Machine-learning-accelerated high-throughput materials screening: Discovery of novel quaternary Heusler compounds

Kyoungdoc Kim, Logan Ward, Jiangang He, Amar Krishna, Ankit Agrawal, and C. Wolverton

Phys. Rev. Materials **2**, 123801 — Published 4 December 2018

DOI: 10.1103/PhysRevMaterials.2.123801

Machine Learning Accelerated High-Throughput Materials Screening: Discovery of Novel Quaternary Heusler Compounds

Kyoungdoc Kim[a, 1], Logan Ward[a, b, 1], Jiangang He[a], Amar Krishna[c], Ankit Agrawal[c], C. Wolverton[a, *]

[a]Department of Materials Science and Engineering, Northwestern University, Evanston, Illinois 60208, United States

[b]Computation Institute, University of Chicago, Chicago, Illinois 60637, United States

[c]Department of Electrical Engineering and Computer Science, Northwestern University, Evanston, Illinois 60208, United States

[1]Authors contributed equally

[*]E-mail: c-wolverton@northwestern.edu

Abstract

Discovering novel, multicomponent crystalline materials is a complex task owing to the large space of feasible structures. Here, we demonstrate a method to significantly accelerate materials discovery by using a machine learning (ML) model trained on Density Functional Theory (DFT) data from the Open Quantum Materials Database (OQMD). Our ML model predicts the stability of a material based on its crystal structure and chemical composition, and we illustrate the

effectiveness of the method by application to finding new Quaternary Heusler (QH) compounds. Our ML-based approach can find new stable materials at a rate 30 times faster than undirected searches and we use it to predict 55 previously- unknown, stable QH compounds. We find the accuracy of our ML model is higher when trained using the diversity of crystal structures available in the OQMD than when training on well-curated datasets which contain only a single family of crystal structures (i.e., QHs). The advantage of using diverse training data shows how large datasets, such as OQMD, are particularly valuable for materials discovery and that we need not train separate ML models to predict each different family of crystal structures. Compared to other proposed ML approaches, we find that our method performs best for small ($<10^3$) and large ($>10^5$) training set sizes. The excellent flexibility and accuracy of the approach presented here can be easily generalized to other types of crystals.

1. Introduction

Since Fritz Heusler discovered the Cu_2MnAl phase in 1903 [1], the Heusler structure ($Fm\bar{3}m$, X_2YZ) has been widely used in a variety of applications, such as for superconductors, thermoelectrics, shape memory alloys, etc. [2] Beyond the original "full" Heusler structure, the Heusler family contains half Heusler (F-43m, XYZ), inverse Heusler (F-43m, XY_2Z), and quaternary Heusler (F-43m, XX'YZ) [3], which all are decorations of a body-centered-cubic lattice. The larger space of possible numbers of compounds afforded by the fourth element in the Quaternary Heusler structure makes these materials particularly interesting for materials design. It is very likely to find QHs that exhibit superior properties than their 3-component analogues (e.g. ultralow thermal conductivity and high dielectric constants [4, 5]) provided this additional

degree of freedom. However, the large possible space of QHs (larger than 3.2 million compounds) frustrates materials discovery by experimental synthesis and characterization.

High-Throughput (HT) atomistic Density Functional Theory (DFT) calculations have proven to be powerful tools for discovering new crystalline compounds [6-10]. For example, Kirklin et al. [11] performed high-throughput DFT calculations to predict potential compounds used as effective strengthening precipitates for a variety of different alloys matrices using the Open Quantum Materials Database (OQMD) [8, 12]. Aykol et al. [13] screened materials in the OQMD to find hydrofluoric-acid barrier coatings for Li-ion batteries using high-throughput DFT. Balluff et al. [14] also performed high-throughput DFT calculations in order to search new antiferromagnetic ternary Heusler compounds through the automatic flow for materials discovery library (AFLOWLIB) library. However, even high-throughput DFT searches have limitations. Over 3 million unique QH compounds can be formed from the 73 metallic elements and evaluating all of them is clearly impractical, as that search would be larger than the 3 largest DFT databases put together. What we need is a tool to quickly identify the materials that are the most likely to be stable.

As shown by many recent studies, machine learning algorithms offer the ability to quickly identify promising materials in intractable search spaces [15-21]. Machine learning models are composed of three parts: (i) a resource of training data, (ii) a representation to convert that materials data into a form compatible with ML, and (iii) a ML algorithm to learn a function that describes how the representation relates to the material properties described in the training set. Besides wide applications of ML in chemical reaction[19, 21], drug research[20], material band

gap[22], and atom potential[23], several examples illustrate the successful use of DFT energetic data to construct ML models, and the use of these methods to find new materials. For example, Curtarolo et al. [24] predicted new binary compounds using a partial least squares regression model based on correlations among *ab initio* energies of structures from a pre-defined library of structure types. Meredig et al. [25] developed a decision tree model based on over 15,000 DFT energies from OQMD to predict the formation energy of a material based on its composition and used it to identify 4500 new ternary compounds. More recently, Faber et al. [26] employed ML to identify 100 new Elpasolite compounds out of a space of millions of possible candidates using a kernel-ridge-regression trained the energies of only 10^4 Elpasolite structures. These studies, among several others [27-42], demonstrate the promise of ML in accelerating the search for new materials.

In this work, we create, validate, and benchmark a machine learning model for identifying new Quaternary Heusler materials. Our approach is distinct from previous examples in that we utilize the entire OQMD for training, which is an order of magnitude larger dataset and includes a wide diversity of structure types, not merely the Heusler structure. Our model works by training a model on the entire OQMD using the Voronoi tessellation method of Ward *et al*. [27] We then employed our model to predict the stability of all 3.2 million possible structures, and identified 303 compositions that are especially likely to yield stable QH compounds. We confirm that 55 are indeed stable through DFT calculaitons, and hence represent predictions of new, T=0K stable compounds that are promising for experimental synthesis. Lastly, we compare our method against other techniques in the literature, and also investigate the role of the training set data in the accuracy of the resulting ML model. We find that the models, perhaps unsurprisingly, get

more accurate as more QH structures are used in the training set, but more surprisingly find that the models also become more accurate as more non-QH data (i.e., data for other structures types) is used in training. We also devise a general workflow for selecting the best ML strategies depends on training data set size. We propose that this workflow can be used to accelerate the discovery of crystalline compounds for a broad variety of different crystal structures and compositions of materials.

2. Methodology – Constructing the Machine Learning Model

Machine learning models are composed of training data, a representation, and a machine learning algorithm. Here, we describe each of these three components of our machine learning model in turn.

