Spring 5-19-2022

Design Of Polar Materials Using Materials Informatics and First-Principles Calculations

Jon Okenfuss
Mechanical Engineering

WASHINGTON UNIVERSITY IN ST. LOUIS

McKelvey School of Engineering

Department of Mechanical Engineering and Materials Science

Thesis Examination Committee:
Rohan Mishra, Chair
Elijah Thimsen
Sanghoon Bae

Design of Polar Materials Using Materials Informatics and First-Principles Calculations

By
Jon Okenfuss

A thesis presented to
the McKelvey School of Engineering
of Washington University in St. Louis in
partial fulfillment of the
requirements for the degree
of Master of Science

May 2022
St. Louis, Missouri

Table of Contents

List of Figures

Acknowledgments

I would like to acknowledge the many people who helped me on the way to this ending of my academic career. My thanks goes first to my advisor in this work, Professor Mishra, and to my reviewer Zhaohan Zhang. Both have worked tirelessly to guide me into the world of materials informatics.

I would also like to thank the other members of the MCube group, who have all offered their assistance in understanding the field of materials science.

Many other members of the Washington University community contributed to bringing me to this point, especially professors Guy Genin and Ruth Okamoto.

My family has been a large source of support, with my parents, brother, and wife pushing me through and believing in me unceasingly.

Jon Okenfuss

Washington University in St. Louis

May 2022

ABSTRACT OF THE THESIS

Design of Polar Materials Using Materials Informatics and First-Principles Calculations

by

Jon Okenfuss

Master of Science in Mechanical Engineering

Washington University in St. Louis, 2022

Research Advisor: Professor Rohan Mishra

Polar materials have a dipole moment or polarization. Ferroelectrics are a special class of polar materials wherein the polarization can be switched with an external electric field. Because of their characteristics, ferroelectrics are especially useful in adjustable capacitors, non-volatile memories, and sensors. Materials databases like Materials Project contain a large number of materials, but identifying new metastable polar, and more specifically ferroelectric materials can be time consuming. In this project, we train a machine learning model to distinguish between binary compounds having a wurtzite structure—which have a permanent polarization—and their nonpolar zincblende and rock salt polymorphs. We use this model to predict a large number of ternary materials that we find to have high chances to be most stable as a wurtzite. These compounds serve as a smaller dataset that is tractable for high-throughput DFT calculations to calculate their stability in the wurtzite phase, their polarization, and switching barriers with higher accuracy. In the second chapter, we first attempt to identify factors or descriptors that can screen ferroelectric materials from a materials database. Using these descriptors, we predict new ternary oxides that can be potential ferroelectrics and evaluate these predictions using first-principles density-functional theory (DFT) calculations.

Chapter 1: Introduction

Polar materials lack on inversion symmetry along one or more axes of their crystal structure [1]. The asymmetry results in an imbalance in position of the positively and negatively charged ions in a certain direction, which cause finite electric dipoles or electric polarization. Materials with a polar structure tend to have useful features, like pyroelectricity or ferroelectricity.

Ferroelectric materials have a spontaneous polarization that is retained without the application of any electric field. Furthermore, the direction of their polarization can be reversed using an applied electric field. Thus, their two polarization states can act as a binary memory. Ferroelectrics are therefore used to make non-volatile random access memories (Fe-RAM) for information processing. The change in polarization in ferroelectrics with the applied electric field shows a non-linear behavior. So, ferroelectrics are also used as adjustable capacitors. Ferroelectricity was discovered in Rochelle salt in 1920[2], and since then ferroelectricity has been reported in only a handful of materials. Of these, the most attractive ferroelectric materials, $PbTiO_3$ and lead-zirconium titanate, $Pb(Zr,Ti)O_3$, contain toxic lead. There has been a rise in search of new ferroelectric materials in the recent years [3,4]. Unfortunately, despite falling computing costs there is still a high price to run calculations on all possible materials to check for ferroelectricity, so we will attempt to create a workflow that can accelerate and reduce the cost of this process.

Recent work in this area has tended towards large-scale screening of extant materials or machine-learning-only approaches. The large-scale screening work depends on filtering the materials databases before running calculations on all of the materials contained therein [5]. Such work is able to find previously-overlooked or unnoticed ferroelectric materials from the database, but does not expand into new materials. An approach that dives into new materials [6] with machine learning to focus it can require feedback from involved experiments in order to gain a reasonable accuracy.

A combination of materials informatics and density functional theory calculations helps us to overcome shortcomings of prior work and make predictions for future developments. To identify new ferroelectrics, we combine structural databases, machine learning, symmetry analysis, and data-mined substitutions to focus our simulations.

In the process of finding ferroelectric materials, we found that some wurtzite-structured materials have been seen to show ferroelectricity [7] and that the separation of wurtzites from other structures, such as zinc blende and rock salt, is an outstanding problem. We work to improve the machine learning process on finding separators between wurtzite, zincblende, and rock salts, then compile a list of likely wurtzites. In the process of building a dataset for that classification, calculations were done on alternate phases of materials that were already in structure databases (i.e., a wurtzite would also be calculated as zincblende and rock salt). These found that in the structure database is incomplete in its calculations of materials in these structures, so 123 new, more-stable materials can be added.

Chapter 2 of this thesis gives an overview of the computational materials science and machine learning methods we use to obtain our results. This begins with the first-principles

atomistic modeling density functional theory. This is followed by a brief overview of the machine learning methods tested and their settings. Chapter 3 reports on the prediction of wurtzite structures from atomic features, which includes the training of a machine learning model to make predictions on a new set of materials. Chapter 4 presents our search for ternary ferroelectric materials, walking through the workflow that takes us from the acquisition of materials through machine learning model training and to the identification of target materials for density functional theory calculations.

Chapter 2: Methods

2.1 High-throughput Density Functional Theory Calculations

Density-functional theory (DFT) calculations presented in this thesis were performed using the Vienna Ab initio Software Package [8] (VASP). This section is intended to give an overview of DFT and specify how we used it.

2.1.1 Quantum mechanical equations

DFT is, generally, an attempt to compute the electronic structure and predict materials properties given an atomistic model that describes the interactions between atoms in a system. The location of an atom is defined by the location of both its nucleus and its electrons. As electrons move much more rapidly than the heavier nucleus, the problem is split using the Born-Oppenheimer [9] approximation so that the ground state of electrons can be determined. The Schrödinger equation 2.1 [10]

$$\left[-\frac{h^2}{2m}\sum_{i=1}^{N}\nabla_i^2+\sum_{i=1}^{N}V(r_i)+\sum_{i=1}^{N}\sum_{j<i}U(r_i,r_j)\right]\Psi=E\Psi \qquad (2.1)$$

is able to describe this problem. Here, m is electron mass, the bracketed terms are, respectively, kinetic energy of the electrons, interaction energy between each electron and the external field generated by the atomic nuclei, and the interaction energy between electrons. Ψ is the electronic wave function, and E is the ground state energy of the electrons. As the number of electrons N increases, $\Psi(\mathrm{r}_1, r_2, .., r_N)$ contains $3N$ spatial variable, where r_i is the position of i^{th} electron. An analytical solution become impossible for systems involving more than a few atoms.

2.1.2 Hohenberg-Kohn Theorems

An alternate way is to represent an electronic system using electron charge density $n(r)$, which only have 3 spatial variables, as shown in Eqn. 2.2:

$$n(r) = \sum_{i=1}^{N} |\psi_i(r)|^2 \quad (2.2)$$

where the $\psi_i(r)$ are the wavefunctions of the i^{th} electron.

This is supported by Kohn and Hohenberg's [11] first theorem: "The ground state energy from Schrödinger's equation is a unique functional of the electron density." A functional is a function of a function. Here the ground state energy E_0 is a functional of ground state electron density as $E_0[n_0(r)]$. This functional is defined by Hohenberg and Kohn's second theorem: "The electron density that minimizes the energy of the overall functional is the true electron density corresponding to the full solution of the Schrödinger equation." This gives the simple goal of minimizing the energy according to:

$$E[n] = \sum_{i=1}^{N} \int d^3\boldsymbol{r}\varphi_i^*(\boldsymbol{r})\left(-\frac{\hbar^2}{2m_e}\nabla^2\right)\varphi_i(\boldsymbol{r}) + \frac{e^2}{2}\int d^3\boldsymbol{r}\int d^3\boldsymbol{r}'\frac{n(\boldsymbol{r})n(\boldsymbol{r}')}{|\boldsymbol{r}-\boldsymbol{r}'|} + \int v(r)n(r)d^3r + E_{XC}[n] \quad (2.3)$$

which is a sum of electron kinetic energies, interactions between electron pairs, interactions between electrons and the nuclei, and a correction term for exchange and correlation effects (E_{XC}). [12]

2.1.3 Kohn-Sham Formulation

Kohn and Sham [13] showed how to express the Schrödinger equation as a set of equations which each has only a single electron interacting with the electron density in equation 2.4:

$$\left[-\frac{\hbar^2}{2m}\nabla^2 + V(r) + V_H(r) + V_{XC}(r)\right]\psi_i(r) = \varepsilon_i\psi_i(r) \quad (2.4)$$

The three potentials, V, V_H, and V_{XC}, are respectively the known interaction between an electron and the nuclei; the Hartree potential, which describes the Coulomb repulsion between the considered electron and all others; and the exchange and correlation contributions. In order to self-consistently solve the Kohn-Sham equations, an iterative process is used: First, an initial electron density n(r) is selected. Second, Kohn-Sham equations are used to find the $\psi_i(r)$. Third calculate the electron density defined by the equations in step 2. Repeat from the second step with an updated electron density until reaching convergence. [12]

2.1.4 Exchange-Correlation Functional

The exchange-correlation functional, E_{XC}, must be defined as an approximation for the equations to function. For our work, we use a generalized gradient approximation (GGA) which uses information about the local electron density and the local gradient of that density. Various options for a functional are able to be selected, but the GGA used in our work is widely used so it is easy to compare against other work. [12]

2.1.5 Projector augmented-wave potentials

Another important specification for a DFT calculation is the choice of pseudopotential. The pseudopotential smooths the electron density for core electrons in a manner that matches the important properties of the true ion core while still reducing computational complexity. In our work, we use the projector-augmented wave (PAW) [14] pseudopotentials. The PAW pseudopotentials use a frozen core of electrons with a modified ionic potential accounting for the core [15].

2.1.6 Reciprocal space

DFT calculations generally take place on the primitive cell, the supercell with the minimum number of atoms necessary to define a periodic material. In reciprocal space, that flipped space where larger lattice vectors in real space correspond to shorter lattice vectors in reciprocal space, the primitive cell is called the Brillouin zone. The most important point in reciprocal space (or *k* space) is the point where k=0, the gamma point. This matters because DFT involves evaluating a number of integrals like this in equation 2.5:

$$\bar{g} = \frac{V_{cell}}{(2\pi)^3} \int_{BZ} g(k)\, dk \qquad (2.5)$$

Which integrates over values of k in the Brillouin Zone. Monkhorst and Pack [16] developed the scheme that we use for choosing the correct number of k points in order to optimize the convergence of the DFT equations with computational costs. [15]

2.1.7 DFT Settings

The materials simulated for this thesis were made based on DFT calculations performed using VASP. The calculations used the PAW potentials with the Perdew-Burke-Ernzerhof (PBE) exchange-correlation functional. The GGA functional PBE uses information about the local electron density and local gradient in electron density, which is more accurate than previous functionals and is well-used enough to be standard. [12,17]

DFT calculations were done in two stages. The first stage, geometric relaxation, was done to allow atoms to settle into their final structure. After that, a single static calculation was completed with higher precision to be sure the calculation was complete.

2.2 Machine Learning

The material search space is huge. The computational resources necessary to complete DFT for all material combinations do not presently exist, so an effort must be made to learn from desirable materials to focus future searches onto the most promising possibilities. A rapidly growing field is attempting to use machine learning (ML) to predict material properties before taking the costly steps of DFT or synthesis [18,19]. In this section, I will give a brief overview of the background of the various models considered.

In general, each ML model used will take as input a set of training data then output a trained model. The supervised learning training data can be labeled for use in either a regression or classification, depending on whether a continuous value or clearly delineated class is desired. The output trained model can be used to predict the label of the input data. The primary computational cost of ML is the time spent training a model. After training, prediction using the trained model is quick.

In determining which ML model to use, cross validation and test set scores are compared. The hyperparameters of each model are also tuned.

2.2.1 K Nearest Neighbors

K Nearest Neighbors (KNN)[20] predicts the class or property value of a sample purely based on votes from the *k* nearest neighbors, where *k* is the number of neighbors considered. It is an outlier among ML models: there is no training time, so all computing time is taken by searching the training set for the closest *k* neighbors of each sample being predicted. The neighbors considered have a known target value.

Tuning of KNN can be done with a variety of settings, but the ones considered here are number of neighbors, weighting, and metrics. A range of three to fifteen nearest neighbors is checked, with only odd numbers of neighbors being considered so that voting cannot be tied. Higher numbers of neighbors reduce noise in the classification. We consider weighting the classifier uniformly, where each neighbor contributes equally, and by distance, where closer neighbors have a greater influence than farther ones. We also consider both Euclidean and Manhattan distance metrics, where Euclidean distance is measured in a straight line and Manhattan distance is measured along each dimension separately. Manhattan distance is useful for high dimensional applications [21].

2.2.2 Decision Tree

A decision tree predicts the value of a target by a series of simple decision rules. Starting from a single node, a tree will split repeatedly as it makes decisions until it reaches a leaf, at which point a prediction can be made. [22]

Tuning of decision trees can be done with a variety of settings, but the ones considered here are the split quality measure, max tree depth, number of samples to split an internal node, and number of samples to make a leaf node. The quality of a split can be measured with either the Gini impurity or by information gain [23]. The Gini impurity measures the frequency of mislabeling of an element at a node, ranging from a pure node, where all elements are of one class, to a random node with even chances of any class. Information gain is measured by entropy, which shows the disorder of features.

The max depth of a tree needs to be controlled to prevent overfitting, as if a tree is allowed to expand so that every leaf node is pure it could perfectly fit to training data while making bad predictions on test data. Therefore, a range of tree depths should be considered. For a similar

reason, the minimum number of samples to split an internal node and at each leaf node is tested across a range of values to address overfitting.

2.2.3 Random Forest

A random forest is so named because it is a random collection of decision trees. Taking a collection of decision trees, which tend towards overfitting their training sets with a high variance, can significantly improve generalization. The decision trees considered are grown on subsets of the dataset taken with replacement. After training the forest of decision trees, for any given unseen sample the classification can be done via a majority vote or an averaging of values. [24]

Tuning of random forests can be done with a variety of settings, but the ones considered here are the same split quality measure and tree depth as in decision trees, the number of trees in the forest, and the number of features. The number of trees in the forest is typically on the order of hundreds of trees. The number of features considered when looking for the best split can be set as an integer, but we consider it as a function of the number of features in the dataset, either taking the square root or the log of the number of features.

2.2.4 Support Vector Machine

The support vector machine (SVM) [25] classifies data by creating a high-dimensional hyperplane that attempts to maximize the distance between itself and the vectors to the points that support it. In two dimensions, the hyperplane is one dimension as shown in Figure 1:

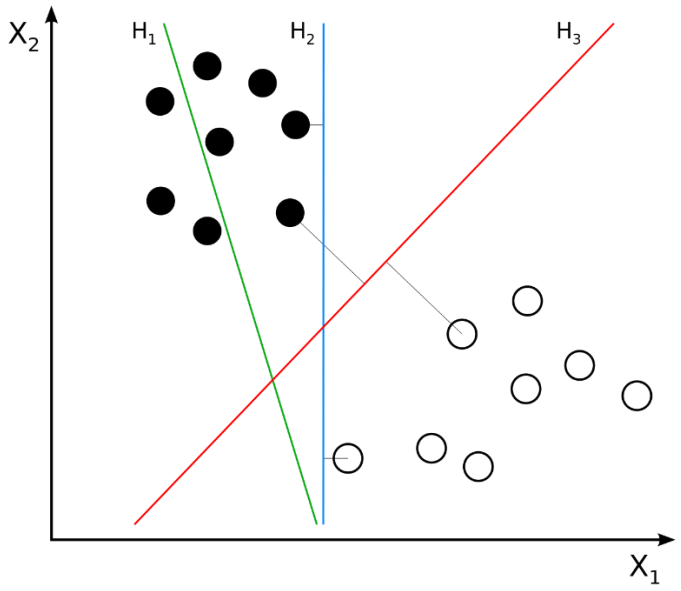

Figure 1 SVM hyperplanes separating two-dimensional data

https://commons.wikimedia.org/w/index.php?curid=22877598

When it moves to higher dimensions, the SVM can redefine variables in terms of a kernel function k (x, y).

Tuning of SVMs can be done with a variety of settings, but the ones considered here are different kernel functions, kernel coefficients, kernel degrees, and regularization parameters. We consider the linear, polynomial, radial basis function, and sigmoid kernels. The linear kernel performs best when data is linearly separable. This is the most basic kernel but runs fast. The polynomial kernel shows the similarity of vectors over polynomials of the original values, allowing for non-linearity. The equation for polynomial kernels is $(\gamma\langle x, x'\rangle + r)^d$, where the degree *d* and coefficient need to be optimized. The radial basis function, $\exp(-\gamma\|x - x'\|^2)$, has a value that depends on the distance between the points considered. The sigmoid kernel, $\tanh(\gamma\langle x, x'\rangle + r)$, comes from neural networks and works similarly to a small two-layer perceptron. [26]

2.2.5 Perceptron

Like an SVM, the perceptron is a linear separator. It functions as a quite simple neural network. Each feature is an input node which has a weight applied. A sum of those weighted features is taken as the output. In training, the weights are varied so that overall error is minimized. [22]

Tuning of perceptrons can be done with a variety of settings, but the ones considered here are the penalty function, penalty constant, and maximum number of iterations. The penalty function can be L1, L2, or a combination of the two, elasticnet. They each penalize mistakes, but L1 penalizes the squared magnitude of mistakes while L2 penalizes the absolute magnitude. That penalty is strongly affected by a regularization constant, so a range of those are checked as well. The maximum number of iterations impacts the cost of training the model, so the minimum number of iterations for a target accuracy is preferred.

2.2.6 Neural Network

A neural network is a system inspired by the brain's connecting neurons. Like the perceptron, features and their weights are the inputs to a neural network. Where this diverges from the perceptron is that instead of a single node taking a sum, there is one or more hidden layers before the output node is reached. In these hidden layers, a nonlinear activation function transforms the value before it proceeds to the next hidden layer or output. [27]

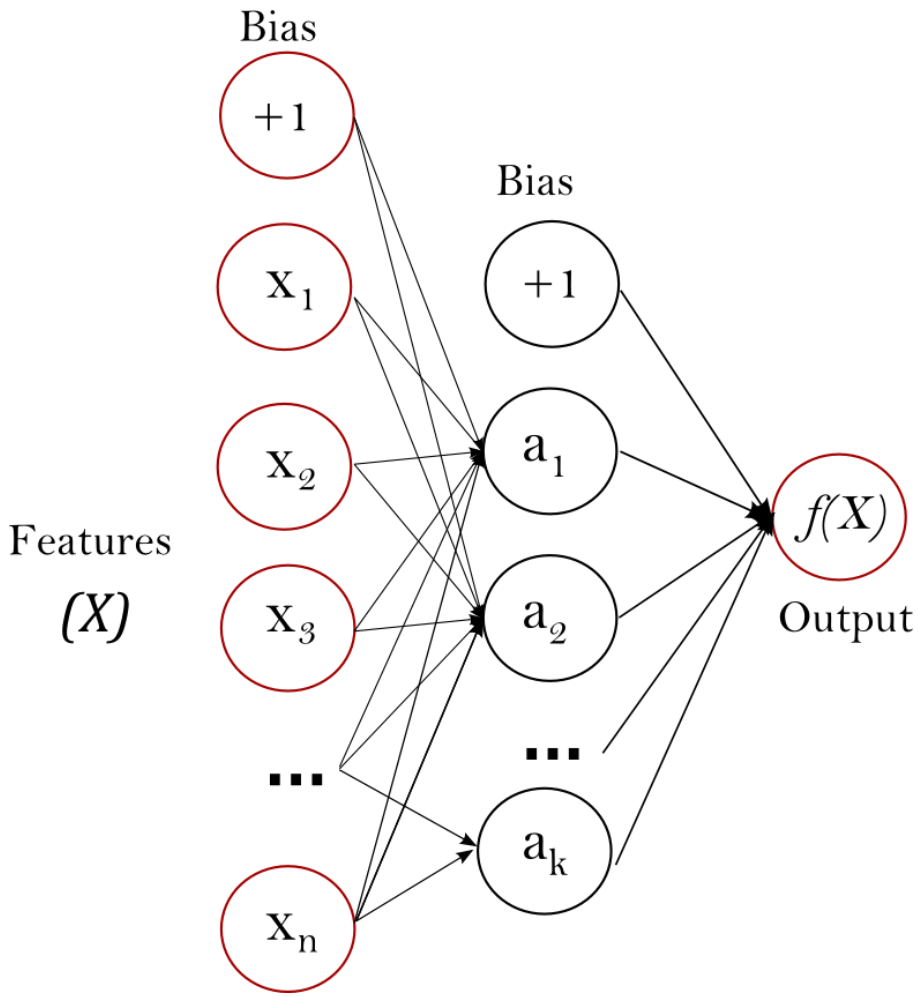

Figure 2 Neural network representation

Tuning of neural networks needs to be done with a variety of settings, but the ones considered here are activation functions, hidden layer size, solvers, regularization, learning rate, and max iterations. The activation functions we consider are the logistic sigmoid function, f(x) = 1 / (1 + exp(-x)), and the rectified linear unit function, f(x) = max (0, x). Hidden layers are where the neural network can find hidden trends in the data, so it makes sense to experiment with varying numbers of nodes in the hidden layer and with various numbers of hidden layers. Various solvers are available, so stochastic gradient descent, adam (a stochastic gradient-based optimizer), and lbfgs (a quasi-Newton optimizer) are compared. Adam tends to work better on

large datasets and lbfgs on small sets. The L2 penalty is used to penalize mistakes, so its coefficient needs to be determined. As training is done on the neural network, the weight is updated either constantly or adaptively (in the adaptive case, the learning rate decreases when two consecutive epochs fail to decrease training loss). The maximum number of iterations impacts the cost of training the model, so the minimum number of iterations for a target accuracy is preferred.

2.2.7 Logistic Regression with Lasso

A logistic regression is a linear classification that models the probability of a binary dependent variable using a logistic function. With the L1 Lasso regularization, the logistic regression solves this optimization: $\min_{w,c} \|w\|_1 + C \sum_{i=1}^{n} \log(\exp(-y_i(X_i^T w + c)) + 1)$. [28] Tuning of the logistic regression can be done with a variety of settings, but the ones considered here are the intercept and the solver. The intercept (or bias) can be used or not set. The solver can be either liblinear (tends to be better for small sets) or saga (better for large sets).

