ADVANCED MATERIALS
www.advmat.de

Phase Selection Rules of Multi-Principal Element Alloys

*Lin Wang and Bin Ouyang**

Computational prediction of phase stability of multi-principal element alloys (MPEAs) holds a lot of promise for rapid exploration of the enormous design space and autonomous discovery of superior structural and functional properties. Regardless of many plausible works that rely on phenomenological theory and machine learning, precise prediction is still limited by insufficient data and the lack of interpretability of some machine learning algorithms, e.g., convolutional neural network. In this work, a comprehensive approach is presented, encompassing the development of a complete dataset that contains 72 387 density functional theory calculations, as well as a predictive global phenomenological descriptor. The phase selection descriptor, based on atomic electronegativity and valence electron concentration, significantly outperforms the widely used valence electron concentration, excelling in both accuracy (with an f1 score of 63% compared to 47%) and its ability to predict the HCP phase (0.48 recall compared to 0). The comprehensive data mining on the global design space of 61 425 quaternary MPEAs made from 28 possible metals, together with the phenomenological theory and physical interpretation, will set up a solid computational science foundation for data-driven exploration of MPEAs.

1. Introduction

Multiprincipal element alloys (MPEAs) are formed by mixing multiple elements at similar concentrations. In contrast to traditional alloys, the high configurational entropy introduced by the multiple elements leads to the stabilization of MPEA as single-phase solid solutions in many situations.[1–3] Since the first report in 2004 by two groups simultaneously,[3,4] MPEAs have been extensively studied as both structural alloys and novel catalysts. Like typical metals, many single-phase MPEAs exhibit body-centered cubic (BCC), face-centered cubic (FCC) or hexagonal closed-packed (HCP) structures. Different crystal structures exhibit distinct mechanical,[5–8] magnetic,[9,10] and catalytic properties[11,12] due to variations in their atomic arrangements and bonding characteristics.[13] Understanding the driving force behind phase stability and designing alloys with enhanced properties is an ongoing pursuit.

L. Wang, B. Ouyang
Department of Chemistry and Biochemistry
Florida State University
Tallahassee, FL 32304, USA
E-mail: bo22b@fsu.edu

The ORCID identification number(s) for the author(s) of this article can be found under https://doi.org/10.1002/adma.202307860

DOI: 10.1002/adma.202307860

The compositional space of MPEAs is enormous and impossible to enumerate with combinatorial experiments and computations. Therefore, one will need a quantitative model to predict the most likely phase for MPEAs. In the past, phenomenological phase selection rules have been attempted with valence electron concentration (VEC),[14,15] atomic radius differences,[16,17] electronegativity difference,[18] melting points (T_m),[19] mixing enthalpy and entropy,[16] pairwise mixing enthalpies,[15,20] etc. In addition to phenomenological theory, machine learning[21–31] has been widely used for predicting the phase selection for specific MPEAs. There are two remaining challenges for predicting the most stable phase for MPEAs. First, the training data sets that establish the phenomenological and machine learning models are relatively limited, usually including hundreds of data points and sometimes a combination of data with different fidelity. When applied to the universal combinatorial space of MPEAs, these models may fail or experience a drop in performance due to either insufficient training or incomplete physics to capture global trends. On the other hand, most machine learning models reported are convoluted maps between basic features and stability metrics based on deep neural networks, often lacking interpretability.[32–34]

In this work, we have presented a comprehensive approach, encompassing the development of a complete dataset and predictive global phenomenological theory for phase selection rules of MPEAs. Our work involves the construction of a density functional theory (DFT) dataset, which covers all MPEAs with equal molar concentrations of up to four constituent metals. The dataset includes the entire combinatorial space of 28 metals and three typical phases: BCC, FCC, and HCP. Through the analysis of 72387 DFT relaxations, we establish the physical understanding of phase selection among all possible MPEA compositions. Furthermore, we developed a phenomenological theory that predicts the lowest energy phase among BCC, FCC, and HCP structures. With the application of symbolic machine learning, we discovered that the phenomenological physical descriptor $\tau = \mathrm{VEC} \times (\mathrm{VEC}_{\mathrm{std}} + \mathrm{VEC}) + \sqrt[3]{\log(\chi_{\mathrm{ave}})}$ shows extraordinary capability for predicting lowest energy phase among BCC, FCC, and HCP. More specifically, we find that BCC is stable at $\tau < 63.31$, HCP is stable at $63.31 < \tau < 81.21$ and FCC is stable at $\tau > 81.21$. Our presented model achieves an average f1 score of 0.63, relative to 0.47 provided by VEC rule.[14,35] Moreover, VEC

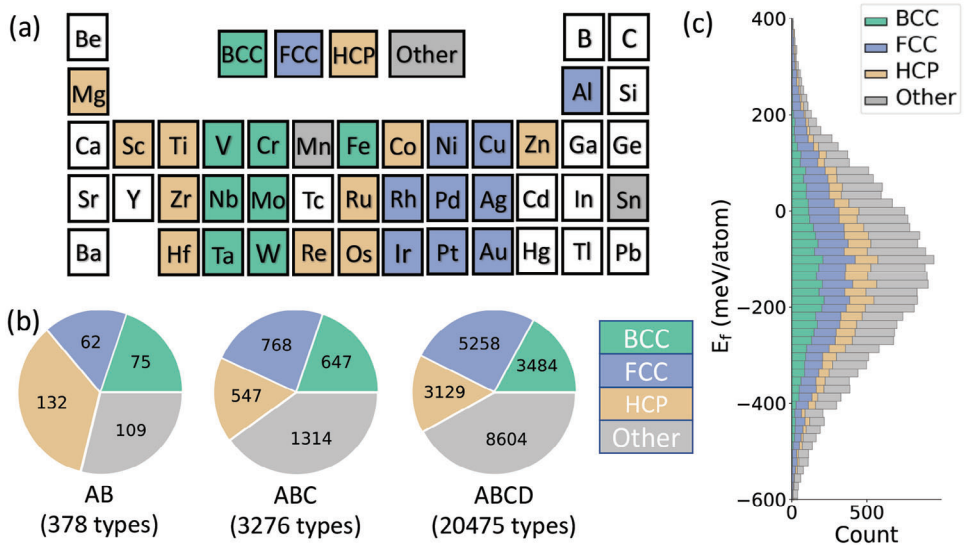

Figure 1. a) The distribution of selected metals in periodic table, the color indicates the room temperature phases as elemental metal.[36] The room temperature phases of Mn and Sn are none of the three phases (BCC/FCC/HCP) we emphasize here so we denote them as "other". b) The distribution of lowest energy phases across all equal-molar combinations of 28 metals, leads to 378 binary and 3276 ternary and 20475 quaternary alloys. c) The stacking histogram of the formation energy (E_f) all the most stable structural form among three computed structures for all 24129 equal-molar alloy compositions.

rule tends to miss-predict all HCP phases, but our model shows 0.46 precision and 0.48 recall for predicting HCP ground states. By establishing universal phase selection rules based on a comprehensive dataset that covers the complete elemental combinatorial space up to quaternary systems, our work offers new insights into the understanding of MPEAs by integrating machine learning with physical intuition. Furthermore, it establishes a solid data science foundation for data-driven design of MPEAs.

