

Data-Driven Discovery of Photoactive Quaternary Oxides Using First-Principles Machine Learning

Daniel W. Davies,[†] Keith T. Butler,[‡] and Aron Walsh^{*,†,§}

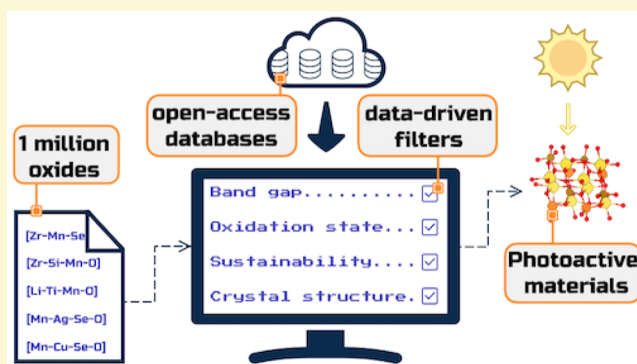
[†]Department of Materials, Imperial College London, Exhibition Road, London SW7 2AZ, U.K.

[‡]SciML, Scientific Computing Division, Rutherford Appleton Laboratory, Harwell Oxford, Didcot, Oxfordshire OX11 0QX, U.K.

[§]Global E³ Institute and Department of Materials Science and Engineering, Yonsei University, Seoul 120-749, Korea

Supporting Information

ABSTRACT: We present a low-cost, virtual high-throughput materials design workflow and use it to identify earth-abundant materials for solar energy applications from the quaternary oxide chemical space. A statistical model that predicts bandgap from chemical composition is built using supervised machine learning. The trained model forms the first in a hierarchy of screening steps. An ionic substitution algorithm is used to assign crystal structures, and an oxidation state probability model is used to discard unlikely chemistries. We demonstrate the utility of this process for screening over 1 million oxide compositions. We find that, despite the difficulties inherent to identifying stable multicomponent inorganic materials, several compounds produced by our workflow are calculated to be thermodynamically stable or metastable and have desirable optoelectronic properties according to first-principles calculations. The predicted oxides are $\text{Li}_2\text{MnSiO}_5$, $\text{MnAg}(\text{SeO}_3)_2$, and two polymorphs of $\text{MnCdGe}_2\text{O}_6$, all four of which are found to have direct electronic bandgaps in the visible range of the solar spectrum.



INTRODUCTION

The past decade has seen the construction of extensive databases for computed materials properties from quantum mechanical calculations.^{1–6} These databases have enabled the virtual screening of thousands of compounds for new target properties in the fields of photovoltaics,^{7–9} solar fuels,^{10–14} thermoelectrics,^{15–17} and others.^{18,19} They are also facilitating the move toward predictive materials design using data-mining and machine learning (ML). A growing infrastructure of ML tools has enabled its application to complex problems across many areas of molecular and materials science.²⁰ This includes building models that relate readily available descriptors to desirable properties including bandgap,^{21–24} thermodynamic stability,^{25–27} thermal transport properties,^{28,29} and the probability for crystal structure types to form.^{30,31} These approaches constitute computationally affordable ways to explore the vast chemical space that is otherwise intractable to high-throughput first-principles computation.³²

While the development of more advanced statistical techniques for chemical and materials science continues,³³ it is already possible to add ML models to the list of tools that can be used in materials design workflows. In this paper, we present a virtual high-throughput screening process in which ML joins the ranks of other data-driven models and density functional theory (DFT) calculations to constitute a hierarchy of filtering stages. The overall workflow is capable of translating from a compositional search space of over 1 million quaternary

oxides ($\text{A}_w\text{B}_x\text{C}_y\text{O}_z$) to compounds predicted to have target optoelectronic properties by explicit quantum-mechanics calculations.

Our workflow consists of five steps. In the first, which deals with the largest number of configurations, a ML model is used to screen for compositions predicted to have a bandgap within a window for potential applications for solar energy conversion. The next stage of filtering, illustrated in Figure 1, combines multiple low-cost data-driven approaches to further reduce the search space. We make use of the Herfindahl Hirschman index of resource availability (HHI_R)³⁴ to focus on the most sustainable element compositions. Two established models are used to assign high-ranking compositions to likely crystal structures³⁵ and then assess the feasibility of these new compounds in terms of oxidation states.³⁶ Finally, automated electronic structure calculations are carried out in order to accurately predict the thermodynamic stability and bandgap of candidate materials. We demonstrate the overall process by screening 1.1 million quaternary oxide compositions to identify four new compounds with suitable bandgaps for solar energy applications comprised of earth-abundant elements. These data-driven approaches are used to drastically reduce the

Received: April 16, 2019

Revised: August 23, 2019

Published: August 26, 2019

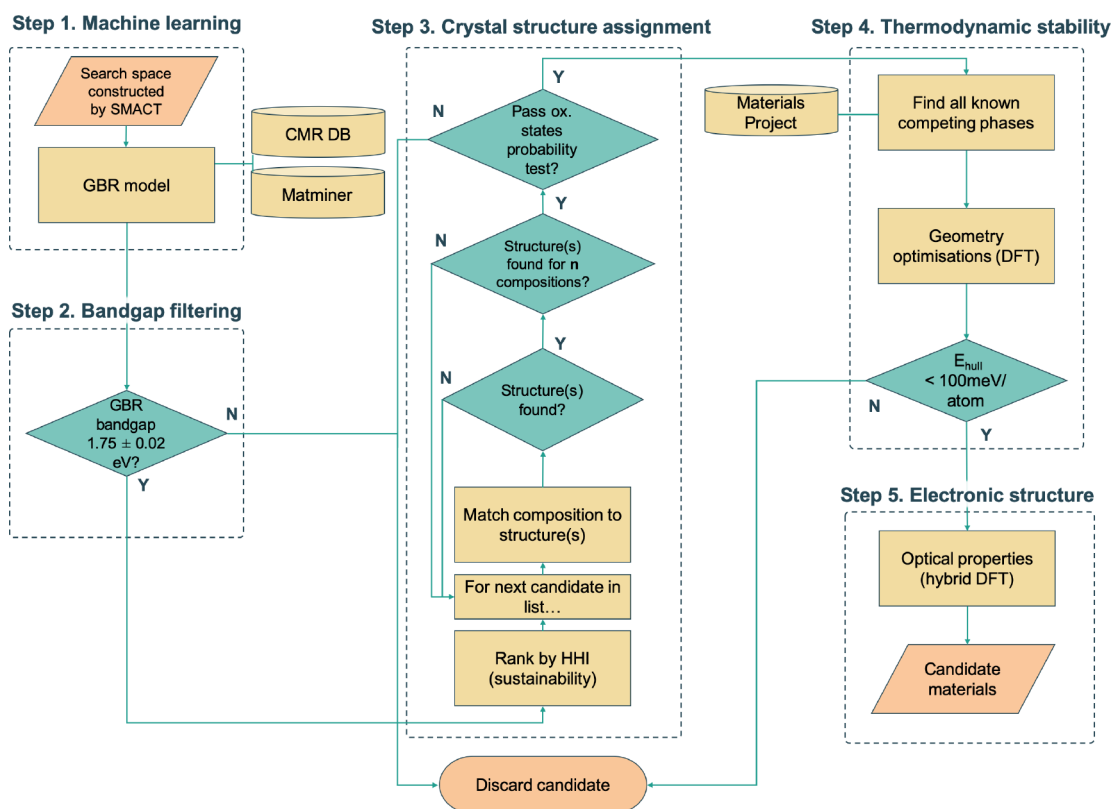


Figure 1. Computer-aided design workflow. Data from the Computational Materials Repository (CMR) database is used in conjunction with *Matminer*³⁷ to construct a gradient boosting regression (GBR) model (step 1), which is then used as a bandgap filter (step 2). Compositions are ranked using the Herfindahl Hirschman index (HHI_R),³⁴ appropriate structures are generated with a structure substitution algorithm,³⁸ and a probabilistic oxidation state model filters out unlikely species combinations (step 3). Thermodynamic stability (step 4) and bandgaps (step 5) are calculated from first-principles using semilocal density functional theory (DFT) and nonlocal hybrid DFT.

required computational resources compared to a brute-force first-principles investigation.

