

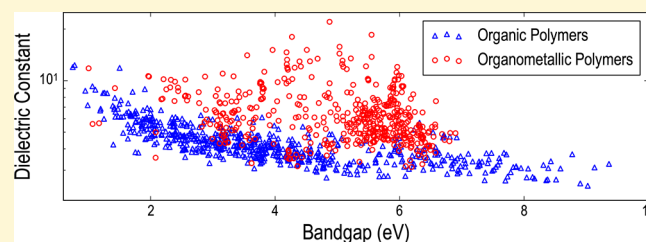
Mining Materials Design Rules from Data: The Example of Polymer Dielectrics

Arun Mannodi-Kanakkithodi,¹ Tran Doan Huan, and Rampi Ramprasad*

Department of Materials Science and Engineering, Institute of Materials Science, University of Connecticut, 97 North Eagleville Road, Storrs, Connecticut 06269, United States

S Supporting Information

ABSTRACT: Mining of currently available and evolving materials databases to discover structure–chemistry–property relationships is critical to developing an accelerated materials design framework. The design of new and advanced polymeric dielectrics for capacitive energy storage has been hampered by the lack of sufficient data encompassing wide enough chemical spaces. Here, data mining and analysis techniques are applied on a recently presented computational data set of around 1100 organic polymers, organometallic polymers, and related molecular crystals, in order to obtain qualitative understanding of the origins of dielectric and electronic properties. By probing the relationships between crucial chemical and structural features of materials and their dielectric constant and band gap, design rules are devised for optimizing either property. Learning from this data set provides guidance to experiments and to future computations, as well as a way of expanding the pool of promising polymer candidates for dielectric applications.



INTRODUCTION

For a long time, empirical data has helped build chemical intuition and scientific insights and supported the formulation of chemical and physical laws. Some classic examples cited in this regard are Hume-Rothery’s set of rules for miscibility in a solid solution¹ and the Hall-Petch relationship between material strength and grain size.^{2,3} Analysis of data procured from meticulous experimentation was key to developing these rules, showing that data-driven approaches have sometimes been a great ally of the materials scientist. Experimental data, while invaluable, could suffer from being time-intensive, nonuniform, and possibly irreproducible; on the other hand, computational methods provide the means to generate data much faster at a uniform level of theory. Indeed, modeling of materials on a computer before a single experiment is even conceived is steadily becoming the preferred route of materials design.^{4–28} Today, one can find several impressive examples of materials design via high-throughput computation,^{4–9} converting computational data into useful design rules and prediction models,^{10–24,29} computation-guided experimental design,^{25–28} and development of enormously useful materials databases.^{30–33} Data, whether experimental or computational, can be mined to unearth the role of chemistry, structure, and other crucial factors in determining the properties of materials. **Figure 1** tries to capture the mechanism of converting materials data into learning in the form of laws, rules, and models.

Recently, we implemented a computation-guided, data-driven strategy for the rational design of new and advanced polymer dielectrics for capacitive energy storage applications.^{25,34–46} Density functional theory^{47–53} was used as the computational workhorse to estimate the dielectric and

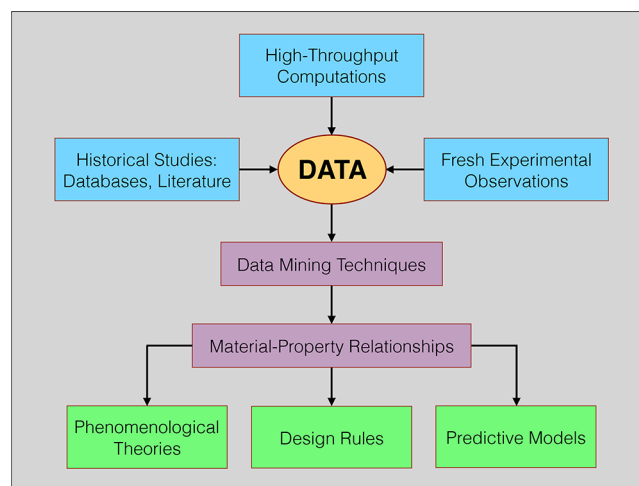


Figure 1. Data-driven materials design philosophy.

electronic properties of several organic and organometallic polymers, leading to the synthesis and characterization of many candidates^{35–40} that could potentially replace the current polymer standard for capacitor dielectrics, biaxially oriented polypropylene (BOPP). This work also led to the generation of a comprehensive computational data set of ~1100 polymers and related materials, recently presented in its entirety³⁴ and collected in the form of a “live” online database and polymer

Received: May 17, 2017

Revised: October 3, 2017

Published: October 3, 2017

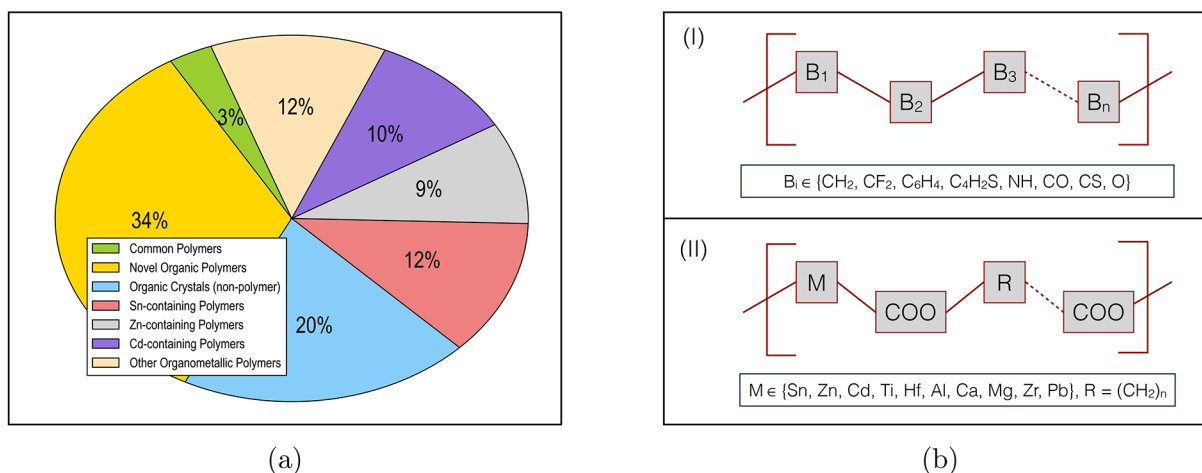


Figure 2. (a) Distribution of different classes of polymers and related materials present in the data set. (b) Chemical space of (I) organic and (II) organometallic polymers in the data set.

recommendation engine.⁵⁴ Specifically, this data set comprised of the computed electronic band gaps and dielectric constants (two properties that provide initial screening criteria for attractive capacitor dielectrics) of many commonly known polymers, novel organic polymers, newly proposed metal containing polymers, and various molecular crystals.

In this work, we attempt to mine this data to obtain a critical understanding of how factors such as the chemical composition and coordination environment affect the dielectric constant and band gap. A systematic way of doing this involves representing all the materials in terms of *fingerprints*,^{5,14,24,55–69} which are then mapped to the properties of interest. Fingerprinting is typically performed using easily attainable physical and chemical characteristics (often referred to as descriptors or feature vectors) such as composition, elemental properties, and coordination environment, in a way that is unique, general, and reproducible.^{55,70,71} Mapping the fingerprint to the property helps reveal the correlation (or lack thereof) between any fingerprint component and the property in question. We fingerprint the polymeric (and related) materials using a scheme that encodes compositional and connectivity information by quantifying the fraction of different types of atoms and the different types of chemical bonds present.⁷²

Our analysis has revealed several key guidelines (many of them nonobvious) for designing dielectric polymers. For instance, we find that while chemical bonds between 4-fold C atoms and H atoms or 2-fold O atoms enhance the band gap, bonds between 3-fold C atoms and S atoms increase the electronic component of the dielectric constant while adversely affecting the band gap. Further, we observe significant differences in the properties of organics and organometallics, with the latter showing notably higher dielectric constants for a given large band gap. The identity of the metal atom, its concentration by volume, and the coordination environment it adopts are each seen to influence the dielectric constant. While the commonly seen 4-fold metal coordination has less of an effect, 6-fold coordinated metals such as Sn, Zn, and Cd when bonded to electronegative atoms like O and F cause the greatest enhancement in ionic contribution to the dielectric constant (and consequently, the total dielectric constant).

