Journal of Applied Physics

Machine learning strategies for high-entropy alloys

Cite as: J. Appl. Phys. **128**, 221101 (2020); https://doi.org/10.1063/5.0030367
Submitted: 22 September 2020 . Accepted: 21 November 2020 . Published Online: 10 December 2020

J. M. Rickman, G. Balasubramanian, C. J. Marvel, H. M. Chan, and M.-T. Burton

AIP Publishing

Journal of Applied Physics TUTORIAL scitation.org/journal/jap

Machine learning strategies for high-entropy alloys

Cite as: J. Appl. Phys. **128**, 221101 (2020); doi: 10.1063/5.0030367
Submitted: 22 September 2020 · Accepted: 21 November 2020 ·
Published Online: 10 December 2020

J. M. Rickman,[1,2,a)] G. Balasubramanian,[3] C. J. Marvel,[2] H. M. Chan,[2] and M.-T. Burton[2]

AFFILIATIONS

[1]Department of Physics, Lehigh University, Bethlehem, Pennsylvania 18015, USA
[2]Department of Materials Science and Engineering, Lehigh University, Bethlehem, Pennsylvania 18015, USA
[3]Department of Mechanical Engineering and Mechanics, Lehigh University, Bethlehem, Pennsylvania 18015, USA

[a)]**Author to whom correspondence should be addressed:** jmr6@lehigh.edu

ABSTRACT

The study of high-entropy (HE) alloys has seen dramatic growth in recent years as, in some cases, these systems can exhibit exceptional properties, including enhanced oxidation resistance, superior mechanical properties, and desirable magnetic properties. The identification of promising HE alloys is, however, extremely challenging due to the extraordinarily large number of distinct systems that may be fabricated from the available palette of elements. For this reason, machine learning strategies have been employed to reduce the size of the associated chemistry/composition space. In this review, we outline several computational strategies that have led to the identification of useful alloys and discuss the relative merits and shortcomings of these approaches. We also present short tutorials illustrating the use of selected computational approaches to HE characterization and design.

I. INTRODUCTION

High-entropy (HE) alloys are a new class of materials that, by contrast with conventional alloys, comprise multiple principal elements (typically ≥ 5) and are therefore associated with a larger configurational entropy.[1–3] The initial focus of study in this field was on equimolar, or near equimolar, alloys while, in recent years, a more general term, namely, multi-principal element alloys (MPEAs), has been employed to describe "baseless" alloys having more complex compositions and, in some cases, multi-phase microstructures.[4] Indeed, the initial report on MPEAs by Cantor[5] provided the impetus for unprecedented attention from the research and industrial communities,[6] with the underlying motivation being the exceptional properties exhibited by certain MPEA compositions.[7,8]

As suggested above, the MPEA element-composition space that must be explored to identify promising alloys is vast indeed. More specifically, if one considers a limited palette of, say, 25 readily available elements, then the number of distinct five-element MPEAs one can construct is $\binom{25}{5} 10^4 \approx 5.3 \times 10^8$, assuming a decimation of the composition scale. If one restricts attention to only equimolar alloys, the number is still rather large (over 50 000). Given this prohibitively large space, it is therefore essential to identify first a smaller subspace that contains the most promising candidate alloys prior to any high-throughput investigation or fabrication. The advent of data science methods, known collectively as materials informatics, has enabled workers to reduce effectively the range of systems for exploration and permitted computational studies that identify energetically favorable alloy configurations.

Materials informatics, which is the application of data science to materials science and engineering problems, has emerged as a powerful tool for materials discovery and design.[9] In particular, recent applications of materials informatics to the field of HE alloys have led to the discovery of new alloys having enhanced hardness and a deeper understanding of the connection among materials descriptors and mechanical properties.[10] In addition, Troparevsky *et al.*[11] employed high-throughput density-functional theory (DFT) to highlight those combination of elements that are thermodynamically favored to form complex alloys. Moreover, Senkov *et al.*[12] used a CALPHAD-based combinatorial approach to screen numerous candidate metal alloys for those with a propensity to form solid solutions and Sarker *et al.*[13] proposed using a new entropy descriptor to highlight HE alloys possessing high hardnesses.

In this review, we describe recent machine learning (ML) and other computational strategies that have permitted the

identification of useful HE alloys, which have led to a better understanding of the thermodynamics and kinetics of these systems and guided high-throughput experimental design. In Sec. II, we first present an overview of a number of computational techniques that have been especially useful in this context. In Sec. III, we focus more specifically on a combination of computational methodologies, namely, a canonical correlation analysis (CCA) used in conjunction with multi-objective optimization to down-select promising alloys via a genetic algorithm (GA).[14] This methodology was used to identify successfully MPEAs with superior mechanical properties.[10] In Sec. IV, we describe a versatile and robust nature-inspired metaheuristic algorithm, namely, the cuckoo search (CS) and its variants[15,16] that can be used in conjunction with molecular dynamics (MD) atomistic simulation to optimize elemental compositions for HE alloys with targeted properties. As will be seen below, the CS is a possible alternative to the more conventional GA for performing optimization. Section V contains a brief discussion and outlook. Finally, in the Appendixes, we present short tutorials describing the practical implementations of two techniques featured here, namely, the CCA and CS. Our aim is to provide a practical guide to the use of these techniques to aid in HE alloy design and to discuss their relative merits and shortcomings.

II. METHODOLOGY OVERVIEW

There have been several applications of ML to the analysis of phase formation and property prediction in HE alloys. Many of these approaches are based on a physically motivated set of parameters that depend on the thermo-mechanical characteristics of the elemental constituents of these alloys. These elemental characteristics include the atomic radius, r; the melting temperature, T_m; the Young's modulus, E; and the valence electron concentration, VEC. In addition, for pairs of elements i and j, this list is often supplemented by the mixing enthalpy $(\Delta H_{mix})_{ij}$. For alloys comprising N elements (where in this context typically $N \geq 4$) in which an element i has composition c_i, the relevant parameters include those summarized as follows:[17]

Radius asymmetry, $\delta = \sqrt{\sum_{i=1}^{N} c_i\left(1 - \frac{r_i}{\bar{r}}\right)^2}$,
Enthalpy of mixing, $\Delta H_{mix} = 4\sum_{i=1, j\neq i}^{N} (\Delta H_{mix})_{ij} c_i c_j$,
Ideal entropy of mixing, $\Delta S_{mix} = -R\sum_{i=1}^{N} c_i \ln c_i$,
Mean melting temperature, $T_m = \sum_{i=1}^{N} c_i (T_m)_i$,
Entropy/enthalpy ratio, $\Omega = \frac{T_m \Delta S_{mix}}{|\Delta H_{mix}|}$,
Young's modulus asymmetry, $\mathcal{E} = \sqrt{\sum_{i=1}^{N} c_i\left(1 - \frac{E_i}{\bar{E}}\right)^2}$,
Valence electron concentration, $\mathrm{VEC} = \sum_{i=1}^{N} c_i \mathrm{VEC}_i$.

In these expressions, a bar denotes an average over the available data.

These parameters can serve as candidate features in ML models that predict phase formation and observed properties in HE alloys. For example, Li and Guo[18] proposed a support-vector machine (SVM) for distinguishing among various phases of as-cast alloys. In addition, this approach was employed to search for new alloys, some of which have desirable properties (e.g., refractories). Wen *et al.*[19] used this parameter set as descriptors in the ML-assisted design of HE alloys with desirable mechanical properties. More specifically, the authors combine machine learning with experimental design algorithms to identify alloys in the Al-Co-Cr-Cu-Fe-Ni system having high hardness values. Finally, Zhou *et al.*[20] provided an appraisal of phase design rules that have been employed with various ML algorithms. This assessment was supplemented by the experimental design of a series of alloys in the Fe-Cr-Ni-Zr-Cu system.

