

# Crystal structure prediction via particle-swarm optimization

Yanchao Wang, Jian Lv, Li Zhu, and Yanming Ma\*

*State Key Laboratory of Superhard Materials, Jilin University, Changchun 130012, China*

(Received 8 May 2010; published 28 September 2010)

We have developed a method for crystal structure prediction from “scratch” through particle-swarm optimization (PSO) algorithm within the evolutionary scheme. PSO technique is different with the genetic algorithm and has apparently avoided the use of evolution operators (e.g., crossover and mutation). The approach is based on an efficient global minimization of free-energy surfaces merging total-energy calculations via PSO technique and requires only chemical compositions for a given compound to predict stable or metastable structures at given external conditions (e.g., pressure). A particularly devised geometrical structure parameter which allows the elimination of similar structures during structure evolution was implemented to enhance the structure search efficiency. The application of designed variable unit-cell size technique has greatly reduced the computational cost. Moreover, the symmetry constraint imposed in the structure generation enables the realization of diverse structures, leads to significantly reduced search space and optimization variables, and thus fastens the global structure convergence. The PSO algorithm has been successfully applied to the prediction of many known systems (e.g., elemental, binary, and ternary compounds) with various chemical-bonding environments (e.g., metallic, ionic, and covalent bonding). The high success rate demonstrates the reliability of this methodology and illustrates the promise of PSO as a major technique on crystal structure determination.

DOI: [10.1103/PhysRevB.82.094116](https://doi.org/10.1103/PhysRevB.82.094116)

PACS number(s): 61.50.Ah

## I. INTRODUCTION

Crystal structure occupies a central and often critical role in materials science, particularly when establishing a correspondence between material performance and its basic composition since properties of a solid are intimately tied to its crystal structure. Experimentally, structural determination through x-ray diffraction technique has been developed extremely well, leading to numerous crystal structures solved. However, it happens frequently that experiments fail to determine structures due to the obtained low-quality x-ray diffraction data, particularly at extreme conditions (e.g., high pressure). Here, the theoretical prediction of crystal structures with the only known information of chemical composition independent of previous experimental knowledge is greatly necessary. However, this is extremely difficult as it basically involves in classifying a huge number of energy minima on the lattice energy surface. 20 years ago John Maddox even published an article in Nature to question the predictive power provided with only the knowledge of chemical composition.<sup>1</sup> His words still remain largely true, as evidenced by poor results of the latest blind test for crystal structure prediction.<sup>2</sup>

Owing to significant progress in both computational power and basic materials theory, it is now possible to predict the crystal structure at 0 K using the quantum-mechanical methods, some of which are simulated annealing,<sup>3,4</sup> genetic algorithm,<sup>5–9</sup> basin hopping,<sup>10,11</sup> metadynamics,<sup>12,13</sup> random sampling,<sup>14–16</sup> and data mining methods.<sup>17</sup> Simulated annealing, basin hopping, and metadynamics have focused on overcoming the energy barriers and are successful in many researches,<sup>4,18,19</sup> particularly, when the starting structure is close to the global minimum. The data mining method relies heavily on the existence of an extensive database of good trial structures and is incapable of generating new crystal structure types in the absence of in-

formation on similar compounds. Random sampling method seems “simple” in principle but nontrivial in practice and works well in many applications.<sup>14–16,20</sup> The genetic algorithm (GA) starts to use a self-improving method and thus is successful in predicting many high-pressure structures.<sup>21–28</sup>

We here have proposed a methodology for crystal structure prediction based on the particle-swarm optimization (PSO) technique within the evolutionary scheme. With the PSO implementation, the computational expense of first-principles density-functional calculations has been significantly reduced. This is stemmed from that PSO is a highly efficient global optimization method. We have successfully applied this method to the prediction of various known systems, such as elemental, binary, and ternary compounds. The high success rate demonstrates the reliability of this methodology and illustrates the promise of PSO as a major tool on crystal structure determination.

This paper is arranged as follows. In Sec. II, the method and implementation of PSO algorithm will be discussed in details. A short overview of results obtained from our method is presented in Sec. III followed by the summary in Sec. IV.

## II. METHOD AND IMPLEMENTATION

PSO is a branch of evolutionary methodology but quite different with GA. In particular, the major evolution operations of “crossover” and “mutation” in GA have been avoided. PSO was first proposed by Kennedy and Eberhart in the mid 1990s.<sup>29,30</sup> As a stochastic global optimization method, PSO is inspired by the choreography of a bird flock and can be seen as a distributed behavior algorithm that performs multidimensional search. According to PSO, the behavior of each individual is affected by either the best local or the best global individual to help it fly through a hyperspace. Moreover, an individual can learn from its part expe-

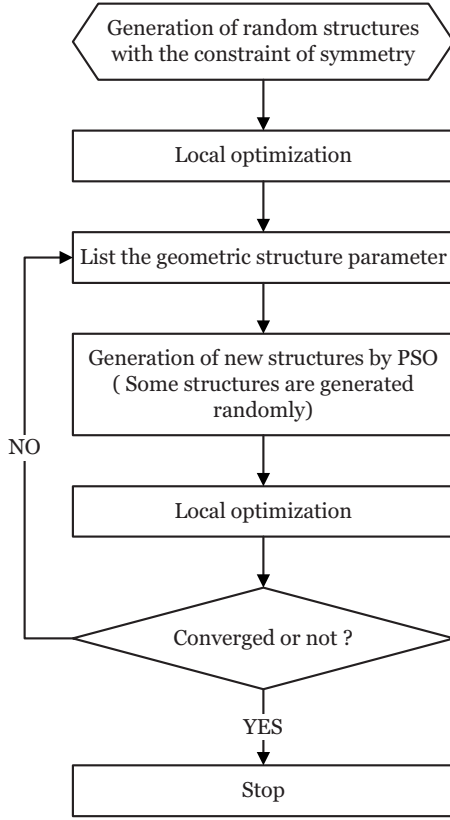


FIG. 1. The flow chart of CALYPSO.

riences to adjust its flying speed and direction. Therefore, all the individuals in the swarm can quickly converge to the global position and near-optimal geographical position by the behavior of the flock and their flying histories. PSO has been verified to perform well on many optimization problems.<sup>31–35</sup> We have implemented PSO algorithm on crystal structure prediction in crystal structure analysis by particle swarm optimization (CALYPSO) code.<sup>36</sup>

Our global minimization method through CALYPSO code for predicting crystal structures comprises mainly four steps as depicted in the flow chart of Fig. 1: (1) generation of random structures with the constraint of symmetry; (2) locally structural optimization; (3) postprocessing for the identification of unique local minima by geometrical structure parameter; (4) generation of new structures by PSO for iteration.