2.1. Training data

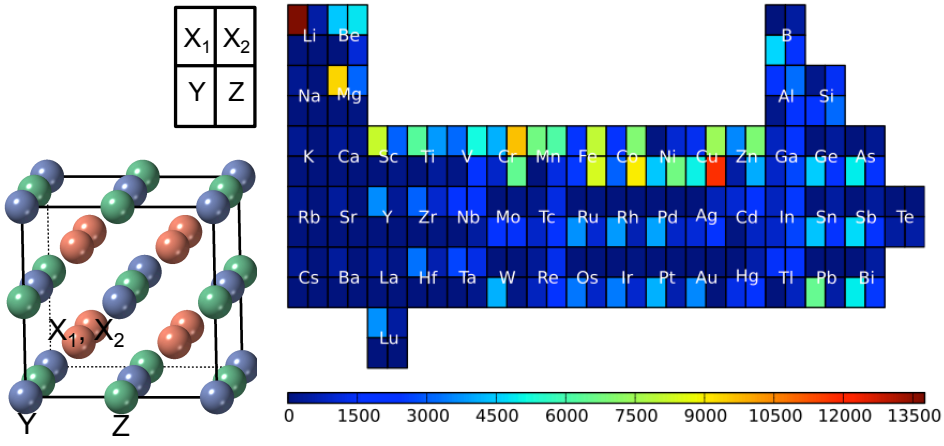

Figure 1. (Left) Crystal structure of quaternary Heusler compound (X_1X_2YZ). Red atoms represent X_1 and X_2 sites on the top right, blue atoms represent Y sites, and orange atoms represent Z sites, respectively. In case of quaternary Heusler compounds, X_1 and X_2 sites are decorated by different species. So, one composition has three symmetry-distinct crystal structure of QHs. (Right) Frequency of chosen elements in QHs training data sets. Red (blue) color indicates higher (lower) frequency.

Our training data is taken from the OQMD (http://www.oqmd.org) [8, 12]. At the time of writing this paper, the OQMD contained 426,148 compounds, and all of these were used for ML training. The OQMD has a large number of Heusler compounds already in the dataset, due to previous high-throughput searches for novel materials with diverse applications, such as strengthening precipitates, thermoelectrics, spintronics, etc [9, 11, 43]. The OQMD dataset contains 184,094 ternary Heuslers (all possible combinations of 73 elements in a ternary full Heusler structure), 96,189 quaternary Heuslers (only ~3% of all possible combinations of 73 elements in the QH structure), as well as a large array of 145,866 non-Heusler compounds.

Fig. 1 shows the frequency of chosen elements in the QH training data sets, which features certain elements more prevalently due to previous use of the OQMD by other studies. For example, the QH dataset is biased towards compounds containing transition metals and Li due to most of this data coming from a study of Li-containing QHs [43]. The Li-containing QH study and others also focused on 18-electron compounds, which means that 17k of the 96k QHs in our training set contain 18 valence electrons. Other previous studies that employed the OQMD

focused on finding stable QHs based valence electron counting heuristics [2, 44-46], which means that 24 valence electrons QHs are also well-represented in our training set.

In the OQMD, all first-principles DFT calculations are performed with the Vienna Ab-initio Simulation Package (VASP) [47-49] and the PBE parameterization of the generalized gradient approximation (GGA-PBE) [50]. Elements with incompletely-filled *f* shells are modeled using pseudo potentials that treat the f electrons as core electrons. For all structures, gamma-centered k-point meshes were constructed to achieve 8,000 k-points per reciprocal atom. For the total energy calculations, high quality static calculation of the relaxed structure is performed with an energy cutoff of 520 eV. More detailed information about the DFT calculations in OQMD is described in http://www.oqmd.org/documentation/vasp and Refs. [8, 12].

We seek to use this data to predict whether a new QH will be energetically stable at T=0K, which can be accomplished by predicting several different quantities. First, we could predict the raw DFT energy of the structure, which requires significant postprocessing to determine if the structure is stable. Alternatively, we could predict the formation energy, which requires less postprocessing, or the distance from the convex hull (a measure of stability) directly. The convex hull distance is the difference between the energy of a compound and the convex hull of all entries of the OQMD at the same composition, which is the lowest-energy combination of phases at that composition. For our task, we use the convex hull formed by all of the entries except the QH compounds to compute the convex hull energy. Using this definition, a stable QH compound will have a negative convex hull distance (i.e., where the distance is given by the difference in

energy between the QH compound and the convex hull at the same composition before including the QH compound).

2.2. Material Representation

The representation of a material is a set of quantitative attributes which serve as the input into a machine learning method. We used the Voronoi-tessellation-based representation of Ward et al. [27] to describe each material in our training set and the search space. Specifically, the Ward et al. [27] representation describes crystal structures based on a set of 271 attributes derived from the composition of the crystal (e.g., the mean atomic number, mean absolute deviation of the melting temperature) and from the Voronoi tessellation of the crystal structure (e.g., coordination numbers, ordering parameters, differences in electronegativity between neighbors). We employ Magpie to compute this representation for each entry [51] (https://bitbucket.org/wolverton/magpie.git).

The Voronoi representation of Ward *et al.* has two advantages for our purpose: (i) it allows for using the entire OQMD as a training set, and (ii) it has the best accuracy of formation energy prediction on the full OQMD among available machine learning methods to date. The Ward Voronoi representation can be trained on the entire OQMD because it can be computed for any crystal structure. In contrast, some prior work uses representations [26, 28, 37], which requires all materials have the same crystal structure. As we will demonstrate, there is a significant advantage to being able to use training data from crystal structures types besides Heuslers. Additionally, as of the time this work was performed, the machine learning models created using

the Ward Voronoi method achieved the best accuracy for predicting the formation enthalpy using the full OQMD dataset (~80 meV/atom in 10-fold cross-validation), making it a strong choice for our application.

2.3. Machine learning algorithm

In this work, we adopt the Random Forest algorithm proposed by Breiman [52] because of its robustness to over fitting and ability to be trained using large datasets. Random Forest works by training many different decision trees models and combining their predictions together to create a single, more-accurate model. This "ensemble" approach is the source of the speed and accuracy advantages of Random Forest. First of all, each individual decision tree is trained on a different randomly-selected subset of the training data. Between the fact that no tree sees the entire dataset and that each tree is trained using a different subset, the aggregate Random Forest is less likely to be overfit than a single model. Secondly, the inherent parallelism of training many smaller models combined with the fact that the training time for each tree scales with only $N \log N$, where N is the dataset size, make it computationally feasible to train Random Forest models on datasets with $>10^5$ entries. In our work, the models train in under an hour on typical hardware. Unless otherwise noted, we use Weka [53] to train Random Forest models.

