[illegible]

Mayeul d'Avezac* and Alex Zunger†
Na i al Re ewable E e g Lab a , G lde , C l ad 80401, USA
(Received 11 September 2007; revised manuscript received 24 April 2008; published 4 August 2008)

We examine how the two different mechanisms proposed historically for biological evolution compare for the determination of crystal structures from random initial lattice configurations. The Darwinian theory of evolution contends that the genetic makeup inherited at birth is the one passed on during mating to new offspring, in which case evolution is a product of environmental pressure and chance. In addition to this mechanism, Lamarck surmised that individuals can also pass on traits *ac i ed* during their lifetime. Here we show that the minimum-energy configurations of a binary $A_{1-}B$ alloy in the full $0 \leq \leq 1$ concentration range can be found much faster if the conventional Darwinian genetic progression—mating configurations and letting the lowest-energy (fittest) offspring survive—is allowed to experience Lamarckian-style fitness improvements during its lifetime. Such improvements consist of $A \leftrightarrow B$ transmutations of some atomic sites (not just atomic relaxations) guided by "virtual-atom" energy gradients. This hybrid evolution is shown to provide an efficient solution to a generalized Ising Hamiltonian, illustrated here by finding the ground states of face-centered-cubic Au_{1-} Pd using a cluster-expansion functional fitted to first-principles total energies. The statistical rate of success of the search strategies and their practical applicability are rigorously documented in terms of average number of evaluations required to find the solution out of 400 independent evolutionary runs with different random seeds. We show that all *e ac* ground states of a 12-atom supercell (2^{12} configurations) can be found within 330 total-energy evaluations, whereas a 36-atom supercell (2^{36} configurations) requires on average 39 000 evaluations. Thus, this problem cannot be currently addressed with confidence using costly energy functionals (e.g., density-functional theory (DFT) based) unless it is limited to $\leq$20 atoms. The computational cost can be reduced at the expense of accuracy: Searching for all *a i a e*-minimum-energy configurations (within 3 meV) of a 12- or 36-atom supercell requires on average 30 or 580 total-energy evaluations, respectively. Thus it could be addressed even by costly energy functionals such as density-functional theory.

DOI: 10.1103/PhysRevB.78.064102 PACS number(s): 02.30.Zz, 31.15.−p, 71.15.−m

. T T

At the heart of solid-state physics and structural inorganic chemistry is the form/function relationship between crystal structure and crystal properties.[1–3] This recognition has led to continued efforts in measuring and cataloging crystal structures[4] and, more recently, to systematic efforts in the theoretical prediction of crystal structures, either from inductive "Pauling-esque" approaches[2,3] or from explicit quantum-mechanical total-energy minimization approaches.[5] There are generally two sets of structural degrees of freedom that need be determined in crystals—the unit-cell lattice structureoptimization problems can be distinguished according to the type and amount of information assumed at the outset regarding these degrees of freedom. In the first class of problems (hereafter referred to as type (i)), the system's lattice structure (e.g., fcc) and the decoration of the lattice sites by different atomic types are assumed. Left to optimize are cell-internal degrees of freedom not specified by the space group and the cell-external degrees of freedom, such as volume or the tetragonal

$\frac{c}{a}$ ratio.[6–8] In the second type of problems (hereafter referred to as type (ii)), one attempts to populate

minimum energy of three-dimensional Ising models[33] is nondeterministic polynomial-time (NP) hard; i.e., the number of trial structures of which the energy is evaluated during the search grows exponentially with the number of atoms, N, in the system.[34] The key question about any such search procedure is therefore how many evaluations of the total energy are needed to obtain the correct minimum-energy configurations with a given degree of confidence.