Step 1: Machine Learning Model of Oxide Bandgaps.

Supervised ML can be used to build statistical models that relate input values (features) to target values (labels) for a set of training samples. These models can then be used predictively given new data. There exists a wide variety of supervised ML approaches, many of which are being applied to numerous problems relating to first-principles materials modeling.²⁰ We now provide a brief outline of the key concepts and training procedure needed to build a gradient boosting regression (GBR) model, which is employed in this work to predict bandgaps from chemical compositions. The GBR model is trained and subsequently applied using the *scikit-learn* Python library.³⁹

Model Structure. In GBR, an ensemble of individual weak learners (usually decision trees) is used. By *weak* learners, we mean that each individual learner has poor predictive power if applied in isolation. When building decision trees, the goal is to predict the value of sample labels by learning simple decision rules from the sample features. Individual trees are constructed using the classification and regression trees (CART) algorithm.⁴⁰ In brief, for a given node of a decision tree (Figure 2a), the sample space is split into two parts that are as homogeneous as possible according to their labels. A decision rule involving one of the sample features is selected to best achieve this goal, i.e. to minimize the impurity of the node. This process is carried out recursively until some stopping criterion is met. For regression, the mean value of the ground

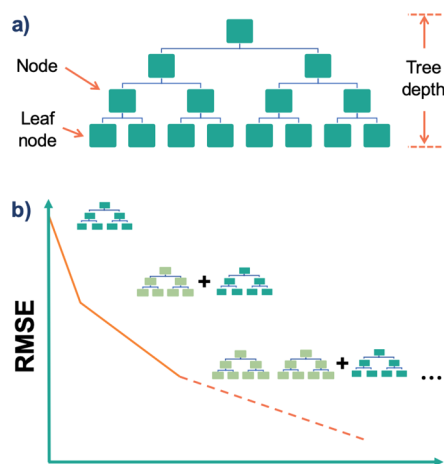


Figure 2. Schematic of the gradient boosting regression (GBR) model. (a) A decision tree splits the sample space recursively at nodes based on feature values, grouping samples into leaf nodes. (b) Multiple decision trees are constructed during the GBR process with each consecutive tree trained on the residuals of the existing model, minimizing the root-mean-squared error (RMSE).

truth labels at a given leaf node is taken as the prediction of the model for samples at that node.

One problem with decision trees—and indeed the reason that they fall into the category of *weak* learners—is that, by splitting the sample space on the basis of one feature at each node, they fail to include predictive power from multiple,

overlapping regions of feature space. As such, decision trees of a small depth tend to ignore valuable information from unused features, while those of a large depth are likely to be fit to random noise in the data set. This shortcoming is countered within GBR by constructing multiple decision trees sequentially. As depicted in Figure 2b, the overall model is built by adding trees in a forward, stagewise fashion with each consecutive tree trained not on the sample labels but on the residuals of the current model. The result is that each consecutive tree can consider the whole sample space and serves to improve the overall performance of the model by minimizing a chosen loss function, in this case the root-mean-squared error (RMSE).

Data Representation. The target property that we wish to predict is the bandgap calculated using the GLLB-sc functional.⁴¹ One possible drawback of this choice of functional is that it is only formulated in the spin unpolarized case, so it may not perform well for materials containing magnetic ions. Nonetheless, it has been shown to give more reliable estimations of bandgap than semilocal DFT functionals that operate within the generalized-gradient approximation (GGA) for a range of chemistries.^{11,13} The bandgap estimates obtained using this functional are closer to those from expensive hybrid DFT and GW frameworks than they are to those from GGA functionals, and at a fraction of the computational cost. We direct the interested reader to ref 13 for details on the accuracy of this method.

The bandgap values produced by Castelli et al. are used as a training set¹³ and are available from the Computational Materials Repository (CMR) database.³ This set is comprised of 2289 inorganic materials, 799 of which are oxides (i.e., contain oxygen and at least one other element), which are used as training data. Larger data sets than this that contain oxide bandgaps do exist but are primarily the result of GGA calculations. This data set therefore represents an appropriate compromise between quality of data and quantity of data points.

The compositions of the materials are represented using the element properties from the *Magpie* package.⁴² The features used are the minimum, maximum, range, mean, mode, and mean absolute deviation (MAD) of atomic number, Mendeleev number, atomic mass, melting temperature, and electronegativity, among others (see the Supporting Information of ref 42 for the full list). This set also includes spacegroup number, which refers to the spacegroup number of the element in its 0 K ground state. It should be noted that spacegroups are given numeric labels in which chemical information may be embedded given that they are directly related to element coordination via its crystal structure type. However, features produced by calculating the mean, range, etc., of these numeric labels are arguably not as meaningful as for other properties such as the boiling point and atomic radius, which are more physically motivated. We include them in this study so as to make use of the full *Magpie* feature set and note that a more full investigation into the effect of different feature sets on this type of model would constitute interesting further work.

In addition, we use the number of valence electrons, the elemental frontier orbital energies calculated from neutral atoms with DFT, and the bandgap center position calculated using the geometric mean of electronegativities as demonstrated by Nethercot.⁴³ All of the 149 features are generated using the *Matminer* Python library³⁷ and can all be calculated using knowledge of the elemental composition

along with the element properties present within the same library.

Finally, it is important to note that predicting the bandgap of a chemical composition is not a well-defined problem; due to the fact that various polymorphs can exist for a given composition, no unique solution can be given in the absence of structural information. The training data set used contains ground state polymorphs only, so a model built using this data will specifically relate chemical composition to the bandgap of the lowest energy phase. The extent to which polymorphism limits the ability to predict bandgap from composition is discussed further in the following section.

Model Training. While the model parameters are set automatically during the learning process as described above, several key *hyperparameters* must be chosen at the start. For GBR, as well as the self-explanatory tree-specific hyperparameters, there are three key boosting parameters (Table 1). The fraction of compounds to fit each tree dictates the

Table 1. Hyperparameter Values Used in the Final GBR Model for Oxide Bandgap Prediction

tree-specific parameters	
min. compounds needed to split nodes	65
max. depth of tree	20
min. compounds required at leaf nodes	1
max. features considered per tree	86
boosting parameters	
fraction of compounds to fit each tree	0.9
learning rate	0.01
number of decision trees	1000

maximum number of samples in the training set that any individual tree can use, introducing some level of diversity into the ensemble, which helps to mitigate against overfitting. The number of decision trees and learning rate refer to the number of boosting stages used in the final ensemble and the factor by which the contribution of each new tree is multiplied, respectively. The overall model is given by

$$F(x) = \sum_{m=1}^M \gamma_m h_m(x) \quad (1)$$

where M is the total number of decision trees, $h_m(x)$ are the individual trees, and γ_m is the learning rate.

The total error in ML approaches comes from a combination of bias, variance, and irreducible errors. Gradient boosting reduces the bias of individual trees but runs the risk of increasing the variance (error from sensitivity to noise in the training data). Upon changing a given hyperparameter, it is crucial to check how the model performs on unseen data, even if the fit to the training data appears to be improving (see Figure 3). Each time a model is built using a trial set of hyperparameters, 10-fold cross-validation (CV) is performed whereby the model is trained on 90% of the data and then tested on the remaining 10%. This process is repeated such that every 10% chunk of data is used for testing; then, the mean RMSE is calculated.