2.4. Search space

Our goal is to identify stable QH compounds formed from any possible combination of metallic elements. We used the 73 metallic elements available in the OQMD: Ac, Ag, Al, As, Au, B, Ba,

Be, Bi, Ca, Cd, Ce, Co, Cr, Cs, Cu, Dy, Er, Eu, Fe, Ga, Gd, Ge, Hf, Hg, Ho, In, Ir, K, La, Li, Lu, Mg, Mn, Mo, Na, Nb, Nd, Ni, Np, Os, Pa, Pb, Pd, Pm, Pr, Pt, Pu, Rb, Re, Rh, Ru, Sb, Sc, Si, Sm, Sn, Sr, Ta, Tb, Tc, Te, Th, Ti, Tl, Tm, U, V, W, Y, Yb, Zn, Zr. Considering there are three crystallographically-unique polymorphs at each composition, our search space contains a total of 3,265,290 ($73C_4$ x 3) variations.

All the data and software necessary to recreate the methods employed in this search are available in the Supplementary Information and on GitHub: https://github.com/WardLT/ml-quaternary-heuslers

3. Results and Discussions

3.1. Training set selection

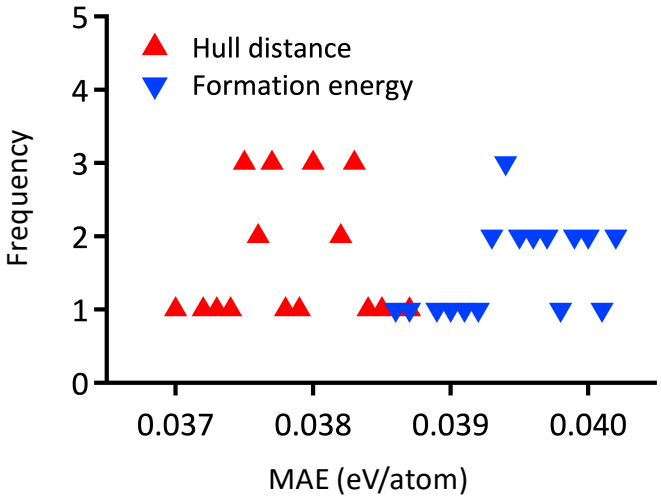

Figure 2. Comparison of the accuracy of models trained on the formation enthalpy and convex hull distance across 25 repetitions of a cross-validation test. For each test, a model was trained on all of the available data in the OQMD minus 10% of the Quaternary Heusler entries, which were withheld for testing. We find the Mean Absolute Error (MAE) of the model trained on the hull distance is better for all iterations of the cross-validation test.

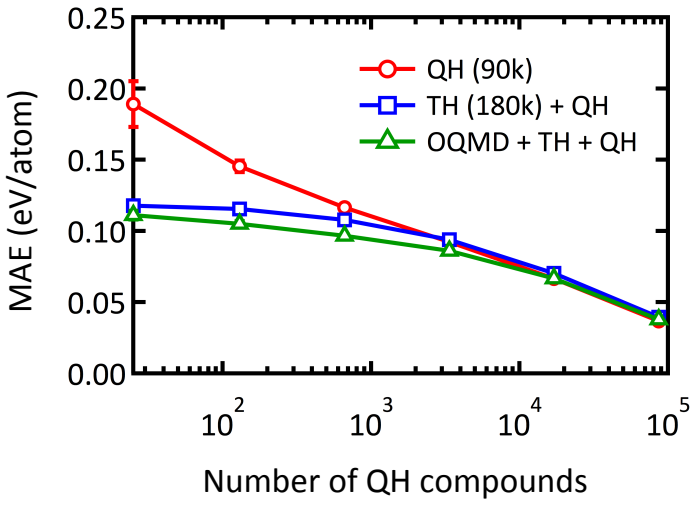

Figure 3. Cross-validation accuracy (mean absolute error, eV/atom) of machine learning models trained on different datasets: (red circle) only Quaternary Heusler (referred to here as QH) compounds, (blue square) ternary (referred to here as TH) and QH compounds, (green triangle) the full OQMD dataset including TH and QH. We show the accuracy as a function of the number of QH compounds in the training set. The full-OQMD model achieves the best accuracy because it learns more generally how stability and crystal structure are related using data for crystal structures other than QH. Consequently, it takes over examples of a 3k QH structure for the QH

model to reach the accuracy of the OQMD model with only 1 QH example. At 90 k examples of QHs, all three models are equivalent.

Our first task was to select an appropriate training set for training our model. As alluded to in Section 2.1, we have several options for training data: what to use as the output variable and what subset of our available data to use for training.

We tested training a model using either the formation energy or hull distance as an output by using a leave-10%-out cross-validation test. In this test, we remove 10% of the QH entries from the OQMD dataset, train a model on the remaining entries, and then measured the Mean Absolute Error (MAE) of the predictions of the model on the 10% of the QH that were initially removed. We repeated the CV test using both the formation energy and hull distance as output variables using the same test data; and repeated the entire test with a different test split a total of 25 times. The convex hull distances are the equivalent to a formation enthalpy plus an unknown constant, which means that the MAE between the two datasets are comparable. As shown in Fig. 2, the model trained on the convex hull distance has a lower MAE for all 25 iterations of the test. From this, we conclude hull distance is a better choice for output variable.

We also needed to decide between training the model on all available data or limiting our training set to only those materials with similar structures to our search space: QHs. For training the ML model, we investigated three different training sets: 1) training on only QH data (90 k compounds), 2) using both quaternary and ternary Heusler (TH) data, and 3) training on all available data in the OQMD, including both QH and TH data. We refer to these three sets as

"QH", "TH+QH", and "OQMD+TH+QH", respectively. To decide which method is best, we performed a cross-validation test where we included gradually larger numbers of QHs in the training set and measured the accuracy of the model against the withheld QHs. We chose to withhold a randomly-selected 10% of our QH data as a test set and repeated each test 5 times using a different test set each time.

Fig. 3 represents the accuracy of trained ML models based on the three different datasets as a function of training set size. The choice of training set is the most relevant when small number of QH data was in the training set. It takes over 3k QH training entries in order for the QH-only model to reach the comparable accuracy of the OQMD model with only 1 QH. With larger number of QH data, models behave accurately regardless of the choice of training set. For the largest training set size we tested (87k QH entries), the CV result of ML model trained using only the QH data (red in Fig. 2) shows MAE of 0.037 ± 0.001 eV/atom. The model using both the QH data and that from the ternary Heuslers (THs, as blue in Fig. 2) shows 0.039 ± 0.001 eV/atom, and the model trained by the full OQMD dataset (as green in Fig. 1) shows 0.038 ± 0.001 eV/atom. The superior accuracy of the "all OQMD" model at small dataset sizes demonstrates the advantage of re-using data. What this suggests is that the data about QH compounds is important, but not necessary for being able to build an accurate ML model. Considering that the accuracy of all three models are equivalent for large training set sizes, we conclude the ML model trained on the entire OQMD is the best option for identifying new, stable QHs.

