

## Stochastic replica voting machine prediction of stable cubic and double perovskite materials and binary alloys

T. Mazaheri, Bo Sun, and J. Scher-Zagier

*Department of Physics, Washington University in St. Louis, Missouri 63160, USA*

A. S. Thind

*Institute of Materials Science and Engineering, Washington University in St. Louis, Missouri 63130, USA*

D. Magee

*Department of Physics, Washington University in St. Louis, Missouri 63160, USA  
and MIT Lincoln Laboratory, Lexington, Massachusetts 02421, USA*

P. Ronhovde

*Department of Physics, Washington University in St. Louis, Missouri 63160, USA  
and Department of Physical Sciences, The University of Findlay, 1000 N. Main St., Findlay, Ohio 45840, USA*

T. Lookman

*Theoretical Division, Los Alamos National Laboratory, MS-B262, Los Alamos, New Mexico 87545, USA*

R. Mishra\*

*Department of Mechanical Engineering and Materials Science, and Institute of Materials Science and Engineering,  
Washington University in St. Louis, Missouri 63130, USA*

Z. Nussinov†

*Department of Physics, Washington University in St. Louis, Campus Box 1105, 1 Brookings Drive, St. Louis, Missouri 63130, USA*



(Received 14 November 2017; revised manuscript received 5 April 2019; published 19 June 2019)

A machine-learning approach that we term the “stochastic replica voting machine” (SRVM) algorithm is presented and applied to a binary and a three-class classification problem in materials science. Here, we employ SRVM to predict candidate compounds capable of forming stable perovskites and double perovskites and further classify binary ( $AB$ ) solids. The results of our binary and ternary classifications compared well to those obtained by SVM and neural network algorithms.

DOI: [10.1103/PhysRevMaterials.3.063802](https://doi.org/10.1103/PhysRevMaterials.3.063802)

### I. INTRODUCTION

Recently, there has been a flurry of activity involving the use of machine learning, an important subfield of artificial intelligence, in the study of materials and complex physical systems, e.g., [1–10]. Data mining techniques enable a rapid search through millions of candidate compounds in order to identify promising technological materials and to potentially predict their detailed properties. Such a task may require far more significant efforts when performed experimentally [11,12] via the traditional trial and error approach. Machine learning can make such searches far more efficient by systematically pointing to promising materials that may then be fabricated and tested experimentally. In this paper, we will focus on two material types: perovskites and binary alloys.

Perovskites (named after the Russian nobleman and mineralogist Lev Perovski) are a large class of compounds having an  $ABX_3$  stoichiometry, where  $A$  and  $B$  are cations and  $X$  is an anion [13,14]. Numerous technologically important materials display the perovskite structure shown in Fig. 1. Some examples include certain high-temperature superconductors, semiconductors for high-efficiency photovoltaic cells [15], light-emitting diodes, lasers, and solid-oxide fuel cells (see, e.g., Refs. [16,17]).

In the examples that we will study here,  $X$  will be an oxygen anion. Following a standard convention, the  $A$  atoms are defined to be the larger of the two cations. An ideal perovskite has a cubic crystal structure that is formed by corner-sharing  $BO_6$  octahedra as seen in Fig. 1. As is seen in this figure,  $A$  ions lie at the corners of the cube while the  $B$  and  $O$  ions are, respectively, located at the body center and face centers of the cube. The  $BO_6$  octahedra can distort and neighboring octahedra can tilt and rotate. This lowers the cubic crystal symmetry but accommodates a large combination of cations from the periodic table.

\*rmishra@wustl.edu

†zohar@wuphys.wustl.edu

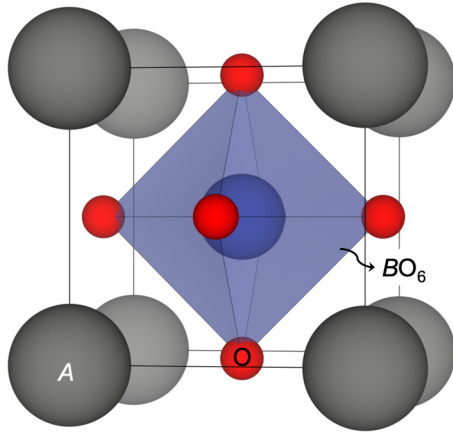


FIG. 1. The structure of an  $ABO_3$  perovskite. A quintessential material having this structure is  $\text{CaTiO}_3$ .

To ensure stability, the relative size of the  $A$  and  $B$  cations must, typically, satisfy certain criteria [18]. (Additional illuminating relations between the atomic radii and structure are found in [19].) More complex double perovskites [11] (see Fig. 2) exhibit the same architecture yet with a larger unit cell; these compounds are of the generic chemical composition  $A'A''B'B''O_6$  (or, more generally,  $A'_yA''_{2-y}B'_zB''_{2-z}O_6$  with  $0 < y', z' < 2$ ). Figure 2 provides an illustration of a double perovskite with two different  $A$  cations and  $B$  cations. There are over 80 elements having, at least, one stable nuclide. Thus, *a priori*, numerous combinations of these elements may potentially realize stable perovskite or double perovskite structures. To experimentally determine the stability of the vast number of these candidate perovskites would be an arduous if not impossible task. In recent years, materials scientists have turned to machine-learning models, often combined with first-principles total-energy calculations based on density-functional theory (DFT), to predict the stability of new theoretical compounds. Along similar lines, in this work, we will introduce an algorithm that takes in different elements as inputs and predicts whether or not their combination will result in a stable perovskite or double perovskite.

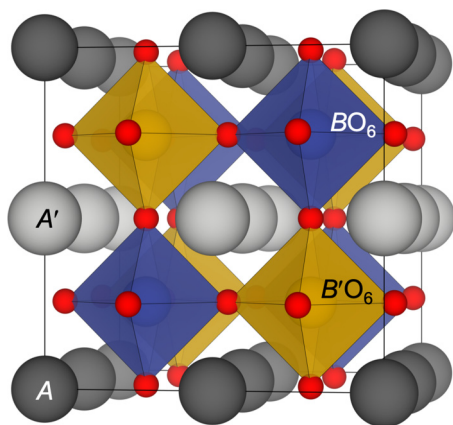


FIG. 2. A double perovskite structure with two different cations at both the  $A$  site and the  $B$  site.

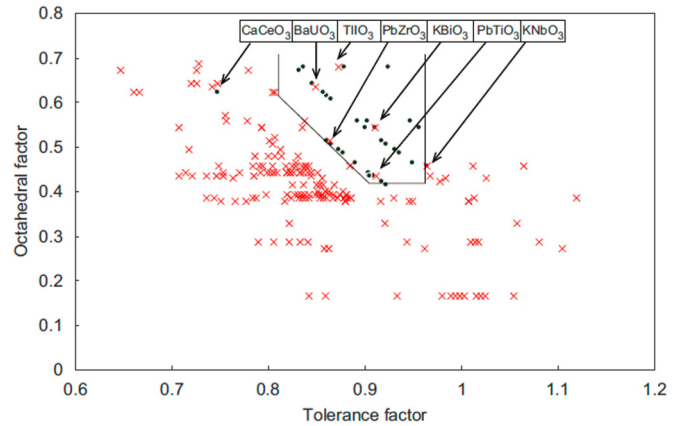


FIG. 3. Reproduced from [18]. Classification of cubic perovskite oxides. Candidate perovskite compounds are displayed according to their tolerance and octahedral factors. (Black) dots indicate stable perovskite compounds; compositions marked by (red) crosses indicate unstable compounds. A goal of our work is to predict which materials might be stable.

Our algorithm takes, as an input, the data from known combinations of  $A$  and  $B$  cations that are capable of forming a perovskite structure (see, e.g., [18] and Fig. 3 for perovskite formability, Sec. I of the Supplemental Material [20], and Fig. 2 for double perovskite). From these data, the algorithm learns which conditions should be met for the different elements in order to allow them to form the perovskite structure. This is the so-called “training process.” Following the training phase, the algorithm may test other combinations of the ions; the algorithm will then yield a “yes” answer for a predicted stable perovskite structure and yield “no” for compositions that are predicted to form an unstable perovskite structure. In the parlance of machine learning, we are training a new binary classifier over a set of known data. Once this training is complete, we then apply the trained classifier to investigate hitherto unknown chemical compositions in order to assess their formability as stable perovskites. We will follow the prevalent practice of classifying the stability of candidate perovskite materials by two well-studied ratios: (i) the “tolerance factor” and (ii) the “octahedral factor.” We will further study other features including electronegativity.

In this work, we will introduce and summarize our algorithm (general details are further discussed in Ref. [21]), and demonstrate its utility for the classification (viable formability) of (1) perovskite-type compounds and (2) binary octet alloys. In both cases, we achieve high accuracy. Our method enables the prediction of stable perovskites and the properties of binary compounds. Other works (e.g., [22–24]) study various aspects of perovskites with existing machine-learning algorithms. In this work, we employed a very general machine-learning algorithm (whose details will be reported on in [21]) and delineated phase boundaries in the two classification problems that we investigated.

Our bare binary classifier can be trivially extended to nonbinary (multiclass) problems via, e.g., the “one-versus-rest” approach [25]. We will detail a three-class problem when investigating binary (“ $AB$ ”) alloys.

The remainder of this work is organized as follows: In Sec. II, we provide a description of our algorithm. In Secs. II A and II B therein, we will, respectively, outline the Gaussian and multinomial variants of our algorithm. We will next explain (Sec. III) how we train the algorithm to ascertain perovskite formability. In Sec. IV, we apply a neural network analysis based version of our approach to the study of double perovskites when only the tolerance and octahedral factors are provided. In Sec. V, we will apply both neural network and Gaussian kernel based versions of our algorithm to the study of double perovskite formability when five features (that further incorporate ionic electronegatives) are included. With this analysis we may make predictions as to which candidate systems might be stable perovskites. In Sec. VI, we contrast the SRVM predictions for stable double perovskites with detailed DFT calculations for the enthalpy of formation. Finally, in Sec. VII, we will invoke the “one-versus-all” approach to a ternary classification problem involving  $AB$  solids [26].

## II. STOCHASTIC REPLICA VOTING MACHINE ALGORITHM

As befits its name, our “stochastic replica voting machine” (SRVM) algorithm relies on a voting procedure among stochastically generated classifiers. As we will explain, these individual classifiers are defined by a kernel that may be of any type: e.g., a sum of Gaussians or a multinomial. Initially, we “train” the system to predict the correct answer. The trained system may then subsequently predict the outcome given initial inputs. Training is performed by adjusting the kernel of each individual classifier such that it reproduces known results. The voting of classifiers is then given new data and a vote is taken amongst the predictions of the individual classifiers.

