Using machine learning to identify factors that govern amorphization of irradiated pyrochlores

Ghanshyam Pilania,[†] Karl R. Whittle,[‡] Chao Jiang,[¶] Robin W. Grimes,[§] Christopher R. Stanek,[†] Kurt E. Sickafus,[‖] and Blas Pedro Uberuaga[*,†]

[†]*Materials Science and Technology Division, Los Alamos National Laboratory, Los Alamos, NM, USA*

[‡]*School of Engineering, University of Liverpool, Liverpool, UK*

[¶]*Fuels Modeling and Simulation Department, Idaho National Laboratory, Idaho Falls, ID, USA*

[§]*Department of Materials, Imperial College London, London, UK SW7 2AZ*

[‖]*Department of Materials Science and Engineering, University of Tennessee, Knoxville, TN, USA*

E-mail: blas@lanl.gov

Abstract

Structure-property relationships is a key materials science concept that enables the design of new materials. In the case of materials for application in radiation environments, correlating radiation tolerance with fundamental structural features of a material enables materials discovery. Here, we use a machine learning model to examine the factors that govern amorphization resistance in the complex oxide pyrochlore ($A_2B_2O_7$). We examine the fidelity of predictions based on cation radii and electronegativities, the oxygen positional parameter, and the energetics of disordering and amorphizing the material. No one factor alone adequately predicts amorphization resistance. We find that, when multiple families of pyrochlores (with different B cations)

1

are considered, radii and electronegativities provide the best prediction but when the machine learning model is restricted to only the B=Ti pyrochlores, the energetics of disordering and amorphization are optimal. This work provides new insight into the factors that govern the amorphization susceptibility and highlights the ability of machine learning approaches to generate that insight.

Introduction

Designing materials for advanced or next-generation applications requires understanding of how properties are related to structure, thatis, identifying so-called structure-property relationships. Having such relationships guides the search for new materials with enhanced performance by identifying regions of structure and composition space that exhibit superior properties. For nuclear energy materials, a key performance metric is tolerance against radiation damage. Pyrochlores ($A_2B_2O_7$) have been extensively studied for their potential application as nuclear waste forms[1–10] and have been incorporated into some compositions of the SYNROC waste form.[11] In this context, significant effort has been directed toward understanding how the chemistry of the pyrochlore – the nature of the A and B cations – dictates the amorphization susceptibility of the compound. In particular, several experimental efforts[12–16] have been focused on determining the critical amorphization temperature, T_C, the temperature at which the material recovery rate is equal to or faster than the rate of damage, as summarized in Fig. 1. Typically, these experiments were performed in an electron microscope equipped with an ion source, such that samples were simultaneously irradiated with electrons and 1 MeV Kr ions. Though the value of T_C is expected to vary depending on ion irradiation conditions,[17] 1 MeV Kr ion irradiation results should be comparable.

As a consequence, a number of "features" – or basic structural and energetic properties – have been identified that provide insight into the radiation response of pyrochlores. These include the radii and electronegativities of the A and B cations;[8,13] the x parameter, which describes how the oxygen sublattice deviates from ideality;[4,8,13] the enthalpy of formation of
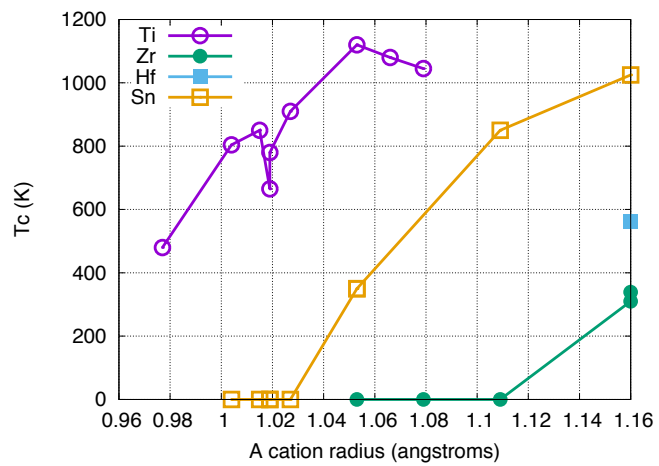
2

Figure 1: Experimentally measured values of T_C, ordered as a function of A cation radius, for several different pyrochlores.

the pyrochlore;[6,18] and the energy to disorder the pyrochlore to a disordered fluorite structure.[1,19] Further, there has been discussion on the extent of the disordered phase field in the phase diagram and its relationship to amorphization resistance.[7] Most of these features have been only heuristically correlated with amorphization resistance or only applied to a subset of pyrochlore chemistries. We are only aware of one attempt to quantify the relationship between these types of features and a prediction of T_C. In that work, Lumpkin and co-workers established a relationship between T_C and lattice constants, electronegativies, disordering energetics, and oxygen positional parameter.[8] While their model provided a significant advance in describing the structure-property relationships of pyrochlores, here we demonstrate how, through the use of machine learning, greater insight can be extracted. In particular, while they considered the disordering energy as one of their features, they used data from atomistic potentials that does not adequately describe all of the chemistries in the experiments. Further, they did not have access to data describing the amorphous state of these compounds. Finally, modern machine learning methods, applied to materials science, offer new avenues to examine the structure-property relationships in these types of systems.

Here, we use machine learning methods to demonstrate how a set of features, for a range of pyrochlore chemistries, can be used to predict T_C. We use both structural parameters such as cation radius and electronegativity supplemented by energetics calculated with density functional theory (DFT) to build a database of features as a function of pyrochlore chemistry. We analyze this database, building machine learning models that predict T_C as a function of pyrochlore chemistry based on a systematic collection of features. We consider pyrochlore chemistries for which experimental data exists for T_C, which includes pyrochlores where B=Ti, Zr, Hf, and Sn. (Here, and in what follows, when we refer to the B cation, we mean the B chemistry of the pyrochlore (the nature of the 4+ species).) We find that, when considering the full range of chemistries, the two features that best predict T_C are the ratio of the radii and the difference in electronegativities of the A and B cations. However, to predict more subtle dependencies of T_C with pyrochlore chemistry characteristic of a given B

chemistry, the energies to disorder and amorphize the compound provide a better prediction of T_C. Importantly, the energy to disorder the compound alone does not correlate with T_C and must be complemented by information about the energy to amorphize the material.

