



# Scoping the polymer genome: A roadmap for rational polymer dielectrics design and beyond

Arun Mannodi-Kanakkithodi<sup>1,2,†</sup>, Anand Chandrasekaran<sup>1,†</sup>, Chiho Kim<sup>1</sup>,  
Tran Doan Huan<sup>1</sup>, Ghanshyam Pilania<sup>3,4</sup>, Venkatesh Botu<sup>5</sup>, Rampi Ramprasad<sup>1,\*</sup>

<sup>1</sup> Department of Materials Science and Engineering, Institute of Materials Science, University of Connecticut, 97 North Eagleville Road, Storrs, CT 06269, USA

<sup>2</sup> Center for Nanoscale Materials, Argonne National Laboratory, 9700 Cass Avenue, Lemont, IL 60439, USA

<sup>3</sup> Materials Science and Technology Division, Los Alamos National Laboratory, Los Alamos, NM 87545, USA

<sup>4</sup> Fritz-Haber-Institut der Max-Planck-Gesellschaft, Faradayweg 4-6, 14195 Berlin, Germany

<sup>5</sup> Science & Technology Division, 1 Science Center Dr., Corning Research & Development Corp., Corning, NY 14831, USA

The Materials Genome Initiative (MGI) has heralded a sea change in the philosophy of materials design. In an increasing number of applications, the successful deployment of novel materials has benefited from the use of computational methodologies, data descriptors, and machine learning. Polymers have long suffered from a lack of data on electronic, mechanical, and dielectric properties across large chemical spaces, causing a stagnation in the set of suitable candidates for various applications. Extensive efforts over the last few years have seen the fruitful application of MGI principles toward the accelerated discovery of attractive polymer dielectrics for capacitive energy storage. Here, we review these efforts, highlighting the importance of computational data generation and screening, targeted synthesis and characterization, polymer fingerprinting and machine-learning prediction models, and the creation of an online knowledgebase to guide ongoing and future polymer discovery and design. We lay special emphasis on the fingerprinting of polymers in terms of their genome or constituent atomic and molecular fragments, an idea that pays homage to the pioneers of the human genome project who identified the basic building blocks of the human DNA. By scoping the polymer genome, we present an essential roadmap for the design of polymer dielectrics, and provide future perspectives and directions for expansions to other polymer subclasses and properties.

## Introduction

### *The evolving materials research ecosystem*

In the 21st century, smart phones, tablets, and personal laptops are almost as mainstream as hand-held radio sets and 50 cent movies were in the middle of the 20th century. One can no longer imagine life without the transformative power of computers and the internet, upon which we rely for driving directions, shopping for clothes, checking the weather, and many other important things. A crucial driving force for many of these

innovations has been the development of new and complex materials which are capable of remarkable functionalities. Materials science—which involves studying the relationships between the structure and properties of metals, ceramics, or polymers—is unrecognizable from the times of Hall and Petch. While it was no mean feat by these researchers to relate the grain size of crystals to the strength of the material to develop the enormously useful Hall–Petch relationship [1,2], the mathematical and computational sophistication we can handle today makes the discovery of much more complex relationships, laws, and design rules a realistic possibility.

Indeed, the advent of computational power and methodologies in computational physics and chemistry have rendered

\* Corresponding author.

E-mail address: Ramprasad, R. ([rampi.ramprasad@uconn.edu](mailto:rampi.ramprasad@uconn.edu)).

† These authors contributed equally to this work.

trial-and-error-based experimental discovery of new materials for batteries, solar cells, and capacitors largely unnecessary. Computational screening or guidance is now increasingly being applied prior to laboratory synthesis and testing. Major theoretical advancements in classical and quantum mechanics, formulations of force-field simulations [3,4], and first principles-based methods like density functional theory (DFT) [5,6] are rapidly becoming critical components of modern materials research and discovery portfolios. DFT is a particularly popular method in this regard, given its widely accepted reliability and accuracy in studying the electronic structure and estimating the mechanical and thermodynamic properties of metals, inorganic compounds, molecules, and polymers [7–14]. Today, dozens of examples can be found of successful materials design driven by DFT calculations, such as new cathode materials for Li batteries [15], novel Ni–Ti shape memory alloys [16], and previously unknown ABX-type thermoelectrics and conductors [17].

One of the great pillars of scientific research, rekindled in recent years, is the analysis/mining of ‘data’. The precursor to knowledge is, more often than not, the presence of systematic data. Kepler’s laws of planetary motion, Mendeleev’s periodic table, and Hume-Rothery’s rules for miscibility of metal atoms [18] bear testimony to the fact. It is fair to state that materials scientists have always relied on sometimes scarce to often vast amounts of robust, experimental data to unearth crucial relationships and make conclusions. In the age of computation-driven materials research, data mining and machine-learning-based approaches have added another important dimension to the discovery process. This is not strictly a new development, as the fields of cheminformatics [19] and bioinformatics [20] have long relied on data storage, retrieval, and analysis for drug discovery and DNA sequencing. A new field called *materials informatics* is burgeoning, and focuses on the development of databases of experimental and computational data, as well as the application of state-of-the-art machine-learning techniques to convert data into easily accessible models.

#### *Polymer genome and the Materials Genome Initiative*

Polymers find wide applicability in the world today owing to their chemical resistance, light weight, easy processability, and ability to act as thermal and electrical insulators. They are used in daily applications like plastic packaging, pot and pan handles, and soft drink bottles, to more sophisticated applications like bullet-proof vests, automotive and aerospace components, and energy storage devices. Polymers (natural like DNA and cellulose, and man-made like Bakelite) are identified by their constitutional repeat unit, forming a 3, 2, or 1 dimensional network, and by a wide range of possible associated molecular weights.

Just like the *human genome* represents the nucleic acid sequence (pictured in Figure 1a) that makes up the DNA of a human being, we use the term *polymer genome* to (uniquely) identify the basic chemical build-up of a polymer. The polymer genome could be expressed in terms of the repeat unit (such as  $[\text{CH}_2\text{--CH}_2]_n$  for polyethylene (PE)), the chemical blocks (such as  $-\text{CH}_2-$  for PE) or the constituent atomic or molecular fragments (such as 4-fold C and 1-fold H atoms for PE). Some example polymers constituted of commonly seen chemical building units are pictured in Figure 1b. Not only does the polymer gen-

ome enable an easy representation of a polymer, it provides a path to correlations with properties, and a possibility of estimating them as a function of the genome. The presence of sufficient data on a variety of polymers is crucial here. If every polymer is represented in the form of the genome, which is essentially a string of numbers that map to the property values of interest, machine learning could be implemented to generate predictive models that surrogate actual experiments or computations by taking only the polymer genome as necessary input. Such models would facilitate rapid property prediction on new polymers spanning wide chemical subspaces, and consequently lead to accelerated design of new functional polymers [21].

This philosophy is aligned well with the goals of the Materials Genome Initiative (MGI) [22,23], launched by the United States government in 2009 ‘to discover, manufacture and deploy advanced materials twice as fast, at a fraction of the cost’. Since it was launched, the MGI has seen computational and informatics approaches enormously benefit experimental studies and drive new discoveries; our recent review of this field covers much of the major successes [24]. Converting past and freshly generated (experimental and computational) data on the properties of polymers into useful databases, and applying machine-learning techniques to deliver polymer design models is what we hope to achieve using the polymer genome approach as we elaborate on here.

#### *Polymers as capacitor dielectrics*

Recently, there has been a rising demand for high energy density capacitors due to the on-going electrification of transportation, communication, and military and civilian systems [7–9]. A capacitor, consisting of a polarizable dielectric material in between two conductive metal plates, can rapidly discharge its stored energy. The maximum amount of energy that can be stored in the capacitor is proportional to the dielectric constant of the material, as well as the (square of) electric field at which it undergoes electrical or mechanical breakdown. While inorganic compounds like  $\text{BaTiO}_3$  and  $\text{TiO}_2$  provide the benefit of massive dielectric constants, polymers are preferred capacitor dielectrics for energy storage because of their easy processability, flexibility, high resistance to external chemical attacks, and most importantly, propensity for graceful failure. The current state-of-the-art polymer dielectric is Biaxially Oriented Polypropylene (BOPP), which has a modest dielectric constant (2.2) and low operating temperature (85 °C), but a very high dielectric breakdown strength ( $E_{\text{bd}} \approx 700 \text{ MV/m}$ ) and a small area ( $1 \text{ cm}^2$ ) [8,9]. This leads to an energy density of  $5 \text{ J/cm}^3$ . However, BOPP has quite a few limitations; while it can function at high electric fields, its low dielectric constant certainly imposes a restriction on the energy density. BOPP also suffers from significant dielectric losses at high temperatures due to electronic conduction and dielectric degradation [8,9]. Thus, there have been experimental as well as computational efforts in improving over BOPP as the dielectric polymer candidate.

