(Check for updates

Accelerating computational modeling and design of high-entropy alloys

Rahul Singh¹, Aayush Sharma¹, Prashant Singh², Ganesh Balasubramanian³ and Duane D. Johnson^{2,4}

High-entropy alloys, with *N* elements and compositions $\{c_{\nu=1,N}\}$ in competing crystal structures, have large design spaces for unique chemical and mechanical properties. Here, to enable computational design, we use a metaheuristic hybrid Cuckoo search (CS) to construct alloy configurational models on the fly that have targeted atomic site and pair probabilities on arbitrary crystal lattices, given by supercell random approximates (SCRAPs) with *S* sites. Our Hybrid CS permits efficient global solutions for large, discrete combinatorial optimization that scale linearly in a number of parallel processors, and linearly in sites *S* for SCRAPs. For example, a four-element, 128-site SCRAP is found in seconds—a more than 13,000-fold reduction over current strategies. Our method thus enables computational alloy design that is currently impractical. We qualify the models and showcase application to real alloys with targeted atomic short-range order. Being problem-agnostic, our Hybrid CS offers potential applications in diverse fields.

omplex solid-solution alloys (CSAs), which have a subset of near-equiatomic high-entropy alloys¹⁻⁷, show remarkable properties for number of elements $N \ge 4$ and set of elemental compositions $\{c_{\nu=1,N}\}^8$, and even for medium-entropy (N=3) alloys⁹. Such findings have encouraged research into CSAs for use in extreme-environment technologies, such as aerospace and energy generation, for example by adding refractory elements for higher operational temperatures. In refractory CSAs, vacancy defects-which are ubiquitous when processing-can have a profound influence on stability and phase selection¹⁰, thereby adding another design 'element'. CSAs thus have a vast design space to create materials with novel or improved properties (for example, resistance to fatigue, oxidation, corrosion and wear), but many (especially bulk) properties, including resistivity, thermoelectricity, elasticity and yield strength^{6,7-15}, can alter rapidly with small compositional changes. As such, accurate, rapidly generated CSA models are needed to enable computational design and to identify trends in $\{c_{\nu}\}$ -derived properties and thermal stability. Yet, models of CSAs have a design space that grows exponentially with number of elements N, number of pairs $\frac{1}{2}N(N-1)$ and number of sites S-a type of NP-hard (NP, nondeterministic polynomial time) combinatorial problem.

To make computational alloy design practical, we employ a metaheuristic cuckoo search (CS)^{16,17} that follows the brood parasitism of female cuckoo birds in which they mimic the color and pattern of eggs for a few host species—an evolutionary algorithm (EA)¹⁸. The advantages of CS are as follows: (1) it has a global convergence success that is higher than other EAs, (2) its local and global searches are controlled by a switching parameter and (3) Lévy flights scan the solution space more efficiently, with no random walks, so it is better than a Gaussian process^{16,17,19,20}. A CS yields approximate solutions ('nests') for intractable or gradient-free problems²¹ with little problem-specific knowledge—often only a 'fitness' function²². For complex cases, fitness can be discontinuous (non-differentiable to noisy). Related methods¹⁹ include simulated-annealing²³, genetic-algorithm²⁴, particle-swarm²⁵, ant-colony²⁶ and bat²⁷ methods. Inspired by CS successes, including in materials design²⁸, we introduce a Hybrid CS that is more efficient and establish this for standard functions, where CS already bests most common EAs^{16,17}. Our Hybrid CS employs Lévy flights for global searches and Monte Carlo (MC) for local explorations of large multimodal space, and scales linearly with the number of processors (that is, doubling the number of processors in parallel halves the compute time). Selecting a best nest at each iteration (or cycle) ensures that solutions ultimately converge to optimality, while diversification via randomization avoids stagnation (that is, being trapped in local minima).

As CSA properties can vary rapidly with composition $\{c_{\nu}\}$, Hybrid CS enables on-the-fly optimal model generation with a substantial reduction in solution times, scaling linearly with system size (in addition to the number of processors inherent to the CS). Hybrid CS constructs pseudo-optimal (discrete) supercell random approximates (SCRAPs) for S sites occupied by N elements to mimic CSAs (Fig. 1) with target one-site $\{c_{\nu}\}$ and two-site (pair) probabilities for a crystal symmetry (for example, body- (bcc) or face- (fcc) centered cubic). The pair probabilities are atomic short-range order (SRO) parameters that qualify a model's fitness, and these can be measured^{29,30}. Each SRO pair takes values over R neighbor shells (say, 1–5) with Z_R atoms per shell, leading to a total number of parameters per site of $\frac{1}{2}N(N-1) \sum_R Z_R$ over which to optimize.

Solution spaces grow rapidly with *N* and *S* (see 'Solution size and fitness' section). For a four-element, 128-atom model (a 10^{73} solution space), Hybrid CS SCRAP is optimal in 0.8 min—a reduction of more than 13,000-fold over current strategies. A five-element alloy with a 250- (500-) atom model (with a space of 10^{169} (10^{415})) is optimal in only 1.6 (4.9) min. Thus, Hybrid CS optimizes large problems with a substantial reduction in time over current methods (Table 1), enabling computational design that is currently impractical.

After establishing the bona fides of Hybrid CS, we define the fitness and associated physical (and discrete model) bounds to eliminate stagnation of MC searches. Hybrid CS-generated SCRAPs are presented for CSAs with targeted SRO in different crystal structures to show that solution times scale linearly with size and with

¹Mechanical Engineering, Iowa State University, Ames, IA, USA. ²Ames Laboratory, US Department of Energy, Iowa State University, Ames, IA, USA. ³Mechanical Engineering & Mechanics, Lehigh University, Bethlehem, PA, USA. ⁴Materials Science & Engineering, Iowa State University, Ames, IA, USA. ^{See}-mail: ddj@iastate.edu

ARTICLES



Fig. 1| Hybrid CS-optimized 250-atom cell for a bcc equiatomic ABCDE solid-solution alloy with zero SRO for three neighbor shells around each atom. See Table 1 for timings.

number of processors, enabling rapid creation of optimal Hybrid CS SCRAPs ('nests'). We then showcase alloy design and model assessment using electronic density-functional theory (DFT) to predict properties. For any random alloy, we discuss a symmetry requirement that permits a reduction in the number of DFT calculations for an alloy design. However, the Hybrid CS is problem-agnostic, so potentially offers optimization applications in diverse areas.

Results

Hybrid CS versus CS. Hybrid CS reaps the benefits of MC for local optimization alongside those of CS for the global optimum using

multiple-nest explorations via Lévy flight. A 'nest' represents, for example, a function value or an alloy configuration (SCRAP). A global CS discards a fraction of nests, p_a , with the worst fitness (that is, the probability of finding an alien nest¹⁶). In the Methods, we replace the local search in CS Algorithm 1 by standard MC and create a Hybrid CS Algorithm 2.

To show efficacy, we applied Hybrid CS and CS to one-dimensional (1D) benchmark functions that are used in applied math (defined in the Methods). In Fig. 2, simulations (mean over 100 runs) are shown for both algorithms versus iterations to reach the optimum. The algorithms converge to optimal values but at different rates. Hybrid CS outperforms CS in all cases, reducing evaluations by factors of 1.75 (Michalewicz) to 8 (Rastrigrin). MC is thus a more efficient search of local minima than Lévy flights alone³¹.

