Global space-group optimization problem: Finding the stablest crystal structure without constraints

Giancarlo Trimarchi and Alex Zunger
National Renewable Energy Laboratory, Golden, Colorado 80401, USA
(Received 13 November 2006; published 26 March 2007)

Finding the most stable structure of a solid is one of the central problems in condensed matter physics. This entails finding both the lattice type (e.g., fcc, bcc, and orthorhombic) and (for compounds) the decoration of the lattice sites by atoms of types *A*, *B*, etc. ("configuration"). Most approaches to this problem either assumed that both lattice type and configuration are known, optimizing instead the cell volume and performing local relaxation. Other approaches assumed that the lattice type is known, searching for the minimum-energy decoration. We present here an approach to the global space-group optimization (GSGO) problem, i.e., the problem of predicting both the lattice structure and the atomic configuration of a crystalline solid. This search method is based on an evolutionary algorithm within which a population of crystal structures is evolved through mating and mutation operations, improving the population by substituting the highest total-energy structures with new ones. The crystal structures are not represented by bit strings as in conventional genetic algorithms. Instead, the evolutionary search is performed directly on the atomic positions and the unit-cell vectors after a similarity transformation is applied to bring structures of different unit-cell shapes to a common basis. Following this transformation, we can define a crossover operation that treats, on the same footing, structures with different unit-cell shapes. Once a new structure has been generated by mating or mutation, it is fully relaxed to the closest local total-energy minimum. We applied our procedure for the GSGO in the context of pseudopotential total-energy calculations to the semiconductor systems Si, SiC, and GaAs and to the metallic alloy AuPd with composition Au_8Pd_4. *Starting from random unit-cell vectors and random atomic positions,* the present search procedure found for all semiconductor systems studied the correct lattice structure and configuration. In the case of Au_8Pd_4, the search retrieved the correct underlying fcc lattice, but energetically closely spaced ($\sim$2 meV/at.) alloy configurations were not resolved. This approach to GSGO opens the way to predicting unsuspected structures by direct optimization using, in the cases noted above, an order of 100 total-energy *ab initio* calculations.

DOI: 10.1103/PhysRevB.75.104113 PACS number(s): 61.50.Ah, 61.66.−f

I. INTRODUCTION

A central feature of the solid-state physics of crystalline solids is the existence of a very rich diversity of stable crystal forms,[1,2] distinguished by their lattice type (e.g., fcc, bcc, hcp, and monoclinic) and by the pattern of occupation of the lattice sites by atom types ("configuration" or "decoration"). Indeed, a central theme in theoretical physics of crystalline solids is the quest for prediction of the stable crystal structures of a *given A-B* (or more complex) periodic solid.[3–7] While early attempts were based on correlating observed structures with elemental scales such as electronegativity,[4] orbital radii,[5,6] or electron-atom ratio,[7] modern attempts are all based on *optimizing the quantum-mechanical total energy* of a solid as a function of its structural degrees of freedom. One can recognize three basic types of searches for stable crystal forms distinguished by the level of restriction imposed on the structural degrees of freedom being optimized.

Type-I optimization involves cases where both the lattice type and the configuration (decoration of sites by *A*- or *B*-type atoms) are assumed at the outset. This includes optimization of the unit-cell volume in known structures,[8–10] optimization of cell-external degrees of freedom (e.g., c/a tetragonal ratio), and symmetry-undetermined cell-internal atomic relaxations in solids of known structure types.[10,11] The optimization methods used for such type-I problems are continuous position-space optimization such as gradient-guided searches[12] and conventional[13] and Car-Parrinello[14] molecular dynamics. These methods have a good local vision but are not concerned with global optimization. Essentially, one looks for the nearest local minimum of a more or less known structural topology. More complex examples of type-I optimization include relaxation of surface structures for given semiconductor topologies[15] and local relaxation around a point defect in semiconductors.[16] A solution to type-I optimization problems has led to the establishment of the equation of state of numerous materials in known structures[8,10,11] as well as to the discovery of surface reconstruction patterns of given semiconductor surface topologies[15] or to the local geometry of the leading defect structures in semiconductors.[16]

Type-II optimization involves cases where the lattice type is given, but its specific decoration by *A*, *B*, etc., atoms (configuration) is unknown. A classic example is finding the ground states among 2^N candidate configurations created occupying the *N* lattice sites of a *fixed underlying Bravais* lattice by either *A*-type or *B*-type atoms.[17] Here, the optimization methods of choice are not *continuous* positional space techniques such as needed in type-I problems, but rather *discrete* methods, including exhaustive enumeration,[18] linear programing,[19] and discrete sampling.[20] Molecular dynamics is not the method of choice here because of its lack of global vision, the slowness of *swapping* a large number of atoms in a concerted fashion, and the difficulty of overcoming a large

number of local activation barriers. Solution to type-II optimization problems has led to the discovery of many previously unknown ground-state configurations in fcc-based[21] and bcc-based[22,23] alloys via search of lowest-energy cluster expansion (Ising type) Hamiltonians.[24]

Type-III optimization problems involve cases where neither the underlying lattice type nor the atomic configurations are known. Examples involve cases where the presumed lattice-type (e.g., fcc) changes to another fundamental type. For example, whereas solid Cu and Pd are fcc in their elemental form, their 50-50 compound has a bcc symmetry.[1] Another class of examples is represented by structures which exhibit dynamic (i.e., phonon) instabilities that carry the system into a new lattice type and configurational arrangement.[25] This includes the instabilities of the initially NaCl-type,[25] CsCl-type,[26] or beta-Sn-type[25] binary semiconductors, which transform spontaneously under pressure into different lattice types. In this type of optimization problems, one has a set of lattice vectors $\{\mathbf{a}_1,\mathbf{a}_2,\mathbf{a}_3\}$ and a set of N atomic position vectors $\{\mathbf{r}_i\}$, where $i=1,\ldots,N$, and these are altered simultaneously, starting from a random arrangement, in search of a minimum of the total energy. It is this type of global space-group optimization (GSGO) problems that we discuss in the present paper.