Much of the work in this regard has taken place with Polyvinylidene fluoride (PVDF), and related modifications. BOPP and most of the other dielectric polymer candidates are nonpolar polymers; atomic and electronic polarizations alone cannot contribute sufficiently to increasing the dielectric constant. PVDF

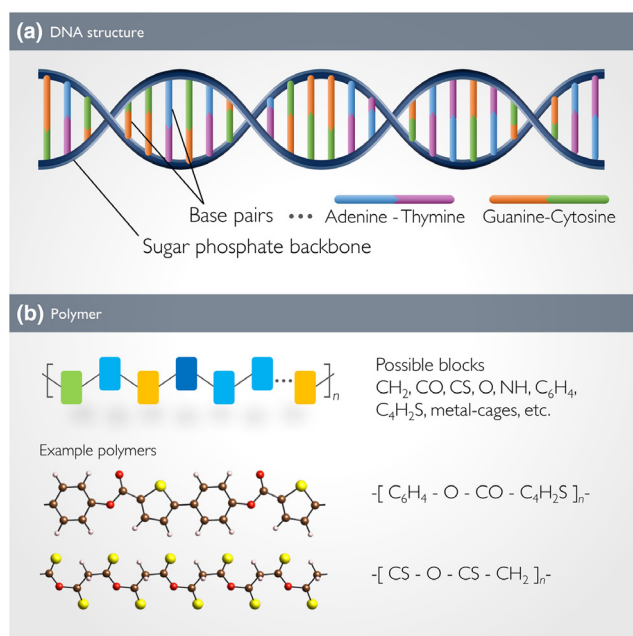


FIGURE 1

(a) The nucleic acid sequence that makes up the human DNA and formed the basis of the Human Genome Project. (b) The polymer genome concept expressed in terms of building blocks that make up a polymer chemical structure.

was thus pursued, given that its orientational polarization and high dipole density could be exploited for high energy densities. Variants of PVDF, as well as PVDF with inorganic fillers added to the matrix, have been studied and recommended for polymer dielectrics, with high dielectric constants of 10 and energy densities of  $30 \text{ J/cm}^3$  achieved [25]. However, a major problem with PVDF and its derivatives was their ferroelectric nature, which results in significant D-E loop hysteresis. This causes heavy energy losses as compared to a paraelectric material, and makes the polymer unsuitable as a dielectric for energy storage. Thus, PVDF and related polymers have been explored as possible dielectrics for energy storage applications, but came up short because of their ferroelectric behavior [9].

Polymer dielectrics for modern power electronics applications require not only high energy densities, but also high temperature capabilities and miniaturization, without significantly affecting the cost. Each of the current possible choices for polymer dielectric applications suffers from one shortcoming or the other. There is a pressing need to expand the pool of polymer dielectric candidates so that novel polymers with the optimal mix of relevant properties can be designed and gradually improved upon. However, there are significant challenges associated with this, none bigger than the vastness of the polymer chemical universe, and how little of it has been experimentally studied till date. The synthesis and property measurement of polymers in a case-by-case manner, leading (one hopes) to the eventual identification of desirable systems, is a very involved and expensive process. This makes a computation-driven treatment appropriate here: it is much faster to study many materials in a computer, and apply initial screening criteria to down-select polymers that can then be studied experimentally. Therefore, computations when

combined with experiments in a rational manner can result in the quick and efficient design of new and improved polymer dielectrics.

### Goals and outline of this paper

In this article, we discuss the application of all the ideas described earlier—computational modeling, guided experiments, materials informatics, and the polymer genome—toward the design of new and advanced polymer dielectrics for energy storage capacitor applications. In Section ‘Rational co-design of polymer dielectrics’, we discuss high-throughput DFT computations that were performed on a few hundred selected polymers, which led to the successful experimental realization of several attractive candidates. In Section ‘Machine learning strategy for accelerated design of polymer dielectrics’, we describe a machine-learning strategy to convert the computational dataset of polymers into predictive models for different properties, thus creating a framework for accelerated polymer dielectrics design. The results and successes highlighted in Sections ‘Rational co-design of polymer dielectrics’ and ‘Machine learning strategy for accelerated design of polymer dielectrics’ led to the creation of the Polymer Genome Application (PGA), a web-based recommendation engine for the ‘on-demand’ design of novel polymeric materials; we describe the details of PGA in Section ‘The Polymer Genome Application: An online knowledgebase’. Following that, we describe ongoing and future expansions of chemical, property and machine-learning spaces in Section ‘Going forward: expansion of chemical, property and machine learning spaces’. In Section ‘Challenges and opportunities beyond capacitor dielectric applications’ we consider opportunities that will allow us to go beyond the polymers and applications studied so far.

### Rational co-design of polymer dielectrics

In the past, many scientific discoveries and technological innovations have been credited to seemingly ‘trial-and-error’ practices and serendipity. Thanks to the famous story about the great American innovator Thomas Edison going through more than a thousand possibilities of materials suitable for the light-bulb filament, this approach has been referred to as the Edisonian approach. We use the term here primarily to distinguish it from rational co-design, which we describe as a materials design approach involving a first stage of computational screening followed by down-selection of promising candidates and targeted synthesis and testing. *Co-design* indicates the collaboration between computations and experiments; one is incomplete or inefficient without the other, and together they constitute a paradigm that can significantly reduce costs, provide enormous insights, and accelerate the materials design process.

Of course, the rational co-design approach only works if the problem is amenable to rapid high-throughput computations, and if it is possible to explicitly state the initial screening criteria in terms of calculable properties. It so happens that with regard to the polymer dielectrics design problem as explained earlier, it is possible to frame the chemical subspace of polymers in terms of chemical structural units (blocks, atoms, or both), leading to a list of combinatorial possibilities. DFT was chosen as the computational workhorse, and the dielectric constant and the band gap

were selected for screening purposes. The latter served as a proxy to the dielectric breakdown strength (which, along with the dielectric constant, determines the energy storage capacity), and is much easier to compute using DFT. High-throughput computations performed on a few hundred polymers led to a shortlist of potentially useful candidate materials, which was further reduced during synthesis. The workflow adopted in the rational polymer dielectric co-design process is captured in Figure 2.

#### Organic polymers as dielectrics: computational data

The application of high-throughput DFT to a selected polymer chemical subspace first involved determining the appropriate DFT formalisms for property computation. Density functional perturbation theory (DFPT) [26] is a powerful technique where the dielectric constant of a material is computed by studying the system responses to external perturbations, in this case, applied electric fields. The band gap can be computed using the hybrid Heyd-Scuseria-Ernzerhof (HSE06) electronic exchange–correlation functional [27], which corrects for the band gap underestimation associated with conventional DFT. Within these formalisms, computed dielectric constants and band gaps have been shown to match up very well with experimentally measured results for inorganic compounds as well as common polymers [14]. For this reason, these methods were selected for performing the high-throughput DFT computations.

While polymers are known to be either amorphous or semi-crystalline in nature, a crucial assumption made here was to consider a closely packed crystalline model. Although crystallographic information is available for many well-known polymers like polyethylene, PVDF, and polyacetylene, there is not sufficient diversity within the family of such common polymers to cover a large enough space for maximum payoff in terms of dielectric properties. To overcome this issue, new chemical spaces had to be devised using some of the most pervasive chemical units as polymer building blocks.

Out of the vast polymer chemical universe, a controlled subspace was selected that was ripe for high-throughput computational study. This chemical subspace consisted of polymers containing the following 7 building blocks: CH<sub>2</sub>, CO, NH, C<sub>6</sub>H<sub>4</sub>, C<sub>4</sub>H<sub>2</sub>S, CS, and O. This set of building blocks was chosen based on how ubiquitous they are in well-known polymer systems, and was deemed to be suitable for performing a controlled computational study of organic polymers [28,7,13,14,29]. For our computations, a dataset of purely 4-block polymers was considered here. A total of 406 symmetry-unique 4-block polymers can be formed using the 7 building blocks, of which only 284 were subjected to DFT computations. Chemical intuition and prior knowledge dictates that some combinations of adjoining chemical blocks make for unstable systems, leading to the elimination of all polymers consisting of O–O, CS–CS, CO–CO, and NH–NH pairs; hence the reduced number. Some examples of these polymers are pictured in Figure 1b.