The CS has two parameters: (1) number of nests *n* and (2) discard probability p_a . Hybrid CS has two more: (3) fraction of top nests (#Top-Nests) chosen for an MC step and (4) number of MC steps (#MCiters). The number of iterations to convergence versus parameter values is tested in Fig. 3. Iterations are roughly constant after n = 20-30 (Fig. 3a) and increase linearly with p_a (Fig. 3b), while #MCiters increases roughly linearly after n = 10-20 (Fig. 3c) and the #Top-Nests passed in MC, with the rest untouched, has little effect on iterations (Fig. 3d). So, to achieve the least iterations to optimum, these results suggest $n \ge 15$, #MCiters $\le n$ and $0.1 < p_a < 0.4$. The parameters were fixed for tests in Fig. 2 (n = 15, $p_a = 0.25$, #MCiters = 15 and #Top-Nests = 0.3).

For any function with appropriate fitness, Hybrid CS outperforms CS, which already bests most common EAs^{16,17}. We thus employ optimal Hybrid CS SCRAPs for materials design using the parameters found above.

Туре	Species, N	No. of atoms, S	Hybrid CS timings for 1 processor (min)	Hybrid CS timings for (12) 24 processors (min)	Hybrid CS scaling for (12) 24 processors	MC-only timings via ATAT (min)	Factor over MC-only for (1) 24 processors
bcc	2	16	0.06	(0.008) 0.006	(7.5) 10.0	0.4	(6.7) 66.7
bcc	2	32*	0.09	(0.010) 0.008	(9.0) 11.3	0.5	(5.6) 62.5
bcc	2	40*	0.11	(0.013) 0.008	(8.5) 13.8	105.2	(956) 13,150
bcc	2	54*	3.16	(0.47) 0.21	(6.7) 15.0	-	-
bcc	3	54*	5.50	(0.71) 0.30	(7.7) 18.3	>1,440	(262) 4,800
bcc	4	128	23.59	(1.75) 0.79	(13.5) 29.9	>10,000	(424) 12,658
bcc	5	250	49.26	(3.57) 1.60	(13.8) 30.8	Unknown	-
bcc	6	432	87.82	(6.56) 2.91	30.2	-	-
bcc	7	686	143.50	(10.62) 4.73	30.3	-	-
bcc	8	1,024	222.21	(16.52) 7.38	30.1	-	-
bcc	9	1,468	319.67	(23.94) 10.82	29.5	-	-
bcc	10	2,000	446.53	(33.74) 15.48	28.8	-	-
fcc	2	32	0.12	(0.012) 0.008	(10.0) 15.0	-	-
fcc	2	108	18.15	(1.38) 0.62	(13.4) 29.3	-	-
fcc	3	108	26.24	(1.99) 0.86	(13.2) 30.5	-	-
fcc	4	108	31.12	(2.30) 1.01	(13.5) 29.8	-	-
fcc	4	256	70.49	(5.28) 2.30	(13.4) 30.6	-	-
fcc	5	500	149.71	(10.99) 4.89	(13.6) 30.6	-	-
AXB ₃	3	10	1.44	(0.29) 0.15	(4.9) 9.6	-	-

Table 1 | Hybrid CS timing and scaling with 1, 12 or 24 processors and a comparison to the popular MC-based ATAT code

Hybrid CS timings to generate optimal cells using 24 nests are solved with 1, 12 or 24 processors to show scaling in solution time for parallel cases, using the same cluster (an Intel dual Xeon Gold 6130 at 2.1 GHz/16-cores). CS parameters were set to 10 iterations (converged in 3–5) and each iteration had up to 1,000 global and 750 MC iterations. For simplicity, cell sizes were set as $S = A \cdot N^3$ so SRO can be exactly zero for bcc (fcc) (A = 2(4)). For an ABX₃ cubic lattice, A atoms (organic cation) are at the body center, X (mixed inorganic cation, Pb or Sn) are at the cube corners and B (halide) are at the face centers. Although shells can be included to a range that avoids correlation from periodic boundaries, SRO was optimized over three shells (* denotes only two shells permitted). Comparisons are made to popular MC-based ATAT code¹³, but, due to the excessive computational demands, multinary results could not be provided. For MC-only with larger cases, a sense of time is garnered (distributions were not assessed; MC often stagnates to a non-Gaussian state). Bolded numbers are those system sizes of typical interest in model design.

ARTICLES



Fig. 2 | Hybrid CS and CS function values versus iterations (objective evaluations) to reach the optimum for six functions. a-f, Function (f) values versus number of iterations for Michalewicz (a), Rosenbrock (b), De Jong (c), Ackley (d), Rastrigin (e) and Easom (f) functions. Each iteration is averaged over 100 runs. Error (±1s.d.) is denoted by the shaded areas around each line.

Solution size and fitness. To assess Hybrid CS (pseudo)optimal SCRAPs, as in Fig. 1, we need a fitness and size estimate of the solution spaces in terms of S sites and N elements. We illustrate this with C cells built from cubes so that $S = A \cdot C^3$, with A = 2(4) atoms for bcc (fcc). If C = N (the number of elements), SRO parameters can be exactly zero (homogeneously random) in a smallest-cell solution. For a bcc equiatomic ternary (ABC), quaternary (ABCD) and quinary (ABCDE), S is a 54-, 128- and 250-atom cell, respectively. In terms of combinatorial coefficient, ${}^{S}C_{S/N}$, the estimated configurations for site occupations are

- •
- $\begin{array}{l} \text{ABC:} \ {}^{54}\mathcal{C}_{18} \times {}^{36}\mathcal{C}_{18} \approx 10^{23} \\ \text{ABCD:} \ {}^{128}\mathcal{C}_{32} \times {}^{96}\mathcal{C}_{32} \times {}^{64}\mathcal{C}_{32} \approx 10^{73} \\ \text{ABCDE:} \ {}^{250}\mathcal{C}_{50} \times {}^{200}\mathcal{C}_{50} \times {}^{150}\mathcal{C}_{50} \times {}^{100}\mathcal{C}_{50} \approx 10^{169} \\ \text{ABCDE:} \ {}^{500}\mathcal{C}_{100} \times {}^{400}\mathcal{C}_{100} \times {}^{300}\mathcal{C}_{100} \times {}^{200}\mathcal{C}_{100} \approx 10^{415} \end{array}$

A cell with bigger S at fixed N alters compositions in discrete but more refined ways, as evinced by the two N=5 cell sizes shown above, but the times to render an optimal cell and use it are more challenging. The solution space increases as the number of pairs $\frac{1}{2}N(N-1)$ grows, requiring a solution for each atom and its pairs over the range R (1–5 shells).

