

Accelerated Search for BaTiO₃-Based Ceramics with Large Energy Storage at Low Fields Using Machine Learning and Experimental Design

Ruihao Yuan, Yuan Tian, Dezhen Xue,* Deqing Xue, Yumei Zhou, Xiangdong Ding,* Jun Sun, and Turab Lookman*

The problem that is considered is that of maximizing the energy storage density of Pb-free BaTiO₃-based dielectrics at low electric fields. It is demonstrated that how varying the size of the combinatorial search space influences the efficiency of material discovery by comparing the performance of two machine learning based approaches where different levels of physical insights are involved. It is started with physics intuition to provide guiding principles to find better performers lying in the crossover region in the composition–temperature phase diagram between the ferroelectric phase and relaxor ferroelectric phase. Such an approach is limiting for multidopant solid solutions and motivates the use of two data-driven machine learning and design strategies with a feedback loop to experiments. Strategy I considers learning and property prediction on all the compounds, and strategy II learns to preselect compounds in the crossover region on which prediction is carried out. By performing only two active learning loops via strategy II, the compound (Ba_{0.86}Ca_{0.14})(Ti_{0.79}Zr_{0.11}Hf_{0.10})O₃ is synthesized with the largest energy storage density $\approx 73 \text{ mJ cm}^{-3}$ at a field of 20 kV cm^{-1} , and an insight into the relative performance of the strategies using varying levels of knowledge is provided.

1. Introduction

A key challenge in developing better materials is the size of the space to be searched for the optimal chemistries and/or processing conditions. The number of experimentally syn-

thesized compounds available to learn from is often a tiny fraction of this total allowed space and this “needle in a haystack” problem requires efficient means of sampling the space to hone in rapidly on regions of potential interest.^[1–3] We have customarily used our intuition and physics-based understanding to guide the search for materials with better response.^[4,5] For example, the susceptibility of a material can be significantly enhanced in the proximity to a phase transition region because the instabilities in the order parameters are most acute there.^[6,7] However, this kind of search is still limited especially in complex systems because of the large search space associated with varying chemistry, processing conditions and microstructures, as well as the resultant size of the search space.

There has recently been much interest in employing machine learning (ML) and optimization methods to guide experimental synthesis to find materials with

targeted properties.^[8–18] The state-of-art in this field is to use an iterative approach, which is largely data-driven, starting from a set of features or material descriptors based on material knowledge to construct a surrogate model learned from data.^[19,20] Ideas and methods from decision theory and experimental design then provide the means to make optimal decisions of the experiments or materials to test next.^[19–22] The strategy has been quite successful in finding alloys and ceramics with enhanced properties.^[11–14,23,24] Even though Bayesian based algorithms using utility functions, such as expected improvement to maximize or minimize the objective or property, have proved especially successful in finding compounds with desired properties, the search space is often too large and leads to excessive exploration. This is a reflection of a multidimensional search space, and what is needed is to isolate regions in this space containing a relatively large number of extrema so that the probability of finding a compound with the desired targeted response is much higher than in the original search space. The physics based understanding of the problem can certainly help by effectively narrowing down the search space. To this end, we compare two machine learning based strategies using varying levels of

R. Yuan, Y. Tian, Prof. D. Xue, D. Xue, Prof. Y. Zhou, Prof. X. Ding, Prof. J. Sun State Key Laboratory for Mechanical Behavior of Materials Xi'an Jiaotong University Xi'an 710049, China E-mail: xuedezhen@xjtu.edu.cn; dingxd@xjtu.edu.cn

R. Yuan, Dr. T. Lookman Theoretical Division Los Alamos National Laboratory Los Alamos, NM 87545, USA E-mail: txl@lanl.gov

 The ORCID identification number(s) for the author(s) of this article can be found under <https://doi.org/10.1002/advs.201901395>.

© 2019 The Authors. Published by WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

DOI: 10.1002/advs.201901395

physics input to control the search space and difficulty of the task to be performed.

We will focus on the problem of finding compounds with better dielectric energy storage density from as few experiments as possible from a given training data in the easily synthesized BaTiO₃-based ceramics. Dielectrics capable of storing and releasing charges are the essential elements in modern electronics and electrical applications such as hybrid electric vehicles, portable electronic devices as well as power pulse devices because of their high power densities owing to the much faster charge–discharge speed than the electrochemical counterparts.^[25–28] Studies of energy storage in devices have largely focused on performance at high electric fields aimed at increasing breakdown dielectric strengths.^[25] For example, polymer-based dielectrics and thin films have attracted much interest in achieving superior energy storage density at very high electric fields where the break down field E_b can reach 6000 kV cm⁻¹ or more.^[29,30] The increase in applied electric field, however, is challenging to the supporting insulation system in the device. It often limits the applications using miniaturized components as well as portable or wearable electronic devices demanding high level of integration. Therefore, it is desirable to search for energy storage materials at relatively small electric field strengths with manageable larger energy storage density. For example, Gao et al. synthesized Ba(Ti_{0.895}Sn_{0.105})O₃ with an energy storage density of ≈55 mJ cm⁻³ at 20 kV cm⁻¹ based on operating in the region of tricritical behavior in the phase diagram. This exceeds most of the ferroelectric materials at the same field strength.^[31,32] In the present study, we introduce several dopants into the BaTiO₃ prototype to search for solid solutions with enhanced energy storage density at a field of 20 kV cm⁻¹. The Ba²⁺ is substituted by Ca²⁺, Sr²⁺, and the Ti⁴⁺ by Zr⁴⁺, Sn⁴⁺, and Hf⁴⁺, which gives rise to a vast search space of almost 9 million candidate choices as the concentration of each cation can be controlled to 0.01 in the synthesis process. It is too large for an Edisonian-based approach. Hence, our focus will be on finding solid-solutions with large dielectric energy storage density from this ≈9 million space of possible compounds using machine learning to guide experiments.