The crystal structures of all 284 4-block polymers were determined using the minima hopping method [30], with the necessary total potential energies and atomic forces computed using DFT. With the stable 3-dimensional arrangements of polymer chains determined for all 284 polymers, their relevant properties

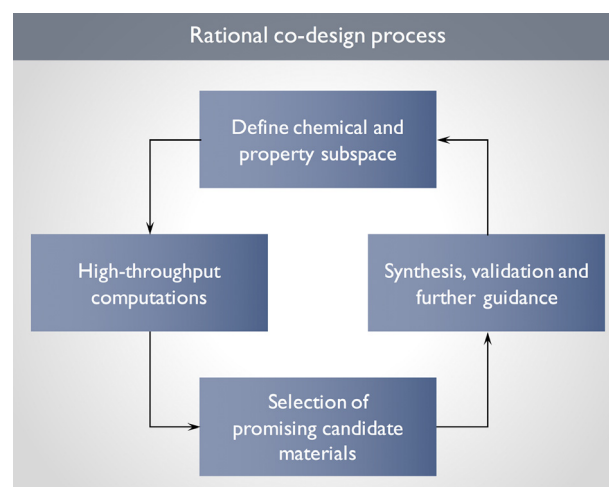


FIGURE 2

Steps involved in a rational co-design approach.

were calculated. In particular, the band gap ( $E_{\text{gap}}$ ) was computed using the HSE06 functional, and the electronic ( $\epsilon_{\text{elec}}$ ), ionic ( $\epsilon_{\text{ion}}$ ) and total ( $\epsilon_{\text{tot}}$ ) dielectric constants were computed within the DFPT formalism. In Figure 3,  $\epsilon_{\text{elec}}$ ,  $\epsilon_{\text{ion}}$ , and  $\epsilon_{\text{tot}}$  are plotted against  $E_{\text{gap}}$ . While  $\epsilon_{\text{elec}}$  is inversely correlated with  $E_{\text{gap}}$ , the ionic dielectric constant  $\epsilon_{\text{ion}}$ , which is decoupled from the electronic response, shows no correlation with  $E_{\text{gap}}$ . In fact, this observation has motivated the development of metal-containing polymers (discussed later), through which  $\epsilon_{\text{ion}}$  can be improved without compromising  $E_{\text{gap}}$ , as discussed in Section ‘Going forward: expansion of chemical, property and machine learning spaces’. The best polymers for energy storage capacitors would occupy the ‘high dielectric constant, large band gap’ area, which we defined as band gap > 3 eV and dielectric constant > 4. This region was seen to be predominately composed of at least one of the polar units (NH, CO, and O, which boost the ionic component) and at least one of the aromatic rings (C<sub>6</sub>H<sub>4</sub> and C<sub>4</sub>H<sub>2</sub>S, which boost the electronic component), which provided us with some vital leads.

#### Experimental realization and major synthetic successes

Based on the insights from computations, three polymers were considered for synthesis: a polyurea-type polymer –NH–CO–NH–C<sub>6</sub>H<sub>4</sub>–, a polyimide-type polymer –CO–NH–CO–C<sub>6</sub>H<sub>4</sub>–, and a polythiourea-type polymer –NH–CS–NH–C<sub>6</sub>H<sub>4</sub>– [7]. Polymer synthesis via a stepwise mechanism was followed by traditional characterization techniques to confirm the chemical and crystalline structures. UV–vis was applied to determine the optical band gap and time-domain dielectric spectroscopy (TDDS) was employed to obtain the dielectric constants over a range of frequencies. The experimentally measured and DFT computed  $E_{\text{gap}}$  and  $\epsilon_{\text{tot}}$  are listed in Figure 4, and we refer to these polymers as the ‘first generation of rationally co-designed polymers’. The agreement between experiments and DFT is impressive considering the synthesized polymers are different from their idealized perfectly crystalline computational renditions.  $\epsilon_{\text{tot}}$  for the three polymers lay between 4 and 6, while  $E_{\text{gap}}$  was greater than 3 eV, which is a promising combination of properties for capacitive energy storage.

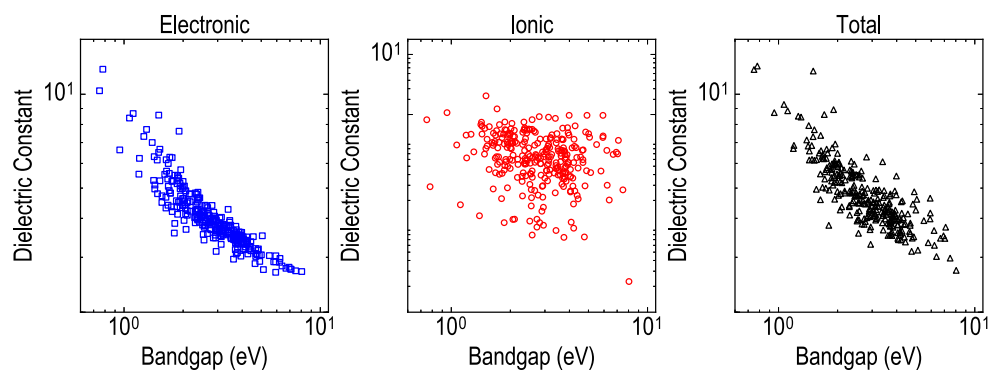


FIGURE 3

Computed dielectric constants and band gaps of the 4-block polymers.

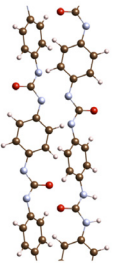
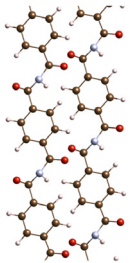
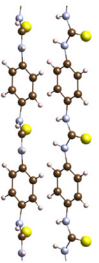
| Polymer class                   | Polyurea   | Polyimide  | Polythiourea   |
|---------------------------------|--|--|--|
| Atomic structure (DFT computed) |  |  |  |
| Repeat unit                     | NH-CO-NH-C <sub>6</sub> H <sub>4</sub>   | CO-NH-CO-C <sub>6</sub> H <sub>4</sub>   | NH-CS-NH-C <sub>6</sub> H <sub>4</sub>   |
| Dielectric constant             | 5.8 (4.9)  | 4.8 (4.9)  | 6.7 (6.0)  |
| Band gap (eV)                   | 3.9 (3.4)  | 4.0 (3.3)  | 3.1 (2.5)  |

FIGURE 4

The repeat units, DFT computed structures, and the measured and computed (shown in brackets) properties of the ‘first generation of rationally co-designed polymers’.

However, issues of solubility and thin-film processability could not be avoided with the three polymers presented above. As a result, a host of new polymers belonging to the polythiourea, polyurea, polyurethane, and polyimide classes were synthesized based on repeat units derived from computational guidance. Some of these polymers were successfully processed into free-standing films, and their properties were measured as before using UV-vis and TDDS. Further, the breakdown strength  $E_{bd}$  was determined through both direct breakdown measurement and electric displacement-electric field (D-E) loop measurement. The dielectric performances of different polymers could be compared using the measured  $\epsilon_{tot}$ ,  $E_{bd}$  and the releasing energy density; the details of the necessary techniques and results obtained are a major part of the discussion in Ref. [8].

A glimpse of the properties of the three best polymers thus obtained, and referred to here as the ‘second generation of rationally co-designed polymers’, is provided in Figure 5, along with the respective free-standing films. For comparison, the properties of BOPP are also shown in Figure 5. The polythiourea PDTc-HDA

and the polyimide BTDA-HDA showed nearly double the dielectric constant as BOPP and a comparable breakdown strength, leading to an energy density twice as high ( $\sim 10 \text{ J/cm}^3$ ). Further, the polyimide BTDA-HK511 showed a very high  $\epsilon_{tot}$  of  $\sim 8$ , leading to an energy density 3 times as high as BOPP ( $\sim 15 \text{ J/cm}^3$ ). Thus, guided by computations on selected representative polymer systems considering only the dielectric constant and the band gap, at least three very attractive dielectric polymer candidates could be identified for capacitive energy storage.

## Machine-learning strategy for accelerated design of polymer dielectrics

### Machine learning in materials science

In the Introduction, we touched upon the importance of data analysis, or learning from prior data, in formulating new theories, laws, and models. While materials scientists have historically taken advantage of data and their many opportunities, the last dozen or so years have seen this paradigm taken to the next level. With the increasing application of machine-learning techniques to materials science problems, a field called *materials informatics* was born and has since flourished. In a recent review paper [24], we cover many of the major examples of successes in this field. The philosophy of materials informatics can be stated as follows: given a critical amount of reliable materials data (this could be previously documented or freshly generated, either experimentally or using the best available computational schemes), one could learn the underlying characteristics that best explain their behavior or properties. This could lead to the creation of “surrogate models” that make forecasts for new materials based on the existing data, without the need to perform the same experiments or computations that went into generating said data.