2. Results

2.1. Phase Stability Map of all MPEAs

To enumerate all possible equal-molar alloys, we exclude all alkali and alkaline earth metals except for Mg, as well as eliminating Ga, In, Cd, Hg, and Tl because of their low melting points, which hinders the formation of entropy-stabilized solid solutions. Tc is also excluded because it is radioactive. As a result, 28 metals were selected to generate 378 binary, 3276 ternary, and 20475 quaternary equal-molar MPEA compositions. The room temperature phases for these metals are illustrated in **Figure 1**a with different colors.[36] For each composition, we generated special quasi-random structure (SQS) for BCC, FCC, and HCP structure, which yields 72387 DFT calculations (details in the Experimental Section). After full DFT relaxation on both cell shape and atomic position, some structures have significant lattice distortion while others are transformed into different structures. Structure matching is thus performed to map all DFT relaxed structures back into the closest structural phase with rigid lattice sites.[37] More specifically, all relaxed structures into BCC lattice, FCC lattice, HCP lattice or other lattices as being classified by a strict tolerance of off-lattice displacement and unit cell deformation (Note S1, Supporting Information). For the structures that cannot be mapped back into the three considered phases, we group and name them as "Other".

Given that three basic structures (BCC, FCC, and HCP) have been relaxed by DFT for all compositions, we can compare the DFT computed energy to derive the lowest energy phase. The distributions of the lowest energy phase for binary, ternary, and quaternary MPEAs are demonstrated in Figure 1b. It is interesting to see there are 109 binary alloys, 1314 ternary alloys and 8604 quaternary alloys with the lowest energy relaxed structures failing to be matched into any of the three phases. To show the stability of these MPEAs after being classified into different phases, the formation energy of the lowest energy phase of all binary, ternary, and quaternary alloys are visualized with a stacking histogram as shown in Figure 1c. It can be seen that most MPEAs have very negative formation enthalpy (E_{Form}). Most of them have a formation enthalpy around −150 meV/atom while some of them can be as low as −600 meV/atom. Such distribution implies that our analysis based on alloy compositions shows reasonable stability and synthetic accessibility.

2.2. Structural Origin of Mechanical Instability

As being demonstrated in Figure 1b, in many situations, the most stable phase is classified as "Other" as the initial atomic positions are not even mechanically stable. As a result, DFT optimization will transform them into something else other than the three considered phases. More detailed statistics of relaxation-induced structural evolutions of quaternary alloys are demonstrated in **Figure 2**a. Specifically, the off-diagonal cubes indicate the number of structures that evolve into very different structures from the initial point. It can be inferred from Figure 2a

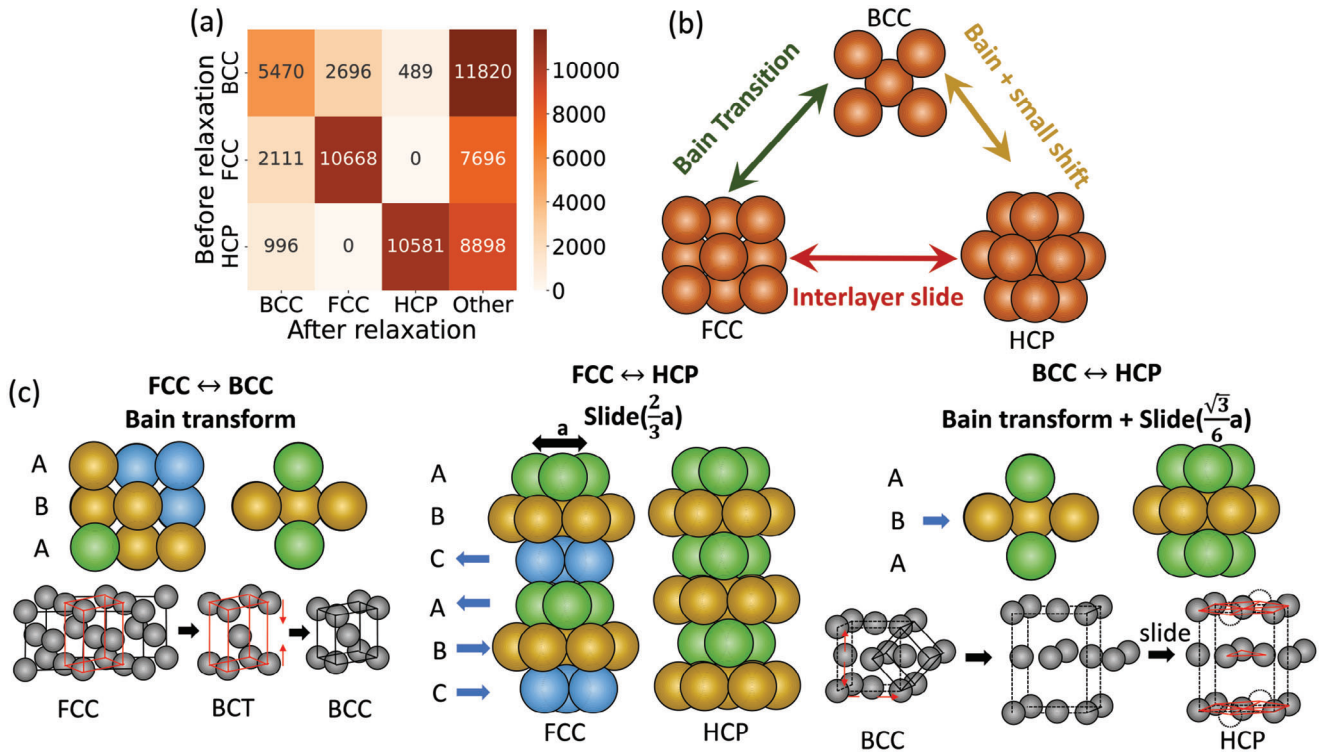

Figure 2. a) The statistics of structures that have phase transformation after DFT relaxation; b) Schematic summary of possible phase transformation mechanism among FCC/BCC/HCP structures; c) Atomic structure that illustrate the structural correlation and transformation pathway among FCC/BCC/HCP phases. Blue arrows represent sliding directions. Here we use shortest atomic distance in the unit cell (denoted as (a)) as a measure of sliding distance.

that BCC ↔ FCC and BCC ↔ HCP transitions have occurred in many compositions while there is no observation of any relaxation-induced FCC ↔ HCP transition. Moreover, there have been a substantial number of structures that cannot be mapped back into BCC/FCC/HCP lattices for all three types of initial states.