Our finding that a large, diverse training set leads to the most accurate machine learning models agrees generally with lessons learned in other materials informatics studies. For example, Browning et al. demonstrated optimized training sets for molecular machine learning models have maximum difference between training entries[54]. Furthermore, it is well known in the potential fitting community that the training sets that span the entire space of structures to be sampled[55]. What we find, in our case, is that denser sampling over composition space is needed to achieve optimal prediction performance for finding new intermetallic materials. The superior accuracy of the "all OQMD" model demonstrates the utility of large training sets (e.g., the QM-9 dataset in quantum chemistry[56]) towards materials discovery.

3.2. Validating machine learning model

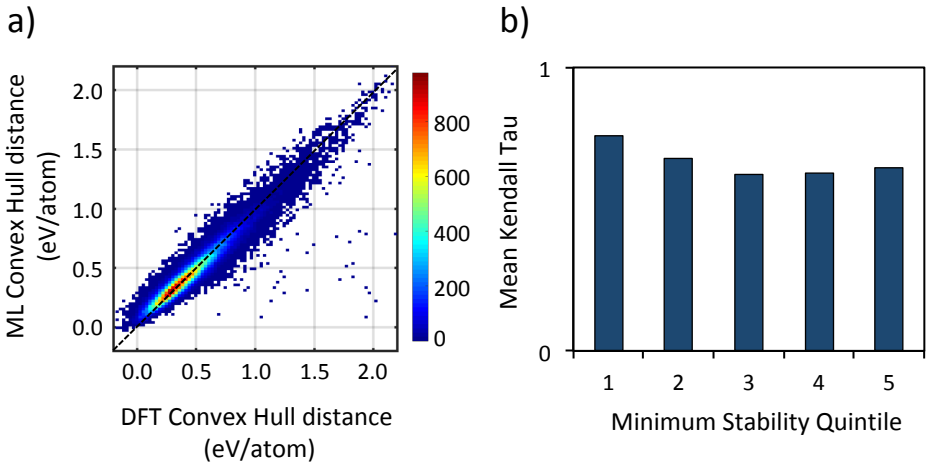

Figure 4. (a) A comparison between ML-predicted and DFT-calculated convex hull distance (eV/atom) from a grouped-validation test where we measured the performance of the model to predict the stability of Quaternary Heusler compounds at compositions not included in the

training set. The chart is a 2D-dimensional histogram where the color of the pixel corresponds to the number of entries with a certain DFT and ML convex hull distance. (b) Measurements of the ability of our model to rank polymorphs at a certain composition in the correct order, as measured using the average Kendall Tau statistic (τ). We measure the average τ for entries in different quintiles based on the minimum stability of all polymorphs at each composition (quintile 1 contains the 20% compositions which are the most stable in DFT). Note that a Kendal score of 1 means all polymorphs at compositions in that quintile were ranked in the correct order and a score of 0 indicates that, on average, the order was random.

As our objective is to find new stable materials, our model needs to be able to predict the stability of a structure at compositions not included in the training set. As our training set includes multiple structures at each composition (there exist 3 possible QH polymorphs per composition), we employed a "grouped" cross-validation test to evaluate the performance of the model. In contrast to conventional cross-validation techniques which randomly partition entries between test and training sets, we first group entries by composition and then randomly partition these groups of entries during cross-validation. This grouped cross-validation procedure ensures that the same composition never appears in both the training set and test set, which provides a more rigorous test of the predictive ability of our model.

Specifically, we performed a 10-fold grouped cross-validation test. We first grouped QH entries by composition and then randomly split the list of groups into 10 sets. We then trained our model using 9 of these 10 sets and all of the remaining data from the OQMD, and then evaluated its performance on the entries on the withheld set. We repeated this test 10 times by iteratively

withholding each of the 10 QH sets to use as the test set, and measured the performance of the model over all test sets.

As shown in Fig. 4 (a), our model achieved strong predictive accuracy during this grouped cross-validation test. The correlation coefficient (R^2) for the model was 0.91, and the MAE was 53 meV/atom. Our training sets included, on average, 86569 QH entries, which is comparable to the largest training set size in Fig. 3. Consequently, we can compare the performance scores and note that our performance here is slightly worse than that observed in the test in Fig. 1, which is not surprising given the increased difficulty of the task. However, the performance still is quite strong, the Root Mean Squared Error (RMSE) of our model is less than 30% of the training set standard deviation. From this, we conclude that our model is capable of being used to predict the stability of Heuslers at compositions not included in our original dataset.

Another desired performance characteristic of our model is the ability to distinguish which QH polymorph at a certain composition is the most stable. For a given composition (e.g., ABCD), there are 3 different ways to arrange these elements in the QH structure that are crystallographically distinct. For an ABCD composition on the X_1X_2YZ sites of QH, the following arrangements are distinct: ABCD, ACBD, and ADBC. All other permutations of these elements are symmetrically equivalent to one of these three structures. To determine whether our model is capable of correctly ordering polymorphs based on stability, we first identified the 5813 compositions where we evaluated all three polymorphs. In these cases, our model determines the correct ordering 58% of the time and the average Kendall τ rank correlation coefficient over all 5813 tests is 0.68. Considering that a model which ranked polymorphs randomly would select

the correct ordering with probability 1/6 (17%) of the time and would have a τ of 0, these scores are quite promising. As shown in Fig. 4 (b), our model achieves a strong ranking performance coefficient (τ = 0.76) for the compositions we are most interested in: those with more stable polymorphs. That said, given that the ranking is not perfect, we still recommend computing the stability of all three prototypes when searching for new QHs regardless to ensure the correct polymorph is found.