#### A. Step 1: Generation of random structures with the constraint of symmetry

Two types of variables are necessary to define a crystal structure: lattice parameters and atomic coordinates. There are six lattice parameters: three angles and three lattice vectors. Each atom has three coordinates coded as a fraction of the corresponding lattice vectors. The first step of our approach is to generate random structures symmetrically constrained within 230 space groups. Once a particular space group is selected, the lattice parameters are then confined within the chosen symmetry. The corresponding atomic coordinates are generated by the crystallographic symmetry op-

erations through matrix-column pairs  $(W, w)$ ,<sup>37</sup> where the point operation  $W$  is a  $3 \times 3$  matrix and the translation operation  $w$  is one column. Using the matrix-column pairs, one obtains new coordinates by matrix multiplication,

$$\tilde{x}_1 = W_{11}x_1 + W_{12}x_2 + W_{13}x_3 + w_1,$$

$$\tilde{x}_2 = W_{21}x_1 + W_{22}x_2 + W_{23}x_3 + w_2,$$

$$\tilde{x}_3 = W_{31}x_1 + W_{32}x_2 + W_{33}x_3 + w_3.$$

This can be written in an abbreviated form:  $\tilde{x} = Wx + w$ . Within this matrix-column pairs operation, one random atomic coordinate can be used to generate other symmetrically related coordinates. For different symmetries, the operation rules within the matrix-column pairs for generating symmetrically related atomic positions are different, ensuring unduplicated generation of different structures.

The generation of random structures ensures unbiased sampling of the energy landscape. The explicit application of symmetric constraints leads to significantly reduced search space and optimization variables, and thus fastens global structural convergence. For example, in the case of monoclinic crystals, the symmetric constraint limits the range of fractional atomic coordinates within 0–0.5, namely, only half of the search space, at the same time, the optimization variables are reduced to five thanks to a fixed lattice angle ( $90^\circ$ ). Moreover, we have applied a symmetry checking technique, in which the appearance of identical symmetric structures is strictly forbidden. This allows the generation of diverse structures, which are crucial for the efficiency of global minimization.

Since crystal structure prediction is performed on a blind base, the choice of the simulation cell sizes (not prior known) is critical to target the global minimum structure. In practice, one can choose all possible cell sizes to perform the separate simulations and then compare all the resulting structures to derive the global stable structure. However, this procedure is extremely computational costly and for some particular cases, is not affordable. Here, the fast learning ability of PSO technique has allowed us to implement a variable cell-size technique enabling the intelligent selection of the correct cell sizes during the structural evolution, and thus the computational cost has been significantly reduced.

#### B. Step 2: Local optimization

The potential-energy surface can be regarded as a multi-dimensional system of hills and valleys with saddle points connecting them. The valleys are the local basin of attractions on the potential-energy surface. The local optimization (such as line minimization, steepest descents, conjugate gradient algorithm, or Broyden-Fletcher-Goldfarb-Shanno algorithm) can drive the structural energy to the local minimum, which may or may not be the global minimum. The approach of locally optimizing every candidate has been used with great success. Local optimization increases the cost of each individual, but effectively reduces the noise of the landscape, enhances comparability between different structures, and provides locally optimal structures for further use. We use

free energy (here, at  $T=0$  K, free energy reduces to enthalpy) of the locally optimized structure as fitness function throughout the simulation. Both the atomic coordinates and lattice parameters are locally optimized. Among the locally optimized structures, a certain number of worst ones are rejected, and the remaining structures participate in creating new structures through PSO for the next generation.

### C. Step 3: Postprocessing for the identification of unique local minima

Our method which solves the packing problem contains a critical step where a large number of preliminary trial structures are generated and then structurally optimized. At this step, many newly generated structures are very similar or even identical. The direct use of these similar structures to generate next generation will significantly slow down the convergence to the global minimization solution.<sup>38</sup> It is thus highly beneficial to remove these duplicates to accelerate the search process. We have designed a method to identifying structural similarity named as geometrical structure parameter on the basis of interatomic distances, which are calculated according to the so called “bond types.” For example, if there are two types (A and B) of atoms in the simulation cell, three bond types, i.e., A-A, A-B, and B-B, will be evaluated. In order to avoid the problems raised by the periodic boundary conditions,  $4 \times 4 \times 4$  supercells are constructed to calculate the interatomic distances. This procedure ensures a safe calculation of first and second nearest neighbors. Then, one can determine the distances and number of the first and second nearest neighbors for different bond types, which are then listed in the matrices as backup for future comparison. Once a new structure is generated, the geometrical structure parameter is applied to check the similarity of this structure with those in the saved matrices within specified tolerances. Specifically, if this structure shares the same number of bonds with one structure in the matrices, the deviation of bond length is then calculated according to the equation  $\Delta d = \sqrt{\sum_{i,j} (L_i - L'_j)^2 \delta_{i,j}}$ , where  $L_i$  and  $L'_j$  are the bond lengths in the two structures, respectively, and  $\delta_{i,j}$  is the delta function. If the deviation ( $\Delta d$ ) is less than the preset threshold, the two structures are considered to be equivalent. Thus, the newly generated structure will be discarded. Otherwise, it is kept and documented in the matrices. The matrices containing all the structure information are updated after local optimization at every generation and used in next generation.