The mechanical instability and consequential transformation can be better understood by examining the structural correlation among the three investigated phases, as illustrated in Figure 2b,c. In summary, as depicted in Figure 2b, the structural transition between BCC and FCC phases can occur through lattice distortion alone, following the Bain transformation.[38] Such transformation is diffusionless, meaning it does not require an activation barrier and can occur even during DFT relaxation. Furthermore, BCC ↔ HCP transformation is close to Bain transformation, with the exception that it involves a slight shift of atom planes. This shift can potentially occur without an activation barrier if there is a significant lattice distortion present. On the other hand, FCC ↔ HCP transformation necessitates substantial interlayer sliding, which cannot occur without an activation barrier or during DFT relaxation. Such a simple framework can completely explain the facts in Figure 2a that (1) There is an absence of FCC ↔ HCP transformation; (2) Much fewer structures have gone through the BCC ↔ HCP transformation compared with BCC ↔ FCC transformation. The structural correlation among these three considered phases can be better elaborated in Figure 2c. The left panel of Figure 2c introduces the structural correlation between BCC and FCC phases. The BCC lattice can be derived from the FCC lattice by simply compressing the c/a ratio of the BCT lattice from 1 to 0.71. Such compressing of lattice does not necessitate passing through a high-energy transition state. Instead, it can be facilitated statistically with off-lattice distortion that is commonly observed in HEA. The panel in the middle demonstrates the structural correlation between FCC and HCP phases. They are both closed-packed structures but with different stacking sequences, e.g., ABCABC stacking sequence for FCC and ABAB stacking for HCP, respectively. As demonstrated by the middle panel of Figure 2c, such transformation requires the creation of a high-density stacking fault, which dictates a significant activation barrier and thus cannot be observed with just DFT relaxations (Figure 2a). The right panel of Figure 2c presents the structural correlation between the BCC and HCP phases. The transformation from BCC to HCP is regarded as a two-stage process. In the first stage, the bcc structure elongates in $[110]_{bcc}$ and $[1\bar{1}0]_{bcc}$ directions, respectively. In the second stage, half of the layers need to slide $\sqrt{3}/6$ of the distance between the two nearest atoms to transform into HCP phase.[39] The most significant structural change that happens during BCC ↔ HCP evolution is in analog to Bain transformation.[38] While a slight sliding of atom planes is also involved, it requires much smaller fraction of atoms and shorter sliding distances compared to the FCC ↔ HCP transformation. As a result, this transformation can be observed during DFT relaxation, as demonstrated in Figure 2a. The smaller scale of atomic movement makes it more feasible for the transformation to occur within the constraints of DFT relaxation.

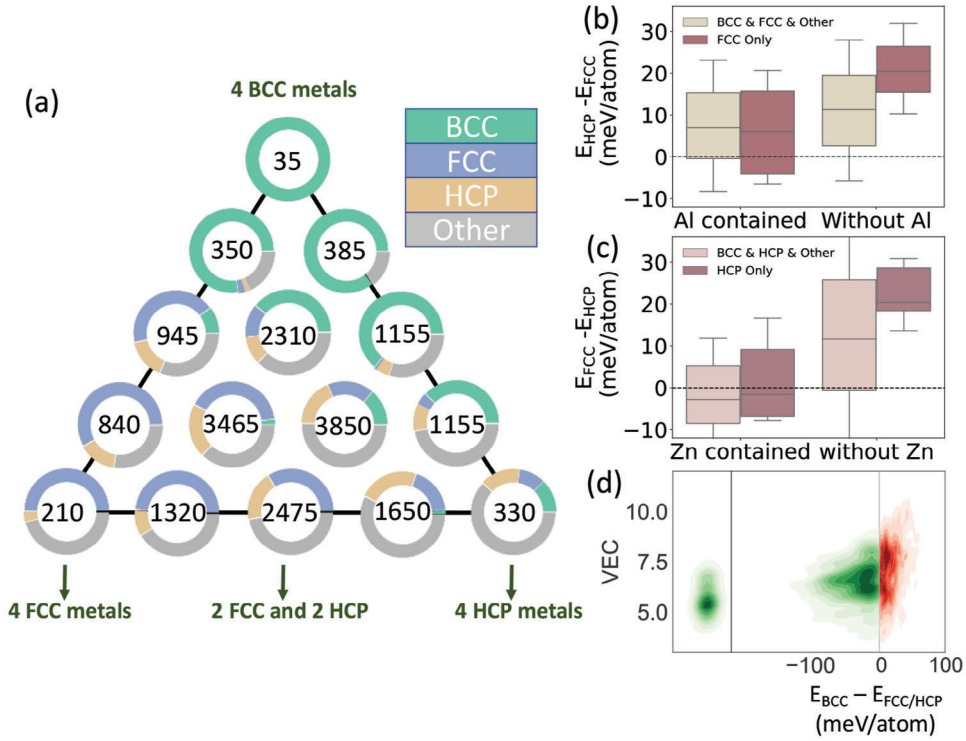

Figure 3. a) The donut plot triangle that shows the amount of lowest energy structural phase distribution with different constituents of elements. The lowest energy structural phase of constituent elements is used to distribute the donut plots in the triangular axis; b) Demonstration of the impact of Al on energy difference between HCP and FCC; c) Demonstration of the impact Zn on energy difference between FCC and HCP; d) Demonstration of the impact of valence electron concentration (VEC) on stabilizing BCC as the ground state. The green region indicates MPEAs that favor BCC phase, whereas red region indicates MPEAs that favor FCC or HCP phase. For MPEAs with all BCC, FCC, and HCP relaxed into BCC structure, an arbitrary −250 was given to emphasize they are stable in BCC phase.

2.3. Elemental Origin of Phase Selection

A triangle plot (**Figure** 3a) is developed to show the influence of elemental phase ground state on the phase selection of the quaternary MPEAs. (ternary and binary alloys are shown in Figure S1 in the Supporting information). In this plot, the triangular axes represent the fractions of FCC, HCP, or BCC elemental ground states for the constituent elements of the MPEA. The three vertices of the triangle illustrate the quaternary MPEAs that have four FCC, HCP, or BCC elements respectively, while the three edges indicate the cases when quaternary MPEAs have a mixing of FCC/HCP elements (bottom edge), mixing of FCC/BCC elements (left edge) and mixing of BCC/HCP elements (right edge). Additionally, the three-donut plot in the middle indicates the compositions that simultaneously contain BCC, FCC, and HCP elements. It should be noted that Sn and Mn do not have their ground states being other phases than the three considered phases, in such analysis, we label them by one of the three phases that shows the lowest energy, e.g., Mn as HCP and Sn as FCC respectively.

Each of the donut plots in Figure 3a has demonstrated the amount of lowest energy structures with color. Here is a summary of all interesting observations: (1) When MPEA constitutes of only FCC metals, FCC, HCP or "Other" group shows lowest energy but not BCC group; (2) When MPEA is only formed by HCP metals, all four groups of structures are observed to show lowest energy; (3) When MPEA is only formed by BCC elements, the lowest energy structure always be BCC; (4) When there are at least 2 BCC metals in the quaternary MPEA, BCC is the mostly likely phase to form statistically. Such trend is absent for FCC and HCP elements.

