7
Crystal Structure Prediction Using Evolutionary Approach

Andriy O. Lyakhov, Artem R. Oganov, and Mario Valle

If a mathematician had to formulate a problem that we face in crystal structure prediction, it would probably sound like this: ''Find the global minimum on a very noisy landscape in a multidimensional space.'' The search space is so huge that one could not do an exhaustive search in any reasonable time even using all the computing power available for the mankind in a near future. This obviously nontrivial problem is further complicated by the fact that we do not know the exact form of the landscape. All we can do is to calculate the ''height'' (free energy) given the coordinates (structure parameters, such as atom positions) and relax the system to the nearest local minimum. This problem is in fact so hard that 15 years ago the answer on the question ''Are the structures predictable?'' was a clear ≪No≫: just by writing this concise statement, in what would be the first one-word paper in the chemical literature, one could safely summarize the present state of affairs'' [1]; see also [2]. However mathematicians have been trying to solve similar problems for over half a century – since first computers were created. And actually they found a plethora of methods that are able to provide us with reasonably good solutions. Some of these methods are problem specific and their success is based on problem constrains and symmetries. But some approaches are very general and thus could be applied to crystal structure prediction (some of them are described in other chapters of this book). One of the most general and powerful techniques that at the same time is adaptable for any specific problem is the evolutionary algorithm approach.

Before going deeper into the details, we would like to mention that no method is currently available which is able to solve a *general* global optimization problem or even prove that the solution found by any means is in general case a global optimum. All algorithms are essentially heuristic – they are able to found a reasonable solution in a reasonable time without a guarantee that this solution is the best one or even close to it. We can only hope that it is. However, the situation is not as grim as one may think. Years of practice show us that usually solutions are quite good and quite often you can indeed find the best solution or a really close one. Evolutionary algorithms are not an exception. They were used in many different fields and produced great results where other methods failed. For example, an evolutionary approach could give a reasonably good solution for the

Modern Methods of Crystal Structure Prediction. Edited by Artem R. Oganov

ISBN: 978-3-527-40939-6

traveling salesman problem with 10 000 cities [3], where the number of possible solution candidates is far beyond astronomical ($\sim 10^{35\,000}$).

7.1 Theory

The name of the approach – evolutionary – indicates that it uses mechanisms inspired by biological evolution: reproduction, mutation, recombination, and selection. The candidate solutions to the optimization problem are individuals in a population that evolves under the repeated application of the above mechanisms, and a fitness function determines how much any of them survives. In more detail, the basic steps of the evolutionary technique (see Figure 7.1) are as follows:

1) The evolutionary approach starts by choosing the adequate representation for the problem: a one-to-one correspondence between the point in the search space and a set of numbers. This is a very important yet neglected step and the quality of the representation has direct impact on the effectiveness of the algorithm.

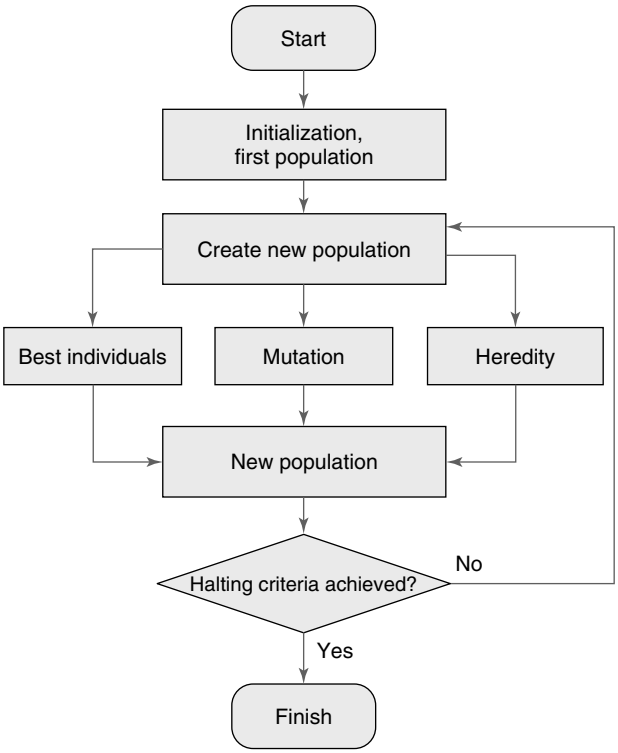

Figure 7.1 Block scheme of the typical evolutionary algorithm.

2) Initialization of the first generation, that is, a set of points in the search space that satisfy the problem constrains.
3) Determination of the quality for each member of the population using the so-called fitness function.
4) Selection of the ''best'' members from the current generation as parents from which the algorithm creates new points (offspring) in the search space by applying specially designed variation operators to them.
5) Evaluation of the quality for each new member of the population.
6) Selection of the ''best'' offspring to build the new generation of the population.
7) Repeat steps 4–6 till some halting criteria is achieved.

One can see that some biological terms are used to describe this approach:

- *Population*. A set of points in the search space (which we will further call solutions) that are analyzed as possible candidates for the optimal solution.
- *Parents*. A set of solutions that are used to create new candidates for the optimum solution.
- *Offspring*. A set of solutions created from parents using variation operators.
- *Selection*. A process that divides solutions into ones that have to ''die'' and ones that have to ''survive'' to build the next generation.

Two main classes of variation operators also have biological names:

- Heredity operators use a few parent solutions to build one offspring solution.
- Mutation operators use a single-parent solution to produce a single child.

Why are evolutionary algorithms so effective? No one actually knows, and as we wrote earlier no one could prove that the best solution found by the algorithm is in general close to the optimum one. However there are two things that inspired people to develop this class of algorithms and made us believe that this approach is indeed a good one. First of all in a correctly designed evolutionary algorithm the quality of the best solution(s) in each new generation (quantified by fitness function) is at least not worse than in the previous one. We can simply keep the best solution or any given number of best solutions, if none of the offspring solutions are better than them. And, secondly, you could see the success of the algorithm executed by Mother Nature if you look into the mirror or hug your pet.

There is one thing to keep in mind during the design and execution of every evolutionary algorithm. You have to deal with the trade-off: diversity of the population versus convergence to the optimum solution. Higher diversity of the population means that you can explore your search space better. However it slows down the process of finding the minimum even if you have a few structures in its basin of attraction. On the other hand, using elitist approaches and reducing the diversity helps your population collapse into a local minimum faster while the risk of omitting the global optimum is higher. The desired balance between diversity and convergence can usually be achieved only if you can tune it ''on the fly'' when you reveal more information about the system and your search space.

7.1.1 Search Space, Population, and Fitness Function

Crystal structure prediction requires us to find the global minimum on the free-energy landscape (that we will call the search space) for system of a given stoichiometry. Each point on this landscape (solution) represents a crystal structure with certain atomic positions and lattice vectors. The set of locally optimized solutions we will call a population.

One of the features of evolutionary algorithms, which is very helpful for the crystal structure prediction problem, is their ability to find metastable states – good local minima on the energy landscape that are clearly separated from the global minimum.

Fitness function describes the quality of each solution and allows us to compare them. Naturally, the free energy would be the relevant fitness function for a crystal structure prediction algorithm. Lower free energy will correspond to better solution and the most stable structure under given conditions will have the lowest fitness function value.

7.1.2 Representation

As we have already mentioned, the choice of the right representation is crucial for the effectiveness of the algorithm. One of the reasons why first attempts to design an evolutionary algorithm for crystal structure prediction ([4, 5]; see also [6]) had small success was counterproductive representation choice. In these approaches, a discrete grid of atom positions was used. The structural variables of the crystal were represented by a binary string, and standard evolution operators for binary strings were applied. These operators were not physically meaningful and, therefore, the algorithm was basically performing a random walk in the search space. Obviously it could reliably find the global minimum only in the simplest systems, and even in this case the number of optimizations it has to do is comparable with the random search methods. Later algorithms use the real-number representation for atom positions and lattice parameters [7, 8]. This representation requires more sophisticated variation operators that are better ''optimized'' for their task and allow researchers to build powerful methods for crystal structure prediction.

In the rest of this chapter, we will describe our evolutionary algorithm Universal Structure Predictor: Evolutionary Xtallography (USPEX) and results that were achieved with it [8, 9]. USPEX represents the coordinates of atoms in the unit cell and lattice vectors by real numbers. Therefore, the search space is continuous and not discrete like in algorithms with binary string representation. The difficulty of the problem is increased, but this is more than compensated by the possibility to construct physically intuitive and powerful variation operators. Using real number representation, we also have less risk to omit some shallow local minima.

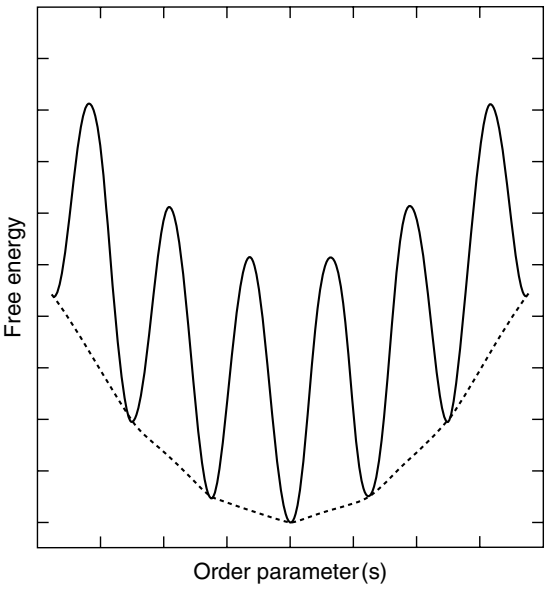

Figure 7.2 Reducing the noise in the complex search space after local optimization.

7.1.3 Local Optimization and Constrains

Each structure produced by USPEX and similar modern algorithms is locally optimized. Local optimization is a crucial part of the technique because otherwise it would be impossible to find the lowest free-energy structure for any reasonable complex system. It also makes the search space much more "smooth" (see Figure 7.2) and reduces the intrinsic dimensionality of the energy landscape,[1] see Table 7.1.

Since complexity of the problem has exponential dependence on the dimensionality, local optimization greatly simplifies the task of finding the global minimum on the energy landscape. USPEX uses external tools for local optimization. So far it can use VASP [12] and SIESTA [13] for first-principles optimization and GULP [14] for semiclassical simulations. It is also relatively easy to add support of other energy-computing engines, if needed.

We would like to mention that not every structure on the energy landscape is a chemically and physically feasible solution. If we try, for example, to optimize the structure where two atoms are located in the same spot then the calculation may

1) Intrinsic dimensionality is the minimum number of variables that are needed to represent the search space. Local optimization, by creating correlations between atomic positions (such as favoring the formation of some bonds and disfavoring others), decreases the intrinsic dimensionality and dramatically simplifies the crystal structure prediction problem.

Table 7.1 Intrinsic vs. extrinsic dimensionality[a].

System	Extrinsic dimensionality	Intrinsic dimensionality
Au_8Pd_4	39	10.9
$Mg_{16}O_{16}$	99	11.6
$Mg_4N_4H_4$	39	32.5

[a]The numbers were produced using the molecular visualization toolkit STM4 [10] using the Grassberger–Procaccia algorithm with a suitably adapted Camastra sampling correction [11].

''explode'' the unit cell and we would not get any meaningful results. Therefore, we have to apply constrains to discard unfeasible solutions. The structure is considered unfeasible if:

1) The distance between any two atoms is smaller than threshold determined by user (e.g., there are no known bonds shorter than 0.5 Å). One can set different thresholds for different pairs of atom types separately; for example, the sum of correspondent atom radii.
2) One of the lattice vectors is too small. User can determine the threshold value; for example, it can be set to the diameter of the largest atom present in the system.
3) The angle between two lattice vectors is too small or the angle between the lattice vector and the diagonal of the parallelogram formed by other two lattice vectors is too small, see Figure 7.3. One can always choose the lattice vectors in such a way that the angle between any of them is in the $(60^\circ, 120^\circ)$ range. To do this, one can replace the longer vector (let's say **a**) in a pair violating the constrain by $\mathbf{a}' = \mathbf{a} - ceil\left(|\mathbf{a}\cdot\mathbf{b}|/\|\mathbf{b}\|^2\right)\cdot sign(\mathbf{a}\cdot\mathbf{b})\cdot\mathbf{b}$ see [20] for more details. Our experience shows that this procedure speeds up local optimization and improves the efficiency of the algorithm as a whole[2)].

