PHYSICAL REVIEW B **80**, 092101 (2009)

Predicting stable stoichiometries of compounds via evolutionary global space-group optimization

Giancarlo Trimarchi and Arthur J. Freeman
Department of Physics and Astronomy, Northwestern University, 2145 Sheridan Road, Evanston, Illinois 60208, USA

Alex Zunger*
National Renewable Energy Laboratory, Golden, Colorado 80401, USA

(Received 18 February 2009; published 3 September 2009)

Whereas the Daltonian atom-to-atom ratios in ordinary molecules are well understood via the traditional theory of valence, the naturally occurring stoichiometries in intermetallic compounds A_pB_q, as revealed by phase-diagram compilations, are often surprising. Even equal-valence elements A and B give rise to unequal (p,q) stoichiometries, e.g., the 1:2, 2:1, and 3:1 ratios in $\mathrm{Al}_p\mathrm{Sc}_q$. Moreover, sometimes different stoichiometries are associated with different lattice types and hence rather different physical properties. Here, we extend the fixed-composition global space-group optimization (GSGO) approach used to predict, via density-functional calculations, fixed-composition lattice types [G. Trimarchi and A. Zunger, J. Phys.: Condens. Matter **20**, 295212 (2008)] to identify simultaneously all the minimum-energy lattice types *throughout the composition range*. Starting from randomly selected lattice vectors, atomic positions and stoichiometries, we construct the $T=0$ "convex hull" of energy vs composition. Rather than repeat a set of GSGO searches over a fixed list of stoichiometries, we minimize the distance to the convex hull. This approach is far more efficient than the former one as a single evolutionary search sequence simultaneously identifies the lowest-energy structures at each composition and among these it selects those that are ground states. For Al-Sc we correctly identify the stable stoichiometries and relative structure types: AlSc_2-$\mathrm{B8}_2$, AlSc-B2, and $\mathrm{Al}_2\mathrm{Sc}$-C15 in the $N_{at}=6$ periodic cells, and $\mathrm{Al}_2\mathrm{Sc}_6$-$\mathrm{D0}_{19}$, AlSc-B2, and $\mathrm{Al}_3\mathrm{Sc}$-$\mathrm{L1}_0$ in the $N_{at}=8$ periodic cells. This extended evolutionary GSGO algorithm represents a step toward a fully *ab initio* materials synthesis, where compounds are predicted starting from sole knowledge of the chemical species of the constituents.

DOI: 10.1103/PhysRevB.80.092101 PACS number(s): 61.50.Ah, 61.66.−f

Dalton's hypothesis[1,2] that compounds manifest characteristic atom-to-atom ratios explained Proust's phenomenological "law of definite proportions" and set the stage for the subsequent development of the concepts of valence[3] as an explanation of such hitherto mysterious atomic ratios. Still, whereas these classic constructs of valence naturally explain the Daltonian stoichiometries of many molecules and covalent solids,[4] they have yet to demystify the atomic ratios even in the simple intermetallic compounds such as those made of elements of equal nominal valence, e.g., the occurrence of 1:2, 1:1, 2:1, and 3:1 ratios in Al-Sc,[5] or the existence of $\mathrm{Na}_2\mathrm{K}$ but not of NaK_2.[6] Indeed, inspection of recent compilations of all measured crystal structures[5,7] reveals numerous instances of Daltonian ratios that are unsuspected on the basis of the traditional theory of valence and bonding. At the same time, knowledge of the ratio between the different atomic constituents in a solid is needed for understanding its basic electronic and magnetic properties.

Total-energy minimization based on first-principles density-functional theory is a natural route to identifying stable, $T=0$ structures and stoichiometries. Yet, an objective search of Daltonian indices can be problematic. Such a search consists, in general, of two steps: first, finding the lowest-energy configuration at a fixed composition, and second, allowing competition between all of the fixed-composition lowest-energy structures. This is illustrated in Fig. 1 for a binary A_xB_{1-x} system: at the nth step of approximation (dashed blue line) one finds structures α, β, and γ as the lowest-energy structures at concentrations x, x', and x'', respectively, whereas in the $(n+1)$st step (continuous red line) the structure $\alpha(x)$ was recognized as having a higher energy then a combination of the two other structures $\beta(x')$ plus $\gamma(x'')$, and thus α was eliminated from the ground state line described by the convex hull of energy vs composition (the continuous red line in Fig. 1).