3.3. Computational discovery of new quaternary Heuslers

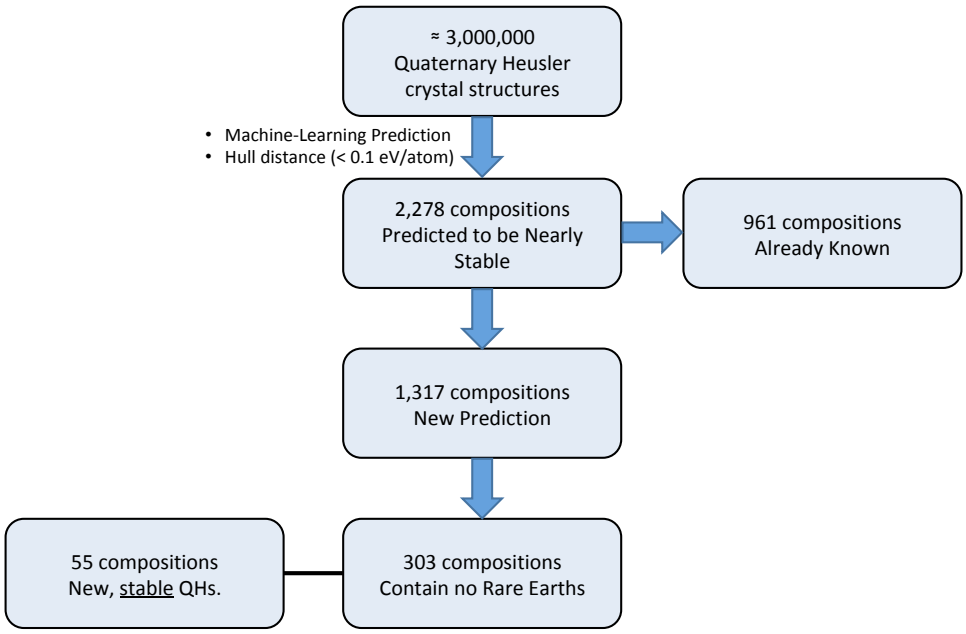

Figure 5. Summary of the machine-learning-assisted search for stable quaternary Heusler compounds (QH). The ML predicted 2,278 compositions where at least one compound has a convex hull distance of less than 0.1 eV/atom. Among those 2,278 compositions, 1,317 compositions are not in training data and 303 compositions among those 1,317 do not contain

rare earth elements. As seen here, we find DFT-stable 55 QH compositions among 303 using ML prediction in this work.

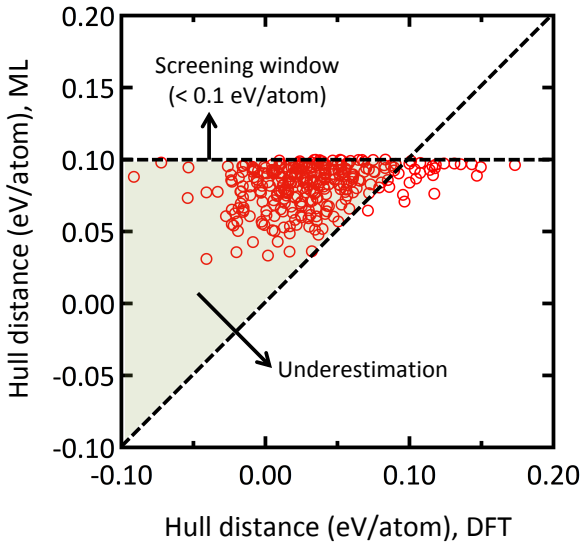

Figure 6. Plot of hull distances (eV/atom) by machine learning (ML) vs. DFT for the 303 QH compositions (Fig. 4) identified using a machine-learning-based screening. Most predictions from the ML model underestimate the formation energy (see the shaded region). As such, we ranked the predicted compounds based on hull distance for choosing order of priority for DFT calculations with the upper limit of 0.1 eV/atom.

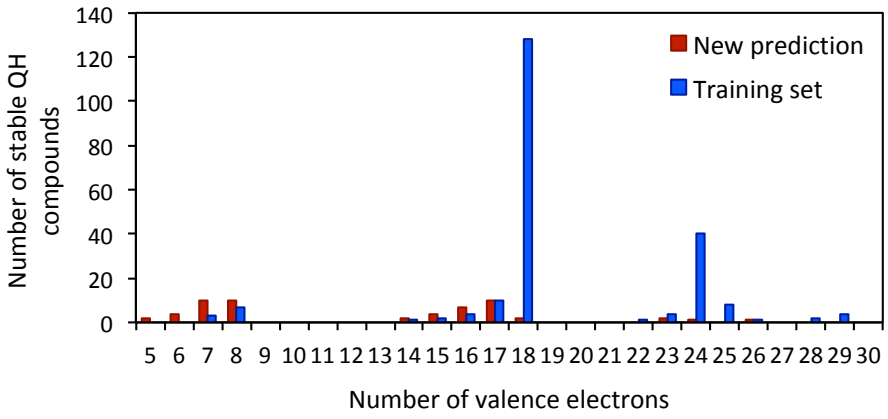

Figure 7. Histogram of number of stable quaternary Heusler compounds (QHs) vs. number of valence electrons in the compound. As seen here, most of stable phases in the training set have 18 or 24 valence electrons. Interestingly, we find many stable phases predicted by ML (and verified by DFT) that do not follow the electron counting rules.

Fig. 5 summarizes the results of our ML-based predictions for searching new stable QH compounds. Our ML model predicted that compounds at 2,278 distinct compositions have convex hull distances that are less than 0.1 eV/atom above the convex hull. Among the 2,278 compositions, 961 compositions were in the training set (OQMD) and their stability had already been calculated by DFT. As such, 1,317 compositions represent newly predicted phases that have not yet been investigated by DFT. Many of these materials contain rare-earth elements and, because rare-earth-containing systems historically have not been extensively explored, discovering a new compound containing these elements is relatively less challenging than those without any rare earth elements. For that reason, we removed any compounds which contain at least one rare earth from consideration. We then performed DFT calculations of all three polymorphs for each of the 303 non-rare-earth-containing compositions where we predicted at

least 1 QH to be stable. Based on our DFT calculations, we find that the 55 of our candidate QHs are indeed stable. As shown by Fig. 6, that the ML model tends to underestimate the convex hull distance and, consequently, we conclude our approach of screening based on a stability threshold of 0.1 eV/atom is practical.

The efficiency of a ML+DFT approach is much higher than conventional unconstrained high throughput screening and even searches guided by electron-counting rules. In a conventional HT-DFT screening, we found 204 stable compounds by performing 96k DFT calculations. When only selecting compounds with 18 or 24 valence electrons per cell, we found 180 out of 17k compounds to be stable: a rate of approximately one stable material per 100 calculations. We found 55 stable compounds using only 909 DFT calculation based on our predictions from the machine learning model, which corresponds to a rate of one stable compound per 16 calculations – around 30 times faster than when using conventional HT.

Many of our predictions contain elements from similar groups. 43 of our 55 predictions contain Li. In retrospect, a large number of Li predictions is somewhat unsurprising given that Li is the element that appears most frequently in the stable QHs in our training set. In 38 of these 43 Li-containing QHs, the compound also contains at least 1 element from the Ni- or Cu-group. Untangling whether the large number of predictions with Li and a group 10/11 metal could provide either insight on how the biases in a training set affect ML predictions or, perhaps, on the stability of QHs, which could be an intriguing opportunity for further study.

The large majority of the stable compounds we found did not follow the 18 or 24 electron-counting heuristic, as shown in Fig. 7. The lack of stable materials with 18 or 24 electrons cannot be explained by the absence of any 18 or 24 electrons in the training set because there are 22876 compositions with 18 or 24 electrons and no rare-earth elements remaining in the search space. Our model simply does not predict many more 18/24 electron QHs to be stable. Rather, we find more compounds with 17 or 7 electrons to be stable in the QH structure. Both of these results suggest that other factors besides the number of valence electrons are important in predicting stability, and that our machine learning model was capable of finding and using them to predict new stable compounds.