7.1.4 Initialization of the First Generation

Common way to create the first generation in evolutionary approach is uniform random sampling, aimed at achieving high diversity of the population. In fact, if you do not have any *a priory* information about how the possible optimal solution could look like, random sampling is the most reasonable way to do the unbiased

2) The 60–120 degree conditions are a simplified version of the cell constraints – more complete conditions are that a unit cell vector (and also cell diagonals) should not have projections onto other cell vectors that are greater by absolute value than half of the latter vector [20].

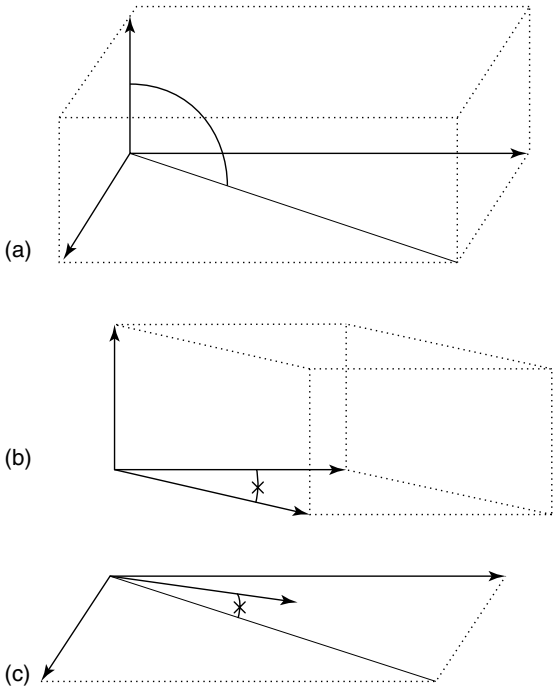

Figure 7.3 Cell angles constrains [20]. (a) All constrains satisfied, (b) angle between lattice vectors is too small, and (c) angle between lattice vector and the diagonal of the parallelogram formed by other two lattice vectors is too small.

initialization. You just pick up random points in the search space and check them for feasibility. Feasible solutions are added to the first generation. This process is repeated until the desired number of trial solutions is reached. In our case it means that we randomly choose the lattice vectors and then randomly drop the atoms into the unit cell. If some information about the optimum solution is known, for example, unit cell volume or lattice parameters, this can be used as constraints for optimization. In this case one would, for example, fix the lattice parameters and vary only atomic positions within the unit cell. User can also ''seed'' the first generation with the structures that seem reasonable (e.g., those known for similar compounds, or for the same compound at different conditions, or coming from previous structure prediction runs) and fill the rest of it with random ones.

High diversity of the first generation is the key to the success of the algorithm. If we do not have a structure in the basin of attraction (so-called funnel, see Figure 7.4) of the global minimum, then the probability to find that optimal solution can be low.

Random initialization, however, poses a problem relevant for large systems. When the number of atoms in the unit cell rises, randomly generated structures become more and more similar [15] from chemical point of view to each other

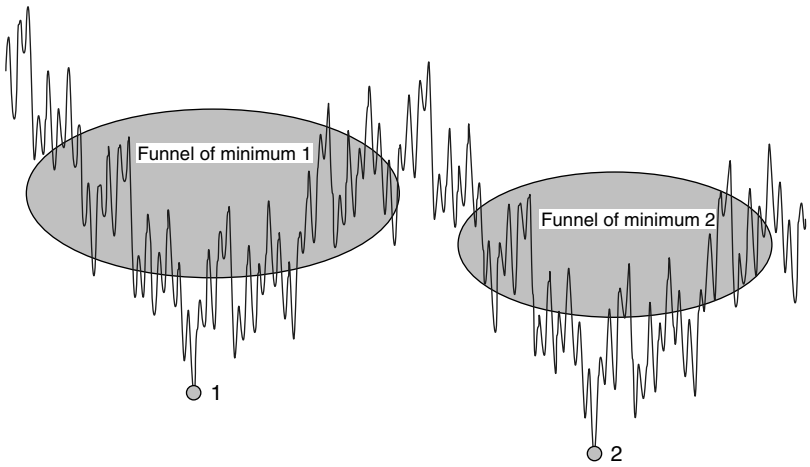

Figure 7.4 Basin of attraction (funnel) of a good local minimum is a set of surrounding local minima. Different funnels typically contain very distinct types of structures and are separated by high-energy barriers.

and to a disordered system. This can be visualized as trying to build a crystal by replicating small volumes of liquid. Thus purely random sampling cannot be used for effective crystal structure prediction in this case. However, there is a good method that combines relatively high diversity of the first population with a high degree of randomness. We call it unit cell splitting: large cell is split into smaller subcells that are filled with atoms randomly and then replicated to fill the full cell, see Figure 7.5a. Such structures have a higher translational symmetry and are more ordered, usually leading to a more diverse population. To help to break this

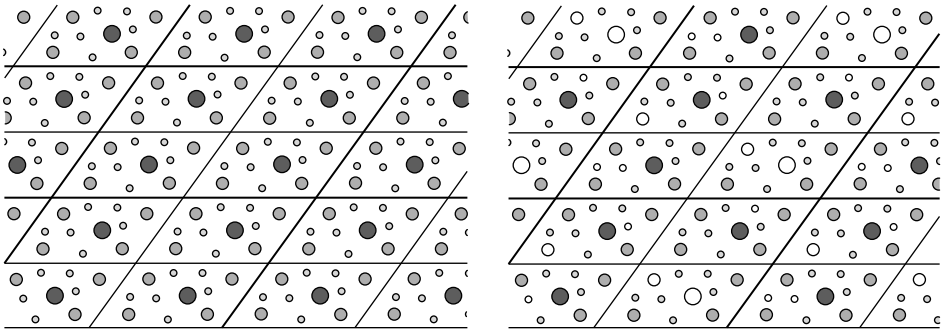

Figure 7.5 Schematic illustration of the (a) subcell and (b) pseudo-subcell splitting for compositions (a) $A_4B_{16}C_{20}$ and (b) $A_3B_{14}C_{16}$ (atoms A – large black circles, B – medium dark-gray circles, C – small gray circles, empty circles – vacancies). Thick lines show the true unit cell, split into four (pseudo-)subcells.

undesired ''additional'' symmetry by variation operators (described below) as well as further increase the diversity of the first population, different structures have to be split into different number of subcells.

For cells where the number of atoms (e.g., a prime number) is not good for splitting into a small number of identical subcells, the algorithm creates random vacancies to keep the correct number of atoms in the whole unit cell, see Figure 7.5b. In this case, no additional symmetry is induced and nontrivial solutions can be found. We also believe that this pseudo-subcell method is able to improve conventional random sampling methods [16, 17] when dealing with large systems.

7.1.5
Variation Operators

In general, the choice of variation operators follows naturally from the representation. Mutation operators usually randomly distort the numbers from the set that represents the solution, while heredity operators combine different parts of these sets from different parent solutions into one child solution. For real-number crystal structure representation, we use two different types of mutation operators – lattice mutation and atom permutation.

Lattice mutation applies strain matrix with zero-mean Gaussian random strains to the lattice vectors:

$$L_i' = (I + \varepsilon)L_i = \begin{pmatrix} 1+\varepsilon_1 & \varepsilon_6/2 & \varepsilon_5/2 \\ \varepsilon_6/2 & 1+\varepsilon_2 & \varepsilon_4/2 \\ \varepsilon_5/2 & \varepsilon_4/2 & 1+\varepsilon_3 \end{pmatrix} L_i$$

Lattice mutation is shown in Figure 7.6. The position of atoms (their fractional coordinates within the lattice) remains unchanged. This operator allows the algorithm

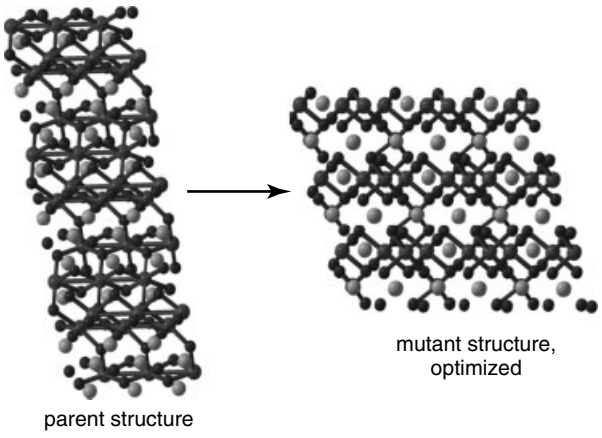

Figure 7.6 Lattice mutation applies strain to the lattice vectors. From Oganov *et al*. [9].

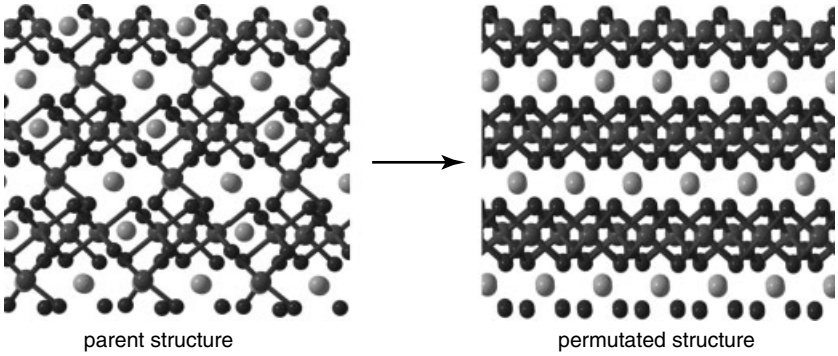

Figure 7.7 Permutation operator swaps identities of a few atom pairs. From Oganov *et al*. [9].

to investigate the neighborhood of good individuals. Also sometimes structures of similar quality differ essentially in lattice and that is why the premature convergence of the lattice to some ''optimal'' value reduces the effectiveness of the algorithm. Lattice mutation operator increases the diversity of the lattices in the population.

Atom permutation operator swaps chemical identities of atoms in randomly selected pairs, see Figure 7.7, while lattice remains unchanged. Such swaps provide the algorithm with steps in the search space that are very far and difficult in Euclidean distance. This operator is especially useful for systems, where chemically similar atoms are present.

Heredity operators are vital part of any evolutionary algorithm. If only mutation operators are present, then the algorithm is identical to a sophisticated version of a random search (such as D. Wales's basin hopping method [18]). Heredity operators are responsible for utilizing and refining the information about the system that we gather during the execution of the algorithm. Since properties of the crystal are determined by spatial arrangement of atoms in the unit cell, the most physically meaningful way to build a heredity operator is to conserve the information from parents by using spatially coherent pieces (spatial heredity).

To create a child from two parents, the algorithm first randomly chooses the lattice vector and a point on that vector. Then the unit cells of the parent structures are cut by the plane parallel to other vectors that goes through this point. Planar slices are matched, see Figure 7.8, and the number of atoms of each kind is adjusted. For big cells, it is possible to use more than two structures as parents and combine slices from all of them into a single child structure. The lattice of the child structure is a weighted averaged of parent lattices.

Altogether three variation operators described above explore the search space while preserving and refining the good spatial features through generations. It can be visualized by comparing the best structures from different generations, see, for example, Figures 7.9 and 7.10.

