TECHNISCHE UNIVERSITÄT
BERGAKADEMIE FREIBERG
The University of Resources. Since 1765.

Development of an evolutionary algorithm for crystal structure prediction

By the Faculty of Chemistry and Physics
of the Technische Universität Bergakademie Freiberg

approved

Thesis

to attain the academic degree of

doctor rerum naturalium
(Dr. rer. nat.)

submitted by **Dipl. Nat. Silvia Bahmann**

born on the 7. August 1983 in Weißenfels

Assessor: Prof. Dr. Jens Kortus
Prof. Dr. Eva Zurek

Date of the award:Freiberg, 15th April 2014

Versicherung

Hiermit versichere ich, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe; die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht.

Die Hilfe eines Promotionsberaters habe ich nicht in Anspruch genommen. Weitere Personen haben von mir keine geldwerten Leistungen für Arbeiten erhalten, die nicht als solche kenntlich gemacht worden sind. Die Arbeit wurde bisher weder im Inland noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

15. April 2014 Dipl. Nat. Silvia Bahmann

Declaration

I hereby declare that I completed this work without any improper help from a third party and without using any aids other than those cited. All ideas derived directly or indirectly from other sources are identified as such.

I did not seek the help of a professional doctorate-consultant. Only those persons identified as having done so received any financial payment from me for any work done for me. This thesis has not previously been published in the same or a similar form in Germany or abroad.

15th April 2014 Dipl. Nat. Silvia Bahmann

Contents

1. Introduction

Computational materials design aims at finding new materials with superior properties. Envisioned is also the ability to guide experimental research in a particular direction. Possibly, some day one can design materials to satisfy very specific needs via calculations and afterwards show also concrete paths to their realisation. Unfortunately, such a procedure is not feasible today yet parts of it can be achieved.

The Freiberg High Pressure research centre (FHP) aimed at the ambitious goal to engineer new materials that are produced and/or can be used at very high pressures. Special focus lay on the finding of alternative materials for drill bits in deep drilling especially in hard rock where hardness as well as the resistance to high temperatures and corrosion is essential. For this application, polycrystalline diamond is widely used today as it exhibits a very high hardness but is prone to oxidation at high temperatures. New materials should be superior in this regime and preferably less cost intensive.

In the aspiring vision of the FHP an ideal material development process would consist of the following parts:

- Theoretical physics proposes crystal structures of new materials that are stable at high pressures and additionally very hard.
- Chemists and high pressure experts synthesise the material and produce considerable amounts of it.
- Materials researchers provide a detailled analysis of the material, its microstructure and properties. Furthermore, they test the material under extreme conditions such as pressure, temperature and speed.
- Engineers develop and optimise the tools needed for appropriate applications.
- The last step would be the transfer of the material to the industrial use.

The work presented here covers the first item of this list: Finding and evaluating new crystal structures.

One approach to crystal structure search is completely based on understanding or chemical and crystallographic intuition. The danger of this methodology is that it may be biased towards already known structures. As a consequence, certain classes of new structures may be overlooked. Treating crystal structure search as a pure optimisation problem on the other hand avoids such prejudices.

Mathematically, this is a global optimisation problem of a function that has $3N + 3$ variables[1] where N is the number of atoms per unit cell. The value of this multidimensional function is the free energy of the crystal structure. The solution to this problem is complicated by three factors: First, the dimensionality of the search space is high and increases with the number of atoms N in the unit cell. Second, no analytical form is known for the objective function which must therefore be evaluated numerically (e.g. using DFT, which is computationally expensive, see chapter 2). Third, there may exist many local minima yet the most stable structure of a given composition occupies the global minimum of the energy landscape at certain conditions (pressure, temperature).

Several approaches to general global optimisation are known and many of them are also applied to crystal structure search (see chapter 3). The particular method used here is an evolutionary algorithm which mimics natural evolution: Individuals are then the crystal structures; several of them form a population. Within a population offspring are produced by recombination and mutation and favourable ones are chosen in a selection process. The method and the concrete realisation will be explained in detail in chapter 4. EVO, the program developed for crystal structures search in this work, features the search in 3-dimensional periodic boundary conditions and is designed for an easy usage as well as further development. It allows for an automated search of crystal structures where the only necessary input is the numbers and species of atoms in the unit cell. A successful search with EVO yields several crystal structures that occupy minima in the energy landscape and are thus suitable candidates for stable structures.

EVO implements several features to minimise the number of evaluations of the objective function to reduce computational cost. At the same time, the search has to stay thorough. It proved to effectively explore the energy landscape and to find known structures reliably. Moreover, previously unknown structures are detected.

The detailled evaluation of discovered structures complements the pure search to a prediction of crystal structures. Its purpose is to assess indicators for stability which are computationally too expensive to be included into the search procedure. Additionally, many properties can be theoretically predicted and examined for stable structures. For possible applications, mechanical or electronic properties may be of special interest.

Some results of the whole process of crystal structure prediction, namely search and evaluation, are presented in chapter 6. The first structure is crossed graphene, a new and very interesting modification of carbon. It is also chosen to illustrate the prediction procedure. Other discoveries are germanium nitrofluoride as a new ternary system and a previously unknown boron sheet structure.

[1] These are composed of the six parameters that describe the unit cell (three lengths and three angles) and $N - 1$ atoms that each need three coordinates.

2. Theoretical background and computational tools

The crystal structure is the most important information one can have about a material. This is mainly due to the fact that, nowadays, theoretical physics is able to extract many properties and material constants from this information alone. For an accurate description of solids it is necessary to characterise the bonding situation within the material. Since bonds are built of electrons, the precise knowledge of the electronic ground state of a given crystal structure gives insight into stability and ground state properties. Owing to the great importance of the electronic ground state there are different methods that tackle the challenge to describe the electronic ground state from first principles where Density Functional Theory (DFT) is the most famous and widely used today. Therefore, the following chapter shall deal with the fundamental concepts of DFT. However, many details will be omitted to achieve brevity and an overview shall be given on the properties of materials and how they are assessed in practice. The information will be far from complete and we will rather concentrate on the calculations used within this thesis. An estimation of the computational costs will also be given to some extent. Furthermore, there are many excellent textbooks covering all the specifics. The one explicitely recommended and excessively used here is 'Electronic structure' by R. M. Martin [1] that not only covers fundamentals and methods but also describes many findings along with a vast pool of valuable references.

Moreover, the present chapter will have a closer look at the methods and programs utilised for the calculations of the results mainly introduced in chapter 6. Those methods rely on the precise determination of the electronic ground state which will therefore be the starting point. For completeness, a small part will also cover visualisation programs used.

2.1. Theoretical foundations

2.1.1. Fundamentals of DFT

Schrödingers equation is the starting point for our argument as it is the basis for electronic structure theory which fully describes each quantum-mechanical system. The

equation is time-dependent but for the charactisation of the electronic ground state its time-independent form is sufficient[1]:

$$\hat{H}\psi = E\psi$$

where the Hamilton operator $\hat{H}$ is applied to the wave function ψ yielding the energy E of the system. Thus, solving this equation and getting the wave function enables us to fully describe the whole system. It has to be noted that $\psi(\vec{r})$ itself has no physical meaning only $|\psi|^2$ is a function denoting the probability density to find a particle in a selected position.

Depending on the actual Hamiltonian for the specific system Schrödingers equation can be solved analytically for small systems (e.g. free particle, harmonic oscillator, hydrogen atom) but must be solved numerically for bigger systems which can become computationally demanding. Materials – even when considering 3-dimensional periodic boundary conditions – are far from being small systems: They have to be described by a many-body Hamiltonian including all the interactions between the particles. Taking into account that even the lightest nuclei are much heavier than electrons the Born-Oppenheimer approximation can be applied. It states that the electrons move much faster compared to the nuclei and thus are in their ground state instantaneously with respect to the positions of the nuclei. Following this assumption rigorously the electrons move in a lattice of fixed nuclei which removes the kinetic energy of the nuclei as well as the nuclei-nuclei interaction from the Hamiltonian. This leads to:

$$\hat{H} = \hat{T} + \hat{V} + \hat{U}$$

where $\hat{T}$ is the kinetic energy, $\hat{V}$ describes the electrostatic potential working on the electrons due to the positively charged nuclei and $\hat{U}$ covers the electron-electron interaction.

The major clue to DFT and also origin of the denomination is the substitution of the wave function dependent on the positions of all electrons by one depending on the density n alone:

$$\psi(\vec{r}_1, \vec{r}_2, \ldots, \vec{r}_N) = \psi(n(\vec{r}))$$

which reduces dimensions dramatically from 3N to 3. Of course, the number of particles N_{part} needs to be preserved:

$$\int n(\vec{r})\mathrm{d}\vec{r} = N_{part}$$

Replacing the electrons by a joint electron density is only possible according to the theorem of Hohenberg and Kohn [2]. They proved that the density $n(\vec{r})$ is unique to a

[1] We will therefore stick to the time-independent formalism in the following.

given external potential (i.e. atomic lattice).This also means that there cannot be two distinct potentials $v_1 \neq v_2$ resulting in the same density for the ground state and that the ground state density solely determines the external potential.

Moreover, Hohenberg and Kohn introduced a universal functional for the energy $E[n]$ and showed that the exact ground state energy E_0 is the global minimum for any given external potential. Also, the exact ground state density n_0 minimises the functional:

$$E_0[n_0] \leq E[n] \qquad \forall \quad n_0 \neq n$$

The equality in the above equation only holds for degenerate ground states.

To practically solve the problem requires a further step: Kohn and Sham [3] introduced an auxiliary system of non-interacting particles that produces the same ground state density as the original interacting electron system. The wavefunctions of these non-interacting particles are denoted as $\phi_i(\vec{r})$. In other words, the sum of the the single-particle densities $|\phi_i(\vec{r})|^2$ yields the actual ground state density.

$$n(\vec{r}) = \sum_{i=1}^{N} |\phi_i(\vec{r})|^2$$

Furthermore, Kohn and Sham established the Kohn-Sham equation which uses an auxiliary Hamiltonian that leads to the one-particle Schrödinger equation:

$$\left(-\frac{\hbar}{2m}\nabla^2 + V_{eff}(\vec{r})\right)\phi_i(\vec{r}) = \varepsilon_i \phi_i(\vec{r})$$

where the first term is the kinetic energy of the particle and V_{eff} contains all other terms as the Coulomb interaction of the electron density with itself (self-interaction), the external potential and the exchange-correlation potential; ε_i are the energy eigenvalues of the Hamiltonian. All the complicated many-body physics is transferred to the exchange-correlation potential by this approach. Unfortunately, it is infeasible to treat exchange and correlation exactly yet several methods have been introduced to solve the problem. We will resign here from explaining the different approaches and rather refer to [1] or one of the other available textbooks on electronic structure theory.

The typical computational procedure to determine the ground state density is the self-consistent solution of the Kohn-Sham equations: It starts with an initial guess of $n(\vec{r})$ used to calculate the effective potential V_{eff}. Incorporating this in the Hamiltonian the single-particle wave functions can be obtained by solving the Kohn-Sham equation. Summing their densities yields a new electron density $n_{new}(\vec{r})$ which is compared to the initial guess. If they differ by more than a certain (small) threshold value, $n_{new}(\vec{r})$ reenters the procedure for the calculation of a new effective potential. Otherwise, self-consistency is reached and the ground state density has been found which can then be

used to get certain output values: energy, forces and stresses, for example. A single determination of the ground state density is called self-consistent field (scf) calculation.

Information that can be gained directly from the ground state density using some minor postprocessing steps are all kinds of electronic properties (of the ground state) as for example: band structure, electronic density of states, charge densities or the Electron Localisation Function (ELF) [4].

2.1.2. Forces and stresses

The ability to determine the electronic ground state of a given crystal structure (i.e. an atomic lattice) is essential for all further calculations. However, in many applications it is also necessary to have access to the 'atomic ground state' or, in other words, the atomic configuration that would be adopted if the atoms were free to move and/or the unit cell free to deform. This corresponds to a (local) optimisation in terms of a minimisation of the energy by reducing forces between the atoms and stresses on the cell. Of course, one could also undirectedly change the atomic positions or unit cell parameters and hope for lower energies. Yet, far more effective would be the use of the information of forces and stresses. So, naturally, there were efforts to accurately describe the two quantities within DFT.

The determination of forces is quite straightforward: As stated by the so-called Hellmann-Feynman or simply force theorem they are defined to be the derivative of the total energy E with respect to the atomic positions R_i:

$$F_i = -\frac{\partial E}{\partial R_i}$$

For the calculation of stresses $\sigma_{\alpha\beta}$ a similar approach was introduced [5]:

$$\sigma_{\alpha\beta} = -\frac{1}{V}\frac{\partial E}{\partial \epsilon_{\alpha\beta}}$$

where V is the volume of the cell and ϵ is strain matrix which sends all coordinates $\vec{r}_i$ including atomic positions and unit cell vectors to new ones $\vec{r}_i{}'$ so that $\vec{r}_i{}' = (1+\epsilon)\vec{r}_i$. α and β are the cartesian indices. In this framework, pressure is defined to be one third of the negative trace of σ: $P = -\frac{\mathrm{Tr}\,\sigma}{3}$. For a detailled derivation of the quantum-mechanical theory of stresses and forces refer to [6] or relevant textbooks.

The importance of the ability to calculate stresses and forces is reasonable when trying to achieve equilibrium crystal structures. Those structures need to fulfil two criteria:

- The atoms need to be force-free.
- The stress σ has to equal the external stress on the cell. For example, if one wants to apply a certain hydrostatic pressure, an external pressure has to be defined. In

the hydrostatic case, the stress σ only has non-zero values on the main diagonal which are required to be identical.

With the knowledge of forces and stresses a directional optimisation towards an equilibrium structure can be realised using standard gradient methods. Naturally, these can only reach the minimum of the zone of attraction where the starting structure is located but they are highly efficient. Many sophisticated electronic structure program packages include such minimisation techniques where the computational cost of each iteration step is comparable to a single scf-calculation. The terms used in the following are:

- relaxation in the case of pure force minimisation, i.e. the form of the unit cell remains unchanged and
- variable-cell relaxation for minimisation of forces and optimisation of stress at the same time. In the vicinity of the global minimum this leads to the most stable structure at a given pressure. Possibly found local minima are termed to be metastable.

Moreover, stresses are of great use for the determination of elastic constants which can then be obtained from the first derivative of the stress-strain curve. A practical tool for the determination of the elastic constants is the ElaStic code recently published; the accompanying article [7] also explains the particulars in a more detailled way and contains lots of references to the original publications. The elastic constants reported in chapter 6 were gained by a similar procedure.

2.1.3. Calculation of phonons and Raman spectra

The study of vibrational spectra is subject of a large volume of experimental research and a standard tool for the charactisation of materials nowadays. Naturally, there are also methods in theory that provide access to the vibrational properties of crystals (see [8] for the fundamental physics). We stick here to the description of phonons which are collective oscillations of atoms in a crystal lattice exhibiting a distinct frequency. More graphically, it is the 'response' of the lattice to a displacement of atoms. The application of the Born-Oppenheimer approximation determines them to be a ground state property since the electrons adopt their ground state instantaneously to a given lattice. For the practical determination of the phonon frequencies there are two approaches available. As before, a more detailled description can be found in [1] and contained references. However, rough principles shall be explained here.

A very popular method is called 'frozen phonons' which is very easily applied since it directly calculates the total energy/forces on the atoms of the lattice as a function of the atomic positions. Therefore, it can be conducted using every program for electronic

structure calculations and no special methodology or implementation is needed. The interatomic force constants can be gained by calculating the numerical derivatives from finite differences for various values of the perturbation (changed atomic position). This is quite straightforward for phonons at the Γ point where the modes are defined with the finite displacement of all the atoms in a single cell. However, for phonons at an arbitrary q-point necessary for the calculation of phonon dispersions the situation becomes more difficult. One has to use supercells 'whose size depends on commensurability of the perturbation with the unperturbed periodic cell' [9]. In other words, it requires a supercell 'having $\vec{q}$ as a reciprocal lattice vector' [10] which leads to large supercells for small $\vec{q}$. While big supercells are a drawback of the approach, the inclusion of anharmonic effects may be beneficial.

The second method is density functional perturbation theory (DFPT): A linear response approach which does not cover anharmonic effects but enables the calculation of phonons at arbitrary q-points with approximately constant computational costs. Briefly summarised, DFPT offers a framework based on DFT to calculate the derivatives of the energy by applying perturbation theory to the standard DFT equations. Baroni and coworkers [10] and Gonze and Lee [9] explain the method in detail deducing all the formulas to a set of equations for the perturbed system analogous to the Kohn-Sham equations of DFT. Their solution is again found self-consistently from which a dynamical matrix can be derived for each q-point. The full phonon dispersion curve is then calculated via the interatomic force constants in real space computed by a discrete Fourier transform from the dynamical matrices (see [9] for the particulars).

A phonon calculation for one q-point covering all atomic displacements requires computational time in the order of $3N$ scf-cycles where N is the number of atoms per unit cell. However, no supercells are needed. A valuable side result is the set of eigenvectors of the atomic motion associated with every phonon, which is experimentally not accessible but may permit further insight into the dynamics of the structure.

In 2003, Lazzeri and Mauri [11] presented an efficient way to calculate Raman intensities in periodic systems. It is designed as an add-on to DFPT and calculates the second-order derivatives of the electronic density matrix with respect to an applied electric field. In combination with DFPT yielding the phonon frequencies the Raman intensities can be used to simulate Raman spectra. This feature is not commonly available (i.e. offered by many programs) but may lead to a better interpretation of experimental Raman spectra or even the prediction of those.

2.2. Tools for calculations and visualisation

Throughout this work several programs were used for the various tasks. The general-purpose utility programs such as Python for implementation of EVO, Grace and Gnu-

plot for data plotting, Inkscape for illustrations or LaTeX are not covered here. In the following, we shall rather describe the special programs for calculations or visualisation and introduce their peculiarities. The purpose is to give an overview of the possibilities and limitations - in particular, when dealing with the calculational tools. Most of the calculations within this work have been done using DFT as implemented in QuantumESPRESSO which is not only freely available but provides many features extremely useful for determination of stability and properties of materials. The main visualisation software in use was XCrySDen as an easy-to-use but nevertheless powerful graphical interface to visualise the results of the calculations. Furthermore, WIEN2k implementing DFT and GULP applying classical force fields have been used for calculations and VESTA for crystal structure display.

2.2.1. QuantumESPRESSO

QuantumESPRESSO originates from PWscf and has been developed for about 20 years. The new name contains the acronym ESPRESSO meaning opEn-Source Package for Research in Electronic Structure, Simulation, and Optimization. It evolved and has been continuously maintained over the years as a free software package and is now a widely used GNU GPL licensed program package for materials research. The program has been intensively described in [12] and the contained references.

QuantumESPRESSO implements DFT in the framework of plane waves (PWs) and pseudopotentials (PPs) and uses 3-dimensional periodic boundary conditions. However, it is not restricted to infinite systems by boundary conditions since finite systems such as single molecules, cluster, layers, or surfaces can be approximated by large cells including a vacuum in the appropriate directions. Of course, the vacuum must be sufficient to eliminate interactions between the adjacent 'images' but apart from that there are no further drawbacks.

The application of plane waves to DFT for infinite crystals is descriptive considering that they map the periodicity of the structure[2]. Yet, plane waves require rather smooth potentials and thus are inappropriate for describing the rapidly divergent potential at the atomic core. Pseudopotentials solve this problem by providing a description of the ion-electron interaction that separates the core states of the ions from the valence electrons. The valence electrons are then treated with the plane-wave approach. Nowadays, many pseudopotentials are available for QuantumESPRESSO featuring traditional norm-conserving ones, ultrasoft PPs and projector augmented wave (PAW) data sets along with various exchange-correlation functionals. For more information see [12] or the website [13].

[2]For derivation see again [1] who also argues that the method may as well be effective for finite systems.

Results obtained by QUANTUMESPRESSO besides the ground state energy are electronic band structures and (projected) densities of states (DOS) as well as data for the visualisation of charge densities or the electron localisation function, among others.

Apart from the standard calculations, QUANTUMESPRESSO offers many valuable tools that allow for simulations of properties and judging the stability of structures. One of most important for this work was the implementation of a variable-cell relaxation to minimise forces and stresses (see section 4.3.4). It includes the possibility to apply constraints (e.g. keeping the volume fixed). This provides an essential and highly efficient way to locally optimise crystal structures. Another non-standard feature is the implementation of DFPT enabling the calculation of phonons at arbitrary points in the reciprocal space to full phonon dispersion curves. Quite unique is the ability to simulate Raman cross sections [11].

Furthermore, the code runs on many architectures and offers a high level of parallelisation using MPI as well as openMP. This is extremely valuable when working on systems of different sizes and complexities using varying computational resources. Moreover, most parts of the code are well documented and the input format is useable for different kinds of calculations. Also, many postprocessing routines for analysis of the results are available with the program.

2.2.2. WIEN2k

For the verification of the results of section 6.1 a full-potential implementation of DFT was used. WIEN2K [14, 15] applies the linearised augmented plane-wave (LAPW) method which splits the unit cell into atomic spheres and an interstitial region. The atomic spheres are represented by spherical harmonics scaled with radial functions while the interstitial region is treated by a plane-wave approach[3].