We must define a fitness function for optimization. An N-component CSA is characterized uniquely by N-1 one-site (occupation) probabilities $p_{i\nu}^{i}$ for species ν and by $\frac{1}{2}N(N-1)$ two-site (pair) probabilities per neighbor shell, with the following definitions (and sum rules):

$$\hat{p}_{\nu} = c_{\nu} = \frac{1}{S} \sum_{i=1}^{S} p_{\nu}^{i} \quad \left(\sum_{\nu=1}^{N} c_{\nu} = 1 \right)$$
(1)



Fig. 3 | Hybrid CS performance, a-d, Hybrid CS performance for 100 runs, showing iterations versus number of nests (a), discard probability p_{a} (**b**), number of local iterations (MC) (**c**) and fraction of top nests (**d**). Box and whisker plots show median values (horizontal line) enclosed in box indicating data from first-to-third quartile (box) range, outlier data are beyond the whiskers, defining an upper/lower bound 1.5 times the interquartile (box) range.

$$p_{\nu\beta}^{ij} = p_{\nu}^{i} p_{\beta}^{j} [1 - \alpha_{\nu\beta}^{ij}] \quad \left(\sum_{\beta=1}^{N} p_{\nu\beta}^{ij} = p_{\nu}^{i}\right)$$
(2)

Here, average compositions ($\hat{p}_{\nu} = c_{\nu}$) are given by the sum over all sites S (with all species conserved). The SRO parameters, $\alpha^{y}_{\nu\beta}$, dictate pair probabilities $p_{nu\beta}^{ij}$ with ν (β) atoms at site *i* (*j*), and their values are bounded³²:

$$-\frac{\min\left(\mathbf{p}_{\nu},\mathbf{p}_{\beta}\right)^{2}}{p_{\nu}p_{\beta}} \le \alpha_{\nu\beta}^{ij} \le +1$$
(3)

where $\alpha < 0$ indicates ordering-type SRO (increased pair probabilities), whereas $\alpha > 0$ indicates clustering of like pairs (decreased pair probabilities). The final SRO for all sites and pairs qualifies a model and so serves as the fitness. So, a SCRAP must be optimized with constraints for target SRO values, and, to avoid stagnation of solutions and senseless iterations (wasted computing), we place 'stop' conditions on MC searches when the SRO falls below discrete bounds set by N and S (Methods). Discrete limits on floor/ceiling SRO values are exemplified in the Supplementary Information for a non-cubic SCRAP for a bcc equiatomic quinary.

Hybrid CS versus MC-only models. With MC stagnation addressed, Hybrid CS enables on-the-fly generation of optimal SCRAPs to model CSAs with arbitrary concentrations, structures and targeted atomic distributions. For ease of plotting, we first use a ternary (N=3) with S=54 sites (no SRO) to compare Hybrid CS with MC-only (Fig. 4). Cells with up to 2,000 sites and 10 elements are timed as shown in Table 1. Background on the MC-only generated cells is provided in the Discussion.

There is a substantial difference in timing for Hybrid CS (0.3 min) versus MC-only (1,440 min, or one day) from the ATAT code³³ (Table 1), with it increasing dramatically with larger S and N. Hybrid CS in

NATURE COMPUTATIONAL SCIENCE

ARTICLES

Fig. 4 was successful in every attempt to find the global (pseudo)optimum—zero SRO for all pairs over three shells for every site—irrespective of the initialization, albeit the iteration count varied. MC-only failed to reach an optimum from stagnation in all cases but one (a random event). For Hybrid CS in larger cells with N=3, 4 and 5, the SCRAPs have the targeted SRO and the distributions are Gaussian (Supplementary Information)—a general result for all system sizes.

Hybrid CS SCRAPs timings and scaling. The timings for Hybrid CS-created SCRAPs (Table 1) show markedly reduced times compared to MC-only, which suffer stagnation as the values of *S* and *N* increase. For a binary 40-atom cell, MC-only needed 105 min while Hybrid CS in serial mode required 0.11 min (<0.01 min in parallel).

The Hybrid CS timings demonstrate that the algorithm scales linearly with the number of processors $(n_{\text{proc}} \leq n_{\text{nests}})$ used for parallel solution (often called strong scaling in computational sciences); that is, we reduce the solution time by a factor of two for every doubling of processors used. In Table 1, we report timings for doubling from 12 to 24 processors (fifth column). By taking their ratio, the reduction factor is found. As an example, for a bcc five-element, 250-atom cell with 12 (24) processors, the solution time is 3.57 (1.60) min and the ratio is 3.57/1.60-a 2.22 (222%) decrease by doubling the number of processors. For 24 nests using 24 processors, we find a reduction of 30 (sixth column), rather than the 24 given by linear scaling. (In the following we show that linear scaling is expected by limiting *R*, so this factor depends slightly on the number of competing internal nests.) We plot t versus n_{proc} in Supplementary Fig. 1 up to 24 (the total nests used here), showing these reduction factors graphically. If we check the ratio for any size SCRAP in Table 1, this reduction is confirmed. The parameters for the Hybrid CS were set to 10 optimization steps (typically converged in 3-5), with a solution for each step having up to 1,000 Lévy and 750 MC searches (iterations).

As the SRO qualifies the alloy and our solutions, we note that all the multinary SCRAPs in Table 1 have specified a value of zero for three shells about all sites (worst error of $<10^{-3}$ for two pairs in the third shell in the 10-element, 2,000-site case). The Hybrid CS optimizes models in minutes (0.25 to 5 min on the processors used here) for the cell sizes (S=54-500) typically considered.

Let us assess the scaling of execution times with a limit on the range R of the SRO. Typically, the range is limited in a solid solution, except near a phase transition where the SRO diverges (or if electronic Fermi-surface nesting operates¹⁵). Optimization at each site over a few shells with Z_{tot} atoms is then usually valid and should scale as the number of atoms S. For ease of analysis, we again use bcc (fcc) SCRAPs with $S = A \cdot N^3$ so that the SRO parameters for each pair (section 'Solution size and fitness') can be exactly zero (target value set and achieved in Table 1). The relative time for different symmetries (for example, bcc and fcc) with fixed N and R can then be estimated as $t_{\rm rel}^N = t_{\rm fcc}^N / t_{\rm bcc}^N \approx \frac{A_{\rm fcc} \times Z_{\rm tot}^{\rm Lc}}{A_{\rm bcc} \times Z_{\rm tot}^{\rm bcc}} = \frac{4}{2} \times \frac{42}{26} = 3.2$ for SRO over three shells. Checking the ratio of timings in Table 1, the ratio is indeed about 3. Similarly, a relative timing for two sizes of fixed symmetry (say bcc) cell should scale as $t_{\text{rel}}^{\text{sym}} = t_{\text{bcc}}^{N_2} / t_{\text{bcc}}^{N_1} \approx \frac{S_2}{S_1} \times \left[\frac{N_2}{N_1}\right]^{1/4}$, as can be verified from the timings in Table 1. So, the Hybrid CS SCRAP optimization scales linearly with S, in addition to linearly with $n_{\rm proc}$, as is inherent to the Hybrid CS.

Hybrid CS SCRAPs are obtained rapidly to address concentration-dependent CSA properties. Six bcc 250-atom, 5-element (quinary) SCRAPs, such as $A_x(BCDE)_{1-x}$ versus the *x* composition along a line in composition space, for example, are found in minutes. Furthermore, any SRO values may be targeted, as SRO in alloys can lower the enthalpy or drive elemental surface enrichment. Smaller *S*-atom cells with larger *N* can be obtained, but zero SRO will not be possible at all compositions.