Variable cell molecular dynamics[27,28] is not usually adequate to address the global optimization problem in solids due to the existence of substantial energy barriers and the slow rate of transition between different configurations. In the recently introduced metadynamics approach,[29] local potential-energy wells are subjected to gradual deformation that makes them shallower, thus allowing the simulation of structural transformations which are characterized by high-energy barriers.

The difficult part in any type-III optimization methods is to identify the lowest-energy structure out of a dense manifold of structures that have similarly low energies and are separated by significant activation barriers and complicated paths in configuration space. Different material types differ in the identity of the low-energy structural excitations. In general, in covalently bonded systems, different *configurations* (e.g., anions bonded by anions vs anions bonded by cations) have rather different energies (since violations of the octet rule are associated with large energy penalties), but different *lattice types* (e.g., wurtzite vs zinc blende) often represent low excitation energies.[4–6] Thus, in covalent systems, the challenging problem (resolving low excitations) is often finding the equilibrium *lattice type*. In contrast, in many metal alloy systems, the configurational energy associated with swapping A and B atoms is rather small,[21,23,30] but different lattice types (e.g., fcc-type $L1_0$ vs bcc-type $B2$) represent higher excitation energies. Thus, in metallic alloys, the challenging optimization problem is often finding the equilibrium atomic configuration.

In this paper, we discuss an approach to tackle type-III (i.e., GSGO) problems. This method is based on an evolutionary algorithm,[31] following earlier works of Deaven and Ho,[32] Abrahams and Probert,[33] Oganov *et al.*,[34] and Oganov and Glass.[35] Evolutionary algorithms have a global vision of the search space and are less prone to get trapped in local minima. We have applied this search algorithm to semiconductor materials and to metallic alloys, starting in all cases with a random set of lattice vectors and a random set of atomic positions. Using less than 100 *ab initio* total-energy calculations, we recover for semiconductors Si, GaAs, and SiC the correct structures (both lattice types and configuration, where applicable). For AuPd, we correctly identify the lattice type, but energetically closely spaced ($\sim$2 meV/atom) configurations are not resolved during $\leqslant$10 simulation steps.

II. GENERAL METHODOLOGY: THE APPROACH TO GLOBAL SPACE-GROUP OPTIMIZATION

Here, we describe the main ideas of the method for searching the structure of lowest energy of a solid, including finding both the lattice type and the decoration of its sites by different atom types. Several of the methods used previously for solving the GSGO problem explore the configuration space of the solid by sampling one equilibrium structure at the time and hopping from one structure to an adjacent structure. This class of methods includes, e.g., simulated annealing,[36] basin hopping,[37,38] and minima hopping.[39] The method at hand searches, instead, for the global minimum of the potential-energy surface of a solid using an evolutionary algorithm that simultaneously explores several parts of configuration space. This procedure is inspired by evolutionary algorithms based on the principle of the survival of the fittest.[31] One defines a set ("population") of crystal structures ("individuals"), each representing a possible realization of a lattice configuration. For each such individual in the population, one defines its fitness score as the total energy of that lattice configuration. The population is allowed to evolve by replacing a given number of the highest-energy lattice configurations by new ones, generated from those present in the population by the operations of "mating" and "mutation." New "generations" of the population are created iteratively until some stopping criterion is met (e.g., the allocated number of generations is exhausted). Replacing the highest-energy structures in the population with new ones aims at sampling new parts of the configuration space associated with local minima whose energy may be lower than that of the local minima previously found. The problems in using such *standard* genetic algorithm approaches for type-III structural optimization are threefold.

First, if each structure is described as a bit-string representation, then there is no simple way to mate structures whose unit cells have different shapes. This pertains to mating different cell shapes belonging to the same Bravais lattice (as in type-II problems) or to mating different Bravais lattices, or lattices with different Wyckoff positions or lattices with different space groups (as in type-III problems). Instead of attempting to encode the atomic structure into a bit string of 1's and 0's, we represent here each structure using directly its atomic positions and unit-cell vectors. To enable the crossover of structures that have different cell shapes, we subject them to a transformation,[33] which brings different shapes to a common basis (see Sec. III A).

Second, in a bit-string-based genetic algorithm, it may happen that the mating operation does not allow certain

structural motifs present in the parents to propagate to the children. To favor the propagation of such local structural motifs to the offspring, we mate two crystal structures using a cut-and-splice procedure which generates a child structure by matching real-space segments of two parent lattice configurations (see Sec. III B). For mutation, given a structure, we generate a new one altering randomly the atomic positions and/or the cell vectors (see Sec. III C). In multispecies systems, we also explore the configuration that corresponds to a given lattice structure by atom swaps (see Sec. III D).