Performing both types of searches can be readily accomplished if all compounds in the $0<x<1$ concentration domain have the same underlying lattice type. In such a case, the problem of finding the structures belonging to the convex hull amounts to determining the lowest-energy decorations of a given lattice type by, say, A and B atoms. Both steps of the search noted above can then be effectively conducted within the cluster-expansion method[8] (see, for instance, Ref. 9). Here, *ab initio* formation energies of a few configurations are first mapped onto a generalized Ising Hamiltonian[10] which is then searched to determine the ground-state struc-

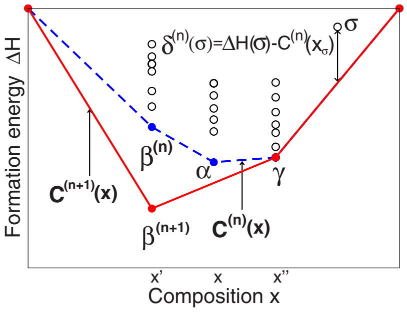

FIG. 1. (Color online) Schematic of the construction of the approximate convex hull $C^n(x)$ during the evolutionary search at the nth and $(n+1)$st generations. The fitness function $\delta^{(n)}(\sigma)$ of a structure σ is defined as the difference between its formation energy $\Delta H(\sigma)$ and the actual (approximate) convex hull $C^n(x_\sigma)$ at the nth generation.

tures among an astronomically large number of possibilities via, e.g., exhaustive evaluation[11] or genetic algorithms.[12,13] Such searches for fixed-lattice type systems have validated known Daltonian stoichiometries and have predicted unexpected stoichiometries, the latter being illustrated by "continuously adaptive structures" of Au-rich Cu-Au[14] or Fe-Co.[15]

Sometimes, however, the A and B constituent solids of binary alloys have different underlying lattice types (e.g., fcc Pd, bcc V or hcp Cd, fcc Pt), often leading to intermediate A_pB_q compounds with difficult-to-guess lattice types (e.g., the non-Bravais-based A15 structure of PdV_3 and the bcc-like CuPd emerging from fcc Cu and fcc Pd). In such cases, e.g., Cu-Pd, one can perform independent cluster expansions for each lattice type and then superpose the ensuing convex hulls, constructing the overall lowest. Alternatively, one can perform individual total-energy calculations of many guessed A_pB_q compounds (e.g., with structures and stoichiometries taken from experimental compilations of related materials[16]) and determine the convex hull from such a more limited list of candidates. More recently, the problem of finding unsuspected lattice types was addressed by solving the "global space-group optimization" (GSGO) problem,[17–19] in which the crystal structure is found in an unbiased way by starting from random lattice vectors and atomic positions, using evolutionary algorithms. Our approach to GSGO has proven to be effective in predicting nonintuitive crystal structures[20] of systems such as Al_2Sc_6, Al_6Sc_2, Al_4Sc_4, Cd_2Pt_6, Cd_4Pd_4, and Pd_6Ti_2. However, such a GSGO approach[17–19,21] is currently performed at fixed composition and stoichiometry A_pB_q, and so Daltonian proportions are not predicted. Here, we extend the *ab initio* evolutionary GSGO algorithm by lifting constraints on composition and by optimizing the formation energy ΔH convex hull for supercells of a given number of atoms N_{at}. As a test, we choose the Al-Sc system that manifests nonintuitive stoichiometries and stoichiometry-dependent lattice types. Independently started evolutionary sequences were able to identify the ground-state compositions and structures: (i) $AlSc_2$-$B8_2$, (ii) AlSc-B2, (iii) Al_2Sc-C15, (iv) Al_2Sc_6-$D0_{19}$, and (v) Al_3Sc-$L1_2$ (see Fig. 2). Here, the compounds of an alloy are predicted fully *ab initio* by an optimization algorithm that concurrently explores structures of all possible compositions and has as starting information only the chemical types of the constituent elements.