We will address this problem in two ways, (A) we will use a physics approach and consider the phase diagram of the BaTiO₃ family from a different perspective than considered previously, namely, we will define a crossover region in the composition–temperature phase diagram between the ferroelectric phase and relaxor ferroelectric phase. Our hypothesis is that the energy storage density is enhanced in this crossover region in the composition–temperature diagram and use this as a design principle in (B). We will provide experimental results on BaTi_{1-x}Zr_xO₃, BaTi_{1-x}Hf_xO₃, and BaTi_{1-x}Sn_xO₃, to validate this hypothesis. (B) We will use two data-driven ML strategies. Strategy I requires us to search across a relatively large search space exceeding 9 million possible compounds. We will use regression to build a surrogate model and then perform experimental design iteratively to make the best selections for synthesis. Strategy II will constrain the initial search space by using input from (A) to search only within a subclass of compounds in the crossover region to increase the probability of finding those compounds likely to lead to enhanced properties. We will do this by first using classification learning to predict

compounds belonging to this subclass only, and then perform regression and experimental design on these compounds to predict the one to consider next for synthesis. We compare in this work the performance of the different strategies. We will show that formulating the materials design question appropriately, as in strategy II, not only minimizes the number of experiments required (two) to find compounds with better energy storage density than the best in our training data, but also leads to compounds with even better energy storage density compared to the ML strategy I or the approach of (A). The best compound (Ba_{0.86}Ca_{0.14})(Ti_{0.79}Zr_{0.11}Hf_{0.10})O₃ we have found in this work shows superior energy storage density ≈73 mJ cm⁻³ at a field of 20 kV cm⁻¹, particularly accompanied with excellent energy storage efficiency (≈90%). This exceeds the performance of most of the ferroelectric materials known today at the same field strength.

2. Results

Physics Insights from the Phase Diagram: In general, the energy storage density U_{re} of ferroelectrics (shaded area of the polarization–electric field curves in **Figure 1a**) is given by $U_{re} = \int_{P_r}^{P_{max}} E dP$, where P_{max} is the maximum polarization and P_r is the remnant polarization.^[33] Therefore, U_{re} is strongly dependent on P_r , P_{max} , and E , where the maximum of E is limited by the dielectric breakdown strength (E_b). To improve U_{re} , a low P_r and a high P_{max} can potentially be realized by chemically modification and designed by utilizing the phase diagram of ferroelectrics

A schematic phase diagram showing how the U_{re} is influenced by varying the dopant content is shown in **Figure 1a**. As it possesses several symmetry allowed degenerate polarization states separated by large energy barriers, a long-range ordered ferroelectric exhibits a square polarization–electric field (P – E) loop.^[34] Thus even though the P_{max} is large, a large P_r is also retained, leading to low U_{re} and efficiency. Inserting more point defects into a normal ferroelectric can generate an abnormal disordered polarization state, usually referred to as the relaxor ferroelectric state.^[35–37] In contrast to the large ordered domains of micrometer size in normal ferroelectrics, the microstructure of relaxors is characterized by polar nanoregions (PNR) embedded in a paraelectric matrix.^[35] The state can be “kinetically frozen” with only short-range order and its response to electric stimuli is similar to linear dielectrics, that is, small P_r and P_{max} , leading to a low U_{re} .^[38,39] However, in the compositional crossover region between normal ferroelectrics and relaxor ferroelectrics, the state is still disordered but not fully frozen. This affords an opportunity to be induced into long-range ordered state with large P_{max} and to recover the initial disordered polarization state on removal of the external stimuli (a small P_r).^[38,39] Thus, the compositions in the crossover region potentially possess a fairly large P_{max} and small P_r . Therefore our design recipe, after building a phase diagram containing both normal and relaxor ferroelectrics, is to select compositions in the crossover region where the U_{re} is expected to be optimized.

Accordingly, we synthesized and established phase diagrams for three different systems, BaTi_{1-x}Zr_xO₃, BaTi_{1-x}Hf_xO₃, and BaTi_{1-x}Sn_xO₃. These systems are representative ferroelectrics and have been intensely investigated previously. The top