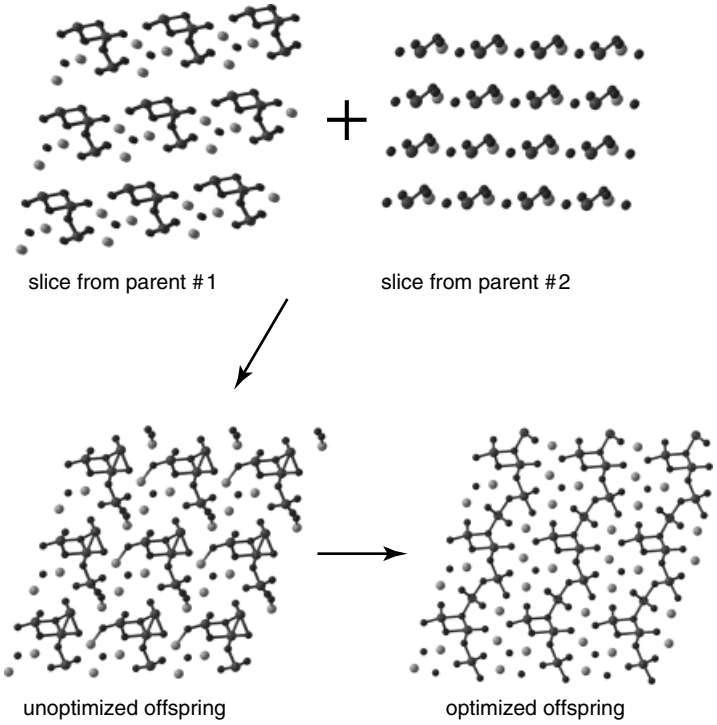

Figure 7.8 Heredity operator combines spatial slices from different parent structures to form an offspring structure. From Oganov *et al*. [9].

7.1.6
Survival of the Fittest and Selection of Parents

When the offspring structures are produced, we have to create a new generation. Usually it is a good idea to select the best few structures from the previous generation into the new one. This will make sure that we would not lose the best structures found during the algorithm execution and is also helpful if we want to search for low-energy metastable states, in addition to the global minimum. New generation is filled with the offspring structures, which are obtained from parent structures using variation operators. But which structures are to be chosen as parents? We rank the solutions by their free energy, worst 40% of structures are discarded (this threshold can be changed by user), and the probability for structure among best 60% to be chosen as a parent is proportional to its fitness rank. Stochastic selection is usually slower than deterministic one, when we choose only the best offspring to pass selection. However, it maintains a relatively high diversity of the population, which saves the simulation from being ''trapped'' in the basin of attraction of a good local minimum that is not the global one.

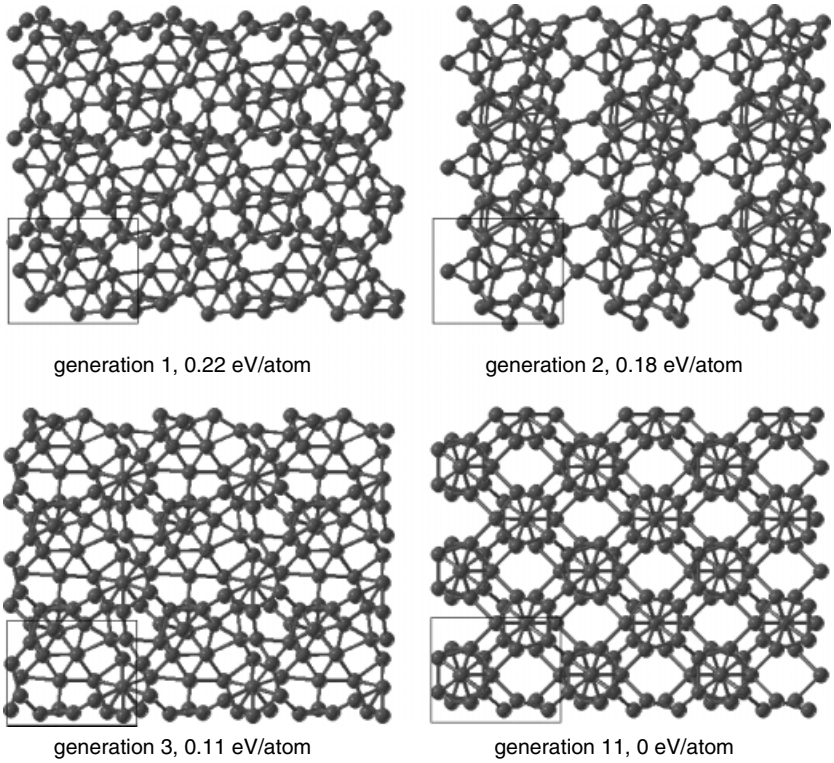

Figure 7.9 Best structures in generation from the evolutionary search for boron structure at 1 atm. The energy per atom shows the deviation from the ground state. One can see that parts of the icosahedra that build the most stable structure started to appear in the very first generations and were slowly refined during the run. From Oganov *et al*. [19].

7.1.7 Halting Criteria

After the new generation is produced and structures are relaxed, algorithm applies selection and variation operators to create the next one. This process is repeated until some halting criteria are achieved. Structure with the best enthalpy is considered as a candidate for the ground state of the system under given conditions. One can repeat the simulation a few times to increase the confidence of the results.

Obvious halting criterion would be reaching a certain number of subsequent generations where the best structure is not changing. However, sometimes we may need more sophisticated criteria. For example, we can continue calculations till the certain level of diversity is maintained and stop only when the population is filled with structures from a around the best structure. Alternatively, we may know something about the desired structure, for example, symmetry group or some other

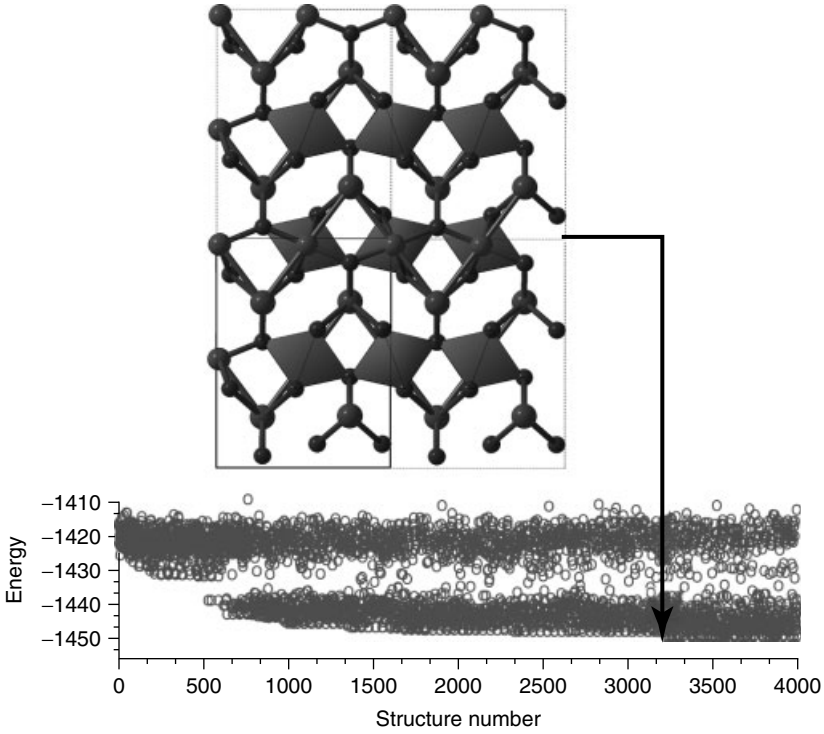

Figure 7.10 An example of a large test: prediction of the crystal structure of $MgSiO_3$ with 80 atoms in the supercell. The lower panel shows that the global-minimum structure was found within ~3200 attempts. From Oganov and Glass [20].

information. In this case, we should continue calculation until the best structure has desired properties or the whole population is converged to a single funnel.

7.1.8
Premature Convergence and How to Prevent It: Fingerprint Function

As we already mentioned earlier, evolutionary algorithms always risk to be ''trapped'' in a funnel around some local minimum that is not a global optimum. This happens because good structures tend to produce children in their vicinity, filling the population with its own replicas and structures from their basin of attraction, reducing the diversity of the population. Such algorithm behavior (that we call ''cancer growth phenomena'') is especially common for energy landscapes with many good local minima surrounded by relatively low-energy areas that are separated from each other by high barriers. Sadly, this type of energy landscape is rather common. Therefore, a method that allows us to control the diversity of

the population and avoid its premature convergence into one basin of attraction would increase the effectiveness of the algorithm.

The first question that arises when you try to develop such a method is ''how can we detect similar structures and measure the degree of similarity between them?'' Direct comparison of atomic coordinates will not work because they are represented in lattice vectors units and there are many equivalent ways to choose a unit cell. Free energy difference will also fail to describe the similarity between the structures because the search space is too noisy and not monotonic. We need some method that does not depend on unit cell choice and is determined solely by structure (and not representation). Ideally, at the same time, it should distinguish the structures where two atoms of different sorts are swapped in their positions. It should also be robust to small numerical errors that are unavoidable in real calculations. Nowadays scientists start to explore such methods. The most common approach to capture the essential geometrical properties of the structure is to use a specially constructed crystal structure descriptor, for example, radial distribution function (RDF).

In our algorithm, we use so-called fingerprint function [21, 22] to describe a crystal structure. It is a function, related to RDF and diffraction spectra, defined as

$$f(R) = \sum_i \sum_{j \neq i} \frac{Z_i Z_j}{4\pi R_{ij}^2} \frac{V}{N} \delta(R - R_{ij})$$

Here Z_i is the atomic number for atom i, R_{ij} is the distance between atoms i and j, V is the unit cell volume, and N is the number of atoms in the unit cell. We would like to note that R is a variable, not a parameter. The index i goes over all atoms in the unit cell, while index j goes over all atoms within some cutoff distance from the atom i. To remove fingerprint dependency from cutoff distance, the function is normalized as follows:

$$f_n(R) = \frac{f(R)}{\sum_{i,j} Z_i Z_j N_i N_j} - 1$$

Here N_i is the number of atoms in the unit cell with atomic number Z_i and the two sums go over all distinct Z values.

Fingerprint function as a method to describe the crystal structure has all the desired properties listed above. First of all, it does not depend on absolute atomic coordinates, but only on interatomic distances. Therefore, the choice of unit cell will not influence $f(R)$. Small perturbations of atomic positions will influence fingerprint function only slightly. Using atomic numbers as weighting coefficients allows us to take into account atom ordering. One could also measure the similarity between structures by computing the distance between their fingerprint functions. In our algorithm we used cosine distance, but one could use other metrics as well, for example, Cartesian distance or Minkowski norm [21].

To simplify and speed up the calculations, we discretize the fingerprint function and represent it as a vector *FP*, called fingerprint.

$$FP_i(R) = \frac{1}{D}\int_{iD}^{(i+1)D} f_n(R)dR$$

Cosine distance between two fingerprints for structures i and j is then defined as

$$d_{ij} = 0.5\left(1 - \frac{FP_i FP_j}{\|FP_i\|\,\|FP_j\|}\right)$$

Having provided the fingerprint's space with a distance measure, we could group fingerprints, and thus structures, using a ''similarity'' or ''almost equality'' criteria.

7.1.9 Improved Selection Rules and Heredity Operator

The ability to measure the degree of similarity between structures allows us to improve the selection rules and variation operators described above [15]. First of all, determining the structures that will participate in building the next generation we ignore all similar and choose only different ones. Two structures are considered ''similar'' if the distance between their fingerprints is less than some user-defined threshold. This will increase the diversity of the offspring population and preclude premature convergence into a single funnel.

We already mentioned that letting a few best structures to survive and become members of the next generation is helpful if we search for metastable states and other low-energy (but not ground state) solutions. This can also be extremely useful for improving global optimization, as it increases the learning power of the algorithm. However, to avoid trapping in the local minimum we have to limit the number of surviving structures per basin of attraction. In most cases, we do not want more than a few or even just one structure per funnel. Ability to measure the similarity between structures allows us to set an additional ''similarity'' threshold and not letting structures with distance less than this threshold survive for the next population.

Heredity operator can also be improved using fingerprints to reduce the number of poor offsprings. In most cases if we take two good parents from different funnels, their offspring will be extremely bad. This phenomenon is schematically explained in Figure 7.11. This is especially useful for calculations where we let many good structures from different funnels survive, thus increasing the probability of choosing too different parents. The simplest way to avoid this is to introduce a threshold for maximum distance between parents that are allowed to produce a child. This trick is similar to ''niching,'' [23] used in cluster structure prediction calculations (though we use more universal and nonempirical way of niching). For clusters, niching proved to be a major improvement.