The number of configurations for which the total energy must be evaluated in order to identify the lowest-energy configurations with a given degree of confidence depends on the extent to which a given material system is dominated by the NP-hard combinatorial issue of site decorations, or by simpler noncombinatorial factors. For example, from a complexity-theory point of view, it is not clear whether lattice-type and unit-cell optimizations are NP hard.[35,36] In metal alloys whose constituents are both nearly isovalent and have the same lattice type (e.g., fcc Au_{1-} Pd ,[28] fcc Au_{1-} Cu ,[32] or bcc Mo_{1-} Ta (Ref. 30)), alteration of the lattice type corresponds to easily resolvable high-energy excitations, whereas combinatorial $A \leftrightarrow B$ cross substitutions on the same lattice type (type (ii) optimization) cost little energy and thus correspond to difficult-to-resolve low-energy excitations. In contrast, cross substitutions of cations and anions in valence compounds[1,2] (e.g., As-on-Ga "antisite defect" in GaAs or anion/cation cross substitutions in Al_2O_3 and $MgSiO_3$) correspond to easily resolvable high-energy excitations, whereas alteration of the lattice type corresponds to a low-energy excitation.[8,37] As a result, determining the lattice type becomes the bottleneck for structure optimization of valence compounds. Previous

tures," e.g., different $A_{N-x}B_x$ decorations of the N periodic lattice sites. It was shown empirically[11] that the number of ICSs evolves as $\sim N^{2/3}$, whereas the number of all same-shape structures (reduced by the symmetries of the lattice) with N atoms evolves as $\sim Ae^{0.6N}$. In order to find the minimum-energy configurations of a binary alloy on a fixed lattice, we follow the strategy outlined in Ref. 11, whereby each ICS is sampled individually. There are 243 ICSs for $N \leq 20$ and 1282 ICSs for $N \leq 32$ which must be explored to search the full configuration space.

In many of the searches performed below, we will look for the ground states in a single inequivalent cell shape, e.g., one $\times$ $\times$ supercell, and sample its decorations. Obviously, obtaining the ground states of the system requires examining all ICSs, one by one. This is not attempted in this paper. The exact algorithmic details for such a procedure can be found in Ref. 11.

$$\cdot \quad \in [0,1]$$

In lattice statistical mechanics,[12,25,45] the ground-state line is defined via the convex hull C

We will show that the performance of a search is highly dependent upon the exact nature of the question asked. We present in the following four searches with four different objectives:

Q e i 1. Find the *e ac dee e* ground state of a binary alloy on fixed lattice for a given × × supercell e.g., for one ICS.

Q e i 1′. Find the *a i a e dee e* ground state of a binary alloy on fixed lattice for a given × × supercell.

Q e i 2. Find *all e ac* ground states of a binary alloy on fixed lattice for a given × × supercell.

Q e i 2′. Find *all a i a e* ground states of a binary alloy on fixed lattice for a given × × supercell.

The four questions are separated according to whether one seeks multiple answers, e.g., questions 1 and 1′, for which we search for a number of ground states simultaneously, or singular answers, e.g., questions 2 and 2′, for which we search for a single ground state or an approximate ground state. Furthermore, we differentiate between exact and fuzzy searches. For instance, we define an answer to question 1′ as any configuration less than 3 meV from the exact deepest ground state, as obtained from answering question 1. In a similar fashion, we formulate question 2′ as the search for *all a i a e* ground states. More explicitly, we search for a convex-hull line $C_{3\ \mathrm{meV}}$ which is no more than 3 meV from the exact convex hull line C obtained here from answering question 2 for a given × × supercell, e.g., $\forall \in 0,1$, $C_{3\ \mathrm{meV}} - C \leq 3$ meV. In practice, one does not know C beforehand. Hence, one cannot know without solving question 2 when question 1′ has been solved. Nevertheless, this formulation allows us to compare the expense of absolute convergence with approximate convergence, as well as compare exact versus fuzzy search goals. Note that each of these questions is termed with respect to a single ICS. The result for all configurations with N lattice sites is recovered by searching each of the $N^{2/3}$ ICSs.

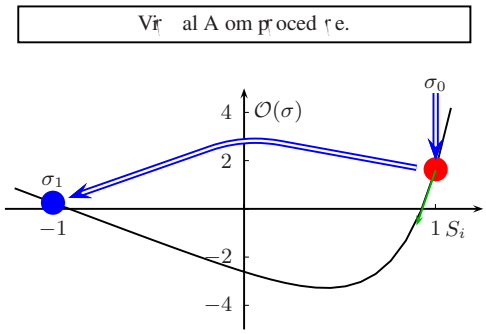


FIG. 3. (Color online) Description of the virtual-atom jump minimizer. The algorithm evaluates $\mathcal{O}$ only at physical configurations and its gradients only between first-neighbor physical points: (i) A physical configuration σ_0 is chosen as a starting point. (ii) The gradient at one lattice site i is evaluated. (iii) If this gradient is negative, then $\mathcal{P}(\sigma_1)$ is evaluated, and the search proceeds from the better structure σ_0 or σ_1. (iv) If, on the other hand, the gradient is negative, the gradient at the next lattice site j is evaluated, and the search proceeds from there.