The most general form of a materials informatics problem is when there is a dataset of materials (the input) and their properties (the output), and a need to make estimates of the properties of new materials not present in the dataset. If every material could be reduced to a unique, representative string of numbers (referred to in the community as fingerprint, feature vector, or descriptors), a mapping could be formed between this representation and the property, leading to a prediction model that takes any material as input and serves the property as output. This prediction could be qualitative (for instance, whether a


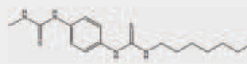
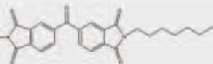
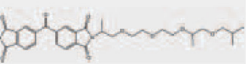




| Polymer name                        | BOPP  | PDTC-HDA<br>(Polythiourea)  | BTDA-HDA<br>(Polyimide)  | BTDA-HK511<br>(Polyimide)   |
|-------------------------------------|---|---|--|---|
| Repeat unit                         |                  |  |  |  |
| Synthesized polymer                 | <br>(Metallized) |  |  |  |
| Dielectric constant                 | 2.2   | 3.7   | 3.6  | 7.8   |
| Breakdown field (MV/m)              | 700   | 685   | 812  | 676   |
| Energy density (J/cm <sup>3</sup> ) | ~5  | ~9  | ~10  | ~16   |

FIGURE 5

The repeat units, free-standing films and measured properties of the “second generation of rationally co-designed polymers in comparison with those of BOPP.

material will have a high band gap or not, or whether a material will exist in a particular crystal structure or not), or quantitative (for example, what is the band gap or dielectric constant of this material?), and can be trained using any of the widely used regression or classification algorithms such as neural networks, ridge regression, decision trees, etc. Such a strategy has been successfully applied by materials scientists in a variety of studies, like predicting the intrinsic electrical breakdown strength of solids [31], classifying binary inorganic compounds into different crystal structure types [32,33], predicting the band gap of insulators by training on available experimental data [34], estimating the mechanical and thermal properties of polymers based on group contributions [35], and rapidly calculating forces on atoms in any given configuration for accelerating molecular dynamics [4].

It should be emphasized that while the actual algorithms for performing machine learning are well studied and manifold, it is the materials ‘fingerprinting’ that poses the greatest challenge, and is perhaps the most crucial cog in the wheel of materials informatics. A ‘fingerprint’ is a vectorial representation of a material/environment, wherein each component of this vector attempts to capture different aspects of the system. In the many examples listed above, fingerprints ranging from the chemical/atomic composition to elemental properties (like electronegativity or HOMO-LUMO gaps) to molecular fragments to radial distribution functions have been applied, with varying levels of accuracy. When it comes to polymers, it has been shown that the key features that determine the properties are the polymer repeat unit and the molecular weight distribution [35]. If we assume a very large molecular weight, the polymer chemical structure is reduced to its repeat unit, or more specifically, the number and ordering of various chemical components (these could be blocks like CH<sub>2</sub> and C<sub>6</sub>H<sub>4</sub>, or atoms like an sp<sup>3</sup> C) that make up the polymer. Indeed, this idea was fundamental to the important work of Van Krevelen and coworkers [36,35], who reduced properties like glass transition temperature, elastic modulus, and refractive index to a linear

function of different chemical groups and combinations of groups in the polymer repeat unit. This was one of the first treatments of a polymer in terms of its ‘genome’, as we explained in the Introduction when talking about the origins and importance of the Polymer Genome.

#### Machine-learning models for polymer property prediction

As mentioned earlier, the most crucial step in developing a machine-learning model is to utilize an appropriate fingerprint/descriptor. Fingerprints can capture gross-level information [31,24,37] such as density and band gap, or, they may represent sub-angstrom-level features such as atomic forces [4,38–41] and potential energy landscapes [3]. For the particular case of polymers, the most appropriate type of fingerprint would be a molecular fragment-type fingerprint which describes the connectivity/topology of the atoms constituting the polymer. Such fingerprints are commonly used in cheminformatics and QSPR modeling to analyze biological activity and physico-chemical properties of hundreds of thousands of molecules [42–44].

Van Krevelan and coworkers [36,35] fingerprinted polymers in terms of single blocks or specific combinations of blocks to devise relationships for different properties. However, a more rigorous application of the group additivity theory along with more sophisticated machine-learning techniques can enable the formulation of more accurate and more general non-linear relationships for the prediction of polymer properties. To this effect, we recently, successfully employed two types of fingerprints for the data-driven learning of polymer properties from high-throughput DFT data. These two types of fingerprints are schematically depicted in Figure 6. Our first fingerprint [13,45,46,29] represented the polymer repeat unit through a vector of molecular fragment “singles”, “doubles”, or “triplets”. As described earlier, our initial polymer database was constructed from a combination of 7 building blocks: CH<sub>2</sub>, CO, NH, C<sub>6</sub>H<sub>4</sub>, C<sub>4</sub>H<sub>2</sub>S, CS, and O. Hence, for example, the fraction of the molecular fragment ‘-CH<sub>2</sub>-CO-NH-’ would be one particular component of the triplets fingerprint vector, and the complete

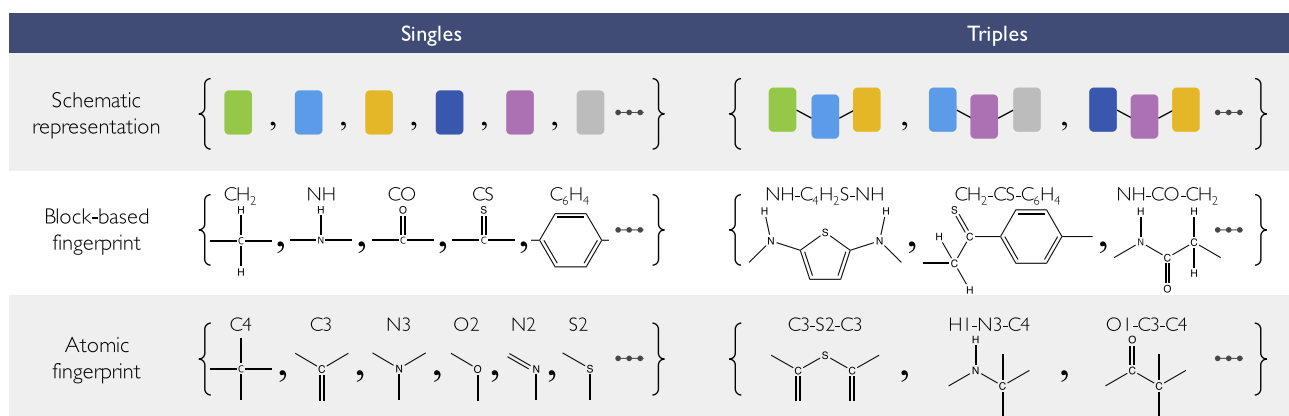


FIGURE 6

Block/atom-level fragments that constitute the polymer fingerprint.

vector would contain the fractions of every possible triplet. The advantage of the triples fingerprint is that it captures many of the long-range features/bonding that are important for certain polymer properties such as the ionic dielectric constant [46]. However, due to its inherent construction, it cannot be easily extended to new classes of polymers containing molecular fragments different from our initial 7 building blocks, even if they are composed of the same variety of atomic-level fragments discussed below.

The second type of fingerprint [47,29,48], uses atomic fragments of triples (e.g.  $C_4-O_2-C_3$ ) rather than molecular-fragments. The atomic building blocks are represented by  $X_n$ , where  $X$  is the identity of the atom and  $n$  is its coordination number. This fingerprint has the advantage that it can be easily applied to any arbitrary class of polymers and provides a greater degree of flexibility. Going forward, it is likely that hierarchical fingerprints covering different length-scales [49,50] will lead to the ideal combination of accuracy and flexibility.

Once the fingerprints have been identified, it is then necessary to choose an appropriate machine-learning model to learn the data. Our technique of choice to decode the polymer genome was Kernel Ridge Regression. Using the ‘kernel-trick’, this method allows us to learn non-linear relationships based on the (dis) similarity of the fingerprint vectors. We fingerprinted the entire computational dataset of polymers presented in Figure 3, and applied kernel ridge regression (KRR) on the data to obtain prediction models for three properties:  $\epsilon_{elec}$ ,  $\epsilon_{ion}$  and  $E_{gap}$ . The performances of these models are shown in Figure 7; an average relative prediction error of  $\sim 10\%$  was seen. We used the KRR models to instantly predict the properties of several thousand enumerated 6-block and 8-block polymers, thus achieving a substantial chemical space expansion. More details of KRR and its application to our polymer data are provided in Refs. [13,51].