2.2.8 Gaussian Process

Gaussian process is a supervised learning method that is based on a collection of random variables using a gaussian distribution. In classification problems it makes a probabilistic prediction, which allows for the display of confidence intervals for predictions. [29]

Tuning of gaussian process can be done with a variety of settings, but the ones considered here are the number of restarts, the maximum iterations, and how multiclass classifications are done. The number of restarts of the optimizer allows for more internal tuning of the kernel's hyperparameters, with more restarts taking more resources. The maximum number of iterations is also tested, as a lower number of required iterations would be computationally cheaper. Multiclass classifications are tested in the one-vs-rest scheme, in which case a classifier is fitted

to separate each class from the whole dataset, or in one-vs-one, where pairs of materials are compared to draw out their differences.

2.2.9 Model Evaluation and Tuning

We use several measures for evaluating model performance, including cross validation score, test set score, and confusion matrices. Cross validation (CV) is a scheme to prevent overfitting in parameter tuning [30]. In a k-fold cross validation, the training data is randomly split into k-folds, the model is fit on k-1 folds of the data, and an accuracy score is taken on the remaining one fold, then that is repeated k times so that all the data is used. To get an unbiased measure for the accuracy of a model, a held-out test set is also used.

To inspect where a model may be making mistakes, a confusion matrix is used. In a confusion matrix (shown below), the comparisons between predicted and actual can be visualized, which helps to detect if a minority class of materials is being overlooked.

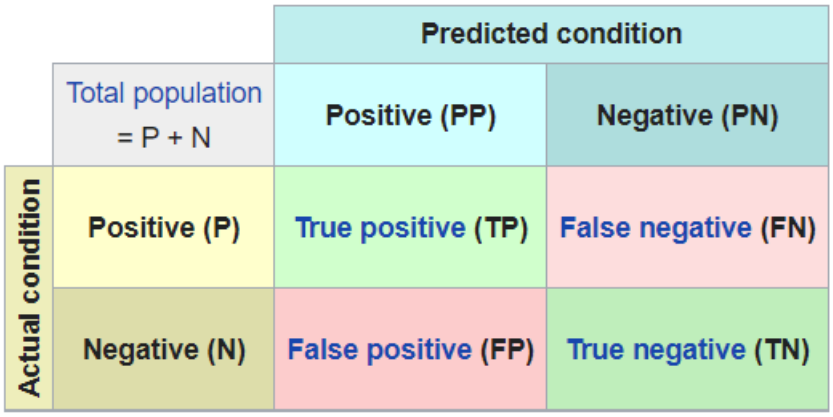

Figure 3 Confusion matrix explanation [31]

The CV score is used for a grid search over each model's parameters [32]. In this search, each combination of model parameters is tested to determine which parameters are ideal for the given dataset, then the fitted model with ideal parameters is returned.

Chapter 3: Machine Learning Classification of Wurtzite, Zincblende, and Rock Salt Structures

3.1 Introduction

Wurtzite and zincblende structures are similar to each other and to rock salt, but have significant differences in application, so the problem of differentiating them without running an expensive calculation is worth addressing. Figure 4 shows the three structures. In this work, we combine crystal structure databases and DFT calculations to train a machine learning model to distinguish between wurtzite, zincblende and rock salt structures of equimolar binary compounds. In constructing our data set, we found 123 new stable wurtzite, zincblende, or rock salt structures, and the model achieves a 72% accuracy in separating the structures based on atomic features. Predictions with that model on a set of ternary materials constructed from charge-balanced permutations of the binary training set lead to a library of a large number of ternary materials that we find to have significantly higher chances to be a wurtzite than a zincblende or rock salt.

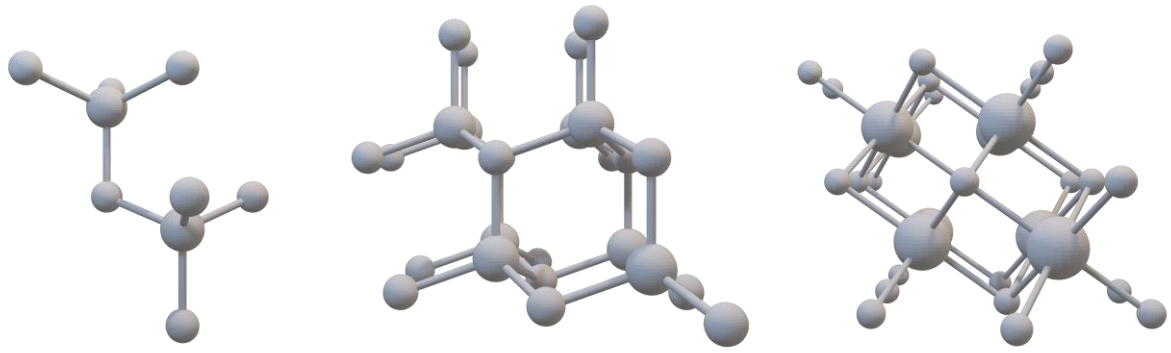

Figure 4 From left to right: wurtzite, zincblende, and rock salt structures

Wurtzite is a hexagonal close packed polar structure that tends to show piezoelectricity and pyroelectricity and has been observed in ferroelectrics [7]. It is named for a polymorph of zincblende, a zinc and iron sulfide mineral. One of the Zn – S bonds is somewhat longer than the rest of the bonds, which will disrupt the mirror symmetries in order to give a polarization.

Zincblende, named for the main ore of zinc sulfide, has an alternating diamond cubic structure. Many semiconductors can be found in a zincblende structure. It is similar to the wurtzite structure, but instead of having a shortened bond all of its bond lengths are equivalent, leading to a nonpolar centrosymmetric structure that has multiple mirror symmetries.

Rock salt structure has each of the two atom types forming interpenetrating face-centered cubic lattice. This is seen in sodium chloride, and it is somewhat more likely when cations are smaller than anions. As the bonds are all the same length, this too is nonpolar.

Binary octet crystals, solids with the chemical composition A^nB^{8-n}, separate into a handful of different structures, and the separation of these structures has been studied for years. As wurtzite and zincblende are some of the most common structures of binary octet semiconductors and since they are incredibly similar [33], quick prediction of the separation of wurtzite and zincblende structures is important for their usage in semiconductors and other industrial applications. Most prior work [34] [19] has focused on separating the rock salt structure from the combined grouping of zincblende and wurtzite, as the gap between the six-fold coordinated rock salt and the four-fold coordinated zincblende and wurtzite structures is somewhat easier for machine learning to draw [35].

Saad *et al*. [36] demonstrated the use unsupervised learning in attempting to separate wurtzite from zincblende. They first identified two descriptors that were able to separate the majority of their binary octet compounds, the St. John-Bloch [37] feature pair r_σ and r_π. Here, we wanted to confirm their work for a larger dataset of binary equimolar compounds and use the identifying factors to discover new wurtzites. Saad *et al.* identifies the two factors separating the binary octets as $r_\pi = [r_P(A) - r_S(A)] + [r_P(B) - r_s(B)]$ and $r_\sigma = \left[r_S(A) + r_p(A)\right] -$

$[r_s(B) + r_p(B)]$. The separation on the dataset used by Saad *et al.* is shown below in Figure 5, with wurtzite shown as W, zincblende as ZB, and rock salt as RS.

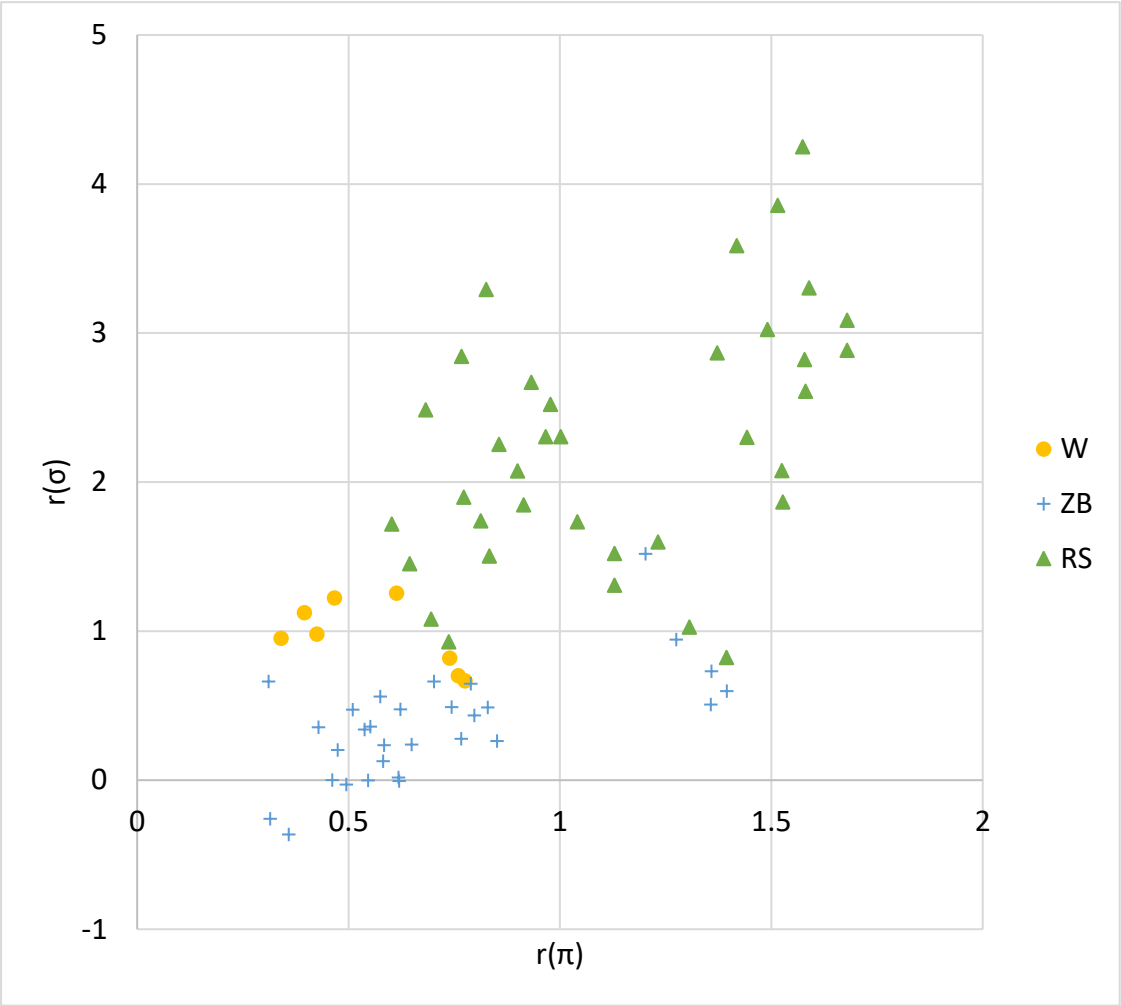

Figure 5 adapted with permission from Reference 36. © 2012 American Physical Society

The goal of this work is to identify a factor differentiating wurtzites from zincblendes and rock salts that can be applied to screen candidates for subsequent DFT calculations. To this end, all the binary wurtzites, zincblendes, and rock salts from Materials Project, as well as some calculated materials, are fed into a machine learning model to identify prominent features to use those features to predict new materials to explore. In the process of that, 123 structures were calculated to be more stable than extant structures and 48 new ternary wurtzites were predicted.

3.2 Methods

In order to create a training set, we filtered materials from the structure database Materials Project [38]. Out of the 63,876 structures currently in the database, 19,282 of these have two elements, 1,679 of those are in wurtzite (space group #186) or zincblende (space group #216), and 693 of those have the formula of AB, so will be included in our training set. A single generic formula of AB was selected to make comparisons between materials more straightforward. Around 620 of these materials were not available in all three structures, (i.e., only available in wurtzite, so zincblende and rock salt were created), so those were listed for calculation using DFT.

High-throughput DFT was done to expand the training set. As these materials are fairly small, a low amount of total computation time was needed on XSEDE[39] Expanse and Stampede2 clusters. The opposite structure for each material (i.e., zincblende for materials that were only available in MP as a wurtzite) was created as a POSCAR file, then relaxation and static calculations were performed. The calculation parameters are detailed in Chapter 2. Method section 1. At the conclusion of these calculations, 247 materials were found to have an energy above hull of 0 meV, indicating that they are stable. 123 of these materials were initialized into a

wurtzite structure, but over the course of the calculation relaxed their way into the hexagonal space group $P6_3/mmc$ (Hermann-Mauguin) #194 (International Tables for Crystallography), 55 were wurtzites #186, 56 were zincblendes #216, and 12 were rock salt #225. A table of these materials is presented in Appendix 4. While this quantity of calculations was doable in the context of hundreds of materials, the expanded search space for new combinations means an ML approach is needed for focusing further calculations.

In order to make a dataset for classification, the material of each formula with the lowest energy was taken to be the target, then various features relating to the property of each atom were attached. With these descriptors, the chart from reference 36 could be recreated below. Here, we have 42 composition that prefer wurtzite phase, 36 in zincblende phase, and 69 in rock salt phase (this set differs in number from the full set, as the descriptors Saad used are not available for all elements). While the feature pair r_σ and r_π do not appear to separate these three phases effectively, we try to overlap our dataset with the original plot from Saad, as shown in Figure 6.

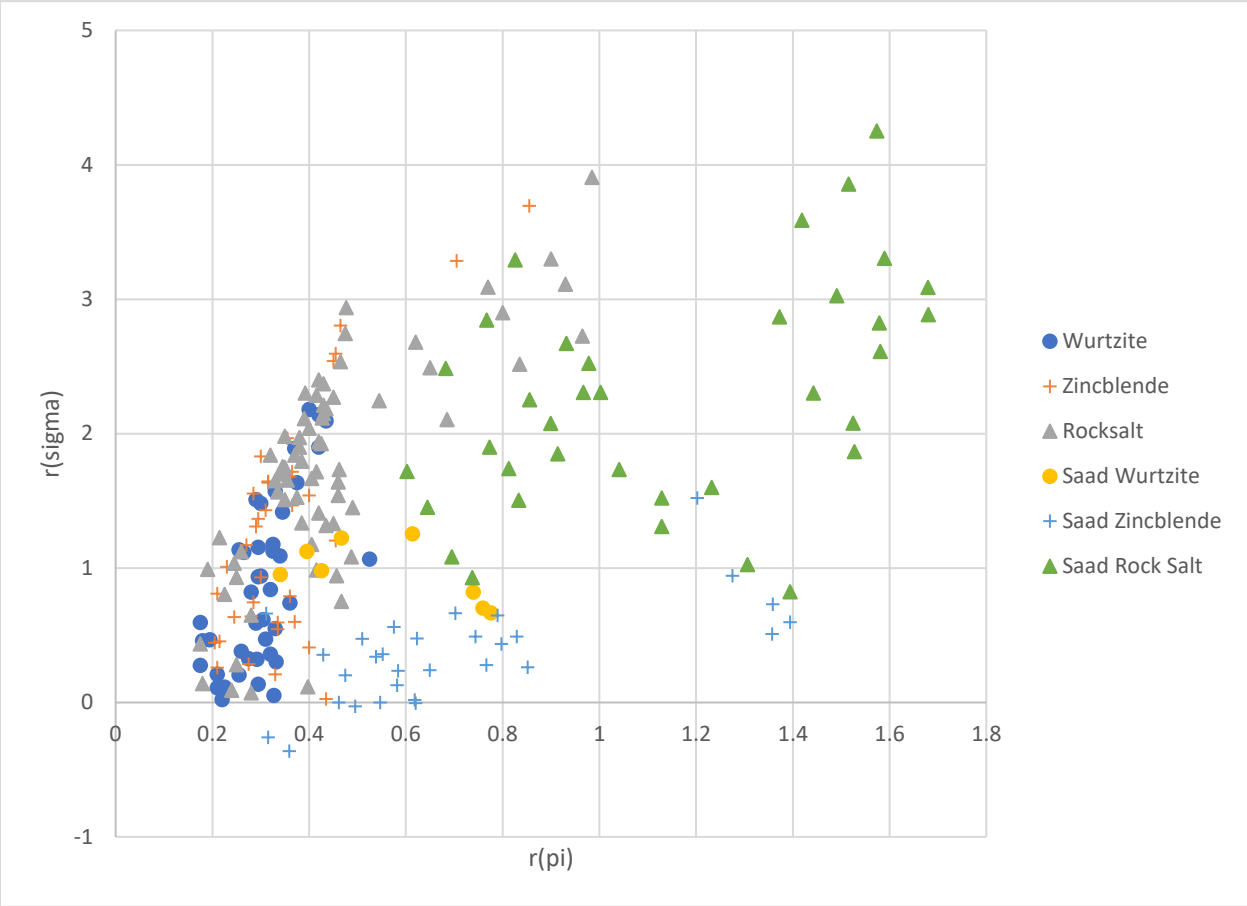

Figure 6 Combination of Saad et al.'s original data with our data

As our new dataset could not be effectively classified using the descriptors Saad *et al.* had used, we used it to fit new classification models. Our test set of 423 materials is set up with a train-test split of 90% train and 10% test, with labels assigned for an ML classification.

The classification program runs a variety of different ML algorithms then compares them on a cross validation test and on a held-back test set. The algorithms from scikit-learn [22] considered are run through a grid search to determine the best parameters and the best model on the given set. Those parameters are detailed in Figure 7.

K Nearest Neighbors	• 3-15 neighbors • uniform or distance weights • Euclidean or Manhattan metrics
Decision Tree	• Gini or entropy criterions to measure quality of a split • a range of max tree depths • a range of minimum samples to split an internal node • a range of minimum samples required at each leaf
Random forest	• Gini or entropy criterions to measure quality of a split • a range of max tree depths • a range of numbers of trees in the forest • max features = sqrt(n_features) or =log2(n_features)
Support vector machines	• Linear, polynomial, radial basis function (rbf), and sigmoid kernels • C, the regularization parameter, in 10x steps from 0.1 to 1000 • Gamma (the kernel coefficient for RBF, polynomial, and sigmoid kernels) in 10x steps from 0.0001 to 1, scale (1 / (n_features * X.var ())), or auto (1/n_features) • Degree of the polynomial kernel in a range from 2 to 6
Neural networks	• Logistic and relu activation functions for the hidden layer • hidden layer size/configurations of (10,30,10), (20), and (20,20,20) • solvers of stochastic gradient descent, adam (stochastic gradient-based optimizer), or lbfgs (quasi-Newton optimizer) • L2 regularization penalty of 0.0001 or 0.05 • learning rate for weight updates either constant or adaptive • max iterations in a range from 100 to 1000
Perceptron	• Penalty function of L1, L2, elasticnet, or none • Regularization constant of 0.0001 or 0.05 • max iterations in a range from 100 to 1000
Lasso	• Fit intercept set to true or false • solver of liblinear or sage
Gaussian Process	• Number of restarts of the optimizer of 0, 1, 2, or 5 • max iteration during predict in a range from 50 to 500 • multiclass classifications are done as one-vs-rest or one-vs-one

Figure 7 Grid search parameters

3.3 Results and Discussion

The classifier we present separates the combined grouping of wurtzite (86 materials) and zincblende (86 materials) from rock salt (251 materials). With all 40 features we gathered, (the

list of features is in Appendix 2, and the list of materials is in Appendix 7) a random forest classifier can separate wurtzite/zincblende from the rock salt with 81% accuracy on a cross validation set and 84% accuracy on a test set. Spending an additional order of magnitude of computing time (22717 seconds (about 6 and a half hours) vs 2204 seconds (about 36 and a half minutes)) to test various neural networks did not result in an improved result. Cutting down the number of features (using the random forest classifier feature reduction described in Chapter 3) shows that, even with only 15 features, accuracy only suffers an acceptable drop to 79% on both cross validation and test set on a random forest classification. The most important features, as identified by the random forest classification, were the anion (*X*) site lattice constant and Mendeleev number. The lattice constant helps to determine geometry of an atom, which is clearly important with two structures as similar as zincblende and wurtzite. Mendeleev number [40] orders the chemical space so that elements with similar properties take neighboring places in a sequence.

When comparing all three classes, accuracy declines relative to the wurtzite and zinc blende classes. The random forest classifier drops from 81% cross validation accuracy with two classes to 69% with three, with a similar drop in test accuracy. To address bias towards the largest class, that of rock salts, a balanced random forest classifier was used. That classifier randomly under-samples the majority class of each tree separately to account for unbalanced data. While overall accuracy was diminished, the confusion matrix in Figure 8 shows that the accuracy for the minority classes are improved by reducing their weight towards rock salt. The most important features for this classification were Mendeleev and Pettifor [40] numbers, which are similar measures.

Predicted ZB	Predicted W	Predicted RS
6	3	3
1	3	3
2	0	22

Random Forest Classification

True ZB
True W
True RS

Predicted ZB	Predicted W	Predicted RS
7	4	1
1	4	2
4	1	19

Balanced Random Forest Classification

Figure 8 Comparison of random forest and balanced random forest confusion matrices separating zinc blende (ZB), wurtzite (W), and rock salt (RS).

When comparing wurtzite and zincblende only, the dataset shrinks so much that overfitting becomes an issue. With all 40 features, the cross-validation score can reach near 70% but against a test set it tends to pick wurtzites very poorly. However, using a random forest classifier to reduce the number of features can dramatically improve testing accuracy while reducing the damage from overfitting. With the top 15 features, cross validation set accuracy stays consistent but the accuracy on the test set is increased, and with the top 10 features, cross validation accuracy again drops to 67% for the top ML method but test set accuracy stays high at 78%.

As a result of the ML selections, predictions can be made on a set of ternary compounds. These compounds were created by taking the training set used and combining two different *A*-site elements with a single *X*-site element. Charge balance is checked, so only charge neutral ternary compounds are output. 25,530 potential materials are selected, after removing duplicates. To set the material properties, the descriptors associated with the two *A*-site elements are averaged into one feature.

3.4 Conclusion

This list is then run through the fitted models created in the earlier machine learning test. 48 compounds stand out as worthy of further investigation. In all of these, both balanced random forest and a trained neural network agree that they are likely to be wurtzite with a greater than 75% degree of certainty. This list is attached in Appendix (ternaryPredictionsAllFeatures.xlsx) information. Future work will simulate a selection of these materials to check that they are in the wurtzite structure.

Chapter 4: Ferroelectric screening and ML-assisted search

4.1 Introduction

Ferroelectric materials have a spontaneous electric polarization or a permanent dipole moment, even after removing an applied electric field. Moreover, the direction of the dipole moment can be switched with an external electric field. Ferroelectric materials are used in capacitors [41], RAM [42], and other sensor applications [43], so their discovery has widespread technological relevance. However, it has been a challenge to design ferroelectric materials due to the costs of experimental creation of materials or of DFT [5]. In this chapter, we attempt to identify key factors needed for new ferroelectrics by finding nonpolar pairs to polar materials from a crystal structure database and evaluating new materials for stability. Our workflow is presented in Figure 9.