The input (“training set”) data for  $N$  items that need to be classified is given in terms of a set of vectors  $\{\vec{v}_i\}_{i=1}^N$  defining the features of the items and their corresponding classification  $\rho_i$ . If the classification is amongst  $q$  different groups, then classification function is a Potts spin variable whose value  $\rho_i = 1, 2, \dots, q$  denotes the group that item  $i$  correctly belongs to. Potts variables may be generally used as a classification index in numerous arenas, e.g., [27–30]. The features of each item are combined into a vector  $\vec{v} = (v_1, v_2, \dots, v_d)$ . Thus, the Cartesian components of each vector  $\vec{v}_i$  are equal to the values of all parameters of the input data associated with item  $i$  (e.g., the values of the individual atomic radii of the ions forming in a candidate perovskite material). If numerous features are given for each data point  $i$ , then the dimensionality ( $d$ ) of the vectors  $\vec{v}_i$  will be high. The goal of machine learning is to make an educated guess (a “prediction”) as to what the corresponding classification outcome will be for a new vector  $\vec{v}$  for which there is no *a priori* correct classification known outcome. Since no additional information is available, the predicted outcome  $\rho$  can only be some function  $F$  of all supplied input: the features defining  $\vec{v}$  and all known training set data. That is, the underlying assumption of any machine-learning approach is that

$$\rho(\vec{v}) = F(\vec{v}; \{\vec{v}_i\}_{i=1}^N, \{\rho_i\}_{i=1}^N). \quad (1)$$

The natural question is as follows: “How may we determine the correct or ‘optimal’ function  $F$ ”? Numerous machine-learning approaches exist. We briefly comment on two of these. In one important subclass of these, known as “support vector machines” (SVM) (e.g., [31,32]),  $F$  is implicitly ascertained by inequalities applied to assumed specific function types. In neural network based machine learning [33,34], in particular in “deep learning” [35], the function  $F$  is formed by a particular hierarchal recursive structure. Our approach (SRVM) is, in some regards, more rudimentary yet, as we will explain, may extend these and other prevalent models. To illustrate its basic premise, we will consider the binary (i.e.,  $q = 2$ ) classification problem. Here,  $\rho_i = 1, 2$  and thus  $\tilde{\rho}_i \equiv (2\rho_i - 3) = \pm 1$  naturally classifies any data point  $\vec{v}$  into one of two groups (labeled by  $\tilde{\rho}_i = 1$  and  $\tilde{\rho}_i = -1$ ). We define  $\tilde{F} \equiv (2F - 3)$  and initially consider  $\tilde{F}$  to be an outcome of a vote amongst the predictions of a large *voting* of general continuous stochastic functions  $\{G_a\}_{a=1}^r$  (that we need not be of different types) where  $r$  is the number of “replicas” in this voting. We will, principally, focus on two broad types of stochastic functions. We will examine what occurs (a) if  $\{G_a\}_{a=1}^r$  are expressible as sums of random basis functions. We then turn to (b) functions  $\{G_a\}_{a=1}^r$  generated by a weighted averages of “neural network type function”; by the latter, we allude to functional composition of linear and Fermi function (also known as a “sigmoid” function in the machine-learning community). Apart from SVM, related machine-learning approaches include “voting” methods [36], which employ randomly generated data sets, “boosting” [37] which aims to combine different weaker algorithms into a stronger learners, “decision tree learners” [38], and “random forest” methods [39] which employ decision trees to combine the results of various classifiers and often do so while choosing different subspaces of the features or given data. While our algorithm enjoys many commonalities with these and other approaches, we underscore that its essential character is that of a stochastic average over different randomly generated functions. Basically, we perform (a discrete sum version of) “functional integration” over different randomly generated fits to the given data set in order to suggest the most likely outcomes for given vectors  $\vec{v}$ . In performing these “functional integration” averages, the SRVM algorithm that we briefly introduce below does not discard any data points nor features to generate lower-dimensional subspaces on which a voting of classifiers is trained on.

We start with stochastic functions of type (a) and expand in terms of simple (equivalent type) nonorthogonal randomly chosen basis functions. A simple choice for the functions  $G_a$  (that will be investigated in this work) is one in which they are a sum of  $R$  random Gaussian functions [40]. Thus, we set

$$G_a = \sum_{j=1}^R c_{ja} e^{-(\vec{v}-\vec{v}_{ja})^2/(2\sigma_{ja}^2)}, \quad (2)$$

where  $\{c_{ja}\}_{j=1}^R$  are coefficients that we will discuss momentarily. In the most minimal form of  $G_a$ , all standard deviations  $\sigma_{ja}$  are set to a uniform fixed value  $\sigma_{ja} = \sigma$ . The centers  $\{\vec{v}_{ja}\}$  of the Gaussians are randomly chosen in the volume spanned by the feature space. Thus, for each of the  $r$  functions  $\{G_a\}_{a=1}^r$ , we randomly choose  $R$  “anchor points” in the feature space

volume to be  $\{\bar{v}_{ja}\}_{j=1}^R$ . The location of these anchor points differs from replica to replica. That is, we define each replica “A” by a different stochastic set of vectors  $\{\bar{v}_{ja}\}_{j=1}^R$ . More comprehensive than the specific choice of random Gaussians in Eq. (2), the functions  $G_a$  may be generally chosen to be of the form

$$G_a = \sum_{j=1}^R c_{ja} K_a^j(\bar{v}). \quad (3)$$

Here, the kernel (or basis) functions  $K_a^j$  could be any arbitrary stochastic functions. For the Gaussian form of Eq. (2),  $K_a^j(\bar{v}) = e^{-(\bar{v}-\bar{v}_{ja})^2/(2\sigma_{ja}^2)}$ . Other general kernels  $K$ , different from a Gaussian function, may, of course, be considered. For instance, another natural (yet typically computationally expensive) choice for the kernel  $K_a$  that we will return to in this work (reasonable when the outcome likelihood is expected to be analytic as a function of the features) is that of multinomials in the Cartesian components of  $\bar{v}$ .

During the training phase, we optimize the values of the coefficients  $\{c_{ja}\}_{j=1}^R$  given the known outcome for the training points  $i = 1, 2, \dots, N$  such that  $G_a(\bar{v}_i)$  matches the correct classification  $\bar{\rho}_i$ . The number  $R$  of the coefficients required in order to achieve high prediction accuracy is typically smaller than the number of training data points  $R < N$  (in most instances, in fact,  $R \ll N$ ). The optimal value of  $R$  depends on the nature of data as well as the size of data and should be chosen carefully to avoid overfitting. For each replica  $a = 1, 2, \dots, r$ , the given training data set translates into linear equations for  $\{c_{ja}\}_{j=1}^R$ . Thus, Eq. (3) explicitly reads as  $G_a(\bar{v} = \bar{v}_i) = \sum_{j=1}^R K_a^{ij} c_{ja}$ , where  $K_a^{ij} \equiv K_a^j(\bar{v} = \bar{v}_i)$ . This embodies a set of overdetermined (since  $N > R$ ) linear equations for the coefficients  $\{c_{ja}\}$ . For each of the replicas  $a = 1, 2, \dots, r$ , the above relation can be trivially cast as an explicit matrix equation  $\hat{G}_a = \hat{K}_a \hat{c}_a$ . Here,  $\hat{G}_a$  and  $\hat{c}_a$  are two column vectors of, respectively, lengths  $N$  and  $R$  whose entries are, respectively,  $\{G_a(\bar{v} = \bar{v}_i)\}_{i=1}^N$  and  $\{c_{ja}\}_{j=1}^R$ . The elements of the rectangular  $N \times R$  dimensional matrix  $K_a$  are, as defined above, given by  $(\hat{K}_a)^{ij} \equiv K_a^j(\bar{v} = \bar{v}_i)$ . The coefficients  $\hat{c}_a$  minimizing the cost function or “energy” defined by the square sum  $\|\hat{G}_a - \hat{K}_a \hat{c}_a\|^2$  are given by

$$c_{ja} = \sum_i (\hat{K}_a^{-1})_{ji} G_{ia}. \quad (4)$$

Here, the rectangular matrix  $\hat{K}_a^{-1}$  [with elements  $(\hat{K}_a^{-1})_{ji}$ ] is the pseudoinverse of  $\hat{K}_a$ . Thus, in the training phase, the goal is to find the coefficient vectors  $\hat{c}_a$  for each of the replicas  $a = 1, 2, \dots, r$ . With the above values of  $c_{ja}$  in tow, we may now predict the classification of a new “test” item  $\bar{v}$  different from all prior training data points (i.e.,  $\bar{v} \neq \bar{v}_i$  for  $1 \leq i \leq N$ ). That is, we may compute the classification of  $\bar{v}$  as predicted by the  $r$  independent replicated stochastic functions  $\{\text{sgn}(G_a(\bar{v}))\}_{a=1}^r$  (where  $\text{sgn}$  denotes the sign function) and then perform a vote amongst all of these classifiers. The vote then yields the final prediction of the SVRM,

$$\bar{\rho}(\bar{v}) = \text{sgn}\left(\sum_{a=1}^r \text{sgn}(G_a(\bar{v}))\right). \quad (5)$$

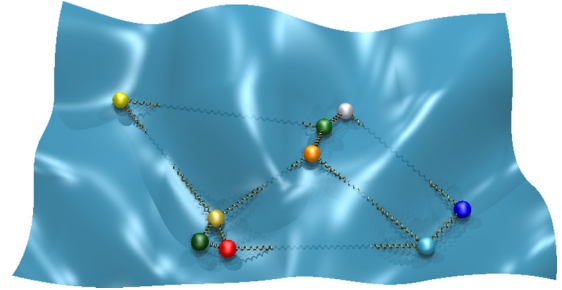


FIG. 4. A schematic representation of replica “interactions.” The spheres depict individual replicas that navigate an “energy landscape” looking for stable minima while simultaneously interacting with one another. As compared to a single solver, in the simplest setting, these “interactions” may correspond to a vote amongst their predictions and notably when seeing at which parameter values the predictions of the replicas are most uniform and thus robust. These “interacting” replicas may more readily avoid false minima and converge on the stable low-energy solutions leading to more stable and accurate predictions. In the algorithm that we outline in this work, the simplest multireplica vote of Eq. (5) is employed. We will further optimize the function types to fit by seeing when the replica predictions are most uniform.

For the  $q = 2$  classification problem that we have considered thus far, the inner  $\text{sgn}$  function in Eq. (5) may be replaced by other appropriately chosen symmetric functions  $\bar{\rho}(\bar{v}) = W(\{G_a(\bar{v})\})$  with the condition that may only assume the two values  $\pm 1$  (corresponding to the two possible classes to which an item  $\bar{v}$  may belong to). The voting in Eq. (5) emulates a more general multireplica “interaction” sketched in Fig. 4. In this schematic, each sphere denotes an individual replica. Together, the voting of replicas may better hone in on optimal predictions for the classification of  $\bar{v}$ . Putting all of the pieces together, Eqs. (3)–(5) nearly completely define the SRVM program. The kernels  $K_a^j$  may, *a priori*, stochastically be chosen to be of any particular functional form. Of course, if a theory exists, then the functional form of  $G_a$  may be more efficiently designed. In the absence of any such information, one may simply examine the predictions for random kernels  $K_a^j$ . There are three remaining inter-related natural questions:

- (1) Is there a particular metric to determine the confidence with which the results are predicted?
- (2) How do we determine the “optimal” number  $r$  of the replicas to be used?
- (3) Similarly, what sets the number  $R$  of kernel functions in Eq. (3)?