### The Polymer Genome Application: An online knowledgebase ( [www.polymergenome.org](http://www.polymergenome.org) )

The Polymer Genome Application (PGA) is a web-based recommendation engine for the rapid design and discovery of polymeric materials, powered by quantum mechanical computations, experimental data, and machine learning. In response to a user query on a particular polymer, whether exist-

ing or hypothetical, the PGA returns a variety of properties, such as band gap, dielectric constant, structural information, dielectric strength, dielectric loss, stabilities, etc. The properties will be either from the database of computations based on DFT, experimental data from the literature, or, from machine-learning forecasts if neither computations nor experimental data are available.

A user-friendly interface has been developed to navigate the PGA. As shown in Figure 8, this interface consists of three components: search form, polymer table, and polymer details. The search tool accepts a variety of query formats, such as chemical formula, name of polymer, polymer repeat unit, SMILES string, and 3-D structures (POSCAR) [52]. Using advanced search options, a combination of multiple searching criteria is also possible. For example, the user can specifically search for all ‘-NH-’ containing polymers with a band gap larger than 3 eV.

Comprehensive information for the material selected from the search result table is summarized in a one-page report. This information includes the name of the polymer, polymer class, crystalline structure with 3-D visualization, electronic structure properties, data acquisition details (e.g. VASP parameters), and machine-learning predictions for several properties.

In principle, there is no limitation on the size of the dataset, types of properties, and learning algorithms. Thus PGA is truly expandable by continuous further efforts in populating the dataset with new polymer properties obtained via computation, experiments, and data scraping.

### Going forward: expansion of chemical, property, and machine-learning spaces

#### Chemical space expansion

Following the success of applying a rational co-design strategy to a chemical space of purely organic polymers, our attention was diverted to an all new subspace of polymers with metal-containing units in their backbone. While the idea of metal-based fillers in polymer matrices is not new [9], inserting metal units in the polymer backbone to change the repeat unit is a novel ‘molecular composite’. The motivation here is to boost the ionic dielectric constant  $\epsilon_{ion}$  of the polymers by incorporating suitable metal-containing bonds, which are generally highly

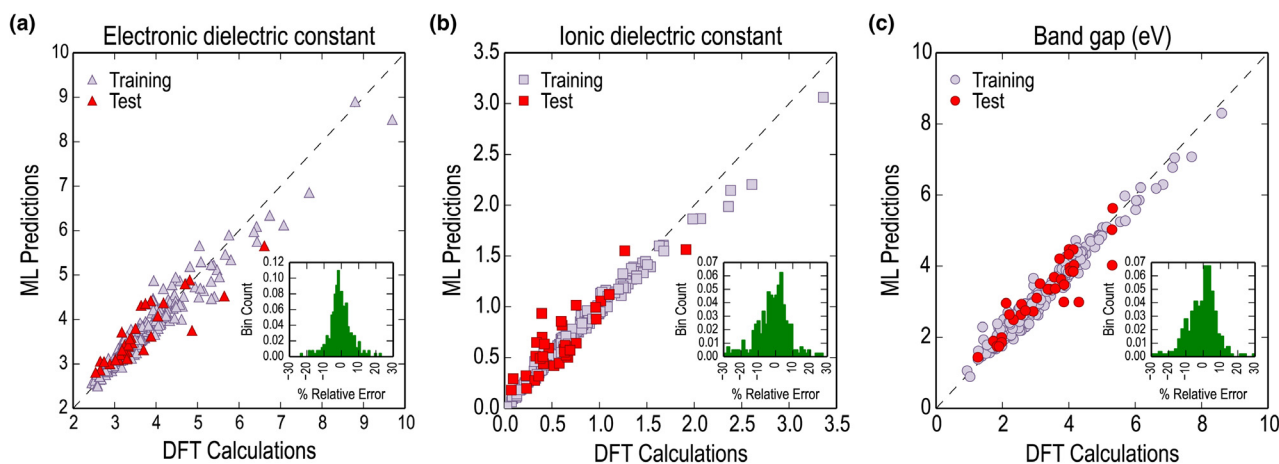


FIGURE 7

The performance of the machine-learning models for (a) Electronic dielectric constant (b) Ionic dielectric constant and (c) Band gap [13].

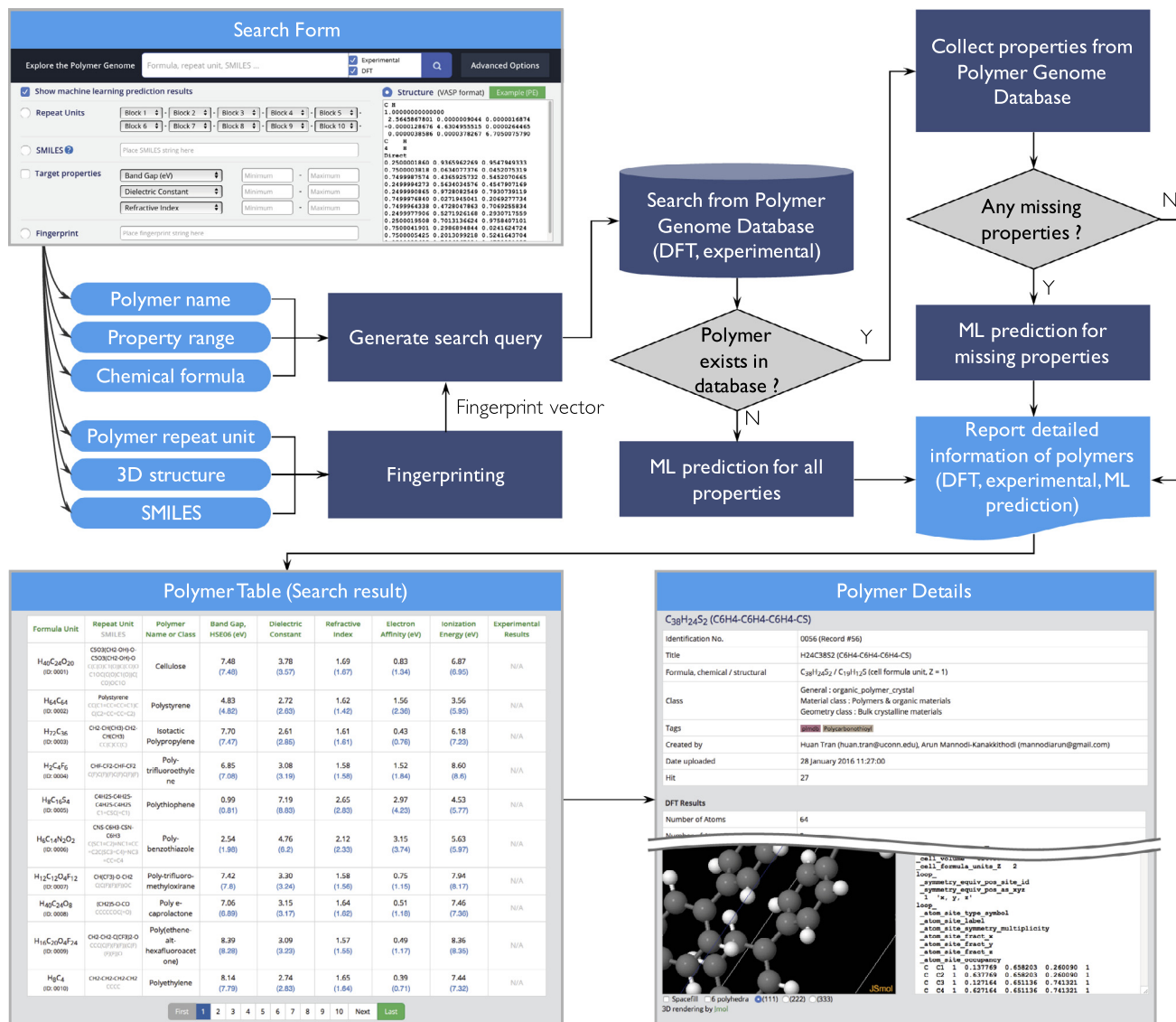


FIGURE 8

Overview of the Polymer Genome Application workflow. This online application may be found at [www.polymergenome.org](http://www.polymergenome.org).



polar. This idea was confirmed by our computations of the dielectric constant of 15 binary compounds involving the dihydrides, difluorides, and dichlorides of the group 14 elements, i.e., C, Si, Ge, Sn, and Pb. Ge, Pb, and especially Sn containing compounds show a remarkable increase in  $\epsilon_{\text{ion}}$ , and consequently, in  $\epsilon_{\text{tot}}$ , as compared to the compounds of C, while maintaining band gaps higher than 4 eV. It was proposed that inserting Sn-based units in the backbone of polyethylene or other common organic polymers would help us to obtain the right mix of high dielectric constant and large band gap.