### D. Step 4: Generation of new structures by PSO

In the next generation, a certain number of new structures (the best 60% of the population size) are generated by PSO. Within the PSO scheme, a structure (an individual) in the searching phase space is regarded as a particle. A set of individual particles is called a population or a generation. During the evolution Eq. (1) is used to update the positions of particles,

$$x_{i,j}^{t+1} = x_{i,j}^t + v_{i,j}^{t+1}. \quad (1)$$

It is necessary to note that the velocity plays an important role on determination of the speed and direction of particle

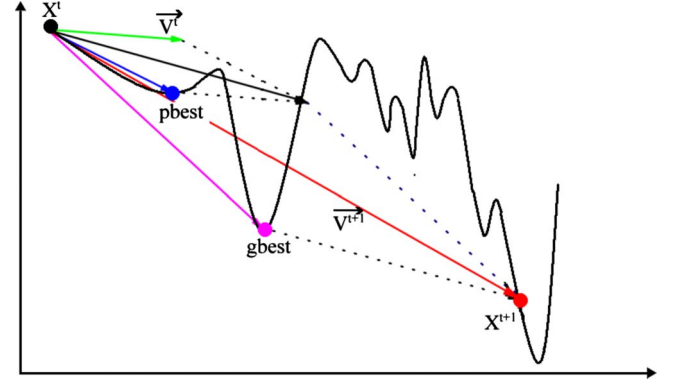


FIG. 2. (Color online) Depiction of the velocity and position updates in PSO. The black solid line approximates a typical energy landscape. Arrows represent either the velocity or the relative positions of a particle (here, a structure). An initial velocity  $V^t$  develops to  $V^{t+1}$  after one PSO step through Eq. (2). A high-energy structure ( $X^t$ ) evolves to a much lower-energy structure ( $X^{t+1}$ ) after one PSO operation through Eq. (1). This illustrates the efficiency of PSO in exploration of low-energy structures.

movement. The new velocity of each individual  $i$  at the  $j$ th dimension is calculated based on its previous location ( $x_{i,j}^t$ ) before optimization, previous velocity ( $v_{i,j}^t$ ), current location ( $pbest_{i,j}^t$ ) with an achieved best fitness, i.e., lowest enthalpy, of this individual, and the population global location ( $gbest^t$ ) with the best fitness value for the entire population according to Eq. (2). The initial  $v_{ij}$  was generated randomly. The position of each particle is updated using its velocity vector as depicted in Fig. 2,

$$v_{i,j}^{t+1} = \omega v_{i,j}^t + c_1 r_1 (pbest_{i,j}^t - x_{i,j}^t) + c_2 r_2 (gbest^t - x_{i,j}^t), \quad (2)$$

where  $j \in \{1, 2, 3\}$ ,  $\omega$  (in the range of 0.9–0.4)<sup>30</sup> denotes the inertia weight and controls the momentum of the particle. High settings of  $\omega$  as 0.9 facilitate global search, and lower settings facilitate rapid local search. In our methodology,  $\omega$  is dynamically varied and decreases linearly from 0.9 to 0.4 during the iteration according to Eq. (3),

$$\omega = \omega_{\max} - \frac{\omega_{\max} - \omega_{\min}}{iter_{\max}} \times iter, \quad (3)$$

where  $\omega_{\max}$  and  $\omega_{\min}$  equals to 0.9 and 0.4, respectively.  $c_1$  is a self-confidence factor and expresses how much the particle trusts its own past experience while  $c_2$  is a swarm confidence factor and expresses how much it trusts the swarm. Earlier study<sup>39</sup> has demonstrated that  $c_1 = c_2 = 2$  gave the best overall performance. Accordingly, in our implementation, we keep  $c_1$  and  $c_2$  as constant 2.  $r_1$  and  $r_2$  are two separately generated random numbers and uniformly distributed in the range  $[0, 1]$ . The velocity update formula includes random parameters ( $r_1$  and  $r_2$ ) to ensure good coverage of the searching space and avoid entrapment in local optima. As shown in Eq. (2), it is quite obvious that the movement of particles in the search space is dynamically influenced by their individual past experience ( $pbest_{i,j}^t, v_{i,j}^t$ ) and successful experiences attained by the whole swarm ( $gbest^t$ ). Thus the velocity makes

TABLE I. Some of standard input parameters of CALYPSO code. The minimal interatomic distances must be sufficient to ensure that there is no pathological overlap of pseudopotential core regions. The best 60% of the population size are generated by PSO.  $N_{pop}$  is short for the population size and should be big for large systems.

Parameters	Value
Minimal interatomic distances	0.8 (Å)
Proportion of the structures generated by PSO	0.6
$N_{pop}$	20–50
Self-confidence factor ( $c_1$ )	2.0
Swarm confidence factor ( $c_2$ )	2.0

the particles to move toward to global minimum and accelerates the convergence speed. Moreover, to overcome explosion and divergence, the magnitudes of the velocities are necessary to be confined within the range of  $[-0.1, 0.1]$ .

In order to improve the efficiency of the procedure, only the low-energy structures (the best 60% of the previous generation) which are on the most promising area of the configuration space, are selected to produce the next generation by PSO. In order to keep the population diversity, a certain number of structures (40% of the population size) whose symmetries must be distinguished from any of previously generated ones, are generated randomly. More specifically, we first set up a list for 230 space groups functioned as a comparison basis. Second, if the generated structures have already gone beyond all 230 space groups, another list including the crystallographic atomic sites is built and used to compare with the newly generated structures. The identical structure is discarded.

### III. APPLICATION AND RESULTS

Here we benchmark our methodology on systems with known structures and explore unknown high pressure structures for Li. All the calculations were performed in the framework of density-functional theory within the all-electron projector-augmented wave (PAW) method<sup>40</sup> as implemented in the VASP code.<sup>41</sup> *Ab initio* total-energy calculations are performed at 0 K and thus the free energy is reduced to enthalpy. Some basic parameters used in CALYPSO code can be found in Table I. An overview of the benchmark systems including elements, binary compounds, and ternary compounds with known structures can be found in Tables II and III.

