

Structure and Stability Prediction of Compounds with Evolutionary Algorithms

Benjamin C. Revard, William W. Tipton, and Richard G. Hennig

Abstract Crystal structure prediction is a long-standing challenge in the physical sciences. In recent years, much practical success has been had by framing it as a global optimization problem, leveraging the existence of increasingly robust and accurate free energy calculations. This optimization problem has often been solved using evolutionary algorithms (EAs). However, many choices are possible when designing an EA for structure prediction, and innovation in the field is ongoing. We review the current state of evolutionary algorithms for crystal structure and composition prediction and discuss the details of methodological and algorithmic choices. Finally, we review the application of these algorithms to many systems of practical and fundamental scientific interest.

Keywords Structure prediction · Genetic algorithm · Crystal structure · Energy landscape · Heuristic optimization · Phase diagram

Contents

1	Introduction	182
1.1	Potential Energy Landscape	183
1.2	Evolutionary Algorithms	184
2	Details of the Method	186
2.1	Representation of Structures	186
2.2	Initial Population	188
2.3	Fitness	189
2.4	Selection	189
2.5	Promotion	191

2.6	Mating	191
2.7	Mutation	195
2.8	System Size	196
2.9	Development and Screening	197
2.10	Maintaining Diversity in the Population	198
2.11	Order Parameters	200
2.12	Frequency of Promotion and Variations	200
2.13	Convergence Criteria: Have We Found the Global Minimum?	201
3	Phase Diagram Searching	202
4	Energy Calculations and Local Relaxation	204
5	Summary of Methods	205
6	Applications	208
6.1	Elemental Solids	208
6.2	Hydrogen-Containing Compounds	210
6.3	Intermetallic Compounds	213
6.4	Minerals	214
6.5	Molecular Crystals	214
6.6	Inorganic Compounds	215
7	Conclusions	215
	References	217

1 Introduction

Many of the most crucial technological challenges today are essentially materials problems. Materials with specific properties are needed but unknown, and new materials must be found or designed. In some cases experiments can be performed to search for and characterize new materials [1], but these methods can be expensive and difficult. Thus, computational approaches can be complementary or advantageous. Theoretical prediction of many materials properties is possible once the atomic structure of a material is known, but structure prediction remains a challenge. However, a number of new methods have been proposed in recent years to address this problem [2–8]. These techniques are often faster and less expensive than experimental work, they dispense with the need to work with sometimes toxic chemicals, and they can be used to explore materials systems under conditions that are still inaccessible to experiment, such as very high pressures.

Unless kinetically constrained, materials tend to form structures that are in thermodynamic equilibrium, i.e., have the lowest Gibbs free energy. Thus, in order to predict a material's structure, we must find the arrangement of atoms that minimizes the Gibbs free energy, given by

$$G = U - TS + pV$$

Here, U is the internal energy, p the pressure, V the volume, T the temperature, and S the entropy. The entropy is comprised of three contributions: electronic, vibrational, and configurational. The vibrational and configurational components are expensive to calculate, and much of the error introduced by neglecting the

entropy vanishes when taking energy differences [9, 10]. For these reason, the entropy is often neglected, effectively constraining the search to the $T = 0$ regime. That is, the enthalpy $H = U + pV$ is frequently used to approximate the Gibbs free energy. Finite temperature effects can be included as a post-processing step once particularly promising structures have been identified. We note that, at high temperatures, anharmonic contributions to the vibrational entropy can stabilize phases that are mechanically or dynamically unstable at low temperature [11]. However, in order to search for stable materials at low temperature and fixed composition, the function we need to minimize, known as the objective function, is the enthalpy per atom.

A thermodynamic ensemble is not always used as the objective function. Bush et al. devised an objective function based on Pauling's valence rules and only performed energy calculations on the best structures identified thereby [2, 12]. Although this approach is computationally efficient, it is limited to ionic materials and is not as reliable as a direct search over the correct thermodynamic quantity.

1.1 Potential Energy Landscape

Given the atomic structure, there exist efficient methods for approximating the enthalpy. A complete description of a crystal structure includes six lattice parameters and $3N - 3$ atomic coordinates, where N is the number of atoms in the unit cell. Thus, the function we seek to minimize can be thought of as a surface in a $3N + 3$ -dimensional space. These surfaces are referred to as energy landscapes. The lowest enthalpy structure is located at the deepest, or global, minimum of the energy landscape. In this way, the physical problem of predicting a material's atomistic structure is expressed as a mathematical optimization problem. In order to understand the search for the global minimum of an energy landscape, it is helpful to examine some general properties of energy landscapes of materials, as follows:

- Much of the configuration space corresponds to structures with unphysical small interatomic distances. These areas of the configuration space can be neglected.
- The energy landscape is effectively partitioned into basins of attraction by the use of a local optimization routine. The local optimizer takes any two structures in the same basin into the local minimum located at the bottom of the basin.
- The number of local minima on the energy landscape scales exponentially with the dimensionality of the search space, i.e., with the number of atoms in the cell [13]. Venkatesh et al. calculated the number of local minima as a function of system size for clusters containing up to 14 Lennard–Jones particles, illustrating this exponential trend [14].
- Deeper basins tend to occupy larger volumes in the multidimensional space. Specifically, a power law distribution describes the relationship between the depth of a basin and its hyper-volume. Combined with our capability for local minimization, this greatly simplifies the search for the optimum structure [15].

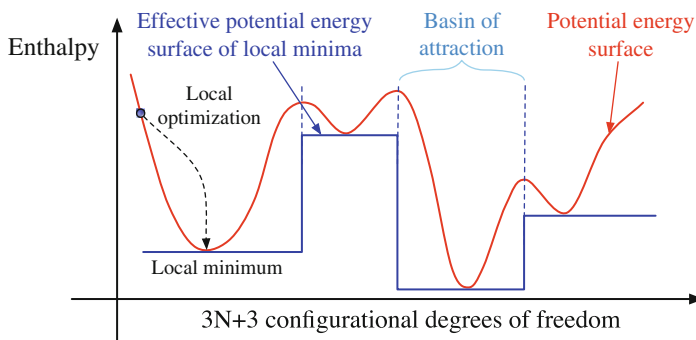


Fig. 1 Potential energy surface. The use of local optimization simplifies the search problem by dividing the continuous solution space into basins of attraction

- The barrier to reach a neighboring basin is usually low if that basin has a deeper minimum than the current basin. This is a consequence of the Bell–Evans–Polanyi principle [7].
- Low-energy minima in the landscape usually correspond to symmetrical structures [7].
- Low-lying minima are usually located near each other on the energy landscape. This tendency gives the landscape an overall structure that can be exploited while searching for the global minimum [16].

No analytical form exists for the enthalpy as a function of atomic configuration. We can only sample the enthalpy and its derivatives at discrete points on the energy landscape using methods such as density functional theory (DFT). Thus, one often resorts to heuristic search methods. One such class of methods that has proven successful is the evolutionary algorithm. This approach draws inspiration from biological evolution. Efficient local optimization utilizes the derivative information and is very beneficial in the solution of the optimization problem (see Sect. 4). Figure 1 illustrates how local optimization transforms the continuous potential energy landscape into a discrete set of basins of attractions, which dramatically simplifies the search space.

1.2 Evolutionary Algorithms

In nature, genetic information is carried in organisms. It is maintained in a population’s gene pool if it is passed on from parents to offspring. New information can be introduced through mutation events, but these are rare (and usually lethal). The success that an organism has in passing on its genes is called the organism’s fitness.

The fitness of an organism is not universal but depends on its environment. Many species which are very successful in their native habitats would do poorly in other

environments. More subtly, there is variance of traits within a single species. In some cases these differences can lead to a difference in the organisms' fitness. The genes of low-fitness individuals are less likely to be passed on, so traits of the high-fitness individuals are likely to be more common in subsequent generations. In this way, populations (but not individuals) evolve to be well suited to their environment. This assumes, of course, that relevant traits are passed on, to varying degrees, from parents to offspring. The correlation between a trait in a parent and that in an offspring is known as the heritability of a trait. In order for environmental pressure to cause quick evolution of a trait, that trait must have high heritability.

Evolutionary algorithms leverage the power of this process to "evolve" solutions to optimization problems. Initial efforts to apply evolutionary algorithms to the structure prediction problem were aimed at finding the lowest energy conformation of large organic molecules [17–21]. Evolutionary search techniques were also successfully applied to atomic clusters [22, 23], and soon the method was extended to 3D periodic systems [2, 12, 24].

It has been observed that evolutionary algorithms (EAs) are well suited to the structure prediction problem for several reasons [25]. First, they can efficiently find the global minimum of multidimensional functions. EAs require little information and few assumptions about the lowest energy structure, which is advantageous when searching for structures about which little is known a priori. Finally, if designed correctly, an EA can take advantage of the structure of the energy landscape discussed in Sect. 1.1.

The evolutionary approach to structure prediction is modeled after the natural process. Each crystal structure is analogous to a single organism. In nature, the fitness of an organism is based on how well its phenotype is suited to its environment and, in particular, how successful it is in reproducing. In an evolutionary algorithm, fitnesses are assigned to the organisms based on their objective function values, and they are allowed to reproduce based on those fitnesses. Pressures analogous to those which force species to adapt to their environments will thus lead to crystal structures with lower energies.

Organisms in an EA are often grouped into generations. The algorithm proceeds by creating successive generations. The methods by which an offspring generation is made from parents are called variation operations or variations and include operations that are analogous to genetic mutation and crossover. A single offspring organism can be created from either one or two parents, depending on which variation is used. Every offspring organism must meet some minimum standards to be considered viable, analogous to the "growing up" process in nature. This is known as the development process. The algorithm terminates when some user-defined stopping criteria are met (see Sect. 2.13).

Improvements are made to the biological analogy when possible. In particular, we would rather not let the optimal solution worsen from one algorithmic iteration to the next. To prevent this, a promotion operator is used to advance some of the best organisms from one generation directly to the next. Also, mutations in nature are usually detrimental. When searching for structures, one might try to use mutation variations that are likely to introduce valuable new information to the gene pool.

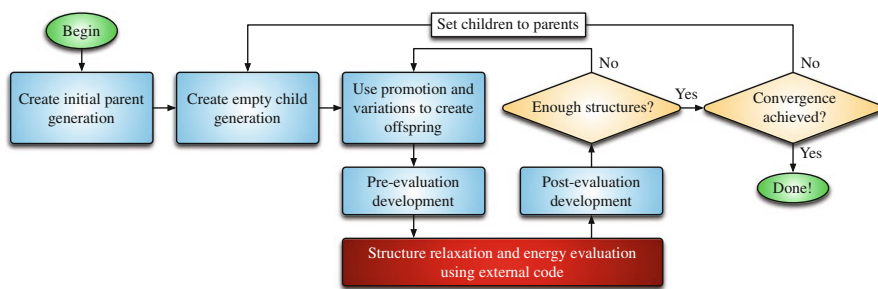


Fig. 2 Outline of evolutionary algorithm for structure prediction

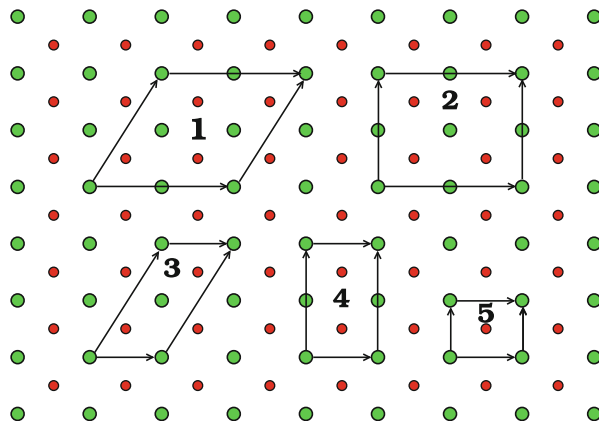
Figure 2 outlines how a typical evolutionary algorithm for structure prediction proceeds, such as that implemented by Tipton et al. [26, 27]. The EA starts by creating an initial population (Sect. 2.2) and calculating the fitness of each organism in it (Sect. 2.3). Organisms are then selected to act as parents (Sect. 2.4) or to be promoted to the next generation (Sect. 2.5). The parents create offspring structures via mating (Sect. 2.6) or mutation (Sect. 2.7). The offspring are developed (Sect. 2.9) and the energy of each offspring organism is calculated using some external energy code, followed by a post-evaluation development step. If successful, the offspring structures are then added to the next generation, and their fitnesses are calculated. Unless the EA has converged (Sect. 2.13), the current children become parents in the next generation.

2 Details of the Method

2.1 Representation of Structures

A total of $3N + 3$ dimensions describe the atomic coordinates and lattice vectors of a crystal structure. Additionally, the number of atoms, N , itself must be determined for ab initio structure predictions. However, these degrees of freedom are not all truly independent. Alternate choices of lattice vectors provide infinitely many ways to represent the same crystal structure, as illustrated in Fig. 3. Additionally, for molecular crystals, the dimensionality of the search space is effectively reduced since the molecular units typically stay intact in these crystal structures. This is due to the separation of energy scales, with strong intramolecular covalent interactions and much weaker intermolecular van der Waals interactions. In this case, structure search algorithms can take advantage of this trait by treating complete molecules, instead of individual atoms, as indivisible structural building blocks [28, 29]. Since the solution space is somewhat more complicated than it has to be, the task of searching that space is also more complicated than necessary. This difficulty may be addressed by attempting to standardize the way structures are represented in the computer.