As we will describe, the answer to all questions may be determined by examining the overlap of the predictions of the different stochastic replica functions  $\{G_a\}_{a=1}^r$ . Throughout this work, we will employ a simple variant of the overlap  $\mathcal{O}(\bar{v})$  associated with any point  $\bar{v}$  whose classification is predicted by the  $r$  replicas  $\{G_1, G_2, \dots, G_r\}$ , namely,

$$\mathcal{O}(\bar{v}) \equiv \frac{1}{r} \left| \sum_{a=1}^r \text{sgn}(G_a(\bar{v})) \right|. \quad (6)$$

With this definition, we first explicitly turn to question (1). If all replicas yield identical predictions (and thus  $\mathcal{O}$  is close to unity), then (as is intuitively natural and we verified by

numerical experiments), this common predicted answer is likely correct. Analogously, if the replicas are far from a unanimous agreement about the correct classification (and, consequently,  $\mathcal{O}$  is much smaller than one), then the predicted answer cannot be trusted with high confidence. The above rule of thumb enables us to scan the parameters  $r$  and  $R$  to find values that are likely to yield optimal accuracy [questions (2) and (3) above]. Typically as the number of replicas  $r$  increases, so does the accuracy. However, larger values of  $r$  entail increasing computational costs with no real benefit. We thus seek sufficiently large  $r$  that enable high accuracy. By contrast, when the number of anchor points (or more general basis functions)  $R$  becomes too large, overfitting leads to increasing errors. There are optimal values of  $R$  that are sufficiently large to capture the characteristics of the data yet not so big that overfitting occurs. In reality, we may fix  $r$  and  $R$  to specific values and examine the replica overlap to ascertain whether the predicted values may be trusted [21]. That is, when the overlap  $\mathcal{O}$  is averaged over all new data points  $\vec{v}$  (whose correct classification is not *a priori* known and that needs to be classified by the algorithm) is high, then the consensus reflected by the average  $\mathcal{O}$  will suggest that the current parameters  $r$  and  $R$  defining Eqs. (3) and (6) enable a correct prediction of the classification problem.

A variant that we will touch on later is that of “an expansion in a box.” For typical basis functions  $K_a^j$ , the functional form of Eq. (3) assumes that the outcome is a generally smooth function of  $\vec{v}$ . If the system exhibits “phase transitions” as a function of the features ( $v_1, v_2, \dots, v_d$ ), then such an assumption is void. Instead, one may fit the training data with a particular function of the form of Eq. (3) with specific coefficients  $\{c_{ja}\}$  only when  $\vec{v}$  lies in a particular volume  $\bar{v} \in \Omega$ ; different regions will correspond to different functional forms (i.e., the coefficients  $\{c_{ja}\}$  may change from one region of  $\vec{v}$  space to another). Here, the expansion will be valid only in a particular “box.” The function  $G_a$  will be allowed to change as  $\vec{v}$  goes from being in one domain  $\Omega$  to another. Thus, in each of the domains  $\{\Omega_b\}$  comprising the system (in which the system is assumed to be “analytic”) there will be a different function  $G_{ab}$  (specified by coefficients  $c_{jab}$ ). In these cases, a natural question is how to ascertain phase transitions and effectively employ the existence of these volumes. Our approach here is once again that of noting when the overlap between different replicas is highest. That is, given a particular test point  $\vec{v}$ , we may train the system with all data that lie in a volume  $\Omega$  (a box) that encloses  $\vec{v}$ . We then see when, as a function of the size  $|\Omega|$ , the overlap  $\mathcal{O}(\vec{v})$  between the replicas for the predicted outcome at point  $\vec{v}$  will be the highest. We employed this approach when the overlap between the various replicas was small and our original classification outcome was less certain.

The accuracy of machine-learning classification algorithms is typically tested by randomly fitting a fraction  $z$  of the known data (i.e., using these data for “training”) and then seeing how well the algorithm correctly predicts the classification of the remaining data that are not used as training but rather supplied to the algorithm only as new vectors  $\vec{v}$  whose correct classification is known yet not given to the user but is to be predicted by the algorithm. This process [or training with

a fraction  $z$  of the data and *testing* the predictions on the remaining fraction of  $(1 - z)$ ] is repeated over and over again with different ways of splitting the known data into two subgroups of relative numbers set by a parameter  $z$ :

$$\begin{aligned} \text{training data points} &: \text{test data points} \\ &= z : (1 - z). \end{aligned} \quad (7)$$

The accuracy of the predicted classification is then averaged over the many ways of splitting the data with this ratio between the size of the number of training data points and the tested points kept fixed. In the accuracy tests that we will report on, we will follow the prevalent practice of choosing  $z = 0.8$ .

### A. Gaussian kernels

In what follows, we provide an explicit example in which the value of  $R$  (the number of basis functions) is determined. In the current context, we seek to find the optimal number  $R$  of anchor points for the Gaussian of Eq. (2). Toward this end, we may plot the average overlap  $\mathcal{O}$  between different replicas as a function of the number of replicas  $r$  and the number of anchor points  $R$ . This overlap enables us to determine the optimal values of  $r$  and  $R$  for which  $\mathcal{O}$  obtains its maximum (or, more generally, its maxima). In general, optimal parameter values (such as  $R$ ) used in the SRVM model (also for function types are Gaussian kernels) may be found by examining when the predictions of the different replicas are most robust. SRVM does not merely average over the predictions of different replicas. Rather, as sketched in Fig. 4, the replicas “interact” with one another so as to make their collective predictions uniform. In the current setting, this corresponds to a choice of  $R$  that maximizes the inter-replica overlap  $\mathcal{O}$ .

To explicitly illustrate the basic premise, we examine the data of the perovskite classification problem that we will turn to in greater detail later on. For the time being, we probe how the average of the overlap  $\mathcal{O}$  varies as a function of the number of basis functions used [or anchor points in the case of the Gaussian kernel of Eq. (2)]. (As remarked earlier [see discussion after Eq. (2)], the anchor points are randomly placed in the feature space.) As Fig. 5 illustrates, the overlap between different replicas is maximal for  $R \approx 60$  anchor points. Since the inter-replica overlap is maximal for this value of  $R$ , we suspect using this number of anchor points would result in the optimal accuracy. The average accuracy that we reached with the Gaussian kernel for determining stable perovskite oxides was 94.19%. This accuracy may be contrasted with the performance of a current state-of-the-art SVM package [41] employing radial basis (i.e., Gaussian) functions; the SVM method yielded a mean accuracy of 92.53%.

### B. Multinomial kernels

As we alluded to earlier, another set of kernels in Eq. (3) is afforded by a  $d$ -component vector  $\vec{j}$  defining monomials

$$K^{\vec{j}}(\vec{v}) = v_1^{j_1} v_2^{j_2} \dots v_d^{j_d}. \quad (8)$$

Here,  $v_k$  are the Cartesian components ( $1 \leq k \leq d$ ). There are a variety of ways to produce multinomial-based replica. For

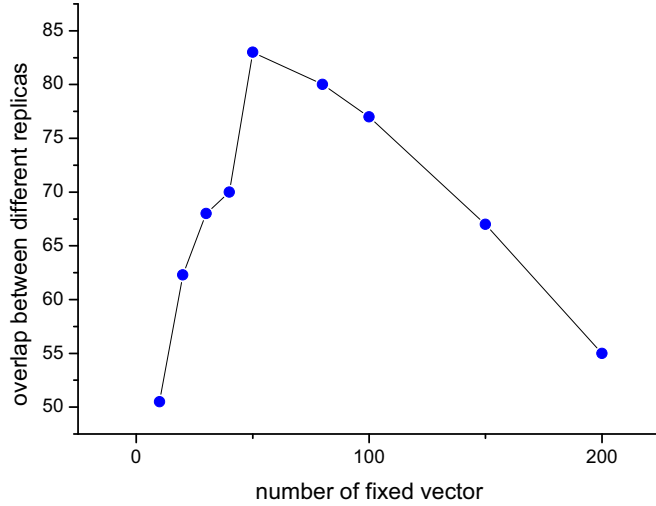


FIG. 5. The overlap of Eq. (6) averaged over  $r = 11$  replicas and different test points  $\vec{v}'$  that appear in different cross-validation partitions for the stability of candidate perovskites. Here, this average overlap (given in percentage) is computed for the predictions of different Gaussian-type [Eq. (2)] replicas having  $R$  fixed anchor points. The average inter-replica overlap is largest when  $R \approx 60$  (suggesting that the most accurate predictions are obtained for such values of  $R$ ).

instance, different rotations in parameter space may lead to independent multinomials. A general rotation  $v_k \rightarrow U_{kk'}^h v_{k'} \equiv v_{kh}$  with  $U^a$  a random rotation matrix will transform the monomial of Eq. (10) into a multinomial in which the sum of all powers in each of the individual monomials

$$J \equiv \sum_{k=1}^d j_k \quad (9)$$

is unchanged relative to its value in Eq. (10). Thus, if we choose a basis of monomials  $\{K_a^{\vec{j}}(\vec{v})\}$  with  $0 \leq j_k \leq p$  (with a general natural number  $p$ ) for all  $1 \leq k \leq d$  in one coordinate system  $v_k$ , then an independent basis of monomials is afforded by

$$K_a^{\vec{j}} = v_{1a}^{j_1} v_{2a}^{j_2} \dots v_{da}^{j_d}, \quad (10)$$

with  $j_k \leq p$ . This is so as the highest power of each of the Cartesian coordinates is  $p < J$ . In Eq. (10),  $\{v_{hk}\}_{k=1}^d$  are the coordinates in the rotated basis generated by  $U^a$ . Equation (3) may be used to concoct several replica functions  $G_a = \sum_{j=1}^R c_{ja} K_a^{\vec{j}}(\vec{v})$ .