#### A. Elements

Lithium is a simple metal at ambient pressure but exhibits complex phase transitions under compression. Experimentally, it has been demonstrated that lithium takes the phase transition sequence of  $bcc \rightarrow fcc \rightarrow hR1 \rightarrow cI16$ ,<sup>42–45</sup> above which new phases are observed but remain unsolved. In theory, the complex post-*cI16* structures above 70 GPa, such as *Cmca*-24, *C2*, *Aba2*, and *P4<sub>2</sub>/mbc*, are proposed by using molecular dynamics, genetic algorithm and random sampling method, respectively.<sup>46–49</sup> Here, we used the PSO

TABLE II. Systems on elements with known structures on which calculations were performed by CALYPSO. All the experimental structures are reproduced within the given generations and population sizes. Note that our choice of population sizes is based on experience and it is highly possible that the use of smaller sizes could also result in the correct structures, with lower computational cost and should be bigger for larger systems.

Systems	Pressure (GPa)	Structures	Generation	$N_{pop}$
Li	0	bcc <sup>a</sup>	1	30
	0	9R <sup>b</sup>	3	30
	10	fcc <sup>c</sup>	1	20
	40	<i>hR1</i> <sup>d</sup>	4	30
	70	<i>cI16</i> <sup>d</sup>	7	30
C	0	Graphite <sup>e</sup>	30	30
	0	Diamond <sup>f</sup>	6	30
Si	2	<i>bc8</i> <sup>g</sup>	6	30
	10	cd <sup>h</sup>	1	20
	10	sh <sup>i</sup>	2	20
	10	$\beta$ -Sn <sup>j</sup>	3	20
	10	<i>Imma</i> <sup>k</sup>	4	20
	40	<i>Cmca</i> <sup>l</sup>	2	20
	40	hcp <sup>l</sup>	4	20
	80	fcc <sup>m</sup>	1	20
Mg	0	hcp <sup>n</sup>	6	30
	100	bcc <sup>o</sup>	4	30

<sup>a</sup>Reference 42.

<sup>b</sup>Reference 43.

<sup>c</sup>Reference 44.

<sup>d</sup>Reference 45.

<sup>e</sup>Reference 51.

<sup>f</sup>Reference 52.

<sup>g</sup>Reference 53.

<sup>h</sup>Reference 54.

<sup>i</sup>Reference 55.

<sup>j</sup>Reference 56.

<sup>k</sup>Reference 57.

<sup>l</sup>Reference 58.

<sup>m</sup>Reference 59.

<sup>n</sup>Reference 60.

<sup>o</sup>Reference 61.

method through CALYPSO code to predict the stable structures at 0, 10, 40, 70, 80, 100, 120, 200, and 300 GPa, and found all the experimental and theoretical structures mentioned above at certain pressure ranges. It is remarkable that for all the simulation of these complex structures, only less than 300 generated structures are needed to derive the correct results. For example, the *cI16* structure is successfully identified at the seventh generation with a population size ( $N_{pop}$ ) of 30, i.e., only 210 structures are generated and locally optimized. Here, we present a newly discovered orthorhombic *Cmca*-56 structure at 200 GPa, which is most stable among all earlier theoretical structures. Our *ab initio* structure relaxations were performed using density-functional theory within the local-density approximation. The PAW potential with 1.2 a.u. core radius was used, and the 1s and 2s electrons were treated as valence. A basis-set cutoff energy of 1000 eV was used for the plane-wave expansion of the wave functions. A Monkhorst-Pack<sup>50</sup> grid of  $12 \times 12 \times 12$  was used to ensure that the enthalpy calculations are well converged to better than 1 meV/atom. Here, the complex *Cmca*-56 structure contains 56 atoms per unit cell, which can be viewed as alterna-



TABLE III. Systems on binary and ternary compounds with known experimental structures on which calculations were performed by CALYPSO. All the experimental structures are reproduced within the given generations and population sizes. Note that our choice of population sizes is based on experience and it is highly possible that the use of smaller sizes could also result in the correct structures, with lower computational cost and less generated structures.

System	Pressure (GPa)	Structures	Generation	$N_{pop}$
SiO <sub>2</sub>	0	$\alpha$ -quartz <sup>a</sup>	5	20
	20	Stishovite <sup>b</sup>	1	20
	70	CaCl <sub>2</sub> -type <sup>c</sup>	5	30
	100	$\alpha$ -PbO <sub>2</sub> -type <sup>d</sup>	4	20
	500	Pyrite-type <sup>e</sup>	15	20
SiC	0	Zinc blende <sup>f</sup>	2	12
	0	Moissanite <sup>f</sup>	3	30
	150	Rocksalt <sup>g</sup>	2	30
ZnO	12	Rocksalt <sup>h</sup>	2	30
TiH <sub>2</sub>	0	$I4/mmm$ <sup>i</sup>	2	20
	0	$Fm\bar{3}m$ <sup>j</sup>	3	20
MoB <sub>2</sub>	0	$R\bar{3}m$ <sup>k</sup>	1	30
TiB <sub>2</sub>	0	AlB <sub>2</sub> -type <sup>l</sup>	1	30
MgSiO <sub>3</sub>	120	$Cmcm$ <sup>m</sup>	5	20
CaCO <sub>3</sub>	0	Calcite <sup>n</sup>	13	30

<sup>a</sup>Reference 63.

<sup>b</sup>Reference 64.

<sup>c</sup>Reference 65.

<sup>d</sup>Reference 66.

<sup>e</sup>Reference 67.

<sup>f</sup>Reference 68.

<sup>g</sup>Reference 69.

<sup>h</sup>Reference 70.

<sup>i</sup>Reference 71.

<sup>j</sup>Reference 72.

<sup>k</sup>Reference 73.

<sup>l</sup>Reference 74.

<sup>m</sup>Reference 75.