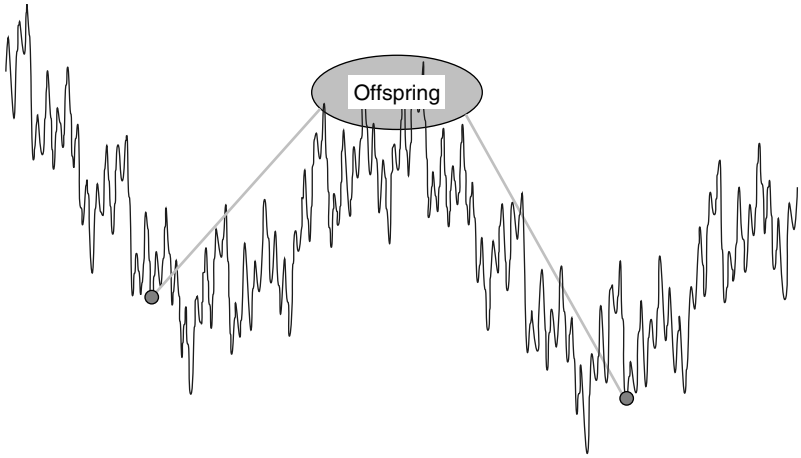

Figure 7.11 Two parents from different basins of attraction that have quite different structures usually produce a poor offspring in the high-energy areas between the funnels.

7.1.10
Extension to Molecular Crystals

In case of many organic crystals, assembling the most stable structure from single atoms will give ... a mixture of H_2O, CO_2, and perhaps other simple molecules, for a simple reason that most complex organic substances are metastable rather than thermodynamically stable. Those complex organic structures are possibly thermodynamically stable only under constraint of fixed (or partially fixed, with some conformational freedom) molecules as building blocks.

Unlike atoms, molecules are no longer point particles with spherical symmetry. Molecular handling capabilities have been implemented in USPEX [24], and are a straightforward extension of the atomic case. For variation operators molecules are now represented by a center of mass and orientation angles relative to a Cartesian coordinate frame. Offspring structures partially inherit the molecular orientations and center-of-mass positions. In addition to normal variation operators, we apply rotational mutation (whereby a randomly selected molecule within the unit cell is rotated as a whole along one or more axes). Special case must now be taken to avoid molecular overlap that may destroy the molecules upon relaxation. Plus, at least during the first stages of local optimization, the molecules should be kept fixed.

7.1.11
Adaptation to Clusters

Schönborn *et al*. [25] have translated the original version of the method [8, 26] to cluster structure prediction. The resulting algorithm was probably more effective than the earlier pioneering work by Deaven and Ho [27], and outperformed the minima hopping method in its original formulation [28]. However, with some new

developments the latter algorithm was able to show similar, and in some cases superior, performance to our evolutionary algorithm. It remains to be seen how the relative performance will change if additional ingredients (fingerprints, niching) are incorporated in the cluster prediction method. Our first results in this direction are extremely encouraging. Adaptation to Clusters.

7.1.12
Extension to Variable Compositions: Toward Simultaneous Prediction of Stoichiometry and Structure

Another major extension of the method is to enable simultaneous prediction of all stable stoichiometries and structures (in a given range of compositions). We would like to mention the pioneering study by Jóhannesson *et al.* [29], who succeeded in predicting stable stoichiometries of alloys within a given structure type. For the completely unconstrained search for both the stoichiometry and structure, a preliminary method outline was proposed in Ref. [30] and implemented in Refs. [31, 32]. Here, the basic ideas are as follows:

1) Start with a population randomly (and sparsely) sampling the whole range of compositions of interest,
2) Allow variation operators to change chemical composition (we lift chemistry-preserving constraints in the heredity operator and, in addition to the permutation operator, introduce a ''chemical transmutation'' operator),
3) Evaluate the quality of each structure not by its (free) energy, but the (free) energy per atom minus the (free) energy of the most stable isochemical mixture of already sampled compounds. This means that this fitness function depends on history of the simulation.

Such an approach seems to work (Figure 7.12), but requires further major developments. While [31] introduced a constraint that in each simulation the total number of atoms in the unit cell is fixed, our method has no such constraint, and this proves beneficial and very convenient. An example of a (very difficult) system is given in Figure 7.13. Odd as it may seem, a binary Lennard–Jones system with a 1 : 2 ratio of radii (see caption to Figure 7.13 for details of the model) exhibits a large number of ground states – including the exotic $A_{14}B$ compound and the well-known AlB_2-type structure, and several marginally unstable compositions (such as A_8B_7, $A_{12}B_{11}$, A_6B_7, A_3B_4, AB_2). The correctness of these predictions is illustrated by the fact that a fixed-composition simulation at AB_2 stoichiometry produced results (gray square in Figure 7.13a) perfectly consistent with the variable-composition runs.

Figure 7.14 shows preliminary results for the Fe–Mg system at pressures of the Earth's inner core. In agreement with a recent work ([33], who arrived at this conclusion using different methods), we find that addition of Mg stabilizes the bcc structure and many of the intermediate compositions are bcc-based alloys, even though pure Fe has an hcp ground state at this pressure.

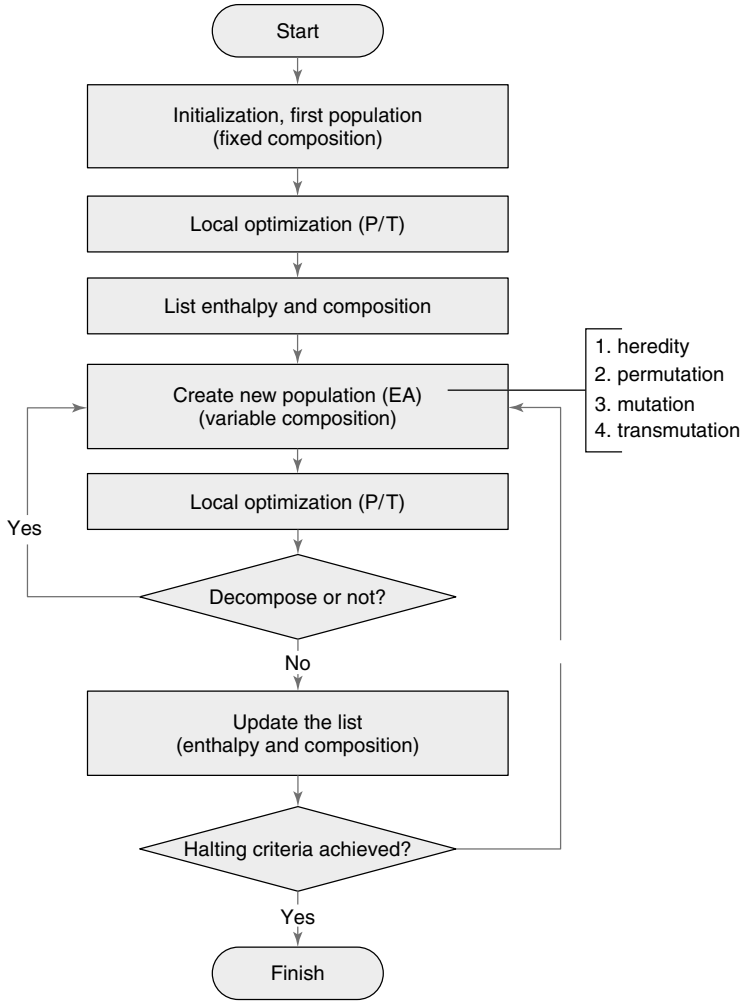

Figure 7.12 Flowchart of variable-composition USPEX. From Wang and Oganov [30].

7.2
A Few Illustrations of the Method

Any method is worth as much as its applications are, and here we review only a small selection of results obtained with USPEX, with the purpose of demonstrating its utility and with the desire to mention some interesting physics uncovered by it. All calculations described here were performed within the generalized gradient approximation (GGA; [34]) and the PAW method [35, 36], using the VASP code [12] for local optimization and total energy calculations. The predicted structures correspond to the global minima on the approximate free energy landscapes – that is, true ground states in cases where the GGA gives an adequate description of the system.

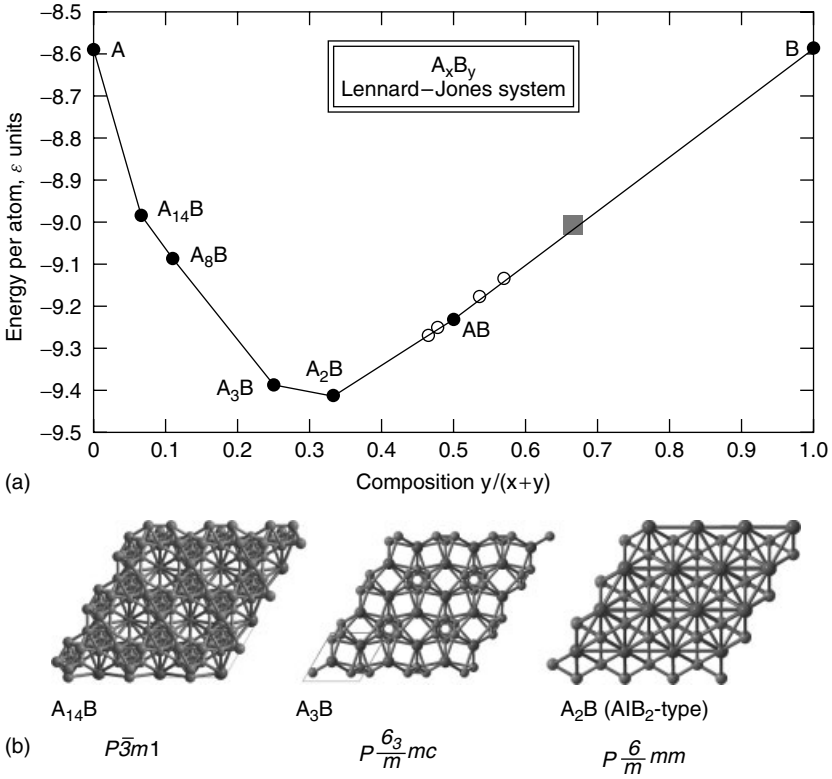

Figure 7.13 Variable-composition USPEX simulation of the A_xB_y binary Lennard–Jones system. (a) Filled circles – stable compositions, open circles – marginally unstable compositions ($A_8B_7, A_{12}B_{11}, A_6B_7, A_3B_4$). Gray square – fixed-composition result for AB_2 stoichiometry, finding a marginally unstable composition in agreement with the variable-composition results. (b) Some of the stable structures. While the ground state of the one-component Lennard–Jones crystal is hexagonal close packed (hcp) structure, ground states of the binary Lennard–Jones system are rather complex (e.g., $A_{14}B$). Note that the structure found for A_2B belongs to the well-known AlB_2 structure type. The potential is of the Lennard–Jones form for each atomic *ij*-pair: $U_{ij} = \varepsilon_{ij}\left[\left(\frac{R_{\min,ij}}{R}\right)^{12} - 2\left(\frac{R_{\min,ij}}{R}\right)^{6}\right]$, where $R_{\min,ij}$ is the distance at which the potential reaches minimum, and ε is the depth of the minimum. In these simulations, we use additive atomic dimensions: $R_{\min,BB} = 1.5R_{\min,AB} = 2R_{\min,AA}$ and nonadditive energies (to favor compound formation): $\varepsilon_{AB} = 1.25\varepsilon_{AA} = 1.25\varepsilon_{BB}$. From Oganov *et al.* [72].

7.2.1 Elements

7.2.1.1 Boron: Novel Phase with a Partially Ionic Character

Boron is perhaps the most enigmatic element: at least 16 phases were reported in the literature, but most are believed or suspected to be compounds (rather than forms of the pure element), and until recently the phase diagram was unknown.

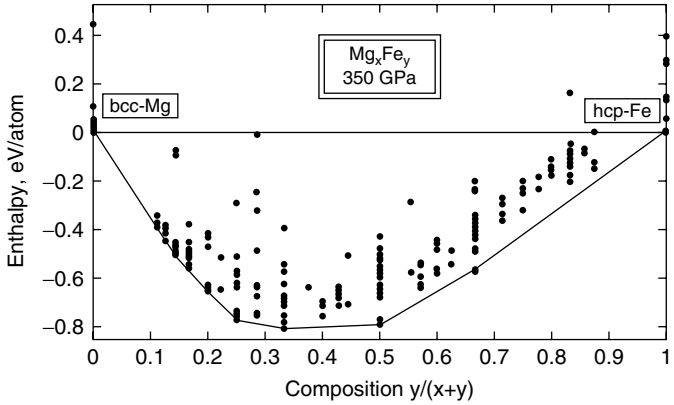

Figure 7.14 Mg–Fe system at 350 GPa. Circles indicate structures (their compositions and enthalpies) sampled, relative to the isochemical mixture of hcp-Fe and bcc-Mg. Clearly, Fe and Mg form very stable alloys (many of which have bcc-type structures) at this pressure. This variable-composition calculation was performed at the GGA level of theory [34].