Machine learning has also been employed in combination with other techniques to model complex alloys at the atomic level. For example, in a recent computational study,[21] active-learning machine-trained potentials informed by *ab initio* calculations were employed in a Monte Carlo (MC) study of HE alloys. This approach was used to explore the phase stability and probe short-range order in prototypical body-centered cubic (bcc) alloys. In a similar vein, Liu *et al.*[22] developed a learned effective Hamiltonian trained on a NbMoTaW family of HE alloys that provides a good representation of the alloy over a large portion of configuration space. Extending this idea a bit further, Grabowski *et al.*[23] have used machine-learning potentials to compute anharmonic, vibrational free energies using thermodynamic integration. They validated their approach by computing the free energy of a quinary refractory alloy.

The use of ML in conjunction with thermodynamic information embodied in a phase-diagram dataset has also proven useful in navigating the large chemistry/composition space associated with complex alloys. For example, Qi *et al.*[24] devised a procedure in which parameters inspired by experimentally compiled phase diagrams were used in a ML strategy to classify over 600 HE alloys. In particular, they employ a collection of binary phase diagrams to assess the phase stability of multi-component alloys with the assumption that such diagrams encode basic structural and chemical information about more complex alloys. Other approaches leverage existing software for the calculation of phase diagrams, such as CALPHAD. Miracle *et al.*[25] developed a HEA design strategy by combining high-throughput experiments and complementary computations in which a feedback loop facilitates validation. In this approach, CALPHAD is used to predict phase equilibria with the results used to suggest high-throughput experiments on materials libraries. As noted above, Senkov *et al.*[12] also used a CALPHAD-based combinatorial approach to identify promising alloys having targeted properties. They found, somewhat surprisingly, that the prevalence of solid solutions decreases with increasing alloy complexity (i.e., more elements).

III. MULTIVARIATE STATISTICAL ANALYSES AND GENETIC ALGORITHMS

We next provide a more detailed review of two computational methodologies, namely, a CCA and a GA, that have been used recently to identify promising HE alloys and thereby guide experimental alloy design. After a brief description of these techniques, we outline their complementary usage in the down-selection of alloys having superior mechanical properties. Building upon this discussion, Appendix A contains a short tutorial describing the implementation of a CCA.

A. Canonical correlation analysis

In a variety of different contexts, including the characterization of HE alloys, it is of interest to explore systematically

correlations between a set of input (i.e., predictor) variables that embody, for example, structural and chemical information and a set of output variables that describe observable properties (e.g., hardness). For this purpose, a canonical correlation analysis (CCA) is a very general approach for quantifying relationships between these two variable sets. It is especially useful in data analytics as a dimensional reduction technique that identifies a relatively few linear combinations of variables in a complex, multidimensional parameter space that is maximally correlated. Indeed, most parametric tests of significance are, in fact, essentially special cases of CCA,[26] and this approach has been employed recently in other materials contexts to explore, for example, structure/processing correlations in ceramic powder processing.[27,28]

As noted above, the aim of the CCA is to identify those linear combinations of predictor variables that are maximally correlated with a set of outcome variables. This is accomplished via the construction of operators from the blocks of a correlation (or covariance) matrix whose eigenvalues, λ_i ($i = 1, 2, \ldots, nv$, where nv is the number of variate pairs), reflect the degree of correlation between predictor and outcome variables, and whose associated eigenvectors determine the relative importance of the variables.[26,28,29] More specifically, $\sqrt{\lambda_i}$ is the correlation coefficient associated with a given variate i with the associated variability $\lambda_i / \sum_{j=1}^{nv} \lambda_j$. The basic elements of the CCA methodology are outlined below. We note that this analysis has been recently generalized to the case of correlations between non-linear combinations of predictor and outcome variables.[28]

Consider M experiments for which one has a set of input variables $\{x_1, x_2 \ldots, x_{N_{in}}\}$ and a corresponding set of output variables $\{y_1, y_2 \ldots, y_{N_{out}}\}$, where N_{in} (N_{out}) are the number of input (output) variables. The aim of this procedure is to identify combinations (known as canonical variates) $V = \sum_i \alpha_i x_i$ and $W = \sum_i \beta_i y_i$, where α_i and β_i are known as canonical weights, such that these linear combinations are maximally correlated. This is accomplished by constructing a correlation (or associated covariance) matrix $\mathbf{\Sigma}$[29,30] for which one defines $\mathbf{X}$ and $\mathbf{Y}$, with components X_{ij} and Y_{ij}, respectively, where $j = 1, 2, \ldots, M$. The correlation matrix $\mathbf{\Sigma}$ may then be written in block form as

$$\mathbf{\Sigma} = \begin{bmatrix} \mathbf{\Sigma}_{\vec{X}\vec{X}} & \mathbf{\Sigma}_{\vec{X}\vec{Y}} \\ \mathbf{\Sigma}_{\vec{Y}\vec{X}} & \mathbf{\Sigma}_{\vec{Y}\vec{Y}} \end{bmatrix}. \tag{1}$$

The goal of a CCA is to calculate $\{\alpha_i\}$ and $\{\beta_i\}$ such that the correlation between W and V is maximized. Thus, one seeks to maximize the correlation

$$corr(W, V) = \frac{\text{cov}(W, V)}{\sqrt{\text{var}(W)\text{var}(V)}}, \tag{2}$$

where the denominator is the square root of the product of the variances of W and V, respectively. This may be accomplished by finding the eigenvectors and corresponding eigenvalues of two operators, $\boldsymbol{\sigma}_1$ and $\boldsymbol{\sigma}_2$, that are constructed from matrix blocks of $\mathbf{\Sigma}$

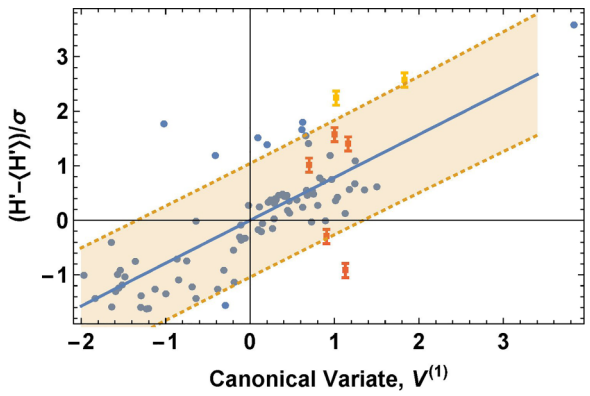

FIG. 1. The canonical variate, $V^{(1)}$ vs the canonical variate, $W^{(1)}$ (a measure of hardness) for each alloy (blue circle) and the corresponding regression line (solid blue). The dotted lines delimit a shaded 90% (single-observation) prediction band. The other colored circles are data for alloys fabricated in our laboratory using a combined CCA/GA strategy. Reproduced from Rickman *et al.*, Nat. Commun. **10**, 2618 (2019). Copyright 2019 Springer Nature.

and given by

$$\begin{aligned} \boldsymbol{\sigma}_1 &= \mathbf{\Sigma}_{\vec{X}\vec{X}}^{-1}\mathbf{\Sigma}_{\vec{X}\vec{Y}}\mathbf{\Sigma}_{\vec{Y}\vec{Y}}^{-1}\mathbf{\Sigma}_{\vec{Y}\vec{X}}, \\ \boldsymbol{\sigma}_2 &= \mathbf{\Sigma}_{\vec{Y}\vec{Y}}^{-1}\mathbf{\Sigma}_{\vec{Y}\vec{X}}\mathbf{\Sigma}_{\vec{X}\vec{X}}^{-1}\mathbf{\Sigma}_{\vec{X}\vec{Y}}. \end{aligned} \tag{3}$$

These operators have the same eigenvalues, and the square roots of these eigenvalues, sqrteig, are the so-called canonical correlations. In addition, the α_i and β_i are constructed from the eigenvectors of these operators. In other words, if $\vec{\alpha}'$ and $\vec{\beta}'$ are eigenvectors of $\boldsymbol{\sigma}_1$ corresponding to its maximum eigenvalue, then the associated canonical variates are maximally correlated.[29] An illustration of the use of this methodology is provided in the short tutorial in Appendix A.