In the following sections, we briefly explain the computational procedure for data generation and discuss the details of fingerprint development. Next, we present all the data in the

form of plots between properties and make observations about clearly visible trends. Following that, we draw correlations between the fingerprint components and the properties, which lead to the aforementioned insights. We end by providing some guidelines for property optimization in the present classes of polymers and related materials, and some concluding remarks.

METHODOLOGY

Computational Data Generation. All the computational data was generated over a series of studies related to new and advanced polymer dielectrics.^{25,34,36,37,41,55} Figure 2a shows the distribution of various material classes, like organic crystals, common polymers (for example, polyethylene, PVDF, and polystyrene), novel organic polymers, and different types of organometallic polymers, across the data set. It can be seen that nearly 80% of the data is from polymeric materials; several molecular crystals were considered for computations in order to add variety to the data set in terms of chemistry and coordination environments. For the molecular crystals and common polymers, crystal structure information was obtained from various sources in the literature, as explained in the past work.³⁴ However, novel organic and organometallic polymers form the bulk of the data set and the lack of any structural information related to these materials necessitated the application of a structure prediction algorithm.⁷³ A glimpse of the two major polymer classes in the data set (making up the 34% novel organic polymers and the 43% organometallic polymers shown in Figure 2a, respectively) is provided in Figure 2b in terms of their constituent chemical building units. While the organic polymers were built from linear combinations of commonly seen chemical blocks like CH_2 and C_6H_4 ,^{41,55} the organometallic polymers contained metal–ester units (and in the case of organo-Sn polymers, dimethyl tin ester $(Sn(CH_3)_2(COO)_2)$ ^{36,37} and tin halide $(SnF_2/SnCl_2)$ ⁷⁴ units) flanked by linker CH_2 units, with the metal atom chosen from a set of a few selected metals.^{34,36,37}

All the data was prepared using density functional theory (DFT) as implemented in the Vienna Ab-Initio Software Package (VASP).⁷⁵ We would like to refer the readers to ref 32 for more specific computational details than mentioned here, such as the exchange-correlation functionals, the energy cut-offs, and other settings used. It should be noted that the criteria were kept uniform across the data set. Van der Waals corrections were explicitly taken into account given the importance of such interactions in polymeric crystals.^{76,77} Ground state crystal structure information was obtained either from the literature or by using the Minima Hopping algorithm.^{78,79} While the HSE functional⁵² was used to compute the band gaps given their known accuracy as compared to PBE band gaps,⁵³ the dielectric constants were computed using the density functional perturbation theory (DFPT)^{50,51} formalism at the PBE level of theory.

Fingerprinting Technique. Our fingerprinting scheme follows from past work on purely organic polymers and molecular crystals.^{55,56} It has been hypothesized that the chemical structure (which comprises of the repeat unit, branches, cross-links, etc.) of a polymer is one of its most fundamental characteristics and determines all its properties.⁸⁰ A chemo-structural fingerprint was required to account for contributions from various chemical constituents (building blocks, atoms, etc.) and from combinations of these constituents that exist in the material. Thus, we used a fingerprint that encodes compositional and connectivity information by quantifying the fraction of different types of atoms in the system and the different types of chemical bonds they form.

Atom types were defined by their chemical identities (such as C, O, Sn, etc.) along with their coordination number (like 3-fold, 4-fold, 6-fold, etc.). For instance, C4 referred to a C atom forming 4 different bonds with neighboring atoms, and Sn6 to an Sn atom forming 6 bonds, each based on predefined bond length ranges. It should be noted that while characteristic bond lengths are well documented in the case of purely organic compounds (for instance, a C4–H1 bond length will be ~ 1 Å and a C4–C4 bond length will be ~ 1.5 Å⁵⁶), the distribution of possible bond lengths is more diverse in the organometallic systems. As an example, the Sn–O bond length (known experimentally to be ~ 2 Å in tin oxide⁸¹) is seen to range from 2 to 3 Å in the organo-Sn systems in our data set. Defining appropriate bond length cut-offs is thus of utmost importance for obtaining meaningful fingerprints; the cut-offs employed in this work for all possible bonding combinations are listed in Table 1.

Table 1. Bond Length Cut-Offs Defined for Fingerprinting Purposes

bond	maximum bond length (Å)	bond	maximum bond length (Å)
C–C	1.7	Zn–C	2.2
C–H	1.3	Cd–O	2.5
C–O	1.6	Cd–C	2.2
C–F	1.6	Al–O	2.5
O–H	1.4	Al–C	2.2
Sn–C	2.3	Mg–O	2.5
Sn–O	2.6	Mg–C	2.2
Sn–F	3.0	Ca–O	2.5
Sn–Cl	3.0	Ca–C	2.2
C–N	2.0	Hf–O	2.5
N–H	1.4	Hf–C	2.2
N–O	2.0	Pb–O	3.0
C–S	2.0	Pb–C	2.2
Ti–O	2.5	Zr–O	3.0
Ti–C	2.2	Zr–C	2.2
Zn–O	2.5	Al–Cl	2.5

In refs 53 and 54, we described a hierarchy of fingerprints, going from singles to doubles to triples: these refer to the complexity and dimensionality of the fingerprint. Here, three kinds of fingerprints were considered again with increasing amounts of information and increasing complexity:

1. When only the count of each atom type (C4, O2, Sn4, etc.) is considered, the fingerprint is called “singles”. The fingerprint dimensionality is equal to the distinct types of atoms, m , present across the data set ($m = 54$ for the present data set).
2. When the count of each pair of atom types (C4–C3, Sn6–O3, Zn4–O2, etc.) is considered, it is called “doubles”. Fingerprint dimensionality would typically be m^2 , but it is ~ 150 here once the components that are zero throughout (that is, the pairs that do not exist in any of the materials) are eliminated. Note here that the chemical bonds are being taken into account along with the atoms.

3. When the count of each triplet of atom types (C4–C3–H1, Sn6–O3–C3, Zn4–O2–C3, etc.) is considered, it is called “triples”. Fingerprint dimensionality would typically be m^3 , but it is ~ 500 here once the components that are zero throughout (that is, the triplets that do not exist in any of the materials) are eliminated. Note that here, along with the atoms and chemical bonds, chemical conjugation⁸² is also being taken into account.

In this fashion, we could keep going higher up in n -tuple combinations of atoms types, with increasing amount of information added but with the caveat of increased dimensionality and complexity. Figure 3 shows a few examples of the different types of singles, doubles, and triples components that exist in the materials that constitute the computational data set. It should be noted that, in each type of fingerprint, the count is normalized with respect to the total number of atoms in the system, which means every fingerprint component is essentially a fraction. Periodicity is accounted for, so that a system if doubled or tripled in size would have the same fingerprint. While this is an elegant way of fingerprinting polymers, it should be noted that structural and morphological considerations are ignored here. Indeed, what makes BOPP, the state-of-the-art polymeric capacitor dielectric, so attractive is the crystalline changes brought about by biaxial orientation to improve the stiffness, strength, and optical properties; our fingerprint would not distinguish between regular polypropylene and BOPP. However, insights gained in terms of molecular structure are very valuable, as we demonstrate in the following sections.

RESULTS AND DISCUSSION

Computational Data: Visualization and Initial Observations. Figure 4 shows all the computational data in the form of plots between band gap and dielectric constant (divided into two distinct components, electronic and ionic).^{25,41} While the data set can be broadly divided into organics and organometallics, further subdivisions were created as shown in Figure 4 in order to see correlations between the presence of specific atoms and the corresponding properties. The organics were divided into systems containing only atoms C and H (Organics-1), systems containing C, H, O, N, and F (Organics-2), and systems containing C, H, O, N, and S (Organics-3), and the organometallics were divided into 10 subsets based on the identity of the constituent metal atom, like Organo-Sn, Organo-Zn, etc.