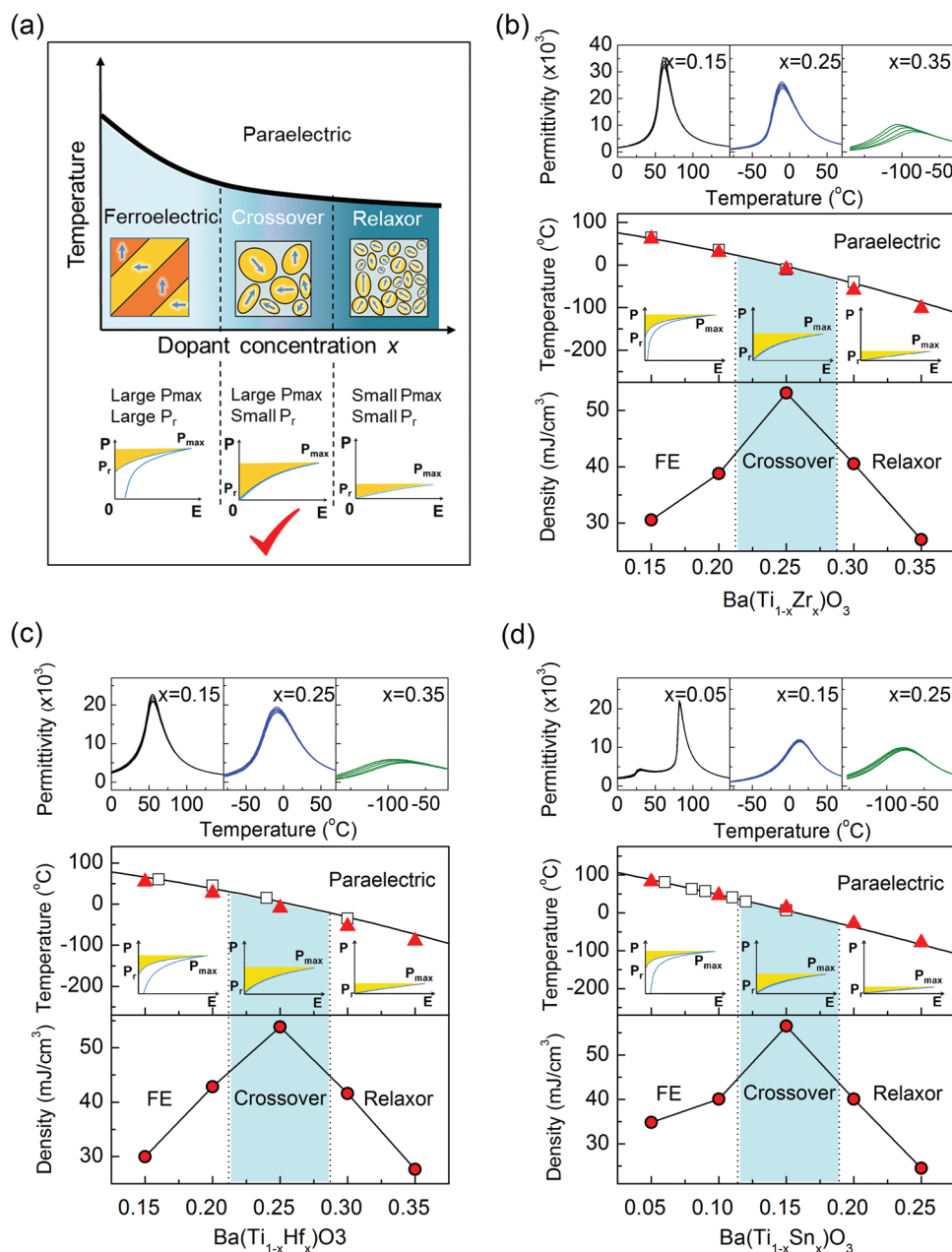


Figure 1. Schema shows U_{re} of different regions and validation in three simple systems. a) Schema from physical perspective shows that compounds located in the crossover region have higher U_{re} , b–d) three simple systems, $\text{BaTi}_{1-x}\text{Zr}_x\text{O}_3$, $\text{BaTi}_{1-x}\text{Hf}_x\text{O}_3$, and $\text{BaTi}_{1-x}\text{Sn}_x\text{O}_3$ show compounds in the crossover region indeed have higher U_{re} compared to relaxors and ferroelectrics, as highlighted by the shaded area. Moreover, our composition–temperature phase diagrams (red triangles) show a good agreement compared to studies in the literatures (squares), frequency dependence of permittivity of typical samples are also presented.

panels of Figure 1b–d show typical permittivity versus temperature ($\epsilon-T$) curves at various frequencies for the three systems, respectively. A sharp transition permittivity peak without frequency dispersion characterizes the normal ferroelectric at low dopant concentration and a smeared permittivity peak with frequency dispersion is the principal feature of the relaxor ferroelectric at high dopant concentration. The further identification of the phase transition type using modified Curie–Weiss law can be seen in Figure S1 (Supporting Information). The middle panels of Figure 1b–d give the composition–tempera-

ture phase diagrams, which agree well with those reported in the literatures,^[40–42] with three distinct composition regions for normal ferroelectrics, crossover, and relaxor ferroelectrics. The concentration dependence of the U_{re} calculated from $P-E$ curves (Figure S1, Supporting Information) is plotted in the bottom panels of Figure 1b–d corresponding to each phase diagram. A peak value indeed appears in all three systems at the crossover region.

The U_{re} obtained in the above simple systems ($\approx 55 \text{ mJ cm}^{-3}$ at the crossover composition) is not particularly large, and

chemical modification to both A site and B site of the perovskite structure should be performed to further optimize the U_{re} . Although this design approach with insights from the phase diagram works well for the above systems, it is rather difficult to apply this recipe to multicomponent compounds, for example, $(\text{Ba}_{1-x-y}\text{Ca}_x\text{Sr}_y)(\text{Ti}_{1-u-v-w}\text{Zr}_u\text{Sn}_v\text{Hf}_w)\text{O}_3$ as the combination of different ions results in numerous possible pseudobinary phase diagrams. We constrain the mole fractions x , y , u , v , and w by $1 - x - y \geq 0.6$, $x \leq 0.4$, $y \leq 0.3$, $1 - u - v - w \geq 0.6$, $u \leq 0.3$, $v \leq 0.3$, and $w \leq 0.3$, respectively. The bounds are set by considering the solubility of the dopants and ensures the transition temperature is not too low to be detected. With an unexplored space of ≈ 9 million possible compounds, we will use two ML and global optimization strategies within an iterative feedback loop to search for materials with enhanced U_{re} .

Strategy I: Searching from All Compounds in Phase Diagram: We employed a purely data-driven strategy with a training data set of 182 compounds and a search space of ≈ 9 million possible compounds. The training data set was obtained from the same lab using traditional solid-solid reaction. The data file can be found in Supporting Information. The inner loop in **Figure 2** schematically shows the feedback loop, including key ingredients: i) a training data $\approx 0.002\%$ of the whole search space, each with the targeted U_{re} measured; ii) a ML model that uses the training data to learn the relationship between features, x_i and property, U_{re} , i.e., $U_{re} = f(x_i)$; iii) the trained model is applied to the search space of ≈ 9 million unexplored compositions to predict U_{re} with associated model uncertainty; iv) utility function

used to maximize the expected utility aids in the selection of the next candidate material; and v) feedback from experiments with subsequent improvement of the ML model.

Each compound is described in terms of features, x_i , which include four different attributes. The change of Curie temperature due to the addition of dopants to the BaTiO_3 matrix is important, which is abbreviated as NCT. If the doped element decreases the Curie temperature, NCT is assigned a value of -1 ; if the doped element increases the Curie temperature, NCT is assigned $+1$; and if the doped element has no effect on the Curie temperature, NCT is assigned a value of 0 . For example, the addition of Zr element into the BaTiO_3 matrix will decrease the Curie temperature, thus NCT for Zr is -1 . The “tolerance factor” (t), the ionic displacement of B-site elements (DB), and the ratio of the element polarizability of A-site and B-site elements (P). The feature for a given compound is then calculated as the weighted mean using the mole fraction of each element. The detailed definition and selection of features are given in Section S2 (Supporting Information). The property U_{re} was obtained by integrating the P - E loop measured under an electric field of 20 kV cm^{-1} for each sample in the training data.

We built and compared 4 different ML models to estimate $U_{re} = f(\text{NCT}, t, \text{DB}, P)$ and selected the support vector machine with a radial-based kernel function (SVR.rbf) based on their mean squared and cross-validation errors. The details of model selection are given in Section S3 (Supporting Information). The performance of this SVR.rbf model comparing the predicted mean values and experimental measured values is plotted in **Figure 3a**. We used cross-validation to find the hyperparameters for a robust SVR.rbf model, balancing the trade-off between overfitting and under-fitting. The mean values μ of the predicted property together with the associated standard deviation σ are obtained by the “bootstrap” method of statistics by randomly sampling 1000 times with repeats from the training data of 182 compounds to build 1000 models. Thus the SVR.rbf model predicts the value of U_{re} of unexplored compounds and the associated uncertainties.