Fig. 3 Five alternate representations of a single physical crystal are shown. Cells 2 and 4 are Niggli reduced versions of cells 1 and 3, respectively. They are also 2×2 and 1×2 supercells, respectively, of the primitive cell 5



2.1.1 Standardization of Representation

Two techniques are employed to standardize the representation of structures. The first and most widely used method is to impose hard constraints on the structures. These constraints include minimum interatomic distances and lattice parameter magnitudes. Limits on the maximum interatomic distances and lattice vector magnitudes are sometimes enforced as well [26, 27, 30, 31]. In addition, most authors constrain the range of angles between the lattice vectors [26, 27, 30–33]. If the algorithm varies the number of atoms per cell, this value is also constrained [27]. A restriction on the total volume of the cell is an additional possibility [30, 31, 34].

Physical considerations must be taken into account when choosing the constraints. For the constraint of the minimum interatomic distances, choosing 80% of the typical bond length or of the sum of the covalent radii of the two atoms under consideration has been proposed [30, 33]. The minimum lattice length has been chosen by adding the typical bond length and the diameter of the largest atom in the system [30, 32, 33]. Bahmann et al. set the maximum lattice vector length to the sum of the covalent diameters of all atoms in the cell [30]. Several authors limit angles between lattice vectors to lie between 60° and 120° [31, 32], although a more liberal range of $45\text{--}135^\circ$ has also been used [30, 33]. Ji et al. fix the volume of the cell during the search [34], and Lonie et al.'s algorithm can be set to use either a fixed cell volume or to constrain the volume to a user-specified range [31]. In the work of Bahmann et al. the cell volume is constrained to the range defined by the volume of the close-packed structure and four times that value [30].

Additional constraints may be used when one wishes to limit the search to a particular geometry. For example, Bahmann et al. restrict the allowable atomic positions and increase the maximum allowable lattice length in one direction to facilitate a search for two-dimensional structures [30]. Woodley et al. use an EA to search for nanoporous materials by incorporating “exclusion zones,” or regions in which atoms are forbidden to reside [35].

Several advantages are gained by using these constraints. Many energy models behave poorly when faced with geometries with very small interatomic distances, so enforcing this constraint from the start helps prevent failed structural relaxations and energy calculations. As mentioned in Sect. 1.1, large regions of the potential energy surface correspond to unphysical structures, and constraints help limit the search to regions that do contain physical minima.

Additionally, they help to ensure that structures are represented similarly. Removing as much redundancy as possible from the space of solutions makes the problem easier without limiting our set of possible answers or introducing any a priori assumptions as to the form of the solution. On the other hand, it is more dangerous to remove merely unlikely regions of the space from consideration, since doing so would bring into question both the validity of results and the claim to first-principles structure prediction.

The second method used to help standardize structure representation involves transforming the cells of all organisms to a unique and physically compact representation when possible. One way to do this is the Niggli cell reduction [27, 36]. There is a Niggli cell for any lattice that is both unique and has the shortest possible lattice lengths. Figure 3 illustrates the representation problem and the Niggli cell reduction. A similar transformation is used by Lonie et al. and Oganov et al. [31, 37]. In addition to simplifying the space that must be searched, removing redundancy by standardizing the representation of structures usually helps to increase the quality of the offspring produced by the mating variation, as is discussed in Sect. 2.6.

2.2 Initial Population

If no experimental data are available for the system under study, then the organisms in the initial population are generated randomly, subject to the constraints discussed above. The initial population generated in this way should sample the entire potential energy landscape within the constraints. If experimental data is available, such as from X-ray diffraction analysis, it can be used to seed the initial population with likely organisms. If one is predicting an entire phase diagram (see Sect. 3), the correct elemental and binary phases may already be known experimentally and can be used in the initial population. The use of pre-existing knowledge has the potential to decrease significantly the time needed to find the global minimum

When searching for molecular crystals, one typically places coherent molecular units instead of individual atoms into the structure [28]. Zhu et al. made an additional modification to the generation of the initial population to facilitate their study of molecular solids [29]. Instead of placing molecules at completely random locations within the cell, structures are built from randomly selected space groups. The authors found that this provides the algorithm with a more diverse initial population and improves the success of the search.

2.3 Fitness

The fitness of an organism is the property on which evolutionary pressure acts, and it depends on the value of the objective function. It is defined so that better solutions have higher fitness, and thus minimizing the energy means maximizing the fitness. It is usually defined as a linear function of the objective function, relative to the other organisms in the population. Exponential and hyperbolic fitness functions have also been used as an alternative way to introduce more flexibility into the selection algorithm (see Sect. 2.4) [38, 39]. In one frequently-applied scheme, an organism with a formation energy per atom, E_f , is assigned a fitness

$$f = \frac{E_f - E_f^{\min}}{E_f^{\max} - E_f^{\min}},$$

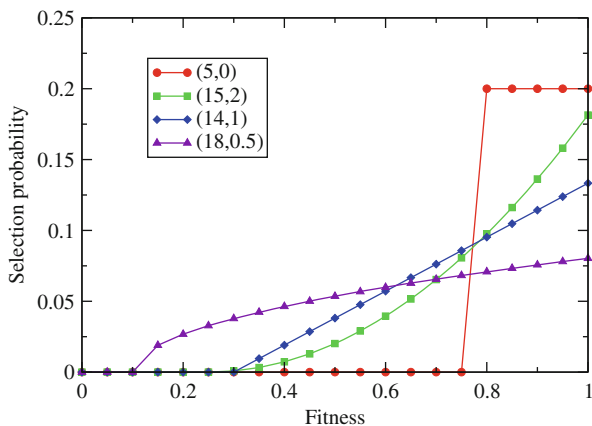
where E_f^{\max} and E_f^{\min} are the highest and lowest formation energies per atom, respectively, in the generation [26, 27, 31, 38]. In this case, the organism with the lowest energy in the generation is assigned a fitness of 1, and the organism with the highest energy has a fitness of 0. An alternative approach is to rank the organisms within a generation by their objective function values. The fitness of an organism is then defined as its rank [32, 34]. For cases when the stoichiometry and number of atoms in the cell is fixed, the fitness can simply be defined as the negative of the energy of the organism [30, 33].

2.4 Selection

The selection method determines which organisms will act as parents. Generally, structures with higher fitnesses are more likely to reproduce. The selection method is a crucial component of the search because it is the only way that the algorithm applies pressure on the population to improve towards the global minimum. Three commonly used strategies are elitist (or truncated) selection, roulette wheel selection, and tournament selection. In elitist selection the top several organisms are allowed to reproduce with equal probability while the rest are prevented from mating [39]. In roulette wheel selection, a random number d between the fitnesses of the best and worst organisms is generated for each organism, and if d is less than the fitness of the organism, it is allowed to reproduce [38]. In this way, it is possible for any organism except the worst one to reproduce, but it is more likely for organisms of higher fitness. Finally, in tournament selection, all of the organisms in the parent generation are randomly divided into small groups, usually pairs, and the best member of each group is allowed to reproduce.

Tipton et al. employ another approach to selection which is essentially a generalization of the three outlined above. Organisms are selected on the basis of a probability distribution over their fitnesses [26, 27]. Two parameters are used to

Fig. 4 Tuning the selection probability distribution ($N; p$) described by the number of potential parents N and power law p allows Tipton et al. to adjust the aggressiveness with which an EA seeks to converge



describe the distribution: the number of potential parents and an exponent which determines the shape of the probability distribution. The number of parents specifies how many of the best organisms in the current generation have nonzero probabilities of acting as parents. The exponent describes a power law. This method allows fine-grained control over the trade-off between convergence speed and the probability of finding the ground state. An aggressive distribution that puts a lot of pressure on the population to improve leads to faster convergence, but the algorithm is more likely to converge to only a local minimum. On the other hand, a less aggressive distribution will probably take more time to converge, but the algorithm has a better chance of finding the global minimum because a higher degree of diversity is maintained in the population. Several choices of selection probability distribution are illustrated in Fig. 4.

Authors employ these strategies in a variety of ways. One approach is to use elitist selection to remove some fraction of the parent generation, and then grow the resulting group back to its original size by creating offspring organisms from the remaining parent organisms with equal probability [30, 33, 34] or with a linear or quadratic probability distribution over their fitnesses [32]. Abraham et al. use roulette wheel selection. When the number of offspring organisms equals the number of parent organisms, either roulette wheel or elitist selection are used on the combined pool of structures to determine which organisms will make up the next generation. Elitist selection was found to be preferable in the final step [38]. Lonie et al. employ a linear probability distribution over the fitnesses to select organisms to act as parents. A continuous workflow is used instead of a generational scheme, so that offspring organisms are immediately added to the breeding population when they are created [31].

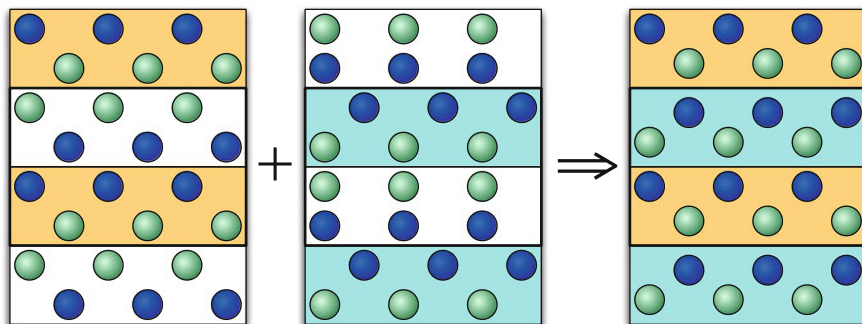


Fig. 5 Schematic illustration of how the mating operator in the evolutionary algorithm combines two crystal structures. For clarity, 1×2 supercells of the child and parent structures are shown

2.5 Promotion

A new generation is created from the structures in the previous one by applying selection in conjunction with promotion and variation. The promotion operation places some of the organisms in the old generation directly into the new generation without undergoing any changes. This is used to ensure that good genetic material is maintained in the population. Many authors use elitist selection to choose which organisms to promote. Lyakhov et al. refined selecting for promotion by only promoting structures whose fingerprints were significantly different (see Sect. 2.11 for fingerprinting) [40]. This was done to prevent loss of population diversity due to promoting similar organisms.

2.6 Mating

The goal of the mating variation is to combine two parents and preserve their structural characteristics in a single offspring organism. In its most basic form, mating consists of slicing parent organisms (cells) into two sections each and then combining one from each parent to produce an offspring organism. This is illustrated in Fig. 5.

It is important that the mating operation be designed so that traits which are important to the energy minimization problem have high heritability. The most energetically-important interactions in materials come from species located close to one another. This suggests that there is some amount of spatial separability in the energy-minimization problem, with the energy depending primarily on the local structure. The mating variation works by exploiting this feature of the problem. The slicing mating variation maintains much of the local structure of each parent in a very direct way, and we will now detail this operation.

2.6.1 Slice Plane Location and Orientation

After two parent organisms have been selected, the next step is choosing the planes along which to slice them. In order to mate organisms that do not have identical lattices, a fractional space representation is used. The positions of atoms in a cell are expressed in the coordinate system of the cell's lattice vectors. As a result, the fractional coordinates of all atoms within the cell have values within the interval $[0;1)$.

Authors have used various techniques to choose the orientation and location of the slice plane. In one method, a lattice vector A and a fractional coordinate s along A are randomly chosen [31, 32]. All atoms in one of the parent organisms with a fractional coordinate greater than s along A and all atoms in the second parent with fractional coordinate less than or equal to s are copied to the new child. Restricting the range of allowed values of s can be used to specify the minimum contribution by each parent to the offspring organism [31].

Alternatively, one may randomly select two planes that are parallel to a randomly chosen facet of the cell. Atoms that lie between these planes are then exchanged between the two parents. By "exchanged" we mean that each atom in the offspring has the same species-type and fractional coordinates as in the corresponding parent. This approach is equivalent to performing a translation operation on the atoms in the cell and then using the single slice plane method outlined above. Both of these methods choose slice planes that are parallel in real space to one of the cell facets of the parent structures [33, 34, 38]. Tipton et al. selects the two slice planes slightly differently: the fractional coordinate corresponding to the center of the sandwiched slab is randomly selected. The width of the sandwiched slab is then randomly chosen from a Gaussian distribution, and the two slice planes are placed accordingly [26, 27]. Another approach is simply to fix the locations of the two slice planes. For example, Abraham et al. specify that the two cuts be made at fractional coordinates of $1/4$ and $3/4$ along the chosen lattice vector [38].

Abraham et al. introduced a periodic slicing operation [38]. In this case, the value s described above becomes a cell-periodic function of the fractional coordinates along the cell lattice vectors other than A . A sine curve is often used, with the amplitude and wavelength drawn from uniform distributions. The wavelength is commonly constrained to be larger than the typical interatomic distance and smaller than the dimensions of the cell. The amplitude should also be small enough to ensure that no portion of the slice exceeds the boundaries of the cell. Abraham et al. found that periodic slicing improved the mean convergence time of their algorithm over planar slicing [38].