A visual examination of the plots in Figure 4 reveals that while the electronic dielectric constant (ϵ_{elec} , a function of atomic polarizabilities) correlates inversely with the band gap (E_{gap}), the ionic dielectric constant (ϵ_{ionic} , coming from IR-active zone-center vibrational modes) does not and can thus lead to a total dielectric constant ($\epsilon_{\text{total}} = \epsilon_{\text{elec}} + \epsilon_{\text{ionic}}$) that fails to correlate with E_{gap} , especially for points possessing a high ϵ_{ionic} . The organics, which span the entire expanse of the ϵ_{elec} spectrum but generally show very low ϵ_{ionic} have ϵ_{total} values that inversely correlate with E_{gap} . All the organometallics, on the other hand, show a similar trend in ϵ_{elec} but clearly surpass the organics in ϵ_{ionic} and thus ϵ_{total} .

In a past computational study of the dielectric properties of isolated single chain renditions of metal containing polymers, it was shown that while ϵ_{elec} is directly proportional to the atomic polarizabilities of the constituent atoms, and ϵ_{ionic} correlates directly with the dipole moments in the system and inversely with rotational barriers around bonds, i.e., the ease with which they are allowed to swing and stretch.⁷⁴ Table 2 lists all the different constituent atoms in the present data set along with their measured and documented polarizabilities⁸³ and electro-negativities.⁸⁴ From Figure 4, it can be seen that while Organics-1 (containing only C and H atoms) show the lowest

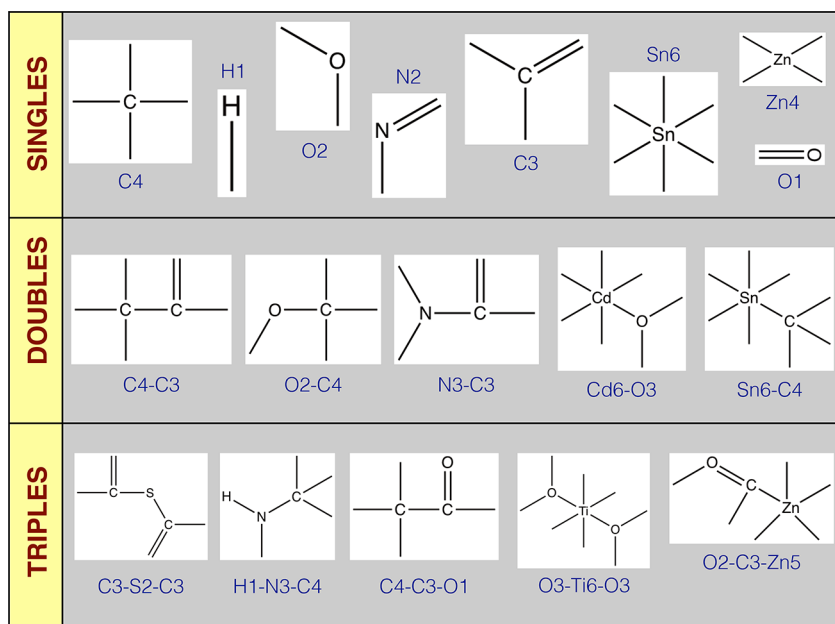


Figure 3. Fingerprinting technique, showing examples of various types of singles, doubles, and triples components found in our polymer data set.

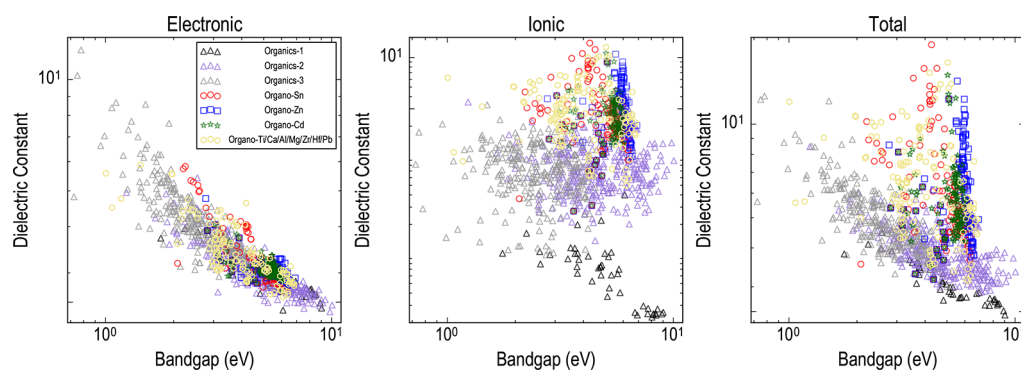


Figure 4. Electronic, ionic, and total dielectric constants plotted against the band gaps for the entire data set, divided into Organics-1 (only C and H atoms), Organics-2 (C, H, O, N, and F atoms), Organics-3 (C, H, O, N, and S atoms), and various Organo-M subdata sets, depending on the identity of the metal M present in the polymer. Organo-Sn, Zn, and Cd are plotted with different colors because there are more than 100 systems for each, while Organo-Ti, Ca, Al, Mg, Zr, Hf, and Pb are all plotted with the same color.

ϵ_{total} with nearly negligible ϵ_{ionic} owing to the closeness of electronegativities of C and H, Organics-2 (C, H, O, N, and F atoms) show higher ϵ_{ionic} values due to the presence of dipoles formed by highly electronegative atoms like O, N, and F bonding with C and H. Organics-3 (S atoms included) show the highest ϵ_{elec} values among the organics because of the higher relative atomic polarizability of S as listed in Table 2, which also leads to their lower E_{gap} , especially when the concentration of S atoms is higher.

Further, Table 2 shows that all the metal atoms are much more polarizable compared to the organic atoms, potentially leading to a high ϵ_{elec} which is then brought down by the presence of C and H atoms throughout the organometallics. The electropositive nature of the metal atoms and the high electronegativities of O and F lead to the presence of large dipoles in the organometallics. Aside from being polar, these bonds also display stretching and wagging vibrational modes that are soft in nature, leading to higher IR intensities at low frequencies^{85–87} and, thus, the highest ϵ_{ionic} values. The organometallics contain no clear demarcations between the different subsets, with high ϵ_{ionic} as well as generally large E_{gap}

shown by all. However, the actual concentration of the metal atom and the coordination environment it adopts play a crucial role here, as discussed later.

The above observations provide an example of how patterns can be extracted from data simply by visual analysis; however, the problem today is that the rate of data generation far surpasses our intrinsic ability to process the data. Consequently, advanced machine learning and data mining techniques are needed. This involves converting materials to numerically representative fingerprints and developing models by mapping them to the properties. As explained in the previous section, our basis for fingerprinting is the abundance of different types of atoms along with their respective connectivity information. While factors such as atomic polarizability, dipole moment and softness of bonds are not explicitly defined in the fingerprint, they would nevertheless be taken into account while correlating fingerprint vectors with the properties. For instance, a fingerprint component that is a metal-electronegative atom pair (like Sn–Cl or Zn–O), if strongly correlated with ϵ_{ionic} would prove the same thing observed above: a larger concentration of polar bonds would lead to a higher ionic

Table 2. All the Constituent Atoms Across the Entire Dataset, the Respective Subsets That Contain Them, Their Polarizabilities, and Electronegativities^a

atom	presence in data set	polarizability (C m ² /V)	electronegativity (Pauling)
C	all organics, all organometallics	11.0	2.55
H	all organics, all organometallics	4.5	2.1
O	Organics-2, Organics-3, all organometallics	6.04	3.44
N	Organics-2, Organics-3, all organometallics	7.43	3.04
F	Organics-2, Organo-Sn	3.76	3.98
S	Organics-3	19.6	2.58
Cl	Organo-Sn	14.7	3.16
Sn	Organo-Sn	52	1.96
Zn	Organo-Zn	39.2	1.65
Cd	Organo-Cd	46.3	1.69
Pb	Organo-Pb	46	2.33
Hf	Organo-Hf	109	1.3
Zr	Organo-Zr	121	1.33
Mg	Organo-Mg	71.7	1.31
Ca	Organo-Ca	160	1.0
Al	Organo-Al	56.3	1.61
Ti	Organo-Ti	99	1.54

^aValues for the latter two are taken from refs 80 and 81, respectively.

contribution to the dielectric constant. In the following section, the origins of polymer properties will be explained using the fingerprint components.