In three following studies [10–12], Sn-containing polymers (Sn esters) were synthesized and characterized and, at the same time, computationally studied. The Sn esters were composed of the general repeat unit  $-\text{OOC}-\text{Sn}(\text{CH}_3)-\text{COO}-(\text{CH}_2)_n-$  ( $n = 0 - 11$ ). Figure 9 is devoted to the computed and measured  $E_g$  and  $\epsilon_{\text{tot}}$  of these polymers, which feature simultaneously high  $E_g$  (> 6 eV) and high  $\epsilon_{\text{tot}}$  (5 – 8). The general trend is that  $E_g$  increases and  $\epsilon_{\text{tot}}$  decreases with larger  $n$ . Three structural motifs, i.e.,  $\alpha$ ,  $\beta$  and  $\gamma$ , which were computationally predicted and experimentally confirmed [10,11], clearly demonstrate the structural diversity offered by the metal-containing environments and the opportunities provided by the new chemical subspace being explored. This was the first time a mix of such high dielectric constants and band gaps was achieved for any of our polymers, making Sn ester-based polymers very promising indeed. Experimentally, blends and copolymers were prepared using different Sn-ester polymers as well as by combining them with purely organic polymers, paving the way to better film qualities and reasonable breakdown strengths. While the energy density of such polymers is comparable with BOPP, efforts are currently underway to further optimize the Sn-ester-based polymers [53].

The computation-driven discovery of novel Sn-based organometallic polymers led to a sweeping new exploration of polymers containing different metals chosen from the periodic table. A comprehensive dataset of organometallic polymers involving 10 different metals, along with the organic polymers from Section ‘Rational co-design of polymer dielectrics’, was prepared in Ref. [14] and has been summarized in Figure 10. The organometallic polymers clearly surpass their pure organic counterparts in terms of high  $\epsilon_{\text{tot}}$  for given values of  $E_{\text{gap}}$ . The fundamental reason behind this increase, which is also the motivation of the expansion to the organometallic subspace, is that the metal-containing bonds are generally highly polar, and the swinging and stretching motions of these polar bonds are generally low in frequency, leading to an increased contribution to  $\epsilon_{\text{ion}}$ . Detailed analysis [48] revealed that the identity of the metals, the content of the metals, and the coordination environment of the metal atoms are among the crucial factors for the optimization of the polymers for energy storage applications. Experimental efforts for Zn- and Cd-based polymers have already borne fruit [54], and organometallic polymers containing various other metals are being tested.

### Property space

We are currently in the process of expanding our polymer properties database beyond the dielectric constant and band gap. For instance, properties such as the dielectric loss and the breakdown field are important for capacitor dielectrics, but are not easily

obtained using first-principles calculations. By including such data from molecular dynamics and experimental sources, the capability for materials design is significantly enhanced. One particular property of immense importance to all polymer applications is the glass transition temperature ( $T_g$ ). Our currently implemented atom-based triples fingerprint is already capable of predicting the glass transition temperature of polymers with a root mean square error (RMSE) of 25 K. By including other parameters such as the length of the side-chain and the fraction of ring atoms, we are able to account for the stiffness of the polymer chain, and a significant improvement in the model’s performance was observed. A more in depth perspective of fingerprinting and machine-learning techniques for experimental properties of polymers will be addressed soon in a future work. The motivation to include specific properties to our database and machine-learning algorithms stems from the challenges in designing polymers for state-of-the-art technological applications as detailed in Section ‘Challenges and opportunities beyond capacitor dielectric applications’.

### Machine learning

**Uncertainty quantification:** The statistical nature of machine-learning predictions implies that uncertainties are expected to be provided for every prediction. The uncertainty, in this case, would be a quantitative description of the degree to which a test case falls within/outside the domain of the training dataset. Gaussian process regression (GPR) [55] is one particular machine-learning scheme, where predictions for any given case also naturally contains the uncertainty information. This is because, in this approach, the data are assumed to be represented by a Gaussian distribution of models. The mean and variance at any given point, within these spectra of models, are nothing but the prediction and uncertainty of the prediction, respectively. We see this as a natural way forward, especially considering the similarity between GPR and KRR. An additional advantage of GPR, is that the model’s hyperparameters can be obtained by gradient-ascent on the so-called marginal likelihood function [56]. Other, more computationally expensive, methods to quantify uncertainty rely on probing the sensitivity of the model to ‘perturbations’ in the dataset [57] (bootstrapping) or examining a distribution of predictions for a set of conceptually different model architectures [3].

**Adaptive design:** The uncertainty quantification has the added benefit of allowing us to systematically expand and improve an existing model. In particular, one can use the uncertainties in the model predictions at any given stage of the learning process to further evolve and improve the machine-learning model. This is done via selecting the next training candidates, which when included in the current training set would lead to a maximal improvement in the model’s performance, evaluated with respect to its present performance.

In addition to providing an active learning framework for iterative model improvement via systematic training set selection, an adaptive design framework can also be directly used to naturally balance the trade-offs between ‘exploration’ and ‘exploitation’ [58,59]. At any given stage a number of candidates may be predicted to have given properties with uncertainties. In Figure 11, this is illustrated using an example where the goal is to

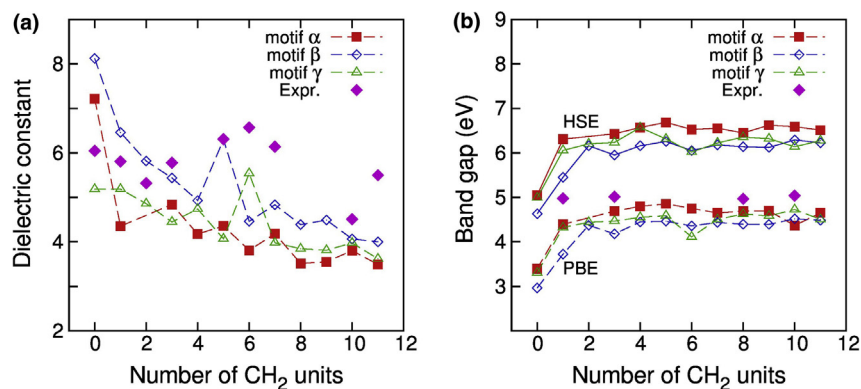


FIGURE 9

Measured and computed properties of Sn-ester-based polymers [8].

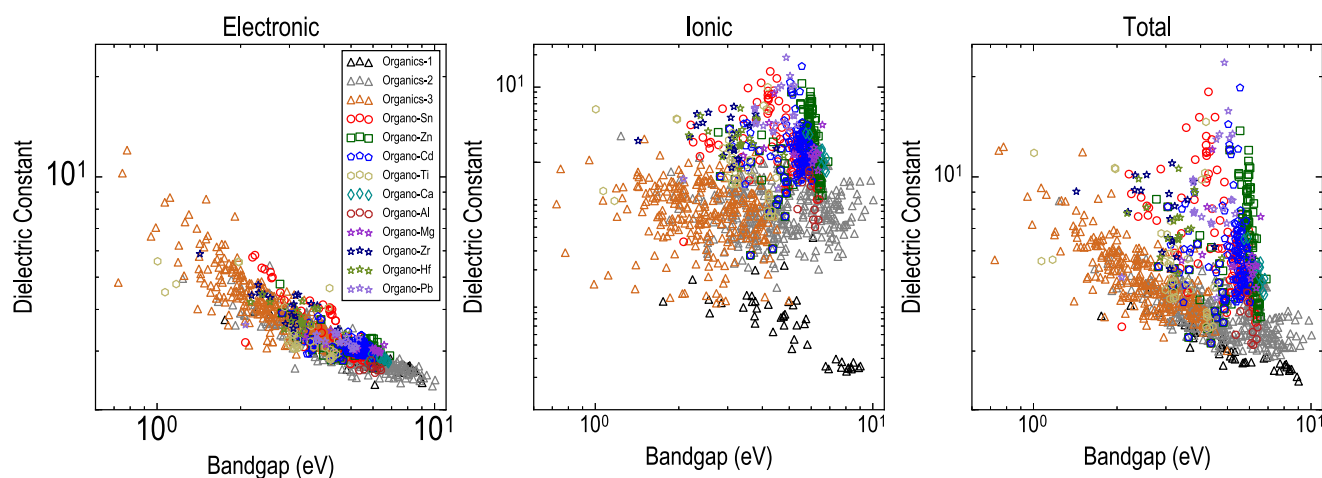


FIGURE 10

The electronic, ionic, and total dielectric constants plotted against the band gaps for the entire dataset.