The program is designed for 3-dimensional periodic boundary conditions and strongly features the use of crystal symmetries when available. Thus, it also includes some tools for symmetry recognition of crystal structures. For the calculation of electronic properties WIEN2K also allows for the determination of the character of the bands in a band structure (e.g. s or p character). Furthermore, it exhibits some strengths in the calculation of optical spectra. To calculate phonons, PHONOPY [16, 17] is recommended which makes use of WIEN2K for the force calculations within the frozen phonon method.

WIEN2K is not freely available to every user, however the fee for the use in academic institutions is quite low. Hence, it is used by many researchers as an accurate program for electronic structure calculations.

[3] There are, of course, many publications dealing with the method but we will refer here to [1] and the references therein.

2.2.3. GULP

Despite the increasing computational power and advanced parallelisation schemes, DFT calculations stay expensive in terms of CPU time, especially for big systems and/or when it is necessary to do a vast number of calculations. Using interatomic potentials that empirically describe the interaction between the atoms in a structure most calculations can be done analytically. The potentials are generated by the fitting of models to (experimental) data available for the corresponding system and are thus only valid for the given atomic interactions[4].

Material simulations using force fields provide a fast way of dealing with large structures. GULP [18, 19] implements the approach in a program free for academic use. It strives to avoid numerical solutions by utilising e.g. analytical derivatives as far as possible. This leads to very fast computations. For example, the optimisation of the structure of crossed graphene (see section 6.1) at constant pressure along with a phonon DOS calculation took about two seconds. (The structure optimisation alone with QUANTUMESPRESSO needed about 16 minutes CPU time.) Moreover, the calculation of most properties such as phonons or elastic constants just requires some keyword to be set and is afterwards done automatically in only a fraction of the time compared to DFT calculations. Due to this, it is of course tempting to rely on the results obtained so easily. Unfortunately, predictions using GULP do not in every case yield results matching that of DFT which is why the influence of an inappropriate potential can never be ignored.

2.2.4. XCrySDen

Visualisation of structures and properties has become an essential tool in theoretical materials research which enhances understanding of results as well as illustrates important issues. Especially in the field of crystal structures one aims at easily displaying structures but also has to ensure that results in close relation to the crystal structure (for example, charge densities) can be depicted in the same image. XCRYSDEN which stands for X-Window Crystal Structures and Densities meets that purpose in a very easy to use and intuitive way. It was developed by Anton Kokalj [20, 21, 22] and visualises extended structures in terms of unit cells. It also works for finite systems.

XCRYSDEN comes with its own file format describing crystal structures. Conveniently, it supports input and output file formats of different programs for electronic structure calculation. Built-in support is available for QUANTUMESPRESSO as well as WIEN2K, among others and for many more structural display is possible with the aid of converting routines. Thus, one just needs the output file of a calculation to easily

[4] In contrast to pseudopotentials, for example, the empirical potentials are not transferrable. I.e. empirical data derived for silicon in SiO_2 must not be used for calculation of silicon nitride since they only describe the interaction of silicon and oxygen.

interpret the results. The GUI (Graphical User Interface) enables the manipulation of the images by interactively determining the perspective of the structure, changing the number of cells to be shown or modifying various other display parameters such as colours or radii of atoms before generating an image file. The user can therefore tune the output until it fits the needs. It is also a flexible tool for measuring bond lengths or angles in a structure.

A major reason for the extensive use of XCrySDen throughout this work is the excellent integration with QuantumESPRESSO. The when user may not only visualise crystal structures but also make animations of the progress of a relaxation or a molecular dynamics run without delay from the output files. There is also the possibility to show forces acting on atoms as well as 3-dimensional data grids for the display of vector fields, charge densities or ELFs. The appropriate file format for the latter is directly produced by the postprocessing routines of QuantumESPRESSO.

When doing calculations with QuantumESPRESSO XCrySDen is the visualisation tool of choice. It is GNU GPL licensed, easy to use in general and particularly convenient combined with QuantumESPRESSO.

2.2.5. VESTA

A quite new and more sophisticated program for visualisation of crystal structures is VESTA, Visualisation for Electronic and STructural Analysis [23]. It is capable of displaying structures accepting numerous file formats though not the one of QuantumESPRESSO. VESTA is also suitable for making quick and easy crystal structure images and is quite strong when dealing with symmetries: For example, importing a cif-file (which is the standard file format for the description of crystal structures) the user is free to choose between different settings of the same space group. VESTA also offers many possibilities to manipulate the graphical output.

Further valuable options are the detection and display of polyhedra in a structure, crystal morphologies as well as the representation of volumetric data from the output of various programs. This list is far from being complete (see [24]) since VESTA is a powerful visualisation program and free for non-commercial users as well. However, the usage requires some practice and massive consultation of the extended documentation.

3. Crystal structure search

3.1. Struggle for the minimum

'One of the continuing scandals in the physical science is that it remains in general impossible to predict the structure of even the simplest crystalline solids from a knowledge of their chemical composition.' [25]

Maddox' famous quotation can often be found when reading about the prediction of crystal structures and still holds for many materials. In the meantime quite a number of methods have been proposed and successfully used for the prediction of crystal structures. Despite the increasing computational power, highly parallelised electronic structure codes and sophisticated search strategies it is by no means common to predict crystal structures from first principles. But 'in general' it is possible to do so.

This chapter gives an overview of the different methods used for crystal structure prediction. Also it explains their characteristics and shows some applications. Before starting with the actual methods some general remarks on optimisation and multidimensional search spaces with non-trivial landscapes are made. Without loss of generality this work reduces all optimisation problems to a minimisation of the objective function.

3.1.1. Energy landscape in crystal structure prediction

In crystal structure prediction one aims at finding stable structures that are defined as being minima of the free energy/enthalpy landscape. The overall – global – minimum at given temperature and pressure is then the most stable structure whereas the other/local minima are called metastable ones. The (objective) function to be minimised is the free energy F

$$F = U + PV - TS$$

where U is the inner energy, P the pressure, V the volume, T the temperature and S the entropy of the crystal. While the other terms are well defined, the determination of the entropy is far from trivial. One possibility is to apply the quasi-harmonic approximation [26] where the entropy enters due to the vibrational degrees of freedom. The temperature influence is then reduced to the thermal expansion of the crystal emerging

in the phonon frequency dependence on the crystal volume. Technically, one has to compute the phonon spectra as a function of the volume of the crystal and can then get the free energy as a function of volume and temperature. Since the accurate calculation of phonon spectra using DFT is computationally very expensive, the enthalpy (free energy at 0 K) is often used as the objective function which permits the expected great number of necessary evaluations in the first place. Yet, some approaches that sample the energy landscape include entropic effects via molecular dynamics.

The energy landscape can formally be described as a function that maps the parameter space of the system to the real space, $f(\vec{x})$, $f : \mathbb{R}^n \to \mathbb{R}$. In crystal structure prediction the $\vec{x}$ has $3N + 3$ components[1] where N is the number of atoms in the unit cell. Thus, the search space for a system containing 30 atoms in the unit cell has 93 dimensions contributing to the challenges for global search.

3.2. Local optimisation — conventional search strategies

Naturally, there are lots of optimisation techniques that find the nearest local minimum with respect to a starting point in a landscape which will here be referred to as conventional search strategies. Most of the methods are widely used for local optimisation, easy to implement and efficient for a variety of problems. Some prominent examples are the conjugate gradient (CG) [27] or the Broyden-Fletcher-Goldfarb-Shanno (BFGS) [28, 29, 30, 31] methods which both use derivatives to find the stationary point of a function. Of course, there are many more techniques that allow for local optimisation. An overview can be found in [32], for example.

In materials research these algorithms are of particular use for the relaxation of structures where one aims at minimising the forces on the atoms and/or stresses on the unit cell. They have been developed to reliably find the local minimum structure in the zone of attraction where the starting point resides.

3.3. Global optimisation

For crystal structure prediction finding the global minimum is essential for describing the system but metastable structures are also of interest. Other problems/examples for global optimisation are:

- Travelling salesman problem, design of microchips
- Worst case analysis

[1] These components are composed of the six parameters that describe the unit cell (three lengths and three angles) and $N - 1$ atoms that each need three coordinates.

- Protein structure prediction

As one of the earliest methods for global optimisation the Monte Carlo method [33] has been introduced. It is a purely stastistical approach that samples the landscape by evaluating the objective function at random points. The greatest advantage of the method is that its probability of finding the global minimum is independent of the shape of the landscape and scales directly with the number of evaluations of the objective function. However, though easy to implement[2] the Monte Carlo method is often too expensive regarding the number of objective function evaluations.

This section shall deal with the many methods used and developed for structure prediction in the fields of clusters, molecular crystals, surfaces and crystalline solids. In principle, the methods are transferrable to the other subjects but may require changes in the implementation and/or constraints used in the search. Yet, most of the methods were first applied to and developed for some special purpose in crystal structure prediction.

All approaches aim at efficiently sampling the search space since the evaluation of the energy of the system using DFT is by far the most expensive part of the calculation. This problem is reduced when easily computable potentials can be applied to determine the quality of a candidate structure. However, the description of increasingly complicated systems drives the pursuit of even better and more effective search strategies.

Several methods that were developed and successfully used shall be explained in the following sections. Most of the techniques mentioned here have been reviewed in [34] often by the first researchers successfully using the presented method for crystal structure prediction. It also collects many references to the original publications and a great number of applications valuable for further reading.

3.3.1. Random search

Basically, what is termed as 'random search' in crystal structure prediction is just a variant of the Monte Carlo method combined with the local optimisation (relaxation) of each candidate structure. For this, every candidate structure is generated randomly though they also have to obey some constraints (e.g. minimal atom distances). It has to be mentioned that some constraints are used in almost all search strategies since they facilitate the relaxation and energy/enthalpy calculation thus allowing to effectively explore the local minima of the energy landscape. The procedure consists of the following steps [34, 35]:

1. create random structure

[2]Neglecting the difficulties of generating true random numbers.

2. optimise locally using standard relaxation techniques

3. check for convergence

The constraints on the structures to be relaxed are rather general: There can be minimum and maximum values for the lattice parameters, minimal nearest neighbour distances for the atoms and/or restrictions on volume or density of the cell. Options to specify symmetries or structural features based on experimental information or chemical intuition can be included as well. Since there are infinitely many possible representations of a unit cell, transferring them to a normalised form for comparability [34, p. 59] enhances the success of the search by avoiding redundancies. Generation and relaxation of the random structures is easily parallelised since the individual trials do not depend on each other. For eventual comparison just a list of structures has to be shared.

To terminate the search, it is possible to limit the maximum number of structures to be generated. Also one could proceed until no further improvement can be detected or the best solution is found repeatedly [34].

As one of the greatest advantages of random search Tipton and Hennig argue in [34] that little programmer time is needed to implement such an algorithm. However, they also estimate that it is only reliable for binary systems and in other small search spaces. Nevertheless, Pickard and Needs used their Ab Initio Random Structure Searching (AIRSS) successfully for prediction in many different systems as high pressure phases of hydrogen [36] and nitrogen [37]. Other researchers dealt with Li-Be alloys [38]. For further applications refer to [35].

3.3.2. Simulated annealing

Inspired by technical annealing where slowly cooling down may lead from unordered melt to an ordered structure and possibly the global energy minimum, Kirkpatrick, Gelatt and Vecchi [39] proposed a simulated annealing strategy for various optimisation problems. In 1990, Pannetier and coworkers [40] first applied simulated annealing to crystal structure prediction.

More thoroughly, the method has been described by Schön and Jansen [41, 42, 34]. In [42] the main procedure of simulated annealing for crystal structure prediction is introduced and summarised as a 'weighted random walk through space'. From a randomly initalised structure the algorithm moves to a new configuration according to a specified moveclass (see below). If this yields a lower energy, the move is accepted and the next move starts from the new configuration. If a higher energy is reached, the move is accepted with probability $e^{-\frac{E_{new}-E_{old}}{k_B T}}$ (k_B is the Boltzmann constant and T the temperature.) and otherwise rejected. In the latter case, the old configuration is reused

as a starting point. During the search, the temperature is reduced thus mimicking an annealing process and the system is expected to end up in a low lying minimum.

Further developments in the method have been reviewed in the according chapter of [34]. As choices for the moveclass for crystals the displacement or exchange of atoms or groups of atoms and the change of the unit cell form are proposed [41, 42]. Also a change of the composition of the cell is possible. Performing the steps can also be done via discrete step walkers (cf. section 3.3.3) or via a constant temperature molecular dynamics (MD) run which exhibits similarities to minima hopping (cf. section 3.3.4) [34]. In the last case, the temperature reduction would be done by lowering the kinetic energy of the atoms in the MD step.

Other peculiarities of simulated annealing can also be tuned to make the search more successful or faster: A temperature schedule can be chosen to be cyclic or adaptive, to reach a certain step acceptance ratio. Different acceptance criteria or a taboo search that excludes already visited minima may also be helpful.

Simulated annealing has up to now found many applications and inspired similar approaches (cf. next two sections) which is also due to the fact, that simple versions are easy to implement. A modification is conformation-family Monte Carlo [43] where structures are not dealt with individually but as families. A family then consists of different representations of the same structure.

The usage of simulated annealing ranges from clusters, polymers and proteins (for references see [34]) to the prediction of many solid state phases, e.g. the prediction of NaN_3 [44] or various phases of boron nitride [45].

3.3.3. Basin hopping

Introduced by Wales and Doye, basin hopping [46] is a method that searches for local minima as well as transition states and pathways between them. The name of the method describes the principle quite well: It hops from one basin to another. Naturally, it is not trivial to reliably find an adjacent minimum and thus step by step explore the energy landscape. After randomly generating or suggesting a starting structure and, if necessary, relaxing it, the basin hopping run contains the following steps.

1. propose move from starting point resulting in new structure
2. optimise locally using standard relaxation techniques
3. a) accept if $\begin{cases} E_{new} < E_{old} & \text{or} \\ E_{new} > E_{old} \wedge \mathrm{e}^{\frac{E_{old}-E_{new}}{k_B T}} > x & x \in [0,1] \end{cases}$

   b) reject otherwise
4. start with new (3a)/old (3b) structure from 1.

It has to be mentioned that the acceptance criterion (3a) in the case of $E_{new} > E_{old}$ is exactly the same as the one in the previous section on simulated annealing. However, we used the formulations as presented in [34].

Displacements from the starting point in basin hopping are realised by choosing a random value the uniform distribution $[-max_d, max_d]$ where max_d denotes a maximal displacement parameter. max_d can also dynamically be adjusted to obtain a certain acceptance ratio of the new structures. Thus, the parameter is raised if too many trial structures are rejected, and reduced otherwise. This enables the algorithm to escape extended minima as well as sample a search space with many small basins. Other variants include reseeding (start with a new random structure), a taboo list [47], imposing symmetries on the structures, or a combination of them [34, p. 39].

A crucial parameter of the search is the temperature T in the acceptance criterion introducing thermodynamical concepts. In [48] it is stated that an inappropriately chosen value can prevent the search from finding the global minimum.

Wales and Doye report on several findings gained using basin hopping. A lot of references can be found in the corresponding chapter of [34], some applications are the structures of Lennard-Jones clusters [46, 49] or transition metal clusters [50]. In combination with NMR-shifts basin hopping was also used for protein structure prediction [51].

3.3.4. Minima hopping

By name, minima and basin hopping seem to be very similar. Indeed, minima hopping as proposed and thouroughly described by Goedecker [48] shares the basic principles: moving to create new structure and local optimisation followed by an acceptance test. Taking a detailled look, reveals greater differences included by Goedecker to account for problems observed in other methods. For example, the revisiting of already known basins was identified to hinder or delay convergence. A possible solution for this is flooding the minimum, e.g. lifting the potential to penalise a revisit. Due to the unknown form of the energy landscape one does not know 'where' to lift. Also, this would permit crossings of a basin thus eventually blocking parts of the energy landscape which is why another strategy was chosen. The next paragraph shall deal with the details of the algorithm before presenting some applications.

Starting with a 'current minimum', e.g. some relaxed structure, a short molecular dynamics (MD) run is used to find a new structure. The MD is stopped as soon as the n^{th} minimum along the trajectory is reached which is then locally optimised to find the nearest local minimum M. If this equals an already known one, a new MD ist started from the 'current minimum'. Otherwise, M is accepted if its energy is less than a given E_{diff} higher than the current minimum. M then becomes the new 'current minimum' and the procedure starts over again.

Dynamically adjusted parameters enable the search process to adapt to certain regions of the energy landscape. E_{diff} as well as the kinetic energy of the atoms in the MD run are adjusted to yield a defined acceptance ratio of the found minima. This enhances the possibilities to escape minima with a large zone of attraction as well as thouroughly sampling the search space. The fact that minima hopping introduces a history to the search by comparing found solutions to the already known ones removes redundancies. Goedecker [48] also states that the behaviours of basin and minima hopping without history are quite similar.

To date minima hopping has found a lot of applications, though in the beginning the focus lay on the search for clusters, for example silicon [52, 53] and gold clusters [54]. Moreover, minima hopping has been applied to protein folding and structure prediction [55] and prediction of periodic solids, e.g. cold compressed graphite [56].

3.3.5. Evolutionary algorithms

Evolutionary algorithms for crystal structure prediction complement the many methods mentioned here and for completeness a brief introduction on the concepts is given. The following chapter will deal exclusively and extensively with evolutionary algorithms for crystal structure prediction.

As the name indicates evolutionary algorithms use the features of natural evolution to sample the search space. The main terms that are ubiquitous in natural evolution are also present in the optimisation method. An individual $I := \vec{x} \in \mathbb{R}^n$ consists of all the variables that specify a point in the search space (unit cell and atomic positions) and the value of the objective function, mostly the energy/enthalpy of the system. A population P then designates a group of μ individuals, $P := I_1, \ldots I_\mu$.

Recombination and mutation as the concepts that introduce diversity in populations are as important in optimisation as they are in natural evolution. In recombination two (or more) individuals are combined to form offspring which can lead to large jumps in the search space. Small steps can be obtained by incorporating mutation in the reproduction process which slightly alters the variables of a single individual.

As in the above methods the evaluation of the structures is done with some electronic structure program using DFT or an approach using potentials. Often a local optimisation through structure relaxation is used to speed up the search.

3.3.6. Other approaches

Apart from the above mentioned, in the following three further methods shall be introduced briefly: metadynamics, data mining and particle swarm optimisation.

Metadynamics Originally, metadynamics [57, 58] was developed to search for the microscopic mechanisms taking place in phase transistions but it was found to also have predictive power. It emerged from variable-cell MD [59] and bypasses the problem, that the MD timescale is too small to simulate phase transistions. Metadynamics decouples the 'slow degrees of freedom' of a phase transition, e.g. the form of unit cell (termed order parameter) from the fast ones (atomic positions) [57]. The order parameter is treated by metadynamics while the atom positions can be covered with molecular dynamics or Monte Carlo simulations. Variation of the order parameter is done by using the stress tensor that needs to be provided by the MD code and includes 'history' to push the system out of the local minimum [34]. The external MD code is also used to equilibriate the atomic positions.

A great advantage of metadynamics is the inclusion of entropic effects via MD. Thus it is also able to find phases stabilised by entropy which are omitted by other approaches which mostly take only enthalpy into account. It is crucial to provide a suitable starting structure and to make an informed choice regarding the order parameter [60].

Data mining Extracting rules from existing data and using them for prediction is the aim of data mining. The application of data mining to crystal structure prediction was introduced by Ceder and coworkers in 2003 [61]. Their method is called data mining of quantum calculations (DMQC) and uses an empirical observation in binary alloys: 'The energies of different crystal structures are strongly correlated between different chemical systems.' To apply data mining it is necessary to have a library of energies of different relaxed crystal structures. Finding new structures is then done by finding a linear combination of known data which allows to identify approximate dependencies and thus most probable ground states. The energies of those predictions are calculated with DFT and afterwards added to the library before searching again.

In [61, 62] it was shown that the usage of data mining for crystal structure prediction works well in systems of binary alloys. Yet, a structure type must already be present in the library otherwise it cannot be found.

Particle swarm optimisation Another method for global optimisation was first proposed by Kennedy and Eberhart for optimisation of nonlinear functions [63] and was inspired by flying of bird flocks or fish schooling. The behaviour of an individual is influenced by the best local or global one and can be regarded as a modification of flying speed and direction.

CALYPSO (Crystal structure AnaLYsis by Particle Swarm Optimisation) was introduced in 2010 [64] and applies this approach to crystal structure prediction. Starting with locally optimised randomly generated structures a first local minimum is identified. The current local/global minimum along with the history of the single individual influences the direction in which the present individual is altered to generate a new one

[65]. The random structures are created from the 230 spacegroups and may therefore obey some symmetry constraints. A filter for similar individuals using bond length deviations is also implemented thus enhancing diversity of the flock.

Despite being a young method in crystal structure prediction particle swarm optimisation has recently proposed high pressure phases of lithium [66] and bismuth telluride [67].

4. Evolutionary algorithms

The concept of evolution in biology was first introduced by Charles Darwin and Alfred Wallace [68] in 1858. In their joint publication they focus on the natural selection as the driving force for evolution and provide several examples for their hypothesis. The scientific community was initially unwilling to accept the new theory but it was gradually adopted during the following decades. At the beginning of the 20th century the Danish botanist and geneticist Wilhelm Johannsen established two terms to describe the disagreement between the genetic information contained in the cell nucleus and the observable characteristics of an individual: genotype and phenotype (figure 4.1). The genotype is encoded in the DNA (DeoxyriboNucleic Acid) that is located in the nucleus of each cell of an individual. The DNA was not discovered until the middle of the 20th century and finally manifested the concept of evolution. The phenotype is the actual appearance of the individual, the full amount of its observable characteristics.

