

CHAPTER 9

Rational Design of Polymer Dielectrics: An Application of Density Functional Theory and Machine Learning

A. MANNODI-KANAKKITHODI AND R. RAMPRASAD*

University of Connecticut, Materials Science and Engineering Department,
N Eagleville Road, Storrs, CT-06269, USA

*Email: rampi.ramprasad@uconn.edu

9.1 Introduction

9.1.1 General Background

Throughout human history, every age and every culture has perhaps been best defined by the materials they used. Prehistoric humans carved tools out of bone and wood and used them for hunting. The stone age, which started nearly 3 million years ago and lasted till around 3000 BCE, was characterized by the use of stone in collecting food and building shelters. While much of the less advanced parts of the world remained in the stone age for a long time, the advent of metallurgy kick-started the bronze-age in eastern and southern Asia around 7000 BCE, before it made its way to Europe. Iron in its native metallic state was already being used during the bronze age, but the true iron age is said to have started around 1000 BCE as humans found the means to smelt iron ore. With metal-working now commonplace, the next 2000 or so years saw marked improvements in production and processing of

Computational Materials Discovery

Edited by Artem R. Oganov, Gabriele Saleh and Alexander G. Kvashnin

© The Royal Society of Chemistry 2019

Published by the Royal Society of Chemistry, www.rsc.org

metals and alloys, woodworking, paper, glasses, ceramics and polymers, ultimately leading to the industrial revolution in Europe in the 18th century. 1

The pace of progress in the 20th century was more dramatic than ever before, and for this, scientists had years of documented knowledge and vast swathes of data pertaining to failed and successful experiments to thank. 5
The advent of high-powered special purpose machinery and mass factory production saw stainless steel become mainstream, and incredible advances in transportation, building and communication. One area where iterative experiments and past data majorly benefited materials design was alloys: it was realized that with additions of different amounts of carbon, chromium, 10
nickel, manganese and molybdenum, the properties of steel can be tailored.¹ Solid solutions of aluminium with copper found applications in the aeronautical industry,² and NiTi-based alloys found amazing shape-memory applications.³ This period also saw the development of some of the most important phenomenological models in materials science, such as the 15
Hume–Rothery rules⁴ and the Hall–Petch relationship,⁵ which emerged from experimental documentations on solid solutions and mild steels, respectively.

In the latter half of the 20th century, materials research was taken over by a romantic notion: that of designing materials on a computer before a single 20
laboratory experiment is performed. The accumulation of data *via* experiments, while invaluable, was seen to be time intensive and prone to human observational errors. Today, massive parallel supercomputers with thousands of processors are being used the world over in weather forecasting, oil and gas exploration, and molecular modelling. The advent of super- 25
computers along with theoretical advancements in classical mechanics^{6,7} and quantum mechanics,^{8,9} formulations of force-field simulations and molecular dynamics,^{10–12} and the development of quantum mechanics based methods like density functional theory (DFT)^{13,14} formally kick-started the era of computational materials science.¹⁵ 30

Quantum mechanics, which provided a fundamental look at the structure and properties of materials in the smallest available length and time scales, made for accurate (but computationally expensive) solutions of many materials science problems. Perhaps the most popular approach in this regard is DFT, where Schrodinger’s equation is solved for a many electron system by 35
converting it into an effective one-electron problem. The accuracy of DFT in investigating the electronic structure of atoms, molecules and condensed phases has been well demonstrated, and it is being widely used today to study the mechanical, electronic, dielectric and thermodynamic properties of metals, inorganic compounds, molecules and polymers.^{16–22} One significant transformation that computational materials science underwent 40
over the last 50 years was the evolution of methods like DFT from being merely *post hoc* (*i.e.*, being applied to study materials and explain observations post-experiment) to driving rational materials design by eliminating guesswork from experiments.²³ In the literature, many glittering examples 45
can be found of DFT-driven experiments leading to the accelerated design of

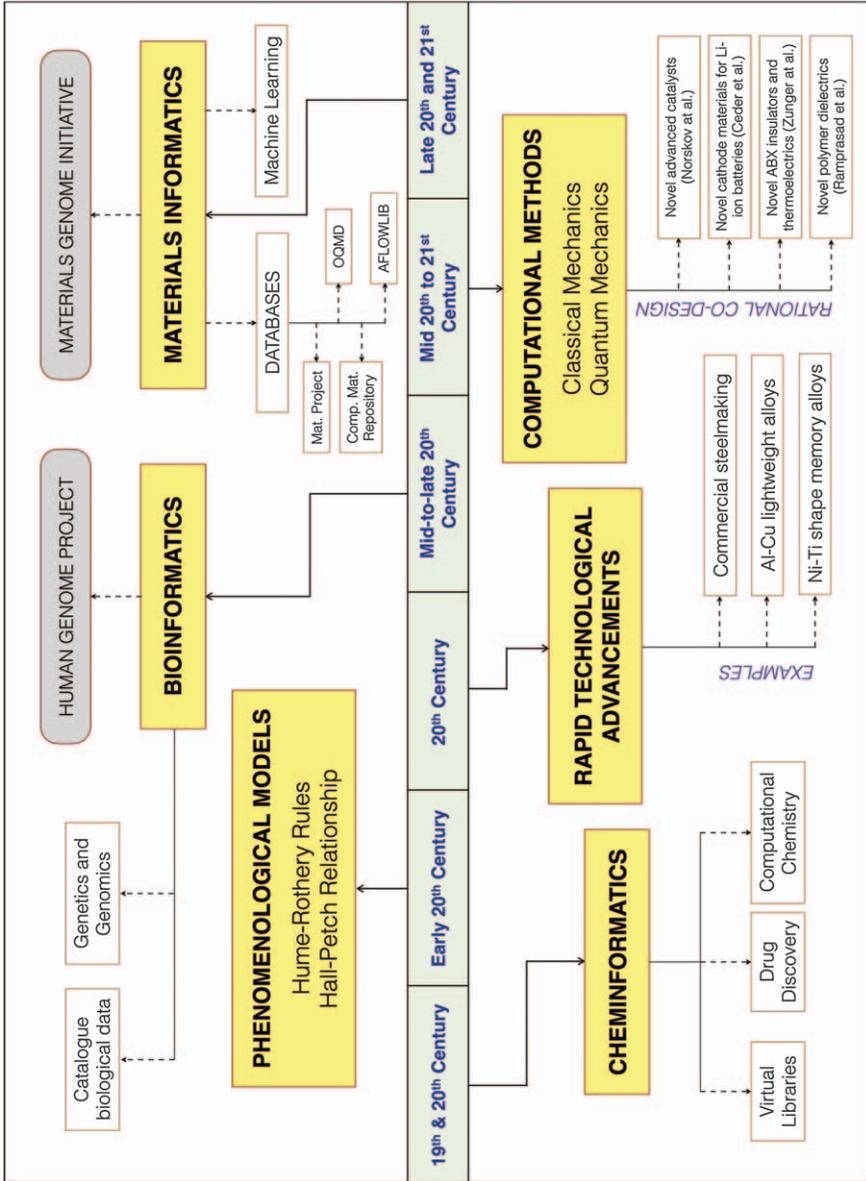
new materials, such as the identification of new cathode materials for Li batteries,²⁴ the design of novel NiTi shape-memory alloys,^{25,26} and the discovery of previously unknown ABX type thermoelectrics and conductors.²⁷

It must be emphasised that *data* more than anything has been the great ally of the scientist in driving innovation and the discovery of physical and chemical laws. While approximate or phenomenological models enable the quick screening and design of materials, precise theories facilitate the generation of robust materials data which can in turn lead to newer, more reliable phenomenological models. Indeed, data generation, storage, retrieval and analysis has been of key importance in the fields of cheminformatics²⁸ and bioinformatics²⁹ over the last century or so, and in the last few years, in *materials informatics*.^{30,31} The latter is a blossoming field in materials science today, focusing on the development of experimental and computational databases and on the application of modern machine learning or data mining methods that help convert the data into easily accessible models.

Figure 9.1 tries to capture a rough timeline of developments in materials science and related fields over the years, in the form of experiment-driven phenomenological models such as the Hume–Rothery rules, computational theories such as classical and quantum mechanics, and data-driven fields in chemistry (cheminformatics), biology (bioinformatics) and materials science (materials informatics). In recent years, there has been further recognition of the power of computations and databases in guiding the rational experimental design of materials in the form of the *Materials Genome Initiative*²³ (along the same lines as the Human Genome Project³²), announced by the US government “to discover, manufacture and deploy advanced materials twice as fast, at a fraction of the cost”. High-performance computing, efficient computational approaches and machine learning based methods provide great promise in accelerating the pace of discovery and deployment of new materials in practice.