To see how this methodology may be implemented in practice, we sought to quantify correlations among the predictor parameters for HE alloys summarized above and the (output) mechanical properties of these alloys. For simplicity, our initial focus was on the Vickers hardness, H, of the alloys.[10] The CCA employed a database comprising 82 HE alloys for which mechanical property information is available to highlight new alloys having potentially high hardnesses. It was found that one pair of canonical coordinates (labeled 1) evinced a relatively high degree of correlation. Figure 1 displays the values for the predictor and output variates, $V^{(1)}$ and $W^{(1)} = (H' - \langle H' \rangle)/\sigma$, respectively, for each of the alloys and, in addition, a regression line that quantifies the relationship between the variates. (The prime here denotes a value normalized by the reference alloy, the angle brackets denote an average over a dataset comprising the 82 HE alloys and σ denotes the standard deviation in H' over the dataset.) Also shown are data for alloys fabricated in our laboratory using the regression line and a genetic algorithm described in Sec. III B.

B. Genetic algorithm

A genetic algorithm (GA) is a method to obtain the solution to a problem, as represented by an individual having genetic

content, by processes that mimic those central to natural selection. From this point of view, trial solutions (a population) evolve from one generation to the next in a Darwinian competition with survivors determined by their fitness.[31,32] Following this evolutionary analogy, the parameters of a given problem are encoded as genes comprising a "chromosome" and represented by a binary string. From an initial population, offspring are generated by mixing and recombining these genes. From this gene manipulation via processes emulating mating and mutation, "stronger" individuals will be produced in a survival-of-the-fittest competition. It should be noted that while a GA is useful in finding optimal solutions for some problems, it can also be used to identify groups of superior solutions that correspond to an evolved population after many generations. In the current context, a GA is employed to obtain potential candidate alloys where survival to a successive generation depends upon a measure of fitness (e.g., having outstanding mechanical properties).[10] (We note that in Sec. IV an optimization method known as a cuckoo search, a useful alternative to a GA, is presented.)

In the context of the HE alloy design space, a GA has been employed to identify promising alloys having superior properties.[10] More specifically, the CCA regression line described in Sec. III A was used as a fitness function, f_i, for the ith chromosome in a GA to find candidate alloys having high hardness values. The GA proceeds from one generation to the next producing fitter (i.e., harder) offspring (i.e., alloys) by favoring those alloys that result in higher hardness values as determined by the regression line shown in Fig. 1. Thus, alloys associated with large, positive values of the canonical variate are selected. Since these values are extrapolated from existing data, in practice we limit the degree of extrapolation to obtain reliable predictions.

To construct candidate solutions, five-element alloys, a sixteen-element palette and sixteen molar compositions per element were used to represent the independent alloys in the GA. The calculation began with a number (500) of randomly selected chromosomes, each represented by a 40-bit string that encodes alloy chemistry and composition. As outlined above, successive generations were produced via a series of evolutionary processes, and one observes substantial improvement in the overall fitness (i.e., hardness) of the alloys over many generations. Based on this analysis, several alloys were selected for fabrication and the results are also shown in Fig. 1.

IV. METAHEURISTIC SEARCH BASED ON CUCKOO SEARCH OPTIMIZATION

It is important to realize that not all multicomponent alloys can be classified as MPEAs or HE alloys, because they all do not form solid solutions or, more specifically, random solid solutions. Consequently, given the extremely large compositional landscape, most contemporary approaches to complex alloy thermodynamics have resorted to trial-and-error experimentation in the absence of phase diagrams of higher order elemental mixtures. Thus, a key problem in this field is the prediction of specific sets of elements that form solid solution phases that may be subsequently synthesized and characterized. It should be recognized that simulations provide some structural, mechanical, and thermodynamic data for compositions that have been examined experimentally, as well as several other, hypothetical compositions. Thus, data assimilation strategies for a simulational/mathematical predictive scheme must be able to record, manage, and integrate the large volumes of heterogeneous data and diverse sets of results generated via both simulation and experiment. Additionally, for the case of nature-inspired search heuristics for alloy discovery, sweeps across several tens of thousands of simulations are typically required. In short, data analytics leverages applied mathematics and modern computationally intensive strategies for the systematic interrogation and design of MPEAs.

A. Cuckoo search

Another nature-inspired scheme based on a metaheuristic algorithm facilitates the rapid prediction of approximate solutions to intractable or gradient-free optimization problems.[33–36] While genetic algorithm (GA) and particle swarm optimization (PSO) methods are often applied in manufacturing and design contexts,[37–40] the advent of a recently developed technique, namely, the cuckoo search (CS), has led to significant advances in the design of MPEAs with novel combinations of structures and properties. CS is primarily an evolutionary algorithm and is perhaps most useful for optimization problems. It satisfies global convergence requirements[41–43] and has been shown to be superior to PSO and GA using benchmark comparisons. This methodology has both local and global search capabilities that are controlled by a switching probability, permitting a thorough and efficient exploration of the search space; consequently, global optimality can be found with a higher probability. Hence, CS is able to embrace the richness and diversity of the available data and thereby identify optimal MPEA compositions. The preference for CS vis-a-vis PSO and GA is attributed to its superior performance in benchmark comparisons.[44,45] In summary, the relative advantages of CS include (1) a higher probability of global convergence, (2) local and global search capabilities controlled via a switching parameter, and (3) the use of Lévy flights rather than standard random walks to scan the design space more efficiently than a simple Gaussian process.[46]

The cuckoo search takes its name from a biological analogy. More specifically, CS imitates obligate brood parasitism of some female cuckoo species that specialize in mimicking the color and behavioral pattern of certain host birds. A cuckoo lays her eggs (design parameters) in the nests (a generation of possible solutions) of other birds that nurture the eggs as if they are their own. Upon discovering that the egg is not its own (selection criteria), the host may destroy the egg or build a new nest elsewhere. When the young cuckoo hatches, it may destroy the eggs of the host, thereby imposing dominance in that nest.

From a methodological standpoint, a generation consists of a set of possible solutions that are iteratively improved to satisfy the objective functions. The poorest fit solution(s) are eliminated, but together with the best fit ones they are used to generate newer solution(s) by a random modification. When a new solution is of better quality than another randomly chosen one from the current generation, the old solution is replaced with the new one. Thus, the best solutions in each generation successively progress forward in the

optimization process, subject to the constraint that the population of each generation is constant. In addition, in every generation, new solutions are inserted with a probability corresponding to the time spent in searching for the best fit solution (local search) and the remaining time is invested in exploring solutions far away from the current samples (global search). Details of the computational scheme are presented in Fig. 2. This implementation mitigates against the possibility that the CS algorithm converges locally to sub-optimal solutions while ensuring that the predicted designer materials are in accord with physically realizable samples.

In the CS optimization, each egg in a nest represents a solution governed by the following three ideal criteria:[16]

1. Each cuckoo lays one egg in a randomly chosen nest at any given time.
2. Only the nest with the highest quality eggs is carried over to the next generation.
3. The probability that the host bird discovers the cuckoo egg is $0 \leq p_a \leq 1$ for a fixed number of available host nests. If/when discovered, the host bird can either eliminate the cuckoo egg or build a completely new nest.