As compared to Ti, Hf, or Zr, Sn is a chemically very different element. It, like Ti, is multivalent, but unlike Ti, has a much stronger prevalence to adopt a charge state other than 4+. Further, as discussed below, it has a significantly higher electronegativity than the other B cations, producing a more covalent bond. This implies that Sn pyrochlores should be less amorphization resistant.[20] However, experiments have shown Sn pyrochlores to be more amorphization resistant than other pyrochlores.[5] This all suggests that Sn pyrochlores are electronically much more complex than the other pyrochlore families, which is one reason that we use DFT to determine the energetics of disordering and amorphization, as DFT can account for the varied valence of the Sn cations. Further, the inclusion of Sn pyrochlores in this analysis, precisely because the behavior is counter-intuitive, provides a more stringent test of the methodology.

Results

DFT Energetics

Figure 2a provides the energetics for disorder and amorphization of a given pyrochlore, as found using DFT, as a function of the chemistry of the pyrochlore. These are ordered by A cation radius. Focusing first on the energetics to disorder, there is a general trend that as the A cation radius increases, the energy associated with disordering the pyrochlore to a disordered fluorite also increases, consistent with previous results using DFT.[21] This is particularly true of the B=Zr, Hf and Sn families of pyrochlores. For the B=Ti family, there is a peak in the disorder energy for the A=Gd composition, again consistent with previous DFT and empirical potential calculations.[19,21]

Figure 2a highlights the apparent contradiction between experimental observations and

the notion that the disordering energy correlates with amorphization resistance. If only disordering energetics dictated the response of the pyrochlore to irradiation, then one would expect that Zr pyrochlores would generally exhibit higher amorphization resistance than Ti pyrochlores (which they do) but also that Sn pyrochlores would be less resistance to amorphization than Ti pyrochlores, which they are not. Thus, other factors must also be important. We propose that the energy of the amorphous phase is one of those factors.

The energy differences between ordered pyrochlore and an amorphous structure are also provided in Fig. 2a. In the case of the B=Hf and Zr families, these are again relatively monotonic with increasing A cation radius. However, the behavior of the B=Ti and Sn families is more complex. In particular, for the B=Ti family, the amorphous energy is non-monotonic with A cation radius, but the peak is for a different chemistry than was the disordering energy. In the B=Ti family, the amorphous energy is greatest for A=Y and generally is high for A=Dy and Tb. The B=Sn family exhibits even more complicated behavior. There is a peak in the amorphous energy for A=Gd and a minimum for A=Ho.

Finally, the shaded regions in Fig. 2a highlight the energy gap between the disordered and amorphous states. The variation of this gap with A cation radius is very different for the different families of pyrochlores. For the B=Zr and Hf pyrochlores, the gap slowly but steadily decreases with A cation radius. For the B=Ti pyrochlores, the gap first increases slightly and then decreases to essentially zero for the A=Nd chemistry. The gap for B=Sn pyrochlores first decreases, then increases, and then decreases again. Further, the gap is smallest for the B=Ti pyrochlores and, overall, largest for the B=Zr and Sn pyrochlores, at least for some A chemistries.

Figure 2b provides the volume changes between the ordered phase and both the disordered and amorphous phases, as determined from the DFT calculations. For nearly all of the cases, a transformation from the ordered to disordered phase results in a volume expansion while the formation of the amorphous phase contracts the lattice. The exceptions are the B=Zr and Hf pyrochlores with small A cations, which exhibit very little change in volume upon

disordering. On the other hand, $Nd_2Ti_2O_7$ (assumed to be cubic here) exhibits very little change upon amorphization.
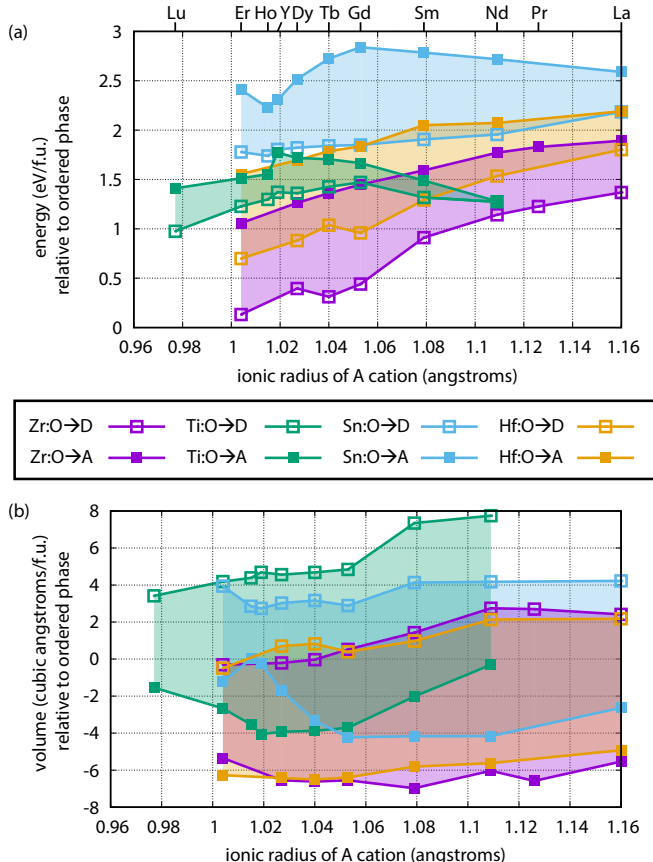

Figure 2: DFT results for the (a) energetics and (b) volume changes associated with an order-to-disorder (O→D, open symbols) and an order-to-amorphous (O→A, closed symbols) transformation for four families of pyrochlores in which B=Ti (green), Zr (purple), Hf (yellow), and Sn (cyan). The shaded regions highlight the differences between the disordered and amorphous structures.