Wang *et al.*[9] In their "linear combination of atomic potentials," each of the substitutional sites i of a backbone molecule is decorated by a fictitious atom (or chemical group) with a fictitious potential V_i, where $V_i = {}_iV_A + (1-{}_i)V_B$ is the concentration-weighted linear average of the end-point potentials V_A and V_B. Each lattice site, or molecule site, is occupied by a different virtual atom. By defining a virtual fitness $\delta^{(1)}(\sigma)$ using these virtual atoms, the gradients $\partial\delta^{(1)}(\sigma)/\partial {}_i$ can be introduced which represent the "chemical appeal" of an $A \leftrightarrow B$ transmutation at site i.

The original approach of Wang *et al.*[9] consists of minimizing the virtual total energy with respect to the occupations ${}_i$, often resulting in a nonphysical minimum where the sites are occupied by fictitious virtual atoms. Unfortunately, there is no clear link between the virtual minimum and the physical minimum-energy decoration. Furthermore, it is difficult to enforce physicality constraints such as $S_i^2 = \pm 1$ using either Lagrangian multipliers or penalty functions. Generally, "constrained" minimization procedures work by first finding a region of space where the constraints are satisfied and then looking for the minimum within this region. In our case, each region spans a discrete point of the configuration space. Hence, a constrained minimization approach will yield only the closest physical point from the starting point, rather than the physical minimum.[50] The virtual-atom "jump" strategy we have adopted (Fig. 3) circumvents these difficulties. We evaluate $\delta^{(1)}(\sigma)$ and first-order derivatives with respect to $A \leftrightarrow B$ transmutation only at physical points. Proceeding from a starting *physical* configuration σ_0, the gradient $\partial\delta^{(1)}(\sigma_0)/\partial {}_i$ at a random site i is evaluated. A negative gradient indicates that transmuting the atom at site i could lead to a lower energy. In that case, the depth $\delta^{(1)}(\sigma_1)$ of this neighboring configuration σ_1 is evaluated, and the procedure iterates from the better configuration σ_0 or σ_1. Otherwise, if the gradient is positive, the procedure iterates with a different lattice site j. The order in which the sites are evaluated is random. Convergence is deemed achieved when every lattice site has been explored without predicting or finding a better neighboring configuration. A similar approach was introduced recently by Keinan *et al.*[10]

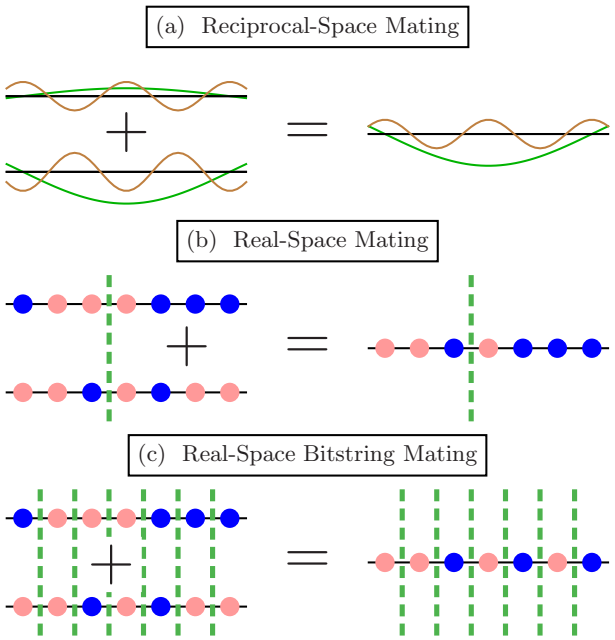


FIG. 4. (Color online) In (a) reciprocal-space mating, individual configurations are identified as waves of A/B material (represented by sinusoidal lines). A new individual is constructed from the A/B material waves of two parents. In (b) real-space mating, the two parents are cut in half by a plane (dashed line in one-dimensional (1D) figure) and the two half structures are spliced together into a new individual. With bit-string mating (c), each lattice site is mapped as a bit on a bit string; this is equivalent to a real-space mating, where each atom is at the center of its own cut-and-splice region.