### C. Deep neural network type functions

We next turn to functions of the sort associated with deep neural networks [33,34]. Following the standard recipe for constructing deep neural networks from individual neuronal activation functions [33–35,42,43], the individual random kernels underlying the SRVM method can be combined to construct more complex composite functions. Herein, the activation function of each neuron in the network is one of our randomly chosen functions (similar to those discussed in the earlier subsections). In the current work, when studying

deep neural network type constructs, we will employ common sigmoid (or Fermi function type) activation functions in simple neural networks of different architectures and contrast their results. More generally, reliable candidate network architecture and activation function types may (similar to the determination of other hyperparameter values in SRVM) be ascertained by noting when the highest inter-replica overlaps appear. The input to the node  $k$  in layer  $\alpha = 0, 1, 2, \dots, N_f$  is given by a linear combination

$$\begin{aligned} \epsilon_k^{(\alpha)} &= L_\alpha(\{\zeta_{k'}^{(\alpha-1)}\}, \{c_k^{(\alpha-1)}\}) \\ &\equiv -\left(\sum_{k'} w_{kk'}^{(\alpha-1)} \zeta_{k'}^{(\alpha-1)}\right) - c_k^{(\alpha-1)}, \end{aligned} \quad (11)$$

with  $w_{kk'}^{(\alpha-1)}$  and  $c_k^{(\alpha-1)}$  being constants. For  $N_f > \alpha' > 0$ , the function  $\zeta_{k'}^{(\alpha')}$  is of the Fermi (or sigmoid) type

$$\zeta_{k'}^{(\alpha')} = f(\epsilon_{k'}^{(\alpha')}) = \frac{1}{1 + e^{\epsilon_{k'}^{(\alpha')}}}. \quad (12)$$

The variables  $\zeta_k^{(0)} = v_k$  are input features. All  $(N_f - 1)$  “layers”  $\alpha \neq 0, 1$  are often termed “hidden layers.” In the last output layer there is only one value of  $\epsilon^{(N_f)}$  (i.e.,  $k = 1$  only); here,  $\zeta^{(N_f)} = \theta(-\epsilon^{(N_f)})$  with  $\theta(z) = [1 + \text{sgn}(z)]/2$  being the Heaviside function. That is, a value of  $\zeta^{(N_f)} = 1$  corresponds to a positive classification in the scheme of the earlier subsection [ $\tilde{\rho}(\vec{v}) = 1$ ] while  $\zeta^{(N_f)} = 0$  corresponds to a prediction of an assignment to the “no” class [ $\tilde{\rho}(\vec{v}) = -1$ ]. The coefficients  $w_{kk'}^{(\alpha)}$  form rectangular matrices with a number of rows equal to the number of “nodes” or “neurons”  $n_\alpha$  (i.e.,  $k = 1, 2, \dots, n_\alpha$ ) in layer  $\alpha$  and a number of columns set by the number of nodes  $n_{\alpha-1}$  in the preceding layer. The initial layer is that of the input values (i.e.,  $v_k = \zeta_k^{(0)}$  are the features). In typical neural nets, the single  $\zeta^{(N_f)}$  appearing in the final layer provides the prediction sought after in a classification problem. Specifically, in binary classification problems, if  $\zeta_f > 0.5$ , then the predicted classification of  $\vec{v}$  is of one type (e.g., “yes”) and if  $\zeta_f < 0.5$  the prediction is that  $\vec{v}$  belongs to the other class. Equations (11) and (12) schematically correspond to the composite of  $2N_f$  individual  $f$  and  $L$  functions

$$\theta(L_{N_f}(f(L_{N_f-1}(f(\dots L_1(\vec{v}) \dots))))). \quad (13)$$

All  $(N_f - 1)$  “layers” apart from the first transformation on the initial input vector  $\vec{v}$  are often termed “hidden layers.” Starting from random initial parameters  $\{w_{kk'}^{(\alpha)}\}$  and  $\{c_k^{(\alpha)}\}$ , we will optimize the parameters of such neural net functions to fit the data. Different initializations do not necessarily lead to a unique set of these parameters when iteratively optimized via backpropagation [44] to fit the training data. Different neural architectures having a different number of layers  $N_f$  and the set  $(n_1, n_2, \dots, n_{N_f})$  specifying the number of nodes in each iterative layer also constitute different functions of the same variety. As in Sec. II A, we may find the “optimal” neural net function types (including contending architectures) by seeing which types of architecture lend themselves to uniform consistent predictions amongst different replicas.

#### D. Ternary and multiclass SRVM

Thus far, we focused on binary classification [wherein the sign, Eq. (5), decided to which of two categories a particular point  $\vec{v}$  should belong to]. There is, of course, more to life than only binary classification. In order to classify  $\vec{v}$  into one of  $p > 2$  groups, various constructs are possible. One, very rudimentary, design is to iteratively classify as a point  $\vec{v}$  as belonging (or not) to any one of the classes  $q = 1, 2, \dots, p$ . Such a rudimentary approach emulates the well known “one-versus-all” (OVR) [25] technique; this is the what we will adopt in this work when we will classify  $AB$  solids into one of  $p = 3$  groups (Sec. VII). Specifically, we will start by predicting the results of an input vector  $\vec{v}$  for each of the possible output values with the SRVM algorithm that we introduced in the earlier subsections. Similar to the binary classification, in order to construct the pseudoinverse for the  $i$ th output value bivariate algorithm, we will set the result of a data point as +1 if it outputs a classification of type  $q$  and -1 otherwise. Instead of just taking the sign of the outputted results, we compared the raw values from results. That is, if the output associated with the vector  $\vec{v}_i$  as tested against candidate classes  $q = 1, 2, \dots, p$  had the highest incidence of positive values for a particular class  $q = q'$  then the vector  $\vec{v}_i$  was classified as belonging to group  $q'$ .

#### III. CUBIC PEROVSKITE FORMABILITY

In this section we employ SRVM to predict whether candidate  $ABO_3$  compounds form stable perovskite structures. The training data that we used [18] has  $d = 2$  features briefly noted in the Introduction: (i) The “tolerance factor”

$$v_1 \equiv \frac{r_A + r_X}{\sqrt{2}(r_B + r_X)}, \quad (14)$$

where  $r_{i=A,B,X}$  denote, respectively, the radii of the  $A$ ,  $B$ , and  $X$  ions, and (ii) the “octahedral factor” defined as the ratio

$$v_2 \equiv \frac{r_B}{r_X}. \quad (15)$$

The data in [18] consist of 223 candidate compounds of the  $ABO_3$  type. Of these compositions, 34 correspond to stable perovskite structures and the rest are unstable structures. (After removing duplicate compounds that share the same tolerance factor and octahedral factor, 188 data points remain, 29 of which form stable cubic perovskite structure.) Once the training is performed with input data, we use it to make the binary prediction regarding the stability of the contending perovskite compounds. Following Eq. (7), we repeatedly partitioned the data into two random subgroups with  $z = 0.8$ . Several partitions with this ratio were generated by the standard cross-validation method in which the data are divided into nearly five equal parts. Subsequently, four of these five sets are then used together to train the algorithm and the remaining one fifth of the data is used as a resource of test data to see how accurate the predictions of the algorithm are. The set that is used as the test data is cycled through (being chosen to be all of the five nearly equal parts of the data). The accuracy is then averaged over the predictions made over the five groups when these are used as test data. The accuracy is further averaged over different random partitions into five groups. Both for comparison as well as in order to obtain

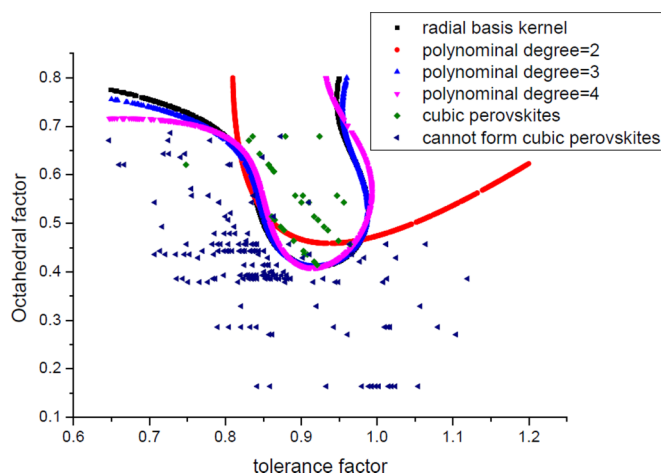


FIG. 6. Classification results using different SVM kernels employing the LIBSVM-3.0 package [41].

a more comprehensive picture, aside from our own SRVM algorithm, we also employed both the standard Gaussian and polynomial kernels in the well-known SVM method [31,32]. In Fig. 6, we provide the results that we obtained by applying the SVM algorithm for different kernels. In this figure, the region above the drawn curves (associated with individual SVM kernels) is predicted to correspond to stable perovskite structures; in the parameter region below these curves, no stable perovskite materials are anticipated. In line with our main thesis (that of inferring a likely outcome from multiple independent kernels), the region that is above all drawn curves corresponds to a domain in the  $v_1 v_2$  plane in which we may expect (with high confidence) stable perovskite structures. Similarly, in Figs. 7 and 8, we display the results obtained

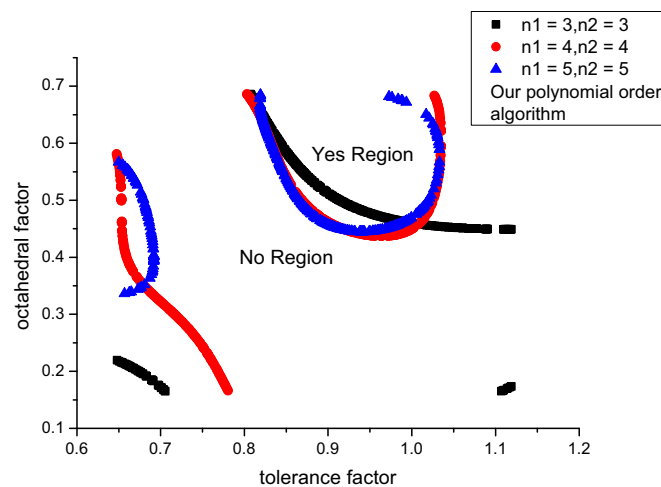


FIG. 7. The viable region in the tolerance factor ( $v_1$ ) and octahedral factor ( $v_2$ ) plane for materials that may form cubic perovskite structure as ascertained by a multinomial order kernel in the SRVM method. Here, we employed multinomials of three different orders (3, 4, and 5). The common region in which all multinomials predict formability of a cubic perovskite structure is marked as the “yes region.” Similarly, the region where all three replicas predict the lack of stable perovskites is denoted as the “no region.”