9.1.2 Polymers as Capacitor Dielectrics

In this chapter, we discuss the application of all the ideas described above—computational modelling, guided experiments and materials informatics—towards the design of new and advanced polymer dielectrics for energy storage capacitor applications. 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.^{33–36} A capacitor, consisting of a polarizable dielectric material in between two conductive metal plates (a schematic is shown in Figure 9.2), 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 BaTiO₃ and TiO₂ provide the benefit of massive dielectric constants, polymers are preferred capacitor



1

5

10

15

20

25

30

35

40

45

dielectrics for energy storage because of their easy processability, flexibility, high resistance to external chemical attacks and most importantly, propensity for graceful failure.^{33,37-40}

The state-of-the-art polymer that is used as a dielectric in high energy density metallized film capacitors is biaxially oriented polypropylene (BOPP), which shows a large breakdown field of $> 700 \text{ MV m}^{-1}$, a low dielectric loss and a small area ($\sim 1 \text{ cm}^2$), but also a low dielectric constant (~ 2.2) and low operating temperature ($85 \text{ }^\circ\text{C}$).⁴¹ While this leads to a respectable energy density of $\sim 6 \text{ J cm}^{-3}$, the low dielectric constant of BOPP and the dielectric losses due to electronic conduction that it suffers at higher temperatures makes it appropriate to search for a desirable alternative. In the past, most of these efforts were concerned with polyvinylidene fluoride (PVDF) and related polymers, which provided the immediate advantage of orientational polarization and high dipole density, unlike BOPP. Many derivatives of PVDF such as defect-modified PVDF, PVDF-HFP (hexafluoropropylene) and PVDF-CTFE (chlorotrifluoroethylene) have been studied as polymer dielectrics and seen to possess dielectric constants ~ 10 and energy densities $\sim 30 \text{ J cm}^{-3}$.⁴²⁻⁴⁶ However, the ferroelectric behaviour of these polymers led to heavy energy losses. Further, highly polar linear polymers like polyurea and polythiourea were studied, motivated by the known high dipole moments of the urea and thiourea groups. Aromatic polyurea thin films showed dielectric constants > 4 and breakdown strength $\sim 700 \text{ MV m}^{-1}$, as well as an impressive energy density of 9 J cm^{-3} and an efficiency $> 95\%$.^{47,48} An aromatic polythiourea further improved upon the dielectric constant and breakdown strength, leading to even better energy density.^{49,50} Some other efforts to improve upon BOPP include polar group-modified polycarbonates,^{51,52} polysulfones,^{53,54} polyethylene terephthalate (PET)⁵⁵ and nanocomposites where inorganic materials like BaTiO_3 are embedded into the polymer matrix.^{56,57}

Although various alternatives for BOPP were thus devised, all of them were seen to suffer from one shortcoming or another. Today, 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. 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 to date.^{58,59} This makes a computation-driven treatment appropriate here, and a general framework for rationally designing new polymer dielectrics was laid out as presented in Figure 9.3. Such a strategy allows the systematic study of selected chemical subspace(s) of polymers and helps guide experiments in a rational manner.

Figure 9.1 A timeline of major developments in materials science and related fields over the last couple of centuries. Along the same lines as the Human Genome Project (initiated in the 1990s to determine the DNA sequence of the entire human genome), the Materials Genome Initiative was launched a few years ago to accelerate the design and deployment of new and advanced materials.

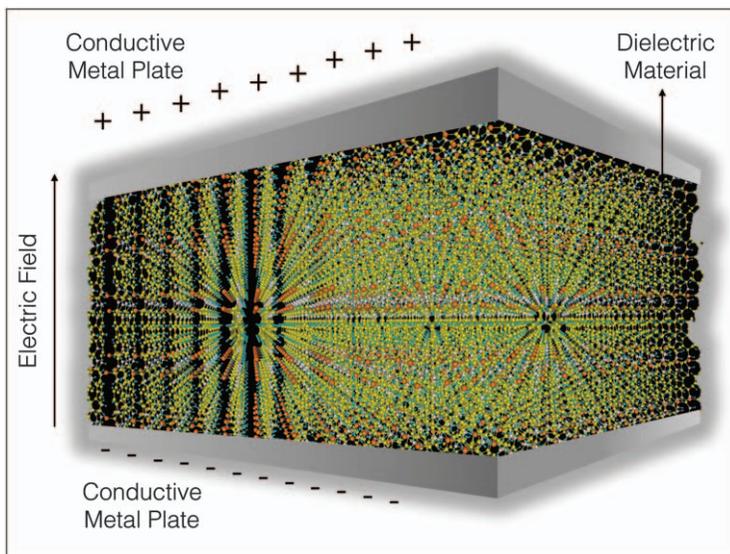


Figure 9.2 Schematic of a capacitor with the metal plates, dielectric material (polymer in this work) and applied electric field labelled.

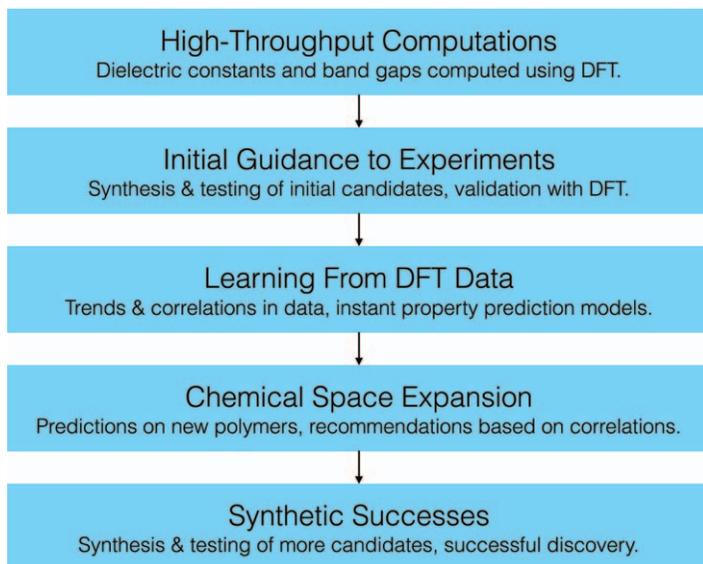


Figure 9.3 The polymer dielectrics design strategy, involving computational guidance, targeted experiments and machine learning.

The dielectric constant and the band gap (known to show correlations with the dielectric breakdown field⁶⁰) were chosen as the two properties that provide an adequate first stage of screening for potential energy storage capacitor dielectrics, as large values of both properties are likely to lead to

high energy densities.^{58,59,61} DFT was chosen as the ideal computational treatment to study the ground state structures, electronic properties and dielectric behaviour of polymeric materials.

The first step in the computation-guided design strategy was performing high-throughput DFT computations (“high-throughput” implying the use of computational resources in an automated manner over a long period) to estimate the relevant properties of polymers belonging to a selected chemical space and screening for promising candidates.^{58,62} Initial recommendations were made for synthesis, and experimental measurements of the same properties provided validation for the DFT results.^{58,59} While the two steps together constitute *rational discovery*, the design process went far beyond to include “learning” from the DFT data: this involved looking for correlations between properties and crucial attributes of the polymers, as well as training machine learning models to facilitate property predictions for newer polymers. This learning was applied to perform chemical space expansion, *i.e.*, to predict the properties of thousands of new polymers without the need to perform more expensive computation.⁶² These predictions provided further recommendations for experiments and fresh computation, paving the way to a successful data-driven design of polymer dielectrics. In the following sections, the computation-guided design strategy is described in detail, in the form of high-throughput computational work on organic and organometallic polymer chemical spaces, the synthetic successes that followed the initial computations, and learning from the computational data that led to useful design rules and prediction models.

9.2 Organic and Organometallic Polymers as Dielectrics

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)^{63–65} is a powerful technique where the dielectric constant of a material is computed by studying the system responses to external perturbations, in this case, electric fields. The band gap can be computed using the hybrid Heyd–Scuseria–Ernzerhof HSE06 electronic exchange-correlation functional,^{66,67} which corrects for the band gap underestimation associated with standard DFT. Dielectric constants and band gaps computing using DFPT and the HSE06 functional respectively have been shown to match up very well with experimentally measured results for inorganic compounds as well as common polymers.^{58,68} Thus, these methods were selected for performing the high-throughput DFT computations.

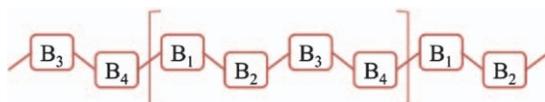
While polymers are known to be either amorphous or semi-crystalline in nature, a crucial assumption made here was to consider a closely packed crystalline model. Although crystal structural information (lattice parameters and bond lengths) is available for many well-known polymers like

polyethylene, PVDF and polyacetylene, there isn't 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.

9.2.1 High-throughput DFT on an Organic Polymer Chemical Space

An organic polymer chemical space (shown in Figure 9.4) consisting of seven basic building blocks—CH₂, NH, CO, C₆H₄, C₄H₂S, CS and O—was selected for initial high-throughput computations. Any *n*-block polymer here was generated by linearly connecting *n* blocks with each of them drawn from the seven possibilities. If *n* was restricted to be *four*, there were ~400 possible unique permutations, of which ~300 remained when chemically unfeasible block pairings (such as CO–CO and NH–NH) were eliminated.^{58,59,62} The final list of 4-block polymers so obtained contained 284 members, and DFT calculations were carried out for all these systems. Crystal structure prediction for so many polymers is no trivial task, especially with scant information available in the literature given that most of these polymers would be hypothetical systems (at least at the first stage). However, recipes for computational prediction of polymeric crystal structures have been well studied in the past.^{69,70} In this work, a structure prediction algorithm known as minima hopping^{71,72} was applied to determine the lowest energy relative packing arrangement of polymer chains (with all energies computed using DFT) in a unit cell, which was then taken to be the ground state crystal structure for the given polymer and used for property computation.

The DFT computed dielectric constants and band gaps for the 284 polymers are plotted against each other in Figure 9.5. From DFPT, the dielectric constant is computed as two separate components: the electronic part, which depends on atomic polarizabilities, and the ionic part, which comes from the IR-active vibrational modes present in the system. The total dielectric constant is expressed as a sum of the electronic and the ionic parts. The casual observer would note right away that the electronic dielectric constant appears to be constrained by some sort of an inverse relationship



$$B_i \in \{\text{NH}, \text{CO}, \text{C}_6\text{H}_4, \text{C}_4\text{H}_2\text{S}, \text{CS}, \text{O}, \text{CH}_2\}$$

Figure 9.4 The chemical subspace of polymers generated by linear combinations of seven basic chemical units.

Reproduced from ref. 58 with permission from John Wiley and Sons, © 2016 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim.

