Chelikowsky JR, Andreoni W, *Phys Rev B* 2012; 85:104104.

[93] Rupp M, Tkatchenko A, Muller K, von Lilienfeld OA, *Phys Rev Lett* 2012; 108:058301.

[94] Hautier G, Fisher CC, Jain A, Mueller T, Ceder G, *Chem Mater* 2010; 22:3762-3767.

[95] Hansen K, Montavon G, Biegler F, Fazli S, Rupp M, Scheffler M, Anatole von Lilienfeld O, Tkatchenko A, Müller K, *J Chem Theory Comput* 2013; 9(8):3404-3419.

[96] Hammett LP, *J Am Chem Soc* 1937; 59(1):96-103.

[97] Hansch C, Muir RM, Fujita T, Maloney PP, Geiger F, Streich M, *J Am Chem Soc* 1963; 85:2817-2824.

[98] Hansch C, Fujita T (eds), Classical and three-dimensional QSAR in agrochemistry. American Chemical Society symposium series. American Chemical Society, Washington, 1995.

[99] Taft RW, *J Am Chem Soc* 1952; 74:3120-3128.

[100] Katritzky AR, Lobanov VS, Karelson M, *Chem Soc Rev* 1995; 24:279-287.

[101] Martin R, *Electronic Structure: Basic Theory and Practical Methods* Cambridge University Press, New York, 2004.

[102] Pilania G, Zhu H, Ramprasad R, *Applications of modern density functional theory to surfaces and interfaces, in A Matter of Density: Exploring the Electron Density Concept in the Chemical, Biological, and Materials Sciences* Ed. N. Sukumar, Wiley, 2012.

[103] Krukau AV, Vydrov OA, Izmaylov AF, Scuseria GE, *J Chem Phys* 2006; 125:224106.

[104] Baroni S, Giannozzi P, Testa A, *Phys Rev Lett* 1987; 58:1861.

[105] Gonze X, *Phys Rev A* 1995; 52:1086.

[106] Gonze X, *Phys Rev A* 1995; 52:1096.

[107] Schlom DG, Guha S, Datta S, *MRS Bulletin* 2008; 33:1017.

[108] Choi JH, *et al*, *Materl Sci Engr R* 2011; 72:97.

[109] Choy TC, *Effective Medium Theory: Principles and Applications* (Oxford University Press, Oxford, 1999).

[110] Osborn JA, *Phys Rev* 1945; 67:351.

[111] Stoner EC, *Phil. Mag.* 1945; 36:803.

[112] Landau LD, Lifshitz EM, Pitaevskii LP, Electrodynamics of Continuous Media, 2nd ed., Course of Theoretical Physics Vol. 8 (Pergamon, Oxford, 1984).

[113] Asaki MLT, Redondo A, Zawodzinski TA, Taylor AJ, *J Chem Phys* 2002; 116:10377.

[114] Billmeyer FW, *J Appl Phys* 1957; 28:1114.

[115] Bicerano J, *Prediction of Polymer Properties*, 3rd ed. (Marcel Dekker, New York, 2002).

[116] Jones N, Quiz-playing computer system could revolutionize research. Nature News, doi:10.1038/news.2011.95 (2011).

[117] MacLeod N, Benfield M, Culverhouse P, *Nature* 2010; 467:154.

[118] Crutchfield JP, *Nature* 2012; 8:17.

[119] Chittka L, Dyer A, *Nature* 2012; 481:154.

[120] Abu-Mostafa YS, *Sci. Am.* 1995; 272:64.

[121] Abu-Mostafa YS, *Sci. Am.* 2012; 307:78.

[122] Bucholz EW, Kong CS, Marchman KR, Sawyer WG, Phillpot SR, Sinnott SB, Rajan K, *Tribol Lett* 2012; 47:211-221.

[123] Hautier G, Fischer C, Ehrlacher V, Jain A, Ceder G, *Inorg Chem* 2011; 50:656-663.

[124] Zakutayev A, Zhang X, Nagaraja A, Yu L, Lany S, Mason TO, Ginley DS, Zunger A, *J Am Chem Soc* 2013; 135:10048-10054.

[125] Dey P, Bible J, Datta S, Broderick S, Jasinski J, Sunkara M, Menon M, Rajan K, *Computational Materials Science* 2014; 83:185-195.

[126] Yuan X, Matsuyama Y, Chung TCM, *Macromolecules* 2010; 43:4011-4014.

[127] Chung TCM, *Green and Sustainable Chemistry* 2012; 2:29-37.

[128] Gupta S, Yuan X, Chung TCM, Kumar S, Weiss RA, *Macromolecules* 2013; 46:5455-5463.

[129] Gonze X, and Lee C, *Phys. Rev. B* 1997; 55:10355.

[130] Gonze X, *Phys. Rev. B* 1997; 55:10337.

[131] Giannozzi P, and Baroni S, *J. Chem. Phys.* 1994; 100:8537.

[132] Porezag D, and Pederson MR, *Phys. Rev. B* 1996; 54:7830.

[133] Misra M, Agarwal M, Sinkovits DW, Kumar S, Wang CC, Pilania G, Ramprasad R, Weiss RA, Yuan X, Chung TCM, *Macromolecules* 2014; 47:1122.

[134] Baldwin AF, Ma R, Wang CC, Ramprasad R, Sotzing GA, *J Appl Polym Sci* 2013; 130:1276-1280.

[135] Oganov AR, Glass CW, *J Chem Phys* 2006; 124:244704.

[136] Lyakhov AO, Oganov AR, Valle M, *Comput Phys Comm* 2010; 181:1623-1632.

[137] Oganov AR (Editor). *Modern Methods of Crystal Structure Prediction*. Wiley-VCH, Berlin. ISBN: 978-3-527-40939-6, 2010.

[138] Goedecker S, *J Chem Phys* 2004; 120:9911.

[139] Amsler M, Goedecker S, *J Chem Phys* 2010; 133:224104.

[140] Pannetier J, Bassas-Alsina J, Rodriguez-Carvajal J, Caignaer V, *Nature* 1990; 346:343-345.

[141] Banihashemi A, Hazarkhani H, Abdolmaleki A, *J Polym Sci Part A: Polym Chem* 2004; 42:2106-2111.

[142] Jenekhe SA, *Macromolecules* 1990; 23:2848-2854.

[143] Kishimoto T, Uraki Y, Ubukata M, *Org Biomol Chem* 2005; 3:1067-1073.

[144] Lorenzini RG, Kline WM, Wang CC, Ramprasad R, Sotzing GA, *Polymer* 2013; 54:3529-3533.

[145] Hastie T, Tibshirani R, Friedman J, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, (Springer, New York, USA, 2009).

[146] Muller K-R, Mika S, Ratsch G, Tsuda K, Scholkopf B, *IEEE Trans Neural Netw* 2001; 12:181-201.

[147] Geerlings P, Boon G, Van Alsenoy C, De Proft F, *Int. J Quantum Chem* 2005; 101:722-732.

[148] Mezey PG, *J Chem Inf Comput Sci* 1999; 39:224-230.

[149] Bader RFW, *Atoms in molecules: a quantum theory*, (Oxford University Press, Oxford, UK, 1990).

[150] Bultinck P, Girones X, Carbo-Dorca R, *Rev Comp Ch*, 2005; Vol 21.

[151] Vafaie H, De Jong K, *Fourth International Conference on Tools with Artificial Intelligence* 1992; IEEE Computer Society Press, pp. 200-203.

[152] Paszkowicz W, *Materials and Manufacturing Processes* 2009; 24:174-197.

[153] Chua ALS, Benedek NA, Chen L, Finnis MW, Sutton AP, *Nature Mater* 2010; 9:418-422.

[154] Robertson J, *Solid-State Electron* 2005; 49:283-293.

[155] Zhu H, Ramprasad R, *Phys Rev B* 2011; 83:081416.

[156] Cook WJ, Cunningham WH, Pulleyblank WR, Schrijver A, *Combinatorial Optimization*, (Wiley-Interscience, 1997).

[157] Feo T, Resende MGC, *Journal of Global Optimization* 1995; 6:109-133.

Appendix A

The rational design of polyurea & polyurethane dielectric materials

ARTICLE IN PRESS

Polymer xxx (2013) 1–5

Contents lists available at SciVerse ScienceDirect

Polymer

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

The rational design of polyurea & polyurethane dielectric materials

R.G. Lorenzini [a], W.M. Kline [b], C.C. Wang [b], R. Ramprasad [b], G.A. Sotzing [a,c,*]

[a] Department of Chemistry & The Polymer Program, University of Connecticut, 97 North Eagleville Road, Storrs, CT 06268, United States
[b] Department of Materials Science & Engineering, University of Connecticut, 97 North Eagleville Road, Storrs, CT 06268, United States
[c] Department of Physics, University of Connecticut, 97 North Eagleville Road, Storrs, CT 06268, United States

ARTICLE INFO

Article history:
Received 18 January 2013
Received in revised form
28 April 2013
Accepted 1 May 2013
Available online xxx

Keywords:
Polymer dielectrics
Dielectric spectroscopy
Capacitors

ABSTRACT

When designing polymeric capacitors, it is important to understand the structure–property relationship between chemical functionalities and dielectric properties to tailor materials for specific applications in terms of dielectric constant, dielectric loss, band gap, breakdown strength, etc. Herein, we report a clear structure–property relationship with dielectric constant and dielectric loss in a series of polyurea and polyurethane thin films. We demonstrate that the dielectric constant systematically increases and the dielectric loss decreases as the number of carbons between polarizable functional groups decreases. Our syntheses are guided with data obtained from high-throughput density functional theory calculations. By modeling 382 polymer systems, we have determined that a dielectric constant >4 is achieved with at least one aromatic group and at least one of the following moieties: –NH–, –C(=O)–, or –O–.

1. Introduction

The development of high energy density capacitors is driven by their importance in various applications, including hybrid electric vehicles, the electric propulsion of ships [1] and electromagnetic railguns [2]. In terms of polymeric materials, biaxially oriented polypropylene (BOPP) is the current industrial standard. BOPP boasts a remarkably high electrical breakdown strength and large band gap, in addition to its ease of processability, high breakdown field and graceful failure [3]. BOPP has a dielectric constant at room temperature of ~2.2 across a broad frequency range and a dielectric loss in the neighborhood of 10^{-3}–10^{-4} [4]. To add to the existing body of work in this field [5–8], we aim to understand the structure–property relationship between chemical functionalities and dielectric properties, through the paradigm of rational capacitor design. Ultimately, this level of understanding should allow for the tailoring of materials for specific electronic applications.

In the present study, we consider a large number of polymer systems with different functionalities, with the intent of identifying new polymer dielectric materials that have better dielectric properties than BOPP without sacrificing its already remarkable insulating properties. In an attempt to identify promising polymer subclasses we use first principles computations based on density functional theory (DFT) to perform an initial combinatorial screening. This initial screening has helped us to identify organic building blocks such as aromatic rings, –NH–, –CO–, –O–, etc., (i.e., those that make up ureas and urethanes) as promising ones. Polymers containing these functionalities are synthesized and their dielectric properties are characterized and compared with the DFT results.

2. Computational guidance

As a first step, we used DFT computations in a high-throughput mode to identify promising cases that may lead to a high dielectric constant and high band gap; we note that the former property leads to a larger energy density, and the latter is an indicator of a good insulator (and high breakdown strength). Fig. 1 captures the computational model adopted in our strategy. In essence, we consider an all-trans single polymer chain containing four independent blocks with periodic boundary conditions along the chain axis. We allow each block in the polymer backbone to be one of following units: $–CH_2–$, –NH–, –C(=O)–, $–C_6H_4–$ (benzene), $–C_4H_2S–$ (thiophene), –C(=S)–, and –O–. Different combinations of these units form traditional polymers, including polyesters, polyamides, polyethers, polyureas, *etc*. This scheme results in 382 symmetry unique systems.

Density functional theory (DFT), as implemented in the Vienna ab initio simulation package (VASP) [9] was used to determine the structural and electronic properties of the 382 polymer chains. The Perdew, Burke, and Ernzerhof functional (PBE) [10,11], projector-

* Corresponding author. Department of Physics, University of Connecticut, 97 North Eagleville Road, Storrs, CT 06268, United States.
E-mail address: g.sotzing@uconn.edu (G.A. Sotzing).

0032-3861/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.polymer.2013.05.003

Please cite this article in press as: Lorenzini RG, et al., The rational design of polyurea & polyurethane dielectric materials, Polymer (2013), http://dx.doi.org/10.1016/j.polymer.2013.05.003

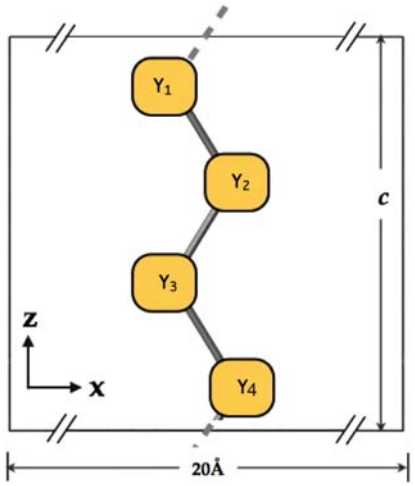


Fig. 1. Model showing our computational strategy, in which we substitute different functionalities into the four blocks. These calculations operate at the atomic scale, and therefore neglect interchain interactions.

augumented wave (PAW) frozen-core potentials [12] and a cutoff energy of 400 eV for the plane wave expansion of the wave-functions were used. The optimized geometry was then used to determine the electronic (high frequency, or optical) and total (low frequency, or static) dielectric constant tensor of the polymer chains density functional perturbation theory (DFPT) [13], followed by extrapolation of these results to correspond to the volume actually occupied in real polymeric systems using methods developed recently [14,15].

Fig. 2 shows the relationship between the electronic (Fig. 2(a)) and total (Fig. 2(b)) dielectric constant and the band gap for the 382 polymer systems. A near perfect inverse pareto-optimal front relationship between the band gap and the electronic dielectric constant can be seen from Fig. 2(a), which imposes a theoretical limit on the electronic part of the dielectric constant that one may be able to achieve (a limit that may be understood by regarding the electronic part of the dielectric response as a sum over electronic transitions from occupied to unoccupied states). Fig. 2(b) shows the variation of the total dielectric constant with the band gap. The total dielectric constant derives its contributions from the electronic and ionic contributions. The latter is, in principle, not constrained by the band gap, and is related to the dipole moment values of the functional groups and the flexibility of bonds that can allow the dipoles to respond to an electric field. From Fig. 2(b), a set of the most promising polymers may be selected based on the criteria that the band gap >3 eV, and total dielectric constant >4. This "promising" region is shown in Fig. 2(b), and corresponds to systems composed of at least one aromatic group and at least one of the following three groups: –NH–, –C(=O)–, –O–.

The DFT-recommended groups provided useful constraints for determining the best functionalities to incorporate into a polymer backbone. We understood that polymers synthesized for electronic applications should be as pure as possible; we sought to eliminate impurities that were both catalytic and inherent in nature. To this end, we ultimately decided on polyureas and polyurethanes derived from the step polymerization of an asymmetrically substituted aromatic diisocyanate and various diols and diamines (Fig. 3) – these systems were harmonious with the results of the DFT calculations, produced no byproducts, and required little or no catalyst.

3. Polymer synthesis

3.1. Materials

Reagents and solvent were purchased from Acros Organics or Sigma–Aldrich, with the exception of the 2,2′-oxybis-ethaneamine, which was purchased from Huntsman. Liquids were purified either by ambient pressure distillation or vacuum distillation and were stored under nitrogen.