While the DNA holds all information and with it the procedures that are needed to build the individual, the actual processes and pathways are quite complicated and would fill another work of this size. Figure 4.2 shall illustrate the two endpoints of the process: The DNA forming the chromosomes and the actual phenotype of a lion individual.

Evolutionary processes in nature work completely on the genotype of the species, directly on the DNA. This leads to genetic changes which eventually account for differences in the phenotypes. Those differences may lead to some kind of advantage over fellow individuals with respect to breeding, food supply or predators. This may enhance the chances of the individual to breed offspring with the same superior characteristics which are more likely to survive. In the long run (and assuming a non-changing environment), a certain genetic variety is chosen by natural selection. Thus, natural evolution can be considered as an optimisation process.

After giving a short overview of the early branches of evolutionary algorithms this chapter focuses on the methodology of evolution strategies. It gives details on each step of the algorithm, including the realisation in other implementations as well as in EVO – the code for evolutionary structure prediction devoloped in this work.

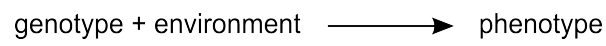


Fig. 4.1.: Relation between genotype and phenotype of an individual.

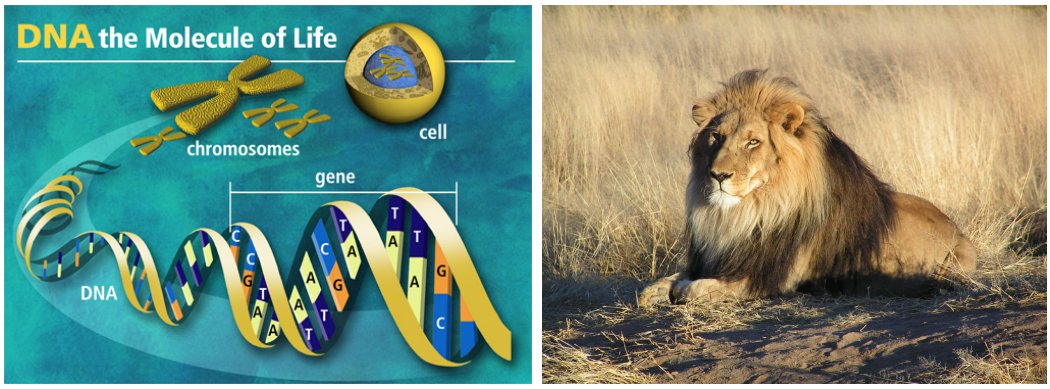


Fig. 4.2.: Illustration of the genetic information contained in the DNA located in the nucleus of each cell (picture from [69]) and the whole individual (picture from [70]).

4.1. Brief history of evolutionary algorithms

Natural evolution inspired the development of evolutionary algorithms which began in the 1960s and 1970s. In search for efficient optimisation techniques for highdimensional search spaces with several minima, Holland [71] and Rechenberg/Schwefel [72] developed genetic algorithms and evolution strategies, respectively. Though the two terms are used synonymously now or under the more generic term evolutionary algorithms, historically the two approaches differ fundamentally.

Genetic algorithms – close to nature – distinguish between a genotype and a phenotype (figure 4.3) of the individuals. In this context, the phenotype holds all the values that describe the individual as would be the form of the unit cell and the atomic positions. On the other hand, the genotype used for the operators of the genetic algorithm, namely recombination/crossover and mutation, is the binary representation of the phenotype. Main evolutionary operator that creates diversity within the population is the crossover (see figure 4.3). In the beginning, there were also different types of selection distinguishing between genetic algorithms and evolution strategies.

Nowadays, most evolutionary algorithms operate directly on the phenotype and should consequently be named as evolution strategies. This work will use the two term evolutionary algorithm and evolution strategies synonymously since we use a real

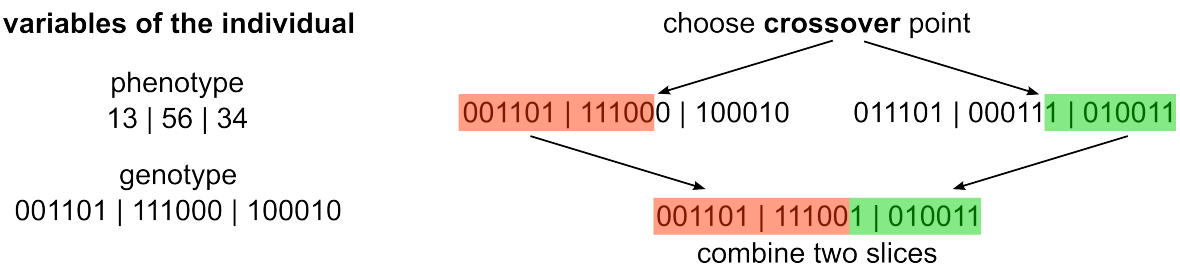


Fig. 4.3.: Representation of the variables and visualisation of a one-point crossover that realises recombination in genetic algorithms.

(phenotypic) representation of the variables of the individuals. It should be mentioned that there are a number of publications that nicely review evolutionary algorithms, point out similarities and differences and contain a lot of useful references. A mathematical/theoretical approach is chosen by Bäck [73] while Nissen [74] approaches the subject in a more application-oriented way.

4.2. Evolutionary algorithms for crystal structure prediction

Most evolutionary algorithms follow the same methodology and share the use of the phenotypical representation (i.e. direct use of the variables, without binary encoding). Unless stated otherwise, this assumption holds for the following sections. The description of the main steps (figure 4.4) of the algorithm will focus on their implementation in EVO [75].

This will be followed by a discussion of the peculiarities of the several other programs available, published or at least used for crystal structure prediction. Unfortunately, only few programs used nowadays are well documented in literature. Two of them should be highlighted here: USPEX [76, 77], first published by Glass, Oganov and Hansen in 2006, has till then been extensively used and also developed further. Especially, the cooperation with high pressure crystallography brought USPEX some publicity, see [78, 79] among others. However, the program code is not freely available to the community, it can be only obtained by contacting the developers. In contrast, XtalOpt published in 2011 [80] is like EVO under the GNU General Public License and thus can be freely used, shared and modified. Apart from the two above mentioned, there are several other evolutionary codes for crystal structure prediction which will be reviewed section 4.5.

4.3. Methodology as implemented in EVO

4.3.1. Individual and population

In the style of natural evolution, early researches on evolutionary algorithms adopted the terms individual and population. In the case of crystal structure prediction, the individual I contains all the information necessary to describe a crystal structure and is characterised by

$$I = (cell, atoms, age, fitness)$$

In this definition, *cell* denotes the form of the unit cell, namely the lengths of the axes and the angles between them which inherently determines the volume. *atoms*

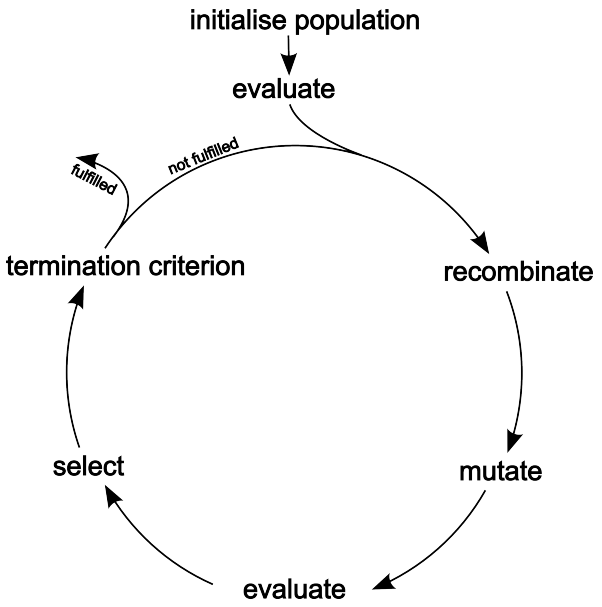


Fig. 4.4.: Schematic overview of the process steps of EVO. Most other evolutionary algorithms follow the same or a very similar flow. One passage through the cycle corresponds to a generation.

describes the positions of the atoms together with the respective atomic species and the *fitness* value is the calculated enthalpy of the structure. Since the fitness value in the algorithm is independent of its computation, there is no general limitation – one could also optimise for energy, bulk modulus or other properties that can be determined computationally from the crystal structure.

Furthermore, EVO assigns an integer *age* parameter to each individual which is a maximal life span [72] thus combining the two traditional selection strategies (see section 4.3.7 and [73, p. 78]). The value defines the number of generations, an individual is allowed to reproduce and pass on its structural features. If it outlives this life span, it is fated to die.

A population P is a group of individuals that can mate and reproduce:

$$P = (I_1, \ldots, I_\mu)$$

The number of individuals in the population μ is an important parameter of the evolution strategy since it (among others) determines the diversity within the structure search.

4.3.2. Initialisation

For the starting generation a set of crystal structures is randomly generated. In principle, it is possible that arbitrarily small or large unit cells are initialised in which the atoms are too near or too far away to form a solid. To avoid unphysical structures EVO applies the following constraints (partly adapted from [81]):

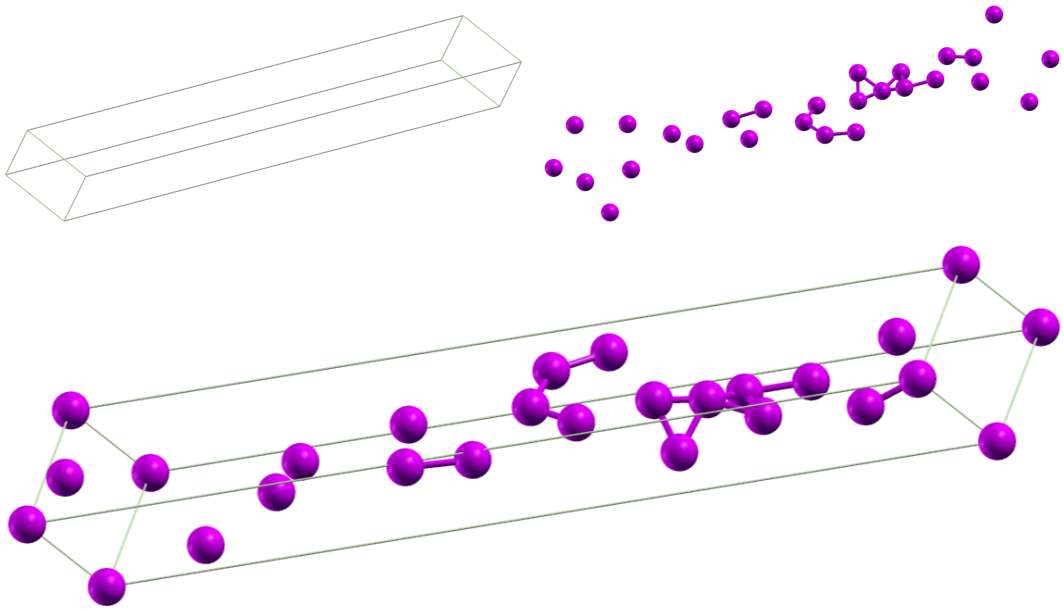


Fig. 4.5.: Randomly created crystal structure for the first generation. Cell and atoms separately, and whole structure below.

- The length of the axes of the unit cell are limited according to the atoms it should hold: The smallest possible value is the diameter of the largest atom present in the structure. The sum of diameters of all atoms defines the upper limit of the lengths.

- The angles between the unit cell axes can vary between 45° and 135° by which the whole space of possible unit cell forms can be reached.

- There also exist limitations on the volume: The densest phase possible in nature is the close packed structure which thus represents the lower limit on the cell volume. The default upper limit is the quadruple of it. Of course, the user may also specify the limits of the allowed volume manually.

The constraints above allow to generate all kinds of unit cells, e.g. extremely flat ones. Since there is an infinite number of possible unit cell representations, we apply a cell transformation which will be explained in detail in the next section.

After generating and transforming, the unit cell is filled with atoms of the given kind(s). One by one, a position is determined randomly, tested against too short bond lengths and then included in the structure. A bond length is defined to be too short if the distance between two atoms falls below 80% of the sum of the covalent radii of the two corresponding atoms. Figure 4.5 shows an individual with 20 atoms in the unit cell.

function SEARCHTRANSFORMATIONMATRIX(C)
  $M \leftarrow \mathbb{1}$
  repeat
    $M' \leftarrow$ DESCENTSTEP(M, C) ▷ Search neighbouring matrices
    $\delta \leftarrow \|M - M'\|$ ▷ New matrix differs?
    $M \leftarrow M'$ ▷ Update matrix
  until $\delta = 0$ ▷ Until no further progress
  $n_1, n_2, n_3 \leftarrow \|MCe_1\|, \|MCe_2\|, \|MCe_3\|$ ▷ Norms of transformed cell vectors
  $i_1, i_2, i_3 \leftarrow$ SORTINDICES(n_1, n_2, n_3) ▷ Indices for descending sort order
  $R \leftarrow [e_{i_1} e_{i_2} e_{i_3}]$ ▷ Permutation matrix
  return RM ▷ Combined tranformation and permutation
end function
function DESCENTSTEP(M, C)
  $M' \leftarrow M$
  for all $M'' \leftarrow M + \{-1, 0, +1\}^{3\times3}$ **do** ▷ Adjust all matrix elements by [-1,0,1]
    if COMPARE(M', M'', C) $\wedge \det M'' \neq 0$ **then** ▷ Found better matrix?
      $M' \leftarrow M''$
    end if
  end for
  return M'
end function
function COMPARE(M, M', C)
  $n_1, n_2, n_3 \leftarrow \|MCe_1\|, \|MCe_2\|, \|MCe_3\|$ ▷ Norms of transformed cell vectors
  $i_1, i_2, i_3 \leftarrow$ SORTINDICES(n_1, n_2, n_3) ▷ Indices for descending sort order
  $n'_1, n'_2, n'_3 \leftarrow \|M'Ce_1\|, \|M'Ce_2\|, \|M'Ce_3\|$ ▷ Norms of transformed cell vectors
  $i'_1, i'_2, i'_3 \leftarrow$ SORTINDICES(n'_1, n'_2, n'_3) ▷ Indices for descending sort order
  return $\forall k : n_{i_k} \leq n'_{i'_k} \wedge \exists k : n_{i_k} < n'_{i'_k}$ ▷ After sorting: None worse, one better?
end function
function SORTINDICES(n_1, n_2, n_3)
  return i_1, i_2, i_3 with $n_{i_1} \geq n_{i_2} \geq n_{i_3}$ ▷ Sorting indices of sequence
end function

Fig. 4.6.: Algorithm of the search for the transformation matrix. Only input is the cell vectors, here denoted by C.

4.3.3. Cell transformation

In crystallography, most conventional unit cells of higher symmetry have one or more angles that are 90°. The cell transformation implemented in EVO thus transforms or normalises the generated unit cell to an equivalent cell, i.e. a cell that describes exactly the same crystal structure, bringing the angles as close to 90° as possible. To do this,

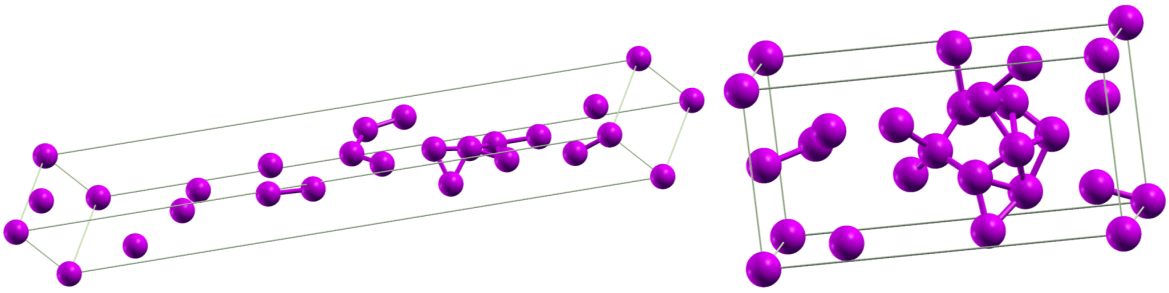


Fig. 4.7.: Crystal structure from Fig. 4.5 and transformed equivalent.

EVO first searches for the matrix that transforms the unit cell only regarding the unit cell vectors. This is done by iteratively scanning invertible matrices if they shorten the lengths of the cell vectors as an indicator for increasing rectangularity (figure 4.6). A permutation step, sorting the cell vectors according to their lengths is also included which enhances comparability of the structures.

The final matrix can then be used to transform unit cell vectors as well as atomic positions easily by matrix multiplication. Figure 4.7 shows starting and transformed unit cell.

4.3.4. Evaluation

Though there is no principal limitation on the choice of the fitness function for evolutionary crystal structure prediction, most implementations use the energy/enthalpy of the structure as calculated by various electronic structure codes. Furthermore, considering that a necessary condition for (meta)stability of a structure is its localisation in a (local) minimum of the energy landscape this choice seems natural.

EVO currently uses QUANTUMESPRESSO [12] and GULP [18] for the evaluation of the enthalpy of the individuals. Both programs feature a variable-cell relaxation which corresponds to a local optimisation. Thus, it is possible to easily explore the minima of the energy landscape in which zone of attraction the starting structure resides. In practice, this is realised by a local minimisation of the stresses on the cell and of the forces between the atoms. Woodley and Catlow [82] refer to this as a Lamarckian kind of evolution and show that it greatly enhances the speed of the search. When using EVO it is not mandatory but highly recommended to use relaxation techniques. In this case, the crystal structure of the individual is updated with the locally optimised structural parameters after evaluating the enthalpy.

The inclusion of other programs for the evaluation is quite straightforward and explained along with other technical details in the appendix.

4.3.5. Recombination and mutation

Creating individuals for a new generation is mainly done by mating two (or more) parents from the subsequent generation. The main evolutionary operator is called recombination and combines the features of the parents which is close to natural evolution. The recombination in evolutionary strategies works directly on the phenotype of the individuals. The recombination operator describes the 'breeding' of λ offspring from the μ individuals of the actual population. Though a ratio of $\frac{\mu}{\lambda} = \frac{1}{7}$ is suggested by the early researches [73], there is no general rule on the values of the population size and the number of offspring. For crystal structure prediction typical ratios are larger thus reducing the costly number of offspring evaluations.

At present, EVO implements purely algebraic and graphic crossover operators which will be explained here. For all recombination operators a predefined number of parents I^p (usually $n_p = 2$) is chosen randomly from the current population P. These are then used to build an offspring according to the operators below.

The two algebraic recombination mechanisms are inspired by the traditional operators used in evolutionary strategies - discrete and intermediate recombination [73, pp. 73]. When applying discrete recombination each variable of the offspring is determined by taking the same variable from one of the parents. Concrete variables x_i in this case are the single atomic positions and the vectors spanning the unit cell.

$$x'_i = x_i^p \qquad p \in I^p$$

For intermediate recombination the offspring is formed by the (weighted) averages of the corresponding variables of the parents.

$$x'_i = \frac{1}{n_p} \sum_{p \in I^p} x_i^p$$

Since the two mentioned algebraic recombination schemes are easy to understand but little descriptive, EVO also uses a more graphic way to combine two parents. For similar approaches see [83, 76, 80]. It is called crossover operator, which is derived from the crossover implemented in historical genetic algorithms (see figure 4.3). The two parent cells are cut along a random direction on the randomly chosen cutting point (taking at least 20% of each cell) and afterwards 'glued' together. The cell vectors are an accordingly weighted average of the parents vectors. Figure 4.8 shows two cut parent cells and the corresponding offspring cell. If the number of atoms then differs from the defined number for the run of EVO, it is adjusted by randomly adding or removing atoms until the right number of atoms is reached.

Mutation is traditionally an important part in evolutionary algorithms [73]. Following [84] it is realised by

$$x_{new} = x_{old} + N(0, \sigma)$$

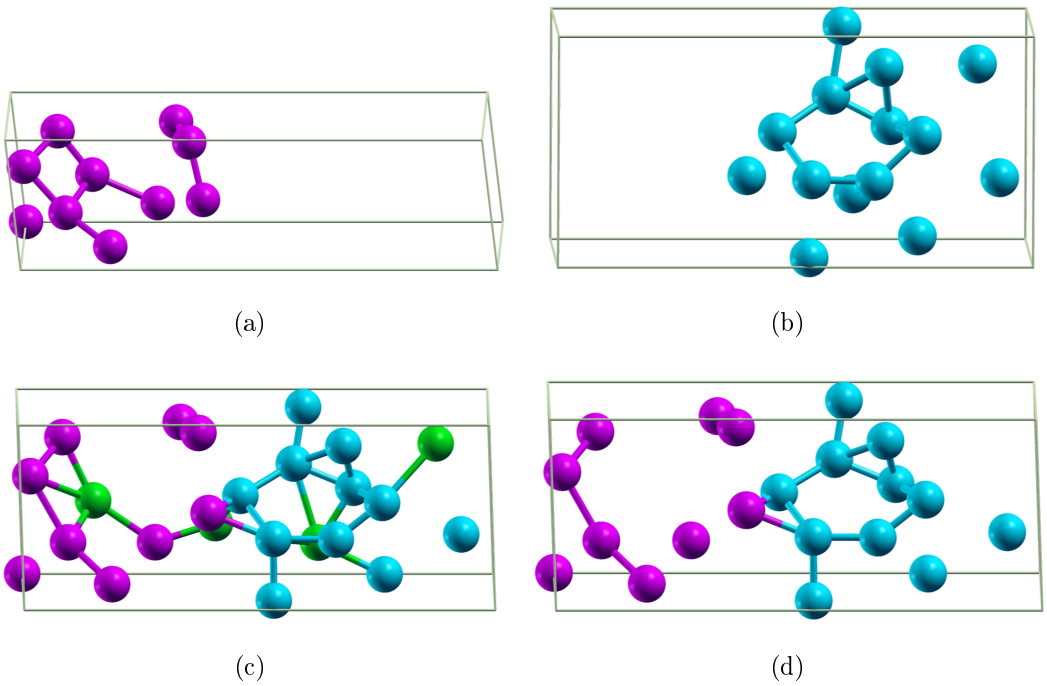


Fig. 4.8.: Two parent structures (a, b) and offspring generated by crossover. The intermediate structure (c) has too many atoms (20 were specified) in the cell. The green atoms have been chosen to be eliminated yielding the final offspring (d).

where x_{new} is the mutated variable, x_{old} the initial one and $N(0, \sigma)$ a normal distribution with expectation zero and standard deviation σ. Therefore, small changes of the variables are more probable than large ones resulting in a kind of local optimisation. Since relaxation techniques provide a very efficient way to locally optimise crystal structures, mutation becomes redundant in evolutionary crystal structure prediction. EVO offers the possibility to include mutation by specifying a non-zero σ. However, this is only recommended when relaxation is not applicable.