Constraining the degree of contribution of each parent by, for example, stipulating a minimum parental contribution can help prevent the mating operation from reproducing one or other of the parent structures essentially unchanged. Once the contribution of each parent has been determined, lattice vectors must be chosen for the offspring structure. Frequently, a randomly weighted average of the parents' lattice vectors is assigned to the offspring [31–33]. Simply averaging the lattice vectors of the parent organisms, i.e., fixing the weight at 0.5, is another common choice [26, 27, 34].

2.6.2 Number of Atoms and Stoichiometry of Offspring

An offspring organism produced via mating as described so far may have a different number of atoms or a different composition than its parents. This presents a difficulty if one wishes to perform a search with a fixed cell size or at a single composition. The simplest way to deal with these issues is simply to reject all offspring organisms that do not meet the desired constraints [33, 38]. Alternatively, nonconforming offspring can be made acceptable by the addition or removal of atoms. It may be best to add and remove atoms from locations near the slice plane [34]. These corrections minimize disruption to the structure transmitted from the parent organisms. Glass et al. use a slightly different approach: atoms to be removed or added are selected randomly from the discarded fragments of one of the parent organisms [31, 32]. Atomic order parameters have also been used to decide which excess atoms to remove (see Sect. 2.11) [40]. Those with the lower degrees of local order are more likely to be removed.

2.6.3 Modifications to the Mating Variation

Several additional modifications of the mating variation have been explored. The first involves shifting all the atoms in a cell by the same amount before mating [32, 41]. These shifts may happen with different probabilities along the axis where the cut is made and an additional random axis. This removes any bias caused by the implicit correlation between the coordinate s on the axis A in one crystal with the coordinate s on the axis A in the other. A similar effect may be obtained by selecting a random vector and shifting all atoms by this vector prior to making the cut [31]. In practice, these shifts help repeat good local structures to other parts of the cell.

In a further innovation, the parent organisms are subjected to random rotations and reflections prior to mating. This procedure removes bias toward any given orientation [31]. Additionally, an order parameter may be used to inform the choice of contribution from each parent (see Sect. 2.11). Several trial slabs of equal thickness are cut from the parents at random locations, and the slabs with the highest degree of order parameters are passed to the offspring [40, 41].

The simple slicing mating operator is not appropriate for molecular crystals, since it does not respect the integrity of the molecular units. Zhu et al. adapted the mating variation to search for molecular crystals [29]. In their scheme, each molecule is treated as an indivisible unit, and the location of the geometric center of a molecule is used to determine its location for the purposes of the mating operator.

2.6.4 Shortcomings of the Slicing Mating Variation

The mating variation acts directly on the particular representation of a structure in the computer. Since it is performed in fractional space, the mating variation can be applied to any two parent organisms, regardless of their cell shapes. However, the offspring structure may not always be successful. An offspring organism that has little in common with either parent can be produced if their representations (in particular the lattice parameters or number of atoms in the cells) are sufficiently different. As a result, the offspring will often have low fitness. Thus the mating variation is most successful when the parents are represented similarly because this increases the heritability of important traits.

The constraints and cell transformations discussed in Sect. 2.1 combat this issue through standardization of structure representation. Another method to increase the similarity of representation prior to mating is to use a supercell of one of the parent structures during mating [26, 27]. If one of the parent structures contains more than twice as many atoms as the other, a supercell of the smaller parent is used in the mating process. This technique ensures that both parent organisms are approximately the same size before mating, which aids in the creation of successful offspring.

Lyakhov et al. use an additional technique to help increase the viability of the offspring. If the distance between the parent organisms' fingerprints (see Sect. 2.10) exceeds a user-specified value, the would-be parents are not allowed to mate. The rationale behind this stipulation is that if two parents are from different funnels in the energy landscape, then their offspring would likely be located somewhere between those funnels and therefore have low fitness [40, 41].

2.6.5 Other Mating Operations

Not all evolutionary algorithms employ the previously described slicing method for mating. Bahmann et al. use a general recombination operation instead, where an offspring structure is produced by combining the lattice vectors and atomic positions of the two parent organisms [30]. This can be done in two ways: intermediate recombination takes a weighted average of the parents' values, and discrete recombination takes some values from each parent without changing them. Smith et al. used binary strings to represent structures on a fixed lattice, with each character in the string indicating the type of atom at a point on the lattice [24]. Mating was carried out by splicing together the strings of two parent structures. Jóhannesson et al. used a similar approach to search for stable alloys of 32 different metals [117]. Although all of these methods combine traits from each of the parents, they may not be as successful in passing the important local structural motifs of parents to the offspring.

2.7 Mutation

The goal of the mutation operation is to introduce new genetic material into the population. Its utility lies in its ability to explore the immediate vicinity of promising regions of the potential energy surface that have been found via the mating variation. The most common mutation entails randomly perturbing the atomic positions or lattice vectors of a single parent organism to produce an offspring organism. Some approaches call for mutating both the lattice vectors and the atomic positions [26, 27, 33], while others affect only one type of variable. To apply a mutation to the lattice vectors, they are subjected to a randomly generated symmetric strain matrix of the form

$$S = \begin{pmatrix} 1 + e_1 & e_6/2 & e_5/2 \\ e_6/2 & 1 + e_2 & e_4/2 \\ e_5/2 & e_4/2 & 1 + e_3 \end{pmatrix}, \quad (1)$$

where the components e_i are taken from uniform or Gaussian distributions [31–33].

Mutations of the atomic positions are achieved in a similar fashion. Each of the three spatial atomic coordinates is perturbed by a random amount, often obtained from a uniform or Gaussian distribution [26, 27]. To keep the size of these perturbations reasonable, either an allowed range or a standard deviation is set by the user. Most formulations do not mutate every atom in the cell but instead specify a probability that any given atom in the cell will be displaced. The approach of Abraham et al. combines mutation with the mating variation, perturbing atomic positions after mating has been performed [38]. However, most authors treat mutation as a separate operation.

Glass et al. claim that randomly mutating atomic positions is not necessary because enough unintentional change occurs during mating and local optimization to make it mostly redundant [32]. However, in later work, Lyakhov et al. use a “smart” mutation operation where atoms with low-order local environments (see Sect. 2.11) are shifted more [40, 41]. A further refinement to mutation has been made by shifting all the atoms along the eigenvector of the softest phonon mode [40, 41].

Permutation is another mutation-type operation for multi-component systems that swaps the positions of different types of atoms in the cell. Generally, the user specifies which types of atoms can be exchanged, and the algorithm performs a certain number of these exchanges each time the permutation variation is used on a parent organism [26, 27, 37]. The extent to which exchanging atomic positions affects the energy is strongly system-dependent. For ionic systems, exchanging an anion with a cation is likely to result in a much larger energy change than exchanges between two different types of cations or anions. In metals, on the other hand, the change in energy under permutation corresponding to anti-site defects is generally small. It is often helpful to use a permutation variation when studying these systems in order to find the minimum among several competing low energy configurations.

The number of swaps carried out can be random within a specified range, or can be pulled from a user-specified distribution. Randomly exchanging all types of atoms has the drawback that many energetically unfavorable exchanges may be performed, especially in ionic systems. If the number of atoms in the cell is small, Trimarchi et al. do an exhaustive search of all possible ways to place the atoms on the atomic sites [33].

Lonie et al. employ a “ripple” variation, in which all atoms in the cell are shifted by varying amounts [31]. First, one of the three lattice vectors is randomly chosen and then atomic displacements are made parallel to this axis. The amount by which each atom is shifted is sinusoidal with respect to the atom’s fractional coordinates along the other two lattice vectors. This produces a ripple effect through the cell. Lonie et al. argue that this variation makes sense because many materials display ripple-like structural motifs. Combining the ripple variation with other variations such as the lattice vector mutation and the permutation leads to hybrid variations that can improve the performance of the EA by reducing the number of redundant structures encountered in the search [31].

Zhu et al. employ an additional mutation when searching for molecular solids. Since molecules are not usually spherically symmetric, a rotational mutation operator was introduced, in which a randomly selected molecule is rotated by a random angle [29].

2.8 System Size

The number of atoms per cell, N , is an important parameter that needs to be considered. If N is fixed to a value which is not a multiple of the size of the ground state primitive cell of the material, the search cannot identify the correct global minimum. However, N is a difficult parameter to search over. In the case of other degrees of freedom for the solution, such as interatomic distances and cell volume, the local optimization performed by the energy code helps to find the best values. No analogous operation is possible in the case of N . Furthermore, the energy hypersurface is not particularly well behaved with respect to this parameter. It is likely that values of N surrounding the optimum will lead to structures quite high in energy while values of N further from the ideal may lead to closer-to-ideal structures.

Several approaches exist to search over this parameter. The first is simply to “guess” the correct value of N [31–33]. Guessing N can make it easy to miss the global minimum, especially for systems about which little is known a priori. To increase the chances of finding the right number, searches can be performed at several different values of N , but this is inefficient. A second technique is to allow the cell size of candidate solutions to vary during the search. This can be done passively through the mating variation by not enforcing a constraint on the number of atoms in the offspring structure [27, 34, 38]. Incorporating a mutation-type variation specifically designed for varying N is an additional option.

Another way to aid the search for the correct number of atoms per cell is to use large cells. Large supercells effectively allow several possible primitive cell sizes to be searched at once because the cells can be supercells of multiple smaller cells. For example, a search with a 50-atom supercell is capable of finding ground state structures with primitive unit cells containing 1, 2, 5, 10, 25, and 50 atoms. However, because the number of local minima of the energy landscape increases exponentially with N [13], and because individual energy calculations are much more expensive for larger structures, efficiency suffers.

Lyakhov et al. describe another difficulty with the large supercell approach that arises when generating the initial population. Randomly generated large cells almost always have quite poor formation energies, and disordered glass-like structures dominate. This discovery implies that there exists an upper limit to the size of randomly generated structures that can provide a useful starting point for the search. Starting an evolutionary algorithm with a low-diversity initial population comprised of low fitness structures provides a small chance of finding the global minimum [40, 41]. To obtain reasonably good large cells for the initial population, Lyakhov et al. generate smaller random cells of 15–20 atoms, and then take supercells of these [40, 41]. In this way, the organisms in the initial population can still contain many atoms, but they possess some degree of order and therefore tend to be more successful.

An alternative approach is to start with smaller supercells and encourage them to grow through the course of the search [26, 27]. This is achieved by occasionally doubling the cell size of one of the parents prior to performing the mating variation. The speed of cell growth in the population can be controlled through the frequency of the random doubling. The advantage of this technique is that it searches over N while still gaining (eventually) the benefits of large supercells. In addition, considering smaller structures first ensures that the quicker energy calculations are performed early in the search, and the more expensive energy calculations required for larger cells are only carried out once the algorithm has already gained some knowledge about what makes good structures.

2.9 Development and Screening

After a new organism has been created by one of the variations, it is checked against the constraints described in Sect. 2.1 and tested for redundancy (see Sect. 2.10). At this stage, many EAs scale the atomic density of the new organisms using an estimate of the optimal density [26, 27, 32]. Starting from an initial guess of the optimal density ρ_0 , the density estimate is updated each generation by taking a weighted average of the old best guess ρ_i and the average density of the best few structures in the most recent generation ρ_{ave} :

$$\rho_{i+1} = w\rho_{\text{ave}} + (1 - w)\rho_i,$$

where w is the density weighting factor. Then any time a new organism is made it is scaled to this atomic density before local relaxation. The primary reason for the density scaling is a practical one. Many minimization algorithms are quite time consuming if the initial solution is far from a minimum. This scaling is an easy first pass at moving solutions towards a minimum. Because the density scaling of an organism alters the interatomic distances, etc., the constraints checks are performed after the scaling of the density.

2.10 *Maintaining Diversity in the Population*

As the evolutionary algorithm searches the potential energy surface, equivalent structures sometimes occur in the population. If a pair of structures mates more than once, they are likely to create similar offspring. If the set of best structures does not change from generation to generation due to promotion, the set of parents, and thus the resulting set of children, can also be very similar. In addition, as the generation as a whole converges to the global minimum, all the organisms are likely to become more similar. What is worse, once a couple of low energy, often selected organisms are in the population, they can reproduce and similar structures will effectively fill up the next generations.

Duplicate structures hinder progress for several reasons. The most computationally expensive part of the algorithm is the energy calculations, and performing multiple energy calculations on the same structure is wasteful. However, this is exactly what happens if duplicate structures are not identified and removed from the population. Furthermore, low diversity in the population makes it difficult for the algorithm to escape local minima and to explore neighboring regions of the potential energy surface. This leads to premature convergence which is in practice indistinguishable from convergence to the correct global minimum. For these reasons, it is desirable to maintain the diversity of the population by identifying and removing equivalent structures. This is not a trivial task because, as discussed in Sect. 2.1, there exist infinitely many ways to represent a structure. Numerical noise adds to the difficulty of identifying equivalent structures.

Some authors directly compare atomic positions to determine whether two structures are identical. Lonie et al. developed an algorithm for this purpose, and it correctly identified duplicate structures that had been randomly rotated, reflected, or translated, and had random cell axes [42].

