DEVELOPMENT OF AN EVOLUTIONARY ALGORITHM FOR THE AB INITIO DISCOVERY OF TWO-DIMENSIONAL MATERIALS

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Benjamin Charles Revard

August 2017

DEVELOPMENT OF AN EVOLUTIONARY ALGORITHM FOR THE AB INITIO DISCOVERY OF TWO-DIMENSIONAL MATERIALS

Benjamin Charles Revard, Ph.D.

Cornell University 2017

Crystal structure prediction is an important first step on the path toward computational materials design. Increasingly robust methods have become available in recent years for computing many materials properties, but because properties are largely a function of crystal structure, the structure must be known before these methods can be brought to bear. In addition, structure prediction is particularly useful for identifying low-energy structures of subperiodic materials, such as two-dimensional (2D) materials, which may adopt unexpected structures that differ from those of the corresponding bulk phases. Evolutionary algorithms, which are heuristics for global optimization inspired by biological evolution, have proven to be a fruitful approach for tackling the problem of crystal structure prediction. This thesis describes the development of an improved evolutionary algorithm for structure prediction and several applications of the algorithm to predict the structures of novel low-energy 2D materials.

The first part of this thesis contains an overview of evolutionary algorithms for crystal structure prediction and presents our implementation, including details of extending the algorithm to search for clusters, wires, and 2D materials, improvements to efficiency when running in parallel, improved composition space sampling, and the ability to search for partial phase diagrams. We then present

several applications of the evolutionary algorithm to 2D systems, including InP, the C-Si and Sn-S phase diagrams, and several group-IV dioxides.

This thesis makes use of the Cornell graduate school's "papers" option. Chapters 1 and 3 correspond to the first-author publications of Refs. [131] and [132], respectively, and chapter 2 will soon be submitted as a first-author publication. The material in chapter 4 is taken from Ref. [144], in which I share joint first-authorship. In this case I have included only my own contributions.

BIOGRAPHICAL SKETCH

Benjamin Revard was born in Bellevue, WA on August 1, 1989. He was homeschooled by his parents, and he took several classes at the Treasure Valley Community College during his last two years of high school. In 2007, he began his undergraduate studies at Washington State University, where he met Katie Brunell. Ben majored in materials science and engineering and graduated in 2012.

During his final year as an undergraduate, Ben became increasingly interested in the computational side of materials science. After graduating, he drove from Pullman, WA to Ithaca, NY to pursue this interest under the supervision of Richard Hennig at Cornell University. This thesis describes much of his graduate work as a member of the Hennig group. Ben will move to Oregon after graduating to join the computational lithography effort at Intel.

To my parents.

ACKNOWLEDGEMENTS

First I would like to thank Professor Richard Hennig for his outstanding mentorship. His patience and encouragement made all the difference, especially toward the beginning as I was learning how to code and struggling with my projects. Whatever obstacle I faced, I could always count on his support and perennial optimism. He also provided me a great deal of flexibility in my research, which allowed me to pursue my interests and gain a range of experiences and skills.

I would like to thank Professors R. Bruce van Dover and David Bindel for serving on my committee, and Sven Rudin for his mentorship during my summer at Los Alamos. I would also like to thank Richard Hennig, Sven Rudin and Susan Sinnott for being my references and taking the time to write letters for me during my job search.

The fellow students I have met in graduate school have made the time fun as well as academically stimulating. I would like to thank Will for teaching me about genetic algorithms and for his patience, and Arunima for her kindness and valuable collaborations. I am grateful to Kiran for many interesting late night discussions, and also for his sound advice. David, Joe and Ben were great study partners during my first year at Cornell, and I have many happy memories with them. I have also had the pleasure of getting to know several students at UF, including Michael, Biswas, and both Joshuas, and I wish them all the best in their future pursuits.

My family has been a constant source of support and encouragement. I am

forever indebted to my parents for the countless hours they invested in my early education and for their unparalleled dedication to create a rich and personalized learning environment. This Ph.D. would not have been possible without the foundation they provided.

Finally, I cannot thank Katie enough for putting up with me this whole time. She has been amazingly patient as I traveled around the country for my graduate studies, and I am forever grateful for her love and support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

STRUCTURE AND STABILITY PREDICTION OF COMPOUNDS WITH EVOLUTIONARY ALGORITHMS

The content of this chapter was previously published in Ref. [131].

1.1 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 [169], 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 to address this problem in recent years [125, 176, 103, 122, 42, 23, 22]. These techniques are often faster and less expensive than experimental work, they preclude 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 or-

der 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 entropy. The vibrational and configurational components are expensive to calculate, and much of the error introduced by neglecting the entropy cancels when taking energy differences [135, 35]. For these reasons, 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 [146]. 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 [22, 180]. 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.

1.1.1 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 follow.

- Much of the configuration space corresponds to structures with unphysically-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 [149]. 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 [170].

- 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 [105].
- 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 [125].
- Low-energy minima in the landscape usually correspond to symmetrical structures [125].
- 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 [117].

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

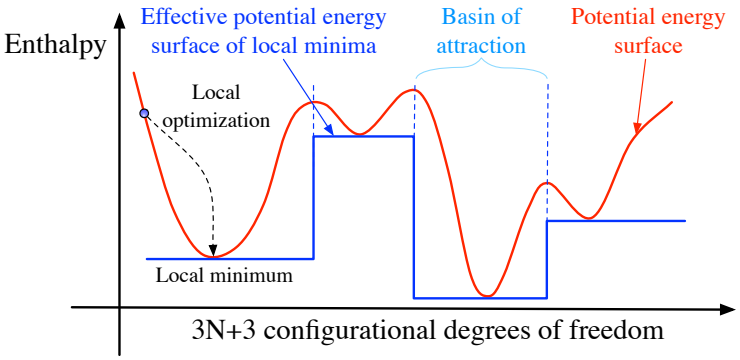

Figure 1.1: Potential energy surface. The use of local optimization simplifies the search problem by dividing the continuous solution space into basins of attraction.

from biological evolution. Efficient local optimization utilizes the derivative information and is very beneficial in the solution of the optimization problem (see Section 1.4). Figure 1.1 illustrates how local optimization transforms the continuous potential energy landscape into a discrete set of basins of attractions, which dramatically simplifies the search space.

1.1.2 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.

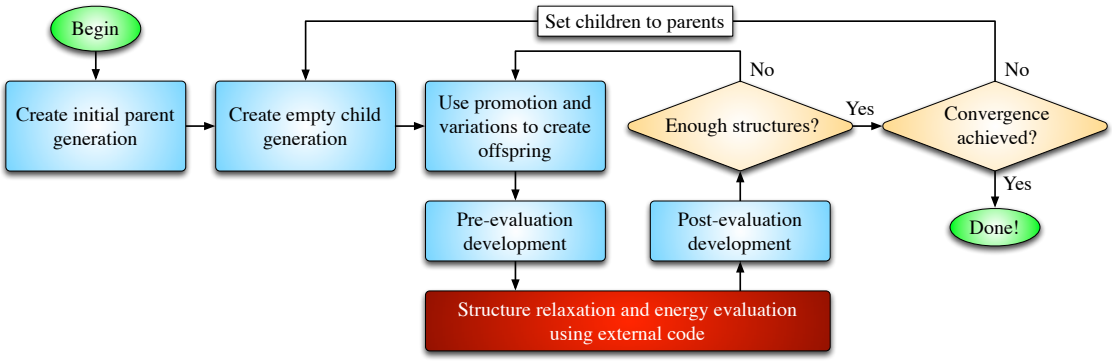

Figure 1.2: Outline of evolutionary algorithm for structure prediction.

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 [92, 151, 106, 26, 21]. Evolutionary search

techniques were also successfully applied to atomic clusters [187, 30], and soon the method was extended to 3-D periodic systems [145, 22, 180].

It has been observed that evolutionary algorithms (EAs) are well suited to the structure prediction problem for several reasons [115]. First, they can efficiently find the global minimum of multidimensional functions. EAs require little information or 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 Section 1.1.1.

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 stage. The algorithm terminates when some user-defined stopping criteria are met (see Section 1.2.13).

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 number 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.

Figure 1.2 outlines how a typical evolutionary algorithm for structure prediction proceeds, such as the one implemented by Tipton et al. [159]. The EA starts by creating an initial population (Section 1.2.2) and calculating the fitness of each organism in it (Section 1.2.3). Organisms are then selected to act as parents (Section 1.2.4) or to be promoted to the next generation (Section 1.2.5). The parents create offspring structures via mating (Section 1.2.6) or mutation (Section 1.2.7). The offspring are developed, (Section 1.2.9) 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 (Section 1.2.13), the current children become parents in the next generation.

1.2 Details of the method

1.2.1 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 Figure 1.3. 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 [126, 204]. 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.

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 [159, 89, 10]. In addition, most authors constrain the range of angles between the lattice vectors [159, 89, 10, 164, 41]. If the algorithm varies the number of atoms per cell, this value is also constrained [159]. A restriction on the total volume of the cell is an additional possibility [89, 63, 10].

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 [164, 10]. The minimum lattice length has been chosen by adding the typical bond length and the diameter of the largest atom in the system [164, 41, 10]. Bahmann *et al.* set the maximum lattice vector length to the sum of the covalent diameters of all atoms in the cell [10]. Several authors limit angles between lattice vectors to lie between 60 and 120 degrees [89, 41], although a more liberal range of 45 to 135 degrees has also been used [164, 10]. Ji *et al.* fix the volume of the cell during the search [63], 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 [89]. 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 [10].

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 [10]. Woodley *et al.* use an EA to search for nan oporous materials by incorporating "exclusion zones", or regions in which atoms are forbidden to reside [182]. 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 Section 1.1.1, 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

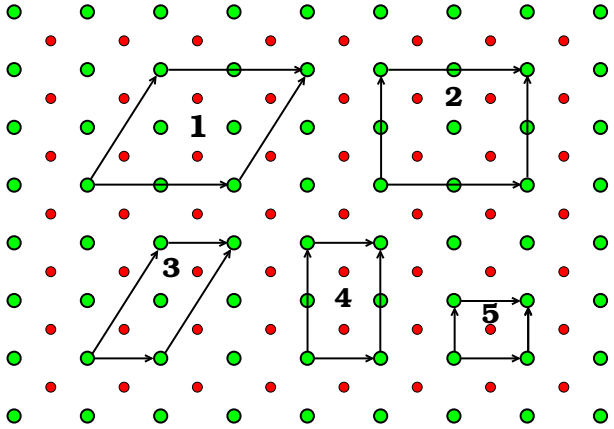

Figure 1.3: Five alternate representations of a single physical crystal are shown. Cells 2 and 4 are Niggli reduced versions of cells 1 and 3, respectively. They are also 2×2 and 1×2 supercells, respectively, of the primitive cell 5.

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 [76, 159]. There is a Niggli cell for any lattice that is both unique and has the shortest possible lattice lengths. Figure 1.3 illustrates the representation problem and the Niggli cell reduction. A similar transformation is used by Lonie *et al.* and Oganov *et al.* [89, 116]. 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 Section 1.2.6.

1.2.2 Initial population

If no experimental data is 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 Section 1.3), the correct elemental and binary phases may be already known experimentally and can be used in the initial population. The use of pre-existing knowledge has the potential to significantly decrease 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 [126]. Zhu *et al.* made an additional modification to the generation of the the initial population to facilitate their study of molecular solids [204]. 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.

1.2.3 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 Section 1.2.4) [1, 65]. In one frequently-applied scheme, an organism with a formation energy per atom, E_f, is assigned a fitness

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

where $E_\mathrm{f}^\mathrm{max}$ and $E_\mathrm{f}^\mathrm{min}$ are the highest and lowest formation energies per atom, respectively, in the generation [159, 89, 1]. 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 [63, 41]. 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 [164, 10].

1.2.4 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 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 [65]. In roulette wheel selection, a random number d between the fitnesses of the best and worst organisms fitness is generated for each organism, and if d is less than the fitness of the organism, it is allowed to reproduce [1]. 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 [159]. Two parameters are used to 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

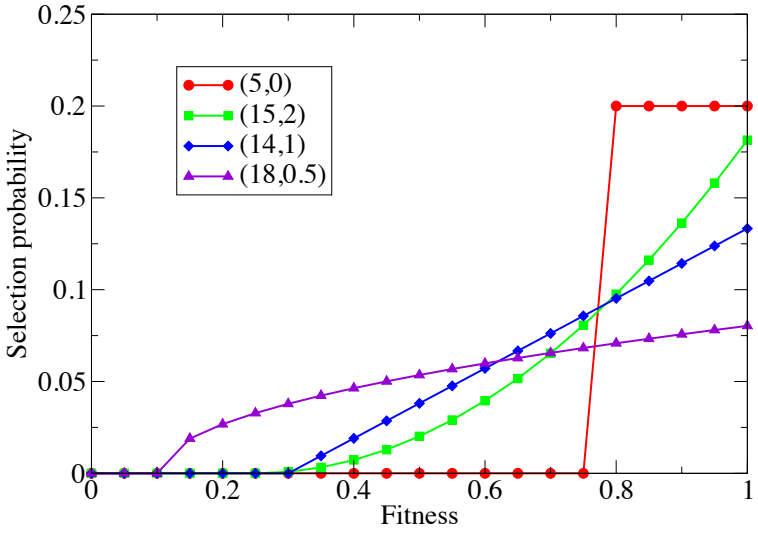

Figure 1.4: Tuning the selection probability distribution (N, p) described by the number of potential parents N and power law p allows Tipton *et al.* to adjust the aggressiveness with which an EA seeks to converge.

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 Figure 1.4.

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 [164, 63, 10] or with a linear or quadratic probability distribution over their fitnesses [41]. 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 [1]. 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 [89].

1.2.5 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 Section 1.2.11 for fingerprinting) [95]. This was done to prevent loss of population diversity due to promoting similar organisms.

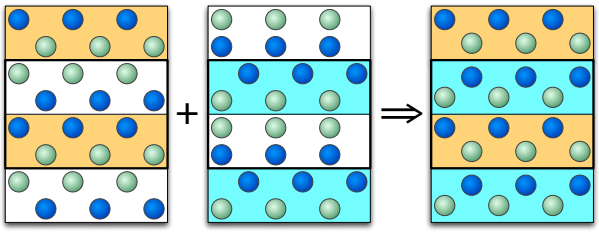

Figure 1.5: Schematic illustration of how the mating operator in the evolutionary algorithm combines two crystal structures. For clarity, 1×2 supercells of the child and parent structures are shown.

1.2.6 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 Figure 3.3.

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-significant 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.

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 frac tional 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 [89, 41]. 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 [89].

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 [164, 63, 1]. 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 [159]. Another approach is to simply fix the locations of the two slice planes. For example, Abraham *et al.* specify that the two cuts be made at fractional coordinates of $\frac{1}{4}$ and $\frac{3}{4}$ along the chosen lattice vector [1].

Abraham *et al.* introduced a periodic slicing operation [1]. 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 [1].

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 [89, 164, 41]. Simply averaging the lattice vectors of the parent organisms, i.e. fixing the weight at 0.5, is another common choice [159, 63].

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 to simply reject all offspring organisms that do not meet the desired constraints [164, 1]. 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 [63]. These corrections minimize disruption to the structure transmitted from the parent organisms. Glass *et al.* use a slightly different approach: atoms to be removed are selected randomly, and atoms to be added are selected randomly from the discarded fragments of one of the parent organisms [89, 41]. Atomic order parameters have also been used to decide which excess atoms to remove (see Section 1.2.11) [95]. Those with the lower degrees of local order are more likely to be removed.

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 [41, 113]. 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 [89]. 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 [89]. Additionally, an order parameter may be used to inform the choice of contribution from each parent (see Section 1.2.11). Several trial slabs of equal thickness are cut from the parents at random locations, and the slabs with the highest degree of order are passed to the offspring [95, 113].

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 [204]. 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.

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 Section 1.2.1 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 [159]. 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 Section 1.2.10) 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 [95, 113].

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 [10]. 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 [145]. 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 [64]. Although 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.

1.2.7 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 off-

spring organism. Some approaches call for mutating both the lattice vectors and the atomic positions [159, 164], 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

$$S = \begin{pmatrix} 1 + e_1 & e_6/2 & e_5/2 \\ e_6/2 & 1 + e_2 & e_4/2 \\ e_5/2 & e_4/2 & 1 + e_3 \end{pmatrix}, \tag{1.1}$$

where the components e_i are taken from uniform or Gaussian distributions [89, 164, 41].

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 [159]. 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 [1]. 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 [41]. However, in later work, Lyakhov *et al.* use a "smart" mutation operation where atoms with low-order local environments (see

Section 1.2.11), are shifted more [95, 113]. A further refinement to mutation has been made by shifting all the atoms along the eigenvector of the softest phonon mode [95, 113].

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 [159, 116]. 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 it 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 over all possible ways to place the atoms on the atomic sites [164].

Lonie *et al.* employ a "ripple" variation, in which all atoms in the cell are

shifted by varying amounts [89]. 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 [89].

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 [204].

1.2.8 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 can not 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 to simply "guess" the correct value of N [89, 164, 41]. 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 [63, 1]. 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 [149], 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 [95, 113]. To obtain reasonably good large cells for the initial population, Lyakhov *et al.* generate smaller random cells of 15 to 20 atoms, and then take supercells of these [95, 113]. 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 [159]. 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.

1.2.9 Development and screening

After a new organism has been created by one of the variations, it is checked against the constraints described in Section 1.2.1 and tested for redundancy (see Section 1.2.10). At this stage, many EAs scale the atomic density of the new organisms using an estimate of the optimal density [159, 41]. 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.

1.2.10 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 be very similar also. In addition, as the generation as a whole converges to the global minimum, all of 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 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 Section 1.2.1, 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 if 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 [90].

Tipton *et al.* also use a direct comparison of structures, with a slight modification [159]. 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 [176]. 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 [10]. 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 [65, 30]. 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 [90]. Lonie *et al.* expanded the fingerprint function to include three parameters: energy, space group, and the volume of the cell [89]. 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 [90].

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 [167, 168, 113]. 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 Sec-

tion 1.2.11). 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 [204].

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 [2].

1.2.11 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 [168, 95, 113]. They used it to guide the algorithm at various points, as mentioned previously.

1.2.12 Frequency of promotion and variations

The user-specified parameters of an evolutionary algorithms 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 [164, 63, 41]. 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 [89]. 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 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 [159].

1.2.13 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 [47]. 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 versus 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.* employ a statistically more relevant approach to quantifying an EA's performance in which the median, the 10^{th} percentile, and the 90^{th} percentile energies of the best structures are plotted [159]. The 10^{th} and 90^{th} 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 1.6 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 [159]. These plots provide insights 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

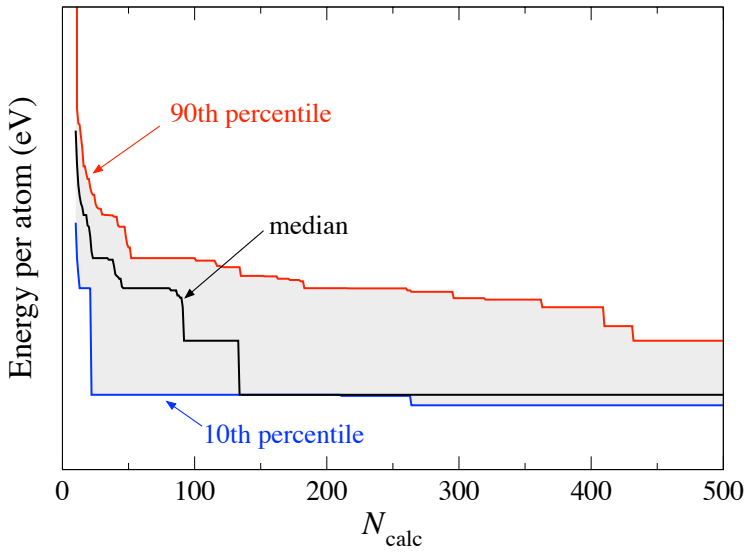

Figure 1.6: Performance distribution plot for 100 structure searches at fixed composition for Zr_2Cu_2Al using an embedded atom model potential [159]. The energy of the 90th percentile best structure is shown in red, the 10th percentile best structure in blue, and the median best structure in black.

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 [89]. 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 [159, 89, 10]. Stopping once an allocated amount of computational resources have 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 [10].

Although most authors use one of the fairly simple convergence criteria mentioned above, a quantitative statistical approach has been proposed by Venkatesh *et al.* [170]. 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.

1.3 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 [194]. First, as discussed in Section 1.2.8, 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 if 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.

In order to perform a phase diagram search, we make use of the convex hull construction [35]. The formation energies of all structures with respect to the elemental constituents are plotted versus 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 1.7 is an example of a convex hull for the Li-Si binary system [158].

Two approaches have been used to construct the convex hull. The first is to perform fixed-stoichiometry searches at many compositions [35, 215]. The lowest-energy structure found in each search is then placed on an energy versus composition plot, and the convex hull is constructed. However, this method is computationally expensive because it requires many separate searches to adequately

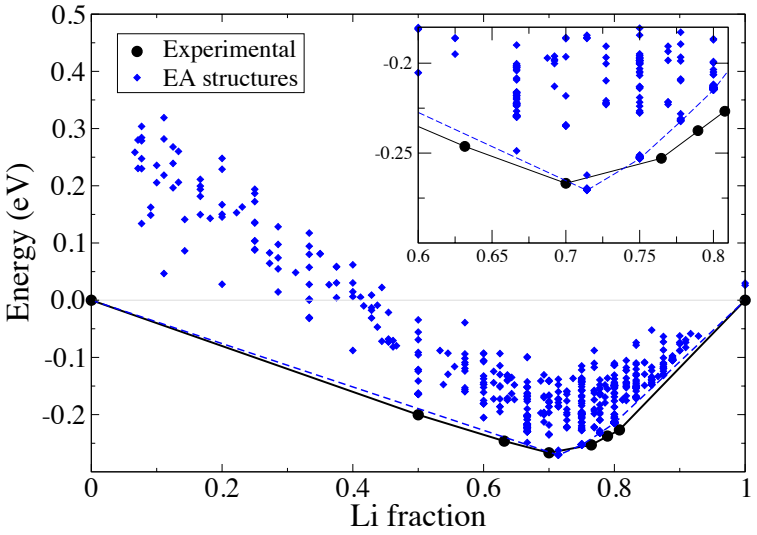

Figure 1.7: A phase diagram search of the Li-Si binary system by Tipton *et al.* [158] using the method of Trimarchi *et al.* [163] showed that a search for relatively small unit cell structures could approximate the structural and energetic characteristics of the known very large experimental structures 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.

sample the composition space [159].

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 by simply giving the initial population random stoichiometries and removing stoichiometry constraints on offspring structures [163]. 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 [159, 163]. 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 Section 1.2.4. As the search progresses, the true convex hull of the system is approached.

As discussed in Section 1.2.8, the global minimum can not 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 (Section 1.2.8) 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 [163]. 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 [159]. The first is to modify the selection criteria in such a way that mating between parents with similar compositions is encouraged. The second approach 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.

1.4 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 Section 1.1.1, the potential energy surface is divided into basins of attraction by the local structure optimization or relaxation available in most energy codes (see Figure 1.1). 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 find most of the local minima in it [125]. 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 [41]. 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 [181].

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 [123]. 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 misguiding 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 [125].

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 [62]. Lennard-Jones potentials were used for most of the energy evaluations, but *ab initio* calculations were performed periodically and the parameters of the Lennard-Jones potentials were fit 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.

1.5 Summary of methods

Tables 1.1 and 1.2 list the salient details of several implementations of evolutionary algorithms for structure prediction. The codes listed in Table 1.1 are production codes available to other users, the codes in Table 1.2 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 [159]. 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 Evolu-

Table 1.1: Comparison of the methods implemented into evolutionary algorithms in various available production codes.

Name Authors	GASP Tipton *et al.* [159]	XTALOPT Lonie *et al.* [89]	USPEX Glass *et al.* [41]	EVO Bahmann *et al.* [10]	MAISE Kolmogorov *et al.* [68]
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 [41], yes [95]	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 [89], direct comparison [90]	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 mutations, continuous scheme without generations	Use of order parameters	Age limit, alternative mating operation	

tionary 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 as soon as they are created [89]. 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 [41, 95]. 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 [10]. 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 [68].

Trimarchi *et al.* developed an evolutionary algorithm for structure prediction that is interfaced with VASP [164]. They later extended the algorithm to include phase diagram searching [163]. 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 [1]. 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 [63]. Bush *et al.* developed an evolutionary algorithm that incorporates a surrogate objective function [22].

1.6 Applications

Evolutionary algorithms have been used to solve the structures of many types of systems including molecules, clusters, surfaces, nanowires and nanoporous

Authors	Ji *et al.* [63]	Trimarchi *et al.* [164]	Bush *et al.* [22]	Abraham *et al.* [1]
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

Table 1.2: Comparison of the methods implemented into evolutionary algorithms in various research codes.

materials [65, 25, 91, 183]. Here we focus on applications of EAs to bulk, 3-D 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 inor-

ganic compounds. Tables 1.3 through 1.8 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.

1.6.1 Elemental solids

Table 1.3 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 of 19 - 89 GPa [114]. A search of carbon under high pressures led to the prediction that the bc8 structure is more stable than diamond above 1 TPa [115]. Oganov *et al.* predict several new superconducting phases of calcium at pressures up to 120 GPa [118]. A study of hydrogen at pressures up to 600 GPa predicted that it remained a molecular solid throughout this pressure range [115]. More on the interesting case of hydrogen under pressure will be 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) that has been experimentally observed for cesium, but predict a new cubic phase of lithium above 300 GPa [99]. 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 [96]. They have also reported a monoclinic, metallic, molecular phase of oxygen in the range between 100 and 250 GPa whose calculated XRD diffraction pattern is in agreement with experiment [97].

Bi *et al.* searched for phases of europium at pressures up to 100 GPa and predicted several nearly degenerate structures in the range of 16 to 45 GPa, which may help explain the mixed phase structure observed experimentally in this pressure regime [13] and the occurrence of superconductivity and magnetism in these phases [14]. 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 [123].

1.6.2 Hydrogen containing compounds

Table 1.4 summarizes EA structure searches performed on hydrogen containing compounds. Ashcroft suggested in 1968 that hydrogen could become a high-temperature superconductor under pressure [4] 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 [34].

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

was predicted to be stable above 110 GPa [215], and a stable phase of NaH_9 was predicted to metallize at 250 GPa [9]. Stable rubidium polyhydride phases were predicted to metallize at pressures above 200 GPa [57]. A stable, superconducting phase of MgH_{12} was identified under pressure, with a predicted T_c of 47-60 K at 140 GPa [88]. Lonie *et al.* also identified a new phase of BeH_2 above 150 GPa, as well as a stable superconducting phase of BaH_6, with a T_c of 30-38 K at 100 GPa [54].

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 [39]. 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 [102]. 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 [37]. 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 [38]. Zaleski *et al.* performed evolutionary structure searches for plumbane (PbH_4) under pressure and predicted that it forms a stable non-molecular solid at pressures greater than 132 GPa [191]. Wen *et al.* predicted five low energy three-dimensional structures of graphane in the pressure range of 0 - 300 GPa, and each was either

semiconducting or insulating [179].

Zhou *et al.* investigated platinum hydrides under pressure and predicted a superconducting hexagonal phase of PtH to be stable above 113 GPa [200]. Hu *et al.* searched for $LiBeH_3$ at pressures up to 530 GPa and predicted two new insulating phases [58]. Zhou *et al.* found two new tetragonal structures of $Mg(BH_4)_2$ under pressure whose densities, bulk moduli, and XRD patterns match experimentally measured values [201].