Every offspring structure undergoes the cell transformation and evaluation.

4.3.6. Similarity test

By using structure relaxation it is very likely to create very similar structures since all structures initially lying in the zone of attraction of the same minimum converge to the actual minimum. Those redundancies limit the search to few minima of the energy landscape. EVO thus enhances diversity by eliminating too similar structures from the population. Different methods for this can be found in literature: Oganov and Valle use a so-called fingerprint function [85, 86, 87] to compare structures which is build on a radial distribution function [88] mapping the atomic distances. Lonie and Zurek [89] transform the structure in a special way enabling a direct comparison of the structures. Both procedures will be explained in detail in the sections that deal with the particular

programs. For the sake of completeness, the bond characterisation matrix applied by Wang and collaborators [65] has to be mentioned, too.

The similarity test of EVO works with the atomic distances in a more intuitive way similar to [90]. A big supercell of the structure including the cartesian coordinates p_{cart} is built for every individual. Pairwise comparison of the structures is done as depicted in Fig. 4.9: The distances of the atoms of the supercell to the defined central atom (located at the point of origin) are calculated and sorted. *dist_ diff* then gives the absolute values of the atom distance differences the first n of which are compared to 20 % of the sum of the covalent radii (bond length) of the involved atoms. If the maximum is lower than the predefined threshold, the two compared structures are assumed to be similar and to lie in the zone of attraction of the same minimum of the energy landscape. The individual with the higher enthalpy value is then removed from the population.

function SIMILAR(a,b)
  for all $atom \in \{\mathrm{H}, \mathrm{He}, \mathrm{Li}, \ldots\}$ **do**
    $sorted_a \leftarrow$ SORTEDDISTANCES$(a, atom)$
    $sorted_b \leftarrow$ SORTEDDISTANCES$(b, atom)$
    $dist\_diff \leftarrow \|sorted_a - sorted_b\|_\infty$ ▷ Maximum of differences
    if $dist\_diff > 0.2 \times bond\_length$ **then** ▷ Compare maximum to threshold
      return False
    end if
  end for
  return True
end function
function SORTEDDISTANCES(struct,atom)
  $p \leftarrow$ POS$(struct, atom)$ ▷ Get positions of atom sort in supercell
  $dist \leftarrow \mathrm{diag}(p^T p)$ ▷ Compute vector of distances from central atom
  $sorted \leftarrow \mathrm{sort}(dist)$ ▷ Sort in ascending order
  return $sorted[0:n]$ ▷ Return first n elements of sorted vector
end function

Fig. 4.9.: Algorithm for the similarity comparison of two individuals a and b.

4.3.7. Selection

As selection is a major step to drive the search towards the minimum, many selection procedures for evolutionary algorithms can be found in literature. For an overview containing lots of further references see [73, pp. 163]. Genetic algorithms use traditionally the proportional, linear ranking, or tournament selection. Evolution strategies on

the other hand at first concentrated on the $(\mu + \lambda)$ and (μ, λ) selection. In the first, the parents for the next generation are chosen from the joint pool of parents and offspring by taking the μ fittest individuals. By this, the survival of the best individuals is guaranteed though risking to get trapped in a local minimum. The (μ, λ) selection otherwise chooses the parents solely from the offspring, avoiding to get caught in a minimum. However, favourable individuals have only one generation to pass over their structural features. As a consequence, EVO uses a mix of the two mentioned strategies giving an *age* to all of the individuals [72]. This maximal life span allows the fit individuals to reproduce over a defined number of generations.

EVO applies the *age* parameter supplementing it by some features specific to crystal structure prediction. The joint pool of parents and offspring are first scanned for too similiar structures (see section 4.3.6) and redundancies are removed. Those of the remaining individuals whose *age* parameter exceeds a given value are eliminated. By doing this, no (good) structure in a local minimum can dominate the population, since it is fated to die after reproducing for a certain number of generations. So, the search can propagate to new minima. (The information of the dying structures is not lost, since the best of them are stored for future analysis.) The *age* parameter of all other individuals is incremented by one.

Finally, the selection is done by choosing the μ fittest individuals of the pool, i.e. those with the lowest enthalpy, to be the parents for the next generation.

4.3.8. Termination criteria

In principle, an evolutionary algorithm can produce generation after generation since it is an infinite loop when ignoring termination criteria. Other evolutionary algorithms for crystal structure prediction rarely implement termination criteria apart from a limiting number of generations. EVO on the other hand utilises different dynamic criteria as well as a static maximum number of generations:

- If a defined number of generations has been calculated, the algorithm is terminated. This guarantees termination.

- All individuals newly created by recombination/mutation are tested for similarity (section 4.3.6). If the diversity of the offspring is too low, e.g. too few of them are dissimilar, EVO possibly was trapped in a minimum and is terminated.

- Diversity of population and list of the best individuals is also tested by fitness comparison. Fitness values that are nearly equal point towards location of all structures in equivalent minima.

- EVO also terminates if there has been no fitness improvement over the last five generations, assuming that the global minimum is reached or EVO is irrevocably stuck in a local minimum.

All those termination criteria enable EVO not only to run unattended but also to come to an end depending on the course of the search.

4.4. Search for layered structures

All of the procedures described in the last section apply to crystal structures in regular three dimensional periodic boundary conditions. A special feature of EVO is the ability to search for layered structures. Since those are also treated in 3D periodic boundary conditions, it requires no major changes of the code and enables the use of the described evaluation methods. Figure 4.10 shows an exemplary structure and unit cell of such a layer.

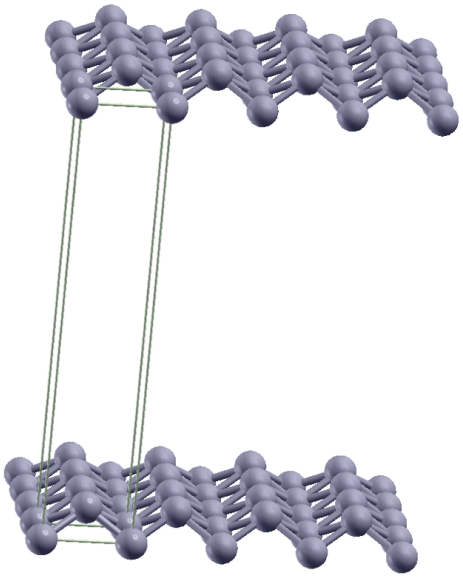


Fig. 4.10.: Sheet structure and illustration of unit cell setting. A large c-axis of the cell guarantees vanishing interaction between the layers preserving 3D boundary conditions.

In order to maintain the shown cell setting, some additional constraints are introduced:

- Per definition, the sheets lie in the a-b-plane of the unit cell which corresponds to a vacuum layer extended in c-direction. For this the constraint on the length of the c-axis of the cell is changed to a value of 1.2 times *layer distance* which is an option of EVO specifying the minimal distance between two adjacent layers.

- The atoms are constrained to a certain (small) part of the unit cell in c-direction. This *layer height* has to be chosen carefully regarding the *layer distance* to avoid interactions between the layers. By constraining sheet input like this, the layers neither have to be monatomic nor flat.

- The cell transformation (section 4.3.3) is not executed in the same way as above: The pure unit cell (without atoms) of the first generation will be transformed (regarding the constraints above) but afterwards we avoid the transformation to prevent the atoms from being moved out of the a-b-plane of the cell.

- Crossover recombination will only be conducted in a- or b-direction since in the c-direction no change of the layer would occur.

4.5. Other evolutionary algorithms for crystal structure prediction

With increasing computational power evolutionary structure prediction was blooming in the last decade. As early as 1995, Deaven and Ho [83] presented a genetic algorithm (binary encoding of the structures) to search for atomic clusters and successfully reproduced the structure of C_{60} (buckminsterfullerene). They also proposed the crossover operator which produces an offspring by cutting the clusters through their centers of mass and combining them. In the same year, Bush, Catlow and Battle [91] used also binary encoded individuals and classical operators from genetic algorithms (two-point crossover, mutation) to predict the structure of Li_3RuO_4. In this and also a later publication [92], they divided the search process with respect to fitness evaluation into two stages: For the genetic algorithm the fitness was evaluated with a fast cost function and then only the best individuals were locally optimised. Further development of the approach led to the insight that phenotype representation along with relaxation of all individuals (Lamarckian scheme) yields more reliable results and is more efficient [82].

Since 2006 more implementations have emerged: Abraham and Probert [93] proposed their algorithm including a periodical cut for crossover and a roulette wheel selection. In [94] they describe the inclusion of a similarity measure into the fitness value of each individual so that similar ones are penalised with respect to their fitness and their selection probability lowers. The evolution strategy proposed by Glass and Oganov [76] will be discussed in detail in the following. In 2006, Trimarchi and Zunger [81] also presented an implementation of the 'global space-group optimisation' problem and applied it to AuPd alloys. Their approach was complemented by an extension to search for arbitrary compositions of a compound [95] which is rated with its distance to the 'convex hull' which marks the formation energy of stable compounds with respect to

composition. By that, is was possible to find several structures in the Al-Sc system [95] and several Na_nN_2 structures [96].

Kolmogorov and coworkers developed MAISE (Module for Ab Initio Structure Evolution) and searched mainly for borides [97, 98]. So far, there is no publication describing the method and details of the implementation and only general information can be found [99]. Another evolutionary algorithm for crystal structure prediction has only been mentioned to be used in the research on high pressure phases of Eu [100]. The last algorithm considered here is XtalOpt which is extensively documented and will be explained in detail in the following.

4.5.1. USPEX

In 2006 Oganov and Glass published a first description of their 'Universal Structure Predictor: Evolutionary Xtallography' USPEX [76] followed by more details in [77]. Since then the algorithm has been steadily improved and new features were added. Particularly, the collaboration with high pressure experimentalists led to several fascinating findings such as a new transparent sodium phase [79], a high pressure boron structure [78], and a new polymorph of carbon [101], among others. In the following, only the methodical peculiarities of USPEX compared to EVO shall be discussed, instead of its applications.

The initialisation of the first generation is done randomly with only some minimal differences in the constraints [77] to the ones described in section 4.3.2. An additional feature is the volume rescaling in which every cell is rescaled to a user-defined volume that can change during the search. A cell transformation step [102] as well as split-cell initialisation [103] for large systems or the usage of symmetric structures in the first generation [104] has been implemented.

Creating individuals for the following generations is done via 'heredity' which is equivalent to recombination and different forms of mutation. More specifically, the crossover [77] is similar to the version described above (section 4.3.5), while the mutation works only on the lattice vectors by applying a strain matrix. The proposed permutation [77] exchanges the positions of two atoms of different types. More recently, a 'non-blind coordinate mutation' and the softmutation have been introduced [103]. The first is built on the 'local order' [103] of an atom, moving atoms with low order more than those with higher order. Softmutation moves the atoms collectively along the eigenvector of the softest phonon mode for which the dynamical matrix is approximated[1]. After creation, every new structure is rescaled to a volume specific for that generation before the evaluation including relaxation is carried out.

[1] *Ab-initio* determination of dynamical matrices is very expensive and hardly applicable for the many structures present in evolutionary crystal structure prediction. To avoid this, the dynamical matrix is calculated using bond hardness coefficients, see [105].

USPEX compares structures and groups the results of the search by using the crystal structure fingerprint. Each crystal structure is characterised by its fingerprint that is related to a radial distribution function of the atomic distances and reduces the multidimensional search space (i.e. energy landscape) to less dimensions. A detailed explanation and description can be found in [85, 86, 87]. By comparing the fingerprints of two structures their similarity is rated and redundancies can be removed (which is also referred to as 'niching'). The 'local order' mentioned in the previous paragraph is also based on fingerprints.

For the selection of the parents for the following generation, initial versions of USPEX just chose the best offspring to survive [76] while more recently a modified ranking selection was introduced [106]. When choosing the parents for crossover, a constraint is made on them to be not too different [103] which is expected to respect the 'funnel'[2] structure of the energy landscape. Ageing or death of the individuals is a fairly new feature [104] implemented by penalising the fitness of the structures.

Termination criteria are not documented in literature. As a rule of thumb, [76] mentioned that the ground state for a system with 20 atoms should be found in 20 generations though there were no information on the population size or the number of offspring.

Unfortunately, neither the code of USPEX is freely available nor are there articles describing the methodology in detail. It is distributed over a number of articles as can be seen from the extracted information above and still remains partial. Thus, it seems to be difficult to understand or reproduce the many successes in structure prediction obtained by USPEX.

4.5.2. XtalOpt

In 2011 Lonie and Zurek presented XtalOpt, another evolutionary algorithm for evolutionary structure prediction which has been released under the GNU Public License. The code is freely available via the CPC Program Library [107] and a detailed description of the method can be found in [80]. The following paragraphs outline the steps of the algorithm and refer to the mentioned article unless stated otherwise.

Individuals are initialised randomly, though seed structures may be specified for the first generation. All cells must obey some constraints to avoid unphysical unit cells that lead to problems during relaxation. A cell transformation/normalisation is also provided.

[2]The term 'funnel' was introduced by Goedecker [48]: 'We will denote as a super-basin the union of several neighboring basins. If one can arrive from any point in such a super-basin at the lowest minimum without crossing barriers that are very high compared to the average difference in energy between local minima, it will be called a funnel.' For a simplified illustration of a two-funnel landscape see [34, p. 154].

XtalOpt uses a broad set of evolutionary operators, that are divided into pure and hybrid ones. A crossover is done quite similar to the one described in section 4.3.5 in analogy to [83, 76] and strain can be applied to the lattice vectors deforming the unit cell [76]. 'Ripple' is a newly developed operator that sends cosine waves through the crystal thus changing the positions of the atoms along a predefined axis. The exchange is synonymous to the permutation used by Glass and Oganov [77] and swaps positions of different atomic species. Additionally, there are two hybrid operators: 'Stripple' combines strain and ripple and 'permustrain' creates an offspring by applying atom exchange and permutation.

The concept of generations is completely abandoned by XtalOpt [108] for the evaluation of individuals. This accounts for the fact that local optimisation (relaxation of the structures) may have very different durations for the distinct individuals. So, though the parallelisation of this step is straightforward, the further processing is delayed if one individual takes very long to relax. To avoid this, XtalOpt creates a new individual from the pool of continuously updated parents which allows for saturation of the computational resources[3].

Testing for too similar structures has been shown to enhance diversity and to significantly speed up the search for the global minimum. XtalOpt refers to that concept as 'niching' as well and at first used the three values of enthalpy, volume and spacegroup as a fingerprint to compare structures. Especially in populations containing spacegroups with nearly no symmetry, this may lead to false positive duplicates. Consequently, XtalComp [89] has been developed and added to the code. It implements a direct comparison of the crystal structures after transforming them to a defined form and applying rotations, translations, and reflections. Extensive tests are also documented.

Selecting individuals for the pool of parents is carried out using a modified proportionate selection strategy. The probability for an individual to be chosen as a parent is not proportional though dependent on its enthalpy.

XtalOpt lacks automatic termination and relies on the user to end the search. But an extended part of statistics on the behaviour of the algorithm available in [80] may guide the user towards finding appropriate parameters and rating of the results.

[3] This implies that the user has a defined number of processors for his own use. However, standard queueing systems used on large clusters do not reserve such resources completely to one user, but grant computation time dynamically. Thus, on heavily used systems it can not be expected that the calculation is starting immediately after submitting it to the cluster.

4.6. Final remarks on evolutionary algorithms for crystal structure prediction

Despite using the same method and principles, there are lots of variants and peculiarities when looking at different realisations of evolutionary algorithms for crystal structure prediction. The recent chapter aimed at giving an overview of the different approaches and collects many references for further reading. It also demonstrated that there are many ideas worth noticing and developments in this field are ongoing.

A main focus lay on the methodology used in EVO which has been explained and illustrated in detail. Several technical details were omitted for brevity and will be discussed in the appendix. Some of the information are also contained in [75] and can, of course, fully be found in the documentation of the program.

Mostly by USPEX there have been many important findings throughout the past years. It has to be mentioned that it looks back on several years of continuous development. Compared to that, other programs implementing the same method are quite young but have nevertheless shown to be valuable for crystal structure prediction. One can be sure that further research may yield to interesting new results and also methodological improvements over time.

EVO is under continuous development to provide new evolutionary operators, support other programs, become more user-friendly or meet the challenges of bigger systems. However, it has already shown its applicability. The next two chapters will demonstrate the behaviour of EVO and present some structures found by the algorithm along with further results concerning stability and properties.

5. Testing the implementation

Successful crystal structure prediction is based on reliability of the search algorithm. Since global optimisation in search spaces containing more than one minimum is far from trivial, there is no guarantee for success, i.e. for locating the global minimum. However, in some cases of crystal structure prediction it can be assumend that nature already found the most stable structure at certain conditions some of which where therefore chosen to be test cases for EVO. An example for this is carbon with the two solid state phases graphite and diamond of which graphite is the global and diamond a local minimum under ambient conditions. It is therefore considered to be a valid test case for the main functionality of EVO: searching the energy landscape for the global minimum and possible local ones.

Actually, the book edited by Oganov in 2010 [34] holds the 'First blind test of inorganic crystal structure prediction methods' which compares the performances of different methods. Only three different approaches faced the challenge: evolutionary algorithm, simulated annealing and random search (see section 3.3 for information on the methods). The tests were mainly done by the authors who contributed a methodical chapter to [34] as well. The search was conducted for very complex hypothetical structures with up to 60 atoms per unit cell using ASP (Vienna Ab-initio Simulation Package, a non-free implementation of DFT) or force fields with GULP. The results favour the evolutionary algorithm USPEX as being the most reliable program. Unfortunately, the complete parameter sets for the force fields or DFT used in the calculations are not documented. This is most likely the reason why there is no other algorithm to our knowledge on which results of the test have been published. It is also the motive that kept EVO from being compared directly. However, the search in the carbon energy landscape demonstrates the reliability of EVO.

The following chapter shall show results of several tests that explore the parameter space and illustrate the behaviour of the algorithm with respect to different input values. The tests have been done using DFT (QUANTUMESPRESSO) or force fields (GULP), the latter mainly for the parameter tests. First of all the results of a systematic investigation of carbon are presented. The focus lies here one the performance of the algorithm and its steadiness in finding the most stable structure. A detection of a new carbon structure and the study of its stability will be covered in section 6.1 which is fully devoted to the research after EVO finished its work.

Furthermore, evolutionary runs of Si, SiO_2 and $MgSiO_3$ were conducted with DFT. The results shall be analysed and compared with respect to their speed of convergence and needed time for computations. Similar aspects are of interest for the extended parameter tests done on the SiO_2 system. They shall show the influence of different choices for the input values of EVO and illustrate important parameters. Thus, it may be possible to derive hints for an appropriate choice of the parameters which can even affect finding the global minimum[1]. Of course, individual optimisation of the parameters for different systems may fasten convergence or even affect success. However, for directed optimisation one would need to know the result so that both aims (i.e. searching unknown energy landscapes and fast convergence) can not be reached simultaneously.

5.1. Carbon energy landscape

One of the first test cases of the early version of EVO has been carbon since it exhibits two clearly distinct phases with diamond and graphite. Furthermore, the unit cell that can represent both structures consists only of eight atoms and the calculational parameters for convergence (e.g. cutoff energy) are affordable. This allows for manageable computation times for the tests. The known minima provide a search environment that is well suited for a 'proof of principle'. Apart from the reliability of finding the global minimum of the energy landscape the search with EVO should simultaneously find local minima as well. An important expectation to be matched was therefore the finding of graphite and diamond in one search run.

Interestingly, Glass and Oganov [76] also chose carbon as a first test case for their evolutionary algorithm USPEX since it was mentioned to be challenging for crystal structure prediction [25]. The reported results focus on the finding of structures at different pressures up to 2000 GPa and they found the structures known to exist in this pressure range and also show some new ones. However, there are only scarce information on the behaviour of the algorithm and nearly no description of the parameters in use. For the calculations at 0 GPa unit cell sizes of four and eight atoms per cell were used. It also does not become clear if the results were found in one run for each parameter set (e.g. pressure and number of atoms) or if there were several of them.