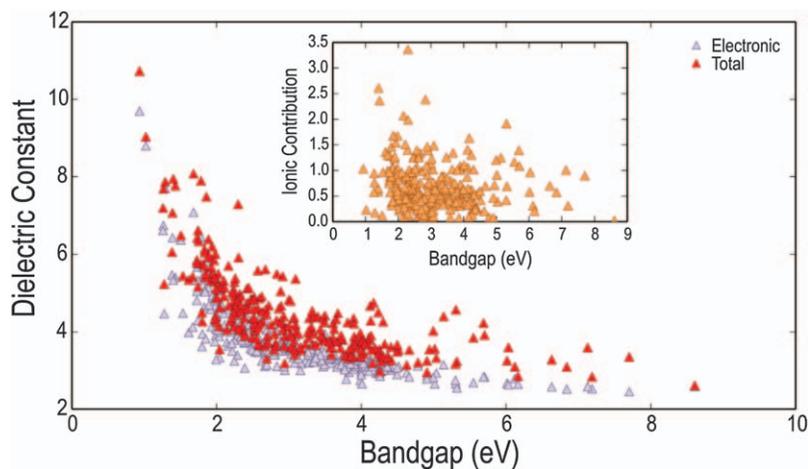


Figure 9.5 The dielectric constants (divided into electronic and ionic parts) and band gaps of 284 polymers computed using DFT.

with the band gap, whereas the ionic dielectric constant shows little or no correlation with the band gap. This effect translates to an inverse relationship in the plot between total dielectric constant and band gap as well, given the larger range of values of the electronic part compared to the ionic part.

It is also interesting that a wide spectrum of dielectric constant values (~ 2 to ~ 12) and band gap values (~ 1 eV to ~ 9 eV) were covered by this chemical subspace of polymers, which led to only roughly 10% of the total points populating the shaded *high dielectric constant, large band gap* region. Regardless, this region provided a few promising candidates for initial experiments as well as leads on the most profitable building block combinations for simultaneously enhancing the two properties. For instance, it was observed that polymers containing urea ($-\text{NH}-\text{CO}-\text{NH}-$), thiourea ($-\text{NH}-\text{CS}-\text{NH}-$) or imide ($-\text{CO}-\text{NH}-\text{CO}-$) linkages alongside an aromatic ring such as $-\text{C}_6\text{H}_4-$ or $-\text{C}_4\text{H}_2\text{S}-$ were present in abundance in the shaded region;^{58,59} subsequently, a few such polymers were considered for experimental studies.

9.2.2 Initial Guidance to Experiments

Three polymers belonging to three distinct polymer classes—polyurea, polyimide and polythiourea—were selected out of the shaded region in Figure 9.5 and synthesized in the laboratory.⁵⁹ Appropriate monomers and reaction schemes were adopted here to yield satisfactory quantities of each polymer, following which ultraviolet-visible spectroscopy (UV-Vis) was performed to estimate the band gaps and time domain dielectric spectroscopy (TDDS) to measure the dielectric constants. As seen from Table 9.1, the experimental results matched quite well with the computational results, providing not only a validation for the high-throughput DFT scheme, but also

Table 9.1 Experimentally measured properties for initial recommendations (listed using the polymer repeat units) from high-throughput DFT, and a comparison with DFT computed values.

Polymer	DFT band gap (eV)	Expt. band gap (eV)	DFT dielectric constant	Expt. dielectric constant
$-\text{[NH-CO-NH-C}_6\text{H}_4\text{]}_n-$	~3.5	~3.9	~4.9	~5.6
$-\text{[CO-NH-CO-C}_6\text{H}_4\text{]}_n-$	~4.1	~4.0	~5.7	~X.5
$-\text{[NH-CS-NH-C}_6\text{H}_4\text{]}_n-$	~2.7	~3.1	~5.8	~6.2

three novel promising polymer dielectric candidates for energy storage capacitor applications. However, it was seen that these initial polymers had solubility issues and could not be processed into thin films, which is an important capacitor dielectric requirement. To overcome these issues, newer, longer chain polymers belonging to the same and related polymer classes were pursued; this is discussed more in Section 9.3.

9.2.3 Moving Beyond Pure Organics: An Organometallic Polymer Chemical Space

While interesting new organic polymer motifs were identified as potential capacitor dielectrics, the low ionic dielectric constants seen throughout for the pure organics hinted at a missed opportunity. The lack of correlation between the ionic dielectric constant and the band gap suggested that the former could perhaps be enhanced without adversely affecting the latter.⁵⁸ Studies carried out for the oxides and halides of group 14 elements showed that Pb, Sn and Ge based compounds have much higher dielectric constants than their C or Si counterparts, as well as band gap values around or greater than 4 eV.^{73,74} This led to the following thought experiment: if metal based units were inserted in the backbone of an otherwise organic polymer (for instance, polyethylene, $-(\text{CH}_2)_n-$), there could potentially be an increase in the dielectric constant compared to the pure organic, while maintaining a large band gap. Metal-organic frameworks (MOFs), which are compounds containing metal clusters surrounded by organic ligands, are commonly used for gas storage, catalysis and supercapacitors.⁷⁵ Along similar lines, a metal-organic polymer framework was proposed wherein the organic polymer chain would be interrupted by a metal containing unit. For initial study, Sn was chosen as the metal atom over the poisonous Pb or expensive Ge. Polymer repeat units were generated by introducing tin difluoride ($-\text{SnF}_2-$), tin dichloride ($-\text{SnCl}_2-$) and dimethyltin-ester ($-\text{COO-Sn}(\text{CH}_3)_2-\text{COO}-$) units in polyethylene chains in varying amounts.⁷⁶⁻⁷⁹ DFT calculations showed that these systems indeed display superior dielectric constants compared to organics for a given band gap value; this caused much excitement in terms of prospective experiments, and the Sn-ester based polymers were duly synthesized and tested, as described in detail in the next section.

The computation-driven discovery of novel Sn-based organometallic polymers paved the way for a sweeping exploration of polymers containing different metals chosen from the periodic table. In Figure 9.6, DFT computed results are presented for organometallic polymers constituted of (respectively) 10 different metal atoms;^{60,80} also, shown for comparison are all the organics discussed in Section 9.2.1. The metal-based systems clearly surpass the pure organics in terms of high dielectric constants for given values of band gap. The primary reason behind this increase is the enhanced polarity of chemical bonds in the organometallics because of bonding between electropositive metal atoms and highly electronegative atoms such as O, F and Cl. The swinging and stretching of these polar bonds at low frequencies cause fluctuations in polarization under electric fields, which means they will contribute more to ionic or dipolar parts of the dielectric constant.^{60,68,74} As seen from Figure 9.6, this effect is more pronounced in some organometallics than others: it was observed that the higher the amount of metal in the system, the higher is the dielectric constant. The identity of the metal atom itself and its coordination environment were other crucial factors at play here.⁸⁰

The next section describes all the parallel experimental efforts that brought the computer-modelled, potentially game-changing materials to life. These include the second generation of computation-guided organic

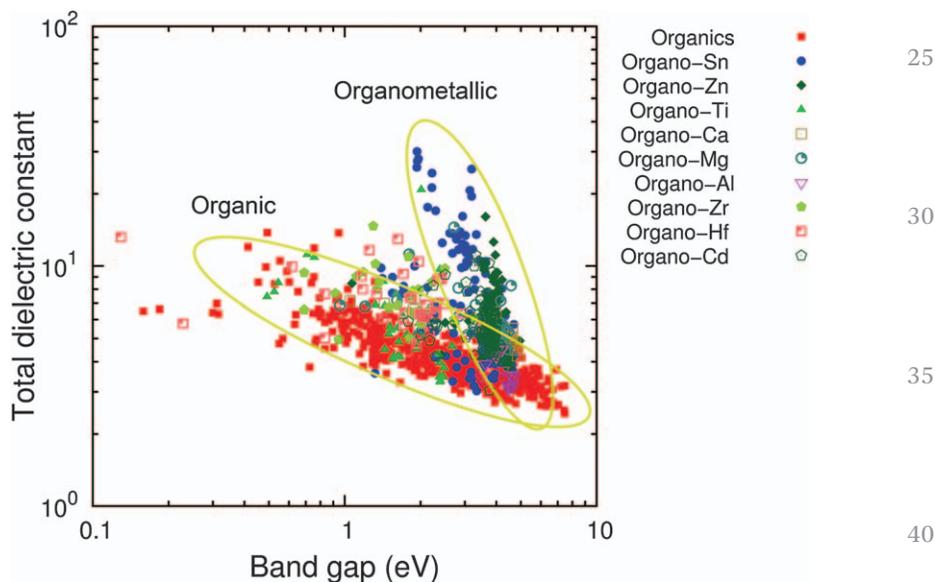


Figure 9.6 DFT computed band gaps and dielectric constants for all organic and organometallic polymers. The organometallics show higher dielectric constants than the organics for a given band gap.

Reproduced from ref. 58 with permission from John Wiley and Sons, © 2016 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim.

polymers that followed from those first described in Section 9.2.2, as well as an entire series of Sn-ester based polymers.

9.3 Synthetic Successes

Without the knowledge attained from modelling polymers on a computer, the polymer chemist might end up lost in a sea of possibilities, much like an explorer setting sail on a rudderless ship. The computational models may be viewed as the GPS to the experimentalist, telling him or her about potentially promising directions to take. The study of organic polymers as described in Section 9.2.1 revealed that NH-CO-NH, NH-CS-NH or CO-NH-CO linkages accompanied by aromatic rings were particularly useful in boosting the dielectric constants and band gaps. This led to the initial synthesis and characterization of three new polymers as described in Section 9.2.2, which helped validate the DFT computations.^{58,59} However, the processability and solubility concerns inspired a foray into a second generation of organic polymer motifs: several new polymers belonging to generic polymer classes—polyureas, polythioureas, polyurethanes and polyimides—were thus synthesized and tested,^{81–84} as pictured in Figure 9.7. Free-standing films were made from most of these polymers, and their dielectric constants, band gaps, dielectric breakdown strengths and loss characteristics, among other properties, were experimentally measured.