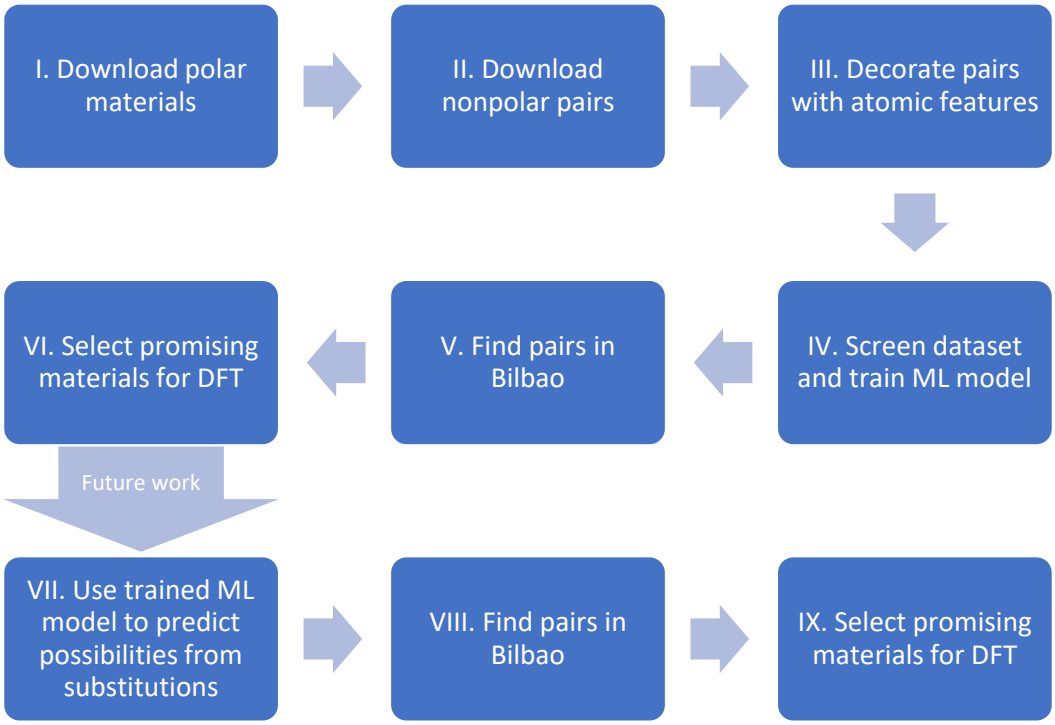

Figure 9 Workflow of ferroelectric search

4.2 Methods

First, in step I of the workflow we query the Materials Project [44] structural database to generate a dataset by finding pairs of polar and nonpolar ternary oxides (and materials near oxides: S, Se, or Te). We gathered ternary compounds that follow the three rules: 1) they must adopt one of ten polar point groups: *1 (C_1), 2 (C_2), m (C_{1h}), mm2 (C_{2v}), 4 (C_4), 4mm (C_{4v}), 3 (C_3), 3m (C_{3v}), 6 (C_6), and 6mm (C_{6v})*; 2) containing elements O, S, Se, or Te; 3) excluding lanthanides, actinides, and artificial elements. This results in 5,847 compounds. Then in the second step II, we match them with paired nonpolar materials, which have the same formula but a centrosymmetric structure, which gives us 8,276 polar – nonpolar. Note that some of the polar

materials have multiple nonpolar matches, so this number of pairs is higher than the number of polar materials. In the next step, we combine this dataset and ML techniques to accelerate the discovery of ferroelectric materials.

To develop an ML model, it is necessary to represent the dataset numerically. Numerical representations of the data are called descriptors or features. In step III, elemental features are collected including orbital radii, orbital energies, atomic radius, Pauling electronegativity, number of electrons and their locations, and the first and second ionization energy for elements. Orbital radii are combined in math operations to create composite features. This is done by adding each of the *s*, *p*, and *d* orbitals in pairs and all together, taking the difference between pairs, taking the ratio between pairs, taking the ratio between one radii vs all radii, taking the ratio of the difference over the sum of pairs, taking the ratio between differences in radii vs all the radii, and taking the ratio of a sum of pairs vs all the radii. The descriptors are listed in Appendix 6.

We treat this problem as a binary classification problem. To determine targets, a polar-non-polar pair is considered to be a potential ferroelectric if it has an energy difference of less than 70 meV per atom and energy above hull of the polar phase less than 0.04 eV per atom, which results in 2016 ferroelectric candidates that we refer to as target 1, and 998 nonferroelectrics outside of that restrictive range that we refer to as target 0. (5031 pairs with a polar band gap of less than 0.1 or an energy above hull of greater than 0.05 eV per atom were dropped before this, in an attempt to identify the features separating materials that are close to those that are good), which are not believed to be candidates. The cutoff point comes from Smidt *et al*'s work [5] in that it closely fits the list of materials they have published as promising ferroelectrics.

One feature considered is the percentage prediction from Xie and Grossman's Crystal Graph Convolutional Neural Networks [45] (CGCNN). We treat it as a feature instead of as an independent ML model so that we can see how impactful the relaxed crystal structure is on a prediction relative to other features. The CGCNN is a convolutional neural network built on top of a crystal graph of crystal structures. This feature is somewhat costly to construct, as it requires each material to include a crystallographic information file (CIF) then trains on these. This neural network is trained separately based on code from Xie and Grossman. The input is the CIF file of each material in the training set, and the output is the probability that each tested material is in the target group. The following kernel density estimate (KDE) plot in Figure 10 visualizes the separation between materials in the target group 1 in red versus those in target group 0 in blue.

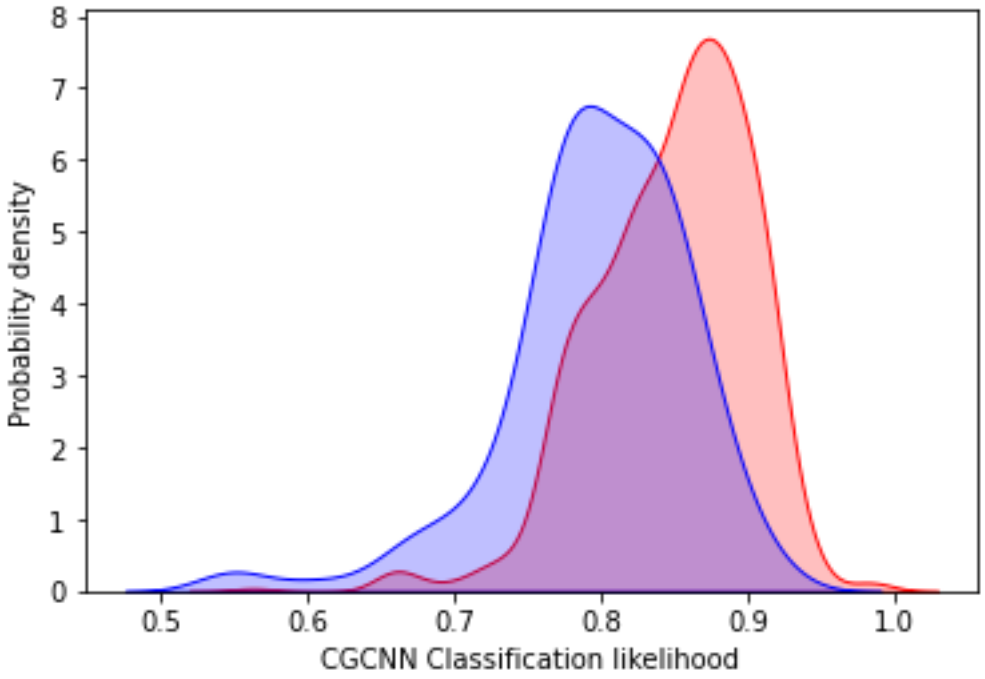

Figure 10 KDE plot showing that CGCNN can effectively separate ferroelectric candidates from non-candidates. The red curve is for targeted ferroelectric candidates.

The predictions from CGCNN are evaluated for inclusion in a logistic regression. This feature is only trained on the polar materials due to a limitation in the code.

A logistic regression with cross-validation is conducted on the collected pairs of materials. The 3014 labelled materials, starting from their 155 columns of features, have an additional 123 columns of features added using matminer's [46] magpie data featurizer and one column of CGCNN's probabilistic predictions. Once strings, energy above hull, spacegroup information, and other strings and calculated data are removed (see notebook in Appendix 9) there are 268 columns of features left to consider. To pare down the list of features, a random forest classifier running in a loop is utilized. For each iteration, a random forest classifier with 100 trees, unlimited max depth, and minimum sample split of 2 is fit to the labelled data. After the fit, an ordered list of features according to their importance is created. From this list, irrelevant features are removed for future fits. After the irrelevant features are all removed, the top fifteen features, listed in Appendix 5, are saved for use in the following logistic regression. Test and training sets for the logistic regression are split using scikit-learn's [22] train_test_split tool with a test set size of 20%. This results in a test set of 2411 pairs and a training set of 603 pairs. A logistic regression, as implemented in scikit-learn's LogisticRegressionCV, [22] with L1 penalty, liblinear solver, and balanced class weights is run repeatedly on this set. After each run, feature importance is plotted so that very low importance features can be removed. After those removals, the top 14 collected material descriptors, including CGCNN, can achieve a CV score of 0.74 with the training set and an accuracy of 0.75 on the test set. In the absence of the CGCNN

prediction, the same set of descriptors can reach CV score of 0.68 and an accuracy of 0.70. Shown below in Figure 11 are the top features used, listed by their impact on the results of the logistic regression, where a greater coefficient denotes higher importance.

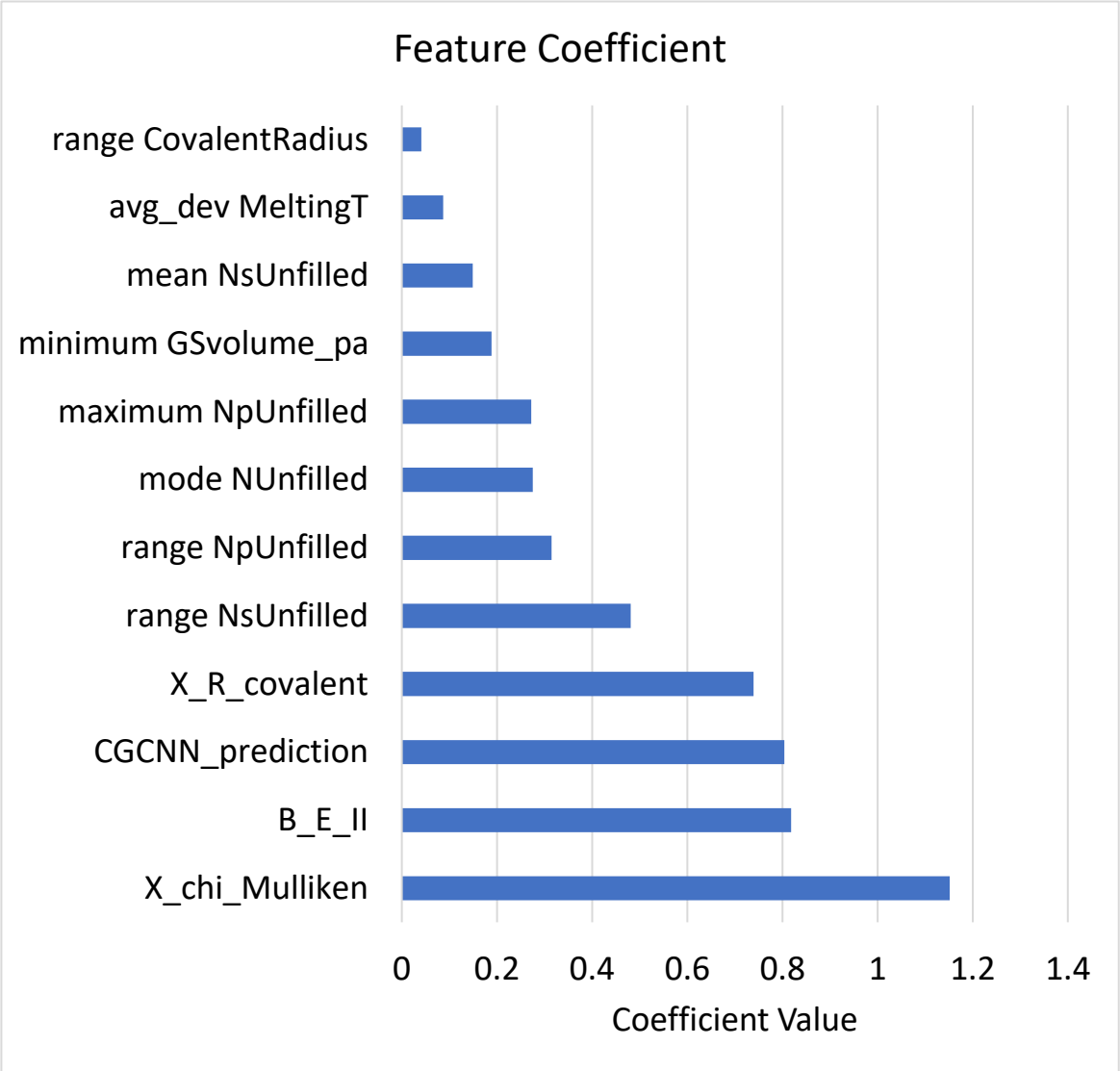

Figure 11 Feature coefficients in a logistic regression of ferroelectric candidates

The features above are, the following features from matminer's Magpie [18]: range of number of unfilled s valence orbitals, maximum of number of unfilled p valence orbitals, range of the

covalent radius of each element, average deviation of the melting temperature, mean of the number of unfilled s valence orbitals, minimum of the DFT volume per atom of T = 0K ground state, mode of the number of unfilled valence orbitals, and the range of the number of unfilled p valence orbitals, as well as X site single bond covalent radius by Pyykko [47] from Mendeleev, CGCNN's probability that a material is a ferroelectric candidate [45], B site second ion energy, and X site Mulliken [48] electronegativity taken from the Mendeleev package [49].

We combine the set of polar materials that are predicted to have a high likelihood of being ferroelectrics with the Bilbao [50] tools made by the University of the Basque Country to generate new nonpolar pseudosymmetries and Brigham Young University's Isodistort [51] to demonstrate the distortion path taken in step V. This pseudosymmetry search is available at https://www.cryst.ehu.es/cryst/pseudosymmetry.html, and the distortion visualization is at https://stokes.byu.edu/iso/isodistort.php. Starting from the polar structure, we search Bilbao for minimal supergroups with a tolerance of 0.5 Å (higher will increase the likelihood of finding pyroelectrics instead of ferroelectrics). The resulting space groups can be compared, and the lowest energy nonpolar structure is the preferred pseudosymmetry. With the relaxed polar and nonpolar structures, Isodistort shows the mode decomposition of a distorted structure with the choice of multiple probable bases, then can generate a movie of transitional steps between the polar and nonpolar structures. From that, a chart of eV/atom can be used to tell the difference between pyroelectrics and ferroelectrics.

Because Bilbao and Isodistort do not have available APIs, there is no straightforward way to automate the usage of these tools for large datasets. In order to solve that problem to check thousands of materials we had to automate form-filling with Mechanize [52], a programmatic web

browsing tool in Python. We also used the Beautiful Soup [53] HTML parser to help convert between Bilbao's internal format for crystal structures and CIF. This integrates cleanly into our workflow, which takes place in Jupyter notebooks.

4.3 Results and Discussion

We plot the prediction results of the logistic regression on a 20% held-out test set as a confusion matrix, as shown in Figure 12. In the confusion matrix, "0" is the label for a non-candidate and "1" for a ferroelectric candidate. Along the diagonal, we have 113 non-candidate samples and 347 candidate samples where the prediction agrees with the actual label, while on the top-right and left-bottom, the model gives the wrong prediction, which results in a true-positive rate of 0.85, true negative rate of 0.59 and an overall accuracy of 0.76. Both neural network classifier and K nearest neighbors classifier can improve predictions, in the best case by up to 8%. As those classifiers are less interpretable, the list of important descriptors is obtained from the logistic regression.

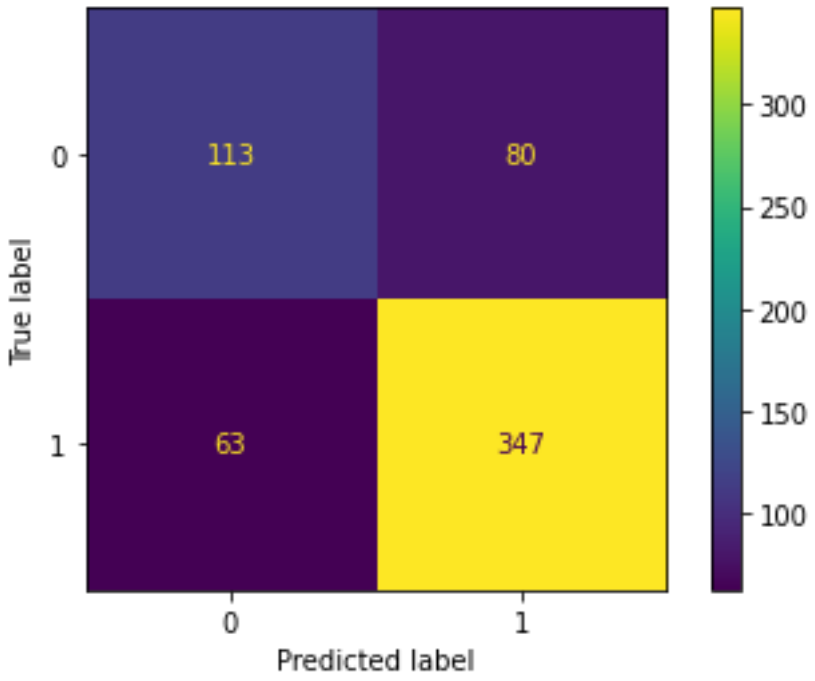

Figure 12 Confusion matrix of logistic regression of ferroelectric candidates, where 1 is a candidate

To be sure that a transition path from the polar to the non-polar structure is possible, we use the Bilbao pseudosymmetry tool [50] to check the symmetry relations. This tool was developed to check if individual polar structures can be related to a non-polar configuration, which identifies them as suitable candidate ferroelectrics. For every one of the polar materials that was pulled from Materials Project, a pseudosymmetry match is searched. The resulting list of material pairs contains 401 possible candidate ferroelectric materials.

Out of 401 pairs that passed the pseudosymmetry analysis, a couple of the resulting material pairs were selected for further inspection. This list is available in Appendix 8. These pairs were based on the structure type of $AgBiS_2$ [54] and $TaFeO_4$ [55]. $AgBiS_2$ can make the transition from *Cc* (Hermann-Mauguin # 9 to *C*2/*c* #15; $TaFeO_4$ can transition from $I4_1md$ #109 to $I4_1/amd$ #141. Substitutions of the cations were done on each based upon the most likely substitutions as

identified by Hautier *et al.* [56]. In their work, Hautier *et al.* present a predictive model that assesses the likelihood of ionic charged elements replacing one another. This model was trained on experimental databases of crystal structures. For Ag in AgBiS2, Hautier *et al.*'s model recommends substitutions of Cu, Au, Li, Na, and Tl. For TaFeO4, the model recommends substitutions of V, Nb, Sb, and As for Ta, with a substitution of Hf for Ta and of Fe^{+4} for the Fe^{+3}. Each of the 12 material pairs was relaxed in DFT. The computation details are included in Chapter 2.1.7. With the relaxed structures, we can use Isodistort [51] to display the interpolation path taken between the polar and nonpolar.

We provide the computed chemical formula, space group, band gap, and energy above hull of the polar-nonpolar pairs in Figure 13.

Material	Polar Band gap (eV)	Polar E_above_hull (meV)	NonPolar Band Gap (eV)	NonPolar E_above_hull (meV)
NaBiS2	1.26	97.7	0.96	99.6
AgBiS2	0.71	116	0.29	123.9
TlBiS2	1.27	144	0.75	145
AuBiS2	1.06	159	0.009	312.35
CuBiS2	0.2592	233.65	0	340.65
AsFeO4	0	164	0	483
SbFeO4	0	78	0	613
NbFeO4	1.81	296	0	762
TaFeO4	2	281	0.1	784
VFeO4	0.4	405	0	712

HfFeO4	0	490	0	949

Figure 13 Results of DFT calculations on distorted polar-nonpolar pairs

Two of the better-looking possibilities here, having acceptable stability and band gaps and not having toxic elements (Tl), NaBiS2 and AgBiS2, have been reported as having been synthesized and seen to exhibit ferroelectric properties (by Rosales [57] and Dutta [58], respectively). Experimental band gap measures from Rosales line up with that calculated using DFT. In each of these cases, a significant amount of intuition and domain knowledge was needed to know where to look for new materials; in this work, the process of focusing a material search can be mostly automated.

4.4 Conclusion

We have seen that our combined machine learning and automated pseudosymmetry searches are able to lead to promising ferroelectric candidates. In future work, we plan to create a new set of materials via substitution, use the previously trained ML model to identify which of those materials are likely to be ferroelectric, then use Bilbao to identify pseudosymmetries and DFT to model polar-nonpolar pairs.

References

(1) Flack, H. D.; Schwarzenbach, D. On the Use of Least-Squares Restraints for Origin Fixing in Polar Space Groups. *Acta Crystallogr. A* **1988**, *44* (4), 499–506. https://doi.org/10.1107/S0108767388002697.

(2) Fousek, J. Joseph Valasek and the Discovery of Ferroelectricity. In *Proceedings of 1994 IEEE International Symposium on Applications of Ferroelectrics*; 1994; pp 1–5. https://doi.org/10.1109/ISAF.1994.522283.

(3) Batra, R.; Tran, H. D.; Johnson, B.; Zoellner, B.; Maggard, P. A.; Jones, J. L.; Rossetti, G. A.; Ramprasad, R. Search for Ferroelectric Binary Oxides: Chemical and Structural Space Exploration Guided by Group Theory and Computations. *Chem. Mater.* **2020**, *32* (9), 3823–3832. https://doi.org/10.1021/acs.chemmater.9b05324.

(4) Pielnhofer, F.; Diehl, L.; Jiménez-Solano, A.; Bussmann-Holder, A.; Schön, J. C.; Lotsch, B. V. Examination of Possible High-Pressure Candidates of SnTiO3: The Search for Novel Ferroelectric Materials. *APL Mater.* **2021**, *9* (2), 021103. https://doi.org/10.1063/5.0029968.

(5) Smidt, T. E.; Mack, S. A.; Reyes-Lillo, S. E.; Jain, A.; Neaton, J. B. An Automatically Curated First-Principles Database of Ferroelectrics. *Sci. Data* **2020**, *7* (1), 72. https://doi.org/10.1038/s41597-020-0407-9.

(6) Balachandran, P. V.; Kowalski, B.; Sehirlioglu, A.; Lookman, T. Experimental Search for High-Temperature Ferroelectric Perovskites Guided by Two-Step Machine Learning. *Nat. Commun.* **2018**, *9* (1), 1668. https://doi.org/10.1038/s41467-018-03821-9.

