

# Combinatorial screening for new materials in unconstrained composition space with machine learning

B. Meredig,<sup>1,\*</sup> A. Agrawal,<sup>2,\*</sup> S. Kirklin,<sup>1</sup> J. E. Saal,<sup>1</sup> J. W. Doak,<sup>1</sup> A. Thompson,<sup>1</sup> K. Zhang,<sup>2</sup>  
A. Choudhary,<sup>2</sup> and C. Wolverton<sup>1</sup>

<sup>1</sup>*Department of Materials Science and Engineering, Northwestern University, Evanston, Illinois 60208, USA*

<sup>2</sup>*Department of Electrical Engineering and Computer Science, Northwestern University, Evanston, Illinois 60208, USA*

(Received 2 November 2013; published 14 March 2014)

Typically, computational screens for new materials sharply constrain the compositional search space, structural search space, or both, for the sake of tractability. To lift these constraints, we construct a machine learning model from a database of thousands of density functional theory (DFT) calculations. The resulting model can predict the thermodynamic stability of arbitrary compositions without any other input and with six orders of magnitude less computer time than DFT. We use this model to scan roughly 1.6 million candidate compositions for novel ternary compounds ( $A_x B_y C_z$ ), and predict 4500 new stable materials. Our method can be readily applied to other descriptors of interest to accelerate domain-specific materials discovery.

DOI: [10.1103/PhysRevB.89.094104](https://doi.org/10.1103/PhysRevB.89.094104)

PACS number(s): 71.15.Dx, 65.40.-b, 71.15.Mb, 81.05.-t

## I. INTRODUCTION

Rational, data-driven materials discovery would be an immense boon for research and development, making these efforts far faster and cheaper. In such a paradigm, [1,2] computer models trained to find patterns in massive chemical datasets would rapidly scan compositions and systematically identify attractive candidates for technological applications, such as new batteries, lightweight alloys, solar cells, and so on. Indeed, machine learning models in particular have been used to address a number of important physics problems, including crystal structure prediction (CSP) [3], molecular atomization energy prediction [4], and approximation of density functionals [5]. Nonetheless, routine computational materials discovery remains an elusive goal. Why?

The two greatest barriers to facile computational materials discovery with a workhorse quantum mechanical tool such as density functional theory (DFT) [6,7] are (a) missing atomistic structural information (a necessary DFT input) and (b) computer resource limitations. As a result, notable high-throughput discovery efforts to date have had to constrain either the structural search space, compositional search space, or both [8–13]. To address the structure issue, much recent effort in the computational materials field has focused on CSP [14–17]. CSP traditionally assumes a target composition, or binary composition range [18], and then performs a global optimization of crystal unit cell parameters with respect to energy until the algorithm locates a suitable energy-minimizing configuration. While CSP is quite successful for many individual materials of interest [19,20], it is computationally costly. This point represents the second barrier to computational materials discovery: thoroughly searching through the combinatorial explosion of ternary or higher-order compositions with DFT and CSP is entirely infeasible, and will likely remain so well into the future.

In the present work, we describe an extremely fast approach for identifying *compositions* with desired materials properties, while removing structure- and resource-based constraints on

computational materials screening. Our approach relies on training both heuristic and machine learning (ML) models on a large database of quantum mechanical calculations on known materials, and then combining these two models into a unified prediction method. We describe in more detail the specific algorithms we employed in the Supplemental Material [21]. Our combined heuristic-ML framework evaluates a single composition in roughly a millisecond on today's hardware, may be trivially parallelized across the search space, and does not use crystal structure as input (though, of course, it has been initially trained on several thousand DFT calculations of ground-state crystal structures); hence, it is able to scan millions to even trillions of candidate compositions in reasonable time. We use as a case study the prediction of undiscovered stable ternary compounds, and produce a ranked list of about 4500 compositions that are likely to represent undiscovered materials.

Mining information from “big data” has recently garnered increased interest in both the popular [22] and scientific literature [23,24], as society grapples with information overload: Our ability to generate new data has far outstripped our ability to make sense of it. Science, and in particular computational science, is a perfect example of the asymmetry between our capacities for data production and data analysis. In the materials field, DFT calculations have become commoditized to the point where individual research groups can calculate large fractions of all known bulk crystalline materials [25,26]. By far the more urgent matter is, when we want to discover a new material, knowing *where* in composition space to point our computer power.

## II. METHODOLOGICAL DETAILS

Our ML-based approach answers this question with a predictive computational tool, trained on the results of previous calculations, which may then be used to suggest new chemistries of interest. In other words, the outcomes of *all prior calculations* should inform what we calculate next. ML concepts have been previously used with success to predict crystal structure [3,13,14]. These works describe a

\*These authors contributed equally to this work.