Table 9.2 provides a glimpse of three newly designed organic polymers with the best characteristics and compares their (experimentally measured) properties with the state-of-the-art polymer dielectric, BOPP. The three polymers are a polythiourea named PDTC-HDA, a polyimide named BTDA-HDA and another polyimide named BTDA-HK511, where PDTC stands for *para*-phenylene diisothiocyanate, HDA stands for hexane diamine, BTDA stands for benzophenone tetracarboxylic dianhydride and HK511 is a jeffamine-containing ether. Apart from the properties listed earlier, the recoverable energy densities were also estimated for all the polymers using electric displacement–electric field (D–E) loop measurements. Apart from forming free-standing films, each polymer displayed an energy density two to three times higher than BOPP. In this fashion, (at least) three new organic polymers were successfully designed that can potentially replace BOPP in capacitor applications.⁵⁸ The rationale for pursuing these kinds of polymers came from computational guidance; however, the choice of the specific polymer repeat units was determined by the polymer chemists using their experience and knowledge of chemical feasibility, solvent considerations and film formability. The experimental data thus obtained further bolsters the polymer dataset and even provides vital leads on newer chemical blocks to introduce in polymers for future computational studies.

Following the fruits yielded by the computation-driven work on organic polymers, attention was diverted to the exciting new field of organometallic polymers. Synthesis of the organo-Sn polyesters proved to be challenging, but the polymer chemists were able to make 12 such polymers containing a

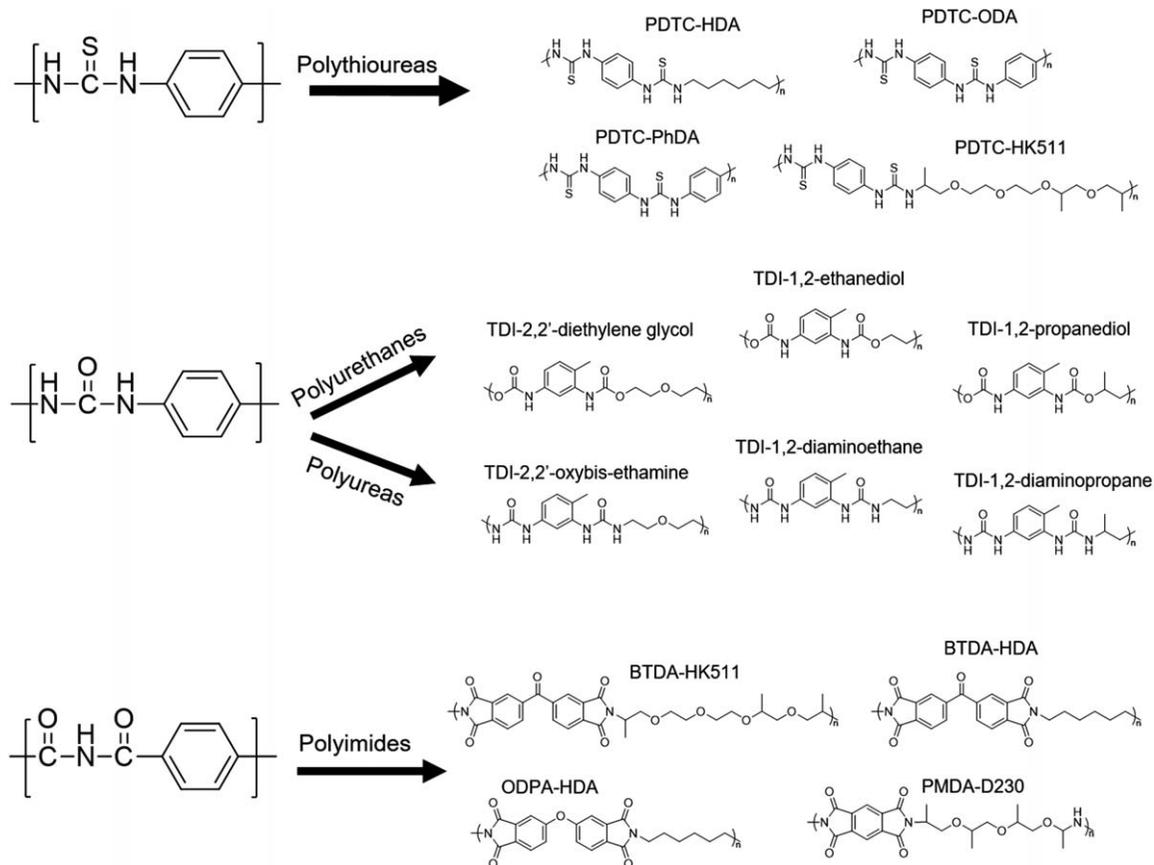


Figure 9.7 The second generation of organic polymers (belonging to different generic polymer classes) that were synthesized based on guidance from computations and the first-generation organics discussed in Section 9.2.2. Reproduced from ref. 58 with permission from John Wiley and Sons, © 2016 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim.

Table 9.2 Measured properties for PDTC–HDA, BTDA–HDA and BTDA–HK511, three of the best novel organic polymer dielectrics designed using computational guidance and targeted experiments. Also, shown for comparison are properties for BOPP (bi-axially oriented polypropylene).

Polymer name	Polymer class	Dielectric constant	Breakdown field (MV m ⁻¹)	Energy density (J cm ⁻³)
BOPP	Polypropylene	~2.2	~700	~5
PDTC–HDA	Polythiourea	~3.7	~685	~9
BTDA–HDA	Polyimide	~3.6	~812	~10
BTDA–HK511	Polyimide	~7.8	~676	~16

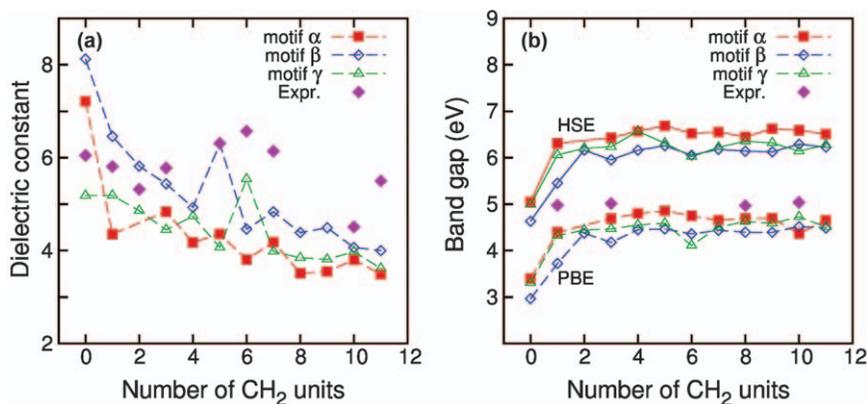


Figure 9.8 Computational and experimental dielectric constants and band gaps for a series of organo-Sn polyesters as a function of the number of linker $-\text{CH}_2-$ units.

Reproduced from ref. 58 with permission from John Wiley and Sons, © 2016 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim.

varying number of linker $-\text{CH}_2-$ units placed between the Sn-based units, yielding the repeat unit $-\text{[Sn}(\text{CH}_3)_2(\text{COO})_2-(\text{CH}_2)_n\text{]}-$, where n changes from 0 to 11. Dielectric constants and band gaps were measured for all these polymers; DFT computations on these systems (this data is part of the organometallic polymers plotted in Figure 9.6) revealed three kinds of low energy crystal structural motifs, and properties were computed for each motif of each polymer. The computed and experimentally measured properties of the entire series of organo-Sn polyesters are shown in Figure 9.8.

To ensure the band gaps are sufficiently large (which could lead to a high breakdown field), both the PBE and HSE band gaps^{66,67} are presented, and seen to be around or above 4 eV and 6 eV respectively for all the polymers. On the other hand, the dielectric constant displayed a general decrease with increasing number of linker $-\text{CH}_2-$ units in the polymer; however, high dielectric constants of >6 were observed for systems with an intermediate number of linker $-\text{CH}_2-$ units (five, six or seven).^{77,79} The remarkable combination of high dielectric constant and large band gap put the organo-Sn

polymers a notch above all the organic polymers studied so far. However, issues of solubility and film formability brought them a notch down again. Co-polymerizing the Sn-ester based polymers with one another (as well as with attractive polyimide or polythiourea based units) led to cast films,^{78,85} for which the measured energy densities were roughly the same as BOPP. Regardless, the work on organo-Sn polymers revealed the true promise of the organometallic chemical space, providing motivation for ongoing efforts to further optimize the polymers and obtain next generation capabilities.

Data on all the polymers studied computationally and experimentally (structural information and computed/measured properties) so far was collected in the form of an online “materials knowledgebase” known as *Khazana*,⁸⁶ which will be discussed further in Section 9.6. Archiving of data in this manner is imperative for the following reasons: (a) to facilitate easy storage and retrieval of information, (b) to avoid duplication of efforts in future studies, (c) to guide any synthetic polymer chemist in the study of dielectrics, and (d) to *learn* from the data and unearth important factors contributing to the properties. The prospects of learning from the data to further accelerate polymer dielectrics design are explored in the following section.

9.4 Learning From Computational Data

First principles computations undoubtedly accelerate the materials design process, but are quite time intensive and could benefit from statistical learning approaches. The substantial computational dataset of polymers created in this work can be mined to *learn* about how the important physical and chemical attributes of a polymer contribute to its properties, and thus make qualitative or quantitative forecasts on the behaviour of newer polymers. This section explains the utility of machine learning—the ability of a computer to learn rules from data—in opening avenues that lead to hitherto unexplored areas of the chemical space without resorting to repeated high-throughput DFT.