This section shall present results of extended tests with carbon and EVO all done at zero pressure and will deal with the convergence behaviour of the algorithm as well as with the structures found. Therefore, several search runs were conducted with identical input apart from the number of atoms which were chosen to be 8, 12 and 16 atoms per unit cell. All other evolutionary parameters were not specified in the input so that

[1] Due to the unknown energy landscapes the finding of the global minimum can not be guaranteed in case of finite calculation times.

the default values were used. This leads to a wide range of possible volumes and unit cell dimensions which induces nearly no restrictions on the search even though the known phases would allow for narrower limits. Thus, the tests were done as unbiased as possible.

All calculations have been done using DFT as implemented in QUANTUMESPRESSO with an ultrasoft pseudopotential and appropriate calculational parameters. Since first tests of an early EVO-version yielded an exciting new phase, crossed graphene (see section 6.1), we will also present results of those tests. Additionally, there has also been one search for all numbers of atoms using the current version of EVO which includes the cell transformation step and the crossover recombination not present before.

First tests in the carbon system Although the rest of the tests reported here have been done using the recent version of EVO, this paragraph is devoted to some older data which led to the identification of crossed graphene. It seems consequent to also present these results which are the basis for the results presented in section 6.1.

Figure 5.1 shows the emergence of the fitness of the best individual of each generation for the two searches with 8 atoms per cell and one with 16 atoms per cell. Yet, two searches have been done using 8, 12 and 16 as numbers of atoms. The discrepancy to the picture (only two runs with 8 atoms and one with 16 atoms) is due to a particularity of the early version of EVO: It was found that the population size can decrease rapidly in some cases. The needed number of individuals is generated and calculated in the first place. However, some (or sometimes many) of them are eliminated if the calculation does not converge or they violate the constraints after the relaxation. This may lead to very few valid individuals. If this happens in the first generation creating parents randomly, the diversity of offspring may decrease below the threshold for automatic termination. This led to the termination of the two test runs using 12 atoms and one of two runs using 16 atoms of carbon done with the early version after one or two offspring generations, respectively.

The two searches with eight atoms did not suffer from a too small diversity: Starting with relatively high enthalpies both runs achieve their lowest fitness as early as in the second offspring generation. This also leads to the automatic termination of EVO due to the fact that no improvement of fitness could be detected for five generations which is easily understood since the lowest-enthalpy structure corresponds to a supercell of graphite. In the case of the 16-atom-search graphite emerged in the eighth generation persisting to the end of the run due to no fitness improvement. When considering 16 atoms per cell, the search space has 51 dimensions which is nearly doubled compared to the 8-atom searches (27 dimensions). This is the reason the bigger unit cell takes more generations to find graphite. None of the searches contributing to the figure approached the too-low-diversity criterion thus proving that it is essential to have a

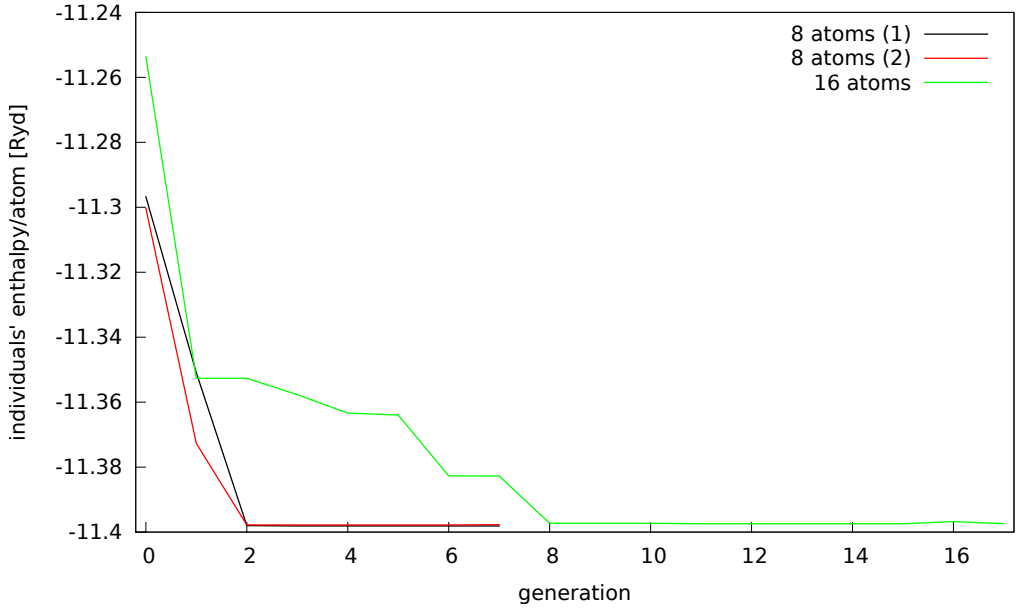


Fig. 5.1.: Evolution of the fitness of the best individuals for three searches in the carbon energy landscape. Two of them used eight atoms per unit cell and one 16.

certain amount of diversity in the first generation. Given that, the produced offspring are able to cover wide parts of the search space.

It has to be mentioned that the overall diversity in the best structures of the run is not that big. By visual inspection many structures have been found to be very similar to graphite but were not detected by the similarity test. Possible reasons for this lie in the very tight thresholds of the test that shall prohibit false positives: If distinct structures are judged to be similar, this unacceptably biases the search. It is rather accepted that similar structures may survive undetectedly. Furthermore, graphite layers are hold together by Van-der-Waals forces which are not described by DFT. Thus, different stackings result in different structures with practically no enthalpy difference.

Also, these few tests could demonstrate that EVO reliably found graphite as the lowest enthalpy structure at 0 GPa in all cases. In addition, a thorough examination of an eight-atom run revealed the existence of diamond and crossed graphene present in the best structures. Figure 5.2 shows the structures found. While graphite exhibits nearly perfect stacking, the diamond structure is found with a stacking fault. Crossed graphene is a newly discovered structure.

Apart from the shown ones no other structures were found. It seems that these three are the dominant minima given the number of atoms per cell. Considering that EVO also finds the graphite supercell is an evidence of its reliability. The simultaneous finding of three distinct structures accounts for the functional though not sophisticated similarity test. Also, the structure of crossed graphene reveals one of the advantages of

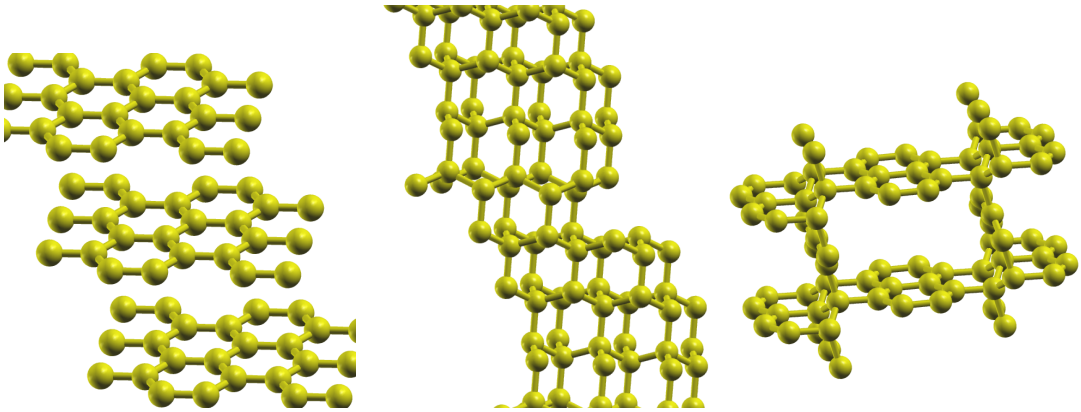


Fig. 5.2.: The three distinct structures found in one run of EVO started with 8 atoms per unit cell: Graphite, diamond with a stacking fault and crossed graphene, from left to right.

evolutionary algorithms for crystal structure prediction: The combination of favourable structural features may yield interesting new modifications.

Results of the recent version of EVO Since the algorithm has been improved gradually, the most recent version differs sufficiently from the one yielding the results of the last paragraph. This is reason enough to conduct the tests reported in the previous paragraphs anew which may give insight to the impacts produced by added features. These are mainly:

- The cell transformation algorithm enhances comparability of the structures and thus allows for finding similar ones more reliably.
- Recombination scheme is extended by structure crossover.
- (The more thorough relaxation of the best individuals during postprocessing is directly coupled to space group recognition.)

At first there shall be a detailled analysis of the test run using 8 atoms per unit cell which directly connects to the results presented before. It has been done at 0 GPa and using default input parameters as above.

The test terminated automatically after creating six offspring generations since there was no fitness improvement for five generations. This implies that graphite as the lowest energy structure showed up in the first generation and thus one generation earlier than in the test with the old version. More interesting is the look at the diversity of the populations which is caused by the enhanced similarity recognition and nearly doubles throughout the whole search.

This affects the best individuals of the run as well. At first, a new postprocessing procedure was applied: The more thorough relaxation with increased computational accuracy has been directly linked to a space group identification mechanism. Of course,

picture label	visual identification	space group number	energy/atom [Ryd]
5.3(a)	graphite	2	-11.398
5.3(b)	lonsdaleite	194	-11.387
5.3(c)	crossed graphene	12	-11.383
5.3(d)	carbon foam	71	-11.376
5.3(e)	unknown	12	-11.376

Tab. 5.1.: Data of the best structures found in the 8-atom test. The structures are shown in figure 5.3.

due to the automatic process there is no guarantee for optimal results. Yet, it can help a lot for the judgement of the results and does not require further manual effort.

Figure 5.3 shows the structures of the best individuals and table 5.1 holds the according names, detected space group numbers and energies per atom. Out of the ten best individuals of the run five are truly unique. The other five are imperfect duplicates of graphite or crossed graphene that were not identified by the similarity test. In the case of graphite this is mostly due to different stacking behaviours. Moreover, the correct spacegroup (186) is not found automatically though the structure can be spotted visually. A promising result is that the space groups of lonsdaleite and crossed graphene are identified correctly. Apart from the natural structures and the previously predicted crossed graphene, we found also a 'carbon foam' (figure 5.3(d)). Similar structures have been constructed from chemical intuition before [109, 110] with different pore sizes. Following the naming convention used there, the found structure is a 1,2-armchair carbon foam. For the first time, EVO found such a structure in an unbiased way, i.e. without including chemical knowledge. Furthermore, another structure (figure 5.3(e)) has been found that can be described as highly distorted graphite sheets that connect at some atoms to form interlayer bonds. Considering this, it may be related to some structures presented in [110].

Compared to the early test using carbon and eight atoms per unit cell the diversity within the best individuals nearly doubled. Graphite as the global minimum was successfully found in all cases. Also, with the carbon foam a completely new structure appeared in the evolutionary search that has neither been found in EVO before nor was reported by Glass and Oganov [76]. The first proposal of carbon foams based upon combinatorial principles alone while this is the first appearance in automatic crystal structure prediction to our knowledge.

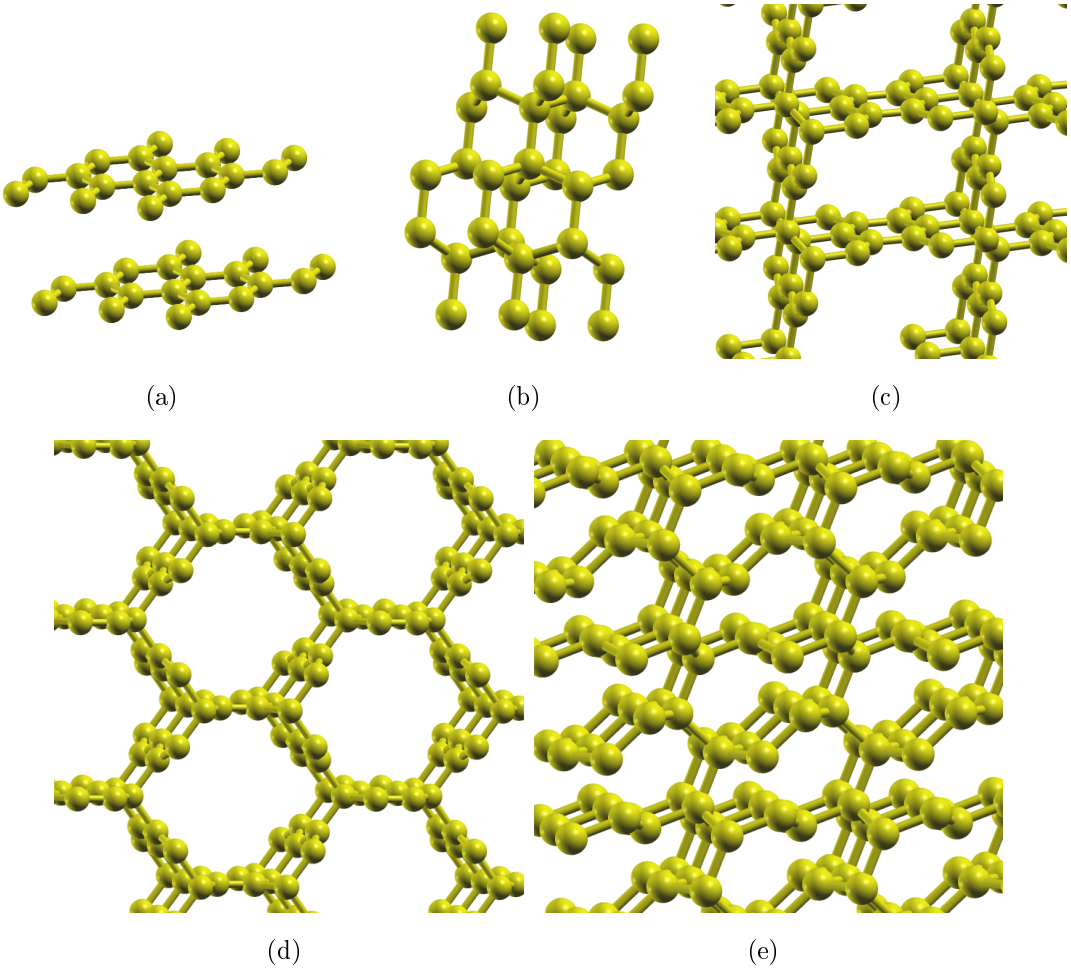


Fig. 5.3.: Best structures of the 8-atom test with the recent version of EVO: a) graphite, b) lonsdaleite or hexagonal diamond, c) crossed graphene, d) carbon foam and e) unknown structure showing irregular pores.

5.2. Systematic exploration of structures with increasing complexity

In order to further investigate the behaviour of EVO a series of chemical compositions known from nature has been chosen. Silicon and oxygen belong to the most abundant elements on earth and are not reluctant to react. They form silicon dioxide SiO_2 which can be found to take various crystal structures[2] when referring to the ICSD [112]. This is due to the fact that there are many possibilities to build a network of tetrahedrally oxygen-coordinated silicon of which the most important form is α-quartz. However, there exists a variety of other modifications that form under higher pressures and/or temperatures and are metastable under ambient conditions. Moreover, there is a huge

[2]For a detailled theoretical study of the most important phases see [111].

Test system	formula units	parents	offspring	max. generations	runs
Si	8				3
SiO_2	3	10	50	50	5
$MgSiO_3$	2				3

Tab. 5.2.: Input parameters for the tests exploring the systems of Si, SiO_2 and $MgSiO_3$.

number of minerals that include SiO_2. As a third test case in the row of silicon and silicon dioxide $MgSiO_3$ was chosen.

For these three materials existing crystal structures are known. However, this section does not explicitely aim at producing and identifying the known structures which would be vastly complicated due to the large number of modifications of SiO_2 or its very flat minimum[3]. Moreover, this issue has been covered in the last section. Therefore, we will only compare the results of EVO by energy and/or volume to the expected stable structures and omit a full identification.

The test shall reveal the dependency of increasingly complex structures on the performance of the algorithm. For a better comparability, the structures were chosen to have similar numbers of atoms for the input of the search. All other input values of EVO remained unchanged. Also, computational parameters of the calculations with QUANTUMESPRESSO were aligned which enhances comparability in terms of CPU time. Table 5.2 holds the input parameters chosen for the EVO-runs with the different systems, for an explanation of the values see section 4.3 and the appendix for a full documentation. The results of the calculations shall be shown and explained starting with the ones for the individual systems including a comparison.

Silicon supercell with 8 atoms Silicon crystallises in the cubic crystal structure of diamond type so that each primitive cell holds two atoms. The supercell of 8 atoms was chosen to approximate the number of atoms present in the quartz unit cell. This ensures comparability with respect to the dimensions of the search space.

Figure 5.4 shows the fitness of the best individual of the according generation. Run 1 and 2 terminated early due to the fact, that the offspring of the according generation were too similar. This automatic termination criterion is met when only three out of 50 offspring are dissimilar. It implies that the algorithm has been trapped in a local minimum. Thus, a further search is not likely to yield further valuable results since the chance is very small that the recombination of the available parents introduces new features. Actually, the third run did not terminate early but only slightly failed to meet the same criterion in many generations.

[3] Quartz is expected to be a difficult example for crystal structure prediction: It has been shown that its energy-volume curve is extremely flat which makes finding of the correct minimum numerically challenging [113].

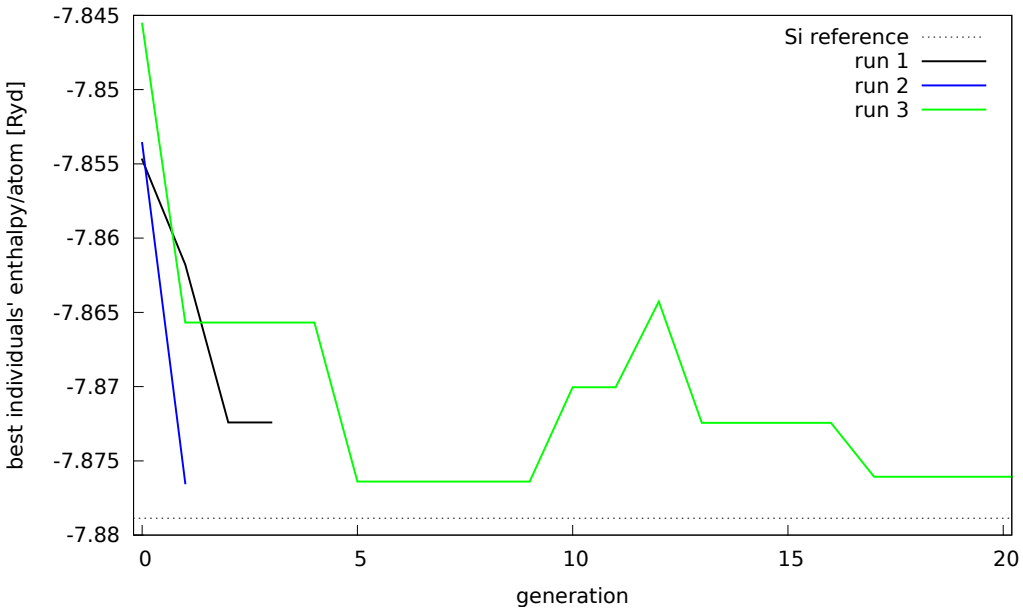


Fig. 5.4.: Evolution of the fitness of the best individual of the three identical runs using 8 silicon atoms.

However, it is also shown, that the fitness gradually decreases and approaches the value of the ideal silicon crystal structure which has been obtained using the same parameters in the input file for the DFT calculations as in the evolutionary runs. As can be seen, the energies per atom of EVO are comparable to those of the idealised structure. It is thus deduced that the best structures found correspond to the expected one.

Silicon dioxide In the case of silicon dioxide, five identical evolutionary runs were done. This accounts for the structural diversity [112, 111] found in the system with many different phases present which can be obtained by applying pressure and/or temperature but are metastable at ambient conditions. Thus, an energy landscape containing lots of minima is expected which complicates crystal structure search.

Interestingly, three out of five searches (figure 5.5) found their minimum of the run during the first 10 generations. However, they were left by EVO in pursuit of possible better ones. During their time of existence, these best individuals were not able to produce offspring with similar good fitness. Yet, the energies of the found structures agree very well with the quartz reference calculation done using the same computational parameters. Marginal lower energies lie within the range of numerical accuracy and do not automatically imply a more favourable structure. Further investigations show that the low energy cutoff is responsible for this behaviour. Choosing a higher energy cutoff yields the expected lower energy for the quartz reference.

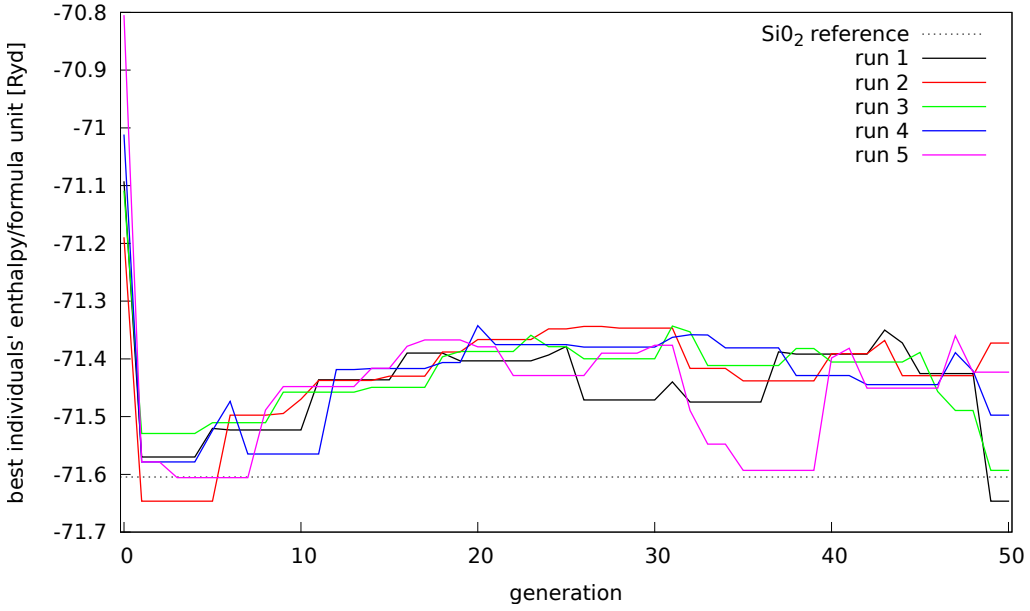


Fig. 5.5.: Evolution of the fitness of the best individual of the five identical runs using three formula units SiO_2.