Correlation of Features with Amorphization Resistance

The DFT results reveal that there are significant differences in the energetics of disorder and amorphization in pyrochlores as a function of both A and B chemistry. We use a machine learning approach to quantify the correlations between these energetics, as well as other features associated with pyrochlores, and the amorphization resistance, as characterized by T_C. The features considered here are r_A/r_B, the ratio of the ionic radii of the A and B cations;

$\Delta X = X_B - X_A$, the difference in electronegativity of the A and B neutral metal atoms (X_A and X_B, respectively); x, the oxygen positional parameter, which measures the deviation of the oxygen sublattice from an ideal (fluorite-like) simple cubic sutlattice; $E_{O \to D}$, the energy difference between the disordered and ordered phases; and $E_{D \to A}$, the energy difference between the amorphous and disordered phases. These features were chosen because (a) they have been shown to correlate to some degree in previous studies and (b) our DFT results indicate that the energetics depend strongly on the A and B chemistry of the pyrochlore, suggesting they may provide a strong descriptor of each compound. We did not consider the enthalpy of formation, proposed by other authors as a factor in radiation tolerance,[6,18] as a feature because data was not available for all compounds.

However, before we examine the results of the machine learning model, it is instructive to examine how the selected features correlate with T_C. Figure 3 provides simple plots of each feature against T_C. The values for T_C, summarized in Table 1, are taken from Refs.[12–16] Figure 3 reveals that while there are rough correlations between T_C and some of the features, there is not one feature that provides a quantitative capability of predicting T_C. For example, overall, r_A/r_B correlates well with T_C over a wide range of B chemistries; however, it does not capture subtleties associated with variations in T_C with a given family of pyrochlores. ΔX, on the other hand, discriminates between pyrochlores with B=Sn and the other families but does not correlate directly with T_C. Similarly, x shows an overall correlation with T_C but again the details are lost. $E_{O \to D}$, on the other hand, seems to correlate reasonably well for pyrochlores within a given family but does not describe variations of T_C between families. Finally, $E_{O \to A}$, similar to x and ΔX, seems to generally correlate separately for B=Sn pyrochlores and the other families of pyrochlores. Thus, while there are rough trends indicating some insight from each of these features, there is certainly not enough of a correlation in any case for a quantitative prediction. However, this suggests, as noted by other authors,[8] that combinations of these features may provide predictive capability. Hence, we use a machine learning approach to quantify this.

Table 1: Values of T_C used in the machine learning model, as reported by the corresponding references. In the case of $Y_2Ti_2O_7$ (*), as there were two significantly different values in the literature, we took an average value of T_C: 723 K. In the case of $La_2Zr_2O_7$ (†), there were also two literature values, but as one was within the error bars of the other, we simply took $T_c = 310$ K for this compound. The error bars were not used in the machine learning model as, owing to the limited data, ML models never reached accuracies where the error bars were relevant. For those compounds were no T_C was measured (i.e. $T_C < 25$ K) we simply used a value of $T_C = 12.5$ K for numerical reasons.

Compound	T_C (K)	Error (K)	Ref.
$Sm_2Ti_2O_7$	1045	n/a	12
$Eu_2Ti_2O_7$	1080	n/a	12
$Gd_2Ti_2O_7$	1120	n/a	12
$Tb_2Ti_2O_7$	970	n/a	12
$Dy_2Ti_2O_7$	910	n/a	12
$Y_2Ti_2O_7$ *	780	n/a	12
$Y_2Ti_2O_7$ *	665	33	15
$Ho_2Ti_2O_7$	850	n/a	12
$Er_2Ti_2O_7$	804	n/a	12
$Lu_2Ti_2O_7$	480	n/a	12
$La_2Zr_2O_7$ †	339	49	13
$La_2Zr_2O_7$ †	310	n/a	12
$Nd_2Zr_2O_7$	< 25	n/a	12
$Sm_2Zr_2O_7$	< 25	n/a	12
$Gd_2Zr_2O_7$	< 25	n/a	12
$La_2Hf_2O_7$	563	22	13
$La_2Sn_2O_7$	1025	n/a	12
$Nd_2Sn_2O_7$	850	n/a	12
$Gd_2Sn_2O_7$	350	n/a	12
$Dy_2Sn_2O_7$	< 25	n/a	12
$Ho_2Sn_2O_7$	< 25	n/a	12
$Y_2Sn_2O_7$	< 25	n/a	12
$Y_2Sn_2O_7$	< 50	n/a	14
$Er_2Sn_2O_7$	< 25	n/a	12

Results of the Machine Learning Model

We use a machine learning (ML) approach to quantify the correlations between the five features described in the previous section and T_C. More specifically, we employed kernel ridge regression (KRR)[22–24]—an algorithm that works on the principle of similarity and is capable of extracting complex non-linear relationships from data in an efficient manner—with a Gaussian kernel to learn and quantify trends exhibited by T_C in the feature space discussed above. A randomly selected 90%/10% training/test split of the available data was used for statistical learning and testing the performance of the trained model on previously unseen data. A leave-one-out cross validation is used to determine the model hyper-parameters to avoid any overfitting of the training data that may lead to poor generalizability. The trained model can subsequently be used to make an interpolative prediction of T_C for a new material (i.e., not used in the model training) i by comparing its distance in feature space d_i (suitably
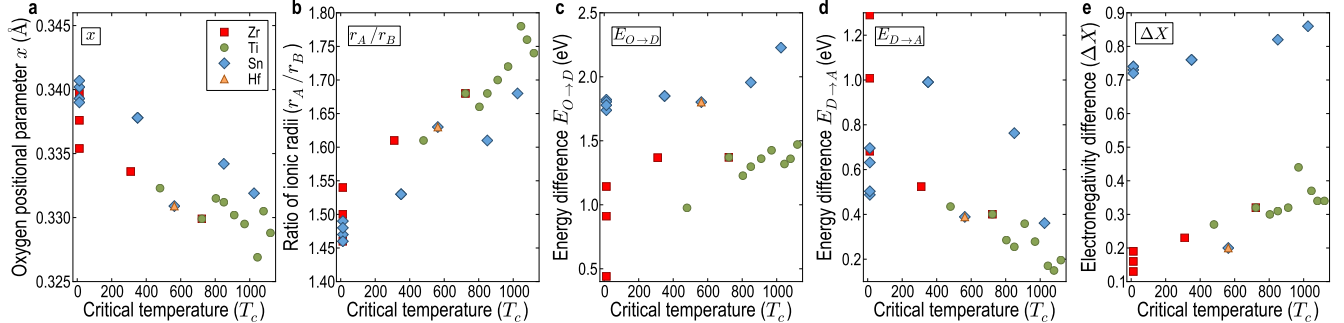
Figure 3: Correlations between experimental measurements of T_C and 5 features describing pyrochlores. (a) x, the oxygen positional parameter. (b) r_A/r_B, the ratio of cation radii. (c) $E_{O \to D}$, the energy difference between disordered fluorite and ordered pyrochlore. (d) $E_{D \to A}$, the energy difference between an amorphous structure and disordered fluorite. (e) ΔX, the difference in electronegativity between the A and B cation. The different symbols in the figures indicate the nature of the B cation (squares=Zr, circles=Ti, diamonds=Sn, triangles=Hf).

defined by a distance measure; in our case the Euclidean norm was used) with those of a set of reference training cases for which the T_C values are known. Further details of our KRR-based ML models are provided in the Methods section.