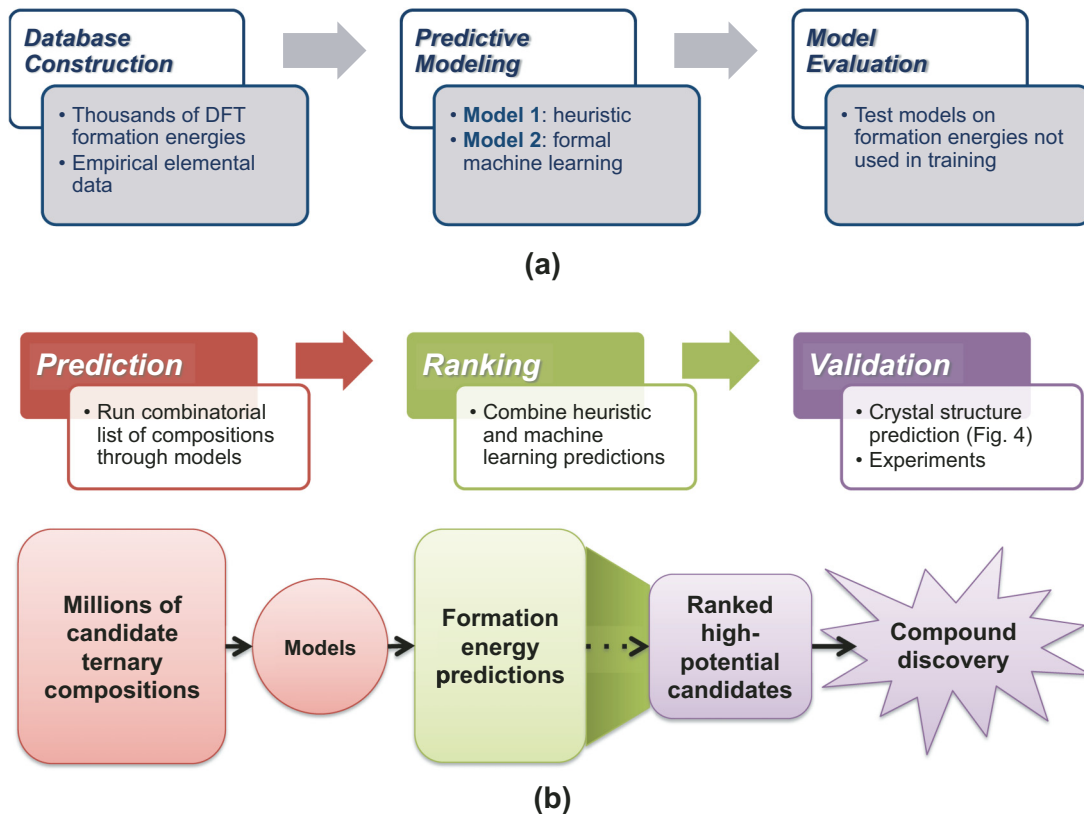


FIG. 1. (Color online) Schematic of our data-driven approach to materials discovery. (a) The creation and evaluation of heuristic and machine learning models from input quantum mechanical energetics. (b) Our method for identifying likely new ternary compounds. We ran an extremely large combinatorial list of ternary compositions through both heuristic and machine learning models, and combined these two models to rank the compositions based on how likely they are to form compounds in nature.

classification model that suggests likely crystal structures for an input composition based on data-mined chemical similarities between elements; to obtain energetics, one must perform DFT calculations on the best candidate structures. In a similar vein, Bayesian statistics has been used to more efficiently construct cluster expansions, which in turn can elucidate ground-state decorations in fixed-lattice mixing problem [27]. Our approach differs from this previous work in a few key ways: It immediately provides *quantitative* property predictions at arbitrary compositions, such that its outputs can be used directly for modeling without recourse to DFT or CSP; it is structure independent, so that we may identify promising compositions even if their eventual structures do not already exist in our input database; and it could be trained on *any* property of interest, not just thermodynamic stability.

Figure 1 depicts our overall DFT calculation, model building, and prediction scheme. Our first set of steps [Fig. 1(a)] results in the construction of a large database of DFT calculations, two predictive formation energy models (one heuristic, one ML based), and a quantitative measure of the models' accuracies. With validated models in hand, we proceed to the second set of steps [Fig. 1(b)]: scanning millions of ternary compositions with the models, ranking the resulting list by *combining* the two models, and presenting this list as a prioritized and nearly exhaustive survey of possible ternaries. In the rest of this paper, we discuss the details and results of these processes.

The input to our prediction machinery is a large database of (composition, formation energy) data points based on outputs from over 15 000 DFT calculations on various materials, spanning a wide variety of chemical systems. We discretize composition space in all binary  $A$ - $B$  elemental systems and use as training data the formation energies at each grid point; for ternary compositions, we include as ML input only formation energy values associated with stable compounds (i.e., no intermediate compositions associated with *mixtures* of stable compounds). The structures associated with all calculations are drawn from the Inorganic Crystal Structure Database (ICSD). The Supplemental Material contains more details on how we built our thermodynamic database [21].