(7) Moriwake, H.; Yokoi, R.; Taguchi, A.; Ogawa, T.; Fisher, C. A. J.; Kuwabara, A.; Sato, Y.; Shimizu, T.; Hamasaki, Y.; Takashima, H.; Itoh, M. A Computational Search for Wurtzite-Structured Ferroelectrics with Low Coercive Voltages. *APL Mater.* **2020**, *8* (12), 121102. https://doi.org/10.1063/5.0023626.

(8) Kresse, G.; Hafner, J. Ab Initio Molecular Dynamics for Liquid Metals. *Phys. Rev. B* **1993**, *47* (1), 558–561. https://doi.org/10.1103/PhysRevB.47.558.

(9) Born, M.; Oppenheimer, R. Zur Quantentheorie Der Molekeln. *Ann. Phys.* **1927**, *389* (20), 457–484. https://doi.org/10.1002/andp.19273892002.

(10) Schrödinger, E. An Undulatory Theory of the Mechanics of Atoms and Molecules. *Phys. Rev.* **1926**, *28* (6), 1049–1070. https://doi.org/10.1103/PhysRev.28.1049.

(11) Hohenberg, P.; Kohn, W. Inhomogeneous Electron Gas. *Phys. Rev.* **1964**, *136* (3B), B864–B871. https://doi.org/10.1103/PhysRev.136.B864.

(12) What is Density Functional Theory? In *Density Functional Theory*; John Wiley & Sons, Ltd, 2009; pp 1–33. https://doi.org/10.1002/9780470447710.ch1.

(13) Kohn, W.; Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.* **1965**, *140* (4A), A1133–A1138. https://doi.org/10.1103/PhysRev.140.A1133.

(14) Blöchl, P. E. Projector Augmented-Wave Method. *Phys. Rev. B* **1994**, *50* (24), 17953–17979. https://doi.org/10.1103/PhysRevB.50.17953.

(15) Nuts and Bolts of DFT Calculations. In *Density Functional Theory*; John Wiley & Sons, Ltd, 2009; pp 49–81. https://doi.org/10.1002/9780470447710.ch3.

(16) Monkhorst, H. J.; Pack, J. D. Special Points for Brillouin-Zone Integrations. *Phys. Rev. B* **1976**, *13* (12), 5188–5192. https://doi.org/10.1103/PhysRevB.13.5188.

(17) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77* (18), 3865–3868. https://doi.org/10.1103/PhysRevLett.77.3865.

(18) Ward, L.; Agrawal, A.; Choudhary, A.; Wolverton, C. A General-Purpose Machine Learning Framework for Predicting Properties of Inorganic Materials. *Npj Comput. Mater.* **2016**, *2* (1), 1–7. https://doi.org/10.1038/npjcompumats.2016.28.

(19) 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* (10), 105503. https://doi.org/10.1103/PhysRevLett.114.105503.

(20) Altman, N. S. An Introduction to Kernel and Nearest-Neighbor Nonparametric Regression. *Am. Stat.* **1992**, *46* (3), 175–185. https://doi.org/10.1080/00031305.1992.10475879.

(21) Aggarwal, C. C.; Hinneburg, A.; Keim, D. A. On the Surprising Behavior of Distance Metrics in High Dimensional Space. In *Database Theory — ICDT 2001*; Van den Bussche, J., Vianu, V., Eds.; Goos, G., Hartmanis, J., van Leeuwen, J., Series Eds.; Lecture Notes in Computer Science; Springer Berlin Heidelberg: Berlin, Heidelberg, 2001; Vol. 1973, pp 420–434. https://doi.org/10.1007/3-540-44503-X_27.

(22) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, E. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.

(23) Decision Trees: Gini vs Entropy ⋆ Quantdare. *Quantdare*, 2020.

(24) Breiman, L. Random Forests. *Mach. Learn.* **2001**, *45* (1), 5–32.

(25) Cortes, C.; Vapnik, V. Support-Vector Networks. *Mach. Learn.* **1995**, *20* (3), 273–297.

(26) Kernel Functions for Machine Learning Applications – César Souza.

(27) 1.17. Neural network models (supervised) https://scikit-learn/stable/modules/neural_networks_supervised.html (accessed 2022 -03 -15).

(28) 1.1. Linear Models https://scikit-learn/stable/modules/linear_model.html (accessed 2022 -03 -16).

(29) 1.7. Gaussian Processes https://scikit-learn/stable/modules/gaussian_process.html (accessed 2022 -03 -16).

(30) Cawley, G. C.; Talbot, N. L. C. On Over-FItting in Model Selection and Subsequent Selection Bias in Performance Evaluation. 29.

(31) Fawcett, T. An Introduction to ROC Analysis. *Pattern Recognit. Lett.* **2006**, *27* (8), 861–874. https://doi.org/10.1016/j.patrec.2005.10.010.

(32) 3.2. Tuning the hyper-parameters of an estimator https://scikit-learn/stable/modules/grid_search.html (accessed 2022 -03 -16).

(33) Mazaheri, T.; Sun, B.; Scher-Zagier, J.; Thind, A. S.; Magee, D.; Ronhovde, P.; Lookman, T.; Mishra, R.; Nussinov, Z. Stochastic Replica Voting Machine Prediction of Stable Cubic and Double Perovskite Materials and Binary Alloys. *Phys. Rev. Mater.* **2019**, *3* (6), 063802. https://doi.org/10.1103/PhysRevMaterials.3.063802.

(34) Pilania, G.; Gubernatis, J. E.; Lookman, T. Structure Classification and Melting Temperature Prediction in Octet AB Solids via Machine Learning. *Phys. Rev. B* **2015**, *91* (21), 214302. https://doi.org/10.1103/PhysRevB.91.214302.

(35) Pilania, G.; Gubernatis, J. E.; Lookman, T. Classification of Octet AB-Type Binary Compounds Using Dynamical Charges: A Materials Informatics Perspective. *Sci. Rep.* **2015**, *5* (1), 17504. https://doi.org/10.1038/srep17504.

(36) Saad, Y.; Gao, D.; Ngo, T.; Bobbitt, S.; Chelikowsky, J. R.; Andreoni, W. Data Mining for Materials: Computational Experiments with A B Compounds. *Phys. Rev. B* **2012**, *85* (10), 104104. https://doi.org/10.1103/PhysRevB.85.104104.

(37) John, J.; Bloch, A. N. Quantum-Defect Electronegativity Scale for Nontransition Elements. *Phys. Rev. Lett.* **1974**, *33* (18), 1095–1098. https://doi.org/10.1103/PhysRevLett.33.1095.

(38) Ong, S. P.; Cholia, S.; Jain, A.; Brafman, M.; Gunter, D.; Ceder, G.; Persson, K. A. The Materials Application Programming Interface (API): A Simple, Flexible and Efficient API for Materials Data Based on REpresentational State Transfer (REST) Principles. *Comput. Mater. Sci.* **2015**, *97*, 209–215. https://doi.org/10.1016/j.commatsci.2014.10.037.

(39) Towns, J.; Cockerill, T.; Dahan, M.; Foster, I.; Gaither, K.; Grimshaw, A.; Hazlewood, V.; Lathrop, S.; Lifka, D.; Peterson, G. D.; Roskies, R.; Scott, J. R.; Wilkins-Diehr, N. XSEDE: Accelerating Scientific Discovery. *Comput. Sci. Eng.* **2014**, *16* (05), 62–74. https://doi.org/10.1109/MCSE.2014.80.

(40) Pettifor, D. G. A Chemical Scale for Crystal-Structure Maps. *Solid State Commun.* **1984**, *51* (1), 31–34. https://doi.org/10.1016/0038-1098(84)90765-8.

(41) Dawber, M.; Chandra, P.; Littlewood, P. B.; Scott, J. F. Depolarization Corrections to the Coercive Field in Thin-Film Ferroelectrics. *J. Phys. Condens. Matter* **2003**, *15* (24), L393–L398. https://doi.org/10.1088/0953-8984/15/24/106.

(42) Arimoto, Y.; Ishiwara, H. Current Status of Ferroelectric Random-Access Memory. *MRS Bull.* **2004**, *29* (11), 823–828. https://doi.org/10.1557/mrs2004.235.

(43) Lee, Y.; Park, J.; Cho, S.; Shin, Y.-E.; Lee, H.; Kim, J.; Myoung, J.; Cho, S.; Kang, S.; Baig, C.; Ko, H. Flexible Ferroelectric Sensors with Ultrahigh Pressure Sensitivity and Linear Response over Exceptionally Broad Pressure Range. *ACS Nano* **2018**, *12* (4), 4045–4054. https://doi.org/10.1021/acsnano.8b01805.

(44) 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* (1), 011002. https://doi.org/10.1063/1.4812323.

(45) Xie, T.; Grossman, J. C. Crystal Graph Convolutional Neural Networks for an Accurate and Interpretable Prediction of Material Properties. *Phys. Rev. Lett.* **2018**, *120* (14), 145301. https://doi.org/10.1103/PhysRevLett.120.145301.

(46) Ward, L.; Dunn, A.; Faghaninia, A.; Zimmermann, N. E. R.; Bajaj, S.; Wang, Q.; Montoya, J.; Chen, J.; Bystrom, K.; Dylla, M.; Chard, K.; Asta, M.; Persson, K. A.; Snyder, G. J.; Foster, I.; Jain, A. Matminer: An Open Source Toolkit for Materials Data Mining. *Comput. Mater. Sci.* **2018**, *152*, 60–69. https://doi.org/10.1016/j.commatsci.2018.05.018.

(47) Pyykkö, P.; Atsumi, M. Molecular Single-Bond Covalent Radii for Elements 1–118. *Chem. – Eur. J.* **2009**, *15* (1), 186–197. https://doi.org/10.1002/chem.200800987.

(48) Mulliken, R. S. A New Electroaffinity Scale; Together with Data on Valence States and on Valence Ionization Potentials and Electron Affinities. *J. Chem. Phys.* **1934**, *2* (11), 782–793. https://doi.org/10.1063/1.1749394.

(49) Mentel, Ł. *Mendeleev – A Python Resource for Properties of Chemical Elements, Ions and Isotopes*; 2014.

(50) Capillas, C.; Tasci, E. S.; Flor, G. de la; Orobengoa, D.; Perez-Mato, J. M.; Aroyo, M. I. A New Computer Tool at the Bilbao Crystallographic Server to Detect and Characterize Pseudosymmetry. **2011**, *226* (2), 186–196. https://doi.org/10.1524/zkri.2011.1321.

(51) Campbell, B. J.; Stokes, H. T.; Tanner, D. E.; Hatch, D. M. *ISODISPLACE* : A Web-Based Tool for Exploring Structural Distortions. *J. Appl. Crystallogr.* **2006**, *39* (4), 607–614. https://doi.org/10.1107/S0021889806014075.

(52) mechanize — mechanize 0.4.7 documentation https://mechanize.readthedocs.io/en/latest/index.html (accessed 2022 -03 -08).

(53) Beautiful Soup Documentation — Beautiful Soup 4.9.0 documentation https://www.crummy.com/software/BeautifulSoup/bs4/doc/ (accessed 2022 -04 -19).

(54) Persson, K. Materials Data on AgBiS2 (SG:9) by Materials Project, 2016. https://doi.org/10.17188/1282902.

(55) Persson, K. Materials Data on TaFeO4 (SG:109) by Materials Project, 2016. https://doi.org/10.17188/1291884.

(56) Hautier, G.; Fischer, C.; Ehrlacher, V.; Jain, A.; Ceder, G. Data Mined Ionic Substitutions for the Discovery of New Compounds. *Inorg. Chem.* **2011**, *50* (2), 656–663. https://doi.org/10.1021/ic102031h.

(57) Rosales, B. A.; White, M. A.; Vela, J. Solution-Grown Sodium Bismuth Dichalcogenides: Toward Earth-Abundant, Biocompatible Semiconductors. *J. Am. Chem. Soc.* **2018**, *140* (10), 3736–3742. https://doi.org/10.1021/jacs.7b12873.

(58) Dutta, M.; Sarkar, D.; Biswas, K. Intrinsically Ultralow Thermal Conductive Inorganic Solids for High Thermoelectric Performance. *Chem. Commun.* **2021**, *57* (39), 4751–4767. https://doi.org/10.1039/D1CC00830G.

(59) Ong, S. P.; Richards, W. D.; Jain, A.; Hautier, G.; Kocher, M.; Cholia, S.; Gunter, D.; Chevrier, V. L.; Persson, K. A.; Ceder, G. Python Materials Genomics (Pymatgen): A Robust, Open-Source Python Library for Materials Analysis. *Comput. Mater. Sci.* **2013**, *68*, 314–319. https://doi.org/10.1016/j.commatsci.2012.10.028.

(60) The VASP Manual - Vaspwiki https://www.vasp.at/wiki/index.php/The_VASP_Manual (accessed 2022 -03 -17).

Appendix:

1. DFT Full results

a. https://d.docs.live.net/610939716c225021/Documents/Grad_School/Thesis%20Project/DFT_calculations%20(sorted).xlsb.xlsx

2. XSEDE Statement:

a. This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1548562. This work used Stampede2 at TACC and Expanse at SDSC through allocation DMR160007.

3. Features for Wurtzite dataset:

a. 'A_{wt}', 'X_{wt}', 'M_A', 'M_X', 'A_{Mend}', 'X_{Mend}', M_{A_wt}', 'M_{X_wt}', 'A_{Mend_wt}', 'X_{Mend_wt}', A/X', 'Xsite_Zunger_r_p', 'Xsite_Zunger_r_s', 'A_AN', 'X_AN', 'A_atomic_radius', 'X_atomic_radius', 'A_mass ', 'X_mass ', 'A_mendeleev_number ', 'X_mendeleev_number ', 'A_vdw_radius ', 'X_vdw_radius ', 'A_en_pauling ', 'X_en_pauling ', 'A_melting_point', 'X_melting_point', 'A_lattice_constant', 'X_lattice_constant', 'A_pettifor_number', 'X_pettifor_number', 'X_electron_affinity', 'A_ionization_potential', 'X_ionization_potential', 'A_electrons', 'X_electrons', 'A_dipole_polarizability', 'X_dipole_polarizability', 'dipole A/X'

4. Stable DFT Results for Wurtzite chapter:

Material id	formula	Spacegroup symbol	Space group point group	Space group number	Energy above hull	Formation energy (meV)

dft-016	NdN	AB	P6_3mc	6mm	186	0	-1650.17
dft-019	SnSe	AB	P6_3mc	6mm	186	0	-686.696
dft-021	GeSe	AB	P6_3mc	6mm	186	0	-471.628
dft-023	VH	AB	P6_3mc	6mm	186	0	-1226.4
dft-024	PrN	AB	P6_3mc	6mm	186	0	-1588.71
dft-025	GeTe	AB	P6_3mc	6mm	186	0	-318.848
dft-030	SnAs	AB	P6_3mc	6mm	186	0	-342.493
dft-032	TlF	AB	P6_3mc	6mm	186	0	-2038.79
dft-033	AgBr	AB	P6_3mc	6mm	186	0	-719.267
dft-042	TiH	AB	P6_3mc	6mm	186	0	-591.253
dft-043	CrN	AB	P6_3mc	6mm	186	0	-1718.06
dft-045	MnTe	AB	P6_3mc	6mm	186	0	-910.027
dft-052	PuN	AB	P6_3mc	6mm	186	0	-1485.77
dft-061	PuC	AB	P6_3mc	6mm	186	0	-63.7764
dft-063	CuS	AB	P6_3mc	6mm	186	0	-515.501
dft-069	HgTe	AB	P6_3mc	6mm	186	0	-510.951
dft-071	HgSe	AB	P6_3mc	6mm	186	0	-552.56
dft-073	CeC	AB	P6_3mc	6mm	186	0	-306.326
dft-081	CuBr	AB	P6_3mc	6mm	186	0	-600.625
dft-082	VN	AB	P6_3mc	6mm	186	0	-1998.06
dft-083	ErH	AB	P6_3mc	6mm	186	0	-793.798
dft-115	CoN	AB	P6_3mc	6mm	186	0	-1124.07
dft-117	FeN	AB	P6_3mc	6mm	186	0	-1734.19
dft-125	WC	AB	P6_3mc	6mm	186	0	-2026.08
dft-134	BeSe	AB	P6_3mc	6mm	186	0	-1138.15
dft-138	HRh	AB	P6_3mc	6mm	186	0	-243.641
dft-146	FeH	AB	P6_3mc	6mm	186	0	-1606.37
dft-158	ReN	AB	P6_3mc	6mm	186	0	-222.506
dft-161	NiC	AB	P6_3mc	6mm	186	0	-468.702
dft-168	NaI	AB	P6_3mc	6mm	186	0	-1583.69
dft-175	InTe	AB	P6_3mc	6mm	186	0	-501.105
dft-200	ErN	AB	P6_3mc	6mm	186	0	-2031.24
dft-201	SmN	AB	P6_3mc	6mm	186	0	-1768.87
dft-206	PmN	AB	P6_3mc	6mm	186	0	-1719.74
dft-209	HoN	AB	P6_3mc	6mm	186	0	-1996.36
dft-211	CoH	AB	P6_3mc	6mm	186	0	-1077.59
dft-245	DyN	AB	P6_3mc	6mm	186	0	-1960.11
dft-253	ThC	AB	P6_3mc	6mm	186	0	-495.055
dft-255	NiH	AB	P6_3mc	6mm	186	0	-1438.57
dft-284	SnSb	AB	P6_3mc	6mm	186	0	-267.188
dft-291	TbN	AB	P6_3mc	6mm	186	0	-1921.09
dft-301	SbPb	AB	P6_3mc	6mm	186	0	-170.463
dft-314	SnTe	AB	P6_3mc	6mm	186	0	-508.005
dft-332	AgCl	AB	P6_3mc	6mm	186	0	-802.319
dft-343	BeTe	AB	P6_3mc	6mm	186	0	-704.912
dft-345	CuF	AB	P6_3mc	6mm	186	0	-1480.25
dft-346	BeS	AB	P6_3mc	6mm	186	0	-1401.2
dft-354	GaGe	AB	P6_3mc	6mm	186	0	-186.86
dft-358	GaBi	AB	P6_3mc	6mm	186	0	-314.57
dft-370	InBi	AB	P6_3mc	6mm	186	0	-311.202

dft-379	AlBi	AB	P6_3mc	6mm	186	0	-257.338
dft-380	TlAs	AB	P6_3mc	6mm	186	0	-225.452
dft-381	AgSe	AB	P6_3mc	6mm	186	0	-254.135
dft-382	SiSn	AB	P6_3mc	6mm	186	0	-198.337
dft-383	TlSb	AB	P6_3mc	6mm	186	0	-203.371
dft-003	NdBi	AB	P6_3/mmc	6/mmm	194	0	-1055.28
dft-005	NaF	AB	P6_3/mmc	6/mmm	194	0	-3027.87
dft-006	HoAs	AB	P6_3/mmc	6/mmm	194	0	-1635.83
dft-012	ScN	AB	P6_3/mmc	6/mmm	194	0	-2161.96
dft-014	HoP	AB	P6_3/mmc	6/mmm	194	0	-1718.53
dft-015	CeP	AB	P6_3/mmc	6/mmm	194	0	-1500.67
dft-027	PuBi	AB	P6_3/mmc	6/mmm	194	0	-383.189
dft-035	RbI	AB	P6_3/mmc	6/mmm	194	0	-1816.92
dft-036	NbC	AB	P6_3/mmc	6/mmm	194	0	-680.628
dft-038	SnP	AB	P6_3/mmc	6/mmm	194	0	-301.064
dft-039	ZrP	AB	P6_3/mmc	6/mmm	194	0	-1432.05
dft-040	BaSe	AB	P6_3/mmc	6/mmm	194	0	-2289.43
dft-044	GeSb	AB	P6_3/mmc	6/mmm	194	0	-173.631
dft-046	LaSb	AB	P6_3/mmc	6/mmm	194	0	-1371.92
dft-048	ZrAs	AB	P6_3/mmc	6/mmm	194	0	-1181.85
dft-053	VC	AB	P6_3/mmc	6/mmm	194	0	-1505.73
dft-065	MgBi	AB	P6_3/mmc	6/mmm	194	0	-298.614
dft-066	MgSb	AB	P6_3/mmc	6/mmm	194	0	-404.439
dft-074	MnSb	AB	P6_3/mmc	6/mmm	194	0	-455.69
dft-076	PrBi	AB	P6_3/mmc	6/mmm	194	0	-1065.16
dft-088	CsF	AB	P6_3/mmc	6/mmm	194	0	-2942.61
dft-089	MgSn	AB	P6_3/mmc	6/mmm	194	0	-159.591
dft-090	TmAs	AB	P6_3/mmc	6/mmm	194	0	-1617.16
dft-092	CrC	AB	P6_3/mmc	6/mmm	194	0	-1273.51
dft-098	MoC	AB	P6_3/mmc	6/mmm	194	0	-2027.29
dft-104	MnSn	AB	P6_3/mmc	6/mmm	194	0	-731.243
dft-113	NaBr	AB	P6_3/mmc	6/mmm	194	0	-1904.67
dft-124	TbBi	AB	P6_3/mmc	6/mmm	194	0	-991.227
dft-135	TaC	AB	P6_3/mmc	6/mmm	194	0	-756.043
dft-139	YN	AB	P6_3/mmc	6/mmm	194	0	-1950.85
dft-142	MoP	AB	P6_3/mmc	6/mmm	194	0	-2176.15
dft-157	YSb	AB	P6_3/mmc	6/mmm	194	0	-1273.25
dft-163	ErBi	AB	P6_3/mmc	6/mmm	194	0	-934.104
dft-187	CsI	AB	P6_3/mmc	6/mmm	194	0	-1857.54
dft-190	CsBr	AB	P6_3/mmc	6/mmm	194	0	-2098.4
dft-198	TmP	AB	P6_3/mmc	6/mmm	194	0	-1712.75
dft-203	LaBi	AB	P6_3/mmc	6/mmm	194	0	-1153.75
dft-207	PrP	AB	P6_3/mmc	6/mmm	194	0	-1636.95
dft-210	HoSb	AB	P6_3/mmc	6/mmm	194	0	-1219.36
dft-215	LiH	AB	P6_3/mmc	6/mmm	194	0	-733.246
dft-216	LuN	AB	P6_3/mmc	6/mmm	194	0	-2125.44
dft-217	TbP	AB	P6_3/mmc	6/mmm	194	0	-1716.27
dft-218	LuSb	AB	P6_3/mmc	6/mmm	194	0	-1141.91
dft-220	SmP	AB	P6_3/mmc	6/mmm	194	0	-1683.86
dft-221	AgF	AB	P6_3/mmc	6/mmm	194	0	-1263.55