Fig. 4 | Hybrid CS versus MC-only optimization for a 54-atom bcc SCRAP. a,b, Monte-Carlo-only method (**a**) and Hybrid CS (**b**) errors for distinct runs versus iterations. Runs (averages) are denoted by (dashed) colored lines. Global interations (#iter) for Hybrid CS are indicated at top of plot.

Note that DFT methods typically scale (Methods) as $S^3 \cdot K_{pt}$, where K_{pt} is the number of symmetry-distinct *k*-points used to solve the electronic DFT eigenvalue problem. DFT solutions (on the processors used here) for a 54-atom cell take ~6 min for a $2 \times 2 \times 2 \ k$ -mesh (K_{pt} =4) and 1.9 h for a $5 \times 5 \times 5 \ k$ -mesh (K_{pt} =63), with different iterations (10–40) to converge the charge densities and total energies. From Table 1, the MC-only solution for a ternary 54-atom cell takes ~1,440 min (24 h) to get a model for one composition. Similarly, for a quaternary 128-atom cell, the solution time is ~10,000 min (one week), whereas each DFT iteration (with K_{pt} =4) takes ~77 min. Thus, MC-only model generation is more time-consuming and becomes worse with larger *S* or *N*. By contrast, Hybrid CS requires 0.3 (0.79) min to yield an optimal 54-atom (128-atom) cell, so DFT is the design bottleneck.

Real alloy applications. We constructed SCRAPs to assess the formation energy (E_{form}) versus SRO parameters (observed or trial $\alpha_{\mu\nu}^{\text{shell}}$). We assessed the relative energy (*E*) versus lattice constants (*a*) and equilibrium values (\bar{a}), along with atomic displacement {*u*,} distributions for binaries to quinaries. We employed an all-electron Korringa–Kohn–Rostoker (KKR) Green's function method^{34,35} and the pseudopotential Vienna ab-initio simulation package (VASP)^{36,37} to obtain E_{form} versus SRO and \bar{a} , compared to experiments. For DFT, we used a Perdew–Burke–Ernzerhof (PBE) exchange-correlation functional³⁸ and Monkorst–Pack meshes for Brillouin zone integrations³⁹. See Methods for details and Supplementary Information for supporting results.

CSAs with SRO. Hybrid CS works for any $\frac{1}{2}N(N-1)$ SRO pairs, so we use fcc Cu₃Au (N=2) as an example for ease of presentation (one CuAu SRO value per shell, α^{s}) and because there are experimental data available. SCRAPs with specified SRO (each optimized in 0.6 min, Table 1) are used to mimic (1) a homogeneously random state at 495 °C ($\alpha^{s}=0$), as well as alloys with observed α^{s} values³⁰ at (2) 450 °C and (3) 405 °C. Figure 5 presents a plot of E_{form} versus SRO. For any N and S, atomic displacements from ideal sites have zero mean (Supplementary Figs. 2 and 3).

KKR and the experiment with no SRO agree well $(3-5 \text{ meV atom}^{-1} \text{ difference})$. Both methods show similar trends, but KKR includes known alloying core-level shifts, explaining the higher values observed with VASP. A gain of 30 meV atom⁻¹ gain is found with SRO (lower entropy). The KKR \bar{a} without SRO is 3.765 Å, which agrees with the 3.749 Å observed⁴⁰⁻⁴², with only a small 0.43% mismatch. With a disordered alloy with SRO this value is 3.755 Å, closer to the ordered alloy value of 3.743 Å. The VASP \bar{a} with no SRO is 3.823 Å (with SRO it is 3.816 Å), that is, a 2% mismatch.



Fig. 5 | E_{form} versus SRO for Cu_{0.75}Au_{0.25} in 108-site SCRAP. E_{form} is shown in meV atom⁻¹ and was obtained using the KKR (circles) and VASP (squares) methods. The measured α values (three shells)³⁰ are as follows: at 495 °C, SRO₀ = {0.0, 0.0, 0.0}; at 450 °C, SRO₁ = {-0.195, +0.215, +0.003}; at 405 °C, SRO₂ = {-0.218, +0.286, -0.012}. For SRO₀, a measured range⁴⁰ of E_{form} is shown (dashed horizontal band).

Distributions and averages. To simplify presentation, we assessed VASP E versus a and displacements $\{u_i\}$ for NbMoTa 54-atom SCRAP (Fig. 6a-c). At \bar{a} , the energy is -64.4 meV atom⁻¹ when volume-relaxed (ideal sites) and -80.5 meV atom⁻¹ when atom-relaxed (a reduction of 16 meV). Vector displacements $\{u_{x,y,z}\}$ sum to zero individually and are Gaussian distributed, as required by CSA symmetry, giving \bar{a} as the diffraction value. Mean-squared displacements determine the Debye-Waller factor (Supplementary Information), which describes the attenuation of X-ray, neutron or electron scattering caused by thermal motion, providing background diffuse intensity from inelastic scattering. Diffraction on 'large' samples (for example, 1 cm³) gives 'self-averaged' properties, as the Avogadro's number for each local configuration is sampled simultaneously. We find similar results for any N and S (Fig. 6d-g). For quaternary TaNbMoW, volume-relaxed (-63.3 meV atom⁻¹) and atom-relaxed (-74.5 meV atom⁻¹) energies show an 11 meV atom⁻¹ reduction from displacements. For quinary TaNbMoWV, volume-relaxed (-105.5 meV atom⁻¹) and atom-relaxed (-126.3 meV atom⁻¹) energies show a larger gain in stability from displacements with vanadium addition (-21 meV). Displacements increase with complexity, but more with vanadium alloying (Fig. 6d,g), enhancing the stability, lattice distortions and mechanical behavior, as discussed in the following.

Configurations. SCRAPs provide good averages if a cell is large enough ('infinite' is exact, but impractical); otherwise, configurational averaging may be warranted. In principle, thermodynamically, all configurations should be sampled ('good', 'bad' and 'ugly', leading to an average ideal lattice)-not just relaxed, low-energy (good) ones, as are often chosen in the literature, but higher-energy, unfavorable (bad) and metastable (ugly) ones too. SCRAPs (before relaxations) are used for arbitrary choices of atomic site occupations, so relate to just one representative configuration out of many, so a model must be qualified. To complete a model, atom types Nb, Mo or Ta must be assigned to A, B or C sites. For example, the formation energy after relaxations may indicate that the structure is stable (favorable negative values), but phonons may exhibit lattice instabilities, as indicated by phonon frequencies ω , making it dynamically unstable. (Phonons with $\omega^2 \ge 0$ are stable and those with $\omega^2 < 0$ are unstable—with $\omega \propto \sqrt{-1}$ (that is, 'imaginary')—such as occurs when a pencil is stood on its point (unstable) as opposed to being held like a pendulum by its point (stable).)