Third, candidate structures obtained by mating and mutation may represent chemically unreasonable configurations (e.g., pathologically short bond lengths). Progressing with genetic algorithm (GA) evaluation on such structures may thus take very long to converge. To accelerate convergence, two operations are performed: (i) we screen the offspring, rejecting the newly generated structures that have unphysically short atomic distances or unit cells with cell vectors that are too short or cell angles that are too small (see Sec. III E); (ii) we perform total-energy structural relaxation on the mated and/or mutated lattice configurations to obtain the closest local equilibrium structure (see Sec. III F).

III. THE METHOD IN MORE DETAIL

A population of N_{pop} crystal structures is defined, where each structure $\mathcal{S}=\{\mathbf{a}_1,\mathbf{a}_2,\mathbf{a}_3|\mathbf{r}_1,\ldots,\mathbf{r}_N\}$ contains in the unit cell N atoms of N_{type} atomic types, and is specified by the triad of lattice vectors $\{\mathbf{a}_1,\mathbf{a}_2,\mathbf{a}_3\}$ and by the atomic positions $\{\mathbf{r}_i\}$, where $i=1,\ldots,N$. The population evolves through a series of generations, each one consisting in replacing the N_{rep} highest-energy structures with new ones produced by mating or mutation through the following steps.

A. Creating common shapes: Similarity transformation

As a preliminary step to both the mating and the mutation operations, the atomic positions $\mathbf{r}_i$ are transformed to the fractional coordinates[33] $\mathbf{s}_i$ defined by the relation $\mathbf{r}_i=A\mathbf{s}_i$, where $A=[\mathbf{a}_1,\mathbf{a}_2,\mathbf{a}_3]$ is the 3×3 matrix of the lattice vectors. If B is the matrix defined by

$$B=[\mathbf{b}_1,\mathbf{b}_2,\mathbf{b}_3]=A^{-1}, \tag{1}$$

then the relation

$$\mathbf{s}_i=B\mathbf{r}_i \tag{2}$$

holds between the fractional coordinates $\mathbf{s}_i$ and $\mathbf{r}_i$. Such a transformation maps the original cell shape onto a cubic cell whose cell vectors have unitary length, whereas the reduced coordinates $\{s_i^{(1)},s_i^{(2)},s_i^{(3)}\}$ are between 0 and 1. The representation of the crystal structures in terms of fractional coordinates makes it easier to formulate a procedure for mating (see Ref. 33) two structures ("parents") that have different cell shapes and generating a new ("child") structure.

B. Real-space mating: Cut-and-splice crossover

Let us consider the two parent geometries $\mathcal{S}_1^{(p)}$ and $\mathcal{S}_2^{(p)}$ chosen randomly among the structures in the population: the child structure $\mathcal{S}^{(c)}$ is created using a procedure based on the idea of cut and splice of Deaven and Ho,[32] adapted to periodic systems by Abraham and Probert in Ref. 33, by Oganov *et al.* in Ref. 34, and by Oganov and Glass in Ref. 35. After the similarity transformation of Eq. (2) has been performed, two cuts are defined in the unit cells of $\mathcal{S}_1^{(p)}$ and $\mathcal{S}_2^{(p)}$ (see Fig. 1). These cuts are parallel to each other and perpendicular to one of the lattice vectors of the cell. As a result, the unit cells of both parent structures are divided into two slabs. The new structures are produced swapping the slabs defined in the supercell of the two parents, as shown in Fig. 1. One of the newly produced sets of atomic positions (given in fractional coordinates) is then chosen as child structure. The last step of the crossover assigns to the child the triad of cell vectors. We assign to the new structure a triad of cell vectors $\{\mathbf{a}_i\}$ defined as a linear combination of the lattice vectors of the parents with a weight factor that is chosen randomly between 0 and 1.

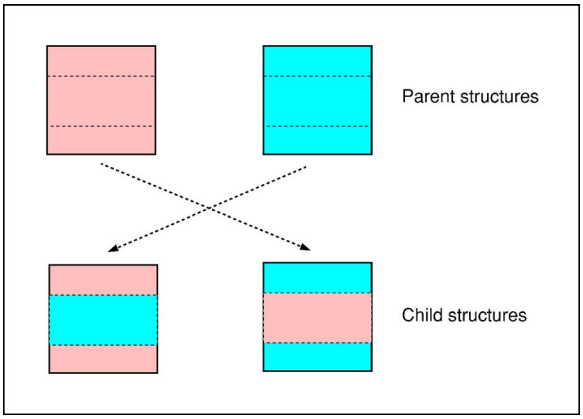

FIG. 1. (Color online) Generation of the "child" structures by the cut-and-splice procedure applied to two "parent" structures (see Sec. III B). The similarity transformation is performed before the cut-and-splice operation.

C. Mutation

The mutation operation consists in subjecting the atomic positions to a random change of maximal amplitude δ_{mut}. To mutate the cell-vector matrix $A=[\mathbf{a}_1,\mathbf{a}_2,\mathbf{a}_3]$, we use the transformation defined by a symmetric matrix[35] Ω, whose elements are chosen randomly between -1 and 1. The matrix A' of the transformed unit-cell vectors is then obtained by the relation

$$A'=(\mathbf{1}+\Omega)A, \tag{3}$$

where $\mathbf{1}$ is the unit matrix. In a multispecies system, the mutation of the cell-internal and the cell-external degrees of freedom is followed by a mutation of the configuration (see Sec. III D).