dft-222	SmBi	AB	P6_3/mmc	6/mmm	194	0	-1034.33
dft-224	LuBi	AB	P6_3/mmc	6/mmm	194	0	-864.424
dft-225	NdP	AB	P6_3/mmc	6/mmm	194	0	-1654.88
dft-227	YP	AB	P6_3/mmc	6/mmm	194	0	-1738.76
dft-229	TlI	AB	P6_3/mmc	6/mmm	194	0	-995.728
dft-233	SrTe	AB	P6_3/mmc	6/mmm	194	0	-1968.97
dft-234	TmBi	AB	P6_3/mmc	6/mmm	194	0	-908.215
dft-236	KBr	AB	P6_3/mmc	6/mmm	194	0	-2071.91
dft-237	LaP	AB	P6_3/mmc	6/mmm	194	0	-1704.15
dft-238	TbAs	AB	P6_3/mmc	6/mmm	194	0	-1646.19
dft-241	YAs	AB	P6_3/mmc	6/mmm	194	0	-1668.57
dft-242	CsH	AB	P6_3/mmc	6/mmm	194	0	-457.401
dft-243	TiN	AB	P6_3/mmc	6/mmm	194	0	-1968.03
dft-244	ZrC	AB	P6_3/mmc	6/mmm	194	0	-897.343
dft-248	NdAs	AB	P6_3/mmc	6/mmm	194	0	-1623.2
dft-256	DyAs	AB	P6_3/mmc	6/mmm	194	0	-1642.1
dft-258	ErSb	AB	P6_3/mmc	6/mmm	194	0	-1203.19
dft-260	PrAs	AB	P6_3/mmc	6/mmm	194	0	-1613.66
dft-264	LuAs	AB	P6_3/mmc	6/mmm	194	0	-1593.26
dft-265	HoBi	AB	P6_3/mmc	6/mmm	194	0	-953.849
dft-266	KH	AB	P6_3/mmc	6/mmm	194	0	-508.488
dft-268	YbP	AB	P6_3/mmc	6/mmm	194	0	-880.787
dft-269	RbH	AB	P6_3/mmc	6/mmm	194	0	-468.884
dft-270	TlCl	AB	P6_3/mmc	6/mmm	194	0	-1339.73
dft-271	ScAs	AB	P6_3/mmc	6/mmm	194	0	-1514.65
dft-272	EuP	AB	P6_3/mmc	6/mmm	194	0	-1131.54
dft-273	CeAs	AB	P6_3/mmc	6/mmm	194	0	-1438.99
dft-274	PbSe	AB	P6_3/mmc	6/mmm	194	0	-826.475
dft-276	EuN	AB	P6_3/mmc	6/mmm	194	0	-1045.02
dft-278	SmSb	AB	P6_3/mmc	6/mmm	194	0	-1276.64
dft-279	TlBr	AB	P6_3/mmc	6/mmm	194	0	-1203.8
dft-280	TiC	AB	P6_3/mmc	6/mmm	194	0	-912.815
dft-282	CaTe	AB	P6_3/mmc	6/mmm	194	0	-1886.64
dft-286	YBi	AB	P6_3/mmc	6/mmm	194	0	-1015.24
dft-288	DyBi	AB	P6_3/mmc	6/mmm	194	0	-973.259
dft-290	TbSb	AB	P6_3/mmc	6/mmm	194	0	-1248.99
dft-294	PrSb	AB	P6_3/mmc	6/mmm	194	0	-1291.05
dft-296	YbTe	AB	P6_3/mmc	6/mmm	194	0	-1880.33
dft-299	BaTe	AB	P6_3/mmc	6/mmm	194	0	-2023.83
dft-300	ScBi	AB	P6_3/mmc	6/mmm	194	0	-812.184
dft-304	TmN	AB	P6_3/mmc	6/mmm	194	0	-2072.1
dft-306	ScP	AB	P6_3/mmc	6/mmm	194	0	-1619.98
dft-307	CsCl	AB	P6_3/mmc	6/mmm	194	0	-2262.16
dft-309	LaAs	AB	P6_3/mmc	6/mmm	194	0	-1690.44
dft-312	PuS	AB	P6_3/mmc	6/mmm	194	0	-1880.38
dft-315	ScSb	AB	P6_3/mmc	6/mmm	194	0	-1075.49
dft-316	DySb	AB	P6_3/mmc	6/mmm	194	0	-1235.05
dft-318	KI	AB	P6_3/mmc	6/mmm	194	0	-1802.5
dft-319	HfC	AB	P6_3/mmc	6/mmm	194	0	-1025.82
dft-321	ErAs	AB	P6_3/mmc	6/mmm	194	0	-1629

dft-323	SmAs	AB	P6_3/mmc	6/mmm	194	0	-1637.02
dft-325	LuP	AB	P6_3/mmc	6/mmm	194	0	-1699.21
dft-326	NdSb	AB	P6_3/mmc	6/mmm	194	0	-1287.06
dft-329	RbBr	AB	P6_3/mmc	6/mmm	194	0	-2055.14
dft-330	ErP	AB	P6_3/mmc	6/mmm	194	0	-1717.79
dft-331	TmSb	AB	P6_3/mmc	6/mmm	194	0	-1180.75
dft-334	NaCl	AB	P6_3/mmc	6/mmm	194	0	-2127.69
dft-335	CeBi	AB	P6_3/mmc	6/mmm	194	0	-826.353
dft-337	NaH	AB	P6_3/mmc	6/mmm	194	0	-515.879
dft-341	DyP	AB	P6_3/mmc	6/mmm	194	0	-1718.61
dft-347	SiP	AB	P6_3/mmc	6/mmm	194	0	-305.877
dft-349	TaC	AB	P6_3/mmc	6/mmm	194	0	-756.041
dft-350	ScP	AB	P6_3/mmc	6/mmm	194	0	-1619.98
dft-351	MoC	AB	P6_3/mmc	6/mmm	194	0	-2027.29
dft-352	NbC	AB	P6_3/mmc	6/mmm	194	0	-680.628
dft-355	ZrC	AB	P6_3/mmc	6/mmm	194	0	-897.624
dft-357	HfC	AB	P6_3/mmc	6/mmm	194	0	-1025.82
dft-359	MnP	AB	P6_3/mmc	6/mmm	194	0	-1213.03
dft-362	MnSn	AB	P6_3/mmc	6/mmm	194	0	-731.239
dft-363	MnAs	AB	P6_3/mmc	6/mmm	194	0	-827.039
dft-364	MoP	AB	P6_3/mmc	6/mmm	194	0	-2176.15
dft-372	ZnNi	AB	P6_3/mmc	6/mmm	194	0	-940.433
dft-373	MnSb	AB	P6_3/mmc	6/mmm	194	0	-219.5
dft-376	YN	AB	P6_3/mmc	6/mmm	194	0	-2142.86
dft-384	CaSe	AB	P6_3/mmc	6/mmm	194	0	-2259.18
dft-385	GeP	AB	P6_3/mmc	6/mmm	194	0	-243.804
dft-386	GeSb	AB	P6_3/mmc	6/mmm	194	0	-173.632
dft-387	TiN	AB	P6_3/mmc	6/mmm	194	0	-2160.04
dft-388	ScN	AB	P6_3/mmc	6/mmm	194	0	-2353.95
dft-464	LaN	AB	F-43m	-43m	216	0	-1807.85
dft-467	NiC	AB	F-43m	-43m	216	0	-493.319
dft-469	TePb	AB	F-43m	-43m	216	0	-566.135
dft-473	CuH	AB	F-43m	-43m	216	0	-285.604
dft-478	KF	AB	F-43m	-43m	216	0	-3212.63
dft-479	NaI	AB	F-43m	-43m	216	0	-1579.43
dft-480	PrH	AB	F-43m	-43m	216	0	-723.856
dft-488	AgH	AB	F-43m	-43m	216	0	-212.5
dft-490	NiI	AB	F-43m	-43m	216	0	-1307.82
dft-491	InTe	AB	F-43m	-43m	216	0	-472.932
dft-492	CrH	AB	F-43m	-43m	216	0	-1264.06
dft-493	RbF	AB	F-43m	-43m	216	0	-3147.46
dft-497	ZnBi	AB	F-43m	-43m	216	0	-112.615
dft-498	AgCl	AB	F-43m	-43m	216	0	-1098.07
dft-500	NiCl	AB	F-43m	-43m	216	0	-1758.35
dft-503	ErN	AB	F-43m	-43m	216	0	-2066.68
dft-505	ScSe	AB	F-43m	-43m	216	0	-1619.51
dft-506	FeS	AB	F-43m	-43m	216	0	-1998.68
dft-515	SmN	AB	F-43m	-43m	216	0	-1849.85
dft-517	ZnP	AB	F-43m	-43m	216	0	-241.945
dft-518	SrF	AB	F-43m	-43m	216	0	-3234.48

dft-523	MgH	AB	F-43m	-43m	216	0	-236.568
dft-527	PmN	AB	F-43m	-43m	216	0	-1818.61
dft-529	YbF	AB	F-43m	-43m	216	0	-3380.69
dft-536	CaF	AB	F-43m	-43m	216	0	-3203.46
dft-541	CoH	AB	F-43m	-43m	216	0	-1173.9
dft-543	VO	AB	F-43m	-43m	216	0	-2771.18
dft-544	WS	AB	F-43m	-43m	216	0	-2050.35
dft-546	DyN	AB	F-43m	-43m	216	0	-2011.46
dft-558	ThC	AB	F-43m	-43m	216	0	-420.354
dft-570	SbPb	AB	F-43m	-43m	216	0	-136.404
dft-577	CrCl	AB	F-43m	-43m	216	0	-1073.23
dft-582	TbN	AB	F-43m	-43m	216	0	-1980.34
dft-584	FeO	AB	F-43m	-43m	216	0	-2394.57
dft-587	NiO	AB	F-43m	-43m	216	0	-1948.77
dft-594	NiH	AB	F-43m	-43m	216	0	-1528.91
dft-603	HoN	AB	F-43m	-43m	216	0	-2040.18
dft-604	TlF	AB	F-43m	-43m	216	0	-2267.84
dft-605	SnAs	AB	F-43m	-43m	216	0	-242.236
dft-606	HRh	AB	F-43m	-43m	216	0	-339.458
dft-608	GeTe	AB	F-43m	-43m	216	0	-192.7
dft-609	VH	AB	F-43m	-43m	216	0	-1323.05
dft-610	GeSe	AB	F-43m	-43m	216	0	-334.63
dft-611	SnSe	AB	F-43m	-43m	216	0	-553.67
dft-612	NdN	AB	F-43m	-43m	216	0	-1759.94
dft-613	PuN	AB	F-43m	-43m	216	0	-1710.77
dft-615	CeC	AB	F-43m	-43m	216	0	-245.717
dft-616	ErH	AB	F-43m	-43m	216	0	-828.673
dft-617	TiH	AB	F-43m	-43m	216	0	-720.787
dft-619	FeH	AB	F-43m	-43m	216	0	-1539.95
dft-620	AcN	AB	F-43m	-43m	216	0	-1589.01
dft-621	PmH	AB	F-43m	-43m	216	0	-752.984
dft-622	LiI	AB	F-43m	-43m	216	0	-1582.51
dft-623	LiCl	AB	F-43m	-43m	216	0	-2474.06
dft-624	LiBr	AB	F-43m	-43m	216	0	-1917.04
dft-627	LiF	AB	F-43m	-43m	216	0	-3513.92
dft-389	PmH	AB	Fm-3m	m-3m	225	0	-734.151
dft-393	WS	AB	Fm-3m	m-3m	225	0	-1758.37
dft-424	AcN	AB	Fm-3m	m-3m	225	0	-1401.12
dft-426	AlP	AB	Fm-3m	m-3m	225	0	-669.321
dft-429	AlSb	AB	Fm-3m	m-3m	225	0	-297.8
dft-432	GaSb	AB	Fm-3m	m-3m	225	0	-306.993
dft-434	GaAs	AB	Fm-3m	m-3m	225	0	-378.275
dft-436	SiGe	AB	Fm-3m	m-3m	225	0	-6.41653
dft-437	PrH	AB	Fm-3m	m-3m	225	0	-730.581
dft-445	NiI	AB	Fm-3m	m-3m	225	0	-1149.09
dft-448	ZnBi	AB	Fm-3m	m-3m	225	0	-182.771
dft-453	FeS	AB	Fm-3m	m-3m	225	0	-1565.23

5. **Top 15 features for logistic regression:**
    a. ('MagpieData minimum GSvolume_pa', 'MagpieData range CovalentRadius', 'MagpieData mean NsUnfilled', 'MagpieData mode NUnfilled', 'X_$\\chi_{Mulliken}$', 'B_IE_{II}', 'MagpieData avg_dev MeltingT', 'X_$R_{covalent}$', 'MagpieData avg_dev MendeleevNumber', 'MagpieData maximum NpUnfilled', 'MagpieData minimum NfUnfilled', 'MagpieData maximum NfUnfilled', 'MagpieData range NpUnfilled', 'MagpieData range NsUnfilled', 'MagpieData maximum GSvolume_pa', 'CGCNN_prediction')

6. All descriptors for ferroelectric search:

TARGET	B_N_{pVe}	B_$r_{d/p}$	MagpieData minimum NdValence
npolar_material_id	X_N_{pVe}	X_$r_{d/p}$	MagpieData maximum NdValence
npolar_pretty_formula	A_N_{dVe}	A_$r_{s/s+p+d}$	MagpieData range NdValence
npolar_e_above_hull	B_N_{dVe}	B_$r_{s/s+p+d}$	MagpieData mean NdValence
npolar_band_gap	X_N_{dVe}	X_$r_{s/s+p+d}$	MagpieData avg_dev NdValence
npolar_formation_energy_per_atom	A_IE_{I}	A_$r_{p/s+p+d}$	MagpieData mode NdValence
npolar_final_energy_per_atom	B_IE_{I}	B_$r_{p/s+p+d}$	MagpieData minimum NfValence
npolar_nsites	X_IE_{I}	X_$r_{p/s+p+d}$	MagpieData maximum NfValence
npolar_spacegroup.point_group	A_IE_{II}	A_$r_{d/s+p+d}$	MagpieData range NfValence
npolar_spacegroup.number	B_IE_{II}	B_$r_{d/s+p+d}$	MagpieData mean NfValence
polar_material_id	X_IE_{II}	X_$r_{d/s+p+d}$	MagpieData avg_dev NfValence
polar_pretty_formula	composition	A_$r_{s+p/s+p+d}$	MagpieData mode NfValence
polar_e_above_hull	MagpieData minimum Number	B_$r_{s+p/s+p+d}$	MagpieData minimum NValence
polar_band_gap	MagpieData maximum Number	X_$r_{s+p/s+p+d}$	MagpieData maximum NValence
polar_formation_energy_per_atom	MagpieData range Number	A_$r_{s+d/s+p+d}$	MagpieData range NValence
polar_final_energy_per_atom	MagpieData mean Number	B_$r_{s+d/s+p+d}$	MagpieData mean NValence
polar_nsites	MagpieData avg_dev Number	X_$r_{s+d/s+p+d}$	MagpieData avg_dev NValence
polar_spacegroup.point_group	MagpieData mode Number	A_$r_{p+d/s+p+d}$	MagpieData mode NValence
polar_spacegroup.number	MagpieData minimum MendeleevNumber	B_$r_{p+d/s+p+d}$	MagpieData minimum NsUnfilled
Energy_diff_meV	MagpieData maximum MendeleevNumber	X_$r_{p+d/s+p+d}$	MagpieData maximum NsUnfilled
M_A	MagpieData range MendeleevNumber	A_$r_{s-p/s+p+d}$	MagpieData range NsUnfilled
M_B	MagpieData mean MendeleevNumber	B_$r_{s-p/s+p+d}$	MagpieData mean NsUnfilled
M_X	MagpieData avg_dev MendeleevNumber	X_$r_{s-p/s+p+d}$	MagpieData avg_dev NsUnfilled
A_{Mend}	MagpieData mode MendeleevNumber	A_$r_{s-d/s+p+d}$	MagpieData mode NsUnfilled
B_{Mend}	MagpieData minimum AtomicWeight	B_$r_{s-d/s+p+d}$	MagpieData minimum NpUnfilled
X_{Mend}	MagpieData maximum AtomicWeight	X_$r_{s-d/s+p+d}$	MagpieData maximum NpUnfilled
M_{A_wt}	MagpieData range AtomicWeight	A_$r_{p-d/s+p+d}$	MagpieData range NpUnfilled
M_{B_wt}	MagpieData mean AtomicWeight	B_$r_{p-d/s+p+d}$	MagpieData mean NpUnfilled
M_{X_wt}	MagpieData avg_dev AtomicWeight	X_$r_{p-d/s+p+d}$	MagpieData avg_dev NpUnfilled
A_{Mend_wt}	MagpieData mode AtomicWeight	A_$r_{s-p/s+p}$	MagpieData mode NpUnfilled
B_{Mend_wt}	MagpieData minimum MeltingT	B_$r_{s-p/s+p}$	MagpieData minimum NdUnfilled
X_{Mend_wt}	MagpieData maximum MeltingT	X_$r_{s-p/s+p}$	MagpieData maximum NdUnfilled
A_AN	MagpieData range MeltingT	A_$r_{s-d/s+d}$	MagpieData range NdUnfilled
B_AN	MagpieData mean MeltingT	B_$r_{s-d/s+d}$	MagpieData mean NdUnfilled
X_AN	MagpieData avg_dev MeltingT	X_$r_{s-d/s+d}$	MagpieData avg_dev NdUnfilled
A_r_s	MagpieData mode MeltingT	A_$r_{p-d/p+d}$	MagpieData mode NdUnfilled

B_r_s	MagpieData minimum Column	B_$r_{p-d/p+d}$	MagpieData minimum NfUnfilled
X_r_s	MagpieData maximum Column	X_$r_{p-d/p+d}$	MagpieData maximum NfUnfilled
A_r_p	MagpieData range Column	A_ΔE	MagpieData range NfUnfilled
B_r_p	MagpieData mean Column	B_ΔE	MagpieData mean NfUnfilled
X_r_p	MagpieData avg_dev Column	X_ΔE	MagpieData avg_dev NfUnfilled
A_r_d	MagpieData mode Column	A_R_{atomic}	MagpieData mode NfUnfilled
B_r_d	MagpieData minimum Row	B_R_{atomic}	MagpieData minimum NUnfilled
X_r_d	MagpieData maximum Row	X_R_{atomic}	MagpieData maximum NUnfilled
A_r_{s+p}	MagpieData range Row	A_$R_{covalent}$	MagpieData range NUnfilled
B_r_{s+p}	MagpieData mean Row	B_$R_{covalent}$	MagpieData mean NUnfilled
X_r_{s+p}	MagpieData avg_dev Row	X_$R_{covalent}$	MagpieData avg_dev NUnfilled
A_r_{s+d}	MagpieData mode Row	A_$\chi_{Pauling}$	MagpieData mode NUnfilled
B_r_{s+d}	MagpieData minimum CovalentRadius	B_$\chi_{Pauling}$	MagpieData minimum GSvolume_pa
X_r_{s+d}	MagpieData maximum CovalentRadius	X_$\chi_{Pauling}$	MagpieData maximum GSvolume_pa
A_r_{p+d}	MagpieData range CovalentRadius	A_$\chi_{Mulliken}$	MagpieData range GSvolume_pa
B_r_{p+d}	MagpieData mean CovalentRadius	B_$\chi_{Mulliken}$	MagpieData mean GSvolume_pa
X_r_{p+d}	MagpieData avg_dev CovalentRadius	X_$\chi_{Mulliken}$	MagpieData avg_dev GSvolume_pa
A_r_{s+p+d}	MagpieData mode CovalentRadius	A_χ_{MB}	MagpieData mode GSvolume_pa
B_r_{s+p+d}	MagpieData minimum Electronegativity	B_χ_{MB}	MagpieData minimum GSbandgap
X_r_{s+p+d}	MagpieData maximum Electronegativity	X_χ_{MB}	MagpieData maximum GSbandgap
A_r_{s-p}	MagpieData range Electronegativity	A_$E_{affinity}$	MagpieData range GSbandgap
B_r_{s-p}	MagpieData mean Electronegativity	B_$E_{affinity}$	MagpieData mean GSbandgap
X_r_{s-p}	MagpieData avg_dev Electronegativity	X_$E_{affinity}$	MagpieData avg_dev GSbandgap
A_r_{s-d}	MagpieData mode Electronegativity	A_$P_{affinity}$	MagpieData mode GSbandgap
B_r_{s-d}	MagpieData minimum NsValence	B_$P_{affinity}$	MagpieData minimum GSmagmom
X_r_{s-d}	MagpieData maximum NsValence	X_$P_{affinity}$	MagpieData maximum GSmagmom
A_r_{p-d}	MagpieData range NsValence	A_N_e	MagpieData range GSmagmom
B_r_{p-d}	MagpieData mean NsValence	B_N_e	MagpieData mean GSmagmom
X_r_{p-d}	MagpieData avg_dev NsValence	X_N_e	MagpieData avg_dev GSmagmom
A_$r_{p/s}$	MagpieData mode NsValence	A_N_{Ve}	MagpieData mode GSmagmom
B_$r_{p/s}$	MagpieData minimum NpValence	B_N_{Ve}	MagpieData minimum SpaceGroupNumber
X_$r_{p/s}$	MagpieData maximum NpValence	X_N_{Ve}	MagpieData maximum SpaceGroupNumber
A_$r_{d/s}$	MagpieData range NpValence	A_N_{sVe}	MagpieData range SpaceGroupNumber
B_$r_{d/s}$	MagpieData mean NpValence	B_N_{sVe}	MagpieData mean SpaceGroupNumber
X_$r_{d/s}$	MagpieData avg_dev NpValence	X_N_{sVe}	MagpieData avg_dev SpaceGroupNumber
A_$r_{d/p}$	MagpieData mode NpValence	A_N_{pVe}	MagpieData mode SpaceGroupNumber