1.6.3 Intermetallic compounds

Table 1.5 summarizes searches for intermetallic compounds. Trimarchi *et al.* have studied many intermetallics. Their algorithm identified the correct lattice of Au_2Pd, which is known to be fcc, but it failed to find the lowest energy atomic configuration because the system exhibits several nearly degenerate structures [164]. In another study by these authors, a new phase of IrN_2 was discovered [165]. 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 [163]. Xie *et al.* discovered two new superconducting phases of $CaLi_2$, one stable from 35 to 54 GPa and the other stable from 54 to 105 GPa. Furthermore, they predict that $CaLi_2$ is unstable with respect to dissociation into the constituents at pressures greater than 105 GPa [188].

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 [29]. In contrast to real space mating operations, these authors also employed a reciprocal space mating scheme [29]. 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.

1.6.4 Minerals

Table 1.6 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 [116], and the structure of the post-aragonite phase of $CaCO_3$, stable from 42 - 137 GPa, was solved [117]. 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 [12]. 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 [120]. Wu *et al.* predicted a new low temperature post-perovskite phase of SiO_2 with the Fe_2P-type structure [184].

1.6.5 Molecular crystals

Table 1.7 summarizes searches for molecular crystals. Applications for molecular solids include high-energy materials, pharmaceuticals, pigments and metal-organic frameworks [3, 128, 8]. 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 [204]. Ji *et al.* searched for ice at terapascal pressures and predicted three new phases [62]. 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 [48]. Oganov *et al.* predicted that the β-cristobalite structure is the most stable for CO_2 between 19 and 150 GPa [116].

1.6.6 Inorganic compounds

Table 1.8 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 [158]. 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 [52, 51]. Hermann *et al.* also predicted a new stable phase of LiBeB at ambient pressure and several additional phases under pressures up to 320 GPa [50]. 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 [189]. 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 [203]. 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 [22]. Li *et al.* resolved the structure of superhard BC_2N and found it to have a rhombohedral

lattice [82].

1.7 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 Section 1.4 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 Section 1.1.1. 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 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 selec-

tion 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 Section 1.2.

In Section 1.6, 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 Section 1.5. 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

Section 1.3.

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, 1-D and 2-D 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 proven a useful global optimization method for addressing the structure prediction problem.

Element	Pressure [GPa]	Number of atoms	Energy code	Reference
Al	0	4, 8 ,12	PWSCF	[10]
B	0, 100, 300	2, 3, 4, 6, 8, 9, 16, 12, 24, 26, 28, 30, 32	VASP	[114]
	0	24, 25, 26, 27, 28, 29, 30, 31, 32	VASP	[63]
Ba	0-300	Variable, up to 15	VASP	[156]
C	10-100	2, 4, 6, 8	VASP	[81]
	0	Not specified	CASTEP	[1]
	0	8, others tried	PWSCF	[10]
	0-2000	8	VASP	[115]
Ca	20-600	3, 4, 6, 8, 9, 12, 16	VASP	[118]
Cl	100	8	VASP	[115]
Eu	0-90	Variable, up to 30	VASP	[13]
F	50, 100	8	VASP	[115]
Fe	350	8	VASP	[115]
Ga	0	8	PWSCF	[10]
H	Up to 600	2, 3, 4, 6, 8, 12, 16	VASP	[115]
In	0	4	PWSCF	[10]
K	40-300	4, 6, 8, 12, 16	VASP	[99]
Li	30-1000	4, 6, 8, 12, 16, 24	VASP	[99]
Mo	0	Variable, up to 40	LAMMPS	[123]
N	100-350	8, 12, 16, 32	VASP	[98]
	100	6, 8, 12, 16	VASP	[115]
Na	0-1000	Not specified	VASP	[96]
O	100-250	4, 8	VASP	[97]
	25, 130, 250	4, 6, 8, 12, 16	VASP	[115]
Rb	40-200	4, 6, 8, 12	VASP	[99]
S	12	3, 4, 6, 8, 9, 12	VASP	[115]
Si	0	8	VASP	[164]
	10, 14, 20	8	VASP	[115]
Tl	0	4	PWSCF	[10]
Xe	200, 1000	8	VASP	[115]

Table 1.3: Application of evolutionary algorithms to single element systems.

Compound	Pressure [GPa]	Formula units per cell	Energy code	Reference
Li_mH, $m = 2-9$	60, 80, 100	2, 4	VASP	[56]
LiH_n, $n = 2-8$	0-300	2, 3, 4	VASP	[215]
NaH_n, $n = 6-12$	100, 300	2, 3, 4	VASP	[9]
KH_n, $n = 2-12$	10, 100, 250	2, 4	VASP	[55]
RbH_n, $n = 2-14$	2-250	2	VASP	[57]
CsH_n, $n = 2-9, 16$	30-200	2, 3, 4	VASP	[137]
BeH_n, $n = 2-5$	50, 150, 200	2, 4	VASP	[54]
MgH_n, $n = 2-16$	0-250	2, 3, 4, 5, 6, 8	VASP	[88]
BaH_n, $n = 2-13$	50-200	2, 4	VASP	[54]
WH_n, $n = 1-6, 8$	25, 150	1, 2, 3, 4	VASP	[78]
CH_4	20-800	2, 3, 4	VASP	[39]
	11	Not specified	VASP	[204]
SiH_4	40-300	1, 2, 3, 4, 6	VASP	[102]
GeH_4	50-250	1, 2, 3, 4	VASP	[37]
SnH_4	30-250	1, 2, 3, 4	VASP	[38]
PbH_4	0-500	2, 4	VASP	[191]
NaH	0, 29.3	4	PWSCF, VASP	[89]
CH	0-300	Fixed, up to 8	VASP	[179]
PtH_x, $x = \frac{1}{2}, 1, \frac{3}{2}, 2$	120	6, 10, 12	VASP	[200]
FeH_x, $x = \frac{1}{4}, \frac{1}{3}, \frac{1}{2}, 1, 2, 3, 4$	300, 400	2, 3, 4, 6, 8	VASP	[12]
$LiBeH_3$	50-550	2, 4	VASP	[58]
$LiNH_2$	0-360	2, 4	VASP	[127]
$NaAlH_4$	0-20	1, 2	VASP	[199]
$NaPtH_2$	0	Not specified	VASP	[177]
$NaNH_2$	0, 10, 20	1, 2, 3, 4	VASP	[196]
$SrFeH_4$	0	Not specified	VASP	[177]
$Mg(BH_4)_2$	0-20	2, 4	VASP	[201]

Table 1.4: Application of evolutionary algorithms to hydrogen-containing compounds.

Compound	Pressure [GPa]	Formula units per cell	Energy code	Reference
Al-Sc system	0	8 atoms	VASP	[165]
	0	6, 8 atoms	VASP	[163]
	0	8 atoms	VASP	[63]
$Al_{13}K$	0	1	VASP	[116]
Au_2Pd	0	4	VASP	[164]
	0	4	VASP	[165]
$CaLi_2$	10-250	1, 2, 3, 4, 6, 8	VASP	[188]
$CdPt_3$	0	2	VASP	[165]
CuPd	0	4	VASP	[165]
Na-Ca system	50	Not specified	VASP	[156]
$PdTi_3$	0	2	VASP	[165]

Table 1.5: Application of evolutionary algorithms to intermetallic compounds.

Mineral	Pressure [GPa]	Formula units per cell	Energy code	Reference
Al_2O_3	300	4	VASP	[115]
Al_2SiO_5	10	4	GULP	[205]
$CaCO_3$	50, 80, 150	1, 2, 4	VASP	[115]
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	[12]
FeS	56, 120, 400	2, 3, 4, 6, 8	VASP	[120]
$FeSi_x$, $x = \frac{1}{3}, \frac{1}{2}, 1, \frac{5}{3}, 2, 3$	300, 400	8, 9, 12, 16, 18, 24 atoms	VASP	[192]
$MgCO_3$	110, 150	1, 2, 4, 6	VASP	[115, 116]
Mg-O system	Up to 850	Up to 20 atoms	VASP	[206]
$MgSiO_3$	250	32	GULP	[205]
	80, 120, 1000	4, 8	VASP	[115]
Na-Cl system	0-250	Up to 16 atoms	VASP	[193]
SiO_2	500, 2000	1 - 8	PWSCF	[184]
	0	3	VASP	[115]
	10	24	GULP	[205]
TiO_2	0	2, 4, 8	GULP	[181]
	0	16	GULP	[89]
	Not specified	4	VASP	[115]

Table 1.6: Application of evolutionary algorithms to minerals.

Compound	Pressure [GPa]	Formula units per cell	Energy code	Reference
Benzene C_6H_6	0 - 300	Not specified	VASP	[179]
	0, 5, 10, 25	4	VASP	[204]
Ice H_2O	1000-4000	8	LAMMPS, PWSCF	[62]
	0	4	VASP	[204]
	0	4	VASP	[115]
CO_2	50, 100, 150	1, 2, 3, 4, 6, 8	VASP	[116]
	12 - 20	2	VASP	[204]
	50	1, 2, 3, 4, 6, 8	VASP	[115]
NH_3	5, 10, 25, 50	4	VASP	[204]
Glycine	1, 2	2, 3, 4	VASP	[204]
Butane-1,4-diammonium dibromide	0	2	VASP	[204]
Urea $CO(NH_2)_2$	0	2	VASP	[115]

Table 1.7: Application of evolutionary algorithms to molecular solids.

Compound	Pressure [GPa]	Formula units per cell	Energy code	Reference
$Al_{12}C$	0	1	VASP	[116]
BeB_2, BeB_3, BeB_4, $BeB_{2.75}$	0, 160	4, 8	VASP	[49]
CsI	5 - 300	2, 3, 4, 8	VASP	[189]
Fe-B system	0	Up to 15 atoms	VASP	[68]
GaAs	0	4	VASP	[164]
Li-B system	0 - 320	1, 2	VASP	[52, 51]
Li-Si system	0	Up to 20 atoms	VASP	[158]
MgB_2	0 - 500	1, 2, 3, 4, 6	VASP	[100]
SiC	0	4	VASP	[164]
WN_2	0, 60	1, 2, 3, 4	VASP	[173]
XeF_2	0 - 200	Up to 4	VASP	[77]
Xe-O System	5 - 220	Up to 36 atoms	VASP	[203]
$ATiO_2$, A = Ba, Be, Ca, Mg, Sr	0	Not specified	VASP	[178]
$NaPtF_2$	0	Not specified	VASP	[178, 177]
$SrFeC_4$	0	Not specified	VASP	[178, 177]
$CaRhO_3$	Not specified	6	VASP	[138]
Li_3RuO_4	0	Not specified	GULP	[22]
$SrTiO_3$	0	10	GULP	[89]
BC_2N	30, 100	1, 2, 4	VASP	[82]
LiBeB	0 - 320	Not specified	VASP	[50]
Si_2N_2O	0	2	VASP	[115]
$SrSiN_2$	0	4	VASP	[115]

Table 1.8: Application of evolutionary algorithms to inorganic compounds.

CHAPTER 2

AN IMPROVED EVOLUTIONARY ALGORITHM FOR THE PREDICTION OF CLUSTERS, WIRES, TWO-DIMENSIONAL MATERIALS AND FULL AND PARTIAL PHASE DIAGRAMS

2.1 Introduction

A material's properties are largely a function of its crystal structure, so the first step in computational materials design is structure determination. However, predicting which structure a material will adopt given only its composition (or range of allowed compositions) is a difficult global optimization problem. In the past decade, much success has been achieved by using evolutionary algorithms to solve this optimization problem [41, 163, 68, 89, 159, 10]. Evolutionary algorithms draw their inspiration from biological evolution, and they model a group of candidate solutions (structures) as a population. Structures in the population produce offspring through a mating operation in which slabs are sliced from the cells of two parents and combined together to form an offspring structure. Better structures are chosen to mate more frequently, and to the extent that good structural features are passed from parents to offspring, the population as a whole tends toward the global minimum over time. See Refs. [159, 131] for more detailed descriptions of evolutionary algorithms for structure prediction.

Here we present an improved evolutionary algorithm for structure prediction.

Our algorithm is based on the Genetic Algorithm for Structure and Phase Predcition (GASP) by Tipton *et al.* [159], and we have made several important changes to improve its computational efficiency and extend its capabilities. The improved algorithm is implemented in Python and makes extensive use of pymatgen [119], a Materials Project open-source Python library for materials analysis. We have also implemented reasonable default values for most of the parameters used by the algorithm to minimize the amount of input that must be given by the user. The code is freely available under the MIT license, and the latest version can be obtained from the Github repository at https://github.com/henniggroup/GASP-python.

Sec. 3.2 describes the details of the new features we have added to the evolutionary algorithm for structure prediction. In Sec. 3.3, we demonstrate the algorithm's new features, including improved composition space sampling in Sec. 2.3.1, partial phase diagram searching in Sec. 2.3.2, and searching for the structures of subperiodic systems, such as clusters, wires and 2D materials in Secs. 2.3.3, 2.3.4 and 2.3.5, respectively.

2.2 Methods

Our algorithm is modeled on the original GASP code [159]. Here we describe the important new features we have implemented.

2.2.1 Continual Structure Creation

The original GASP code organizes the evolving population of organisms into generations, where an organism is made up of a structure and an energy. The organisms comprising a generation are used to create offspring, which are then grouped together to form the next generation. When running in parallel (more than one energy calculation at once), this generation-based scheme can lead to inefficient use of computational resources because the energy of every organism in a given generation must be computed before that generation can create offspring. So even if we have the capacity to run, e.g., 20 total energy calculations simultaneously, we must wait for the slowest one to complete before moving on. This can lead to a significant slowdown of the search, especially if using DFT for the total energy calculations, because there often exists a large variation in the time needed to relax different structures.

To alleviate this problem, we abandon the generational scheme in favor of having the algorithm make offspring organisms continually. In our new formulation, we call the collection of organisms comprising the current breeding population the pool. The pool is composed of two parts: the promotion set and the queue. The promotion set contains the best few organisms seen by the algorithm so far, and its purpose and function are analogous to those of the promotion operator in the previous version of the algorithm [159]. The queue contains the rest of the organisms in the pool, and it corresponds to a first in, first out (FIFO) data structure. When an organism is added to the back of the queue, the organism at the

front of the queue is popped off and removed from the pool. In this way the pool maintains a fixed size, and each organism has an equal number of chances to be selected to create offspring before being removed from the pool (except for the organisms in the promotion set; they get more opportunities). See Fig. 2.1 for an illustration of the algorithm, including the mechanics of adding an organism to the pool.

Lonie and Zurek implemented a similar scheme in their evolutionary algorithm for structure prediction, XTALOPT [89]. Our design differs from theirs in that we allow suboptimal solutions to enter the pool by being added to the queue, as described above. In XTALOPT, the pool only contains the best structures seen by the algorithm.

The pool described above allows offspring organisms to be continually created and therefore naturally lends itself to parallelization. For example, if N total energy calculations are to be run at a time, the algorithm first makes N offspring organisms and calculates the energy of each on a separate thread. The algorithm then repeatedly checks the N threads, and as soon as one finishes its energy calculation, the relaxed organism is added to the pool and the selection probabilities of all organisms in the pool are updated. A new offspring organism is then created and submitted for an energy calculation on its own thread, and the algorithm goes back to checking for dead threads. This process is outlined in the pseudocode below, where the N threads are stored in a list named `threads`.

```
while not converged:
   for index, thread in enumerate(threads):
      if not thread.is_alive():
         relaxed_org = thread.get_relaxed_org()
         pool.add_relaxed_org(relaxed_org)
         pool.compute_selection_probabilities()
         new_org = pool.make_offspring()
         new_thread = Thread(
            target=compute_energy,
            args=new_org)
         new_thread.start()
         threads[index] = new_thread
```

2.2.2 Constraints

For a variety of reasons, it is advantageous to impose constraints on the parameters of structures considered by the evolutionary algorithm. For example, constraining the minimum distance between atoms has the effect of excluding structures containing unphysically short bond lengths, and constraints on lattice vector lengths and angles help standardize the way structures are represented, which is important for the performance of the algorithm. See Refs. [159, 131] for a more thorough discussion of constraints and the motivations behind them.

The constraint on the minimum distance between atoms is particularly important because it excludes unphysical regions of phase space and prevents computational resources from being spent relaxing structures with extremely unfavorable starting configurations. However, when searching for structures containing more than one atom type, a single constraint on the minimum distance between atoms is often insufficient. This is because atoms of different types have different equilibrium bond lengths. For example, consider a system containing carbon and titanium. A C-C bond in diamond has a length of 1.54 Å, while the shortest distance between Ti atoms in the ground state hcp crystal structure is about 2.64 Å. Therefore, setting the minimum interatomic distance constraint to a value that is appropriate for Ti-Ti bonds would cause equilibrium C-C bonds to be misidentified as too short. On the other hand, setting the constraint with only carbon in mind would permit structures containing Ti atoms that are unphysically close to each other. To address this problem, we employ species-specific minimum interatomic distance constraints. That is, a different value of the constraint may be set for each distinct pair of atom types. By default, the algorithm sets the mininim interatomic distance constraint for a pair of atom types to 60% of the sum of the radii of the atoms.

Constraints are also important when searching for non-bulk structures. In particular, they are used to limit the size of the non-bulk geometry. See section 2.2.4 for more details.

2.2.3 Phase Diagram Search

The lowest energy structures exhibited by a system at various compositions are frequently of interest. To find these structures, we use the genetic algorithm to search for a system's phase diagram, based on the method of Trimarchi *et al.* [163]. Two main modifications to the algorithm are needed to search for phase diagrams. The first is the objective function: instead of simply using the enthalpy per atom, we must take the composition of a candidate structure into account. To do so, we set the objective function to the distance from the computed convex hull. As the algorithm finds better structures, the convex hull is updated, along with the objective function values of the organisms in the population. The second modification is simple: instead of constraining the algorithm to only create structures at a fixed composition, we allow the algorithm to create structures with any composition within the range under consideration. See Refs. [159, 131] for more thorough discussions of the motivation for phase diagram searching and details of convex hull construction. Here we discuss some of the challenges associated with searching for phase diagrams and our modifications to the algorithm to address those challenges.

Sampling Composition Space

When searching for phase diagrams, it's important to sample the composition space as uniformly as possible because in general, for an arbitrary multicompo-

nent system, the distribution of the ground states with respect to composition is not known *a priori*. Of course, the ground states may turn out to be distributed nonuniformly with respect to composition. In this case, the algorithm will naturally spend more time in regions of composition space with higher densities of ground states. This source of nonuniformity in the sampling of the composition space is a normal part of the exploration-exploitation tradeoff, and it is desirable that the algorithm should spend more resources on promising regions of phase space.

However, bias in sampling the composition space may arise through other, unwanted causes and can lead to suboptimal performance. The species-specific minimum interatomic distance constraints (see section 2.2.2 above) introduce a significant source of bias. This is because, all other things being equal, structures with more of the smallest type of atom are more likely to satisfy the constraints (and therefore not be discarded during development) than structures with more of the larger types of atoms. The greater the difference in size (and therefore minimum interatomic distance constraint) between the atom types, the more bias is introduced. The net result is that many more energy evaluations are performed on structures with higher concentrations of the smallest atom type than on those with higher concentrations of the largest atom type.

To ameliorate this problem, we employ composition-dependent volume scaling during pre-evaluation development. In particular, the decomposition of each candidate offspring structure is computed relative to the current convex hull.

The volume (per atom) of the candidate offspring structure is then scaled to the weighted average of the volumes per atom of the structures on the convex hull into which its stoichiometry decomposes. This counteracts the bias caused by species-specific minimum interatomic distance constraints because structures with higher concentrations of the largest atom type tend to have larger volumes (per atom). Volume scaling in this way has the added benefit of quickly moving a candidate structure closer to the nearest minimum of the energy landscape before starting the (potentially expensive) energy calculation, reducing the number of iterations that must be performed by the external energy code's local minimizer.

We note that both of these techniques (species-specific minimum interatomic distance constraints and composition-dependent volume scaling) must be used in conjunction to prevent bias. For example, if volume scaling is performed, but only a single minimum interatomic distance constraint is used, the search will then be biased toward the region of the composition space containing higher concentrations of the largest atom type. This is because, as far as the constraint is concerned, all the atoms are identical, but the structures with more of the largest type are scaled to a larger volume, making them more likely to satisfy the constraint and survive development.

In addition to the non-symmetric bias (i.e., favoring one endpoint over the others) caused by the species-specific minimum interatomic distance constraints, symmetric bias with respect to composition also occurs when searching for phase diagrams. The first source of this bias lies with the number of atoms contained

within a structure. Evidently, the set of compositions that a structure may adopt is necessarily limited by the number of atoms in its cell. The problem arises when considering a group of structures of varying sizes: the compositions that can be sampled with such a group are not uniformly distributed. Fig. 2.2 is a histogram showing which compositions can be accessed in a binary system, and the (normalized) numbers of distinct ways to access them, with structures containing up to 10, 20 and 30 atoms. As can be seen in the figure, the accessable compositions are clustered toward the middle of the composition space, leaving gaps near the endpoints and on either side of the center. Note that as the maximum system size increases, the accessable compositions tend to a more uniform-like distribution.

Symmetric bias is also caused by the cut-and-splice mating operation. Mating usually produces offspring structures with compositions between those of the two parents, and this causes the population of structures to drift toward the middle region of the composition space over time. The result is that the algorithm oversamples the central part of the composition space relative to other areas.

We suggest two ways to address the problem of symmetric bias in the sampling of the composition space. Based on Fig. 2.2, it is clear that one approach is to increase the number of atoms in the structures considered by the algorithm. This can be accomplished by simply increasing the size (by taking supercells) of the reference structures provided to the algorithm at the endpoints of the composition space. We also encourage the algorithm to increase the size of structures in the population over time by doubling one of the parents before mating with some

non-zero probability, and also by specifying a mutation operator that adds atoms on average [159]. This allows the algorithm to discover the structural motifs correlated with low energy using small cells at the beginning before considering larger, more complex structures later in the search.

Increasing the number of atoms comes at a computational cost. This may not be a severe issue when using empirical potentials for the energy computations, but it can be when using DFT. For this reason, we find that it is best to start searches with relatively small structures (up to 8 or 10 atoms) and then grow them gradually over the course of the run, as previously described. Contrasted with a more naive approach of simply forcing the algorithm to only consider larger structures from the very beginning (via a constraint on the minimum number of atoms in the cell), our approach performs the cheaper energy calculations on smaller cells early in the search, while still eventually accessing the improved composition space sampling distributions provided by larger cells.

The second technique we use to reduce symmetric bias is motivated by the desire to prevent offspring organisms from having predominantly intermediate compositions. It involves modifying the selection probabilities such that two parent organisms are likely to have similar compositions, which causes their offspring to also have a similar composition. See Appendix A for details of our implementation. We note that this strategy affects the trade-off between exploration and exploitation with respect to the parameter of composition, favoring the latter over the former. This is because the first parent organism is selected only on

the basis of its objective function value, and the second parent and resulting offspring are likely to have compositions similar to that of the first parent. Therefore, the compositions of the best few organisms will tend to be overrepresented in the population. To limit this effect, when selecting the second parent, we place a moderate weight on its composition (relative to that of the first parent), so that parents with dissimilar compositions will still mate reasonably often. An added benefit of this approach is that structures with similar compositions are more likely to have similar structural motifs, which increases the odds of producing good offspring.

Even with the solutions discussed above, some non-uniformity in the composition sampling will usually be present. If one is interested in only a certain region of the composition space, performing a partial phase diagram search (see Section 2.2.3) may provide a better sampling of the relevant compositions.

Searching Endpoint Compositions

In many cases, the ground state structures at the endpoints of the composition space are known *a priori*. In this situation, there is no need to search for structures at the endpoint compositions, and we have made searching for structures at the endpoint compositions optional, both in the initial random population and in the subseqent offspring. However, the ground state structures at the endpoint compositions must be provided to the algorithm in the initial population.

Two pitfalls should be avoided when assuming the ground state structures at

the endpoint compositions. The first has to do with how well the external energy code approximates reality, since the algorithm searches the energy landscape defined by the energy code. For example, empirical potentials have been known to stabilize incorrect structures [159]. So even if the ground state structures at the endpoint compositions are known experimentally, it can be worth while to include the endpoint compositions in the search to see if the energy landscape of the empirical potential contains false minima at the endpoint compositions. The second pitfall occurs when searching systems for which some uncertainty exists concerning the lowest energy structures at the endpoint compositions. This could be the case, for example, when searching for non-bulk structures or systems under pressure. In these cases, it's best to include the endpoints in the overall phase diagram search.

Partial Phase Diagram Search

In some cases, we are interested only in a certain region of a phase diagram and would prefer not to expend computational resources searching areas outside of this region. Investigating the hydrogen-rich side of a high-pressure phase diagram [215, 9] is an example of this situation. To facilitate these types of searches, we have implemented searching for partial phase diagrams; that is, phase diagrams with one or more non-elemental endpoint compositions.

Care should be taken when searching for phase diagrams with non-elemental endpoints and interpreting the results of such searches. In particular, unless all

the endpoint compositions correspond to points on the convex hull of the entire system, it will not be possible for the algorithm to find the correct partial phase diagram (i.e., the thermodynamic ground states).

2.2.4 Non-Bulk Structures

Much progress has been made in recent years synthesizing free-standing materials with sizes in the nanometer and sub-nanometer range, including clusters, wires and 2D materials. Such materials can be described as subperiodic; that is, they lack periodicity in one or more dimensions. Due to their small size, these materials possses extremely large surface area to volume ratios. In fact, the fraction of atoms located at or near a surface is so large that these materials sometimes adopt structures that differ from those of their bulk conterparts. These different structures, together with the effects of quantum confinement, can lead to unique optical, electronic and mechanical properties. Since structure prediction is an important first step to identify such materials, we have implemented searching for clusters, wires and 2D materials in our evolutionary algorithm.

Thermodynamic stability is an important criterion when considering subperiodic materials. In general, the relevant quantity is the difference between the energy (per atom) of the proposed structure and the energy of the competing ground state bulk phase. We note that due to surface energies, all subperiodic materials are unstable relative to the bulk phases. However, if they display sufficiently large

metastability (i.e., low formation energy relative to the bulk), these materials may be kinetically stabilized and persist in the free-standing state. We therefore take the energy per atom as the objective function when searching for such structures.

It should be noted that formation energy is not the only choice of objective function for subperiodic materials. Tosatti *et al.* proposed minimizing the line tension of nanowires instead [162], and surface energy has been suggested as an alternative objective function for 2D materials [132]. We have not yet implemented these alternative objective functions in our evolutionary algorithm.

Because most external energy codes assume periodic boundary conditions in all dimensions, care must be taken when using them to compute the energies of subperiodic structures. In addition, the mating operator must be slightly modified to work properly with these structures, and additional constraints may be imposed on their size. Below we outline the changes we made to the algorithm to address these issues for each type of subperiodic structure.

0D Materials (Clusters)

By definition, clusters lack periodic boundary conditions in all directions. Therefore, the lattice vectors contain no physically meaningful information, and we are free to replace them with vectors of our choice. For simplicity, the algorithm replaces the lattice vectors of clusters with a set of orthogonal vectors lying parallel to the three Cartesian coordinates. To minimize the interation between

clusters and their periodic images, vacuum padding is added in each dimension prior to the energy calculation; it is removed from the relaxed structure when the energy calculation is finished. We note that the external energy code should not be permitted to relax the lattice vectors during the energy calculation. Instructions for achieving this with GULP, LAMMPS and VASP are included in the manual.

A few changes to the mating operator are needed to properly treat clusters. For bulk materials, the algorithm shifts the atoms in each parent cell by a random amount in the direction of the lattice vector that has been randomly chosen to be cut [41]. The purpose of such shifts is to remove any implicit correlation between the representations of the two parent structures. However, for non-bulk structures, shifting the atoms in a non-periodic direction would change the structure itself, not just its representation, leading to the loss of important information. Since clusters lack periodicity in all three directions, the algorithm does not shift the atoms of the parent structures prior to making the cut, regardless of which lattice vector has been chosen.