We now describe our construction of two types of models, which we eventually combine into a single stability-prediction framework. The first is based on a simple but powerful metallurgical heuristic, which states that the formation energy in a ternary system can be estimated from a composition-weighted average of that ternary's constituent *binary* formation energies [28,29]. For example, the heuristic would estimate the formation energy of a ternary compound  $ABC$  as an equal weighting of binary  $AB$ ,  $BC$ , and  $AC$  formation energies, whose compositions are determined via a simple geometrical construct we describe further in the Supplemental Material [21]. We find that, when we apply this heuristic to our data, it significantly underestimates the ternary formation energy—but that this underestimation is remarkably systematic. Using

linear regression to a fitting set of ternary compounds, we modify the simple composition-weighted formation energy (FE) heuristic as follows:

$$FE_{\text{modified}} = 1.50 FE_{\text{heuristic}} - 0.020 \text{ eV/atom.}$$

The second model we use is based on formal ML techniques, and in particular consists of ensembles of decision trees trained to predict formation energies of arbitrary compositions.

We provide in the Supplemental Material a detailed description of these ML methods [21].

### III. RESULTS

We illustrate the performance of the heuristic and ML models in Fig. 2. For a true test of predictive power, we fit to 4000 ternaries and withhold  $\sim 8600$  others from either of the fitting procedures, and compare DFT formation energies