The aforementioned switching parameter, p_a, controls the selection between a local random walk and a global-explorative random walk. A local random walk is represented as

$$x_i^{(t+1)} = x_i^t + \alpha s \otimes H(p_a - \epsilon) \otimes (x_j^{(t)} - x_k^{(t)}), \tag{4}$$

where ϵ is a random number, $H(u)$ is the Heaviside function, s is the step size, $x_j^{(t)}$ and $x_k^{(t)}$ are two different solutions selected arbitrarily by random permutation, and $\otimes$ is the entry-wise product. By contrast, a global random walk, or Lévy flight, is represented as

$$x_i^{(t+1)} = x_i^{(t)} + \alpha L(s, \lambda), \tag{5}$$

where $L(s, \lambda) = \lambda\Gamma(\lambda)\sin(\pi\lambda/2)/(\pi s^{(1+\lambda)})$ $(s \gg s_0 \gg 0)$ and $\alpha > 0$ is a step-size factor related to the scale of the problem. Γ is the usual gamma function.[47] The convergence rate in CS optimization is insignificantly dependent on the choice of key parameters such as the number of nests (n) and p_a.[40,48] Trial simulations on test functions, such as the Ackley and 6-hump camel back, show $p_a = 0.2$ to be appropriate for most cases where $5 \leq n \leq 100$.[44] This approach limits local search time consumption to about 1/5th of the total time and permits extensive exploration of the global design space within reasonable computing times.

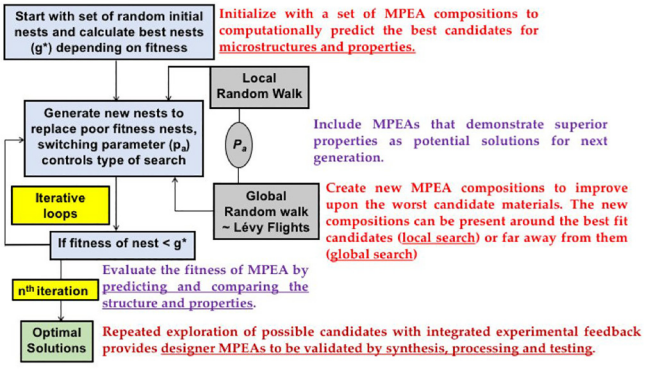

FIG. 2. The design methodology employing an integrated computational/mathematical CS optimization scheme for determining MPEA compositions with targeted structure and properties.

B. Combining a cuckoo search with molecular dynamics

Molecular dynamics (MD) simulations may be embedded within the CS algorithm to, for example, survey an alloy's composition space. The results of such simulations to determine the elemental composition for a set of model MPEAs with targeted properties[16] are presented in Fig. 3. In this work, a CS-MD coupled computational mathematics framework is constructed, implemented, and verified to optimize user-declared design variables (such as the elemental alloy concentration) for the desired property, the ultimate tensile strength.[49–52] Each cycle (generation) of CS comprises the evaluation of different solutions (nests) to preserve the best candidate while replacing all unfavorable solutions with fresher alternatives predicted via the global and local exploratory searches in the design space. For every cycle, only the alloy with the elemental concentration that yields the maximum ultimate tensile strength (UTS) is retained among all the available solutions, namely, the best nest (solution) (g^*).

The convergence rate in CS optimization has been often found to be least dependent on the choice of key parameters, such as n and p_a,[53] and all MPEA simulations under the CS-MD framework described above used $p_a = 0.2$. A comparison of the computational times between the CS and a strictly local search of the design space suggested that the latter consumed more computational resources/time given its relatively higher probability of trapping in local minima. From these considerations, CS was found to be more efficient at exploring a design space at both the local and global search scales.[54,55] We note that the CS-MD code was tested for robustness by varying input concentrations from 2.5% to 97.5%, in increments of 2.5%. Our procedure provided the correct elemental concentrations (input) and the corresponding maximum strengths for all cases tested.

We first considered an application of the CS-MD mathematical framework to the case of a binary Al-Fe alloy, followed by ternary and quinary combinations. The selected optimization parameters for the binary, ternary, and quinary alloys are listed in Table I, including the upper and lower bounds of the design variable for each material. For the binary case, the Al elemental % (atomic) was selected to be the design variable, while the objective function was constructed to lead to an increase in the ultimate tensile strength (UTS) for the alloy. The associated design space illustrated in Figs. 3(a), 3(c), and 3(e) shows the elemental composition (in at. %) of the design variable along the z-axis varying with the number of iterations and number of nests (i.e., the number of solutions considered) displayed along the y- and the x-axes, respectively. The total number of evaluations of the objective function is simply the product of the number of nests and the number of

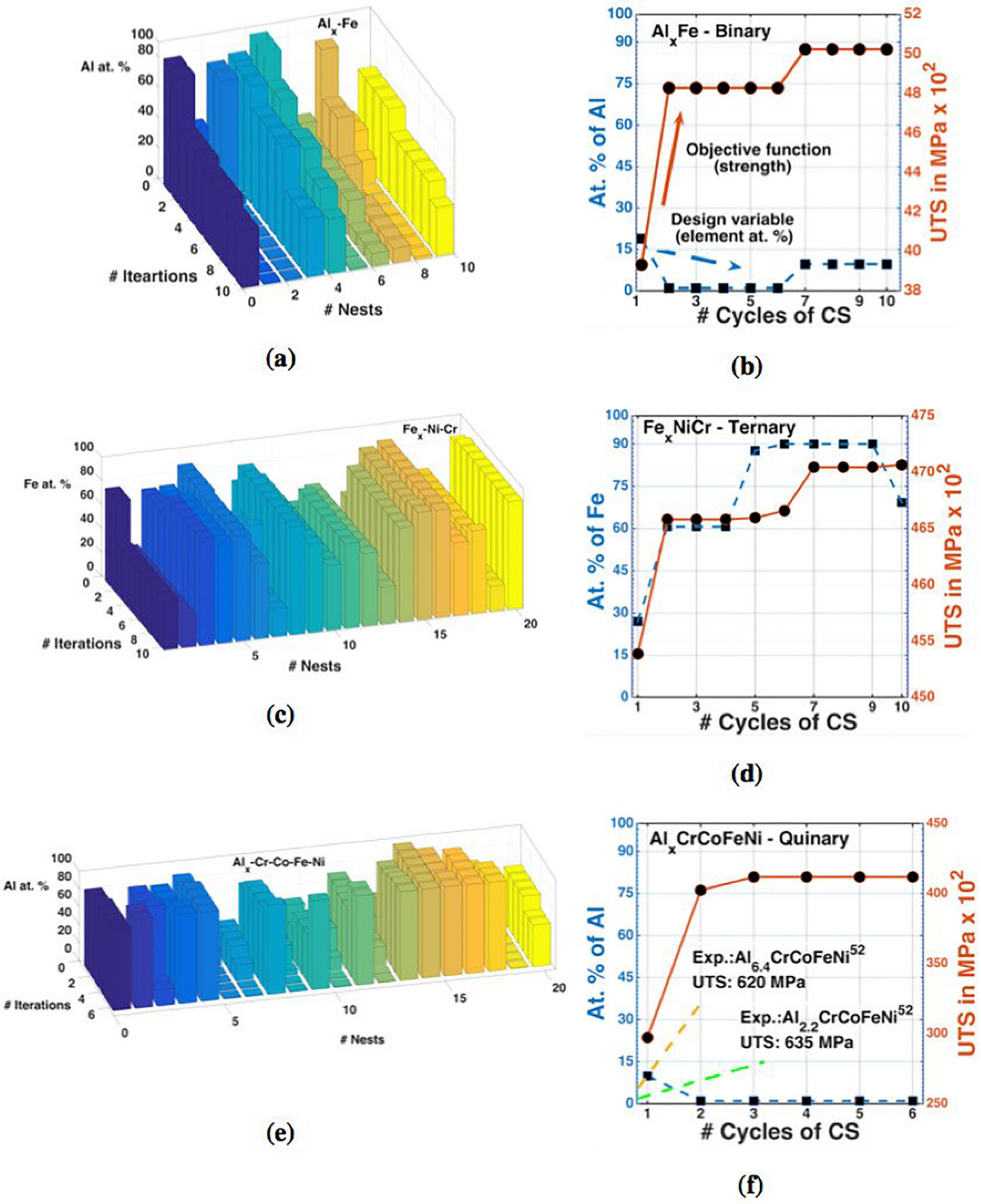

FIG. 3. Design-space-exploration map [(a), (c), (e)], and design variable and objective function (UTS) variation with each cycle of CS [(b), (d), (f)], for different MPEAs—binary Al-Fe, ternary Fe-N-Cr, and quinary Al-Cr-Co-Fe-Ni. Reproduced from Sharma *et al.*, Scr. Mater. **130**, 292 (2017). Copyright 2017 Elsevier.