Since clusters are completely non-periodic, the orientation of a cluster relative to the lattice vectors is arbitrary and contains no physical information. To avoid favoring any particular orientation, the algorithm randomly rotates each parent cluster about each Cartesian axis before making the cut [129, 65].

In addition to contraints on the maximum and mimimum numbers of atoms in a cluster, constraints on the maximum and minimum size of a cluster may also be imposed, where the size of a cluster is defined as the maximum distance between

atoms within the cluster.

1D Materials (Wires)

Wires display periodicity in one dimension, but not in the other two. To capture this feature, the algorithm represents wires as lying along the Cartesian z-axis. A new structure is first rotated such that its $\vec{c}$ lattice vector is parallel to the z-axis, and then lattice vectors $\vec{a}$ and $\vec{b}$ are set parallel to the Cartesian x and y directions, respectively. Before doing an energy calculation, vacuum padding must be added to the $\vec{a}$ and $\vec{b}$ lattice vectors to prevent the wire from interacting with its periodic images. This padding is removed upon completion of the energy evaluation. Analogous to the case of clusters, the external code used to compute the total energies should not be permitted to relax the $\vec{a}$ and $\vec{b}$ lattice vectors.

The mating operator must be modified slightly to work with wires. As discussed for clusters above, the atoms in the parent structures must not be shifted along a non-periodic direction prior to mating. Given our representation of wires above, this means that random shifts should only occur along the $\vec{c}$ lattice vector. In addition, since the orientation of the wires relative to the $\vec{a}$ and $\vec{b}$ lattice vectors is arbitrary, they are randomly rotated about the $\vec{c}$ lattice vector (Cartesian z-axis) before the cut is made.

The user may impose constraints on the maximum and minimum size of wires considered by the algorithm. If a wire is aligned along the Cartesian z-axis and

its atoms are projected onto the $x - y$ plane, its size is taken to be the maximum distance between the projected atoms.

2D Materials

The evolutionary algorithm treats 2D materials in a way analogous to clusters and wires, described above. See Ref. [132] for a complete description.

2.3 Results and Discussion

Here we evaluate the performance of the algorithm and demonstrate its new features. In Sec. 2.3.1, we evaluate our methods for reducing bias in the composition space sampling by applying the algorithm to search for the binary Zr-Cu phase diagram predicted by an empirical potential using the embedded-atom method (EAM). To demonstrate the partial phase diagram searching capability of the algorithm, we apply it to the partial ternary Al-CuAl-ZrAl system predicted by an EAM potential in Sec. 2.3.2. In Secs. 2.3.3 and 2.3.4, we use the algorithm to search for the structures of Au clusters and wires, respectively, predicted by an EAM potential. To demonstrate searching for 2D materials, we apply the algorithm to search for 2D structures of Al_2O_3 in Sec. 2.3.5.

With the exception of the search for 2D Al_2O_3, we used empirical potentials for the energy computations in all the systems considered here. This choice was

made primarily in the interest of computational speed. We note however that the genetic algorithm is interfaced to the DFT code VASP, and to provide more accurate predictions, we generally use DFT when investigating energy landscapes.

2.3.1 Zr-Cu Phase Diagram

To test the effect of the techniques discussed in Sec. 2.2.3 on the algorithm's sampling of composition space, we apply the algorithm to the Zr-Cu binary system, using the empirical EAM potentials of Cheng *et al.* [24].

We first evaluate our techniques for mitigating nonsymmetric bias in the composition space sampling. In particular, we consider the effects of species-specific minimum interatomic distance constraints and volume scaling, discussed in Sec. 2.2.3. To do so, we collect three data sets, each obtained by running 100 genetic algorithm searches, where each search used 1000 objective function evaluations. At the beginning of each search, we provide the algorithm with the elemental ground state structures: fcc Cu and hcp Zr. Fig. 2.3 shows the distributions over the composition sampled by the algorithm. In the first set of runs, we employ species-specific minimum interatomic distance constraints, but not volume scaling. The data from this set are shown with blue diamonds in the figure. As can be seen, the composition sampling is skewed toward the side of the composition space with smallest type of atom (Cu in this case). In the second set of runs, we use volume scaling, but not species-specific minimum interatomic distance con-

straints, shown with red squares in the figure. This setting causes the algorithm to oversample the side of the composition space with the largest type of atom (Zr). In the third set of searches, both volume scaling and species-specific minimum interatomic distance constraints were employed. The data from these searches are shown in black circles in the plot. Using both techniques together removes bias toward either end of the composition space.

Next, we evaluate our approaches to minimizing symmetric bias in the composition space sampling, discussed in section 2.2.3. We again collect three data sets, each one from 100 genetic algorithm searches, where each search used 1000 objective function evaluations, and the ground state structures of Cu and Zr were given to the algorithm at the beginning of each search. Fig. 2.4 shows the distributions over the composition space sampled by the algorithm. In the first set of runs, we provide the algorithm with larger supercells of the elemental reference structures, containing eight atoms each, instead of smaller ones containing only four. The data from this set are shown with blue diamonds in the figure. In the second set of runs, we use an additional measure of fitness to preferentially select second parent structures that have compositions similar to those of the first parents, but provide the algorithm with smaller elemental reference structures containing only four atoms. These data are shown with red squares. In the third set of searches, both larger reference structures and composition-dependent fitness were used.

From Fig. 2.4, we see that employing both techniques together produces the best result, reducing the oversampling of the middle of the composition space.

Table 2.1: Compositions, space groups and atoms per primitive cell of the ground states predicted with the EAM potentials. The new phases are given in bold. Careful relaxation revealed the ground state structures at compositions $ZrAl_5$ and $ZrAl_6$ to posses higher symmetry than previously reported.

Composition	Space group		Atoms per cell
Al	194	$P6_3/mmc$	2
CuAl	221	$Pm\bar{3}m$	2
$CuAl_2$	65	$Cmmm$	3
$CuAl_5$	15	$C2/c$	12
ZrAl	65	$Cmmm$	4
$ZrAl_2$	191	$P6/mmm$	3
$ZrAl_3$	63	$Cmcm$	8
$ZrAl_5$	164	$P\bar{3}m1$	6
$ZrAl_6$	166	$R\bar{3}m$	7
$ZrAl_8$	12	$C2/m$	9
$ZrAl_{13}$	1	$P1$	14
$ZrAl_{15}$	44	$Imm2$	16
$ZrAl_{17}$	1	$P1$	18
$ZrCuAl_2$	123	$P4/mmm$	4

However, a significant amount of bias toward the center still exists, and the regions near the endpoints remain undersampled. These effects are consequences of searching for moderately sized cells (less than 20 atoms on average). If it is important to thoroughly explore an area of the composition space that tends be undersampled in full-range searches, performing a partial phase diagram search is the best way to achieve additional sampling in the region of interest.

2.3.2 Al-CuAl-ZrAl Partial Phase Diagram

To illustrate searching for partial phase diagrams, we apply the evolutionary algorithm to the ternary Al-CuAl-ZrAl system, using the same empirical potential as in section 2.3.1. Tipton *et al.* searched the full ternary composition space (Al-Cu-Zr) with a previous version of GASP, and the phase diagram predicted with the EAM potential contained many ground state structures in the Al-rich region of the composition space [159]. Using the new version of the algorithm with partial phase diagram searching capability, we explore this part of the composition space further. We provide the algorithm with structures at the endpoints of the composition space (Al, CuAl, ZrAl) predicted by Tipton *et al.* and run the search for 10,000 objective function evaluations.

Fig. 2.5 shows the convex hull of the EAM potential predicted by the genetic algorithm. In this region of the ternary phase diagram, the system exhibits several binary ground states, and also a single ternary one with composition Al_2CuZr. Table 2.1 contains the compositions and space groups of the predicted EAM ground state structures. The algorithm recovered the previously predicted ground states, and it also found new ones with compositions $CuAl_5$ and $ZrAl_n$, for $n = 13, 15, 17$. The new ground state with composition $CuAl_5$, shown with a green marker in Fig. 2.5, reveals that the previously predicted ground state with composition $CuAl_4$ actually lies 25 meV/atom above EAM convex hull. For pure Al, Tipton *et al.* found that the EAM potential stabilizes the hcp structure over the experimental fcc structure [159]. The new EAM ground states found here with compositions

$ZrAl_n$ ($n = 13, 15, 17$) correspond to hcp Al with substitutional Zr impurity atoms; these are shown with red markers in Fig. 2.5. These structures indicate that Zr impurity atoms lower the energy relative to the true competing phases, which is likely an artifact of the parametrization of the EAM potentials. Developing empirical potentials that don't stabilize unphysical ground states remains an extremely difficult problem, particularly for multi-component systems, and evolutionary algorithms for structure prediction can be used to identify such issues during the design phase of the potentials, when they can be more easily corrected [60, 123].

2.3.3 Au Clusters

To demonstrate the cluster searching capability of the evolutionary algorithm, we apply it to search for Au clusters. The EAM empirical potential of Foiles *et al.* is used for the energy calculations [36], and we run 16 genetic algorithm searches for clusters with fixed numbers of atoms, ranging from 5 to 20. It has been observed that the number of local minima in the energy landscapes of Lennard-Jones clusters increases exponentially with the number of atoms in the cluster [53, 170], In case a similar relationship holds for EAM clusters, we increase the computational budget (number of objective function evaluations) exponentially with cluster size, ranging from 100 LAMMPS calculations for 5 atom clusters to 3000 LAMMPS calculations for 20 atom clusters. In each search, the genetic algorithm identified the minimum energy cluster well before exhausting the computational budget.

Fig. 2.6 shows the minimum energy clusters found by the algorithm, containing from 5 to 20 atoms. The clusters with 5, 6, 7 and 8 atoms formed trigonal bipyramidal, octahedral, pentagonal bipyramidal and bicapped octehedral structures, respectively. The clusters containing from 9 to 11 atoms form a small series, which can be approximately described as a 7 atom close-packed planar flake capped with 2, 3 and 4 atoms, respectively. (Note: the 7 atom flake is not actually planar, but distorted by the capping atoms.) The 12 atom cluster corresponds to an icosahedron missing a single atom, and the 13 atom cluster forms a perfect icosahedron. The 14 and 15 atom clusters are similar to an icosahedron, but with one and two rings containing 6 atoms instead of 5, respectively. The cluster containing 16 to 18 atoms form a series based on the 15 atom structure: instead of a single capping atom on each end, these structures have 2, 3 and 4 capping atoms on one end, and just one on the other. The 19 atom structure can be described as an 8-atom ring surrounding a central atom, with each side capped by a square pyramid. We observe that the EAM potential stabilizes this 19 atom cluster over the 19 atom double icosahedral structure by 18 meV/atom. The 20 atom cluster does not display a high degree of symmetry.

To determine the thermodynamic stability of the clusters, we compare their energies to that of the ground state bulk phase of Au, which has the fcc crystal structure. The formation energy relative to the bulk ground state is given by

$$\Delta E_{\mathrm{f}} = \frac{E_{\mathrm{cluster}}}{N_{\mathrm{cluster}}} - \frac{E_{\mathrm{bulk}}}{N_{\mathrm{bulk}}}, \tag{2.1}$$

where E_{cluster} and E_{bulk} are the total energies of the cluster and bulk respectively, and N_{cluster} and N_{bulk} are the numbers of atoms in each structure. Fig. 2.7 shows the formation energies of the clusters predicted by the EAM potential. The formation energy monotonically decreases as the cluster size increases, although the rate is significantly lower for larger cluster sizes.

To evaluate a cluster's stability relative to those containing one atom more and less, the second finite difference of the total energy is often considered [66, 32, 40, 46, 174]. For a cluster containing n atoms, this quantity is defined as

$$\Delta_2 E(n) = E_{n-1} + E_{n+1} - 2E_n, \tag{2.2}$$

where E_{n-1}, E_n and E_{n+1} are the total energies of the clusters containing $n-1$, n and $n+1$ atoms, respectively. Fig. 2.8 is a plot of the second finite difference energies for the clusters predicted by the EAM potential. According to this measure, the 13 atom icosahedral cluster is particularly stable relative to its neighbors in n-space. We also observe that odd values of n are preferred, for $n > 11$.

Many authors have studied Au and other noble metal clusters using empirical potentials. Michaelian *et al.* and Darby *et al.* used genetic algorithms to investigate the Au clusters predicted by the many-body Gupta potential [108, 28, 44]. Several of the cluster geometries they predict are equivalent to those found here with the EAM potential of Foiles *et al.*, including the clusters containing 6, 7, 10 and 13 atoms. Doye and Wales used a basin hopping algorithm to explore the energy

landscapes of Au clusters defined by the Sutton-Chen potential [32, 154]. The minimum energy clusters they predict for $n = 9 - 11, 13$ are the same as those found here. Karabacak *et al.* used molecular dynamics with slow quenching to study the minimum energy goemetries of Pd clusters predicted by the Voter-Chen EAM potential [66, 172]. The Pd clusters containing 5-8, 13, 15, and 17 atoms have the same geometries as the corresponding Au clusters found here.

It should be noted that several DFT studies have predicted that Au clusters adopt planar geometries at sizes less than about 7 atoms [46, 43, 174, 195, 186]. Embedded atom empirical potentials, including the one studied here, tend to prefer high-coordination configurations, and Garcia *et al.* have noted that they usually do not describe very small clusters correctly [40]. The EAM potentials of Foiles *et al.* were designed to model bulk fcc crystals, and we do not interpret their apparent shortcomings in treating clusters as a reflection of their quality.

2.3.4 Au Wires

We next apply the evolutionary algorithm to search for the structures of Au nanowires. The EAM potential of Foiles *et al.* [36] is again used for the energy calculations, and we run seven searches, with the maximum wire diameter constraint ranging from 3 to 9 Å. The number of objective function evaluations and the constraint on the maximum number of atoms ranged from 1000 to 2500 and 10 to 50, respectively.

Table 2.2: Maximum diameters and formation energies of the Au wires predicted with EAM potential.

Maximum diameter (Å)	Formation energy (meV/atom)
3	716
4	547
5	433
6	345
7	342
8	292
9	233

Fig. 2.9 illustrates the lowest energy structures found in each search. In the search for the thinnest wires, the algorithm identified the triangular structure shown in (a) to have the lowest energy. A similar structure was found in the search for slightly thicker wires, except it contains four columns of atoms instead of three, shown in (b). The structure shown in (c) is composed of 5 atom rings surrounding a central linear chain of Au atoms. We note that this structure corresponds to the icasohedral structure reported for 13 atom clusters in Sec. 2.3.3, except extended indefinitely in one dimension. The next structure, labeled (d) in the figure, is similar to the previous one, except the outer rings contain 6 atoms each instead of 5. This structure is similarly related to the minimum energy 15 atom cluster found in Sec. 2.3.3. The last three structures, shown in (e), (f) and (g), are for thicker wires. We observe that for these thicker wires, the EAM potential stabilizes close-packed fcc-like structures over the ring structures found for the thinner wires.

Several authors have experimentally synthesized Au wires with sub-

nanometer diameters. Kondo *et al.* report core-shell Au nanowires with diameters as small as 0.8 nm, with the core adopting the fcc structure (with the [110] direction oriented along the wire axis), and the shell displaying the hcp structure [69]. In a later study, Kondo *et al.* report Au nanowires with a helical multi-shell structure, in which each shell corresponds to a planar triangular lattice folded into a tube [70]. Identical structures were also reported by Oshima *et al.* [121]. Bilalbegović used molecular dynamics simulations with an EAM potential [33] to study both finite length [15] and infinite [16] Au nanowires with diameters of 0.9 and 1.2 nm. The wires were initially configured in the fcc structure, with the axis oriented along the [111], [110] and [100] directions. During the MD simulations, the wires oriented along the [111] and [100] directions were observed to adopt a multi-shell structure, while those oriented along the [110] direction displayed a core-shell morphology, with the core region retaining the fcc structure. Tosatti *et al.* used empirical potentials and DFT to study several different multi-shell Au nanowires, and find that the experimentally observed helical multi-shell structures minimize the string tension of the wires [162].

The helical multi-shell structures predicted by Oshima *et al.* have helical pitches of at least 11 nm [121]. Wires with such large repeat lengths contain many more atoms than were permitted in the structures considered by genetic algorithm, and so did not reside in the search space. However, the smaller structures found by the algorithm form shells around the central axis of the wire, indicating a motif similar to those of the helical structures. We observe that for wires with diameters greater than 6 Å, the EAM potential favors close-packed structures over

Table 2.3: Structural information, including space group, lattice parameters a, b and γ, Wyckoff positions, and formation energy, ΔE_f for the predicted 2D phases of Al_2O_3. We have used 3D space groups to describe these finite-thickness 2D structures that lack periodicity in the direction normal to the 2D sheet. In the representations given here, the $\vec{c}$ lattice vector is normal to the plane of the 2D sheet. The vertical components of the general Wyckoff positions are given as distances from the central plane, and the in-plane components are given as fractions of the $\vec{a}$ and $\vec{b}$ lattice vectors. Symmetry information was obtained with the FINDSYM software package [150].

	Space group	a, b (Å)	γ (°)	Atomic positions		ΔE_f (meV/atom)
hexagonal	P3m1 (156)	2.939	120	Al 1(a)	z = 1.538 Å	186
				Al 1(b)	z = -1.462 Å	
				O 1(a)	z = -0.180 Å	
				O 1(b)	z = 2.137 Å	
				O 1(c)	z = -2.137 Å	
triclinic	P$\bar{1}$ (2)	2.819, 5.214	95.13	Al 2(i)	x = 0.412	180
					y = 0.146	
					z = -1.602 Å	
				Al 2(i)	x = -0.171	
					y = -0.364	
					z = -1.279 Å	
				O 2(i)	x = -0.332	
					y = 0.269	
					z = 0.000 Å	
				O 2(i)	x = 0.305	
					y = 0.464	
					z = -2.112 Å	
				O 2(i)	x = -0.118	
					y = -0.042	
					z = -2.048 Å	

shell structures, at least for the solution space explored here.

2.3.5 2D Al_2O_3

To illustrate the ability of the evolutionary algorithm to discover novel structures of two-dimensional materials, we apply it to search for structures of 2D Al_2O_3. We use the Vienna *ab initio* simulation package (VASP) [75, 71, 73, 74] with the projector augmented wave (PAW) method to compute DFT total energies [18]. The Perdew-Burke-Ernzerhof (PBE) generalized-gradient exchange-correlation functional is used to perform the structural relaxations [124], and we use a plane-wave energy cutoff of 283 eV and a k-point mesh density of 25 k-points per $Å^{-1}$. To limit the computational cost, a single search was performed, and 500 objective function evaluations were used by the algorithm. We constrained the layer thickness of the 2D structures to 5 Å.

The two lowest energy structures found by the algorithm are shown in Fig. 2.10. The first structure, labeled (a) in the figure, contains 5 atoms in the unit cell and corresponds to a buckled hexagonal monolayer stacked on top of a single layer with the 1T structure. The buckled hexagonal structure has been reported for silicene [171] and predicted for several group III-V monolayers [214], and some 2D materials have been shown to adopt the 1T structure, including SnS_2 [152, 212]. The second structure in the figure contains 10 atoms in the unit cell, and it can be described as two slightly distorted buckled hexagonal sheets connected by a central sublayer of oxygen atoms. Table 2.3 contains the structural information for these two predicted 2D phases.

The two predicted structures shown in Fig. 2.10 are nearly degenerate in energy: we compute their formation energies relative to the ground state bulk structure of Al_2O_3 to be within 6 meV/atom of each other, shown in Table 2.3. 2D materials that have been synthesized as free-standing monolayers usually posses formation energies below 200 meV/atom relative to their corresponding bulk phases [143]. Although the formation energies of the 2D phases of Al_2O_3 predicted here lie toward the top of that range, their synthesis as free-standing films may be feasible.

For additional examples of applying the genetic algorithm to predict the structures of 2D materials, see Refs [132, 144].

2.4 Conclusion

We developed an improved evolutionary algorithm for structure prediction capable of searching for partial phase diagrams and the structures of clusters, wires and 2D materials. For searches in which the composition is allowed to vary, we have improved the sampling of the composition space. We have also redesigned the algorithm to increase efficiency when running in parallel. Our algorithm uses default parameter values to mininmize the input needed by the user, and it takes advantage of existing open-source libraries for structure manipulation and analysis. We demonstrated the capabilities of the algorithm by applying it to find the Al-CuAl-ZrAl partial phase diagram predicted by an empirical potential, as well

as the structures of Au clusters and wires predicted by an EAM potential. We also use the algorithm, coupled with *ab initio* total energy calculations, to predict two new low-energy structures of 2D Al_2O_3. Based on its efficiency and extended capabilities, we believe the improved evolutionary algorithm is a useful addition to the toolbox of computational materials design.

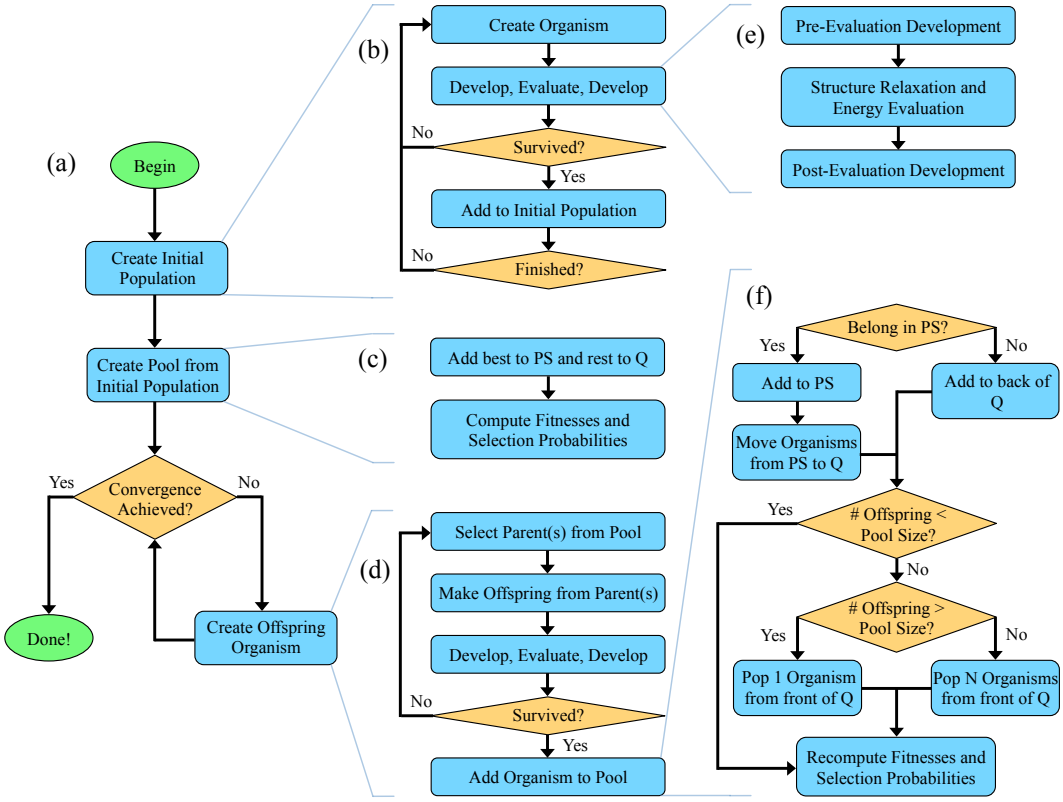

Figure 2.1: Outline of the evolutionary algorithm, neglecting parallelization. The flowchart in (a) is a high-level description of the algorithm, where the initial population is the starting group of structures, either generated randomly or provided to the algorithm, and the pool is the collection of organisms that are allowed to reproduce, representing the current population. In (b), the process of assembling the initial population is shown. The algorithm creates a structure, evaluates its energy, and adds it to the initial population. This continues until the initial population is full. The pool is comprised of the promotion set (PS) and the queue (Q), and how it is made from the initial population is shown in (c). The best few organisms are placed in the promotion set, and the rest are added to the queue. The fitnesses and selection probabilities of all the organisms in the pool are then computed. The flowchart in (d) illustrates the steps to make an offspring organism. One or two parent organisms are selected and used to make an offspring organism through one the variations (e.g., mating, mutation). Once a viable offspring organism has been made, it's energy is evaluated and it is added to the pool. In (e), the steps of development and energy evaluation are shown. Development ensures the candidate offspring organism is viable (by checking the constraints, among other things). If the candidate offspring organism survives development, it is passed to an external code for structural relaxation and energy evaluation, after which it is developed once more. The flowchart in (f) illustrates the steps for adding an organism to the pool. The algorithm first checks if an organism is good enough to go in the promotion set. If so, it is added to the promotion set, and one or more organisms previously in the promotion set may be demoted and moved to the back of the queue. Otherwise, the offspring organism is added to the back of the queue. The algorithm then checks how many total offspring organisms have been added to the pool so far. If that number is less than the specified size of the pool, then the algorithm is still filling up the pool and doesn't remove any organisms from it. If the number of added offspring is equal to the specified size of the pool, the algorithm removes the initial population from the pool by popping as many (N) organisms off the front of the queue as were in the initial population. If the number of added offspring exceeds the specified size of the pool, then the initial population has already been removed, and only one organism is removed from the pool (by popping it off the front of the queue), therefore maintaining the pool's size. The fitnesses and selection probabilities of all organisms in the pool are then recomputed.

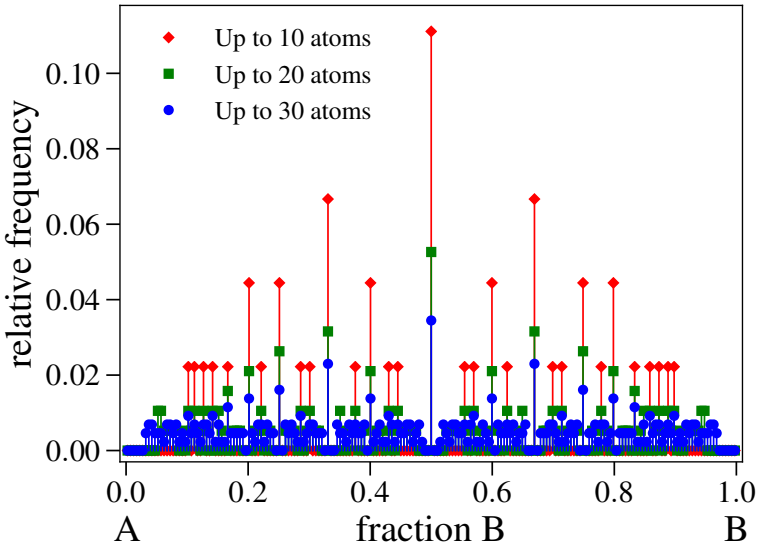

Figure 2.2: Histogram of the compositions in a binary system that can be sampled with a group of structures containing up to 10, 20 and 30 atoms, excluding the endpoints.

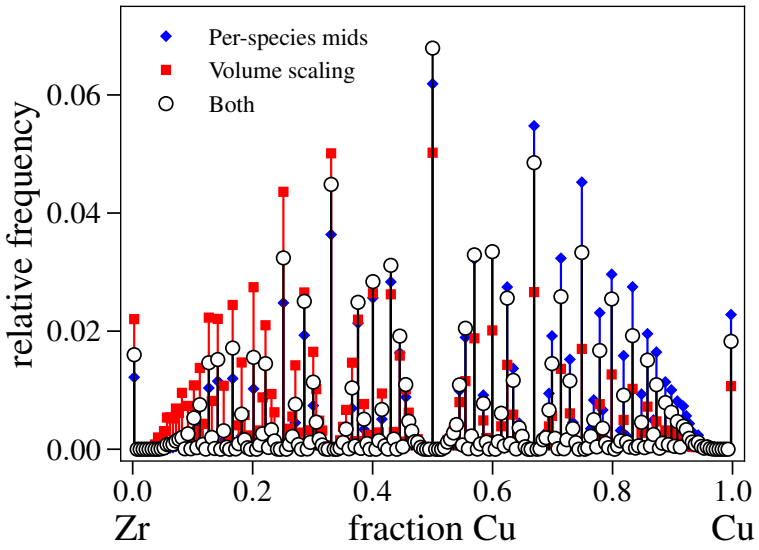

Figure 2.3: Histogram showing the distribution of compositions sampled by the offspring structures created by the genetic algorithm. Using both species-specific minimum interatomic distance constraints and composition-dependent volume scaling reduces non-symmetric bias in the sampling of the composition space.