Next, within the KRR ML model, we aim to identify the best feature combination that exhibits highest prediction performance, quantified by its ability to accurately predict T_C of the test set compounds. We do this in a comprehensive manner by building KRR ML models using all possible combinations of Ω features with $\Omega \in [2,5]$. Performance of each of these models was evaluated separately on the entire data set as well as on a reduced set that only included the Ti pyrochlores. The root mean square (rms) errors for the T_C predictions on training and test sets for various models is presented in Fig. 4. In order to account for model prediction variability associated with randomly selected training/test splits, Fig. 4 reports the rms errors averaged over 100 different randomly selected training/test splits for each of the models. The 2D models that lead to the lowest rms errors on the test set data have been marked with a '⋆' in Fig. 4a (when taking the entire data) and Fig. 4b (for the Ti pyrochlores). It is interesting to note that, for both cases, going beyond the best performing 2D models does not lead to a significant improvement in the model prediction performance.

For instance, while the best binary feature pair $(r_A/r_B, \Delta X)$ leads to a test set rms error of 101.2 K in T_C, the ML models built on the best 4D and 5D (taking all 5 features considered) feature vectors only result in nominal improvements leading to rms errors of 97.8K and 98K, respectively. Since, as a general rule, higher model complexity often leads to poor generalizability, in case of a comparable prediction performance, a simpler model (*i.e.*, built on a lower dimensional feature set) should always be preferred over a more complex one. Therefore, henceforth we focus our attention on the the best performing 2D models.
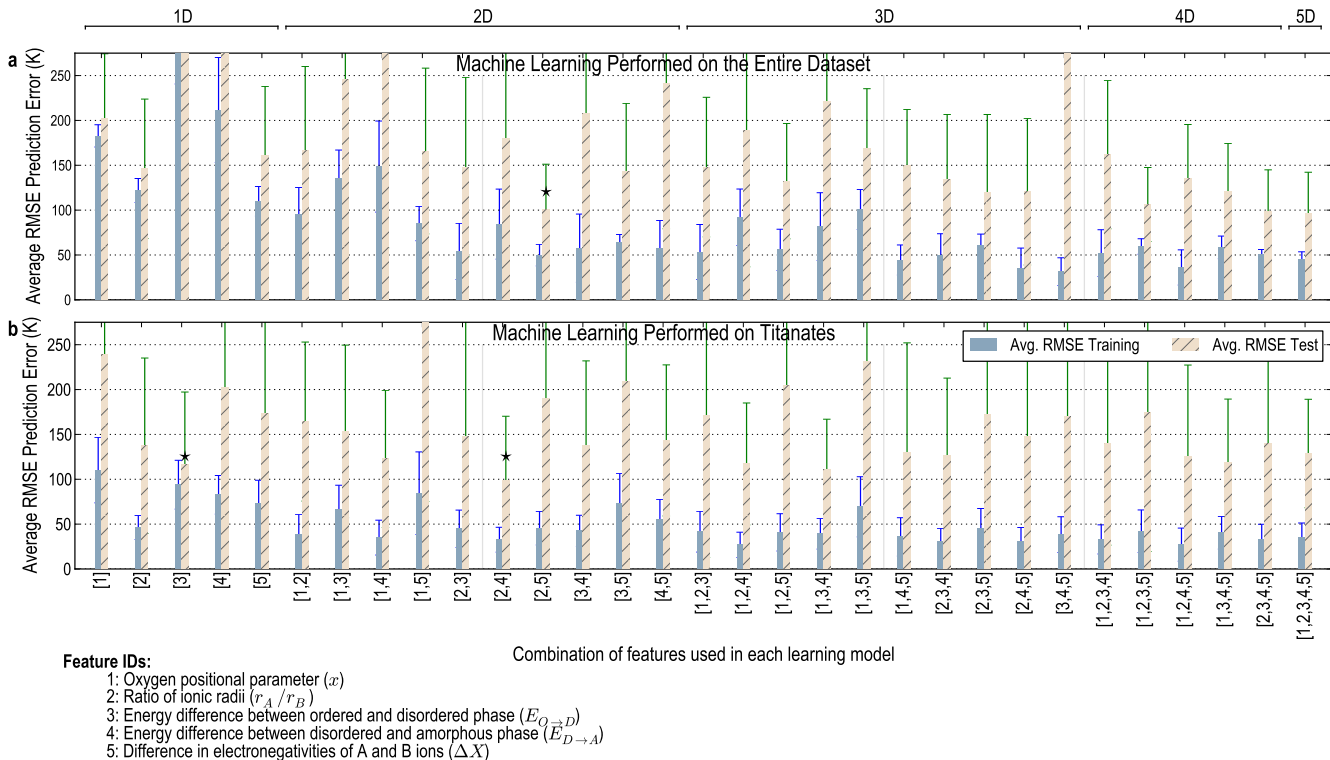

Figure 4: Results from the machine learning model. (a) Results when applied to the entire set of pyrochlores and (b) results when applied only to the titanate family. The solid bars indicate the average RMS error for the training data while the hashed bars indicate the error for the test data. The indices on the abscissa indicate the features used in that particular model. Models including 2, 3, 4 or all 5 features were considered. The best 2D feature set is indicated with the star for both cases. Error bars represent the standard deviations for the rms error in predict T_C, computed over the 100 different training/test set splits.

The superior performance exhibited by the $(r_A/r_B, \Delta X)$ feature pair is not entirely unexpected and can be understood by looking at Fig. 3b and e. As alluded to previously,

while r_A/r_B helps capture the overall T_C trends among different chemistries, ΔX allows for an effective separation between different chemistries (especially, between the Sn-based compounds and rest of the dataset), while still capturing relative T_C trends between these subgroups. The best performing feature pair for the titanate pyrochlores dataset, however, is constituted by $E_{O\rightarrow D}$ and $E_{D\rightarrow A}$. While the (r_A/r_B, ΔX) feature pair performs much poorer on this subset than the overall dataset, the performance of (r_A/r_B, $E_{O\rightarrow D}$) feature pair is also found comparable to that of the best 2D feature pair.