Tipton et al. also use a direct comparison of structures, with a slight modification [26, 27]. During the search, two lists of previously-observed structures are maintained. The first contains all the structures, relaxed and unrelaxed, that the algorithm has seen. If a new unrelaxed offspring structure matches one of the structures in the list, it is discarded. The assumption is that if it was good enough

to keep the first time, it was promoted, and if not, there is no reason to spend more effort on it. A second list contains the relaxed structures of all the organisms in the current generation. If a new relaxed offspring structure matches one of the structures in this list, it is discarded to avoid having duplicate structures in the generation. This approach both minimizes the number of redundant calculations performed and prevents the population from stagnating.

Wang et al. employ a bond characterization matrix to identify duplicate structures in the population [8]. The components of the matrix are based on bond lengths and orientations, and the types of atoms participating in the bond. Bahmann et al. identify duplicate structures by choosing a central atom in each organism and comparing the bond lengths between the central atom and the other atoms in a supercell [30]. These authors introduced an additional technique to help prevent the population from stagnating by stipulating an organism age limit. If an organism survives unchanged (via promotion) for a user-specified number of generations, it is removed from the population. This feature is meant to prevent a small number of good organisms from dominating the population and reducing its diversity.

Another method involves defining a fingerprint function which describes essential characteristics of a structure. When two organisms are found to have the same fingerprints, they are likely identical, and one is discarded. Several fingerprint functions have been used. The simplest is just the energy [22, 39]. The logic is that if two structures are in fact identical, they should have the same energies to within numerical noise. An interval is chosen to account for the noise. However, the size of the interval is fairly arbitrary and system-dependent, and this method is prone to false positives. Eliminating good unique organisms from the population can be even more detrimental to the search than not removing any organisms at all [42]. Lonie et al. expanded the fingerprint function to include three parameters: energy, space group, and the volume of the cell [31]. Again, intervals were set on the volume and energy. This is an improvement over simply using the energy as a fingerprint, but it can still occasionally fail, especially for low-symmetry structures or when atoms are displaced slightly from their ideal positions. Lonie et al. found that their direct comparison algorithm outperformed their fingerprint function at identifying duplicate structures [42].

Valle et al. employ a fingerprint function that is based on the distributions of the distances between different pairs of atom types in an extended cell [41, 43, 44]. For example, a binary system contains three interatomic distance distributions. The fingerprint function takes all three distributions into account. They used this fingerprint function to define an order parameter (see Sect. 2.11). Zhu et al. modified this fingerprint function slightly when searching for molecular solids; since distances between atoms within molecules do not change significantly, these distances are not considered when calculating structures' fingerprints [29].

Discarding duplicate structures from the population is not the only method employed to maintain diversity. Abraham et al. use a fingerprint function to determine how similar all the structures in a generation are to the lowest energy structure in the generation. Instead of simply removing similar structures, a modified fitness function is used which penalizes organisms based on their similarity to this best structure [45].

2.11 Order Parameters

Order parameters give a measure of the degree of order of an entire structure and also of the local environment surrounding individual atoms. Since energy is often correlated with local order, this can be a useful tool. Valle et al. extended their fingerprint function by using it to define an order parameter [40, 41, 44]. They used it to guide the algorithm at various points, as mentioned previously.

2.12 Frequency of Promotion and Variations

The user-specified parameters of an evolutionary algorithm affect its performance. However, running hundreds or thousands of structure searches to optimize these parameters can be prohibitively expensive, especially if an ab initio energy model is used. Furthermore, optimal values depend on the system under study. Physical and chemical intuition can be used to specify some of the parameters, such as the minimum interatomic distance constraint, but there exists no clear way to determine many of the others without performing enough searches to obtain reliable statistics.

Many authors arbitrarily choose how much each variation contributes to the next generation [32–34]. Lonie et al. performed thousands of searches for the structure of TiO_2 using empirical potentials to determine the best set of parameters for their algorithm [31]. They found that the relative frequency of the different variations did not significantly affect the success rate of the algorithm. However, the parameters associated with each variation did. For example, the lattice mutation variation was found to produce more duplicate structures when the magnitude of the mutation was small. This is likely to be due to structures relaxing back to their previous local minima when only slightly perturbed.

There is an important distinction between the relative frequency with which a given variation is called by the algorithm and the actual proportion of organisms in the next generation that are produced by that variation. The difference arises because not all the variations have the same likelihood of creating viable offspring. For example, mating is more likely to give good offspring structures than mutation of atomic positions because the latter will more frequently produce offspring that violate the minimum interatomic distance constraint. For this reason, the researcher's intention may be more clearly communicated if the proportion of offspring created by each variation is specified rather than the frequency that each variation is called by the algorithm.

Sometimes a situation arises in which it is not possible for one of the variations to produce a viable offspring organism. This could happen, for example, if the variation increases the size of the structures in the system, but all the potential parent organisms are already close to the maximum allowed cell size. In this case, the search will stop unless there is some way for the algorithm to get around the user-specified requirement that a certain percentage of the offspring come from this variation. Setting an upper limit on the number of failed attempts per variation is one way to achieve this [26, 27].

2.13 *Convergence Criteria: Have We Found the Global Minimum?*

When searching for an unknown structure, there is no known criterion that guarantees that the best structure encountered by the evolutionary algorithm is in fact the global minimum. One common technique to analyze the success rate of a heuristic search algorithm was given by Hartke [46]. In this method, many independent structure searches are performed on the same system, using the same set of parameters for the algorithm. For each search the energy of the best structure in each generation is recorded. These values are then used to create a plot of the energy vs generation number (or the total number of energy evaluations) that contains three curves: the energy of the highest-energy best structure, the energy of the lowest-energy best structure, and the average energy of the best structures. One shortcoming of the Hartke plot is that the lowest and highest best energies encountered are outliers, and in practice they depend strongly on the choice of the number of independent structure searches.

Tipton et al. employs a statistically more relevant approach to quantifying an EA's performance in which the median, the 10th percentile, and the 90th percentile energies of the best structures are plotted [27]. The 10th and 90th percentiles offer a better characterization of the distribution of results and are less susceptible to outliers and the number of independent structure searches performed to characterize the efficiency of the algorithm.

Figure 6 shows an example of a performance distribution plot for Zr_2Cu_2Al that was obtained by performing 100 independent runs of an EA with an embedded atom model potential [27]. These plots provide insight into the expected performance of the algorithm for the given material system and parameter settings and enable statistical comparisons of the performance of different methodologies or parameterizations of an EA. Of course, the strength of these conclusions depends on how many searches were used to construct the performance distribution plot, and the algorithm must be tested on systems with known ground state structures to be certain when the search was successful.

Lonie et al. showed that a decaying exponential fits the average-best energy curve of a Hartke plot well for a system with a known ground state structure [31]. The half-life of the exponential fit provides a measure of how fast the algorithm converges and can be used to determine a stopping criterion for the search. However, not all of the searches find the global minimum, so allowing a search to run for many half-lives still does not guarantee that the global minimum will be found, but it does increase confidence in the result.

A more common approach is to stop the search after a user-specified number of generations has elapsed without improvement of the best organism [26, 27, 30, 31]. Stopping once an allocated amount of computational resources has been expended is a popular alternative. Bahmann et al. have determined convergence when population diversity falls below a certain threshold or when all the organisms have very similar energies [30].

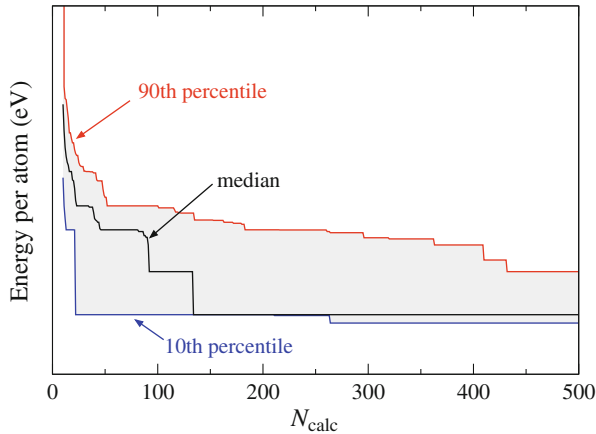


Fig. 6 Performance distribution plot for 100 structure searches at fixed composition for $\text{Zr}_2\text{Cu}_2\text{Al}$ using an embedded atom model potential [27]. The energy of the 90th percentile best structure is shown in *red*, the tenth percentile best structure in *blue*, and the median best structure in *black*

Although most authors use one of the fairly simple convergence criteria mentioned above, a quantitative statistical approach has been proposed by Venkatesh et al. [14]. Using Bayesian analysis, they determined the distribution of local minima based on the number found by a random search. This distribution was then used to calculate how many attempts would be required to find the global minimum with a specified probability.

3 Phase Diagram Searching

Even when one can say with a reasonable degree of confidence that the evolutionary algorithm has converged to the global minimum of the potential energy landscape, the result might still not represent the lowest energy structure that would be observed in nature. Skepticism is justified for several reasons [47]. First, as discussed in Sect. 2.8, unless the number of atoms in the cell is correctly guessed or allowed to vary, the EA cannot find the global minimum. Second, the structure identified as the global minimum might not be mechanically or dynamically stable, which would be reflected in energy-lowering imaginary phonon modes for the proposed global minimum crystal structure. Third, the reported global minimum might actually represent a metastable phase that decomposes into two or more structures with different stoichiometries. To determine whether this is the case, a phase diagram search must be performed. In addition to predicting the decomposition of structures into phases of other stoichiometries, phase diagrams are of great interest for many practical applications.

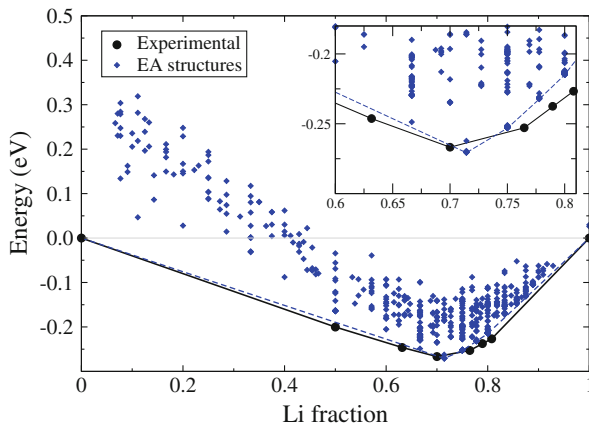


Fig. 7 A phase diagram search of the Li–Si binary system by Tipton et al. [48] using the method of Trimarchi et al. [49] showed that a search for relatively small unit cell structures could approximate the structural and energetic characteristics of the known very large experimental structures and thus be used to predict the voltage characteristics of a Li–Si battery anode. The search also identified a previously unknown member of the low-temperature phase diagram with composition Li_5Si_2

In order to perform a phase diagram search, we make use of the convex hull construction [9]. The formation energies of all structures with respect to the elemental constituents are plotted vs the composition. To determine the elemental references, one can either refer to the literature or perform preliminary searches. The smallest convex surface bounding these points is the convex hull, and the lowest energy facet for each composition is of physical interest. Thus, the convex hull is a graphical representation of the lowest energy a system can attain at each composition, and the points that lie on the convex hull correspond to stable structures. Figure 7 is an example of a convex hull for the Li–Si binary system [48].

Two approaches have been used to construct the convex hull. The first is to perform fixed-stoichiometry searches at many compositions [9, 50]. The lowest energy structure found in each search is then placed on an energy vs composition plot and the convex hull is constructed. However, this method is computationally expensive because it requires many separate searches to adequately sample the composition space [27].

The second approach entails modifying the evolutionary algorithm to search over composition space in the course of a single run. This requires two changes to the standard algorithm. The first is that the stoichiometry of structures the algorithm considers must be allowed to vary. This can be achieved simply by giving the initial population random stoichiometries and removing stoichiometry constraints on offspring structures [49]. The second modification involves the objective function. The algorithm constructs a current convex hull for each generation of structures, and a structure's objective function is defined as its distance from the current convex hull [26, 27, 49]. In this way, structures that lie on the current convex hull

have the highest fitness, and those above the convex hull have lower fitnesses. Selection then acts on this value in the standard way described in Sect. 2.4. As the search progresses, the true convex hull of the system is approached.

As discussed in Sect. 2.8, the global minimum cannot be found if the cell does not contain an integer multiple of the correct number of atoms. Since the structures lying on the convex hull often do not contain the same numbers of atoms, allowing the number of atoms to vary (Sect. 2.8) during the phase diagram search helps the algorithm find the correct convex hull. An alternative approach is to perform several composition searches with different, fixed system sizes. Each search generates a convex hull, and these hulls can be overlaid to obtain the overall lowest convex hull [49]. It should also be noted that the stoichiometries accessible to the algorithm are constrained by the number of atoms in the cell. For example, a cell containing four atoms in a binary system provides the algorithm with only five possible compositions (0, 25, 50, 75, and 100% A or B). The use of larger system sizes may be necessary for the algorithm to find the correct convex hull.

Another difficulty with phase diagram searches is inadequate sampling of the entire composition range. Mating between parents with different stoichiometries tends to produce offspring structures of intermediate composition. Because of this, over time the population as a whole may drift toward the middle region of the composition range, making it difficult to sufficiently sample more extreme compositions. Two solutions to this problem have been proposed [26, 27]. The first is to modify the selection criteria in such a way that mating between parents with similar compositions is encouraged. The second is to divide the composition range into sections and perform separate searches over each section. Agglomerating the results from all the sections gives the overall convex hull.