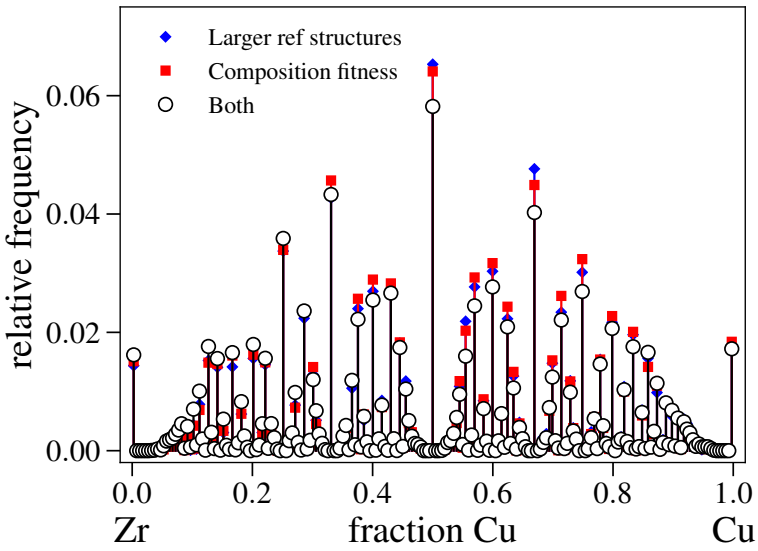

Figure 2.4: Histogram showing the distribution of compositions sampled by the offspring structures created by the genetic algorithm. Using both larger endpoint reference structures and composition-dependent fitness reduces symmetric bias in the sampling of the composition space.

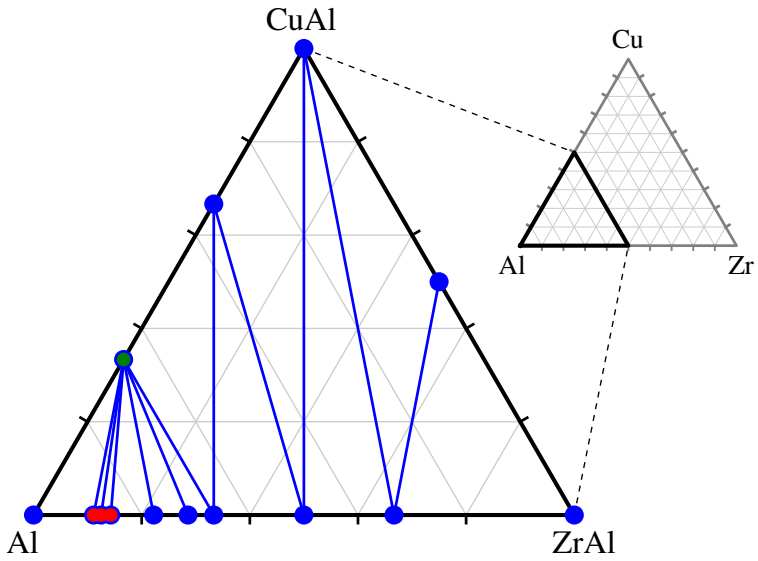

Figure 2.5: Al-CuAl-ZrAl partial ternary phase diagram predicted by the EAM potentials.

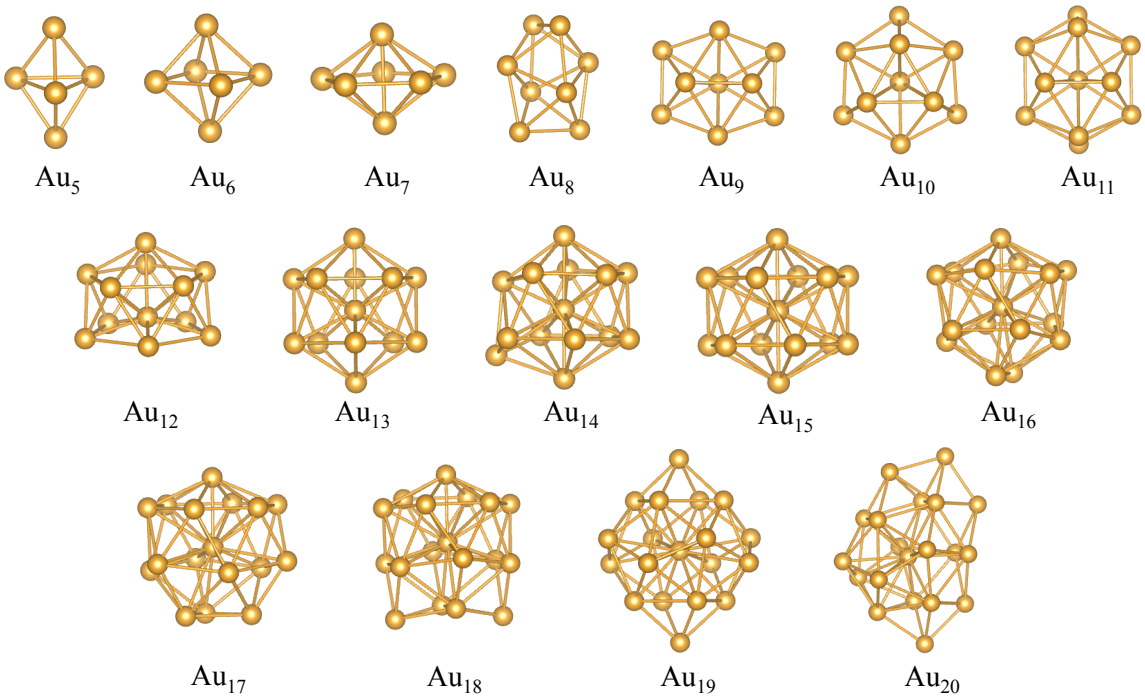

Figure 2.6: Lowest energy Au clusters containing from 5 to 20 atoms, predicted by the EAM potential.

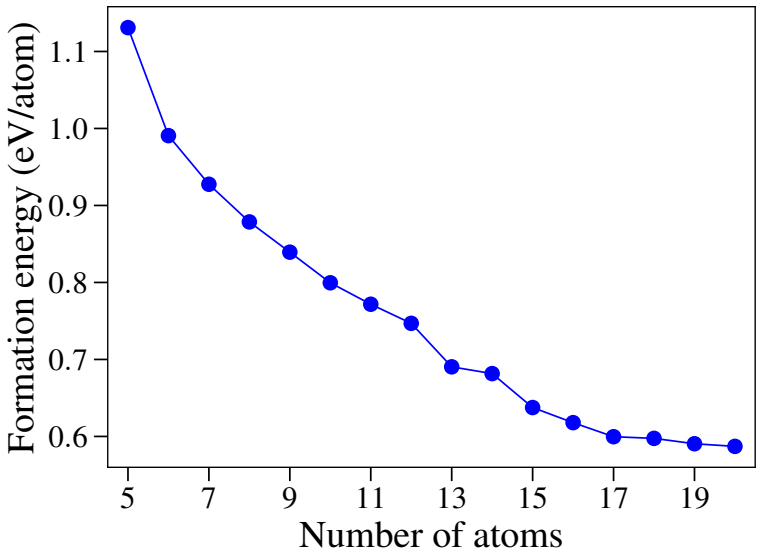

Figure 2.7: Formation energies (relative to fcc Au) of the lowest energy clusters found by the genetic algorithm, containing from 5 to 20 atoms.

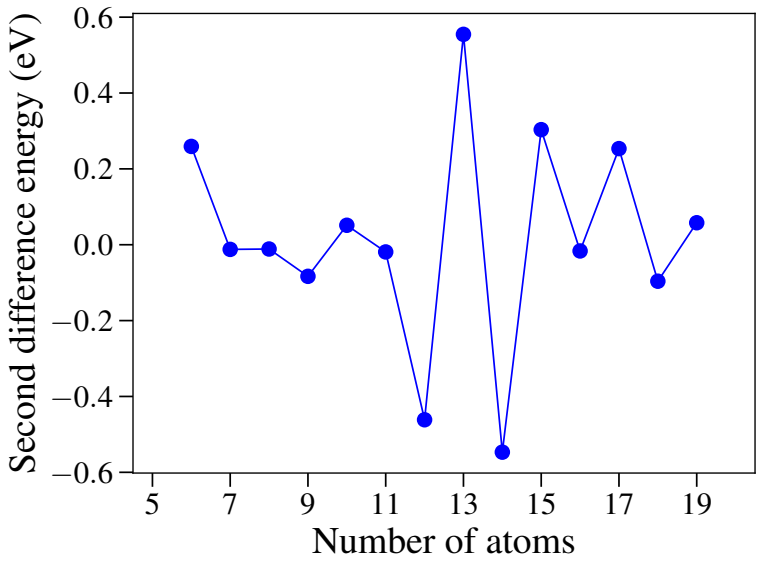

Figure 2.8: Second finite difference energies of the lowest energy clusters found by the genetic algorithm.

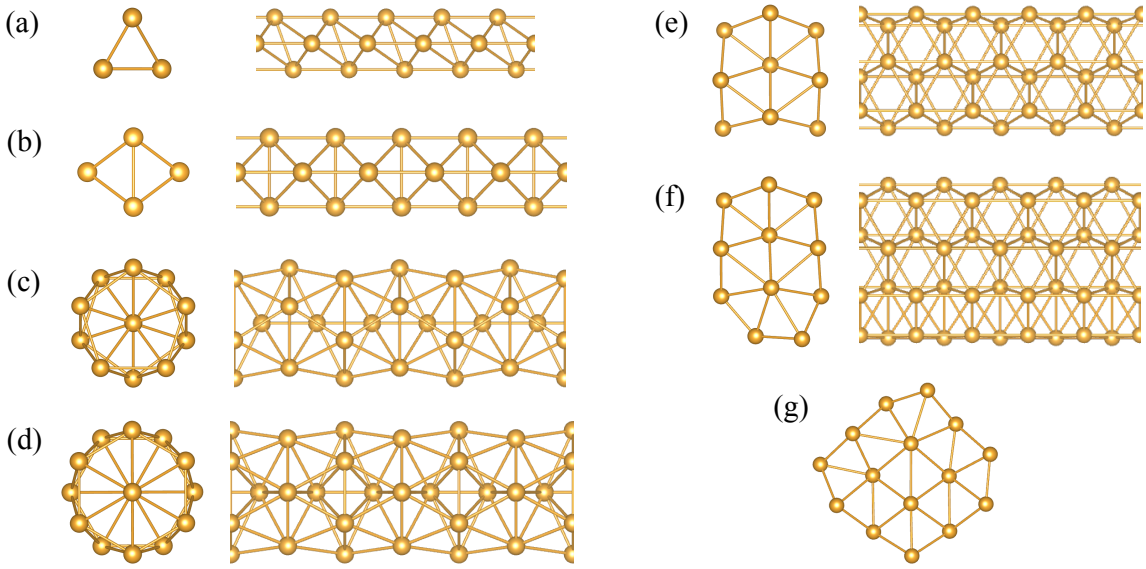

Figure 2.9: The lowest energy Au wire structures predicted with the EAM potential with maximum wire diameter constraints of (a) 3, (b) 4, (c) 5, (d) 6, (e) 7, (f) 8 and (g) 9 Å.

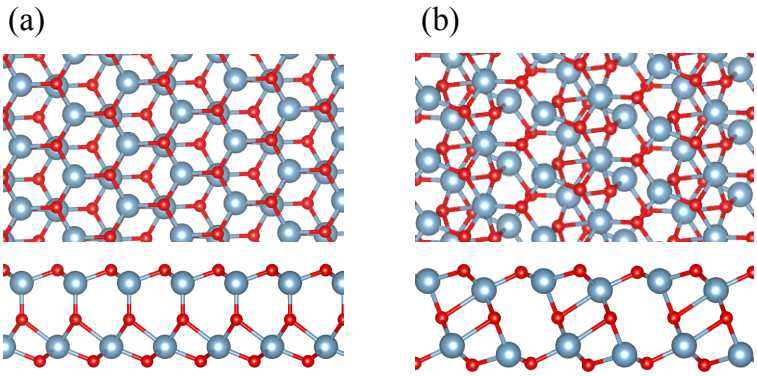

Figure 2.10: Top and side views of the 2D phases of Al_2O_3 predicted by the genetic algorithm. The structures shown in (a) and (b) display hexagonal and triclinic symmetries, respectively.

CHAPTER 3

PREDICTING TWO-DIMENSIONAL MATERIALS WITH A GRAND-CANONICAL EVOLUTIONARY ALGORITHM

The content of this chapter was previously published in Ref. [132].

3.1 Introduction

Two-dimensional (2D) materials possess properties that are distinct from those of their bulk counterparts. For example, single-layer MoS_2 exhibits a wider and more direct band gap than bulk MoS_2 [101, 148], and single-layer SnS_2 has interesting photocatalytic properties lacked by the bulk material [152, 212]. Such unique physical properties have motivated intense research interest in the field of 2D materials in the past decade. Proposed applications for 2D materials include optical sensors [101], nanoelectronics [130], and photocatalysts for water splitting [152, 212, 143].

Several new single-layer materials have been computationally predicted in recent years [136, 6, 208, 80]. In one common approach to 2D structure prediction which relies on chemical intuition and substitution, the lattice sites of known 2D crystal structures are decorated with different atomic species [208]. Another method involves mining databases of bulk crystal structures to identify those with layered motifs, from which a single layer could potentially be exfoliated [80]. These approaches are quite useful, but since they are based on previously known

structures, they make assumptions about the structures and compositions of 2D materials. These assumptions may unnecessarily constrain the search, leaving other potentially synthesizable single-layer materials awaiting discovery.

Recently, much practical success in the prediction of the structures of clusters and three-dimensional crystals has been achieved with global optimization techniques, such as evolutionary algorithms [125, 41, 159, 163, 30, 65, 131] and particle swarm optimization [175]. Some of these methods have been extended to search for 2D materials. Bahmann and Kortus developed an evolutionary algorithm that can search for 2D crystals [10], and Zhou *et al.* extended an evolutionary algorithm to search for 2D structures [198]. Both search for structures with fixed stoichiometry and number of atoms per cell. Luo *et al.* extended a particle swarm optimization algorithm to search for fixed stoichiometry 2D structures that are both completely planar [94] and have finite layer thicknesses [93]. In this work, we extend our grand-canonical evolutionary algorithm [159] to enable unbiased searching for the *compositions* and *structures* of 2D materials with low formation energies and finite layer thicknesses.

The details of the grand-canonical evolutionary algorithm for 2D materials are described in Sec. 3.2. In Sec. 3.3, we demonstrate that the algorithm enables the prediction of 2D materials structures by applying it to a system at fixed composition (2D-InP) in Sec. 3.3.1 and then to the composition space of the binary Sn-S and C-Si materials systems in Secs. 3.3.2 and 3.3.3. The algorithm successfully identifies the known low-energy 2D structures in these materials systems. In addition,

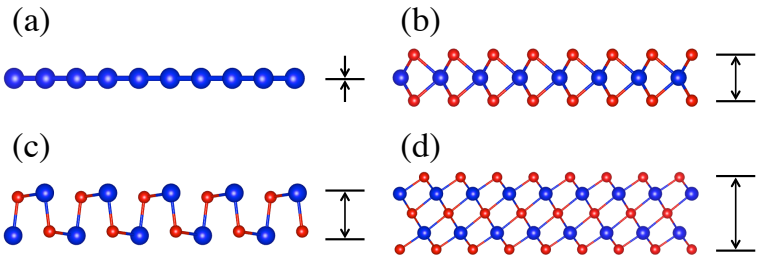

Figure 3.1: Layer thicknesses of several 2D structures: (a) graphene with a thickness of zero, (b) MoS_2, (c) PbO, and (d) the slightly thicker structure of Bi_2Te_3. The structures are shown from the side, and the layer thickness is the maximum vertical distance between atoms in the cell.

it discovers several different 2D structures of InP and a different 2D structure of C_6Si with formation energy below that of previously known 2D structures in this system.

3.2 Methods

3.2.1 Search space and objective function

The search for novel 2D materials requires the definition of a search space and an objective function. We define a 2D material to be a crystal with a structure that is periodic in two dimensions and has a finite extent in the third dimension. The layer thickness of a 2D material is the maximum vertical distance between atoms in the structure, as illustrated in Fig. 3.1. Two-dimensional materials display a range of layer thicknesses, from completely planar structures such as graphene

and hexagonal boron nitride to thicker structures with multiple sublayers, such as metal dichalcogenides (three sublayers), group-IV monochalcogenides (four sublayers), and Bi_2Te_3 (five sublayers) [101, 148, 140, 157]. To include 2D structures with different layer thicknesses in our search, we employ a finite-thickness constraint. Hence, the search space for 2D materials consists of the configuration space of all materials that are periodic in two dimensions and exhibit a finite thickness in the third dimension.

For the objective function, we propose the formation energy of a 2D material relative to the bulk ground state phase, as defined by

$$\Delta E_{\mathrm{f}} = \frac{E_{\mathrm{2D}}}{N_{\mathrm{2D}}} - \frac{E_{\mathrm{3D}}}{N_{\mathrm{3D}}}, \tag{3.1}$$

where E_{2D} and E_{3D} are the total energies of the 2D and 3D structures, respectively, and N_{2D} and N_{3D} are the numbers of atoms in the 2D and 3D unit cells [209]. Common to all synthesized free-standing 2D materials is a low formation energy with respect to the bulk ground state phase. Three well-known 2D materials illustrate this point: the formation energy of graphene is only about 56 meV/atom relative to graphite [147], that of single-layer MoS_2 is 77 meV/atom [207], and we calculated the formation energy of phosphorene to be 112 meV/atom relative to bulk phosphorous.

Figure 3.2 summarizes the formation energies of many 2D materials, both predicted and synthesized. We observe that all 2D materials that have been synthesized as free standing films possess formation energies less than 200 meV/atom [143]. Therefore, we take this as an empirical rule of thumb to de-

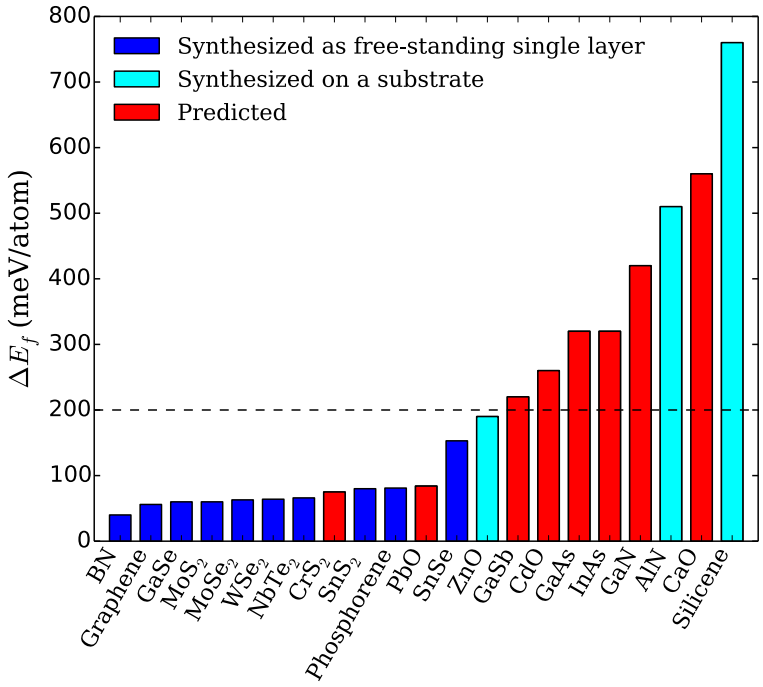

Figure 3.2: Calculated formation energies relative to the bulk of several 2D materials, both predicted and synthesized. All 2D materials that have been synthesized as free standing films have formation energies below 200 meV/atom, illustrated by the horizontal dashed line. The formation energies were calculated as follows: graphene [147] and phosphorene [139] with quantum Monte Carlo; BN [17], MoS_2 [17], $MoSe_2$ [17], WSe_2 [17], $NbTe_2$ [17] and PbO [17] with the random-phase approximation (RPA) method; GaSe [211], CrS_2 [213], SnS_2 [212] and SnSe [140] with the van der Waals density functional (vdW-DF-optB88); CdO [210], CaO [210], ZnO [210], GaAs [214], GaSb [214], InAs [214], GaN [214] and AlN [214] with the PBE functional; and silicene [209] with the local density approximation (LDA) exchange-correlation functional.

termine whether a 2D material has a realistic chance of being experimentally synthesized without the need for stabilization by a substrate [141, 142]. In Sec. 3.3.1, we discuss an alternative choice of objective function for 2D structure prediction.

3.2.2 2D Evolutionary Algorithm

To search for low-energy 2D crystal structures, we modify our grand canonical genetic algorithm for structure and phase prediction (GASP) [159] code by constraining the layer thickness of the crystal structures considered by the algorithm. In the following, we provide a brief description of the general evolutionary algorithm, and then focus in detail on the changes made for the 2D search. For a complete description of GASP and some of its applications, see Refs. [159, 131, 158, 160]. The GASP code is freely available under the GPL v3 license at http://gasp.mse.ufl.edu.

Overview. The evolutionary algorithm starts by generating a population of random structures that represent a broad sampling of the solution space. Structures are evaluated based on their relative formation energies per atom, with lower energy solutions being more favorable [163]. An offspring generation is then populated by probabilistically selecting lower energy structures to "reproduce" through biologically inspired operators such as mutation and mating. The mutation operator randomly perturbs the atomic positions and lattice vectors of the parent to create an offspring structure, and the mating operator combines spatially coherent pieces of two parent structures [22, 30, 180, 65, 1, 41, 112], as illustrated in Fig. 3.3. When enough child structures have been created, they are evaluated for their fitness and then they make an offspring generation of their own using the same evolutionary operators. This process continues until some user-defined stopping criteria are met. Since better solutions are selected more frequently to

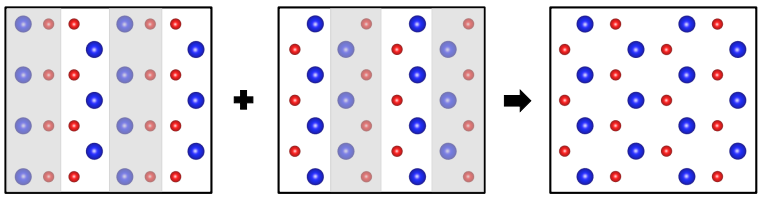

Figure 3.3: Illustration of the mating operator. Sections are sliced from each parent structure, shown on the left, and combined to form an offspring structure. Supercells are shown for clarity.

reproduce, structural traits associated with low-energy crystals are propagated in subsequent generations.

Constraints. It is generally advantageous to apply some loose constraints to the parameters of structures considered by the algorithm. Constraints can guide the search for structures with particular characteristics of interest, and they can apply prior knowledge about which structures are likely to have good objective function values. For example, we know that crystal structures with atoms quite near to each other (less than about 75% of the equilibrium bond length) will have very high formation energies, and this knowledge is put to use by enforcing a constraint on the minimum interatomic distance. Constraints on the maximum and minimum lattice vector lengths can help prevent structures with unphysically large or small aspect ratios from entering the population [126]. The composition, number of atoms in the unit cell, lattice vector angles and symmetries may also be constrained.

Structure representation. An evolutionary structure search will be more successful if structures are represented in the computer in a standardized fashion. This is

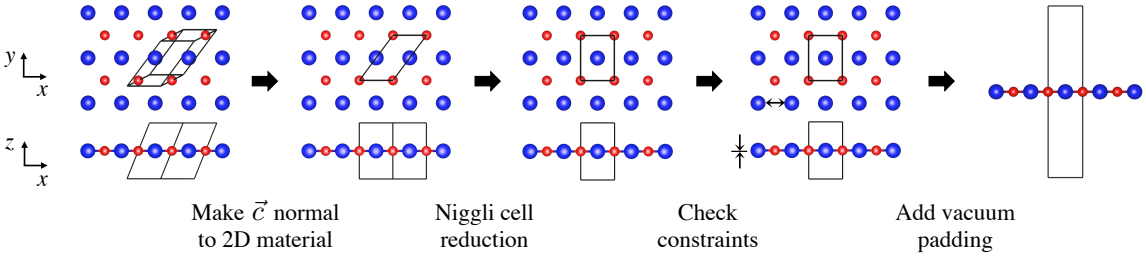

Figure 3.4: Modifications made to GASP to prepare a candidate 2D structure for an energy calculation. First the structure is rotated to lie within the *x*-*y* plane and a $\vec{c}$ lattice vector normal to the plane is chosen. The algorithm then applies Niggli cell reduction to obtain the most cubic representation of the structure. The reduced structure is checked against the constraints, including the layer thickness constraint. If it passes, vacuum padding is added to the cell before it is relaxed and its energy evaluated by an external code. Afterwards, the vacuum padding is removed from the relaxed cell, Niggli cell reduction is applied, and the constraints are checked again before it is added to the offspring generation.

because the mating operator is more likely to produce viable offspring if the parent structures are represented similarly [159]. In addition, it is easier to identify duplicate structures if each structure has a unique representation. The application of constraints (discussed above) to the lattice vector lengths and angles is one way to help standardize how structures are represented. Another useful method is the Niggli cell reduction algorithm [76], which essentially transforms the representation of a structure such that its unit cell is closest to a cube.

Modifications for 2D structure search. To facilitate searching for 2D structures, we make three modifications to our evolutionary algorithm: (i) We impose a constraint on the layer thickness of structures, (ii) we modify the structure representation to be suitable for 2D materials, and (iii) we add a vacuum layer to the structures for the energy evaluation. Figure 3.4 illustrates these modifications made to

GASP to search for 2D structures.

First, a new constraint is imposed on the layer thickness of structures, which limits the search to the 2D regime. This constraint is an input parameter, so the user may change it to search for thinner or thicker structures as considered appropriate for the given material system. The algorithm checks all structures it generates against this new constraint, both before and after structural relaxation.

The second modification involves how 2D structures are represented. Once a new structure is created, the algorithm rotates the cell such that lattice vector $\vec{a}$ is parallel to the cartesian x-axis and lattice vector $\vec{b}$ lies in the (x, y) plane. Since 2D materials are not periodic in the third dimension and have a finite thickness, no lattice vector along the third dimension is required. However, for compatibility with energy codes that are designed for 3D crystal structures, it is convenient to select a $\vec{c}$-lattice vector that is normal to the $(\vec{a}, \vec{b})$ plane. This is achieved by replacing the $\vec{c}$-lattice vector with its component along the cartesian z axis. The $\vec{c}$-lattice vector is fixed during structural relaxation to reduce the computational cost of the relaxation, avoid collapse of the vacuum region, and to prevent a sheet from sliding relative to its periodic images during relaxation, which could lead to spurious minima.

Our choice of the $\vec{c}$-lattice vector also improves the success of the mating operator, which combines slices of unit cells that are taken parallel to one of the cell facets [159, 131]. When the $\vec{c}$-lattice vector is normal to the 2D material structure, the slice plane is always either parallel or perpendicular to the plane of the

2D sheet in both parent structures, increasing the chances that the result will also correspond to a valid 2D structure.