Optimal hyperparameter values for this GBR model were found by Bayesian optimization and are listed in Table 1. This was achieved using the *scikit-optimize* Python library⁴⁴ and involves approximating the model using Gaussian processes. The next set of hyperparameters to trial is chosen by

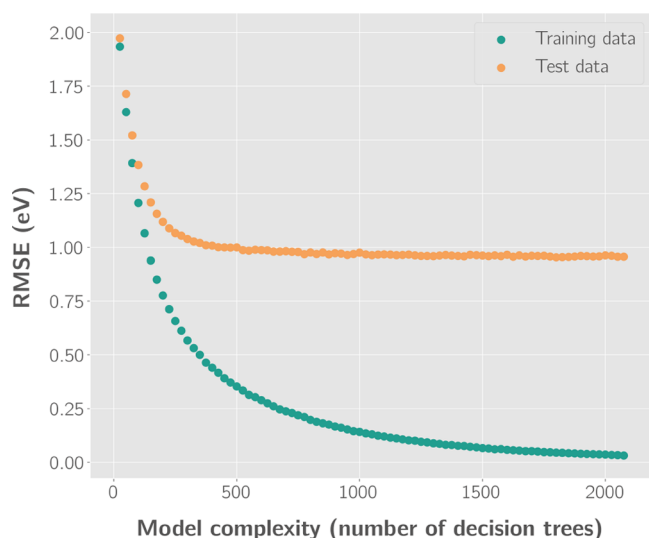


Figure 3. Effect of the number of decision trees (boosting stages) on model performance for bandgap prediction. At a certain threshold (~ 250), increasing the number of trees in the model ceases to improve its performance on unseen test data, even though it appears to better fit the training data.

an acquisition function over the Gaussian prior, which is cheaper to evaluate than the model itself. A more detailed explanation of this approach can be found in ref 45. Using these parameters, as well as removing oxide gases such as CO_2 and SO_2 and complex anions containing uncommon oxidation states such as phosphites and perphosphates, yields a final model with an RMSE of 0.95 eV. We note that the scope of the model will be limited by excluding these compounds containing uncommon oxidation states, to ensure that they are also excluded from the search space when the model is used predictively.

Finally, it might be assumed that the correlation between bandgap calculated using GGA exchange-correlation functionals, which tend to be consistently underestimated, and that calculated using GLLB-sc could be high enough to use predictively. If this were the case, a ML model could be trained using a larger database, such as the Materials Project (MP), which contains $\sim 86,000$ inorganic structures with bandgaps calculated using the PBE functional.⁴⁶ We find that, while the expected linear relationship is observed between bandgaps calculated using PBE and GLLB-sc, there is significant deviation from the relationship and that this is larger in general for oxides (see Figure S1 in the Supporting Information). The standard deviation is 0.85 eV; thus, for a two-step approach to be advantageous, the RMSE of the model trained on the large data set of PBE bandgaps would have to be unreasonably low (< 0.1 eV).

Model Performance and Limitations. Features representing the crystal structures of inorganic compounds to ML algorithms are the subject of much recent development.^{47–50} The use of such features has been shown to improve the predicted properties of inorganic solids beyond compositional representations alone. As such, the accuracy of our model is limited because atomic connectivity is not accounted for. This effect is particularly prevalent for oxides, as their structural diversity results in a wide variety of local bonding arrangements. We have quantified this phenomenon by showing that the unscreened Madelung site potential of the oxide anion—a

quantity that reflects the electrostatic potential of an ion in a crystal by approximating ions as point charges—varies across all binary metal oxides with a striking range of 16 V.⁵¹

As discussed previously, the training data contains ground-state polymorphs only, which means that a perfect model would have a RMSE of 0 eV, predicting the bandgap of the ground state polymorph perfectly. However, for the more general practical case of mapping composition to bandgap, polymorphism will provide an inherent limit to the RMSE that can be achieved. The distribution of the maximum (PBE) bandgap difference between polymorphs for all oxide compositions in the MP database is shown in Figure 4.

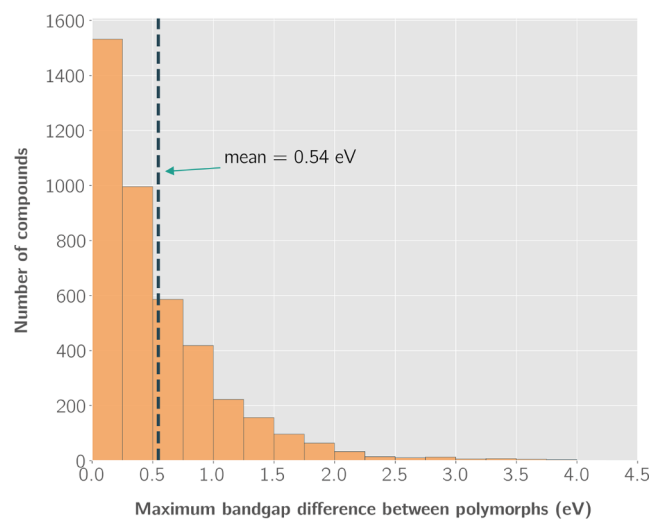


Figure 4. Distribution of the maximum bandgap difference between polymorphs for oxides in the Materials Project database that exhibit polymorphism. Only compounds with an energy above the convex hull of < 0.1 eV and a maximum bandgap difference of > 0.05 eV are included. Bandgaps are calculated in the Materials Project using the GGA-PBE functional.

While for a large number of oxides, polymorphism results in a bandgap difference of < 0.5 eV, the difference can be as large as 4.18 eV (e.g., LiFePO_4) and the mean difference is 0.57 eV. While it is not possible to translate this result directly to an optimal RMSE value of a model trained using the GLLB-sc functional, this analysis highlights the extent to which crystal structure plays a role in dictating bandgap. A model that considers chemical composition alone can only be used as a prescreening filter, and in this context, a composition-only model with an RMSE of 0.95 eV is reasonable.

The model is also limited by the number of available training samples. The data set used constitutes a compromise between quality of data (the GLLB-sc DFT functional is used as opposed to a GGA functional) and quantity of data points, so it is likely that a lower RMSE could be achieved if more training samples were available (see Figure S2). We note also that inclusion of the nonoxide compounds from the CMR data set did not improve the predictive power of the model for oxide bandgaps. There are a number of emerging approaches that could be examined in the future to be used to negate the data quality/quantity compromise. Several recent studies have demonstrated the application of multifidelity data sets, for example, combining theoretical and experimental data or theoretical data from several levels of theory. Gaussian processes have been applied to fuse data sets for prediction

of absorption energies,⁵² and “combined quantum machine learning” was shown to be able to incorporate a hierarchy of theoretical models for enhanced predictive power on atomization energies.⁵³ In the field of neural networks, transfer learning has proved exceptionally popular and successful. Active transfer learning was recently applied to fuse DFT and a small set of CCSDT data to obtain a potential model with accuracy approaching the CCSDT level on a test set.⁵⁴

It is also instructive to compare this approach to existing heuristic methods. For example, the solid state energy (SSE) scale,^{55,56} derived from the relationship between electron affinity (EA) and ionization potential (IP) and bandgap for a selection of binary closed-shell inorganic semiconductors and insulators, can be used to estimate bandgaps for new compounds.⁵⁷ The SSE has knowledge only of the EA and IP values of the constituent cations and anions, respectively. The range and standard deviation of IP values for the 56 binary oxides used in the construction of the SSE model are 4.9 and 1.44 eV, respectively, giving O the largest associated uncertainty of all of the anions featured. For this reason, there is no correlation between the bandgap predicted using the SSE scale and the GLLB-sc bandgap of the 799 oxides in the training data set (see Figure S3). By taking into account more information about the constituent elements, the GBR model we developed is able to predict bandgaps to a higher level of accuracy.