The evolutionary algorithm of Refs. 17–19 entails two steps which we will change in what follows: (i) setup a population of structures, all at the same, target composition which is kept fixed during the evolutionary search; and (ii) define the fitness score of a structure σ equal to its total energy $E(\sigma)$. We change these two steps as follows:

(i) The population of structures consists of an ensemble of any of the allowed compositions $A_pB_{N_{at}-p}$ with $1 \leq p \leq N_{at}-1$. The formation energy per atom of a binary A_pB_q structure at $T=0$ K is defined as

$$\Delta H(A_pB_q) = E(A_pB_q) - (1-x)E(A) - xE(B), \qquad (1)$$

where $E(A_pB_q)$, $E(A)$, and $E(B)$ are, respectively, the equilibrium total energies per atom of the given binary structure and of the end-point solids A and B; $x=\frac{q}{p+q}$ is the concentration (mole fraction) of B. $\Delta H(A_pB_q)$ corresponds to the $T=0$ Gibbs free energy of the A_pB_q configuration minus the chemical potentials of the A and B constituents (i.e., their equilibrium $T=0$ total energies per atom) each taken with a weight equal to the mole fraction in the A_pB_q configuration. Mating is performed also between structures of different compositions. Both the mating and mutation operations are allowed to produce child structures with compositions that differ from those of the parents, thereby making possible for the evolutionary search to survey simultaneously structures across the whole composition range.

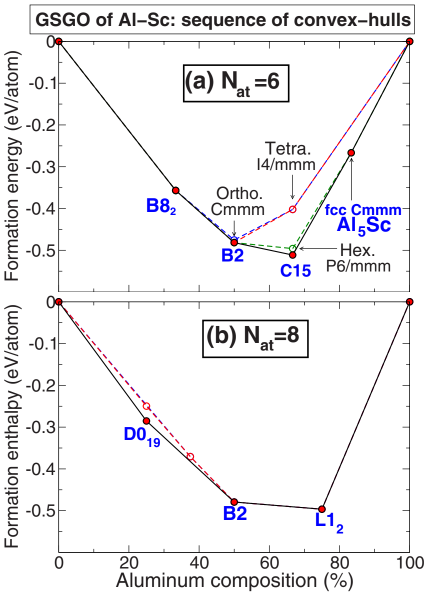

FIG. 2. (Color online) Ground-state search of Al-Sc with (a) N_{at}=6 and (b) 8. Open markers and the dashed lines connecting them depict, respectively, the approximate ground-state structures and the corresponding convex-hull evolving toward the optimal convex hull (solid markers and line).

(ii) The fitness score is defined as the difference between the formation energies ΔH of an $A_{1-x}B_x$ structure and the current convex hull. This is done as follows: for a structure α to be a ground state its $\Delta H(\alpha)$ must lie below the tie line between the formation energies of any pair of structures β and γ with $x_\beta < x_\alpha < x_\gamma$, i.e., the structure must be stable against disproportionation into structures at two neighboring concentrations. This means[11] (see Fig. 1)

$$\Delta H(\alpha) \leq \Delta H(\gamma)\frac{x_\alpha - x_\beta}{x_\gamma - x_\beta} + \Delta H(\beta)\frac{x_\gamma - x_\alpha}{x_\gamma - x_\beta}. \qquad (2)$$

The vertices of the line so obtained correspond to the ground-state structures defining the convex hull, line $C(x)$. Since the free energies per atom of all structures sampled in the evolutionary search are referred to the chemical potentials of the pure elements, the stoichiometric balance between the ΔH of a given structure and the phase-separation term on the right-hand side of Eq. (2) makes no sum of chemical potentials of the elemental constituents appear in this relation. To search for the ground-state structures (not only for the lowest formation energy structures at a given

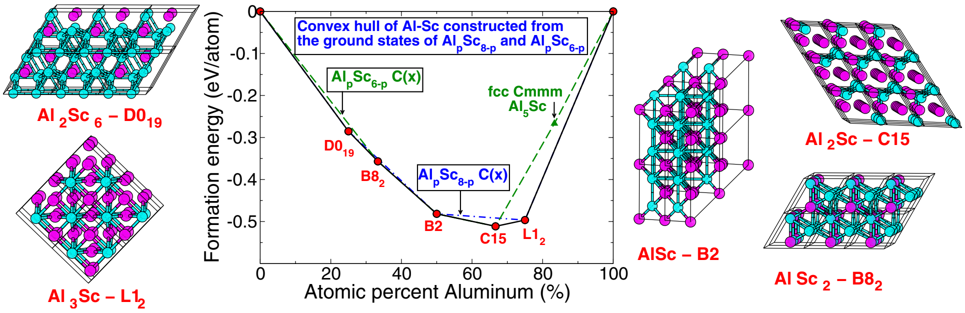

FIG. 3. (Color online) Ground-state convex hull of Al-Sc (solid line) and corresponding ground-state structures determined by combining the convex-hull lines of Al-Sc periodic cells with $N_{at}=6$ and 8 atoms [respectively, green (dashed) and blue (dot-dashed) lines].