Following experimental findings of J. Chen and V.L. Solozhenko (both arrived independently at the same conclusions in 2004) of a new phase at pressures above 10 GPa and temperatures of 1800–2400 K, the structure could not be determined from experimental data alone. We found the structure by using USPEX. We named this phase γ-B_{28} (because it contains 28 atoms/cell). Its structure (Figure 7.15) consists of icosahedral B_{12} clusters and B_2 pairs in an NaCl-type arrangement and exhibits sizable charge transfer from B_2 pairs to B_{12} clusters (Figure 7.16), quite unexpected for a pure element (see [19], for details).

γ-B_{28} can be represented as a "boron boride" $(B_2)^{\delta+}(B_{12})^{\delta-}$. While the exact value of the charge transfer δ depends on the definition of an atomic charge, all

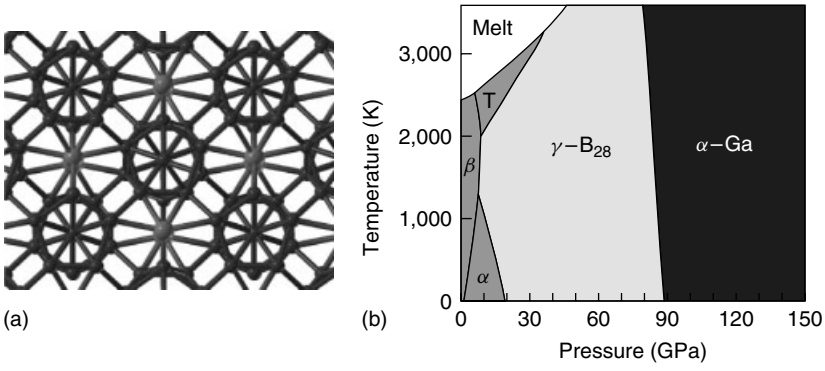

Figure 7.15 Boron: (a) structure of γ-B_{28} (B_{12} icosahedra and B_2 pairs are marked by different shades). (b) Phase diagram of boron, showing a wide stability field of γ-B_{28}. From Oganov *et al*. [19].

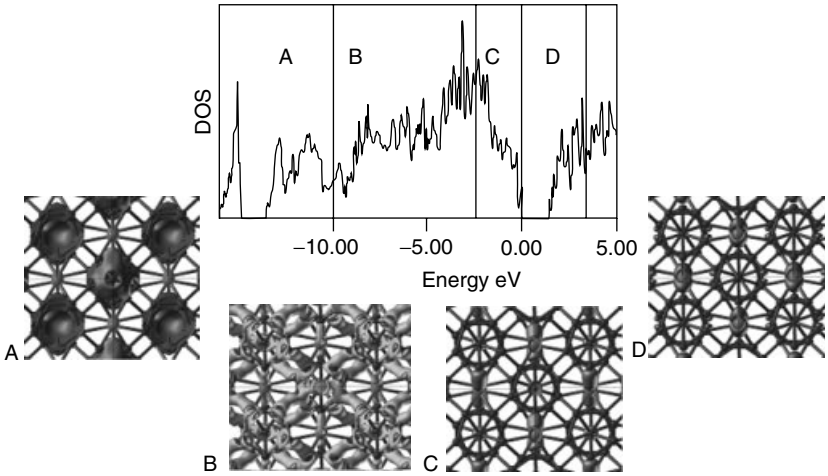

Figure 7.16 γ-B_{28}: total electronic DOS and energy-decomposed electron densities. Lowest-energy valence electrons are dominated by the B_{12} icosahedra, while top of the valence band and bottom of the conduction band (i.e., holes) are localized on the B_2 pairs. This is consistent with atom-projected DOSs and the idea of charge transfer $B_2 \rightarrow B_{12}$. From Oganov *et al*. [73].

definitions give the same qualitative picture. Our preferred definition, due to Bader [37], gives $\delta \sim 0.5$ [19]. Based on the similarity of synthesis conditions and many diffraction peaks, the same high-pressure boron phase may have been observed by Wentorf [38], though Wentorf's material was generally not believed to be pure boron (due to the sensitivity of boron to impurities and lack of chemical analysis or structure determination in his work) and its diffraction pattern was deleted from Powder Diffraction File database. γ-B_{28} is structurally related to several compounds – for instance, B_6P or $B_{13}C_2$, where the two sublattices are occupied by different chemical species (instead of interstitial B_2 pairs there are P atoms or C–B–C groups, respectively). Significant charge transfer can be found in other elemental solids, and observations of dielectric dispersion [39], implying LO–TO splitting, suggest it for β-B_{106}. The nature of the effect is possibly similar to γ-B_{28}. Detailed microscopic understanding of charge transfer in β-B_{106} would require detailed knowledge of its structure, and reliable structural models of β-B_{106} finally begin to emerge from computational studies [40–42]. γ-B_{28} is a superhard phase, with a measured Vickers hardness of 50 GPa [43], which puts it among the half a dozen hardest materials known to date.

7.2.1.2 **Sodium: A Metal that Goes Transparent under Pressure**

Sodium, a simple s-element at normal conditions, behaves in highly nontrivial ways under pressure. The discovery of an incommensurate host–guest structure [44], followed by the finding of several other complex phases [45] just below the pronounced minimum of sodium's melting curve, and the very existence of that extremely deep minimum in the melting curve at about 110 GPa [46] – all points

to some unusual changes in the physics of sodium. Later, it was also shown that the incommensurate host–guest structure is a 1D metal [47], where conductivity is mainly within the guest atom chains.

Yet another unusual phenomenon was predicted using USPEX and later (but within the same paper – [48]) verified experimentally: on further compression sodium becomes a wide-gap insulator! This happens at ~190 GPa, and Figure 7.17 shows the crystal structure of the insulating "hP4" phase, its enthalpy relative to other structures, and the electronic structure. The structure contains two inequivalent Na positions, both six-coordinated: Na1 and Na2, which have the octahedral and trigonal-prismatic coordination, and the hP4 structure can be described as the elemental analog of the NiAs structure type. Calculations suggest that sodium is no longer an s-element; instead, the outermost valence electron has significant s-, p-, and d-characters (Figure 7.17c). In other words, sodium can be considered as a transition metal because of its significant d-character.

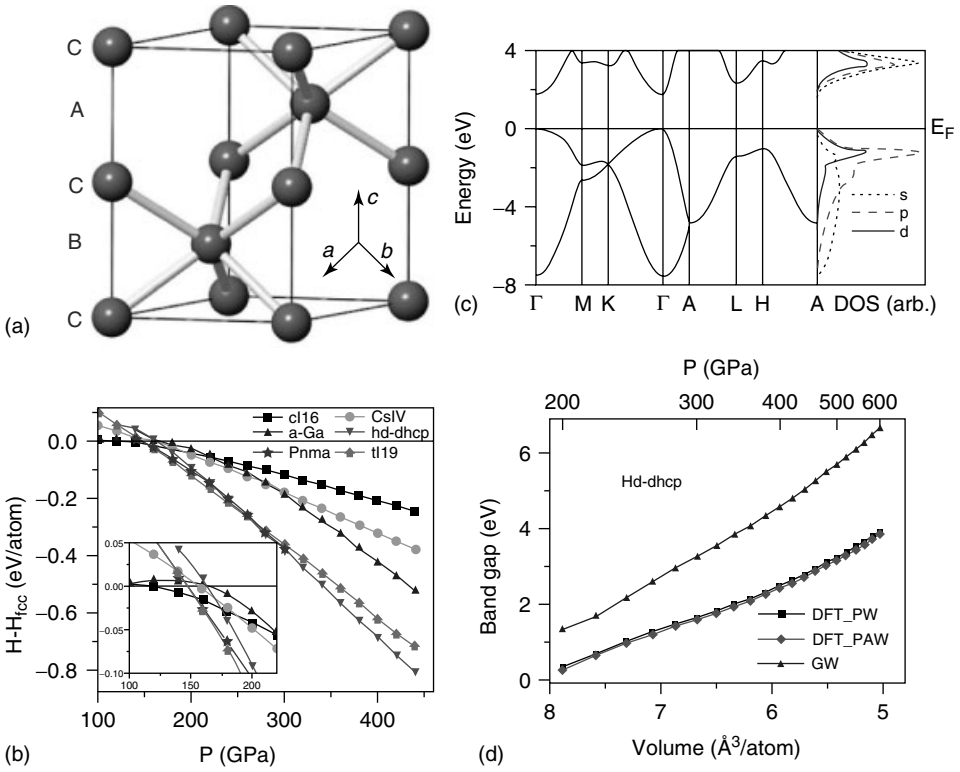

Figure 7.17 Transparent hP4 phase of sodium: (a) its crystal structure, (b) enthalpies of competing high-pressure phases (relative to the fcc structure), (c) band structure, and (d) pressure dependence of the band gap, indicating rapid increase in the band gap on compression. From Ma *et al.* [48].

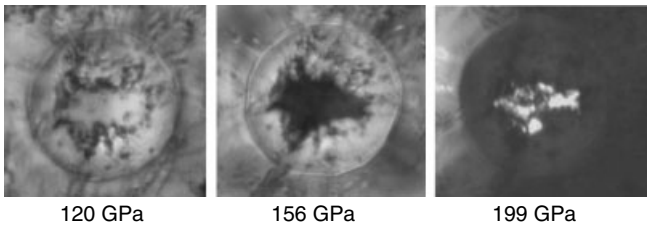

Figure 7.18 Photographs of sodium samples under pressure. At 120 GPa, the sample is metallic and highly reflective, at 156 GPa the reflectivity is very low, and at 199 GPa the sample is transparent. From Ma *et al*. [48]. (Please find a color version of this figure on the color plates.)

The band gap is direct and increases with pressure. At 200 GPa, the band gap calculated with the GW approximation is 1.3 eV, and increases to 6.5 eV at 600 GPa. These predictions implied that above 200 GPa, sodium will be red and transparent, and at ∼300 GPa it will become colorless and transparent! This has indeed been confirmed in experiment [48], see Figure 7.18. The insulating behavior is explained by the extreme localization of the valence electrons in the interstices of the structure, that is, the ''empty'' space (Figure 7.19). These areas of localization are characterized by surprisingly high values of the electron localization function (nearly 1.0) and maxima of the total electron density. The number of such maxima is half the number of sodium atoms, and therefore in a simple model we can consider Na atoms as completely ionized (Na^+), and interstitial maxima as containing one electron pair. The hP4 structure can also be described as a Ni_2In-type structure, where Na atoms occupy positions of Ni atoms, and interstitial electron pairs in

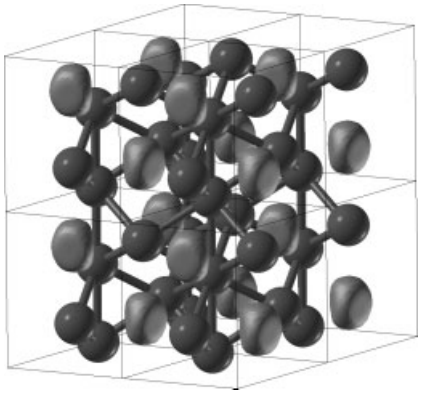

Figure 7.19 Crystal structure and electron localization function (isosurface contour 0.90) of the hP4 phase of sodium at 400 GPa. Interstitial electron localization is clearly seen. (Please find a color version of this figure on the color plates.)

hP4-Na sit on the same positions as In atoms in Ni_2In. At first counterintuitively, the degree of localization of the interstitial electron pairs increases with pressure, explaining the increase in the band gap (Figure 7.17d). hP4-Na can be described as an electride, that is, an ionic ''compound'' formed by ionic cores and localized interstitial electron pairs. The very fact that sodium, one of the best and simplest metals, under pressure becomes a transparent insulator with localized valence electrons is remarkable and forces one to reconsider classical ideas of chemistry.