We emphasize that this ML model constitutes the first computationally cheap step in a hierarchy of screening steps, in which we specifically target properties in a structure-agnostic fashion in the first instance. As such, the RMSE of 0.95 eV obtained, the improvement upon existing models such as the SSE scale, and the improvement upon randomly selecting oxides (shown in Figure 6 in the next section) makes it a suitable approach for this purpose.

Finally, we can inspect which features are most important in the final GBR model using the mean decrease impurity method (or gini importance).⁴⁰ Figure 5 shows the mean absolute deviation (MAD) of covalent radius is the most important

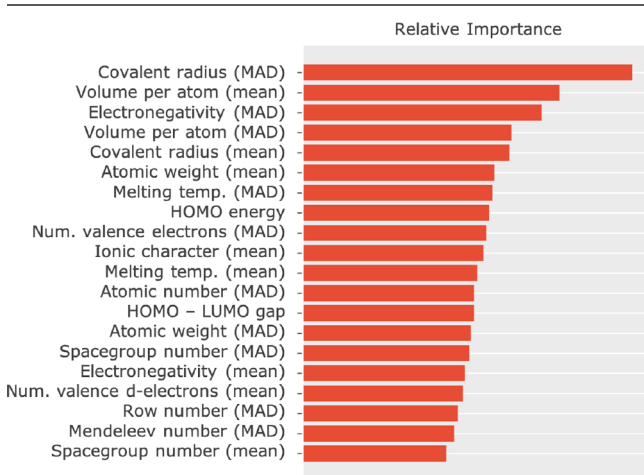


Figure 5. Relative importance of the 20 most important features in the final gradient boosting regression model. HOMO and LUMO energies refer to highest occupied and lowest unoccupied molecular orbital energies as calculated within DFT, respectively (taken directly from the Matminer Python library). Ionic character refers to Pauling’s empirical ionic character between pairs of atoms calculated using electronegativities.⁵⁸

feature. The mean value for volume per atom and MAD of melting temperature are also relatively important. The extent to which this can be interpreted as meaningful depends on how highly correlated the features are. For example, covalent radius and volume per atom are strongly correlated, which makes it harder to decouple their contributions to the overall model. In general, a number of features contribute significantly to the final model. Investigation into the effect of systematically removing correlated features, and retraining the model, is an avenue for further study and a means of extracting physically intuitive relationships.

Step 2: Bandgap Filter. We now use the trained GBR model to search for promising candidates from a large search space. A pool of 1.1 million hypothetical quaternary oxide compositions was generated using the SMACT Python library, implementing the heuristic chemical rules employed in that code.³² The target bandgap range of 1.0–2.5 eV will capture a wide range of photoactive materials. Smaller gaps may be more suitable for single-junction photovoltaic applications, while wider gaps could be used in tandem systems or solar fuel processes.^{59,60}

To test the efficacy of our model we perform a baseline comparison against a random sampling of materials from the data set. The distribution of errors obtained using the GBR model is shown in Figure 6a. Materials predicted to have a bandgap at the center of the target range (1.75 eV) have a 60% probability of having a GLLB-sc bandgap within the range. In

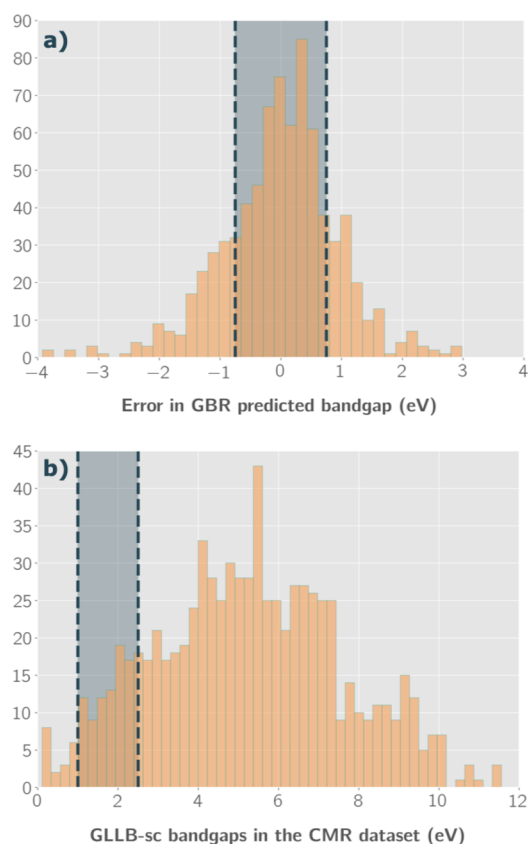


Figure 6. (a) Distribution of error in predicted bandgap by the trained GBR model. The shaded region corresponds to an error of ± 0.75 eV and encloses 60% of all predictions. (b) Distribution of GLLB-sc bandgaps for oxides in the CMR training data set. The shaded region corresponds to a bandgap of 1.75 ± 0.75 eV.

contrast, Figure 6b shows the distribution of bandgaps of all oxides in the CMR data set and the probability of choosing one at random with a bandgap in the target range is just 8%.

We filter the 1.1 million candidates that do not have a predicted bandgap of 1.75 ± 0.02 eV. A narrow window is chosen because we do not aim to capture all of the hypothetical compositions that fall within the target bandgap range. Rather, those compositions that are most likely to have bandgaps within the range according to the GBR model are targeted. If we assume the errors in Figure 6a follow a normal distribution, then the materials with predicted bandgaps in this window have a 60% probability of having a GLLB-sc bandgap within the target range, as already stated. As we widen the window, this probability decreases. For example, for materials with predicted bandgaps of 1.35 and 2.15 eV, the probability is reduced to 55%, and for 1.0 and 2.5 eV, it is 45%. While this is not a very large reduction in probability, we choose to minimize false positive results as opposed to minimizing false negative results. In the language of classification models, we prioritize precision over recall. This decision is in keeping with the need to drastically reduce the search space so that a manageable number of compositions are taken forward to subsequent screening stages and leads to a greater than 60-fold reduction of the search space, leaving 17,833 candidates.

Step 3: Crystal Structure Assignment. The surviving 17,833 compositions are ranked by sustainability using the HHI_R scale.³⁴ Starting with the most sustainable composition, chemically plausible quaternary oxide crystal structures are constructed using the structure substitution algorithm developed by Hautier and co-workers.³⁸ This algorithm uses a statistical model and relies on a database of known compounds including oxidation state information: A combination of species (elements in a given oxidation state) of interest is substituted onto lattice sites in known structures from the data set of known materials. Each species substitution is associated with a certain probability, which comes from a statistical model trained on the compounds that already exist in the Inorganic Crystal Structure Database (ICSD). If the overall probability for a given set of substitutions is above a certain threshold, it is added to a list of possible structures. We use a threshold of 1×10^{-5} as recommended in the original paper.³⁸

For each predicted structure, a second probabilistic model⁵⁷ is applied as an additional filter, to check that the combination of species in structure is chemically plausible. This model ascribes probabilities to the formation of hypothetical compounds given the oxidation states of the constituent species and was trained on a set of 16,700 ordered, inorganic compounds. The probabilities are derived from the observed frequency of occurrence of cation–anion pairs, and only the most electronegative anion is considered in each compound. A low probability threshold of 0.005 is used. As per the results in the original paper,⁵⁷ this has the effect of eliminating only very unlikely species combinations, i.e., those that appear less than once or twice in the training data. We also choose to eliminate Ti^{3+} compounds due to the d^1 electronic configuration being linked to fast electron–hole recombination for solar applications.

The above procedure was repeated until 235 candidate materials were generated, corresponding to 61 unique chemical compositions. This pool is small enough to allow for explicit first-principles calculations, and we take these candidate materials forward to calculate their thermodynamic stability. All of these crystal structures are available in the Supporting

Information data repository in the form of Pymatgen Structure objects saved in a json file.