maximize the targeted property and two distinct candidate choices (*i.e.*, choice 1 and choice 2) are available in the  $n + 1^{\text{th}}$  cycle to choose from for further model/property improvement. Note that both the choices have higher mean predicted values for the targeted property as compared with the current best. However, the predicted uncertainties for the two choices are qualitatively different. While the learning model is quite confident about the predictions for choice 1, the predictions for choice 2 have large uncertainties. The trade-off is between exploiting the results by choosing to perform the next computation on the material choice 1 predicted to have the improved target property with a high level of confidence (*i.e.*, low uncertainty) or further improving the model by performing the calculation on the material choice 2 where the predictions have the largest uncertainties. By choosing the former, one tries to harness the predictive power of the current model to look for materials with desired properties. On the other hand, the latter choice leads toward further model improvements and expansion of its domain of applicability, which will influence the results of the next iteration in the loop. For the particular case of dielectric polymers, the adaptive design framework allows us to address practical questions such as: At a given stage, do we expand our

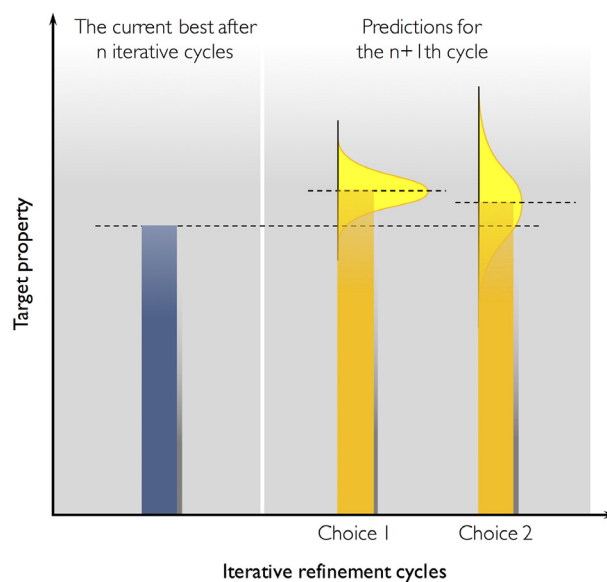


FIGURE 11

The exploration versus exploitation trade-off during the iterative model refinement via active learning.

training set in a domain with large uncertainties or do we look more closely in the vicinity of polymers, which already show simultaneous large band-gaps and dielectric constants. This decision can be made rigorously using well-established information theoretic selector frameworks such as the knowledge gradient [60,61] and efficient global optimization based on the concept of expected improvement [62]. While, exploration/exploitation trade-off-based algorithms are still an active area of research [63], the utilization of the probabilistic description in-built in a Bayesian framework such as GPR offers potential to develop a truly adaptive machine-learning strategy to decode the polymer genome.

### Challenges and opportunities beyond capacitor dielectric applications

So far, we have demonstrated the success of rationally designing polymer dielectrics by combining computations, experiments, and machine learning. From a practical point of view, the challenges of producing a novel polymeric material for a given application are twofold. The first is the identification of promising chemistry, which was the subject of this review, whereas the second, often ignored challenge, is devising potential processing routes to allow easy industrial manufacturing. The decoupling of these two facets of the materials design problem tremendously increases the time it takes to go from research to development. Questions about solvents that allow polymer synthesis, the tendency of a polymer to crystallize, and viable processing routes remain largely unanswered in the context of polymers identified computationally or studied in a university laboratory.

Further, the true description of a polymeric material is far more complex than the ideal models used to depict them. Large amorphous regions in polymers may render computational results on pure crystals unhelpful, especially if the free volume leads to contributions to the ionic part of the dielectric constant. Further, there may occur significant cross-linking between polymer chains, or the polymer may be constituted of various fragments of different molecular weights, each of which can be viewed as a physical defect that affects the electronic and dielectric properties in ways first principles computations cannot currently capture. Indeed, the incorporation of such complexities within the polymer fingerprinting aspect of the machine-learning framework could provide a path toward training and prediction of experimentally relevant properties such as the glass transition temperature, solubility parameter, and bulk modulus.

Although there are some key limitations, the materials design paradigm presented here is, nevertheless, a major step forward in the faster identification of candidates that are worthy of being pursued. The same strategy can be applied to a wide range of applications and scenarios, some of which are highlighted below along with the associated properties of interest:

- Lithium-polymer solid electrolyte batteries – Polymer electrolytes are sought because of their thermal stability and low cost. However, due to limitations in their ionic conductivity, there is a scope to apply the polymer design strategy described in this review to optimize properties such as the Li-ion diffusivity, conductivity, and glass transition temperature for this application [64].
- Photovoltaics – Semiconducting polymers are used in organic photovoltaic devices because of their low cost and compatibility with roll-to-roll printing technologies [65]. In addition to the band gap, other properties of interest here are the relative positions of the HOMO-LUMO states of the donor and acceptor, as well as the exciton binding energy.
- Flexible electronics – The interplay between mechanical and electronic properties of polymers is crucial for the field of flexible electronics [66,67]. In addition to relevant electronic properties (e.g., the band gap and dielectric constant), other properties like the elastic tensor (which may be obtained using computations) and the tensile strength (easier from well-established experimental methods) are important.
- Optoelectronics – High refractive index polymers have been extensively developed over the past few years for applications in advanced display devices, LEDs, and lenses [68]. The refractive index (which is the square root of the electronic part of the dielectric constant) is easily amenable for high-throughput investigation using computational methods.

### Conclusions

The design of novel polymer dielectrics for capacitive energy storage was performed using a combination of high-throughput computation, targeted experimentation, and machine learning. Several candidates thus identified were synthesized as free-standing films and showed attractive dielectric characteristics, and several more are currently being optimized in terms of processability and solubility. The systematically accumulated computational and experimental data [69,70], and the prediction models trained on them were crystallized in the form of an online knowledgebase of polymers. This facilitates an efficient search for existing polymers with desirable properties, as well as the visualization of hypothetical materials likely to serve a specific application.

The idea of the polymer genome was central to this work, right from the selection of common chemical building blocks to the description of polymer fingerprints in terms of fractions of various constituent atomic and molecular fragments that provided a pathway toward polymer–property correlations. The polymer genome language has the potential to break barriers between the computational scientist and the polymer chemist; while the former may make recommendations in terms of atomic units, chemical blocks or repeat units, the latter can bring them to life and also provide domain expertise in the form of appropriate chemical subspaces, properties, and reliable experimental data. Indeed, this claim would be put to its greatest test in the coming years and decades, as the materials design paradigm presented here is extensively applied to newer chemical spaces, properties, and applications. The advent of massive high-performance computational clusters, state-of-the-art experimental facilities, and modern statistical learning techniques are a great marker for potential future success stories.

### Acknowledgment

The authors acknowledge financial support for their high energy density capacitor materials research from the Office of Naval Research, through a regular grant (Award Number N00014-16-1-2580) and through a Multi-disciplinary University

Research Initiative grant (Award Number N00014-10-1-0944). GP acknowledges the Alexander von Humboldt Foundation.