<sup>n</sup>Reference 76.

tively stacking of *Cmca*-24 (Ref. 46) and  $P4_2/mbc$  (Ref. 37) structures. The history of the CALYPSO search performed on Li is shown in Fig. 3. The complex *Cmca*-56 structure was identified at fourth generation.

Other elements, such as carbon,<sup>51,52</sup> silicon,<sup>53–59</sup> and magnesium<sup>60,61</sup> (Table II), were also tested and the simulations quickly reproduced all the experimental structures. Particularly, several metastable structures (Fig. 4) proposed earlier by other theoretical methods<sup>7,62</sup> of carbon were predicted at 0 GPa and the *bc8* structure (metastable phase) of silicon is also predicted at 2 GPa. This indicates that our method can be used to predict metastable structures.

### B. Binary compounds

Silica is a binary semiconductor which exhibits many novel polymorphs at elevated pressures. We have successfully reproduced the experimental  $\alpha$ -quartz, stishovite, CaCl<sub>2</sub>-type,  $\alpha$ -PbO<sub>2</sub>-type, and pyrite-type structures<sup>63–67</sup> at the certain pressure regimes by the structural prediction through CALYPSO code. Again, all the structures rapidly converge to the global minimum with less than 300 local opti-

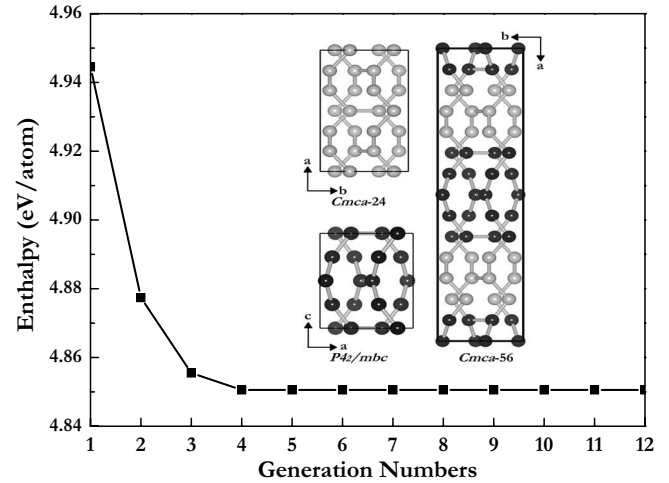


FIG. 3. Main figures: the history of CALYPSO structure search on Li at 200 GPa. Insets: the predicted *Cmca*-56 structure at the fourth generation. For the *Cmca*-56 structure at 220 GPa, the lattice parameters are  $a=15.687$  Å,  $b=3.855$  Å, and  $c=3.847$  Å with atomic positions at 16g (0.14302, 0.83356, 0.43863), (0.21482, 0.10157, 0.15158), (0.92849, 0.86370, 0.87551), and 8f (0.00000, 0.44292, 0.83303), respectively.

mizations. Specifically, the structural search easily found the  $\alpha$ -quartz structure at the fifth generation with only 120 structures at 0 GPa. The history plot of the simulations by CALYPSO code for silica at 70 GPa is shown in Fig. 5(a). The experimental CaCl<sub>2</sub>-type structure was found at the fifth generation.

Other binary compounds<sup>36,68–74</sup> are also benchmarked as listed in Table III. All the simulations show fast convergence to the experimental structures.

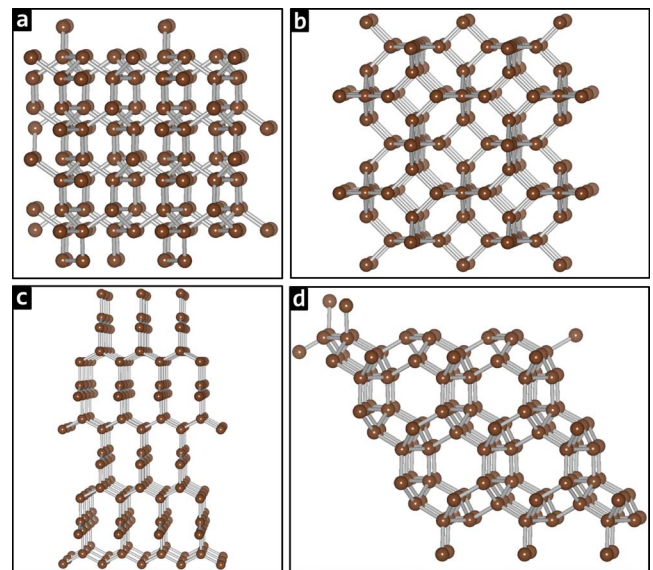


FIG. 4. (Color online) The metastable structures of carbon predicted by CALYPSO. (a) *bc8* structure predicted at 2000 GPa. (b)  $C_6$   $Im\bar{3}m$  structure predicted at 0 GPa. (c)  $\beta$ -Sn structure predicted at 0 GPa. (d) Chiral framework  $P6_122$  or  $P6_522$  structure predicted at 0 GPa.