3.2. Polyurea and polyurethane syntheses

To a nitrogen-filled, flame-dried 25 mL 3-neck flask equipped with a gas inlet, rubber and glass stoppers and a magnetic stirbar, 10 mL of dimethylsulfoxide was added via cannula needle. Next, 5 mmol of diamine/diol was cannulated; the materials used were: 1,2-diaminoethane, 1,2-diaminopropane, 1,3-diaminopropane, 1,6-diaminohexane, 2,2′-oxybis-ethanamine, 1,2-ethanediol, 1,2-

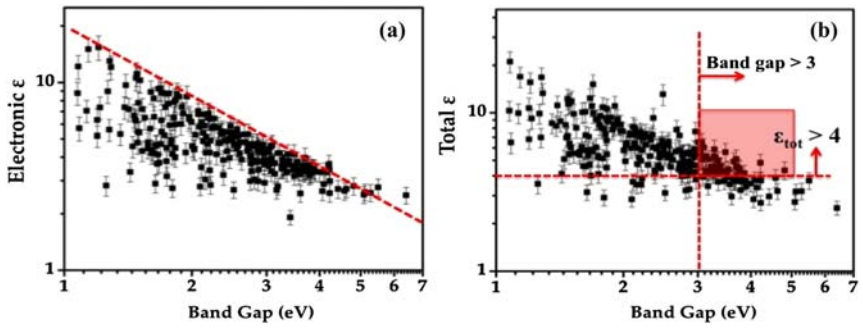


Fig. 2. DFT computed (a) electronic and (b) total dielectric constant along the polymer chain axis as a function of the band gap. The axes are in logarithmic scale.

Please cite this article in press as: Lorenzini RG, et al., The rational design of polyurea & polyurethane dielectric materials, Polymer (2013), http://dx.doi.org/10.1016/j.polymer.2013.05.003

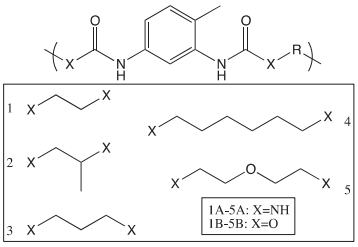


Fig. 3. Structures of synthesized polyureas and polyurethanes.

propanediol, 1,3-propanediol, 1,6-hexanediol, diethylene glycol. In the case of the diols, 1 wt% of dibutyltin dilaurate was added as a catalyst. Cannulation of 5 mmol of 2,4-toluene diisocyanate (TDI) initiated an exothermic event. When the temperature from the initial mixing subsided, an oil bath was used to heat the mixture to 60 °C for 2 h. The resultant polymers were precipitated in methanol, and were allowed to stir in the methanol for 1 h to remove any excess monomer/catalyst, after which they were filtered using a Büchner funnel and dried *in vacuo* for 24 h at 160 °C. This cleaning and heating protocol ensured complete removal of solvent, excess monomer and methanol-soluble oligomers, as confirmed by a GC/MS polymer desorption method. In all cases, the polymers were obtained in nearly quantitative yields.

3.3. Film preparation for TDDS

Polymer samples were prepared for films as follows: 5 wt% solutions of polymer/DMSO were prepared by heating, stirring and filtration through a 0.45 μm PTFE syringe filter. Polymer films were drop cast on stainless steel shims, and were left on the bench overnight to allow a bit of solvent to evaporate off, leaving a "tacky" film. These tacky samples were then heated in a vacuum oven for 160 °C overnight to ensure complete removal of solvent. During all steps, care was taken to prevent the introduction of dust from the laboratory, such as the use of petri dishes for solvent evaporation. Polymer films prepared in this way were determined to be 200–400 nm thick.

Table 1
The dielectric constant and dielectric loss of the polyureas at 60 Hz and 1 kHz at various temperatures.

Polyureas						
Polymer	Room temp.		50 °C		100 °C	
	ε	ε	ε	ε	ε	ε
	60 Hz	1 kHz	60 Hz	1 kHz	60 Hz	1 kHz
1A	5.24	5.18	5.31	5.23	6.63	5.85
2A	4.35	4.29	4.50	4.38	5.92	5.13
3A	3.54	3.47	3.55	3.48	3.45	3.38
4A	2.17	2.08	2.23	2.13	2.20	2.11
5A	6.72	6.19	7.03	6.44	7.22	6.87
Polymer	tan(δ)	tan(δ)	tan(δ)	tan(δ)	tan(δ)	tan(δ)
	60 Hz	1 kHz	60 Hz	1 kHz	60 Hz	1 kHz
1A	0.00683	0.00758	0.00883	0.00862	0.116	0.0539
2A	0.00941	0.00889	0.0179	0.0135	0.116	0.0675
3A	0.0117	0.0173	0.0130	0.0176	0.0141	0.0124
4A	0.0222	0.0312	0.0259	0.0323	0.0261	0.0242
5A	0.0342	0.0429	0.0322	0.0543	0.0266	0.0444

Table 2
The dielectric constant and dielectric loss of the polyurethanes at 60 Hz and 1 kHz at various temperatures.

Polyurethanes						
Polymer	Room temp.		50 °C		100 °C	
	ε	ε	ε	ε	ε	ε
	60 Hz	1 kHz	60 Hz	1 kHz	60 Hz	1 kHz
1B	6.47	6.35	6.71	6.52	9.75	8.43
2B	6.87	6.74	6.84	6.73	6.90	6.74
3B	5.96	5.81	4.72	4.54	6.37	5.63
4B	4.22	4.09	4.29	4.15	4.52	4.37
5B	10.8	10.5	11.0	10.6	13.8	12.0
Polymer	tan(δ)	tan(δ)	tan(δ)	tan(δ)	tan(δ)	tan(δ)
	60 Hz	1 kHz	60 Hz	1 kHz	60 Hz	1 kHz
1B	0.0113	0.0126	0.0178	0.0157	0.1190	0.0650
2B	0.00896	0.0154	0.00874	0.0136	0.0138	0.0140
3B	0.0158	0.0139	0.0211	0.0210	0.0980	0.0608
4B	0.0184	0.0156	0.0167	0.0180	0.0220	0.0220
5B	0.0151	0.0188	0.0168	0.0201	0.135	0.0615

3.4. TDDS spectroscopy

The dielectric spectra were obtained on an IMASS time domain dielectric spectrometer at the University of Connecticut Electrical Insulation Research Center. Measurements were taken at room temperature, 50 °C and 100 °C between silicone rubber guarded electrodes.

3.5. Polymer characterization

^{1}H NMR data was obtained on a Bruker DMX-500, and IR spectroscopy was performed on a Nicolet Magna-IR 560 spectrometer. Thermogravimetric analysis was conducted on a TA

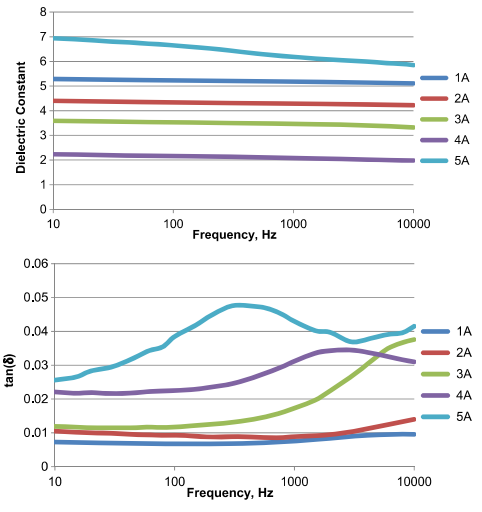


Fig. 4. Dielectric constant (top) and tan(δ) (bottom) at room temperature for the different polyureas.

Please cite this article in press as: Lorenzini RG, et al., The rational design of polyurea & polyurethane dielectric materials, Polymer (2013), http://dx.doi.org/10.1016/j.polymer.2013.05.003

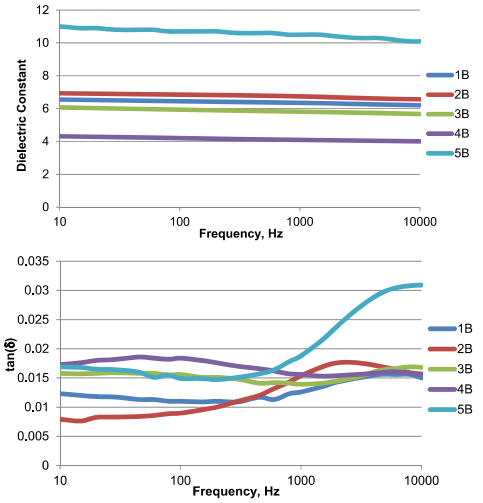


Fig. 5. Dielectric constant (top) and tan(δ) (bottom) at room temperature for the different polyurethanes.

Instruments TGA Q500 by heating 10 °C/min from room temperature to 500 °C in oxygen; the data presented represents the temperature at 3% weight loss. Differential scanning calorimetry was performed on a TA Instruments DSC Q100: first, the sample was heated at 10 °C/min to just before its weight loss onset temperature, and was held isothermally for 5 min. It was then cooled to −50 °C at 10 °C/min. Data was obtained on the next cycle, when the sample was heated to the 3% degradation temperature at 10 °C/min. Gel permeation chromatography was conducted using dimethylacetamide as the mobile phase and polystyrene standards with the following instrumentation: Waters 515 HPLC pump, Waters 2414 refractive index detector, two mixed bed Jordi Gel DVB columns. Samples were filtered twice prior to GPC analysis: first with a Kimwipe-plugged pipette to remove large insoluble materials, and second with a 0.45 μm PTFE syringe filter.

Table 3
Table containing the glass transition temperature (T_g), melting temperature (T_m) and 3% degradation temperature (T_d) for each of the sample polymers. All temperatures are reported in °C.

Polymer	T_g	T_m	T_d
Polyureas			
1A	N/O	242	246
2A	N/O	228	234
3A	N/O	234	227
4A	N/O	224	263
5A	N/O	201	264
Polyurethanes			
1B	130	N/O	237
2B	75	N/O	251
3B	72	215	251
4B	95	213	277
5B	91	167	249

Table 4
Table containing the number-average molecular weight (M_n), the weight-average molecular weight (M_w), and the polydispersity index (PDI). All molecular weights are reported in g/mol.

Polymer	M_n	M_w	PDI
Polyureas			
1A	Not soluble		
2A	Not soluble		
3A	37,500	54,400	1.45
4A	31,400	41,200	1.31
5A	29,100	39,500	1.36
Polyurethanes			
1B	24,300	37,800	1.56
2B	19,600	28,900	1.47
3B	28,700	36,400	1.27
4B	21,700	30,800	1.42
5B	37,900	49,400	1.30

4. Results & discussion

4.1. TDDS spectroscopy results

For all of the polymers, the dielectric constant and dielectric loss at 60 Hz and 1 kHz are tabulated (Tables 1 and 2). In addition, the room temperature dielectric spectra are overlaid (Figs. 4 and 5).

4.2. Structure–property relationship: polymer classes, alkyl length, dielectric constant & loss

The dielectric behavior of these materials can be explained through several paradigms. Most generally, polyurethanes have higher dielectric constants than polyureas, which is reasonable based on electronegativity arguments. These data suggest that for both classes of polymers, increasing the number of carbons in the backbone lowers the dielectric constant. This can be attributed to the increase in free volume, which decreases the number of polarizable groups per unit volume [16,17]. Inclusion of a backbone ether moiety (5a&b) drastically increases the dielectric constant in both cases, in agreement with literature values for polyether polyurethanes; more ether oxygen atoms present in the backbone result in higher dielectric constants [18]. The dielectric loss tan(δ) is also observed to follow the aforementioned trend: smaller alkyl spacers yield lower loss values. These conclusions are in agreement with generally accepted means of increasing dielectric constant, including maximizing polarizability and imparting a low degree of free volume [19].

4.3. Polymer characterization

Due to the structural similarity of the polymers being tested, differences in the NMR and IR spectra were difficult to discern. However, we used these data to both confirm the purity of the materials and the absence of reactive isocyanate end groups, which would manifest themselves in the IR as a sharp peak at ~2250 cm^{-1}. The TGA (3% degradation) and DSC data are presented below (Table 3), demonstrating high thermal stability well beyond the operating temperature of BOPP. The low polydispersity (Table 4) observed in these polymers likely has to do with the workup conditions — stirring in methanol removed some of the lower weight chains. A study comparing the dielectric constants and losses of polypropylene glycols with varying molecular weights shows that the dielectric properties saturate between 1000 and 2000 g/mol [20].

5. Conclusions & summary

This study was primarily intended to span the chemical space, searching for the paramount influences of structure on dielectric

Please cite this article in press as: Lorenzini RG, et al., The rational design of polyurea & polyurethane dielectric materials, Polymer (2013), http://dx.doi.org/10.1016/j.polymer.2013.05.003

constant. Our synthetic efforts were guided by DFT calculations, which revealed that polymers with high dielectric constants contain at least one aromatic group, and one of three functionalities: –NH–, –C(=O)–, or –O–; the majority of the synthesized polymers had dielectric constants >4 as predicted by said calculations. These dielectric data will be conglomerated with data from parallel synthetic efforts in our group, to be fed into a quantitative structure–activity relationship model (QSAR) to be published at a later date. Our ultimate aim is to investigate materials with more desirable electronic properties (higher dielectric constant, higher band gap, comparable loss) compared with the current industry standard BOPP, and we consider this work to be generally successful toward our goal. Most of the aforementioned materials can operate at or beyond BOPP's failure temperature, which successfully operates between 60 and 80 °C [21]. BOPP has a dielectric constant of 2.2; the dielectric constants reported herein are either equal to or higher than BOPP. BOPP has loss values in the 10^{-3}–10^{-4} regime, whereas our lowest loss polymer (1A) is towards the upper-end of this range ($6.8 * 10^{-3}$).

Acknowledgments

We would like to thank Ms. JoAnne Ronzello and Dr. Steven A. Boggs of the University of Connecticut Electrical Insulation Research Center for dielectric testing and helpful suggestions. In addition, we would like to express gratitude for financial support from the Office of Naval Research through the Multidisciplinary University Research Initiative (MURI).

References

[1] Flynn PF. Meeting the energy needs of future warriors. Washington D.C.: National Academics Press; 2004.
[2] Electromagnetic railgun. Office of naval research fact sheet. http://www.onr.navy.mil/Media-Center/Fact-Sheets/Electromagnetic-Railgun.aspx; Feb. 2012.
[3] Sarjeant WJ, Zirnheld J. Handbook of low and high dielectric constant materials and their applications. UK: Academic Press; 1999.
[4] Ho J, Jow R. Characterization of high temperature polymer thin films for power conditioning capacitors. Army Research Laboratories; Jul. 2009. ARL-TR-4880.
[5] Zhang Q, Zhu D, Su F, Xie Y, Ma Z, Shen J. J Biomed Mater Res A 2012;100A(7):1868–76.
[6] Wang Y, Zhou X, Chen Q, Chu B, Zhang QM. IEEE Trans Dielectr Electr Insul 2010;17(4):1036–42.
[7] Oprea S, Potolinca O, Oprea V. High Perform Polym 2011;23(1):49–58.
[8] Adjokatse SK, Mishra AK, Waghmare UV. Polymer 2012;53:2751–7.
[9] Kresse G, Furthmuller J. Phys Rev B 1996;54:11169.
[10] Perdew JP, Burke K, Ernzerhof M. Phys Rev Lett 1996;77:3865.
[11] Blochl PE. Phys Rev B 1994;50:17953.
[12] Kresse G, Joubert D. Phys Rev B 1999;59:1758.
[13] Baroni S, de Gironcoli Stefano, Dal Corso Andrea. Rev Mod Phys 2001;73:515.
[14] Pilania G, Ramprasad R. J Mater Sci 2012;47:7580.
[15] Wang CC, Pilania G, Ramprasad R. Submitted for publication.
[16] Simpson JO, St.Clair AK. Fundamental insight on developing low dielectric constant polyimides. Thin Solid Films 1997;308–309:480–5.
[17] Hwang H-J, Li C-H, Wang C-S. Polymer 2006;47:1291–9.
[18] North AM, Reid JC. Dielectric relaxation in a series of heterophase polyether polyurethanes. Eur Polym J 1972;8:1129–38.
[19] Deligoz H, Ozgumus S, Yalcinyuva T, Yildirim S, Deger D, Ulutas K. Polymer 2005;46:3720–9.
[20] Sarode AV, Kumbharkhane AC. J Mol Liq 2011;160:109–13.
[21] Zou C, Zhang Q, Zhang S, Kushner D, Zhou X, Bernard R, et al. PEN/Si_3N_4 bilayer film for DC bus capacitors in power converters in hybrid electric vehicles. J Vac Sci Technol B 2011;29:061401.