3.4. Comparing our machine learning model with other ML approaches

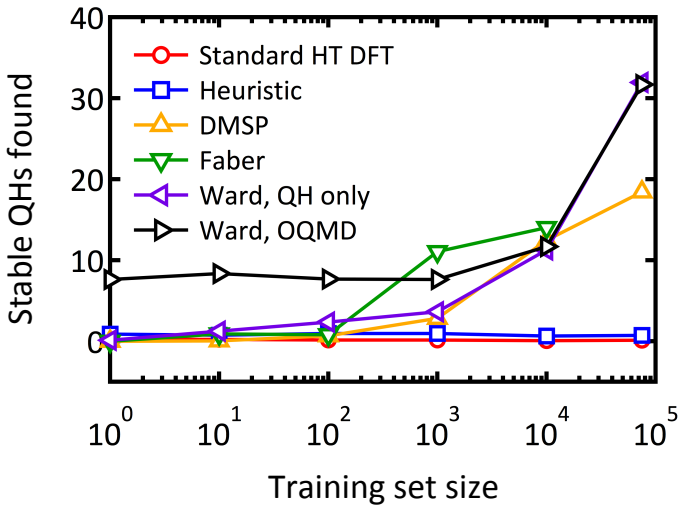

Figure 8. Performance of several strategies for identifying stable crystalline compounds. The test was to identify 50 stable quaternary Heuslers (QH) out of a list of 20 k given a training set of

between 1 and 75000 quaternary Heuslers. We plot the number of stable compounds found after 50 guesses as a function of training set size. We compare machine learning models trained with the Voronoi tessellation method of Ward et al. [27] using only QH data (Ward, QH) and a training set supplemented with data from the OQMD (Ward), the Kernel-Ridge-Regression-based method of Faber et al. [26], the Data Mining Structure Predictor (DMSP) method [57], and the electron counting heuristics of [2, 44, 45]. This figure shows the number of compounds within the 50 predictions that were actually stable.

As we have just demonstrated, our model is certainly capable of identifying new stable compounds at a substantially accelerated rate. However, our Voronoi-tessellation-based approach is not the only method for identifying new stable compounds. So, in this section, we evaluate the effectiveness of our approach (the Ward Voronoi tessellation method) against several other methods: random search, chemical heuristics, the machine learning approach used by Faber et al. [26] and the Data Mining Structure Prediction (DMSP) method of Fischer et al. [57] The approach by Faber works by learning a Kernel Ridge Regression (KRR) model based on the distance between the row and group number of elements on each site in a crystal. The DMSP method builds a classification model of whether a material will be stable based on which other crystal structure types (e.g., ternary Heusler) are stable in a chemical system.

To compare these methods, we employed a test where we randomly selected between 1 and 75000 quaternary Heuslers to train a model and then used the model to identify the 50 compounds most likely to be stable from a search space of 20000 compounds not included in the training set. We scored the performance of each method by measuring how many of the selected

compounds were indeed stable, and report the average of this metric over 25 tests (each with a different training set and search space).

As shown in Fig. 8, the best-performing algorithm changes as more QH data is made available. At the smallest training set size (only 1 QH training example), the model created using the entire OQMD dataset and the Ward Voronoi method (i.e., the method used in previous sections) achieves the best performance: 8.2 ± 0.4 compounds in the 50 choices were actually stable. At this small training set size, the second-best method was preferentially selecting compounds with 18 or 24 electrons, which found 0.8 ± 0.2 compounds. While less than 1 stable compound per 50 guesses is a low success rate, the heuristic is still a factor of three better than random selection, which further confirms that electron counting rules are helpful. Given the lack of training data, the machine learning methods that only use QH data (e.g., DMSP) perform equivalent to random selection. These results suggest that either using known heuristics (e.g., electron counting rules) or ML methods that use data from other crystalline compounds are the best choice when few examples of a new structure are available.

At a moderate training set size of 1000 QH entries, we find that the model created using the Faber method [26] has the best performance. The Faber method finds 11.2 ± 0.9 stable compounds, which is clearly better than the score of 8.0 ± 0.2 for Ward Voronoi method trained on the entire OQMD. To further understand why the Faber method performs better, we tested a model created using the same KRR algorithm but with Ward Voronoi representation to define the distance between structures. We observed a performance of 4.88 ± 0.43 for this model when training using only the QH data – better than Ward representation using a Random Forest

algorithm, but worse than the same algorithm and the Faber representation. This lower performance suggests that both the algorithm and representation used by Faber are the keys to its strong performance. In contrast to the Ward Voronoi method, the Faber method uses a representation that only supports a single structure type, which apparently describes the QH data better than a general representation and leads to more accurate models even though it uses a smaller training set.

We find that models created using the Ward method have the best performance at the largest training set size of 75000 QH examples. At such large training set sizes, we are unable to test the Faber method due to large time required to generate the kernel matrix. That said, given the trajectory of the performance of the Faber method with system size, it is possible that it may perform worse than the Ward method for large training set sizes. We find the models created with the Ward method perform equivalently well whether the model was trained using the entire OQMD or just the QH datasets; the models find 31.7 ± 0.8 and 32.0 ± 0.8 stable compounds per 50 guesses, respectively. In comparison, the next best method, DMSP (as implemented in Magpie), achieves a score of 18.4 ± 0.6 compounds. We therefore conclude the Ward method is the best choice for large training sets ($>10^4$ entries).

Given these results, we propose that different methods should be used depending on the amount of training data available. For small training set sizes, the ability of the Voronoi-tessellation-based method of Ward et al. [27] to use data from crystal structures of different types allows it to make predictions about a new structure type with few examples. For larger training sets, we recommend using the method of Faber et al. [26], which – despite using only data describing a

single structure type – achieves better predictive performance for datasets greater than 1000 entries in size. For even larger dataset sizes, the Faber method can become computationally intractable and we recommend using the Ward method. The threshold at which switching from the Ward method to the Faber method will likely be different for different types of crystals, and we propose that one should employ a similar cross-validation scheme as the one demonstrated in this section to determine which method is better suited for a particular problem. With the combination of these two methods and data from the OQMD, it is possible to find new stable materials at a rate up to 30x faster than random searching regardless of how many training examples are available.

4. Conclusions

In this work, we established a ML model that predicts the stability of crystalline materials and used to them to discover 55 new quaternary Heuslers. Our model, which is trained using the entire OQMD, achieves excellent accuracy on predicting the stability of quaternary Heuslers even with little training data describing compounds with that structure. After using this model to evaluate the stability of 3.2M candidate materials, we used the predictions to discover 55 new stable quaternary Heusler phases with only 909 DFT energetic calculations – a rate of discovering stable compounds 30x higher than what we previously achieved without machine learning. We then benchmarked our machine learning method against other approaches, and found that our method performs best for small ($<10^3$ entries) training sets and large training set sizes ($>10^5$), the method of Faber *et al.* 2016 is best for moderate dataset sizes (10^3-10^4 entries).