4 Energy Calculations and Local Relaxation

The potential energy landscape over which an evolutionary algorithm searches is defined by the code used for the energy calculations. These energy codes approximate the true potential energy landscape of the system, so the global minimum found by an EA will only represent the true global minimum of the system insofar as the approximate Hamiltonian accurately represents the physics of the system.

As discussed in Sect. 1.1, the potential energy surface is divided into basins of attraction by the local structure optimization or relaxation available in most energy codes (see Fig. 1). In order to find the global minimum, we must only sample a structure that resides in its basin of attraction, and the local optimizer will do the rest. This tremendously reduces the effective size of the space that must be searched; a relatively sparse sampling of a region can identify most of the local minima [7]. Local optimization is therefore crucial to the success of the search. Although the method depends on the energy code used, the local optimization problem is relatively well understood and its solutions are generally stable.

Glass et al. observed that the energies of a relaxed and unrelaxed structure are only weakly correlated [32]. This implies that the energy of an unrelaxed structure is not a reliable indicator of how close that structure is to a minimum in the potential energy landscape. Although omitting local optimization is computationally cheaper, an evolutionary search performed this way is unlikely to be successful. Woodley et al. compared the performance of an evolutionary algorithm with and without local relaxation and found that locally relaxing every structure greatly improved the efficiency and success rate of the algorithm [51].

Both empirical and ab initio energy codes have successfully been used in evolutionary algorithms to perform energy calculations and local relaxations. Due to their approximate nature, empirical potential energy landscapes often contain unphysical minima [52]. In addition, the cut-off distances imposed in many interatomic potentials leave discontinuities in the energy landscape, which can impede local relaxation. Although they mimic the true potential energy landscape more accurately than empirical potentials, density functional theory (DFT) calculations are also capable of misleading the search if care is not taken. Pickard et al. found that insufficiently dense k -point sampling can lead to false minima, and for calculations at high pressures, pseudopotentials with small enough core radii must be used to give accurate results [7].

Many EA implementations are interfaced with multiple energy codes, and more than one type of energy calculation may even be used in a single search. Ji et al. employed both empirical potentials and DFT calculations when searching for structures of ice at high pressures [53]. Lennard-Jones potentials were used for most energy calculations, but ab initio calculations were performed periodically and the parameters of the Lennard-Jones potentials were fitted to the DFT results. In this way, the empirical potential improved as the search progressed, and fewer computational resources were consumed than if ab initio methods alone had been used.

5 Summary of Methods

Tables 1 and 2 list the salient details of several implementations of evolutionary algorithms for structure prediction. The codes listed in Table 1 are production codes available to other users; the codes in Table 2 are research codes. In the following we summarize some of the distinguishing features of these evolutionary algorithms.

The Genetic Algorithm for Structure Prediction (GASP) is interfaced with VASP, GULP, LAMMPS, and MOPAC and has phase diagram searching capability [26, 27]. In addition, GASP can perform searches with a variable number of atoms in the cell, and it implements a highly tunable probability distribution for selecting organisms for mutation, mating, and promotion. The Open-Source Evolutionary Algorithm for Crystal Structure Prediction (XTALOPT) is interfaced with VASP, PWSCF, and GULP. It incorporates a unique ripple mutation, as well as hybrid mutations. It does not use a generational scheme but rather allows offspring structures to act as parents

Table 1 Comparison of the methods implemented into evolutionary algorithms in various available production codes

Name	GASP	XTALOPT	USPEX	EVO	MAISE
Authors	Tipton et al. [26, 27]	Lonie et al. [31]	Glass et al. [32]	Bahmann et al. [30]	Kolmogorov et al. [54]
Selection strategy	Probability distribution	Linear distribution	Elitist, linear, or quadratic probability distribution	Elitist	
Mutation of lattice vectors	Yes	Yes	Yes	Yes	Yes
Mutation of atomic positions	Yes	No	No [32], yes [40]	No	Yes
Permutation of atomic positions	Yes	Yes	Yes	No	Yes
Promotion	Yes	No	Yes	Yes	
Volume scaling	Yes	Yes	Yes	No	
Number of atoms in the cell	Variable	Fixed	Fixed	Fixed	
Cell reduction	Yes	Yes	Yes	No	
Diversity protection	Direct comparison	Fingerprinting [31], direct comparison [42]	Fingerprinting	Fingerprinting	
Phase diagram searching	Yes	No	No	No	No
Energy codes	VASP, MOPAC, GULP, LAMMPS	VASP, PWSCF, GULP, CASTEP	VASP, SIESTA, GULP, DMACRYS, CP2K, PWSCF	PWSCF, GULP	VASP
Unique features	Flexible probability distribution for selection	Ripple and hybrid scheme without generations	Use of order parameters	Age limit, alternative mating operation	

Table 2 Comparison of the methods implemented into evolutionary algorithms in various research codes

Authors	Ji et al. [34]	Trimarchi et al. [33]	Bush et al. [2]	Abraham et al. [38]
Selection strategy	Elitist	Not specified	Elitist	Roulette wheel, elitist
Mutation of lattice vectors	No	Yes	No	No
Mutation of atomic positions	No	Yes	Yes	Yes
Permutation of atomic positions	Yes	Yes	No	No
Promotion	Yes	Yes	Yes	Yes
Volume scaling	No	No	No	No
Number of atoms in the cell	Fixed	Fixed	Fixed	Variable
Cell reduction	No	No	No	No
Diversity protection	None	None	None	None
Phase diagram searching	No	Yes	No	No
Energy codes	VASP	VASP	GULP	CASTEP
Unique features	Cells constrained to constant volume		Surrogate objective function	Periodic slicing, mutation only after mating

as soon as they are created [31]. The Universal Structure Predictor: Evolutionary Xtallography (USPEX) code is interfaced with VASP, SIESTA, PWSCF, GULP, DMACRYS, and CP2K. It incorporates a unique order parameter, both for cells and individual atoms, that is used to help guide the search [32, 40]. The Evolutionary Algorithm for Crystal Structure Prediction (EVO) is interfaced with PWSCF and GULP. It applies an age limit to structures encountered in the search, and it also employs a mating operation that is different from the cut-and-splice technique used in most other evolutionary structure searches [30]. Finally, the Module for Ab Initio Structure Evolution (MAISE) is interfaced with VASP. Both planar and periodic slices can be used during mating, and it has the option to perform mating and mutation in a single variation [54].

Trimarchi et al. developed an evolutionary algorithm for structure prediction that is interfaced with VASP [33]. They later extended the algorithm to include phase diagram searching [49]. Abraham et al. designed an evolutionary algorithm with several unique features, including periodic slicing during the mating operation and mutation of atomic positions only after mating [38]. The algorithm also accepts offspring structures with different numbers of atoms than the parents. It is interfaced with CASTEP. The evolutionary algorithm of Ji et al. is interfaced with VASP, and it constrains the structures it considers to a constant volume

[34]. Bush et al. developed an evolutionary algorithm that incorporates a surrogate objective function [2].

6 Applications

Evolutionary algorithms have been used to solve the structures of many types of systems including molecules, clusters, surfaces, nanowires, and nanoporous materials [39, 55–57]. Here we focus on applications of EAs to bulk, 3D periodic systems. Within this constraint, we have made an effort to provide a comprehensive review of prior applications. We grouped the application into six categories based on the type of material studied: pure elements, hydrogen-containing compounds, intermetallics, minerals, molecular solids, and other inorganic compounds. Tables 3, 4, 5, 6, 7, and 8 correspond to these categories and list the applications of the method. For each study, we indicate the system studied, the number of atoms in the configuration space searched over, the energy code used, and the lead author.

6.1 Elemental Solids

Table 3 describes searches for elemental solids. Some elemental phase diagrams are still not fully characterized, especially under extreme conditions such as high pressure. Several elements have been predicted to display unusual properties at high pressure, such as superconductivity. Ma and Oganov studied several different elements under pressure. They found a new phase of boron with 28 atoms in the unit cell that is predicted to be stable in the pressure range 19–89 GPa [58]. A search of carbon under high pressures led to the prediction that the bc8 structure is more stable than diamond above 1 TPa [25]. Oganov et al. predict several new superconducting phases of calcium at pressures up to 120 GPa [61]. A study of hydrogen at pressures up to 600 GPa predicted that it remained a molecular solid throughout this pressure range [25]. The interesting case of hydrogen under pressure will be further described below in the discussion of hydrogen-containing compounds. Ma et al. predict that potassium and rubidium follow the same sequence of phase transitions under pressure (40–300 GPa) observed experimentally for cesium, but predict a new cubic phase of lithium above 300 GPa [63]. Ma et al. also studied nitrogen under pressure, predicting new polymeric insulating phases above 188 GPa, and studied sodium at pressures up to 1 TPa, predicting a new optically transparent, insulating phase above 320 GPa [65]. They have also reported a monoclinic, metallic, molecular phase of oxygen in the range of 100–250 GPa whose calculated XRD diffraction pattern is in agreement with experiment [66].

Bi et al. searched for phases of europium at pressures up to 100 GPa and predicted several nearly degenerate structures in the range 16–45 GPa, which

Table 3 Application of evolutionary algorithms to single element systems

Element	Pressure (GPa)	Number of atoms	Number of atoms	References
Al	0	4, 8, 12	PWSCF	Bahmann et al. [30]
B	0,100,300	2, 3, 4, 6, 8, 9, 16, 12, 24, 26, 28, 30, 32	VASP	Oganov et al. [58]
Ba	0	24, 25, 26, 27, 28, 29, 30, 31, 32	VASP	Ji et al. [34]
	0–300	Variable, up to 15	VASP	Taillon et al. [59]
C	10–100	2, 4, 6, 8	VASP	Li et al. [60]
	0	Not specified	CASTEP	Abraham et al. [38]
	0	8, others tried	PWSCF	Bahmann et al. [30]
	0–2,000	8	VASP	Oganov et al. [25]
Ca	20–600	3, 4, 6, 8, 9, 12, 16	VASP	Oganov et al. [61]
Cl	100	8	VASP	Oganov et al. [25]
Eu	0–90	Variable, up to 30	VASP	Bi et al. [62]
F	50, 100	8	VASP	Oganov et al. [25]
Fe	350	8	VASP	Oganov et al. [25]
Ga	0	8	PWSCF	Bahmann et al. [30]
H	Up to 600	2, 3, 4, 6, 8, 12, 16	VASP	Oganov et al. [25]
In	0	4	PWSCF	Bahmann et al. [30]
K	40–300	4, 6, 8, 12, 16	VASP	Ma et al. [63]
Li	30–1,000	4, 6, 8, 12, 16, 24	VASP	Ma et al. [63]
Mo	0	Variable, up to 40	LAMMPS	Park et al. [52]
N	100–350	8, 12, 16, 32	VASP	Ma et al. [64]
	100	6, 8, 12, 16	VASP	Oganov et al. [25]
Na	0–1000	Not specified	VASP	Ma et al. [65]
O	100–250	4, 8	VASP	Ma et al. [66]
	25, 130, 250	4, 6, 8, 12, 16	VASP	Oganov et al. [25]
Rb	40–200	4, 6, 8, 12	VASP	Ma et al. [63]
S	12	3, 4, 6, 8, 9, 12	VASP	Oganov et al. [25]
Si	0	8	VASP	Trimarchi et al. [33]
	10, 14, 20	8	VASP	Oganov et al. [25]
Tl	0	4	PWSCF	Bahmann et al. [30]
Xe	200, 1,000	8	VASP	Oganov et al. [25]

may help explain the mixed phase structure observed experimentally in this pressure regime [62] and the occurrence of superconductivity and magnetism in these phases [109]. In a novel application of evolutionary algorithms, Park et al. used an EA to verify that a new modified embedded atom potential for molybdenum accurately reproduced the energy landscape of molybdenum [52].