Interstitial charge localization can be described in terms of (s)-p-d orbital hybridizations, and its origins are in the exclusionary effect of the ionic cores on valence electrons: valence electrons, feeling repulsion from the core electrons, are forced into the interstitial regions at pressures where atomic cores begin to overlap [50].

7.2.1.3 **Superconducting ξ-Oxygen**

Oxygen shows many unusual features under pressure. The metallic (superconducting at very low temperatures, [51]) ζ-phase, stable above 96 GPa, was discovered in 1995 [52], and its structure remained controversial for a long time. Neutron diffraction showed [53] that already in the ε-phase (at 8 GPa) there is no long-range magnetic order and likely even no local moments. The disappearance of magnetism is a consequence of increasing overlap of molecular orbitals with increasing pressure. Ultimately, orbital overlap leads to metallization.

Evolutionary simulations at 130 and 250 GPa uncovered two interesting structures with $C2/m$ and $C2/c$ space groups [54]. These have very similar enthalpies; the $C2/m$ structure is slightly lower in enthalpy and matches experimental X-ray diffraction and Raman spectroscopy data very well, better than the $C2/c$ structure [54]. Recently, a single-crystal X-ray diffraction study [55] confirmed our predicted $C2/m$ structure of ζ-oxygen. This structure is isosymmetric with the lower pressure ε-phase; both structures are shown in Figure 7.20.

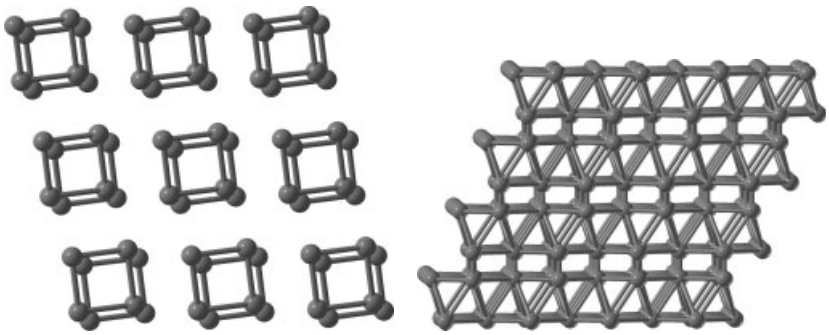

Figure 7.20 High-pressure structures of oxygen: (a) experimentally found ε-O_8 structure at 17.5 GPa [56]. (b) Predicted and confirmed $C2/m$ structure of the ζ-phase at 130 GPa [54]. Contacts up to 2.2 Å are shown as bonds.

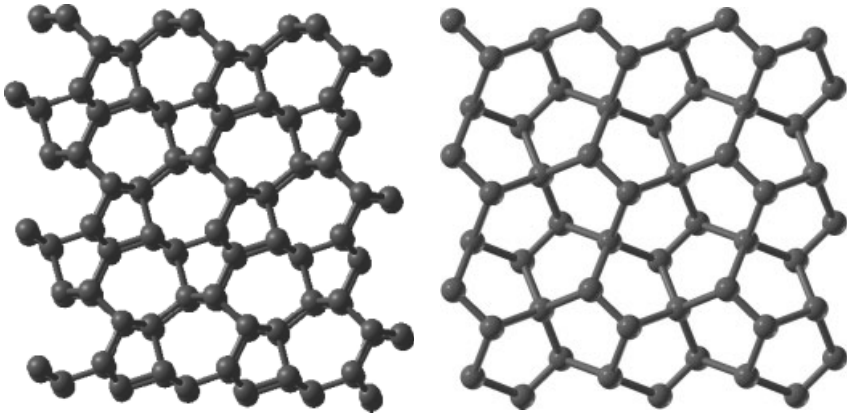

Figure 7.21 Structures of carbon: (a) novel metastable form with a 3D-framework and (b) hypothetical 2D structure, possible at interfaces. From Oganov *et al*. [9] and Oganov and Glass [26].

7.2.1.4 **Briefly on Some of the (Many) Interesting Carbon Structures**

While thermodynamically stable phases of carbon are few and simple (but still far from boring!), the metastable ones are almost inexhaustible and worth exploring both computationally and experimentally. USPEX is an attractive method for predicting low-energy metastable structures (i.e., not only the global minimum, but also the nearby local ones) because of the construction of the method, focusing to an increasing degree on low-energy parts of the energy landscape as the simulation progresses.

A number of interesting metastable carbon structures derived by USPEX was mentioned in Ref. [26], and here we just wish to mention a few novel twists in the story. One of the low-energy structures found in Ref. [26] was the three-dimensional "5 + 7" structure shown in Figure 7.21a. Later, it was found [57] that this structure matches the observed properties of the so-called superhard graphite, a new phase of carbon obtained by cold compression of graphite to pressures above 15 GPa, that is, beyond the stability field of graphite [58]. The "5 + 7" structure contains corrugated hexagonal layers reminiscent of the graphene layers; its predicted density and hardness are similar to diamond's, and its diffraction pattern closely agrees with experimental one for "superhard graphite" of Ref. [58].

Another set of interesting carbon structures was found [9] in 2D-space (hence, such structures might appear on surfaces and in interfaces) – looking for possible stable compounds in the Xe–C system at high pressures, we consistently found layered structures with Xe and C atoms segregated from each other, indicative of the tendency to phase separation. While the Xe layers had the expected close-packed configuration, the layers of carbon atoms had very creative arrangements made of three- and four-connected carbon atoms. One of such layers is shown in Figure 7.21b. While no stable Xe–C compounds were found at pressures of up to

200 GPa, we did observe in the segregated structures a considerable Xe–C bonding. It is this bonding that we believe to be responsible for the large reconstructions in the carbon layers. Such layered carbon structures should have interesting and highly tunable (by means of modifying their interaction with the substrate, in this case Xe layers) properties, and it must be possible to prepare them under special conditions.

7.2.2 Compounds and Minerals

7.2.2.1 Insulators by Metal Alloying?

Based on the electronic ''jellium'' model, the existence of a very interesting class of *insulating* materials (e.g., $Al_{13}K$ and $Al_{12}C$), based on icosahedral aluminum clusters, has been hypothesized (e.g., in Ref. [59]). Recent work done with USPEX [20] has refuted this suggestion. For instance, for $Al_{12}C$ a much more stable structure exists – fcc structure with stacking faults, where carbon atoms occupy octahedrally coordinated sites (Figure 7.22). This also suggests that addition of carbon to aluminum may promote the formation of stacking faults, a conclusion of some interest to metallurgy.

7.2.2.2 MgB_2: Analogy with Carbon and Loss of Superconductivity under Pressure

MgB_2 is a superconductor with a very high $T_c = 39$ K [60], highest among the experimentally studied conventional superconductors. In its structure, Mg atoms are sandwiched between graphene-like layers of boron atoms. Using USPEX, [49] found that at high pressure (190 GPa) MgB_2 adopts another structure (Figure 7.23) with B atoms forming a distorted hexagonal diamond structure. In both phases, holes are the main charge carriers, but the high-pressure phase is a much poorer

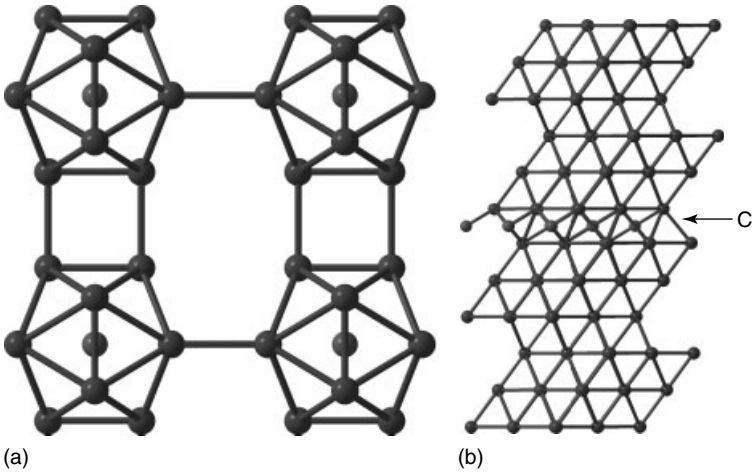

Figure 7.22 Structures of $Al_{12}C$ with 13 atoms/cell: (a) with icosahedral clusters and (b) more stable stacking-fault structure. From Oganov and Glass [20].

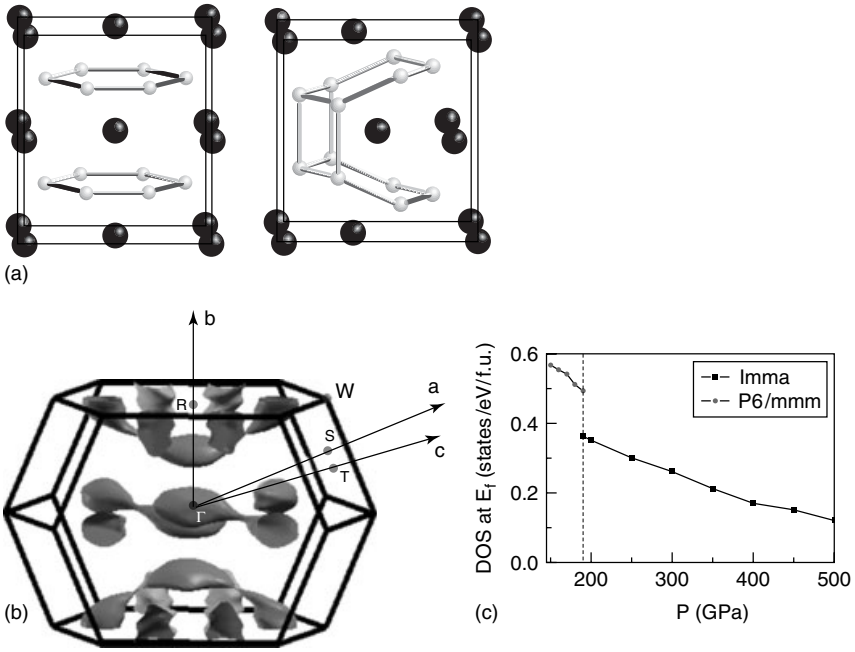

Figure 7.23 New phase of MgB_2: (a) its structure (right) and relationship to the structure known at 1 atm (left), (b) its Fermi surface, and (c) electronic density of states at the Fermi level as a function of pressure. From Ma *et al.* [49]. (Please find a color version of this figure on the color plates.)

metal and is not a superconductor. The structural trend parallels that of carbon: graphite is a much better conductor than diamond.

7.2.2.3 Hydrogen-Rich Hydrides under Pressure, and Their Superconductivity

Recently, three studies were published, where USPEX was used to find high-pressure metallic ground states of hydrogen-rich compounds – GeH_4 [61], SiH_4 [62], and LiH_2, LiH_6, and LiH_8 [63]. In line with the recent proposals of N.W. Ashcroft, we indeed found that hydrides adopt superconducting states more readily than pure hydrogen does.

Already at 100 GPa, LiH_n ($n > 1$) metallic compounds attain thermodynamic stability. Their structures (Figure 7.24) contain the ''semimolecular'' H_2 units with bonds slightly longer than 0.74 Å of the free H_2 molecule. Zurek *et al.* [63] have explained that it is charge transfer from Li to the unoccupied (antibonding) levels of the H_2 molecules that simultaneously creates the metallic character and weakens bonding within the H_2 units. We also note that the very existence (and stability!) of such compounds violate the traditional chemical valency concepts. LiH_2, by the way, has a structure closely related to the host–guest structures known for Ca, Sr, and Ba (H_2units occupy the guest sublattice, with Li + H being the host).