7. Full Wurtzite data:

material_id	Pretty formula	band_gap	Spacegroup symbol	Spacegroup point_group	Space group number	Final energy per atom	Formation energy per atom	E above hull
mp-1216	YbO	3.5311	Fm-3m	m-3m	225	-6.52962	-3.63699	0
dft-627	LiF		F-43m	-43m	216		-3.51392	0
dft-529	YbF		F-43m	-43m	216		-3.38069	0
mp-2605	CaO	3.6919	Fm-3m	m-3m	225	-6.43897	-3.3164	0
mp-644481	ScO	0	Fm-3m	m-3m	225	-8.5372	-3.24814	0.066777
mp-10688	CeO	0	Fm-3m	m-3m	225	-8.33119	-3.24181	0
dft-518	SrF		F-43m	-43m	216		-3.23448	0
dft-478	KF		F-43m	-43m	216		-3.21263	0
mp-1206710	LaO	0	Fm-3m	m-3m	225	-7.80281	-3.21198	0.024107
dft-536	CaF		F-43m	-43m	216		-3.20346	0
dft-493	RbF		F-43m	-43m	216		-3.14746	0
mp-21394	EuO	0	Fm-3m	m-3m	225	-10.3976	-3.12874	0
mp-1611	SmO	0	Fm-3m	m-3m	225	-7.59694	-3.11526	0.110978
mp-2542	BeO	7.4631	P6_3mc	6mm	186	-7.10304	-3.1105	0
mp-2472	SrO	3.4487	Fm-3m	m-3m	225	-6.05232	-3.08474	0
mp-754545	NdO	0	Fm-3m	m-3m	225	-7.58789	-3.08098	0.077956
mp-1265	MgO	4.638	Fm-3m	m-3m	225	-5.98371	-3.06243	0
mp-11344	PrO	0	Fm-3m	m-3m	225	-7.56723	-3.05395	0.066283
mp-7870	GdO	0	F-43m	-43m	216	-12.2036	-3.04273	0.211081
mp-1206987	ThO	0	Fm-3m	m-3m	225	-8.83027	-3.0005	0.282985
mp-682	NaF	6.0951	Fm-3m	m-3m	225	-4.32427	-2.95788	0
dft-472	PuO		F-43m	-43m	216		-2.86052	0.043308
mp-1784	CsF	5.281	Fm-3m	m-3m	225	-4.01502	-2.85705	0
mp-1342	BaO	2.2561	Fm-3m	m-3m	225	-5.9136	-2.83128	0
mp-2735	PaO	0	Fm-3m	m-3m	225	-9.6621	-2.78194	0.250076
dft-543	VO		F-43m	-43m	216		-2.77118	0
mp-2664	TiO	0	Fm-3m	m-3m	225	-8.71257	-2.64221	0.273613
mp-7830	UO	0	Fm-3m	m-3m	225	-10.3847	-2.61615	0.211428
mp-1820	YbS	2.444	Fm-3m	m-3m	225	-5.08501	-2.57872	0
dft-107	ZrO		P6_3mc	6mm	186		-2.57178	0.334495
mp-6927	NpO	0	Fm-3m	m-3m	225	-11.1487	-2.55196	0.156594
mp-2350	LaS	0	Fm-3m	m-3m	225	-6.69934	-2.49484	0
mp-1672	CaS	2.5596	Fm-3m	m-3m	225	-5.21944	-2.48321	0
mp-1087	SrS	2.5619	Fm-3m	m-3m	225	-5.05653	-2.4753	0
dft-623	LiCl		F-43m	-43m	216		-2.47406	0
mp-20587	EuS	0	Fm-3m	m-3m	225	-9.35268	-2.47017	0
mp-1500	BaS	2.2091	Fm-3m	m-3m	225	-5.10623	-2.41025	0

dft-584	FeO		F-43m	-43m	216		-2.39457	0
mp-510402	GdS	0	Fm-3m	m-3m	225	-11.1671	-2.39255	0
mp-1096	CeS	0	Fm-3m	m-3m	225	-7.08054	-2.3775	0
mp-1232235	PmS	0	Fm-3m	m-3m	225	-6.47621	-2.36444	0
mp-1748	NdS	0	Fm-3m	m-3m	225	-6.47648	-2.35591	0
mp-1269	SmS	0	Fm-3m	m-3m	225	-6.45034	-2.35501	0
mp-2495	PrS	0	Fm-3m	m-3m	225	-6.47432	-2.34738	0
mp-1610	TbS	0	Fm-3m	m-3m	225	-6.39213	-2.33845	0
mp-1534	YS	0	Fm-3m	m-3m	225	-7.29536	-2.32583	0
mp-2470	DyS	0	Fm-3m	m-3m	225	-6.36159	-2.3217	0
mp-1240	HoS	0	Fm-3m	m-3m	225	-6.33006	-2.30236	0
mp-1623	ErS	0	Fm-3m	m-3m	225	-6.30242	-2.28221	0
dft-604	TlF		F-43m	-43m	216		-2.26784	0
mp-1766	TmS	0	Fm-3m	m-3m	225	-6.23207	-2.25805	0
mp-23193	KCl	5.0448	Fm-3m	m-3m	225	-3.4307	-2.24857	0
mp-573697	CsCl	5.0172	Fm-3m	m-3m	225	-3.31992	-2.24529	0
mp-23295	RbCl	5.0162	Fm-3m	m-3m	225	-3.3524	-2.2352	0
mp-656	LuS	0	Fm-3m	m-3m	225	-6.20047	-2.2035	0
mp-1476	ScS	0	Fm-3m	m-3m	225	-7.09516	-2.19243	0
mp-503	ThS	0	Fm-3m	m-3m	225	-7.61167	-2.16824	0
mp-22862	NaCl	5.1451	Fm-3m	m-3m	225	-3.38812	-2.10509	0
mp-2857	ScN	0.325	Fm-3m	m-3m	225	-9.24656	-2.10409	0
mp-286	YbSe	2.0325	Fm-3m	m-3m	225	-4.60353	-2.08577	0
mp-1102	LuN	0.3793	Fm-3m	m-3m	225	-8.30622	-2.06951	0
dft-503	ErN		F-43m	-43m	216		-2.06668	0
mp-2758	SrSe	2.2914	Fm-3m	m-3m	225	-4.65502	-2.06232	0
dft-578	AlO		F-43m	-43m	216		-2.0528	0.810312
dft-544	WS		F-43m	-43m	216		-2.05035	0
mp-1253	BaSe	2.2121	Fm-3m	m-3m	225	-4.75502	-2.04758	0
dft-603	HoN		F-43m	-43m	216		-2.04018	0
dft-125	WC		P6_3mc	6mm	186		-2.02608	0
mp-1415	CaSe	2.1362	Fm-3m	m-3m	225	-4.77307	-2.02539	0
mp-21009	EuSe	0	Fm-3m	m-3m	225	-8.91914	-2.02516	0
mp-1975	TmN	0.306	Fm-3m	m-3m	225	-8.22769	-2.01393	0
dft-546	DyN		F-43m	-43m	216		-2.01146	0
dft-506	FeS		F-43m	-43m	216		-1.99868	0
dft-082	VN		P6_3mc	6mm	186		-1.99806	0
mp-1161	LaSe	0	Fm-3m	m-3m	225	-6.21027	-1.99431	0
mp-19006	MnO	0.1993	Fm-3m	-3m	225	-7.85671	-1.99345	0
dft-582	TbN		F-43m	-43m	216		-1.98034	0
dft-164	TaO		P6_3mc	6mm	186		-1.95493	0.383703
dft-587	NiO		F-43m	-43m	216		-1.94877	0

mp-2828	HfN	0	Fm-3m	m-3m	225	-10.8935	-1.93864	0
dft-624	LiBr		F-43m	-43m	216		-1.91704	0
dft-147	NbO		P6_3mc	6mm	186		-1.91078	0.375025
mp-492	TiN	0	Fm-3m	m-3m	225	-9.83144	-1.90768	0
mp-2114	YN	0.2858	Fm-3m	m-3m	225	-9.08833	-1.87906	0
mp-1352	ZrN	0	Fm-3m	m-3m	225	-10.1202	-1.87008	0
dft-515	SmN		F-43m	-43m	216		-1.84985	0
mp-510404	GdSe	0	Fm-3m	m-3m	225	-10.6338	-1.8478	0
mp-1674	PrSe	0	Fm-3m	m-3m	225	-5.97478	-1.83637	0
mp-1453	NdSe	0	Fm-3m	m-3m	225	-5.96234	-1.83031	0
mp-1232234	PmSe	0	Fm-3m	m-3m	225	-5.95037	-1.82714	0
mp-1009545	PaN	0	Fm-3m	m-3m	225	-10.5543	-1.82073	0
dft-527	PmN		F-43m	-43m	216		-1.81861	0
mp-2563	CeSe	0	Fm-3m	m-3m	225	-6.52919	-1.81469	0
mp-1447	SmSe	0	Fm-3m	m-3m	225	-5.91703	-1.81023	0
dft-464	LaN		F-43m	-43m	216		-1.80785	0
mp-2133	ZnO	0.7316	P6_3mc	6mm	186	-4.55122	-1.79866	0
mp-1000	BaTe	1.8555	Fm-3m	m-3m	225	-4.32347	-1.79234	0
mp-571222	CsBr	4.4247	Fm-3m	m-3m	225	-3.05206	-1.7859	0
mp-940	GdN	0	Fm-3m	m-3m	225	-12.7846	-1.7703	0
mp-834	ThN	0	Fm-3m	m-3m	225	-9.4472	-1.76403	0
mp-1561	TbSe	0	Fm-3m	m-3m	225	-5.8282	-1.76306	0
mp-23251	KBr	4.5151	Fm-3m	m-3m	225	-3.13531	-1.76165	0
mp-22867	RbBr	4.3724	Fm-3m	m-3m	225	-3.07032	-1.76158	0
mp-2637	YSe	0	Fm-3m	m-3m	225	-6.74156	-1.76056	0
dft-612	NdN		F-43m	-43m	216		-1.75994	0
mp-1315	MgS	2.9642	Fm-3m	m-3m	225	-4.29425	-1.75931	0
dft-500	NiCl		F-43m	-43m	216		-1.75835	0
mp-814	DySe	0	Fm-3m	m-3m	225	-5.79425	-1.7429	0
mp-1958	SrTe	2.0435	Fm-3m	m-3m	225	-4.15908	-1.74269	0
dft-117	FeN		P6_3mc	6mm	186		-1.73419	0
dft-043	CrN		P6_3mc	6mm	186		-1.71806	0
mp-464	HoSe	0	Fm-3m	m-3m	225	-5.75474	-1.71558	0
dft-613	PuN		F-43m	-43m	216		-1.71077	0
mp-10109	NpS	0	Fm-3m	m-3m	225	-9.90868	-1.69829	0
mp-2493	CeN	0	Fm-3m	m-3m	225	-8.63326	-1.69048	0
mp-2491	ErSe	0	Fm-3m	m-3m	225	-5.71901	-1.68735	0
mp-542583	EuTe	0	Fm-3m	m-3m	225	-8.3985	-1.68083	0
mp-1925	ZrS	0	Fm-3m	m-3m	225	-7.68349	-1.67314	0.092123
mp-1779	YbTe	1.5273	Fm-3m	m-3m	225	-4.00703	-1.66558	0
dft-496	GdCl		F-43m	-43m	216		-1.66465	0.169543

mp-2822	TmSe	0	Fm-3m	m-3m	225	-5.6385	-1.65302	0
mp-1519	CaTe	1.6041	Fm-3m	m-3m	225	-4.22278	-1.65139	0
mp-994	YP	0	Fm-3m	m-3m	225	-7.56418	-1.6245	0
dft-505	ScSe		F-43m	-43m	216		-1.61951	0
mp-2014	DyP	0	Fm-3m	m-3m	225	-6.61946	-1.60943	0
mp-19091	CrO	0	Fm-3m	m-3m	225	-5.97664	-1.60919	1.95069
mp-744	HoP	0	Fm-3m	m-3m	225	-6.6068	-1.60895	0
mp-1144	ErP	0	Fm-3m	m-3m	225	-6.59747	-1.60711	0
mp-645	TbP	0	Fm-3m	m-3m	225	-6.6309	-1.60707	0
dft-146	FeH		P6_3mc	6mm	186		-1.60637	0
mp-7171	TmP	0	Fm-3m	m-3m	225	-6.54583	-1.60166	0
mp-510401	GdP	0	Fm-3m	m-3m	225	-11.3407	-1.59603	0
mp-661	AlN	4.0536	P6_3mc	6mm	186	-7.44493	-1.5959	0
mp-1560	LaTe	0	Fm-3m	m-3m	225	-5.63322	-1.59357	0
dft-620	AcN		F-43m	-43m	216		-1.58901	0
dft-024	PrN		P6_3mc	6mm	186		-1.58871	0
mp-414	LuSe	0	Fm-3m	m-3m	225	-5.59355	-1.58512	0
mp-10192	LuP	0	Fm-3m	m-3m	225	-6.55132	-1.5842	0
dft-168	NaI		P6_3mc	6mm	186		-1.58369	0
mp-22916	NaBr	4.0895	Fm-3m	m-3m	225	-3.05725	-1.58268	0
dft-622	LiI		F-43m	-43m	216		-1.58251	0
mp-2423	US	0	Fm-3m	m-3m	225	-8.96112	-1.57892	0
mp-2384	LaP	0	Fm-3m	m-3m	225	-6.74292	-1.56827	0
mp-710	SmP	0	Fm-3m	m-3m	225	-6.63218	-1.5667	0
mp-708	LaAs	0.0414	Fm-3m	m-3m	225	-6.36125	-1.56399	0
mp-931	ThP	0	Fm-3m	m-3m	225	-7.96706	-1.55349	0
mp-2183	ThSe	0	Fm-3m	m-3m	225	-7.00426	-1.54938	0
mp-933	YAs	0	Fm-3m	m-3m	225	-7.11119	-1.5489	0
mp-2288	PuSe	0	Fm-3m	m-3m	225	-10.4282	-1.54638	0
mp-614603	CsI	3.8556	Fm-3m	m-3m	225	-2.75564	-1.54593	0
dft-456	CaBr		Fm-3m	m-3m	225		-1.53711	0.266233
mp-510374	GdAs	0	Fm-3m	m-3m	225	-10.9001	-1.53274	0
mp-2823	NdP	0	Fm-3m	m-3m	225	-6.62308	-1.53237	0
mp-2640	TbAs	0	Fm-3m	m-3m	225	-6.1779	-1.53147	0
dft-594	NiH		F-43m	-43m	216		-1.52891	0
mp-2627	DyAs	0	Fm-3m	m-3m	225	-6.15906	-1.52641	0
mp-1738	SmAs	0	Fm-3m	m-3m	225	-6.20852	-1.52042	0
mp-295	HoAs	0	Fm-3m	m-3m	225	-6.13923	-1.51877	0
mp-1688	ErAs	0	Fm-3m	m-3m	225	-6.12454	-1.51157	0
mp-601	PrP	0	Fm-3m	m-3m	225	-6.60737	-1.51028	0
mp-22903	RbI	3.9541	Fm-3m	m-3m	225	-2.7586	-1.50633	0
mp-2602	NdAs	0	Fm-3m	m-3m	225	-6.21689	-1.50356	0
mp-1865	UN	0	Fm-3m	m-3m	225	-11.1235	-1.50157	0

mp-2154	CeP	0	Fm-3m	m-3m	225	-7.17381	-1.50062	0
mp-1101	TmAs	0	Fm-3m	m-3m	225	-6.06467	-1.49789	0
mp-10622	PrAs	0	Fm-3m	m-3m	225	-6.21116	-1.49146	0
mp-22898	KI	4.0433	Fm-3m	m-3m	225	-2.80816	-1.49096	0
mp-2807	ScP	0	Fm-3m	m-3m	225	-7.36255	-1.48968	0
dft-345	CuF		P6_3mc	6mm	186		-1.48025	0
mp-2017	LuAs	0	Fm-3m	m-3m	225	-6.05823	-1.46851	0
mp-2532	PrTe	0	Fm-3m	m-3m	225	-5.39978	-1.43768	0
mp-2748	CeAs	0	Fm-3m	m-3m	225	-6.72533	-1.42953	0
mp-570	NdTe	0	Fm-3m	m-3m	225	-5.37672	-1.421	0
mp-574283	GdTe	0	Fm-3m	m-3m	225	-10.0211	-1.41142	0
mp-22408	CoO	0.8425	F-43m	-42m	216	-6.15002	-1.41003	0
dft-346	BeS		P6_3mc	6mm	186		-1.4012	0
mp-1753	ThAs	0	Fm-3m	m-3m	225	-7.4323	-1.39611	0
mp-1639	BN	4.6482	F-43m	-43m	216	-8.71146	-1.39553	0.077321
dft-555	GaO		F-43m	-43m	216		-1.39323	0.505245
mp-342	SmTe	0	Fm-3m	m-3m	225	-5.31276	-1.38226	0
mp-1132	CdO	0	Fm-3m	m-3m	225	-3.95664	-1.3807	0
mp-2052	ScAs	0	Fm-3m	m-3m	225	-6.86856	-1.37308	0
dft-513	InO		F-43m	-43m	216		-1.36916	0.302658
mp-1525	CeTe	0	Fm-3m	m-3m	225	-5.90099	-1.3628	0
mp-930	ZrP	0	Fm-3m	m-3m	225	-8.30902	-1.32852	0.045579
dft-609	VH		F-43m	-43m	216		-1.32305	0
mp-569639	TlCl	2.5976	Fm-3m	m-3m	225	-3.12258	-1.31483	0
mp-1708	YTe	0	Fm-3m	m-3m	225	-6.11255	-1.30786	0.013135
dft-490	NiI		F-43m	-43m	216		-1.30782	0
mp-1176	TbTe	0	Fm-3m	m-3m	225	-5.19571	-1.30688	0.002196
mp-2596	NpN	0	Fm-3m	m-3m	225	-11.7559	-1.30581	0
mp-2159	DyTe	0	Fm-3m	m-3m	225	-5.14999	-1.27494	0
mp-1065	LaSb	0	Fm-3m	m-3m	225	-5.79973	-1.26722	0
dft-492	CrH		F-43m	-43m	216		-1.26406	0
mp-13031	MgSe	2.5473	F-43m	-43m	216	-3.7963	-1.2499	0
mp-7592	AgF	0	Fm-3m	m-3m	225	-3.37506	-1.24852	0
mp-1187440	ThCl	0	P6_3mc	6mm	186	-5.57615	-1.2423	0.406614
mp-919	HoTe	0	Fm-3m	m-3m	225	-5.10381	-1.24095	0
mp-1280	ErTe	0	Fm-3m	m-3m	225	-5.06138	-1.20601	0
mp-926	PuP	0	Fm-3m	m-3m	225	-11.0342	-1.19367	0
mp-1785	PrSb	0	Fm-3m	m-3m	225	-5.64016	-1.18522	0
mp-1586	NdSb	0	Fm-3m	m-3m	225	-5.63063	-1.18206	0
dft-541	CoH		F-43m	-43m	216		-1.1739	0
mp-2281	SmSb	0	Fm-3m	m-3m	225	-5.59475	-1.17141	0
mp-237	TmTe	0	Fm-3m	m-3m	225	-4.97273	-1.16355	0.034557

mp-510403	GdSb	0	Fm-3m	m-3m	225	-10.2652	-1.16268	0
mp-215	YSb	0	Fm-3m	m-3m	225	-6.45687	-1.15934	0
mp-10695	ZnS	2.3504	F-43m	-43m	216	-3.50934	-1.14312	0.000387
mp-2724	TbSb	0	Fm-3m	m-3m	225	-5.5224	-1.14071	0
dft-134	BeSe		P6_3mc	6mm	186		-1.13815	0
mp-1007	DySb	0	Fm-3m	m-3m	225	-5.49225	-1.12436	0
dft-115	CoN		P6_3mc	6mm	186		-1.12407	0
dft-105	TaN		P6_3mc	6mm	186		-1.11633	0.23625
mp-1556	PuTe	0	Fm-3m	m-3m	225	-9.81903	-1.11347	0
mp-2050	HoSb	0	Fm-3m	m-3m	225	-5.46163	-1.10593	0
dft-498	AgCl		F-43m	-43m	216		-1.09807	0
mp-1254	LuTe	0	Fm-3m	m-3m	225	-4.91983	-1.08771	0.041976
mp-2191	ErSb	0	Fm-3m	m-3m	225	-5.43441	-1.0862	0
mp-10054	ZrAs	0	Fm-3m	m-3m	225	-7.67651	-1.0734	0.069181
dft-577	CrCl		F-43m	-43m	216		-1.07323	0
mp-627	NpSe	0	Fm-3m	m-3m	225	-9.28458	-1.06274	0
mp-2548	PuAs	0	Fm-3m	m-3m	225	-10.5245	-1.06132	0
mp-2520	TmSb	0	Fm-3m	m-3m	225	-5.36234	-1.06031	0
mp-387	CeSb	0	Fm-3m	m-3m	225	-6.08827	-1.05723	0
mp-1580	NbN	0	Fm-3m	m-3m	225	-10.0796	-1.05269	0.207021
mp-22926	LaBi	0	Fm-3m	m-3m	225	-5.45856	-1.04735	0
mp-15694	PaAs	0	Fm-3m	m-3m	225	-8.13326	-1.04667	0
mp-20300	EuP	0	Fm-3m	m-3m	225	-8.87412	-1.02145	0
mp-516	LuSb	0	Fm-3m	m-3m	225	-5.33927	-1.0143	0
mp-20340	EuN	0	Fm-3m	m-3m	225	-10.1244	-1.00218	0
mp-672	CdS	1.1162	P6_3mc	6mm	186	-3.16201	-0.97241	0
mp-2011	UP	0	Fm-3m	m-3m	225	-9.32015	-0.9678	0
mp-1187584	TmBr	0	P6_3mc	6mm	186	-4.01663	-0.96064	0.260192
mp-23171	PrBi	0	Fm-3m	m-3m	225	-5.2875	-0.95385	0
mp-23300	NdBi	0	Fm-3m	m-3m	225	-5.27111	-0.94383	0
mp-21075	HfC	0	Fm-3m	m-3m	225	-10.5318	-0.93982	0
mp-1213702	Cs2O2	0	Fm-3m	m-3m	225	-3.5076	-0.93707	0.294001
mp-549	ScSb	0	Fm-3m	m-3m	225	-6.16561	-0.93487	0
mp-973976	LuBr	0	P6_3mc	6mm	186	-4.00742	-0.92848	0.280979
mp-2086	PuS	0	Fm-3m	m-3m	225	-9.79248	-0.92207	0.851885
mp-22873	SmBi	0	Fm-3m	m-3m	225	-5.22308	-0.92103	0
dft-045	MnTe		P6_3mc	6mm	186		-0.91003	0
mp-614481	GdBi	0	Fm-3m	m-3m	225	-9.88795	-0.90668	0
mp-1209332	Rb2O2	0	Fm-3m	m-3m	225	-3.51949	-0.90639	0.327169
mp-1751	ThSb	0	Fm-3m	m-3m	225	-6.67349	-0.90206	0