Step 4: Thermodynamic Stability. Competing phases are identified using the chemical potentials from the MP database. Then, full geometry optimization is carried out on candidate compounds and all competing phases using DFT at the GGA (PBEsol) level, with an equivalent computational setup. This is done in high-throughput using the *Atomate*⁶¹ and *Fireworks*⁶² Python libraries. The DFT total energies are used to determine thermodynamic stability via the distance from the 3D convex hull of the quaternary phase diagram.

The significance of thermodynamic (meta)stability has come to the forefront of materials discovery research recently, with various studies using existing materials databases to assess the limits of metastability and its relationship to the likelihood of compounds to be realized experimentally.^{63,64} For example, Sun et al.⁶³ estimate that around half of all compounds in the ICSD do not lie on the convex hull and are therefore thermodynamically metastable. Interestingly, they show that there is a large distribution of unobserved, low-energy hypothetical polymorphs within the energy spectrum of observed polymorphs. This means that metastability alone is not a sufficient criterion to predict whether a compound can be synthesized. Crucially, they also find that there is an approximately exponentially decreasing probability distribution of metastability versus frequency of known compounds, so defining some upper limit of metastability can be useful in screening workflows.

We use a metastability window of 100 meV/atom above the convex hull. This window covers 90% of compounds in the Materials Project database that represent fully characterized structures in the ICSD and is consistent with the findings of Sun et al.⁶³ This limit has been used previously in the discovery of ternary halides,³⁶ chalcogenides,⁵⁷ and transparent conductors of various chemistries.⁶⁵

Of the 235 compounds, 27 are calculated to be within the predefined metastability window of the convex hull. Four of the 27 compounds were found to be structurally identical to one other compound in the set, leaving 23 unique compounds, corresponding to 8 distinct compositions. The presence of identical structures can occur when different parent structures are found for one composition using the structure substitution algorithm, which then ultimately yield the same crystal structure following geometry optimization.

The relatively small proportion of stable and metastable compounds is unsurprising given the existence of a large number of stable binary and ternary oxides that act as competing phases. The energies above the convex hull for all 23 compounds are given in Table S1. Only one compound, $\text{Li}_2\text{MnSiO}_5$, has been previously reported in the MP database but has not been synthesized experimentally to the authors' knowledge. Shown in Figure 7a, the compound $\text{ZrMnSi}_2\text{O}_7$ is the only one predicted to be thermodynamically stable, while a second polymorph of $\text{ZrMnSi}_2\text{O}_7$ along with a $\text{Li}_2\text{TiMnO}_4$ structure are predicted to be <10 meV/atom above the convex hull, as shown in Figure 7b and c, respectively.

While three polymorphs of $\text{Li}_2\text{TiMnO}_4$ are in the MP database, including one that has been investigated as a possible active material for Li-ion battery applications,⁶⁶ none of the crystal structures adopted by the candidate compounds have previously been reported. The new phase of $\text{Li}_2\text{TiMnO}_4$ differs from the three previously reported polymorphs, as the metals are in tetrahedral environments as opposed to octahedral. It

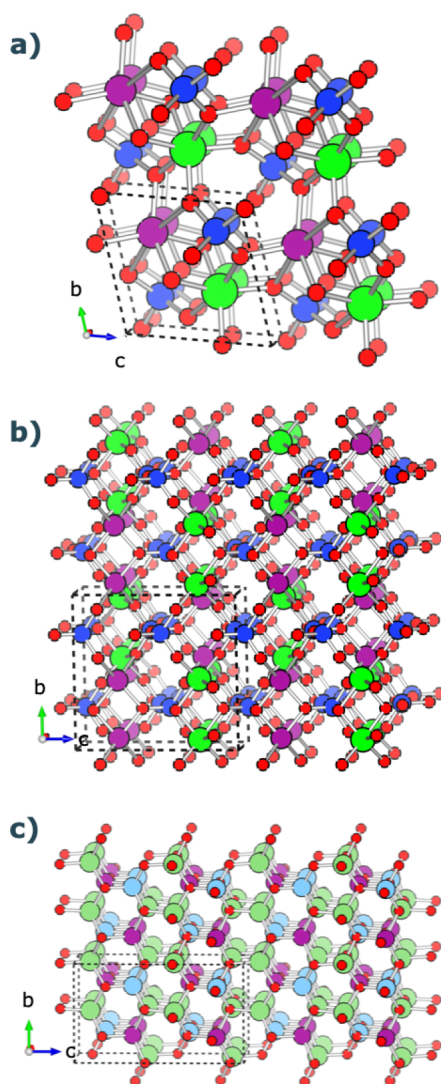


Figure 7. Three most stable compounds identified by the workflow. (a and b) Different polymorphs of $\text{ZrMnSi}_2\text{O}_7$ in which Si, Zr, and Mn atoms are depicted as blue, green, and purple circles, respectively. (c) A $\text{Li}_2\text{TiMnO}_4$ structure in which Li, Ti, and Mn atoms are depicted as green, blue, and purple circles, respectively. O atoms are red circles in all three structures.

also has a wide electronic bandgap of 4.21 eV, as calculated using a hybrid DFT functional in the following section, whereas the previously reported compounds all have PBE-calculated bandgaps of less than 0.4 eV. To the best of our knowledge, no compounds have previously been reported for any of the other seven compositions.

Step 5: Electronic Structure. The electronic structures of the 23 remaining candidate compounds were calculated with the HSE06 hybrid functional.^{67,68} While being more computationally demanding, this approach yields more accurate electronic structure information than what is available in current materials databases. The majority of compounds have a calculated bandgap of >4 eV, which is well outside the target bandgap window (see Table S1). The bandgaps of the 23 candidates are distributed closer to the bandgap range of interest (1–2.5 eV) than the oxides in the training data given in Figure 6, with a range of 1.76–5.12 eV vs 0–12 eV, which is promising. Four of the compounds are calculated to have bandgaps within the target window and are listed in Table 2.

The most thermodynamically stable compound with a bandgap within the target range is $\text{MnAg}(\text{SeO}_3)_2$ and is shown in Figure 8.

Table 2. Summary of Compounds Found to Have (Predicted HSE06/DFT) Bandgaps That Fall within the Target Window of 1.0–2.5 eV

formula	space group	E_{hull} (meV/atom)	bandgap (eV)
$\text{Li}_2\text{MnSiO}_5$	$P4/nmm$	86	2.24
$\text{MnCdGe}_2\text{O}_6$	$P2_1/c$	99	2.47
$\text{MnCdGe}_2\text{O}_6$	$C2/c$	99	1.76
$\text{MnAg}(\text{SeO}_3)_2$	$Pna2_1$	36	2.31

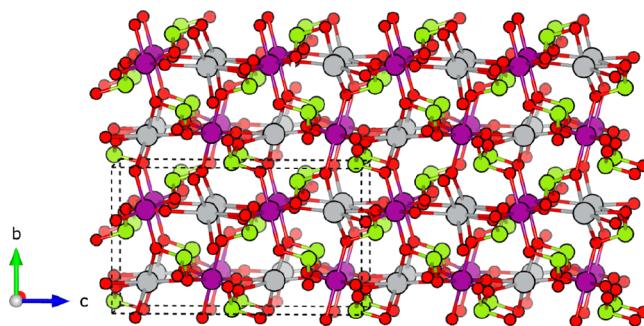


Figure 8. The most stable compound identified by the high-throughput workflow with a bandgap within the target window, $\text{MnAg}(\text{SeO}_3)_2$. Mn, Ag, Se, and O atoms are depicted as purple, silver, green, and red circles, respectively.

Encouragingly, the four compounds with bandgaps in the target range include three different compositions. Since the original GBR model is trained on composition alone, this indicates a coarse 37.5% success rate, based on the fact that only eight distinct compositions were represented by the final 23 candidate compounds. While the success rate is not as high as the original 60%, as indicated by the 10-fold CV results, the latter should be considered a maximum achievable success rate when using this model predictively. Cross-validation can give some indication of model performance, but there are limited options to glean further insight before applying the model predictively where existing data is scarce.