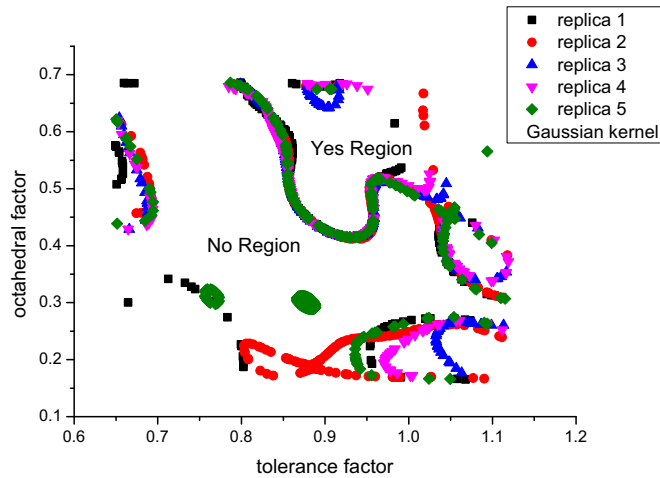


FIG. 8. The predicted formability of the cubic perovskite structure as provided by the Gaussian kernels of Eq. (2) for five different replicas. These different replicas are produced by randomly choosing  $R = 50$  fixed vectors in Eq. (2) (see text).

by our SRVM algorithm for, respectively, the multinomial and Gaussian kernels, respectively [see Sec. II and the discussion following Eq. (2) for a description of replicas in the Gaussian case]. The designations of “yes” and “no” reflect the predictions of the algorithm regarding the viability of putative compounds of an  $ABO_3$ -type composition to form stable perovskite structures. In Fig. 9, we overlay (with different levels of resolution in the two panels) the predictions of the SVM method and our SRVM algorithm with multiple kernels/replicas. The shaded region in Fig. 9 is the one in which all methods/replicas/functions predict that stable perovskite structures should form. With this region in hand, all candidate  $ABO_3$  materials (of the correct chemistry to allow such a composition) with tolerance and octahedral factors that lie in the shaded area are predicted to be stable perovskites. Some compositions lie near the boundary and do not enable (insofar as our approach is concerned) a definite prediction regarding perovskite structures that do not appear in the data set that we used for training and validation [18]. Two such candidates are  $\text{EuZrO}_3$  ( $v_1 = 0.857$ ,  $v_2 = 0.514$ ) and  $\text{EuHfO}_3$  ( $v_1 = 0.861$ ,  $v_2 = 0.507$ ). The location of these potential stable perovskite structures is highlighted in Fig. 9(b). With  $z = 0.8$ , the SVM algorithm achieved an accuracy of 92.52%. By contrast, the SRVM algorithm obtained an accuracy of 94.14% with a multinomial kernel [here two different multinomials (where different order multinomials were used as replicas) and we further employed the “expansion in the box” construction]; SRVM achieved an accuracy of 94.19% with a Gaussian kernel (here we employed 11 replicas, each having randomly chosen anchor points). These two candidates are indeed known to constitute stable perovskites [45,46].

#### IV. FORMABILITY OF BISMUTH-BASED DOUBLE PEROVSKITES USING TOLERANCE AND OCTAHEDRAL FACTORS

We next turn to predictions regarding the formability of more complex double perovskites (see, e.g., Fig. 2). Specifi-

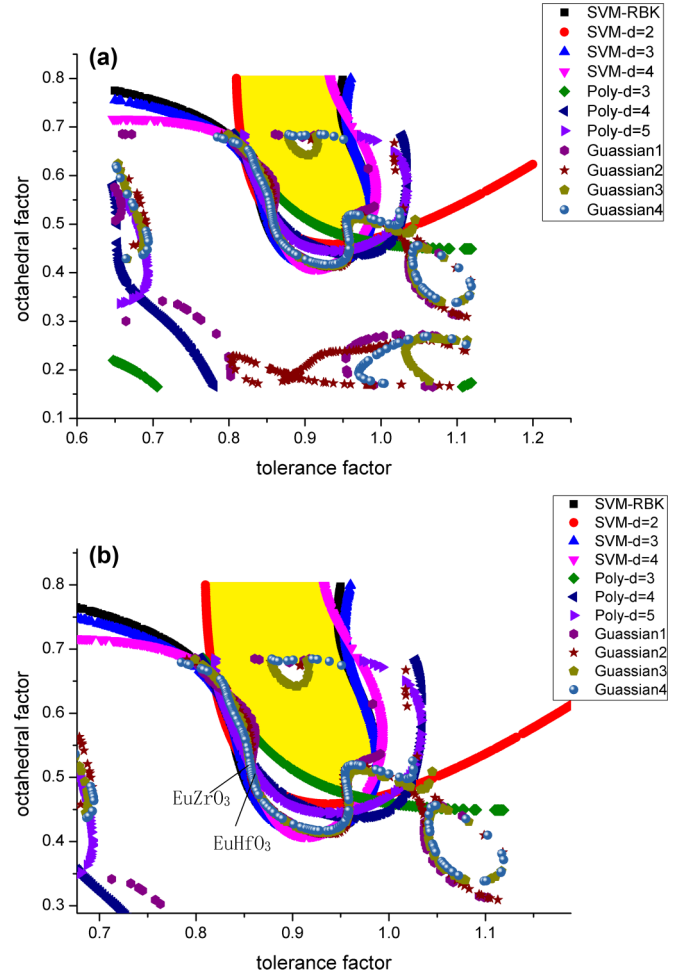


FIG. 9. The formability of cubic perovskite structure at two different resolutions. The yellow region is that in which all methods (SVM, SRVM with both multinomial and Gaussian kernels) predict that cubic perovskite structures will form. In (a), we show the entire region of measured tolerance and octahedral ratios. (b) Provides a zoomed view. Two possible candidate compounds from cubic perovskite structure are highlighted:  $\text{EuHfO}_3$  and  $\text{EuZrO}_3$ .

cally, we are interested in finding bismuth-based oxide double perovskites that can be described with a general formula of  $A'A''B'\text{BiO}_6$  with Bi occupying half of the  $B$ -site cations. Our interest in Bi-based oxide double perovskites is motivated by the goal to achieve stable and nontoxic alternatives for lead-halide perovskites that have recently emerged as high-performance semiconductors with applications in solar cells and light-emitting diodes, but are plagued with instability and toxicity issues [47–49].

For the double perovskites, we define the octahedral and tolerance factors by Eqs. (14) and (15):

$$r_A \equiv \frac{r_{A'} + r_{A''}}{2},$$

$$r_B \equiv \frac{r_{B'} + r_{B''}}{2}. \quad (16)$$

Toward this end, we will examine what occurs if the functions  $\{G_a\}_{a=1}^r$  are either a sum of randomly generated Gaussians or neural network type functions. We will examine both



(i) a data set involving only double perovskites as well as (ii) combined data for both cubic and double perovskites. The double perovskite data set that we first study (provided in Sec. I of the Supplemental Material [20]) consist of 72 candidate compositions. Of these, 57 are stable and 15 are unstable double perovskites. Similar to Sec. III, these data have only two features: the tolerance and octahedral factors  $v_{1,2}$ . These two factors are calculated from the Shannon radii  $r$  of the ions forming the double perovskite. The octahedral and tolerance factors of double perovskites are defined just as they were for the perovskites [Eqs. (14) and (15)]. Now, however, we replace the radius of the  $A$ -site ( $B$ -site) cation with the average radii of the  $A$ -site and  $A'$ -site ( $B$ -site and  $B'$ -site) cations involved.

### A. Neural net analysis of combined double perovskite and cubic perovskite data

We examine the 295 instances that include data on both (i) double perovskite data of Sec. I of the Supplemental Material (SM) [20] to this work, and (ii) the ternary perovskite formability data [18] (that we employed in Sec. III). Additional details regarding these data sets have been relegated to the SM [20]. We study these materials using the “neural net” library [50–54]. We randomized the input data and performed a 10-cross validation analysis for a neural net consisting of one layer and three nodes. We run an SRVM version of this neural net: the initial input weights (i.e., those set before back propagation is applied) are chosen randomly each time that we run the neural net model. Further details concerning this analysis are provided in Sec. II of the SM [20].

### B. SRVM–neural net predictions for new stable double perovskites

Armed with a proof of principle of the reliability of the ensemble of neural network functions for the combined formability data [18] and the double perovskite data (appearing in the Sec. A of the SM [20]) we now apply it to suggest hitherto unexplored candidate perovskite compounds. Predictions for the stability of the candidate double perovskite compounds are reported, rather expansively, in Table VI of the SM [20]. Herein, the stability probabilities for the various compounds are computed as equal to the ratio of (the number of replicas that yield a positive outcome)/(the total number of replicas employed).

## V. DOUBLE PEROVSKITE ANALYSIS WITH ELECTRONEGATIVITIES AND IONIC SHANNON RADII

We now turn to a complete double perovskite data set of the SM [20] and study these double perovskites alone (i.e., we do not include the ternary perovskite data in our analysis). In addition to the tolerance and octahedral factors of Eqs. (14) and (15), we will now add basic electronegativity features that directly touch on the interactions underlying the stability of these systems. Two of these new features are the electronegativity of the  $X$  (in the materials that we consider, oxygen anion) sites relative to average electronegativity  $n_A$

TABLE I. Space-group symmetries and octahedral tilt patterns for  $A'A''B'B''O_6$  compounds.

Index	Space group	Tilt system
2	$P\bar{1}$	$a^- b^- c^-$
4	$P2_1$	$a^- a^- c^+$
5	$C2$	$a^- b^0 c^+$
11	$P2_1/m$	$a^- a^- c^0$
13	$P2_1/c$	$a^+ b^- c^0$
16	$P222$	$a^- b^- c^-$
49	$Pccm$	$a^+ b'^0 c^0$
81	$P4$	$a^+ a^+ c^-$
85	$P4/n$	$a^0 a^0 c^-$
90	$P42_12$	$a^0 a^0 c^+$
111	$P42m$	$a^+ a^+ c^0$
129	$P4/nmm$	$a^0 a^0 c^0$

and  $n_B$  of, respectively, the ions at the  $A$  and  $B$  sites:

$$v_3 = n_X - \frac{n_{A'} + n_{A''}}{2} \equiv n_X - n_A,$$

$$v_4 = n_X - \frac{n_{B'} + n_{B''}}{2} \equiv n_X - n_B. \quad (17)$$

In practice, in order to compute  $n_{A,B}$  we need to know the four electronegativities  $n_{A'}$ ,  $n_{A''}$ ,  $n_{B'}$ , and  $n_{B''}$ . We will employ the two features of Eq. (17) since these may be extended also to the ternary perovskites  $ABO_3$ . Since the  $B-O$  bond is covalent while the  $A-O$  interactions are more ionic, we anticipate  $v_3$  to be more important than  $v_4$ . A new feature suggested by [55] (dubbed therein as  $\tau$ ) which we included was a nontrivial function of the average electronegativity of the  $A$  ions and the average Shannon radii  $r_{A,B}$  [or, equivalently, the tolerance and octahedral factors of Eqs. (14) and (15)] with the substitution of Eq. (16):

$$v_5 = \tau \equiv \frac{r_X}{r_B} - n_A \left( n_A - \frac{r_A/r_B}{\ln(r_A/r_B)} \right). \quad (18)$$

The input features of known stable and unstable double perovskites and our results are, respectively, provided in Tables I and II of the SM [20]. Our corresponding predictions are

TABLE II. Space-group symmetries and octahedral tilt patterns for  $A_2B'B''O_6$  compounds.