When applying the Niggli cell reduction [76] to 2D structures, the reduced cell must correspond to a 2D structure. This means that the algorithm should only transform the $\vec{a}$ and $\vec{b}$ lattice vectors and, furthermore, $\vec{a}$ and $\vec{b}$ should remain in the *x-y* plane. To achieve this in practice, we simply increase the magnitude of the $\vec{c}$ lattice vector to an arbitrary large value before passing the cell to the Niggli reduction algorithm. After reduction, the $\vec{c}$ lattice vector is returned to its original magnitude.

The final modification is related to the evaluation of the energy of the 2D structures using codes that are designed for 3D bulk structures and employ periodic boundary conditions. Before passing a newly created 2D structure to an external code for relaxation and energy calculation, the algorithm sets the $\vec{c}$ lattice parameter of the unit cell such that the spacing between periodic images of the 2D material are sufficiently separated by vacuum to prevent spurious interactions between the periodic images.

3.2.3 Density-functional Calculations

To accurately relax the 2D candidate structures and determine their energy, we perform density-functional theory (DFT) calculations with the Vienna *ab initio* simulation package (VASP) [75, 74]. The interactions between valence elec-

Table 3.1: Parameters of the DFT calculations to converge the energy within 1 meV/atom for the three materials systems under consideration.

System	Cutoff energy (eV)	k-points density (Å)	Vacuum padding (Å)
InP	300	35	16
Sn-S	550	55	12
C-Si	500	50	12

trons and ionic cores are described by the projector-augmented wave (PAW) method [18, 72]. The core electron states described by the PAW potentials are $1s^2$ for C, $1s^2 2s^2 2p^6$ for Si, P, and S, [Kr]$4d^{10}$ for Sn, and [Kr] for In. Table 3.1 summarizes the values of the DFT parameters used for the three considered materials systems. For each system, the energy was converged to within 1 meV/atom for each parameter relative to the energy obtained with a maximum reference value of the parameter. The maximum reference values for the cutoff energy, plane-wave basis set and vertical vacuum padding were 700 eV, a k-point density of 70 per Å^{-1}, and 20 Å, respectively. All structural relaxations during the evolutionary algorithm search used the Perdew-Burke-Ernzerhof (PBE) approximation for the exchange-correlation functional [124]. Following the structure search, we perform structural relaxations and calculate the energies of all 2D structures found by the algorithm, as well as of the known ground-state bulk structures, using the computationally more demanding non-local vdW-DF-optB88 exchange-correlation functional [31, 134, 67]. This accurately accounts for the dispersion interactions that are important for the layered bulk structures of some of the materials. The $\vec{c}$ lattice vector is kept fixed during all structural relaxations.

For all structure searches, the number of atoms in the simulation cell was allowed to vary. Equilibrium bond lengths were obtained by relaxing the known ground state bulk structures in each system, and these were used to calculate values for the minimum interatomic distance constraints, as discussed in Sec. 3.2.2. For the InP 2D structure searches, we fixed the stoichiometry of all structures encountered in the search to In:P = 1:1 and employed an upper bound of 12 atoms in the cell to limit the size of the search space and the computational cost of the DFT calculations. To explore the energy landscape as a function of layer thickness, we carried out a total of five evolutionary algorithm structure searches for InP, each with a successively larger layer thickness constraint, ranging from 2 to 6 Å. In each search, the algorithm was stopped after 500 successful structural relaxations and subsequent energy calculations. For the Sn-S and C-Si structure searches, we allowed the stoichiometry to vary between the pure elements, and 1000 relaxations and energy calculations were performed in the structure search for each of these materials systems. For the Sn-S search, we chose a layer thickness constraint of 4 Å, which is slightly larger than the layer thicknesses of the known 2D structures in this system. Structures were permitted with up to 15 atoms in the cell. For the C-Si search, we used a fairly conservative layer thickness constraint of 2 Å because the prevously predicted 2D structures in this system have all been nearly or completely planar. The known elemental 2D structures of planar hexagonal graphene and buckled hexagonal silicene, with layer thicknesses of 0.00 Å and 0.45 Å, respectively, were provided to the algorithm in the initial generation. Up to 12 atoms per cell were allowed in the C-Si search.

Table 3.2: Structural parameters and formation energies of six low-energy 2D InP materials and the lowest energy C-Si material found by the evolutionary algorithm. We have used 3D space groups to describe these finite-thickness 2D structures that lack periodicity in the direction normal to the 2D sheet. In the representations given here, the $\vec{c}$ lattice vector is normal to the plane of the 2D sheet. Symmetry information was obtained with the FINDSYM software package [150].

	Space group	a (Å)	b (Å)	c (Å)	Atomic positions		ΔE_f (meV/atom)
InP (i)	$P3m1$ (156)	4.209	4.209	3.146	In 1(b)	$z = -0.413$	571
					P 1(c)	$z = 0.413$	
InP (ii)	$Pmn2_1$ (31)	4.095	6.057	5.210	In 2(a)	$y = 0.290$, $z = -0.447$	484
					P 2(a)	$y = -0.388$, $z = 0.250$	
InP (iii)	$Abm2$ (39)	4.902	6.770	5.547	In 4(c)	$y = 0.214$, $z = 0.299$	421
					P 4(c)	$y = 0.340$, $z = -0.234$	
InP (iv)	$P\bar{3}m1$ (164)	4.240	4.240	6.317	In 2(d)	$z = 0.322$	356
					P 2(d)	$z = -0.206$	
InP (v)	P-1 (2)	5.039	5.979	7.414	In 2(i)	$x = 0.459$, $y = -0.316$, $z = -0.175$	322
					In 2(i)	$x = 0.461$, $y = 0.179$, $z = -0.186$	
					P 2(i)	$x = -0.047$, $y = 0.424$, $z = -0.374$	
					P 2(i)	$x = -0.209$, $y = 0.165$, $z = -0.488$	
InP (vi)	$C2/m$ (12)	7.115	12.572	8.278	In 8(j)	$x = 0.322$, $y = 0.351$, $z = -0.343$	310
					In 4(i)	$x = 0.375$, $z = -0.244$	
					P 8(j)	$x = 0.311$, $y = 0.341$, $z = 0.335$	
					P 4(i)	$x = 0.289$, $z = 0.157$	
C_6Si (ii)	$Cmm2$ (35)	16.600	2.456	2.842	C 4(d)	$x = 0.400$, $z = 0.404$	414
					C 4(d)	$x = -0.229$, $z = 0.282$	
					C 4(d)	$x = -0.142$, $z = 0.346$	
					Si 2(b)	$z = -0.282$	

3.3 Results and Discussion

3.3.1 Indium Phosphide

Bulk indium phosphide is a direct gap semiconductor and occurs in the zincblende crystal structure [61]. A novel tetragonal structure of 2D InP was recently proposed by Zhuang *et al.*, and its formation energy was calculated to be comparable to that of the 2D buckled hexagonal structure previously proposed

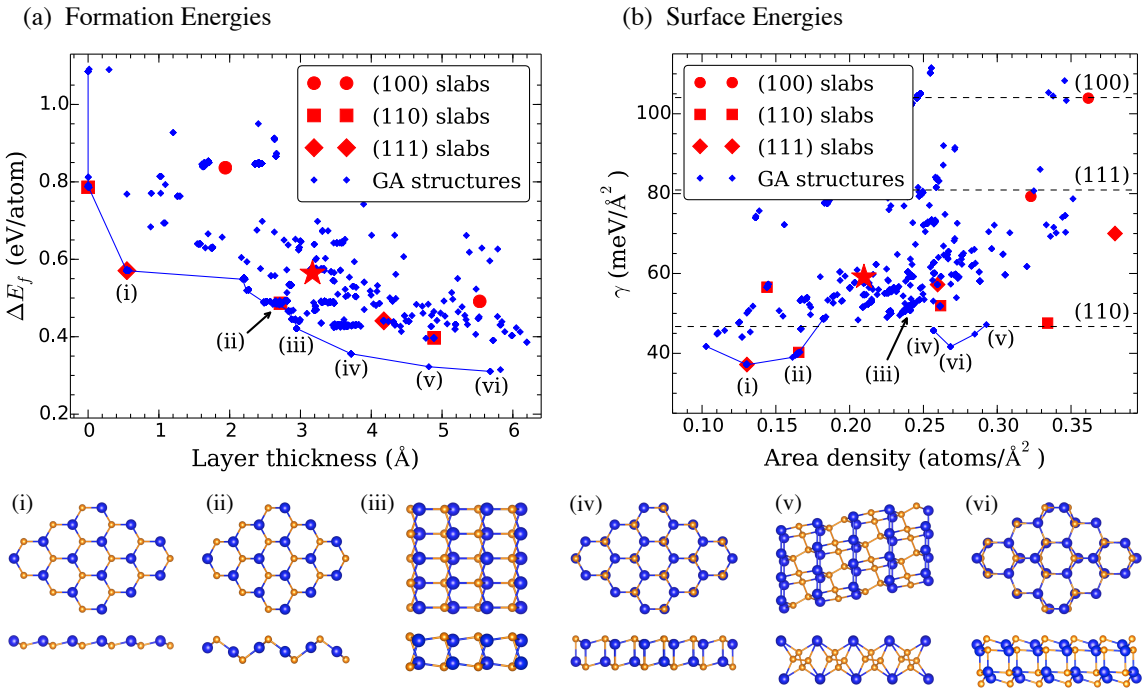

Figure 3.5: Structure search for 2D InP. In both plots, the small blue diamonds indicate 2D structures found by GASP, and the larger red symbols correspond to 2D structures that were obtained by taking slabs from high symmetry planes of the bulk InP structure. The large red star near the center of each plot represents the previously proposed tetragonal structure [214]. In (a), the formation energies of the 2D structures with respect to the bulk phase is plotted vs their layer thicknesses. The blue line connects the lowest energy structures found by the algorithm at various thicknesses. Several structures that lie on this line are labeled (i) through (v) in both plots. In (b), the surface energies of the 2D structures are plotted vs their area densities. The three horizontal dashed lines designate the calculated surface energies of three facets of the bulk material. Top and side views of the low-energy 2D structures (i) to (vi) found by the algorithm are shown below the plots.

by Şahin *et al.* [214, 136]. Tong *et al.* [161] predicted a new 2D structure of InP to be lower in energy than both the buckled hexagonal and tetragonal structure, and Susuki [155] recently reported a novel 2D structure of InP with the same layer

thickness as the structure reported by Tong *et al.* and nearly degenerate formation energy. To determine if other low-energy 2D InP structures exist, we employ the evolutionary algorithm to do several 2D structure searches with varying layer thickness constraints. Figure 3.5 displays the results of these 2D structure searches.

Since we are searching for 2D materials, we can treat the layer thickness as a second objective function that we seek to minimize (that is, in addition to the formation energy). The set of solutions that optimize the tradeoff between layer thickness and formation energy form the Pareto front for this system. These structures are connected by a blue line in Fig. 3.5(a), which shows the formation energy versus layer thickness.

Several low formation energy structures that lie on the Pareto front are shown in Fig. 3.5 below the plots, labeled (i) through (vi). The buckled hexagonal structure shown in Fig. 3.5(i) was previously proposed by Şahin *et al.* [136] and was also recovered by relaxing a monolayer from the (111) plane of the bulk crystal. The orthorhombic structure shown in Fig. 3.5(ii) has been reported by Tong *et al.* [161] and is also obtained by relaxing a bilayer from the (110) plane of the bulk crystal. The structure labeled Fig. 3.5(iii) has orthorhombic symmetry and eight atoms in the cell, and the hexagonal bilayer structure in Fig. 3.5(iv) is equivalent to two stacked buckled hexagonal structures, with one inverted. The triclinic structure shown in Fig. 3.5(v) has eight atoms per cell, and Fig. 3.5(vi) is a more complex bilayer structure with monoclinic symmetry and 12 atoms in the cell. Table 3.2

contains the parameters for these six structures. The previously proposed tetragonal structure [214] is also shown, and it is located above the Pareto front.

We compare the results of the structure searches with an alternative approach to generating candidate 2D materials, in which thin slabs from the bulk structure are relaxed. Slabs of the bulk zinc-blende structure from the (100), (110), and (111) planes were padded with vacuum and relaxed, and the resulting 2D structures are shown as red symbols in Fig. 3.5. The low energy of several of the relaxed slabs of the bulk material shows that this simple approach can efficiently provide useful candidates for the prediction of novel 2D structures – it recovered two low formation energy structures lying on the Pareto front of this system. However, this technique misses the lowest energy structures available at higher thicknesses, labeled (iii) - (vi) in Fig. 3.5(a).

All of the 2D InP structures found by the evolutionary algorithm have formation energies greater than 300 meV/atom, indicating that synthesizing them as free standing films is unlikely [143]. However, it may be possible to stabilize some of them on appropriate substrates [141].

As can be seen in Fig. 3.5(a), the lowest formation energy found by the algorithm continuously decreases as the layer thickness increases. In other words, thicker slabs tend to be thermodynamically favored, and the bulk structure is the thermodynamic ground state. We expect most materials systems to follow this general pattern. For this reason, formation energy relative to the bulk is not an ideal objective function when searching for physically-realizable 2D materials,

and we must enforce a layer thickness constraint (see Sec. 3.2 B) to prevent the algorithm from finding ever-thicker structures.

A possible solution to this issue is to use the surface energy of 2D structures as the objective function. The surface energy of a 2D material is given by

$$\gamma = \frac{N_{2\mathrm{D}}}{2A}\Delta E_{\mathrm{f}}, \tag{3.2}$$

where N_{2D} is the number of atoms in the 2D cell, $A = |\vec{a} \times \vec{b}|$ is the area of the $(\vec{a}, \vec{b})$ facet of the 2D cell, and ΔE_f is the formation energy relative to the bulk of the 2D material. Our definition of surface energy is motivated by considering slabs cut from a bulk crystal. For thick slabs, γ corresponds to the surface energy of the bulk material. For thinner slabs (with layer thickness less than about 10 Å), surface reconstructions are no longer constrained to match the underlying bulk structure, and additional reconstructions can occur through the interactions between the top and bottom surfaces. The 2D structure with the most optimal objective function value, as defined in Eq. (2), corresponds to the lowest-energy reconstruction.

Figure 3.5(b) shows the surface energies of all 2D InP structures found by the algorithm versus their area densities. The choice of area density is motivated by experimental considerations as it is an experimentally accessible growth parameter, in contrast to the layer thickness. For this reason, and to provide an additional perspective of the 2D InP structures, we plot the surface energy as a function of area density rather than layer thickness. First of all, the six low-energy structures, i.e., Figs. 3.5(i) - 3.5(vi), also display low surface energies. Second, as can be seen in the figure, there are two minimums in surface energy with respect to area density,

one around 0.14 atoms/Å^2 and a second one at about 0.27 atoms/Å^2. The structures near the low-density minimum correspond to single-layer structures, while the higher-density minimum is composed of structures that are better described as bilayer structures.

We note that the lowest surface energies in Fig. 3.5(b) are about 40 meV/Å^2, which is roughly four times the values calculated by T. Björkman *et al.* for many transition metal dichalcoginides, whose bulk structures comprise weakly bonded layers [17]. The structure of bulk InP forms a 3D bonded network and lacks layered motifs, which is likely responsible for the high surface energies of the 2D structures. Due to their high surface energies, it may not be feasible to synthesize these 2D InP structures as free-standing films, although perhaps stabilization on substrates is viable.

Overall, the surface energy provides an alternative objective function that does not require imposing a thickness constraint during the evolutionary algorithm search for 2D materials. This objective function could be extended to variable composition 2D structure searches as well.

3.3.2 Tin-Sulfur 2D Phase Diagram

The grand-canonical evolutionary algorithm enables the search for both composition and structure of unknown materials. To test this capability, we apply the algorithm to the binary Sn-S system, which exhibits two previously studied 2D

materials, SnS and SnS_2, at intermediate composition. A two-dimensional structure of SnS_2 with potential application as a photocatalyst for water splitting has recently been synthesized by Sun *et al.* [152] and computationally characterized by Zhuang *et al.* [212]. Tritsaris *et al.* and Singh *et al.* predicted the electronic properties of single layer SnS, and calculated the binding energy between adjacent layers to be only 56 meV per unit cell [166] and the formation energy to be 144 meV/atom [140].

To find the lowest energy structures across the composition range, we use the phase diagram searching mode of the evolutionary algorithm, as described by Tipton *et al.* [159]. Two main modifications to the algorithm are required to search for low-energy structures and compositions across a phase diagram. First we permit structures with arbitrary stoichiometries in the search. Second, the objective function is replaced; instead of using the energy per atom, the objective function is now defined as a structure's vertical distance from the lowest convex hull known to the algorithm.

Figure 3.6 shows the energies of the 2D structures in the Sn-S system found by the evolutionary algorithm relative to the energies of the ground state structures of bulk Sn and S. To visualize the formation energy of the 2D structures with respect to the bulk Sn-S phases at all compositions, the convex hull of the bulk structures is shown as well. The evolutionary algorithm recovers both the high [133] and low [190, 202] buckled hexagonal structures of stanene, labeled (i) and (ii), respectively. The known 2D structures of SnS and SnS_2 were also recov-

ered by the algorithm. 2D SnS occurs in the distorted rocksalt structure [166, 140] and 2D SnS_2 displays the 1*T* structure common to many transition-metal dichalcogenides [212]. These two structures have the lowest formation energy relative to the bulk phase of any of the 2D structures found by the algorithm in this system. The lowest energy structure of pure S found by the algorithm is not a 2D structure at all, but rather consists of rows of 1D polymeric chains lying in a plane. This structure is shown in Fig. 3.6(v).

3.3.3 Carbon-Silicon 2D Phase Diagram

Graphene has been the subject of intense research since it was first successfully synthesized by Novoselov *et al.* [110]. Two-dimensional silicon, or silicene, has also garnered attention more recently [171, 7, 45, 86]. In addition, nanosheets of SiC have been experimentally reported [85], and two different structures of buckled SiC monolayers have been predicted by Menon *et al.* [107]. Li *et al.* [83] and Zhou *et al.* [197] have predicted completely planar structures of single-layer SiC_2. Because of the potential existence of several 2D structures in the C-Si system, we studied it using a 2D phase diagram search. We seeded the algorithm only with the known structures of planar hexagonal graphene and buckled hexagonal silicene in the initial generation. No other known 2D structures were provided.

Figure 3.7 shows the energies of the 2D structures in the C-Si system found by the evolutionary algorithm relative to the energies of the ground state structures

of bulk C and Si. Similar to Fig. 3.6 for the Sn-S system, we show the convex hull of the bulk and 2D C-Si structures. Planar hexagonal graphene is labeled as Fig. 3.7(i). The lowest energy C-Si structure found by the algorithm, labeled Fig. 3.7(ii), consists of nanoribbons of graphene joined by rows of Si atoms. The parameters for this structure are given in Table 3.2.

The structures labeled Figs. 3.7(iii) - 3.7(vi) have all been previously reported in the literature and were all recovered by our evolutionary algorithm. Figure 3.7(iii) is a structure of C_2Si, which consists of Si atoms bonded to four C atoms in a plane; this structure was predicted by Li *et al.* [83]. A lower energy planar hexagonal structure of C_2Si, labeled Fig. 3.7(iv), was predicted by Zhou *et al.* [197]. A planar hexagonal structure of CSi is labeled as Fig. 3.7(v); Menon *et al.* [107] predicted a slightly buckled version of this structure. We found that the buckled version does not correspond to a local minimum in the energy landscape defined by the Hamiltonian (VASP-PBE) used in our study. Another lower energy planar hexagonal structure of CSi, consisting of alternating C and Si atoms, is labeled as Fig. 3.7(vi); Lin [85] reported the synthesis of nanoflakes consisting of a few layers of this structure. The structure of buckled hexagonal silicene is labeled as Fig. 3.7(vii).

With the exception of graphene, all of the 2D structures shown in Fig. 3.7 lie at least 400 meV/atom above the convex hull of the bulk system. We also note that the convex hull for 2D structures in this system is significantly skewed with respect to the bulk convex hull. A comparison of the calculated formation ener-

gies of the structures at the endpoints of the 2D convex hull illustrates this point: graphene has a formation energy of only 69 meV/atom, while silicene has a formation energy of 754 meV/atom. Due to this gradient in thermodynamic instability across the composition range, we predict that carbon-rich 2D structures in this system hold the most promise for experimental synthesis.

Examining more closely the carbon-rich 2D structures found by the evolutionary algorithm, we identify three families of low-energy structures in this region of the phase diagram; representative members are illustrated in Fig. 3.8. Each of the three families corresponds to a different Si defect in graphene, with the members of a family having different defect densities. Figures 3.8(a) and 3.8(b) show that the first two defects are substitutional Si atoms, either isolated or arranged in rows. Figure 3.8(c) shows that the third defect consists of 1D chains of fourfold coordinated Si atoms; each Si atoms is bonded to two other Si atoms along the straight chain and two C atoms. This defect interrupts the hexagonal graphene lattice and leads to out-of-plane distortions.

To analyze the formation of these defects, we consider the defect formation reaction

$$C_n + Si_m \rightleftharpoons C_nSi_m, \tag{3.3}$$

where C_n represents n carbon atoms forming graphene, Si_m represents m silicon atoms forming bulk diamond cubic silicon, and C_nSi_m stands for the defective graphene structure, containing n carbon atoms and m silicon atoms. Given the

reaction above, we determine the defect formation energy per Si atom as

$$\Delta E^{f}_{\mathrm{def}} = \frac{E(\mathrm{C}_n\mathrm{Si}_m) - mE(\mathrm{Si}) - nE(\mathrm{C})}{m}, \tag{3.4}$$

where $E(\mathrm{C}_n\mathrm{Si}_m)$ is the total energy of the defective graphene structure, $E(\mathrm{Si})$ is the energy per atom of bulk silicon, and $E(\mathrm{C})$ is the energy per atom of graphene.

Figure 3.9 shows the formation energies per Si atom as a function of Si fraction for the three types of defects in graphene identified with the evolutionary algorithm. Using the leftmost two data points in each curve, we estimate the formation energies of the isolated defects by linearly extrapolating to zero Si fraction. The resulting estimates of the defect formation energies are 4.4 eV/Si for an isolated substitutional Si atom, 3.6 eV/Si for a row of substitutional Si atoms, and 1.8 eV/Si for a chain of fourfold coordinated Si atoms. Interestingly, the formation energies of both 1D Si defects are lower than that of isolated Si substitutions, confirming the previously observed tendency of Si defect atoms to cluster in graphene [153]. In addition, the 1D defect formed by a chain of four-fold coordinated Si atoms exhibits a significantly lower formation energy than the substitutional Si defects. These results demonstrate that the evolutionary algorithm can be employed to discover defect structures in 2D materials.

3.4 Conclusion

We developed a grand canonical evolutionary algorithm for discovering low-energy structures in the emergent class of 2D materials. The algorithm enables both fixed and variable composition structure searches for 2D materials with finite thickness. The constraint on the layer thickness of the 2D structures in the search is tunable by the user. We applied the algorithm to search for 2D structures of InP, and it recovered the known buckled hexagonal structure, as well as several other bilayer structures with lower formation energies. We further carried out variable composition searches on the Sn-S and C-Si binary systems, and the algorithm recovered the previously reported 2D structures in both of these systems. For the C-Si system, the algorithm also finds several structures corresponding to Si defects in graphene, including two 1D defects with formation energies below that of a substitutional Si atom in graphene. Based on these successes, we believe the evolutionary algorithm for structure prediction is a useful tool to take the first step toward the computational discovery and design of novel 2D materials.

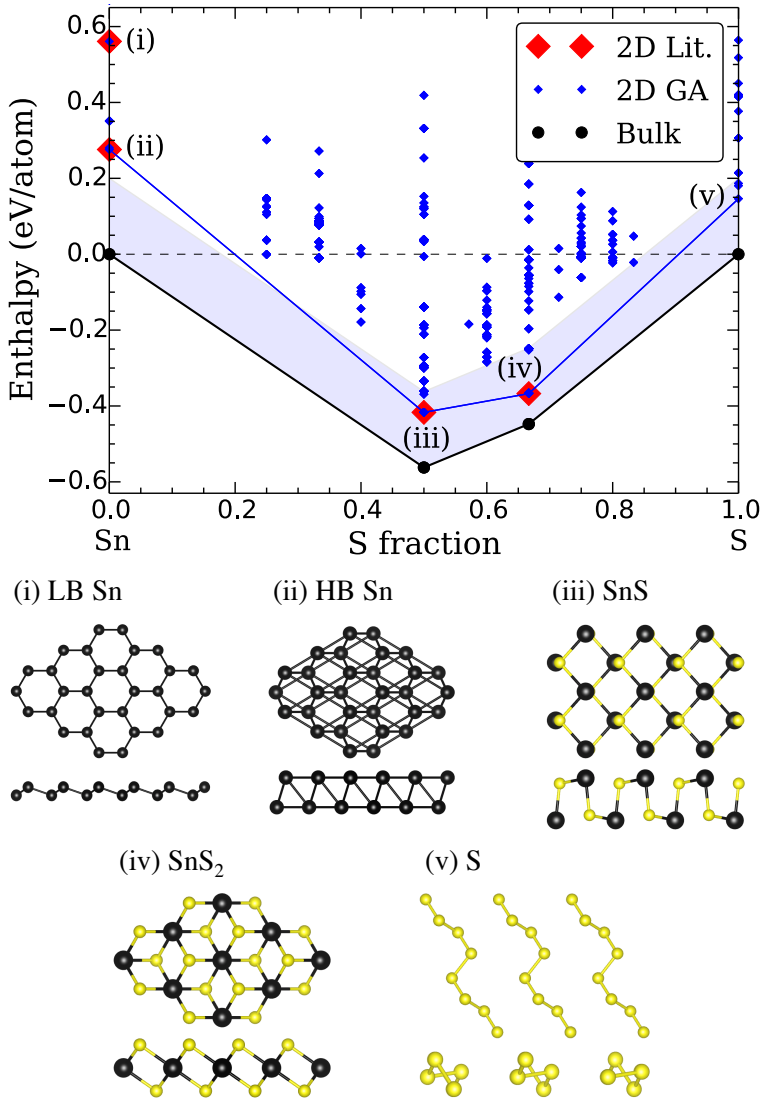

Figure 3.6: Results of the search for 2D structures in the Sn-S system. The black circles denote the ground state bulk structures, and the lines connecting them form the convex hull for the bulk system. The light blue shading indicates the region less than 200 meV/atom above the bulk convex hull. The blue diamonds denote 2D structures found by the evolutionary algorithm, and the blue line segments form the convex hull for the 2D structures. The red diamonds, labeled (i) - (iv), denote 2D structures that have previously been reported in the literature. Top and side views of the structures labeled in the plot are shown below.