Choosing proper candidates for next experiment is then crucial to efficiently search for the materials with better U_{re} . We used efficient global optimization (EGO), which balances the exploration and exploitation aspects of the search to select next candidates.^[13,22,43] EGO evaluates the “expected improvement” (EI), a measure of the possible improvement with respect to the best performer in the training data. Also, $\text{EI} = E[\max(\gamma - \mu^*, 0)] = \int_{\mu^*}^{\infty} (\gamma - \mu^*)P(\gamma | x') d\gamma = \sigma[\phi(z) + z\Phi(z)]$, where $\gamma - \mu^*$ is the possible improvement for certain γ or energy density, $P(\gamma | x')$ is the distribution of the predicted γ and is assumed to be normal, and $z = (\mu - \mu^*)/\sigma$, where μ^* is the maximum value of the energy density (U_{re}) in the training data, μ and σ are the predicted value of U_{re} and the associated uncertainty, respectively. The $\phi(z)$ and $\Phi(z)$ are the standard normal density and distribution functions, respectively. Maximizing EI provides an optimal means to balance trade-off between exploitation ($\sigma \rightarrow 0$) and exploration ($\sigma \rightarrow \infty$).^[22] This selector considers both the predicted value and its associated uncertainty to choose the next experiment, in contrast to merely choosing the largest predicted value. As we can synthesize four compounds at a time, we predicted four compounds, which were synthesized and characterized and the training data augmented.

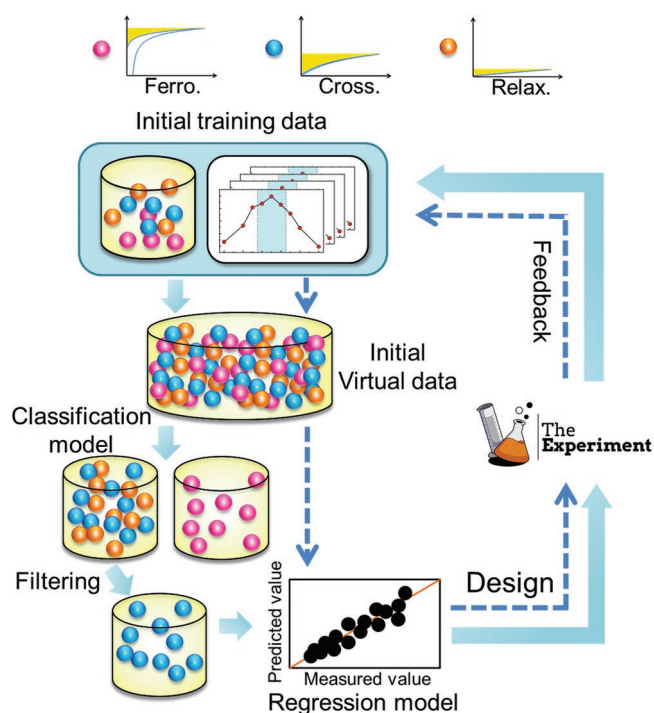


Figure 2. Two closed feedback loops. Inner loop with dashed arrows includes only regression ML model used in strategy I. Outer loop combines classification with regression ML models and will be used in strategy II. In each iteration, experimental results were accumulated into the training data and new models were trained and used for the next iteration.

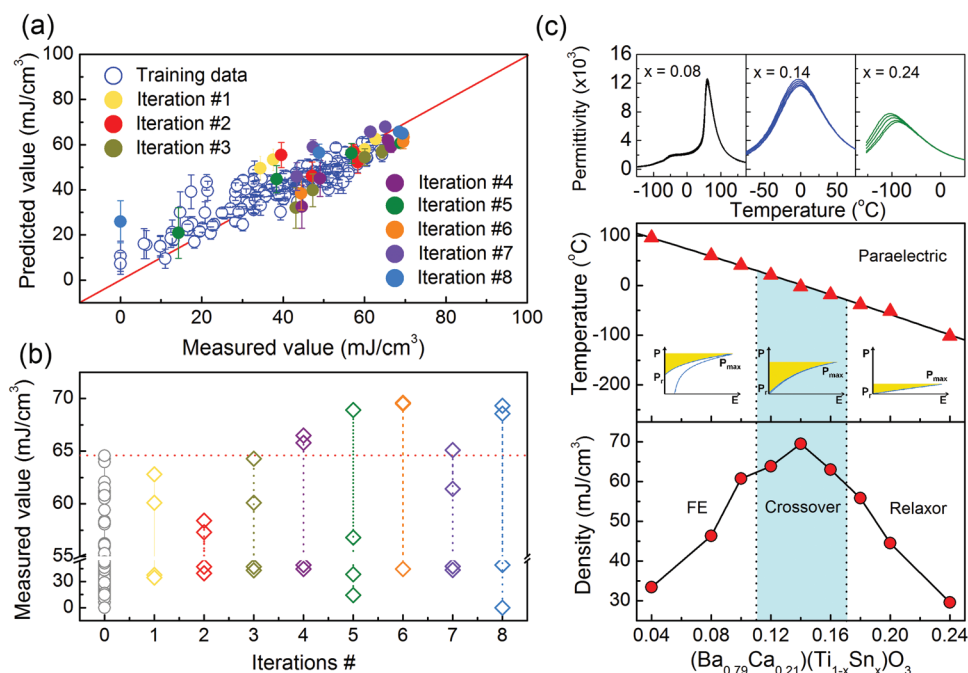


Figure 3. The regression model performance and the new compounds found in strategy I. a) Predicted values obtained from ML model as a function of measured values, error bar is from bootstrap sampling. b) Measured values as a function of iteration #, red line represents the best U_{re} in the training data, best material is found in the 6th iteration. c) Typical permittivity spectra from which the composition–temperature phase diagram of $(\text{Ba}_{0.79}\text{Ca}_{0.21})(\text{Ti}_{1-x}\text{Sn}_x)\text{O}_3$ is established. Crossover region is shadowed and shows better U_{re} .