Mating is the key to a successful GA: it must be able to identify and pass on favorable traits and patterns from parents to offspring. Previous genetic algorithms for type (iii) general space-group optimization[19,21–23,38,51] and type (ii) configuration-search[11] mate individuals directly in *real space* (see Fig. 4). The former approaches[19,21–23,38,51] generally define a plane in real space which cuts the configuration of each parent into two (see Fig. 4(b)). Two "half structures" are then chosen and spliced together to form a new individual. The objective is to integrate into one individual the real-space patterns present in each half of the parents. Unfortunately, this procedure also produces patterns straddling the cutting plane which are not inherited but are rather an artifact of the mating procedure. As such, it is likely that these patterns are not particularly fit, i.e., that they do not correspond to low-energy structures. Type (ii) configuration search proceeds by swapping atom types between two parents, e.g., by performing a standard bit-string mating where the occupation of each site i in the ICS is identified by a "spin variable" $S_i = \pm 1$ (where $S_i = 1$ corresponds to an occupation of site i by an A atom and $S_i = -1$ to that by a B atom). It can also be viewed as an extreme version of real-space mating where

each parent is cut into as many “cut-and-splice” regions as there are atoms in the supercell (see Fig. 4(c)). In addition to the cut-and-splice cross-over operation described above, one also introduces a real-space mutation operator where randomly chosen atoms in the unit cell are flipped from an *A* to a *B* atom. The rate with which the mutation operator is applied and the number of sites to flip per mutation are controlled via two parameters. We present here a mating procedure for type (ii) optimizations, reciprocal-space mating, which does not suffer from the drawbacks of cut-and-splice operation and incorporates crossover and mutations in a single operation.

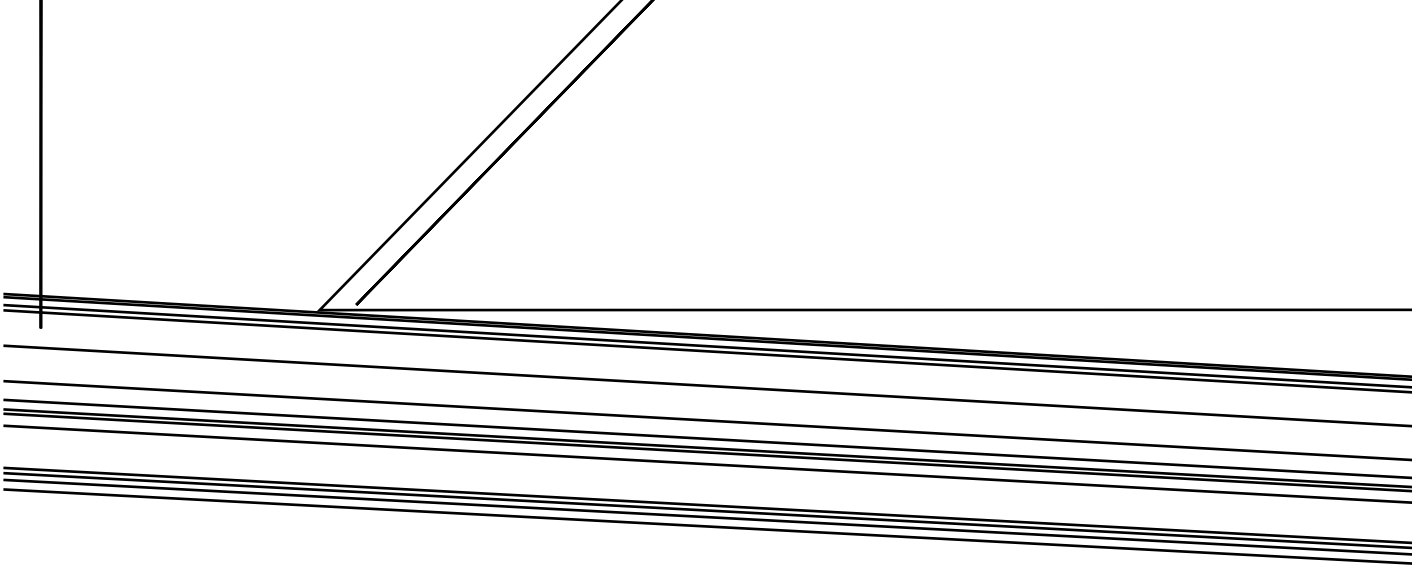
Reciprocal-space mating (Fig.wiig.wiig.