TABLE I. Optimization parameters used for the different alloys.

Alloy composition	Design variable	Upper bound (at. %)	Lower bound (at. %)	Switching parameter (p_a)	No. of nests	No. of cycles
Binary: AlFe	Al	99	1	0.2	20	10
Ternary: FeNiCr	Fe	90	10	0.2	20	10
Quinary: AlCrCoFeNi	Al	99	1	0.2	20	6

iterations. Thus, the predictive landscape shown in Figs. 3(a), 3(c), and 3(e) represents the exploratory walks performed by the different nests/solutions (20) of the CS algorithm. In this figure, the same colored histograms represent the variation in the design parameter (atomic concentration) values with each iteration of the CS cycle while fluctuations in the histograms during the different iterations indicate that the algorithm is employing different design values to arrive at a global optimum, thereby achieving the desired increased alloy strength. In each iteration, the complete exploration of the design variable by the different cuckoo nests effectively identifies favorable Al elemental compositions leading to strength enhancements of the binary Al_xFe alloy.

The CS-MD results suggest that as the Al composition decreases in an alloy, its strength increases from 4000 MPa for 20% Al to a final value of 5000 MPa for 9 at. % Al for 100 objective function evaluations. Each objective function evaluation involves a complete MD simulation where, for a particular composition, a nanostructure of the alloy is examined by quenching from 2200 K to 300 K, followed by the uniform tensile deformation in the ⟨100⟩ direction. The predictions of the CS-MD approach for the strain deformation of Al_xFe alloys concur with observations in the literature, whereas an increase in Fe composition promotes higher tensile strength in quenched alloys.[56–60] The same approach was employed to understand deformation in ternary and quinary multicomponent alloys. The results of the computational scheme reveal the correlation between the concentration of a single element (design variable) and UTS (objective function) that are qualitatively in agreement with earlier experimental measurements. In summary, this technique accelerates the selection of elements and compositions for a MPEA with desired structures and properties and provides a robust computational framework for exploring the vast materials landscape for such complex structures.

C. A hierarchy of increasingly complex optimization problems

In the simplest case, as outlined above, the goal is to determine a set of MPEA compositions that satisfy the objective functions associated with structural or mechanical properties, such as elastic constants or atomistic elemental distributions, where both are obtained from simulations, such as MD. The optimization framework highlights possibly useful compositions by iteratively predicting newer solutions that are assessed via simulation. Those samples having superior performance may then be synthesized and characterized experimentally. The method is able to treat additional complexities by, for instance, also incorporating temperature and grain-size distributions of the processed samples to construct a multi-modal optimization problem to interrogate various sets of MPEAs that are physically reproducible and measurable. Thus, with this mathematical framework, one can incorporate several different constraints and parameters to achieve designer MPEAs with targeted material phases as well as desired structures and thermo-mechanical properties. While many optimization heuristics typically encounter issues in fitting into design exploration paradigms given their unpredictability and randomness, the CS algorithm is not plagued by these drawbacks and can be efficiently employed to design MPEAs from heterogeneous datasets. In Appendix B, we present a brief tutorial on the practical application of the CS method to materials optimization problems.

D. Improving global convergence by adaptive cuckoo search (ACS)

We conclude this discussion by identifying algorithmic improvements that accelerate convergence. The m-dimensional search space for the ith cuckoo can be written as $X_i = (x_i^1, x_i^2, \ldots, x_i^m)$ for $i = 1, 2, \ldots, N$. For this case, at time t, the new search space for the ith cuckoo is

$$X_i(t+1) = X_i(t) + \alpha \,\text{Lévy}(\lambda), \tag{6}$$

where

$$\text{Lévy}(\lambda) = \left| \frac{\Gamma(1+\lambda)\,\sin(\pi\lambda/2)}{\lambda\, 2^{(\lambda-1)/2}\,\Gamma\left(\frac{1+\lambda}{2}\right)} \right|^{(1/\lambda)}, \tag{7}$$

and where α is the step-size that sets the scale of random search patterns and facilitates exploring deviations away from the local optima via effective far-field randomization. In a typical CS algorithm, α is assigned a constant value (usually 1) and $1 < \lambda < 3$. The Lévy process is a random walk comprising a series of instantaneous jumps chosen from a heavy-tailed probability density function. More specifically, a random walk of the next generation depends on the current one, as derived from a Markov chain and the Lévy distribution.[61]

The standard CS technique lacks the control over the step-size that is required to reach global maxima or minima.[61,62] By linking the step-size with the fitness of an individual nest in the search space, one can essentially remove the dependency on α. The required relation between the fitness and the step-size for generation t is given in terms of the fitness of the ith nest in the tth

generation, $\mathrm{fit}_i(t)$, by

$$\mathrm{Step}_i(t+1) = \frac{1}{t}^{\epsilon(t)}, \tag{8}$$

where

$$\epsilon(t) = \left|\frac{\mathrm{bestfit}(t) - \mathrm{fit}_i(t)}{\mathrm{bestfit}(t) - \mathrm{worstfit}(t)}\right|, \tag{9}$$

and where bestfit(t) [worstfit(t)] is the best (worst) fitness value in the tth generation. Thus, the step-size is adaptively determined from the fitness value such that the step-size decreases as the number of cycles increases. The governing equation for the search space exploration for this adaptive CS[61] becomes then $\mathrm{X}_i(t+1) = \mathrm{X}_i(\mathrm{t}) + \mathrm{randn}\ \mathrm{Step}_i(\mathrm{t}+1)$, where randn is a uniformly distributed random number on [0,1). In summary, the main advantages of the ACS implementation result from (1) eliminating the dependency of the predictions on the step-size and the attendant parameter-independent optimization and (2) potentially accelerating the global convergence rate relative to that of a traditional CS implementation.

V. OUTLOOK AND DISCUSSION

We have described here several techniques that have proven useful in the prediction and subsequent characterization of HE alloys. More recently, novel ML strategies continue to find application to the discovery of these complex alloys. We conclude this review and tutorial with a brief survey of a few of these newer applications. There are several advances that are particularly noteworthy. For example, with regard to phase prediction, Pei *et al.*[63] employed a ML-informed approach to the prediction of HE alloys that generalized the empirical Hume-Rothery rules. They examined a rather large dataset comprising over 1000 multicomponent alloys and demonstrated that their improved set of rules enabled the accurate prediction of solid-solution behavior. Huang *et al.*[64] also explored phase formation rules using ML techniques as applied to a rather large alloy dataset consisting of approximately 400 HEs. More specifically, three different ML algorithms, including K-nearest neighbors, an artificial neural network, and a SVM, were used for the classification of solid solutions, intermetallic phases, etc.

There have also been new applications of ML techniques to identify HE alloys having superior properties. In this regard, Chang *et al.*[65] used ML to find new compositions of AlCoCrFeMnNi-based alloys that exhibit relatively high hardness values. They also quantify correlations among composition, hardness, and microstructure. Newer methodologies having application to structural materials with outstanding mechanical properties have been summarized recently by Sparks *et al.*[66] Some HE alloys have also garnered interest as potential catalysts, and ML tools and related methodologies have been employed, for example, to find optimal compositions and structures for nanoparticle catalysts for fuel cells.[67] In addition, Pedersen *et al.*[68] examined the use of HE alloys as catalysts for CO_2 and CO reduction reactions. Finally, it has been suggested that grain-boundary complexions in HE alloys may be utilized to promote stability in nanocrystalline alloys.[69–71]

These recent contributions are a small sample of the ongoing efforts to apply ML and data analytics to the discovery and characterization of HE alloys. Other important work involves the use of hybrid methodologies that combine first-principles or atomic-level simulation with the ML toolkit. In summary, given the increasing interest in these complex alloys, it is expected that ML will continue to play an important role in future developments.