Please cite this article in press as: Lorenzini RG, et al., The rational design of polyurea & polyurethane dielectric materials, Polymer (2013), http://dx.doi.org/10.1016/j.polymer.2013.05.003

Appendix B

Structure-property relationship of polyimides based on pyromellitic dianhydride and short-chain aliphatic diamines for dielectric material applications

Applied Polymer SCIENCE

Structure-Property Relationship of Polyimides Based on Pyromellitic Dianhydride and Short-Chain Aliphatic Diamines for Dielectric Material Applications

Aaron F. Baldwin,[1] Rui Ma,[1] Chenchen Wang,[2] Rampi Ramprasad,[2] Gregory A. Sotzing[1,3]
[1]The Polymer Program, Institute of Materials Science, University of Connecticut, Storrs, Connecticut 06269
[2]Department of Chemical, Materials and Biomolecular Engineering, Institute of Materials Science, University of Connecticut, Storrs, Connecticut 06269
[3]Department of Chemistry, University of Connecticut, Storrs, Connecticut 06269
Correspondence to: G. A. Sotzing (E-mail: sotzing@mail.ims.uconn.edu)

ABSTRACT: Most polyolefins that are used for dielectric materials exhibit a low dielectric constant and operating temperatures up to 70°C. Polyimides offer a means to a higher dielectric constant material by the introduction of a polar group in the polymer backbone and are thermally stable at temperatures exceeding 250°C. A common dianhydride, pyromellitic dianhydride (PMDA), is reacted with various short-chain diamines to produce polymers with high imide density. Homopolymers and copolymers synthesized had dielectric constants ranging from 3.96 to 6.57. These materials exhibit a dielectric constant twice that of biaxially oriented polypropylene and therefore a twofold increase in capacitance as well as maintaining low dissipation factors that are acceptable for this application. The experimental dielectric constants of these materials are also compared to density functional theory calculations and exhibit a close relationship. © 2013 Wiley Periodicals, Inc. J. Appl. Polym. Sci. 000: 000–000, 2013

KEYWORDS: dielectric properties; polyimides; structure-property relations

Received 10 January 2013; accepted 26 February 2013; published online
DOI: 10.1002/app.39240

INTRODUCTION

There is a current push by the industrial complex to move away from traditional pneumatic, hydraulic, and mechanical devices to electrochemical ones which are controlled electrically.[1] Thus, there is a need for a high-energy storage system to operate these new devices. The best solution to this problem is the incorporation of high-energy density capacitor banks in which the electrical energy can be stored over long charging times and released as needed through short pulses.[2] For this system to operate efficiently and to reduce the size of the overall capacitor bank, an insulating material is needed that possesses a large dielectric constant while still maintaining a low dielectric loss. The two benchmark polymer dielectric materials, biaxially oriented polypropylene (BOPP) and polyvinylidene fluoride (PVDF), have dielectric constants of 2.2 and 10, respectively.[3–6]

Polyimides are attractive materials for this application as they have a more polar backbone and thus a higher dielectric constant than polyolefins. The fact that make polyimides even more desirable is that they are thermally stable at temperatures exceeding 250°C, twice the operating temperature of most common dielectric materials, making them able to withstand the heat generation that these capacitor banks will give rise to and there is less need for cooling the entire capacitor bank.[7] Most common polymers used as dielectric materials exhibit severe decrease in dielectric strength around 70°C.[8]

Interestingly, a search of recent literature yields little work being done in this field. Instead, polyimides are being researched as a possible replacement of silicon dioxide in applications such as the insulating material in semiconductors, printed microelectronics, and so forth.[9–13] Most studies of polyimides in electronic applications have looked at ways to reduce the dielectric constant of the material. The reduction of the dielectric constant has been controlled by lowering the polarizability of the polymer through modification of the backbone with the incorporation of bulky, space-filling groups, the replacement of hydrogen with fluorine atoms, or a combination of both. The introduction of space-filling groups, such as aromatics, increases the free volume and thus decreasing the dipolar and atomic polarizability, whereas fluorine replacement will reduce the total polarizability through the heightening of the hydrophobicity of the polymer.[14–22]

In our research, we focus on the opposite end of the polyimide chemical space in which an increase of the dielectric constant of the polymer is sought. This is done by the polymerization of a common aromatic dianhydride with various alkyl diamines producing an overall reduction of the free volume as well as maintaining a high-imide functional group density in the polymer backbone. The experimental results are combined with theoretical dielectric calculations done through density functional theory (DFT) to give a better understanding of the structure–property relationship of these polyimides. A comparison of theoretical and experimental data will then lead to a next generation of polyimides for the ones reported here.

EXPERIMENTAL

Materials

PMDA, 1,3-diaminopropane (1,3-DAP), 1,2-diaminopropane (1,2-DAP), dimethylformamide (DMF), and *N*-methyl-2-pyrrolidone (NMP) were purchased from Aldrich Chemical (St. Louis, Missouri). Jeffamines (Figure 1) EDR-104, HK511, and D230 were provided by the Huntsman (The Woodlands, Texas). PMDA was recrystallized from acetic anhydride, whereas the diamines were used as received. Shim stocks were purchased from McMaster Carr (Princeton, New Jersey) with a diameter of 2″, a thickness of 0.01″, and ASTM A666 stainless steel.

General Procedure for Homopolymerization

A three-neck flask equipped with a Teflon™-coated magnetic stir bar was fitted with a reflux condenser, a glass stopper, and a glass apparatus for trapping evolved water. The entire apparatus is flame dried under vacuum to remove moisture and placed under an inert atmosphere. To the flask, the appropriate amount of diamine, 0.02 mL of isopropylamine (IPAm) and 50 mL of NMP, was added. The solution was stirred for 30 min at room temperature to disperse the amine equally in solution. To the solution, the appropriate amount of dianhydride was added and this mixture was stirred at 50–100°C for 2 h. Once all of the solids were dissolved, the temperature was increased to 180°C and the solution was stirred for 10 h. The polyimide was precipitated out of solution by the addition of methanol. The solvent was decanted off and the polymer was dried at 75°C *in vacuo*.

Polymer 4. FTIR (cm^{-1}): 1774, 1717, 1664, 1395, and 1363. Differential scanning calorimetry (DSC) results: T_g = not observed; T_m = not observed. Thermogravimetric analysis (TGA) (10°C min^{-1}): N_2 (onset): 258°C.

Polymer 7. ^{1}H NMR ($(CD_3)_2S{=}O$, ppm): δ = 1.23 (m, 12H), 2.94 (m, 11H), 4.23 (m, 2H), 8.11 (m, 2H). FTIR (cm^{-1}): 2976, 2878, 1771, 1716, 1460, 1380, 1355, 1265, 1105, and 917. DSC results: T_g = 75°C; T_m = not observed. TGA (10°C min^{-1}): N_2 (onset): 262°C.

General Procedure for Copolymerization

A three-neck flask equipped with a Teflon™-coated magnetic stir bar was fitted with a reflux condenser, a glass stopper, and a glass apparatus to trap evolved water and a rubber septum stopper. The apparatus was then flame dried under vacuum to remove moisture and placed under an inert atmosphere. To the flask, the appropriate amount of the two diamines, 0.04 mL of IPAm and 25 mL of DMF, was added. The solution was stirred for 30 min at room temperature to disperse the diamines equally in DMF. The appropriate amount of PMDA was added to an Erlenmeyer flask and stored under inert atmosphere. To the flask, 75 mL of DMF was added. The flask is heated slightly until all of the PMDA is dissolved. The amine solution is heated to 50°C and the PMDA/DMF solution is then canulated into the three-neck flask. After the addition of PMDA/DMF solution is complete, the temperature is raised to 150°C and the solution is stirred until the polyamic acid is dissolved. After complete dissolution of the polyamic acid, the temperature is raised to 180°C and the solution is refluxed for 12–24 h to form the polyimide.

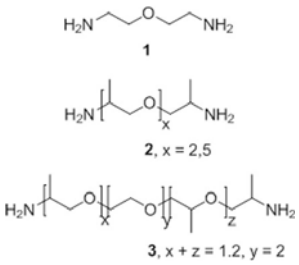


Figure 1. Structure of Jeffamines EDR-104 (top), D230 (middle), and HK511 (bottom).

Polymer 9. ^{1}H NMR ($(CD_3)_2S{=}O$, ppm): δ = 1.01 (m, 10H), 1.28 (m, 8H), 3.64 (m, 10H), 4.43 (m, 3H), 8.14 (m, 4H). FTIR (cm^{-1}): 2975, 2936, 2873, 1771, 1717, 1458, 1358, 1266, 1157, 1110, 1067, 1040, 916, 828, and 731. DSC results: T_g = 54°C; T_m = not observed. TGA (10°C min^{-1}): N_2 (onset): 349°C.

Polymer 10. ^{1}H NMR ($(CD_3)_2S{=}O$, ppm): δ = 1.08 (m, 16H), 4.18 (m, 10H), 8.17 (m, 4H). FTIR (cm^{-1}): 2975, 2936, 2873, 1771, 1717, 1458, 1358, 1266, 1157, 1110, 1067, 1040, 916, 828, and 731. DSC results: T_g = 54°C; T_m = not observed. TGA (10°C min^{-1}): N_2 (onset): 349°C.

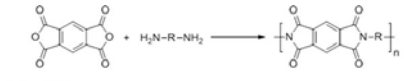
Polymer 11. ^{1}H NMR ($(CD_3)_2S{=}O$, ppm): δ = 0.93 (m, 10H), 1.28 (m, 12H), 3.56 (m, 10H), 4.43 (m, 4H), 8.15 (m, 4H). FTIR (cm^{-1}): 2867, 1770, 1716, 1355, 1105, 1036, and 728. DSC results: T_g = 58°C; T_m = not observed. TGA (10°C min^{-1}): N_2 (onset): 350°C.

Instrumentation

Solution ^{1}H NMR was performed on a Bruker DMX500 high-resolution digital NMR spectrometer. All chemical shifts were referenced to $(CD_3)_2S{=}O$ (^{1}H = 2.50 ppm). DSC was done with a TA instruments DSC Q series with a heating rate of 10°C min^{-1}. Samples were sealed in an aluminum pan with an empty pan used as a reference. TGA was performed using a TA instruments TGA Q500. Dielectric measurements were made with an Imass Time Domain Dielectric Spectrometer.

General Procedure for Generation of Polyimide Thin Film

An appropriate amount of polyimide and DMF is added to a scintillation vial to produce a 10% wt/wt mixture. The mixture is heated until the polyimide dissolves and the solution is then filtered. The filtered solution is drop casted onto a stainless steel



4: R = -$(CH_2)_3$-
5: R = -$CH_2CH(CH_3)$-
6: R = -$(CH_2)_2O(CH_2)_2$-
7: R = -$(CH(CH_3)CH_2O)_xCH_2CH(CH_3)$-
8: R = -$(CH(CH_3)CH_2O)_x(CH_2CH_2O)_y(CH_2CH(CH_3)O)_zCH_2CH(CH_3)$-
9: R = -$[(CH_2)_3]_{0.2}$- . -$[(CH(CH_3)CH_2O)_xCH_2CH(CH_3)]_{0.8}$-
10: R = -$[(CH_2)_3]_{0.2}$- . -$[(CH(CH_3)CH_2O)_x(CH_2CH_2O)_y(CH_2CH(CH_3)O)_zCH_2CH(CH_3)]_{0.8}$-
11: R = -$[CH_2CH(CH_3)]_{0.2}$- . -$[(CH(CH_3)CH_2O)_xCH_2CH(CH_3)]_{0.8}$-
12: R = -$[CH_2CH(CH_3)]_{0.5}$- . -$[(CH(CH_3)CH_2O)_xCH_2CH(CH_3)]_{0.5}$-
13: R = -$[CH_2CH(CH_3)]_{0.5}$- . -$[(CH(CH_3)CH_2O)_x(CH_2CH_2O)_y(CH_2CH(CH_3)O)_zCH_2CH(CH_3)]_{0.5}$

Figure 2. Structures of synthesized polyimides.

shim stock and covered with a glass petri dish slightly opened to the air and allowed to dry overnight to produce a tacky film. The shim stock with the tacky film is then placed into a vacuum oven and heated at 165–175°C for 24 h to evaporate all of the solvents. Isothermal TGA is run at 170°C on a dried part of the film to determine the level of solvent remaining

RESULTS AND DISCUSSION

Homopolymers

Solubility and Thermal Properties. The homopolymers shown in Figure 2 were all synthesized from the same condensation procedure. Polymers **4**, **5**, and **6** proved to be insoluble in most organic solvents with the exception of *m*-cresol. However, even after extensive drying under vacuum, the minute amount of *m*-cresol remaining caused the films casted from this solvent to be too conductive. Polymers **7** and **8** proved to be more soluble stemming from the increase in distance between rigid aromatic functionalities in the polymer backbone.

The DSC data summarized in Table I show that polymers **4**, **5**, and **6** do not exhibit a glass transition temperature even down to −90°C. A melting temperature is either absent or is above the point of thermal degradation and thus not observed. The incorporation of a longer alkyl ether chain between aromatic imide groups gives rises to a T_g but again there is an absence of a melting temperature.

Dielectric Properties. The insoluble polymers were pressed into a pellet to test the dielectric properties of the material with polymer **4** being cohesive enough and did not fall apart when placed in the sample holder. Figure 3 (top) shows the dielectric spectra of polymer **4**. The dielectric constant is stable above 5.5, whereas the loss factor remains below 2% even as the temperature is increased to 100°C. Polymer **4**, exhibits a dielectric constant that is 2.5 times larger than BOPP while still being well under the 5% dielectric loss threshold which constitutes capacitor failure.[23] When the longer ether chains are added instead of the short alkyl chains, the dielectric constant is reduced as there is an increase in free volume stemming from the protruding methyl groups. The loss factor, which is below 1% at room temperature, increases significantly when the polymer is heated to T_g and the instrument short circuits as soon as the polymer is heated above T_g.

Copolymers

As the polymers that contain the longer ether chains are more soluble in high boiling solvents, it was decided to copolymerize them with the shorter alkyl diamines. The introduction of the shorter alkyl diamine served a twofold purpose (1) by increasing the percent functionality of imide there should be an increase in dielectric constant and (2) the loss factor should also be reduced through an increase in rigidity of the polymer, that is less response by the polymer to the switching electric field.

Thermal Properties and Film Formation. Combining the short alkyl diamine with the long ether diamines caused a reduction in the T_g versus a polymer solely made with the ether linkage. The solubility of these copolymers, made with 20% of the alkyl and 80% of the ether diamines, in high boiling solvents was maintained compared to the homopolymers and the films produced looked like a sheet of glass over the shim stock. Changing the feed of the diamines to a 50/50 mixture created polymers with reduced solubility in DMF. Instead of a smooth, clear film, a dispersion in DMF was obtained at a much reduced concentration (1–5%). Upon cooling the film, cracks appeared and the film peeled off of the shim stock.