choose to first fit total energies computed from first principles onto a cluster-expansion functional. We will use a cluster-expansion functional of Au_{1-} Pd which has been previously fitted to a few density-functional-theory (DFT) total energies.[28] The resulting cluster-expansion functional allows us to compute formation enthalpies outside the fitting set to within 3 meV of DFT energies at a fraction of the cost in computer resources. Indeed, beyond the need for a fast functional to perform the ground-state search detailed in this paper, the cluster-expansion functional can also be used to obtain the formation enthalpy of the random alloy, or even -T phase diagrams using the Monte Carlo method.[52]

Configurational searches can be performed on explicit Born-Oppenheimer energy surfaces $E_{\mathrm{direct}}(\sigma,\ )$ evaluated on the fly[53,54] or by first parametrizing $E_{\mathrm{direct}}(\sigma,\ _{\mathrm{min}})$, where $\ _{\mathrm{min}}$ is the set of relaxed atomic positions of configurations σ. The latter can be given by cluster expansion, which fits a few (50) total-energy calculations obtained from DFT to a generalized Ising model,

$$\Delta H_{\mathrm{CE}}(\sigma) = J_0 + \frac{1}{N}\left[\sum_i J_i S_i + \sum_{ij} J_{ij} S_i S_j + \cdots\right] + \sum \Delta E_{\mathrm{CS}}(\ ,\ _{\sigma})(S)\ (^2F)(\ ,\ _{\sigma}), \qquad (4)$$

where the occupation of each lattice site i by an A or a B atom in configuration



tions 1, 2, 1′, and 2′ correspond, respectively, to Figs. 5–8. The top three panels of each figure show the average number of evaluations of the cluster-expansion functional required to solve the figure's particular question with an expectation for success ϵ in $2\times2\times8$ supercells. In the lower panels, we report the number of evaluation required on average by 2 $\times2\times$ supercells of increasing sizes and degree of confidence ϵ=95%. Statistics were obtained using a minimum of 400 independent evolutionary runs for each search strategy.

A. [illegible] $n\times m\times p$ [illegible]

1. Real-space GA versus reciprocal-space GA

Figures 5(a) and 5(b) compare the use of real-space bit-string and reciprocal-space matings in GA. Figure 5(a) reports the average number of evaluations required to achieve a given expectation for success, or degree of confidence, within a $2\times2\times8$ supercell. We find that the reciprocal-space mating performs better than the real-space bit-string mating for any degree of confidence ϵ. The reciprocal-space mating generally needs a smaller population size, leading to a shorter "learning period," during which the search strategies evolve the starting population to the region of space containing the deepest ground state, hence resulting in a much lower value at ϵ=0%. For most of the degree of confidence range $\epsilon\in$(0%,100%), the curve is linear. This means that higher success rates come at an exponential increase in the required number of evaluations. Figure 5(b) reports the number of evolutions required to find the exact deepest ground state of $2\times2\times$ supercells versus with a degree of confidence of ϵ=95% for real-space bit-string and reciprocal-space GAs. We find again that the reciprocal-space mating is much more efficient. Interestingly, both curves are sawlike, with the number of evaluations required by $2\times2\times$ supercells with even smaller than those with odd . The cluster expansion of Au_{1-} Pd is clustering, e.g., the first-neighbor pair interaction favors Au-Pd patterns over Au-Au or Pd-Pd patterns. This type of arrangement is not possible for odd , leading to

with respect to the degree of confidence ϵ (panel (c)) or the size of the configuration space (panel (d)), $\mathcal{B}$-LGA performs only as well as the virtual-atom approach. It would seem that the computing cost inherent to performing a virtual-atom (VA) minimization overwhelms the Lamarckian-GA benefit of mating local minima only. On the other hand, we find in panels (e) and (f) that reciprocal-space mating performs at least as well with a Lamarckian evolution (-LGA) than it does with a Darwinian evolution (-GA). Overall, the reciprocal-space Lamarckian, -LGA, is the most effective method for finding the exact deepest ground state.