composition), we iteratively refine this approximate convex hull, $C^{(n)}(x)$, by using the objective function introduced by d'Avezac and Zunger in Ref. 13,

$$\delta^{(n)}(\sigma) = \Delta H(\sigma) - C^{(n)}(x_\sigma), \tag{3}$$

i.e., the distance of the formation energy $\Delta H(\sigma)$ of σ from the convex hull $C^{(n)}(x)$ at the nth generation. Figure 1 depicts the construction of the approximate convex hull and the calculation of the objective function. The evolving approximate ground-state structures are on the convex hull at each generation and have $\delta^{(n)}(\sigma)=0$. The structures with the largest values of $\delta^{(n)}(\sigma)$ are those farthest away from the convex hull and are replaced at each generation of the evolutionary sequence, thereby driving the evolutionary search toward the lowest-energy structures at each composition–which furthers the refinement of the convex hull.

The evolutionary search proceeds as follows: a population of N_{pop} candidate structures with N_{at} atoms per cell is created. The initial population is determined by randomly selecting lattice vectors, atomic positions, and composition $A_pB_{N_{at}-p}$. The population is evolved through a sequence of generations where at each generation, the current convex hull $C^{(n)}(x)$ is constructed and the N_{rep} structures farthest away from it are replaced by new ones generated by mating or mutation. Before any crossover or mutation, the parent crystal structures are subjected to a similarity transformation[17] which maps the atomic positions onto fractional coordinates. The crossover of two structures picked from the current population is done via real-space cut and splice[17] where, corresponding slabs in the two parents are swapped to generate the child structures. The structures obtained from mating and mutation are subjected to full local structural relaxation and the relaxed structures are included in the $(n+1)$st generation of the population. The sequence of generations is carried out until a given stopping criterion is met, e.g., the allocated number of generations or of local structural relaxations is exhausted.

The evolutionary search described here does not require the number of atoms N_{at} per supercell to be fixed. Indeed, at a concentration x one can have structures of composition A_pB_q with any $N_{at}=p+q$ so that $\frac{q}{p+q}=x$. Furthermore, having N_{at} vary during the search allows one to survey structures in a denser set of concentrations $x=q/N_{at}\in[0,1]$, and, correspondingly, a more complete space of periodic arrays of A and B atoms, than would be possible by keeping N_{at} constant. Here, however, we test the variable-composition evolutionary algorithm by applying it to few numbers of atoms N_{at} per periodic cell, one at a time. All concentrations $x=q/N_{at}$ with $1<q<N_{at}-1$ are simultaneously surveyed by a single evolutionary sequence, as opposed to a "do-loop" approach that would require $N_{at}-1$ independent search runs, one for each of the $N_{at}-1$ binary concentrations $x=q/N_{at}$ in N_{at} atom cells.

One might wonder what forces the variable-composition GSGO to consider the whole composition range. Indeed, the variable composition algorithm will concentrate (for a completely flat energetic landscape) on the region around $x=0.5$, because that is where the largest number of decorations/structures reside. As the energy landscape is not flat, the variable-composition GSGO will compromise between the region with the most structures, and that with the lowest-energy ones, i.e., doing its job correctly. Another way to put it is that neither mutation nor crossover produce structures with lower or higher energy than the parent ones, i.e. they do not play the role of Maxwell's demon. Then, offspring structures are more likely to be created in the largest combinatorial region.