D. Exploring the configuration space by atom swaps

To explore the set of configurations in a multispecies system, one needs to scan a range of different occupations (e.g., by A or B) of the lattice sites. This implies searching all

decorations of the sites. If the numbers of sites per cell is small, one could enumerate all configurations and evaluate the fitness for each of them. If the number of sites is large, we start from the configuration that a newly generated individual inherits by mutation or mating and perform on it a series of atom swaps, i.e., a series of permutations of the position of pairs of unlike atoms, to produce a new configuration. Usually, in the calculations that are presented in the following, we performed up to ten atom swaps per newly generated structure.

E. Screening of the offspring

A structure produced by crossover or mutation is rejected whenever it contains the wrong number of atoms or does not have the composition that is considered for that compound. Moreover, it might also happen that the structures that result from the crossover or the mutation contain atoms that are too close to each other or have the unit cell with angles that are too small or cell vectors that are too short. Therefore, a newly generated structure is included in the population only if (i) the nearest-neighbor atomic distances are longer than a minimum distance d_{min} that is about 80% of the typical bonding length in the system, (ii) the cell angles are between 45° and 135°, and (iii) the cell-vector lengths are not shorter than the typical bonding distance. Whenever a structure generated by crossover or mutation does not meet these geometrical requirements, the crossover or the mutation is repeated until a valid configuration is obtained.

F. Refinement

After the child structure is accepted in the population, its cell shape and the atomic positions are relaxed to the nearest total-energy local minimum and the equilibrium total-energy is assigned to it as fitness score. In the present work, the total energy is represented by the pseudopotential plane-wave approach to the local-density approximation[40] with computational parameters described in the Appendix. To fully relax the new structure, we employ a conjugate-gradient algorithm that uses the gradients with respect to the atomic positions (i.e., atomic forces[40]) and the gradients with respect to the cell vectors (i.e., components of the stress tensor[41]). This relaxation[32] performed on each individual in the population makes the present optimization procedure more efficient in surveying the search space than algorithms based only on the generation of new structures by mating and mutation alone. Indeed, after relaxation, an individual is associated with an entire zone of the configurational space for which it represents the local equilibrium structure rather than just one particular realization of the system. The replacement of a significant number of the highest-energy structures in the population by new ones allows the evolutionary search to explore at each generation new parts of the configurational space in a parallel fashion. This is an important aspect of the search algorithm that avoids trapping in local minima of the total energy. Moreover, the fact that the search proceeds by the generation of new structures from locally optimized ones allows the ordered motifs that might have appeared in some individuals to propagate to the new generation.

IV. SIZE OF THE SEARCH SPACE

To estimate the number of realizations of the crystal structure included in the search space of a monoatomic system or of a binary alloy, we follow the simple argument proposed in Ref. 35. Let us define within the supercell a regular mesh of points with resolution δ and make the simplifying assumption that the atomic positions allowed are the points of the mesh. If V is the volume of the supercell, the number M of points of the mesh is equal to $\frac{V}{\delta^3}$. Reasonably, one can take δ equal to the typical value of the bond distance in the material under study. As a result, the number Q of crystal structures of a monoatomic system is

$$Q = \binom{M}{N}. \tag{4}$$

Now, let us consider a binary system A_nB_{N-n}: in such a case, yet under the assumption of discrete positions, one has to count for each realization of the crystal lattice of N sites all the possible ways to decorate it with n atoms of type A and $N-n$ atoms of type B, i.e., all the allowed configurations corresponding to the given atomic positions. Therefore, one has

$$Q = \binom{M}{N}\binom{N}{n}. \tag{5}$$

As an example, we can consider the A_5B_5 system represented by a cubic supercell having edges 10 Å long and define in such a cell a regular grid of resolution δ=1.0 Å. Under these assumptions the number Q of possible configurations is 6.6×10^{25}. In Sec. V, we estimate, following the simple argument given above, the size of the search space for each system for which a GSGO has been performed.

V. RESULTS OF THE GSGO

In the GSGO searches described in this section, we used the total-energy pseudopotential method as implemented in the VASP (Refs. 42 and 43) code to fully relax the unit-cell vectors and the atomic positions of all the structures that were generated (with parameters given in the Appendix for all systems studied).

A. Silicon

The first material for which we perform a GSGO search is silicon, which represents a nontrivial case for its tendency to form amorphous structures. Upon amorphization, three- and fivefold coordinated atoms appear in the GSGO simulation, while the fourfold coordinated atoms are at the center of distorted tetrahedra that form a disordered array. One can distinguish structural excitations at two different energy scales: high-formation-energy structures containing three- or fivefold coordinated Si atoms or low-formation-energy structures exhibiting distorted tetrahedra centered around fourfold coordinated Si atoms. Therefore, dealing with silicon, the difficult task is retrieving the ordered array of tetrahedrally coordinated atoms. In our GSGO search, we used eight atoms per supercell and performed a full relaxation of each