ACKNOWLEDGMENTS

The authors acknowledge support from the Office of Naval Research under Grant Nos. N00014-18-1-2181 and N00014-18-1-2484. They also wish to acknowledge support from the Nano/Human Interface initiative at Lehigh University. G.B. thanks the National Science Foundation for support through award No. 1944040.

APPENDIX A: PERFORMING A CANONICAL CORRELATION ANALYSIS (CCA)

In these Appendixes, we provide explicit instructions to perform two useful tasks outlined in this paper, namely, a canonical correlation analysis (Appendix A) and a cuckoo search (Appendix B). The purpose of these tutorials is to provide the details required to employ these techniques in a research setting.

As noted above, a canonical correlation analysis (CCA) is a very general approach for quantifying relationships between two variable sets, namely, an input (predictor) set and an output set, and it provides a convenient dimensional reduction strategy that identifies linear combinations of variables in a complex, multidimensional parameter space that are maximally correlated. To illustrate the utility of this methodology, we apply a CCA to a small, synthetic dataset for which correlations are *a priori* known. While the relevant quantities can be calculated directly from Eqs. (1) and (3), one can also use built-in function in MATLAB,[72,73] such as "canoncorr," or R.

Consider the three-variable predictor and output datasets for $M = 10$ experiments, denoted by X_{ij} and Y_{ij} ($i = 1, 2, 3$; $j = 1, 2, \ldots, 10$) respectively, given in Table II. Upon constructing the operators $\boldsymbol{\sigma}_1$ and $\boldsymbol{\sigma}_2$ [see Eq. (3)], the set of correlation

TABLE II. The three predictor variables $X_{i.}$ and output variables $Y_{i.}$ for $M = 10$ hypothetical experiments. (The dot denotes one of the experiments.) The variable sets are related by $Y_{3.} = 1.5\ X_{1.} - 0.2\ X_{2.} + 0.35{*}X_{3.}$.

$X_{1.}$	$X_{2.}$	$X_{3.}$	$Y_{1.}$	$Y_{2.}$	$Y_{3.}$
1	3	2	4	8	1.6
2	1	5	3	4	4.55
4	2	3	1	4	6.65
2	5	4	3	3	3.4
7	3	3	6	1	10.95
7	2	4	4	7	11.5
1	5	5	1	1	2.25
2	1	7	8	7	5.25
6	6	4	3	8	9.2
8	7	5	2	7	12.35

coefficients, sqrteig, are the square roots of the eigenvalues, λ_i, and are summarized in Table III. (Note that both operators have the same spectrum.) The corresponding canonical coefficients are obtained from the eigenvectors of $\boldsymbol{\sigma}_1$ and $\boldsymbol{\sigma}_2$, namely, $\vec{\alpha}'$ and $\vec{\beta}'$, and are found to be

TABLE III. The correlation coefficients corresponding to the three canonical variates from the CCA.

Canonical variates	1	2	3
sqrteig	1.0	0.573	0.048

$$\begin{aligned}
&\text{Variate 1: } \vec{\alpha} = (-0.3768, 0.0502, -0.0879) \quad \vec{\beta} = (0, 0, -0.2512),\\
&\text{Variate 2: } \vec{\alpha} = (0.0716, -0.4277, 0.3294) \quad \vec{\beta} = (0.4705, -0.0515, -0.0199),\\
&\text{Variate 3: } \vec{\alpha} = (-0.0469, 0.2515, 0.6347) \quad \vec{\beta} = (-0.0249, 0.3746, -0.0506).
\end{aligned} \tag{A1}$$

These coefficients are obtained from the eigenvectors via the normalization

$$\begin{aligned}
\vec{\alpha} &= \vec{\alpha}' / \sqrt{\vec{\alpha}'^{T} \cdot \boldsymbol{\Sigma}_{\vec{X}\vec{X}} \cdot \vec{\alpha}'},\\
\vec{\beta} &= \vec{\beta}' / \sqrt{\vec{\beta}'^{T} \cdot \boldsymbol{\Sigma}_{\vec{Y}\vec{Y}} \cdot \vec{\beta}'},
\end{aligned} \tag{A2}$$

where T denotes a matrix transpose. It should be noted that upon dividing $\vec{\alpha}$ and $\vec{\beta}$ for variate 1 by $\beta_3 = -0.2512$ we obtain the expected (i.e., built-in) relationship among the parameters.

These results can be obtained in MATLAB by invoking the function "canoncorr." More specifically, after defining the array elements X_{ij} and Y_{ij}, one enters, for example,

$$[\text{A B sqrteig}] = \text{canoncorr(X, Y)}.$$

The columns of the output matrices A and B comprise $\vec{\alpha}$ and $\vec{\beta}$, respectively, and sqrteig contains the canonical correlations. We note that this analysis has been recently generalized to the case of correlations between non-linear combinations of predictor and outcome variables.[28]

Once a CCA has been performed, one can then use hypothesis testing to determine which correlations are statistically significant and which are simply attributable to chance.[29,74,75] The null hypothesis for such a test is that all of the canonical correlations are essentially zero (i.e., $\lambda_i = 0\ \forall i$). If the null hypothesis can be rejected for any λ_i, then the corresponding variates are correlated. Because multiple variables are present, a multivariate approach such as Wilks' Lambda (Λ) must be used. One first defines

$$\Lambda = \prod_{j=1}^{nv} (1 - \lambda_j), \tag{A3}$$

where the λ_i are arranged in decreasing order and λ_1 is the largest eigenvalue. To determine statistical significance, one calculates the test statistic

$$Z = -\left[M - \frac{1}{2}(N_{in} + N_{out} + 3)\right] \log \Lambda, \tag{A4}$$

where Z is distributed approximately as a χ^2 distribution with $N_{in} \cdot N_{out}$ degrees of freedom. If the corresponding p-value is below a chosen threshold, then the null hypothesis is rejected for λ_1. The largest eigenvalue is then removed, and the hypothesis is tested again by calculating Z for the remaining eigenvalues. This process is repeated until one fails to reject the null hypothesis. All correlations corresponding to eigenvalues that were removed from the set are considered to be statistically significant, while those corresponding to the remaining eigenvalues are not. We note that Λ provides insight into the strength of correlation between the predictor and output variables; $\Lambda \approx 0$ indicates that most of the variance in the output variables is explained by the predictor variables. In the contrived example presented here $\Lambda = 0$, and so the output and predictor variable sets are (perfectly) correlated for $i = 1$.

APPENDIX B: PERFORMING A CUCKOO SEARCH (CS) TO PREDICT STRUCTURAL ORDERING

Short-range order (SRO) influences the physical properties of materials, especially in complex and nano-engineered structures such as HE alloys or MPEAs. More specifically, SRO dictates the thermodynamics, electronic transport (including electrical resistivity), magnetic properties, and the mechanics of materials, including, but not limited to, dislocation motion in solid solutions.[76–79] The Warren-Cowley parameters, which are employed to describe the SRO in a structure through a statistical model,[80] are defined as

$$\varpi^{AB}_{lmn} = 1 - \frac{P^{AB}_{lmn}}{c_B} = 1 - \frac{P^{BA}_{lmn}}{c_A}, \tag{B1}$$

where P^{AB}_{lmn} is the probability of finding an atom B at position $\vec{r}_{lmn} = l\vec{a}_1 + m\vec{a}_2 + n\vec{a}_3$ and c_A and c_B are the atomic fractions of A and B, respectively, in the alloy. In this notation, $\vec{a}_1$, $\vec{a}_2$, and $\vec{a}_3$ are lattice vectors and l, m, and n are fractional coordinates from an origin situated on an atom A.