Fingerprint–Property Relationships. Here, we draw linear correlations between various fingerprint components and three properties respectively: ϵ_{elec} , ϵ_{ionic} , and E_{gap} . If for a data set of n members, any fingerprint component column is denoted as $\mathbf{x} = [x_1, x_2, \dots, x_n]$, and any property is denoted as $\mathbf{y} = [y_1, y_2, \dots, y_n]$; the Pearson coefficient of linear correlation⁸⁸ between \mathbf{x} and \mathbf{y} is defined as

$$r(\mathbf{x}, \mathbf{y}) = \left(\frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \right) \quad (1)$$

This coefficient can go from 1 (maximum positive correlation, when \mathbf{x} is identical to \mathbf{y}) to -1 (maximum negative correlation, when \mathbf{x} is exactly the negative of \mathbf{y}). Correlation coefficients were calculated for each component of the singles fingerprint and the doubles fingerprint with the properties over the entire data set; the triples fingerprint was avoided at this stage for ease of discussion.

Figure 5 shows the correlation coefficients between all the components of the singles and the three properties (in the form of bar charts), as well as between some selected components of the doubles and the properties (in the form of heat maps). The singles components correlating the most positively and the most negatively with the properties have been highlighted; the most relevant doubles components are shown in the middle covering up the singles that show little or no correlation. Any component of the heat map refers to a specific pair of atom types and the color shows the amount of positive or negative (or no) correlation.

The most positive correlations to ϵ_{elec} come from 3-fold C and 1- or 2-fold S atoms, while the negative correlations are provided by 4-fold C and H, and 2- and 3-fold O atoms. The correlations to E_{gap} follow exactly the opposite trend. These

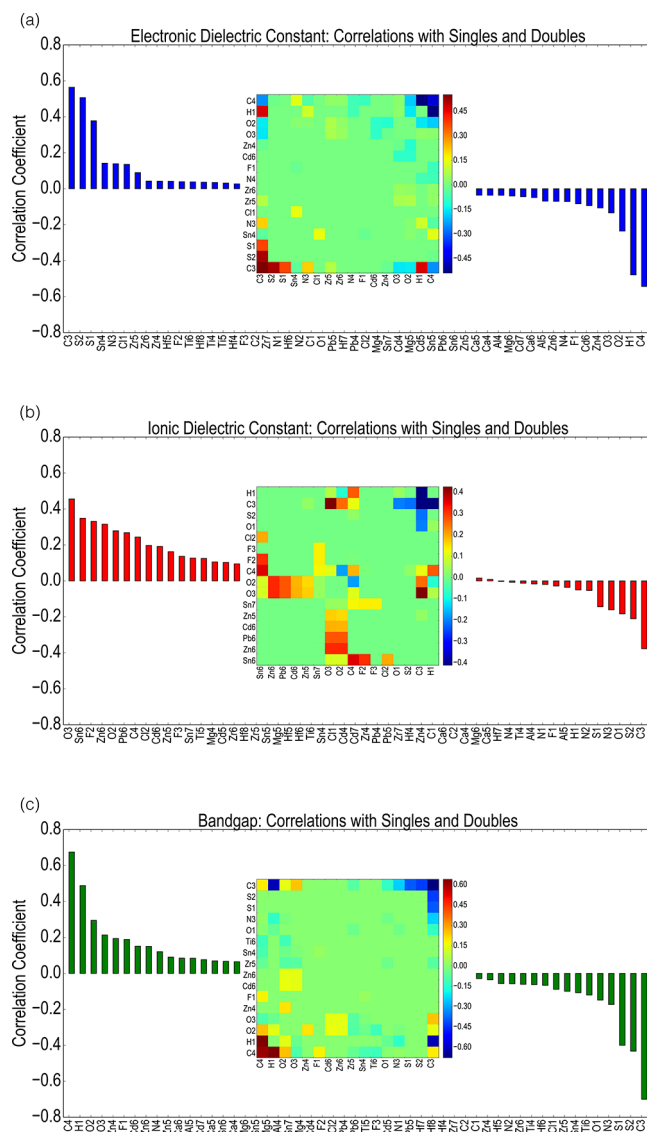


Figure 5. Correlations between the three properties and different components of the singles (bar plots) and doubles (heat maps) fingerprints, respectively.

observations can be explained with the help of the atomic polarizabilities listed in Table 2 as well as the frequencies of occurrence of different types of atoms. While the metal atoms have by far the largest polarizabilities, they do not contribute as much to ϵ_{elec} because their relative concentrations compared to C, H, and O atoms are very small. Meanwhile, S atoms are present in comparatively higher concentrations in Organics-3, and their effect in increasing ϵ_{elec} and decreasing E_{gap} is considerable.

The bonds that contribute to large ϵ_{elec} are C3–C3, C3–S2, C3–S1, and C3–H1. The presence of S atoms in Organics-3 is in the form of thiophene ($-\text{C}_4\text{H}_2\text{S}-$) groups and thiol ($-\text{CS}-$) groups, both of which contain 3-fold C atoms forming bonds with S atoms, thus explaining the results in Figure 5a. It can be seen from the heat map in Figure 5c that the same chemical bonds decrease E_{gap} . C4–C4, C4–H1, and C4–O2 bonds, on the other hand, increase E_{gap} and decrease ϵ_{elec} . This is owing to all the data subsets other than Organics-3, in which C and H as well as O atoms exist in abundance and S atoms do not.

Figure 5b shows that the largest positive correlations to ϵ_{ionic} are dominated by high coordination number (CN) metal atoms like Sn6, Zn6, Pb6, and Cd6 and by atoms with high electronegativity like O, F, and Cl, as evident from Table 2. The heat map shows that it is indeed chemical bonds between high CN metals and O2, O3, F2, or Cl2 atoms that, owing to the dipole moments they introduce in the system, as touched upon earlier, contribute the highest to ϵ_{ionic} . Bonds between C3 and O atoms as well as bonds between C4 and Sn6 atoms also show positive correlations, mainly because of the high abundance of these bonds in organometallic systems.

Negative correlations with ϵ_{ionic} are shown by C3, O1, and the S atoms, owing to relatively lower polarities in systems where 3-fold C atoms are singly bonded to H, doubly bonded to O, and singly or doubly bonded to S. It is interesting to note that while O2 and O3 increase ϵ_{ionic} greatly, O1 has the opposite effect; this is because of the abundance of 2- and 3-fold O atoms bonded either to C or to metal atoms in the organometallic systems, which show much higher ϵ_{ionic} values than the Organics-2 where double bonds between C and O1 atoms are common. Moreover, the C3–O1 double bond stretching mode of vibration occurs at a higher frequency than the C4/C3–O2/O3 single bond mode, as shown previously,³⁶ and the former also shows a slightly lower dipole moment owing to a shorter bond length; this leads to the C4/C3–O2/O3 bonds contributing to higher ϵ_{ionic} .

Guidelines for Property Optimization. The analysis presented in the previous section can be utilized to engineer novel polymers that are likely to show desirable properties for capacitor applications, that is, high dielectric constants and large band gaps. In the past, we found the right mix of chemical constituents in pure organic polymers to optimize both properties simultaneously,^{25,41,55} resulting in the synthesis and testing of polymer repeat units like $-\text{[NH-CO-NH-C}_6\text{H}_4\text{]}-$ and $-\text{[NH-CS-NH-C}_6\text{H}_4\text{]}-$, which showed ϵ_{total} as high as 5 and 6, respectively, and E_{gap} greater than 3 eV. Perhaps more exciting is the substantial increase in ϵ_{ionic} seen in organometallic polymers where metal atoms like Sn, Zn, Cd, and Pb adopt an octahedral coordination surrounded by electronegative atoms like O and F. This led to the successful synthesis of a series of Sn-ester based polymers (with the repeat unit $-\text{[(COO)Sn(CH}_3\text{)}_2\text{COO-(CH}_2\text{)}_n\text{]}-$, $n \in 0, 11$; this is pictured in Figure 2b),^{36,37} with impressive correspondence between experimentally measured and computed properties.