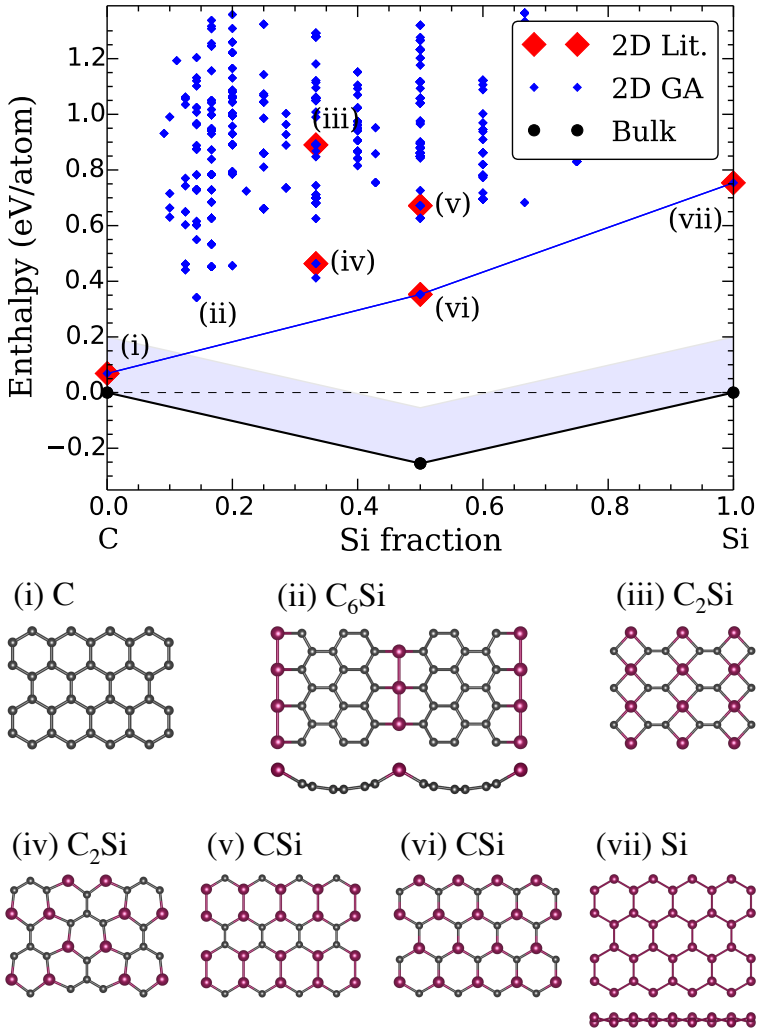

Figure 3.7: Results of the search for 2D structures in the C-Si system. The symbols have the same meanings as in Fig. 3.6. The structures labeled (i) - (vii) in the plot are illustrated below. For the completely planar structures (i) and (iii)-(vi), side views are omitted.

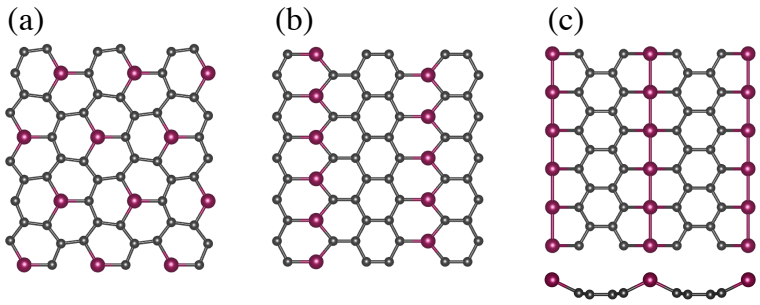

Figure 3.8: Examples of three types of Si defects in graphene found by the evolutionary algorithm. The structures in (a) and (b) are completely planar and can be described as graphene with substitutional Si atoms. In (a), the substitutional Si atoms are located as far from each other as possible, while in (b) they are arranged in rows. In (c), the defect consists of 1D chains of fourfold coordinated Si atoms that are bonded to each other and two C atoms,with a slight distortion of the planarity of the graphene sheet.

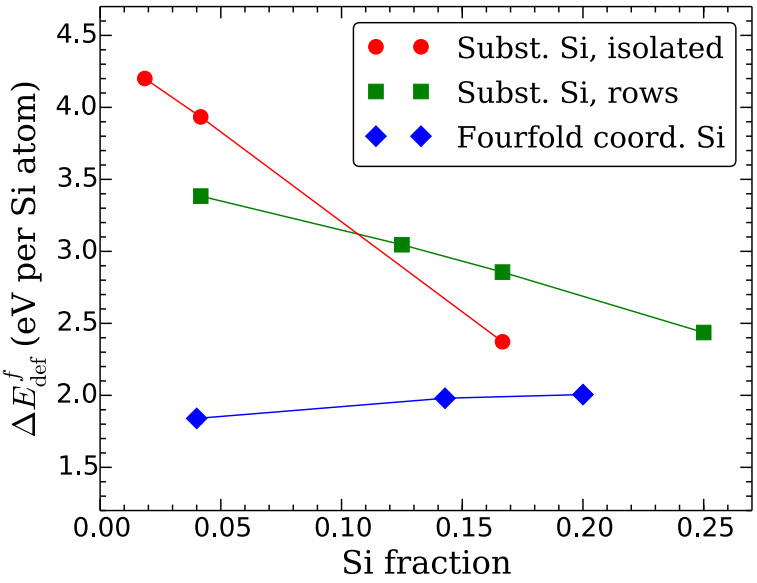

Figure 3.9: Defect formation energies per Si atom for three types of defects found by the evolutionary algorithm in graphene, as a function of Si concentration.

CHAPTER 4

GENETIC ALGORITHM PREDICTION OF TWO-DIMENSIONAL GROUP-IV DIOXIDES

Much of the content of this chapter was previously published in Ref. [144].

4.1 Introduction

Two-dimensional (2D) crystalline materials exhibit a periodic structure in two dimensions and a finite extent in the third dimension [208, 143, 132]. These materials are of great interest as they maximize their surface area, display large quantum confinement, and exhibit different symmetries compared to their bulk counterparts [208, 109]. Quantum confinement generally increases the band gap of 2D materials compared to corresponding bulk materials [208, 109]. This leads, for example, to increased photocatalytic activity in 2D SnS_2 compared to the bulk material [152, 212] and enhanced photoluminescence in single-layer MoS_2 due to its direct band gap [101, 148]. Several non-piezoelectric bulk materials lose their inversion symmetry when reduced to 2D form and thus become piezoelectric [19, 185, 213]. As a consequence, 2D materials potentially have a wide variety of applications in opto-electronic devices, sensing applications, and energy conversion technologies [111, 104, 190].

Many materials systems exhibit metastable 2D phases. Computational techniques such as density-functional theory (DFT) offer a way to rapidly determine

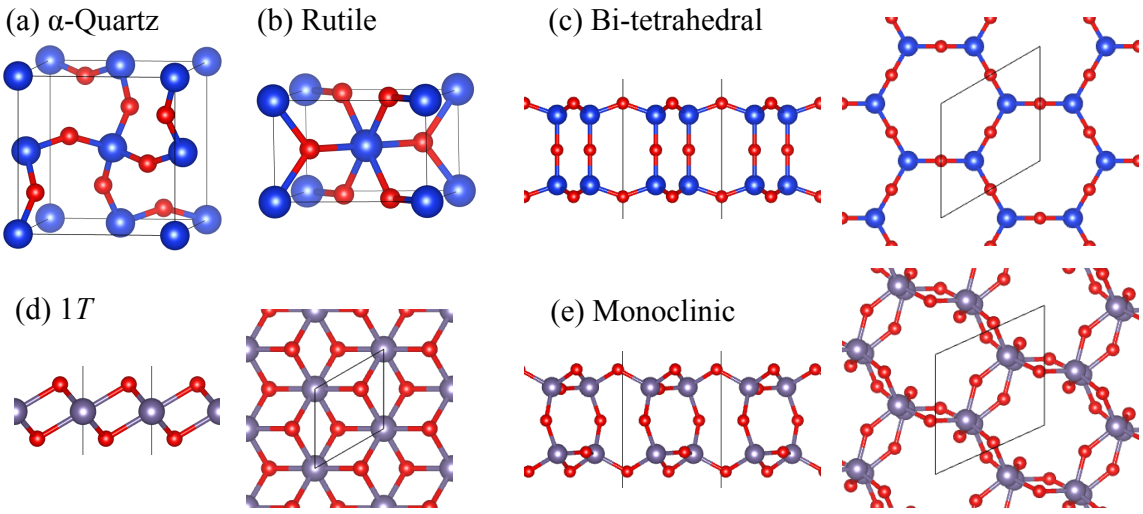

Figure 4.1: Bulk crystal structures of (a) α-quartz SiO_2 and (b) rutile AO_2 (A = Ge, Sn, Pb) and 2D structures of (c) bi-tetrahedral SiO_2, (d) 1T AO_2 (A = Ge, Sn, Pb) and (e) monoclinic GeO_2.

the stability of hypothetical 2D materials and characterize their properties to identify potentially useful 2D phases. As an example, DFT previously predicted the 2D phase of GaN and its stability and structural relaxation on various substrates [214, 141, 142]. Recently, 2D GaN was synthesized in experiments and shown to exhibit the predicted buckled structure [11].

A key step in the discovery of 2D materials is structure determination. A common technique for identifying the structure of a 2D material is to isolate a slab from a low energy plane of the material's bulk structure [208, 5]. This approach works very well for materials with layered structural motifs, such as van der Waals bonded layered materials, where the 2D structure is often just a single layer of the bulk phase, but it fails for a material such as SiO_2, which has recently been synthesized in both crystalline and amorphous 2D forms [84, 87, 59]. Figs. 4.1

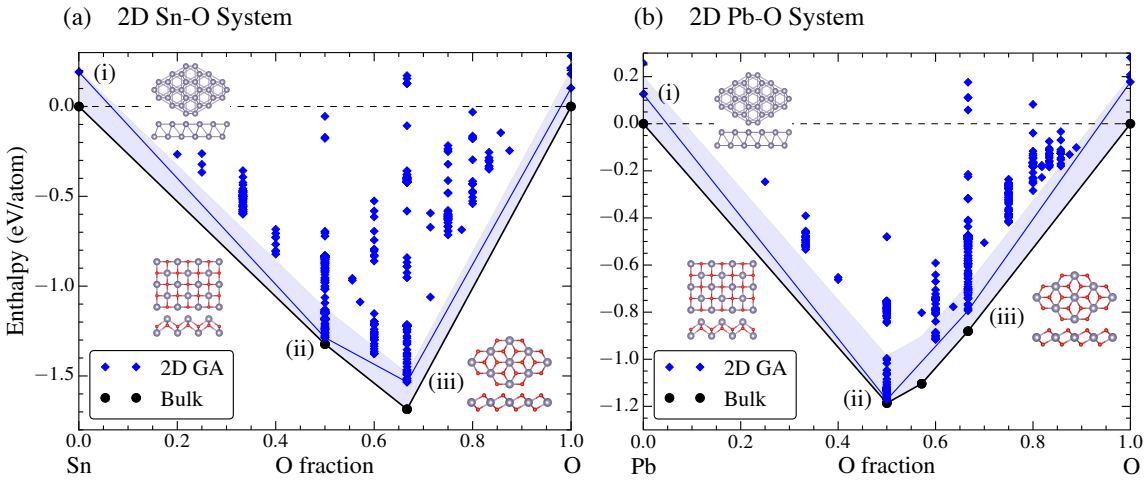

Figure 4.2: 2D structure searches for the (a) Sn-O and (b) Pb-O systems. The black circles denote the ground state bulk structures, and the lines connecting them form the bulk convex hull. The light blue shading indicates the region less than 200 meV/atom above the bulk convex hull. The blue diamonds denote 2D structures found by the genetic algorithm, and the blue line segments form the convex hull for the 2D structures. The algorithm found the same three low-energy structures on the 2D convex hull of each system, shown in each plot and labeled (i) - (iii).

(a) and (c) show that the structure of 2D SiO_2 is fundamentally different from that of α-quartz [59].

In this work, we identify the low-energy 2D structures of group-IV dioxides using the Genetic Algorithm for Structure Prediction (GASP) code [159, 131, 132]. Coupled with accurate DFT methods, GASP efficiently explores a material's multidimensional potential energy surface. GASP has been successfully applied to predict low-energy structures for various bulk and 2D systems including In-P, Sn-S and C-Si [132, 158, 160]. We show that the 2D group-IV dioxides AO_2 (A = Ge, Sn, Pb) exhibit monoclinic and 1T structures with low formation energies relative to competing bulk phases.

Figs. 4.1(a) and (b) show the ground state bulk structures of the group-IV dioxides. The thermodynamically stable phase for bulk SiO_2 is α-quartz (space group $P3_221$) [79], while GeO_2, SnO_2, and PbO_2 occur in the rutile structure (space group $P4_2/mn$) [20, 27]. In α-quartz, Si is four-fold and O is two-fold coordinated. In the rutile phase, the cations are six-fold and O is three-fold coordinated.

4.2 Methods

4.2.1 Density-Functional Calculations

We employ the Vienna Ab-initio Simulation Package (VASP) [75, 71, 73, 74] for all DFT calculations using the projector augmented wave (PAW) method [18]. The structural relaxations are performed with the Perdew-Burke-Ernzerhof (PBE) generalized-gradient exchange-correlation functional [124]. A plane-wave energy cutoff of 500 eV and a k-point mesh density of 30 k-points per $Å^{-1}$ ensures convergence of the energy to 1 meV/atom. For the 2D materials, a vacuum spacing of 10 Å reduces the interaction between layers to about 1 meV/atom.

4.2.2 2D Structure Search

We use the genetic algorithm for structure and phase prediction (GASP) [159, 132] to identify the low-energy 2D structures of the AO_2 compounds (A = Si, Ge, Sn, Pb). The genetic algorithm starts with an initial population of random structures that broadly samples the solution space. The structures are relaxed and low-energy structures are preferentially selected as parents to create child structures using genetic operators such as mutation and mating. When enough child structures have been created, they in turn are selected to make offsprings of their own. This process continues until some user-defined stopping criteria are met. The GASP code is freely available under the GPL v3 license at https://github.com/henniggroup.

In the 2D structure searches, the number of atoms is allowed to vary, and we use an upper limit of 15 atoms per cell. The layer thickness of the 2D materials is constrained to 4 Å. For the Sn-O and Pb-O systems, we employ the phase diagram searching mode of the algorithm, which allows the stoichiometry to vary, and we stop the searches after 1000 structure relaxations. For the 2D SiO_2 and GeO_2 searches, we fix the stoichiometry and use a stopping criterion of 500 structure relaxations.

For the 2D SiO_2 and GeO_2 systems we perform secondary structure searches in which the initial population is seeded with the bi-tetrahedral structure, as well as the low-energy structures found in the first searches. For the second GeO_2

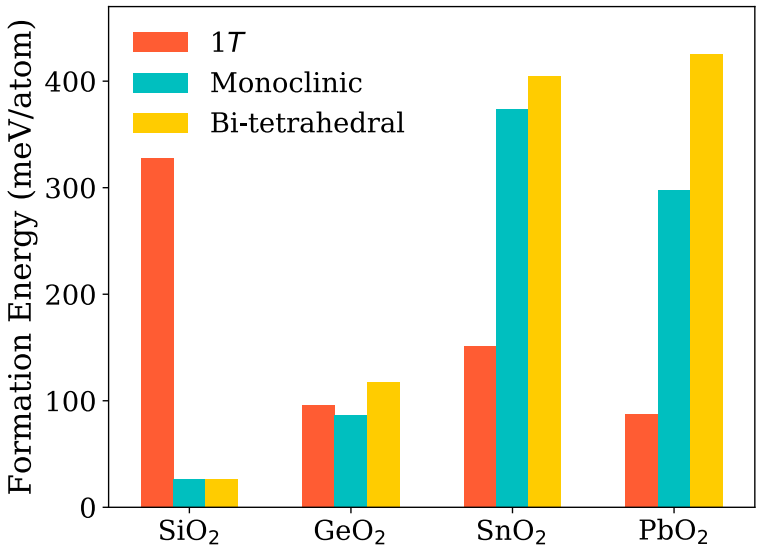

Figure 4.3: Formation energies of the 1T, bi-tetrahedral and monoclinic 2D structures for SiO_2, GeO_2, SnO_2 and PbO_2. The monoclinic structure of SiO_2 relaxes to the bi-tetrahedral structure.

search, we also seed the initial population with the low-energy structures found in the second SiO_2 search. We impose a more liberal layer thickness constraint of 6 Å for these secondary searches, and to speed up the structural relaxations, we use the default cutoff energies and a k-point mesh density of only 20 k-points per $Å^{-1}$. We then re-relax the best third of the resulting structures with the converged parameters (500 eV cutoff energy and 30 k-points per $Å^{-1}$ k-point mesh density).

Table 4.1: Structure information, including lattice parameter, a, b and γ, space group, Wyckoff positions, formation energy, ΔE_f, and cation Bader charge, Q for the 2D group-IV dioxide structures. We have used 3D space groups to describe these finite-thickness 2D structures that lack periodicity in the direction normal to the 2D sheet. In the representations given here, the $\vec{c}$ lattice vector is normal to the plane of the 2D sheet. The vertical components of the general Wyckoff positions are given as distances from the mirror plane, and the in-plane components are given as fractions of the $\vec{a}$ and $\vec{b}$ lattice vectors. Symmetry information was obtained with the FINDSYM software package [150].

	Space group	a, b (Å)	γ (°)	Atomic positions		ΔE_f (meV/atom)	Q (e)
Bi-tetrahedral SiO_2	$P6/mmm$ (191)	5.325, 5.325	120.0	Si 4(h)	z = 1.627 Å	26	3.22
				O 2(d)			
				O 6(i)	z = 2.162 Å		
Monoclinic GeO_2	$C2/m$ (12)	5.746, 8.897	90.0	Ge 8(j)	x = 0.529	86	2.23
					y = 0.189		
					z = −1.620 Å		
				O 4(h)	y = 0.775		
				O 4(i)	x = 0.372		
					z = 1.533 Å		
				O 8(j)	x = 0.757		
					y = 0.290		
					z = 2.477 Å		
1T SnO_2	$P\bar{3}m1$ (164)	3.225, 3.225	120.0	Sn 1(b)		151	2.22
				O 2(d)	z = 1.015 Å		
1T PbO_2	$P\bar{3}m1$ (164)	3.408, 3.408	120.0	Pb 1(b)		87	1.74
				O 2(d)	z = 1.075 Å		

4.3 Results

4.3.1 2D Structure Prediction

Fig. 4.2 shows the energies of the 2D structures found by the genetic algorithm in the Sn-O and Pb-O binary systems relative to the ground state structures of bulk Sn or bulk Pb, respectively, and O_2. The genetic algorithm identifies the same three structures on the 2D convex hulls in both systems: (i) the high buckled hexago-

nal Sn and Pb structure, which was previously reported [133], (ii) the tetragonal litharge structure, which was previously predicted for both SnO and PbO [140], and (iii) the 1*T* structure, which is the lowest energy structure of several other 2D dichalcogenides, such as 2D SnS_2 [212], but not yet reported for 2D oxides.

In the first 2D GeO_2 structure search, the algorithm finds the 1*T* structure to have the lowest energy, Fig 4.1 (d). However, in the second search, in which the initial population is seeded with the bi-tetrahedral and other low-energy structures, the algorithm finds a lower energy structure with monoclinic symmetry, shown in Fig. 4.1(e). Some other low energy GeO_2 structures with energies only slightly larger than that of the monoclinic structure are also identified, however, the phonon dispersion of these structures reveals that they are dynamically unstable (see supplementary materials).

Similarly, in the first 2D SiO_2 search, the lowest energy structure found by the algorithm is not the experimentally observed bi-tetrahedral structure, but rather an orthorhombic structure with higher energy (see supplementary materials). In the second, seeded search, the algorithm uncovers two other low-energy 2D SiO_2 structures, that are energetically unfavorable compared to the bi-tetrahedral structure and are also dynamically unstable (see supplementary materials).

As mentioned above, the genetic algorithm did not succeed in finding the lowest energy 2D structures of SiO_2 and GeO_2 in the first searches. We speculate that due to their open character, these structures pose a difficult challenge for the genetic algorithm. In particular, it is unlikely for features as large as the open

vacuum regions in the bi-tetrahedral structure, which have a diameter of about 6 Å, to appear in randomly generated structures because the fractional atomic coordinates are drawn from a uniform distribution. Furthermore, such a large irreducible feature will probably not easily arise in the subsequent populations. The genetic operators work by combining and perturbing local structural motifs of parents to create offspring, so if no parent structure contains a particular irreducible trait, it is unlikely to appear in the offspring.

We have shown that if prior knowledge about the system of interest is available, it can be used to overcome the difficulty of finding structures with large irreducible features by seeding the initial population with known structures possessing those features.

4.3.2 Energetic Stability

We examine the energetic stability of all 2D group-IV dioxides, SiO_2, GeO_2, SnO_2 and PbO_2, in the two low-energy 2D structures discovered by the genetic algorithm, 1T and monoclinic, and the experimental bi-tetrahedral SiO_2 structure by comparing their formation energies relative to the competing bulk phases, $\Delta E^{\mathrm{f}} = E_{\mathrm{2D}}/N_{\mathrm{2D}} - E_{\mathrm{3D}}/N_{\mathrm{3D}}$, where E and N are the energies of and numbers of atoms in the respective phases. Fig. 4.3 illustrates the computed formation energies of these 2D materials. We note that the lowest energy structure of 2D SiO_2 is the experimental bi-tetrahedral structure. The lowest energy structure of GeO_2 is the

monoclinic structure and that of SnO_2 and PbO_2 is the 1T structure, both found by the genetic algorithm. In the monoclinic structure, Ge is four-fold coordinated, unlike the bulk rutile structure, which is six-fold coordinated. For the other group-IV dioxides, the coordination of the cations in the lowest-energy 2D structures is the same as in their bulk counterparts.

Table 4.1 summarizes the formation energies and structural parameters of the lowest energy 2D structure of each dioxide. The monoclinic structure of GeO_2 and the 1T structures of SnO_2 and PbO_2 all have formation energies below 200 meV/atom, similar to those of already synthesized 2D materials [132, 143]. This indicates promise for experimental synthesis of 2D GeO_2, SnO_2 and PbO_2 as free-standing films.

4.4 Conclusion

In summary, we have used a genetic algorithm for structure prediction coupled with density-functional theory calculations to identify the structures of the 2D group-IV dioxides. We confirm that 2D SiO_2 is most stable in its experimentally determined bi-tetrahedral structure and predict that 2D SnO_2 and PbO_2 exhibit the 1T structure. For 2D GeO_2, we predict a new monoclinic structure, as well as several other nearly degenerate structures. Their low formation energies ($<$ 151 meV/atom) indicate that it should be possible to synthesize them as free standing layers or on substrates.

CHAPTER 5

CONCLUSIONS

We have presented a review of evolutionary algorithms for structure prediction. Many choices are possible when designing evolutionary algorithms, and we have reviewed several of them and discussed how they can affect performance. We also presented a summary of the available codes that implement evolutionary algorithms for structure prediction and reviewed their applications to various bulk materials systems.

We then described our improved evolutionary algorithm for structure prediction, with an emphasis on the extended capabilities and improvements we have implemented. These include searching for structures with subperiodic geometries, including clusters, wires and 2D materials. We have also improved the load balancing of the algorithm when running in parallel by adopting a scheme in which offspring structures are created continually. This approach makes more efficient use of computational resources and significantly reduces the wall time needed to search the energy landscapes of *ab initio* Hamiltonians. For phase diagram searches, we have improved the algorithm's sampling of the composition space, reducing bias both toward one end or another of the allowed range of compositions, and also toward the center of the composition space. We have also extended the algorithm to enable searching for partial phase diagrams. This has the potential to substantially reduce the computational cost of phase diagram searches in cases where the researcher is interested in only a restricted range of

compositions.

Next we described several applications of the evolutionary algorithm to 2D materials. Coupled with *ab initio* total energy calculations, we have used the algorithm to explore the 2D systems of InP, C-Si and Sn-S phase diagrams, and group-IV dioxides. The algorithm identified several new 2D structures of InP, and it found two 1D Si defects in graphene with formation energies below that of isolated substitutional Si atoms. When applied to 2D group-IV dioxides, the algorithm predicted that 2D PbO_2 and SnO_2 adopt the 1T structure, while 2D GeO_2 adopts a new structure with monoclinic symmetry. All three of these predicted 2D structures have low formation energies (< 200 meV/atom) relative to the ground state bulk structures. This is significant because, in contrast to the bulk structures of several known 2D materials (e.g., graphene), bulk group-IV dioxides do not posses layered structural motifs. This study therefore demonstrated the usefulness of the evolutionary algorithm in predicting low-energy 2D materials in cases where the corresponding 3D bulk structures do not provide obvious clues as to what structure a single layer might adopt.

There exist several areas of potential improvement. By far the most computationally expensive part of an *ab initio* structure search are the DFT calculations. We are currently implementing a machine learning surrogate model that utilizes more of the information produced in the course of a DFT relaxation, such as the structure and energy at each ionic step. The idea is to train the surrogate model during the search and use it to obtain cheap estimates of candidate structures' fi-

nal energies. The more unfavorable candidates can then be screened out to avoid wasting expensive DFT relaxations on them. Another avenue of development are alternative objective functions for non-bulk systems. For inherently metastable wires and 2D materials, string tension and surface energy, respectively, may be better choices for the objective function. More complex interfaces, such as surface reconstructions and grain boundaries, present further areas of potential development.

APPENDIX A

COMPOSITION-DEPENDENT FITNESS FOR IMPROVED COMPOSITION SPACE SAMPLING

In phase diagram searches, the cut-and-splice mating operation tends to produce offspring structures with compositions between those of the two parents. Over time, this causes the population of structures to drift toward the middle regions of the composition space, leaving the more extreme compositions (those closer the endpoints of the composition space) under-sampled. To address this problem, we modify the selection probabilities such that two parent structures are likely to have similar compositions.

Before we can select parent structures with similar compositions, we need a general method of determining the distance between points in composition space. That is, we need a distance metric for composition spaces with an arbitrary number of dimensions. We observe that a phase diagram with n endpoints may be considered to be the portion of the L_1 norm unit sphere that lies in the positive orthant of an n-dimensional space, where each dimension corresponds to the fraction of an endpoint of the composition space. Fig. A.1 illustrates this idea for binary and ternary phase diagrams.

We note that the L_1 norm satifies the requirements of a distance metric in composition space. For example, consider a ternary phase diagram with endpoint compositions A, B and C, as shown in Fig. A.1(b). Any point in the composition

space may be expressed in terms of the fractions of A, B and C, and the endpoints themselves can be defined by the following vectors:

$$\mathbf{A} = (1, 0, 0) \qquad \mathbf{B} = (0, 1, 0) \qquad \mathbf{C} = (0, 0, 1) \tag{A.1}$$

Consider the distance between **A** and **B**, that is, the distance between two of the endpoints of the composition space. Using the L_1 norm as the distance metric, we have

$$||\mathbf{A} - \mathbf{B}||_1 = ||(1, -1, 0)||_1 = |1| + |-1| + |0| = 2$$

Similarly, the distance between **A** and **C** is

$$||\mathbf{A} - \mathbf{C}||_1 = ||(1, 0, -1)||_1 = |1| + |0| + |-1| = 2$$

Now consider the distance between **A** and a point **P** halfway between **B** and **C**. That is,

$$\mathbf{P} = \frac{1}{2}(\mathbf{B} + \mathbf{C}) = \left(0, \frac{1}{2}, \frac{1}{2}\right)$$

We know that the distance between **A** and **P** must be equivalent to the distance between **A** and **B** (and **A** and **C**) because composition P contains none of endpoint A. That is, **P** is orthogonal to **A** (just like **B** and **C**). Computing the distance between **A** and **P** using the L_1 norm, we have

$$||\mathbf{A} - \mathbf{P}||_1 = \left|\left|\left(1, -\frac{1}{2}, -\frac{1}{2}\right)\right|\right|_1 = |1| + \left|-\frac{1}{2}\right| + \left|-\frac{1}{2}\right| = 2$$

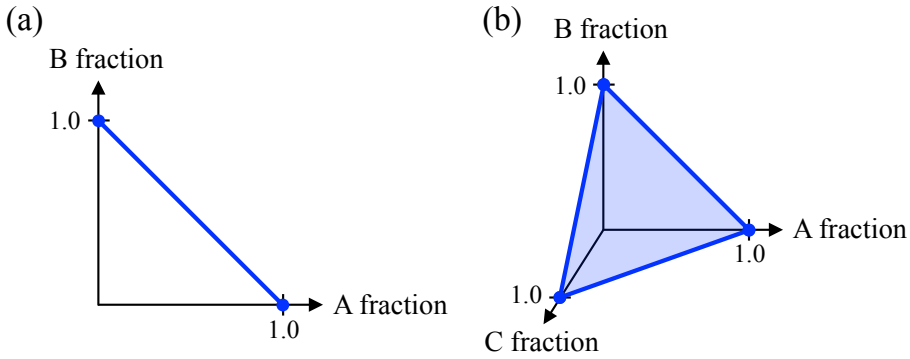

Figure A.1: Illustration of the composition space for (a) binary and (b) ternary systems. In the binary system, the composition space ranges from pure A to pure B, and is shown by the blue line, which is the part of the L_1 norm unit circle that lies in the positive quadrant. In the ternary system, the composition space has three endpoints: pure A, B and C. The three blue lines outline the shaded region that represents the ternary composition space, and each line comprises a binary composition space between two of the endpoints. The ternary composition space corresponds to the part of the three-dimensional L_1 norm unit sphere that lies in the positive octant.

which is the desired result.