Dielectric Properties. Figure 3 (bottom) shows the typical dielectric results for the copolymers. There is an increase in the dielectric constant with copolymers constructed from 1,3-DAP and either Jeffamine **2** or **3**. Using 1,2-DAP instead caused a reduction in the dielectric constant which can be explained by the free volume increase from the methyl group. Even with the incorporation of alkyl diamine, the loss factor remains on the same order as the homopolymer, whereas increasing the temperature to the T_g of the polymer does not cause as significant increase in the loss as compared to the homopolymer. Raising the temperature above the T_g of the polymer causes the loss factor to increase well above 5% at high frequencies and above 10% at low frequencies.

Table I. Thermal and Dielectric Properties of Polyimides

Polymer	T_g (°C)[a]	T_m (°C)	T_d (°C)[b]	k_{calc}	$k_{60\ Hz}$[c]	$k_{1\ kHz}$[c]	tan(δ)$_{60\ Hz}$[c]	tan(δ)$_{1\ kHz}$[c]
4[d]	n/o	n/o	258	4.0 ± 0.5	5.61	5.60	0.000869	0.000553
7	75	n/o	262	3.8 ± 0.4	4.20	4.18	0.00357	0.00518
9	54	n/o	349	TBD[e]	6.57	6.55	0.00362	0.00626
10	58	n/o	350	TBD[e]	5.15	5.12	0.00363	0.00460
11	51	n/o	378	TBD[e]	3.96	3.92	0.00477	0.00608

[a]Values taken from the midpoint, [b]Values taken from onset, [c]Values at room temperature, [d]Pressed pellet, [e]To be determined.

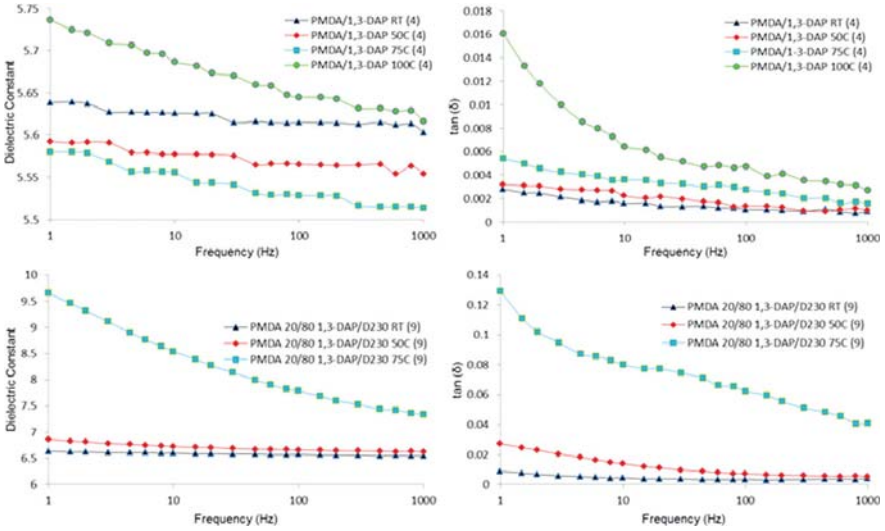


Figure 3. Temperature effects on dielectric properties. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Comparison to Theoretical Calculations

Theoretical Calculations. Given that the synthesized polyimides have complex structures, it is difficult and time-intensive to perform DFT calculations. Therefore, we considered a few polyimides (Figure 4) with simpler but similar structures compared to those synthesized polyimides. In this way, we can give a qualitative view on the structure–property relationship of synthesized polyimides. To mirror the experimental work, the repeat unit of A-1 is exactly the same as synthesized polymer **4**, and polymer A-3 which is very similar to synthesized polymer **7**. By considering A-3, we can identify the effect of incorporating methyl groups and increasing the number of —CH2— units in the repeat unit.

DFT, as implemented in the Vienna *ab initio* simulation package (VASP), was used to determine the structural and electronic properties of the polymers.[24] The Perdew, Burke, and Ernzerhof functional (PBE), projector-augmented wave frozen-core potentials, and a cutoff energy of 400 eV for the plane wave expansion of the wave functions were used.[25–27] The PBE-optimized geometry was then used to determine the dielectric constant tensor using density functional perturbation theory (DFPT).[28] In this study, we considered an isolated infinite chain of the polyimides. The true dielectric permittivity of the polymer chain alone was then extracted by combining the DFPT computation of the supercell (containing a significant amount of vacuum) with effective medium theory, using a recently developed method.[29,30] The band gap and dielectric constant results of polyimides are shown in Figure 5 in which the filled symbol represents the electronic

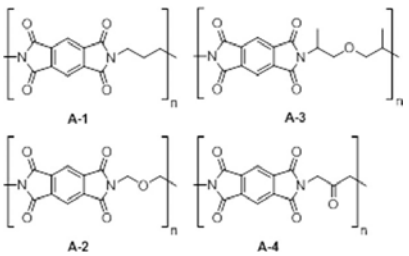


Figure 4. Repeating units of polyimides calculated using DFT.

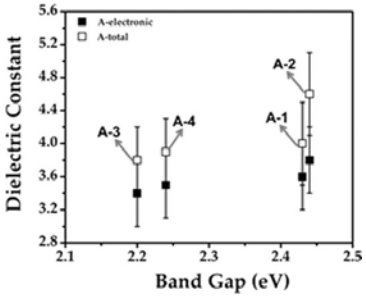


Figure 5. Band gap and dielectric constant results.

dielectric constant, whereas the open symbol is for total dielectric constant. When changing the linker from —CH2— to —O—, the band gap remains almost the same, but both the electronic and the total dielectric constant increase, but when the linker is changed to —CO—, both the band gap and the dielectric constant decrease. Also, in Figure 5, when comparing A-3 with A-2, we can see that incorporating methyl groups and increasing the number of —CH2— units in the repeating unit will decrease the band gap and the dielectric constant.

Polymer **4** has an experimentally higher, about a factor of 1.4, than the calculated dielectric constant, whereas the calculated and experimental dielectric constant is the same for polymer **7**. The difference in accuracy for the two calculations could be owing to how well the polymer chains are ordered as polymer **4** is a pressed pellet and polymer **7** is a thin film. The ordering of polymer chains could be different and this difference is not taken into account in DFT calculations.

CONCLUSIONS

A series of high-functionality polyimide homopolymers and copolymers featuring PMDA with various short-chain aliphatic diamines were synthesized and dielectric properties were measured. The homopolymers with the densest functionality proved to be insoluble, whereas the soluble homopolymers had a large increase in dielectric loss and could not be measured at temperatures near or above T_g. A reduction in loss of the soluble homopolymers is accomplished by copolymerization with a shorter alkyl diamine which also allows for the measurements to be made above T_g, but again the loss is quite significant above this temperature. Still, the dielectric constant of all of the polymers was higher than biaxially oriented polypropylene and possessed operating temperatures well above that of BOPP. Future work will look at processing a film of polymer **4** from the more soluble poly(amic) acid form and then performing thermal imidization.

ACKNOWLEDGMENTS

This study is supported through a multidisciplinary university research initiative (MURI) grant through the Office of Naval Research. The authors acknowledge JoAnne Ronzello for performing all of the dielectric measurements. A. F. Baldwin, R. Ma, and C. Wang drafted the article from acquired data. G. A. Sotzing and R. Ramprasad gave final approval of the submitted version of the manuscript.

REFERENCES

1. Beach, F.; McNab, I. In IEEE Conference on Pulsed Power, Monterey, California, USA, June 13–17, **2005**, 1.
2. Sarjeant, W.; Zirnheld, J.; MacDougall, F. *IEEE Trans. Plasma Sci.* **1998**, *26*, 1368.
3. Barshaw, E.; White, J.; Chait, M.; Cornette, J.; Bustamante, J.; Folli, F.; Biltchick, D.; Borelli, G.; Picci, G.; Rabuffi, M. *IEEE Trans. Magn.* **2007**, *43*, 223.
4. Guan, F.; Pan, J.; Wang, J.; Wang, Q.; Zhu, L. *Macromolecules* **2010**, *43*, 384.
5. Dang, Z. M.; Wang, H. Y.; Xu, H. P. *Appl. Phys. Lett.* **2006**, *89*, 112902.
6. Anderson, E. W.; McCall, D. W. *J. Polym. Sci.: Part A: Polym. Chem.* **1958**, *31*, 241.
7. Rabuffi, M.; Picci, G. *IEEE Trans. Plasma Sci* **2002**, *30*, 1939.
8. MacDougall, F. W.; Ennis, J. B.; Cooper, R. A.; Bates, J.; Seal, K. In 14th IEEE International Pulsed Power Conference Digest of Technical Papers, Dallas, Texas, USA, June 15–18, **2003**, 513.
9. Hergenrother, P. M. *High Perform. Polym.* **2003**, *15*, 3.
10. Rebeiz, G. M. In RF MEMS: Theory, Design, and Technology; Wiley Interscience: Hoboken, New Jersey, **2003**, p 179.
11. Chen, C.-J.; Yen, H.-J.; Chen, W.-C.; Liou, G.-S. *J. Polym. Sci.: Part A: Polym. Chem.* **2001**, *49*, 3709.
12. Lee, C.; Sundar, S.; Kwon, J.; Han, H. *J. Polym. Sci.: Part A: Polym. Chem.* **2004**, *42*, 3621.
13. Gonzalo, B.; Vilas, J. L.; Breczewski, T.; Pèrez-Jubindo, M. A.; De La Fuente, M. R.; Rodriguez, M.; Lèon, L. M. *J. Polym. Sci.: Part A: Polym. Chem.* **2009**, *47*, 722.
14. Chisca, S.; Musteata, V.; Sava, I.; Bruma, M. In Semiconductor Conference, 2010 International, Sinaia, Romania, October 11–13, **2010**, *2*, 325.
15. Houghman, G.; Tesoro, G.; Viehbeck, A.; Chapple-Sokol, J. D. *Macromolecules* **1994**, *27*, 5964.
16. Mo, X.; Wang, C. Y.; Li, G.; Jiang, J. M. *J. Macromol. Sci. B Phys.* **2012**, *51*, 1370.
17. Popovici, D.; Hulubei, C.; Cozan, V.; Lisa, G.; Bruma, M. *High Perform. Polym.* **2012**, *24*, 194.
18. Wang, C.; Zhao, X.; Li, G.; Jiang, J. *Colloid. Polym. Sci.* **2011**, *289*, 1617.
19. Othman, M. B. H.; Ming, N. A. S.; Akil, H. M.; Ahmad, Z. *J. Appl. Polym. Sci.* **2011**, *121*, 3192.
20. Wu. S.-Y.; Yuen, S.-M.; Ma, C.-C. M.; Huang, Y.-L. *J. Appl. Polym. Sci.* **2009**, *113*, 2301.
21. Banerjee, S.; Madhra, M. K.; Kute, V. *J. Appl. Polym. Sci.* **2004**, *93*, 821.
22. Sroog, C. E.; Endrey, A. L.; Abramo, S. V.; Berr, C. E.; Edwards, W. M.; Olivier, K. L. *J. Polym. Sci.: Part A: Polym. Chem.* **1965**, *3*, 1373.
23. MacDonald, J. R.; Schneider, M. A.; Ennis, J. B.; MacDougall, F. W.; Yang, X. H.; In IEEE Electrical Insulation Conference, Montreal, Canada, May 31–June 3, **2009**, 306.
24. Kresse, G.; Furthmuller, J. *Phys. Rev. B* **1996**, *54*, 11169.
25. Perdew, J. P.; Burke, K.; Ernzerhof, M. *Phys. Rev. Lett.* **1996**, *77*, 3865.
26. Blochl, P. E. *Phys. Rev. B* **1994**, *50*, 17953.
27. Kresse, G.; Joubert, D. *Phys. Rev. B* **1999**, *59*, 1758.
28. Baroni, S.; de Gironcoli, S.; Dal Corso, A. *Rev. Mod. Phys.* **2001**, *73*, 515.
29. Pilania, G.; Ramprasad, R. *J. Mater. Sci.* **2012**, *47*, 7580.
30. Wang, C. C.; Pilania, G.; Ramprasad, R. *Phys. Rev. B* **2013**, *87*, 035103.

Appendix C

High througput DFT results of organic polymers

Table C.1 shows the calculated band gap, electronic and total dielectric constant of 267 organic polymers. The dielectric constant is computed using the single-chain approach, while the Heyd-Scuseria-Ernzerhof HSE06 functional is used to obtain the corrected band gap values.

Table C.1: Band gap (E_g), electronic (ϵ_e), and total (ϵ_t) dielectric constant of 267 organic polymers

System	E_g	ϵ_e	error bar	ϵ_t	error bar
CH2-CH2-CH2-CH2	7.18	2.51	0.25	2.51	0.25
CH2-CH2-CH2-NH	5.34	2.58	0.27	2.71	0.29
CH2-CH2-CH2-C6H4	5.08	3.34	0.39	3.37	0.40
CH2-CH2-CH2-C4H2S	4.79	3.21	0.37	3.25	0.37
CH2-CH2-CH2-CS	3.42	2.79	0.29	2.84	0.30
CH2-CH2-CH2-O	6.35	2.53	0.26	2.73	0.30
CH2-CH2-NH-CO	5.87	2.71	0.28	3.21	0.37
CH2-CH2-NH-C6H4	4.28	3.60	0.43	4.09	0.51
CH2-CH2-NH-C4H2S	4.27	3.54	0.43	4.34	0.56
CH2-CH2-NH-CS	4.04	3.05	0.34	3.23	0.37
CH2-CH2-CO-C6H4	4.77	3.47	0.40	4.10	0.50
CH2-CH2-CO-C4H2S	4.21	3.73	0.44	4.32	0.54
CH2-CH2-CO-O	7.04	2.61	0.26	3.19	0.36
CH2-CH2-C6H4-C6H4	4.01	4.20	0.51	4.24	0.52
CH2-CH2-C6H4-C4H2S	3.79	4.17	0.51	4.24	0.52
CH2-CH2-C6H4-CS	3.03	4.01	0.47	4.51	0.55
CH2-CH2-C6H4-O	4.93	3.31	0.38	3.68	0.44
CH2-CH2-C4H2S-C4H2S	3.55	4.33	0.54	4.39	0.55

CH2-CH2-C4H2S-CS	2.68	4.19	0.52	4.60	0.59
CH2-CH2-C4H2S-O	4.93	3.28	0.38	3.67	0.44
CH2-CH2-CS-O	4.16	2.97	0.33	3.29	0.38
CH2-CH2-O-NH	5.66	2.71	0.30	2.97	0.35
CH2-NH-CH2-NH	5.19	2.64	0.29	2.84	0.32
CH2-NH-CH2-CO	4.78	2.68	0.28	3.44	0.41
CH2-NH-CH2-C6H4	4.97	3.43	0.41	3.54	0.43
CH2-NH-CH2-C4H2S	4.70	3.34	0.39	3.69	0.45
CH2-NH-CH2-CS	3.45	3.04	0.34	3.51	0.42
CH2-NH-CH2-O	5.61	2.65	0.29	3.87	0.50
CH2-NH-CO-NH	5.89	2.81	0.30	4.33	0.55
CH2-NH-CO-C6H4	5.15	3.45	0.39	4.05	0.49
CH2-NH-CO-C4H2S	4.36	3.70	0.44	5.00	0.65
CH2-NH-CO-O	6.47	2.77	0.29	3.73	0.45
CH2-NH-C6H4-NH	4.38	3.63	0.44	4.41	0.56
CH2-NH-C6H4-CO	3.44	4.40	0.55	5.81	0.77
CH2-NH-C6H4-C6H4	3.97	4.41	0.55	6.50	0.88
CH2-NH-C6H4-C4H2S	3.76	4.53	0.57	4.92	0.64
CH2-NH-C6H4-O	4.36	3.68	0.44	5.22	0.70
CH2-NH-C4H2S-NH	4.36	3.70	0.46	5.87	0.82
CH2-NH-C4H2S-CO	3.19	4.78	0.61	6.74	0.93
CH2-NH-C4H2S-C6H4	3.66	4.48	0.57	5.81	0.79