We iterated the feedback loop eight times, and the measured versus predicted U_{re} with uncertainties are plotted on the top of the initially trained ML model in Figure 3a with solid points. Figure 3b shows the measured U_{re} of newly made 32 compounds (listed in Table S2 in the Supporting Information) as a function of 8 iterations. The best U_{re} in each iteration decreases first and increases until the peak value shown in the 6th iteration, followed by another decrease. The behavior of the predicted values as a function of iterations can be found in Figure S9 (Supporting Information), which presents a similar tendency to that in Figure 3b. The three best compounds were obtained in the 6th iteration after synthesizing and characterizing 24 compounds. The best U_{re} has a value of 69.5 mJ cm^{-3} , which is about 8% higher than that of the best one in the training data.

To validate the assumption that the best compound we discovered is located within the crossover region, we experimentally established a composition–temperature pseudo-binary phase diagram by varying the Sn^{4+} content with fixed A-site cations in the best performing compound, $(\text{Ba}_{0.79}\text{Ca}_{0.21})(\text{Ti}_{0.86}\text{Sn}_{0.14})\text{O}_3$. The phase diagram of $(\text{Ba}_{0.79}\text{Ca}_{0.21})(\text{Ti}_{1-x}\text{Sn}_x)\text{O}_3$ is shown in Figure 3c. The ϵ – T curves of various Sn^{4+} compounds (Figure S10, Supporting Information) demonstrate that the best compound from strategy I is located in the crossover region. Typical curves are shown in the top panel of Figure 3c and the crossover region is highlighted by the shadow. The U_{re} obtained from P – E curves (Figure S10, Supporting Information) is plotted as a function of Sn^{4+} content in the bottom panel. The best compound $(\text{Ba}_{0.79}\text{Ca}_{0.21})(\text{Ti}_{0.86}\text{Sn}_{0.14})\text{O}_3$ outperforms the surrounding compositions, indicating at least one local maximum of the search space.

Strategy II: Searching from Preselected Compounds in Crossover Region: This strategy only consider the compounds that fall within the crossover regions. As these compounds are likely the extrema of U_{re} , we anticipate that the design optimization will be more favorable and less exploratory in terms of number of iterations required. The strategy is schematically shown by the outer loop in Figure 2. A key difference compared to strategy I is that a binary classification model is first used to down select crossover region compounds. It is noted that a multiclass classifier to “ferroelectric”, “relaxor”, and “crossover” is an alternative way to predict the “crossover” compositions.

The training data for classification includes 183 compounds labeled as either ferroelectric or relaxor (Supporting Information). The training data thus has 56 samples labeled relaxor and 127 samples labeled ferroelectric. Compounds in the crossover region show a little frequency dispersion in the dielectric spectrum and therefore we labeled them as relaxors. We used the mole fraction of each ion as feature to build the classifier, which was a support vector machine with a radial-based kernel function. We calculated the receiver operating characteristic (ROC) curve that uses the probability of prediction to evaluate the performance of a binary classification model. Every point on the line shows how the classifier performs at a given threshold.^[44] The initial labeled 183 compounds were randomly separated into a training set (140 compounds) and a test set (43 compounds). The same algorithm was used to train a model in the training set and then validated on the test set. **Figure 4a** shows the ROC curve based on the test set, where the red dashed line represents the random guess. The area under the ROC curve is 0.94 which ensures a robust classification model. We also evaluated the performance of the classifier by considering

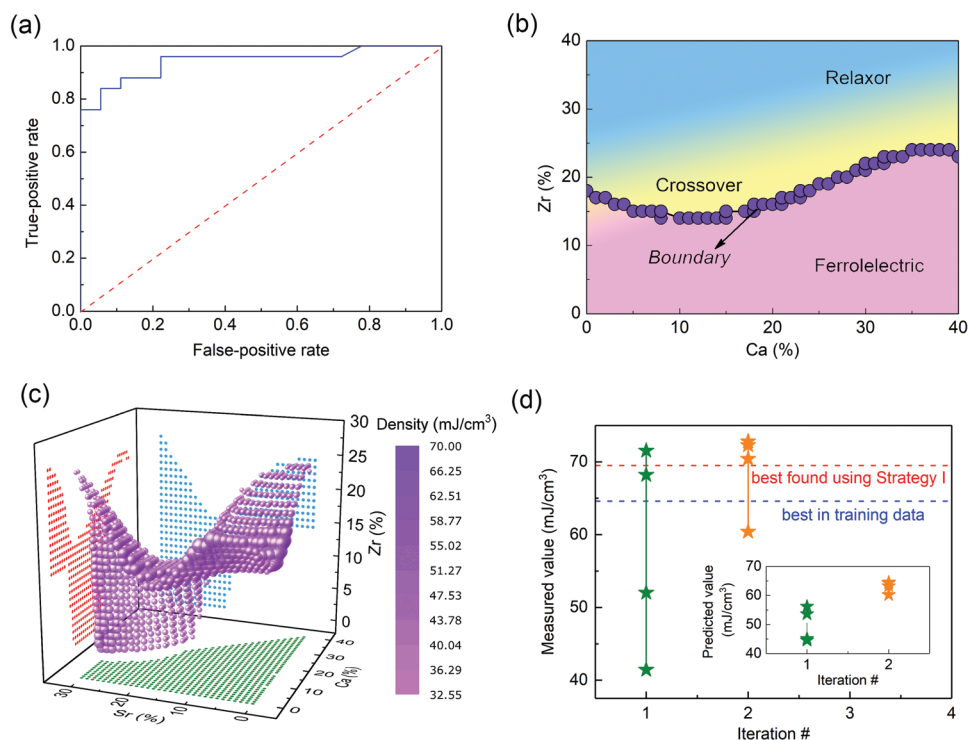


Figure 4. Performance of the classification model and the new 8 compounds synthesized in strategy II. a) ROC curve to validate our classifier. b) A projection of the boundary to a 2D phase diagram featured by elements Zr and Ca. c) A 3D phase diagram to show the boundary between normal ferroelectrics and relaxors (includes crossover region) predicted from classification model; above the boundary are compounds in crossover and relaxor region, below are normal ferroelectrics. d) Measured values as a function of number of iterations; inset shows that the predicted values change with iterations. Red and blue dashed lines indicate the best U_{re} found in strategy I and in training data, respectively.

training, cross-validation and test errors (details can be seen in Figure S11 in the Supporting Information). The misclassification rate for each compound in the training data is also calculated using the bootstrap method and listed in Table S3 (Supporting Information).