We envision that the combination of these two techniques make it possible to quickly identify new stable materials for many types of compounds.

Acknowledgements

K. K. acknowledges Predictive Science and Engineering Design (PSED) graduate program at Northwestern University. J. H. acknowledges support of the Toyota Research Institute. L. W., A. K., A. A., P. V., and C. W. were supported by financial assistance Award 70NANB14H012 from U. S. Department of Commerce, National Institute of Standards and Technology as part of the Center for Hierarchical Materials Design (CHiMaD). This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1548562. Specifically, we used Jetstream at the Texas Advanced Computing Center through allocation CIE170012.

5. References

[1] F. Heusler, W. Starck, E. Haupt, Verh DPG 5 (1903) 220-3.

[2] T. Graf, C. Felser, S.S.P. Parkin, Simple rules for the understanding of Heusler compounds, Progress in Solid State Chemistry 39(1) (2011) 1-50.

[3] J. Drews, U. Eberz, H.-U. Schuster, Optische Untersuchungen an farbigen Intermetallischen phasen, Journal of the Less Common Metals 116(1) (1986) 271-278.

[4] J. He, M. Amsler, Y. Xia, S.S. Naghavi, V.I. Hegde, S. Hao, S. Goedecker, V. Ozoliņš, C. Wolverton, Ultralow Thermal Conductivity in Full Heusler Semiconductors, Physical Review Letters 117(4) (2016) 046602.

[5] J. He, S.S. Naghavi, V.I. Hegde, M. Amsler, C. Wolverton, Designing and Discovering a New Family of Semiconducting Quaternary Heusler Compounds Based on the 18-Electron Rule, Chemistry of Materials 30(15) (2018) 4978-4985.

[6] S. Curtarolo, G.L.W. Hart, M.B. Nardelli, N. Mingo, S. Sanvito, O. Levy, The high-throughput highway to computational materials design, Nature Materials 12 (2013) 191.

[7] H. Chen, G. Hautier, A. Jain, C. Moore, B. Kang, R. Doe, L. Wu, Y. Zhu, Y. Tang, G. Ceder, Carbonophosphates: A New Family of Cathode Materials for Li-Ion Batteries Identified Computationally, Chemistry of Materials 24(11) (2012) 2009-2016.

[8] J.E. Saal, S. Kirklin, M. Aykol, B. Meredig, C. Wolverton, Materials Design and Discovery with High-Throughput Density Functional Theory: The Open Quantum Materials Database (OQMD), JOM 65(11) (2013) 1501-1509.

[9] J. Ma, V.I. Hegde, K. Munira, Y. Xie, S. Keshavarz, D.T. Mildebrath, C. Wolverton, A.W. Ghosh, W.H. Butler, Computational investigation of half-Heusler compounds for spintronics applications, Physical Review B 95(2) (2017) 024411.

[10] A. Jain, K. Persson, G. Ceder, Research Update: The materials genome initiative: Data sharing and the impact of collaborative ab initio databases, APL Materials 4 (2016) 053102.

[11] S. Kirklin, J.E. Saal, V.I. Hegde, C. Wolverton, High-throughput computational search for strengthening precipitates in alloys, Acta Materialia 102 (2016) 125-135.

[12] S. Kirklin, J.E. Saal, B. Meredig, A. Thompson, J.W. Doak, M. Aykol, S. Rühl, C. Wolverton, The Open Quantum Materials Database (OQMD): assessing the accuracy of DFT formation energies, Npj Computational Materials 1 (2015) 15010.

[13] M. Aykol, S. Kim, V.I. Hegde, D. Snydacker, Z. Lu, S. Hao, S. Kirklin, D. Morgan, C. Wolverton, High-throughput computational design of cathode coatings for Li-ion batteries, Nature Communications 7 (2016) 13779.

[14] J. Balluff, K. Diekmann, G. Reiss, M. Meinert, High-throughput screening for antiferromagnetic Heusler compounds using density functional theory, Physical Review Materials 1(3) (2017) 034404.

[15] L. Ward, C. Wolverton, Atomistic calculations and materials informatics: A review, Current Opinion in Solid State and Materials Science 21(3) (2017) 167-176.

[16] R. Ramprasad, R. Batra, G. Pilania, A. Mannodi-Kanakkithodi, C. Kim, Machine learning in materials informatics: recent applications and prospects, npj Computational Materials 3(1) (2017) 54.

[17] A. Jain, G. Hautier, S.P. Ong, K. Persson, New opportunities for materials informatics: Resources and data mining techniques for uncovering hidden relationships, Journal of Materials Research 31(8) (2016) 977-994.

[18] K.T. Butler, D.W. Davies, H. Cartwright, O. Isayev, A. Walsh, Machine learning for molecular and materials science, Nature 559(7715) (2018) 547-555.

[19] Z.W. Ulissi, M.T. Tang, J. Xiao, X. Liu, D.A. Torelli, M. Karamad, K. Cummins, C. Hahn, N.S. Lewis, T.F. Jaramillo, K. Chan, J.K. Nørskov, Machine-Learning Methods Enable Exhaustive Searches for Active Bimetallic Facets and Reveal Active Site Motifs for CO2 Reduction, ACS Catalysis 7(10) (2017) 6600-6608.

[20] A. Mayr, G. Klambauer, T. Unterthiner, M. Steijaert, J.K. Wegner, H. Ceulemans, D.-A. Clevert, S. Hochreiter, Large-scale comparison of machine learning methods for drug target prediction on ChEMBL, Chemical Science 9(24) (2018) 5441-5451.

[21] D.T. Ahneman, J.G. Estrada, S. Lin, S.D. Dreher, A.G. Doyle, Predicting reaction performance in C–N cross-coupling using machine learning, Science 360(6385) (2018) 186-190.

[22] G. Pilania, J.E. Gubernatis, T. Lookman, Multi-fidelity machine learning models for accurate bandgap predictions of solids, Computational Materials Science 129 (2017) 156-163.

[23] T.L. Jacobsen, M.S. Jørgensen, B. Hammer, On-the-Fly Machine Learning of Atomic Potential in Density Functional Theory Structure Optimization, Physical Review Letters 120(2) (2018) 026102.

[24] S. Curtarolo, D. Morgan, K. Persson, J. Rodgers, G. Ceder, Predicting Crystal Structures with Data Mining of Quantum Calculations, Physical Review Letters 91(13) (2003) 135503.

[25] B. Meredig, A. Agrawal, S. Kirklin, J.E. Saal, J.W. Doak, A. Thompson, K. Zhang, A. Choudhary, C. Wolverton, Combinatorial screening for new materials in unconstrained composition space with machine learning, Physical Review B 89(9) (2014) 094104.