CH2-NH-C4H2S-C4H2S	3.25	4.87	0.63	5.89	0.80
CH2-NH-C4H2S-CS	2.46	5.92	0.81	8.04	1.16
CH2-NH-C4H2S-O	4.46	3.49	0.42	4.58	0.60
CH2-NH-CS-NH	4.14	3.26	0.38	4.11	0.53
CH2-NH-CS-CO	2.65	3.34	0.38	4.12	0.51
CH2-NH-CS-C6H4	3.57	3.86	0.46	4.33	0.53
CH2-NH-CS-C4H2S	3.27	4.22	0.53	4.74	0.62
CH2-NH-CS-CS	1.91	3.94	0.48	4.97	0.65
CH2-NH-CS-O	4.83	3.25	0.37	3.89	0.48
CH2-NH-O-CO	5.93	2.99	0.33	3.74	0.46
CH2-NH-O-C6H4	4.99	3.50	0.41	3.78	0.46
CH2-NH-O-C4H2S	4.62	3.61	0.44	4.01	0.50
CH2-NH-O-CS	4.36	3.37	0.40	3.95	0.50
CH2-CO-CH2-CO	4.26	2.68	0.27	3.12	0.34
CH2-CO-CH2-C6H4	4.77	3.42	0.40	3.72	0.45
CH2-CO-CH2-C4H2S	4.67	3.34	0.38	3.66	0.43
CH2-CO-CH2-CS	2.87	2.90	0.31	3.15	0.36
CH2-CO-CH2-O	5.38	2.56	0.26	2.96	0.33
CH2-CO-NH-CO	4.39	2.98	0.32	3.97	0.48
CH2-CO-NH-C6H4	4.62	3.79	0.45	4.39	0.55
CH2-CO-NH-C4H2S	4.30	3.88	0.47	4.40	0.55
CH2-CO-NH-CS	2.94	3.16	0.36	4.11	0.51

CH2-CO-C6H4-CO	4.59	3.65	0.42	4.64	0.58
CH2-CO-C6H4-C6H4	3.94	4.41	0.53	4.82	0.60
CH2-CO-C6H4-C4H2S	3.51	4.66	0.58	5.36	0.69
CH2-CO-C6H4-CS	2.75	4.45	0.53	5.42	0.68
CH2-CO-C6H4-O	4.24	3.79	0.45	4.56	0.57
CH2-CO-C4H2S-CO	4.01	3.88	0.46	5.03	0.64
CH2-CO-C4H2S-C6H4	3.52	4.82	0.60	5.35	0.69
CH2-CO-C4H2S-C4H2S	2.88	5.40	0.70	6.15	0.82
CH2-CO-C4H2S-CS	2.67	4.32	0.54	5.54	0.73
CH2-CO-C4H2S-O	3.81	3.85	0.46	4.80	0.62
CH2-CO-O-CO	4.99	2.80	0.29	5.09	0.65
CH2-CO-O-C6H4	5.04	3.48	0.40	4.16	0.51
CH2-CO-O-C4H2S	4.63	3.60	0.42	4.27	0.53
CH2-CO-O-CS	3.48	2.85	0.30	3.86	0.47
CH2-C6H4-CH2-C6H4	5.26	3.66	0.44	3.73	0.45
CH2-C6H4-CH2-C4H2S	4.96	3.35	0.38	3.43	0.39
CH2-C6H4-CH2-CS	3.57	3.41	0.40	3.54	0.42
CH2-C6H4-CH2-O	5.02	3.44	0.41	3.73	0.46
CH2-C6H4-NH-NH	4.25	3.86	0.48	5.86	0.82
CH2-C6H4-NH-C6H4	4.21	4.20	0.51	4.55	0.57
CH2-C6H4-NH-C4H2S	4.33	3.65	0.43	4.19	0.52
CH2-C6H4-NH-CS	3.82	4.18	0.51	4.55	0.57

CH2-C6H4-NH-O	4.90	3.57	0.42	4.48	0.57
CH2-C6H4-CO-C6H4	4.27	3.76	0.43	4.02	0.47
CH2-C6H4-CO-C4H2S	4.04	3.86	0.45	4.38	0.54
CH2-C6H4-CO-O	4.94	3.62	0.42	4.40	0.54
CH2-C6H4-C6H4-C6H4	3.65	4.70	0.58	4.77	0.59
CH2-C6H4-C6H4-C4H2S	3.38	5.21	0.66	5.33	0.68
CH2-C6H4-C6H4-CS	2.97	4.71	0.59	5.15	0.66
CH2-C6H4-C6H4-O	4.35	4.15	0.50	4.59	0.57
CH2-C6H4-C4H2S-C6H4	3.72	4.70	0.59	4.77	0.60
CH2-C6H4-C4H2S-C4H2S	3.58	4.96	0.64	5.21	0.68
CH2-C6H4-C4H2S-CS	2.76	5.47	0.71	5.86	0.77
CH2-C6H4-CS-C6H4	2.87	3.75	0.44	4.01	0.48
CH2-C6H4-CS-C4H2S	2.78	4.19	0.51	4.57	0.57
CH2-C6H4-CS-O	3.45	4.15	0.50	4.69	0.59
CH2-C6H4-O-C6H4	5.09	3.33	0.37	3.58	0.41
CH2-C6H4-O-C4H2S	4.70	3.55	0.41	3.86	0.46
CH2-C6H4-O-CS	4.05	3.70	0.44	4.16	0.51
CH2-C4H2S-CH2-C4H2S	5.02	3.60	0.43	3.69	0.44
CH2-C4H2S-CH2-CS	3.33	3.54	0.41	3.66	0.43
CH2-C4H2S-CH2-O	4.62	3.38	0.40	4.91	0.66
CH2-C4H2S-NH-C4H2S	4.14	3.70	0.44	3.90	0.48
CH2-C4H2S-NH-CS	3.45	4.33	0.54	4.66	0.59

CH2-C4H2S-NH-O	4.31	3.72	0.46	5.37	0.74
CH2-C4H2S-CO-C4H2S	3.63	4.26	0.52	4.84	0.61
CH2-C4H2S-CO-O	4.55	3.65	0.44	4.58	0.59
CH2-C4H2S-C6H4-C4H2S	3.22	5.40	0.70	5.54	0.72
CH2-C4H2S-C6H4-CS	2.87	5.26	0.68	5.93	0.79
CH2-C4H2S-C6H4-O	4.05	4.27	0.53	4.76	0.61
CH2-C4H2S-C4H2S-C4H2S	2.81	5.78	0.77	5.91	0.79
CH2-C4H2S-C4H2S-CS	2.47	6.22	0.83	6.89	0.94
CH2-C4H2S-C4H2S-O	3.44	4.60	0.59	5.37	0.72
CH2-C4H2S-CS-C4H2S	2.60	4.81	0.61	5.28	0.69
CH2-C4H2S-CS-O	3.46	4.11	0.51	4.92	0.64
CH2-C4H2S-O-C4H2S	4.74	3.45	0.40	3.70	0.44
CH2-C4H2S-O-CS	3.54	3.87	0.47	4.35	0.55
CH2-CS-CH2-CS	0.29	3.39	0.38	8.43	1.19
CH2-CS-CH2-O	3.07	2.75	0.30	2.94	0.33
CH2-CS-NH-CS	0.38	3.61	0.41	8.36	1.16
CH2-CS-NH-O	4.38	3.15	0.35	4.74	0.62
CH2-CS-C6H4-NH	2.49	5.52	0.73	6.98	0.96
CH2-CS-C6H4-CS	1.34	12.15	1.76	21.08	3.17
CH2-CS-C6H4-O	2.79	4.60	0.57	5.34	0.69
CH2-CS-C4H2S-CS	2.09	9.66	1.37	11.11	1.59
CH2-CS-C4H2S-O	2.84	4.46	0.57	5.34	0.72

CH2-CS-O-CS	0.39	3.51	0.40	8.44	1.17
CH2-O-CH2-O	6.53	2.59	0.27	3.21	0.38
CH2-O-NH-CO	6.01	2.82	0.31	3.87	0.49
CH2-O-NH-O	3.13	2.50	0.30	6.35	1.07
CH2-O-C6H4-O	4.90	3.41	0.39	4.82	0.63
CH2-O-C4H2S-C6H4	3.74	4.34	0.54	5.47	0.72
CH2-O-C4H2S-O	4.89	3.42	0.40	4.51	0.58
CH2-O-CS-O	4.81	3.18	0.36	4.16	0.52
CH2-O-O-C6H4	4.46	3.52	0.42	5.39	0.72
CH2-O-O-C4H2S	4.60	2.88	0.33	4.19	0.55
CH2-O-O-CS	5.49	2.19	0.20	2.72	0.29
NH-CO-NH-CO	4.04	3.37	0.38	5.70	0.76
NH-CO-NH-C6H4	3.84	4.99	0.64	6.46	0.88
NH-CO-NH-C4H2S	3.36	5.08	0.66	6.23	0.85
NH-CO-NH-CS	3.52	3.97	0.47	5.79	0.76
NH-CO-NH-O	3.24	2.23	0.24	3.79	0.55
NH-CO-C6H4-CO	4.50	3.70	0.44	7.43	1.04
NH-CO-C6H4-C6H4	3.15	5.99	0.78	6.75	0.90
NH-CO-C6H4-C4H2S	3.78	4.56	0.57	5.39	0.71
NH-CO-C6H4-CS	2.96	4.15	0.51	7.89	1.10
NH-CO-C6H4-O	4.74	3.83	0.45	4.65	0.58
NH-CO-C4H2S-CO	3.54	4.36	0.54	6.19	0.83

NH-CO-C4H2S-C6H4	2.72	6.95	0.94	8.04	1.11
NH-CO-C4H2S-C4H2S	2.32	8.29	1.16	9.75	1.39
NH-CO-C4H2S-CS	2.35	5.33	0.70	8.47	1.21
NH-CO-C4H2S-O	4.22	3.95	0.49	5.25	0.70
NH-CO-O-C6H4	4.38	4.23	0.52	5.87	0.78
NH-CO-O-C4H2S	3.87	4.44	0.56	5.79	0.78
NH-C6H4-NH-C6H4	3.97	4.72	0.62	5.29	0.72
NH-C6H4-NH-C4H2S	3.00	6.49	0.91	7.47	1.07
NH-C6H4-NH-CS	3.26	5.67	0.75	6.75	0.93
NH-C6H4-NH-O	2.50	4.87	0.63	8.98	1.30
NH-C6H4-CO-C6H4	3.02	6.12	0.80	7.40	1.00
NH-C6H4-CO-C4H2S	3.00	5.62	0.73	9.00	1.26
NH-C6H4-CO-O	4.24	4.30	0.53	5.42	0.71
NH-C6H4-C6H4-C6H4	3.25	5.76	0.76	6.34	0.85
NH-C6H4-C6H4-C4H2S	2.46	8.39	1.17	9.39	1.32
NH-C6H4-C6H4-CS	3.66	5.94	0.78	6.68	0.90
NH-C6H4-C6H4-O	2.13	7.23	1.01	8.98	1.30
NH-C6H4-C4H2S-C6H4	3.60	5.19	0.67	8.10	1.14
NH-C6H4-C4H2S-C4H2S	2.24	9.30	1.33	11.01	1.61
NH-C6H4-C4H2S-CS	2.36	8.33	1.16	9.19	1.30
NH-C6H4-C4H2S-O	2.22	5.26	0.74	5.97	0.86
NH-C6H4-CS-C6H4	2.56	7.41	1.01	8.72	1.22

NH-C6H4-CS-O	3.42	4.80	0.61	5.31	0.70
NH-C6H4-O-C6H4	3.40	5.61	0.73	6.61	0.89
NH-C6H4-O-C4H2S	3.42	3.84	0.46	4.62	0.59
NH-C6H4-O-CS	3.79	4.73	0.60	5.90	0.79
NH-C4H2S-NH-C4H2S	2.74	6.54	0.92	7.36	1.05
NH-C4H2S-NH-CS	2.97	5.75	0.77	6.74	0.94
NH-C4H2S-NH-O	2.79	3.84	0.53	4.39	0.63
NH-C4H2S-CO-C4H2S	2.27	9.17	1.31	12.26	1.80
NH-C4H2S-CO-O	3.26	5.17	0.68	8.31	1.19
NH-C4H2S-C6H4-C4H2S	2.40	8.63	1.21	9.40	1.34
NH-C4H2S-C6H4-CS	3.27	4.65	0.59	5.19	0.68
NH-C4H2S-C6H4-O	2.30	6.75	0.95	9.19	1.36
NH-C4H2S-C4H2S-C4H2S	2.05	10.26	1.49	11.14	1.63
NH-C4H2S-C4H2S-CS	2.04	9.56	1.37	10.76	1.56
NH-C4H2S-C4H2S-O	3.35	4.91	0.64	6.64	0.92
NH-C4H2S-CS-C6H4	2.28	7.80	1.08	11.57	1.67
NH-C4H2S-CS-C4H2S	1.70	13.01	1.92	16.73	2.51
NH-C4H2S-CS-O	2.69	6.78	0.96	10.82	1.63
NH-C4H2S-O-C4H2S	4.38	3.85	0.47	4.30	0.54
NH-C4H2S-O-CS	3.28	4.96	0.64	6.18	0.84
NH-CS-C6H4-CS	1.06	20.62	3.14	37.64	5.87
NH-CS-C6H4-O	3.25	4.29	0.52	4.84	0.61

NH-CS-C4H2S-CS	1.35	20.24	3.07	35.84	5.57
NH-CS-C4H2S-O	3.14	4.44	0.57	5.52	0.75
NH-O-C6H4-O	1.75	5.16	0.73	7.41	1.12
NH-O-C4H2S-O	1.86	4.08	0.57	5.09	0.75
CO-C6H4-CO-C6H4	4.13	3.94	0.48	4.59	0.59
CO-C6H4-CO-C4H2S	3.69	4.08	0.49	5.43	0.71
CO-C6H4-CO-O	4.61	3.76	0.43	6.21	0.81
CO-C6H4-C6H4-C6H4	3.31	5.09	0.64	5.49	0.70
CO-C6H4-C6H4-C4H2S	2.87	6.16	0.80	7.02	0.94
CO-C6H4-C6H4-CS	2.29	6.52	0.86	7.69	1.04
CO-C6H4-C6H4-O	3.88	4.47	0.55	5.28	0.68
CO-C6H4-C4H2S-C6H4	2.53	7.24	0.97	7.98	1.08
CO-C6H4-C4H2S-C4H2S	2.59	6.83	0.93	8.43	1.18
CO-C6H4-C4H2S-CS	2.22	7.00	0.94	9.06	1.26
CO-C6H4-C4H2S-O	3.53	4.79	0.61	6.37	0.86
CO-C6H4-CS-C6H4	2.89	4.01	0.50	4.57	0.59
CO-C6H4-CS-C4H2S	2.28	5.02	0.63	6.05	0.79
CO-C6H4-CS-O	3.27	4.22	0.50	6.26	0.82
CO-C6H4-O-C6H4	3.49	5.14	0.65	6.24	0.82
CO-C6H4-O-C4H2S	3.65	4.54	0.56	6.74	0.90
CO-C6H4-O-CS	3.47	3.82	0.45	4.74	0.59
CO-C4H2S-CO-C4H2S	3.00	5.94	0.77	7.73	1.05