While Fig. 4 captures the average performance and variability (taken over 100 different runs) for our best performing 2D models (marked with a $\star$), in Fig. 5a-b we present parity plots comparing the experimental T_C with the ML predictions using the best 2D descriptors found for the entire dataset (Fig. 5a) and the titanates (Fig. 5b), respectively. In each case $\sim 90\%$ of the dataset was used for training (plotted as squares) with the remaining for testing the model performance (plotted as circles). In each case, we used four different ML runs randomly selecting training and test set splits (depicted by different colors). It can be see from the parity plots that our ML models can reasonably predict (within the error bars established in Fig. 4), T_C over the entire dataset. A couple of conclusions can be drawn from these plots. First, visually it can be seen that the model prediction performance is comparable for the training and test sets, indicating that there is no overfitting (a problem when a ML model performs very well on a training set but exhibits a poor performance on a test set). Second, despite their simplicity (given that we are only using a two-dimensional feature in each case), the ML models exhibit good predictive power and stability (predictions do not change drastically over different training/test splits). This highlights the robustness of the model.

To gain a deeper insight into the KRR model's prediction performance, we next construct contour plots for each of the two best performing 2D feature pairs discussed above. In each case, we start with a fine 2D grid in the feature space constituted by the primary features identified above, while still confining ourselves within the boundaries of the original feature
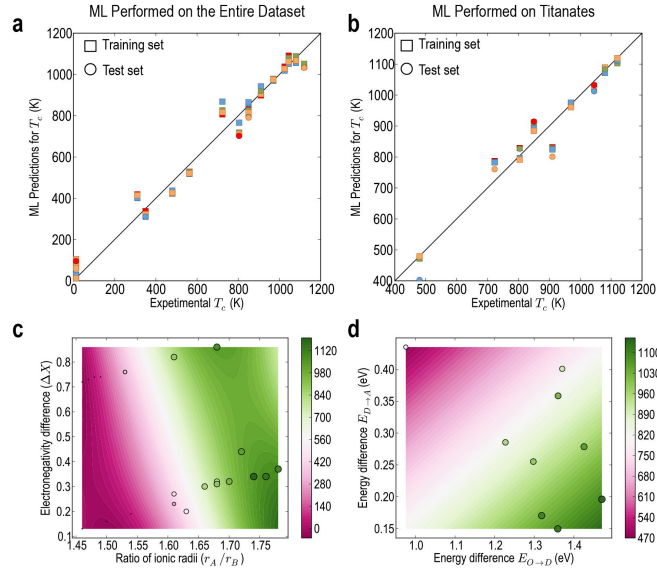
Figure 5:   (a-b) Parity plots of the machine learning results for (a) the entire set of pyrochlores and (b) the titanate family. The squares represent training data while the circles are the test set. The different colors represent different runs with different training/test set splits. (c-d) Results from the machine learning model for (c) all of the pyrochlores considered and (d) just the titanate famliy. The size and color of the circles indicate the experimental T_C while the position of the circles indicates the predicted T_C. The contours indicate predictions of T_C for other values of the feature pairs.

13

space used to train the KRR models. Each point on this grid then, in principle, represents a point in the feature space, which can be used as an input for the respective trained KRR models to make predictions. That is, we map out the predicted value of T_C as a function of the two features over a range of values for each of the two features. Since the ML models are interpolative, one can readily use a fine grid in the 2D feature space to visualize trends in T_C versus the feature values and make predictions of T_C for new chemistries.

Figure 5c shows the best two-feature descriptor for the entire set of pyrochlores considered. Again, in this case, the two features that best correlate with T_C are r_A/r_B and ΔX. This combination of features is able to distinguish the different T_C behavior exhibited by the B=Sn pyrochlores and the other families of pyrochlores, by virtue of the properties of ΔX. In principle, the types of maps provided in Fig. 5 could be used to predict T_C for chemistries of pyrochlores not included in Table 1. However, care must be taken as machine learning models are inherently interpolative in nature. Still, Fig. 5 indicate that small r_A/r_B radius ratio and small electronegativity differences between the A and B cations will maximize amorphization resistance (minimize T_C). Similarly, minimizing $E_{O \to D}$ and maximizing $E_{D \to A}$ will also minimize T_C.

However, as discussed above, this combination of features has an effective uncertainty of ~ 100 K, indicating that it cannot describe the fine features exhibited by the B=Ti family of pyrochlores. For example, T_C is not monotonic with A cation radius (see Fig. 1). As discussed, limiting the model to just the B=Ti pyrochlores results in a different optimal two-feature set, namely $E_{O \to D}$ and $E_{O \to A}$, as shown in Fig. 4b. In particular, as shown in Fig. 5d, this set of features can describe the subtle behavior in which the A=Gd compound has the highest value of T_C, correlating with the fact that it has the highest value of $E_{O \to D}$, while the A=Y compound, which has values of $E_{O \to D}$ similar to the neighboring compounds, exhibits an anomalously low value of T_C. This is a consequence of its rather high value of $E_{D \to A}$, a consequence of the fact that Y is not a rare earth and thus the bonding associated with it is subtly different to the other elements around it.

Discussion and Conclusions

Combining experimental results for T_C for various pyrochlore compounds, DFT calculations of the energetics of disordering and amorphization, and a machine learning model, we conclude that (a) basic ionic properties such as r_A/r_B and ΔX have the qualitative capability of predicting trends in T_C over a wide-range of pyrochlore compounds but that (b) more quantitative predictions that capture the subtleties associated with variations in A cation chemistry require knowledge of both the disordering and amorphization energetics. This generalizes the previous understanding in which rough correlations between amorphization resistance and, for example, disordering energetics were hypothesized based on a few observations.

However, what is clear from the machine learning analysis is that, even with the input of DFT energetics, the predictive capabilities are still limited. Even when limited to the B=Ti family of pyrochlores, the model results in predictive uncertainty of 75 K. This is a consequence of many factors, including the limited amount of experimental data, the uncertainties in what experimental data that there is, and uncertainties in the DFT calculations. To determine an even better predictive model, more experimental data is required. In particular, values of T_C for other families of pyrochlores would enhance the strength of the model. For example, without the B=Sn pyrochlores, the importance of ΔX would likely not have been revealed. Importantly, given the small data set, domain knowledge – experience with the behavior of this system – was important in narrowing down a set of likely relevant features.