To gain insights into the underlying physics of the aforementioned observations, comparisons were made with controlled groups of compounds, as depicted in Figure 3b–d. These comparison groups designed and shown in Figure 3b are to understand the phase stabilization mechanism for HCP phase, particularly when FCC elements dominate the composition. The energy difference between HCP and FCC phase ($E_{HCP} - E_{FCC}$) is used as the y-axis for the plot, while we have developed two groups of MPEAs distributing in *x*-axis. The "BCC & FCC & Other" group includes all MPEAs without HCP metals in their constituent metals, while the "FCC only" group includes MPEAs that have only FCC metals as their constituent metals. It can be inferred from Figure 3b that even if Al is FCC metal, the introduction of Al to MPEAs in both groups will lead to a significant drop in $E_{HCP} - E_{FCC}$. This indicates that the presence of Al in MPEA can energetically favor HCP phase. Previous simulations and experiments also demonstrated that the addition of Al reduces the stacking fault energy between HCP and FCC, and the formation of HCP structure has been observed.[40] In our calculations, 326 MPEAs in the two groups even show neglectable or negative $E_{HCP} - E_{FCC}$ (tabulated in Table S2 in the Supporting Information). All these

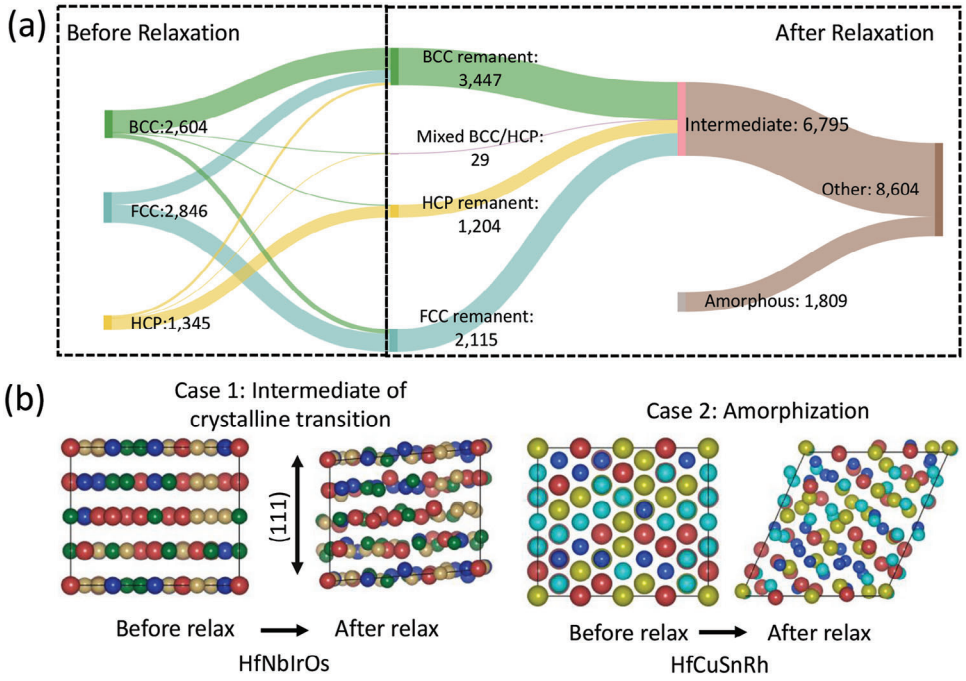

Figure 4. a) The Sankey diagram that demonstrates the relaxation outcome of all lowest energy structures that cannot be mapped back to BCC/FCC/HCP; b) Two examples of the relaxation outcome. The left panel illustrates the case when a structure will become remnants along BCC → HCP transition. The right panel illustrates the case when a structure will be amorphous like after relaxation.

observations lead to the conclusion that Al acts as an "HCP stabilizer" when cooperating with other elements.

To investigate the stabilization mechanism for FCC phases, especially when such MPEA is made from mostly HCP elements as indicated in the lower right section of the donut triangle in Figure 3a, the energy difference between the FCC and HCP phase ($E_{FCC} - E_{HCP}$) is plotted on the *y*-axis in Figure 3c. Similar to the case in Figure 3b, two groups of MPEAs were developed for comparison. The "HCP Only" group includes all the alloys only HCP metals. The "BCC & HCP & Other" group includes all other alloys with no FCC elements. In both selected groups of MPEAs, the presence of Zn tends to result in lower $E_{FCC} - E_{HCP}$ values, which indicates "Zn" could act as an FCC stabilizer. Moreover, it has been found that 135 alloys show negative $E_{FCC} - E_{HCP}$, which is tabulated in Table S3 in the Supporting Information.

To comprehend the stabilization mechanism of BCC phase, the valence electron concentration (VEC) is identified as a good descriptor. We have plotted the energy difference $E_{BCC} - E_{FCC/HCP}$ (the lower energy between FCC and HCP phase was picked) as a function of VEC in Figure 3d. The 2684 alloys that exhibit a negative value for $E_{BCC} - E_{FCC/HCP}$ have been indicated by green. It is worth mentioning that there are 798 MPEAs with all three initial structures (BCC, FCC, and HCP) relaxed into BCC structure. These MPEAs do not possess values for $E_{BCC} - E_{FCC/HCP}$ as FCC and HCP are mechanically unstable. Therefore, an arbitrary value of −250 meV/atom is assigned to represent this group of MPEAs, depicted with the same plotting style in the left panel of **Figure** 4d with green color. It can be observed in Figure 4d that MPEAs with lower VEC in general tend to stabilize BCC phase, whereas the ones with higher VEC tend to stabilize at FCC/HCP phase. However, it is also worth mentioning that VEC does not work perfectly for distinguishing BCC MPEAs from the rest, as many MPEA with $E_{BCC} - E_{FCC/HCP} > 0$ also show comparably low VEC as shown in both Figure 4d and Note S2 in the Supporting Information.