Furthermore, it has to be mentioned that the fitness calculations are not converged since the energy cutoffs needed for a well converged calculation of quartz (e.g. for finding the minimum of the energy-volume curve correctly) are extremely high. If we used these instead, the computational time needed for the search would not have been manageable. Table 5.3 shows energy and computational cost data for the quartz reference cell and the best found individual. At the very high cutoff of 250 Ryd the quartz calculation is converged. For the same cutoff the total energies of the optimised structures compare well while bigger differences occur when looking at the results with different cutoffs for the same structure. As expected, the CPU time needed for the relaxation of the cell increases with higher cutoff energies. However, for the structure created by EVO the increase is much larger since the structure is located further away from the actual minimum. The CPU time of 29 hours for such a calculation is way to high for an evolutionary run regarding the need of about 50 such calculations for the creation of a new generation. This large amount of computer time is unacceptable for a test case but may be tolerated in production runs depending on the computer power available.

However, since the energies obtained with the low cutoff value agree well, the results of the test are deemed trustworthy. Furthermore, it has to be mentioned that the best individuals of each run are relaxed more thoroughly (increased cutoffs and more k-points) as a first step of postprocessing. Visual inspection of the results confirms the assumption: Figure 5.6 shows the quartz structure on the left side and a very similar

	total energy [Ryd]		CPU time [h]	
E_{cut} [Ryd]	40	250	40	250
quartz reference	-71.60	-72.20	0.25	5.33
best individual	-71.46	-72.17	1	29

Tab. 5.3.: Comparison of calculated total energies in Ryd and CPU times in hours for the quartz reference structure and the best individual found in the runs using two different values for the energy cutoff. The total energy is given per formula unit. CPU times are reported for a full cell relaxation, for the quartz reference starting from the experimental structure.

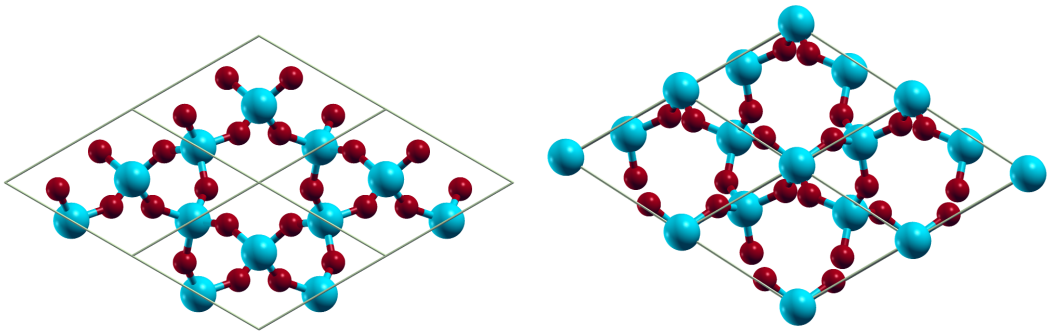


Fig. 5.6.: Left: Quartz reference structure. Right: Very similar structure produced by EVO. The perspective has been chosen to point out the structural similarities. Oxygen is red and silicon is turquoise.

structure proposed by the evolutionary search. The SiO_2 phase found exhibits the SiO_4 tetrehedra expected and illustrates nicely its analogy to α-quartz.

Yet, the structural diversity of the best individuals of the run is very high. A variety of features appear, from the well known fourfold oxygen-coordinated silicon to octahedrally coordinated silicon which can be found in the high-pressure SiO_2 phase stishovite. Figure 5.7 illustrates this by showing the best individuals of a run. Some of them are rather open structures with SiO_4-tetrahedra connected in different ways while others are very dense phases with SiO_6-octahedra. There was no further investigation since a very accurate relaxation is needed for the determination of the space group of the individual structures. However, the test demonstrates nicely that SiO_2 has various minima in its energy landscape. This is also confirmed by the observation that the diversity of offspring remains high throughout the whole test. None of the five searches terminated automatically, they were ended by reaching the maximum number of generations.

Using silicon dioxide as a test case revealed much valuable information about the behaviour of the evolutionary algorithm in complicated energy landscapes. No criterion leading to an early termination (i.e. before reaching the maximal number of generations) applied here. Yet, it is very useful in other systems since it relieves the user from judging the ongoing of the search. This may lead to quite biased results considering that the

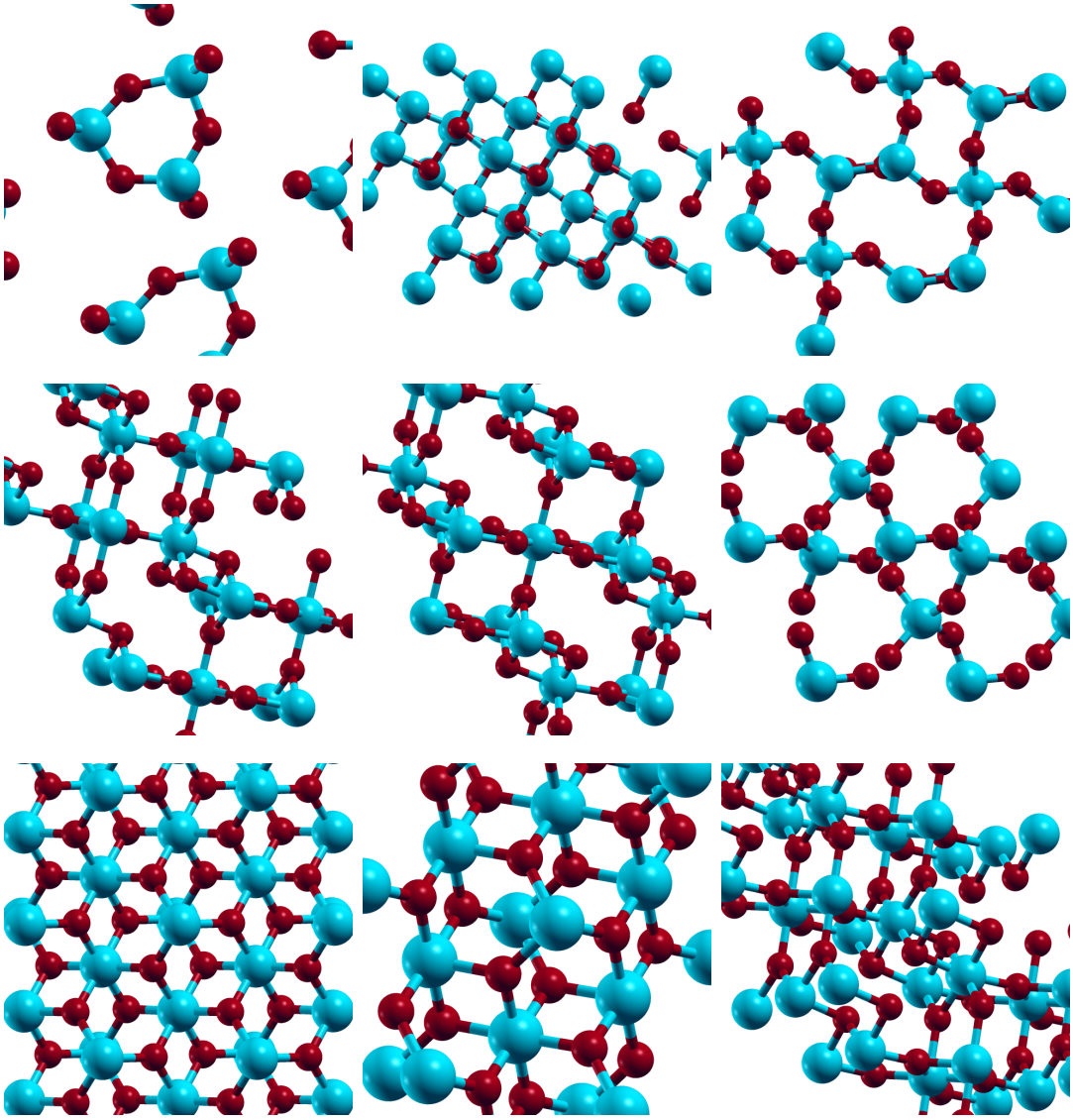


Fig. 5.7.: Illustration of the best individuals found by a single evolutionary search. The great diversity is apparent.

user might stop a search when certain expected structural features appear rather than awaiting new and unexpected results.

Moreover, it was shown that one has to find a balance between the computational cost for each fitness evaluation and the accuracy. CPU times for a fully converged calculation may be unacceptably high for the many evaluations needed during the evolutionary search. Therefore, it is rather recommended to choose a justifiable low cutoff energy for the search and attach a rigorous relaxation for the best found individuals. As demonstrated above, this strategy is suitable for finding a structure comparable to α-quartz which is the global minimum of the energy landscape at zero pressure and temperature. It has to be stressed, that all the results were obtained with EVOs very

liberal default settings. The inclusion of restrictions on the volume, for example, may fasten the search but also may limit the results.

Magnesium silicate $MgSiO_3$ extends the tests of silicon and silicon dioxide to a ternary system. Two formula units per cell have been chosen as input which do not map a cell or supercell of a real phase which has at least five units – 30 atoms – per unit cell. Since these are too many for such a test case, we used only ten atoms for the calculations. Thus, it is not expected to reproduce the structure of $MgSiO_3$ but structural features have to be found. Moreover, the computational time would be much higher and a comparison to the other systems would be impossible with an increased search space.

Figure 5.8 depicts the convergence behaviour of the best individuals of the three test runs. All of them reach the same level of energies quite early causing two of them to terminate after 9 and 16 generations because there has been no improvement of fitness for five generations. In the case of the third run, no termination criterion was reached before the maximal number of generations. However, though very low, fitness variations remained higher than the threshold for termination.

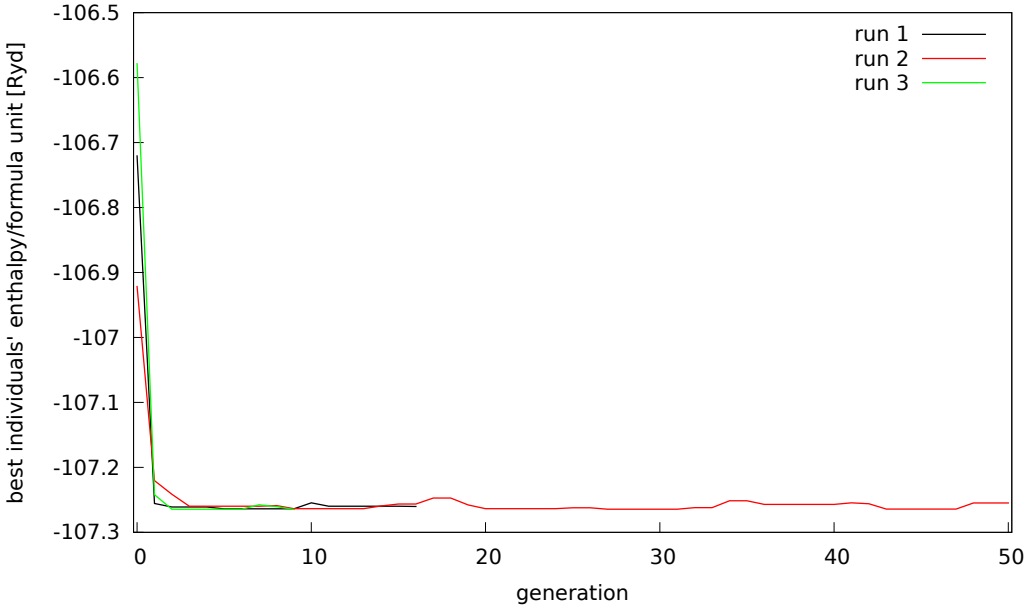


Fig. 5.8.: Evolution of the fitness of the best individual of the three identical runs using two formula units $MgSiO_3$.

The diversity of the offspring stays very high throughout the whole search. Actually, usually 40 out of 50 offspring are dissimilar in each generation. This is also apparent when analysing the parents of the longest run more closely. As shown in figure 5.9, the energy variety within the parents of each generation remains high. However, compared

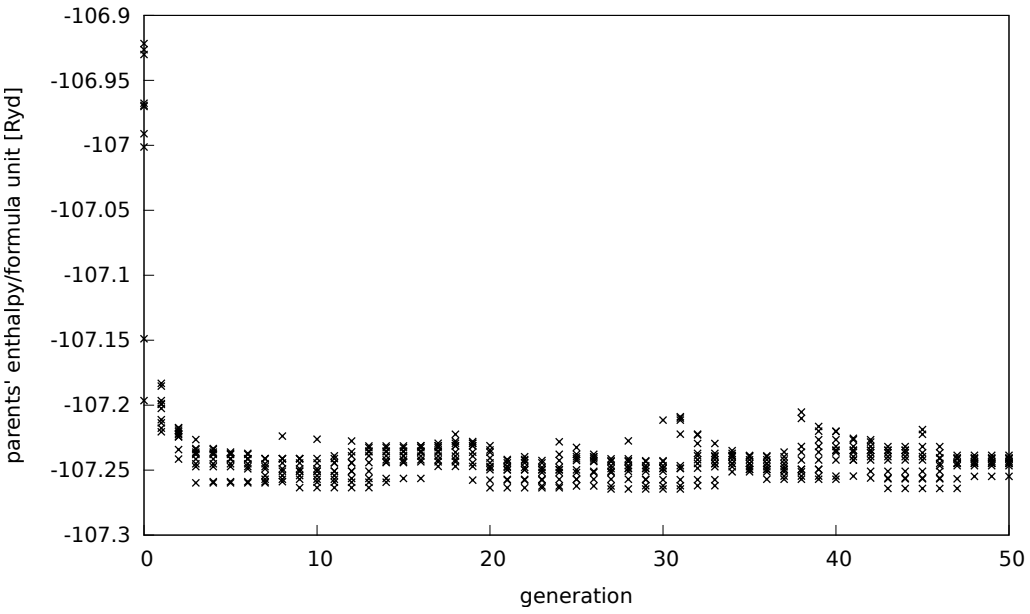


Fig. 5.9.: Evolution of the fitness of the parents of one run of the $MgSiO_3$ search. Corresponds to the red curve in figure 5.8.

to the variations that can be found in the best individuals of SiO_2 (see figure 5.5) the ones in $MgSiO_3$ are marginal.

Despite being expected to be the most difficult energy landscape regarding the number of atoms as well as atomic species, magnesium silicate appears to be less challenging for crystal structure prediction compared to the SiO_2 test. Also, by visually inspecting the obtained results the structures seem less diverse than in the SiO_2 case. This may be caused by an energy landscape containing less minima and/or more pronounced ones or by the fact, that the number of formula units chosen for the test is too small to map the real (possibly more complicated) energy landscape. The small unit cell is also the reason that no comparison to experimental structures can be done.

Summary of the systematic tests The presented tests on the systems of silicon, silicon dioxide and magnesium silicate revealed many issues that are important for crystal structure prediction. First of all, it is stressed that the absolute number of atoms which was nearly equal does not determine the complexity of the search which may be measured by the number of generations to the automatic termination of EVO. This rather depends on the specific energy landscape which can not be estimated for so far unknown systems. However, for SiO_2 which is known to exist in so many modifications all occupying minima in the energy landscape the search proved to be difficult. Yet, structures that are very similar to the global minimum structure, quartz, (see figure 5.6) appear in several runs proving EVOs ability to deal with complicated systems as well.

Moreover, using the SiO_2 example, it has been demonstrated that it is not necessary to use fully converged calculations for the fitness evaluation. For the sake of reasonable computational times one can rather resign from highly accurate calculations. However, this assumption is only valid when suitable postprocessing steps are taken as a proper relaxation with an increased cutoff energy and k-points density. Since this is only done for the best individuals determined automatically by EVO, the costs are manageable. Postprocessing then guarantees the exact location of the various minima and is essential for a full crystal structure prediction including determination of stability and properties which is the subject of the next chapter (chapter 6).

The tests may be analysed even further. For example, it would be possible to add a full structure identification. Yet, this exceeds the scope of this section and valuable insights could be gained on the given basis. In the following, this chapter shall also describe the behaviour of the algorithm with respect to certain input parameters such as the size of the population or the number of atoms per unit cell. All these issues were completely omitted here and will be covered in the next section.

5.3. Parameter tests of EVO

Evolutionary algorithms for crystal structure prediction share their working principle but can become very different when looking at their particulars. Sections 4.3 and 4.5 give insight into some of the details that distinguish between different approaches. Even within one implementation there are many options that can be tuned to aid convergence or fasten the search. Of course, they can also affect the outcome of the search, e.g. finding the global and possible local minima.

Reviewing literature reveals only scarce information on that issue. There are several reports on the convergence behaviour of general evolutionary algorithms which date especially from the early years of development. Those have been determined by the investigation of special test functions. Though, concerning crystal structure prediction most developers did not publish any information on that behalf. Outstanding in this field is XTalOpt [80] providing a very detailled and computationally expensive analysis (i.e. they conducted a total of 4500 test runs of the algorithm) of the different input parameters that tune the convergency of the search. The authors mainly discuss the influence of the various evolutionary operators and their parameters.

In contrast to that, we will focus on the 'classical' parameters in evolutionary strategies such as the size of the population. Moreover, the effect of the size of the supercell or the cell transformation step will be determined. The test searches are done on the SiO_2 system which was also subject of the previous section. Due to the many possible structures the system is a challenge for crystal structure prediction and sufficiently complex for studying the convergence behaviour of EVO. Too trivial energy landscapes

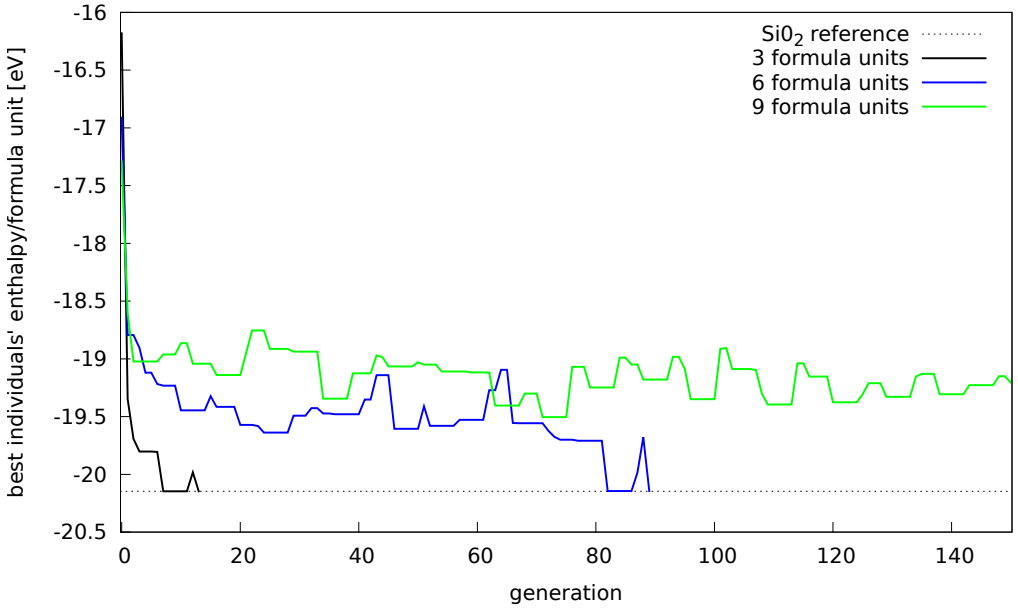


Fig. 5.10.: Evolution of the fitness of the best individual using different numbers of formula units of SiO_2.

formula units	atoms	dimensions of search space	reached reference in generation
3	9	30	7
6	18	57	82
9	27	84	-

Tab. 5.4.: Overview of input data and number of needed generations to find the minimum energy, i.e. quartz structure.

are not qualified for such investigations since the minimum/minima are found easily in any case.

For the sake of reasonable CPU times, all calculations were done using GULP and accompanying potentials which speeds up the single fitness evaluations (e.g. variable cell relaxations) dramatically compared to DFT. All the tests except the first one were done with nine atoms per unit cell thus depicting the conventional quartz unit cell.

Number of formula units The test presented here discusses the influence of supercells with 3, 6 and 9 SiO_2 formula units on the results and ongoing of the search. All multiples were chosen to describe quartz (3 formula units per conventional cell) ensuring comparability. As hard termination criterion a maximal number of generation was fixed to be 150 and all other search parameters remained unchanged.