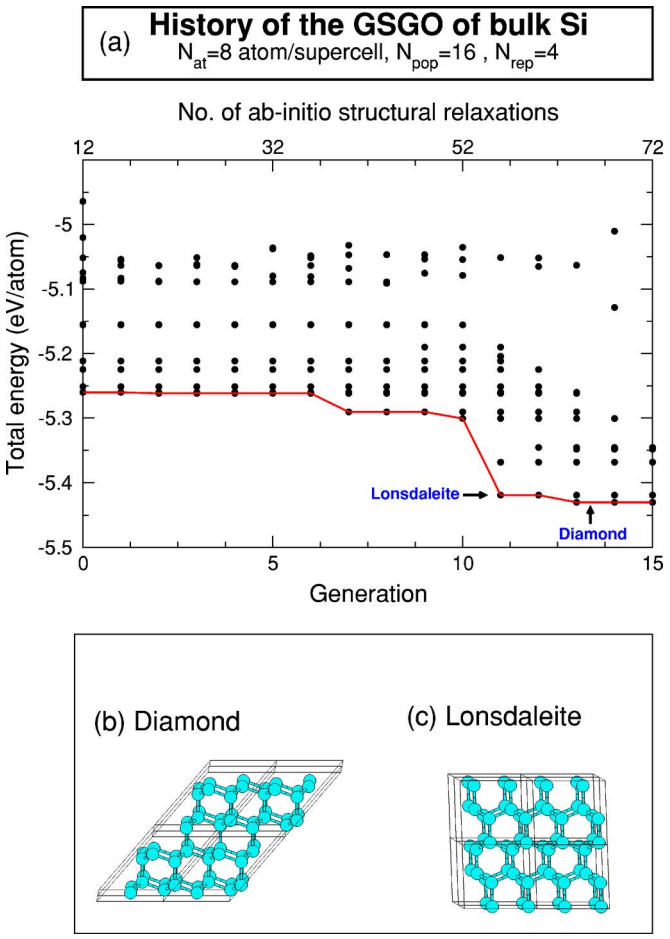

FIG. 2. (Color online) (a) History of the GSGO performed on bulk silicon, with eight atoms per supercell; (b) final structure of the GSGO search: diamond is obtained as the lowest-energy crystal structure; (c) lonsdaleite which represents the first "excited" structure found by GSGO.

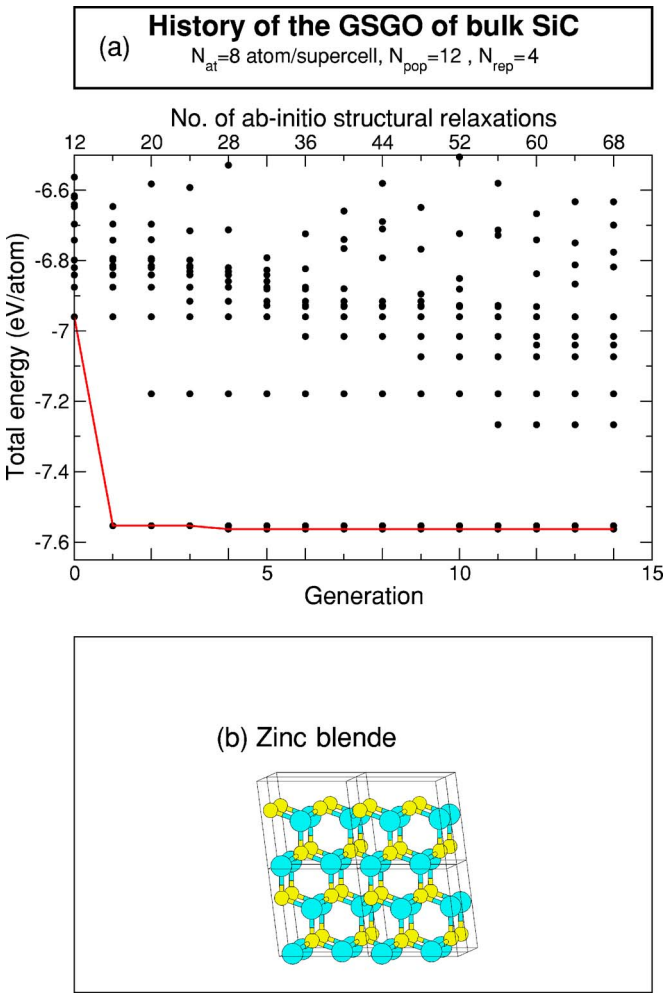

FIG. 3. (Color online) (a) History of the GSGO search performed on bulk SiC, with eight atoms per supercell; (b) final structure of the GSGO search: zinc blende is obtained as the lowest-energy crystal structure.

structure produced. Following Sec. IV, taking the volume $V=160$ Å^3 and grid resolution $\delta=2.0$ Å, we estimate that the number of possible configurations in the search space is $Q=1.3\times10^5$. In Fig. 2(a), we show the history plot of the evolutionary algorithm search. The diamond structure was correctly found as the lowest-energy structure in about 15 generations, i.e., by performing the *ab initio* structural relaxation of about 60 structures. Along with diamond [see Fig. 2(b)], the GSGO search found the lonsdaleite structure [see Fig. 2(c)] as a local minimum with an energy that is about 10 meV/at. higher than that of diamond.

B. SiC

SiC is a binary semiconductor system which is stable toward decomposition in its elemental constituents: this is mostly due to the larger stability of the Si–C bond with respect to the Si–Si and the C–C bonds.[44] The history plot of the GSGO run is shown in Fig. 3(a). In this case, we included eight atoms in the search and took $N_{pop}=12$ with $N_{rep}=4$ structures replaced at each generation. Concerning the size of the search space, taking the unit-cell volume $V=80$ Å^3 and grid resolution $\delta=1.5$ Å one obtains that the number of possible configuration in the search space is $Q=5.1\times10^7$. The GSGO search finds the observed ground-state structure, which is zinc blende [see Fig. 3(b)], performing about 20 *ab initio* structural relaxations.