mp-23241	YBi	0	Fm-3m	m-3m	225	-6.068	-0.89175	0
mp-568560	TlBr	2.3494	Fm-3m	m-3m	225	-2.88351	-0.88422	0
mp-2065	MnS	0	Fm-3m	-3m	225	-7.20003	-0.88269	0
mp-1039	MgTe	2.362	P6_3mc	6mm	186	-3.24743	-0.87733	0
mp-22921	TbBi	0	Fm-3m	m-3m	225	-5.13294	-0.87254	0
mp-10108	NpP	0	Fm-3m	m-3m	225	-10.0518	-0.87131	0
mp-21276	PbS	0.9926	Fm-3m	m-3m	225	-4.45755	-0.86476	0
mp-22907	DyBi	0	Fm-3m	m-3m	225	-5.09819	-0.85159	0
mp-32641	LiS	0	Fm-3m	m-3m	225	-3.52651	-0.83555	0.435733
mp-22866	HoBi	0	Fm-3m	m-3m	225	-5.06337	-0.82896	0
dft-616	ErH		F-43m	-43m	216		-0.82867	0
mp-23285	CeBi	0	Fm-3m	m-3m	225	-5.72457	-0.81482	0
mp-978928	SnCl	0	P6_3mc	6mm	186	-3.4329	-0.80983	0.236509
mp-600	YbAs	0	Fm-3m	m-3m	225	-3.90777	-0.80872	0.210412
mp-631	TiC	0	Fm-3m	m-3m	225	-9.36851	-0.8076	0
mp-23245	ErBi	0	Fm-3m	m-3m	225	-5.0322	-0.80528	0
mp-2795	ZrC	0	Fm-3m	m-3m	225	-9.69222	-0.80498	0
mp-2143	USe	0	Fm-3m	m-3m	225	-8.19623	-0.80257	0.146319
dft-489	CuO		F-43m	-43m	216		-0.79633	0.15411
dft-569	RbO		F-43m	-43m	216		-0.78569	0.447876
mp-23234	TmBi	0	Fm-3m	m-3m	225	-4.95601	-0.77528	0
mp-1093	ThGe	0	Fm-3m	m-3m	225	-6.78524	-0.76956	0
dft-427	ZnSe		Fm-3m	m-3m	225		-0.75867	0.192627
dft-621	PmH		F-43m	-43m	216		-0.75298	0
mp-1876	SnS	0.9549	Fm-3m	m-3m	225	-4.4715	-0.73886	0.039185
mp-2390	YbP	0	Fm-3m	m-3m	225	-4.2104	-0.73396	0
dft-437	PrH		Fm-3m	m-3m	225		-0.73058	0
mp-568059	LuBi	0	Fm-3m	m-3m	225	-4.92723	-0.72354	0
mp-2104	UAs	0	Fm-3m	m-3m	225	-8.69607	-0.72111	0
dft-617	TiH		F-43m	-43m	216		-0.72079	0
dft-033	AgBr		P6_3mc	6mm	186		-0.71927	0
dft-425	GaSe		Fm-3m	m-3m	225		-0.71694	0.110238
dft-343	BeTe		P6_3mc	6mm	186		-0.70491	0
mp-1066	YbN	0	Fm-3m	m-3m	225	-5.44982	-0.70379	0.321439
mp-2186	NpAs	0	Fm-3m	m-3m	225	-9.49675	-0.69361	0
dft-019	SnSe		P6_3mc	6mm	186		-0.6867	0
mp-571102	TlI	2.4361	Fm-3m	m-3m	225	-2.62437	-0.68153	0
mp-1058119	KS	0	Fm-3m	m-3m	225	-2.96854	-0.67684	0.58675
mp-22914	CuCl	0.8018	F-43m	-43m	216	-3.35204	-0.67551	0.002367
dft-426	AlP		Fm-3m	m-3m	225		-0.66932	0
mp-804	GaN	1.7376	P6_3mc	6mm	186	-6.15861	-0.66833	0
mp-570632	ScBi	0	Fm-3m	m-3m	225	-5.76866	-0.65922	0

mp-1058171	NaS	0	Fm-3m	m-3m	225	-3.05118	-0.65858	0.505186
mp-1057437	RbS	0	Fm-3m	m-3m	225	-2.87576	-0.649	0.591158
mp-1916	YbSb	0	Fm-3m	m-3m	225	-3.47617	-0.64188	0.261455
mp-2566	PuSb	0	Fm-3m	m-3m	225	-9.83243	-0.63401	0
mp-2691	CdSe	0.5095	F-43m	-43m	216	-2.82901	-0.62795	0
dft-081	CuBr		P6_3mc	6mm	186		-0.60063	0
dft-263	TePb		P6_3mc	6mm	186		-0.59141	0.02536
mp-807	NpTe	0	Fm-3m	m-3m	225	-8.62548	-0.57994	0.041211
mp-1086	TaC	0	Fm-3m	m-3m	225	-11.1182	-0.57594	0
mp-2201	PbSe	0.9081	Fm-3m	m-3m	225	-4.16491	-0.56066	0
mp-2811	MoN	0	P6_3mc	6mm	186	-9.95768	-0.55863	0
dft-071	HgSe		P6_3mc	6mm	186		-0.55256	0
dft-067	HgO		P6_3mc	6mm	186		-0.54726	0.090823
dft-063	CuS		P6_3mc	6mm	186		-0.5155	0
dft-069	HgTe		P6_3mc	6mm	186		-0.51095	0
dft-314	SnTe		P6_3mc	6mm	186		-0.508	0
dft-534	AuCl		F-43m	-43m	216		-0.50629	0.00618
dft-446	PPt		Fm-3m	m-3m	225		-0.50469	0.044531
dft-175	InTe		P6_3mc	6mm	186		-0.50111	0
dft-253	ThC		P6_3mc	6mm	186		-0.49505	0
dft-467	NiC		F-43m	-43m	216		-0.49332	0
mp-1009130	MnN	0.0945	F-43m	-43m	216	-9.04693	-0.48984	0
mp-2172	AlAs	1.6871	F-43m	-43m	216	-4.68665	-0.48461	0
mp-23703	LiH	3.1751	Fm-3m	m-3m	225	-3.05685	-0.48252	0
mp-1058549	CaN	0	Fm-3m	m-3m	225	-5.45419	-0.47822	0.398165
mp-2176	ZnTe	1.0782	F-43m	-43m	216	-2.67468	-0.4733	0
dft-021	GeSe		P6_3mc	6mm	186		-0.47163	0
mp-406	CdTe	1.2285	F-43m	-43m	216	-2.48847	-0.46371	0
mp-1059621	KSe	0	Fm-3m	m-3m	225	-2.76361	-0.46045	0.481981
mp-1123	HgS	0	F-43m	-43m	216	-2.34757	-0.45926	0
mp-910	NbC	0	Fm-3m	m-3m	225	-10.1232	-0.45913	0.040874
mp-567580	PaC	0	Fm-3m	m-3m	225	-9.81356	-0.44284	0
mp-1061530	RbSe	0	Fm-3m	m-3m	225	-2.67309	-0.43486	0.49336
mp-1059357	NaSe	0	Fm-3m	m-3m	225	-2.82638	-0.42232	0.408052
mp-2269	UTe	0	Fm-3m	m-3m	225	-7.63757	-0.42021	0.104995
mp-345	HfB	0	Fm-3m	m-3m	225	-8.73179	-0.4135	0.349909
mp-1282	VC	0	Fm-3m	m-3m	225	-9.56596	-0.41139	0.087833

mp-2490	GaP	1.6843	F-43m	-43m	216	-4.6293	-0.40862	0
mp-1479	BP	1.4662	F-43m	-43m	216	-6.45369	-0.40735	0
dft-443	RhS		Fm-3m	m-3m	225		-0.40123	0.297945
mp-1009489	RbTe	0	Fm-3m	m-3m	225	-2.45446	-0.39254	0.444758
dft-434	GaAs		Fm-3m	m-3m	225		-0.37828	0
mp-451	ZrB	0	Fm-3m	m-3m	225	-7.9873	-0.37375	0.366766
dft-078	CrTe		P6_3mc	6mm	186		-0.36034	0.025508
mp-22879	PuBi	0	Fm-3m	m-3m	225	-9.43407	-0.35695	0.02333
mp-999540	MnSe	0	P6_3mc	6mm	186	-6.68617	-0.35591	0
mp-10728	PdO	0	Fm-3m	m-3m	225	-5.06027	-0.34919	0.449968
dft-030	SnAs		P6_3mc	6mm	186		-0.34249	0
dft-606	HRh		F-43m	-43m	216		-0.33946	0
dft-025	GeTe		P6_3mc	6mm	186		-0.31885	0
dft-358	GaBi		P6_3mc	6mm	186		-0.31457	0
dft-370	InBi		P6_3mc	6mm	186		-0.3112	0
dft-432	GaSb		Fm-3m	m-3m	225		-0.30699	0
dft-073	CeC		P6_3mc	6mm	186		-0.30633	0
dft-429	AlSb		Fm-3m	m-3m	225		-0.2978	0
dft-554	BaN		F-43m	-43m	216		-0.29545	0.355207
dft-473	CuH		F-43m	-43m	216		-0.2856	0
mp-1058586	SrN	0	Fm-3m	m-3m	225	-5.10167	-0.28069	0.434957
mp-22925	AgI	1.7236	F-43m	-43m	216	-2.45889	-0.28062	0
dft-284	SnSb		P6_3mc	6mm	186		-0.26719	0
mp-23870	NaH	4.0702	Fm-3m	m-3m	225	-2.5427	-0.26672	0
mp-24084	KH	3.5362	Fm-3m	m-3m	225	-2.44001	-0.26494	0
mp-972036	WN	0	P6_3mc	6mm	186	-10.7158	-0.26047	0.290304
dft-379	AlBi		P6_3mc	6mm	186		-0.25734	0
dft-381	AgSe		P6_3mc	6mm	186		-0.25413	0
mp-864928	BrCl	1.3584	P6_3mc	6mm	186	-1.69929	-0.2539	0.096014
dft-517	ZnP		F-43m	-43m	216		-0.24194	0
dft-523	MgH		F-43m	-43m	216		-0.23657	0
mp-20305	InAs	0.305	F-43m	-43m	216	-3.93406	-0.22897	0
dft-380	TlAs		P6_3mc	6mm	186		-0.22545	0
mp-24721	RbH	3.2017	Fm-3m	m-3m	225	-2.33381	-0.22367	0
dft-158	ReN		P6_3mc	6mm	186		-0.22251	0
mp-20969	NpSb	0	Fm-3m	m-3m	225	-8.7585	-0.22012	0
dft-488	AgH		F-43m	-43m	216		-0.2125	0
mp-1057286	CsH	2.5878	Fm-3m	m-3m	225	-2.27715	-0.20958	0
mp-20351	InP	0.4664	F-43m	-43m	216	-4.28829	-0.2058	0
mp-567551	SiC	2.0641	P6_3mc	6mm	186	-7.53058	-0.20549	0.000193

dft-383	TlSb		P6_3mc	6mm	186		-0.20337	0
dft-382	SiSn		P6_3mc	6mm	186		-0.19834	0
dft-354	GaGe		P6_3mc	6mm	186		-0.18686	0
dft-448	ZnBi		Fm-3m	m-3m	225		-0.18277	0
mp-568844	BiSe	0	Fm-3m	m-3m	225	-3.86363	-0.17247	0.150584
dft-301	SbPb		P6_3mc	6mm	186		-0.17046	0
mp-519	USb	0	Fm-3m	m-3m	225	-7.8757	-0.1655	0.071134
mp-1184548	HPd	0	P6_3mc	6mm	186	-4.36936	-0.16125	0
dft-344	AgO		P6_3mc	6mm	186		-0.15395	0.298839
mp-22895	CuI	1.1763	F-43m	-43m	216	-2.95842	-0.14681	0.006762
dft-442	AuS		Fm-3m	m-3m	225		-0.1457	0.034067
mp-10020	ScC	0	Fm-3m	m-3m	225	-7.91826	-0.13864	0.308516
mp-1094188	MgSb	0	Fm-3m	m-3m	225	-2.99902	-0.13607	0.16767
dft-153	PtO		P6_3mc	6mm	186		-0.13396	0.58956
mp-2489	UC	0	Fm-3m	m-3m	225	-10.3838	-0.12472	0
dft-414	VI		Fm-3m	m-3m	225		-0.1145	0.441757
mp-11577	YbZn	0	Fm-3m	m-3m	225	-1.50511	-0.10558	0.220531
mp-20012	InSb	0	F-43m	-43m	216	-3.54304	-0.1027	0
dft-133	TcN		P6_3mc	6mm	186		-0.10249	0.070307
mp-567362	BiTe	0	Fm-3m	m-3m	225	-3.61132	-0.09646	0.103454
mp-22205	InN	0.4777	P6_3mc	6mm	186	-5.44244	-0.09036	0
mp-179	YZn	0	Fm-3m	m-3m	225	-3.95126	-0.08849	0.277711
dft-560	IrSe		F-43m	-43m	216		-0.08244	0.173981
mp-1039447	CaZn	0	Fm-3m	m-3m	225	-1.69923	-0.06976	0.179262
dft-061	PuC		P6_3mc	6mm	186		-0.06378	0
mp-1039170	MgBi	0	Fm-3m	m-3m	225	-2.78086	-0.03921	0.127657
dft-572	TiI		F-43m	-43m	216		-0.03253	0.504054
mp-10044	BAs	1.2522	F-43m	-43m	216	-5.68818	-0.01923	0.080167
mp-1002187	TcB	0	Fm-3m	m-3m	225	-8.52751	-0.0075	0.377637
dft-436	SiGe		Fm-3m	m-3m	225		-0.00642	0
mp-475	SnP	0	Fm-3m	m-3m	225	-4.70545	-0.00266	0.043192
dft-401	RhI		Fm-3m	m-3m	225		0.002577	0.384705
mp-1001785	MoP	0	Fm-3m	m-3m	225	-8.11081	0.018647	0.674294
dft-537	TlH		F-43m	-43m	216		0.021902	0.021902
dft-348	BSb		P6_3mc	6mm	186		0.023197	0.023197
dft-552	GaH		F-43m	-43m	216		0.026257	0.026257
dft-539	SiB		F-43m	-43m	216		0.027797	0.05493
dft-451	TlP		Fm-3m	m-3m	225		0.041035	0.055824

mp-998893	YC	0	Fm-3m	m-3m	225	-7.80315	0.043269	0.35898
dft-526	ZnH		F-43m	-43m	216		0.047872	0.047872
dft-144	RuN		P6_3mc	6mm	186		0.04843	0.04843
dft-487	InH		F-43m	-43m	216		0.051501	0.051501
dft-561	BeH		F-43m	-43m	216		0.062723	0.174591
dft-075	NpC		P6_3mc	6mm	186		0.065425	0.065425
mp-2559	SeS	0	Fm-3m	m-3m	225	-3.41636	0.068093	0.384928
dft-520	HPb		F-43m	-43m	216		0.073564	0.073564
dft-596	CdH		F-43m	-43m	216		0.075269	0.075269
mp-10051	TeAs	0	Fm-3m	m-3m	225	-3.82282	0.078087	0.090693
mp-1094232	MgSn	0	Fm-3m	m-3m	225	-2.70999	0.08461	0.227409
mp-22912	NpBi	0	Fm-3m	m-3m	225	-8.3299	0.087199	0.087199
dft-176	OsN		P6_3mc	6mm	186		0.12752	0.12752
dft-565	BC		F-43m	-43m	216		0.131035	0.172019
mp-1094636	MgGa	0	Fm-3m	m-3m	225	-2.17592	0.136569	0.286582
dft-407	RhBr		Fm-3m	m-3m	225		0.137368	0.705538
mp-1009217	MnSb	0	Fm-3m	m-3m	225	-5.738	0.137561	0.924183
mp-1039157	MgCd	0	Fm-3m	m-3m	225	-1.11174	0.139812	0.248779
mp-9935	GeSb	0	Fm-3m	m-3m	225	-4.23059	0.142667	0.142667
mp-614444	ZrZn	0	Fm-3m	m-3m	225	-4.76056	0.143016	0.427602
mp-1009131	MnAs	0	F-43m	-43m	216	-6.20961	0.175721	0.82364
dft-566	GeB		F-43m	-43m	216		0.180417	0.180417
mp-22886	UBi	0	Fm-3m	m-3m	225	-7.39951	0.189406	0.189406
mp-579	CrC	0	Fm-3m	m-3m	225	-9.24527	0.194636	0.278196
mp-1009135	MnP	0	F-43m	-43m	216	-7.09181	0.195689	0.752009
dft-145	RhN		P6_3mc	6mm	186		0.206457	0.206457
mp-2746	MoC	0	Fm-3m	m-3m	225	-9.82317	0.213031	0.297923
dft-140	TcC		P6_3mc	6mm	186		0.215152	0.215152
mp-1094298	SrMg	0	Fm-3m	m-3m	225	-1.41753	0.225655	0.29666
mp-8097	SiP	0	F-43m	-43m	216	-5.18824	0.230102	0.366284
mp-13116	NiN	0	F-43m	-43m	216	-6.6247	0.241452	0.285175
mp-1009753	ScGe	0	F-43m	-43m	216	-5.22451	0.250476	1.043765
mp-8373	GeP	0	F-43m	-43m	216	-4.7523	0.263096	0.263096
dft-152	RuC		P6_3mc	6mm	186		0.266332	0.266332
mp-1097929	HfMg	0	P6_3mc	6mm	186	-5.48428	0.292765	0.30426

dft-573	PbN		F-43m	-43m	216		0.296246	0.296246
dft-180	IrC		P6_3mc	6mm	186		0.311535	0.311535
dft-162	RhC		P6_3mc	6mm	186		0.313222	0.313222
mp-1018145	MnSn	0	Fm-3m	m-3m	225	-6.2573	0.319701	0.319701
mp-1002227	RbSb	0	Fm-3m	m-3m	225	-2.23255	0.322214	0.741339
dft-154	ReC		P6_3mc	6mm	186		0.333311	0.355621
mp-1094402	YMg	0	Fm-3m	m-3m	225	-3.68823	0.343252	0.455783
mp-1184680	HIr	0	P6_3mc	6mm	186	-5.69135	0.347732	0.347732
dft-576	SnB		F-43m	-43m	216		0.397198	0.397198
mp-23722	HCl	0	Fm-3m	m-3m	225	-1.84897	0.397828	1.280381
mp-1009747	ScSn	0	F-43m	-43m	216	-4.75826	0.404118	1.036778
mp-999485	ZnN	0	F-43m	-43m	216	-4.1799	0.426068	0.460503
mp-1002164	GeC	1.8486	F-43m	-43m	216	-6.48738	0.434763	0.434763
mp-1039010	MgAl	0	Fm-3m	m-3m	225	-2.22699	0.444243	0.471202
dft-171	SnC		P6_3mc	6mm	186		0.455886	0.455886
dft-579	BAu		F-43m	-43m	216		0.472288	0.472288
mp-1183947	CsAc	0	P6_3mc	6mm	186	-2.02664	0.481653	0.481653
mp-1009752	ScSi	0	F-43m	-43m	216	-5.38563	0.492302	1.329897
mp-2765	PuB	0	Fm-3m	m-3m	225	-9.97161	0.501997	0.817701
dft-177	OsC		P6_3mc	6mm	186		0.51217	0.51217
dft-580	RuI		F-43m	-43m	216		0.519405	0.519405
mp-1008987	IrN	0	F-43m	-43m	216	-7.83071	0.564739	0.564739
mp-1183956	CsLa	0	P6_3mc	6mm	186	-2.34566	0.570045	0.570045
dft-416	IrI		Fm-3m	m-3m	225		0.5834	0.877371
mp-1180622	KP	0	Fm-3m	m-3m	225	-2.66557	0.596269	1.008379
dft-183	PtC		P6_3mc	6mm	186		0.609165	0.609165
dft-588	GaC		F-43m	-43m	216		0.626979	0.626979
mp-1179679	RbP	0	Fm-3m	m-3m	225	-2.56947	0.627444	0.987931
mp-1039270	CeMg	0	Fm-3m	m-3m	225	-3.12225	0.64274	0.659219
dft-149	CuN		P6_3mc	6mm	186		0.656091	0.656091
mp-10143	TiB	0	F-43m	-43m	216	-6.56671	0.720507	1.55278

mp-1017982	TlN	0	P6_3mc	6mm	186	-4.43562	0.721442	0.721442
dft-167	PtN		P6_3mc	6mm	186		0.724926	0.724926
dft-150	CdN		P6_3mc	6mm	186		0.773812	0.773812
mp-974469	ReTe	0	P6_3mc	6mm	186	-7.0092	0.784718	0.900916
mp-23903	HBr	0	Fm-3m	m-3m	225	-1.62917	0.80916	1.137211
mp-1059264	KGe	0	Fm-3m	m-3m	225	-2.04078	0.823171	1.027876
mp-1006880	BiB	0.3557	F-43m	-43m	216	-4.45913	0.823773	0.823773
mp-567903	ZnNi	0	F-43m	-43m	216	-2.69214	0.827498	1.082531
mp-973835	PdN	0	F-43m	-43m	216	-5.70522	0.859264	0.859264
dft-184	PdC		P6_3mc	6mm	186		0.890923	0.890923
mp-1060680	RbGe	0	Fm-3m	m-3m	225	-1.90298	0.896048	1.076548
mp-1059612	LiN	0	F-43m	-43m	216	-3.94132	0.989376	1.430361
dft-597	RbN		F-43m	-43m	216		1.021872	1.308242
dft-591	HgN		F-43m	-43m	216		1.051921	1.051921
mp-1009221	NaN	0	Fm-3m	m-3m	225	-3.55008	1.082269	1.343217
dft-592	InC		F-43m	-43m	216		1.08998	1.08998
dft-593	PbC		F-43m	-43m	216		1.125117	1.125117
mp-1213965	CaC	0	Fm-3m	m-3m	225	-4.48029	1.132822	1.132822
mp-1058689	KN	0	Fm-3m	m-3m	225	-3.35891	1.172527	1.462589
mp-1058944	KSi	0	Fm-3m	m-3m	225	-2.09142	1.175475	1.241569
dft-590	BaC		F-43m	-43m	216		1.23853	1.288332
mp-1208639	SrC	0	Fm-3m	m-3m	225	-4.1888	1.269327	1.294735
mp-974474	ReBr	0	P6_3mc	6mm	186	-5.72937	1.311354	1.509988
dft-185	AuN		P6_3mc	6mm	186		1.342381	1.342381
dft-178	AgN		P6_3mc	6mm	186		1.385317	1.385317
dft-600	AuC		F-43m	-43m	216		1.528979	1.528979
dft-191	IN		P6_3mc	6mm	186		1.601441	1.766276
dft-192	BrN		P6_3mc	6mm	186		1.81606	1.81606
dft-601	NCl		F-43m	-43m	216		1.836267	1.871087
mp-1056860	NaB	0	Fm-3m	m-3m	225	-2.14339	1.852412	1.887897
dft-602	HgC		F-43m	-43m	216		1.927244	1.927244
mp-1006887	CdC	0	Fm-3m	m-3m	225	-3.08566	1.980821	1.980821

mp-1059907	KB	0	Fm-3m	m-3m	225	-1.80933	2.085562	2.100309
mp-567465	TlC	0	Fm-3m	m-3m	225	-3.68762	2.106589	2.106589
mp-1060982	CsB	0	Fm-3m	m-3m	225	-1.64125	2.146142	2.146142
mp-1061023	RbB	0	Fm-3m	m-3m	225	-1.65765	2.172314	2.172314
mp-1008653	AgC	0	Fm-3m	m-3m	225	-3.79414	2.23551	2.23551
mp-1057017	KC	0.6809	Fm-3m	m-3m	225	-2.85234	2.316242	2.334293