It is convenient to normalize the distances between points in the composition space such that the maximum possible distance between points is 1. From the examples above, we see that this can be achieved by dividing the L_1 norm distance by 2. So in general, we take the distance between any two compositions X and Y in a composition space to be

$$d_{XY} = \frac{1}{2}\|\mathbf{X} - \mathbf{Y}\|_1$$

We note that this normalization holds for composition spaces of arbitrary dimension.

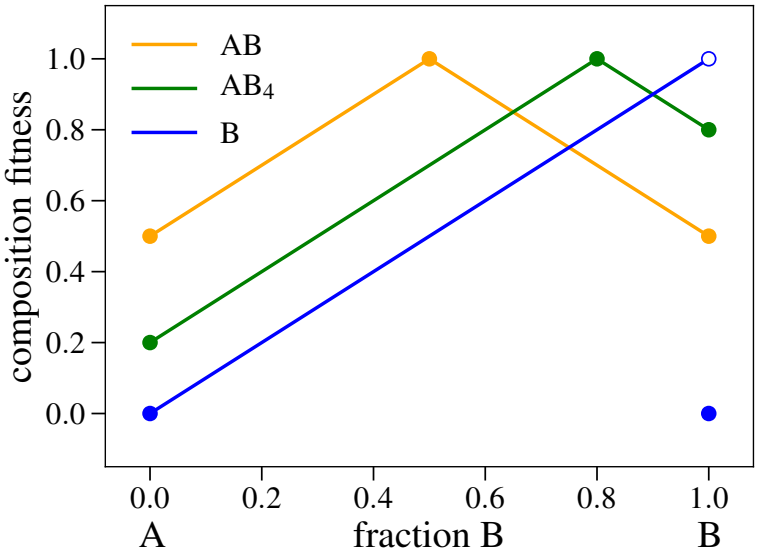

Figure A.2: Composition fitness as a function of composition in a binary system relative to first parent organisms with three different compositions: AB, AB_4 and pure B. We note that when the first parent organism is located at an endpoint of the composition space, the composition fitnesses of other organisms at that endpoint are set to zero.

Now that we have a way to compute the distances between points in composition space, we may return to the original problem of preventing the population from drifting toward intermediate compositions. Our approach is to select the first parent organism in the standard way, and then to modify the selection probabilities of the remaining organisms such that those with compositions closer to that of the first parent are more likely to be selected.

After selecting the first parent organism, we assign fitness values to the remaining organisms in the population relative to the first parent. The fitness of an organism relative to the first parent organism is defined by

$$f_{rel} = w_{comp} f_{comp} + (1 - w_{comp}) f_{reg} \tag{A.2}$$

where f_{reg} is the regular fitness of the organism (i.e., based on it's objective function value), and f_{comp} is the composition fitness of the organism relative to the composition of the first parent:

$$f_{comp} = 1 - d$$

where d is the normalized distance in composition space between the organism and the first parent. We note that like the regular fitness, the composition fitness ranges from 0 to 1. An exception to how the composition fitness is computed arises when the first parent happens to lie at an endpoint of the composition space. In that case, the composition fitnesses of other organisms with the same composition are set to zero. This is done to prevent the algorithm from oversampling the endpoint compositions (the regions that are undersampled tend to be the compositions that are near to the endpoints, but not the endpoints themselves). Fig. A.2 illustrates the composition fitnesses of organisms in a binary system for several different compositions of the first parent organism.

From Equation A.2, we see that the relative fitness is the weighted average of the regular fitness and the composition fitness, where w_{comp} is the weight assigned to the composition fitness and lies in [0, 1]. So in order to compute relative fitnesses, we must choose a value for w_{comp}. If the first parent organism is located near a region of the composition space that tends to be under-sampled, we would like the eventual offspring structure to lie in that region as well. Therefore, we should choose the second parent such that it has a composition close to that of the first, which can be achieved by increasing w_{comp}. On the other hand, if the first

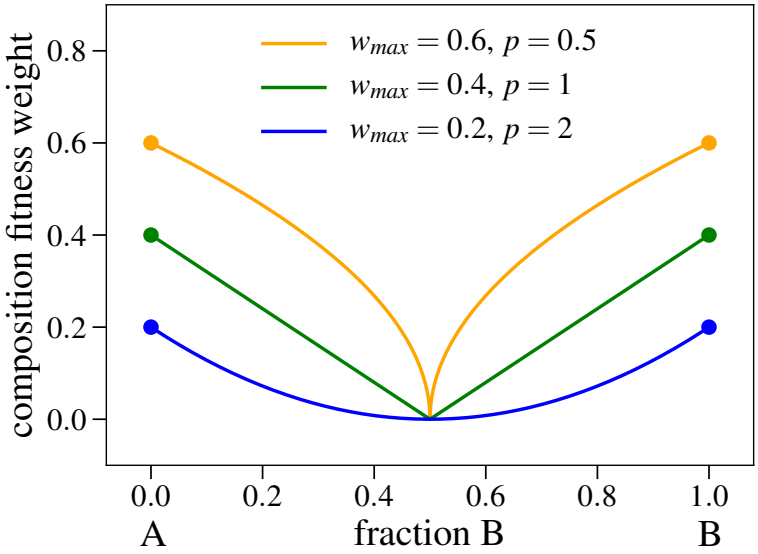

Figure A.3: Weight assigned to the composition fitness as a function of the composition of the first parent organism in a binary system. The weights are shown for three different choices of the parameters w_{max} and p in Equation A.3.

parent organism is located in a region of the composition space that tends to be over-sampled, it is not as important that the offspring structure also be located in that region. In this case we are relatively indifferent to the composition of the second parent, which is expressed by using a small value for w_{comp}.

To apply these considerations, the value we assign to w_{comp} depends on the composition of the first parent organism. In particular, we use a power law relation:

$$w_{comp} = w_{max} d^p \tag{A.3}$$

where w_{max} is the maximum value of w_{comp} (at the endpoints), and d is the distance between the composition of the first parent organism and the center of the compo-

sition space; this distance is normalized so that the distance between an endpoint and the center of the composition space is 1. p is the power that determines how w_{comp} decreases upon moving away from an endpoint of the composition space. Fig. A.3 shows the values of w_{comp} versus composition for several different choices of w_{max} and p in a binary system.

Once the relative fitnesses of all the organisms have been obtained, their selection probabilities are computed in the standard way, except using the relative fitnesses instead of the regular fitnesses. A second parent organism is then chosen based on these new selection probabilities.

BIBLIOGRAPHY

[1] N. L. Abraham and M. I. J. Probert. A periodic genetic algorithm with real-space representation for crystal structure and polymorph prediction. *Physical Review B*, 73(22):224104, 2006.

[2] N. L. Abraham and M. I. J. Probert. Improved real-space genetic algorithm for crystal structure and polymorph prediction. *Physical Review B*, 77:134117, 2008.

[3] J. P. Agrawal. Recent trends in high-energy materials. *Progress in Energy and Combustion Science*, 24(1):1–30, 1998.

[4] N. W. Ashcroft. Metallic hydrogen: a high-temperature superconductor? *Physical Review Letters*, 21:1748–1749, 1968.

[5] Michael Ashton, Joshua Paul, Susan B. Sinnott, and Richard G. Hennig. Topology-scaling identification of layered solids and stable exfoliated 2D materials. *Physical Review Letters*, 118:106101, 2017.

[6] C. Ataca, H. Şahin, and S. Ciraci. Stable, single-layer MX_2 transition-metal oxides and dichalcogenides in a honeycomb-like structure. *The Journal of Physical Chemistry C*, 116(16):8983–8999, 2012.

[7] B. Aufray, A. Kara, S. Vizzini, H. Oughaddou, C. Leandri, B. Ealet, and G. Le Lay. Graphene-like silicon nanoribbons on Ag (110): A possible formation of silicene. *Applied Physics Letters*, 96(18):183102, 2010.

[8] I. A. Baburin, S. Leoni, and G. Seifert. Enumeration of not-yet-synthesized zeolitic zinc imidazolate MOF networks: a topological and DFT approach. *The Journal of Physical Chemistry B*, 112(31):9437–9443, 2008.

[9] Pio Baettig and Eva Zurek. Pressure-stabilized sodium polyhydrides: NaH_n ($n > 1$). *Physical Review Letters*, 106(23):237002, 2011.

[10] Silvia Bahmann and Jens Kortus. EVO-evolutionary algorithm for crystal

structure prediction. *Computer Physics Communications*, 184(6):1618–1625, 2013.

[11] Zakaria Y. Al Balushi, Ke Wang, Ram Krishna Ghosh, Rafael A. Vila, Sarah M. Eichfeld, Joshua D. Caldwell, Xiaoye Qin, Yu-Chuan Lin, Paul A. DeSario, Greg Stone, Shruti Subramanian, Dennis F. Paul, Robert M. Wallace, Suman Datta, Joan M. Redwing, and Joshua A. Robinson. Two-dimensional gallium nitride realized via graphene encapsulation. *Nature Materials*, 2016.

[12] Zulfiya G. Bazhanova, Artem R. Oganov, and Omar Gianola. Fe–C and Fe–H systems at pressures of the Earth's inner core. *Physics-Uspekhi*, 55(5):489, 2012.

[13] W. Bi, Y. Meng, R. S. Kumar, A. L. Cornelius, W. W. Tipton, R. G. Hennig, Y. Zhang, C. Chen, and J. S. Schilling. Pressure-induced structural transitions in europium to 92 GPa. *Physical Review B*, 83(10):104106, 2011.

[14] W. Bi, N. M. Souza-Neto, D. Haskel, G. Fabbris, E. E. Alp, J. Zhao, R. G. Hennig, M. M. Abd-Elmeguid, Y. Meng, R. W. McCallum, K. Dennis, and J. S. Schilling. Synchrotron x-ray spectroscopy studies of valence and magnetic state in europium metal to extreme pressures. *Physical Review B*, 85:205134, 2012.

[15] Goranka Bilalbegović. Structure and stability of finite gold nanowires. *Physical Review B*, 58(23):15412, 1998.

[16] Goranka Bilalbegović. Structures and melting in infinite gold nanowires. *Solid State Communications*, 115(2):73–76, 2000.

[17] Torbjörn Björkman, Andris Gulans, Arkady V. Krasheninnikov, and Risto M. Nieminen. van der Waals bonding in layered compounds from advanced density-functional first-principles calculations. *Physical Review Letters*, 108(23):235502, 2012.

[18] Peter E. Blöchl. Projector augmented-wave method. *Physical Review B*, 50(24):17953, 1994.

[19] Michael N. Blonsky, Houlong L. Zhuang, Arunima K. Singh, and Richard G. Hennig. Ab initio prediction of piezoelectricity in two-dimensional materials. *ACS Nano*, 9(10):9885–9891, 2015.

[20] Adrian A. Bolzan, Celesta Fong, B. J. Kennedy, and Ch J. Howard. Structural studies of rutile-type metal dioxides. *Acta Crystallographica Section B*, 53(3):373–380, 1997.

[21] Tilman Brodmeier and Ernö Pretsch. Application of genetic algorithms in molecular modeling. *Journal of Computational Chemistry*, 15(6):588–595, 1994.

[22] T. S. Bush, C. R. A. Catlow, and P. D. Battle. Evolutionary programming techniques for predicting inorganic crystal structures. *Journal of Materials Chemistry*, 5(8):1269–1272, 1995.

[23] G. Ceder, D. Morgan, C. Fischer, K. Tibbetts, and S. Curtarolo. Data-mining-driven quantum mechanics for the prediction of structure. *MRS Bulletin*, 31(12):981–985, 2006.

[24] Y. Q. Cheng, E. Ma, and H. W. Sheng. Atomic level structure in multicomponent bulk metallic glass. *Physical Review Letters*, 102(24):245501, 2009.

[25] F. C. Chuang, Cristian V. Ciobanu, V. B. Shenoy, Cai-Zhuang Wang, and Kai-Ming Ho. Finding the reconstructions of semiconductor surfaces via a genetic algorithm. *Surface Science*, 573(2):L375–L381, 2004.

[26] Thomas Dandekar and Patrick Argos. Folding the main chain of small proteins with the genetic algorithm. *Journal of Molecular Biology*, 236(3):844, 1994.

[27] Peter D'Antonio and A. Santoro. Powder neutron diffraction study of chemically prepared β-lead dioxide. *Acta Crystallographica Section B*, 36(10):2394–2397, 1980.

[28] Sarah Darby, Thomas V. Mortimer-Jones, Roy L. Johnston, and Christopher Roberts. Theoretical study of Cu-Au nanoalloy clusters using a genetic algorithm. *The Journal of Chemical Physics*, 116(4):1536–1550, 2002.

[29] Mayeul d'Avezac and Alex Zunger. Identifying the minimum-energy atomic configuration on a lattice: Lamarckian twist on Darwinian evolution. *Physical Review B*, 78(6):064102, 2008.

[30] D. M. Deaven and K. M. Ho. Molecular geometry optimization with a genetic algorithm. *Physical Review Letters*, 75(2):288, 1995.

[31] Max Dion, Henrik Rydberg, Elsebeth Schröder, David C. Langreth, and Bengt I. Lundqvist. Van der Waals density functional for general geometries. *Physical Review Letters*, 92(24):246401, 2004.

[32] Jonathan P. K. Doye and David J. Wales. Global minima for transition metal clusters described by Sutton-Chen potentials. *New Journal of Chemistry*, 22(7):733–744, 1998.

[33] F. Ercolessi, M. Parrinello, and E. Tosatti. Simulation of gold in the glue model. *Philosophical Magazine A*, 58(1):213–226, 1988.

[34] Ji Feng, Wojciech Grochala, Tomasz Jaroń, Roald Hoffmann, Aitor Bergara, and N. W. Ashcroft. Structures and potential superconductivity in SiH_4 at high pressure: en route to metallic hydrogen. *Physical Review Letters*, 96(1):017006, 2006.

[35] Ji Feng, Richard G. Hennig, N. W. Ashcroft, and Roald Hoffmann. Emergent reduction of electronic state dimensionality in dense ordered Li-Be alloys. *Nature*, 451(7177):445–448, 2008.

[36] S. M. Foiles, M. I. Baskes, and M. S. Daw. Embedded-atom-method functions for the fcc metals Cu, Ag, Au, Ni, Pd, Pt, and their alloys. *Physical Review B*, 33(12):7983, 1986.

[37] Guoying Gao, Artem R. Oganov, Aitor Bergara, Miguel Martinez-Canales, Tian Cui, Toshiaki Iitaka, Yanming Ma, and Guangtian Zou. Superconducting high pressure phase of germane. *Physical Review Letters*, 101(10):107002, 2008.

[38] Guoying Gao, Artem R. Oganov, Peifang Li, Zhenwei Li, Hui Wang, Tian

Cui, Yanming Ma, Aitor Bergara, Andriy O. Lyakhov, Toshiaki Iitaka, et al. High-pressure crystal structures and superconductivity of stannane (SnH_4). *Proceedings of the National Academy of Sciences*, 107(4):1317–1320, 2010.

[39] Guoying Gao, Artem R. Oganov, Yanming Ma, Hui Wang, Peifang Li, Yinwei Li, Toshiaki Iitaka, and Guangtian Zou. Dissociation of methane under high pressure. *The Journal of Chemical Physics*, 133:144508, 2010.

[40] J. Garcia-Rodeja, C. Rey, L. J. Gallego, and J. A. Alonso. Molecular-dynamics study of the structures, binding energies, and melting of clusters of fcc transition and noble metals using the Voter and Chen version of the embedded-atom model. *Physical Review B*, 49(12):8495, 1994.

[41] Colin W. Glass, Artem R. Oganov, and Nikolaus Hansen. USPEX-evolutionary crystal structure prediction. *Computer Physics Communications*, 175(11):713–720, 2006.

[42] Stefan Goedecker. Minima hopping: an efficient search method for the global minimum of the potential energy surface of complex molecular systems. *The Journal of Chemical Physics*, 120:9911, 2004.

[43] Henrik Grönbeck and Wanda Andreoni. Gold and platinum microclusters and their anions: comparison of structural and electronic properties. *Chemical Physics*, 262(1):1–14, 2000.

[44] Raju P. Gupta. Lattice relaxation at a metal surface. *Physical Review B*, 23(12):6265, 1981.

[45] G. G. Guzmán-Verri and L. C. L. Y. Voon. Electronic structure of silicon-based nanostructures. *Physical Review B*, 76(7):075131, 2007.

[46] Hannu Häkkinen and Uzi Landman. Gold clusters (Au_N, $2 \leq N \leq 10$) and their anions. *Physical Review B*, 62(4):R2287, 2000.

[47] Bernd Hartke. Global geometry optimization of clusters using genetic algorithms. *The Journal of Physical Chemistry*, 97(39):9973–9976, 1993.

[48] Andreas Hermann, N. W. Ashcroft, and Roald Hoffmann. High pressure ices. *Proceedings of the National Academy of Sciences*, 109(3):745–750, 2012.

[49] Andreas Hermann, N. W. Ashcroft, and Roald Hoffmann. Making sense of boron-rich binary Be–B phases. *Inorganic Chemistry*, 51(16):9066–9075, 2012.

[50] Andreas Hermann, B. L. Ivanov, N. W. Ashcroft, and Roald Hoffmann. LiBeB: a predicted phase with structural and electronic peculiarities. *Physical Review B*, 86(1):014104, 2012.

[51] Andreas Hermann, Alexandra McSorley, N. W. Ashcroft, and Roald Hoffmann. From Wade–Mingos to Zintl–Klemm at 100 GPa: binary compounds of boron and lithium. *Journal of the American Chemical Society*, 134(45):18606–18618, 2012.

[52] Andreas Hermann, Ainhoa Suarez-Alcubilla, Idoia G. Gurtubay, Li-Ming Yang, Aitor Bergara, N. W. Ashcroft, and Roald Hoffmann. LiB and its boron-deficient variants under pressure. *Physical Review B*, 86(14):144110, 2012.

[53] M. R. Hoare and J. A. McInnes. Morphology and statistical statics of simple microclusters. *Advances in Physics*, 32(5):791–821, 1983.

[54] James Hooper, Bahadir Altintas, Andrew Shamp, and Eva Zurek. Polyhydrides of the alkaline earth metals: a look at the extremes under pressure. *The Journal of Physical Chemistry C*, 117(6):2982–2992, 2013.

[55] James Hooper and Eva Zurek. High pressure potassium polyhydrides: a chemical perspective. *The Journal of Physical Chemistry C*, 116(24):13322–13328, 2012.

[56] James Hooper and Eva Zurek. Lithium subhydrides under pressure and their superatom-like building blocks. *ChemPlusChem*, 77(11):969–972, 2012.

[57] James Hooper and Eva Zurek. Rubidium polyhydrides under pressure: emergence of the linear H_3^- species. *Chemistry-A European Journal*, 18(16):5013–5021, 2012.

[58] Chao-Hao Hu, Artem R. Oganov, Andriy O. Lyakhov, Huai-Ying Zhou, and J. Hafner. Insulating states of $LiBeH_3$ under extreme compression. *Physical Review B*, 79(13):134116, 2009.

[59] Pinshane Y. Huang, Simon Kurasch, Anchal Srivastava, Viera Skakalova, Jani Kotakoski, Arkady V. Krasheninnikov, Rober Hovden, Qingyun Mao, Jannik C. Meyer, Jurgen Smet, David A. Muller, and Ute Kaiser. Direct imaging of a two-dimensional silica glass of graphene. *Nano Letters*, 12:1081–1086, 2012.

[60] Md Mahbubul Islam, Alireza Ostadhossein, Oleg Borodin, A. Todd Yeates, William W. Tipton, Richard G. Hennig, Nitin Kumar, and Adri C. T. van Duin. ReaxFF molecular dynamics simulations on lithiated sulfur cathode materials. *Physical Chemistry Chemical Physics*, 17(5):3383–3393, 2015.

[61] A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, and K. A. Persson. The Materials Project: A materials genome approach to accelerating materials innovation. *APL Materials*, 1(1):011002, 2013.

[62] Min Ji, Koichiro Umemoto, Cai-Zhuang Wang, Kai-Ming Ho, and Renata M. Wentzcovitch. Ultrahigh-pressure phases of H_2O ice predicted using an adaptive genetic algorithm. *Physical Review B*, 84:220105, 2011.

[63] Min Ji, Cai-Zhuang Wang, and Kai-Ming Ho. Comparing efficiencies of genetic and minima hopping algorithms for crystal structure prediction. *Physical Chemistry Chemical Physics*, 12(37):11617–11623, 2010.

[64] Gisli Holmar Johannesson, Thomas Bligaard, Andrei V. Ruban, Hans Lomholt Skriver, Karsten Wedel Jacobsen, and Jens Kehlet Nørskov. Combined electronic structure and evolutionary search approach to materials design. *Physical Review Letters*, 88(25):255506, 2002.

[65] Roy L. Johnston. Evolving better nanoparticles: genetic algorithms for optimising cluster geometries. *Dalton Transactions*, (22):4193–4207, 2003.

[66] M. Karabacak, S. Özçelik, and Z. B. Güvenç. Structures and energetics of

Pd_n ($n = 2 - 20$) clusters using an embedded-atom model potential. *Surface Science*, 507:636–642, 2002.

[67] Ji říKlimeš, David R. Bowler, and Angelos Michaelides. Van der Waals density functionals applied to solids. *Physical Review B*, 83:195131, 2011.

[68] A. N. Kolmogorov, S. Shah, E. R. Margine, A. F. Bialon, T. Hammerschmidt, and R. Drautz. New superconducting and semiconducting Fe-B compounds predicted with an ab initio evolutionary search. *Physical Review Letters*, 105(21):217003, 2010.

[69] Yukihito Kondo and Kunio Takayanagi. Gold nanobridge stabilized by surface structure. *Physical Review Letters*, 79(18):3455, 1997.

[70] Yukihito Kondo and Kunio Takayanagi. Synthesis and characterization of helical multi-shell gold nanowires. *Science*, 289(5479):606–608, 2000.

[71] G. Kresse and J. Hafner. Ab initio molecular-dynamics simulation of the liquid-metal–amorphous-semiconductor transition in germanium. *Physical Review B*, 49(20):14251, 1994.

[72] G. Kresse and D. Joubert. From ultrasoft pseudopotentials to the projector augmented-wave method. *Physical Review B*, 59:1758–1775, 1999.

[73] Georg Kresse and Jürgen Furthmüller. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Computational Materials Science*, 6(1):15–50, 1996.

[74] Georg Kresse and Jürgen Furthmüller. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Physical Review B*, 54(16):11169, 1996.

[75] Georg Kresse and Jürgen Hafner. Ab initio molecular dynamics for liquid metals. *Physical Review B*, 47(1):558, 1993.

[76] I. Krivy and B. Gruber. A unified algorithm for determining the reduced

(Niggli) cell. *Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography*, 32(2):297–298, 1976.

[77] Dominik Kurzydłowski, Patryk Zaleski-Ejgierd, Wojciech Grochala, and Roald Hoffmann. Freezing in resonance structures for better packing: XeF_2 becomes (XeF+)(F-) at large compression. *Inorganic Chemistry*, 50(8):3832–3840, 2011.

[78] Vanessa Labet, Roald Hoffmann, and N. W. Ashcroft. Molecular models for WH_6 under pressure. *New Journal of Chemistry*, 35(10):2349–2355, 2011.

[79] Yvon Le Page and Gabrielle Donnay. Refinement of the crystal structure of low-quartz. *Acta Crystallographica Section B*, 32(8):2456–2459, 1976.

[80] S. Lebègue, T. Björkman, M. Klintenberg, R. M. Nieminen, and O. Eriksson. Two-dimensional materials from data filtering and *ab initio* calculations. *Physical Review X*, 3:031002, 2013.

[81] Quan Li, Yanming Ma, Artem R. Oganov, Hongbo Wang, Hui Wang, Ying Xu, Tian Cui, Ho-Kwang Mao, and Guangtian Zou. Superhard monoclinic polymorph of carbon. *Physical Review Letters*, 102(17):175506, 2009.

[82] Quan Li, Mei Wang, Artem R. Oganov, Tian Cui, Yanming Ma, and Guangtian Zou. Rhombohedral superhard structure of BC_2N. *Journal of Applied Physics*, 105(5):053514–053514, 2009.

[83] Y. Li, F. Li, Z. Zhou, and Z. Chen. SiC_2 silagraphene and its one-dimensional derivatives: where planar tetracoordinate silicon happens. *Journal of the American Chemical Society*, 133(4):900–908, 2010.

[84] Leonid Lichtenstein, Christin Buechner, Bing Yang, Shamil Shaikhutdinov, Markus Heyde, Marek Sierka, Radoslaw Wlodarczyk, Joachim Sauer, and Hans-Joachim Freund. The atomic structure of a metal-supported vitreous thin silica film. *Angewandte Chemie International Edition*, 51:404–407, 2012.

[85] S. S. Lin. Light-emitting two-dimensional ultrathin silicon carbide. *The Journal of Physical Chemistry C*, 116(6):3951–3955, 2012.

[86] C. C. Liu, W. Feng, and Y. Yao. Quantum spin Hall effect in silicene and two-dimensional germanium. *Physical Review Letters*, 107(7):076802, 2011.

[87] D. Löffler, J. J. Uhlrich, M. Baron, B. Yang, X. Yu, L. Lichtenstein, L. Heinke, C. Büchner, M. Heyde, S. Shaikhutdinov, H.-J. Freund, R. Włodarczyk, M. Sierka, and J. Sauer. Growth and structure of crystalline silica sheet on Ru(0001). *Physical Review Letters*, 105:146104, 2010.

[88] David C. Lonie, James Hooper, Bahadir Altintas, and Eva Zurek. Metallization of magnesium polyhydrides under pressure. *Physical Review B*, 87(5):054107, 2013.

[89] David C. Lonie and Eva Zurek. XtalOpt: an open-source evolutionary algorithm for crystal structure prediction. *Computer Physics Communications*, 182(2):372–387, 2011.

[90] David C. Lonie and Eva Zurek. Identifying duplicate crystal structures: XtalComp, an open-source solution. *Computer Physics Communications*, 183(3):690 – 697, 2012.

[91] Ning Lu, Cristian V. Ciobanu, Tzu-Liang Chan, Feng-Chuan Chuang, Cai-Zhuang Wang, and Kai-Ming Ho. The structure of ultrathin H-passivated [112] silicon nanowires. *The Journal of Physical Chemistry C*, 111(22):7933–7937, 2007.

[92] Carlos B. Lucasius, S. Werten, A. H. J. M. van Aert, Gerrit Kateman, and Marcel J. J. Blommers. Conformational analysis of DNA using genetic algorithms. In *Parallel Problem Solving from Nature*, pages 90–97. Springer, 1991.

[93] Wei Luo, Yanming Ma, Xingao Gong, and Hongjun Xiang. Prediction of silicon-based layered structures for optoelectronic applications. *Journal of the American Chemical Society*, 136(45):15992–15997, 2014.