Crucially, this study represents a small sample size, making it impossible to draw strong conclusions. Qualitatively, it is promising that we have identified four candidate compounds using only 235 first-principles calculations, given the “needle in a haystack” nature of the problem. Without using the data-driven screening stages, computationally prohibitive structure optimization calculations would have been required for each of the top compositions suggested by the ML model. The overall virtual high-throughput screening process constitutes a multi-objective optimization, in which bandgap, sustainability, and stability are all targeted sequentially (see Figure 1). The latter of these, stability, is likely to be a significant bottleneck for any screening of quaternary materials as compared with binary or ternary phases, given the expected lower stability window due to an increase in possible decomposition pathways.

Finally, the model was trained on bandgaps calculated using the GLLB-sc functional, whereas the bandgaps of the new compounds are calculated using the HSE06 functional. In the original work by Castelli et al. in which they calculate the bandgaps used here for training data, they show that bandgaps

calculated using HSE06 and GLLB-sc are generally in good agreement.¹³ However, they also show that, for smaller bandgaps such as those considered here, the GLLB-sc functional has a tendency to be underestimated as compared with the HSE06 functional. This could be another reason for getting a lower success rate and would also explain why no compounds had bandgaps calculated using HSE06 smaller than the target window (<1.0 eV). The availability of a comprehensive database with electronic and thermodynamic properties of materials at a consistent high level of theory would greatly benefit the training of ML models and future data-driven studies.

CONCLUSION

We outlined a multistage computational procedure to reduce a chemical space of over 1 million compositions to four target compounds using a combination of techniques and chemical filters. The majority of the study has been performed on a single-processor workstation. A GBR model was trained to predict bandgaps for quaternary oxide compositions. This model is shown to outperform established chemical heuristics for the ability to predict bandgap and allows for a 60-fold reduction of the initial search space, with an order of magnitude better chance of identifying suitable compounds compared to random filtering. Additional screening based on sustainability, oxidation state combinations, and thermodynamic stability was used, before performing high-quality electronic structure calculations on a pool of 23 candidate materials. Finally, we identified four new quaternary oxides not previously reported or explored for solar energy applications. The workflow that we present here can be a blueprint for using a combination of machine learning and first-principles calculations to allow efficient, targeted screening of the vast chemical structure–composition hyperspace.

COMPUTATIONAL METHODS

Full information on the workflow is available in the [Supporting Information](#). It makes use of the Python libraries [SMACT](#),³² [Pymatgen](#),⁶⁹ [Matminer](#),³⁷ [Scikit-learn](#),³⁹ [Atomate](#),⁶¹ and [Fireworks](#).⁶²

Electronic Structure Calculations. First-principles calculations are carried out using Kohn–Sham DFT with a projector-augmented plane wave basis⁷⁰ as implemented in the Vienna Ab-initio Simulation Package (VASP).^{71,72} We use the PBEsol exchange–correlation functional,⁷³ and a k -point grid is generated for each calculation with a density of 120 \AA^3 in the reciprocal lattice. The kinetic-energy cutoff is set at 600 eV, and the forces on each atom are minimized to below $0.005 \text{ eV \AA}^{-1}$.

Semilocal exchange–correlation treatments such as the PBEsol functional provide an accurate description of crystal structures but tend to underestimate the electronic bandgaps of semiconductors. To overcome this issue, more accurate electronic structure calculations are performed using the hybrid nonlocal functional HSE06,⁶⁸ which includes 25% screened Hartree–Fock exact exchange. Γ -Centered homogeneous k -point grids are used with a density of 64 \AA^3 in the reciprocal lattice, and the kinetic energy cutoff is set at 520 eV.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the [ACS Publications website](#) at DOI: [10.1021/acs.chemmater.9b01519](https://doi.org/10.1021/acs.chemmater.9b01519).

Plot showing PBE vs GLLB-sc calculated bandgap for oxides, plot showing the RMSE of the GBR model after

10-fold cross-validation as a function of training set size, plot showing GLLB-sc bandgap vs GBR-predicted bandgap using the final model, plot showing GLLB-sc bandgap vs SSE-predicted bandgap, and table of the most stable compounds found at the end of the workflow ([PDF](#))

AUTHOR INFORMATION

Corresponding Author

*E-mail: a.walsh@imperial.ac.uk.

ORCID

Daniel W. Davies: 0000-0003-4094-5992

Keith T. Butler: 0000-0001-5432-5597

Aron Walsh: 0000-0001-5460-7033

Notes

The authors declare no competing financial interest.

Code showing all steps of the screening process is available in Jupyter notebook form at <https://doi.org/10.5281/zenodo.2609120>, while the SMACT code is available from <https://doi.org/10.5281/zenodo.2609134>.

ACKNOWLEDGMENTS

Via our membership of the UK's HEC Materials Chemistry Consortium, which is funded by EPSRC (EP/L000202), this work used the ARCHER UK National Supercomputing Service (<http://www.archer.ac.uk>) for all DFT calculations. D.W.D. is supported by the EPSRC via the Doctoral Prize Fellowship, and A.W. is supported by a Royal Society University Research Fellowship. This research was also supported by the Creative Materials Discovery Program through the National Research Foundation of Korea (NRF) funded by Ministry of Science and ICT (2018M3D1A1058536).

REFERENCES

- (1) Setyawan, W.; Curtarolo, S. High-throughput electronic band structure calculations: Challenges and tools. *Comput. Mater. Sci.* **2010**, *49*, 299–312.
- (2) Setyawan, W.; Gaume, R. M.; Lam, S.; Feigelson, R. S.; Curtarolo, S. High-Throughput Combinatorial Database of Electronic Band Structures for Inorganic Scintillator Materials. *ACS Comb. Sci.* **2011**, *13*, 382–390.
- (3) Landis, D. D.; Hummelshøj, J. S.; Nestorov, S.; Greeley, J.; Dullak, M.; Bligaard, T.; Nørskov, J. K.; Jacobsen, K. W. The computational materials repository. *Comput. Sci. Eng.* **2012**, *14*, 51–57.
- (4) Jain, A.; Ong, S. P.; Hautier, G.; Chen, W.; Richards, W. D.; Dacek, S.; Cholia, S.; Gunter, D.; Skinner, D.; Ceder, G.; Persson, K. A. Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL Mater.* **2013**, *1*, 011002.
- (5) Saal, J. E.; Kirklin, S.; Aykol, M.; Meredig, B.; Wolverton, C. Materials design and discovery with high-throughput density functional theory: The open quantum materials database (OQMD). *JOM* **2013**, *65*, 1501–1509.
- (6) The NoMaD Repository. <http://nomad-repository.eu/>, accessed Feb 9, 2017.
- (7) Yu, L.; Kokenyesi, R. S.; Keszler, D. A.; Zunger, A. Inverse Design of High Absorption Thin-Film Photovoltaic Materials. *Adv. Energy Mater.* **2013**, *3*, 43–38.
- (8) Krishnamoorthy, T.; Ding, H.; Yan, C.; Leong, W. L.; Baikie, T.; Zhang, Z.; Sherburne, M.; Li, S.; Asta, M.; Mathews, N.; Mhaisalkar, S. G. Lead-free germanium iodide perovskite materials for photovoltaic applications. *J. Mater. Chem. A* **2015**, *3*, 23829–23832.
- (9) Hinuma, Y.; Hatakeyama, T.; Kumagai, Y.; Burton, L. A.; Sato, H.; Muraba, Y.; Iimura, S.; Hiramatsu, H.; Tanaka, I.; Hosono, H.;

Oba, F. Discovery of earth-abundant nitride semiconductors by computational screening and high-pressure synthesis. *Nat. Commun.* **2016**, *7*, 11962.