The system we choose for this test is the Al-Sc metallic alloy, where Al and Sc are, respectively, fcc and hcp solids. The compounds of Al-Sc known from experiment[5,7] are (i) fcc-based Al_3Sc-$L1_2$, (ii) Al_2Sc-C15, (iii) the bcc-based AlSc-B2 structure, and (iv) the hcp-based $AlSc_2$-$B8_2$ (see structure models in Fig. 3). Curtarolo *et al.*[16] calculated the total energy of 176 candidate structure types (taken from data mining compilations[7]) and confirmed that these four experimentally observed structures are ground states among this limited candidate list. They also predicted an additional ground state at composition $AlSc_3$ with the hcp-based $D0_{19}$ structure, where the experimental phase diagram[5,7] indicates instead the coexistence of the Al and $AlSc_2$-$B8_2$ solid phases at $T=0$ K. Thus, predicting the ground states of the Al-Sc system is a challenging GSGO problem because the C15 structure is not a simple decoration of a parent Bravais lattice and, along with B2, can hardly be inferred from the fcc and hcp lattices of the constituents.

In an earlier GSGO investigation of Al-Sc, Trimarchi and Zunger[20] applied the fixed composition evolutionary GSGO procedure to Al_2Sc_6, Al_4Sc_4, Al_6Sc_2, and successfully retrieved the lowest-energy structures at each of these compositions. Here, the *ab initio* evolutionary ground-state search of Al-Sc[22] is done taking $N_{at}=6$ and 8 atoms per supercell. Each randomly started evolutionary sequence was usually carried out for at least 25 generations and included at least 100 structural evaluations. For $N_{at}=6$, we produced five independent sequences with $N_{pop}=16$ and $N_{rep}=4$. Three ground-state structures of Al-Sc are representable in $N_{at}=6$ periodic cells, i.e., $AlSc_2$-$B8_2$, AlSc-B2, and Al_2Sc-C15. Figure 2(a) summarizes the history of one of the evolutionary runs for $N_{at}=6$. At the initial generation, the convex hull shows three breaking points: $AlSc_2$-$B8_2$, AlSc-B2, and a tetragonal $I4/mmm$ Al_4Sc_2 structure. As the search unfolds, the Al_4Sc_2 structure changes first into a hexagonal $P6/mmm$ structure and, at the 23rd generation (i.e., after 108 local structural relaxations), into C15, with an energy difference of 15 meV/at between these two structures. The $N_{at}=6$ final convex hull has a vertex also at Al_5Sc. All the independent restarts with $N_{at}=6$ found the $AlSc_2$-$B8_2$ and AlSc-B2 ground states, and stable structures at Al_5Sc and Al_4Sc_2. The fcc-based Al_5Sc *Cmmm* structure was always found to be stable in the six-atom periodic cells, while Al_4Sc_2-C15 was obtained in two restarts and in the remaining three restarts the hexagonal $P6/mmm$ was found instead.

Three independent sequences were produced for $N_{at}=8$, with $N_{pop}=40$ and $N_{rep}=8$. Figure 2(b) shows the iteration history of one of these sequences. In the first four generations, the algorithm found Al_4Sc_4-B2, Al_6Sc_2-$L1_2$, and breaking points also at Al_2Sc_6 and Al_3Sc_5 compositions. At the fifth iteration (i.e., after 80 relaxed total-energy evaluations), Al_2Sc_6-$D0_{19}$ was obtained and the Al_3Sc_5 was removed from the convex hull. Hence, this evolutionary sequence found all the ground-state structures that are realizable in Al_pSc_{8-p} periodic cells. All the $N_{at}=8$ runs found Al_4Sc_4-B2 and Al_6Sc_2-$L1_2$. Furthermore, one of the three restarts found Al_2Sc_6-$D0_{19}$ while the other two retrieved Al_2Sc_6-$L1_2$. To determine the final convex hull of Al-Sc, we apply the condition of stability against disproportionation of Eq. (2) to determine the stable structures out of all the ground-state structures corresponding to all values considered for N_{at}. This final step is illustrated by Fig. 3, where the convex-hull lines of the Al_pSc_{6-p} and Al_pSc_{8-p} periodic cells are combined to give the final convex hull of Al-Sc, whose ground states are $AlSc_2$-$B8_2$, AlSc-B2, Al_2Sc-C15, Al_2Sc_6-$D0_{19}$, and Al_2Sc_6-$L1_0$ (see Fig. 3), i.e., all those previously known.