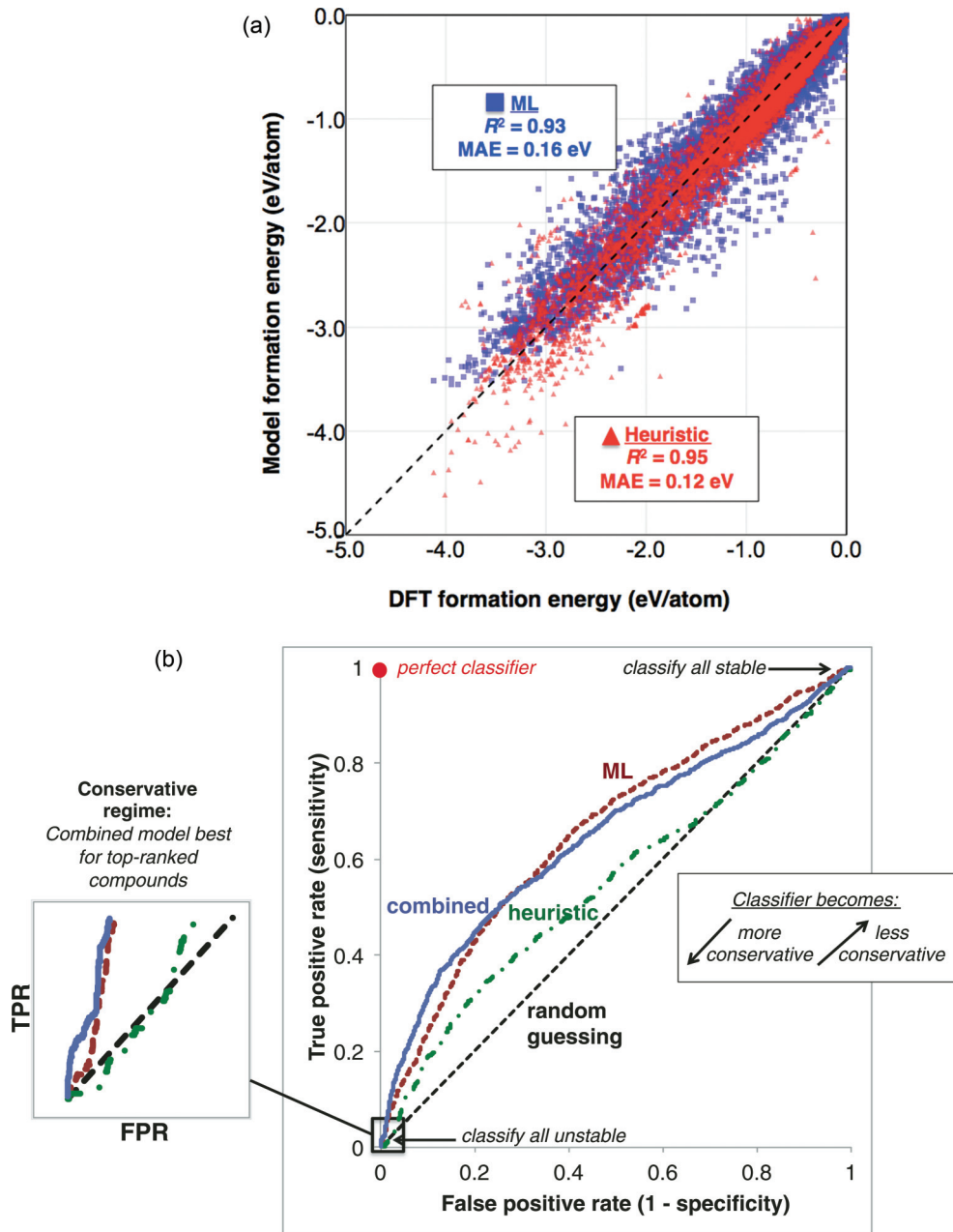


FIG. 2. (Color online) (a) Comparison between DFT-calculated and model-predicted formation energies for  $\sim 8600$  ternary compounds to which the models were not fit. Both models give excellent  $R^2$  scores, and mean absolute errors (MAEs) relative to DFT that are much smaller in magnitude than typical DFT discrepancies with experiment ( $\sim 0.25$  eV/atom) [30]. (b) Receiver operating characteristic (ROC) curves for the heuristic, ML model, and the combined heuristic-ML stability-ranking scheme. ROC curves provide a visual comparison of the various methods' ability to rank input compositions by thermodynamic stability. Better ranking performance causes the curves to rise above the random guessing line toward the perfect classifier in the upper left. The inset magnifies the region in the lower left corresponding to a very conservative classifier—i.e., one that labels only the top few compounds as “very stable.” The combined model performs best in the conservative classifier regime.

to the models' outputs. As shown in Fig. 2(a), both models are excellent at predicting DFT formation energies to which they were not fit, with  $R^2$  scores above 0.9 and mean absolute errors (MAEs) well within DFT's typical agreement with experiment [30]. We note that expected DFT error is a strong function of chemistry, and our MAE values are averaged over an exceptionally wide variety of chemistries, as in Ref. [29]. Our models' performance carries remarkable implications: In principle, these models can make formation energy predictions without any structural input, at six orders of magnitude lower computational expense than DFT, and without sacrificing the accuracy of DFT compared to experiment.

Although our models' numerical predictions of DFT formation energies are important indicators of accuracy, in our work, we are asking the models a slightly different question. We want them to scan vast swaths of ternary composition space and recommend to us the compositions most likely to contain undiscovered but stable compounds. This undertaking amounts to *ranking* millions of candidate ternary compositions by their predicted stability. A so-called receiver operating characteristic (ROC) curve is an established means of visualizing a ranking-based Boolean classifier's performance [31]—in our case, we rank our entire list by predicted driving force for formation, and declare all members of the list above a certain threshold “stable.” The rest, by default, are classified as “unstable.” Because, for discovery purposes, we are most interested in compositions whose predicted energies are *well* below any combinations of known compounds, we modify the binary classification to be “*very* stable” (at least 100 meV/atom below mixtures of currently known compounds) or “*not very* stable” (which could mean simply less stable, or not stable at all).

We depict ROC curves for the heuristic alone, the ML model alone, and their ensemble in Fig. 2(b), classifying the same set of  $\sim 8600$  ternary compounds whose formation energies are known to us but not used to fit models in Fig. 2(a). ROC curves, generally plotted as true positive rate (TPR) vs false positive rate (FPR), illustrate the necessary trade-off between *sensitivity* (i.e., TPR, or stable compounds classified as stable) and *specificity* (i.e.,  $1 - \text{FPR}$ , or unstable compounds classified as unstable). Each point on a model's ROC curve corresponds to a particular choice of stability threshold along the ranked list of compounds; the bottom left of the plot represents declaring *all* compounds unstable, the top right point classifies *all* compounds as stable, and the more relevant intermediate thresholds trace the curves. A perfect classifier (indicated) has a TPR of 1 and an FPR of 0. The  $y = x$  line on the plot gives the result of random guessing, or randomly shuffling the list of compounds and choosing arbitrary stability thresholds along the resulting list; any classifiers *above* that line (all three, in this case) are superior to guessing.

Figure 2(b) demonstrates that, at the 100 meV/atom stability criterion considered here, the ML model alone is superior to the heuristic alone, even though the heuristic gave a better  $R^2$  score when comparing to DFT formation energies [Fig. 2(a)]. However, we find that a *combined* model—formed by averaging the heuristic and ML rankings for a given compound—performs best in the region of ROC space most relevant to us: the bottom left, where the models rank the highest-priority discoveries associated with a very stringent

cutoff on the ranked list of compounds. In general, we expect that only the most promising predictions of our model for new ternary compounds will undergo further investigation, and this observation corresponds to a *conservative* model: one in which only a few top compounds are classified as stable (i.e., attractive for discovery). As indicated in the inset to Fig. 2(b), the combined model rises faster than either model alone, and maintains its performance lead over the ML model until  $\text{FPR} \approx 0.3$ . The superior performance of our hybrid model is a direct result of ensembling, which is a well-established machine learning strategy to create combined models whose predictions are more reliable than any of the constituent methods alone [32].