C. GaAs

GaAs represents a system where the structural excitations of largest energy are the antisite defects, which are atoms that violate the octet rule binding atoms of the same valence. Therefore, as for SiC, a procedure for surveying the lattice decoration during the search is required to find defect-free structures. Figure 4(a) shows the result of the GSGO search. We considered eight atoms per supercell and fully optimized the atomic positions and the cell vectors of every structure generated during the search. To estimate the size of the search space, we take the unit-cell volume $V=180$ Å^3 and the grid resolution $\delta=2.0$ Å and find that the number of structures in the search space is $Q=3.4\times10^7$. A population of $N_{pop}=12$ individuals has been set, and the $N_{rep}=3$ highest-

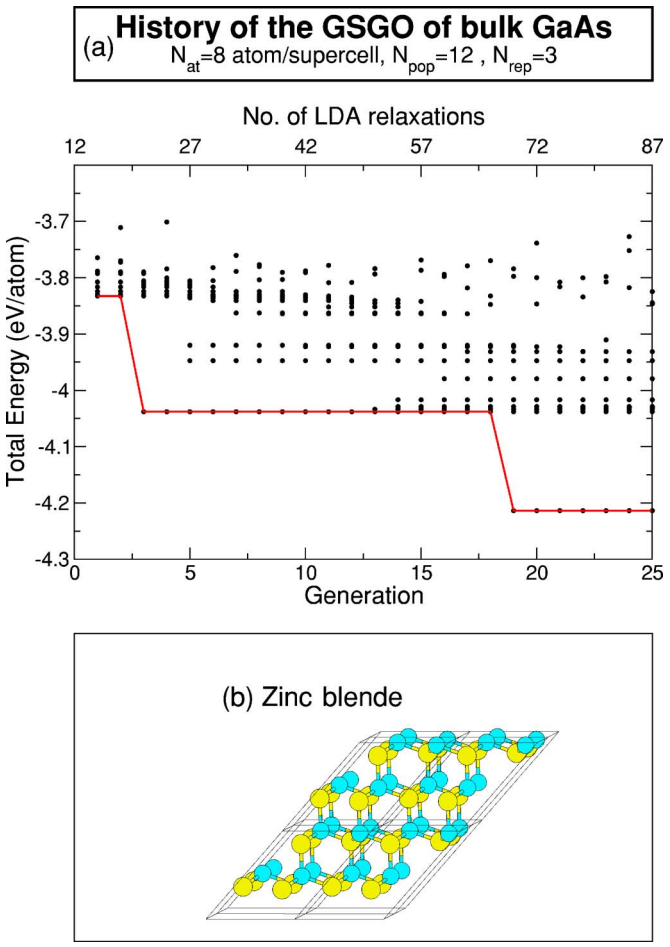

FIG. 4. (Color online) (a) History of the GSGO search performed on bulk GaAs, considering eight atoms per supercell; (b) zinc-blende structure as obtained by GSGO.

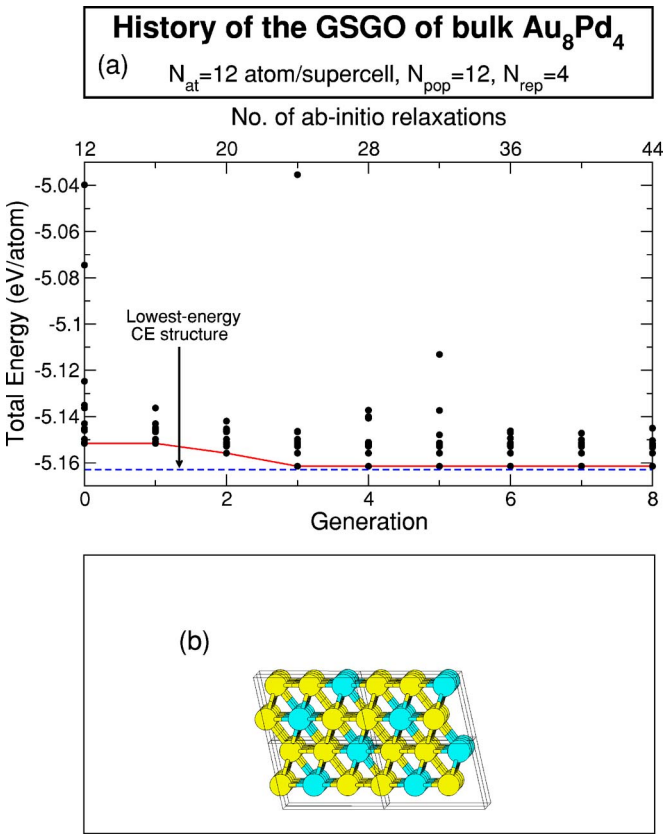

FIG. 5. (Color online) (a) History of the GSGO search performed on bulk Au_8Pd_4: in this plot the (blue) dashed line represents the energy of the ground-state structure for this composition which was found through cluster expansion. (b) Ground-state structure of Au_8Pd_4 found by GSGO.

energy structures have been replaced at each generation. The GSGO search correctly retrieved the zinc-blende structure as the lowest-energy one in less than 20 generations, i.e., about 70 *ab initio* structural relaxations.