The success of the present stoichiometry unconstrained evolutionary algorithm in determining *all* the ground states of the nontrivial Al-Sc system shows that this method has a truly global view of the space of periodic arrangements in binary compounds. Whereas the fixed stoichiometry GSGO (Refs. 17–19) requires one to know in advance the (p,q) the present variable-composition GSGO combines the ability of cluster expansion to find the convex hull over all compositions with the GSGO's ability to explore different lattice types. In the present scheme, the only input is the chemical species of the constituents of the material, so, here, a completely *fully ab initio*, unbiased prediction of the $T=0$ structures of a material is performed.

In summary, we extended the earlier fixed composition GSGO evolutionary algorithm to determine the stoichiometries A_pB_q and the optimal lattice type of the stable phases simultaneously across the whole composition range. As a test, we searched the ground-state structures of the Al-Sc system and successfully determined all its stable $T=0$ compounds.

We thank Mayeul d'Avezac for useful discussions on this subject. Work at Northwestern University was supported by the AFOSR (Grant No. FA9550-07-1-0174). Work at NREL funded by DOE-SC-BES-MSED through NREL under Contract No. DE-AC36-08GO2830.

---

*alex.zunger@nrel.gov

[1] J. Dalton, *A New System of Chemical Philosophye* (1808).

[2] See also: http://web.lemoyne.edu/~giunta/dalton.htm and http://antoine.frostburg.edu/chem/senese/101/atoms/dalton-postulates.shtml

[3] G. N. Lewis, J. Am. Chem. Soc. **38**, 762 (1916).

[4] Ironically, Dalton described water as OH, because he postulated that compounds made of two types of elements will always have 1:1 ratio.

[5] *Binary Alloy Phase Diagrams*, edited by T. B. Massalski (ASM International, Ohio, 1992).

[6] F. Laves and H. J. Wallbaum, Z. Anorg. Allg. Chem. **250**, 110 (1942).

[7] *Pauling File: Inorganic Materials Database and Design System—Binaries Edition*, edited by P. Villars (ASM International, Ohio, 2002).

[8] A. Zunger, in *First Principles Statistical Mechanics of Semiconductor Alloys and Intermetallic Compounds*, edited by P. Turchi and A. Gonis (Plenum, New York, 1994), p. 361.

[9] M. H. F. Sluiter, C. Colinet, and A. Pasturel, Phys. Rev. B **73**, 174204 (2006).

[10] J. M. Sanchez, F. Ducastelle, and D. Gratias, Physica A, **128**, 334 (1984).

[11] L. G. Ferreira, S.-H. Wei, and A. Zunger, Int. J. Supercomput. Appl. **5**, 34 (1991).

[12] G. Trimarchi, P. Graf, and A. Zunger, Phys. Rev. B **74**, 014204 (2006).

[13] M. d'Avezac and A. Zunger, Phys. Rev. B **78**, 064102 (2008).

[14] M. Sanati, L. G. Wang, and A. Zunger, Phys. Rev. Lett. **90**, 045502 (2003).

[15] R. Drautz, A. Diaz-Ortiz, M. Fähnle, and H. Dosch, Phys. Rev. Lett. **93**, 067202 (2004).

[16] S. Curtarolo, D. Morgan, and G. Ceder, CALPHAD: Comput. Coupling Phase Diagrams Thermochem. **29**, 163 (2005).

[17] N. L. Abraham and M. I. J. Probert, Phys. Rev. B **73**, 224104 (2006).

[18] A. R. Oganov and C. W. Glass, J. Chem. Phys. **124**, 244704 (2006).

[19] G. Trimarchi and A. Zunger, Phys. Rev. B **75**, 104113 (2007).

[20] G. Trimarchi and A. Zunger, J. Phys.: Condens. Matter **20**, 295212 (2008).

[21] Y. Yao, J. S. Tse, and K. Tanaka, Phys. Rev. B **77**, 052103 (2008).

[22] Local crystal-structure relaxations were perfomed by using the VASP (Ref. 23) package, with ultrasoft pseudopotentials, GGA to exchange and correlation, and a planewave cutoff of 240 eV. The BZ integration was done with uniform k-point grids with a resolution of $2\pi\times0.04$ and $2\pi\times0.02$ Å^{-1}, respectively, for the structural relaxation and the calculation of the total energy of the relaxed structure.

[23] G. Kresse and J. Furthmuller, Comput. Mater. Sci. **6**, 15 (1996).