We applied the classifier to the unexplored space of ≈ 9 million compounds, with each compound assigned as either ferroelectric or relaxor. The compounds on the phase boundary between ferroelectric and relaxor were determined by monitoring their nearest neighbors. For each compound in the unexplored space, if at least one nearest neighbor is labeled differently from the others, this compound is considered to be on the boundary. To clarify, Figure 4b shows a 2D phase diagram where the axes are the features monitoring the composition of Zr and Ca. The boundary (or the margin) clearly distinguishes ferroelectrics from relaxors. Given that we have a crossover region rather than just a boundary between the ferroelectrics and relaxors, we incorporate compounds within the crossover region by considering compounds up to four neighbors or steps into the relaxor region. As a 5D plot is challenging to visualize, a plot of the boundary between normal ferroelectrics and relaxors in three dimensions is shown in Figure 4c, in terms of the Zr, Ca, and Sr compositions with the relaxors above separated from the ferroelectrics below by the compounds on the boundary. The crossover region includes $\approx 690\,000$ compounds in total and is much smaller compared to the whole unexplored space of ≈ 9 million compounds. Most importantly, the compounds in this smaller search space have the most promising potential to show enhanced U_{re} .

The same regression model based on 182 compounds ($\approx 0.025\%$ of the smaller unexplored space) trained in strategy I was employed to make predictions on this smaller search space. We again used expected improvement as the selection criterion, as shown in Figure 2. After synthesis and characterization, the experimental results were added to the training data for both classification and regression, and subsequent iteration proceeds accordingly. The outer loop in Figure 2 iterates twice and 8 new compounds were synthesized and characterized. The corresponding measured U_{re} (calculated from P - E curves in Figure S12 in the Supporting Information) are listed in Table 1. Five samples show comparable or higher U_{re} than

Table 1. The 8 new compounds synthesized in strategy II (five of these (in bold) have a comparable or higher U_{re} than the best found in strategy I).

Iteration #	Ba	Ca	Sr	Ti	Zr	Sn	Hf	U_{re}
1	0.97	0.03	0.00	0.86	0.00	0.14	0.00	52.1
1	0.61	0.09	0.30	0.90	0.10	0.00	0.00	41.4
1	0.81	0.19	0.00	0.84	0.00	0.16	0.00	68.2
1	0.87	0.13	0.00	0.79	0.11	0.00	0.10	71.5
2	0.86	0.14	0.00	0.79	0.05	0.00	0.16	70.4
2	0.80	0.20	0.00	0.83	0.00	0.17	0.00	60.4
2	0.87	0.13	0.00	0.79	0.10	0.00	0.11	72.3
2	0.86	0.14	0.00	0.79	0.11	0.00	0.10	72.8

the best value found in strategy I (red dashed line), as shown in Figure 4d. The best compound ($\text{Ba}_{0.86}\text{Ca}_{0.14}(\text{Ti}_{0.79}\text{Zr}_{0.11}\text{Hf}_{0.10})\text{O}_3$) found in strategy II shows a U_{re} of $\approx 73 \text{ mJ cm}^{-3}$, improved by 14% compared to the best in the training data. Inset plot indicates that predicted values versus iterations show a similar tendency to the measured values. The ϵ - T curves in Figure S13 (Supporting Information) show a degree of diffusion, confirming that all 5 new compounds are likely located in the crossover region. Thus, invoking the physics associated with the crossover region between normal and relaxor ferroelectrics has allowed us to effectively reduce the search space to significantly accelerate finding high U_{re} compounds.

3. Discussion and Summary

To glean some insights into the two strategies, Figure 5a shows the initial distribution of predictions of the virtual space (≈ 9 million) for strategy I, where the predictions vary from 5 to 63 mJ cm^{-3} with a peak located at a U_{re} around 30 mJ cm^{-3} . The initial distribution of predictions in the virtual space for ≈ 0.7 million compounds in strategy II shifts to higher values, as shown in Figure 5b. All the predictions are larger than 30 mJ cm^{-3} and the range now is from 30 to 63 mJ cm^{-3} , in particular, the peak is located at a U_{re} around 39 mJ cm^{-3} . Thus, compared to strategy I that possesses many local extrema

as shown in Figure 5c, strategy II effectively preselects a subset of compounds (using classification model) from the total space that are likely to be favored, as shown in Figure 5d. This will avoid excess exploration around local extrema.

In summary, we proposed a guiding principle which finds better energy storage properties in the crossover region of the composition-temperature phase diagram between the ferroelectric phase and relaxor ferroelectric phase. As a prior information of phase diagram is needed, this approach is limiting in complex multicomponent systems. We thus have compared the performance of two ML strategies to find multicomponent solid-solutions with enhanced energy storage density, U_{re} in BaTiO_3 -based ceramics using experimental data. Strategy I applied ML model and adaptive design directly to a huge virtual space of ≈ 9 million compounds. After 6 iterations, compositions were found that improved the targeted property by 8% compared to the best in the training data. By using physical insights from the composition-temperature phase diagram that allowed only a subclass of compounds to be considered in the virtual space, strategy II utilized ML in the form of classification and regression models, as well as adaptive design. The search therefore becomes far more efficient as only 2 iterations were needed to obtain 5 compounds with comparable or higher U_{re} than the best found in strategy I. The best compound ($\text{Ba}_{0.86}\text{Ca}_{0.14}(\text{Ti}_{0.79}\text{Zr}_{0.11}\text{Hf}_{0.10})\text{O}_3$) found in the reduced search space (with 5 easy to process elements) has a U_{re}