CO-C4H2S-CO-O	3.98	3.89	0.47	7.20	1.00
CO-C4H2S-C6H4-C4H2S	2.39	8.16	1.11	9.21	1.28
CO-C4H2S-C6H4-CS	2.18	7.29	0.99	9.04	1.27
CO-C4H2S-C6H4-O	2.92	6.09	0.81	7.43	1.02
CO-C4H2S-C4H2S-C4H2S	1.98	10.55	1.50	12.23	1.76
CO-C4H2S-C4H2S-CS	1.97	8.78	1.22	10.30	1.46
CO-C4H2S-C4H2S-O	2.62	6.75	0.92	8.66	1.23
CO-C4H2S-CS-C4H2S	1.99	6.84	0.92	8.20	1.13
CO-C4H2S-CS-O	2.97	4.29	0.54	8.18	1.18
CO-C4H2S-O-C4H2S	2.89	6.16	0.82	8.10	1.14
CO-C4H2S-O-CS	2.86	4.57	0.57	5.60	0.74
CO-O-C6H4-CS	2.95	4.38	0.53	5.97	0.79
CO-O-C6H4-O	5.35	3.44	0.40	4.80	0.62
CO-O-C4H2S-CS	2.29	5.17	0.67	6.96	0.96
CO-O-C4H2S-O	4.26	4.03	0.49	5.43	0.71
C6H4-C6H4-C6H4-C6H4	2.54	8.27	1.12	8.33	1.13
C6H4-C6H4-C6H4-C4H2S	2.48	8.27	1.14	8.37	1.16
C6H4-C6H4-C6H4-CS	2.58	5.59	0.73	7.04	0.96
C6H4-C6H4-C6H4-O	3.57	5.07	0.64	5.42	0.70
C6H4-C6H4-C4H2S-C4H2S	2.27	8.78	1.22	9.11	1.28
C6H4-C6H4-C4H2S-CS	2.31	7.18	0.97	8.05	1.11
C6H4-C6H4-C4H2S-O	2.95	6.41	0.86	7.09	0.96

C6H4-C6H4-CS-C4H2S	2.31	7.20	0.98	8.09	1.12
C6H4-C6H4-CS-O	3.23	4.78	0.60	5.36	0.70
C6H4-C6H4-O-C4H2S	2.93	6.46	0.86	7.16	0.97
C6H4-C6H4-O-CS	3.21	4.83	0.61	5.40	0.70
C6H4-C4H2S-C6H4-C4H2S	2.79	6.76	0.92	6.89	0.94
C6H4-C4H2S-C6H4-CS	2.11	8.00	1.10	8.61	1.20
C6H4-C4H2S-C6H4-O	2.81	6.66	0.89	7.28	0.99
C6H4-C4H2S-C4H2S-C4H2S	1.96	11.11	1.61	11.34	1.64
C6H4-C4H2S-C4H2S-O	3.28	4.98	0.64	6.71	0.92
C6H4-C4H2S-CS-C4H2S	2.10	12.06	1.74	13.26	1.93
C6H4-C4H2S-CS-O	2.57	6.86	0.93	8.06	1.13
C6H4-C4H2S-O-C4H2S	2.90	6.44	0.87	7.05	0.96
C6H4-C4H2S-O-CS	3.08	5.06	0.66	5.97	0.80
C6H4-CS-C6H4-CS	2.69	4.22	0.54	4.67	0.61
C6H4-CS-C6H4-O	2.56	5.08	0.65	5.91	0.78
C6H4-CS-C4H2S-C4H2S	2.37	6.26	0.84	7.15	0.99
C6H4-CS-C4H2S-CS	2.04	5.18	0.67	7.04	0.97
C6H4-CS-C4H2S-O	2.65	5.05	0.64	6.76	0.92
C6H4-O-C6H4-O	4.75	4.00	0.50	4.62	0.60
C6H4-O-C4H2S-O	4.59	3.69	0.43	4.35	0.54
C6H4-O-CS-C4H2S	2.57	6.82	0.93	8.06	1.13
C4H2S-C4H2S-C4H2S-C4H2S	1.60	15.39	2.29	15.55	2.31

C4H2S-C4H2S-C4H2S-CS	1.50	15.04	2.20	16.88	2.49
C4H2S-C4H2S-C4H2S-O	2.39	7.73	1.08	8.43	1.19
C4H2S-C4H2S-CS-O	2.40	7.33	1.02	9.00	1.29
C4H2S-C4H2S-O-CS	2.37	7.42	1.03	9.10	1.30
C4H2S-CS-C4H2S-CS	0.93	9.67	1.36	11.03	1.57
C4H2S-CS-C4H2S-O	2.51	7.22	1.00	9.82	1.42
C4H2S-O-C4H2S-O	3.73	4.71	0.61	5.49	0.74

Appendix D

High througput DFT results of Group 14 element-based hybrid polymers

Table D.1 shows the calculated atomization energy, formation energy, c lattice constant, spring constant, electron affinity, band gap, electronic and total dielectric constant of 175 Group 14 element-based hybrid polymers. The dielectric constant is computed using the single-chain approach, while the PBE functional is used to obtain the band gap values.

Table D.1: Atomization energy (E_a), formation energy (E_f), lattic constant (c), spring constant (k), electron affinity (EA), band gap (E_g), electronic (ϵ_e) and total (ϵ_t) dielectric constant of 175 Group 14 element-based hybrid polymers

System	E_a (eV)	E_f (eV)	c ($\AA$)	k	EA (eV)	E_g	ϵ_e	ϵ_t
CH2-CH2-CH2-CH2	-16.33	-2.79	5.12	119.39	0.57	6.39	2.50	2.50
CH2-CH2-CH2-GeF2	-15.64	-8.00	5.92	86.60	1.49	5.42	2.79	4.01
CH2-CH2-CH2-GeCl2	-14.44	-3.08	6.73	52.52	3.27	3.37	2.47	3.60
CH2-CH2-CH2-SiF2	-16.45	-10.30	5.71	96.80	0.94	5.78	2.52	3.15
CH2-CH2-CH2-SiCl2	-15.30	-5.59	5.74	78.46	0.99	5.75	2.66	3.01
CH2-CH2-CH2-SnF2	-15.29	-7.24	6.63	72.73	2.90	4.09	2.42	6.24
CH2-CH2-CH2-SnCl2	-14.36	-3.40	6.84	47.16	3.12	3.33	2.53	4.26
CH2-CH2-GeF2-GeF2	-14.97	-13.30	6.80	64.06	2.88	4.14	3.72	7.03
CH2-CH2-GeCl2-GeF2	-14.10	-9.69	6.79	50.72	3.18	3.77	3.63	6.10
CH2-CH2-GeCl2-GeCl2	-13.24	-6.10	6.77	37.82	3.47	3.47	3.57	5.39
CH2-CH2-SiF2-GeF2	-15.33	-13.82	6.03	82.76	3.70	3.55	2.35	5.72
CH2-CH2-SiF2-GeCl2	-14.82	-11.65	6.59	56.45	2.33	4.46	3.10	4.67
CH2-CH2-SiF2-SiF2	-16.40	-17.18	6.38	75.48	1.92	4.75	2.80	4.36
CH2-CH2-SiCl2-GeF2	-14.57	-10.62	6.60	56.12	2.20	4.63	3.16	5.18
CH2-CH2-SiCl2-GeCl2	-13.70	-7.04	6.60	42.59	2.48	4.23	3.12	4.54
CH2-CH2-SiCl2-SiF2	-15.29	-12.58	6.40	60.97	2.05	4.65	2.83	3.85
CH2-CH2-SiCl2-SiCl2	-14.17	-7.99	6.43	47.83	2.13	4.60	2.87	3.70

CH2-CH2-SnF2-GeF2	-14.31	-11.31	6.56	72.69	3.63	2.79	3.10	12.40
CH2-CH2-SnF2-GeCl2	-13.84	-9.28	7.14	39.91	3.93	3.09	4.31	11.33
CH2-CH2-SnF2-SiF2	-15.22	-14.02	6.32	90.17	3.45	3.72	2.59	5.00
CH2-CH2-SnF2-SiCl2	-14.09	-9.38	6.44	55.99	3.28	3.36	2.65	5.69
CH2-CH2-SnF2-SnF2	-14.48	-12.65	6.63	50.57	4.44	2.98	4.27	10.46
CH2-CH2-SnCl2-GeF2	-13.40	-7.54	7.08	29.99	4.55	1.89	4.39	14.20
CH2-CH2-SnCl2-GeCl2	-12.55	-3.99	7.44	18.49	4.29	2.11	3.57	7.88
CH2-CH2-SnCl2-SiF2	-14.31	-10.24	6.34	57.21	4.05	2.92	2.76	4.32
CH2-CH2-SnCl2-SiCl2	-13.19	-5.61	6.95	24.37	3.55	2.82	2.80	3.69
CH2-CH2-SnCl2-SnF2	-13.66	-9.21	7.51	32.35	4.16	2.70	5.06	12.84
CH2-CH2-SnCl2-SnCl2	-12.84	-5.82	6.53	19.98	4.21	2.65	3.88	7.31
CH2-GeF2-CH2-GeF2	-14.92	-13.10	6.67	79.13	2.19	5.30	3.00	5.40
CH2-GeF2-CH2-GeCl2	-14.03	-9.39	6.69	66.19	2.21	4.75	3.04	4.75
CH2-GeF2-CH2-SiF2	-15.79	-15.65	6.47	82.39	1.80	5.59	2.80	4.70
CH2-GeF2-CH2-SiCl2	-14.61	-10.80	6.52	72.82	1.76	5.30	2.90	4.38
CH2-GeF2-CH2-SnF2	-14.61	-12.50	6.99	62.28	2.91	4.51	3.24	7.31
CH2-GeF2-CH2-SnCl2	-13.80	-9.15	6.99	35.22	2.83	4.13	3.18	5.75
CH2-GeF2-GeF2-GeF2	-14.26	-18.42	7.68	56.59	3.94	3.33	4.41	11.75
CH2-GeF2-GeCl2-GeF2	-13.40	-14.86	7.63	47.00	4.01	3.13	4.10	9.92
CH2-GeF2-SiF2-GeF2	-14.88	-19.98	7.45	61.18	3.30	3.81	3.70	8.07
CH2-GeF2-SiCl2-GeF2	-13.79	-15.47	7.44	49.64	3.24	3.74	3.56	6.90
CH2-GeF2-SnF2-GeF2	-14.03	-18.17	7.93	45.86	4.37	3.07	4.74	14.84

CH2-GeF2-SnCl2-GeF2	-13.24	-14.89	7.93	38.55	4.33	2.89	4.45	11.60
CH2-GeCl2-CH2-GeCl2	-13.08	-5.48	6.81	56.45	2.56	3.99	3.27	4.49
CH2-GeCl2-CH2-SiF2	-14.89	-11.90	6.50	69.93	2.11	4.83	2.86	4.17
CH2-GeCl2-CH2-SiCl2	-13.65	-6.84	6.66	49.54	2.40	4.17	3.15	4.20
CH2-GeCl2-CH2-SnF2	-13.72	-8.80	7.01	54.76	2.78	4.28	3.28	6.25
CH2-GeCl2-CH2-SnCl2	-12.88	-5.33	7.09	33.87	2.96	3.70	3.40	5.27
CH2-GeCl2-GeF2-GeF2	-13.40	-14.84	7.66	46.16	3.93	3.09	4.26	9.92
CH2-GeCl2-GeF2-GeCl2	-12.50	-11.11	7.69	37.86	4.37	2.74	4.32	8.42
CH2-GeCl2-GeCl2-GeF2	-12.52	-11.22	7.61	35.63	4.25	2.93	3.99	8.70
CH2-GeCl2-GeCl2-GeCl2	-11.66	-7.64	7.64	26.29	4.28	2.62	4.09	7.67
CH2-GeCl2-SiF2-GeF2	-13.98	-16.23	7.45	49.44	3.51	3.64	3.60	6.94
CH2-GeCl2-SiF2-GeCl2	-13.11	-12.65	7.50	42.59	3.52	3.31	3.69	6.85
CH2-GeCl2-SiCl2-GeF2	-12.90	-11.81	7.44	39.21	3.48	3.55	3.50	6.31
CH2-GeCl2-SiCl2-GeCl2	-12.04	-8.23	7.50	31.72	3.57	3.19	3.61	5.87
CH2-GeCl2-SnF2-GeF2	-13.13	-14.44	7.94	37.82	4.65	2.76	4.67	11.69
CH2-GeCl2-SnF2-GeCl2	-12.28	-10.89	7.96	29.78	4.70	2.44	4.71	10.23
CH2-GeCl2-SnCl2-GeF2	-12.38	-11.31	7.93	28.75	4.35	2.64	4.39	10.89
CH2-GeCl2-SnCl2-GeCl2	-11.51	-7.69	7.97	21.14	4.59	2.34	4.49	9.65
CH2-SiF2-CH2-SiF2	-16.64	-18.12	6.27	92.24	1.33	5.96	2.55	3.99
CH2-SiF2-CH2-SiCl2	-15.46	-13.25	6.33	79.30	1.23	5.79	2.68	3.91
CH2-SiF2-CH2-SnF2	-15.48	-15.06	6.78	63.45	2.69	4.57	3.07	5.96
CH2-SiF2-CH2-SnCl2	-14.67	-11.67	6.80	40.09	2.73	4.17	3.04	4.83

CH2-SiF2-GeF2-GeF2	-15.02	-20.55	7.44	59.86	3.39	3.76	3.81	9.57
CH2-SiF2-GeF2-GeCl2	-14.15	-16.93	7.45	48.64	3.52	3.44	3.71	7.34
CH2-SiF2-GeF2-SiF2	-15.76	-22.56	7.21	64.28	3.00	4.12	3.25	8.38
CH2-SiF2-GeCl2-GeF2	-14.17	-17.01	7.41	48.80	3.40	3.57	3.61	7.43
CH2-SiF2-GeCl2-GeCl2	-13.30	-13.39	7.41	39.10	3.60	3.25	3.55	6.65
CH2-SiF2-GeCl2-SiF2	-14.91	-19.05	7.20	53.59	2.94	4.02	3.15	5.80
CH2-SiF2-SiF2-GeF2	-15.64	-22.11	7.22	64.22	3.12	3.90	3.37	6.32
CH2-SiF2-SiF2-GeCl2	-14.78	-18.50	7.23	53.38	3.03	3.73	3.31	5.87
CH2-SiF2-SiF2-SiF2	-16.38	-24.12	6.97	69.06	2.85	3.94	3.00	6.15
CH2-SiF2-SiCl2-GeF2	-14.56	-17.63	7.22	52.69	3.02	3.87	3.27	6.05
CH2-SiF2-SiCl2-GeCl2	-13.69	-14.02	7.24	42.67	3.10	3.66	3.23	5.33
CH2-SiF2-SiCl2-SiF2	-15.30	-19.68	6.99	57.01	2.77	3.94	2.95	4.86
CH2-SiF2-SnF2-GeF2	-14.77	-20.21	7.68	48.67	3.70	3.54	4.06	13.73
CH2-SiF2-SnF2-GeCl2	-13.91	-16.61	7.70	40.43	3.91	3.14	4.00	11.82
CH2-SiF2-SnF2-SiF2	-15.22	-21.09	6.94	73.94	3.49	3.23	2.56	11.15
CH2-SiF2-SnCl2-GeF2	-13.99	-16.94	7.69	41.03	3.73	3.31	3.87	9.27
CH2-SiF2-SnCl2-GeCl2	-13.12	-13.34	7.72	30.73	3.94	2.96	3.86	8.44
CH2-SiF2-SnCl2-SiF2	-14.71	-18.88	7.46	44.42	3.15	3.76	3.33	7.19
CH2-SiCl2-CH2-SiCl2	-14.21	-8.11	6.52	70.47	1.56	5.03	2.98	3.80
CH2-SiCl2-CH2-SnF2	-14.31	-10.26	6.83	56.68	2.62	4.45	3.15	5.67
CH2-SiCl2-CH2-SnCl2	-13.47	-6.74	6.92	37.90	2.91	3.80	3.30	4.85
CH2-SiCl2-GeF2-GeF2	-13.90	-15.92	7.45	49.52	3.37	3.57	3.79	8.40