[94] Xinyu Luo, Jihui Yang, Hanyu Liu, Xiaojun Wu, Yanchao Wang, Yanming Ma, Su-Huai Wei, Xingao Gong, and Hongjun Xiang. Predicting two-dimensional boron–carbon compounds by the global optimization method. *Journal of the American Chemical Society*, 133(40):16285–16290, 2011.

[95] Andriy O. Lyakhov, Artem R. Oganov, and Mario Valle. How to predict very large and complex crystal structures. *Computer Physics Communications*, 181(9):1623 – 1632, 2010.

[96] Yanming Ma, Mikhail Eremets, Artem R. Oganov, Yu Xie, Ivan Trojan, Sergey Medvedev, Andriy O. Lyakhov, Mario Valle, and Vitali Prakapenka. Transparent dense sodium. *Nature*, 458(7235):182–185, 2009.

[97] Yanming Ma, Artem R. Oganov, and Colin W. Glass. Structure of the metallic ζ-phase of oxygen and isosymmetric nature of the ε-ζ phase transition: ab initio simulations. *Physical Review B*, 76(6):064101, 2007.

[98] Yanming Ma, Artem R. Oganov, Zhenwei Li, Yu Xie, and Jani Kotakoski. Novel high pressure structures of polymeric nitrogen. *Physical Review Letters*, 102(6):065501, 2009.

[99] Yanming Ma, Artem R. Oganov, and Yu Xie. High-pressure structures of lithium, potassium, and rubidium predicted by an ab initio evolutionary algorithm. *Physical Review B*, 78(1):014102, 2008.

[100] Yanming Ma, Yanchao Wang, and Artem R. Oganov. Absence of superconductivity in the high-pressure polymorph of MgB_2. *Physical Review B*, 79(5):054101, 2009.

[101] K. F. Mak, C. Lee, J. Hone, J. Shan, and T. F. Heinz. Atomically thin MoS_2: a new direct-gap semiconductor. *Physical Review Letters*, 105:136805, 2010.

[102] Miguel Martinez-Canales, Artem R. Oganov, Yanming Ma, Yan Yan, Andriy O. Lyakhov, and Aitor Bergara. Novel structures and superconductivity of silane under pressure. *Physical Review Letters*, 102(8):087005, 2009.

[103] R. Martoňák, Alessandro Laio, and Michele Parrinello. Predicting crystal structures: the Parrinello-Rahman method revisited. *Physical Review Letters*, 90(7):075503, 2003.

[104] Ruben Mas-Balleste, Cristina Gomez-Navarro, Julio Gomez-Herrero, and

Felix Zamora. 2D materials: to graphene and beyond. *Nanoscale*, 3(1):20–30, 2011.

[105] Claire P. Massen and Jonathan P. K. Doye. Power-law distributions for the areas of the basins of attraction on a potential energy landscape. *Physical Review E*, 75(3):037101, 2007.

[106] D. B. McGarrah and Richard S. Judson. Analysis of the genetic algorithm method of molecular conformation determination. *Journal of Computational Chemistry*, 14(11):1385–1395, 1993.

[107] M. Menon, E. Richter, A. Mavrandonakis, G. Froudakis, and A. N. Andriotis. Structure and stability of SiC nanotubes. *Physical Review B*, 69(11):115322, 2004.

[108] K. Michaelian, N. Rendon, and I. L. Garzón. Structure and energetics of Ni, Ag, and Au nanoclusters. *Physical Review B*, 60(3):2000, 1999.

[109] Pere Miro, Martha Audiffred, and Thomas Heine. An atlas of two-dimensional materials. *Chemical Society Reviews*, 43:6537–6554, 2014.

[110] K. S. Novoselov, A. K. Geim, S. V. Morozov, D. Jiang, Y_ Zhang, S. V. Dubonos, I. V. Grigorieva, and A. A. Firsov. Electric field effect in atomically thin carbon films. *Science*, 306(5696):666–669, 2004.

[111] K. S. Novoselov, D. Jiang, F. Schedin, T. J. Booth, V. V. Khotkevich, S. V. Morozov, and A. K. Geim. Two-dimensional atomic crystals. *Proceedings of the National Academy of Sciences*, 102(30):10451–10453, 2005.

[112] A. R. Oganov and C. W. Glass. Crystal structure prediction using ab initio evolutionary techniques: principles and applications. *The Journal of Chemical Physics*, 124(24):244704, 2006.

[113] Artem R. Oganov. Evolutionary crystal structure prediction and computational materials discovery, 2013.

[114] Artem R. Oganov, Jiuhua Chen, Carlo Gatti, Yanzhang Ma, Yanming Ma,

Colin W. Glass, Zhenxian Liu, Tony Yu, Oleksandr O. Kurakevych, and Vladimir L. Solozhenko. Ionic high-pressure form of elemental boron. *Nature*, 457(7231):863–867, 2009.

[115] Artem R. Oganov and Colin W. Glass. Crystal structure prediction using ab initio evolutionary techniques: principles and applications. *The Journal of Chemical Physics*, 124(24):244704, 2006.

[116] Artem R. Oganov and Colin W. Glass. Evolutionary crystal structure prediction as a tool in materials design. *Journal of Physics: Condensed Matter*, 20(6):064210, 2008.

[117] Artem R. Oganov, Colin W. Glass, and Shigeaki Ono. High-pressure phases of $CaCO_3$: crystal structure prediction and experiment. *Earth and Planetary Science Letters*, 241(1):95–103, 2006.

[118] Artem R. Oganov and Andriy O. Lyakhov. Towards the theory of hardness of materials. *Journal of Superhard Materials*, 32(3):143–147, 2010.

[119] Shyue Ping Ong, William Davidson Richards, Anubhav Jain, Geoffroy Hautier, Michael Kocher, Shreyas Cholia, Dan Gunter, Vincent L. Chevrier, Kristin A. Persson, and Gerbrand Ceder. Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis. *Computational Materials Science*, 68:314–319, 2013.

[120] Shigeaki Ono, Artem R. Oganov, John P. Brodholt, Lidunka Vočadlo, Ian G. Wood, Andriy Lyakhov, Colin W. Glass, Alexander S. Côté, and G. David Price. High-pressure phase transformations of FeS: novel phases at conditions of planetary cores. *Earth and Planetary Science Letters*, 272(1):481–487, 2008.

[121] Yoshifumi Oshima, Akiko Onga, and Kunio Takayanagi. Helical gold nanotube synthesized at 150 K. *Physical Review Letters*, 91(20):205503, 2003.

[122] J. Pannetier, J. Bassas-Alsina, J. Rodriguez-Carvajal, and V. Caignaert. Prediction of crystal structures from crystal chemistry rules by simulated annealing. *Nature*, 346(6282):343–345, 1990.

[123] Hyoungki Park, Michael R. Fellinger, Thomas J. Lenosky, William W. Tipton, Dallas R. Trinkle, Sven P. Rudin, Christopher Woodward, John W. Wilkins, and Richard G. Hennig. Ab initio based empirical potential used to study the mechanical properties of molybdenum. *Physical Review B*, 85(21):214121, 2012.

[124] John P. Perdew, Kieron Burke, and Matthias Ernzerhof. Generalized gradient approximation made simple. *Physical Review Letters*, 77(18):3865, 1996.

[125] C. J. Pickard and R. J. Needs. Ab initio random structure searching. *Journal of Physics: Condensed Matter*, 23(5):053201, 2011.

[126] Chris J. Pickard and R. J. Needs. High-pressure phases of silane. *Physical Review Letters*, 97(4):045504, 2006.

[127] Dasari L. V. K. Prasad, N. W. Ashcroft, and Roald Hoffmann. Lithium amide ($LiNH_2$) under pressure. *The Journal of Physical Chemistry A*, 116(40):10027–10036, 2012.

[128] Sarah L. Price. The computational prediction of pharmaceutical crystal structures and polymorphism. *Advanced Drug Delivery Reviews*, 56(3):301–319, 2004.

[129] Wayne J. Pullan. Structure prediction of benzene clusters using a genetic algorithm. *Journal of Chemical Information and Computer Sciences*, 37(6):1189–1193, 1997.

[130] B. Radisavljevic, A. Radenovic, J. Brivio, V. Giacometti, and A. Kis. Single-layer MoS_2 transistors. *Nature Nanotechnology*, 6(3):147–150, 2011.

[131] Benjamin C. Revard, William W. Tipton, and Richard G. Hennig. Structure and stability prediction of compounds with evolutionary algorithms. In *Prediction and Calculation of Crystal Structures*, pages 181–222. Springer, 2014.

[132] Benjamin C. Revard, William W. Tipton, Anna Yesypenko, and Richard G. Hennig. Grand-canonical evolutionary algorithm for the prediction of two-dimensional materials. *Physical Review B*, 93(5):054117, 2016.

[133] Pablo Rivero, Jia-An Yan, Víctor M. García-Suárez, Jaime Ferrer, and Salvador Barraza-Lopez. Stability and properties of high-buckled two-dimensional tin and lead. *Physical Review B*, 90(24):241408, 2014.

[134] Guillermo Román-Pérez and José M. Soler. Efficient implementation of a van der Waals density functional: application to double-wall carbon nanotubes. *Physical Review Letters*, 103(9):096102, 2009.

[135] Sven P. Rudin, Matt D. Jones, and Robert C. Albers. Thermal stabilization of the hcp phase in titanium. *Physical Review B*, 69(9):094117, 2004.

[136] H. Şahin, S. Cahangirov, M. Topsakal, E. Bekaroglu, E. Akturk, R. Tugrul Senger, and Salim Ciraci. Monolayer honeycomb structures of group-IV elements and III-V binary compounds: first-principles calculations. *Physical Review B*, 80(15):155453, 2009.

[137] Andrew Shamp, James Hooper, and Eva Zurek. Compressed cesium polyhydrides: Cs^+ sublattices and H_3^- three-connected nets. *Inorganic Chemistry*, 51(17):9333–9342, 2012.

[138] Y. Shirako, H. Kojitani, A. R. Oganov, K. Fujino, H. Miura, D. Mori, Y. Inaguma, K. Yamaura, and M. Akaogi. Crystal structure of $CaRhO_3$ polymorph: high-pressure intermediate phase between perovskite and post-perovskite. *American Mineralogist*, 97(1):159–163, 2012.

[139] Luke Shulenburger, Andrew D. Baczewski, Zhen Zhu, Jie Guan, and David Tomanek. The nature of the interlayer interaction in bulk and few-layer phosphorus. *Nano Letters*, 15(12):8170–8175, 2015.

[140] A. K. Singh and R. G. Hennig. Computational prediction of two-dimensional group-IV mono-chalcogenides. *Applied Physics Letters*, 105(4):042103, 2014.

[141] A. K. Singh, H. L. Zhuang, and R. G. Hennig. Ab initio synthesis of single-layer III-V materials. *Physical Review B*, 89(24):245431, 2014.

[142] Arunima K. Singh and Richard G. Hennig. Computational synthe-

sis of single-layer GaN on refractory materials. *Applied Physics Letters*, 105(5):051604, 2014.

[143] Arunima K. Singh, Kiran Mathew, Houlong L. Zhuang, and Richard G. Hennig. Computational screening of 2D materials for photocatalysis. *The Journal of Physical Chemistry Letters*, 6(6):1087–1098, 2015.

[144] Arunima K. Singh, Benjamin C. Revard, Rohit Ramanathan, Michael Ashton, Francesca Tavazza, and Richard G. Hennig. Genetic algorithm prediction of two-dimensional group-IV dioxides for dielectrics. *Physical Review B*, 95(15):155426, 2017.

[145] Richard W. Smith. Energy minimization in binary alloy models via genetic algorithms. *Computer Physics Communications*, 71(1):134–146, 1992.

[146] P. Souvatzis, O. Eriksson, M. I. Katsnelson, and S. P. Rudin. Entropy driven stabilization of energetically unstable crystal structures explained from first principles theory. *Physical Review Letters*, 100:095901, 2008.

[147] Leonardo Spanu, Sandro Sorella, and Giulia Galli. Nature and strength of interlayer binding in graphite. *Physical Review Letters*, 103(19):196401, 2009.

[148] A. Splendiani, L. Sun, Y. Zhang, T. Li, J. Kim, C.Y. Chim, G. Galli, and F. Wang. Emerging photoluminescence in monolayer MoS_2. *Nano Letters*, 10(4):1271–1275, 2010.

[149] Frank H. Stillinger. Exponential multiplicity of inherent structures. *Physical Review E*, 59(1):48, 1999.

[150] Harold T. Stokes and Dorian M. Hatch. FINDSYM: program for identifying the space-group symmetry of a crystal. *Journal of Applied Crystallography*, 38(1):237–238, 2005.

[151] Shaojian Sun. Reduced representation model of protein structure prediction: statistical potential and genetic algorithms. *Protein Science*, 2(5):762–785, 1993.

[152] Yongfu Sun, Hao Cheng, Shan Gao, Zhihu Sun, Qinghua Liu, Qin Liu, Fengcai Lei, Tao Yao, Jingfu He, Shiqiang Wei, and Yi Xie. Freestanding tin disulfide single-layers realizing efficient visible-light water splitting. *Angewandte Chemie International Edition*, 51(35):8727–8731, 2012.

[153] Toma Susi, Jani Kotakoski, Demie Kepaptsoglou, Clemens Mangler, Tracy C. Lovejoy, Ondrej L. Krivanek, Recep Zan, Ursel Bangert, Paola Ayala, Jannik C. Meyer, et al. Silicon–carbon bond inversions driven by 60-keV electrons in graphene. *Physical Review Letters*, 113(11):115501, 2014.

[154] A. P. Sutton and J. Chen. Long-range Finnis–Sinclair potentials. *Philosophical Magazine Letters*, 61(3):139–146, 1990.

[155] Tatsuo Suzuki. Theoretical discovery of stable structures of group III-V monolayers: the materials for semiconductor devices. *Applied Physics Letters*, 107(21):213105, 2015.

[156] Joshua A. Taillon, William W. Tipton, and Richard G. Hennig. Ab initio discovery of novel crystal structure stability in barium and sodium-calcium compounds under pressure using DFT. *arXiv preprint arXiv:1207.3320*, 2012.

[157] D. Teweldebrhan, V. Goyal, and A. A. Balandin. Exfoliation and characterization of bismuth telluride atomic quintuples and quasi-two-dimensional crystals. *Nano Letters*, 10(4):1209–1218, 2010.

[158] William W. Tipton, Clive R. Bealing, Kiran Mathew, and Richard G. Hennig. Structures, phase stabilities, and electrical potentials of Li-Si battery anode materials. *Physical Review B*, 87(18):184114, 2013.

[159] William W. Tipton and Richard G. Hennig. A grand canonical genetic algorithm for the prediction of multi-component phase diagrams and testing of empirical potentials. *Journal of Physics: Condensed Matter*, 25(49):495401, 2013.

[160] William W. Tipton, Catherine A. Matulis, and Richard G. Hennig. Ab initio prediction of the Li_5Ge_2 Zintl compound. *Computational Materials Science*, 93:133–136, 2014.

[161] Chuan-Jia Tong, Hui Zhang, Yan-Ning Zhang, Hao Liu, and Li-Min Liu. New manifold two-dimensional single-layer structures of zinc-blende compounds. *Journal of Materials Chemistry A*, 2(42):17971–17978, 2014.

[162] E. Tosatti, S. Prestipino, S. Kostlmeier, A. Dal Corso, and F. D. Di Tolla. String tension and stability of magic tip-suspended nanowires. *Science*, 291(5502):288–290, 2001.

[163] Giancarlo Trimarchi, Arthur J. Freeman, and Alex Zunger. Predicting stable stoichiometries of compounds via evolutionary global space-group optimization. *Physical Review B*, 80(9):092101, 2009.

[164] Giancarlo Trimarchi and Alex Zunger. Global space-group optimization problem: Finding the stablest crystal structure without constraints. *Physical Review B*, 75:104113, 2007.

[165] Giancarlo Trimarchi and Alex Zunger. 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. *Journal of Physics: Condensed Matter*, 20(29):295212, 2008.

[166] G. A. Tritsaris, B. D. Malone, and E. Kaxiras. Optoelectronic properties of single-layer, double-layer, and bulk tin sulfide: a theoretical study. *Journal of Applied Physics*, 113(23):233507, 2013.

[167] M. Valle and Artem R. Oganov. Crystal structures classifier for an evolutionary algorithm structure predictor. In *Visual Analytics Science and Technology, 2008. VAST '08. IEEE Symposium on*, pages 11–18, 2008.

[168] Mario Valle and Artem R. Oganov. Crystal fingerprint space – a novel paradigm for studying crystal-structure sets. *Acta Crystallographica Section A*, 66(5):507–517, 2010.

[169] Robert Bruce van Dover, L. F. Schneemeyer, and R. M. Fleming. Discovery of a useful thin-film dielectric using a composition-spread approach. *Nature*, 392(6672):162–164, 1998.

[170] Prasana K. Venkatesh, Morrel H. Cohen, Robert W. Carr, and Anthony M. Dean. Bayesian method for global optimization. *Physical Review E*, 55(5):6219, 1997.

[171] Patrick Vogt, Paola De Padova, Claudio Quaresima, Jose Avila, Emmanouil Frantzeskakis, Maria Carmen Asensio, Andrea Resta, Bénédicte Ealet, and Guy Le Lay. Silicene: compelling experimental evidence for graphenelike two-dimensional silicon. *Physical Review Letters*, 108(15):155501, 2012.

[172] A. F. Voter. Embedded atom method potentials for seven fcc metals: Ni, Pd, Pt, Cu, Ag, Au, and Al. *Los Alamos National Laboratory, Unclassified Technical Report No. LA-UR*, pages 93–3901, 1993.

[173] Hui Wang, Quan Li, Yinwei Li, Ying Xu, Tian Cui, Artem R. Oganov, and Yanming Ma. Ultra-incompressible phases of tungsten dinitride predicted from first principles. *Physical Review B*, 79(13):132109, 2009.

[174] Jinlan Wang, Guanghou Wang, and Jijun Zhao. Density-functional study of Au_n ($n = 2-20$) clusters: Lowest-energy structures and electronic properties. *Physical Review B*, 66(3):035418, 2002.

[175] Yanchao Wang, Jian Lv, Li Zhu, and Yanming Ma. Crystal structure prediction via particle-swarm optimization. *Physical Review B*, 82(9):094116, 2010.

[176] Yanchao Wang, Jian Lv, Li Zhu, and Yanming Ma. CALYPSO: a method for crystal structure prediction. *Computer Physics Communications*, 183(10):2063 – 2070, 2012.

[177] Xiao-Dong Wen, Thomas J. Cahill, Nicholas M. Gerovac, Michael J. Bucknum, and Roald Hoffmann. Playing the quantum chemical slot machine: an exploration of ABX_2 compounds. *Inorganic chemistry*, 49(1):249–260, 2009.

[178] Xiao-Dong Wen, Thomas J. Cahill, Roald Hoffmann, and Akira Miura. Tuning of metal–metal bonding by counterion size in hypothetical $AeTiO_2$ compounds. *Journal of the American Chemical Society*, 131(41):14632–14633, 2009.

[179] Xiao-Dong Wen, Louis Hand, Vanessa Labet, Tao Yang, Roald Hoffmann,

N. W. Ashcroft, Artem R. Oganov, and Andriy O. Lyakhov. Graphane sheets and crystals under pressure. *Proceedings of the National Academy of Sciences*, 108(17):6833–6837, 2011.

[180] S. M. Woodley, P. D. Battle, J. D. Gale, and C. R. A. Catlow. The prediction of inorganic crystal structures using a genetic algorithm and energy minimisation. *Physical Chemistry Chemical Physics*, 1(10):2535–2542, 1999.

[181] S. M. Woodley and C. R. A. Catlow. Structure prediction of titania phases: implementation of Darwinian versus Lamarckian concepts in an evolutionary algorithm. *Computational Materials Science*, 45(1):84 – 95, 2009.

[182] Scott M. Woodley. Prediction of inorganic crystal framework structures part 2: using a genetic algorithm and a direct approach to exclusion zones. *Physical Chemistry Chemical Physics*, 6:1823–1829, 2004.

[183] Scott M. Woodley. Engineering microporous architectures: combining evolutionary algorithms with predefined exclusion zones. *Physical Chemistry Chemical Physics*, 9(9):1070–1077, 2007.

[184] Shunqing Wu, Koichiro Umemoto, Min Ji, Cai-Zhuang Wang, Kai-Ming Ho, and Renata M. Wentzcovitch. Identification of post-pyrite phase transitions in SiO_2 by a genetic algorithm. *Physical Review B*, 83:184102, 2011.

[185] Wenzhuo Wu, Lei Wang, Yilei Li, Fan Zhang, Long Lin, Simiao Niu, Daniel Chenet, Xian Zhang, Yufeng Hao, Tony F Heinz, and Zhong Lin Hone, James amd Wang. Piezoelectricity of single-atomic-layer MoS_2 for energy conversion and piezotronics. *Nature*, 514(7523):470–474, 2014.

[186] Li Xiao and Lichang Wang. From planar to three-dimensional structural transition in gold clusters and the spin–orbit coupling effect. *Chemical Physics Letters*, 392(4):452–455, 2004.

[187] Yongliang Xiao and Donald E. Williams. Genetic algorithm: a new approach to the prediction of the structure of molecular clusters. *Chemical Physics Letters*, 215(1):17–24, 1993.

[188] Yu Xie, Artem R. Oganov, and Yanming Ma. Novel high pressure structures and superconductivity of $CaLi_2$. *Physical Review Letters*, 104(17):177005, 2010.

[189] Ying Xu, S. Tse John, Artem R. Oganov, Tian Cui, Hui Wang, Yanming Ma, and Guangtian Zou. Superconducting high-pressure phase of cesium iodide. *Physical Review B*, 79(14):144110, 2009.

[190] Yong Xu, Binghai Yan, Hai-Jun Zhang, Jing Wang, Gang Xu, Peizhe Tang, Wenhui Duan, and Shou-Cheng Zhang. Large-gap quantum spin Hall insulators in tin films. *Physical Review Letters*, 111(13):136804, 2013.

[191] Patryk Zaleski-Ejgierd, Roald Hoffmann, and N. W. Ashcroft. High pressure stabilization and emergent forms of PbH_4. *Physical Review Letters*, 107(3):037002, 2011.

[192] Feiwu Zhang and Artem R. Oganov. Iron silicides at pressures of the Earth's inner core. *Geophysical Research Letters*, 37(2), 2010.

[193] Weiwei Zhang, Artem R. Oganov, Alexander F. Goncharov, Qiang Zhu, Salah Eddine Boulfelfel, Andriy O. Lyakhov, Maddury Somayazulu, and Vitali B. Prakapenka. Unexpected stable stoichiometries of sodium chlorides. *arXiv preprint arXiv:1211.3644*, 2012.

[194] Xiuwen Zhang, Giancarlo Trimarchi, and Alex Zunger. Possible pitfalls in theoretical determination of ground-state crystal structures: the case of platinum nitride. *Physical Review B*, 79:092102, 2009.

[195] Jin Zhao, Jinlong Yang, and J. G. Hou. Theoretical study of small two-dimensional gold clusters. *Physical Review B*, 67(8):085404, 2003.

[196] Yan Zhong, Huai-Ying Zhou, Chao-Hao Hu, Dian-Hui Wang, and Artem R. Oganov. Theoretical studies of high-pressure phases, electronic structure, and vibrational properties of $NaNH_2$. *The Journal of Physical Chemistry C*, 116(15):8387–8393, 2012.

[197] L. J. Zhou, Y. F. Zhang, and L. M. Wu. SiC_2 siligraphene and nanotubes:

novel donor materials in excitonic solar cells. *Nano Letters*, 13(11):5431–5436, 2013.

[198] Xiang-Feng Zhou, Xiao Dong, Artem R Oganov, Qiang Zhu, Yongjun Tian, and Hui-Tian Wang. Semimetallic two-dimensional boron allotrope with massless Dirac fermions. *Physical Review Letters*, 112(8):085502, 2014.

[199] Xiang-Feng Zhou, Xiao Dong, Zhisheng Zhao, Artem R. Oganov, Yongjun Tian, and Hui-Tian Wang. High-pressure phases of $NaAlH_4$ from first principles. *Applied Physics Letters*, 100(6):061905–061905, 2012.

[200] Xiang-Feng Zhou, Artem R. Oganov, Xiao Dong, Lixin Zhang, Yongjun Tian, and Hui-Tian Wang. Superconducting high-pressure phase of platinum hydride from first principles. *Physical Review B*, 84(5):054543, 2011.

[201] Xiang-Feng Zhou, Artem R. Oganov, Guang-Rui Qian, and Qiang Zhu. First-principles determination of the structure of magnesium borohydride. *Physical Review Letters*, 109(24):245503, 2012.

[202] Feng-feng Zhu, Wei-jiong Chen, Yong Xu, Chun-lei Gao, Dan-dan Guan, Can-hua Liu, Dong Qian, Shou-Cheng Zhang, and Jin-feng Jia. Epitaxial growth of two-dimensional stanene. *Nature Materials*, 2015.

[203] Qiang Zhu, Daniel Y. Jung, Artem R. Oganov, Colin W. Glass, Carlo Gatti, and Andriy O. Lyakhov. Stability of xenon oxides at high pressures. *Nature Chemistry*, 5(1):61–65, 2012.

[204] Qiang Zhu, Artem R. Oganov, Colin W. Glass, and Harold T. Stokes. Constrained evolutionary algorithm for structure prediction of molecular crystals: methodology and applications. *Acta Crystallographica Section B: Structural Science*, 68(3):215–226, 2012.

[205] Qiang Zhu, Artem R. Oganov, and Andriy O. Lyakhov. Evolutionary metadynamics: a novel method to predict crystal structures. *CrystEngComm*, 14(10):3596–3601, 2012.

[206] Qiang Zhu, Artem R. Oganov, and Andriy O. Lyakhov. Unexpected

stoichiometries in Mg-O system under high pressure. *arXiv preprint arXiv:1211.6521*, 2012.

[207] H. L. Zhuang and R. G. Hennig. Computational search for single-layer transition-metal dichalcogenide photocatalysts. *The Journal of Physical Chemistry C*, 117(40):20440–20445, 2013.

[208] H. L. Zhuang and R. G. Hennig. Computational discovery, characterization, and design of single-layer materials. *JOM*, 66(3):366–374, 2014.

[209] Houlong L. Zhuang and Richard G. Hennig. Electronic structures of single-layer boron pnictides. *Applied Physics Letters*, 101(15):153109, 2012.

[210] Houlong L. Zhuang and Richard G. Hennig. Computational identification of single-layer CdO for electronic and optical applications. *Applied Physics Letters*, 103(21):212102, 2013.

[211] Houlong L. Zhuang and Richard G. Hennig. Single-layer group-III monochalcogenide photocatalysts for water splitting. *Chemistry of Materials*, 25(15):3232–3238, 2013.

[212] Houlong L. Zhuang and Richard G. Hennig. Theoretical perspective of photocatalytic properties of single-layer SnS_2. *Physical Review B*, 88(11):115314, 2013.

[213] Houlong L. Zhuang, Michelle D. Johannes, Michael N. Blonsky, and Richard G. Hennig. Computational prediction and characterization of single-layer CrS_2. *Applied Physics Letters*, 104(2):022116, 2014.

[214] Houlong L. Zhuang, Arunima K. Singh, and Richard G. Hennig. Computational discovery of single-layer III-V materials. *Physical Review B*, 87(16):165415, 2013.

[215] Eva Zurek, Roald Hoffmann, N. W. Ashcroft, Artem R. Oganov, and Andriy O. Lyakhov. A little bit of lithium does a lot for hydrogen. *Proceedings of the National Academy of Sciences*, 106(42):17640–17643, 2009.