(10) Castelli, I. E.; Landis, D. D.; Thygesen, K. S.; Dahl, S.; Chorkendorff, I.; Jaramillo, T. F.; Jacobsen, K. W. New cubic perovskites for one- and two-photon water splitting using the computational materials repository. *Energy Environ. Sci.* **2012**, *5*, 9034–9043.

(11) Castelli, I. E.; Olsen, T.; Datta, S.; Landis, D. D.; Dahl, S.; Thygesen, K. S.; Jacobsen, K. W. Computational screening of perovskite metal oxides for optimal solar light capture. *Energy Environ. Sci.* **2012**, *5*, 5814–5819.

(12) Wu, Y.; Lazić, P.; Hautier, G.; Persson, K.; Ceder, G. First principles high throughput screening of oxynitrides for water-splitting photocatalysts. *Energy Environ. Sci.* **2013**, *6*, 157–168.

(13) Castelli, I. E.; Hüser, F.; Pandey, M.; Li, H.; Thygesen, K. S.; Seger, B.; Jain, A.; Persson, K. A.; Ceder, G.; Jacobsen, K. W. New light-harvesting materials using accurate and efficient bandgap calculations. *Adv. Energy Mater.* **2015**, *5*, 1400915.

(14) Pandey, M.; Vojvodic, A.; Thygesen, K. S.; Jacobsen, K. W. Two-Dimensional Metal Dichalcogenides and Oxides for Hydrogen Evolution: A Computational Screening Approach. *J. Phys. Chem. Lett.* **2015**, *6*, 1577–1585.

(15) Toher, C.; Plata, J. J.; Levy, O.; de Jong, M.; Asta, M.; Nardelli, M. B.; Curtarolo, S. High-throughput computational screening of thermal conductivity, Debye temperature, and Grüneisen parameter using a quasiharmonic Debye model. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2014**, *90*, 174107.

(16) Sparks, T. D.; Gaultois, M. W.; Oliynyk, A.; Brgoch, J.; Meredig, B. Data mining our way to the next generation of thermoelectrics. *Scr. Mater.* **2016**, *111*, 10–15.

(17) Faghaninia, A.; Yu, G.; Aydemir, U.; Wood, M.; Chen, W.; Rignanes, G.-M.; Jeffrey, S.; Hautier, G.; Jain, A. A computational assessment of the electronic, thermoelectric, and defect properties of bournonite (CuPbSbS₃) and related substitutions. *Phys. Chem. Chem. Phys.* **2017**, *19*, 6743–6756.

(18) de Jong, M.; Chen, W.; Geerlings, H.; Asta, M.; Persson, K. A. A database to enable discovery and design of piezoelectric materials. *Sci. Data* **2015**, *2*, 150053.

(19) Miller, S. A.; Gorai, P.; Aydemir, U.; Mason, T. O.; Stevanović, V.; Toberer, E. S.; Snyder, G. J. SnO as a potential oxide thermoelectric candidate. *J. Mater. Chem. C* **2017**, *5*, 8854–8861.

(20) Butler, K. T.; Davies, D. W.; Cartwright, H.; Isayev, O.; Walsh, A. Machine learning for molecular and materials science. *Nature* **2018**, *559*, 547–555.

(21) Zhuo, Y.; Mansouri Tehrani, A.; Brgoch, J. Predicting the Band Gaps of Inorganic Solids by Machine Learning. *J. Phys. Chem. Lett.* **2018**, *9*, 1668–1673.

(22) Zhu, Z.; Dong, B.; Yang, T.; Zhang, Z.-D. Fundamental Band Gap and Alignment of Two-Dimensional Semiconductors Explored by Machine Learning. 2017, arXiv:1708.04766. arXiv.org e-Print archive. <https://arxiv.org/abs/1708.04766> (accessed August 1, 2019).

(23) Weston, L.; Stampfl, C. Machine learning the band gap properties of kesterite I₂–II–IV–V₄ quaternary compounds for photovoltaics applications. *Phys. Rev. Materials* **2018**, *2*, 085407.

(24) Lee, J.; Seko, A.; Shitara, K.; Nakayama, K.; Tanaka, I. Prediction model of band gap for inorganic compounds by combination of density functional theory calculations and machine learning techniques. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2016**, *93*, 115104.

(25) Legrain, F.; Carrete, J.; van Roekeghem, A.; Madsen, G. K. H.; Mingo, N. Materials Screening for the Discovery of New Half-Heuslers: Machine Learning versus Ab Initio Methods. *J. Phys. Chem. B* **2018**, *122*, 625–632.

(26) Seko, A.; Hayashi, H.; Tanaka, I. Compositional descriptor-based recommender system for the materials discovery. *J. Chem. Phys.* **2018**, *148*, 241719.

(27) Faber, F. A.; Lindmaa, A.; Von Lilienfeld, O. A.; Armiento, R. Machine Learning Energies of 2 Million Elpasolite (ABC₂D₆) Crystals. *Phys. Rev. Lett.* **2016**, *117*, 135502.

(28) Ju, S.; Shiga, T.; Feng, L.; Hou, Z.; Tsuda, K.; Shiommi, J. Designing Nanostructures for Phonon Transport via Bayesian Optimization. *Phys. Rev. X* **2017**, *7*, 021024.

(29) Legrain, F.; Carrete, J.; van Roekeghem, A.; Curtarolo, S.; Mingo, N. How Chemical Composition Alone Can Predict Vibrational Free Energies and Entropies of Solids. *Chem. Mater.* **2017**, *29*, 6220–6227.

(30) Ryan, K.; Lengyel, J.; Shatruk, M. Crystal Structure Prediction via Deep Learning. *J. Am. Chem. Soc.* **2018**, *140*, 10158.

(31) Liu, Y.-H.; van Nieuwenburg, E. P. Discriminative Cooperative Networks for Detecting Phase Transitions. *Phys. Rev. Lett.* **2018**, *120*, 176401.

(32) Davies, D. W.; Butler, K. T.; Jackson, A. J.; Morris, A.; Frost, J. M.; Skelton, J. M.; Walsh, A. Computational Screening of All Stoichiometric Inorganic Materials. *Chem.* **2016**, *1*, 617–627.

(33) Gubernatis, J. E.; Lookman, T. Machine learning in materials design and discovery: Examples from the present and suggestions for the future. *Physical Review Materials* **2018**, *2*, 120301.

(34) Gaultois, M. W.; Sparks, T. D.; Borg, C. K. H.; Seshadri, R.; Bonificio, W. D.; Clarke, D. R. Data-Driven Review of Thermoelectric Materials: Performance and Resource Considerations. *Chem. Mater.* **2013**, *25*, 2911–2920.

(35) Hautier, G.; Ong, S. P.; Jain, A.; Moore, C. J.; Ceder, G. Accuracy of density functional theory in predicting formation energies of ternary oxides from binary oxides and its implication on phase stability. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2012**, *85*, 155208.

(36) Davies, D. W.; Butler, K. T.; Isayev, O.; Walsh, A. Materials discovery by chemical analogy: role of oxidation states in structure prediction. *Faraday Discuss.* **2018**, *211*, 553–568.

(37) Ward, L.; et al. Matminer: An open source toolkit for materials data mining. *Comput. Mater. Sci.* **2018**, *152*, 60–69.

(38) Hautier, G.; Fischer, C.; Ehlacher, V.; Jain, A.; Ceder, G. Data mined ionic substitutions for the discovery of new compounds. *Inorg. Chem.* **2011**, *50*, 656–663.

(39) Pedregosa, F.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.

(40) Breiman, L. *Classification and Regression Trees*; Taylor & Francis Group: Boca Raton, FL, 1984; Vol. 1.

(41) Kuisma, M.; Ojanen, J.; Enkovaara, J.; Rantala, T. T. Kohn-Sham potential with discontinuity for band gap materials. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2010**, *82*, 115106.