Index	Space group	Tilt system
2	$P\bar{1}$	$a^- b^- c^-$
12	$C2/m$	$a^0 b^- b^-$
14	$P2_1/c$	$a^+ b^- b^-$
15	$C2/2$	$a^0 b^- c^-$
48	$Pnnn$	$a^+ b^+ c^+$
86	$P4_2/n$	$a^+ a^+ c^-$
87	$I4/m$	$a^0 a^0 c^-$
128	$P4/mnc$	$a^+ a^+ c^-$
134	$P4_2/nmm$	$a^0 b^- b^+$
148	$R\bar{3}$	$a^- a^- a^-$
201	$Pn\bar{3}$	$a^- a^+ a^+$
225	$Fm\bar{3}m$	$a^0 a^0 a^0$

given in Table VII of the SM. In the SM (Sec. VI) [20], we further provide specifics concerning neural net functions used to fit the data. We return to a simple Gaussian SRVM of Secs. II A and III and apply it to an analysis of the double perovskite data with the above-mentioned five features  $v_{1 \leq i \leq 5}$ . A vote within sets of  $r = 31$  replicas was performed 10 times to yield final predictions. (In Tables X–XVIII of the SM [20], we display  $r = 9$  such constructed replica functions.) Each of the replica functions is of the form of Eq. (2) with  $R = 17$  anchor points. The centers  $\{\bar{v}_{ja}\}$  of the Gaussians and the corresponding coefficients  $\{c_{ja}\}_{j=1}^R$  are listed in Tables XX–XVIII of the SM [20]. All standard deviations  $\sigma_{ja}$  are set to unity,  $\sigma_{ja} = \sigma = 1$ . Unlike the analysis in Sec. III, we have not employed the overlap to optimize the number of anchor points. The resulting fivefold cross-validated accuracy for the double perovskite data of Sec. I of the SM with the  $d = 5$  features using the SRVM-Gaussian algorithm is 92%. In Table I of the SM [20], we provide predictions for the stability of the screened candidate compounds. In Table VII of the SM [20], we also present the results of SVM with radial basis kernels applied and neural networks of different architecture. The fivefold cross-validated accuracy for this model is, similarly, also 92%.

## VI. THEORETICAL FORMATION ENTHALPY OF SELECTED DOUBLE PEROVSKITES

As described in Secs. IV and V, we employed SRVM to predict the formability of hypothetical Bi-based double perovskite oxides with a general formula of  $A'A''B'\text{BiO}_6$ , where Bi occupies half of the  $B$ -site cations. From a simple charge balance, 30 357 total double perovskite oxides of the form  $A'A''B'\text{BiO}_6$  are possible. We considered all the cations up to Bi in the periodic table for  $A'$ ,  $A''$ , and  $B'$  sites, while excluding lanthanides and radioactive Technetium as possible candidates. Moreover, all possible oxidation states for each cation were included during charge balance while keeping the oxidation state of Bi fixed at +3.

Due to the limited number of experimentally synthesized Bi-based double perovskite oxides, we use a twofold approach for training our SRVM/SVM models. First, we devised a quick screening criterion to screen some hypothetical compounds based on the atomic features of the 28 unique  $A'A''B'\text{BiO}_6$  double perovskites (with Bi at the  $B$  site) that have been reported in inorganic crystal structure database (ICSD) [56]. The atomic features that we used for initial screening are the tolerance and octahedral factors of Eqs. (14) and (15). We used Slater's empirical atomic radii [57] for calculating these factors. This was done since the tolerance and octahedral factors are independent of oxidation state and coordination number, which are unknown for a hypothetical perovskite. We found that these factors follow a linear regression model with an  $R$ -squared value of 0.84, for the experimental  $A'A''B'\text{BiO}_6$  compounds. Moreover, we find that 144 hypothetical  $A'A''B'\text{BiO}_6$  compounds adhere to this linear regression model.

Next, by using density-functional theory (DFT), we optimized the crystal structure of each screened compound to calculate  $\Delta H_f$  and evaluate its potential formability. The perovskite framework undergoes cooperative tilting of the

$\text{BX}_6$  octahedra to optimize the coordination environment of  $A$ - and  $B$ -site cations [58]. We consider all viable octahedral tilt patterns for investigating the ground state of a given stoichiometry. These possible tilt patterns and their corresponding space-group symmetries are summarized in Tables I and II [59–61]. For a stoichiometry to be considered stable, its calculated  $\Delta H_f$  should be negative. However, compounds with positive  $\Delta H_f$  are quite common and can be experimentally synthesized by optimizing experimental conditions. For example, a recent survey of the DFT-calculated  $\Delta H_f$  of all existing binary oxides reported  $\sim 90$ th percentile of the compounds lie within 94 meV/atom above the ground-state polymorph [62]. Therefore, we expect compounds with  $\Delta H_f < 100$  meV/atom to be formable under suitable experimental conditions. We label each of the 144 compounds as stable ( $\Delta H_f < 100$  meV/atom) and unstable ( $\Delta H_f > 100$  meV/atom).

We used experimentally synthesized Bi-based double perovskite oxides (with Bi at the  $B$  site), nonperovskite oxides, and the compounds labeled as stable and unstable from initial screening to train our SVM models. Here, non-double-perovskite oxide corresponds to a compound following the stoichiometry of a double perovskite but adopts a nonperovskite crystal structure, listed in the ICSD [56]. As described in Sec. V, we used five different SVM models followed by a voting procedure (SRVM) to predict the formability of a hypothetical Bi–double perovskite oxide. The models employed in this classification process use the following atomic features: the tolerance factor, the octahedral factor, the average  $A$ -site electronegativity, and the average  $B$ -site electronegativity. We calculate the octahedral and tolerance factors using Slater's empirical atomic radii [57] as well as Shannon's ionic radii [63]. We consider a hypothetical Bi–double perovskite oxide to be formable if at least one of the five SVMs predicts it to be stable.

Out of a total 30 357 hypothetical  $A'A''B'\text{BiO}_6$  compounds, our SRVM algorithm voted 9795 compounds to be stable, where 8115 of the 9795 compounds are predicted to be stable by all the five SVM models. We found 1680 out of 9795 compounds are voted as stable by at least one SVM model. To further assess the formability of compounds that were voted by our SRVM algorithm to be stable, we selected a few compounds to perform structural optimization and subsequent formation enthalpy analysis using DFT. At this point, it is not possible to optimize all the 9795 compounds using DFT due to their computational expense. We optimized the crystal structure of the selected compounds using DFT and calculate their  $\Delta H_f$ . As mentioned earlier, the perovskite framework can undergo cooperative tilting of the  $\text{BO}_6$  octahedra to optimize the coordination environment of  $A$ - and  $B$ -site cations [58]. We consider all possible octahedral tilt patterns for investigating the ground state of a given stoichiometry. The DFT calculations were performed using the projector-augmented-wave (PAW) method [64] as implemented in the Vienna *ab initio* simulation package (VASP) [65]. We employed generalized gradient approximation (GGA) as implemented in the Perdew-Burke-Ernzerhof (PBE) functional [66] structure optimization. Layered and rock-salt ordering were imposed at the  $A$  and  $B$  sites, respectively. A plane-wave basis set with a cutoff of 400 eV was used for structure relaxation, and 520 eV

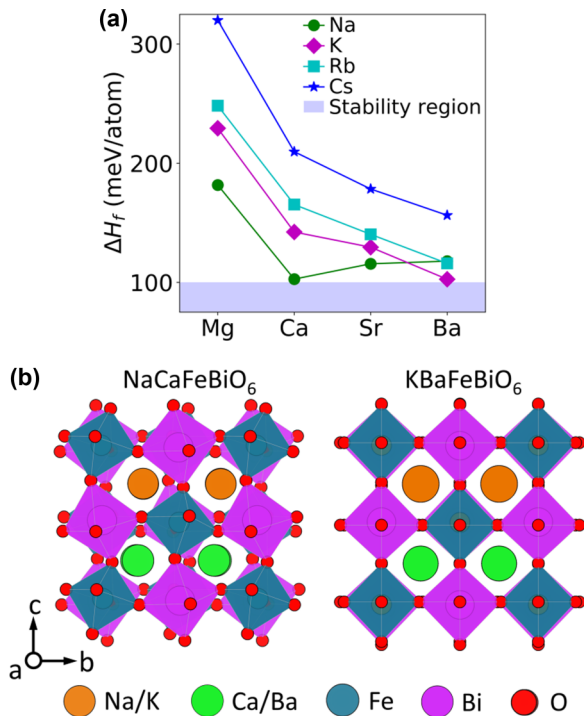


FIG. 10. (a) Formation enthalpy for the ground-state structure of the  $A'A''\text{FeBiO}_6$  family of compounds. The X axis represents the  $A''$ -site cation whereas  $A'$ -site cations are defined in the legend. (b) Schematics showing the [100] projections of ground-state structures of  $\text{NaCaFeBiO}_6$  and  $\text{KBaFeBiO}_6$ .

for the final static total energy calculation step. Gamma-centered Monkhorst-Pack [67]  $k$ -points mesh was used for sampling the Brillouin zone, where  $k$  points per reciprocal atom (KPPRA) were set to be  $\sim 8000$  for relaxation and the single-step static calculation. A Hubbard  $U$  parameter of 4 eV was used to account for localized  $d$ -electron interactions in Fe [68].  $G$ -type antiferromagnetism was imposed during relaxation. The DFT calculations were carried out using the pseudopotentials and other DFT settings employed by the Open Quantum Materials Database (OQMD) [69].

The calculated formation enthalpy for each compound is the difference between the total energy of a given structure with the minimized free energy of chemical reactants, which lie at the convex hull, for the given stoichiometry as implemented within OQMD [69,70]. For example, Eq. (19) shows the chemical reaction path for which the free energy is minimized for the reactants for  $\text{KSrFeBiO}_6$ . The formation enthalpy for  $\text{KSrFeBiO}_6$  [ $\Delta H_f(\text{KBaTeBiO}_6)$ ] is then calculated using Eq. (20), where  $E(\text{KSrFeBiO}_6)$  is the DFT total energy/formula unit of  $\text{KSrFeBiO}_6$ , while  $E(\text{SrFeO}_3)$  and  $E(\text{KBiO}_3)$  are DFT total energies/formula unit of  $\text{SrFeO}_3$  and  $\text{KBiO}_3$  respectively:



$$\Delta H_f(\text{KSrFeSiO}_6) = E(\text{KSrFeBiO}_6) - [E(\text{SrFeO}_3) - E(\text{KBiO}_3)]. \quad (20)$$

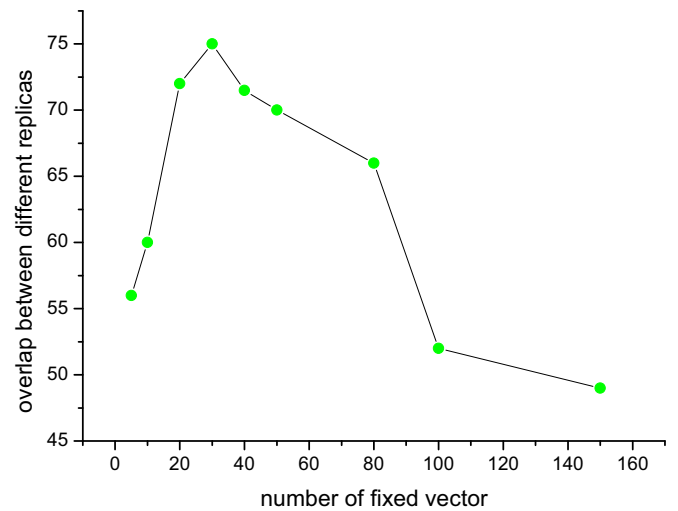


FIG. 11. The average overlap (given as a percentage) as computed by Eq. (22) averaged over the different cross validation between the  $r = 7$  different replicas. Here, replicas of the Gaussian replica model [Eq. (2)] were employed for the ternary classification of the binary ( $AB$ ) solids as a function of fixed anchor points  $R$ . The maximum overlap (suggesting an accurate prediction) appears circa  $R \approx 40$ .