A statistical average governs nature's reality, and an instability is controlled by environments around each atom. So, to eliminate an instability, a larger SCRAP is necessary to improve the statistical 'self-average'; alternately, a simple swap of atom types in a small SCRAP may eliminate a local instability. For example, if we assign Nb, Mo and Ta to A, B and C sites, respectively, we find a minimum energy and stable lattice (that is, positive phonon frequencies) (Fig. 6c), found from the PHONOPY code⁴³ with DFT inputs (Methods). From these results we can assess the alloy properties; for example, the average \bar{a} is 3.248 Å. Yet, with A \leftrightarrow C (that is, Nb \leftrightarrow Ta), we find a higher (+0.05 eV) energy and unstable phonons (Fig. 6c), suggesting that this model is in general too small, and care must be taken.

Lattice distortions. Each atom in a CSA has a different chemical environment that can cause lattice distortion (for example, from atomic size differences)⁴⁴. However, the effect of lattice distortion on the CSA mechanical response has been explored less because of a lack of computationally efficient models. In SCRAPs, lattice distortion in refractory CSAs can be tuned by changing the local environment to enhance the mechanical response (as intimated in Fig. 6a,d,f), an effect that is observed in ultra-strong ternaries⁴⁵. Rather than a size difference, embodied to zeroth-order in a solid solution's electronic bandwidths (the electronic origin of Hume-Rothery's size-effect rule⁴⁶), strength enhancement correlates with the electronegativity difference between elements (on the Allen scale for solids, with vanadium largest), where largest bond distortions occur around vanadium sites (Supplementary Fig. 4).

Discussion

Reducing DFT computational times. Having saved orders of magnitude in model generation, DFT computational time is a major issue as DFT methods typically scale as $S^3 \cdot K_{pp}$, whereas Hybrid CS SCRAPs generation scales as *S*. However, a savings in DFT time is possible. As displacements { u_i } must have zero mean in any disordered alloy, the equilibrium (average) volume must be mathematically identical to that of the 'ideal' (diffraction) lattice. An example of this was shown in Fig. 6a, where ideal and atomically relaxed SCRAPs have identical equilibrium volumes. So, relaxations for any sized multinary SCRAP with any SRO need only be performed at the equilibrium volume (found from SCRAPs with ideal atomic positions) to assess properties and trends.

Limitations. Our Hybrid CS can be built with an arbitrary cell created by using $M_1 \times M_2 \times M_3$ smaller base units, but we must carefully limit the range *R* of the SRO parameters so as not to correlate them directly with distant sites due to periodic boundary conditions (true for any cellular technique). In addition, to exemplify the methods and analysis, we limited the alloy model generation in the implemented code to homogeneously disordered crystal structures (simple cubic, bcc, fcc and hexagonal close-packed (hcp)). However, there is no restriction in general, so, in the near future, we will generalize the code for more complex superstructures (like partially disordered compounds).

Related cellular techniques. A supercell to mimic random alloys is not a new idea. Structural models are often constructed by specifically occupying sites of a finite-sized periodic cell. For Metropolis MC methods⁴⁷, including simulated annealing, potential energies serve as a fitness criterion for acceptance of a trial move, yet solutions for global optima often stagnate, even in problems that are not large³¹. We have already discussed the fitness for SCRAPs configurational optimization, along with floor/ceiling bounds given by each *S* and *N* and SRO value in the MC optimization (a worked example is provided in the Supplementary Information).

The original special quasi-random structure (SQS) used Ising-like MC to find supercells that mimic zero atomic correlations in the alloy by arranging atoms in particular ordered

NATURE COMPUTATIONAL SCIENCE

ARTICLES



Fig. 6 | SCRAPs results for refractory bcc TaNbMo, TaNbMoW and TaNbMoWV. a-**f**, Results for TaNbMo (54-atom) (**a**-**c**), TaNbMoW (128-atom) (**d**,**e**) and TaNbMoWV (250-atom) (**e**,**f**). In **a**,**d**,**f**, components u_i of vector atomic displacements (mean 0) relative to their average (scattering) lattice position are shown. For clarity, only the u_z distribution is shown in **a**, and the average squared displacement (mean of 0.027 Å²) is shown in Supplementary Fig. 3. In **b**, the energy for TaNbMo is shown relative to the lowest energy for the three A, B, C assignments, and, as required, $\bar{a} = 3.248$ A is the same for each. In **c**, acoustic phonons are described for stable NbMoTa (positive-definite frequencies) and unstable TaMoNb (imaginary frequencies, plotted as negative). In **e**,**g**, distributions for u_i are shown for TaNbMoW with $\bar{a} = 3.247$ A (**e**) and TaNbMoWV with $\bar{a} = 3.198$ A (**g**).

layers depending on the number of sites and atom types⁴⁸. In some cases, there was more than one configuration for a fixed number of sites, thus requiring an average. Such SQS did not have proper lattice symmetry (like bcc), so atomic displacements could not sum to zero as required by symmetry, in contrast to SCRAPs. Recently, the SQS algorithm was implemented using a stochastic MC approach³³ to determine a sample configuration allowing a supercell with an arbitrary number of base units, such as $M_1 \times M_2 \times M_3$ bcc two-atom cells, as done for SCRAPs. However, as N or S increase, MC-only times become impractical and solutions stagnate (Table 1). Other implementations of the MC-only approach with an arbitrary number of base units have been done recently, although some of the results were correlated through the use of improper boundary conditions⁴⁹.

In principle, Hybrid CS and MC-only schemas should get the same supercells for specific cases, but our Hybrid CS avoids stagnated solutions, and timings are markedly reduced in serial mode and substantially reduced in parallel mode (Table 1). Moreover, Hybrid CS can rapidly optimize any sized SCRAP for any number of elements and for any targeted disorder, that is, any values of SRO parameters.

Enabling design via machine learning. Our Hybrid CS optimal SCRAPs permit the design of arbitrary complex solid-solution alloys, predicting properties and trends, including for surfaces, catalysis and oxidation, that are currently impractical. To showcase this, we assessed the stability and properties of binary to quinary solid solutions and discussed the qualification of the models. However, notable DFT calculation resources are necessary to generate databases for a broad range of compositions and properties. To further accelerate design, SCRAPs is integrated with high-throughput DFT calculations to produce accurate but limited databases, possibly validated or supplemented with experimental data, then we utilize the data for machine-learning (physics-based) models, an ongoing activity. Finally, Hybrid CS offers potential optimization improvements in other fields, such as manufacturing, commerce, finance, science and engineering, as long as an appropriate fitness can be defined that can be evaluated expeditiously.

Methods

Cuckoo search. The CS is based on the brood parasitism of a female cuckoo bird, which specialize in mimicking the color and pattern of a few host species. For this there are three idealized rules: (1) a cuckoo lays an egg in a randomly selected nest; (2) the nest with highest-quality egg (fitness) survives and is forwarded to the next generation; (3) the host bird can discover the cuckoo egg with a probability $p_a \in (0,1)$ and, once discovered, it dumps either the nest or the cuckoo egg. The key advantages of this process are listed in the main text.