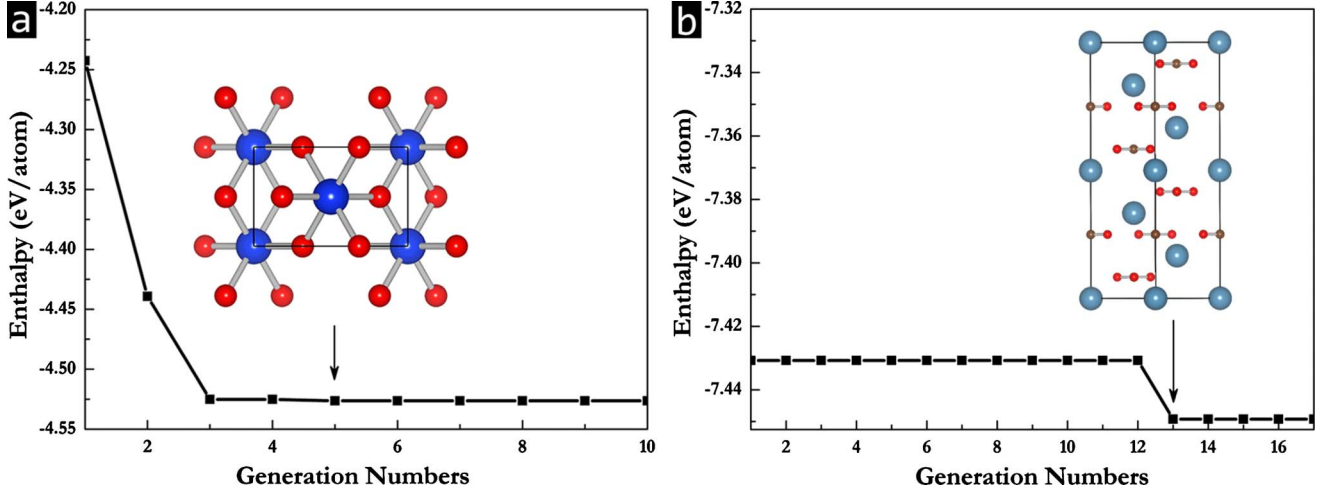


FIG. 5. (Color online) Main figures: enthalpy history of CALYPSO structure search on  $\text{SiO}_2$  (a) at 70 GPa and  $\text{CaCO}_3$  (b) at 120 GPa. Insets: the predicted  $\text{CaCl}_2$  ( $Pnnm$ ) structure of  $\text{SiO}_2$  (a) at the fifth generation and perovskite structure of  $\text{CaCO}_3$  (30 atoms/cell) (b) at the 13th generation.

### C. Ternary compounds

We have successfully identified the most stable structures of  $\text{MgSiO}_3$  and  $\text{CaCO}_3$  under pressure. The postperovskite  $Cmcm$  structure of  $\text{MgSiO}_3$  (Re. 75) was quickly found in the fifth generation at 120 GPa with only 100 local optimizations. In addition, the metastable perovskite phase of  $\text{MgSiO}_3$  was identified in the sixth generation with less than 120 structures. This simulation further illustrates that our method is able to find both stable and metastable structures. Moreover, the experimental calcite phase of  $\text{CaCO}_3$  (Ref. 76) has been reproduced in the 13th generation at 0 GPa. In Fig. 5(b), we show the history plot of the structural search on  $\text{CaCO}_3$  with the only input information of chemical composition. At the 13th generation, the enthalpy shows a pronounced drop, and the examination of lowest enthalpy structure confirmed the convergence to the experimental calcite structure [Fig. 5(b)].

In order to demonstrate the performance of our algorithm, we present a direct comparison with the available data<sup>9</sup> from GA for several systems as listed in Table IV. It can be clearly

seen that all of the predicted structures in our methodology were identified within five generations, i.e., our method only needs fewer optimization steps. The reasons why our method is efficient can be traced back to several true facts: (i) PSO is a highly efficient global minimization algorithm; (ii) the symmetry constraints are imposed to significantly reduce the search space and utilize the structure diversity; (iii) the elimination of similar structures has greatly reduced the number of local optimizations, expediting the global minimization convergence; (iv) a variable cell-size technique was implemented to efficiently reduce the computational cost.

### IV. CONCLUSION

We have developed a systemic methodology for the crystal structure prediction based on PSO technique within evolutionary scheme as implemented in CALYPSO code. Our method could efficiently search the free-energy space of the lattice geometry and atomic configuration of a solid looking for the ground-state and metastable structures in complex

TABLE IV. The comparison between PSO and GA from Ref. 9 for several systems with equal population sizes ( $N_{pop}$ ).  $N_{opt}$  is short for the number of locally optimized structures. In our PSO results, the number of generations and  $N_{opt}$  are averaged out over five different runs. It is noteworthy that the GA data from Ref. 9 is published in 2007 and one should aware that later GA algorithm can be significantly improved. The comparison of PSO with GA is only to illustrate the true facts of PSO performance.

Systems	Algorithm	Pressure (GPa)	Structures	Generations	$N_{pop}$	$N_{opt}$
Si	PSO	0	Diamond	8/5	16	128/5
	GA <sup>a</sup>	0	Diamond	15	16	60
SiC	PSO	0	Zinc blende	8/5	12	96/5
	GA <sup>a</sup>	0	Zinc blende	5	12	20
GaAs	PSO	0	Zinc blende	16/5	12	192/5
	GA <sup>a</sup>	0	Zinc blende	19	12	70

<sup>a</sup>Reference 9.

systems. The key elements of the proposed approach are PSO algorithm, the state-of-art *ab initio* structural optimization based on density-functional theory, the symmetry constraint on the structural generation, and the geometrical structure parameter technique on elimination of the similar structures. We have implemented a variable cell-size technique enabling the intelligent selection of the correct cell sizes, and thus significantly reducing the computational cost. This methodology has been successfully applied to various known experimental structures on elemental, binary, and ternary compounds with, but not limited to, metallic, ionic, and covalent bonding. Our method is proved to be powerful with high efficiency and high success rate. Future development of this efficient PSO technique on prediction of much larger systems (say  $\sim 100$  atoms/cell or above) is foreseeing fea-

sible and thus the predictive power on structure solutions of nanomaterials, surface or thin films, and biomaterials are highly expected. Within this PSO algorithm the doors toward materials design (e.g., design of novel superconductive, thermoelectric, superhard, and energetic materials, etc.) could open.

## ACKNOWLEDGMENTS

We thank M. Probert for helpful discussions. The authors acknowledge funding from the National Natural Science Foundation of China under Grants No. 10874054 and No. 91022029, and the research fund for excellent young scientist in Jilin University (Grant No. 200905003).