While enhancement in ϵ_{total} from organics to organometallics is undeniable, Figure 4b showed that the organometallics themselves cover a wide range of ϵ_{ionic} values. This comes from the varying concentrations of metal atom in the system, an effect we have explored in previous studies.^{37,41} For instance, in ref 34, it was shown that the Sn-polyesters experienced a general decrease in ϵ_{ionic} (and consequently, ϵ_{total}) as the number of linker CH_2 units increased and metal concentration went down. The same trend was observed for organometallics containing any of the other metals, showing that ϵ_{total} generally increases with the metal content. This also follows from Figure 5b, where fingerprint components corresponding to bonds between 6-fold metal atoms and O/Cl/F atoms (where every component is the fraction of the particular type of bond in the material, and as such a measure of the metal content) showed maximum correlation with ϵ_{ionic} . That said, a certain amount of organic linker in the organometallic polymer is beneficial for easy synthesis and processability into free-standing thin films.⁴¹

The fingerprint–property correlations reveal the importance of the metal concentration, metal coordination number, and identity of its surrounding atoms in determining the dielectric constant. Given the great promise offered by the organometallic polymers, we attempted a closer look at the effect of these factors by plotting ϵ_{total} as a function of the identity of metal atom in Figure 6, using different colors and circle sizes to

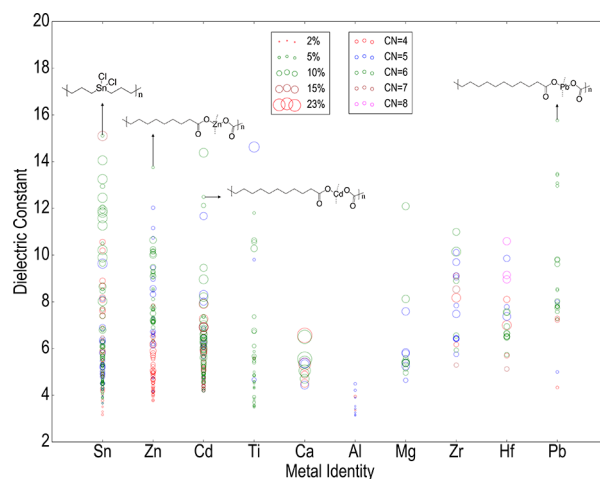


Figure 6. Dependence of dielectric constant on metal identity, volume fraction, and coordination number. As shown in the legends, circle sizes correlate with the metal volume fraction (which ranges from 2% to 23%) whereas different colors correspond to different metal coordination numbers (which vary from 4 to 8). Four polymers that provide the best combination of high dielectric constant and low metal content (restricted to <5%) are pictured in terms of their repeat units (from left to right): $-\text{[(SnCl}_2\text{)}_2\text{(CH}_2\text{)}_6\text{]}-$, $-\text{[Zn(COO)}_2\text{(CH}_2\text{)}_8\text{]}-$, $-\text{[Cd(COO)}_2\text{(CH}_2\text{)}_{10}\text{]}-$, and $-\text{[Pb(COO)}_2\text{(CH}_2\text{)}_{10}\text{]}-$.

distinguish between coordination number and metal content. The latter quantity is the fraction of the metal covalent volume⁸⁹ to the total crystalline volume of the polymer, as estimated from its computationally obtained crystal structure. The metal volume fractions ranged from the lower limit of $\sim 2\%$ to a high of $\sim 23\%$.

It can be seen from Figure 6 that circle size does not uniformly go up with increase in ϵ_{total} , which means metal content is not the only factor affecting the property. In general, the highest ϵ_{total} values were shown by systems where the metal exists in a 6-fold coordination, and systems with 4-fold coordinated metal atoms dominated the lowest ϵ_{total} region. It was further observed that certain polymers containing Sn, Zn, Cd, and Pb display $\epsilon_{\text{total}} > 10$ even with low metal volume fractions of 2–5%. This is an important insight, as it implies that it is possible to boost the dielectric constant of organometallic polymers without the need to insert a very large quantity of metal. Indeed, synthetic and cost concerns dictate that we seek polymers with smaller metal content yet a large ϵ_{total} ; four of the best polymers in this regard are highlighted using their repeat units in Figure 6, one each from the polymers in the Organo-Sn, Zn, Cd, and Pb subdata sets.

The repeat units of the four highlighted polymers, mentioned in the Figure 6 caption, contain metal based units ($\text{SnCl}_2/\text{Zn(COO)}_2/\text{Cd(COO)}_2/\text{Pb(COO)}_2$) flanked by 6 to 10 CH_2 units, which leads to desirably low metal content. It should be noted that, for each of these polymer repeat units, the structure prediction algorithm yielded several competing low energy structures, leading to a whole range of computed dielectric

constants³⁴ (thus justifying the uneven distribution of evenly sized circles in Figure 6). The specific structures we show here are the ones with the respective highest ϵ_{total} , which if experimentally isolated could be very rewarding. The Sn-halide containing polymers were previously recommended for experiments but cast aside owing to difficulty of synthesis,^{36,37,74} while the Pb-ester based polymer may not be the best choice because of the toxicity of Pb. However, a series of Zn- and Cd-ester based polymers (with the repeat unit $-\text{[(COO)-M-COO-(CH}_2\text{)}_n\text{]}_n-$, where $M \in \text{Zn, Cd}$ and $n \in 1, 8$) were recently subjected to synthesis and characterization by our collaborators, and it was seen that not only did their experimentally measured properties compare well with the DFT values, but the relationship between metal content and ϵ_{total} was recovered qualitatively. This will be the subject of a follow up report. Figure 6 can thus prove to be a valuable chart for polymer chemists in their search for the optimal mix of metal atom, coordination number and metal content needed in a polymer backbone for a desired dielectric constant.

Quantitative Prediction of Properties. Studying correlations between different singles and doubles fingerprint components and the properties has helped us qualitatively determine the atoms and bonds one would choose or avoid when looking for a certain desired combination of dielectric and electronic properties. However, an even more valuable utility would be to make a quantitative property prediction for any given polymer based on its fingerprint. In our past work, we developed such models for ϵ_{elec} , ϵ_{ionic} and E_{gap} for a data set of purely organic polymers (pictured in Figure 2a),⁵⁵ using the triples fingerprint based on chemical building blocks (such as CH_2 , NH , C_6H_4 , etc.) as opposed to atom types. Kernel Ridge Regression (KRR)^{4,55} was applied on the data to yield models with prediction errors of less than 10%. Testing on newer polymers proved that these models could, within an acceptable statistical level of accuracy, act as a surrogate for future DFT computations on polymers belonging the same chemical subspaces.

We attempted the same here using the triples fingerprint and employing KRR. Prediction performances for the three properties are shown in the Supporting Information for three distinct subdata sets:

- The data set of purely organic polymers, i.e., all the polymers belonging to the sets Organics-1, Organics-2, and Organics-3.
- The data set of organic and some organometallic polymers, i.e., polymers belonging to the sets Organics-1, Organics-2, Organics-3, Organo-Sn, Organo-Zn, and Organo-Cd.
- All the organic and organometallic polymers presented in this work.

It was seen that KRR prediction performances steadily became worse upon including the organometallic systems, with ϵ_{ionic} especially being harder to predict, owing to the sudden increase in its values, as well as a widening of the range it spans, from organics to organometallics. While relative prediction errors of $\sim 10\%$ are seen for the three properties for pure organics, similar to the performances obtained in ref 55, the errors for the dielectric constants went up to as much as $\sim 30\%$ for the data sets including the organometallics. It can be concluded that more variety in the data set of metal containing polymers is necessary here. Moreover, a stronger dependence of the dielectric constant of organometallic polymers on the

crystal structure and morphology (as evidenced by the wide range of property values sometimes obtained for the same polymer chemical composition) indicates that our fingerprint definition may not be adequate in training predictive models. Factors such as polymer side chains, distances between polymer chains, and percentage of crystallinity could be crucial here and would need to be considered in future studies where quantitative prediction is the goal.