With the robustness of both the heuristic and ML stability-ranking models established, we turn now to discovery with an ensemble of the two methods, which yielded the best performance at our stability-ranking task. We use the combined model to conduct “virtual combinatorial chemistry” in search of heretofore-undiscovered stable ternary compounds of the form  $A_x B_y C_z$ . The elements  $A$ ,  $B$ , and  $C$  are drawn from a list of essentially all technologically relevant elements (83 total). Within the corresponding  $C(83,3) = 91\,881$  ternary systems, we evaluate our stability predictor on compositions that are statistically the most common in the ICSD (though our method also works on arbitrary compositions). For example, our statistical search reveals that the most-abundant ternary composition is  $ABC_3$ , the second is  $ABC_2$ , the third is  $ABC$ , and so on. In the case of compositions containing preferentially ionic elements such as O and F, we require that the resulting composition satisfy charge balance by considering common oxidation states for cationic and anionic species. This enumeration procedure results in a list of 1.6 million physically reasonable candidate compositions for evaluation.

We run these 1.6 million compositions through both the heuristic and ML models, ranking the compositions by the degree to which the models predict them to be thermodynamically preferred to any combination of known compounds. Note that this ranking is possible in minutes with the models we have constructed, in contrast to the tens of thousands of processor years that a DFT-based crystal structure prediction method would require. As described above, we then combine the results of the two models by averaging their rankings for each composition. This combined model ranks the input compositions by the likelihood that they form a compound.

Figure 3 gives four representative slices of our full ternary stability map; brighter colors indicate higher rankings. We emphasize that these heat maps are just four example components of an essentially complete description of ternary energetics across a combinatorial search space of elemental compositions and stoichiometries, which would be entirely impractical to achieve either experimentally or with existing computational methods.

Figure 3 offers a number of compelling discovery results that emerge automatically from our composition scan. First, in keeping with our intuition, novel oxides generally rank highly in our survey; where they remain undiscovered, their formation energies relative to known compounds are expected to be large. In particular, we predict many new oxides containing Tc, Rh, and Pa, which is sensible given that these chemistries