Hybrid CS. Our Hybrid CS schema reaps the benefits of traditional MC for local optimization alongside the CS schema for global optimization utilizing multiple nest explorations via Lévy flight. A global CS removes a fraction of nests, p_a , with worst fitness (a nest represents a lattice configuration), and it signifies the probability of finding an alien nest¹⁰. We replace the local search in CS Algorithm 1 with MC and create the Hybrid CS given in Algorithm 2 (shown in pseudo-code below), where the global search uses multiple-nest explorations. For Hybrid CS, a basic MC worked well, as embodied in Algorithm 2 between 'begin MC' and 'end MC', in performing the MC steps:

- (1) Obtain a nest from the sample of nests.
- (2) Randomly swap a pair of site occupations.
- (3) If Fitness_{new} < Fitness_{old}, Accept Swap; or
- (4) Else Reject; Switch; and Go to (1).

Algorithm 1. Cuckoo Search Algorithm.

Input: Fix input & identify optimization variables					
Output: Optimized solution					
Initialize nests					
while iteration < Global maximum number do					
Create new nests using Lévy Flight (Global Search)					
Calculate fitness F of the nests					
Choose a nest randomly					
if $F_{old} < F_{new}$ then					
replace nest with the new cuckoo					
Discard fraction p_a of worst nests & build new ones					
Keep best nests with the best results					
Rank the solutions & find the current best					
Return the best solutions					

Return the best solutions

ARTICLES

NATURE COMPUTATIONAL SCIENCE

Algorithm 2. Hybrid CS Algorithm.

Input: Fix input and optimization function

Output: Optimized solution

Initialize nests

while iteration < Global maximum number do

Create new nests using Levy Flight (GLOBAL Search) Calculate fitness *F* of the nests

calculate infless i of the flests

Choose fraction of nests with best fitness (top nests) LOCAL Search via Monte Carlo 7D2 \ll begin MC \gg

foreach nests \in top nests **do**

acceptance = 0

rejections = 0

while iteration < Local iterations do

Calculate delta step ($\delta x = \sigma^* randn$)

Perturb nests with $x_i + \delta x$

Calculate fitness, $F(x_i + \delta x)$

Calculate $\delta F = F(x) - F(x_i + \delta x)$

if $\delta F > 0$ then

Perform the switch

acceptance+=1

else

```
rejections+=1
```

if acceptance > mc_1^* Local iterations then $\sigma = \sigma^* a$

if rejections > mc_2^* Local iterations then

 $\sigma = \sigma/b$

Discard fraction p_a of worst nests \ll end MC \gg

Rank the solutions & find the current best

Return the best solutions

'Local' MC iteration chooses a fraction of nests to optimize based on a nest's value of fitness and a fraction equal to *top nests* $\in \{0, 1\}$. Aside from 'top nests', the local MC depends on $mc_1 \in \{0, 1\}$ and $mc_2 \in \{0, 1\}$, which are used to optimize the value of step size, $\delta x = \sigma^* randn$, by altering the value of σ (*randn* is a value from a standard normal distribution). For local optimization, number of acceptances/rejections are counted and, depending on their value, the value of δx alters. The other parameters are a(b) > 1, the increase/decrease increment in σ . By collecting the number of acceptances/rejections, we increase/decrease $|\delta x|$ to get a local optimized value faster.

Standard test functions. The Hybrid CS schema (Algorithm 2) and the CS schema using only Lévy flights (Algorithm 1) were competed using a standard set of 1D benchmark functions, as shown in Fig. 2. The function name, its global optimum $f(x^*)$, which occurs at x^* , and the function definition are given below, where *d* is the dimension of the input parameters:

1. Michalewicz (d=5): $f(x^*) = -4.6876$

 $\begin{aligned} f(x) &= -\sum_{i=1}^{d} \sin(x_i) \sin^{2m} \left(\frac{x_i^2}{\pi}\right) \\ 2. \operatorname{Rosenbrock} (d=16): 0.0 \text{ at } x^* = (1, 1, ..., 1) \\ f(x) &= \sum_{i=1}^{d-1} 100(x_{i+1} - x_i^2)^2 + (x_i - 1)^2 \\ 3. \operatorname{De} \operatorname{Jong} (d=16): 0.0 \text{ at } x^* = (0, 0, ..., 0) \\ f(x) &= \sum_{i=1}^{d} x_i^2 \\ 4. \operatorname{Ackey} (d=16): 0.0 \text{ at } x^* = (0, 0, ..., 0) \\ f(x) &= -20 \exp(-0.2 \sqrt{\frac{1}{d} \sum_{i=1}^{d} x_i^2}) \\ 5. \operatorname{Rastrigin:} 0.0 \text{ at } x^* = (0, 0, ..., 0) \\ f(x) &= 10d + \sum_{i=1} dx_i^2 - 10 \cos(2\pi x_i) \\ 6. \operatorname{Easom:} -1 \text{ at } x^* = (\pi, \pi) \\ f(x) &= -\cos(x_1) \cos(x_2) \exp(-(x_1 - \pi)^2 - (x_2 - \pi)^2) \end{aligned}$

Bounded discrete searches—no stagnation. SCRAPs must be optimized with constraints for target SRO values:

minimize
$$\sum |\hat{\alpha}_{\alpha\beta}^{s} - d_{\alpha\beta}^{s}|$$

subject to $\sum_{\alpha=1}^{N} p_{\alpha}^{i} = 1$ and $\sum_{\beta=1}^{N} p_{\alpha\beta}^{ij} = p_{\alpha}^{i}$ (4)

 $\hat{\alpha}_{\alpha\beta}^s$ refers to the average SRO for the *s*th shell for an (α, β) pair. For $\frac{1}{2}N(N-1)$ pairs, $d_{\alpha\beta}^s$ is the target SRO value. The final SRO for all sites and pairs qualifies the model.

To avoid stagnation of solutions we place 'stop' conditions on MC (local) searches when the SRO falls below the discrete bounds set by the cell N and S.

Such criteria avoid senseless iterations (wasted computing), working well when combined with a CS that guarantees global (pseudo-optimal) convergence in a range R. The discrete limits for SRO parameters from equations (1)–(4) are given by

$$\left(n_{\alpha} - \frac{\left|g_{\alpha\beta}^{s}\right|}{n_{s}c_{\beta}}\right)\frac{1}{n_{\alpha}} \le \alpha_{\alpha\beta}^{s} \le \left(n_{\alpha} - \frac{\left[g_{\alpha\beta}^{s}\right]}{n_{s}c_{j}}\right)\frac{1}{n_{\alpha}}$$
(5)

with the radial distribution function $(S_{\alpha\beta}^{s})$, number of atoms in shell $s(n_{s})$ and type $\alpha(n_{a})$. [] and [] represents the decimal at the lower (floor) and higher (ceiling) integer values, respectively. We use distance of SRO $a_{\alpha\beta}^{s}$ from one of these values for a 'stop' criteria, that is

$$\sum |\bar{\alpha}_{\alpha\beta}^{i} - \left(n_{\alpha} - \frac{|\underline{s}_{\alpha\beta}|}{n_{\kappa}\epsilon_{\beta}}\right)\frac{1}{n_{\alpha}}| \le \epsilon_{1}$$

$$\sum |\bar{\alpha}_{\alpha\beta}^{s} - \left(n_{\alpha} - \frac{|\underline{s}_{\alpha\beta}|}{n_{\kappa}\epsilon_{\beta}}\right)\frac{1}{n_{\alpha}}| \le \epsilon_{2}$$
(6)

where ϵ_1 and ϵ_2 are predefined values. Choosing S and N to set discrete p_a^i and specifying $\frac{1}{2}N(N-1)$ target values for $p_{a\beta}^{ij}$, the final values of SRO for all sites and atom pairs qualify the model fitness. We exemplify discrete limits on floor/ceiling SRO values in a $3 \times 3 \times 5$ bcc supercell for an equiatomic quinary in the Supplementary Information, with the values shown in Supplementary Table 1.