Table 4 Application of evolutionary algorithms to hydrogen-containing compounds

Compound	Pressure (GPa)	Formula units per cell	Energy code	References
Li_mH , $m = 2 - 9$	60, 80, 100	2, 4	VASP	Hooper et al. [67]
LiH_n , $n = 2 - 8$	0–300	0–300	VASP	Zurek et al. [50]
NaH_n , $n = 6 - 12$	100, 300	2, 3, 4	VASP	Baettig et al. [68]
KH_n , $n = 2-12$	10, 100, 250	2, 4	VASP	Hooper et al. [69]
RbH_n , $n = 2-14$	2–250	2	VASP	Hooper et al. [70]
CsH_n , $n = 2-9;16$	30–200	2, 3, 4	VASP	Shamp et al. [71]
BeH_n , $n = 2-5$	50, 150, 200	2, 4	VASP	Hooper et al. [72]
MgH_n , $n = 2-16$	0–250	2, 3, 4, 5, 6, 8	VASP	Lonie et al. [73]
BaH_n , $n = 2-13$	50–200	2, 4	VASP	Hooper et al. [72]
WH_n , $n = 1-6;8$	25, 150	1, 2, 3, 4	VASP	Labet et al. [74]
CH_4	20–800	2, 3, 4	VASP	Gao et al. [75]
	11	Not specified	VASP	Zhu et al. [29]
SiH_4	40–300	1, 2, 3, 4, 6	VASP	Martinez et al. [76]
GeH_4	50–250	1, 2, 3, 4	VASP	Gao et al. [77]
SnH_4	30–250	1, 2, 3, 4	VASP	Gao et al. [78]
PbH_4	0–500	2, 4	VASP	Zaleski-Ejgierd et al. [79]
NaH	0, 29.3	4	PWSCF, VASP	Lonie et al. [31]
CH	0–300	Fixed, up to 8	VASP	Wen et al. [80]
PtH_x , $x = \frac{1}{4}, \frac{1}{3}, \frac{1}{2}, 1, 2, 3, 4$	120	6, 10, 12	VASP	Zhou et al. [81]
FeH_x , $x = \frac{1}{4}, \frac{1}{3}, \frac{1}{2}, 1, 2, 3, 4$	300, 400	2, 3, 4, 6, 8	VASP	Bazhanova et al. [82]
LiBeH_3	50–550	2, 4	VASP	Hu et al. [83]
LiNH_2	0–360	2, 4	VASP	Prasad et al. [84]
NaAlH_4	0–20	1, 2	VASP	Zhou et al. [85]
NaPtH_2	0	Not specified	VASP	Wen et al. [86]
NaNH_2	0, 10, 20	1, 2, 3, 4	VASP	Zhong et al. [87]
SrFeH_4	0	Not specified	VASP	Wen et al. [86]
$\text{Mg}(\text{BH}_4)_2$	0–20	2, 4	VASP	Zhou et al. [88]

6.2 Hydrogen-Containing Compounds

Table 4 summarizes EA structure searches performed on hydrogen-containing compounds. Ashcroft suggested in 1968 that hydrogen could become a high-temperature superconductor under pressure [110] and in 2006 that doping hydrogen to form chemically precompressed hydrogen-rich materials could be a potential route to reduce the pressure required for superconductivity [111].

Hooper, Lonie, and Zurek have performed several studies on polyhydrides of alkali and alkaline earth metals under pressure. A metallic phase of LiH_6

Table 5 Application of evolutionary algorithms to intermetallic compounds

Compound	Pressure (GPa)	Formula units per cell	Energy code	References
Al–Sc system	0	8 atoms	VASP	Trimarchi et al. [89]
	0	6, 8 atoms	VASP	Trimarchi et al. [49]
	0	8 atoms	VASP	Ji et al. [34]
Al ₁₃ K	0	1	VASP	Oganov et al. [37]
Au ₂ Pd	0	4	VASP	Trimarchi et al. [33]
	0	4	VASP	Trimarchi et al. [89]
CaLi ₂	10–250	1, 2, 3, 4, 6, 8	VASP	Xie et al. [90]
CdPt ₃	0	2	VASP	Trimarchi et al. [89]
CuPd	0	4	VASP	Trimarchi et al. [89]
Na–Ca system	50	Not specified	VASP	Taillon et al. [59]
PdTi ₃	0	2	VASP	Trimarchi et al. [89]

Table 6 Application of evolutionary algorithms to minerals

Mineral	Pressure (GPa)	Formula units per cell	Energy code	References
Al ₂ O ₃	300	4	VASP	Oganov et al. [25]
Al ₂ SiO ₅	10	4	GULP	Zhu et al. [91]
CaCO ₃	50, 80, 150	1, 2, 4	VASP	Oganov et al. [25]
FeC _x , $x = \frac{1}{4}, \frac{1}{3}, \frac{1}{2}, 1, 2, \frac{7}{3}, 3, 4$	300, 400	2, 3, 4, 6, 8	VASP	Bazhanova et al. [82]
FeS	56, 120, 400	2, 3, 4, 6, 8	VASP	Ono et al. [92]
FeSi _x , $x = \frac{1}{3}, \frac{1}{2}, 1, \frac{5}{3}, 2, 3$	300, 400	8, 9, 12, 16, 18, 24 atoms	VASP	Zhang et al. [93]
MgCO ₃	110, 150	1, 2, 4, 6	VASP	Oganov et al. [25, 37]
Mg–O system	Up to 850	Up to 20 atoms	VASP	Zhu et al. [94]
MgSiO ₃	250	32	GULP	Zhu et al. [91]
	80, 120, 1,000	4, 8	VASP	Oganov et al. [25]
Na–Cl system	0–250	Up to 16 atoms	VASP	Zhang et al. [95]
SiO ₂	500, 2,000	1–8	PWSCF	Wu et al. [96]
	0	3	VASP	Oganov et al. [25]
	10	24	GULP	Zhu et al. [91]
TiO ₂	0	2, 4, 8	GULP	Woodley et al. [51]
	0	16	GULP	Lonie et al. [31]
	Not specified	4	VASP	Lonie et al. [31]

was predicted to be stable above 110 GPa [50], and a stable phase of NaH₉ was predicted to metallize at 250 GPa [68]. Stable rubidium polyhydride phases were predicted to metallize at pressures above 200 GPa [70]. A stable, superconducting phase of MgH₁₂ was identified under pressure, with a predicted

Table 7 Application of evolutionary algorithms to molecular solids

Compound	Pressure (GPa)	Formula units per cell	Energy code	References
Benzene C ₆ H ₆	0–300	Not specified	VASP	Wen et al. [80]
	0, 5, 10, 25	4	VASP	Zhu et al. [29]
Ice H ₂ O	1,000–4,000	8	LAMMPS, PWSCF	Ji et al. [53]
	0	4	VASP	Zhu et al. [29]
	0	4	VASP	Oganov et al. [25]
CO ₂	50, 100, 150	1, 2, 3, 4, 6, 8	VASP	Oganov et al. [37]
	12–20	2	VASP	Zhu et al. [29]
	50	1, 2, 3, 4, 6, 8	VASP	Oganov et al. [25]
NH ₃	5, 10, 25, 50	4	VASP	Zhu et al. [29]
Glycine	1, 2	2, 3, 4	VASP	Zhu et al. [29]
Butane-1,4- diammonium dibromide	0	2	VASP	Zhu et al. [29]
Urea CO(NH ₂) ₂	0	2	VASP	Oganov et al. [25]

T_c of 47–60 K at 140 GPa [73]. Lonie et al. also identified a new phase of BeH₂ above 150 GPa, as well as a stable superconducting phase of BaH₆, with a T_c of 30–38 K at 100 GPa [72].

Several studies have been performed on the group IV hydrides ranging from methane to plumbane. Gao et al. searched for methane structures under pressure and predicted that it dissociates into ethane and hydrogen at 95 GPa, butane and hydrogen at 158 GPa, and finally carbon and hydrogen at 287 GPa [75]. Martinez et al. looked at silane under pressure and predicted two new phases, one stable from 25 to 50 GPa, and the other from 220 to 250 GPa. The latter was predicted to be superconducting, with a T_c of 16 K at 220 GPa [76]. Gao et al. searched for germane and stannane under pressure. Germane was predicted to be stable with respect to decomposition into pure germanium and hydrogen above 196 GPa, and it was predicted to be superconducting, with a T_c of 64 K at 220 GPa [77]. Two stannane isomers were predicted – one stable from 96 to 180 GPa, and the other occurring above 180 GPa. Both phases were calculated to be superconductors [78]. Zaleski et al. performed evolutionary structure searches for plumbane (PbH₄) under pressure and predicted that it forms a stable non-molecular solid at pressures greater than 132 GPa [79]. Wen et al. predicted five low energy three-dimensional structures of graphane in the pressure range 0–300 GPa, and each was either semiconducting or insulating [80].

Zhou et al. investigated platinum hydrides under pressure and predicted a superconducting hexagonal phase of PtH to be stable above 113 GPa [81]. Hu et al. searched for LiBeH₃ at pressures up to 530 GPa and predicted two new insulating phases [83]. Zhou et al. found two new tetragonal structures of Mg(BH₄)₂ under pressure whose densities, bulk moduli, and XRD patterns match experimentally measured values [88].

Table 8 Application of evolutionary algorithms to various inorganic compounds

Compound	Pressure (GPa)	Formula units per cell	Energy code	References
Al ₁₂ C	0	1	VASP	Oganov et al. [37]
BeB ₂ , BeB ₃ , BeB ₄ , BeB _{2.75}	0, 160	4, 8	VASP	Hermann et al. [97]
CsI	5–300	2, 3, 4, 8	VASP	Xu et al. [98]
Fe-B system	0	Up to 15 atoms	VASP	Kolmogorov [54]
GaAs	0	4	VASP	Trimarchi et al. [33]
Li-B system	0–320	1, 2	VASP	Hermann et al. [99, 100]
Li-Si system	0	Variable, up to 20 atoms	VASP	Tipton et al. [48]
MgB ₂	5–300	1, 2, 3, 4, 6	VASP	Ma et al. [101]
SiC	0	4	VASP	Trimarchi et al. [33]
WN ₂	0, 60	1, 2, 3, 4	VASP	Wang et al. [102]
XeF ₂	0–200	Up to 4	VASP	Kurzydowski et al. [103]
Xe-O System	5–220	Up to 36 atoms	VASP	Zhu et al. [104]
ATiO ₂ , A=Ba, Be, Ca, Mg, Sr	0	Not specified	VASP	Wen et al. [105]
NaPtF ₂	0	Not specified	VASP	Wen et al. [86, 105]
SrFeC ₄	0	Not specified	VASP	Wen et al. [86, 105]
CaRhO ₃	Not specified	6	VASP	Shirako et al. [106]
Li ₃ RuO ₄	0	Not specified	GULP	Bush et al. [2]
SrTiO ₃	0	10	GULP	Lonie et al. [31]
BC ₂ N	30, 100	1, 2, 4	VASP	Li et al. [107]
LiBeB	30, 100	Not specified	VASP	Hermann et al. [108]
Si ₂ N ₂ O	0	2	VASP	Oganov et al. [25]
SrSiN ₂	0	4	VASP	Oganov et al. [25]

6.3 Intermetallic Compounds

Table 5 summarizes searches for intermetallic compounds. Trimarchi et al. have studied many intermetallics. Their algorithm identified the correct lattice of Au₂Pd, which is known to be fcc, but it failed to find the lowest energy atomic configuration [33] because the system exhibits several nearly degenerate structures. In another study by these authors, a new phase of IrN₂ was discovered [89]. They performed a phase diagram search on the Al–Sc system, which exhibit several crystal structures across the composition range, and successfully identified the experimentally known ground state phases [49]. Xie et al. discovered two new superconducting phases of CaLi₂, one stable from 35 to 54 GPa and the other stable from 54 to 105 GPa. Furthermore, they predict that CaLi₂ is unstable with respect to dissociation into the constituents at pressures greater than 105 GPa [90].

Sometimes the underlying lattices for these systems are known empirically, and the search reduces to finding the lowest energy arrangement of atoms on the lattice. For these cases, an efficient method to search over permutations of atomic positions is crucial. D’Avezac et al. employed a virtual atom technique, in which the species type of an atom is “relaxed” to determine if exchanging atom types at that site would likely lead to a lower energy configuration [112]. In contrast to real space mating operations, these authors also employed a reciprocal space mating scheme [112]. With this technique, the structure factors of two parent organisms are combined to form the offspring organism’s structure factor, which is then transformed to real space, giving the offspring organism. A cluster expansion fitted to DFT calculations was leveraged to perform energy calculations.

6.4 Minerals

Table 6 contains a summary of EA searches for the structures of several minerals. Many of these studies were carried out at high pressure in order to simulate the conditions in planetary interiors. Oganov et al. found two new phases of MgCO_3 under pressure. In addition, a new phase of CaCO_3 was reported to be stable above 137 GPa [37], and the structure of the post-aragonite phase of CaCO_3 , stable from 42 to 137 GPa, was solved [16]. Iron carbides of various compositions were explored under pressure, and it was predicted that the cementite structure of Fe_3C is unstable at pressures above 310 GPa, indicating that this phase does not exist in the Earth’s inner core [82]. Zhang et al. searched for iron silicide structures under pressure and predicted that only FeSi with a cesium chloride structure is stable at pressures greater than 20 GPa. Ono et al. investigated the structure of FeS under pressure and reported several new phases up to 135 GPa [92]. Wu et al. predicted a new low temperature post-perovskite phase of SiO_2 with the Fe_2P -type structure [96].

6.5 Molecular Crystals

Table 7 summarizes searches for molecular crystals. Applications for molecular solids include high-energy materials, pharmaceuticals, pigments, and metal-organic frameworks [113–115]. Molecular solids are not always in their thermodynamic ground states. Instead, the system is kinetically trapped and the molecular units are maintained. Zhu et al. made several changes to the standard EA to facilitate searching for molecular crystals, and they applied their algorithm to search for structures of ice, methane, ammonia, carbon dioxide, benzene, glycine, and butane-1,4-diammonium dibromide. Experimentally known structures were recovered by the algorithm [29]. Ji et al. searched for ice at terapascal pressures and predicted three new phases [53]. Lennard-Jones potentials were used to model the system, but

ab initio calculations were periodically performed and the results used to fit the empirical potentials. Hermann et al. performed evolutionary searches for high-pressure phases of ice and predicted that ice becomes metallic at 4.8 TPa [116]. Oganov et al. predicted that the β -cristobalite structure is the most stable for CO_2 between 19 and 150 GPa [37].

6.6 Inorganic Compounds

Table 8 summarizes searches for inorganic compounds. Many of these studies aimed to clarify regions of various phase diagrams. Others sought to identify phases with desirable properties, such as superconductivity.