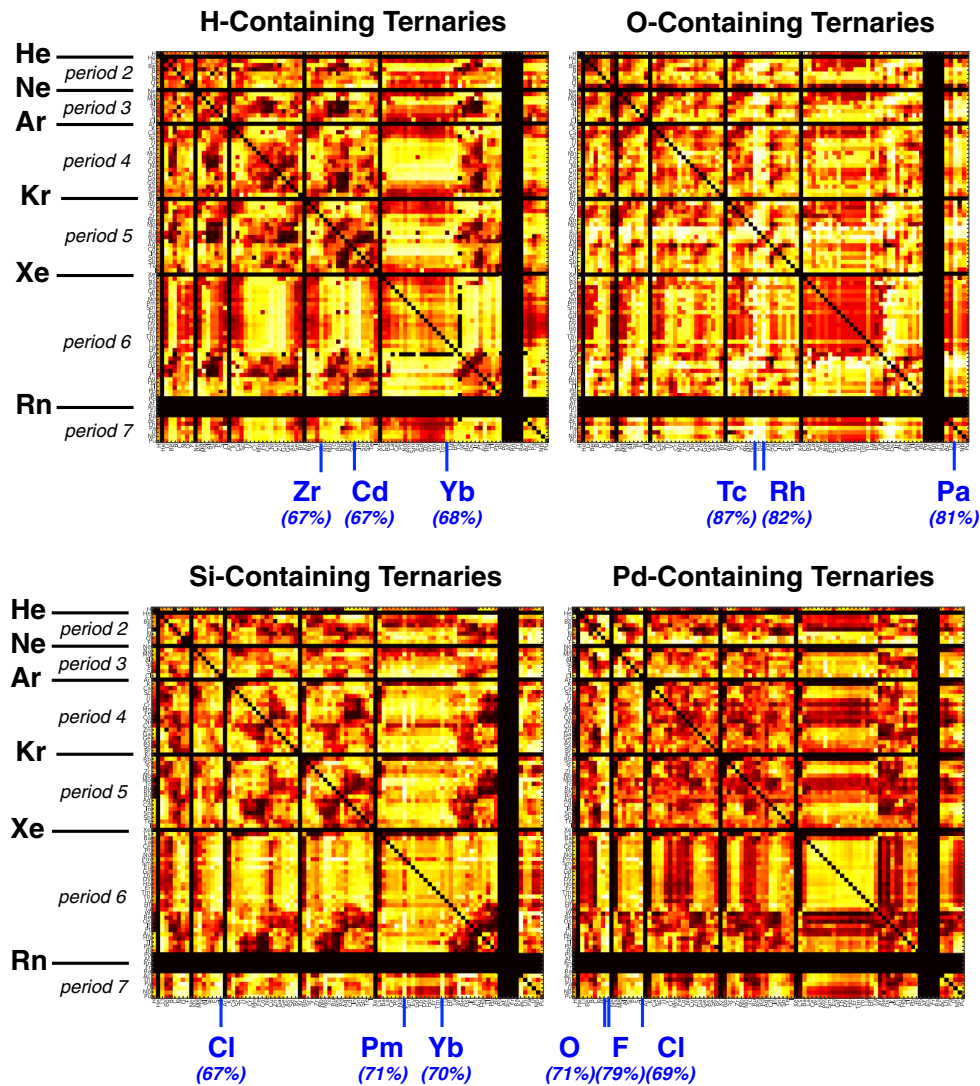


FIG. 3. (Color online) Representative heat maps of ternary compositions' stability rankings, upon fixing one component of the ternary to either H, O, Si, or Pd. Brighter colors imply higher rankings (greater likelihood of stability). Each heat map cell specifies the average ranking of ternary compounds containing the element in the heat map title and the two additional elements given on the  $x$  and  $y$  axes. Because the element labels are necessarily small, we demarcate each period for greater clarity. For each of the four elements we consider here, we list the top three additional elements most conducive to ternary stability, along with their percentiles among the 1.6 million compositions we evaluated. Thus, O-Tc-X ternaries are on average very stable, in the 87th percentile of our ranked list. Black bars on the plot correspond to either noble gases or several exotic heavy elements that we did not consider in our survey.

are likely to be less explored among possible transition metal and rare earth oxides. We also highlight Pd-containing oxides, fluorides, and chlorides as promising candidates for discovery, which may be relevant in catalysis applications and provides further insight into the nobility (or lack thereof) of Pd. Turning to Si-containing ternaries, Si-transition metal (TM)–rare earth (RE) intermetallics represent a particularly bright region of the Si heat map. These unique chemistries, combining  $p$ -block,  $d$ -block, and  $f$ -block elements, could possess unusual electronic and physical properties. Finally, our search for new hydrides yields regions of promise among H-TM-RE and H-RE-semimetal chemistries. A broader understanding of metal hydride systems could aid in the design of effective hydrogen storage materials.

To estimate the total number of possible new compounds highlighted in our study, we conservatively limit the count to just one new material per  $A$ - $B$ - $C$  ternary system, require that *both* the heuristic and ML models predict the compound to be at least 0.2 eV/atom more stable than any possible mixture of known compounds (an energetic threshold greater than the models' MAE values from Fig. 2), and discard any very dilute compositions. Based on these criteria, we identify about 4500 predicted new ternary compounds in the present work; the full list appears in the Supplemental Material [21]. Encouragingly, some of the highest-ranking predictions correspond to compounds that seem intuitively reasonable based on analogy with known compounds, but may not have yet been synthesized in the laboratory: For example,

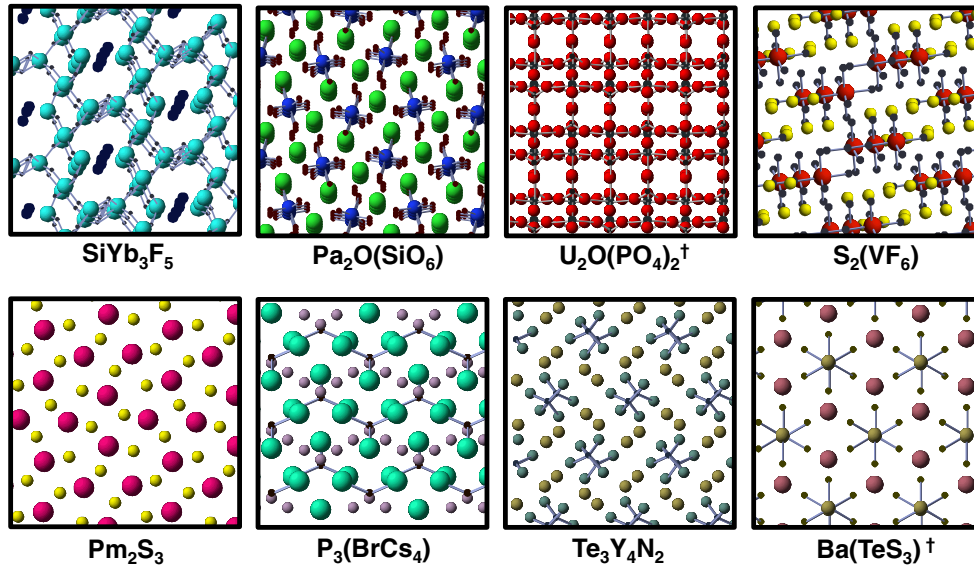


FIG. 4. (Color online) Example high-ranking compositions from our ternary compound screen whose stability we explicitly confirmed with database-search crystal structure prediction. Our method is successful at identifying new stable compounds across a wide variety of chemistries. “†” indicates a model prediction associated with a known stable ternary compound that had been missing from our DFT thermodynamic database; the prediction is thus confirmed, but no crystal structure search was necessary.

the model suggests a set of rare-earth phosphates, sulfates, nitrates, and selenates as very likely to be stable.

Our final objective in this work is to explicitly confirm a chemically diverse set of predictions that have emerged from our approach. We achieve this goal by systematically selecting several high-ranking stability predictions, and investigating them with a full DFT crystal structure search. We consider a prediction successful if we locate a structure at the target composition whose total energy within DFT is lower (i.e., more favorable) than any combination of existing compounds or elements at that composition, implying that a new compound should form there.