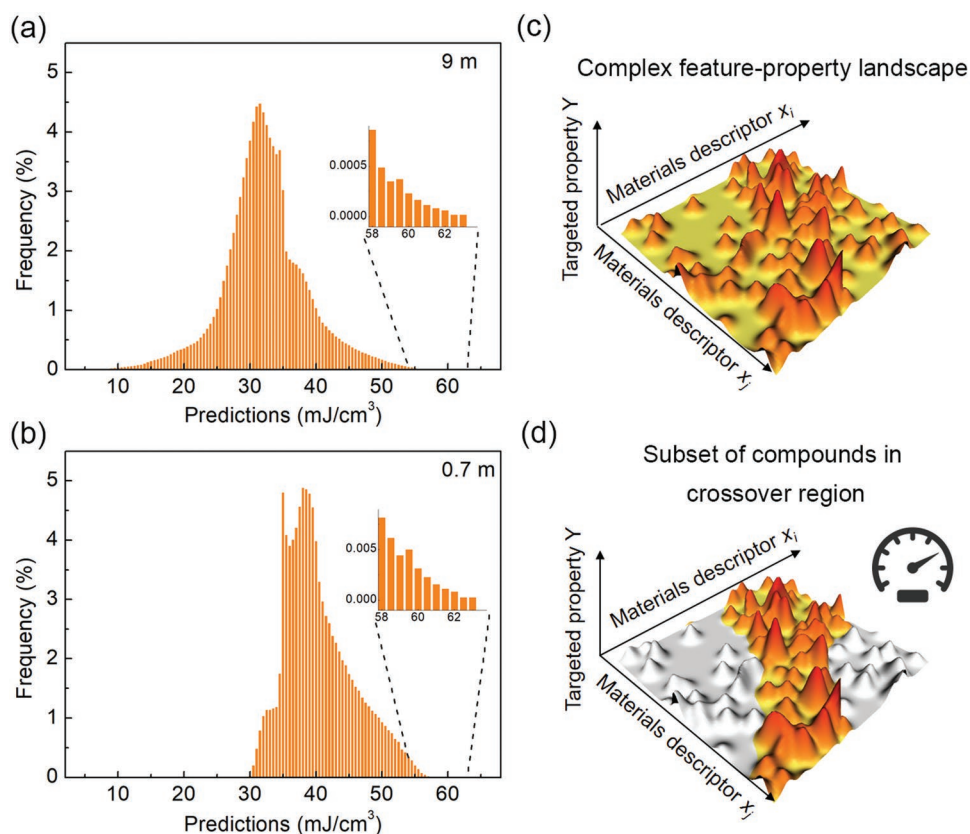


Figure 5. Distribution of predictions of the virtual space used in strategy I and strategy II. a) The distribution of predicted U_{re} of virtual space in strategy I, inset is the amplification of the part with large predicted values. b) The distribution of predicted U_{re} of virtual space in strategy II, inset is the amplification of the part with large predicted values. c) Schema to show a complex landscape with many local extrema. d) Major local extrema are removed from the virtual space using physical understanding, there is a greater probability of finding large U_{re} in this landscape.

of $\approx 73 \text{ mJ cm}^{-3}$ at 20 kV cm^{-1} , 14% better than the best in the training data. The relative performance of the two strategies is revealed from a statistical perspective. We envision the proposed framework can be extended to other complex systems to accelerate the searching process of new materials.

4. Experimental Section

Experimental: $(\text{Ba}_{1-x-y}\text{Ca}_x\text{Sr}_y)(\text{Ti}_{1-u-v-w}\text{Zr}_u\text{Sn}_v\text{Hf}_w)\text{O}_3$ ceramics were fabricated by a conventional solid–solid reaction method with the starting materials of BaCO_3 (99.8%), CaCO_3 (99.9%), SrCO_3 (99.9%), BaZrO_3 (99.9%), SnO_2 (99.9%), HfO_2 (99.8%), and TiO_2 (99.6%). The calcination was performed at $1350 \text{ }^\circ\text{C}$ for 3 h and sintering was done at $1450 \text{ }^\circ\text{C}$ for 3 h in air. All the samples were synthesized under the same conditions to reduce the dependence of targeted property on processing. The sintered samples for ferroelectric and dielectric measurements were polished to obtain parallel sides and coated with silver electrodes. The polarization–electric field (P – E) loops were identified by a ferroelectric workstation at 10 Hz and frequency dependence of permittivity on temperature were measured using a HIOKI 3532-50 LCR meter, all with disk-shaped samples.

Machine Learning Models: Four ML regression models are employed in this study, support vector machine with a radial-based kernel function, ridge regression with a radial-based kernel function, gradient boosting decision tree, and random forest. The former two models use tenfolds cross-validation to choose the best hyperparameters. In the first three models, the predictions and uncertainties are obtained from 1000 models built on 1000 resampling training data with repeats. For the random forest model, 500 000 trees are trained based on the resampling method with repeats. In terms of the classification model, support vector machine with a radial-based kernel function is used, and leave-one-out cross-validation is used to choose the best hyperparameters. The support vector machine with a radial-based kernel function for regression and classification was implemented in the e1017 package within the RSTUDIO environment.

Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

This work was supported by the National Key Research and Development Program of China (2017YFB0702401), National Natural Science Foundation of China (Grant Nos. 51571156, 51671157, 51621063, and 51320105014), and the 111 project 2.0 (BP2018008). The authors are also grateful to the Laboratory Directed Research Development (LDRD) program at LANL for support via project # 20180660ER.

Conflict of Interest

The authors declare no conflict of interest.

Keywords

Bayesian optimization, ceramics, energy storage, machine learning, optimal experimental design