For a truly random atomic distribution in any alloy, $\varpi^{AB}_{lmn} = 0$ (with the exception that, for any material, $\varpi^{AB}_{000} = 1$ by definition). Positive (negative) values of ϖ^{AB}_{lmn} indicate a preference for like (unlike) atoms to be in proximity. Thus, if the elemental distribution in the MPEA is completely random, all ϖ^{AB} will tend to zero while, for any long-range ordered structure, ϖ^{AB} oscillates with a

predictable trend across the different coordination shells. Although the foregoing analysis is related to a binary alloy, its extension to ternary, quaternary, and quinary MPEAs is straightforward. Thus, the pairwise multicomponent short-range order (PM-SRO) is defined by Refs. 80 and 81,

$$\varpi_m^{AB} = \frac{p_m^{AB} - c_B}{\delta^{AB} - c_B}, \tag{B2}$$

where δ^{AB} is the Kronecker delta and p_m^{AB} is the probability of finding atom B around A in the mth shell. This relationship reduces to the Warren-Cowley parameter for a binary alloy.

With this formalism, random walk simulations can be employed to interrogate structural ordering in MPEAs. The general approach is outlined below. The acceptance or rejection of trial moves in the random walk simulation is based on the overall SRO of the alloy, rather than the traditional comparison of potential energies. The objective of the random walk simulation is then to drive the overall system toward an atomic configuration with a particular or predefined SRO, akin to Ref. 82. The general algorithmic steps for the computation are as follows:

- Choose two atoms randomly to swap.
- Calculate the change in SRO, i.e., ΔSRO.
- Accept or reject the swap based on ΔSRO, i.e., accept the swap if the difference between the current SRO and the target SRO decreases.

Thus, from an initial configuration, the above steps are repeated until an elemental configuration is reached that matches the target SRO. In this illustration, the target SRO that will be considered is $\varpi_m^{AB} = 0$, indicating a truly random atomic distribution in the alloy.

Finally, random walk simulations can be integrated into the CS framework (see Algorithm 1), and the coupling leads to faster convergence and an effective design space exploration relative to a simple brute force approach. This unique approach essentially mimics a heuristic exploration of structural configurations with the desired SRO. Thus, atomic arrangements that are not chemically ordered[83] and those that inherently possess realistic chemical environments are available for examining energetics and properties using first-principles and atomistic methods.[84]

Algorithm 1: A metacode outlining the general CS algorithm

```
Data: Input quantities, conditions and optimization variables
Result: Optimal solution
Initialize the nests (possible solutions) with a distribution of the eggs (properties);
while iteration < maximum number of generations do
    Create new possible nests using Levy Flight (Global Search);
    Calculate the fitness F of the nests for the objective function(s);
    Choose a nest at random;
    if F_old < F_new then
        replace old nest with the new nest (solution);
    Reject a fraction p_a of the worst performing nests;
    Replace with new nests built using Levy flight or random walk;
    Keep the best nests (solutions with the best results for the fitness);
    Enumerate the solutions and identify the current best;
Return the best nest with the optimal solution;
```

DATA AVAILABILITY

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

REFERENCES

[1] D. B. Miracle and O. N. Senkov, Acta Mater. **122**, 448 (2017).
[2] E. Pickering and N. Jones, Int. Mater. Rev. **61**(3), 183 (2016).
[3] M.-H. Tsai and J.-W. Yeh, Mater. Res. Lett. **2**(3), 107 (2014).
[4] S. Gorsse, J.-P. Couzinie, and D. B. Miracle, C. R. Phys. **19**(8), 721 (2018).
[5] B. Cantor, I. Chang, P. Knight, and A. Vincent, Mater. Sci. Eng. A **375**, 213 (2004).
[6] O. N. Senkov, D. B. Miracle, K. J. Chaput, and J.-P. Couzinie, J. Mater. Res. **33**(19), 3092 (2018).
[7] A. Gali and E. P. George, Intermetallics **39**, 74 (2013).
[8] C.-Y. Hsu, C.-C. Juan, W.-R. Wang, T.-S. Sheu, J.-W. Yeh, and S.-K. Chen, Mater. Sci. Eng. A **528**(10), 3581 (2011).
[9] J. M. Rickman, T. Lookman, and S. V. Kalinin, Acta Mater. **168**, 473 (2019).
[10] J. M. Rickman, H. M. Chan, M. P. Harmer, J. A. Smeltzer, C. J. Marvel, A. Roy, and G. Balasubramanian, Nat. Commun. **10**, 2618 (2019).
[11] M. C. Troparevsky, J. R. Morris, P. R. C. Kent, A. R. Lupini, and G. M. Stocks, Phys. Rev. X **5**, 011041 (2015).
[12] O. N. Senkov, J. D. Miller, D. R. Miracle, and C. Woodward, Nat. Commun. **6**, 6529 (2015).
[13] P. Sarker, T. Harrington, C. Toher, C. Oses, M. Samiee, J.-P. Maria, D. W. Brenner, K. S. Vecchio, and S. Curtarolo, Nat. Commun. **9**, 4980 (2018).
[14] A. Konak, D. W. Coit, and A. E. Smith, Reliab. Eng. Syst. Saf. **91**(9), 992 (2006).
[15] J. Z. Wang, H. Jiang, Y. J. Wu, and Y. Dong, Energy **81**, 627 (2015).
[16] A. Sharma, R. Singh, P. K. Liaw, and G. Balasubramanian, Scr. Mater. **130**, 292 (2017).
[17] Y. Zhang, Y. J. Zhou, J. P. Lin, G. L. Chen, and P. K. Liaw, Adv. Eng. Mater. **10**, 534 (2008).
[18] Y. Li and W. Guo, Phys. Rev. Mater. **3**, 095005 (2019).
[19] C. Wen, Y. Zhang, C. Wang, D. Xue, Y. Bai, S. Antonov, L. Dai, T. Lookman, and Y. Su, Acta Mater. **170**, 109 (2019).
[20] Z. Zhou, Y. Zhou, Q. He, Z. Ding, F. Li, and Y. Yang, npj Comput. Mater. **5**, 128 (2019).
[21] T. Kostiuchenko, F. Körmann, J. Neugebauer, and A. Shapeev, npj Comput. Mater. **5**, 55 (2019).
[22] X. Liu, J. Zhang, M. Eisenbach, and Y. Wang, "Machine learning modeling of high entropy alloy: The role of short-range order," arXiv:1906.02889.
[23] B. Grabowski, Y. Ikeda, P. Srinivasan, F. Körmann, C. Freysoldt, A. I. Duff, A. Shapeev, and J. Neugebauer, npj Comput. Mater. **5**, 80 (2019).
[24] J. Qi, A. M. Cheung, and S. J. Poon, Sci. Rep. **9**, 15501 (2019).
[25] D. B. Miracle, J. D. Miller, O. N. Senkov, C. Woodward, M. D. Uchic, and J. Tiley, Entropy **16**, 494 (2014).
[26] T. R. Knapp, "Canonical correlation analysis: A general parametric significance-testing system," Psychol. Bull. **85**, 410–416 (1978).
[27] A. Lawrence, J. M. Rickman, M. P. Harmer, and A. D. Rollett, Acta Mater. **103**, 681–687 (2016).
[28] J. M. Rickman, J. M. Y. Wang, A. D. Rollett, M. P. Harmer, and M. P. C. Compson, npj Comput. Mater. **3**, 26 (2017).
[29] J. D. Jobson, *Applied Multivariate Data Analysis* (Springer-Verlag, 1992), Vol. II.
[30] R. Gittins, *Canonical Analysis: A Review with Applications in Ecology* (Springer-Verlag, Berlin, 1985).
[31] K.-F. Man, K.-S. Tang, and S. Kwong, *Genetic Algorithms: Concepts and Designs* (Springer-Verlag, London, 1999).
[32] A. Konak, D. W. Coit, and A. E. Smith, Reliab. Eng. Syst. Saf. **91**, 992–1007 (2006).