Figure 5.10 illustrates the evolution of the best individuals of the searches over the generations. First of all, it shows that the number of generations until the automatic or enforced termination of the search increases rapidly with the number of of formula units

and thus atoms in the unit cell. This is easily explained by the simultaneous increase in the dimensions of the search space, see table 5.4. The dramatic elongation of the search when using the doubled supercell by a factor 11 exemplifies the advantages of using as small cells as possible. The even bigger cell (27 atoms) did not reach the minimum at all within 150 generations. However, the global minimum was found for the 18-atom supercell of the system in reasonable time. It matches expectation that the higher the dimensions of the search space, the more expensive the search gets. Thus, the use of supercell larger than necessary cannot be recommended.

For practical purposes it may be useful to first search the lower dimensional spaces and gradually increase the number of atoms rather than counting on the fact that the algorithm will find the global miminum in a large supercell as well. This is true only in the case of an infinite number of generations.

Number of generations Since in most of the other evolutionary algorithms for crystal structure prediction the number of generations or created individuals is an important value, the following test deals with the influence of the number of generations on the results of the algorithm. With EVOs automatic termination check applying several criteria, this is expected to be of less importance. As a side effect of the previous tests, the duration of the search in terms of needed generations for the different cases could be seen. When looking at these results, it is not possible to derive a rule for a maximal number of generations since this heavily depends on the system and its energy landscape as well as the size of the unit cell.

To more thoroughly investigate this issue an extended collection of tests with the 9-atom cell of SiO_2 and the pair potentials implemented in GULP was conducted which shall complement the results presented so far. There were no variable parameters in this test suite. The maximal number of generations was set to 50.

All of the eight identical EVO-runs terminated before reaching the maximal number of generations. Seven reached the minimum denoted by the reference energy of quartz. All runs were ended because there was no fitness change for five generations. The curves show that most runs found the minimum earlier and left it before returning which demonstrates EVOs ability to leave minima and to avoid being trapped in local ones.

Regarding the successful searches the automatic termination criterion was reached after a mean of 18 generations. The success rate was 88 %. It has to be mentioned that there was no real statistical analysis as the number of samples is to low for such considerations. Yet, the maximal number of generations of 50 for the search of the quartz structure in a 9-atom SiO_2 unit cell can be considered to be well above the generally needed number of generations for this system. Since it is designed to be the criterion reached if no other can be applied before, it fulfils its purpose completely.

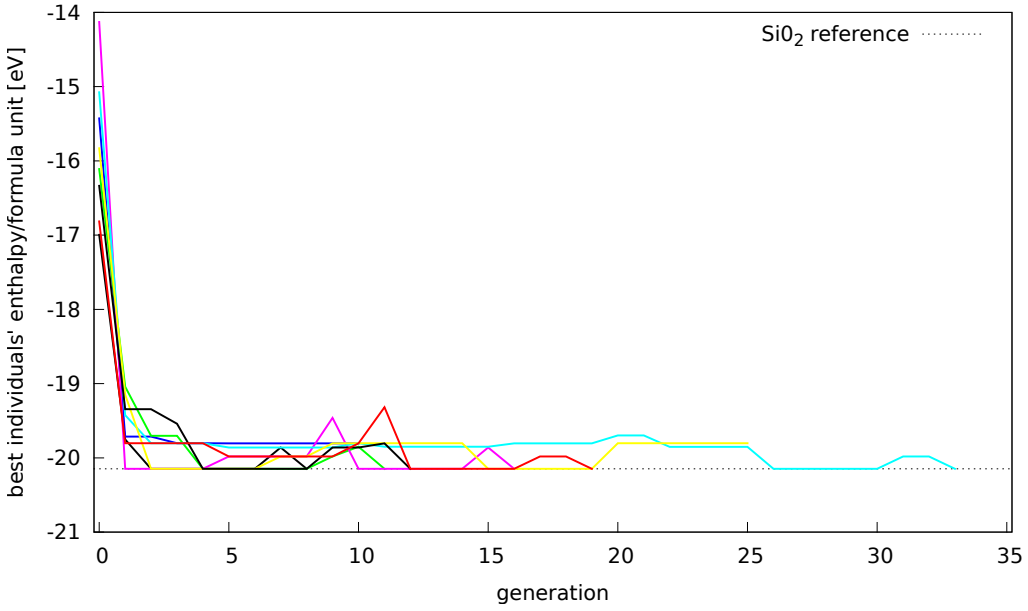


Fig. 5.11.: Evolution of the fitness of the best individual of the SiO_2 search of 8 identical runs with GULP. The dark blue curve did not reach the reference energy.

Another interesting result comes into play when comparing the results here to the ones obtained for the same system and evolutionary parameters with the fitness evaluations done within the framework of DFT as implemented in QUANTUMESPRESSO. Figure 5.5 shows that none of the searches there terminated automatically before reaching the maximal 50 generations though all of them draw near the reference structure. However, it seems that the evaluation method (DFT vs. pair potentials) extensively influences the energy landscape of SiO_2. This is considered to be a hint for the biasing of known structures by pair potentials while DFT is more impartial.

Number of parents and offspring Traditionally, the numbers of parents μ and offspring λ are highly important in evolutionary algorithms. Bäck [73] presents theoretical results on optimal parameters (e.g. μ and λ) for a standard evolutionary strategy based on theoretical and practical findings. Yet, these value cannot be transferred directly to most evolutionary algorithms for crystal structure prediction since they are often used in conjunction with relaxation as a tool for local optimisation within the global optimisation scheme. This alters the search from covering the whole space to rather sampling the minima alone.

Figures 5.12 and 5.13 show the trends of the fitness of the best individuals of each generation with varying numbers of parents and offspring, respectively. Note, that the label of the x-axes now shows the numbers of calculated individuals instead of the generation number. Since the fitness evaluation is in general (at least when using

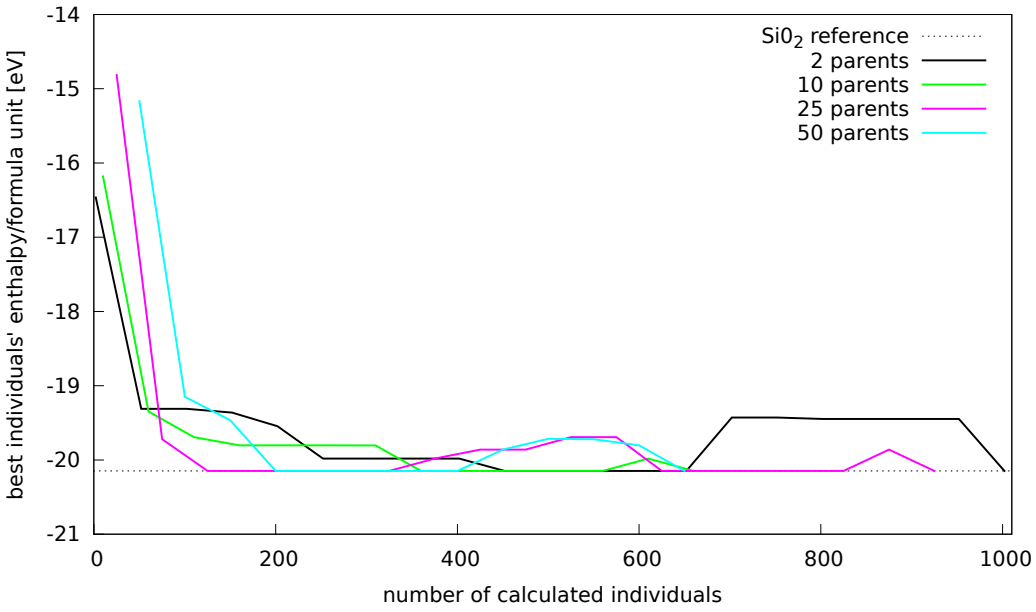


Fig. 5.12.: Evolution of the fitness of the best individual of the SiO_2 search varying the number of parents μ. The offspring number was 50 in all cases.

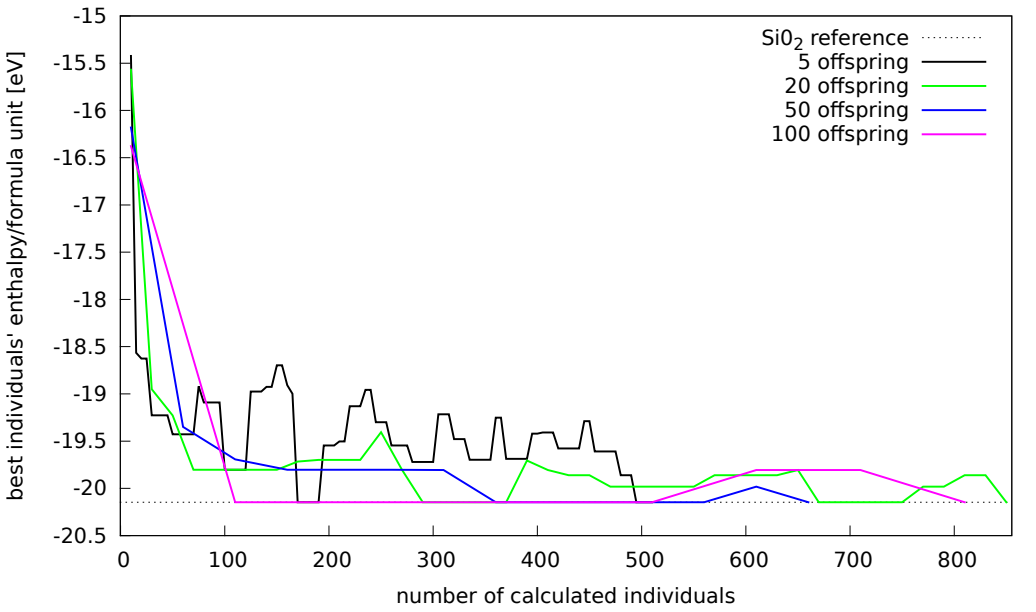


Fig. 5.13.: Evolution of the fitness of the best individual of the SiO_2 search varying the number of offspring λ. The number of parents was 10 in all cases.

DFT) the most expensive part of the search and the number of evaluations varies with changing μ and λ, this label choice has been determined to be more significant.

A modification of the number of parents (figure 5.12) appears to have little influence on the finding of the minimum since all searches reach the α-quartz energy. The first approach to the reference energy happens earlier in the case of higher μ while the search which first found the lowest energy also takes longest to terminate. It also becomes visible that having too few parents ($\mu < 10$) is the worst choice of the ones shown here: The minimal energy is found late and the run also takes very long to terminate automatically. Termination happens due to a too low diversity in the best structures of the run.

When changing the number of offspring (figure 5.13), the curves look quite different: The bigger λ the smoother are the graphs which is only due to different "borders" of generations that along with the selection produce steps in the curves. Neither in the needed times of evaluations nor in the success of the search there can be any conclusions on optimal choices for the input parameters μ and λ. However, interpreting the early finding of the minimum with 100 offspring implies that there is random success in the search. It seems that one only has to produce a big enough number of random oder quasi-random structures to find the lowest energy.

In summary, it is not possible to derive recommendations for the values of μ and λ apart from the fact that the number of parents should be not too low to guarantee diversity. The user may therefore apply the default values of EVO $\mu = 10$ and $\lambda = 50$ and occasionally raise the number of parents.

Cell transformation The effect of the cell transformation feature on the diversity of the results has been shown in section 5.1. In the tests with carbon and 8 atoms per unit cell new structures emerged when using it. Since the comparison was done between two code versions including other changes, the following test guarantees comparability by two runs just differing in the application of the cell transformation.

As shown in figure 5.14 the introduction of the cell transformation positively affects the length of the search as well as on the generation in which the minimum is first found. While the minimum is reached in generation 13 and the search terminates automatically after 22 generations without cell transformation, with the feature the quartz energy is found in generation 7. Moreover, the algorithm terminates after 13 generations.

Along with the insights obtained in the carbon tests cell transformation is highly recommended for crystal structure search within EVO.

Constraining the volume By limiting the volume it is possible to reduce the search space quite dramatically. EVO-structures may take volumes between the close-packing of the atoms and the quadruple of it (see section 4.3.2). For the current test system

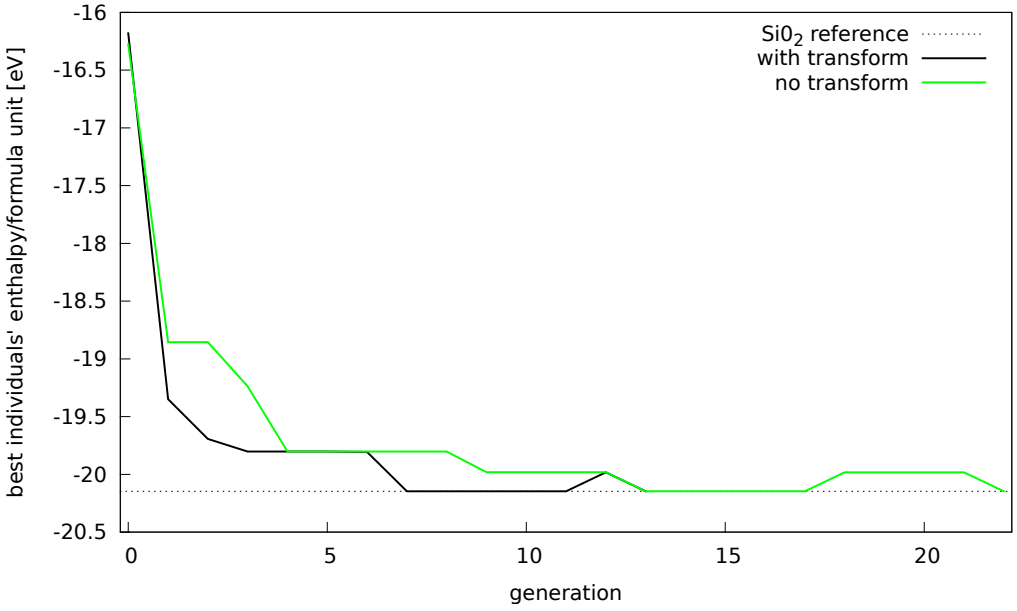


Fig. 5.14.: Evolution of the fitness of the best individual of the SiO_2 search with and without cell transformation.

this means that the volumes of generated structures are allowed to take values between 37 and 148 Å^3. In this range, the reference volume of quartz is located well above the mean at 113 Å^3.

EVO offers the possibility to constrain the volumes according to the needs of the user. It shall be shown here that the usage of this option can be beneficial to the convergency of the search to include such information which may be known about the target structure. Figure 5.15 illustrates the favourable influence of constraining the volume. The tighter limits lead to an accelerated finding of the minimum as well as an early termination compared to the default volume limits.

In actual search situations applying volume limits can speed up the calculations a lot. If one also considers lower computation times due to fewer relaxation steps towards the current local minimum (especially important for DFT calculations), this effect is increased even more. Yet, choosing too tight constraints can prevent the algorithm from finding the global minimum or possible local ones.

5.4. Overview of the tests

The results of the present chapter allowed an insight into the behaviour of the evolutionary algorithm EVO. Therefore, several tests runs were done using both DFT and pair potentials for the fitness evaluation. While the latter approach impresses by calculational speed, it seems to constrict the results according to the used potential.

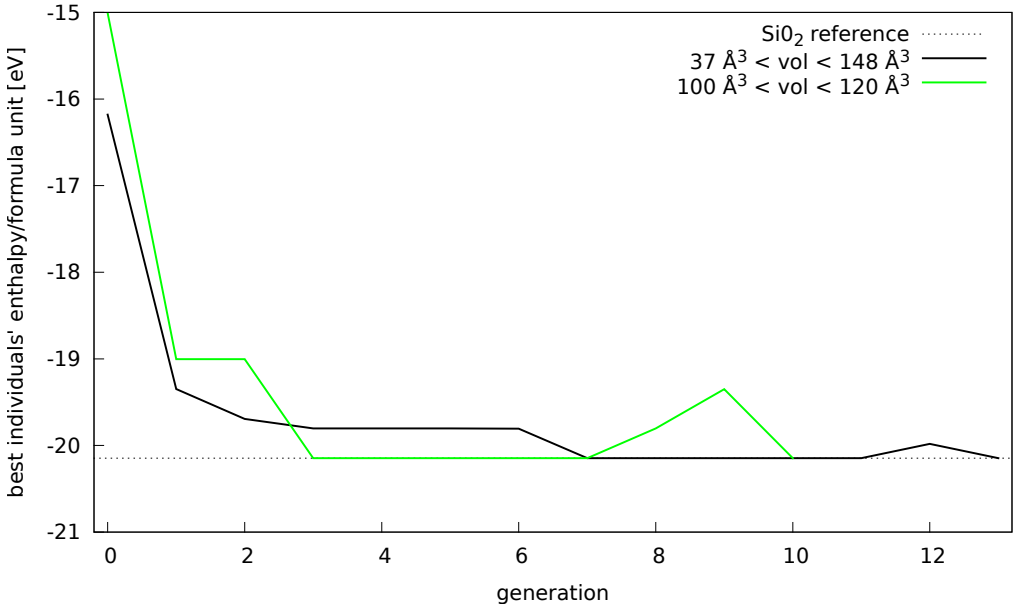


Fig. 5.15.: Evolution of the fitness of the best individual of the SiO_2 search with the standard constraints on the volume (black) and more tight range around the quartz volume (113 Å^3).

The applicability of the potentials to arbitrary structures cannot be guaranteed and is thus not recommended to use in unknown energy landscapes.

Yet, for the determination of suitable input parameters for EVO using the well-known silicon dioxide and pair potentials enabled quick test runs. Though it is not possible to derive universially valid parameters a few recommendations are given here:

- Constrain the volume if possible.
- Too few parents reduce diversity and thus hinder convergency. A value of about 10 is recommended by many tests.
- Cell transformation is extremely valuable for quick search results and can also increases diversity of the found structures (see section 5.1). It should therefore not be switched off.

Furthermore, it has been shown that the number of atomic species does not determine the complexity of the search. Comparing the tests of SiO_2 and $MgSiO_3$ the formed energy landscape and thus the challenge of the search cannot be derived from it. Within the same system, searching for supercells decelerates convergency very much. Thus, cells should be as small as possible to build the target structure.

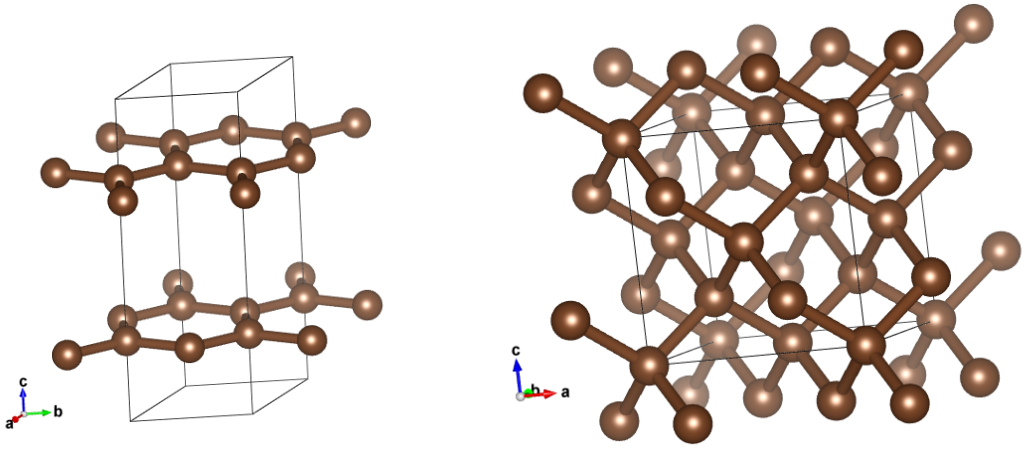
6. Predicting crystal structures

Finding a low energy structure of a certain system unfortunately is only the starting point of crystal structure prediction. From a theoretical point of view there are many other calculations that need to be carried out for confirming or disproving stability of a structure. Also, the impact of a structure can be approximated by evaluating many properties theoretically. The present chapter deals with some new structures proposed by EVO and the further research on their properties and stability. In fact, this second part of crystal structure prediction is much more challenging and most often computationally far more expensive than the search itself. The examples presented here shall give insight into the complex matter of determining stability and properties of materials theoretically.

At first, a so far unreported structure of carbon, named crossed graphene, is shown before further findings of a structure of germanium nitrofluoride and a search for boron sheets are presented. The methods used for the diverse calculations are widely known in theoretical materials research but are explained in the chapter covering DFT (see chapter 2) for integrity.

6.1. Complementing the carbon universe with crossed graphene

The two commonly known solid phases of carbon, namely graphite and diamond, are since long ago of interest for mankind. Diamond is not only a beautiful jewel but a wide-gap insulator exhibiting high hardness while graphite is a higly anisotropic semi-metal and a lubricant. Those complementary properties originate from the chemical bonding found in the two phases (figures 6.1(a) and 6.1(b)): Graphite consists of planes of sp^2 bonded carbon atoms which are held together by the weak Van-der-Waals interaction. In-plane elastic constants and electronic properties are highly interesting. Freestanding single sheets are called graphene the discovery of which led to the Nobel prize awarded to Andre Geim and Konstantin Novoselov in 2010 for 'groundbreaking experiments regarding the two-dimensional material graphene' [114]. They conducted fundamental experiments with graphene and thus opened a wide field of research [115]. In diamond on the other hand each atom is tetrahedrally bonded to four other carbon atoms and shows an outstanding thermal conductivity as well as a very good chemical resistivity.