CONCLUSIONS

A large computational data set of polymers and related materials has been mined to unravel trends and hidden chemistry in their dielectric constants and band gaps. By understanding the role of specific atoms and coordination environments, certain “design rules” were developed for optimizing polymers for dielectric (and related) applications. Some of the main conclusions are listed as follows:

- Organic polymers containing a majority of 4-fold C atoms bonded to H atoms show very large band gaps but low electronic dielectric constants and almost negligible ionic dielectric constants.
- Organic polymers containing S atoms bonded to 3-fold C atoms show very high electronic dielectric constants and low band gaps.
- In organic polymers, O and N atoms lead to slightly higher ionic dielectric constants (compared to polymers with only C and H atoms) via bonds with C while maintaining large band gaps.
- Organometallic polymers far outperform pure organics in terms of simultaneously enhancing the dielectric constant and the band gap. The metal atom, the metal concentration, and its coordination chemistry are the important factors that control the polymer’s dielectric constant.
- 6-fold Sn, Zn, Cd, and Pb atoms bonded with O or F lead to the highest ionic dielectric constants observed in this work, with band gaps in the moderate to high regions. For many of these polymers, modest metal fractions (by volume) of 2–5% were sufficient to obtain dielectric constants greater than 10. The same metal atoms in a 4-fold coordination, meanwhile, led to lower dielectric constants between 3 and 5, highlighting the significance of a higher coordination number.

Although the precise atoms and bonds required to manipulate the polymer properties have been revealed, in practice, choosing the right mix of chemical constituents to optimize different properties simultaneously is no trivial task and is further complicated by issues of experimental viability, processability, and stability. Regardless, the process is made more systematic as a result of these design rules, which can be further reinforced and improved upon by infusion of fresh data. Higher amounts of data are being generated on a regular basis by the community, which when combined with an improved fingerprint could lead to more accurate and more general models for the properties of polymers. Further, while qualitative models as obtained here are more than valuable in making forecasts on the properties of materials, accurate prediction models based on regression techniques would enable quantitative estimates of the properties. Such models have been developed using the “triples” fingerprint for subsets of the data in the past^{55,56} and are now a part of an online on-demand polymer design tool.⁵⁴ The most important utility of such

models is in knowing beforehand the properties of new and hypothetical polymers and in enabling decisions on whether to pursue them for a given application or not.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: [10.1021/acs.chemmater.7b02027](https://doi.org/10.1021/acs.chemmater.7b02027).

Three sets of parity plots between KRR model predictions and DFT results, obtained by training on different subdata sets (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: rampi@uconn.edu.

ORCID

Arun Mannodi-Kanakkithodi: [0000-0003-0780-1583](https://orcid.org/0000-0003-0780-1583)

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This paper is based upon work supported by a Multidisciplinary University Research Initiative (MURI) grant (N00014-10-1-0944) from the Office of Naval Research. Computational support was provided by the Extreme Science and Engineering Discovery Environment (XSEDE), the National Energy Research Scientific Computing Center (NERSC), and Los Alamos National Laboratory (LANL).

■ REFERENCES

- (1) Hume-Rothery, W.; Coles, B. *Atomic theory for students of metallurgy*; London: 1969.
- (2) Hall, E. The Deformation and Ageing of Mild Steel: III Discussion of Results. *Proc. Phys. Soc., London, Sect. B* **1951**, *64*, 747.
- (3) Petch, N. The Cleavage Strength of Polycrystals. *J. Iron Steel Inst. London*. **1953**, *173*, 25–28.
- (4) Pilania, G.; Wang, C.; Jiang, X.; Rajasekaran, S.; Ramprasad, R. Accelerating materials property predictions using machine learning. *Sci. Rep.* **2013**, *3*, 2810.
- (5) Pilania, G.; Mannodi-Kanakkithodi, A.; Uberuaga, B. P.; Ramprasad, R.; Gubernatis, J. E.; Lookman, T. Machine learning bandgaps of double perovskites. *Sci. Rep.* **2016**, *6*, 19375.
- (6) Meredig, B.; Agrawal, A.; Kirklin, S.; Saal, J. E.; Doak, J. W.; Thompson, A.; Zhang, K.; Choudhary, A.; Wolverton, C. Combinatorial screening for new materials in unconstrained composition space with machine learning. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2014**, *89*, 094104.
- (7) Faber, F. A.; Lindmaa, A.; von Lilienfeld, O. A.; Armiento, R. Machine Learning Energies of 2 Million Elpasolite (ABC_2D_6) Crystals. *Phys. Rev. Lett.* **2016**, *117*, 135502.
- (8) Hautier, G.; Fischer, C. C.; Jain, A.; Mueller, T.; Ceder, G. Finding Natures Missing Ternary Oxide Compounds Using Machine Learning and Density Functional Theory. *Chem. Mater.* **2010**, *22*, 3762–3767.
- (9) Pilania, G.; Balachandran, P. V.; Kim, C.; Lookman, T. Finding New Perovskite Halides via Machine Learning. *Front. Mater.* **2016**, DOI: [10.3389/fmats.2016.00019](https://doi.org/10.3389/fmats.2016.00019).
- (10) Ramakrishnan, R.; Dral, P. O.; Rupp, M.; von Lilienfeld, O. A. Big Data Meets Quantum Chemistry Approximations: The \hat{I}^m -Machine Learning Approach. *J. Chem. Theory Comput.* **2015**, *11*, 2087–2096.
- (11) Rupp, M.; Tkatchenko, A.; Müller, K.-R.; von Lilienfeld, O. A. Fast and Accurate Modeling of Molecular Atomization Energies with Machine Learning. *Phys. Rev. Lett.* **2012**, *108*, 058301.

(12) Curtarolo, S.; Setyawan, W.; Wang, S.; Xue, J.; Yang, K.; Taylor, R. H.; Nelson, L. J.; Hart, G. L.; Sanvito, S.; Buongiorno-Nardelli, M.; Mingo, N.; Levy, O. AFLOWLIB.ORG: A distributed materials properties repository from high-throughput ab initio calculations. *Comput. Mater. Sci.* **2012**, *58*, 227–235.

(13) Mueller, T.; Kusne, A. G.; Ramprasad, R. Machine Learning in Materials Science: Recent Progress and Emerging Applications. *Rev. Comput. Chem.* **2016**, 186–273.

(14) Botu, V.; Ramprasad, R. Learning scheme to predict atomic forces and accelerate materials simulations. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2015**, *92*, 094306.

(15) Botu, V.; Ramprasad, R. Adaptive machine learning framework to accelerate ab initio molecular dynamics. *Int. J. Quantum Chem.* **2015**, *115*, 1074–1083.

(16) Botu, V.; Batra, R.; Chapman, J.; Ramprasad, R. Machine Learning Force Fields: Construction, Validation, and Outlook. *J. Phys. Chem. C* **2017**, *121*, 511–522.

(17) Botu, V.; Chapman, J.; Ramprasad, R. A study of adatom ripening on an Al (1 1 1) surface with machine learning force fields. *Comput. Mater. Sci.* **2017**, *129*, 332–335.

(18) Li, L.; Snyder, J. C.; Pelaschier, I. M.; Huang, J.; Niranjana, U.-N.; Duncan, P.; Rupp, M.; Müller, K.-R.; Burke, K. Understanding machine-learned density functionals. *Int. J. Quantum Chem.* **2016**, *116*, 819–833.

(19) Hansen, K.; Montavon, G.; Biegler, F.; Fazli, S.; Rupp, M.; Scheffler, M.; von Lilienfeld, O. A.; Tkatchenko, A.; Müller, K.-R. Assessment and Validation of Machine Learning Methods for Predicting Molecular Atomization Energies. *J. Chem. Theory Comput.* **2013**, *9*, 3404–3419.

(20) Jain, A.; Shin, Y.; Persson, K. A. Computational predictions of energy materials using density functional theory. *Nat. Rev. Mater.* **2016**, *1*, 15004.

(21) Jain, A.; Ong, S. P.; Hautier, G.; Chen, W.; Richards, W. D.; Dacek, S.; Cholia, S.; Gunter, D.; Skinner, D.; Ceder, G.; Persson, K. A. Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL Mater.* **2013**, *1*, 011002.

(22) Fischer, C. C.; Tibbetts, K. J.; Morgan, D.; Ceder, G. Predicting crystal structure by merging data mining with quantum mechanics. *Nat. Mater.* **2006**, *5*, 641–646.