$n \times m \times p$

1. Real-space GA versus reciprocal-space GA

Figure 6 reports the number of evaluations with respect to ϵ and the size of the configuration space required to find all exact ground-states of a given supercell. Panels (a) and (b) show that in contrast with question 1 (Fig. 5), the real-space bit-string $\mathcal{B}$-GA is more efficient than the reciprocal-space -GA for question 2 at larger configuration space and confidence values. Interestingly, we note that the real-space bit-string strategy is more effective at finding *all* ground states (Fig. 6) than it is at finding only the deepest one (Fig. 5). This may seem a paradoxical result. It stems from the different objective functions which are minimized in the two cases. To find the deepest ground state, we chose to minimize the formation enthalpy $\Delta H(\sigma)$ alone. When searching for all ground states, we minimize the distance to the known convex hull, $\Delta H(\sigma) - C( )(_{\sigma})$. The latter objective incorporates more information than is needed $a \quad i \quad i$ to find the exact deepest ground state. Nonetheless, it makes the search more effective. This result is quite general and is discussed in detail in Appendix B.

2. Darwinian evolution versus Lamarckian evolution

Panels (c) and (d) of Fig. 6 compare the real-space Lamarckian $\mathcal{B}$-LGA and Darwinian $\mathcal{B}$-GA. We find again that the real-space bit-string Lamarckian $\mathcal{B}$-LGA behaves much like the virtual-atom search. In fact, the latter does not solve the

(a) and (b) show little difference between the real-space -GA and the reciprocal-space -GA, with the latter having a slight edge. Figure 2 shows that a very large proportion of local minima are within 3 meV of the deepest ground state; hence these are *acceptable* answers to this question. The relative abundance of solutions to the problem may explain why -GA and -GA display similar success rates.

2. Darwinian evolution versus Lamarckian evolution

In panels (c) and (d) of Fig. 7, we show once again that the real-space bit-string Lamarckian evolution -LGA behaves like the simpler VA approach. However, the virtual-atom approach turns out to be very efficient when searching for the approximate deepest ground state. This is not surprising. Indeed, we have shown in Fig. 2 that a large ratio of the local minima lies close to the exact deepest ground state. In fact, it grows larger with the size of the configuration space for those supercells we have studied. Since local minima can be obtained in polynomial time and since the number of local

Furthermore, the complete answer to a ground-state search problem contains in general a number of configurations across the concentration range $\in [0,1]$. We have shown that combining local refinements with a reciprocal-space mating scheme and using as an objective function the distance to the known convex hull yields a search procedure which efficiently solves ground-state problems. Nonetheless, finding all exact ground-state configurations of an $A\,B_{1-}$ alloy of even relatively small systems soon becomes intractable [330 assessed configurations for $N=12$, 2800 for $N=24$, and up to 39 000 for $N=36$] without the use of "indirect" first-principles methods such as cluster expansion. However, this same problem can be solved approximately [within 3 meV] at a much smaller cost in functional evaluations [30 assessed configurations for $N=12$, 380 for $N=24$, and 580 for $N=36$]. Finally, the same search procedure is also quite equal to solving single-valued searches such as the exact deepest ground-state search presented in the paper.

This work was funded by the U.S. Department of Energy, Office of Science, Basic Energy Sciences, under Contract No. DE-AC36-99GO10337 for NREL within NREL's Laboratory Directed Research and Development Program. We further acknowledge the use of the Evolutionary Object library[56] and the Alloy Theoretical Automated Toolkit.[57]

size. Furthermore, the reciprocal-space mating incorporates mutations and crossover in one single operation, without the need for extra parameters. As a result, the Lamarckian reciprocal-space GA is not only the most effective search strategy presented here, it is also the most black-box-like.