While the feature set of ΔX and r_A/r_B have the best predictive capability for distinguishing between the various families of pyrochlores, the reason why Sn pyrochlores are radiation tolerant while exhibiting such high disordering energies is found in examining the amorphization energetics. The gap between the disordering and amorphization energies for the Sn pyrochlores is typically quite large and even if, during the course of irradiation, enough energy is deposited into the lattice such that the structure becomes disordered, it is not enough to amorphize the material. The gap betwen the disordering and amorphization energies is much larger in the Sn pyrochlores than it is in the Ti family and, for some A cations,

larger than for the Hf and Zr families as well. Thus, the origin of the radiation tolerance of some of the Sn pyrochlores comes from the fact that they are extremely difficult to amorphize. Importantly, by incorporating properties such as $E_{O \rightarrow D}$ and $E_{D \rightarrow A}$ that are derived from electronic structure calculations, information about bonding in different environments (e.g. ordered, disordered and amorphous) are naturally included into the model.

The insights gained by the machine learning model apply specifically to pyrochlores and, because of the interpolative nature of these models, to the families of pyrochlores considered here. That said, the features identified as being best able to predict T_C can be justified physically and thus may be applicable to other classes of complex oxides, such as δ-phase,[25] that have fluorite as the parent structure. However, other classes of complex oxides, such as spinel, which have fundamentally different crystal structures may have different dependencies on these features, or require new features to predict behavior. In particular, structural vacancies on the cation sublattice in spinel can facilitate recovery of damage in a way that is not possible in pyrochlore.[26] Further, other factors, such as short-range order, which is known to occur in complex oxides,[27,28] may also play a role. However, we suspect that treating the disordered state as truly random captures much of the behavior of these materials, given the ability of the disordered fluorite structure to predict order-disorder temperatures in these systems.[21,29]

In this work, we have used T_C as a metric for relative amorphization resistance. In reality, the value of T_C encompasses not only thermodynamic properties such as disordering and amorphization energetics, but also kinetic processes of defect annihilation and defect production. Further, T_C depends on irradiation conditions,[30] changing with ion mass, energy and flux. Thus, actually predicting T_C from fundamental defect behavior would be a daunting task, as the computational materials modeling community cannot predict the damage state induced by arbitrary irradiation conditions nor the behavior of defects in pyrochlore as a function of disorder (further, for quantitative predictions, these quantities would need to be known at a DFT level). However, T_C does provide a metric to compare the susceptibility

of amorphization that has been measured for a range of pyrochlore chemistries. In addition, while T_C will change with irradiation conditions, we expect that the general trends exhibited between different pyrochlore chemistries will not. If the trends were very sensitive to the irradiation conditions, then correlations of T_C with e.g. disordering tendencies would not exist, as even trends in T_C would be dominated by external factors. Finally, this work demonstrates an alternative route to developing predictive models of fundamentally complicated properties such as T_C. In principle, with more experimental data, irradiation conditions could be accounted for in the machine learning model via parameters such as ion mass, energy and flux. While a machine learning model does not provide the intuitive physical picture that a fundamental model would, it does allow for predictions of systems where developing a fundamental model is prohibitive.

Finally, this work highlights the utility of machine learning approaches in materials science. In this case, the ML model elucidates those features which provide predictive capability, providing insight into those factors which dictate amorphization resistance in pyrochlores. The model also shows that sets of two features result in optimal predictions; higher-order feature sets do not add significant value. The fact that different combinations of features provide are optimal for predictions for the entire set of pyrochlores (r_A/r_B and ΔX) versus the Ti family ($E_{O \to D}$ and $E_{O \to A}$) reinforces the point that the best set of features depends on the level of detail (here, the error in the predicted T_C) required in the prediction.

Methods

Density Functional Theory

Density functional theory (DFT) calculations were performed using the all-electron projector augmented wave method[31] within the local density approximation (PBE) with the VASP code.[32] A plane-wave cutoff of 400 eV and dense k-point meshes were used to ensure convergence. The lattice parameters and all atomic positions were allowed to relax, though the cells

were constrained to be cubic. The disordered fluorite structure was modeled using the special quasirandom structures (SQS) approach.[33] The SQS structures were generated as described in Ref.[21] The amorphous structures were created by performing *ab initio* molecular dynamics at a very high temperature and then quenching the structures to 0 K. For the B=Zr and Hf families, there is a deviation from true monotonic behavior at A=Tb, in contrast with previous DFT calculations[21] that used the same methodology (pseudopotentials, functional, k-point mesh, and energy cutoff). We assume that the differences from previously published results are due to changes in different versions of VASP.

Machine Learning Model

We used Kernel ridge regression (KRR) with a Gaussian kernel for machine learning. KRR is a similarity-based learning algorithm, which has recently been very successful for a range of materials property prediction problems. Some recent applications of this approach to materials problems include predictions of the properties of molecular[34,35] and periodic systems,[36,37] development of adaptive force fields,[38,39] crystal structure classification,[40] dielectric breakdown strength prediction,[41,42] self-consistent solutions for quantum mechanics,[43] and predictions of bandgaps.[44] The treatment in this paper is, to the best of our knowledge, the first that uses machine learning to anaylze a small experimental data set to build a predictive model. Within KRR, the ML estimate of a target property (in our case the critical temperature T_C) of a new system j, is estimated by a sum of weighted kernel functions (*i.e.*, Gaussians) over the entire training set, as

$$T_{Cj}^{ML} = \sum_{i=1}^{N} w_i \exp\left(-\frac{1}{2\sigma^2}|\mathbf{d}^{ij}|^2\right). \tag{1}$$

where i runs over the systems in the training dataset, and $|\mathbf{d}^{ij}|^2 = ||\mathbf{d}_i - \mathbf{d}_j||_2^2$, the squared Euclidean distance between the feature vectors $\mathbf{d}_i$ and $\mathbf{d}_j$. The coefficients w_is are obtained from the training (or learning) process built on minimizing the expression $\sum_{i=1}^{N}\left(T_{Ci}^{ML} - T_{Ci}^{Exp}\right)^2 +$

$\lambda \sum_{i=1}^{N} w_i^2$, with T_{Ci}^{ML} being the ML estimated critical temperature, T_{Ci}^{Exp} the corresponding experimental value, and the model hyper-parameters σ and λ are optimized within a internal cross-validation loop on a fine logarithmic grid. The explicit solution to this minimization problem is $\boldsymbol{\alpha} = (\mathbf{K} + \boldsymbol{\lambda}\mathbf{I})^{-1}\mathbf{P}^{DFT}$, where $\mathbf{I}$ is the identity matrix, and $K_{ij} = \exp\left(-\frac{1}{2\sigma^2}|\mathbf{d}^{ij}|^2\right)$ is the kernel matrix elements of all materials in the training set. The parameters λ and σ are determined in an inner loop of fivefold cross validation using a logarithmically scaled fine grid.