2.4. Understand the Nonclassical Ground State

To form more insights on the materials categorized into "Other" group, we developed structure-matching criteria to further classify those structures. For classifying the relaxed structures back to BCC/FCC/HCP and "Other" group, we set a strict tolerance to map structures back into BCC/FCC/HCP phases (Table S1, Supporting Information). The outliers that are grouped as "Other" group can be further classified by extending the tolerance in off-lattice displacement (stol) and lattice vector length (ltol). The Sankey diagram shown in Figure 4a provides an overview of the classification process of "Other" group. To form such classifications, we gradually extend the stol and ltol values while regrouping the structures based on the threshold that maps them back into the three basic structures (BCC/FCC/HCP). Initially, we extend stol value up to 0.5, while maintaining ltol at the minimum threshold. Consequently, part of the structures in the "Other" group can be mapped into three basic structures (BCC/FCC/HCP). These structures can be regarded as a remnant of the initial structure, as they largely preserve the crystalline symmetry, albeit with significant off-lattice displacement. An illustration of such a structure (HfNbIrOs) is presented in the left panel of Figure 4b. Subsequently, we attempt to map the remaining structures by extending ltol to 0.5 while keeping

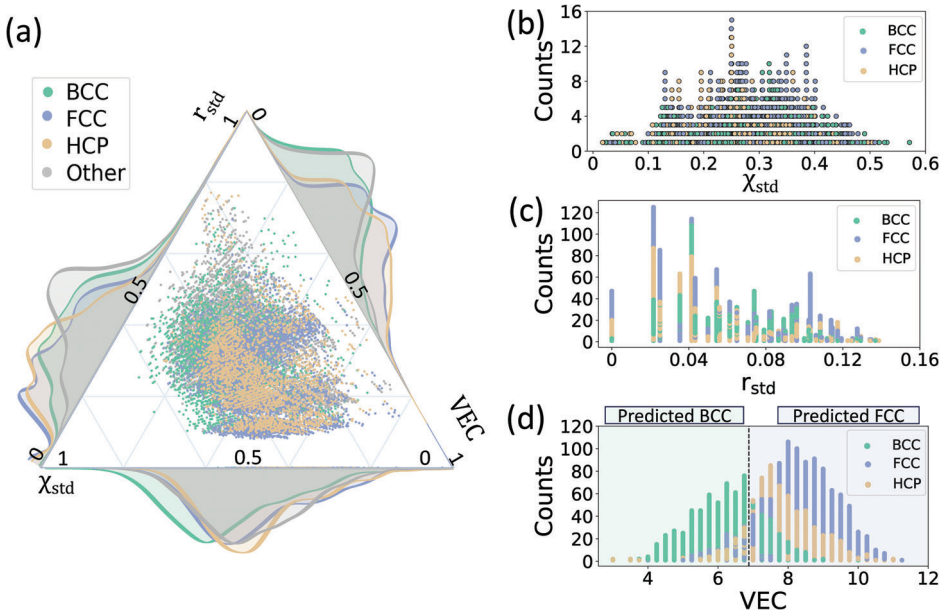

Figure 5. a) The distribution of lowest energy phase as a 3D function of normalized radius, electronegativity, and valence electron concentration plotted in the style of Van Arkle-Ketelaar (VAK) triangle. The normalized details are presented in Note S3 in the Supporting Information; The histogram distribution of normalized data in each axis is also presented at three axes of this triangular plot; b) The distribution of 2395 validation data for decision tree training as a function of standard deviation of electronegativity (χ_{std}), all data are classified as FCC; c) The distribution of 2395 validation data for decision tree training as a function of standard deviation of atomic radius (r_{std}), all data are classified as FCC; d) The distribution of 2395 validation data as a function of VEC, the dashed lines are decision boundaries that separate BCC with FCC from decision tree classification.

stol at its minimum value of 0.16. However, no structure can be mapped back to the BCC/FCC/HCP. Ultimately, by extending both stol and ltol values to 0.5, all the remaining structures can be mapped back into BCC/FCC/HCP. These structures can be regarded as intermediates during amorphization process, as they completely lose their lattice symmetry and periodicity. One example of such a structure is demonstrated in the right panel of Figure 4b.

By applying the classification method mentioned above, 3447, 1204, and 2115 structures are classified as remnants of BCC, HCP, or FCC phase respectively. Additionally, after increasing the stol value, 29 structures were found to match both the BCC and HCP phases. These structures can be regarded as intermediate states of BCC ↔ HCP transformation. The presence of such an intermediate can be explained by the transformation mechanism shown in Figure 2c. It suggests that BCC ↔ HCP transition involves a small energy barrier for shifting atomic layers (Figure 2c), causing DFT relaxation to stop at only an intermediate state that is local energy minimum. Conversely, no structure is observed to map simultaneously to both BCC and FCC or FCC and HCP. This can be understood again by referring to the transformation mechanism demonstrated in Figure 2c: the BCC ↔ FCC transition can occur without a barrier, allowing the DFT relaxation to directly transform the structure; while the FCC ↔ HCP transition requires significant sliding of atoms (large barrier), which inhibits the formation of intermediate states. Lastly, 1809 structures are identified as amorphous intermediates, suggesting that these compositions may not crystallize as a single phase or will crystallize into a completely different phase unrelated to BCC/FCC/HCP.

3. Discussion

3.1. Van Arkle-Ketelaar Triangle of Phase Selection

In addition to the data mining and data analysis of our dataset, we also want to form phenomenological design principles for phase selection in MPEA. Three atomic properties, e.g., standard deviation of atomic radius (r_{std}), standard deviation of electronegativity (χ_{std}), and valence electron concentration (VEC) were selected as the three basic descriptors that are responsible for phase stability. The selection of r_{std} and χ_{std} is based on the intuition of Hume–Rothery rule,[41,42] while VEC is widely used for phase classification of high entropy alloys.[14] **Figure 5a** presents a plot following the fashion of Van Arkle-Ketelaar (VAK) triangle that illustrates the lowest energy structure for each composition. The triangular axes represent the three elemental properties, (e.g., r_{std}, χ_{std}, VEC), while the distributions of the lowest energy phases are plotted (details in Note S3 in the Supporting Information). In addition to this plot, we also include probability distribution of four types of structures as 1D functions of each descriptor, on top of three edges of the triangular plot and Figure S3 in the Supporting Information. From Figure 5a, it is evident that the data points representing four groups of phases are intricately intertwined. Although there appears to be preference of r_{std}, χ_{std}, and VEC, neither Hume–Rothery rules nor VEC alone can effectively separate the four groups of MPEAs.

To assess the predicting capability of the 1D descriptors in determining different phases, we applied the decision tree algorithm[43] on these descriptors individually. For all classification models, the training set consisted of 80% of the quaternary alloys with the lowest energy phases classified as BCC, FCC, or

HCP. The decision tree algorithm is used to identify the boundary, whereas the rest 20% were used to evaluate the performance using the f1 score (Note S4, Supporting Information), the classification metrics are presented in Table S4 in the Supporting Information. The results revealed that when using χ_{std} and r_{std} as descriptors, the decision tree model[43] classified all MPEAs into the FCC phase, with the distribution shown in Figure 5b,c and Figure S4a,b in the Supporting Information. This indicates that the Hume–Rothery rule[41,42] alone cannot be used for predicting phase preference. On the other hand, the decision tree modeling[43] utilizing VEC as a descriptor could reasonably differentiate BCC and FCC phases, with a boundary value of VEC = 6.88. However, such a model will predict all HCP phase into either BCC or FCC (Figure 5d; Figure S4c, Supporting Information) to maximize f1 score. To conclude, none of these three basic descriptors proved to be effective tools for accurately predicting the formed phase of MPEAs.