Received: June 7, 2019

Revised: August 17, 2019

Published online: September 2, 2019

- [1] K. T. Butler, D. W. Davies, H. Cartwright, O. Isayev, A. Walsh, *Nature* **2018**, 559, 547.
- [2] R. Ramprasad, R. Batra, G. Pilania, A. Mannodi-Kanakkithodi, C. Kim, *npj Comput. Mater.* **2017**, 3, 54.
- [3] S. Ramakrishna, T.-Y. Zhang, W.-C. Lu, Q. Qian, J. S. C. Low, J. H. R. Yune, D. Z. L. Tan, S. Bressan, S. Sanvito, S. R. Kalidindi, *J. Intell. Manuf.* **2018**, 30, 2307.
- [4] J. Scott, Electrocaloric materials, *Annu. Rev. Mater. Res.* **2011**, 41, 229.
- [5] J. Rödel, K. G. Webber, R. Dittmer, W. Jo, M. Kimura, D. Damjanovic, *J. Eur. Ceram. Soc.* **2015**, 35, 1659.
- [6] X. Moya, E. Stern-Taulats, S. Crossley, D. González-Alonso, S. Kar-Narayan, A. Planes, L. Mañosa, N. D. Mathur, *Adv. Mater.* **2013**, 25, 1360.
- [7] W. Liu, X. Ren, *Phys. Rev. Lett.* **2009**, 103, 257602.
- [8] G. Pilania, C. Wang, X. Jiang, S. Rajasekaran, R. Ramprasad, *Sci. Rep.* **2013**, 3, 2810.
- [9] P. Raccuglia, K. C. Elbert, P. D. Adler, C. Falk, M. B. Wenny, A. Mollo, M. Zeller, S. A. Friedler, J. Schrier, A. J. Norquist, *Nature* **2016**, 533, 73.
- [10] Z. Li, Q. Xu, Q. Sun, Z. Hou, W.-J. Yin, *Adv. Funct. Mater.* **2019**, 29, 1807280.
- [11] D. Xue, P. V. Balachandran, J. Hogden, J. Theiler, D. Xue, T. Lookman, *Nat. Commun.* **2016**, 7, 11241.
- [12] D. Xue, P. V. Balachandran, R. Yuan, T. Hu, X. Qian, E. R. Dougherty, *Proc. Natl. Acad. Sci. USA* **2016**, 113, 13301.
- [13] R. Yuan, Z. Liu, P. V. Balachandran, D. Xue, Y. Zhou, X. Ding, J. Sun, D. Xue, T. Lookman, *Adv. Mater.* **2018**, 30, 1702884.
- [14] P. V. Balachandran, B. Kowalski, A. Sehirliglu, T. Lookman, *Nat. Commun.* **2018**, 9, 1668.
- [15] F. Ren, L. Ward, T. Williams, K. J. Laws, C. Wolverton, J. Hattrick-Simpers, A. Mehta, *Sci. Adv.* **2018**, 4, eaaq1566.
- [16] R. Yuan, D. Xue, D. Xue, Y. Zhou, X. Ding, J. Sun, T. Lookman, *IEEE Trans. Ultrason., Ferroelectr., Freq. Control.* **2019**, 66, 394.
- [17] S. Sanvito, C. Oses, J. Xue, A. Tiwari, M. Zic, T. Archer, P. Tozoman, M. Venkatesan, M. Coey, S. Curtarolo, *Sci. Adv.* **2017**, 3, e1602241.
- [18] S. Lu, Q. Zhou, Y. Ouyang, Y. Guo, Q. Li, J. Wang, *Nat. Commun.* **2018**, 9, 3405.
- [19] T. Lookman, P. V. Balachandran, D. Xue, J. Hogden, J. Theiler, *Curr. Opin. Solid State Mater. Sci.* **2017**, 21, 121.
- [20] T. Lookman, F. J. Alexander, K. Rajan, *Information Science for Materials Discovery and Design*, Springer, Switzerland **2016**.
- [21] J. Theiler, B. G. Zimmer, *Stat. Anal. Data Min.: ASA Data Sci. J.* **2017**, 10, 211.
- [22] D. R. Jones, M. Schonlau, W. J. Welch, *J. Global Optim.* **1998**, 13, 455.
- [23] D. Xue, D. Xue, R. Yuan, Y. Zhou, P. V. Balachandran, X. Ding, J. Sun, T. Lookman, *Acta Mater.* **2017**, 125, 532.
- [24] C. Wen, Y. Zhang, C. Wang, D. Xue, Y. Bai, S. Antonov, L. Dai, T. Lookman, Y. Su, *Acta Mater.* **2019**, 170, 109.
- [25] B. Chu, X. Zhou, K. Ren, B. Neese, M. Lin, Q. Wang, F. Bauer, Q. M. Zhang, *Science* **2006**, 313, 334.
- [26] P. Khanchaitit, K. Han, M. R. Gadinski, Q. Li, Q. Wang, *Nat. Commun.* **2013**, 4, 2845.
- [27] X. Zhang, Y. Shen, B. Xu, Q. Zhang, L. Gu, J. Jiang, J. Ma, Y. Lin, C.-W. Nan, *Adv. Mater.* **2016**, 28, 2055.
- [28] Q. Chen, Y. Shen, S. Zhang, Q. Zhang, *Annu. Rev. Mater. Res.* **2015**, 45, 433.
- [29] A. Patsidis, K. Kalaitzidou, G. Psarras, *Mater. Chem. Phys.* **2012**, 135, 798.
- [30] V. Ranjan, L. Yu, M. B. Nardelli, J. Bernholc, *Phys. Rev. Lett.* **2007**, 99, 047801.
- [31] J. Gao, Y. Wang, Z. He, Y. Liu, D. Wang, L. Jin, T. Zhao, L. Zhong, X. Ren, *Adv. Funct. Mater.* **2019**, 29, 1807162.

- [32] J. Gao, Y. Wang, Y. Liu, X. Hu, X. Ke, L. Zhong, Y. He, X. Ren, *Sci. Rep.* **2017**, *7*, 40916.
- [33] Z. Yao, Z. Song, H. Hao, Z. Yu, M. Cao, S. Zhang, M. T. Lanagan, H. Liu, *Adv. Mater.* **2017**, *29*, 1601727.
- [34] Y. Xu, *Ferroelectric Materials and Their Applications*, Elsevier, The Netherlands **2013**.
- [35] L. E. Cross, *Ferroelectrics* **1987**, *76*, 241.
- [36] V. Westphal, W. Kleemann, M. Glinchuk, *Phys. Rev. Lett.* **1992**, *68*, 847.
- [37] C. Lei, A. A. Bokov, Z.-G. Ye, *J. Appl. Phys.* **2007**, *101*, 084105.
- [38] F. Li, S. Zhang, D. Damjanovic, L.-Q. Chen, T. R. Shrout, *Adv. Funct. Mater.* **2018**, *28*, 1801504.
- [39] V. K. Thakur, R. K. Gupta, *Chem. Rev.* **2016**, *116*, 4260.
- [40] Z. Yu, C. Ang, R. Guo, A. S. Bhalla, *J. Appl. Phys.* **2002**, *92*, 1489.
- [41] Y. Yao, C. Zhou, D. Lv, D. Wang, H. Wu, Y. Yang, X. Ren, *Europhys. Lett.* **2012**, *98*, 27008.
- [42] T. Mitsui, W. B. Westphal, *Phys. Rev.* **1961**, *124*, 1354.
- [43] P. V. Balachandran, D. Xue, J. Theiler, J. Hogden, T. Lookman, *Sci. Rep.* **2016**, *6*, 19660.
- [44] N. A. Obuchowski, *Radiology* **2003**, *229*, 3.