To evaluate our approach’s ability to identify new compounds across a variety of chemistries, we explore several high-ranking ternary systems according to the following logic. First, we generated charge-balanced compositions having plentiful candidate crystal structures within the four top-ranked ternary systems: Si-Yb-F ( $\text{SiYb}_3\text{F}_5$ ), Si-Pa-O ( $\text{SiPa}_2\text{O}_7$ ), U-P-O ( $\text{U}_2\text{P}_2\text{O}_9$ ), and V-F-S ( $\text{VS}_2\text{F}_6$ ). Next, we consider the top-ranked ternary excluding any O- or F-containing systems, to obtain a very different chemistry: Pm-S-Se (top composition  $\text{Pm}_{12}\text{S}_{19}\text{Se}$ , which we interpret as a very interesting “binary”  $\text{Pm}_2\text{S}_3$  prediction from our ternary model). We then exclude, in order, rare earths (obtaining  $\text{Rb}_2\text{N}_3\text{Cl}$ ), Cl (obtaining  $\text{Cs}_4\text{BrP}_3$ ), Br and P (obtaining  $\text{Te}_3\text{Y}_4\text{N}_2$ ), and N (obtaining  $\text{BaTeS}_3$ ). Again, we emphasize that the intention behind exploring our list in this fashion is to generate an interesting, chemically distinct set of compositions to test; for example, many of the top-ranked compositions correspond to various elements’ phosphates, but for illustrative purposes we only explicitly investigate a uranium phosphate.

Among the nine compositions we identify above, we successfully confirm eight as new ternary compounds with a DFT crystal structure database search; each search involved calculating  $\sim 40$ – $200$  candidate structures. The eight successes

are shown in Fig. 4. Our best-obtained structure for  $\text{Rb}_2\text{N}_3\text{Cl}$  is about 30 meV/atom less stable than a mixture of known compounds; this near miss ( $kT$  at room temperature is 26 meV/atom) suggests that unconstrained crystal structure prediction or an alternative composition may yet yield a stable compound. Our 89% success rate is extremely high, given that only about 11% of ternary systems have *any* known compounds, and only 0.5% of the 1.6 million enumerated compositions we scanned correspond to already-known materials. We also emphasize that the set of 4500 predictions we report in this study would—if experimentally confirmed—represent a more than 10% increase in the total number of ternary compounds discovered to date.

#### IV. SUMMARY

Here we described an approach for an extremely rapid computational materials screening method, based on coupling a physically motivated heuristic and an advanced machine learning model to a large database of quantum mechanical calculations on known materials. Our data-driven approach has learned the rules of chemistry from DFT, makes energetic predictions for new compositions at six orders of magnitude lower computational expense, and requires no knowledge of crystal structure. We took as an example application the evaluation of thermodynamic stability of 1.6 million ternary compositions, which we ranked by their probability to exist as a stable compound in nature; this search revealed a total of about 4500 heretofore-unknown, yet likely stable, ternary materials. Our comprehensive survey greatly expands our understanding of ternary composition space, lays the foundation for entirely new phase diagrams, suggests additional contributions to existing phase diagrams, and provides a large list of interesting new chemistries that may now be mined for technological applications. This survey would be computationally prohibitive

with full quantum mechanical methods, as well as impractical to conduct in the laboratory. We suggest that, beyond guiding ternary compound synthesis, our method could be used in a very wide range of crucial materials discovery tasks, wherein instead of predicting thermodynamic stability, one would predict, e.g., Li capacity (a feature of batteries), band gap (a feature of solar cells), or magnetic moment (a feature of permanent magnets).

#### ACKNOWLEDGMENTS

The authors of this work gratefully acknowledge support from the following sources: the National Defense Science and Engineering Graduate Fellowship (B.M.); DOE under

Grant No. DE-FG02-07ER46433 (B.M. and C.W.), the Center for Electrical Energy Storage: Tailored Interfaces, an Energy Frontier Research Center funded by the US Department of Energy, Office of Science, Office of Basic Energy Sciences (S.K.); NSF Awards No. CCF-1029166 and No. OCI-1144061, DOE Awards No. DE-SC0005340 and No. DE-SC0007456, and AFOSR Award No. FA9550-12-1-0458 (A.A., K.Z., A.C.); the Revolutionary Materials for Solid State Energy Conversion, an Energy Frontier Research Center funded by the US Department of Energy, Office of Science, Office of Basic Energy Sciences under Award No. DE-SC00010543 (J.W.D.); and the Office of Basic Energy Sciences through Grant No. DE-FG02-98ER45721 (J.E.S.). Calculations were performed on Northwestern University's Quest high-performance computing system.