Tipton et al. applied an evolutionary algorithm to investigate Li-Si anode battery materials and carried out a phase diagram search on the Li-Si system. They discovered a new stable phase with composition Li_5Si_2 [48]. Hermann et al. searched for structures in the Li-B system under pressure and found several stable structures. LiB was found to become increasingly stable as the pressure was increased beyond 300 GPa [99, 100]. Hermann et al. also predicted a new stable phase of LiBeB at ambient pressure and several additional phases under pressures up to 320 GPa [108]. Kolmogorov et al. searched for Fe-B structures at several different compositions and reported new phases with compositions FeB_4 and FeB_2 .

Xu et al. discovered a new orthorhombic phase of CsI that is predicted to be stable from 42 GPa up to at least 300 GPa [98]. This material is predicted to metallize at 100 GPa and to become superconducting at 180 GPa. Zhu et al. searched for xenon oxides under pressure and found three stable compounds: XeO above 83 GPa, XeO_2 above 102 GPa, and XeO_3 above 114 GPa [104]. Bush et al. solved the structure of Li_3RuO_4 at zero pressure. They used an alternative fitness function in their EA and only calculated energies at the end of the search [2]. Li et al. resolved the structure of superhard BC_2N and found it to have a rhombohedral lattice [107].

7 Conclusions

Crystal structure prediction is a long-standing challenge in the physical sciences. If we frame the structure prediction problem as one of global optimization, robust and accurate free energy methods such as those described in Sect. 4 can be used as objective functions, which can be minimized to find the thermodynamically stable structure. This minimization problem has been effectively addressed in recent years using evolutionary algorithms. However, the EA is less of a particular algorithm and more of a general problem solving strategy. Thus, many methodological and design choices are possible when creating an EA for structure prediction, and innovation in the field is ongoing.

The method generally begins with a broad sampling of the solution phase space. Information gained from early calculations is used to try to guess new low energy candidate structures. The ability to make such inferences relies on some characterization of or knowledge about the structure of the energy landscape, as described in Sect. 1.1. In the evolutionary approach, we leverage the power of biological evolution to search for low energy structures. Parent structures that are good solutions to the problem are varied and combined in such a way so as to pass down their traits to children. In this way, favorable properties are propagated in the population while unfavorable ones tend to die out. The parent selection and variation operators are very important to the success of this approach, and the most common and successful variation operators take advantage of the partial spatial separability of the energy minimization problem. These and other methodological issues were discussed in Sect. 2.

In Sect. 6 we reviewed many applications of the method to systems of fundamental scientific as well as technological interest. As the field has begun to mature, researchers have had many successes in practical applications. Several publicly available software packages for performing these calculations exist and are described in Sect. 5. However, evolutionary algorithms are not yet a commodity method, and an understanding of the methodology remains helpful for obtaining best results.

A number of challenges remain. Since *ab initio* energy calculations are the most expensive part of the method, reducing the number of these necessary to obtain high quality predictions is the focus of methodological developments. The energy and relaxed structure are a small subset of the information provided by *ab initio* calculations, so the opportunity exists to improve results by making better use of the full set of data. Increasing the efficiency of the search through solution representation and variation operators is also a promising avenue of improvement. Searching over structures' compositional degrees of freedom remains inefficient since there is no local relaxation of these parameters, and the energy landscape has little structure with respect to them. This makes prediction of the number of atoms in the unit cell and phase diagram prediction challenging, as discussed in Sect. 3.

Although they lie beyond the scope of this review and volume, uses of evolutionary algorithms to predict the atomic structures of other systems, such as surfaces, 1D and 2D materials, atomic clusters, or molecules, etc., are also important and active areas of research. A primary goal of computational materials science is to find materials with desirable properties, and structure prediction is a necessary early step in first principles prediction of materials' properties. Evolutionary algorithms have turned out to be a useful global optimization method for addressing the structure prediction problem.

Acknowledgments This work was supported by the National Science Foundation under award number CAREER DMR-1056587 and by the Energy Materials Center at Cornell (EMC2) funded by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences under award number DE-SC0001086. W.W.T. was supported by the NSF IGERT Fellowship Program "A Graduate Traineeship in Materials for a Sustainable Future" under award number DGE-0903653 and the NSF GK12 Program "Grass Roots: Advancing education in renewable energy and cleaner fuels through collaborative graduate fellow/teacher/grade-school student

interactions” under award number DGE-1045513. This research used computational resources of the Texas Advanced Computing Center under Contract Number TG-DMR050028N and of the Computation Center for Nanotechnology Innovation at Rensselaer Polytechnic Institute.

References

1. van Dover RB, Schneemeyer L, Fleming R (1998) Discovery of a useful thin-film dielectric using a composition-spread approach. *Nature* 392(6672):162–164
2. Bush TS, Catlow CRA, Battle PD (1995) Evolutionary programming techniques for predicting inorganic crystal structures. *J Mater Chem* 5:1269–1272
3. Ceder G, Morgan D, Fischer C, Tibbetts K, Curtarolo S (2006) Data-mining-driven quantum mechanics for the prediction of structure. *MRS Bull* 31(12):981–985
4. Goedecker 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
5. Martoňák R, Laio A, Parrinello M (2003) Predicting crystal structures: the Parrinello–Rahman method revisited. *Phys Rev Lett* 90(7):075503
6. Pannetier J, Bassas-Alsina J, Rodriguez-Carvajal J, Caignaert V (1990) Prediction of crystal structures from crystal chemistry rules by simulated annealing. *Nature* 346(6282):343–345
7. Pickard CJ, Needs RJ (2011) Ab initio random structure searching. *J Phys Condens Matter* 23(5):053201
8. Wang Y, Lv J, Zhu L, Ma Y (2012) CALYPSO: a method for crystal structure prediction. *Comput Phys Commun* 183(10):2063–2070
9. Feng J, Hennig RG, Ashcroft NW, Hoffmann R (2008) Emergent reduction of electronic state dimensionality in dense ordered Li–Be alloys. *Nature* 451(7177):445–448
10. Rudin SP, Jones MD, Albers RC (2004) Thermal stabilization of the HCP phase in titanium. *Phys Rev B* 69(9):094117
11. Souvatzis P, Eriksson P, Katsnelson MI, Rudin SP (2008) Entropy driven stabilization of energetically unstable crystal structures explained from first principles theory. *Phys Rev Lett* 100:095901
12. Woodley MS, Battle DP, Gale DJ, Catlow RAC (1999) The prediction of inorganic crystal structures using a genetic algorithm and energy minimisation. *Phys Chem Chem Phys* 1:2535–2542
13. Stillinger FH (1999) Exponential multiplicity of inherent structures. *Phys Rev E* 59(1):48
14. Venkatesh PK, Cohen MH, Carr RW, Dean AM (1997) Bayesian method for global optimization. *Phys Rev E* 55:6219–6232
15. Massen CP, Doye JP (2007) Power-law distributions for the areas of the basins of attraction on a potential energy landscape. *Phys Rev E* 75(3):037101
16. Oganov AR, Glass CW, Ono S (2006) High-pressure phases of CaCO₃: crystal structure prediction and experiment. *Earth Planet Sci Lett* 241(1):95–103
17. Brodmeier T, Pretsch E (1994) Application of genetic algorithms in molecular modeling. *J Comput Chem* 15(6):588–595
18. Dandekar T, Argos P (1994) Folding the main chain of small proteins with the genetic algorithm. *J Mol Biol* 236(3):844
19. Lucasius CB, Werten S, van Aert A, Kateman G, Blommers MJ (1991) Conformational analysis of DNA using genetic algorithms. In: Schwefel HP, Maenner R (eds) *Parallel problem solving from nature*. Springer, Berlin, pp 90–97
20. McGarragh D, Judson RS (1993) Analysis of the genetic algorithm method of molecular conformation determination. *J Comput Chem* 14(11):1385–1395
21. Sun S (1993) Reduced representation model of protein structure prediction: statistical potential and genetic algorithms. *Protein Sci* 2(5):762–785

22. Deaven DM, Ho KM (1995) Molecular geometry optimization with a genetic algorithm. *Phys Rev Lett* 75:288–291
23. Xiao Y, Williams DE (1993) Genetic algorithm: a new approach to the prediction of the structure of molecular clusters. *Chem Phys Lett* 215(1):17–24
24. Smith RW (1992) Energy minimization in binary alloy models via genetic algorithms. *Comput Phys Commun* 71(1):134–146
25. Oganov AR, Glass CW (2006) Crystal structure prediction using ab initio evolutionary techniques: principles and applications. *J Chem Phys* 124(24):244704
26. Tipton WW, Hennig RG (2013) Genetic algorithm for structure and phase prediction. Cornell University. <http://gasp.mse.cornell.edu/>. Accessed 22 Oct 2013
27. Tipton WW, Hennig RG (2013) A grand canonical genetic algorithm for the prediction of multicomponent phase diagrams and testing empirical potentials. *J Phys Cond Matter* 25:495401
28. Pickard CJ, Needs RJ (2006) High-pressure phases of silane. *Phys Rev Lett* 97:045504
29. Zhu Q, Oganov AR, Glass CW, Stokes HT (2012) Constrained evolutionary algorithm for structure prediction of molecular crystals: methodology and applications. *Acta Crystallogr Sect B Struct Sci* 68(3):215–226
30. Bahmann S, Kortus J (2013) EVO – evolutionary algorithm for crystal structure prediction. *Comput Phys Commun* 184(6):1618–1625
31. Lonie DC, Zurek E (2011) Xtalopt: an open-source evolutionary algorithm for crystal structure prediction. *Comput Phys Commun* 182(2):372–387
32. Glass CW, Oganov AR, Hansen N (2006) USPEX – evolutionary crystal structure prediction. *Comput Phys Commun* 175(1112):713–720
33. Trimarchi G, Zunger A (2007) Global space-group optimization problem: finding the stablest crystal structure without constraints. *Phys Rev B* 75:104113
34. Ji M, Wang CZ, Ho KM (2010) Comparing efficiencies of genetic and minima hopping algorithms for crystal structure prediction. *Phys Chem Chem Phys* 12(37):11617–11623
35. Woodley SM (2004) Prediction of inorganic crystal framework structures part 2: using a genetic algorithm and a direct approach to exclusion zones. *Phys Chem Chem Phys* 6:1823–1829
36. Křivý I, Gruber B (1976) A unified algorithm for determining the reduced (Niggli) cell. *Acta Crystallogr A* 32(2):297–298
37. Oganov AR, Glass CW (2008) Evolutionary crystal structure prediction as a tool in materials design. *J Phys Condens Matter* 20(6):064210
38. Abraham NL, Probert MIJ (2006) A periodic genetic algorithm with real-space representation for crystal structure and polymorph prediction. *Phys Rev B* 73:224104
39. Johnston RL (2003) Evolving better nanoparticles: genetic algorithms for optimising cluster geometries. *Dalton Trans* 4193–4207
40. Lyakhov AO, Oganov AR, Valle M (2010) How to predict very large and complex crystal structures. *Comput Phys Commun* 181(9):1623–1632
41. Oganov AR (2013) Evolutionary crystal structure prediction and computational materials discovery. doi: [10.1007/128_2013_508](https://doi.org/10.1007/128_2013_508)
42. Lonie DC, Zurek E (2012) Identifying duplicate crystal structures: Xtalcomp, an open-source solution. *Comput Phys Commun* 183(3):690–697
43. Valle M, Oganov AR (2008) Crystal structures classifier for an evolutionary algorithm structure predictor. In: IEEE symposium on visual analytics science and technology VAST'08, 2008. Columbus, Ohio, pp 11–18
44. Valle M, Oganov AR (2010) Crystal fingerprint space – a novel paradigm for studying crystalstructure sets. *Acta Crystallogr A* 66(5):507–517
45. Abraham NL, Probert MIJ (2008) Improved real-space genetic algorithm for crystal structure and polymorph prediction. *Phys Rev B* 77:134117
46. Hartke B (1993) Global geometry optimization of clusters using genetic algorithms. *J Phys Chem* 97(39):9973–9976