Moreover, it is important to consider the significant roles that temperature and lattice distortion play in determining the selection of phases. In particular, temperature contributes to both vibrational and configurational entropy in MPEAs. Nevertheless, the impact of configurational entropy is generally negligible (except when short-range order varies) since we are primarily discussing phase stability within the same composition. At the same time, the influence of vibrational entropy is quite mysterious and challenging, often regarded to be relatively limited on changing phase stability but there also have been abnormal observations.[44] As for lattice distortion, two models used for characterizing lattice distortion, an average off-lattice displacement (shown in Figure S5a in the Supporting Information) and the atomic size difference parameter δ[16] (shown in Figure S5b in the Supporting Information) are used to characterize distortion. The difficulty to obtain average off-lattice distortion from both DFT relaxation and experimental measurement as well as the unclear boundaries among four ground state structures limit lattice distortion in working as a practical descriptor.

3.2. Symbolic Machine Learning Derived Phase Selection Rules

To develop a new phenomenological theory that can effectively identify the preferred phase for MPEAs, we combined sure independence screening (SIS)[45–47] with the decision tree algorithm[43] to explore a better phenomenological model for searching more meaningful phenomenological descriptor. We performed an 80–20 stratified split of 11871 quaternary alloys data set that are labeled as BCC, FCC, or HCP. Six basic atomic properties (average and standard deviation of VEC, radius and electronegativity, e.g., VEC, VEC_{std}, r_{ave}, r_{std}, χ_{ave}, χ_{std}) are considered together along with twelve mathematical operators ("+", "−", " × ", "÷", "exp", "log", 'x^{-1}', "x^2", 'x^3', "x^{-2}", 'x^{-3}', "$|x|$") to generate a total of 54468418 candidate descriptors using SIS. To make the descriptor physically meaningful, we applied dimensionless treatment to all six basic descriptors (details in Note S5 in the Supporting Information). The top 50 000 1D descriptors generated by SISSO were selected to find a boundary using a decision tree model. The f1 score and boundaries among the three-phase classifications were determined by the decision tree. The best descriptor, τ, with the highest f1 score was selected, with the formula defined as: $\tau = VEC \times (VEC_{std} + VEC) + \sqrt[3]{\log(\chi_{ave})}$, which is the first machine-learnt phenomenological model for MPEA phase selection. According to this descriptor, the stability regions for the three phases are as follows: BCC is stable when $\tau < 63.31$, HCP is stable when $63.31 < \tau < 81.21$, and FCC is stable when $\tau > 81.21$.

The validation dataset consisting of 2375 validation data, the widely used VEC descriptor successfully labeled 72% of the BCC phase and 56% of the FCC phase correctly but did not identify any samples as HCP phase. On the other hand, our descriptor τ correctly identified 74% of the BCC phase, 72% of the FCC phase and 46% of the HCP phase. The notable improvement in predicting the HCP phase, with precision increasing from 0 to 0.46 and recall increasing from 0 to 0.48, indicates the effectiveness of our new descriptor. Additionally, the overall f1 score increased from 0.47 to 0.63 when compared with VEC rules (all classification metrics are listed in Table S4 in the Supporting Information). These results demonstrate that our new descriptor can be serve as an effective phenomenological model for predicting thermodynamically favored phases of MPEAs.

The symbolic machine learning identified descriptor can be interpreted as a combination of impacts from VEC and electronegativity. **Figure 6c** shows the term of $\sqrt[3]{\log(\chi_{ave})}$ coupled with $VEC \times (VEC + VEC_{std})$ can effectively separate most of the BCC phases from HCP phases, while $VEC \times (VEC + VEC_{std})$ can effectively separate FCC from BCC phases (Figure 6d). Both terms are important as $\sqrt[3]{\log(\chi_{ave})}$ alone cannot separate FCC from HCP (Figure S6a, Supporting Information). This implies that the phase preference between BCC and HCP of the same VEC is primarily governed by average electronegativity (Figure S6b, Supporting Information), while the phase selection between FCC and HCP is predominantly related to the valence electron concentration.

At the same time, to avoid the possible misclassification induced by DFT calculation error, we introduced an energy boundary of 10 meV/atom between the most and second stable phases. If the energy difference exceeds 10 meV/atom, it is considered that the DFT calculations can capture the experimental ground state more reliably. We further refine the validation dataset by including only MPEAs with the $E_{lowest} - E_{second\_lowest} < -10$ meV/atom. The reduced dataset consisted of 1002 samples, and the predicted metrics are shown in Table S5 in the Supporting Information. The f1 score remained similar to the case with complete validation data, which validates that our phenomenological should not be influenced by potential DFT errors and will be a reliable predictor for experiments.

To attempt better performance, we also expanded the descriptor to two dimensions, as detailed in Note S6 in the Supporting Information. The best 2D descriptor identified was (τ, $VEC + VEC^2 + \sqrt[3]{\log(\chi_{ave})}$) (Figure S7, Supporting Information), with an overall f1 score 0.66 on the 2395 validation data set. The classification metric of the 2D descriptor can be found in Table S6 in the Supporting Information. This is regarded as only a marginal increase in performance but significantly increases the complexity of both predicting model and decision boundary. Therefore, we conclude that the 1D feature τ, and 1D decision boundary are the best phenomenological model we have found.

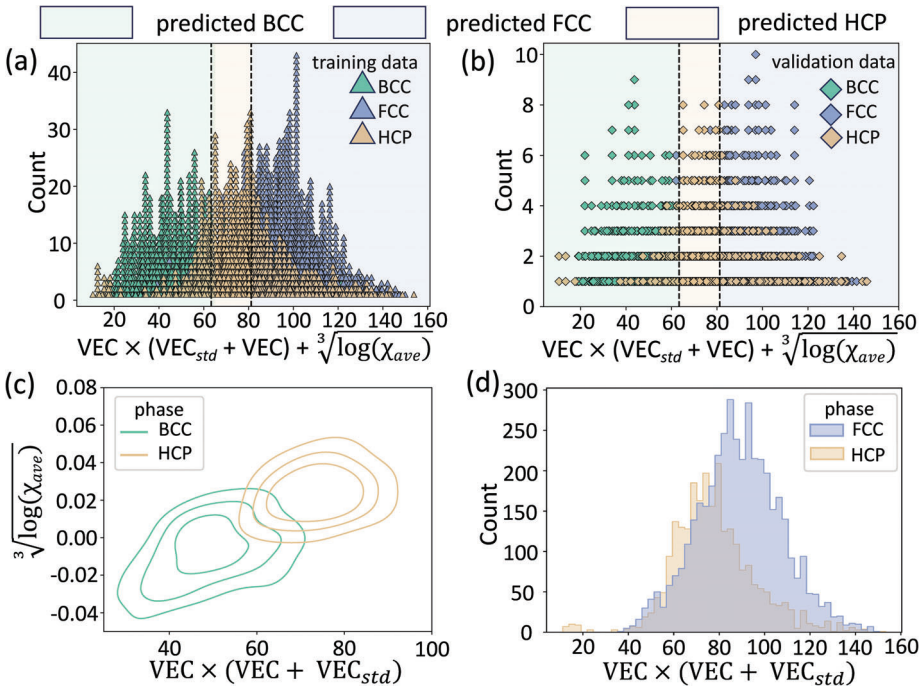

Figure 6. In (a) and (b), a decision tree classifier determines the boundary between BCC and HCP phase, and between HCP and FCC phase. The green, blue and yellow shaded regions are the predicted space for BCC, FCC and HCP phase. a) The distribution of all the training data as a function of descriptor τ. b) The distribution of all the validation data as a function of descriptor τ. c) The distribution MPEAs with BCC and HCP phase as a function of VEC × (VEC + VEC$_{std}$) and $\sqrt[3]{\log(\chi_{ave})}$, showing the data in the central region with higher density. d) The distribution MPEAs with FCC and HCP phase as a function of VEC × (VEC + VEC$_{std}$).