The field of materials science that deals with using machine learning (ML) to accelerate materials design is often referred to as *materials informatics*.^{62,87–92} In recent years, informatics approaches have been used for the prediction and classification of crystal structure types,^{93–95} stability of phases,^{96,97} band gaps,^{89,98,99} elastic moduli,¹⁰⁰ dielectric breakdown^{101,102} and instant atomic forces.^{103–105} The most crucial aspect of materials informatics is *fingerprinting*, or the numerical representation of a material in terms of its most important attributes.^{95,106,107} For instance, if one were to fingerprint the polymers belonging to the chemical subspace shown in Figure 9.4 using their band gap values, one could qualitatively predict a new polymer’s dielectric constant based on the magnitude of its band gap. However, the purpose of fingerprinting materials is to have easily attainable, general and unique vectors that can be mapped to the properties of interest.^{87,95} Materials scientists have used elemental properties such as electronegativity and ionization energy,^{89,101,108} oxidation states,¹⁰⁹ HOMO–LUMO

levels,^{89,110} shape and structural parameters,^{27,111} chemical composition,^{108,112} radial distribution functions^{104,105,113} and Coulomb matrices^{114,115} for fingerprinting materials. What kind of a fingerprint would be ideal for the polymers studied here?

9.4.1 Polymer Fingerprinting

With the idea that the electronic and dielectric properties of a polymer would be dictated by “group contributions” from its basic building elements,¹¹⁶ fingerprinting was performed by quantifying the chemical composition in terms of the basic chemical units (CH₂, C₆H₄ *etc.*, fingerprint type I), and in terms of the basic atomic units (4-fold C atoms, 2-fold O atoms *etc.*, fingerprint type II). Within each type, three fingerprints were defined in a hierarchical manner, as pictorially depicted in Figure 9.9: (a) the singles, counting the number of times each unit appeared in the polymer, (b) the doubles, counting the number of times each pair of units appeared in the polymer, and (c) the triples, counting the number of times each triplet of units appeared in the polymer.^{62,87} While these fingerprints are easy to obtain, and can be generalized for all *n*-block polymers (by normalization of the counts with respect to *n*), they are not necessarily always unique (for instance, the singles are not unique for 4-block polymers and the doubles are not unique for 8-block polymers) and may not possess sufficient information. However, the simplicity of the singles and doubles presented the opportunity to make qualitative assessments of how any given block (atom type) or pair of blocks (atom types) affects the properties.

9.4.2 ML Models Trained using DFT Data

ML techniques were applied to the polymer data within the frameworks of fingerprint types I and II independently.^{62,87} As an example, the results obtained (presented in Figure 9.10) when applying fingerprint type I on the dataset of organic polymers described in Section 9.2.1 are explained here. A linear correlation analysis was performed between the components of singles and doubles respectively and four properties: band gap, electronic dielectric constant, ionic dielectric constant and total dielectric constant. Pearson correlation coefficients for each property with each component of the fingerprints shown in Figure 9.10(a) revealed that while CH₂ and O blocks, and CH₂-CH₂ and CH₂-O pairs lead to the highest band gaps, C₄H₂S and CS blocks and their pairs with each other decrease the band gap the most. The effects on the dielectric constant followed quite the opposite trend, thanks to the inverse relationship between the electronic dielectric constant and the band gap. The ionic dielectric constant, meanwhile, is positively contributed to by NH and CO groups, and NH-CO pairs. Thus, the influence of specific blocks and block pairs on the polymer properties was identified, and a similar analysis using fingerprint type II would reveal the atom types and pairs of atom types that are influential.

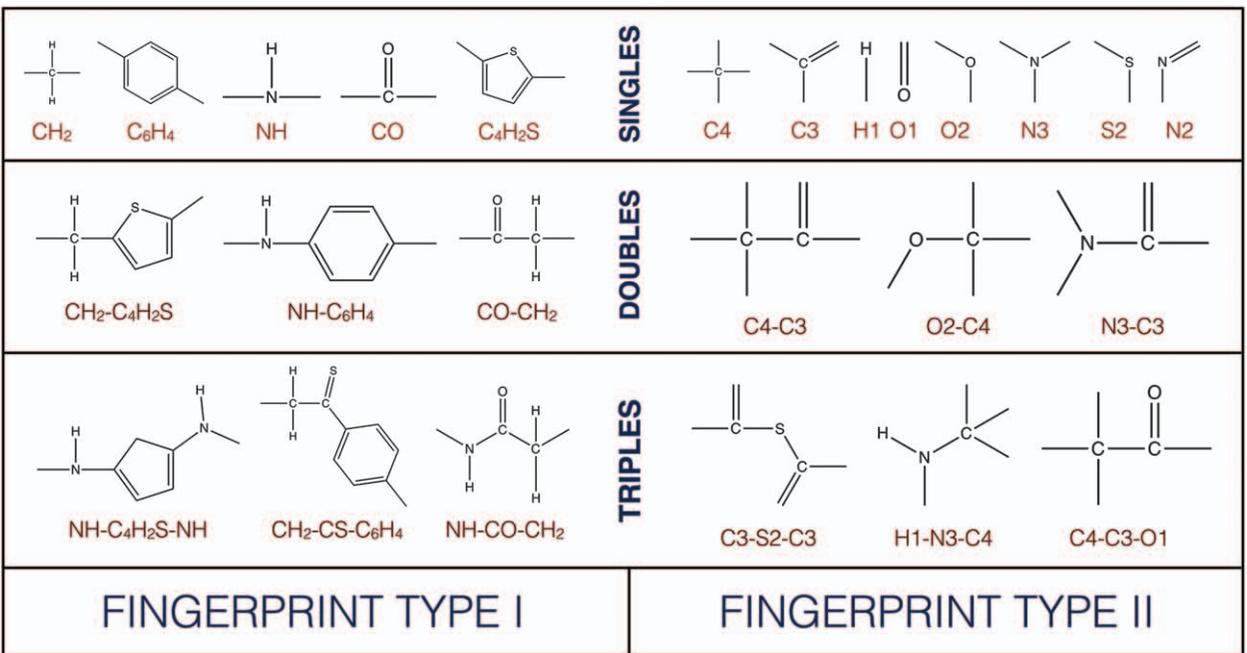


Figure 9.9 Examples of the basic building blocks, building block pairs and building block triplets that help define fingerprint types I (where chemical units such as CH₂ and C₆H₄ are building blocks) and II (where atoms like 4-fold C (C₄) and 2-fold O (O₂) are building blocks).

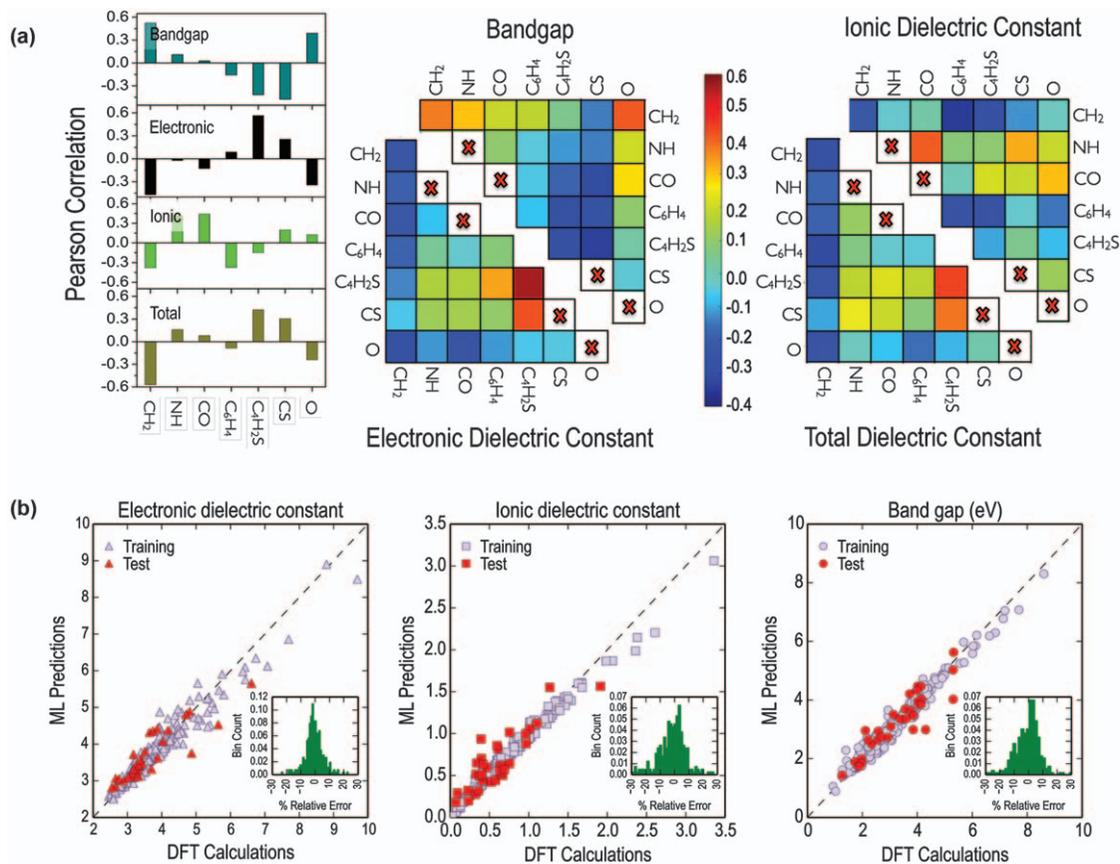


Figure 9.10 Results of machine learning applied on the DFT data. (a) Correlations between different components of singles and doubles of fingerprint type I with the different properties, and (b) performances of property prediction models trained using kernel ridge regression.