Supporting Information

Full five-feature dataset used to develop the kernel ridge regression machine learning model.

Acknowledgments

Author contributions

G.P. developed the machine learning model and applied it to the pyrochlores. K.R.W. compiled the experimental data. C.J. constructed the DFT simulation cells. R.W.G. and K.E.S. performed the original studies that proposed the concept linking radiation tolerance to disordering. C.R.S. helped devise the study. B.P.U. wrote the main manuscript text and performed the DFT calculations. All authors reviewed the manuscript.

Additional information

Competing financial interests: The authors declare no competing financial interests.

References

(1) Sickafus, K. E.; Minervini, L.; Grimes, R. W.; Valdez, J. A.; Ishimaru, M.; Li, F.; McClellan, K. J.; Hartmann, T. Radiation Tolerance of Complex Oxides. *Science* **2000**, *289*, 748–751.

(2) Begg, B. D.; Hess, N. J.; McCready, D. E.; Thevuthasan, S.; Weber, W. J. Heavy-ion Irradiation Effects in $Gd_2(Ti_{2-x}Zr_x)O_7$ Pyrochlores. *J. Nucl. Mater.* **2001**, *289*, 188–193.

(3) Lian, J.; Wang, L.; Chen, J.; Sun, K.; Ewing, R. C.; Farmer, J. M.; Boatner, L. A. The Order-Disorder Transition in Ion-Irradiated Pyrochlore. *Acta Mater.* **2003**, *51*, 1493–1502.

(4) Lian, J.; Chen, J.; Wang, L. M.; Ewing, R. C.; Farmer, J. M.; Boatner, L. A.; He-lean, K. B. Radiation-Induced Amorphization of Rare-Earth Titanate Pyrochlores. *Phys. Rev. B* **2003**, *68*, 134107.

(5) Lian, J.; Helean, K. B.; Kennedy, B. J.; Wang, L. M.; Navrotsky, A.; Ewing, R. C. Effect of structure and thermodynamic stability on the response of lanthanide stannate pyrochlores to ion beam irradiation. *J. Phys. Chem. B* **2006**, *110*, 2343–2350.

(6) Helean, K. B.; Ushakov, S. V.; Brown, C. E.; Navrotsky, A.; Lian, J.; Ewing, R. C.; Farmer, J. M.; Boatner, L. A. Formation Enthalpies of Rare Earth Titanate Py-rochlores. *J. Solid State Chem.* **2004**, *177*, 1858–1866.

(7) Sickafus, K. E.; Grimes, R. W.; Valdez, J. A.; Cleave, A.; Tang, M.; Ishimaru, M.; Corish, S. M.; Stanek, C. R.; Uberuaga, B. P. Radiation-Induced Amorphization Re-

sistance and Radiation Tolerance in Structurally Related Oxides. *Nat. Mater.* **2007**, *6*, 217–223.

(8) Lumpkin, G. R.; Pruneda, M.; Rios, S.; Smith, K. L.; Trachenko, K.; Whittle, K. R.; Zaluzec, N. J. Nature of the Chemical Bond and Prediction of Radiation Tolerance in Pyrochlore and Defect Fluorite Compounds. *J. Solid State Chem.* **2007**, *180*, 1512–1518.

(9) Sattonnay, G.; Sellami, N.; Thomé, L.; Legros, C.; Grygiel, C.; Monnet, I.; Jagielski, J.; Jozwik-Biala, I.; Simon, P. Structural Stability of $Nd_2Zr_2O_7$ Pyrochlore Ion-Irradiated in a Broad Energy Range. *Acta Mater.* **2013**, *61*, 6492–6505.

(10) Li, Y. H. et al. Role of Antisite Disorder on Preamorphization Swelling in Titanate Pyrochlores. *Phys. Rev. Lett.* **2012**, *108*, 195504.

(11) Ringwood, A.; Kesson, S.; Ware, N.; Hibberson, W.; Major, A. *Nature* **1979**, *278*, 219.

(12) Ewing, R. C.; Lian, J.; Wang, L. M. Ion Beam-Induced Amorphization of the Pyrochlore Structure-Type: A Review. *Mater. Res. Soc. Proc.* **2003**, *792*, 37–48.

(13) Lumpkin, G. R.; Whittle, K. R.; Rios, S.; Smith, K. L.; Zaluzec, N. J. Temperature Dependence of Ion Irradiation Damage in the Pyrochlores $La_2Zr_2O_7$ and $La_2Hf_2O_7$. *J. Phys.: Condens. Matter* **2004**, *16*, 8557–8570.

(14) Lumpkin, G. R.; Smith, K. L.; Blackford, M. G.; Whittle, K. R.; Harvey, E. J.; Redfern, S. A. T. Ion Irradiation of Ternary Pyrochlore Oxides. *Chem. Mater.* **2009**, *21*, 2746–2754.

(15) Whittle, K. R.; Blackford, M. G.; Aughterson, R. D.; Lumpkin, G. R.; Zaluzec, N. J. Ion Irradiation of Novel Yttrium/Ytterbium-Based Pyrochlores: The Effect of Disorder. *Acta Mater.* **2011**, *59*, 7530–7537.

(16) Lian, J.; Ewing, R. C.; Wang, L. M.; Helean, K. B. Ion-Beam Irradiation of $Gd_2Sn_2O_7$ and $Gd_2Hf_2O_7$ Pyrochlore: Bond-Type Effect. *J. Mater. Res.* **2004**, *19*, 1575–1580.