DFT. For Cu₃Au, the VASP results used a 108-atom SCRAP with SRO. The structures were relaxed using a 350-eV plane-wave energy cutoff, PBE exchange-correlation functional³⁶ and $3 \times 3 \times 3$ Monkhorst–Pack *k*-mesh³⁹ for Brillouin zone integrations. Total energy calculations were done at a denser $(7 \times 7 \times 7)$ *k*-mesh. By definition, $E_{\text{form}} = E_{\text{tot}} - \sum_{i} n_i E_{i}$, where $E_{\text{tot}}(E_i)$ is the total energy of the alloy (pure elements '*t*') and n_i is the number of sites per element in a supercell. For the same *k*-meshes, KKR³⁴ was also used for E_{form} . Self-consistent charge densities were found using the Green's function by complex-energy (Gauss–Legendre semicircular) contour integration with 24 energies in a spherical-harmonic basis, including *s*, *p*, *d* and *f* orbital symmetries³⁵. Core eigenvalues were from Dirac solutions, and the valence used a scalar-relativistic solution (no spin–orbit coupling).

Typically, DFT methods scale as S^3 , as for any matrix eigenvalue solution. However, electronic DFT must be solved for all points on the *k*-mesh used for convergence—that is, $N_k^{\text{total}} = N_{k_1} \times N_{k_2} \times N_{k_3}$ —which may be reduced to fewer symmetry-distinct points, $K_{\text{pt}} \leq N_k^{\text{total}}$, for a given high-symmetry crystal, like bcc or fcc. Hence, total times depends on $S^3 \cdot K_{\text{pt}}$.

For ternary to quinary alloys, a 54-atom TaNbMo SCRAP was relaxed in VASP using a 350-eV plane-wave energy cutoff, $8 \times 8 \times 8$ k-mesh ($K_{pt} = 256$) and the PBE exchange-correlation functional³⁸. Compared to ternary, the only difference for 128-atom TaNbMoW and 250-atom TaNbMoWV was the k-mesh, that is, $5 \times 5 \times 5$ and $2 \times 2 \times 2$, respectively. After symmetry operations are applied, the 5³ and 2³ meshes have, respectively, 63 and 4 symmetry-distinct K_{pt} points.

For phonons, DFT energy and force convergence criteria were set to be very high $(10^{-7} \text{ eV and } 10^{-6} \text{ eV Å}^{-1}$, respectively). A finite-displacement method (set to 0.03 Å) was employed using PHONOPY⁴³. Phonon dispersion was plotted along the high-symmetry Brillouin zone directions (Γ -H-N- Γ). Unstable (imaginary) frequencies are plotted as negative frequencies for simplicity of presentation.

Data availability

Supporting data for all data plotted in the Figs. 1–6 (as well as Supplementary Figs. 1–4) are available as source data in spreadsheets, in the Supplementary Information (see additional information) and at Code Ocean⁵⁰ and https://github.com/DuaneDJohnson/Hybrid-Cuckoo-Search/. Source data are provided with this paper.

Code availability

Interactive open-source codes are available via Code Ocean for Hybrid-CS SCRAPs⁵⁰ and for Hybrid CS for 1D functions⁵¹. For open-source codes (and data) for Hybrid CS SCRAPs or 1D functions, see https://github.com/DuaneDJohnson/Hybrid-Cuckoo-Search/.

Received: 17 June 2020; Accepted: 29 November 2020; Published online: 14 January 2021

References

- Yeh, J. W. et al. Nanostructured high-entropy alloys with multiple principal elements: novel alloy design concepts and outcomes. *Adv. Eng. Mater.* 6, 299–303 (2004).
- Cantor, B., Chang, I. T. H., Knight, P. & Vincent, A. J. B. Microstructural development in equiatomic multicomponent alloys. *Mater. Sci. Eng. A* 375–377, 213–218 (2004).
- Senkov, O. N., Miller, J., Miracle, D. & Woodward, C. Accelerated exploration of multi-principal element alloys with solid solution phases. *Nat. Commun.* 6, 6529 (2015).