\*Author to whom correspondence should be addressed; mym@jlu.edu.cn

- <sup>1</sup>J. Maddox, *Nature (London)* **335**, 760 (1988).
- <sup>2</sup>G. Day, W. D. S. Motherwell, H. L. Ammon, S. X. M. Boerrigter, R. G. Della Valle, E. Venuti, A. Dzyabchenko, J. D. Dunitz, B. Schweizer, B. P. van Eijck, P. Erk, J. C. Facelli, V. E. Bazterra, M. B. Ferraro, D. W. M. Hofmann, F. J. J. Leusen, C. Liang, C. C. Pantelides, P. G. Karamertzanis, S. L. Price, T. C. Lewis, H. Nowell, A. Torrisi, H. A. Scheraga, Y. A. Arnautova, M. U. Schmidt, and P. Verwer, *Acta Crystallogr., Sect. B: Struct. Sci.* **61**, 511 (2005).
- <sup>3</sup>S. Kirkpatrick, C. D. Gelatt, Jr., and M. P. Vecchi, *Science* **220**, 671 (1983).
- <sup>4</sup>J. Pannetier, J. Bassas-Alsina, J. Rodriguez-Carvajal, and V. Caignaert (unpublished).
- <sup>5</sup>S. Woodley, P. Battle, J. Gale, and C. Catlow, *Phys. Chem. Chem. Phys.* **1**, 2535 (1999).
- <sup>6</sup>N. L. Abraham and M. I. J. Probert, *Phys. Rev. B* **73**, 224104 (2006).
- <sup>7</sup>A. Oganov and C. Glass, *J. Chem. Phys.* **124**, 244704 (2006).
- <sup>8</sup>C. Glass, A. Oganov, and N. Hansen, *Comput. Phys. Commun.* **175**, 713 (2006).
- <sup>9</sup>G. Trimarchi and A. Zunger, *Phys. Rev. B* **75**, 104113 (2007).
- <sup>10</sup>D. Wales and J. Doye, *J. Phys. Chem. A* **101**, 5111 (1997).
- <sup>11</sup>A. Nayeem, J. Vila, and H. Scheraga, *J. Comput. Chem.* **12**, 594 (1991).
- <sup>12</sup>R. Martoňák, A. Laio, and M. Parrinello, *Phys. Rev. Lett.* **90**, 075503 (2003).
- <sup>13</sup>R. Martoňák, A. Laio, M. Bernasconi, C. Ceriani, P. Raiteri, F. Zipoli, and M. Parrinello, *Z. Kristallogr.* **220**, 489 (2005).
- <sup>14</sup>C. Pickard and R. Needs, *Phys. Rev. Lett.* **97**, 045504 (2006).
- <sup>15</sup>C. Pickard and R. Needs, *Nat. Phys.* **3**, 473 (2007).
- <sup>16</sup>C. Pickard and R. Needs, *Nature Mater.* **7**, 775 (2008).
- <sup>17</sup>A. Mujica and R. J. Needs, *Phys. Rev. B* **55**, 9659 (1997).
- <sup>18</sup>D. Wales and H. Scheraga, *Science* **285**, 1368 (1999).
- <sup>19</sup>J. Behler, R. Martoňák, D. Donadio, and M. Parrinello, *Phys. Rev. Lett.* **100**, 185501 (2008).
- <sup>20</sup>C. Pickard and R. Needs, *Nature Mater.* **9**, 624 (2010).
- <sup>21</sup>Y. Ma, M. Eremets, A. Oganov, Y. Xie, I. Trojan, S. Medvedev, A. Lyakhov, M. Valle, and V. Prakapenka, *Nature (London)* **458**, 182 (2009).
- <sup>22</sup>A. Oganov, J. Chen, C. Gatti, Y. Ma, C. Glass, Z. Liu, T. Yu, O. Kurakevych, and V. Solozhenko, *Nature (London)* **457**, 863 (2009).
- <sup>23</sup>Q. Li, Y. Ma, A. Oganov, H. Wang, Y. Xu, T. Cui, H. Mao, and G. Zou, *Phys. Rev. Lett.* **102**, 175506 (2009).
- <sup>24</sup>Y. Ma, Y. Wang, and A. R. Oganov, *Phys. Rev. B* **79**, 054101 (2009).
- <sup>25</sup>G. Gao, A. Oganov, A. Bergara, M. Martinez-Canales, T. Cui, T. Iitaka, Y. Ma, and G. Zou, *Phys. Rev. Lett.* **101**, 107002 (2008).
- <sup>26</sup>Y. Xie, A. Oganov, Y. Ma, and T. However, *Phys. Rev. Lett.* **104**, 177005 (2010).
- <sup>27</sup>G. Gao, A. R. Oganov, P. Li, Z. Li, H. Wang, T. Cui, Y. Ma, A. Bergara, A. O. Lyakhov, T. Litaka, and G. Zou, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 1317 (2010).
- <sup>28</sup>Y. Li, G. Gao, Y. Xie, Y. Ma, T. Cui, and G. Zou, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 15708 (2010).
- <sup>29</sup>J. Kennedy and R. Eberhart, *Particle Swarm Optimization* (IEEE, Piscataway, NJ, 1995), p. 1942.
- <sup>30</sup>R. Eberhart and J. Kennedy, *A New Optimizer Using Particle Swarm Theory* (IEEE, New York, NY, 1995).
- <sup>31</sup>H. Lu, *Dynamic Population Strategy Assisted Particle Swarm Optimization in Multiobjective Evolutionary Algorithm Design* (IEEE Neural Network Society, IEEE NNS Student Research Grants 2002, NY, 2002).
- <sup>32</sup>R. Eberhart and Y. Shi, *Particle Swarm Optimization: Developments, Applications, and Resources* (IEEE, Piscataway, NJ, 2001), p. 81.
- <sup>33</sup>K. Parsopoulos and M. Vrahatis, *Particle Swarm Optimization Method in Multiobjective Problems* (ACM, New York, NY, 2002), p. 603.
- <sup>34</sup>H. Yoshida, K. Kawata, Y. Fukuyama, S. Takayama, and Y. Nakanishi, *IEEE Trans. Power Syst.* **15**, 1232 (2000).
- <sup>35</sup>K. Parsopoulos and M. Vrahatis, *IEEE Trans. Evol. Comput.* **8**, 211 (2004).
- <sup>36</sup>Freely distributed on Academic use, request should be sent to Yanming Ma, mym@jlu.edu.cn
- <sup>37</sup>E. Prince and C. H. Spiegelman, *International Tables for Crystallography* (Kluwer Academic, Dordrecht, 1992).
- <sup>38</sup>N. L. Abraham and M. I. J. Probert, *Phys. Rev. B* **77**, 134117