CH2-SiCl2-GeF2-GeCl2	-13.01	-12.22	7.50	41.20	3.73	3.11	3.87	8.69
CH2-SiCl2-GeF2-SiF2	-14.63	-17.91	7.23	54.61	2.90	4.00	3.27	7.68
CH2-SiCl2-GeF2-SiCl2	-13.45	-13.08	7.32	46.08	3.28	3.64	3.44	6.04
CH2-SiCl2-GeCl2-GeF2	-13.03	-12.30	7.42	38.70	3.72	3.37	3.63	7.39
CH2-SiCl2-GeCl2-GeCl2	-12.16	-8.69	7.48	29.96	3.90	2.93	3.75	6.53
CH2-SiCl2-GeCl2-SiF2	-13.76	-14.31	7.24	42.52	3.02	3.94	3.19	5.67
CH2-SiCl2-GeCl2-SiCl2	-12.63	-9.67	7.33	34.37	3.28	3.49	3.37	5.48
CH2-SiCl2-SiF2-GeF2	-14.47	-17.30	7.24	53.41	3.17	3.79	3.34	6.10
CH2-SiCl2-SiF2-GeCl2	-13.61	-13.69	7.32	45.69	3.27	3.44	3.45	5.74
CH2-SiCl2-SiF2-SiF2	-15.20	-19.26	7.02	57.96	2.85	3.86	3.02	5.45
CH2-SiCl2-SiF2-SiCl2	-14.11	-14.79	7.14	50.88	2.93	3.60	3.18	5.54
CH2-SiCl2-SiCl2-GeF2	-13.43	-13.00	7.25	42.94	3.00	3.75	3.29	5.77
CH2-SiCl2-SiCl2-GeCl2	-12.55	-9.32	7.35	34.23	3.34	3.33	3.41	5.24
CH2-SiCl2-SiCl2-SiF2	-14.18	-15.06	7.05	46.13	2.75	3.87	2.99	4.76
CH2-SiCl2-SiCl2-SiCl2	-13.03	-10.33	7.21	38.80	2.99	3.59	3.17	4.62
CH2-SiCl2-SnF2-GeF2	-13.66	-15.62	7.71	40.34	3.82	3.25	4.10	13.13
CH2-SiCl2-SnF2-GeCl2	-12.78	-11.96	7.77	33.49	4.14	2.80	4.17	9.60
CH2-SiCl2-SnF2-SiF2	-14.36	-17.51	7.46	44.02	3.14	3.80	3.45	9.16
CH2-SiCl2-SnF2-SiCl2	-13.24	-12.86	7.56	36.48	3.51	3.34	3.63	9.14
CH2-SiCl2-SnCl2-GeF2	-12.87	-12.34	7.73	31.46	3.87	3.04	3.95	8.93
CH2-SiCl2-SnCl2-GeCl2	-11.99	-8.69	7.82	23.15	4.22	2.64	4.08	7.97
CH2-SiCl2-SnCl2-SiF2	-13.59	-14.27	7.52	33.61	3.26	3.54	3.40	7.09

CH2-SiCl2-SnCl2-SiCl2	-12.46	-9.61	7.66	26.31	3.67	3.10	3.61	6.70
CH2-SnF2-CH2-SnF2	-14.28	-11.84	7.30	59.07	3.22	4.14	3.44	9.18
CH2-SnF2-CH2-SnCl2	-13.48	-8.51	7.32	47.80	3.12	3.91	3.40	7.35
CH2-SnF2-GeF2-GeF2	-13.99	-17.98	7.93	35.52	4.33	3.07	4.72	17.39
CH2-SnF2-GeF2-GeCl2	-13.13	-14.43	7.94	37.16	4.24	2.84	4.68	15.65
CH2-SnF2-GeF2-SiF2	-14.75	-20.11	7.68	40.15	3.83	3.40	4.16	11.50
CH2-SnF2-GeF2-SiCl2	-13.64	-15.53	7.71	36.39	3.80	3.20	4.20	11.96
CH2-SnF2-GeF2-SnF2	-13.71	-17.55	8.25	43.44	4.60	2.77	5.23	21.61
CH2-SnF2-GeCl2-GeF2	-13.14	-14.45	7.90	36.55	4.39	2.81	4.43	14.29
CH2-SnF2-GeCl2-GeCl2	-12.28	-10.90	7.93	28.67	4.34	2.64	4.43	11.78
CH2-SnF2-GeCl2-SiF2	-13.91	-16.60	7.68	33.65	3.96	3.11	3.97	10.78
CH2-SnF2-GeCl2-SiCl2	-12.79	-12.02	7.72	30.40	3.95	2.93	4.05	10.53
CH2-SnF2-GeCl2-SnF2	-12.87	-14.03	8.24	35.57	4.65	2.53	4.91	17.74
CH2-SnF2-SiF2-GeF2	-14.85	-20.50	7.69	41.22	3.42	3.62	3.94	10.55
CH2-SnF2-SiF2-GeCl2	-13.68	-15.70	7.71	37.09	3.60	3.46	3.88	9.21
CH2-SnF2-SiF2-SiF2	-15.25	-21.20	7.43	43.90	3.39	3.68	3.65	8.09
CH2-SnF2-SiF2-SiCl2	-14.17	-16.75	7.47	42.41	3.29	3.59	3.64	8.73
CH2-SnF2-SiF2-SnF2	-14.21	-18.60	7.99	47.28	3.89	3.37	4.27	14.69
CH2-SnF2-SiCl2-GeF2	-13.45	-14.78	7.70	39.62	3.78	3.44	3.79	8.93
CH2-SnF2-SiCl2-GeCl2	-12.62	-11.33	7.73	32.49	3.71	3.26	3.79	7.81
CH2-SnF2-SiCl2-SiF2	-14.21	-16.88	7.46	40.97	3.42	3.55	3.54	9.18
CH2-SnF2-SiCl2-SiCl2	-13.12	-12.41	7.52	33.95	3.41	3.42	3.58	7.26

CH2-SnF2-SiCl2-SnF2	-13.16	-14.26	8.01	38.01	3.97	3.14	4.11	12.11
CH2-SnF2-SnF2-GeF2	-13.69	-17.44	8.13	36.34	4.80	2.88	4.97	26.04
CH2-SnF2-SnF2-GeCl2	-12.86	-14.01	8.18	30.29	4.72	2.62	4.99	18.35
CH2-SnF2-SnF2-SiF2	-14.51	-19.79	7.86	34.05	4.13	3.16	4.34	16.69
CH2-SnF2-SnF2-SiCl2	-13.41	-15.26	7.93	31.70	4.11	2.98	4.43	18.17
CH2-SnF2-SnF2-SnF2	-13.50	-17.36	8.43	37.71	4.84	2.65	5.37	46.94
CH2-SnF2-SnCl2-GeF2	-12.99	-14.50	8.16	30.58	4.59	2.68	4.71	17.85
CH2-SnF2-SnCl2-GeCl2	-12.14	-10.98	8.23	25.58	4.58	2.45	4.83	14.46
CH2-SnF2-SnCl2-SiF2	-13.73	-16.56	7.91	29.95	4.12	2.98	4.16	16.62
CH2-SnF2-SnCl2-SiCl2	-12.63	-12.02	8.00	25.61	4.18	2.76	4.33	11.36
CH2-SnF2-SnCl2-SnF2	-12.73	-14.12	8.46	31.30	4.80	2.48	5.10	26.25
CH2-SnCl2-CH2-SnCl2	-12.66	-5.08	7.40	39.87	3.25	3.54	3.52	6.10
CH2-SnCl2-GeF2-GeF2	-13.20	-14.72	7.94	32.37	4.25	2.86	4.54	14.38
CH2-SnCl2-GeF2-GeCl2	-12.33	-11.10	7.99	28.30	4.34	2.58	4.65	11.04
CH2-SnCl2-GeF2-SiF2	-13.96	-16.80	7.71	35.79	3.88	3.14	4.02	9.32
CH2-SnCl2-GeF2-SiCl2	-12.83	-12.17	7.78	31.61	4.01	2.88	4.19	8.72
CH2-SnCl2-GeF2-SnF2	-12.93	-14.29	8.27	32.15	4.46	2.63	5.02	19.53
CH2-SnCl2-GeF2-SnCl2	-12.14	-10.98	8.33	26.48	4.54	2.40	5.05	13.68
CH2-SnCl2-GeCl2-GeF2	-12.35	-11.17	7.92	28.92	4.32	2.67	4.30	10.15
CH2-SnCl2-GeCl2-GeCl2	-11.48	-7.57	7.98	22.23	4.47	2.41	4.45	9.78
CH2-SnCl2-GeCl2-SiF2	-13.11	-13.29	7.71	31.09	3.97	2.90	3.88	8.01
CH2-SnCl2-GeCl2-SiCl2	-11.99	-8.65	7.80	23.52	4.18	2.65	4.10	8.06

CH2-SnCl2-GeCl2-SnF2	-12.09	-10.77	8.27	27.52	4.58	2.41	4.78	13.97
CH2-SnCl2-GeCl2-SnCl2	-11.29	-7.46	8.35	19.32	4.68	2.21	4.86	12.38
CH2-SnCl2-SiF2-GeF2	-13.80	-16.20	7.71	36.03	3.44	3.42	3.81	9.84
CH2-SnCl2-SiF2-GeCl2	-12.94	-12.59	7.79	33.11	3.54	3.14	3.93	7.88
CH2-SnCl2-SiF2-SiF2	-14.56	-18.30	7.47	38.88	3.25	3.49	3.54	7.11
CH2-SnCl2-SiF2-SiCl2	-13.44	-13.68	7.57	32.31	3.38	3.26	3.70	6.88
CH2-SnCl2-SiF2-SnF2	-13.51	-15.69	8.03	34.98	3.64	3.20	4.16	12.59
CH2-SnCl2-SiF2-SnCl2	-12.84	-12.86	8.12	30.24	3.76	2.95	4.20	10.47
CH2-SnCl2-SiCl2-GeF2	-12.72	-11.72	7.72	31.89	3.58	3.23	3.71	7.85
CH2-SnCl2-SiCl2-GeCl2	-11.87	-8.18	7.83	24.80	3.80	2.94	3.87	7.05
CH2-SnCl2-SiCl2-SiF2	-13.48	-13.85	7.52	33.50	3.41	3.31	3.48	6.35
CH2-SnCl2-SiCl2-SiCl2	-12.36	-9.22	7.66	25.57	3.62	3.04	3.68	6.14
CH2-SnCl2-SiCl2-SnF2	-12.44	-11.24	8.07	29.89	3.85	2.98	4.05	10.28
CH2-SnCl2-SiCl2-SnCl2	-11.64	-7.91	8.19	21.92	4.03	2.71	4.16	8.90
CH2-SnCl2-SnF2-GeF2	-12.93	-14.29	8.16	30.77	4.69	2.65	4.81	16.44
CH2-SnCl2-SnF2-GeCl2	-12.09	-10.80	8.24	24.57	4.76	2.36	4.99	17.29
CH2-SnCl2-SnF2-SiF2	-13.65	-16.22	7.92	30.20	4.21	2.94	4.24	15.41
CH2-SnCl2-SnF2-SiCl2	-12.56	-11.75	8.02	24.74	4.32	2.67	4.46	15.52
CH2-SnCl2-SnF2-SnF2	-12.67	-13.89	8.48	27.82	4.83	2.48	5.27	26.64
CH2-SnCl2-SnF2-SnCl2	-11.91	-10.70	8.58	20.06	4.87	2.25	5.37	18.88
CH2-SnCl2-SnCl2-GeF2	-12.17	-11.11	8.21	23.15	4.62	2.50	4.65	12.72
CH2-SnCl2-SnCl2-GeCl2	-11.33	-7.63	8.34	17.00	4.75	2.19	4.93	11.60

CH2-SnCl2-SnCl2-SiF2	-12.90	-13.10	7.98	25.63	4.21	2.75	4.12	10.57
CH2-SnCl2-SnCl2-SiCl2	-11.82	-8.64	8.17	17.94	4.46	2.42	4.52	9.89
CH2-SnCl2-SnCl2-SnF2	-11.91	-10.73	8.54	22.30	4.81	2.31	5.11	17.14
CH2-SnCl2-SnCl2-SnCl2	-11.15	-7.56	8.70	15.24	4.91	2.05	5.37	14.36

Appendix E

Chemo-structural feature vector generation - python code

```
import csv
import re

#given a sequence 1122, which is CH2-CH2-CH2-CH2
#1111 contains 4 1-1, 4 1-1-1, overlapping



def Get_Reverse(fragment):
        #to generate fragements for each item in the target seq
        #for example, '123' --> '123' and '321'
        #It is not permutation, but to generate a reverse
        str_frag = [str(c) for c in fragment]
        fragement_reverse = ''.join(list(reversed(str_frag)))
        return fragement_reverse

def function(string, str_to_search_for):
   count = 0
   count_1 = 0
   count_2 = 0
   if str_to_search_for == Get_Reverse(str_to_search_for):
      for x in xrange(len(string) - len(str_to_search_for) + 1):
           if string[x:x+len(str_to_search_for)] == str_to_search_for:
                count += 1
      return count
   else:
      for x in xrange(len(string) - len(str_to_search_for) + 1):
           if string[x:x+len(str_to_search_for)] == str_to_search_for:
                count_1 += 1
      for x in xrange(len(string) - len(str_to_search_for) + 1):
           if string[x:x+len(str_to_search_for)] == Get_Reverse(str_to_search_for):
                count_2 += 1
      return count_1 + count_2



data = [row for row in csv.reader(open('System-Info.csv', 'rU'))]

polymer_list = []
for polymer in data:
  for poly_seq in polymer:
   poly_seq = poly_seq.replace('-', '')
   if 'CH2' in poly_seq:
    poly_seq = poly_seq.replace("CH2", "1")
   if 'NH' in poly_seq:
    poly_seq = poly_seq.replace("NH", "2")
   if 'CO' in poly_seq:
    poly_seq = poly_seq.replace("CO", "3")
   if 'C6H4' in poly_seq:
    poly_seq = poly_seq.replace("C6H4", "4")
   if 'C4H2S' in poly_seq:
    poly_seq = poly_seq.replace("C4H2S", "5")
   if 'CS' in poly_seq:
    poly_seq = poly_seq.replace("CS", "6")
   if 'O' in poly_seq:
    poly_seq = poly_seq.replace("O", "7")
   polymer_list.append(poly_seq)

target_seq = ['3', '4', '5', '6', '7', '11', '12', '14', '15', '16', '17', '23',
              '27', '35', '37', '44', '45', '46', '55', '56', '111', '113', '115',
              '117', '125', '134', '137', '144', '145', '146', '147', '157', '245',
              '246', '247', '255', '256', '344', '347', '356', '444', '447', '455',
              '213', '215', '232', '254', '263', '264', '265', '272', '276', '316',
              '323', '374', '415', '425', '435', '516', '526', '527', '537', '545',
              '565', '627', '646', '656', '717', '727', '747']
```

```
feature_mat = []
for item in polymer_list:
        feature_vec = []
        for item_target in target_seq:
                nblock = len(item)
                double_poly = item + item
                item_target_reverse = Get_Reverse(item_target)
                #pat = re.compile("%s|%s" %(item_target,item_target_reverse))
                n1 = function(item, item_target)
                n2 = function(double_poly, item_target)
                #n1 = len(pat.findall(item))
                #n2 = len(pat.findall(double_poly))
                #n = n1 + (n2-2*n1)
                n = n2 - n1
                n = n / float(nblock)
                feature_vec.append(n)
        feature_mat.append(feature_vec)


with open('feature.csv', 'wb') as f:
 wtr = csv.writer(f, delimiter= ' ')
 wtr.writerows(feature_mat)
```