47. Zhang X, Trimarchi G, Zunger A (2009) Possible pitfalls in theoretical determination of groundstate crystal structures: the case of platinum nitride. *Phys Rev B* 79:092102
48. Tipton WW, Bealing CR, Mathew K, Hennig RG (2013) Structures, phase stabilities, and electrical potentials of Li-Si battery anode materials. *Phys Rev B* 87:184114
49. Trimarchi G, Freeman AJ, Zunger A (2009) Predicting stable stoichiometries of compounds via evolutionary global space-group optimization. *Phys Rev B* 80:092101
50. Zurek E, Hoffmann R, Ashcroft NW, Oganov AR, Lyakhov AO (2009) A little bit of lithium does a lot for hydrogen. *Proc Natl Acad Sci U S A* 106(42):17640–17643
51. Woodley SM, Catlow CRA (2009) Structure prediction of titania phases: implementation of Darwinian versus Lamarckian concepts in an evolutionary algorithm. *Comput Mater Sci* 45(1):84–95
52. Park H, Feller MR, Lenosky TJ, Tipton WW, Trinkle DR, Rudin SP, Woodward C, Wilkins JW, Hennig RG (2012) Ab initio based empirical potential used to study the mechanical properties of molybdenum. *Phys Rev B* 85:214121
53. Ji M, Umemoto K, Wang CZ, Ho KM, Wentzcovitch RM (2011) Ultrahigh-pressure phases of H₂O ice predicted using an adaptive genetic algorithm. *Phys Rev B* 84:220105
54. Kolmogorov AN, Shah S, Margine ER, Bialon AF, Hammerschmidt T, Drautz R (2010) New superconducting and semiconducting Fe–B compounds predicted with an ab initio evolutionary search. *Phys Rev Lett* 105(21):217003
55. Chuang F, Ciobanu CV, Shenoy V, Wang CZ, Ho KM (2004) Finding the reconstructions of semiconductor surfaces via a genetic algorithm. *Surf Sci* 573(2):L375–L381
56. Lu N, Ciobanu CV, Chan TL, Chuang FC, Wang CZ, Ho KM (2007) The structure of ultrathin H-passivated [112] silicon nanowires. *J Phys Chem C* 111(22):7933–7937
57. Woodley SM (2007) Engineering microporous architectures: combining evolutionary algorithms with predefined exclusion zones. *Phys Chem Chem Phys* 9(9):1070–1077
58. Oganov AR, Chen J, Gatti C, Ma Y, Ma Y, Glass CW, Liu Z, Yu T, Kurakevych OO, Solozhenko VL (2009) Ionic high-pressure form of elemental boron. *Nature* 457(7231):863–867
59. Taillon JA, Tipton WW, Hennig RG (2012) Ab initio discovery of novel crystal structure stability in barium and sodium-calcium compounds under pressure using DFT. arXiv preprint arXiv:1207.3320
60. Li Q, Ma Y, Oganov AR, Wang H, Wang H, Xu Y, Cui T, Mao HK, Zou G (2009) Superhard monoclinic polymorph of carbon. *Phys Rev Lett* 102(17):175506
61. Oganov AR, Lyakhov AO (2010) Towards the theory of hardness of materials. *J Superhard Mater* 32(3):143–147
62. Bi W, Meng Y, Kumar RS, Cornelius AL, Tipton WW, Hennig RG, Zhang Y, Chen C, Schilling JS (2011) Pressure-induced structural transitions in europium to 92 GPa. *Phys Rev B* 83(10):104106
63. Ma Y, Oganov AR, Xie Y (2008) High-pressure structures of lithium, potassium, and rubidium predicted by an ab initio evolutionary algorithm. *Phys Rev B* 78(1):014102
64. Ma Y, Oganov AR, Li Z, Xie Y, Kotakoski J (2009) Novel high pressure structures of polymeric nitrogen. *Phys Rev Lett* 102(6):065501
65. Ma Y, Eremets M, Oganov AR, Xie Y, Trojan I, Medvedev S, Lyakhov AO, Valle M, Prakapenka V (2009) Transparent dense sodium. *Nature* 458(7235):182–185
66. Ma Y, Oganov AR, Glass CW (2007) Structure of the metallic ζ -phase of oxygen and isosymmetric nature of the ε - ζ phase transition: ab initio simulations. *Phys Rev B* 76(6):064101
67. Hooper J, Zurek E (2012) Lithium subhydrides under pressure and their superatom-like building blocks. *ChemPlusChem* 77(11):969–972
68. Baettig P, Zurek E (2011) Pressure-stabilized sodium polyhydrides: NaH_n ($n > 1$). *Phys Rev Lett* 106(23):237002
69. Hooper J, Zurek E (2012) High pressure potassium polyhydrides: a chemical perspective. *J Phys Chem C* 116(24):13322–13328

70. Hooper J, Zurek E (2012) Rubidium polyhydrides under pressure: emergence of the linear H_3^- species. *Chemistry* 18(16):5013–5021
71. Shamp A, Hooper J, Zurek E (2012) Compressed cesium polyhydrides: Cs^+ sublattices and H_3^- three-connected nets. *Inorg Chem* 51(17):9333–9342
72. Hooper J, Altintas B, Shamp A, Zurek E (2013) Polyhydrides of the alkaline earth metals: a look at the extremes under pressure. *J Phys Chem C* 117(6):2982–2992
73. Lonie DC, Hooper J, Altintas B, Zurek E (2013) Metallization of magnesium polyhydrides under pressure. *Phys Rev B* 87(5):054107
74. Labet V, Hoffmann R, Ashcroft NW (2011) Molecular models for WH_6 under pressure. *New J Chem* 35(10):2349–2355
75. Gao G, Oganov AR, Ma Y, Wang H, Li P, Li Y, Iitaka T, Zou G (2010) Dissociation of methane under high pressure. *J Chem Phys* 133:144508
76. Martinez-Canales M, Oganov AR, Ma Y, Yan Y, Lyakhov AO, Bergara A (2009) Novel structures and superconductivity of silane under pressure. *Phys Rev Lett* 102(8):087005
77. Gao G, Oganov AR, Bergara A, Martinez-Canales M, Cui T, Iitaka T, Ma Y, Zou G (2008) Superconducting high pressure phase of germane. *Phys Rev Lett* 101(10):107002
78. Gao G, Oganov AR, Li P, Li Z, Wang H, Cui T, Ma Y, Bergara A, Lyakhov AO, Iitaka T et al (2010) High-pressure crystal structures and superconductivity of stannane (SnH_4). *Proc Natl Acad Sci U S A* 107(4):1317–1320
79. Zaleski-Ejgierd P, Hoffmann R, Ashcroft N (2011) High pressure stabilization and emergent forms of PbH_4 . *Phys Rev Lett* 107(3):037002
80. Wen XD, Hand L, Labet V, Yang T, Hoffmann R, Ashcroft N, Oganov AR, Lyakhov AO (2011) Graphane sheets and crystals under pressure. *Proc Natl Acad Sci U S A* 108(17):6833–6837
81. Zhou XF, Oganov AR, Dong X, Zhang L, Tian Y, Wang HT (2011) Superconducting highpressure phase of platinum hydride from first principles. *Phys Rev B* 84(5):054543
82. Bazhanova ZG, Oganov AR, Gianola O (2012) Fe–C and Fe–H systems at pressures of the Earth’s inner core. *Phys Uspekhi* 55(5):489
83. Hu CH, Oganov AR, Lyakhov AO, Zhou HY, Hafner J (2009) Insulating states of LiBeH_3 under extreme compression. *Phys Rev B* 79(13):134116
84. Prasad DL, Ashcroft N, Hoffmann R (2012) Lithium amide (LiNH_2) under pressure. *J Phys Chem A* 116(40):10027–10036
85. Zhou XF, Dong X, Zhao Z, Oganov AR, Tian Y, Wang HT (2012) High-pressure phases of NaAlH_4 from first principles. *Appl Phys Lett* 100(6):061905–061905
86. Wen XD, Cahill TJ, Gerovac NM, Bucknum MJ, Hoffmann R (2009) Playing the quantum chemical slot machine: an exploration of ABX_2 compounds. *Inorg Chem* 49(1):249–260
87. Zhong Y, Zhou HY, Hu CH, Wang DH, Oganov AR (2012) Theoretical studies of highpressure phases, electronic structure, and vibrational properties of NaNH_2 . *J Phys Chem C* 116(15):8387–8393
88. Zhou XF, Oganov AR, Qian GR, Zhu Q (2012) First-principles determination of the structure of magnesium borohydride. *Phys Rev Lett* 109(24):245503
89. Trimarchi G, Zunger A (2008) Finding the lowest-energy crystal structure starting from randomly selected lattice vectors and atomic positions: first-principles evolutionary study of the Au–Pd, Cd–Pt, Al–Sc, Cu–Pd, Pd–Ti, and Ir–N binary systems. *J Phys Condens Matter* 20(29):295212
90. Xie Y, Oganov AR, Ma Y (2010) Novel high pressure structures and superconductivity of CaLi_2 . *Phys Rev Lett* 104(17):177005
91. Zhu Q, Oganov AR, Lyakhov AO (2012) Evolutionary metadynamics: a novel method to predict crystal structures. *CrystEngComm* 14(10):3596–3601
92. Ono S, Oganov AR, Brodholt JP, Vočadlo L, Wood IG, Lyakhov A, Glass CW, Côté AS, Price GD (2008) High-pressure phase transformations of FeS: novel phases at conditions of planetary cores. *Earth Planet Sci Lett* 272(1):481–487

93. Zhang F, Oganov AR (2010) Iron silicides at pressures of the earth's inner core. *Geophys Res Lett* 37(2), L02305
94. Zhu Q, Oganov AR, Lyakhov AO (2012) Unexpected stoichiometries in Mg-O system under high pressure. arXiv preprint arXiv:1211.6521
95. Zhang W, Oganov AR, Goncharov AF, Zhu Q, Bouffelfel SE, Lyakhov AO, Somayazulu M, Prakapenka VB (2012) Unexpected stable stoichiometries of sodium chlorides. arXiv preprint arXiv:1211.3644
96. Wu S, Umemoto K, Ji M, Wang CZ, Ho KM, Wentzcovitch RM (2011) Identification of post-pyrite phase transitions in SiO₂ by a genetic algorithm. *Phys Rev B* 83:184102
97. Hermann A, Ashcroft N, Hoffmann R (2012) Making sense of boron-rich binary Be-B phases. *Inorg Chem* 51(16):9066–9075
98. Xu Y, John ST, Oganov AR, Cui T, Wang H, Ma Y, Zou G (2009) Superconducting high-pressure phase of cesium iodide. *Phys Rev B* 79(14):144110
99. Hermann A, McSorley A, Ashcroft NW, Hoffmann R (2012) From Wade-Mingos to Zintl-Klemm at 100 GPa: binary compounds of boron and lithium. *J Am Chem Soc* 134(45):18606–18618
100. Hermann A, Suarez-Alcubilla A, Gurtubay IG, Yang LM, Bergara A, Ashcroft NW, Hoffmann R (2012) LiB and its boron-deficient variants under pressure. *Phys Rev B* 86(14):144110
101. Ma Y, Wang Y, Oganov AR (2009) Absence of superconductivity in the high-pressure polymorph of MgB₂. *Phys Rev B* 79(5):054101
102. Wang H, Li Q, Li Y, Xu Y, Cui T, Oganov AR, Ma Y (2009) Ultra-incompressible phases of tungsten dinitride predicted from first principles. *Phys Rev B* 79(13):132109
103. Kurzydłowski D, Zaleski-Ejgierd P, Grochala W, Hoffmann R (2011) Freezing in resonance structures for better packing: XeF₂ becomes (XeF₊)(F₋) at large compression. *Inorg Chem* 50(8):3832–3840
104. Zhu Q, Jung DY, Oganov AR, Glass CW, Gatti C, Lyakhov AO (2012) Stability of xenon oxides at high pressures. *Nat Chem* 5(1):61–65
105. Wen XD, Cahill TJ, Hoffmann R, Miura A (2009) Tuning of metal-metal bonding by counterion size in hypothetical AeTiO₂ compounds. *J Am Chem Soc* 131(41):14632–14633
106. Shirako Y, Kojitani H, Oganov A, Fujino K, Miura H, Mori D, Inaguma Y, Yamaura K, Akaogi M (2012) Crystal structure of CaRhO₃ polymorph: high-pressure intermediate phase between perovskite and post-perovskite. *Am Mineral* 97(1):159–163
107. Li Q, Wang M, Oganov AR, Cui T, Ma Y, Zou G (2009) Rhombohedral superhard structure of BC₂N. *J Appl Phys* 105(5):053514–053514
108. Hermann A, Ivanov B, Ashcroft NW, Hoffmann R (2012) LiBeB: a predicted phase with structural and electronic peculiarities. *Phys Rev B* 86(1):014104
109. Bi W, Souza-Neto NM, Haskel D, Fabbri G, Alp EE, Zhao J, Hennig RG, Abd-Elmeguid MM, Meng Y, McCallum RW, Dennis K, Schilling JS (2012) Synchrotron X-ray spectroscopy studies of valence and magnetic state in europium metal to extreme pressures. *Phys Rev B* 85:205134
110. Ashcroft NW (1968) Metallic hydrogen: a high-temperature superconductor? *Phys Rev Lett* 21:1748–1749
111. Feng J, Grochala W, Jaroń T, Hoffmann R, Bergara A, Ashcroft N (2006) Structures and potential superconductivity in SiH₄ at high pressure: en route to metallic hydrogen. *Phys Rev Lett* 96(1):017006
112. dAvezac M, Zunger A (2008) Identifying the minimum-energy atomic configuration on a lattice: Lamarckian twist on Darwinian evolution. *Phys Rev B* 78(6):064102
113. Agrawal J (1998) Recent trends in high-energy materials. *Prog Energ Combust* 24(1):1–30
114. Baburin I, Leoni S, Seifert G (2008) Enumeration of not-yet-synthesized zeolitic zinc imidazolate MOF networks: a topological and DFT approach. *J Phys Chem B* 112(31):9437–9443

115. Price SL (2004) The computational prediction of pharmaceutical crystal structures and polymorphism. *Adv Drug Deliv Rev* 56(3):301–319
116. Hermann A, Ashcroft NW, Hoffmann R (2012) High pressure ices. *Proc Natl Acad Sci U S A* 109(3):745–750
117. Johannesson GH, Bligaard T, Ruban AV, Skriver HL, Jacobsen KW, Nørskov JK (2002) Combined electronic structure and evolutionary search approach to materials design. *Phys Rev Lett* 88(25):255506