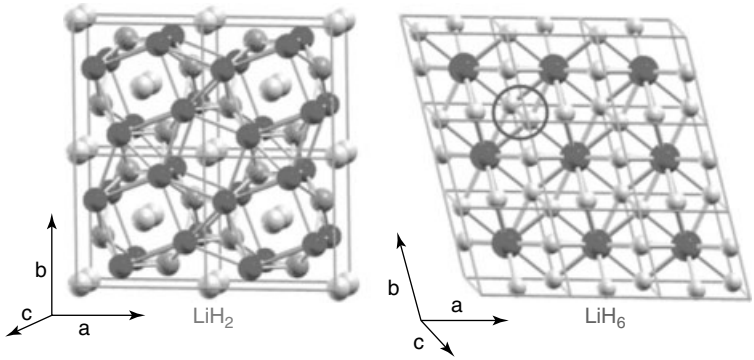

Figure 7.24 Structures of LiH_2 (a) and LiH_6 (b) predicted to be stable at pressures >100 GPa Li atoms are green, "lone" hydrogen atoms are pink, and those in the H_2 units are white. (from [63]). (Please find a color version of this figure on the color plates.)

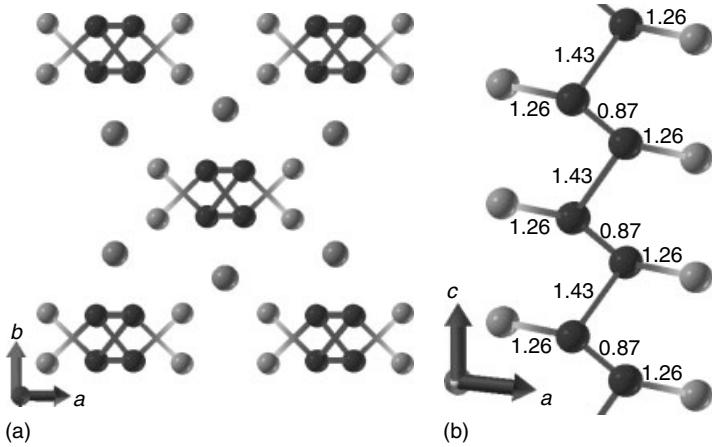

Figure 7.25 Crystal structure of the $C2/c$ phase of GeH_4. (a) Two types of hydrogen atoms are shown by different color (light – "lone" atoms, and dark – forming H_2 semi-molecular units). Panel (b) shows the connectivity of hydrogens in the structure and distances (in Å). (Please find a color version of this figure on the color plates.)

An interesting prediction was made for GeH_4 that it will become thermodynamically stable above 196 GPa [61] and that the stable $C2/c$ should have a very high superconducting T_c (64 K at the pressure of 220 GPa). The structure of this phase (Figure 7.25a) contains a branched chain of H atoms with alternating bond lengths shown in Figure 7.25b.

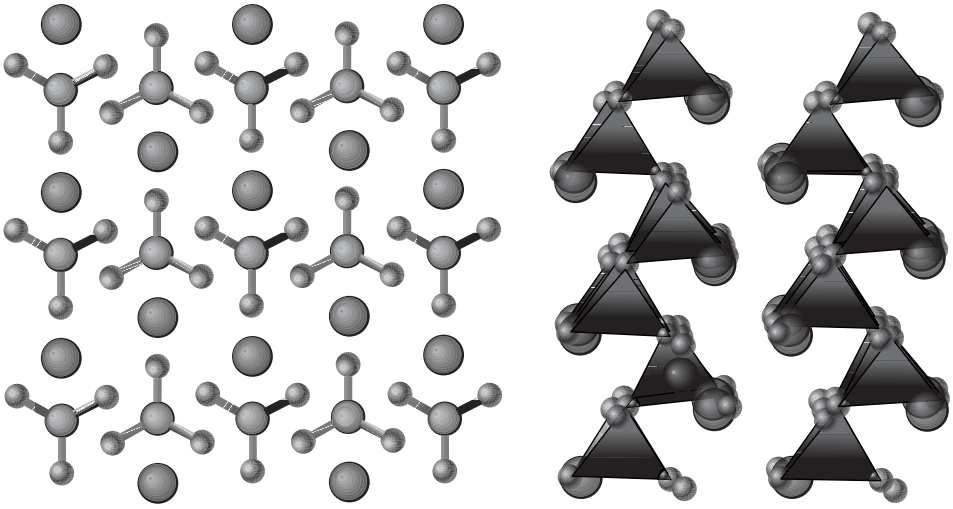

Figure 7.26 $CaCO_3$ at high pressure. (a) Structure of postaragonite phase and (b) $C222_1$ phase. From [65].

7.2.2.4 High-Pressure Polymorphs of $CaCO_3$

High-pressure behavior of carbonates is very important for understanding the Earth's carbon cycle, but remained controversial until recently. Even for CO_2, a chemically simpler compound, major controversies still exist (see [20], for a discussion of results on CO_2). For $CaCO_3$, there is a well-known transition from calcite to aragonite at ~2 GPa, followed by a transition to a postaragonite phase at ~40 GPa [64], the structure of which (Figure 7.26a) was solved [65] using USPEX, and the predicted structure matched the experimental X-ray diffraction pattern well. Now postaragonite structure is also known in $SrCO_3$ and $BaCO_3$ [64, 66].

Furthermore, we have predicted [65] that above 137 GPa a new phase, with space group $C222_1$ and containing chains of carbonate tetrahedral (Figure 7.26b), becomes stable. Recently this prediction was verified by experiments [67] at pressures above 130 GPa. We note that both postaragonite and the $C222_1$ structure (Figure 7.26b) belong to new structure types and could not have been found by analogy with any known structures.

The presence of tetrahedral carbonate ions at very high pressures invites an analogy with silicates, but the analogy is limited. In silicates, the intertetrahedral angle Si–O–Si is extremely flexible [68], which is one of the reasons for the enormous diversity of silicate structure types. Figure 7.27 shows the variation of the energy as a function of the Si–O–Si angle in the model $H_6Si_2O_7$ molecule – method borrowed from [68]. One can see only a shallow minimum at $\angle(\mathrm{Si-O-Si}) = 135^\circ$, but a deep minimum at $\angle(\mathrm{C-O-C}) = 124^\circ$ with steep energy variations for $H_6C_2O_6$ (Figure 7.27). This suggests a much more limited structural variety/flexibility of metacarbonates, compared to silicates. In tetrahedrally coordinated forms of both $CaCO_3$ ($C222_1$ structure) and CO_2 ($\beta-$ cristobalite structure, found to be stable

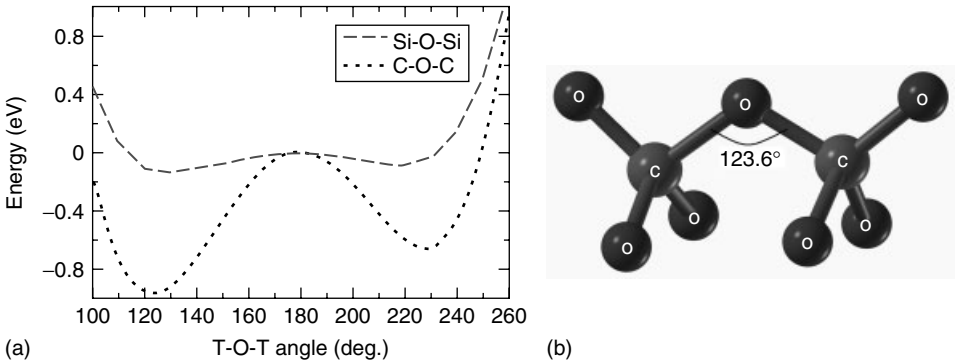

Figure 7.27 (a) Energy variation as a function of the T–O–T angle (red dashed line – T = Si, black dotted line – T = C). Calculations were performed on $H_6T_2O_7$ molecules; at each angle all T–O distances and O–T–O valence angles were optimized. Optimum angle C–O–C = 124°, Si–O–Si = 135°. (b) These calculations were performed with SIESTA code [13] using the GGA functional [34], norm-conserving pseudopotentials and a double-ζ basis set with a single polarization function for each atom. From [69].

above 19 GPa – [69–71]), the $\angle(C-O-C)$ angles are close to 124° in a wide pressure range.

7.3 Conclusions

Evolutionary algorithms, based on physically motivated forms of variation operators and local optimization, are a powerful tool enabling reliable and efficient prediction of stable crystal structures. This method has a wide field of applications in computational materials design (where experiments are time-consuming and expensive) and in studies of matter at extreme conditions (where experiments are very difficult or sometimes beyond the limits of feasibility).

One of the current limitations is the accuracy of today's *ab initio* simulations; this is particularly critical for strongly correlated and for van der Waals systems. Note, however, that the method itself does not make any assumptions about the way energies are calculated and can be used in conjunction with any method, that is able to provide total energies. Most of practical calculations are done at $T = 0$ K, but temperature can be included as long as the free energy can be calculated efficiently. Difficult cases are aperiodic and disordered systems (for which only the lowest energy periodic approximants and ordered structures can be predicted at this moment).

We are suggesting USPEX as the method of choice for crystal structure prediction of systems with up to ~50 atoms/cell, where no information (or just the lattice parameters) is available. Above 50–100 atoms/cell runs become expensive due to the "curse of dimensionality" (although still feasible), eventually necessitating the

use of other ideas within USPEX or another approach. There is, however, hope of enabling structure prediction for a very large (>200 atoms/cell) systems.

USPEX has been applied to many important problems. One expects many more applications to follow, both in high-pressure research and in materials design.

Acknowledgments

ARO and AOL thank Intel Corporation, DARPA (grant 54751), and Research Foundation of Stony Brook University for financial support. We thank the Joint Supercomputer Center (Russian Academy of Sciences, Moscow) for providing supercomputer time. USPEX code is available on request from ARO.

References

1. Gavezzotti, A. (1994) Are crystal structure predictable? *Acc. Chem. Res.*, **27**, 309–314.
2. Maddox, J. (1988) Crystals from first principles. *Nature*, **335**, 201.
3. Tsai, H.K., Yang, J.M., Tsai, Y.F., and Kao, C.Y. (2004) An evolutionary algorithm for large traveling salesman problems. *IEEE Trans. Syst. Man. Cybern. B Cybern.*, **34** (4), 1718–1729.
4. Bush, T.S., Catlow, C.R.A., and Battle, P.D. (1995) Evolutionary programming techniques for predicting inorganic crystal structures. *J. Mater. Chem.*, **5**, 1269–1272.
5. Woodley, S.M., Battle, P.D., Gale, J.D., and Catlow, C.R.A. (1999) The prediction of inorganic crystal structures using a genetic algorithm and energy minimization. *Phys. Chem. Chem. Phys.*, **1**, 2535–2542.
6. Woodley, S.M. (2004) Prediction of crystal structures using evolutionary algorithms and related techniques. *Struct. Bond.*, **110**, 95–132.
7. Bazterra, V.E., Ferraro, M.B., and Facelli, J.C. (2002) Modified genetic algorithm to model crystal structures. I. Benzene, naphthalene and anthracene. *J. Chem. Phys.*, **116**, 5984–5991.
8. Glass, C.W., Oganov, A.R., and Hansen, N. (2006) USPEX – evolutionary crystal structure prediction. *Comput. Phys. Commun.*, **175**, 713–720.
9. Oganov, A.R., Ma, Y., Glass, C.W., and Valle, M. (2007) Evolutionary crystal structure prediction: overview of the USPEX method and some of its applications. *Psi-k Newsl.*, **84**, 142–171.
10. Valle, M. (2005) STM3: a chemistry visualization platform. *Z. Krist.*, **220**, 585–588.
11. Valle, M. and Oganov, A.R. (2010) Crystal fingerprints space, A novel paradigm to study crystal structures sets (in press). Acta Cryst. A.
12. Kresse, G. and Furthmüller, J. (1996) Efficient iterative schemes for *ab initio* total-energy calculations using a plane wave basis set. *Phys. Rev. B*, **54**, 11169–11186.
13. Soler, J.M., Artacho, E., Gale, J.D., Garcia, A., Junquera, J., Ordejon, P., and Sanchez-Portal, D. (2002) The SIESTA method for *ab initio* order-N materials simulation. *J. Phys.: Condens. Matter*, **14**, 2745–2779.
14. Gale, J.D. (2005) GULP: capabilities and prospects. *Z. Krist.*, **220**, 552–554.
15. Lyakhov, A.O., Oganov, A.R., and Valle, M. (2010) How to predict very large and complex crystal structures. *Comput. Phys. Commun.*, **181**, 1623–1632.
16. Lloyd, L.D. and Johnston, R.L. (1998) Modelling aluminum clusters with an empirical many-body potential. *Chem. Phys.*, **236**, 107.