(17) Meldrum, A.; Boatner, L. A.; Weber, W. J.; Ewing, R. C. Amorphization and Recrystallization of the ABO(3) Oxides. *J. Nucl. Mater.* **2002**, *300*, 242–254.

(18) Helean, K. B.; Navrotsky, A.; Lian, J.; Ewing, R. C. Correlation of Formation Enthalpies with Critical Amorphization Temperature for Pyrochlore and Monazite. *Mater. Res. Soc. Proc.* **2004**, *824*, 22–27.

(19) Minervini, L.; Grimes, R. W.; Sickafus, K. E. Disorder in Pyrochlore Oxides. *J. Am. Ceram. Soc.* **2000**, *83*, 1873–1878.

(20) Naguib, H. M.; Kelly, R. Criteria for Bombardment-Induced Structural Changes in Non-Metallic Solids. *Radiat. Eff.* **1975**, *25*, 1–12.

(21) Jiang, C.; Stanek, C. R.; Sickafus, K. E.; Uberuaga, B. P. First-Principles Prediction of Disordering Tendencies in Pyrochlore Oxides. *Phys. Rev. B* **2009**, *79*, 104203.

(22) Witten, I. H.; Frank, E.; Hall, M. A. *Data Mining: Practical Machine Learning Tools and Techniques*; Elsevier, 2011.

(23) Müller, K. R.; Mika, S.; G.Ratsch,; Tsuda, K.; Scholkopf, B. An Introduction to Kernel-based Learning Algorithms. *IEEE Trans. Neural Networks* **2001**, *12*, 181–201.

(24) Hofmann, T.; Scholkopf, B.; Smola, A. J. Kernel Methods in Machine Learning. *Ann. Statist.* **2008**, *36*, 1171–1220.

(25) Stanek, C. R.; Jiang, C.; Uberuaga, B. P.; Sickafus, K. E.; Cleave, A. R.; Grimes, R. W. Predicted Structure and Stability of $A_4B_3O_{12}$ δ-Phase Compositions. *Phys. Rev. B* **2009**, *80*, 174101.

(26) Uberuaga, B. P.; Tang, M.; Jiang, C.; Valdez, J. A.; Smith, R.; Wang, Y.-Q.; Sicka-fus, K. E. Opposite Correlations Between Cation Disordering and Amorphization Resistance in Spinels Versus Pyrochlores. *Nat. Commun.* **2015**, *6*, 8750.

(27) Jiang, C.; Sickafus, K. E.; Stanek, C. R.; Rudin, S. P.; Uberuaga, B. P. Cation Disorder in MgX_2O_4 (X=Al, Ga, In) Spinels from First Principles. *Phys. Rev. B* **2012**, *86*, 024203.

(28) Shamblin, J.; Feygenson, M.; Neuefeind, J.; Tracy, C. L.; Zhang, F.; Finkeldei, S.; Bosbach, D.; Zhou, H.; Ewing, R. C.; Lang, M. Probing Disorder in Isometric Pyrochlore and Related Complex Oxides. *Nat. Mater.* **2016**,

(29) Li, Y.; Kowalski, P. M.; Beridze, G.; Birnie, A. R.; Finkeldei, S.; Bosbach, D. Defect Formation Energies in $A_2B_2O_7$ Pyrochlores. *Scr. Mater.* **2015**, *107*, 18–21.

(30) Wang, S. X.; Wang, L. M.; Ewing, R. C.; Was, G. S.; Lumpkin, G. R. Ion Irradiation-Induced Phase Transformation of Pyrochlore and Zirconolite. *Nucl. Instrum. Methods Phys. Res., Sect. B* **1999**, *148*, 704.

(31) Kresse, G.; Joubert, D. From Ultrasoft Pseudopotentials to the Projector Augmented-Wave Method. *Phys. Rev. B* **1999**, *59*, 1758–1775.

(32) Kresse, G.; Furthmuller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **1996**, *54*, 11169–11186.

(33) Zunger, A.; Wei, S. H.; Ferreira, L. G.; Bernard, J. E. Special Quasirandom Structures. *Phys. Rev. Lett.* **1990**, *65*, 353–356.

(34) Rupp, M.; Tkathenko, A.; Muller, K.-R.; von Lilienfeld, O. A. Fast and Accurate Modeling of Molecular Atomization Energies with Machine Learning. *Phys. Rev. Lett.* **2012**, *108*, 058301.

(35) Huan, T. D.; Mannodi-Kanakkithodi, A.; Ramprasad, R. Accelerated Materials Property Predictions and Design Using Motif-Based Fingerprints. *Phys. Rev. B* **2015**, *92*, 014106.

(36) Pilania, G.; Wang, C.; Jiang, X.; Rajasekaran, S.; Ramprasad, R. Accelerating Materials Property Predictions Using Machine Learning. *Sci. Rep.* **2013**, *3*, 2810.

(37) Mannodi-Kanakkithodi, A.; Pilania, G.; Huan, T. D.; Lookman, T.; Ramprasad, R. Machine Learning Strategy for Accelerated Design of Polymer Dielectrics. *Sci. Rep.* **2016**, *6*, 20952.

(38) Botu, V.; Ramprasad, R. Adaptive Machine Learning Framework to Accelerate Ab Initio Molecular Dynamics. *Int. J. Quantum Chem.* **2015**, *115*, 1074.

(39) Botu, V.; Ramprasad, R. Learning Scheme to Predict Atomic Forces and Accelerate Materials Simulations. *Phys. Rev. B* **2015**, *92*, 094306.

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

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

(42) Kim, C.; Pilania, G.; Ramprasad, R. Machine Learning Assisted Predictions of Intrinsic Dielectric Breakdown Strength of ABX_3 Perovskites. *J. Phys. Chem. C* **2016**, *120*, 14575.

(43) Snyder, J. C.; Rupp, M.; Hansen, K.; Muller, K. R.; Burke, K. Finding Density Functionals with Machine Learning. *Phys. Rev. Lett.* **2012**, *108*, 253002.

(44) Pilania, G.; Mannodi-Kanakkithodi, A.; Uberuaga, B. P.; Ramprasad, R.; Gubernatis, J. E.; Lookman, T. Machine Learning Bandgaps of Double Perovskites. *Sci. Rep.* **2016**, *6*, 19375.

Graphical TOC Entry