We report in Table I the required number of evaluations for answering question 2 with a degree of confidence of 95% using the reciprocal-space Lamarckian GA and with respect to different supercell sizes and populations. We find that population size can have a large effect. Indeed, taking, for instance, a $2 \times 2 \times 5$ supercell (third row), using a population size of 20 individuals will require 5500 evaluations, whereas increasing the size to 40 will obtain the answer in no more than 440 evaluations. When evolving small populations, the search may saturate (despite the diversity constraint), e.g., a not-quite-optimal region of the configuration space is found from which the search cannot easily escape because all the individuals in the population contain the same traits from that region. In that case, the genetic algorithm is stuck until a sufficiently favorable mutation happens along. If one increases the population size too far, then the evolving the population will take longer. As such, there exists a population size which is a compromise between the possibility of saturation and the speed of evolution. It is overall less detrimental to operate with a larger population than to risk saturation.

Table II reports the number of evaluations required to solve question 1 with 95% degree of confidence when using bit-string Darwinian GA with and without mutations. The mutation parameters have been optimized. We find that mutations do have a very large impact on the efficiency of the real-space -GA search. Indeed, the mutation operators generally allow the search to explore a larger manifold by introducing new traits into the population. Although not shown here, we have also added real-space mutations to the reciprocal-space -GA search (e.g., in addition to the intrinsic mutations present within reciprocal-space mating). We find that this addition to -GA do not accelerate the search. It would seem that the intrinsic mutations of the reciprocal-space mating are sufficient. As such, the reciprocal-space GAs are more black box than the real-space GA.

Table III gives the population sizes used for each method
iesiesiesie(es)953(accelerat0a3a3a38ifa7TJ12.44-eTD2.2(he).6(TJ12.4–-eT

A T A T T A T A

One general question regarding the strategy we have chosen for all ground states, namely, to minimize $\Delta H(\sigma) - C(\sigma)$ rather than $\Delta H(\sigma)$ at fixed concentration σ, is that it constitutes a moving target, and thus may actually be more difficult to optimize. We find however that (i) a fixed concentration is a cumbersome constraint during mating operations, especially with the reciprocal-space mating and the Lamarckian refinements described in the paper, and (ii) it requires for N-atom configurations N independent GA minimization. Furthermore, there is evidence that removing the convex hull actually makes the space less complex. Figure 9 plots the required number of evolutions for finding the exact deepest ground state with respect to $2\times2\times$ supercells for all five search strategies presented here. In panel (a) of Fig. 9, the objective function minimized to solve question 1 is, as mentioned earlier in the text, the formation enthalpy $\Delta H(\sigma)$. On the other hand, in panel (b) the objective function becomes the depth to the known convex hull at iteration of the genetic algorithm, e.g., the objective function used previously to find *all* ground states, $\delta^{i}(\sigma) = \Delta H(\sigma) - C^{i}(\sigma)$. Remarkably, although this last objective is more complex and, in fact, changes during the course of the evolutionary run, both Darwinian GAs converge faster in panel (b), and both Lamarckian GAs and the virtual-atom approach converge faster in panel (b) for supercells larger than $2\times2\times8$.