(23) Curtarolo, S.; Morgan, D.; Persson, K.; Rodgers, J.; Ceder, G. Predicting Crystal Structures with Data Mining of Quantum Calculations. *Phys. Rev. Lett.* **2003**, *91*, 135503.

(24) Kim, C.; Pilania, G.; Ramprasad, R. From Organized High-Throughput Data to Phenomenological Theory using Machine Learning: The Example of Dielectric Breakdown. *Chem. Mater.* **2016**, *28*, 1304–1311.

(25) Sharma, V.; Wang, C.; Lorenzini, R. G.; Ma, R.; Zhu, Q.; Sinkovits, D. W.; Pilania, G.; Oganov, A. R.; Kumar, S.; Sotzing, G. A.; Boggs, S. A.; Ramprasad, R. Rational design of all organic polymer dielectrics. *Nat. Commun.* **2014**, *5*, 4845.

(26) Zakutayev, A.; Zhang, X.; Nagaraja, A.; Yu, L.; Lany, S.; Mason, T. O.; Ginley, D. S.; Zunger, A. Theoretical Prediction and Experimental Realization of New Stable Inorganic Materials Using the Inverse Design Approach. *J. Am. Chem. Soc.* **2013**, *135*, 10048–10054.

(27) Xue, D.; Balachandran, P. V.; Hogden, J.; Theiler, J.; Xue, D.; Lookman, T. Accelerated search for materials with targeted properties by adaptive design. *Nat. Commun.* **2016**, *7*, 11241.

(28) Takahashi, K.; Tanaka, Y. Material synthesis and design from first principle calculations and machine learning. *Comput. Mater. Sci.* **2016**, *112*, 364–367.

(29) Mannodi-Kanakkithodi, A.; Pilania, G.; Ramprasad, R. Critical assessment of regression-based machine learning methods for polymer dielectrics. *Comput. Mater. Sci.* **2016**, *125*, 123–135.

(30) Jain, A.; Ong, S. P.; Hautier, G.; Chen, W.; Richards, W. D.; Dacek, S.; Cholia, S.; Gunter, D.; Skinner, D.; Ceder, G.; Persson, K. The Materials Project: A materials genome approach to accelerating materials innovation. *APL Mater.* **2013**, *1*, 011002.

- (31) Curtarolo, S.; Setyawan, W.; Wang, S.; Xue, J.; Yang, K.; Taylor, R. H.; Nelson, L. J.; Hart, G. L.; Sanvito, S.; Buongiorno-Nardelli, M.; Mingo, N.; Levy, O. AFLOWLIB.ORG: A distributed materials properties repository from high-throughput ab initio calculations. *Comput. Mater. Sci.* **2012**, *58*, 227–235.
- (32) Saal, J. E.; Kirklin, S.; Aykol, M.; Meredig, B.; Wolverton, C. Materials Design and Discovery with High-Throughput Density Functional Theory: The Open Quantum Materials Database (OQMD). *JOM* **2013**, *65*, 1501–1509.
- (33) Landis, D. D.; Dulak, M.; Greeley, J.; Nestorov, S.; Hummelshoj, J. S.; Bligaard, T.; Norskov, J. K.; Jacobsen, K. W. The Computational Materials Repository. *Comput. Sci. Eng.* **2012**, *14*, 51–57.
- (34) Huan, T. D.; Mannodi-Kanakkithodi, A.; Kim, C.; Sharma, V.; Pilania, G.; Ramprasad, R. A polymer dataset for accelerated property prediction and design. *Sci. Data* **2016**, *3*, 160012.
- (35) Ma, R.; Baldwin, A. F.; Wang, C.; Offenbach, I.; Cakmak, M.; Ramprasad, R.; Sotzing, G. A. Rationally Designed Polyimides for High-Energy Density Capacitor Applications. *ACS Appl. Mater. Interfaces* **2014**, *6*, 10445–10451.
- (36) Baldwin, A. F.; Ma, R.; Mannodi-Kanakkithodi, A.; Huan, T. D.; Wang, C.; Tefferi, M.; Marszalek, J. E.; Cakmak, M.; Cao, Y.; Ramprasad, R.; Sotzing, G. A. Poly(dimethyltin glutarate) as a Prospective Material for High Dielectric Applications. *Adv. Mater.* **2015**, *27*, 346–351.
- (37) Baldwin, A. F.; Huan, T. D.; Ma, R.; Mannodi-Kanakkithodi, A.; Tefferi, M.; Katz, N.; Cao, Y.; Ramprasad, R.; Sotzing, G. A. Rational Design of Organotin Polyesters. *Macromolecules* **2015**, *48*, 2422–2428.
- (38) Ma, R.; Sharma, V.; Baldwin, A. F.; Tefferi, M.; Offenbach, I.; Cakmak, M.; Weiss, R.; Cao, Y.; Ramprasad, R.; Sotzing, G. A. Rational design and synthesis of polythioureas as capacitor dielectrics. *J. Mater. Chem. A* **2015**, *3*, 14845–14852.
- (39) Lorenzini, R.; Kline, W.; Wang, C.; Ramprasad, R.; Sotzing, G. The rational design of polyurea and polyurethane dielectric materials. *Polymer* **2013**, *54*, 3529–3533.
- (40) Treich, G. M.; Nasreen, S.; Mannodi Kanakkithodi, A.; Ma, R.; Tefferi, M.; Flynn, J.; Cao, Y.; Ramprasad, R.; Sotzing, G. A. Optimization of Organotin Polymers for Dielectric Applications. *ACS Appl. Mater. Interfaces* **2016**, *8*, 21270–21277.
- (41) Mannodi-Kanakkithodi, A.; Treich, G. M.; Huan, T. D.; Ma, R.; Tefferi, M.; Cao, Y.; Sotzing, G. A.; Ramprasad, R. Rational Co-Design of Polymer Dielectrics for Energy Storage. *Adv. Mater.* **2016**, *28*, 6277–6291.
- (42) Facchetti, A. Pi-Conjugated Polymers for Organic Electronics and Photovoltaic Cell Applications. *Chem. Mater.* **2011**, *23*, 733–758.
- (43) Chan, S.-H.; Chen, C.-P.; Chao, T.-C.; Ting, C.; Lin, C.-S.; Ko, B.-T. Synthesis, Characterization, and Photovoltaic Properties of Novel Semiconducting Polymers with ThiophenePhenyleneThiophene (TPT) as Coplanar Units. *Macromolecules* **2008**, *41*, 5519–5526.
- (44) Ling, Q.-D.; Liaw, D.-J.; Zhu, C.; Chan, D. S.-H.; Kang, E.-T.; Neoh, K.-G. Polymer electronic memories: Materials, devices and mechanisms. *Prog. Polym. Sci.* **2008**, *33*, 917–978.
- (45) Yan, H.; Chen, Z.; Zheng, Y.; Newman, C.; Quinn, J. R.; Dotz, F.; Kastler, M.; Facchetti, A. A high-mobility electron-transporting polymer for printed transistors. *Nature* **2009**, *457*, 679–686.
- (46) Mannodi-Kanakkithodi, A.; Pilania, G.; Ramprasad, R.; Lookman, T.; Gubernatis, J. E. Multi-objective optimization techniques to design the Pareto front of organic dielectric polymers. *Comput. Mater. Sci.* **2016**, *125*, 92–99.
- (47) Hohenberg, P.; Kohn, W. Inhomogeneous Electron Gas. *Phys. Rev.* **1964**, *136*, B864–B871.
- (48) Kresse, G.; Hafner, J. Ab initio molecular dynamics for liquid metals. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1993**, *47*, 558–561.
- (49) Blöchl, P. E. Projector augmented-wave method. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1994**, *50*, 17953–17979.
- (50) Baroni, S.; de Gironcoli, S.; Dal Corso, A.; Giannozzi, P. Phonons and related crystal properties from density-functional perturbation theory. *Rev. Mod. Phys.* **2001**, *73*, 515–562.
- (51) Bernardini, F.; Fiorentini, V.; Vanderbilt, D. Polarization-Based Calculation of the Dielectric Tensor of Polar Crystals. *Phys. Rev. Lett.* **1997**, *79*, 3958–3961.
- (52) Heyd, J.; Scuseria, G. E.; Ernzerhof, M. Hybrid functionals based on a screened Coulomb potential. *J. Chem. Phys.* **2003**, *118*, 8207–8215.
- (53) Heyd, J.; Peralta, J. E.; Scuseria, G. E.; Martin, R. L. Energy band gaps and lattice parameters evaluated with the Heyd-Scuseria-Ernzerhof screened hybrid functional. *J. Chem. Phys.* **2005**, *123*, 174101.
- (54) Polymer Genome: A Recommendation Engine for the Rapid Design and Discovery of Polymer Dielectrics Powered by Quantum Mechanical Computations, Experimental Data and Machine Learning. www.polymergenome.org; 2016.
- (55) Mannodi-Kanakkithodi, A.; Pilania, G.; Huan, T. D.; Lookman, T.; Ramprasad, R. Machine Learning Strategy for Accelerated Design of Polymer Dielectrics. *Sci. Rep.* **2016**, *6*, 20952.
- (56) Huan, T. D.; Mannodi-Kanakkithodi, A.; Ramprasad, R. Accelerated materials property predictions and design using motif-based fingerprints. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2015**, *92*, 014106.
- (57) Leach, A. R.; Gillet, V. *An Introduction to Chemoinformatics*; Springer: 2007.
- (58) Hansen, K.; Montavon, G.; Biegler, F.; Fazli, S.; Rupp, M.; Scheffler, M.; von Lilienfeld, O. A.; Tkatchenko, A.; Müller, K.-R. Assessment and Validation of Machine Learning Methods for Predicting Molecular Atomization Energies. *J. Chem. Theory Comput.* **2013**, *9*, 3404–3419.
- (59) Le, T.; Epa, V. C.; Burden, F. R.; Winkler, D. A. Quantitative StructureProperty Relationship Modeling of Diverse Materials Properties. *Chem. Rev.* **2012**, *112*, 2889–2919.
- (60) Faber, F.; Lindmaa, A.; von Lilienfeld, O. A.; Armiento, R. Crystal structure representations for machine learning models of formation energies. *Int. J. Quantum Chem.* **2015**, *115*, 1094–1101.
- (61) Chowdhury, A.; Kautz, E.; Yener, B.; Lewis, D. Image driven machine learning methods for microstructure recognition. *Comput. Mater. Sci.* **2016**, *123*, 176–187.
- (62) Dolgirev, P. E.; Oganov, A. R. Machine learning scheme for fast extraction of chemically interpretable interatomic potentials. *AIP Adv.* **2016**, *6*, 085318.
- (63) Lee, J.; Seko, A.; Shitara, K.; Nakayama, K.; Tanaka, I. Prediction model of band gap for inorganic compounds by combination of density functional theory calculations and machine learning techniques. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2016**, *93*, 115104.
- (64) Seko, A.; Maekawa, T.; Tsuda, K.; Tanaka, I. Machine learning with systematic density-functional theory calculations: Application to melting temperatures of single- and binary-component solids. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2014**, *89*, 054303.
- (65) Snyder, J. C.; Rupp, M.; Hansen, K.; Müller, K.-R.; Burke, K. Finding Density Functionals with Machine Learning. *Phys. Rev. Lett.* **2012**, *108*, 253002.
- (66) Pozun, Z. D.; Hansen, K.; Sheppard, D.; Rupp, M.; Müller, K.-R.; Henkelman, G. Optimizing transition states via kernel-based machine learning. *J. Chem. Phys.* **2012**, *136*, 174101.
- (67) Balabin, R. M.; Lomakina, E. I. Neural network approach to quantum-chemistry data: Accurate prediction of density functional theory energies. *J. Chem. Phys.* **2009**, *131*, 074104.
- (68) Ramakrishnan, R.; Hartmann, M.; Tapavicza, E.; von Lilienfeld, O. A. Electronic spectra from TDDFT and machine learning in chemical space. *J. Chem. Phys.* **2015**, *143*, 084111.
- (69) Vu, K.; Snyder, J. C.; Li, L.; Rupp, M.; Chen, B. F.; Khelif, T.; Müller, K.-R.; Burke, K. Understanding kernel ridge regression: Common behaviors from simple functions to density functionals. *Int. J. Quantum Chem.* **2015**, *115*, 1115–1128.
- (70) Ward, L.; Wolverton, C. Atomistic calculations and materials informatics: A review. *Curr. Opin. Solid State Mater. Sci.* **2017**, *21*, 167–176.