D. AuPd

In metallic alloys, the difficult task is to identify the stable lattice decoration as there is usually a small energy difference between different atomic configurations.[22,23] For this reason, metal alloys were traditionally treated by cluster expansion methods[21–24] that, while restricting the consideration to one lattice type at the time (i.e., type-II problems), retain a high numerical accuracy ($\leqslant$5 meV/at.) over a large range of different configurations. As an example of application of the algorithm for the GSGO to this type of systems, we chose the AuPd alloy. This alloy has been recently studied by a cluster expansion (CE) approach,[45] which has led to the discovery of several unanticipated ground-state structures. In particular, this CE predicted fcc-based ground states at composition Au_8Pd_4, whereas a recent data-mining study[46] had predicted a non-fcc ground state (the C37 structure). Barabash *et al.*[45] studied the structures predicted by the two approaches and found (using similarly refined convergence criteria) that the one found by CE has lower energy than that suggested through data mining. To see whether the current generation of the algorithm for the GSGO can resolve such subtle differences, we performed a study of Au_8Pd_4. Figure 5(a) displays the history plot of the GSGO search, and Fig. 5(b) shows the lowest-energy structure that was identified after 8 GA steps: the fcc lattice was correctly identified starting from random lattice vectors. However, the structure found has an energy of 2 meV/at. higher than the energy of the structure predicted by CE with different atomic decorations. This reflects the difficulty in resolving the closely spaced energies of the $\binom{12}{4}=495$ configurations of Au_8Pd_4.

VI. CONCLUSIONS

We have described an approach to the GSGO problem of crystal structure prediction. The approach to the GSGO problem searches the space of the lattice geometry and atomic configuration of a solid looking for its ground-state structure. This search procedure is based on an evolutionary algorithm within which an "individual" represents a crystal structure. Unlike usual evolutionary algorithms, the present approach to the GSGO problem avoids a bit-string encoding of the crystal structure. Instead, the mating and mutation operations are performed handling directly the crystal structures in real

space. To define a mating operation between structures that have different cell shapes, all cell shapes are brought to a common basis using a similarity transformation. Once a structure is generated by the evolutionary operators, it is refined via a total-energy relaxation which brings it to the nearest total-energy local minimum. The approach to the GSGO problem described in this paper shares several steps with other optimization procedures presented in the recent literature, but it also shows different traits with respect to them. Indeed, in assigning the cell vectors to a child structure, we proceed by a linear combination of the cell vectors of the parents. This favors inheritance by the offspring of the information on the cell shapes of the parent structures. Also, particular attention has been paid to the screening of the offspring so as to avoid crystal structures with unphysical cell shapes and refinement of too close atom pairs. To treat multispecies systems, we scan the lattice-site occupations for surveying the set of possible decorations of the crystal, with the aim of finding the most stable one. We applied the algorithm for the GSGO to selected semiconductor systems and to the AuPd metallic alloy. In the case of all the semiconductor systems considered, the GSGO search retrieved the experimental crystal structure and atomic configuration. In the case of the AuPd alloy, we performed the global search for composition Au_8Pd_4 and obtained a structure that has an fcc lattice type, in agreement with earlier high-resolution cluster expansion study. The atomic configuration, however, differs from that predicted by cluster expansion. While we are encouraged that the current generation of approach to GSGO can resolve the correct lattice and comes close to resolving configurations, it is, at present, not able to mimic for metal alloys the resolution of cluster expansion.

ACKNOWLEDGMENTS

We thank Z. Liu, M. d'Avezac, A. Franceschetti, and P. Graf for useful discussions. This work was funded by the U.S. Department of Energy, Office of Science, Basic Energy Sciences, Materials Science and Engineering, under Contract No. DE-AC36-99GO10337 to NREL.

APPENDIX: PARAMETERS OF THE *AB INITIO* CALCULATIONS

We report in this appendix the parameters of the VASP calculations for all the systems considered in Sec. V. The electron-nucleus interaction was treated using ultrasoft pseudopotentials. For Si, GaAS, and SiC, we used the generalized gradient approximation[47] to the exchange and correlation functional, whereas for AuPd, we employed the local-density approximation.[48]

Si. The basis-set cutoff energy was set equal to 13.84 Ry for the plane-wave expansion of the wave functions. The (4,4,4) and (6,6,6) Monkhorst-Pack **k**-point meshes (excluding the Γ point) were used, respectively, for the structural relaxation and the calculation of the total energy of the relaxed structure. The structural relaxation was stopped when the residual forces on the atoms were lower than 0.025 eV/Å.

GaAs. The pseudopotential used for Ga placed the 3*d* states in the core. A basis-set cutoff energy of 13.26 Ry was used for the plane-wave expansion of the wave functions. A (3,3,3) Monkhorst-Pack **k**-point mesh (which included the Γ point) was used for the structural relaxation. The structural relaxation was stopped when the change in the total energy was less than 2 meV/cell. After the structural relaxation, the total energy of the equilibrium structure was calculated using a (4,4,4) Monkhorst-Pack **k**-point mesh (not Γ centered).

SiC. The basis-set cutoff energy was set equal to 26.34 Ry for the plane-wave expansion of the wave functions. The (4,4,4) and (6,6,6) Monkhorst-Pack **k**-point meshes (excluding the Γ point) were used, respectively, for the structural relaxation and the calculation of the total energy of the relaxed structures. The structural relaxation was stopped when the residual forces on the atoms were lower than 0.025 eV/Å.

AuPd. The plane-wave expansion of the wave functions was extended up to a kinetic energy of 18.28 Ry. We used (3,3,3) and (6,6,6) Monkhorst-Pack **k**-point meshes (not including the Γ point) to perform the structural relaxation and the calculation of the total energy of the relaxed structure, respectively.