3.3. Phase Selections Rules and Design Strategy

To design materials with specific crystal structures, different strategies can be employed as follows. The mixing of BCC metals will always prefer BCC structure energetically. There is no exception observed from our calculation. At the same time, BCC can always be stabilized in the lower VEC region. Therefore, to design a BCC MPEA composition, a straightforward way is to ensure enough BCC elements in the compounds and lower the VEC.

Notably, the strategy for designing MPEA with HCP structure is to contain more natural HCP elements and avoid Zn which works as an FCC stabilizer. Among all HCP metals, Os, Ru, Re, and Co are the top recommended elements, with 1084, 983, 890, and 673 occurrences, respectively, among 3808 binary, ternary, and quaternary alloys with HCP structures. Similarly, the design of MPEA with FCC structure is to combine more natural FCC elements (prioritize elements like Au, Pt, and Pd, with 1640, 1578, and 1553 compositions containing these elements among 6088 computed binary, ternary, and quaternary FCC alloys) and avoid Al (HCP stabilizer).

Most importantly, our phenomenological descriptor from symbolic machine learning can work as a predictive tool that integrates all the above design principles, as it picks up the influences of valence electron concentration (VEC) and electronegative while forming a phenomenological model that can be used by any MPEA to estimate stability with just plug of VEC and electronegativity.

4. Conclusion

We have successfully developed and presented the first comprehensive dataset encompassing all possible equal-molar multiple principal element alloys (MPEAs). This dataset serves as a valuable resource for studying MPEAs and understanding their phase selection behavior. Furthermore, we have conducted a detailed interpretation of the phase selection rules governing these alloys. In addition, we have employed symbolic machine learning techniques to propose a phenomenological model that outperform significantly of the commonly used valence electron concentration criterion. Our model not only achieves better performance in predicting phase preferences but also uncovers new insights into the underlying physics of phase selection in MPEAs.

5. Experimental Section

First-principles total energies calculations were performed with the Vienna ab initio simulation package (VASP) using a plane-wave basis set.[48] Projector augmented-wave[49,50] potentials with a kinetic energy cutoff of 520 eV and the exchange-correlation form in the Perdew–Burke–Ernzerhof generalized gradient approximation (GGA-PBE)[51] were employed in all the structural optimizations and total-energy calculations. For all the calculations, a reciprocal space discretization of 25 k-points per Å^{-1} was applied, and the convergence criteria were set as 10^{-6} eV for electronic iterations and 0.02 eV Å^{-1} for ionic iterations.

Special quasi-random structures (SQS)[52] were used to represent the fully disordered states of all alloys studied. In terms of composition, all possibilities of binary, ternary, and quaternary compositions were enumerated with equal-molar concentrations with from 28 selected metals, with

atomic configurations shown in Figure S8 in the Supporting Information. That leads to 378 binary alloy compositions, 3276 ternary alloy compositions and 20475 quaternary alloy compositions. For each of the alloy composition, the structure models of BCC, FCC, and HCP, respectively, were created. The three lattice vectors for supercell are [111], [1$\bar{1}$0], and [11$\bar{2}$] for BCC and FCC structure respectively. For both BCC/FCC structures, 2 × 4 × 2 supercell of the defined unit cell was created, while for HCP structure, a 4 × 4 × 3 supercell of the conventional HCP cell was used. There were 96 atoms for all three types of crystal structures.

Supporting Information

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

Acknowledgements

The authors acknowledge support from the startup funding from Florida State University. The Computational resources were provided by the Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support (ACCESS), the National Energy Research Scientific Computing Center (NERSC), a DOE Office of Science User Facility supported by the Office of Science and the U.S. Department of Energy under contract no. DE-AC02-05CH11231 and Research Computing Center (RCC) at Florida State University. The computation and data processing were also supported by the supercomputing resources from the Department of Energy's Office of Energy Efficiency and Renewable Energy at the National Renewable Energy Laboratory.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Keywords

multiprincipal element alloys, phase selection, symbolic machine learning

Received: August 4, 2023
Revised: October 15, 2023
Published online:

[1] E. P. George, D. Raabe, R. O. Ritchie, *Nat. Rev. Mater.* **2019**, *4*, 515.
[2] O. N. Senkov, J. D. Miller, D. B. Miracle, C. Woodward, *Nat. Commun.* **2015**, *6*, 6529.
[3] B. Cantor, I. T. H. Chang, P. Knight, A. J. B. Vincent, *Mater. Sci. Eng., A* **2004**, *375*, 213.
[4] J.-W. Yeh, S.-K. Chen, S.-J. Lin, J.-Y. Gan, T.-S. Chin, T.-T. Shun, C.-H. Tsau, S.-Y. Chang, *Adv. Eng. Mater.* **2004**, *6*, 299.
[5] L. Weidong, X. Di, L. Dongyue, Z. Yong, G. Yanfei, K. L. Peter, *Prog. Mater. Sci.* **2021**, *118*, 100777.
[6] G. A. Salishchev, M. A. Tikhonovsky, D. G. Shaysultanov, N. D. Stepanov, A. V. Kuznetsov, I. V. Kolodiy, A. S. Tortika, O. N. Senkov, *J. Alloys Compd.* **2014**, *591*, 11.
[7] W. Woei-Ren, W. Wei-Lin, W. Shang-Chih, T. Yi-Chia, L. Chun-Hui, Y. Jien-Wei, *Intermetallics* **2012**, *26*, 44.
[8] O. N. Senkov, S. V. Senkova, D. B. Miracle, C. Woodward, *Mater. Sci. Eng., A* **2013**, *565*, 51.
[9] H. Shuo, L. Wei, L. Xiaoqing, S. Stephan, B. Lars, H. Erik, V. Lajos Károly, V. Levente, *Mater. Des.* **2016**, *103*, 71.
[10] M. S. Lucas, L. Mauger, J. A. Muñoz, Y. Xiao, A. O. Sheets, S. L. Semiatin, J. Horwath, Z. Turgut, *J. Appl. Phys.* **2011**, *109*, 07E307.
[11] A. A. B. Thomas, K. P. Jack, H. W. Simon, E. C. Ivano, W. J. Karsten, R. Jan, *Joule* **2019**, *3*, 834.
[12] S. Nellaiappan, N. K. Katiyar, R. Kumar, A. Parui, K. D. Malviya, K. G. Pradeep, A. K. Singh, S. Sharma, C. S. Tiwary, K. Biswas, *ACS Catal.* **2020**, *10*, 3658.
[13] O. N. Senkov, J. D. Miller, D. B. Miracle, C. Woodward, *Calphad* **2015**, *50*, 32.
[14] N. Guo, J. Wei, Q. Shu, Y.i Jia, Z. Li, K. Zhang, H. Zhu, K. Wang, S. Song, Y. Xu, D. Wu, *J. Appl. Phys.* **2011**, *109*, 103505.
[15] S. Yang, G. Liu, Y.u Zhong, *J. Alloys Compd.* **2022**, *916*, 165477.
[16] Y. Zhang, Y. J. Zhou, J. P. Lin, G. L. Chen, P. K. Liaw, *Adv. Eng. Mater.* **2008**, *10*, 534.
[17] Y. Zhang, Z. P. Lu, S. G. Ma, P. K. Liaw, Z. Tang, Y. Q. Cheng, M. C. Gao, *MRS Commun.* **2014**, *4*, 57.
[18] F. Shoushi, X. Xueshan, X. Lei, L. Weihuo, D. Yuanda, *J. Non-Cryst. Solids* **2003**, *321*, 120.
[19] W. Chen, A. Hilhorst, G. Bokas, S. Gorsse, P. J. Jacques, G. Hautier, *Nat. Commun.* **2023**, *14*, 2856.
[20] Z. Zhang, M. Li, J. Cavin, K. Flores, R. Mishra, *Acta Mater.* **2022**, *241*, 118389.
[21] W. Huang, P. Martin, H. L. Zhuang, *Acta Mater.* **2019**, *169*, 225.
[22] N. Islam, W. Huang, H. L. Zhuang, *Comput. Mater. Sci.* **2018**, *150*, 230.
[23] Y. Zhang, C. Wen, C. Wang, S. Antonov, D. Xue, Y. Bai, Y. Su, *Acta Mater.* **2020**, *185*, 528.
[24] Z. Zhou, Y. Zhou, Q. He, Z. Ding, F. Li, Y. Yang, *npj Comput. Mater.* **2019**, *5*, 128.
[25] Y. Li, W. Guo, *Phys. Rev. Mater.* **2019**, *3*, 095005.
[26] G. L. W. Hart, T. Mueller, C. Toher, S. Curtarolo, *Nat. Rev. Mater.* **2021**, *6*, 730.
[27] Z. Pei, J. Yin, J. A. Hawk, D. E. Alman, M. C. Gao, *npj Comput. Mater.* **2020**, *6*, 50.
[28] J. F. Durodola, *Prog. Mater. Sci.* **2022**, *123*, 100797.
[29] X. Liu, J. Zhang, Z. Pei, *Prog. Mater. Sci.* **2022**, *131*, 101018.
[30] K. Lee, M. V. Ayyasamy, P. Delsa, T. Q. Hartnett, P. V. Balachandran, *npj Comput. Mater.* **2022**, *8*, 25.
[31] P. Brown, H. Zhuang, *Mater. Today* **2023**, *63*, 18.
[32] C. Rudin, *Nat. Mach. Intell.* **2019**, *1*, 206.
[33] J. A. Keith, V. Vassilev-Galindo, B. Cheng, S. Chmiela, M. Gastegger, K.-R. Müller, A. Tkatchenko, *Chem. Rev.* **2021**, *121*, 9816.
[34] V. Buhrmester, D. Münch, M. Arens, *Mach. Learn. Knowl. Extr.* **2021**, *3*, 966.
[35] S. Guo, C. T. Liu, *Prog. Nat. Sci.: Mater. Int.* **2011**, *21*, 433.
[36] N. N. Greenwood, A. Earnshaw, *Chemistry of the Elements*, Pergamon Press, Oxford Oxfordshire, NY, USA **1984**.
[37] S. P. Ong, W. D. Richards, A. Jain, G. Hautier, M. Kocher, S. Cholia, D. Gunter, V. L. Chevrier, K. A. Persson, G. Ceder, *Comput. Mater. Sci.* **2013**, *68*, 314.
[38] N. Al-Zoubi, B. Johansson, G. Nilson, L. Vitos, *J. Appl. Phys.* **2011**, *110*, 013708.
[39] Z. Lu, W. Zhu, T. Lu, W. Wang, *Modell. Simul. Mater. Sci. Eng.* **2014**, *22*, 025007.
[40] P. Yu, R. Feng, J. Du, S. Shinzato, J.-P. Chou, B. Chen, Y.u-C. Lo, P. K. Liaw, S. Ogata, A. Hu, *Acta Mater.* **2019**, *181*, 491.

[41] W. Hume-Rothery, H. M. Powell, *Z. Kristallogr. Cryst. Mater.* **1935**, *91*, 23.
[42] W. Hume-Rothery, *Indian J. Phys.* **1969**, *11*, 74.
[43] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, *J. Mach. Learn. Res.* **2011**, *12*, 2825.
[44] M. Esters, C. Oses, D. Hicks, M. J. Mehl, M. Jahnátek, M. D. Hossain, J.-P. Maria, D. W. Brenner, C. Toher, S. Curtarolo, *Nat. Commun.* **2021**, *12*, 5747.
[45] R. Ouyang, S. Curtarolo, E. Ahmetcik, M. Scheffler, L. M. Ghiringhelli, *Phys. Rev. Mater.* **2018**, *2*, 083802.
[46] C. J. Bartel, C. Sutton, B. R. Goldsmith, R. Ouyang, C. B. Musgrave, L. M. Ghiringhelli, M. Scheffler, *Sci. Adv.* **2019**, *5*, eaav0693.
[47] B. Ouyang, J. Wang, T. He, C. J. Bartel, H. Huo, Y. Wang, V. Lacivita, H. Kim, G. Ceder, *Nat. Commun.* **2021**, *12*, 5752.
[48] G. Kresse, J. Furthmüller, *Comput. Mater. Sci.* **1996**, *6*, 15.
[49] G. Kresse, D. Joubert, *Phys. Rev. B* **1999**, *59*, 1758.
[50] P. E. Blöchl, *Phys. Rev. B* **1994**, *50*, 17953.
[51] J. P. Perdew, K. Burke, M. Ernzerhof, *Phys. Rev. Lett.* **1996**, *77*, 3865.
[52] A. Zunger, S.-H. Wei, L. G. Ferreira, J. E. Bernard, *Phys. Rev. Lett.* **1990**, *65*, 353.