(71) Ghiringhelli, L. M.; Vybiral, J.; Levchenko, S. V.; Draxl, C.; Scheffler, M. Big Data of Materials Science: Critical Role of the Descriptor. *Phys. Rev. Lett.* **2015**, *114*, 105503.

(72) Callister, W. D. *Materials Science and Engineering: An Introduction*; Wiley: 2007.

(73) Zhu, Q.; Oganov, A. R.; Glass, C.; Stokes, H. Constrained evolutionary algorithm for structure prediction of molecular crystals: methodology and applications. *Acta Crystallogr., Sect. B: Struct. Sci.* **2012**, *68*, 215–226.

(74) Pilania, G.; Wang, C. C.; Wu, K.; Sukumar, N.; Breneman, C.; Sotzing, G.; Ramprasad, R. New Group IV Chemical Motifs for Improved Dielectric Permittivity of Polyethylene. *J. Chem. Inf. Model.* **2013**, *53*, 879–886.

(75) Kresse, G.; Hafner, J. *Ab initio* molecular dynamics for liquid metals. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1993**, *47*, 558–561.

(76) Liu, C.-S.; Pilania, G.; Wang, C.; Ramprasad, R. How Critical Are the van der Waals Interactions in Polymer Crystals? *J. Phys. Chem. A* **2012**, *116*, 9347–9352.

(77) Zhu, Q.; Sharma, V.; Oganov, A. R.; Ramprasad, R. Predicting polymeric crystal structures by evolutionary algorithms. *J. Chem. Phys.* **2014**, *141*, 154102.

(78) Goedecker, S. Minima hopping: An efficient search method for the global minimum of the potential energy surface of complex molecular systems. *J. Chem. Phys.* **2004**, *120*, 9911–9917.

(79) Amsler, M.; Goedecker, S. Crystal structure prediction using the minima hopping method. *J. Chem. Phys.* **2010**, *133*, 224104.

(80) Van Krevelen, D.; Nijenhuis, K. T. *Properties of Polymers*; Elsevier: 2009.

(81) Davies, A. G.; Symes, W. R. Organometallic reactions. Part VI. The addition of the Sn-O bond to the carbonyl group. *J. Chem. Soc. C* **1967**, *0*, 1009–1016.

(82) Dou, L.; Liu, Y.; Hong, Z.; Li, G.; Yang, Y. Low-Bandgap Near-IR Conjugated Polymers/Molecules for Organic Electronics. *Chem. Rev.* **2015**, *115*, 12633–12665.

(83) Schwerdtfeger, P. Atomic Static Dipole Polarizabilities. *Computational Aspects of Electric Polarizability Calculations: Atoms, Molecules and Clusters* **2006**, 1–32.

(84) Barrett, R. L. Electronegativity chart. *J. Chem. Educ.* **1962**, *39*, 251.

(85) Huan, T. D.; Boggs, S.; Teyssedre, G.; Laurent, C.; Cakmak, M.; Kumar, S.; Ramprasad, R. Advanced polymeric dielectrics for high energy density applications. *Prog. Mater. Sci.* **2016**, *83*, 236–269.

(86) Wang, C. C.; Pilania, G.; Ramprasad, R. Dielectric properties of carbon-, silicon-, and germanium-based polymers: A first-principles study. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2013**, *87*, 035103.

(87) Mannodi-Kanakkithodi, A.; Wang, C. C.; Ramprasad, R. Compounds based on Group 14 elements: building blocks for advanced insulator dielectrics design. *J. Mater. Sci.* **2015**, *50*, 801–807.

(88) Stigler, S. M. Francis Galton's Account of the Invention of Correlation. *Statist. Sci.* **1989**, *4*, 73–79.

(89) The Periodic Table by WebElements. <https://www.webelements.com/> (accessed January 2017).