Next, a regression algorithm was used to train a model that converts a fingerprint input to its property, within a statistical accuracy. The benefit of having such a prediction model as opposed to mere correlations is being able to make a quantitative prediction of the dielectric constant and band gap of any new polymer, and consequently enhance the initial computational dataset to include hundreds and thousands of new polymers. Kernel ridge regression (KRR¹¹⁷) was applied to the dataset and predictive models were obtained whose performances are shown in Figure 9.10(b). The triples fingerprint, given its uniqueness for 4- to 8-block polymers and the degree of information it contains, was used for this purpose. KRR is a popular non-linear regression technique where the data points are transformed from the fingerprint space to a kernel space, and the property of interest is defined as a function of the similarity (defined in terms of the Euclidean distance) of a given fingerprint with every other fingerprint in the dataset. The necessary parameters for this functional form are obtained by training the KRR model on a subset of the dataset known as the *training set*, while testing of the model for generality and performance evaluation is done on the *test set*. Here, the training set comprised ~90% of the total data points, upon which 5-fold *cross-validation* was performed to ensure there is no overfitting in the predictions. The latter involves dividing the training data into five further subsets, making predictions for the points in every subset by training models on the remaining four, and obtaining the final optimal parameters as an average over the five cases. More details of the KRR formalism can be obtained from ref. 117 and 118, while details of applying KRR to a polymer dataset can be obtained from ref. 88 and 99.

9.4.3 Validation and Utility of ML Framework

With statistically satisfactory relative errors of less than 10% seen between the ML predictions and the DFT results as shown in Figure 9.10(b), the stage was set for the prediction models to be tested on newer, longer-chain polymers. Nearly 40 random polymers containing six to ten blocks in their repeat units were selected for this purpose; some of them had been studied experimentally, while crystal structure prediction and DFT computation of properties was performed for each polymer. Impressive agreement was seen between the ML, DFT and experimental results as shown in Figure 9.11; the qualitative trends were generally captured, even if a 100% quantitative match did not always occur. It should be noted that having been trained on purely 4-block polymers, the ML models make surprisingly good predictions even for higher-block polymers.⁶²

With “on-demand” predictions now possible for any *n*-block polymer belonging to the chemical space shown in Figure 9.4, the ML models were collected in the form of user-friendly design and prediction tools in the online materials knowledgebase *Khazana*,⁸⁶ as discussed in Section 9.6. However, one needs to be guarded against the limitations of the ML approach. The trained models shown and tested in Figures 9.10 and 9.11 are

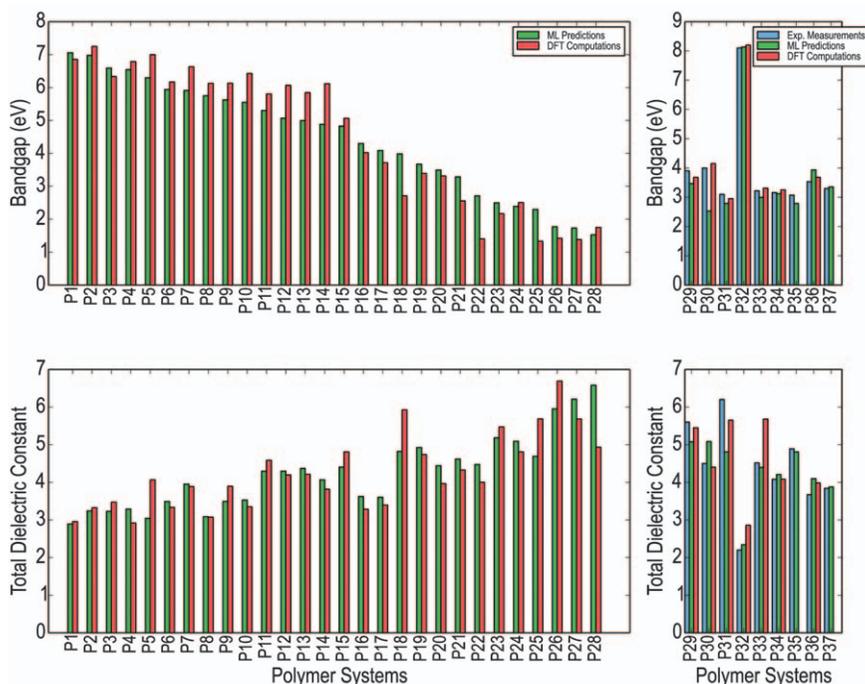


Figure 9.11 Validation of machine learning predictions for nearly 40 polymers with arbitrarily long chain lengths, against their DFT computed and experimentally measured properties.

valid only for polymers containing the same seven basic chemical units. Fingerprint type II provides a more general ML framework by taking atom types, chemical bonds and chemical conjugation into account: similar ML models when trained with this fingerprint (the subject of our work in ref. 87) will be applicable to all polymers containing C, H, O, N or S atoms in their standard coordination environments. Furthermore, it should be noted that both fingerprint types I and II quantify the chemical build-up of the polymer in terms of chemical units or atoms and their respective neighbours, but lack information regarding their crystal structures or conformations. For two polymers with the same repeat unit but different structural arrangements, the fingerprints, and therefore the predicted properties, would be the same, although structural differences could have very realistic consequences.

9.5 Exploring the Polymer Genome

The importance of *data* in driving discovery and innovation puts the onus on scientists to catalogue their computational and experimental results, and whatever insights they may have gained from them, for the benefit of the entire scientific community. This aligns well with the goals and objectives of the *Materials Genome Initiative*,²³ and efforts towards the same are evidenced

by the rise of many materials databases over the last few years.^{119–123} All the polymer data (including computationally obtained ground state structures, and the DFT estimated and experimentally measured properties) and machine learning models presented in this chapter may be found within the “Polymer Genome” search feature within the *Khazana* platform.⁸⁶ Any user searching for a polymer by its repeat unit, chemical name or desired properties will be able to access the relevant experimental or computational data, as well as ML predicted properties, and can utilize this information to make an instant go/no-go decision on whether to pursue it for applications of interest. Fingerprinting a polymer in terms of its basic building block (the chemical unit or atom) is like tracking the polymer “genetic material” or “gene”, which is then utilized for explaining trends in the properties; hence the terminology “the polymer genome”. This knowledgebase is an attempt to unravel the polymer genome, and through the medium of past data and machine learning tools, provide ready access to meaningful spaces of the polymer chemical universe to the community.

AQ-1

An important limitation of the ML approach is that the trained models are always only as good as the training data used. For instance, using the models presented in Section 9.4.2, which were trained on the polymer chemical space shown in Figure 9.4, predictions on polymers containing side chains, fresh chemical units or newer coordination environments may not stand up to a stricter quantum mechanical test. Thus, there would be a requirement of constant data infusion and model retraining to obtain systematic and progressive improvement. Indeed, an *adaptive* learning approach is imperative, wherein fresh computational data on systems sufficiently distinct from previous data (*i.e.*, polymers containing new chemical units and environments) would be added when available and the ML models would be retrained to make fresh, more accurate predictions on newer regions of the chemical space. This process could be repeated in an iterative manner as follows: “ML model → predictions → fresh computations → retrained ML model”, thus establishing a strategy of slowly but surely pushing the boundaries of the polymer chemical space and progressively expanding the predictive regions *via* an adaptive learning framework.

9.6 Conclusions and Outlook

In this chapter, the importance of computation-guided and data-driven strategies for the rational design of materials was highlighted with the example of advanced polymer dielectrics for energy storage capacitor applications. A design strategy involving high-throughput DFT, guided experiments and ML based insights was executed here, culminating in the successful discovery of several novel organic and organometallic polymer dielectric candidates. DFT was used to compute two crucial properties—dielectric constant and band gap—for a few hundred organic polymers, followed by several organometallic polymers. After a first stage of screening yielded promising candidates that were synthesized and tested to provide

validation for the DFT computations, subsequent generations of polymers were experimentally studied to overcome the issues posed by the initial polymers. Further, ML methods were applied on the DFT data to obtain design rules based on correlation analysis and regression-based prediction models, which facilitate quick and easy estimation of the properties of new polymers. All the computational and experimental data generated as part of this work, along with the ML models, were collected in the form of an “online materials knowledgebase”. Such data repositories and design tools are critical to the future of materials design, providing ready guidance to future experiments and computations, consequently leading to faster, more efficient design and discovery.

The synergistic use of computations and experiments in a *rational co-design* formulation enabled the design of new polymer dielectrics much faster than implementing standalone experiments. Any computations are incomplete without accompanying experiments, which provide validation as well as realization of modelled materials; experiments, on the other hand, suffer from a lack of direction without computational insights. A marriage between the two is truly a recipe for success in the modern materials research environment. Further, the ability to learn from uniform, curated (experimental or computational) data, and apply this learning to new materials, is truly transformative in terms of accelerating materials design. This is the rapidly progressing field of *materials informatics*, which deals with developing phenomenological theories, design rules and predictive models based on learning from data, as well as logically determining next computation or experiments that should be performed to improve the models and expand the pool of promising materials. Regular improvements in computing power and the increasing use of machine learning based approaches presents endless possibilities in materials research in the coming years.

Acknowledgements

The authors would like to thank their colleagues Dr Chiho Kim and Dr Huan Tran for partaking in critical discussions and for assistance with figures. The experimental work of the respective research groups of Prof. Gregory Sotzing and Prof. Yang Cao at the Institute of Materials Science in University of Connecticut, specifically related to polymer synthesis and property measurements, is also acknowledged. Finally, the authors acknowledge financial support for their high-energy density capacitor materials research from the Office of Naval Research (ONR), most recently through a multidisciplinary university research initiative (MURI) grant.

References

1. M. F. Ashby and D. R. H. Jones, *Engineering Materials 2*, Oxford, Pergamon Press, 1992.
2. E. Starke and J. Staley, *Prog. Aeronaut. Sci.*, 1996, **32**(2), 131–172.