[33]S. E. K. Fateen and A. Bonilla-Petriciolet, *On the Effectiveness of Nature-Inspired Metaheuristic Algorithms for Performing Phase Equilibrium Thermodynamic Calculations* (Scientific World Journal, 2014).
[34]T. K. Sharma, M. Pant, and M. Singh, Mater. Manuf. Process. **28**(7), 788–802 (2013).
[35]M. N. Shehata, S. E. K. Fateen, and A. Bonilla-Petriciolet, Fluid Phase Equilib. **409**, 280–290 (2016).
[36]K. M. Udayraj, P. Talukdar, A. Das, and R. Alagirusamy, Int. J. Heat Mass Transf. **89**, 359–378 (2015).
[37]L. D. Coelho, C. E. Klein, S. L. Sabat, and V. C. Mariani, Energy **75**, 237–243 (2014).
[38]M. Jamil, H. J. Zepernick, and X. S. Yang, in *IEEE Military Communications Conference (MILCOM)* (IEEE, 2013), pp. 823–828.
[39]J. Z. Wang, H. Jiang, Y. J. Wu, and Y. Dong, Energy **81**, 627–644 (2015).
[40]X. S. Yang and S. Deb, Neural Comput. Appl. **24**(1), 169–174 (2014).
[41]A. H. Gandomi, X. S. Yang, and A. H. Alavi, Eng. Comput. **29**(1), 17–35 (2013).
[42]E. Valian, S. Tavakoli, S. Mohanna, and A. Haghi, Comput. Ind. Eng. **64**(1), 459–468 (2013).
[43]S. Walton, O. Hassan, K. Morgan, and M. R. Brown, Chaos Solitons Fractals **44**(9), 710–718 (2011).
[44]X. S. Yang and S. Deb, *World Congress on Nature & Biologically Inspired Computing* (Nabic, 2009), pp. 210–214.
[45]A. Natarajan, S. Subramanian, and K. Premalatha, Int. J. Bio-Inspired Comput. **4**(2), 89–99 (2012).
[46]X. S. Yang, *Nature-Inspired Optimization Algorithm* (Elsevier, 2014).
[47]G. B. Arfken and H.-J. Weber, *Mathematical Methods for Physicists*, 6th ed. (Elsevier Inc., Amsterdam, 2005).
[48]A. Ouaarab, B. Ahiod, and X. S. Yang, Neural Comput. Appl. **24**(7-8), 1659–1669 (2014).
[49]H. Y. Diao, L. J. Santodonato, Z. Tang, T. Egami, and P. K. Liaw, JOM **67**(10), 2321–2325 (2015).
[50]K. G. Pradeep, N. Wanderka, P. Choi, J. Banhart, B. S. Murty, and D. Raabe, Acta Mater. **61**(12), 4696–4706 (2013).
[51]B. Schuh, F. Mendez-Martin, B. Volker, E. P. George, H. Clemens, R. Pippan, and A. Hohenwarter, Acta Mater. **96**, 258–268 (2015).
[52]C. C. Tasan, Y. Deng, K. G. Pradeep, M. J. Yao, H. Springer, and D. Raabe, JOM **66**(10), 1993–2001 (2014).
[53]D. C. Ma, B. Grabowski, F. Kormann, J. Neugebauer, and D. Raabe, Acta Mater. **100**, 90–97 (2015).
[54]P. P. Bhattacharjee, G. D. Sathiaraj, M. Zaid, J. R. Gatti, C. Lee, C. W. Tsai, and J. W. Yeh, J. Alloys Compd. **587**, 544–552 (2014).
[55]G. R. Holcomb, J. Tylczak, and C. Carney, JOM **67**(10), 2326–2339 (2015).
[56]S. D. Kaloshkin, V. V. Tcherdyntsev, I. A. Tomilin, D. V. Gunderov, V. V. Stolyarov, Y. V. Baldokhin, I. G. Brodova, and E. V. Shelekhov, Mater. Trans. **43**(8), 2031–2038 (2015).
[57]V. V. Tcherdyntsev, S. D. Kaloshkin, E. A. Afonina, I. A. Tomilin, Y. V. Baldokhin, E. V. Shelekhov, D. V. Gunderov, I. G. Brodova, and V. V. Stolyarov, *Diffusion, Segregation, and Stresses in Materials* (Scitec Publications, Ltd., 2003), Vol. 216-2, pp. 313–321.
[58]D. K. Mukhopadhyay, C. Suryanarayana, and F. H. Froes, Metall. Mater. Trans. A **26**(8), 1939–1946 (1995).
[59]F. Cardellini, V. Contini, and G. Mazzone, J. Mater. Sci. **31**(16), 4175–4180 (1996).
[60]F. Cardellini, V. Contini, R. Gupta, G. Mazzone, A. Montone, A. Perin, and G. Principi, J. Mater. Sci. **33**(10), 2519–2527 (1998).
[61]M. Naik, M. R. Nath, A. Wunnava, S. Sahany, and R. Panda, in *IEEE 2nd International Conference on Recent Trends in Information Systems (ReTIS)* (IEEE, 2015), pp. 1–5.
[62]P. Ong, *Adaptive Cuckoo Search Algorithm for Unconstrained Optimization* (Scientific World Journal, 2014).
[63]Z. Pei, J. Yin, J. A. Hawk, D. E. Alman, and M. C. Gao, npj Comput. Mater. **6**, 50 (2020).
[64]W. Huang, P. Martin, and H. L. Zhuang, Acta Mater. **169**, 225–236 (2019).
[65]Y.-J. Chang, C.-Y. Jui, W.-J. Lee, and A.-C. Yeh, JOM **71**, 3433–3442 (2019).
[66]T. D. Sparks, S. K. Kauwe, M. E. Parry, A. M. Tehrani, and J. Brgoch, Annu. Rev. Mater. Res. **50**, 27–48 (2020).
[67]G. Tomboc, T. Kwon, J. Joo, and K. Lee, J. Mater. Chem. A **8**, 14844–14862 (2020).
[68]J. K. Pedersen, T. A. A. Batchelor, A. Bagger, and J. Rossmeisl, ACS Catal. **10**, 2169–2176 (2020).
[69]N. Zhou, T. Hu, and J. Luo, Curr. Opin. Solid State Mater. Sci. **20**, 268–277 (2016).
[70]J. M. Rickman, H. M. Chan, M. P. Harmer, and J. Luo, Surf. Sci. **618**, 88–93 (2013).
[71]J. M. Rickman and J. Luo, Curr. Opin. Solid State Mater. Sci. **20**, 225–230 (2016).
[72]See https://www.mathworks.com/help/stats/canoncorr.html for documentation.
[73]See http://sites.psu.edu/mcnl/files/2017/03/BIOE597_SP17_Lecture13-2fxzc43.pdf for documentation.
[74]A. Sherry and R. K. Henson, J. Pers. Assess. **84**(1), 37–48 (2005).
[75]R. Nandy and D. Cordes, Magn. Reson. Med. **50**, 354–365 (2003).
[76]E. E. Stansbury, C. R. Brooks, and T. L. Arledge, J. Inst. Met. **94**(4), 136 (1966).
[77]H. G. Baer, Z. Metall. **49**(12), 614–622 (1958).
[78]W. Wagner, R. Poerschke, and H. Wollenberger, Philos. Mag. B **43**(2), 345–355 (1981).
[79]J. C. Fisher, Acta Metall. **2**(1), 9–10 (1954).
[80]L. R. Owen, H. Y. Playford, H. J. Stone, and M. G. Tucker, Acta Mater. **115**, 155–166 (2016).
[81]A. V. Ceguerra, M. P. Moody, R. C. Powles, T. C. Petersen, R. K. W. Marceau, and S. P. Ringer, Acta Crystallogr. Sect. A **68**, 547–560 (2012).
[82]R. L. McGreevy and L. Pusztai, Mol. Simul. **1**, 359–367 (1988).
[83]C. Jiang, Acta Mater. **57**(16), 4716–4726 (2009).
[84]R. Singh, A. Sharma, P. Singh, G. Balasubramanian, and D. D. Johnson, "Accelerating computational modeling and design of high-entropy alloys," arXiv:2010.12107.