17. Pickard, C.J. and Needs, R.J. (2006) High-pressure phases of silane. *Phys. Rev. Lett.*, **97**, 045504.
18. Wales, D.J. and Doye, J.P.K. (1997) Global optimization by Basin–Hopping and the lowest energy structures of Lennard–Jones clusters containing up to 110 atoms. *J. Phys. Chem. A*, **101**, 5111.
19. Oganov, A.R., Chen, J., Gatti, C., Ma, Y.-Z., Ma, Y.-M., Glass, C.W., Liu, Z., Yu, T., Kurakevych, O.O., and Solozhenko, V.L. (2009) Ionic high-pressure form of elemental boron. *Nature*, **457**, 863–867.
20. Oganov, A.R. and Glass, C.W. (2008) Evolutionary crystal structure prediction as a tool in materials design. *J. Phys.: Cond. Mattter*, **20**, 064210.
21. Valle, M. and Oganov, A.R. (2008a) Crystal structure classifier for an evolutionary algorithm structure predictor. Proceedings of the IEEE Symposium on Visual Analytics Science and Technology, October 21–23, 2008, Columbus, pp. 11–18.
22. Oganov, A.R. and Valle, M. (2009) How to quantify energy landscapes of solids. *J. Chem. Phys.*, **130**, 104504.
23. Hartke, B. (1999) Global cluster geometry optimization by a phenotype algorithm with Niches: location of elusive minima, and low-order scaling with cluster size. *J. Comput. Chem.*, **20**, 1752.
24. Glass, C.W. (2008) Computational crystal structure prediction. PhD thesis, ETH Zurich.
25. Schönborn, S., Goedecker, S., Roy, S., and Oganov, A.R. (2009) The performance of minima hopping and evolutionary algorithms for cluster structure prediction. *J. Chem. Phys.*, **130**, 144108.
26. Oganov, A.R. and Glass, C.W. (2006) Crystal structure prediction using *ab initio* evolutionary techniques: principles and applications. *J. Chem. Phys.*, **124**, 244704.
27. Deaven, D.M. and Ho, K.M. (1995) Molecular geometry optimization with a genetic algorithm. *Phys. Rev. Lett.*, **75**, 288–291.
28. Gödecker, S. (2004) Minima hopping: An efficient search method for the global minimum of the potential energy surface of complex molecular systems. *J. Chem. Phys.*, **120**, 9911–9917.
29. Jóhannesson, G.H., Bligaard, T., Ruban, A.V., Skriver, H.L., Jacobsen, K.W., and Nørskov, J.K. (2002) Combined electronic structure and evolutionary search approach to materials design. *Phys. Rev. Lett.*, **88**, 255506.
30. Wang, Y. and Oganov, A.R. (2008) Research on the evolutionary prediction of very complex crystal structures. IEEE Computational Intelligence Society Walter Karplus. Summer Research Grant, Final Report. (*ieee-cis.org/_files/EAC_Research_2008_Report_WangYanchao.October 2008*).
31. Trimarchi, G., Freeman, A.J., and Zunger, A. (2009) Predicting stable stoichiometries of compounds via evolutionary global space-group optimization. *Phys. Rev. B*, **80**, 092101.
32. Lyakhov, A.O., and Oganov, A.R. (2010) Simultaneous prediction of chemical formula and crystal structure using an evolutionary algorithm (in press).
33. Kadas, K., Vitos, L., Johansson, B., and Ahuja, R. (2009) Stability of body-centered cubic iron-magnesium alloys in the Earth's inner core. *Proc. Natl. Acad. Sci.*, **106**, 15560–15562.
34. Perdew, J.P., Burke, K., and Ernzerhof, M. (1996) Generalized gradient approximation made simple. *Phys. Rev. Lett.*, **77**, 3865–3868.
35. Blöchl, P.E. (1994) Projector augmented-wave method. *Phys. Rev. B*, **50**, 17953–17979.
36. Kresse, G. and Joubert, D. (1999) From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B*, **59**, 1758–1775.
37. Bader, R. (1990) *Atoms in Molecule: A Quantum Theory*, Oxford University Press, Oxford.
38. Wentorf, R.H. (1965) Boron: another form. *Science*, **147**, 49–50.
39. Tsagareishvili, O.A., Chkhartishvili, L.S., and Gabunia, D.L. (2009) Apparent low-frequency charge capacitance of semiconducting boron. *Semiconductors*, **43**, 14–20.
40. van Setten, M.J., Uijttewaal, M.A., de Wijs, G.A., and de Groot, R.A. (2007) Thermodynamic stability of boron: The

role of defects and zero point motion. *J. Am. Chem. Soc.*, **129**, 2458–2465.
41. Widom, M. and Mikhalkovic, M. (2008) Symmetry-broken crystal structure of elemental boron at low temperature. *Phys. Rev. B*, **77**, 064113.
42. Ogitsu, T., Gygi, F., Reed, J., Motome, Y., Schwegler, E., and Galli, G. (2009) Imperfect crystal and unusual semiconductor: boron, a frustrated element. *J. Am. Chem. Soc.*, **131**, 1903–1909.
43. Solozhenko, V.L., Kurakevych, O.O., and Oganov, A.R. (2008) On the hardness of a new boron phase, orthorhombic$\gamma - B_{28}$. *J. Superhard Mater.*, **30**, 428–429.
44. Hanfland, M., Syassen, K., Loa, I., Christensen, N.E., and Novikov, D.L. (2002) Na at megabar pressures. Poster at 2002 High Pressure Gordon Conference.
45. Gregoryanz, E., Lundegaard, L.F., McMahon, M.I., Guillaume, C., Nelmes, R.J., and Mezouar, M. (2008) Structural diversity of sodium. *Science*, **320**, 1054–1057.
46. Gregoryanz, E., Degtyareva, O., Somayazulu, M., Hemley, R.J., and Mao, H.K. (2005) Melting of dense sodium. *Phys. Rev. Lett.*, **94**, 185502.
47. Lazicki, A., Goncharov, A.F., Struzhkin, V.V., Cohen, R.E., Liu, Z., Gregoryanz, E., Guillaume, C., Mao, H.K., and Hemley, R.J. (2009) Anomalous optical and electronic properties of dense sodium. *Proc. Natl. Acad. Sci.*, **106**, 6525–6528.
48. Ma, Y., Eremets, M.I., Oganov, A.R., Xie, Y., Trojan, I., Medvedev, S., Lyakhov, A.O., Valle, M., and Prakapenka, V. (2009a) Transparent dense sodium. *Nature*, **458**, 182–185.
49. Ma, Y., Wang, Y., and Oganov, A.R. (2009b) Absence of superconductivity in the novel high-pressure polymorph of MgB_2. *Phys. Rev. B*, **79**, 054101.
50. Neaton, J.B. and Ashcroft, N.W. (1999) Pairing in dense lithium. *Nature*, **400**, 141–144.
51. Shimizu, K., Suhara, K., Ikumo, M., Eremets, M.I., and Amaya, K. (1998) Superconductivity in oxygen. *Nature*, **393**, 767–769.
52. Akahama, Y., Kawamura, H., Hausermann, D., Hanfland, M., and Shimomura, O. (1995) New high-pressure structural transition of oxygen at 96 GPa associated with metalization in a molecular solid. *Phys. Rev. Lett.*, **74**, 4690–4693.
53. Goncharenko, I.N. (2005) Evidence for a magnetic collapse in the epsilon phase of solid oxygen. *Phys. Rev. Lett.*, **94**, 205701.
54. Ma, Y.-M., Oganov, A.R., and Glass, C.W. (2007) Structure of the metallic ζ-phase of oxygen and isosymmetric nature of the $\varepsilon - \zeta$ phase transition: *Ab initio* simulations. *Phys. Rev. B*, **76**, 064101.
55. Weck, G., Desgreniers, S., Loubeyre, P., and Mezouar, M. (2009) Single-crystal structural characterization of the metallic phase of oxygen. *Phys. Rev. Lett.*, **102**, 255503.
56. Lundegaard, L.F., Weck, G., McMahon, M.I., Desgreniers, S., and Loubeyre, P. (2006) Observation of an O_8 molecular lattice in the epsilon phase of solid oxygen. *Nature*, **443**, 201–204.
57. Li, Q., Oganov, A.R., Wang, H., Wang, H., Xu, Y., Cui, T., Ma, Y., Mao, H.-K., and Zou, G. (2009) Superhard monoclinic polymorph of carbon. *Phys. Rev. Lett.*, **102**, 175506.
58. Mao, W.L., Mao, H.K., Eng, P.J., Trainor, T.P., Newville, M., Kao, C.C., Heinz, D.L., Shu, J., Meng, Y., and Hemley, R.J. (2003) Bonding changes in compressed superhard graphite. *Science*, **302**, 425–427.
59. Gong, X.G. (1997) Structure and stability of cluster-assembled solid $Al_{12}C(Si)$: a first-principles study. *Phys. Rev. B*, **56**, 1091–1094.
60. Nagamatsu, J., Nakagawa, N., Muranaka, T., Zenitani, Y., and Akimitsu, J. (2001) Superconductivity at 39 K in magnesium diboride. *Nature*, **410**, 63–64.
61. Gao, G., Oganov, A.R., Bergara, A., Martinez-Canalez, M., Cui, T., Iitaka, T., Ma, Y., and Zou, G. (2008) Superconducting high pressure phase of germane. *Phys. Rev. Lett.*, **101**, 107002.
62. Martinez-Canales, M., Oganov, A.R., Lyakhov, A., Ma, Y., and Bergara, A.

(2009) Novel structures of silane under pressure. *Phys. Rev. Lett.*, **102**, 087005.

63. Zurek, E., Hoffmann, R., Ashcroft, N.W., Oganov, A.R., and Lyakhov, A.O. (2009) A little bit of lithium does a lot for hydrogen. *Proc. Natl. Acad. Sci.*, **106**, 17640–17643.

64. Ono, S., Kikegawa, T., Ohishi, Y., and Tsuchiya, J. (2005) Post-aragonite phase transformation in $CaCO_3$ at 40 GPa. *Am. Mineral.*, **90**, 667–671.

65. Oganov, A.R., Glass, C.W., and Ono, S. (2006) High-pressure phases of $CaCO_3$: Crystal structure prediction and experiment. *Earth Planet. Sci. Lett.*, **241**, 95–103.

66. Ono, S. (2007) New high-pressure phases in $BaCO_3$. *Phys. Chem. Miner.*, **34**, 215–221.

67. Ono, S., Kikegawa, T., and Ohishi, Y. (2007) High-pressure phase transition of $CaCO_3$. *Am. Mineral.*, **92**, 1246–1249.

68. Lasaga, A.C. and Gibbs, G.V. (1987) Applications of quantum-mechanical potential surfaces to mineral physics calculations. *Phys. Chem. Miner.*, **14**, 107–117.

69. Oganov, A.R., Ono, S., Ma, Y., Glass, C.W., and Garcia, A. (2008) Novel high-pressure structures of $MgCO_3$, $CaCO_3$ and CO_2 and their role in the Earth's lower mantle. *Earth Planet. Sci. Lett.*, **273**, 38–47.

70. Dong, J.J., Tomfohr, J.K., Sankey, O.F., Leinenweber, K., Somayazulu, M., and McMillan, P.F. (2000) Investigation of hardness in tetrahedrally bonded nonmolecular CO_2 solids by density-functional theory. *Phys. Rev. B*, **62**, 14685–14689.

71. Holm, B., Ahuja, R., Belonoshko, A., and Johansson, B. (2000) Theoretical investigation of high pressure phases of carbon dioxide. *Phys. Rev. Lett.*, **85**, 1258–1261.

72. Oganov, A.R., Ma, Y., Lyakhov, A.O., Valle, M., and Gatti, C. (2010) Evolutionary crystal structure prediction as a method for the discovery of minerals and materials. *Rev. Mineral. Geochem.*, **71**, 271–298.

73. Oganov, A.R., and Solozhenko, V.L. (2009) Boron: a hunt for superhard polymorphs. *J. Superhard Materials*, **31**, 285–291.