3. W. J. Buehler, J. W. Gilfrich and R. C. Wiley, *J. Appl. Phys.*, 1963, **34**(5), 1475–1477. 1
4. W. Hume-Rothery, *Atomic Theory for Students of Metallurgy, The Institute of Metals*, London, 1969.
5. E. Hall, *Proc. Phys. Soc.*, 1951, **64**, 747–753. 5
6. R. Dugas, *A History of Mechanics*, Dover Publications, 1988.
7. F. J. Ragep, *Science in Context*, Cambridge University Press, 2001, **14**, 145–163.
8. J. Mehra and H. Rechenberg, *The Historical Development of Quantum Theory*, New York, Springer-Verlag, 1982, p. 4. 10
9. R. Feynman, R. Leighton and M. Sands, *The Feynman Lectures on Physics*, Addison-Wesley Pub. Co., Reading, MA, 1964, p. 3.
10. T. Dauxois, M. Peyrard and S. Ruffo, *Eur. J. Phys.*, 2005, **26**, S3.
11. B. J. Alder and T. E. Wainwright, *J. Chem. Phys.*, 1959, **31**, 2.
12. A. Rahman, *Phys. Rev.*, 1964, **136**(2A), A405–A411. 15
13. H. P. Komsa and K. W. Ja-cobsen, *Phys. Rev.*, 1964, **136**(3B), 864–871.
14. K. W. Zhang and S. L. Chint, *Phys. Rev.*, 1965, **140**(4A), 1133–1138.
15. G. Ceder and K. Persson, *Sci. Am.*, 2013, **309**, 36–40.
16. K. S. Ágnes Nagy, *Computation*, 2016, **4**, 45.
17. T. H. Jorg Neugebauer, *Comput. Mol. Biosci.*, 2013, **3**, 438–448. 20
18. K. Lejaeghere, V. V. Speybroeck, G. V. Oost and S. Cottenier, *Critical Reviews in Solid State and Materials Sciences*, Taylor & Francis, 2014, **39**, pp. 1–24.
19. A. D. Becke, *J. Chem. Phys.*, 2014, **140**, 18A301.
20. H. S. Yu, S. L. Li and D. G. Truhlar, *J. Chem. Phys.*, 2016, **145**, 130901. 25
21. L. M. Ghiringhelli, in *Many-Electron Approaches in Physics, Chemistry and Mathematics: A Multidisciplinary View*, 2014, 191–206.
22. J. Hafner, C. Wolverton and G. Ceder, *MRS Bull.*, 2006, **31**(9), 659–668.
23. “Materials Genome Initiative (MGI),” 2011. [Online]. Available: <https://www.whitehouse.gov/mgi>. 30
24. G. Ceder, Y. M. Chiang, D. R. Sadoway, M. K. Aydinol, Y. I. Jang and B. Huang, *Nature*, 1998, **392**, 694–696.
25. D. Xue, P. V. Balachandran, J. Hogden, J. Theiler, D. Xue and T. Lookman, *Nat. Commun.*, 2016, **7**, 11241.
26. T. Lookman, P. V. Balachandran, D. Xue, J. Hogden and J. Theiler, *Curr. Opin. Solid State Mater. Sci.*, 2016. 35
27. R. Gautier, X. Zhang, L. Hu, L. Yu, Y. Lin, T. O. L. Sunde, D. Chon, K. R. Poeppelmeier and A. Zunger, *Nat. Chem.*, 2015, **7**, 308–316.
28. A. R. Leach and V. J. Gillet, *An Introduction To Chemoinformatics*, Springer, 2007. 40
29. A. M. Lesk, *Introduction to Bioinformatics*, Oxford University Press, 2008.
30. A. Agrawal and A. Choudhary, *APL Mater.*, 2016, **4**, 5.
31. K. Rajan, *Annu. Rev. Mater. Res.*, 2015, **45**, 153–169.
32. H. Chial, *Nat. Educ.*, 2008, **1**, 219.
33. *Handbook of Low and High Dielectric Constant Materials and Their Applications*, ed. H. S. Nalwa, Academic Press, 1999. 45

34. J. Ennis, F. MacDougall, X. Yang, R. Cooper, K. Seal, C. Naruo, B. Spinks, P. Kroessler and J. Bates, *16th IEEE International Pulsed Power Conference*, Albuquerque, NM, USA, 2007. 1
35. F. E. J. MacDougall, X. H. Yang, R. Cooper, J. Gilbert, J. Bates, C. Naruo, M. Schneider, N. Keller, S. Joshi, T. Jow, J. Ho, C. Scozzie and S. Yen, *IEEE Pulsed Power Conference*, Washington, DC, 2009. 5
36. H. Bluhm, *Pulsed Power Systems: Principles and Applications*, Springer, 2006.
37. J. Ho, R. Ramprasad and S. Boggs, *IEEE Trans. Dielectr. Electr. Insul.*, 2007, **14**, 5. 10
38. G. P. M. Rabuffi, *IEEE Trans. Plasma Sci.*, 2002, **30**, 1939.
39. E. J. Barshaw, J. White, M. J. Chait, J. B. Cornette, J. Bustamante, F. Folli, D. Biltchick, G. Borelli, G. Picci and M. Rabuffi, *IEEE Trans. Magn.*, 2007, **43**(1), 223–225.
40. J. Tortai, N. Bonifaci and A. Denat, *J. Appl. Phys.*, 2005, **97**, 053304. 15
41. J. Ho and T. R. Jow, *Army Research Laboratory*, Adelphi, MD, 2009.
42. L. Yang, J. Ho, E. Allahyarov, R. Mu and L. Zhu, *ACS Appl. Mater. Interfaces*, 2015, **7**, 19894–19905.
43. Q. M. Zhang, V. Bharti and X. Zhao, *Science*, 1998, **280**(5372), 2101–2104.
44. S. Zhang, C. Zou, D. Kushner, X. Zhou, R. J. Orchard, N. Zhang and Q. Zhang, *IEEE Trans. Dielectr. Electr. Insul.*, 2012, **19**, 1158–1166. 20
45. W. Li, L. Jiang, X. Zhang, Y. Shen and C. W. Nan, *J. Mater. Chem. A*, 2014, **2**, 15803–15807.
46. L. Jiang, W. Li, J. Zhu, X. Huo, L. Luo and Y. Zhu, *Appl. Phys. Lett.*, 2015, **106**, 052901. 25
47. Y. Wang, X. Zhou, M. Lin and Q. Zhang, *Appl. Phys. Lett.*, 2009, **94**, 20.
48. Y. Wang, X. Zhou, Q. Chen, B. Chu and Q. Zhang, *IEEE Trans. Dielectr. Electr. Insul.*, 2010, **17**(4), 1036–1042.
49. S. Wu, W. Li, M. Lin, Q. Burlingame and Q. Chen, *Adv. Mater.*, 2013, **25**(12), 1734–1738. 30
50. Q. Burlingame, S. Wu, M. Lin and Q. Zhang, *Adv. Energy Mater.*, 2013, **3**(8), 1051–1055.
51. J. T. Bendler, C. A. Edmondson, M. C. Wintersgill, D. A. Boyles, T. S. Filipova and J. J. Fontanella, *Eur. Polym. J.*, 2012, **48**, 830–840.
52. J. T. Bendler, D. A. Boyles, C. A. Edmondson, T. Filipova, J. J. Fontanella, M. A. Westgate and M. C. Wintersgill, *Macromolecules*, 2013, **46**, 4024–4033. 35
53. J.-K. Tseng, S. Tang, Z. Zhou, M. Mackey, J. M. Carr, R. Mu, L. Flandin, D. E. Schuele, E. Baer and L. Zhu, *Polymer*, 2014, **55**, 8–14.
54. L. Zhu, *J. Phys. Chem. Lett.*, 2014, **5**, 3677–3687. 40
55. J. M. Carr, M. Mackey, L. Flandin, D. Schuele, L. Zhu and E. Baer, *J. Polym. Sci., Part B: Polym. Chem.*, 2013, **51**, 882–896.
56. P. Kim, N. Doss, J. Tillotson, P. Hotchkiss and S. Marder, *ACS Nano*, 2009, **3**(9), 2581–2592.
57. P. Kim, S. Jones, P. Hotchkiss, J. Haddock and B. Kippelen, *Adv. Mater.*, 2007, **19**(7), 1001–1005. 45