Appendix F

Kernel ridge regression (KRR) - matlab code

main.m

```
clear all
clc

% Predicting data using Kernel Ridge Regression model, where we require a
% feature vector with a given set of properties to be predicted. The
% feature vector uniquely identifies the candidates and this case being the
% chemostructural information or the charge density.

% Read the feature vector along with the property to be predicted
Feature = xlsread('input.xls');
% Property = xlsread('Vacancy_formation.xlsx');
Property = xlsread('output.xls');

global Inputs

% Inputs
Inputs.n_test=67;               % Number of Test group
Inputs.cv_group=5;              % k-fold cross-validation

% Parameters
Inputs.lambda = [0.00001 0.0001 0.0002 0.0006 0.0007 0.001 0.002 0.005
    0.006 0.01 0.02 0.07 0.1 0.5 0.8 1];
% Inputs.sigma = [21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39
40 41 42 43 44 45 46 47 48 49 50];
% Inputs.lambda = [0.01:0.01:2];
% Inputs.sigma = [0.01:0.01:1];
% Inputs.sigma = [0.01:0.1:5];
%Inputs.sigma = [0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0 1.1 1.5 1.7 2 2.3
3 3.5 4 4.5 5 5.5 6 6.5 7 7.5 8];
Inputs.sigma = [0.01 0.05 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0 1.1 1.5
    1.7 2 2.3 3 3.5 4 4.5 5 5.5 6 6.5 7 7.5 8 9 10]
R = zeros(100,1);
for seed_value = 1:100,
KRR;
R(seed_value)= r;
end
xlswrite('RESULTS_R.xls',R)

seed_best = find (R==max(R))

KRR_Final;
```

crossvaliation.m

```
function error_k = cross_validation(X,Y,sigma,lambda)

global Inputs

k = Inputs.cv_group;
d = size(X,1);
i = d/k;

%rng(Inputs.mySeed2)
n = randperm(d);
B = Y(n,:);

for j=1:k,
 a_j = 1+i*(j-1);
 b_j = j*i;

P = n(:,a_j:b_j);
T = n;
T(:,a_j:b_j) = [];

A_j_pre = X(P,T);
A_j_train = X(T,T);
B_j_pre = B(a_j:b_j,1);
B_j_train = B;
B_j_train(a_j:b_j,:) = [];

alpha_j = kernel(A_j_train,B_j_train,sigma,lambda);
error_j = cv(A_j_pre,B_j_pre,alpha_j,sigma);

error_k = 0;
error_k = error_k + error_j;

end

end
```

cv.m

```
function error = cv(X,Y,alpha,sigma)
K = exp(-(X.^2)/(2*sigma^2));
Y_pre = K*alpha;
A = Y_pre - Y;
error = sqrt(sum(A.^2)/size(Y,1));
end
```

kernel.m

```
function alpha = kernel(X,Y,sigma,lambda)
    K = exp(-(X.^2)/(2*sigma^2));
    I = eye(size(K,2));
    alpha = (K+lambda*I)\Y;
end
```

KRR.m

```
global Inputs Outputs
%% Obtaining the Training and Test Set


% Create a random number list to determine the training and test set
mySeed = seed_value;
rng(mySeed);
Outputs.length=size(Property,1);
n = randperm(Outputs.length);
Outputs.n_train= Outputs.length-Inputs.n_test;

% Randomize the A and B input data set
A = Feature(n,:);
B = Property(n,:);

% Obtain Training and Test Set by Concatenation
Feature_train = A;
Feature_train(1:Inputs.n_test,:) = [];
Property_train = B;
Property_train(1:Inputs.n_test,:) = [];

Feature_test = A(1:Inputs.n_test,:);
Property_test = B(1:Inputs.n_test,:);


%% Distance Matrix

% Since the distance matrix is symmetrical, the loop below
% calculates an upper triangular matri
for i=1:Outputs.n_train
    for j=i:Outputs.n_train
        Dist(j,i)=sqrt(sum((Feature_train(i,:)-Feature_train(j,:)).^2));
    end
end

%Calculates the final distance matrix
Distance_matrix = Dist+Dist';



%% Model Selection


for i = 1:length(Inputs.lambda)
    for j = 1:length(Inputs.sigma)
 a = Inputs.sigma(j);
 b = Inputs.lambda(i);
error = cross_validation(Distance_matrix,Property_train,a,b)/Inputs.cv_group;
G(i,j) = error;

    end
end
[c,d]=find(G==min(min(G)));


 %% Prediction

lambda = Inputs.lambda(c);
sigma = Inputs.sigma(d);

alpha = kernel(Distance_matrix,Property_train,sigma,lambda);
```

```
K = exp(-(Distance_matrix.^2)/(2*sigma^2));
Property_train_pre = K*alpha;

for i=1:Inputs.n_test,
    for j=1:Outputs.n_train,
       C = Feature_test(i,:)-Feature_train(j,:);
       D(j,1) = alpha(j,1)*exp(-sum(C.^2)/(2*sigma^2));
    end

    Property_pre(i,1) = sum(D);

end

Property_DFT = [Property_train; Property_test];
Property_KRR = [Property_train_pre; Property_pre];
r_all = corr(Property_DFT, Property_KRR);
r = corr(Property_test, Property_pre);
```

KRRFinal.m

```
global Inputs Outputs
%% Obtaining the Training and Test Set


% Create a random number list to determine the training and test set
mySeed = seed_best;
rng(mySeed);
Outputs.length=size(Property,1);
n = randperm(Outputs.length);
Outputs.n_train= Outputs.length-Inputs.n_test;

% Randomize the A and B input data set
A = Feature(n,:);
B = Property(n,:);

% Obtain Training and Test Set by Concatenation
Feature_train = A;
Feature_train(1:Inputs.n_test,:) = [];
Property_train = B;
Property_train(1:Inputs.n_test,:) = [];

Feature_test = A(1:Inputs.n_test,:);
Property_test = B(1:Inputs.n_test,:);


%% Distance Matrix

% Since the distance matrix is symmetrical, the loop below
% calculates an upper triangular matri
for i=1:Outputs.n_train
    for j=i:Outputs.n_train
        Dist(j,i)=sqrt(sum((Feature_train(i,:)-Feature_train(j,:)).^2));
    end
end

%Calculates the final distance matrix
Distance_matrix = Dist+Dist';



%% Model Selection


for i = 1:length(Inputs.lambda)
    for j = 1:length(Inputs.sigma)
 a = Inputs.sigma(j);
 b = Inputs.lambda(i);
error = cross_validation(Distance_matrix,Property_train,a,b)/Inputs.cv_group;
G(i,j) = error;

    end
end
[c,d]=find(G==min(min(G)));


 %% Prediction

lambda = Inputs.lambda(c);
sigma = Inputs.sigma(d);

alpha = kernel(Distance_matrix,Property_train,sigma,lambda);
```

```
K = exp(-(Distance_matrix.^2)/(2*sigma^2));
Property_train_pre = K*alpha;

for i=1:Inputs.n_test,
    for j=1:Outputs.n_train,
       C = Feature_test(i,:)-Feature_train(j,:);
       D(j,1) = alpha(j,1)*exp(-sum(C.^2)/(2*sigma^2));
    end

    Property_pre(i,1) = sum(D);

end

%E_train = Property_train - Property_train_pre;
%E=[E_pre;E_train];
%RMS_test = sqrt(sum(E_pre.*E_pre)/size(E_pre,1));
%RMS = sqrt(sum(E.*E)/size(E,1));
Property_DFT = [Property_train; Property_test];
Property_KRR = [Property_train_pre; Property_pre];
r_all = corr(Property_DFT, Property_KRR)
r = corr(Property_test, Property_pre)
x = 0:7;
y = x;

plot(Property_train, Property_train_pre,'MarkerFaceColor',[0 0 1],
    'MarkerEdgeColor',[0 0 0],...
    'MarkerSize',10,...
    'Marker','square',...
    'LineWidth',2,...
    'LineStyle','none',...
    'Color',[1 0 0],...
    'DisplayName','Training-4Block');
hold on


plot(Property_test, Property_pre,'MarkerFaceColor',[0 1 0],
    'MarkerEdgeColor',[0 0 0],...
    'MarkerSize',10,...
    'Marker','square',...
    'LineWidth',2,...
    'LineStyle','none',...
    'Color',[1 0 0],...
     'DisplayName','Test-4Block');
plot(x,y,'k','LineWidth',2)
```

Appendix G

Genetic algorithm for feature selection - matlab code

GA.m

```
%           Binary Genetic Algorithm
%
%  feature selection using genetic algorithm

clear all
clc

%--------------------------------------------------------
% I. Setup input and output
global Inputs

Attribute = xlsread('input.xls');
Property = xlsread('output.xls');
%--------------------------------------------------------
% II. Stopping criteria
maxit=100; % max number of iterations
max_accuracy=0.98; % minimum cost
%--------------------------------------------------------
% III. GA parameters
popsize=16; % set population size
mutrate=.15; % set mutation rate
selection=0.5; % fraction of population
% kept
keep=floor(selection*popsize); % #population members
% that survive
%--------------------------------------------------------
% Create the initial population
iga=0; % generation counter
% initialized
pop=round(rand(popsize,size(Attribute,2))); % random population of
% 1s and 0s
for i_GA = 1:popsize,
    A_GA = pop(i_GA,:);
    A_GA = A_GA(ones(size(Attribute,1),1),:);
    B_GA = times(A_GA,Attribute);
    B_GA( :, ~any(B_GA,1) ) = [];
    Feature = B_GA;

% Inputs
Inputs.n_test=45;               % Number of Test group
Inputs.cv_group=5;              % k-fold cross-validation

% Parameters
Inputs.lambda = [0.00001 0.0001 0.0002 0.0006 0.0007 0.001 0.002 0.005
                 0.006 0.01 0.02 0.07 0.1 0.5 0.8 1];
%Inputs.sigma = [0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0 1.1 1.5 1.7 2
                 2.3 3 3.5 4 4.5 5 5.5 6 6.5 7 7.5 8];
Inputs.sigma = [1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20]
R = zeros(100,1);
for seed_value = 1:100,
KRR;
R(seed_value)= r_init;
end
seed_best = find (R==max(R));
KRR_Final;
accuracy(i_GA,1) = r;
end

[accuracy,ind]=sort(accuracy,'descend'); % min cost in element 1
pop=pop(ind,:); % sorts population with
% highest accuracy first
maxc(1)=max(accuracy); % maxc contains max of
% population
```

```
meanc(1)=mean(accuracy); % meanc contains mean
% of population
%-------------------------------------------------------
% Iterate through generations
while iga<maxit
iga=iga+1; % increments generation counter
%-------------------------------------------------------
% Pair and mate
M=ceil((popsize-keep)/2); % number of matings
prob=flipud([1:keep]'/sum([1:keep]));% weights
% chromosomes based
% upon position in
% list
odds=[0 cumsum(prob(1:keep))']; % probability distribution function
pick1=rand(1,M); % mate #1
pick2=rand(1,M); % mate #2
% ma and pa contain the indicies of the chromosomes
% that will mate
ic=1;
while ic<=M
for id=2:keep+1
if pick1(ic)<=odds(id) & pick1(ic)>odds(id-1)
ma(ic)=id-1;
end % if
if pick2(ic)<=odds(id) & pick2(ic)>odds(id-1)
pa(ic)=id-1;
end % if
end % id
ic=ic+1;
end % while
%-------------------------------------------------------
% Performs mating using single point crossover
ix=1:2:keep; % index of mate #1
for i_mate=1:size(ix,2)
xp=ceil(rand(1,M)*(size(Attribute,2)-1));
pop(keep+ix(i_mate),:)=
      [pop(ma(i_mate),1:xp(i_mate)) pop(pa(i_mate),xp(i_mate)+1:size(Attribute,2))];
% first offspring
pop(keep+ix(i_mate)+1,:)=
      [pop(pa(i_mate),1:xp(i_mate)) pop(ma(i_mate),xp(i_mate)+1:size(Attribute,2))];
end
%-------------------------------------------------------
% Mutate the population
nmut=ceil((popsize-1)*size(Attribute,2)*mutrate); % total number
% of mutations
mrow=ceil(rand(1,nmut)*(popsize-1))+1; % row to mutate
mcol=ceil(rand(1,nmut)*size(Attribute,2)); % column to mutate
for ii=1:nmut
pop(mrow(ii),mcol(ii))=abs(pop(mrow(ii),mcol(ii))-1);
% toggles bits
end % ii
%-------------------------------------------------------
% The population is re-evaluated for cost

for i_GA = 2:popsize,
    A_GA = pop(i_GA,:);
    A_GA = A_GA(ones(size(Attribute,1),1),:);
    B_GA = times(A_GA,Attribute);
    B_GA( :, ~any(B_GA,1) ) = [];
    Feature = B_GA;

% Inputs
Inputs.n_test=45;               % Number of Test group
Inputs.cv_group=5;              % k-fold cross-validation

% Parameters
Inputs.lambda = [0.00001 0.0001 0.0002 0.0006 0.0007 0.001 0.002
```

```
                0.005 0.006 0.01 0.02 0.07 0.1 0.5 0.8 1];
%Inputs.sigma = [0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0 1.1 1.5
                1.7 2 2.3 3 3.5 4 4.5 5 5.5 6 6.5 7 7.5 8];
Inputs.sigma = [1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20]
R = zeros(100,1);
for seed_value = 1:100,
KRR;
R(seed_value)= r_init;
end
seed_best = find (R==max(R));
KRR_Final;
accuracy(i_GA,1) = r;
end

[accuracy,ind]=sort(accuracy,'descend'); % min cost in element 1
pop=pop(ind,:); % sorts population with
% highest accuracy first
%------------------------------------------------------------
% Do statistics for a single nonaveraging run
maxc(iga+1)=max(accuracy);
meanc(iga+1)=mean(accuracy);
%------------------------------------------------------------
% Stopping criteria
if iga>maxit | accuracy(1)> max_accuracy
break
end
[iga accuracy(1)]
end %iga
%------------------------------------------------------------
% Displays the output
day=clock;
disp(datestr(datenum(day(1),day(2),day(3),day(4),day(5),day(6)),0))
format short g
disp(['popsize = ', num2str(popsize), ' mutrate = ', num2str(mutrate)])
disp(['#generations=', num2str(iga), ' best accuracy=', num2str(accuracy(1))])
disp(['best solution'])
disp([num2str(pop(1,:))])
disp('binary genetic algorithm')
figure(24)
iters=0:length(maxc)-1;
plot(iters,maxc,iters,meanc,'-');
xlabel('generation');ylabel('accuracy');
text(0,maxc(1),'best');text(1,maxc(2),'population average')
```