- (2008).
- <sup>39</sup>R. Eberhart and Y. Shi, *Comparing Inertia Weights and Construction Factors in Particle Swarm Optimization* (IEEE, Piscataway, NJ, 2000), p. 84.
  - <sup>40</sup>P. E. Blochl, *Phys. Rev. B* **50**, 17953 (1994).
  - <sup>41</sup>G. Kresse and J. Furthmuller, *Phys. Rev. B* **54**, 11169 (1996).
  - <sup>42</sup>C. Barrett, *Acta Crystallogr.* **9**, 671 (1956).
  - <sup>43</sup>A. W. Overhauser, *Phys. Rev. Lett.* **53**, 64 (1984).
  - <sup>44</sup>B. Olinger and J. Shaner, *Science* **219**, 1071 (1983).
  - <sup>45</sup>M. Hanfland, K. Syassen, N. E. Christensen, and D. L. Novikov, *Nature (London)* **408**, 174 (2000).
  - <sup>46</sup>R. Rousseau, K. Uehara, D. Klug, and J. Tse, *ChemPhysChem* **6**, 1703 (2005).
  - <sup>47</sup>C. J. Pickard and R. J. Needs, *Phys. Rev. Lett.* **102**, 146401 (2009).
  - <sup>48</sup>Y. Yao, J. Tse, and D. Klug, *Phys. Rev. Lett.* **102**, 115503 (2009).
  - <sup>49</sup>Y. Ma, A. R. Oganov, and Y. Xie, *Phys. Rev. B* **78**, 014102 (2008).
  - <sup>50</sup>H. Monkhorst and J. Pack, *Phys. Rev. B* **13**, 5188 (1976).
  - <sup>51</sup>P. Trucano and R. Chen, *Nature (London)* **258**, 136 (1975).
  - <sup>52</sup>D. Rilby, *Nature (London)* **153**, 587 (1944).
  - <sup>53</sup>R. O. Piltz, J. R. Maclean, S. J. Clark, G. J. Ackland, P. D. Hatton, and J. Crain, *Phys. Rev. B* **52**, 4072 (1995).
  - <sup>54</sup>J. Kasper and S. Richards, *Acta Crystallogr.* **17**, 752 (1964).
  - <sup>55</sup>H. Olijnyk, S. Sikka, and W. Holzapfel, *Phys. Lett. A* **103**, 137 (1984).
  - <sup>56</sup>J. Jamieson, *Science* **139**, 762 (1963).
  - <sup>57</sup>M. I. McMahon, R. J. Nelmes, N. G. Wright, and D. R. Allan, *Phys. Rev. B* **50**, 739 (1994).
  - <sup>58</sup>M. Hanfland, U. Schwarz, K. Syassen, and K. Takemura, *Phys. Rev. Lett.* **82**, 1197 (1999).
  - <sup>59</sup>S. J. Duclos, Y. K. Vohra, and A. L. Ruoff, *Phys. Rev. Lett.* **58**, 775 (1987).
  - <sup>60</sup>G. Raynor and W. Hume-Rothery, *J. Inst. Met.* **65**, 379 (1939).
  - <sup>61</sup>H. Olijnyk and W. B. Holzapfel, *Phys. Rev. B* **31**, 4682 (1985).
  - <sup>62</sup>C. J. Pickard and R. J. Needs, *Phys. Rev. B* **81**, 014106 (2010).
  - <sup>63</sup>L. Levien, C. Prewitt, and D. Weidner, *Am. Mineral.* **65**, 920 (1980).
  - <sup>64</sup>S. Stishov and S. Popova, *Geochem.* **10**, 837 (1961).
  - <sup>65</sup>K. Kingma, R. Cohen, R. Hemley, and H. Mao, *Nature (London)* **374**, 243 (1995).
  - <sup>66</sup>L. S. Dubrovinsky, S. K. Saxena, P. Lazor, R. Ahuja, O. Eriksson, J. M. Wills, and B. Johansson, *Nature (London)* **388**, 362 (1997).
  - <sup>67</sup>Y. Kuwayama, K. Hirose, N. Sata, and Y. Ohishi, *Science* **309**, 923 (2005).
  - <sup>68</sup>A. Addamiano, *Silicon Carbide-1973*, edited by R. C. Marshall, J. W. Faust, Jr., and C. E. Ryan (University of South Carolina Press, Columbia, SC, 1974).
  - <sup>69</sup>T. Sekine and T. Kobayashi, *Phys. Rev. B* **55**, 8034 (1997).
  - <sup>70</sup>C. Bates, W. White, and R. Roy, *Science* **137**, 993 (1962).
  - <sup>71</sup>L. Padurets, Z. Dobrokhotova, and A. Shilov, *Int. J. Hydrogen Energy* **24**, 153 (1999).
  - <sup>72</sup>J. H. Weaver, D. J. Peterman, D. T. Peterson, and A. Franciosi, *Phys. Rev. B* **23**, 1692 (1981).
  - <sup>73</sup>S. Okada, T. Atoda, I. Higashi, and Y. Takahashi, *J. Mater. Sci.* **22**, 2993 (1987).
  - <sup>74</sup>B. Post, F. W. Glaser, and D. Moskowitz, *Acta Metall.* **2**, 20 (1954).
  - <sup>75</sup>A. Oganov and S. Ono, *Nature (London)* **430**, 445 (2004).
  - <sup>76</sup>D. Graf, *Am. Mineral.* **46**, 1283 (1961).