58. A. Mannodi-Kanakkithodi, G. M. Treich, T. D. Huan, R. Ma, M. Tefferi, Y. Cao, G. A. Sotzing and R. Ramprasad, *Adv. Mater.*, 2016, **28**, 30. 1
59. V. Sharma, C. C. Wang, R. G. Lorenzini, R. Ma, Q. Zhu, D. W. Sinkovits, G. Pilania, A. R. Oganov, S. Kumar, G. A. Sotzing, S. A. Boggs and R. Ramprasad, *Nat. Commun.*, 2014, **5**, 4845. 5
60. T. D. Huan, A. Mannodi-Kanakkithodi, C. Kim, V. Sharma, G. Pilania and R. Ramprasad, *Sci. Data.*, 2016, **3**, 160012.
61. C. C. Wang, G. Pilanina, S. Boggs, S. Kumar, C. Breneman and R. Ramprasad, *Polymer*, 2014, **55**(4), 979.
62. A. Mannodi-Kanakkithodi, G. Pilania, T. D. Huan, T. Lookman and R. Ramprasad, *Sci. Rep.*, 2016, **6**, 20952. 10
63. S. Baroni, S. de Gironcoli, A. Dal Corso and P. Giannozzi, *Rev. Mod. Phys.*, 2001, **73**(2), 515.
64. X. Gonze and C. Lee, *Phys. Rev. B*, 1997, **55**, 10355–10368.
65. X. Gonze, *Phys. Rev. B*, 1997, **55**, 10337–10354. 15
66. J. Heyd, G. E. Scuseria and M. Ernzerhof, *J. Chem. Phys.*, 2003, **118**(18), 8207.
67. J. P. Perdew, *Int. J. Quant. Chem.*, 1985, **28**, 497–523.
68. T. Huan, S. Boggs, G. Teyssedre, C. Laurent, M. Cakmak, S. Kumar and R. Ramprasad, *Prog. Matter. Sci.*, 2016, **83**, 236–269. 20
69. Q. Zhu, A. R. Oganov, C. W. Glass and H. T. Stokes, *Acta Cryst.*, 2012, **B68**, 215–226.
70. Q. Zhu, V. Sharma, A. R. Oganov and R. Ramprasad, *J. Chem. Phys.*, 2014, **141**(15), 154102.
71. S. Goedecker, *J. Chem. Phys.*, 2004, **120**(21), 9911. 25
72. M. Amsler and S. Goedecker, *J. Chem. Phys.*, 2010, **133**, 224104.
73. A. Mannodi-Kanakkithodi, C. C. Wang and R. Ramprasad, *J. Mater. Sci.*, 2015, **50**(2), 801.
74. C. Wang, G. Pilania and R. Ramprasad, *Phys. Rev. B*, 2013, **87**, 035103.
75. H. Furukawa, K. E. Cordova, M. O’Keeffe and O. M. Yaghi, *Science*, 2013, **341**, 6149. 30
76. G. Pilania, C. C. Wang, K. Wu, N. Sukumar, C. Breneman, G. Sotzing and R. Ramprasad, *J. Chem. Inf. Model.*, 2013, **53**(4), 879–886.
77. A. F. Baldwin, T. D. Huan, R. Ma, A. Mannodi-Kanakkithodi, M. Tefferi, N. Katz, Y. Cao, R. Ramprasad and G. A. Sotzing, *Macromolecules*, 2015, **48**, 2422–2428. 35
78. A. F. Baldwin, R. Ma, T. D. Huan, Y. Cao, R. Ramprasad and G. A. Sotzing, *Macromol. Rapid Commun.*, 2014, **35**(24), 2082.
79. A. F. Baldwin, R. Ma, A. Mannodi-Kanakkithodi, T. D. Huan, C. C. Wang, M. Tefferi, J. E. Marszalek, M. Cakmak, Y. Cao, R. Ramprasad and G. A. Sotzing, *Adv. Mater.*, 2015, **27**, 346. 40
80. A. Mannodi-Kanakkithodi, T. Huan and R. Ramprasad, Mining materials design rules from data: The example of polymer dielectrics, *Submitted*.
81. A. F. Baldwin, R. Ma, C. Wang, R. Ramprasad and G. Sotzing, *J. Appl. Polym. Sci.*, 2013, **130**, 1276–1280. 45

82. R. Lorenzini, W. Kline, C. Wang, R. Ramprasad and G. Sotzing, *Polymer*, 2013, **54**(14), 3529. 1
83. R. Ma, A. F. Baldwin, C. C. Wang, I. Offenbach, M. Cakmak, R. Ramprasad and G. A. Sotzing, *ACS Appl. Mater. Interfaces*, 2014, **6**, 13, 10445. 5
84. R. Ma, V. Sharma, A. F. Baldwin, M. Tefferi, I. Offenbach, M. Cakmak, R. Weiss, Y. Cao, R. Ramprasad and G. A. Sotzing, *J. Mater. Chem. A*, 2015, **3**, 14845.
85. G. Treich, S. Nasreen, A. Mannodi-Kanakkithodi, R. Ma, M. Tefferi, J. Flynn, Y. Cao, R. Ramprasad and G. Sotzing, *Appl. Mater. Interfaces*, 2016, **8**(33), 21270–21277. 10
86. [Online]. Available: <http://khazana.uconn.edu/>.
87. T. D. Huan, A. Mannodi-Kanakkithodi and R. Ramprasad, *Phys. Rev. B*, 2015, **92**, 014106.
88. G. Pilania, C. C. Wang, X. Jiang, S. Rajasekaran and R. Ramprasad, *Sci. Rep.*, 2013, **3**, 2810. 15
89. G. Pilania, A. Mannodi-Kanakkithodi, B. Uberuaga, R. Ramprasad, J. Gubernatis and T. Lookman, *Sci. Rep.*, 2016, **6**, 19375.
90. T. Mueller, A. G. Kusne and R. Ramprasad, *Reviews Computational Chemistry*, John Wiley & Sons, 2016, p. 29. 20
91. K. Schütt, H. Glawe, F. Brockherde, A. Sanna, K. Müller and E. Gross, *Phys. Rev. B.*, 2014, **89**, 205118.
92. F. Faber, A. Lindmaa, O. V. Lilienfeld and R. Armiento, *Int. J. Quant. Chem.*, 2015, **115**(16), 1094–1101.
93. G. Pilania, J. Gubernatis and T. Lookman, *Sci. Rep.*, 2015, **5**, 17504. 25
94. G. Pilania, J. Gubernatis and T. Lookman, *Phys. Rev. B.*, 2015, **91**, 214302.
95. L. M. Ghiringhelli, J. Vybiral, S. V. Levchenko, C. Draxl and M. Scheffler, *Phys. Rev. Lett.*, 2015, **114**, 105503.
96. J. Hattrick-Simpers, J. Gregoire and G. Kusne, *APL Mater.*, 2016, **4**, 053211. 30
97. A. Kusne, D. Keller, A. Anderson, A. Zaban and I. Takeuchi, *Nanotechnology*, 2015, **26**, 444002.
98. A. Mannodi-Kanakkithodi, G. Pilania, R. Ramprasad, T. Lookman and J. Gubernatis, *Comput. Mater. Sci.*, 2016, **125**, 92. 35
99. A. Mannodi-Kanakkithodi, G. Pilania and R. Ramprasad, *Comput. Mater. Sci.*, 2016, **125**, 123.
100. M. De Jong, W. Chen, R. Notestine, K. Persson, G. Ceder, A. Jain, M. Asta and A. Gamst, *Sci. Rep.*, 2016, **6**, 34256.
101. C. Kim, G. Pilania and R. Ramprasad, *Chem. Mater.*, 2016, **28**, 1304. 40
102. C. Kim, G. Pilania and R. Ramprasad, *J. Phys. Chem. C*, 2016, **120**, 14575.
103. Z. Li, J. Kermode and A. De Vita, *Phys. Rev. Lett.*, 2015, **114**, 096405.
104. V. Botu and R. Ramprasad, *Int. J. Quant. Chem.*, 2015, **115**(16), 1074–1083. 45
105. V. Botu and R. Ramprasad, *Phys. Rev. B.*, 2015, **92**, 094306.

106. M. Rupp, A. Tkatchenko, K. Muller and O. A. von Lilienfeld, *Phys. Rev. Lett.*, 2012, **108**, 058301. 1
107. B. Meredig, A. Agrawal, S. Kirklin, J. E. Saal, J. W. Doak, A. Thompson, K. Zhang, A. Choudhary and C. Wolverton, *Phys. Rev. B*, 2014, **89**, 094104. 5
108. A. Seko, T. Maekawa, K. Tsuda and I. Tanaka, *Phys. Rev. B.*, 2014, **89**, 054303.
109. V. Sharma, G. Pilania, G. Rossetti, K. Slenes and R. Ramprasad, *Phys. Rev. B*, 2013, **87**, 134109.
110. E. I. L. Roman and M. Balabin, *J. Chem. Phys.*, 2009, **131**, 074104. 10
111. C. L. Phillips and G. A. Vothab, *Soft Matter*, 2013, **9**, 8552.
112. G. Hautier, C. C. Fischer, A. Jain, T. Mueller and G. Ceder, *Chem. Mater.*, 2010, **22**, 3762–3767.
113. O. A. v. Lilienfeld, R. Ramakrishnan, M. Rupp and A. Knoll, *Int. J. Quant. Chem*, 2015, **115**, 1084–1093. 15
114. M. Rupp, M. Bauer, R. Wilcken, A. Lange, M. Reutlinger, F. Boeckler and G. Schneider, *PLOS Comp. Biol.*, 2014, **10**, 1.
115. G. Montavon, M. Rupp, V. Gobre, A. Vazquez-Mayagoitia, K. Hansen, A. Tkatchenko, K.-R. Muller and A. V. Lilienfeld, *New J. Phys.*, 2013, **15**, 095003. 20
116. R. L. Miller, *Crystallographic Data and Melting Points for Various Polymers*, John Wiley and Sons Inc., 2003.
117. K. Vu, J. Snyder, L. Li, M. Rupp, B. Chen, T. Khelif, K. Muller and K. Burke, *Int. J. Quant. Chem.*, 2015, **115**(16), 1115–1128.
118. B. Schölkopf, in *Advances in Neural Information Processing Systems 13*, ed. T. K. L. A. T. G. D. A. V. Tresp, MIT Press, Cambridge, MA, 2001, pp. 301–307. 25
119. S. Curtarolo, W. Setyawan, S. Wanga, J. Xue, K. Yang, R. Taylor, L. Nelson, G. Hart, S. Sanvito, M. Nardelli, N. Mingo and O. Levy, *Comput. Mater. Sci.*, 2012, **58**, 227–235. 30
120. G. Pizzi, A. Cepellotti, R. Sabatini, N. Marzari and B. Kozinsky, *Comput. Mater. Sci.*, 2016, **111**, 218–230.
121. A. Jain, S. Ong, G. Hautier, W. Chen, W. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder and K. Persson, *APL Mater.*, 2013, **1**, 011002. 35
122. T. Bligaard, M. Dulak, J. Greeley, S. Nestorov, J. Hummelshoj, D. Landis, J. Norskov and K. Jacobsen, *Comp. Sci. Eng.*, 2012, **14**, 51–57.
123. J. Saal, S. Kirklin, A. Muratahan, B. Meredig and C. Wolverton, *JOM*, 2013, **65**(11), 1501–1509. 40