8. Pairs found in pseudosymmetry analysis

material	Pretty Formula	Polar e above hull
mp-1025567	NbCl2O	0
mp-1078724	LiAsSe2	0
mp-1210002	NaNbS6	0
mp-1216843	V(CrS2)2	0
mp-1221454	Na(VSe2)2	0
mp-15573	Tl3AsO4	0
mp-22965	BiTeI	0
mp-23360	Zn(IO3)2	0
mp-23400	In(IO3)3	0
mp-25250	MoP2O7	0
mp-29146	VSbO4	0
mp-34857	KNO2	0
mp-36262	BrO2F	0
mp-36383	Ag3(PO4)2	0
mp-36383	Ag3(PO4)2	0

mp-38322	Mg(IO3)2	0
mp-504482	AlInS3	0
mp-504951	InGaS3	0
mp-504952	InGaSe3	0
mp-546285	NbI3O	0
mp-549720	NbI2O	0
mp-550070	NbBr2O	0
mp-554980	Mn(IO3)2	0
mp-555874	LiAsS2	0
mp-555903	Al(IO3)3	0
mp-555903	Al(IO3)3	0
mp-561104	Ga(IO3)3	0
mp-561104	Ga(IO3)3	0
mp-5709	Tl3PO4	0
mp-572672	Co(IO3)2	0
mp-752467	NbO2F	0
mp-752955	WOF4	0
mp-752955	WOF4	0
mp-755246	Ni(IO3)2	0
mp-757594	Os(OF2)2	0
mp-760343	FePO4	0
mp-766046	Tl(WO3)3	0
mp-777786	Na2CoO3	0
mp-862787	AlInSe3	0

mp-1223488	KIn5S8	0.000242419
mp-15896	Cu2SiSe3	0.000559023
mp-624234	TePb2O5	0.000691704
mp-29959	TlBSe3	0.000805594
mp-1223478	KCr5Se8	0.000832671
mp-1196124	BaGe4O9	0.000844464
mp-36066	PNO	0.000926678
mvc-11360	BiOF	0.001099528
mvc-11360	BiOF	0.001099528
mp-752852	BaTi8O16	0.001140446
mp-757607	Na(WO3)10	0.001806756
mp-25294	VP2O7	0.00182687
mp-1216612	TlIn5S8	0.001987981
mp-4342	KNbO3	0.002133648
mp-556916	GaAgS2	0.002607307
mp-5724	AlPO4	0.0030932
mp-19903	V2Pb3O8	0.00335464
mp-558582	In(PO3)3	0.003892088
mp-1003318	CaMn16O32	0.004897406
mp-676586	Na(VS2)2	0.00574272
mp-29299	TeSeO4	0.006055697
mp-1003549	LiMn16O32	0.006327228
mp-26630	TiP2O7	0.006500262
mp-1225738	CuBi5S8	0.006662489

mp-753203	VO2F	0.006718884
mp-1216689	TlCr5Te8	0.007917644
mp-1281596	VOF2	0.007925496
mp-752561	SbO2F	0.008646286
mp-22955	LiIO3	0.009031907
mp-22955	LiIO3	0.009031907
mp-765995	Ba(WO3)6	0.00906744
mvc-3279	VPO5	0.009161424
mp-15895	Cu2SiS3	0.009571319
mp-1217790	Ta(Mo2S5)2	0.010963415
mp-1223833	K(WO3)3	0.011855273
mp-32509	V(PO3)4	0.012491394
mp-559112	K2Zn6O7	0.012927885
mp-1221958	MgZn3O4	0.013224873
mp-762225	BaCO3	0.013232805
mp-775277	K(WO3)3	0.014172594
mp-775277	K(WO3)3	0.014172594
mp-1283877	Co2As2O7	0.014789272
mp-1225994	CoTeSe	0.014912353
mp-1220404	NbMoS4	0.015156199
mp-761390	TaFeO4	0.015948565
mp-1223529	KCu7S4	0.016109902
mp-546787	KNO2	0.016135897
mp-756368	Te(WO4)3	0.016841241

mp-26982	MnP2O7	0.016884342
mp-1076718	Sr2Mn2O5	0.018192594
mp-1076718	Sr2Mn2O5	0.018192594
mp-1076718	Sr2Mn2O5	0.018192594
mp-1217910	TaMoS4	0.019086662
mp-1226512	Co(NiSe2)2	0.019218735
mp-1272458	NbVO4	0.019492711
mp-1293437	LiMn4O8	0.020016263
mp-26842	Ni(PO3)3	0.020660499
mp-1288892	NaV2O4	0.020722618
mvc-5693	MoWO6	0.020773747
mp-1003637	LiMn3O6	0.021324518
mp-558497	Cr(PO3)3	0.021608542
mp-25542	LiMnO2	0.02195722
mp-10444	Ca2Al2O5	0.022727661
mp-10444	Ca2Al2O5	0.022727661
mp-1178789	V2ZnO6	0.02287634
mp-29227	TlV5S8	0.023009824
mp-1276433	TiVO4	0.023760478
mp-1276433	TiVO4	0.023760478
mp-1276433	TiVO4	0.023760478
mp-753525	VOF3	0.024944593
mp-1295695	NaCo2O3	0.025101599
mp-1277859	LiMn3O4	0.025965061

mp-675020	Li4MoO5	0.026240082
mp-675020	Li4MoO5	0.026240082
mp-1272170	TiVO4	0.026465839
mp-32028	Mn(PO3)4	0.027779975
mp-1178157	K(WO3)3	0.028392356
mp-1178157	K(WO3)3	0.028392356
mp-1178157	K(WO3)3	0.028392356
mp-1178823	V2CoO6	0.029015417
mp-20435	V(PO3)3	0.029048307
mp-773918	Zn(Bi19O30)2	0.029131098
mp-35320	Ag3(AsO4)2	0.029400737
mp-35320	Ag3(AsO4)2	0.029400737
mp-11000	CaB4O7	0.030734596
mp-754937	LiTi3O6	0.031320743
mp-26252	VP2O7	0.032697511
mp-13108	SrHfO3	0.032896417
mp-26342	SbP2O7	0.034499807
mp-1299651	LiMn3O4	0.034873808
mp-1299651	LiMn3O4	0.034873808
mp-755562	LiMnO2	0.035489804
mp-1280914	MnSbO4	0.035535197
mp-690542	Li4WO5	0.036121758
mp-690542	Li4WO5	0.036121758

mp-1099761	Sr2Cu2O5	0.036180643
mp-1099761	Sr2Cu2O5	0.036180643
mp-19376	TiMnO3	0.036319113
mp-19376	TiMnO3	0.036319113
mp-25587	LiNiO2	0.037632317
mp-1284090	NaNi5O6	0.0379615
mp-38975	PNO	0.039006336
mp-38975	PNO	0.039006336
mp-38975	PNO	0.039006336
mp-996960	AgNO2	0.03960236
mp-26294	MoP2O7	0.039793518
mp-675977	AgBiS2	0.045538545
mp-997022	AgAsO2	0.046350603
mp-756035	LiTi4O8	0.046721848
mp-1068742	SrZrO3	0.048364574
mp-504341	Sb(PO3)4	0.051654048
mp-753367	WO2F	0.052883073
mp-760367	VOF3	0.053653695
mp-760367	VOF3	0.053653695
mp-1182604	BaV13O18	0.054563166
mp-22995	Ag3SI	0.054778279
mp-22995	Ag3SI	0.054778279
mp-22995	Ag3SI	0.054778279
mp-26974	CrP2O7	0.055054524

mp-1222220	Mg(Mo3Se4)2	0.05562031
mp-756084	BaNa4O3	0.055910146
mp-504928	NaCu3Te2	0.057171475
mvc-3271	CrPO5	0.057258622
mp-761177	VOF3	0.058068043
mp-754840	TaAsO4	0.058884394
mp-27032	MnP2O7	0.062028781
mp-1216724	V(FeS2)2	0.062098277
mp-773030	ScP2O7	0.062323723
mvc-12181	Ca(CoO2)4	0.063257888
mp-761589	NbCo3O8	0.063888265
mp-674350	TiPb9O11	0.064662578
mp-1220110	Ni(Mo3Se4)2	0.06537754
mp-676400	TlNO2	0.066425597
mp-1105293	GaAgO2	0.067857991
mp-1105293	GaAgO2	0.067857991
mp-1282563	NaCo2O3	0.067953305
mp-1076014	Ca2Mn2O5	0.069416643
mp-1076014	Ca2Mn2O5	0.069416643
mp-1076014	Ca2Mn2O5	0.069416643
mp-1076048	Na2V2O5	0.07090165
mp-1076048	Na2V2O5	0.07090165
mp-1304173	NaV5O8	0.071277961

mp-752918	NaTi6O12	0.071438871
mp-762679	V(TeO3)4	0.072709938
mp-27044	CoP2O7	0.074534391
mp-1225174	Fe(CoSe2)2	0.075360299
mp-998350	BaZrSe3	0.077542493
mp-1076454	Sr2Ti2O5	0.078962611
mp-1076454	Sr2Ti2O5	0.078962611
mp-675830	FeCu5S4	0.079096446
mvc-7384	CaCr2O7	0.079403461
mp-1178444	Cr(WO4)3	0.08007118
mp-690525	MnCO3	0.081420198
mp-31572	Co(PO3)4	0.082779386
mp-1105075	Al(Si2O5)2	0.084294281
mp-1016144	MgMn3O6	0.084670592
mvc-12238	CaV4O8	0.085101505
mp-1216404	VCl2O	0.085236007
mp-1293177	TiCo3O8	0.085461822
mp-1101449	NbV3O8	0.086131598
mp-770104	LiMn4O8	0.086779177
mp-770104	LiMn4O8	0.086779177
mp-762371	NbOF3	0.086971215
mp-1262087	Al(BiO3)3	0.088793085
mp-1287027	Na(CoO2)3	0.089615542
mp-769218	Ni(PS3)2	0.093298813

mp-756090	BaNa4O3	0.093807286
mvc-12176	CaTi4O8	0.097156701
mp-23069	Fe(IO3)3	0.097462248
mvc-2157	MgV2O6	0.098058858
mp-540424	Cr(PO3)4	0.098357786
mvc-12450	MgTi4O8	0.09840394
mp-1097775	Ba2Cu2O5	0.101116902
mp-1097775	Ba2Cu2O5	0.101116902
mp-765705	NaV5O8	0.102067568
mp-753365	NaV3O6	0.103422587
mp-705019	Cr(PO3)5	0.104996032
mp-26996	CoP2O7	0.10610935
mp-634755	CSO	0.106332544
mp-1296090	TiFeO4	0.106948647
mp-754223	Ta2CuO6	0.106987258
mp-1078919	CsV2O5	0.107120418
mp-1100345	CaSnS3	0.108729064
mp-1100345	CaSnS3	0.108729064
mp-675078	Ba(FeS2)2	0.109016984
mvc-12608	CaMn4O8	0.109917701
mp-759588	Cr(Bi7O12)2	0.111174106
mp-1288520	NaCoO2	0.111866024
mp-1288520	NaCoO2	0.111866024
mp-1288520	NaCoO2	0.111866024

mp-1288520	NaCoO2	0.111866024
mp-1180693	Li(SeO3)2	0.113488357
mvc-2851	Ca(BiO2)4	0.114629319
mp-753154	Li(FeO2)3	0.116453384
mp-1180273	Na(SeO3)2	0.1175313
mp-12681	LiMnTe2	0.120534081
mp-26364	Ni(PO3)4	0.121030308
mvc-12565	Mg(CoO2)4	0.124679447
mvc-10702	Ca(BiO2)4	0.124786477
mp-1225290	Fe(Mo3Se4)2	0.124912313
mvc-10720	Mg(BiO2)4	0.125520394
mvc-7295	Zn(SbO2)2	0.127532171
mp-1080068	CuCO3	0.12931978
mp-1080068	CuCO3	0.12931978
mp-1101465	VCo3O8	0.130428
mvc-12617	Ca(NiO2)4	0.133135594
mp-1076548	Ca2Cu2O5	0.137984259
mp-1076548	Ca2Cu2O5	0.137984259
mp-1076548	Ca2Cu2O5	0.137984259
mp-998363	NiBiO3	0.138744354
mp-1235981	LiNi9O13	0.139593668
mp-27249	CuCO3	0.140410082
mp-1176498	MgFe2O5	0.141107465
mp-27036	CrP2O7	0.141790084

mp-1099875	Na2Mo2O5	0.142420428
mp-1099875	Na2Mo2O5	0.142420428
mp-1217360	Ti(CrS2)2	0.145203506
mvc-2898	Zn(SnO2)4	0.145264882
mvc-2854	Mg(BiO2)4	0.14713087
mvc-2868	Zn(BiO2)4	0.1509015
mp-1076215	K2Nb2O5	0.151590912
mp-1076215	K2Nb2O5	0.151590912
mp-1105229	GaCuO2	0.152009084
mp-1105229	GaCuO2	0.152009084
mvc-10708	Ca(SnO2)4	0.154430958
mp-1208353	Ti(PO4)2	0.155069981
mp-753614	Fe(PS3)2	0.156371309
mp-753614	Fe(PS3)2	0.156371309
mp-753614	Fe(PS3)2	0.156371309
mp-1233601	MgNi9O13	0.157180534
mp-1099805	Mg2Mn2O5	0.157188453
mp-1099805	Mg2Mn2O5	0.157188453
mp-1099805	Mg2Mn2O5	0.157188453
mvc-2834	Mg(SnO2)4	0.157473017
mp-1016145	CaMn3O6	0.157783679
mp-36890	TaBi3O7	0.160553107
mp-25246	FeP2O7	0.161081133
mp-632694	ScMo3O8	0.161570929

mp-632694	ScMo3O8	0.161570929
mp-27754	WCl4O	0.164676256
mp-27754	WCl4O	0.164676256
mp-1235379	LiNi9O13	0.169273352
mp-1079620	Cr(ClO)2	0.173497272
mp-1079620	Cr(ClO)2	0.173497272
mp-1076527	Mg2Cu2O5	0.173803873
mp-1076527	Mg2Cu2O5	0.173803873
mp-1076527	Mg2Cu2O5	0.173803873
mp-753112	LiMn3O6	0.174223869
mvc-12570	Mg(NiO2)4	0.17469891
mvc-10691	Zn(SnO2)4	0.176061749
mp-1075952	Na2Ta2O5	0.176107035
mp-1075952	Na2Ta2O5	0.176107035
mvc-667	Te(WO4)2	0.185274157
mvc-2858	Mg(SbO2)4	0.185909223
mvc-12231	MgV4O8	0.186043768
mvc-5774	CrWO6	0.186366609
mvc-2103	Mg(MoO3)2	0.187093396
mp-1099683	Na2Cr2O5	0.187391077
mp-1099683	Na2Cr2O5	0.187391077
mvc-740	Ta(WO4)2	0.188029195
mp-31703	CrOF3	0.189398073
mp-1076323	Rb2Nb2O5	0.190528917

mp-1076323	Rb2Nb2O5	0.190528917
mp-1096805	AgBrO2	0.192991708
mvc-10699	Ca(SbO2)4	0.195264854
mp-1229330	Bi(W3O10)6	0.198836676
mvc-12580	MgMn4O8	0.199085636
mvc-2162	Zn(MoO3)2	0.199181941
mp-26182	FeP2O7	0.202816948
mp-585418	Fe(PO3)4	0.210075401
mvc-175	Zn(CoO2)4	0.214477284
mp-1244584	CrSiO5	0.215387236
mp-26008	NiP2O7	0.217281251
mp-1076354	Rb2V2O5	0.235041812
mp-1076354	Rb2V2O5	0.235041812
mvc-12609	Zn(NiO2)4	0.245884151
mp-1228770	Au(OF3)2	0.247249522
mp-1099786	Ca2Co2O5	0.24780621
mp-1099786	Ca2Co2O5	0.24780621
mp-1099786	Ca2Co2O5	0.24780621
mp-770187	Cr3TeO8	0.251018822
mp-1097724	Na2W2O5	0.25238501
mp-1097724	Na2W2O5	0.25238501
mp-996984	AgClO2	0.252788259
mvc-3085	Zn(MoO2)4	0.253799309
mp-1076652	Mg2Ti2O5	0.257477167

mp-1076652	Mg2Ti2O5	0.257477167
mp-1224511	K(Mo2O3)4	0.259807669
mp-1224511	K(Mo2O3)4	0.259807669
mp-1224511	K(Mo2O3)4	0.259807669
mp-997026	CuClO2	0.262947456
mp-997026	CuClO2	0.262947456
mp-1244739	MgCo4O9	0.276180491
mp-25260	FeAs2O7	0.282398423
mp-26232	NiP2O7	0.284971764
mp-1244857	Ba2Ti3O7	0.28595912
mvc-5756	CaWO3	0.286731635
mvc-12629	Zn(FeO2)4	0.296635564
mvc-14274	ZnCr4O8	0.297003019
mvc-3110	Zn(WO2)4	0.30063536
mp-998781	TlNiO3	0.304148638
mp-1212414	In(PO4)3	0.310164145
mp-1212414	In(PO4)3	0.310164145
mp-1215996	Zr2P4O17	0.33073998
mp-1075956	K2Mo2O5	0.333335633
mp-1075956	K2Mo2O5	0.333335633
mvc-10669	Mg(SnO2)4	0.33814102
mp-1100988	TlSbSe2	0.340508396
mp-1217487	Te(AsO2)2	0.342802745
mp-675198	BaW6O19	0.342919511

mp-1066872	InOF2	0.344649463
mp-1066872	InOF2	0.344649463
mp-574250	InSiTe3	0.348818967
mp-540440	FeP2O7	0.350291881
mvc-10695	Zn(SbO2)4	0.352913222
mvc-157	Mg(WO2)4	0.37392081
mp-780759	NaV13O18	0.379669061
mp-1099822	Rb2Mo2O5	0.387513587
mp-1099822	Rb2Mo2O5	0.387513587
mvc-12611	CaCr4O8	0.391453028
mp-1226409	Cs3Cl5O	0.391872993
mp-1226409	Cs3Cl5O	0.391872993
mp-1102157	NaNi3O8	0.408893292
mp-1189919	LiPO4	0.413606437
mp-1221810	Na(Mo2O7)2	0.416245397
mp-1103574	NaIO4	0.424288638
mp-1076401	Rb2Cr2O5	0.424938352
mp-1076401	Rb2Cr2O5	0.424938352
mp-1105595	Zn2SiO5	0.435299269
mvc-14243	CaCoO3	0.440541247
mvc-165	Ca(WO2)4	0.457003681
mvc-12639	Mg(FeO2)4	0.468781561
mvc-5801	MgWO3	0.475680957
mp-1097917	TlNiO3	0.500560182

mp-1194474	BaCl2O11	0.509572504
mp-1076319	Ca2Fe2O5	0.509618321
mp-1076319	Ca2Fe2O5	0.509618321
mp-1076391	K2W2O5	0.514713612
mp-1076391	K2W2O5	0.514713612
mp-1179308	Sr(SO5)2	0.517965501
mp-1182310	Ca(SO5)2	0.522239256
mp-1214392	Ca(SO5)2	0.54967312
mp-1180205	NaAsO6	0.567159309
mp-1075937	Rb2W2O5	0.573195842
mp-1075937	Rb2W2O5	0.573195842
mp-1234978	MgNi9O13	0.573761816
mp-1101800	MgSO9	0.581118807
mvc-12189	Ca(FeO2)4	0.597596217
mp-1102637	MgSeO9	0.599917362
mp-1227738	CaV8O24	0.612393064
mvc-11882	Al(CuO2)3	0.629696608
mp-1003766	NaMn2O4	0.670007164
mp-1180304	NClO3	0.711512648
mp-1214130	CaAsO6	0.719420427
mp-1192328	Fe(CO)5	0.722052634
mp-36875	Na(CoO2)3	0.820614754
mp-1101924	NiSO9	0.861449201
mp-758077	Li(Fe2O3)4	0.937318771

mp-1076932	BaTiO3	0.979015368
mp-1080550	KO2F	0.997235746
mp-1199189	SN8O13	1.342532657
mvc-12581	MgCr4O8	1.363799782

9. Jupyter Notebooks

Code will be available at https://github.com/jokenfuss/Ferroelectrics

10. Vasp Settings:

VASP Settings: For both relaxation and static, the following settings were used in Pymatgen's MPRelaxSet or MPStaticSet [59] to set the INCAR and KPOINTS files for VASP:

ALGO=Normal – This selects the algorithm in use to optimize the orbitals. The 'Normal' algorithm is the default in VASP. It is the blocked-Davidson scheme, which optimizes a set of bands simultaneously.

ISYM=2 – This enables the use of an efficient, memory conserving symmetry to reduce memory requirements.

ICHARG=0 – This calculates charge densities from initial wave functions.

NCORE= 16 – Sets the number of compute cores used for each orbital, parallelizing the work to improve performance.

EDIFFG= -0.001 – Defines the break condition for relaxation as when the norm of all forces is smaller than 0.001 eV/Ang.

EDIFF=1E-7 – Similar to EDIFFG, relaxation stops if the total energy change and the band-structure-energy change between steps are smaller than 1E-7 eV. This is a high-precision setting that slows convergence at the expense of greater accuracy.

ENCUT= 520 – Another cutoff, this one specifies the cutoff energy for the plane-wave-basis set at 520 eV.

NELM=100 – This sets the maximum number of electronic self-consistency steps at 100. This is higher than the default of 60 to give a greater chance of marginal setups converging.

SIGMA=0.05 – This specifies the width of smearing as 0.05 eV.

ISMEAR= 0 – This specifies that Gaussian smearing should be done. This is a reasonable default for systems where there is no a priori knowledge of the final type (insulator, metal, etc.) of the system.

NEDOS=2001 – Sets the number of grid points for evaluation of the density of states.

ADDGRID=TRUE – adds an additional support grid for evaluation of augmentation charges to improve accuracy of forces.

ISPIN=2 – This performs spin polarized calculations.

PREC=Accurate – Sets the general precision mode as “accurate,” the highest supported precision in VASP.

For relaxation, the following settings were used:

LSCALAPACK=FALSE – controls use of scalapack

ISIF=3 – Allows for all degrees of freedom, allowing for changes of position, cell shape, and cell volume.

NSW=600 – This sets the maximum number of ionic steps to 600.

user_kpoints_settings= {"grid_density": 10000} – This sets the kpoints mesh for the calculation. It uses a Gamma-centered Monkhorst-Pack approach.

For static calculations, the following settings were used:

LWAVE=TRUE – Writes wave functions to file at the end of the run.

LCHARG=TRUE – Writes charge densities to file at the end of the run.

IBRION= -1 – This calculation does not move the ions, but still performs the outer loops of NSW.

ISIF=2 – Calculates the stress tensor and allows for change of position, but not change of cell shape or cell volume

NSW= 1 – This sets the program to only run a single ionic step.

user_kpoints_settings = {"grid_density": 16000}

[15,60]