(42) 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*, 16028.

(43) Nethercot, A. H. Prediction of Fermi energies and photoelectric thresholds based on electronegativity concepts. *Phys. Rev. Lett.* **1974**, *33*, 1088–1091.

(44) Head, T. Scikit-optimize. <https://scikit-optimize.github.io/>, accessed May 9, 2018.

(45) Rasmussen, C. E.; Williams, C. K. I. *Gaussian Processes for Machine Learning*, 1st ed.; MIT Press: Cambridge, MA, 2006.

(46) Perdew, J. P.; Burke, K.; Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **1996**, *77*, 3865–3868.

(47) Schütt, K. T.; Glawe, H.; Brockherde, F.; Sanna, A.; Müller, K. R.; Gross, E. K. U. How to represent crystal structures for machine learning: Towards fast prediction of electronic properties. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2014**, *89*, 205118.

(48) Ward, L.; Liu, R.; Krishna, A.; Hegde, V. I.; Agrawal, A.; Choudhary, A.; Wolverton, C. Including crystal structure attributes in machine learning models of formation energies via Voronoi tessellations. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2017**, *96*, 024104.

(49) Isayev, O.; Oses, C.; Toher, C.; Gossett, E.; Curtarolo, S.; Tropsha, A. Universal Fragment Descriptors for Predicting Electronic Properties of Inorganic Crystals. *Nat. Commun.* **2017**, *8*, 15679.

- (50) Jain, A.; Bligaard, T. Atomic-position independent descriptor for machine learning of material properties. *Phys. Rev. B: Condens. Matter Mater. Phys.* **2018**, *98*, 214112.
- (51) Walsh, A.; Butler, K. T. Prediction of electron energies in metal oxides. *Acc. Chem. Res.* **2014**, *47*, 364–372.
- (52) Tian, H.; Rangarajan, S. On Deriving Probabilistic Models for Adsorption Energy on Transition Metals using Multi-level Ab initio and Experimental Data. 2019, arXiv:1901.09253. arXiv.org e-Print archive. <http://arxiv.org/abs/1901.09253> (accessed August 1, 2019).
- (53) Hansen, K.; Montavon, G.; Biegler, F.; Fazli, S.; Rupp, M.; Scheffler, M.; von Lilienfeld, O. A.; Tkatchenko, A.; Müller, K.-R. Assessment and Validation of Machine Learning Methods for Predicting Molecular Atomization Energies. *J. Chem. Theory Comput.* **2013**, *9*, 3404–3419.
- (54) Smith, J. S.; Nebgen, B. T.; Zubatyuk, R.; Lubbers, N.; Devereux, C.; Barros, K.; Tretiak, S.; Isayev, O.; Roitberg, A. E. Approaching coupled cluster accuracy with a general-purpose neural network potential through transfer learning. *Nat. Commun.* **2019**, *10*, 2903.
- (55) Pelatt, B. D.; Ravichandran, R.; Wager, J. F.; Keszler, D. A. Atomic solid state energy scale. *J. Am. Chem. Soc.* **2011**, *133*, 16852–16860.
- (56) Pelatt, B. D.; Kokenyesi, R. S.; Ravichandran, R.; Pereira, C. B.; Wager, J. F.; Keszler, D. A. Atomic solid state energy scale: Universality and periodic trends in oxidation state. *J. Solid State Chem.* **2015**, *231*, 138–144.
- (57) Davies, D. W.; Butler, K. T.; Skelton, J. M.; Xie, C.; Oganov, A. R.; Walsh, A. Computer-aided design of metal chalcogenide semiconductors: from chemical composition to crystal structure. *Chemical Science* **2018**, *9*, 1022–1030.
- (58) Pauling, L. *The Nature of the Chemical Bond*, 3rd ed.; Cornell University Press: Ithaca, NY, 1960.
- (59) Bak, T.; Nowotny, J.; Rekas, M.; Sorrell, C. Photoelectrochemical hydrogen generation from water using solar energy. Materials-related aspects. *Int. J. Hydrogen Energy* **2002**, *27*, 991–1022.
- (60) Pinaud, B. A.; Benck, J. D.; Seitz, L. C.; Forman, A. J.; Chen, Z.; Deutsch, T. G.; James, B. D.; Baum, K. N.; Baum, G. N.; Ardo, S.; Wang, H.; Miller, E.; Jaramillo, T. F.; Turner, J. A.; Dinh, H. N. Technical and economic feasibility of centralized facilities for solar hydrogen production via photocatalysis and photoelectrochemistry. *Energy Environ. Sci.* **2013**, *6*, 1983–2002.
- (61) Mathew, K.; et al. Atomate: A high-level interface to generate, execute, and analyze computational materials science workflows. *Comput. Mater. Sci.* **2017**, *139*, 140–152.
- (62) Jain, A.; Ong, S. P.; Chen, W.; Medasani, B.; Qu, X.; Kocher, M.; Brafman, M.; Petretto, G.; Rignanese, G. M.; Hautier, G.; Gunter, D.; Persson, K. A. FireWorks: A dynamic workflow system designed for high-throughput applications. *Concurrency Computation* **2015**, *27*, 5037–5059.
- (63) Sun, W.; Dacek, S. T.; Ong, S. P.; Hautier, G.; Jain, A.; Richards, W. D.; Gamst, A. C.; Persson, K. A.; Ceder, G. The thermodynamic scale of inorganic crystalline metastability. *Science Advances* **2016**, *2*, No. e1600225.
- (64) Aykol, M.; Dwaraknath, S. S.; Sun, W.; Persson, K. A. Thermodynamic limit for synthesis of metastable inorganic materials. *Science Advances* **2018**, *4*, No. eaaq0148.
- (65) Woods-Robinson, R.; Broberg, D.; Faghaninia, A.; Jain, A.; Dwaraknath, S. S.; Persson, K. A. Assessing High-Throughput Descriptors for Prediction of Transparent Conductors. *Chem. Mater.* **2018**, *30*, 8375–8389.
- (66) Kůzma, M.; Dominko, R.; Meden, A.; Makovec, D.; Bele, M.; Jamnik, J.; Gabersček, M. Electrochemical activity of $\text{Li}_2\text{FeTiO}_4$ and $\text{Li}_2\text{MnTiO}_4$ as potential active materials for Li ion batteries: A comparison with $\text{Li}_2\text{NiTiO}_4$. *J. Power Sources* **2009**, *189*, 81–88.
- (67) Heyd, J.; Scuseria, G.; Ernzerhof, M. Hybrid functionals based on a screened Coulomb potential. *J. Chem. Phys.* **2003**, *118*, 8207–8215.
- (68) Krukau, A. V.; Vydrov, O. A.; Izmaylov, A. F.; Scuseria, G. E. Influence of the exchange screening parameter on the performance of screened hybrid functionals. *J. Chem. Phys.* **2006**, *125*, 224106.
- (69) 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.
- (70) Kresse, G.; Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1999**, *59*, 1758–1775.
- (71) Kresse, G.; Furthmüller, J. Efficiency of Ab-initio Total Energy Calculations for Metals and Semiconductors Using a Plane-wave Basis Set. *Comput. Mater. Sci.* **1996**, *6*, 15–50.
- (72) Kresse, G.; Furthmüller, J. Efficient Iterative Schemes for Ab Initio Total-energy Calculations Using a Plane-wave Basis Set. *Phys. Rev. B: Condens. Matter Mater. Phys.* **1996**, *54*, 11169–11186.
- (73) Perdew, J. P.; Ruzsinszky, A.; Csonka, G. I.; Vydrov, O. A.; Scuseria, G. E.; Constantin, L. A.; Zhou, X.; Burke, K. Restoring the Density-Gradient Expansion for Exchange in Solids and Surfaces. *Phys. Rev. Lett.* **2008**, *100*, 136406.