We calculated  $\Delta H_f$  of compounds with a general formula of  $A'A''\text{FeBiO}_6$  using DFT, where  $A' = \text{Na, K, Rb, and Cs}$  and  $A'' = \text{Mg, Ca, Sr, and Ba}$ . This family of compounds was voted to be stable by the SRVM algorithm. We exclude the  $\text{Li}^+$  and  $\text{Be}^{2+}$  cations as possible  $A$ -site candidates due to their extremely small size for the  $A$ -site cubo-octahedral cavities. Of the 16 possible compounds in the  $A'A''\text{FeBiO}_6$  family, 13 compounds are predicted to be stable double perovskites by all the five SVM models, while for the other three compounds ( $\text{NaMgFeBiO}_6$ ,  $\text{KMgFeBiO}_6$ , and  $\text{RbMgFeBiO}_6$ ) SVM-2 and SVM-5 predict them to be unstable with an overall vote predicting them to be stable. From an analysis of the formation enthalpy, we find that the  $A'A''\text{FeBiO}_6$  family of double perovskite oxides is reasonably close to our formability criterion ( $\Delta H_f < 100$  meV/atom). As shown in Fig. 10, we find that the ground state of all the 16 compounds has  $\Delta H_f > 100$  meV/atom. However, 8 of the 16 compounds have  $\Delta H_f$  within (100 to 150) meV/atom. Figure 10(b) shows the [100] projections of ground-state crystal structures of  $\text{NaCaFeBiO}_6$  and  $\text{KBaFeBiO}_6$ .  $\text{NaCaFeBiO}_6$  is predicted to have ground-state crystal structure belonging to  $P2_1$  space group with  $a^-a^-c^+$  tilt pattern, whereas the ground state of  $\text{KBaFeBiO}_6$  is predicted to be the ideal double perovskite structure ( $P4/nmm$ ) without any octahedral tilts. The compounds having Mg at the  $A''$  site exhibit the highest  $\Delta H_f$ . The average  $\Delta H_f$  for compounds having Mg at the  $A''$  site is 245 meV/atom. The small size of Mg makes it unsuitable for the large  $A$ -site cavities, even after octahedral tilts, resulting in a high  $\Delta H_f$ . Also, smaller size at the  $A$  site results in higher degree of octahedral tilting. For instance, the ground state of  $\text{RbMgFeBiO}_6$  has a space-group symmetry of  $P\bar{1}$ , which corresponds to a tilting pattern  $a^-b^-c^-$ , resulting in octahedral tilting along all three crystallographic directions. Despite having  $100 < \Delta H_f < 150$  meV/atom,  $A'A''\text{FeBiO}_6$

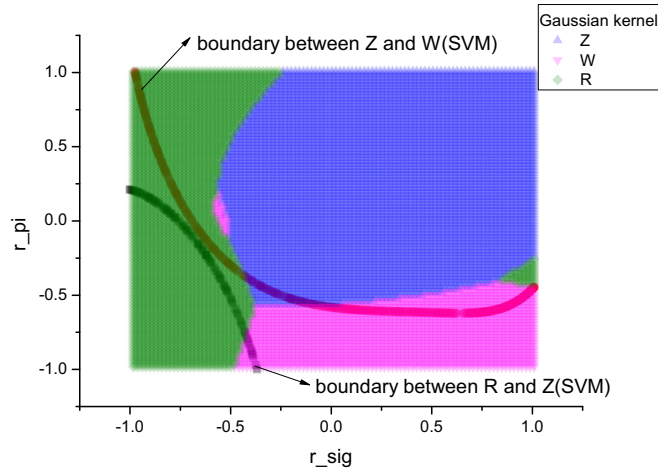


FIG. 12. The phase diagram predicted by both the SVM algorithm (denoted by thick solid lines) and our method when using Gaussian kernels (different color solid regions).

shows a good potential toward thermodynamic stability and we expect some of the compounds within this family can be synthesized by optimizing experimental conditions. We remark that while we achieve very high cross-validation accuracies (the 92% cross-validation accuracy of Sec. V), the success of our method in predicting stable double perovskite compounds (in as much as these can be ascertained via DFT calculations) is far lower.

## VII. TERNARY CLASSIFICATION OF AB SOLIDS

We next turn our attention, using the data of [71], to the classification of binary octet solids [26] (these solids have the chemical composition  $A^n B^{8-n}$  where  $n$  denotes the number of valence electrons) into one of  $q = 3$  groups (denoted W, Z, or R [26,71]). Octet binary alloy crystals include technologically important semiconductors such as GaAs, GaN, and ZnO. The three classes that we analyze correspond to dominant zincblende (Z), wurtzite (W), or rock-salt (R) crystal structures that these alloys typically form. Similar to Sec. III, we applied both the standard SVM technique with our multiclass variant SRVM approach (see Sec. II D) with multinomial and Gaussian kernels. We employed two commonly used figures of merit [72] as features:

$$\begin{aligned} r_\sigma &\equiv r_s^A + r_p^A - r_s^B - r_p^B, \\ r_\pi &\equiv r_p^A - r_s^A + r_p^B - r_s^B. \end{aligned} \quad (21)$$

Here,  $r_s^A$ ,  $r_p^A$ ,  $r_s^B$ , and  $r_p^B$  denote the pertinent radii for an electron bound to the A or B ion that is in an s or p orbital.

In the Gaussian approach, we employed the sum of  $R = 30$  individual Gaussians (associated with different anchor

points). The general behavior of the inter-replica overlap is displayed in Fig. 11 in which it is seen that the overlap becomes maximal at  $R \sim 30$ . The final classification for each data point was determined by the group for which a given data point appeared most frequently out of the five replicas employed. Let  $n_w^{\vec{v}'}$  be the number of replicas that classify point  $\vec{v}'$  into one of the three classes  $w = 0, 1, 2$ . For a given cross-validation run, the average replica overlap for this three-classification problem is defined as

$$\bar{O} \equiv \frac{1}{rN_{\vec{v}'}} \sum_{\vec{v}'} \max_w \{n_w^{\vec{v}'}\}. \quad (22)$$

Here,  $N_{\vec{v}'}$  is the number of test points  $\vec{v}'$ . The average of  $\bar{O}$  over the different cross-validation partition is plotted in Fig. 11 as a percentage. The cross-validation accuracies [as ascertained by Eq. (7) for  $z = 0.8$ ] that our SRVM algorithm obtained for the Gaussian and multinomial kernels were, respectively, 92.72% and 90.90%. These values were lower than the accuracy achieved by an SVM algorithm with a radial kernel (that we found to be 94.54%). In Fig. 12, we provide the phase boundaries (between the W, Z, and R phases) as ascertained by SVM (see the solid curves therein) alongside the boundaries determined by our SRVM method (the domains of the different phases as predicted by SRVM are marked by different colors).

## VIII. CONCLUSION

In summary, we introduced and implemented a classification algorithm to classify various materials and identify promising compounds. In particular, we investigated (1) the formability of cubic and double perovskite type compounds (a binary classification problem) and (2) binary octet alloys (via ternary classification). A more detailed description of our algorithm appears in a companion paper [21]. Using this algorithm, we achieved a high accuracy in both problems. Combining our approach with other machine techniques, we suggest candidate stable perovskites and properties of binary compounds.

*Note added.* Recently, we became aware of a recent work by Park *et al.* [73] that explores, along very different lines, the applicability of machine learning for double perovskites.

## ACKNOWLEDGMENTS

This research was partially supported by the NSF CMMT under Grant No. 1411229. R.M. and A.S.T. acknowledge support from NSF Ceramics program through Grant No. 1806147. Computations in this work benefited from the use of the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by NSF Grants No. ACI-1053575 and No. ACI-1548562.

T.M., B.S., and A.S.T. contributed equally to this work.

[1] T. Mueller, A. G. Kusne, and R. Ramprasad, Machine learning in materials science, in *Reviews in Computational Chemistry*, Vol. 29, edited by A. L. Parrill and K. B. Lipkowitz (Wiley, Hoboken, NJ, 2016), Chap. 4.

[2] *Information Science for Materials Discovery and Design*, Springer Series Materials, Vol. 225, edited by T. Lookman, F. Alexander, and K. Rajan (Springer, Berlin, 2016).