[1] *Pauling file: Inorganic Materials Database and Design System—Binaries Edition*, edited by P. Villars (ASM International, Ohio, 2002).
[2] K. Rabe, MRS Bull. **18**, 2 (1993).
[3] *Structure and Bonding in Crystals*, edited by M. O'Keefe and A. Navrotsky (Academic, New York, 1981).
[4] J. C. Phillips, *Bonds and Bands in Semiconductors* (Academic, New York, 1973).
[5] A. Zunger, Phys. Rev. Lett. **44**, 582 (1980).
[6] A. Zunger, Phys. Rev. B **22**, 5839 (1980).
[7] W. Hume-Rothery and G. Rainor, *The Structure of Metals and Alloys* (Institute of Metals, London, 1954).
[8] M. T. Yin and M. L. Cohen, Phys. Rev. Lett. **50**, 2006 (1983).
[9] S. Froyen and M. L. Cohen, Phys. Rev. B **28**, 3258 (1983).
[10] A. Mujica, A. Rubio, A. Munoz, and R. J. Needs, Rev. Mod. Phys. **75**, 863 (2003).
[11] A. Munoz and K. Kunc, Phys. Rev. B **44**, 10372 (1991).
[12] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes in FORTRAN 77: The Art of Scientific Computing* (Cambridge University Press, Cambridge, 1992).
[13] D. Frenkel and B. Smit, *Understanding Molecular Simulation*, Computational Science Series Vol. 1 (Academic, New York, 2001).
[14] R. Car and M. Parrinello, Phys. Rev. Lett. **55**, 2471 (1985).
[15] M. C. Desjonquere and D. Spanjaard, *Concepts in Surface Physics* (Springer, Berlin, 2002).
[16] D. J. Chadi and K. J. Chang, Phys. Rev. Lett. **61**, 873 (1988).
[17] F. Ducastelle, *Order and Phase Stability in Alloys*, Cohesion and

Structure Vol. 3 (North-Holland, Amsterdam, 1991).
[18] L. G. Ferreira, S.-H. Wei, and A. Zunger, Int. J. Supercomput. Appl. **5**, 34 (1991).
[19] J. Kanamori and Y. Kakehashi, J. Phys. (Paris), Colloq. **38**, C7-274 (1977).
[20] G. Trimarchi, P. Graf, and A. Zunger, Phys. Rev. B **74**, 014204 (2006).
[21] C. Wolverton, V. Ozoliņš, and A. Zunger, Phys. Rev. B **57**, 4332 (1998).
[22] V. Blum and A. Zunger, Phys. Rev. B **69**, 020103(R) (2004).
[23] V. Blum and A. Zunger, Phys. Rev. B **70**, 155108 (2004).
[24] A. Zunger, in *Statistics and Dynamics of Alloys Phase Transformations*, NATO ASI, edited by P. E. A. Turchi and A. Gonis (Plenum, New York, 1994), p. 361.
[25] V. Ozolins and A. Zunger, Phys. Rev. Lett. **82**, 767 (1999).
[26] K. Kim, V. Ozoliņš, and A. Zunger, Phys. Rev. B **60**, R8449 (1999).
[27] M. Parrinello and A. Rahman, Phys. Rev. Lett. **45**, 1196 (1980).
[28] R. M. Wentzcovitch, Phys. Rev. B **44**, 2358 (1991).
[29] A. Laio and M. Parrinello, Proc. Natl. Acad. Sci. U.S.A. **99**, 12562 (2002).
[30] Z. W. Lu, D. B. Laks, S.-H. Wei, and A. Zunger, Phys. Rev. B **50**, 6642 (1994).
[31] A. E. Eiben and J. E. Smith, *Introduction to Evolutionary Computing* (Springer, Berlin, 2003).
[32] D. M. Deaven and K. M. Ho, Phys. Rev. Lett. **75**, 288 (1995).
[33] N. L. Abraham and M. I. J. Probert, Phys. Rev. B **73**, 224104 (2006).
[34] A. R. Oganov, C. W. Glass, and S. Ono, Earth Planet. Sci. Lett. **241**, 95 (2006).
[35] A. R. Oganov and C. W. Glass, J. Chem. Phys. **124**, 244704 (2006).
[36] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, Science **220**, 671 (1983).
[37] D. Wales and J. Doye, J. Phys. Chem. A **101**, 5111 (1997).
[38] J. P. K. Doye and D. J. Wales, Phys. Rev. Lett. **80**, 1357 (1998).
[39] S. Goedecker, J. Chem. Phys. **120**, 9911 (2004).
[40] J. Ihm, A. Zunger, and M. L. Cohen, J. Phys. C **12**, 4409 (1979).
[41] O. H. Nielsen and R. M. Martin, Phys. Rev. B **32**, 3780 (1985).
[42] G. Kresse and J. Hafner, Phys. Rev. B **47**, 558 (1993).
[43] G. Kresse and J. Furthmuller, Phys. Rev. B **54**, 11169 (1996).
[44] J. L. Martins and A. Zunger, Phys. Rev. Lett. **56**, 1400 (1986).
[45] S. V. Barabash, V. Blum, S. Muller, and A. Zunger, Phys. Rev. B **74**, 035108 (2006).
[46] S. Curtarolo, D. Morgan, and G. Ceder, CALPHAD: Comput. Coupling Phase Diagrams Thermochem. **29**, 163 (2005).
[47] J. P. Perdew and Y. Wang, Phys. Rev. B **45**, 13244 (1992).
[48] J. P. Perdew and A. Zunger, Phys. Rev. B **23**, 5048 (1981).