- 
- [1] T. Hey, S. Tansley, and K. M. Tolle, *The Fourth Paradigm: Data-Intensive Scientific Discovery* (Microsoft Research, Redmond, WA, 2009).
- [2] K. Rajan, *Mater. Today* **8**, 38 (2005).
- [3] S. Curtarolo, D. Morgan, K. Persson, J. Rodgers, and G. Ceder, *Phys. Rev. Lett.* **91**, 135503 (2003).
- [4] M. Rupp, A. Tkatchenko, K.-R. Müller, and O. A. von Lilienfeld, *Phys. Rev. Lett.* **108**, 058301 (2012).
- [5] J. C. Snyder, M. Rupp, K. Hansen, K.-R. Müller, and K. Burke, *Phys. Rev. Lett.* **108**, 253002 (2012).
- [6] P. Hohenberg and W. Kohn, *Phys. Rev.* **136**, B864 (1964).
- [7] W. Kohn and L. J. Sham, *Phys. Rev.* **140**, A1133 (1965).
- [8] X. W. Zhang, L. P. Yu, A. Zakutayev, and A. Zunger, *Adv. Funct. Mater.* **22**, 1425 (2012).
- [9] L. Yu and A. Zunger, *Phys. Rev. Lett.* **108**, 068701 (2012).
- [10] J. W. Bennett, K. F. Garrity, K. M. Rabe, and D. Vanderbilt, *Phys. Rev. Lett.* **109**, 167602 (2012).
- [11] A. Roy, J. W. Bennett, K. M. Rabe, and D. Vanderbilt, *Phys. Rev. Lett.* **109**, 037602 (2012).
- [12] S. Kirklin, B. Meredig, and C. Wolverton, *Adv. Energy Mater.* **3**, 252 (2012).
- [13] G. Hautier, C. C. Fischer, A. Jain, T. Mueller, and G. Ceder, *Chem. Mater.* **22**, 3762 (2010).
- [14] C. C. Fischer, K. J. Tibbetts, D. Morgan, and G. Ceder, *Nat. Mater.* **5**, 641 (2006).
- [15] C. W. Glass, A. R. Oganov, and N. Hansen, *Comput. Phys. Commun.* **175**, 713 (2006).
- [16] G. Trimarchi and A. Zunger, *Phys. Rev. B* **75**, 104113 (2007).
- [17] C. Wolverton and V. Ozolins, *Phys. Rev. B* **75**, 064101 (2007).
- [18] G. Trimarchi, A. J. Freeman, and A. Zunger, *Phys. Rev. B* **80**, 092101 (2009).
- [19] Y. Ma, M. Eremets, A. R. Oganov, Y. Xie, I. Trojan, S. Medvedev, A. O. Lyakhov, M. Valle, and V. Prakapenka, *Nature* **458**, 182 (2009).
- [20] A. R. Oganov, J. Chen, C. Gatti, Y.-Z. Ma, Y.-M. Ma, C. W. Glass, Z. Liu, T. Yu, O. O. Kurakevych, and V. L. Solozhenko, *Nature* **457**, 863 (2009).
- [21] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevB.89.094104> for full calculation details and list of predicted materials.
- [22] N. Ampazis, in *Artificial Neural Networks—ICANN 2010*, edited by K. Diamantaras, W. Duch, and L. S. Iliadis, Lecture Notes in Computer Science (Springer, Berlin, 2010), Vol. 6354, pp. 429–434.
- [23] D. N. Reshef, Y. A. Reshef, H. K. Finucane, S. R. Grossman, G. McVean, P. J. Turnbaugh, E. S. Lander, M. Mitzenmacher, and P. C. Sabeti, *Science* **334**, 1518 (2011).
- [24] G. Bell, T. Hey, and A. Szalay, *Science* **323**, 1297 (2009).
- [25] A. Jain, G. Hautier, C. J. Moore, S. P. Ong, C. C. Fischer, T. Mueller, K. A. Persson, and G. Ceder, *Comput. Mater. Sci.* **50**, 2295 (2011).
- [26] K. S. Yang, W. Setyawan, S. D. Wang, M. Buongiorno-Nardelli, and S. Curtarolo, *Nat. Mater.* **11**, 614 (2012).
- [27] T. Mueller and G. Ceder, *Phys. Rev. B* **80**, 024103 (2009).
- [28] M. Hillert, *Phase Equilibria, Phase Diagrams and Phase Transformations: Their Thermodynamic Basis* (Cambridge University Press, Cambridge, 2008).
- [29] Y. M. Muggianu, M. Gambino, and J. P. Bros, *J. Chim. Phys. Phys.-Chim. Biol.* **72**, 83 (1975).
- [30] V. Stevanovic, S. Lany, X. W. Zhang, and A. Zunger, *Phys. Rev. B* **85**, 115104 (2012).
- [31] J. A. Swets, *Science* **240**, 1285 (1988).
- [32] T. G. Dietterich, in *Multiple Classifier Systems: 1st International Workshop, MCS 2000*, edited by J. Kittler and F. Roli, Lecture Notes in Computer Science (Springer, Berlin, 2000), Vol. 1857, pp. 1–15.