

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  $\text{Au}_8\text{Pd}_4$ . *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  $\text{Au}_8\text{Pd}_4$ , 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](https://doi.org/10.1103/PhysRevB.75.104113)

PACS number(s): 61.50.Ah, 61.66.-f



A central feature of the solid-state physics of crystalline solids is the existence of a very rich diversity of stable crystal forms,<sup>1,2</sup> 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.<sup>3-7</sup> While early attempts were based on correlating observed structures with elemental scales such as electronegativity,<sup>4</sup> orbital radii,<sup>5,6</sup> or electron-atom ratio,<sup>7</sup> 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,<sup>8-10</sup> 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.<sup>10,11</sup> The optimization methods used for such type-I problems are continuous position-space optimization such as gradient-

guided searches<sup>12</sup> and conventional<sup>13</sup> and Car-Parrinello<sup>14</sup> 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<sup>15</sup> and local relaxation around a point defect in semiconductors.<sup>16</sup> A solution to type-I optimization problems has led to the establishment of the equation of state of numerous materials in known structures<sup>8,10,11</sup> as well as to the discovery of surface reconstruction patterns of given semiconductor surface topologies<sup>15</sup> or to the local geometry of the leading defect structures in semiconductors.<sup>16</sup>

*Type-II* optimization involves cases where the lattice type

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<sup>21</sup> and bcc-based<sup>22,23</sup> alloys via search of lowest-energy cluster expansion (Ising type) Hamiltonians.<sup>24</sup>

*Type-III* optimization problems involve cases where nei-

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).

child structure  $\mathcal{S}^{(c)}$  is created using a procedure based on the

### III. GENETIC ALGORITHM

A population of  $N_{pop}$  crystal structures is defined, where each structure  $\mathcal{S} = \{ \mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N \}$  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  $i=1, \dots, N$ . 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. Crossover

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

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

then the relation

$$\mathbf{r}_i = B \mathbf{s}_i \quad (2)$$

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. Mating

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

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. $t$

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  $45^\circ$  and  $135^\circ$ , 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. $t$

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<sup>40</sup> 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<sup>40</sup>) and the gradients with respect to the cell vectors (i.e., components of the stress tensor<sup>41</sup>). This relaxation<sup>32</sup> performed on each individual in the population makes the present optimization procedure more efficient in surveying the searchspace.

structure produced. Following Sec. IV, taking the volume  $V = 160 \text{ \AA}^3$  and grid resolution  $\delta = 2.0 \text{ \AA}$ , we estimate that the number of possible configurations in the search space is  $Q = 1.3 \times 10^5$ . 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. C

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.<sup>44</sup> 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  $V$

$= 80 \text{ \AA}^3$  and grid resolution  $\delta = 1.5 \text{ \AA}$  one obtains that the number of possible configuration in the search space is  $Q = 5.1 \times 10^7$ . 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. G A

GaAs represents a system GaAs

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. A**

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.<sup>22,23</sup> For this reason, metal alloys were traditionally treated by cluster expansion methods<sup>21-24</sup> that, while restricting the consideration to one lattice type at the time (i.e., type-II problems), retain a high numerical accuracy ( $\leq 5$  meV/at.) over a large range of

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, par-

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