*mayeul_davezac@nrel.gov
†alex_zunger@nrel.gov

[1] A. F. Wells, *S c al I ga ic Che i* (Oxford University, New York, 1984).
[2] *The Na e f he Che ical B d*, edited by L. Pauling (Cornell University Press, Ithaca, 1960).
[3] W. Hume-Rothery, R. E. Smallman, and C. W. Haworth, *The S c e f Me al a d All* (The Institute of Metals, London, 1969).
[4] *Pea ' Ha db k f C all g a hic Da a f I e e allic Pha e*, edited by P. Villars and L. D. Calvert (American Society for Metals, Metal Park, OH, 1985).
[5] R. M. Martin, *Elec ic S c e: Ba ic The a d P ac ical Me h d* (University Press, Cambridge, 2004).
[6] M. T. Yin and M. L. Cohen, Phys. Rev. Lett. **50**, 2006 (1983).
[7] S. Froyen and M. L. Cohen, Phys. Rev. B **28**, 3258 (1983).
[8] A. Mujica, A. Rubio, A. Muñoz, and R. J. Needs, Rev. Mod. Phys. **75**, 863 (2003).
[9] M. Wang, X. Hu, D. N. Beratan, and W. Yang, J. Am. Chem. Soc. **128**, 3228 (2006).
[10] S. Keinan, X. Hu, D. N. Beratan, and W. Yang, J. Phys. Chem. A **111**, 146 (2006).
[11] G. Trimarchi, P. Graf, and A. Zunger, Phys. Rev. B **74**, 014204 (2006).
[12] F. Ducastelle, *O de a d Pha e S abili i All (C he i a d S c e)* (North-Holland, Amsterdam, 1991).
[13] J. W. D. Connolly and A. R. Williams, Phys. Rev. B **27**, 5169 (1983).
[14] L. G. Ferreira, S.-H. Wei, and A. Zunger, Int. J. Supercomput. Appl. **5**, 34 (1991).
[15] A. Zunger, Proceedings of the NATO ASI on "Statistics and Dynamics of Alloy Phase Transformation (Plenum, New York, 1993), pp. 361–419.
[16] W. Lenz, Phys. Z. **21**, 613 (1920).
[17] E. Ising, Phys. Z. **31**, 253 (1925).
[18] W. L. Bragg and E. J. Williams, Proc. R. Soc. London, Ser. A **145**, 699 (1934).
[19] D. M. Deaven and K. M. Ho, Phys. Rev. Lett. **75**, 288 (1995).
[20] B. Hartke, J. Phys. Chem. **7**, 9973 (1993).
[21] A. R. Oganov and C. W. Glass, J. Chem. Phys. **124**, 244704 (2006).
[22] D. P. Stucke and V. H. Crespi, Nano Lett. **3**, 1183 (2003).
[23] N. L. Abraham and M. I. J. Probert, Phys. Rev. B **73**, 224104 (2006).
[24] G. Trimarchi and A. Zunger, Phys. Rev. B **75**, 104113 (2007).
[25] J. Kanamori and Y. Kakehashi, J. Phys. (Paris), Colloq. **38**, C7 (1977).
[26] M. K. Phani, J. L. Lebowitz, and M. H. Kalos, Phys. Rev. B **21**, 4027 (1980).
[27] A. Finel and F. Ducastelle, Europhys. Lett. **1**, 135 (1986).
[28] S. V. Barabash, V. Blum, S. Müller, and A. Zunger, Phys. Rev. B **74**, 035108 (2006).
[29] D. B. Laks, L. G. Ferreira, S. Froyen, and A. Zunger, Phys. Rev. B **46**, 12587 (1992).
[30] V. Blum and A. Zunger, Phys. Rev. B **72**, 020104(R) (2005).
[31] A. Zunger and S. Mahajan, *Ha db k Se ic d c* (Elsevier, Amsterdam, 1994), Vol. 3, p. 1399.
[32] M. Sanati, L. G. Wang, and A. Zunger, Phys. Rev. Lett. **0**, 045502 (2003).
[33] F. Barahona, J. Phys. A **15**, 3241 (1982).
[34] S. Istrail, Proceedings of the 32nd ACM Symposium on the Theory of Computing (STOC00) (ACM, Portland, OR, 2000), pp. 87–96.
[35] L. T. Wille and J. Vennik, J. Phys. A **18**, L419 (1985).
[36] A. B. Adib, J. Phys. A **38**, 8487 (2005).
[37] J. Crain, G. J. Ackland, and S. J. Clark, Rep. Prog. Phys. **58**, 705 (1995).
[38] C. W. Glass, A. R. Oganov, and N. Hansen, Comput. Phys. Commun. **175**, 713 (2006).
[39] M. d'Avezac and A. Zunger, J. Phys.: Condens. Matter **1** , 402201 (2007).
[40] The search presented in Ref. 39 and in this paper considers type (ii) problems, i.e., *dec a i* of an N-atom unit cell on a *ed* underlying Bravais lattice of $A_{1-}B$, simultaneously throughout the composition range, whereas the search in Ref. 21 is a type (iii) problem; i.e., it optimizes the lattice type, unit cell, and atomic positions (and hence the lattice decoration) of an $A B$ structure at fixed stoichiometry and . In both methods, structures are relaxed with respect to strain.
[41] N. A. Barricelli, Methodos **1 54**, 45.
[42] C. Darwin, *The O igi f S ecie* (Murray, London, 1859).
[43] J.-B. Lamarck, *Phil hie Z l gi e E i i de C -*

ida a i Rela iv e l'Hi i e Na. elle de A i a Dentu, Paris, 1809.

[44] Gradients of the free energy with respect to the fractional occupation of different sites occur in the concentration-wave theory of ordering Refs. 58–61. These gradients allow the determination of phase precursors to the partially ordered structure within