## References

- [1] E.O. Hall, Proc. Phys. Soc. London, Sect. B 64 (9) (1951) 747–753, <https://doi.org/10.1088/0370-1301/64/9/303>.
- [2] N.J. Petch, J. Iron Steel Inst. 174 (1953) 25–28.
- [3] J. Behler, J. Phys. Condens. Matter 26 (18) (2014) 183001, <https://doi.org/10.1088/0953-8984/26/18/183001>.
- [4] V. Botu, R. Ramprasad, Int. J. Quantum Chem. 115 (16) (2015) 1074–1083, <https://doi.org/10.1002/qua.24836>.
- [5] P. Hohenberg, W. Kohn, Phys. Rev. 136 (1964) B864–B871.
- [6] W. Kohn, L. Sham, Phys. Rev. 140 (1965) A1133–A1138.
- [7] V. Sharma et al., Nat. Commun. 5 (2014) 4845, <https://doi.org/10.1038/ncomms5845>.
- [8] A. Mannodi-Kanakithodi et al., Adv. Mater. 28 (30) (2016) 6277–6291, <https://doi.org/10.1002/adma.201600377>.
- [9] T.D. Huan et al., Prog. Mater. Sci. 83 (2016) 236.
- [10] A.F. Baldwin et al., Adv. Matter. 27 (2015) 346.
- [11] A.F. Baldwin et al., Macromolecules 48 (2015) 2422.
- [12] A.F. Baldwin et al., Macromol. Rapid Commun. 35 (2014) 2082.
- [13] A. Mannodi-Kanakithodi et al., Sci. Rep. 6 (2016) 20952, <https://doi.org/10.1038/srep20952>.
- [14] T.D. Huan et al., Sci. Data 3 (2016) 160012, <https://doi.org/10.1038/sdata.2016.12>.
- [15] K. Kang et al., Science 311 (5763) (2006) 977–980.
- [16] X. Huang, G.J. Ackland, K.M. Rabe, Nat. Mater. 2 (5) (2003) 307–311, <https://doi.org/10.1038/nmat884>.
- [17] F. Yan et al., Design and discovery of a novel Half-Heusler transparent hole conductor made of all-metallic heavy elements, arXiv:1406.0872.
- [18] W. Hume-Rothery, R.E. Smallman, C.W. Haworth, The Structure of Metals and Alloys, The Institute of Metals, 1969.
- [19] A. Cherkasov et al., J. Med. Chem. 57 (12) (2014) 4977–5010, <https://doi.org/10.1021/jm4004285>.
- [20] Y. Saeyes, I. Inza, P. Larranaga, Bioinformatics 23 (19) (2007) 2507–2517, <https://doi.org/10.1093/bioinformatics/btm344>.
- [21] D.J. Audus, J.J. de Pablo, ACS Macro Lett. 6 (10) (2017) 1078–1082, <https://doi.org/10.1021/acsmacrolett.7b00228>.
- [22] Materials Genome Initiative, <https://www.mgi.gov/>.
- [23] A. Jain et al., APL Mater. 1 (1) (2013) 011002, <https://doi.org/10.1063/1.4812323>.
- [24] R. Ramprasad et al., NPJ Comput. Mater. 3 (2017) 54.
- [25] B. Chu et al., Science 313 (5785) (2006) 334–336.
- [26] S. Baroni et al., Rev. Mod. Phys. 73 (2) (2001) 515–562, <https://doi.org/10.1103/RevModPhys.73.515>.
- [27] J. Heyd, G.E. Scuseria, M. Ernzerhof, J. Chem. Phys. 118 (18) (2003) 8207–8215.
- [28] C.C. Wang et al., Polymer 55 (2014) 979.
- [29] J. Hill, A. Mannodi-Kanakithodi, R. Ramprasad, B. Meredig, Materials data infrastructure and materials informatics, Computational Materials System Design, Springer International Publishing, 2017, [https://doi.org/10.1007/978-3-319-68280-8\\_9](https://doi.org/10.1007/978-3-319-68280-8_9).
- [30] M. Sicher, S. Mohr, S. Goedecker, Efficient moves for global geometry optimization methods and their application to binary systems, J. Chem. Phys. 134 (2011) 044106, <https://doi.org/10.1063/1.3530590>.
- [31] C. Kim, G. Pilania, R. Ramprasad, Chem. Mater. 5 (2016) 1304–1311, <https://doi.org/10.1021/acs.chemmater.5b04109>.
- [32] G. Pilania, J.E. Gubernatis, T. Lookman, Phys. Rev. B 91 (21) (2015) 214302, <https://doi.org/10.1103/PhysRevB.91.214302>.
- [33] G. Pilania, J.E. Gubernatis, T. Lookman, Sci. Rep. 5 (2015) 17504, <https://doi.org/10.1038/srep17504>.
- [34] L. Ward et al., NPJ Comput. Mater. 2 (2016) 201628.
- [35] D.W. Van Krevelen, K. Te Nijenhuis, Properties of polymers: their correlation with chemical structure; their numerical estimation and prediction from additive group contributions, Elsevier, 2009.
- [36] H. Weyland, P. Hoftyzer, D. Van Krevelen, Polymer 11 (2) (1970) 79–87, [https://doi.org/10.1016/0032-3861\(70\)90028-5](https://doi.org/10.1016/0032-3861(70)90028-5).
- [37] G. Pilania et al., Sci. Rep. 6, 19375. doi: <https://doi.org/10.1038/srep19375>.
- [38] V. Botu, R. Ramprasad, Phys. Rev. B Condens. Matter 92 (9) (2015) 94306.
- [39] V. Botu, J. Chapman, R. Ramprasad, Comput. Mater. Sci. 129 (2017) 332–335.
- [40] V. Botu et al., J. Phys. Chem. C 121 (1) (2017) 511–522, <https://doi.org/10.1021/acs.jpcc.6b10908>.
- [41] T.D. Huan et al., NPJ Comput Mater 3 (2017) 1.
- [42] G. Landrum, RDKit: Open-source cheminformatics, Online. <<http://www.rdkit.org>>, 2012 (accessed 3 (04) (2006)).
- [43] G.W. Adamson, J.A. Bush, Nature 248 (5447) (1974) 406–407, <https://doi.org/10.1038/248406a0>.
- [44] A. Tropsha, P. Gramatica, V. Gombar, QSAR Comb. Sci. 22 (1) (2003) 69–77, <https://doi.org/10.1002/qsar.200390007>.
- [45] G. Pilania, et al., Sci. Rep. 3 (2810). doi: <https://doi.org/10.1038/srep02810>.
- [46] A. Mannodi-Kanakithodi, G. Pilania, R. Ramprasad, Comput. Mater. Sci. 125 (2016) 123–135, <https://doi.org/10.1016/j.commatsci.2016.08.039>.
- [47] T.D. Huan et al., Phys. Rev. B Condens. Matter (2015) 14106, <https://doi.org/10.1103/PhysRevB.92.014106>. 014106.
- [48] R.R.A. Mannodi-Kanakithodi, T.D. Huan, R. Ramprasad, Mining materials design rules from data: The example of polymer dielectrics, Chem. Mater. 29 (21) (2017) 9001–9010, <https://doi.org/10.1021/acs.chemmater.7b02027>.
- [49] K. Wu et al., J. Polym. Sci., Part B: Polym. Phys. 54 (20) (2016) 2082–2091, <https://doi.org/10.1002/polb.24117>.
- [50] A. Menon et al., Mol. Syst. Des. Eng. 2 (3) (2017) 263–273, <https://doi.org/10.1039/C7ME00027H>.
- [51] A.L. Parrill, K.B. Lipkowitz, Reviews in computational chemistry. vol. 29, Chapter 4, p. 452.
- [52] J. Hafner, J. Comput. Chem. 29 (13) (2008) 2044–2078, <https://doi.org/10.1002/jcc.21057>.
- [53] G.M. Treich et al., ACS Appl. Mater. Interfaces 33 (2016) 21270–21277, <https://doi.org/10.1021/acsami.6b04091>.
- [54] S. Nasreen, et al., A Materials Genome Approach to Dielectric Design through Incorporating Zinc and Cadmium in Main Chain Organic Polymers, Submitted.
- [55] T. Hastie, R. Tibshirani, J. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Springer Science & Business Media, 2013.
- [56] C.E. Rasmussen, C.K.I. Williams, Gaussian Processes for Machine Learning, Adaptive Computation and Machine Learning, MIT Press, Cambridge, MA, USA, 2006.
- [57] J. Felsenstein, Bootstrap confidence levels for phylogenetic trees, in: C.N. Morris, R. Tibshirani (Eds.), The Science of Bradley Efron, Springer Series in Statistics, Springer, New York, New York, NY, 2008, pp. 336–343.
- [58] T. Lookman et al., Curr. Opin. Solid State Mater. Sci. 21 (3) (2017) 121–128.
- [59] W.B. Powell, I.O. Ryzhov, Optimal Learning, John Wiley & Sons, 2012.
- [60] W.B. Powell et al., The Knowledge Gradient for Optimal Learning, in: Wiley Encyclopedia of Operations Research and Management Science, John Wiley & Sons Inc, 2010.
- [61] I.O. Ryzhov, W.B. Powell, P.I. Frazier, Oper. Res. 60 (1) (2012) 180–195.
- [62] D.R. Jones, M. Schonlau, W.J. Welch, J. Global Optim. 13 (4) (1998) 455–492, <https://doi.org/10.1023/A:1008306431147>.
- [63] G. Reddy, A. Celani, M. Vergassola, J. Stat. Phys. 163 (6) (2016) 1454–1476, <https://doi.org/10.1007/s10955-016-1521-0>.
- [64] D.M. Pesko et al., Macromolecules 49 (14) (2016) 5244–5255, <https://doi.org/10.1021/acs.macromol.6b00851>.
- [65] S. Holliday, Y. Li, C.K. Luscombe, Recent advances in high performance donor-acceptor polymers for organic photovoltaics (jul 2017). doi: <https://doi.org/10.1016/j.progpolymsci.2017.03.003>.
- [66] A.D. Printz, D.J. Lipomi, Appl. Phys. Rev. 3 (2) (2016) 021302, <https://doi.org/10.1063/1.4947428>.
- [67] M. Zarek et al., Adv. Mater. 28 (22) (2016) 4449–4454, <https://doi.org/10.1002/adma.201503132>.
- [68] T. Higashihara, M. Ueda, Macromolecules 48 (7) (2015) 1915–1929, <https://doi.org/10.1021/ma502569r>.
- [69] J. Brandrup, E.H. Immergut, E.A. Grulke, Polymer handbook, Wiley-Interscience, 1999.
- [70] PolyInfo, <http://polymer.nims.go.jp/indexen.html>.