NATURE COMPUTATIONAL SCIENCE

ARTICLES

- George, E. P., Raabe, D. & Ritchie, R. O. High-entropy alloys. *Nat. Rev. Mater.* 4, 515–534 (2019).
- Gao, M.C., Yeh, J-W., Liaw, P. K., Zhang, Y. (Ed.), *High-Entropy Alloys: Fundamentals and Applications*, 1st ed., Springer Inter. Publishing, Switzerland, 2016, pp. 333–366.
 Singh, P., Smirnov, A. V. & Johnson, D. D. Atomic short-range order
- Singh, P., Smirnov, A. V. & Johnson, D. D. Atomic short-range order and incipient long-range order in high-entropy alloys. *Phys. Rev. B* 91, 224204 (2015).
- 7. Singh, P. et al. Design of high-strength refractory complex solid-solution alloys. *npj Comput. Mater* **4**, 16 (2018).
- 8. Miracle, D. B. & Senkov, O. N. A critical review of high entropy alloys and related concepts. *Acta Mater.* **122**, 448–511 (2017).
- Zhang, Y. et al. Influence of chemical disorder on energy dissipation and defect evolution in concentrated solid solution alloys. *Nat. Commun.* 6, 8736 (2015).
- 10. Singh, P. et al. Vacancy-mediated complex phase selection in high entropy alloys. *Acta Mater.* **194**, 540–546 (2020).
- Karati, A. et al. Ti₂NiCoSnSb—a new half-Heusler type high-entropy alloy showing simultaneous increase in Seebeck coefficient and electrical conductivity for thermoelectric applications. *Sci. Rep.* 9, 5331 (2019).
- Ding, Q. et al. Tuning element distribution, structure and properties by composition in high-entropy alloys. *Nature* 574, 223–227 (2019).
- Li, Z., Pradeep, K. G., Deng, Y., Raabe, D. & Tasan, C. C. Metastable high-entropy dual-phase alloys overcome the strength-ductility trade-off. *Nature* 534, 227–230 (2016).
- 14. Zhang, R. et al. Short-range order and its impact on the CrCoNi medium-entropy alloy. *Nature* 581, 283–287 (2020).
- Singh, P., Smirnov, A. V. & Johnson, D. D. Ta-Nb-Mo-W refractory high-entropy alloys: anomalous ordering behavior and its intriguing electronic origin. *Phys. Rev. Mater.* 2, 055004 (2018).
- Yang, X. S. & Deb, S. Cuckoo search via Lévy flights. In Proc. World Congress on Nature and Biologically Inspired Computing 210–214 (IEEE, 2009).
- Yang, X. S. & Deb, S. Engineering optimisation by Cuckoo Search. Int. J. Math. Model. Numer. Optim. 1, 330–343 (2010).
- Back, T., Fogel, D. & Michalewicz, Z. Handbook of Evolutionary Computation (Oxford Univ. Press, 1996).
- 19. Yang, X. S. Engineering Optimization: An Introduction with Metaheuristic Applications (Wiley, 2010).
- Yang, X. S., Koziel, S. & Leifsson, L. Computational optimization, modelling and simulation: recent trends and challenges. *Procedia Comput. Sci.* 18, 855–860 (2013).
- 21. Blum, C. & Roli, A. Metaheuristics in combinatorial optimization: overview and conceptual comparison. *ACM Comput. Surv.* **35**, 268–308 (2003).
- 22. Ashlock, D. Evolutionary Computation for Modeling and Optimization 1st edn (Springer, 2016).
- Kirkpatrick, S., Gelatt, C. D. Jr & Vecchi, M. P. Optimization by simulated annealing. *Science* 220, 671–680 (1983).
- Holland, J. H. Adaptation in Natural and Artificial Systems 1st edn (MIT Press, 1992).
- Kennedy, J. & Eberhart, R. Particle swarm optimization. In Proc. ICNN'95 International Conference on Neural Networks Vol. 4, 1942–1948 (IEEE, 1995).
- Dorigo, M., Maniezzo, V. & Colorni, A. Ant system: optimization by a colony of cooperating agents. *IEEE Trans. Syst. Man Cybern. B Cybern.* 26, 29–41 (1996).
- Yang, X. S. Bat algorithm for multi-objective optimisation. Int. J. Bio-Inspired Comput. 3, 267–274 (2011).
- Sharma, A., Singh, R., Liaw, P. K. & Balasubramanian, G. Cuckoo searching optimal composition of multicomponent alloys by molecular simulations. *Scrip. Mater.* 130, 292–296 (2017).
- 29. Cowley, J. M. An approximate theory of order in alloys. *Phys. Rev.* 77, 669–675 (1950).
- Moss, S. C. X-ray measurement of short-range order in Cu₃Au. J. Appl. Phys. 35, 3547–3553 (1964).
- Gutowski, M. Lévy flights as an underlying mechanism for global optimization algorithms. Preprint at https://arxiv.org/pdf/math-ph/0106003. pdf (2001).
- 32. Johnson, D. D. in Computation of Diffuse Intensities in Alloys,
- Characterization of Materials (ed. Kaufmann, E.) 346–375 (Wiley, 2012). 33. Van de Walle, A. et al. Efficient stochastic generation of special quasirandom
- structures. Calphad 42, 13–18 (2013).
 34. Johnson, D. D., Nicholson, D. M., Pinski, F. J., Gyorffy, B. L. & Stocks, G. M. Density-functional theory for random alloys: total energy within the coherent-potential approximation. Phys. Rev. Lett. 56, 2088–2091 (1986).

- Alam, A. & Johnson, D. D. Structural properties and stability of (meta)stable ordered, partially ordered, and disordered Al–Li alloy phases. *Phys. Rev. B* 85, 1441202 (2012).
- 36. Kresse, G. & Hafner, J. Ab initio molecular dynamics for liquid metals. *Phys. Rev. B* 47, 558–561 (1993).
- Kresse, G. & Hafner, J. Ab initio molecular-dynamics simulation of the liquid-metal/amorphous-semiconductor transition in germanium. *Phys. Rev.* B 49, 14251–14269 (1994).
- Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized gradient approximation made simple. *Phys. Rev. Lett.* 77, 3865–3868 (1994).
- 39. Monkhorst, H. J. & Pack, J. D. Special points for Brillouin-zone integrations. *Phys. Rev. B* 13, 5188–5192 (1976).
- Orr, R. L. Heats of formation of solid Au–Cu alloys. Acta Metall. 8, 489–493 (1960).
- Flinn, P. A., McManus, G. M. & Rayne, J. A. Elastic constants of ordered and disordered Cu₃Au from 4.2 to 300 °K. J. Phys. Chem. Solids 15, 189–195 (1960).
- 42. Smallmann, R. E. & Nagan, A. H. W. *Modern Physical Metallurgy* 3rd edn (Butterworths, 1970).
- Togo, A. & Tanaka, I. First-principles phonon calculations in materials science. Scrip. Mater. 108, 1–5 (2015).
- Yeh, J.-W. Alloy design strategies and future trends in high-entropy alloys. JOM 65, 1759–1771 (2013).
- Sohn, S. S. et al. Ultrastrong medium-entropy single-phase alloys designed via severe lattice distortion. Adv. Mater. 31, 1807142 (2019).
- Pinski, F. J. et al. Origins of compositional order in NiPt alloys. *Phys. Rev.* Lett. 66, 766–769 (1991).
- Ceguerra, A. V. et al. Short-range order in multicomponent materials. Acta Crystallogr. A 68, 547–560 (2012).
- Zunger, A., Wei, S.-H., Ferreira, L. G. & Bernard, J. E. Special quasirandom structures. *Phys. Rev. Lett.* 65, 353–356 (1990).
- Song, H. et al. Local lattice distortion in high-entropy alloys. *Phys. Rev.* Mater. 1, 023404 (2017).
- Singh, R., Sharma, A., Singh, P., Balasubramanian, G. & Johnson, D. D. SCRAPs: a multicomponent alloy structure design tool; https://doi. org/10.24433/CO.0000024.v1
- Singh, R., Sharma, A., Singh, P., Balasubramanian, G. & Johnson, D. D. Hybrid-CS code for 1D test functions; https://doi.org/10.24433/CO.6419254.v1

Acknowledgements

R.S. was supported in part by D.D.J.'s F. Wendell Miller Professorship at ISU. Work at Ames Laboratory (by R.S., A.S., P.S. and D.D.J.) was funded by the US Department of Energy (DOE), Office of Science, Basic Energy Sciences, Materials Science & Engineering Division. Ames Laboratory is operated for the US DOE by Iowa State University under contract no. DE-AC02-07CH11358. G.B. was funded by the National Science Foundation through award no. 1944040.

Author contributions

D.D.J. proposed and supervised the project. R.S. wrote the SCRAPs generation code using the hybrid CS algorithm. R.S. and A.S. did initial testing. D.D.J. developed the linear-scaling parallel algorithm and scaling analysis. P.S. and R.S. implemented SCRAPs optimization with parallelization and catalogged timings. P.S. completed DFT and phonon calculations, and performed analysis with D.D.J. P.S. got the code running on Code Ocean. R.S., A.S., P.S. and G.B. drafted the initial manuscript, then D.D.J. prepared the final manuscript with approval from all the authors.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s43588-020-00006-7.

Correspondence and requests for materials should be addressed to D.D.J.

Peer review information Fernando Chirigati was the primary editor on this Article and managed its editorial process and peer review in collaboration with the rest of the editorial team. *Nature Computational Science* thanks the anonymous reviewers for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature America, Inc. 2021