[26] F.A. Faber, A. Lindmaa, O.A. von Lilienfeld, R. Armiento, Machine Learning Energies of 2 Million Elpasolite $(AB{C}_{2}{D}_{6})$ Crystals, Physical Review Letters 117(13) (2016) 135502.

[27] L. Ward, R. Liu, A. Krishna, V.I. Hegde, A. Agrawal, A. Choudhary, C. Wolverton, Including crystal structure attributes in machine learning models of formation energies via Voronoi tessellations, Physical Review B 96(2) (2017) 024104.

[28] G. Pilania, P.V. Balachandran, C. Kim, T. Lookman, Finding New Perovskite Halides via Machine Learning, Frontiers in Materials 3(19) (2016).

[29] G. Ceder, D. Morgan, C. Fischer, K. Tibbetts, S. Curtarolo, Data-Mining-Driven Quantum Mechanics for the Prediction of Structure, MRS Bulletin 31(12) (2011) 981-985.

[30] A.O. Oliynyk, E. Antono, T.D. Sparks, L. Ghadbeigi, M.W. Gaultois, B. Meredig, A. Mar, High-Throughput Machine-Learning-Driven Synthesis of Full-Heusler Compounds, Chemistry of Materials 28(20) (2016) 7324-7331.

[31] A. Seko, H. Hayashi, I. Tanaka, Matrix- and tensor-based recommender systems for the discovery of currently unknown inorganic compounds, Physical Review Materials 2(1) (2018) 013805.

[32] G. Hautier, C.C. Fischer, A. Jain, T. Mueller, G. Ceder, Finding Nature's Missing Ternary Oxide Compounds Using Machine Learning and Density Functional Theory, Chemistry of Materials 22(12) (2010) 3762-3767.

[33] G. Pilania, C. Wang, X. Jiang, S. Rajasekaran, R. Ramprasad, Accelerating materials property predictions using machine learning, Scientific Reports 3 (2013) 2810.

[34] M. de Jong, W. Chen, R. Notestine, K. Persson, G. Ceder, A. Jain, M. Asta, A. Gamst, A Statistical Learning Framework for Materials Science: Application to Elastic Moduli of k-nary Inorganic Polycrystalline Compounds, Scientific Reports 6 (2016) 34256.

[35] T.D. Sparks, M.W. Gaultois, A. Oliynyk, J. Brgoch, B. Meredig, Data mining our way to the next generation of thermoelectrics, Scripta Materialia 111 (2016) 10-15.

[36] D. Xue, P.V. Balachandran, J. Hogden, J. Theiler, D. Xue, T. Lookman, Accelerated search for materials with targeted properties by adaptive design, Nature Communications 7 (2016) 11241.

[37] J. Schmidt, J. Shi, P. Borlido, L. Chen, S. Botti, M.A.L. Marques, Predicting the Thermodynamic Stability of Solids Combining Density Functional Theory and Machine Learning, Chemistry of Materials 29(12) (2017) 5090-5103.

[38] J.M. Granda, L. Donina, V. Dragone, D.-L. Long, L. Cronin, Controlling an organic synthesis robot with machine learning to search for new reactivity, Nature 559(7714) (2018) 377-381.

[39] A.O. Oliynyk, A. Mar, Discovery of Intermetallic Compounds from Traditional to Machine-Learning Approaches, Accounts of Chemical Research 51(1) (2018) 59-68.

[40] A. Seko, H. Hayashi, H. Kashima, I. Tanaka, Matrix- and tensor-based recommender systems for the discovery of currently unknown inorganic compounds, Physical Review Materials 2(1) (2018) 013805.

[41] E.O. Pyzer-Knapp, K. Li, A. Aspuru-Guzik, Learning from the Harvard Clean Energy Project: The Use of Neural Networks to Accelerate Materials Discovery, Adv. Funct. Mater. 25 (2016) 6495-6502.

[42] K. Choudhary, B. DeCost, F. Tavazza, Machine learning with force-field-inspired descriptors for materials: Fast screening and mapping energy landscape, Physical Review Materials 2(8) (2018) 083801.

[43] J. He, S.S. Naghavi, V.I. Hegde, M. Amsler, C. Wolverton, Designing and discovering a new family of semiconducting quaternary Heusler compounds based on the 18-electron rule, Unpublished (2018).

[44] I. Galanakis, P.H. Dederichs, N. Papanikolaou, Slater-Pauling behavior and origin of the half-metallicity of the full-Heusler alloys, Physical Review B 66(17) (2002) 174429.

[45] R. Gautier, X. Zhang, L. Hu, L. Yu, Y. Lin, O.L. SundeTor, D. Chon, K.R. Poeppelmeier, A. Zunger, Prediction and accelerated laboratory discovery of previously unknown 18-electron ABX compounds, Nat Chem 7(4) (2015) 308-316.

[46] K. Hem Chandra, F. Claudia, S. Ram, Covalent bonding and the nature of band gaps in some half-Heusler compounds, Journal of Physics D: Applied Physics 39(5) (2006) 776.

[47] G. Kresse, J. Furthmüller, Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set, Physical Review B 54(16) (1996) 11169-11186.

[48] G. Kresse, J. Furthmüller, Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set, Computational Materials Science 6(1) (1996) 15-50.

[49] G. Kresse, D. Joubert, From ultrasoft pseudopotentials to the projector augmented-wave method, Physical Review B 59(3) (1999) 1758-1775.

[50] J.P. Perdew, K. Burke, M. Ernzerhof, Generalized Gradient Approximation Made Simple, Physical Review Letters 77(18) (1996) 3865-3868.

[51] L. Ward, A. Agrawal, A. Choudhary, C. Wolverton, A general-purpose machine learning framework for predicting properties of inorganic materials, Npj Computational Materials 2 (2016) 16028.

[52] L. Breiman, Random Forests, Machine Learning 45(1) (2001) 5-32.

[53] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I.H. Witten, The WEKA data mining software, ACM SIGKDD explorations newsletter 11(1) (2009) 10-18.

[54] N.J. Browning, R. Ramakrishnan, O.A. von Lilienfeld, U. Roethlisberger, Genetic Optimization of Training Sets for Improved Machine Learning Models of Molecular Properties, The Journal of Physical Chemistry Letters 8(7) (2017) 1351-1359.

[55] J. Behler, Constructing high-dimensional neural network potentials: A tutorial review, International Journal of Quantum Chemistry 115(16) (2015) 1032-1050.

[56] R. Ramakrishnan, P.O. Dral, M. Rupp, O.A. von Lilienfeld, Quantum chemistry structures and properties of 134 kilo molecules, Scientific Data 1 (2014) 140022.

[57] C.C. Fischer, K.J. Tibbetts, D. Morgan, G. Ceder, Predicting crystal structure by merging data mining with quantum mechanics, Nat Mater 5(8) (2006) 641-646.