- [3] P. Ronhovde, S. Chakrabarty, M. Sahu, K. F. Kelton, N. A. Mauro, K. K. Sahu, and Z. Nussinov, Detecting hidden spatial and spatio-temporal structures in glasses and complex physical systems by multiresolution network clustering, *Eur. Phys. J. E* **34**, 105 (2011).
- [4] P. V. Balachandran, S. R. Broderick, and K. Rajan, Identifying the 'inorganic gene' for high-temperature piezoelectric perovskites through statistical learning, *Proc. R. Soc. A* **467**, 2271 (2011).
- [5] P. Ronhovde, S. Chakrabarty, D. Hu, M. Sahu, K. K. Sahu, K. F. Kelton, N. A. Mauro, and Z. Nussinov, Detection of hidden structures for arbitrary scales in complex physical systems, *Sci. Rep.* **2**, 329 (2012).
- [6] J. C. Snyder, M. Rupp, K. Hansen, K. R. Muller, and K. Burke, Finding Density Functionals with Machine Learning, *Phys. Rev. Lett.* **108**, 253002 (2012).
- [7] Y. Zhang and Eun-Ah Kim, Quantum Loop Topography for Machine Learning, *Phys. Rev. Lett.* **118**, 216401 (2017).
- [8] J. Carrasquilla and R. G. Melko, Machine learning phases of matter, *Nat. Phys.* **13**, 431 (2017).
- [9] G. Carleo and M. Troyer, Solving the quantum many-body problem with artificial neural networks, *Science* **355**, 602 (2017).
- [10] M. Koch-Janusz and Z. Ringel, Mutual information, neural networks and the renormalization group, *Nat. Phys.* **14**, 578 (2018).
- [11] G. Pilia, A. Mannodi-Kanakkithodi, B. P. Uberuaga, R. Ramprasad, J. E. Gubernatis, and T. Lookman, Machine learning bandgaps of double perovskites, *Sci. Rep.* **6**, 19375 (2016).
- [12] A. Mannodi-Kanakkithodi, G. Pilia, T. D. Huan, T. Lookman, and R. Ramprasad, Machine learning strategy for accelerated design of polymer dielectrics, *Sci. Rep.* **6**, 20952 (2016).
- [13] O. Muller and R. Roy, *The Major Ternary Structural Families* (Springer, New York, 1974).
- [14] A. Navrotsky, Energetics and crystal chemical systematics among ilmenite, lithium niobate, and perovskite structures, *Chem. Mater.* **10**, 2787 (1998).
- [15] N.-G. Park, Perovskite solar cells: An emerging photovoltaic technology, *Mater. Today* **18**, 65 (2015).
- [16] F. S. Galasso, *Perovskites and High Tc Superconductors* (Gordon and Breach, New York, 1990).
- [17] Z. L. Wang and Z. C. Kang, *Functional and Smart Materials: Structural Evolution and Structure Analysis* (Plenum, New York, 1998).
- [18] L. M. Feng, L. Q. Jiang, M. Zhu, H. B. Liu, X. Zhou, and C. H. Li, Formability of ABO<sub>3</sub> cubic perovskites, *J. Phys. Chem. Solids* **69**, 967 (2008).
- [19] A. Khan and S. G. Javed, Predicting regularities in lattice constants of GdFeO<sub>3</sub>-type perovskites, *Acta Crystallogr. Sect. B* **64**, 120 (2008).
- [20] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevMaterials.3.063802> for (1) screened double perovskite data, (2) two-feature neural network analysis of combined single and double perovskite data, (3) probability for candidate double perovskites based on the tolerance and octahedral factor, (4) five-feature empirical data and our predictions for stable double perovskite materials, (5) neural network for five-feature analysis of double perovskites, and (6) specific SRVM-Gaussian functions used for the five predictions of stable double perovskites.
- [21] P. Chao, T. Mazaheri, B. Sun, N. B. Weingartner, and Z. Nussinov, The stochastic replica approach to machine learning: Stability and parameter optimization, [arXiv:1708.05715](https://arxiv.org/abs/1708.05715).
- [22] S. G. Javed, A. Khan, A. Majid, A. M. Mirza, and J. Bashir, Lattice constant prediction of orthorhombic ABO<sub>3</sub> perovskites using support vector machines, *Comput. Mater. Sci.* **39**, 627 (2007).
- [23] G. Hautier, C. C. Fischer, A. Jain, T. Mueller, and G. Ceder, Finding nature's missing ternary oxide compounds using machine learning and density functional theory, *Chem. Mater.* **22**, 3762 (2010).
- [24] G. Pilia, P. V. Balachandran, J. E. Gubernatis, and T. Lookman, Classification of ABO<sub>3</sub> perovskite solids: A machine learning study, *Acta Crystallogr. Sect. B* **71**, 507 (2015).
- [25] R. Rifkin and A. Klautau, In defense of one-vs-all classification, *J. Machine Learning Res.* **5**, 101 (2004).
- [26] P. V. Balachandran, J. Theiler, J. M. Rondinelli, and T. Lookman, Materials prediction via classification learning, *Sci. Rep.* **5**, 13285 (2015).
- [27] S. Wiseman, M. Blatt, and E. Domany, Superparamagnetic clustering of data, *Phys. Rev. E* **57**, 3767 (1998).
- [28] P. Ronhovde and Z. Nussinov, An improved potts model applied to community detection, *Phys. Rev. E* **81**, 046114 (2010).
- [29] P. Ronhovde and Z. Nussinov, Multiresolution community detection for megascale networks by information-based replica correlations, *Phys. Rev. E* **80**, 016109 (2009).
- [30] D. Hu, P. Ronhovde, and Z. Nussinov, A replica inference approach to unsupervised multiscale image segmentation, *Phys. Rev. E* **85**, 016101 (2012).
- [31] V. Vapnik, *The Nature of Statistical Learning Theory*, 2nd ed. (Springer, New York, 1999).
- [32] J. A. K. Suykens and J. Vandewalle, Least squares support vector machine classifiers, *Neural Process. Lett.* **9**, 293 (1999).
- [33] J. Hopfield, Neural networks and physical systems with emergent collective computational abilities, *Proc. Natl. Acad. Sci. USA* **79**, 2554 (1982).
- [34] C. M. Bishop, *Neural Networks for Pattern Recognition* (Clarendon, Oxford, 1995).
- [35] G. E. Hinton, S. Osindero, and Y. Teh, A fast learning algorithm for deep belief nets, *Neural Comput.* **18**, 1527 (2006).
- [36] L. Breiman, Bagging predictors, *Machine Learning* **24**, 123 (1996).
- [37] E. R. Schapire, The strength of weak learnability, *Machine Learning* **5**, 197 (1990).
- [38] L. Rokach and O. Maimon, *Data Mining with Decision Trees: Theory and Applications* (World Scientific, Singapore, 2008).
- [39] Tin Kam Ho, Random decision forests, in *Proceedings of 3rd International Conference on Document Analysis and Recognition* (IEEE, Piscataway, NJ, 1995), Vol. 1, p. 278.
- [40] A trivial extension of these functions is one in which the Cartesian coordinate axes are rotated and dilated (allowing for a sum of general multivariate Gaussian distributions with a nondiagonal kernel). For simplicity, in this work, we will contend ourselves with the simple diagonal covariant form of Eq. (2).
- [41] <https://www.csie.ntu.edu.tw/~cjlin/libsvm/>.
- [42] Y. LeCun, Y. Bengio, and G. Hinton, Deep learning, *Nature* **521**, 436 (2015).

- [43] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning* (MIT Press, Cambridge, MA, 2016).
- [44] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, Learning representations by back-propagating errors, *Nature* **323**, 533 (1986).
- [45] V. Viallet, J.-F. Marucco, J. Saint, M. Herbst-Ghysel, and N. Dragoë, Structural, magnetic and electrical properties of a perovskite containing divalent europium  $\text{EuZrO}_3$ , *J. Alloys Compd.* **461**, 346 (2008).
- [46] P. V. Balachandran, A. A. Emery, J. E. Gubernatis, T. Lookman, C. Wolverton, and A. Zunger, Predictions of new  $\text{ABO}_3$  perovskite compounds by combining machine learning, and density functional theory, *Phys. Rev. Mater.* **2**, 043802 (2018).
- [47] C. C. Stoumpos and M. G. Kanatzidis, Halide perovskites: Poor man's high-performance semiconductors, *Adv. Mater.* **28**, 5778 (2016).
- [48] W.-J. Yin, B. Weng, J. Ge, Q. Sun, Z. Li, and Y. Yan, Oxide perovskites, double perovskites and derivatives for electrocatalysis, photocatalysis, and photovoltaics, *Energy Environ. Sci.* **12**, 442 (2019).
- [49] A. S. Thind, X. Huang, J. W. Sun, and R. Mishra, First-principles prediction of a stable hexagonal phase of  $\text{CH}_3\text{NH}_3\text{PbI}_3$ , *Chem. Mater.* **29**, 6003 (2017).
- [50] <https://www.rdocumentation.org/packages/neuralnet/versions/1.33/topics/neuralnet>.
- [51] M. Riedmiller, Rprop: Description and Implementation Details, Technical report, University of Karlsruhe, 1994 (unpublished).
- [52] M. Riedmiller and H. Braun, A direct adaptive method for faster backpropagation learning: The RPROP algorithm, in *Proceedings of the IEEE International Conference on Neural Networks (ICNN), San Francisco, 1993* (IEEE, Piscataway, NJ, 1993), pp. 586–591.
- [53] A. Anastasiadis *et al.*, New globally convergent training scheme based on the resilient propagation algorithm, *Neurocomputing* **64**, 253 (2005).
- [54] O. Intrator and N. Intrator, Using neural nets for interpretation of nonlinear models, in *Proceedings of the Statistical Computing Section, San Francisco, 1993* (American Statistical Association, Alexandria, VA, 1993), pp. 244–249.
- [55] C. J. Bartel, C. Sutton, B. R. Goldsmith, R. Ouyang, C. B. Musgrave, L. M. Ghiringhelli, and M. Scheffler, New tolerance factor to predict the stability of perovskite oxides and halides, *Sci. Adv.* **5**, eaav0693 (2019).
- [56] Inorganic Crystal Structure Database, <https://icsd.fiz-karlsruhe.de>.
- [57] J. C. Slater, Atomic radii in crystals, *J. Chem. Phys.* **41**, 3199 (1964).
- [58] P. M. Woodward, Octahedral tilting in perovskites. I. Geometrical considerations, *Acta Crystallogr. Sect. B* **53**, 32 (1997).
- [59] M. C. Knapp, and P. M. Woodward, A-site cation ordering in  $\text{AA}'\text{BB}'\text{O}_6$  perovskites, *J. Solid State Chem.* **179**, 1076 (2006).
- [60] C. J. Howard and H. T. Stokes, Structures and phase transitions in perovskites - a group-theoretical approach, *Acta Crystallogr., Sect. A* **61**, 93 (2005).
- [61] C. J. Howard, B. J. Kennedy, and P. M. Woodward, Ordered double perovskites: A group-theoretical analysis, *Acta Crystallogr., Sect. B* **59**, 463 (2003).
- [62] W. Sun *et al.*, The thermodynamic scale of inorganic crystalline metastability, *Sci. Adv.* **2**, e1600225 (2016).
- [63] R. T. Shannon and C. T. Prewitt, Effective ionic radii in oxides and fluorides, *Acta Crystallogr., Sect. B* **25**, 925 (1969).
- [64] P. E. Blochl, Projector augmented-wave method, *Phys. Rev. B* **50**, 17953 (1994).
- [65] G. Kresse and J. Furthmüller, Efficient iterative schemes for *ab initio* total-energy calculations using a plane-wave basis set, *Phys. Rev. B* **54**, 11169 (1996).
- [66] J. P. Perdew, K. Burke, and M. Ernzerhof, Generalized Gradient Approximation Made Simple, *Phys. Rev. Lett.* **77**, 3865 (1996).
- [67] H. J. Monkhorst and J. D. Pack, Special points for Brillouin-zone integrations, *Phys. Rev. B* **13**, 5188 (1976).
- [68] S. L. Dudarev, G. A. Botton, S. Y. Savrasov, C. J. Humphreys, and A. P. Sutton, Electron-energy-loss spectra and the structural stability of nickel oxide: An LSDA+ U study, *Phys. Rev. B* **57**, 1505 (1998).
- [69] S. Kirklin *et al.*, The Open Quantum Materials Database (OQMD): Assessing the accuracy of DFT formation energies, *NPJ Comput. Mater.* **1**, 15010 (2015).
- [70] J. E. Saal, S. Kirklin, M. Aykol, B. Meredig, and C. Wolverton, Materials design and discovery with high-throughput density functional theory: The open quantum materials database (OQMD), *JOM* **65**, 1501 (2013).
- [71] Y. Saad, D. Gao, T. Ngo, S. Bobbitt, J. R. Chelikowsky, and W. Andreoni, Data mining for materials: Computational experiments with  $\text{AB}$  compounds, *Phys. Rev. B* **85**, 104104 (2012).
- [72] J. R. Chelikowsky and J. C. Phillips, Quantum-defect theory of heats of formation and structural transition energies of liquid and solid simple metal alloys and compounds, *Phys. Rev. B* **17**, 2453 (1978).
- [73] H. Park, R. Mall, F. H. Alharbi, S. Sanvito, N. Tabet, H. Bensmail, and F. El-Mellouhi, Exploring new approaches towards the formability of mixed-ion perovskite by DFT and Machine Learning, *Phys. Chem. Chem. Phys.* **21**, 1078 (2019).