(a) (b)

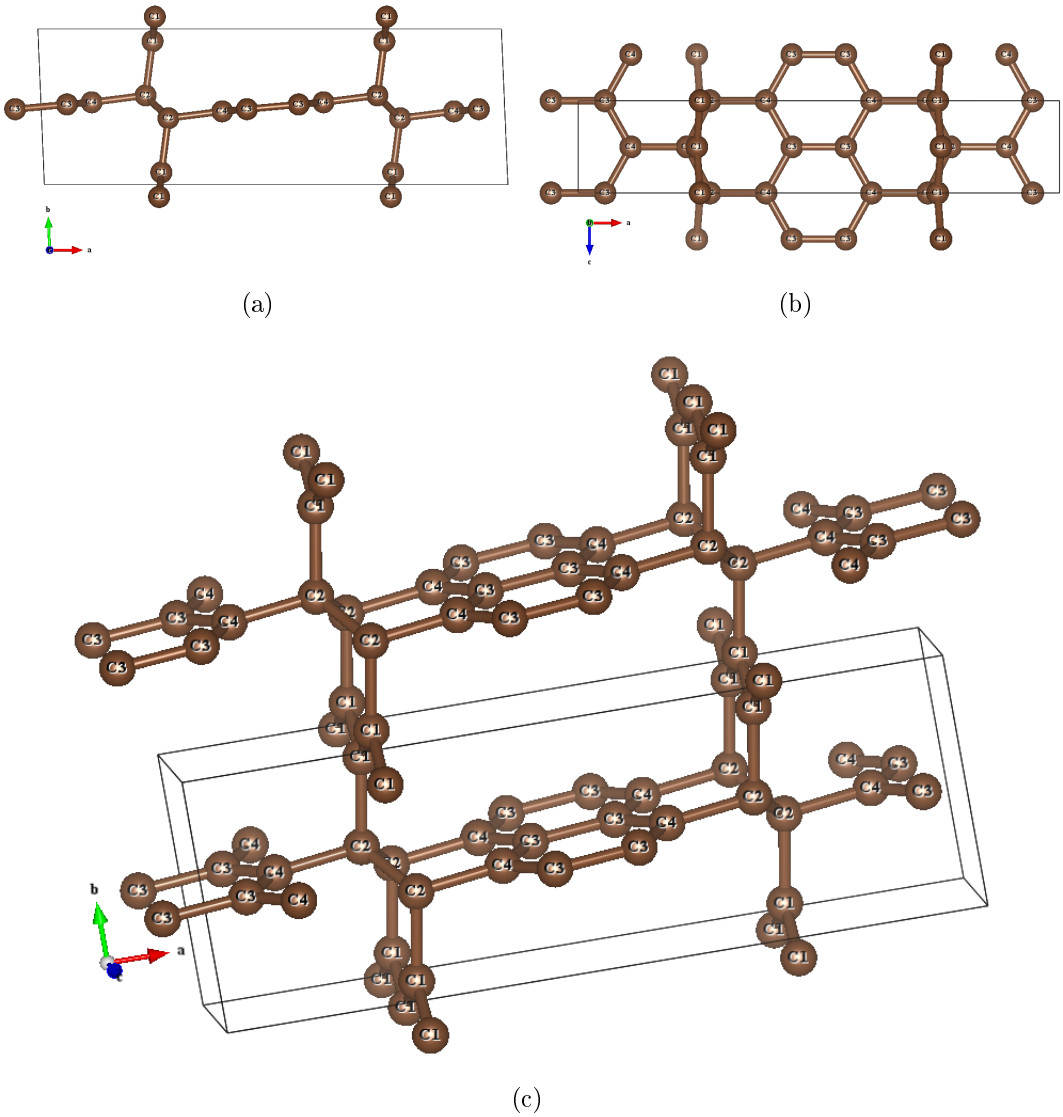
Fig. 6.1.: Crystal structures of (a) graphite and (b) diamond.

Regarding the peculiar but complementary properties of the two natural phases of carbon a purposeful combination of sp^2 and sp^3 bonds could yield very promising materials. Theoretically, a variety of dense phases [116, 117, 118, 119, 120] and so called 'carbon foams' [121, 122, 123, 110, 109] have been proposed over the years. Most of them have never been found in experiments except for a superdense C_8 phase (see [120]) or a high pressure phase made of solvated fullerenes [124], for example. Due to results like this, research on new phases never ceased.

Interestingly, the high pressure modification of *M-carbon* was predicted by the evolutionary algorithm USPEX [76] (see section 4.5.1) in the early test phase of the algorithm. Few years later *M-carbon* [125, 126] was discovered to match *superhard graphite*, a carbon phase which was found in experiments [127] but which structure could back then not be solved.

When developing EVO carbon was used as a test case as well: It should prove that the code can find the two main bulk modifications simultaneously. The test (see section 5.1) was successful and graphite as well as diamond were found as structures with the lowest enthalpies (conducted at 0 GPa) thus confirming the expectation. Surprisingly, a third phase (figure 6.2) appeared in all cases which was termed *crossed graphene* since it is built by crossing graphene sheets that form nearly rectangular tubes.

The present section shall first give details of the actual search and identification of the structure followed by an extensive explanation of the determination of stability and properties. Main parts of the following paragraphs have recently been published [128]. Unless otherwise mentioned, the calculations for the following results have been done using QUANTUMESPRESSO.



(a) (b)

(c)

Fig. 6.2.: Different perspectives of crossed graphene.

6.1.1. Search for carbon phases with EVO

In order to test the implemented evolutionary algorithm, an early version of it (e.g. without cell transformation) was started using different numbers of atoms per unit cell, namely 8, 12 and 16, at zero pressure. All other settings were constant. The focus will lie here on the two runs using 8 atoms per cell which yielded the expected graphite and diamond structure along with a new structure. For general considerations concerning convergence with respect to different input parameters refer to chapter 5. Further information on the carbon test case is given in section 5.1.

Both runs of EVO terminated automatically after 7 generations since the best structure did not change for 5 generations. Figure 6.3 illustrates the ongoing of the crystal structure search by the enthalpy values of all parent individuals of the respective generations. A graphite structure can actually be found as the lowest energy structure in

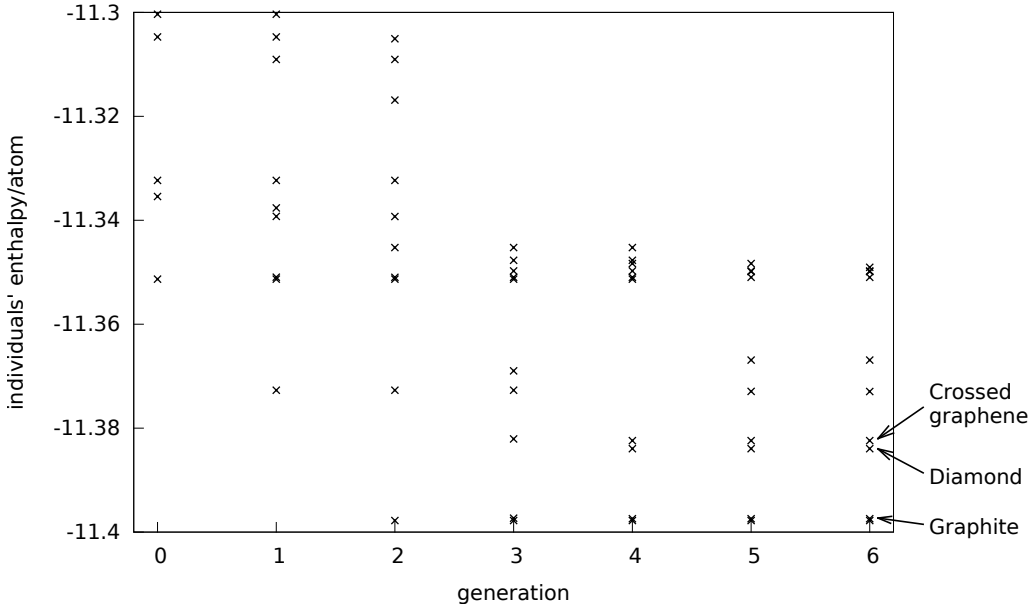


Fig. 6.3.: Evolution of the individuals enthalpy per atom during the generations. Each point denotes the enthalpy of one structure. Best individuals have been identified visually, see legend. Overlapping points mainly denote the same structures. This is due to the unavailable cell transformation which hinders the effective removing of redundancies.

generation two which persists to the end of the search. Also a gradual improvement of the pool of the individuals becomes visible. Due to the early version of the code without the cell transformation step and therefore an ineffective similarity test there are many overlapping points designating very similar structures.

The best structures of a run were examined more carefully. First of all, they are relaxed more throroughly (i.e. increased cutoffs and k-point resolution). This enables the use of programs that determine spacegroups from arbitrary crystal structures which are quite sensitive to accurate atomic positions and form of the unit cell. Two routines were used here for the investigation of crossed graphene:

- the PySPGLib (algorithm after [129]) extension to PythonASE [130] (Atomistic Simulation Environment). It can directly be imported into any python code and just needs the appropriate ASE crystal format as input.
- the sgroup [131] program distributed along with WIEN2k [14]. To use this, a suitably formatted input file has to be provided and the symmetric structure has to be read from the output.

For the interesting structure of crossed graphene, the original output (figure 6.4(a)) of EVO first underwent the cell transformation (see section 4.3.3) implemented some time after the actual search (and was therefore not used during the run). The transformed structure (figure 6.4(b)) already looks nicer but the above mentioned symmetry finding

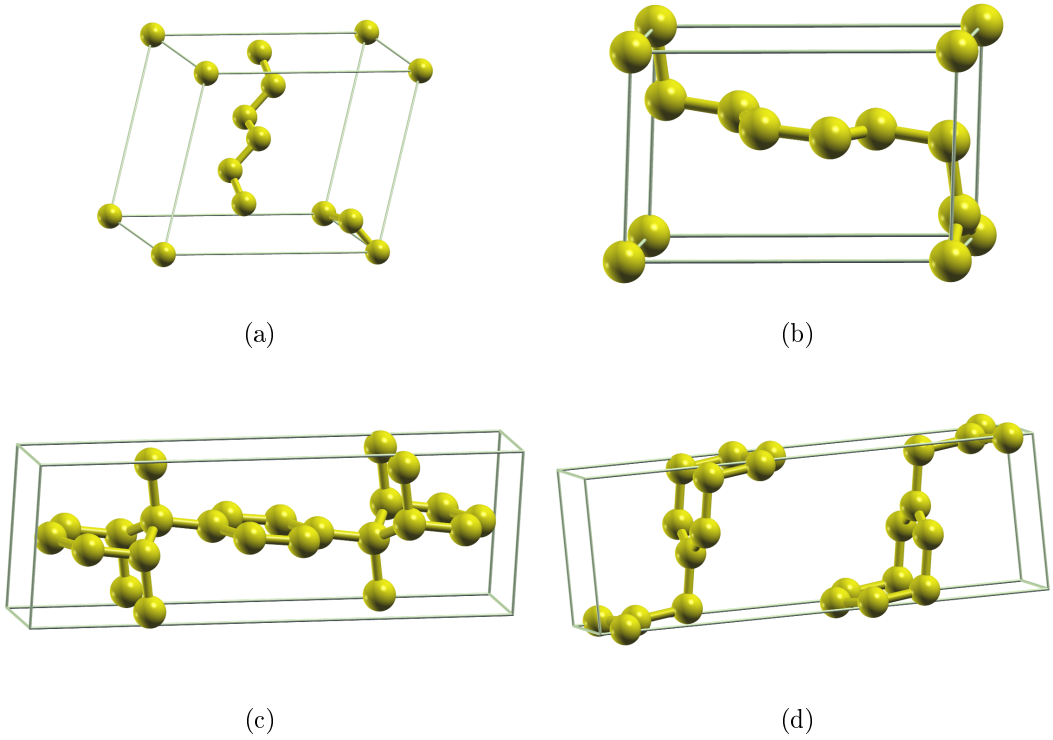


Fig. 6.4.: Finding symmetry in a predicted structure: (a) original output of EVO, (b) same individual after cell transformation as implemented in EVO, (c) PySPGLib finds crystal structure of spacegroup 12 (C2/m), (d) sgroup also detects spacegroup 12 but obviously uses another setting.

routines (tolerance set to 0.001[1]) improve the result even further. Both determine the spacegroup of crossed graphene to be monoclinic base centered with spacegroup number 12 (C2/m) [132] and 16 atoms in the unit cell of which four are non-equivalent. Table 6.1 holds the resulting structural data.

Wyckoff site	x	y	z	Point group
4i	0.2536	0.5787	0.0000	m
4i	0.7265	0.0732	0.0000	m
4i	0.0560	0.0154	0.0000	m
4i	0.3903	0.9704	0.0000	m

Tab. 6.1.: Structural data of crossed graphene in the monoclinic space group (12) C $2/m$, setting B $112/m$ (unique axis c), $a = 12.9371$ Å, $b = 4.3378$ Å, $c = 2.4678$ Å, $\gamma = 92.568°$. All sites are fully occupied.

The origin of the name crossed graphene stems from the obvious analogy between nearly rectangular crossing graphene sheets and the found structure. It can be seen as

[1]There is no general recommendation for the tolerance value in PySPGLib or sgroup. The first finds more symmetry (i.e. higher spacegroup numbers) with increasing tolerance while sgroup unintuitively may yield lower symmetries with higher tolerances.

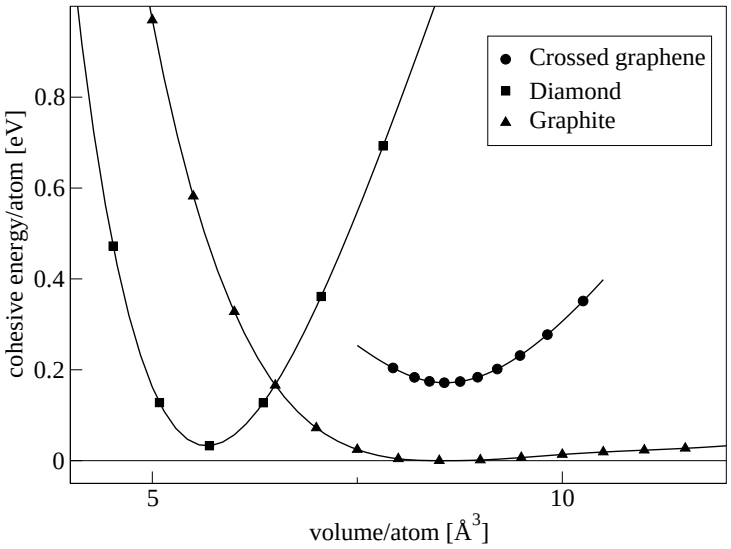


Fig. 6.5.: Energy-volume curve of crossed graphene compared to graphite and diamond. Due to the insufficient description of Van-der-Waals interactions by DFT and in contrast to the ones for diamond and crossed graphene, the data for graphite has been obtained by optimising a and c/a for the different volumes rather than by variable-cell relaxation.

a hybrid diamond-graphite material. The pores formed measure approximately 12 Å× 8 Å and when transferred to the notation introduced by Kuc and Seifert [109] can be denoted as a (3,2) armchair carbon foam. In contrast to the structures presented in [109] where three planes cross at angles of 120°, in crossed graphene four of them are arranged nearly rectangularly. In between the junctions, crossed graphene (figure 6.2) exhibits almost perfect planar sp^2 carbon while the crossing is built by a double line of slightly distorted tetrahedral sp^3 bonded atoms.

As demonstrated, EVO can successfully find a new crystalline phase along with known ones. Further postprocessing steps are necessary to identify and refine the structure and spot contained symmetries. In the following, it shall be shown how stability and properties of the new structure can be assessed.

6.1.2. Determination of stability and properties of crossed graphene

Energy-volume curve A necessary condition for (meta)stability of a material is that its energy regarded as a function of the volume is minimal. An elegant way to calculate it is the usage of the variable-cell relaxation featured by QUANTUMESPRESSO for a user-defined hydrostatic target pressure. The optimal volume changes dependent on the pressure, allowing to indirectly compute energy-volume pairs of values.

Figure 6.5 depicts the energy-volume curve of crossed graphene. The form of a parabola confirms the hypothesis of stability with respect to the change of volume.

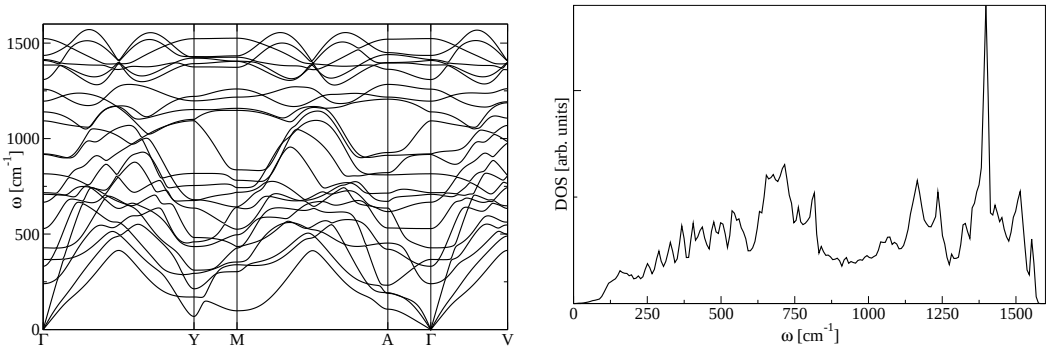
However, comparing the curve to the ones of graphite and diamond shows a higher cohesive energy in the whole volume range thus indicating that crossed graphene cannot be made simply by applying pressure. The very same data can also be used to derive the bulk modulus. It describes the ability of a material to withstand pressure and is related to the curvature of the energy-volume curve. To get the bulk modulus an equation of state (EOS, Birch-Murnaghan 2nd order [133]) is fit to the data yielding a bulk modulus of 217 GPa.

Elastic tensor and moduli Apart from analysing the curvature of the energy-volume curve, there is another way to obtain the bulk modulus. By calculating the elastic tensor C_{ij} (6×6) which fully describes the elastic properties of a crystal it is also accessible. Since the matrix is symmetric, the 36-component tensor can contain up to 21 independent entries for triclinic systems. In the case of the monoclinic cell of crossed graphene this reduces to 13 components. These are determined by extensive stress-strain calculations: The unit cell is distorted in defined directions and the stress tensor of the resulting cell is calculated by DFT. Then, the stresses can be used to obtain the values of the elastic tensor:

$$C_{ij} = \begin{pmatrix} 701 & 23 & 80 & 0 & 0 & 43 \\ & 437 & 43 & 0 & 0 & 19 \\ & & 1084 & 0 & 0 & 9 \\ & & & 195 & 20 & 0 \\ & & & & 300 & 0 \\ & & & & & 65 \end{pmatrix} \text{GPa}$$

Since the elastic tensor is positive definite, we get a second indicator for dynamic stability [134]: Crossed graphene can resist small distortions of the unit cell. Looking more closely reveals a pronounced anisotropy between in-plane and out-of-plane elastic properties which corresponds to the observations and expectations available from graphite/graphene. Moreover, C_{33} describing the c-direction (figure 6.2) pointing along the extended graphene sheets in crossed graphene is in good agreement with the graphite in-plane elastic constant [135]. This analogy implies that the stiffness of the structure in this direction shows nearly no influence of the double row of sp^3 carbon connecting crossing layers.

It is possible to derive the bulk modulus from the elastic tensor along with another value important for materials description: the shear modulus. Both elastic moduli can be regarded as isotropic values derived from the full tensor. The two averaging methods were proposed by Voigt [136] and Reuss [137]. Hill [138] nicely characterises the differences: 'Voigt's procedure contains the assumption that the strain is uniform throughout an aggregate, and Reuss's that the stress is uniform.' Thus, bulk B and shear moduli G according to the methods are:



(a) (b)

Fig. 6.6.: Phonon dispersion and phonon density of states (PDOS) as calculated using supercell simulations (128 atoms) applying WIEN2k. There are no imaginary frequencies present.

$$B_V = 279 \text{ GPa} \quad G_V = 250 \text{ GPa}$$
$$B_R = 230 \text{ GPa} \quad G_R = 139 \text{ GPa}$$

The deviation in the values according to Voigt and Reuss, especially regarding the shear moduli, demonstrates the great influence of anisotropy to those averages. However, the bulk modulus obtained by fitting the energy-volume curve is in good agreement with the bulk modulus after Reuss. Also, when comparing the elastic moduli to that of the structures proposed by Kuc and Seifert [109] we get a remarkably high shear modulus. This may favour crossed graphene over the other carbon foams for realisation in experiments.

Phonon dispersion The lattice vibrations can be understood as a measure of the stability of a structure with respect to perturbed atomic positions: Finding imaginary frequencies in a phonon dispersion indicate the present structure to be unstable to some atomic movement, i.e. appropriate (long) MD simulations would push the structure into the real ground state. For the calculation of the phonon dispersion curve (figure 6.6(a)) or the phonon density of states (PDOS, figure 6.6(b)) supercell calculations with WIEN2k [139, 16, 17] as well as density-functional perturbation theory (DFPT) as implemented in QUANTUMESPRESSO were employed. Both methods yield similar results, figure 6.6 shows the data gathered with WIEN2k. Since there are no imaginary frequencies present, crossed graphene can be regarded to be stable.

Raman spectrum Apart from stability considerations phonons also play their role in the characterisation of materials. Especially when it comes to carbon, Raman spectroscopy is a tool of great importance (cf. corresponding chapter in [140] or for special work on graphene/graphite [141, 142]). The Raman spectrum of crossed graphene was calculated by applying second-order response [11] as implemented in QUANTUM-

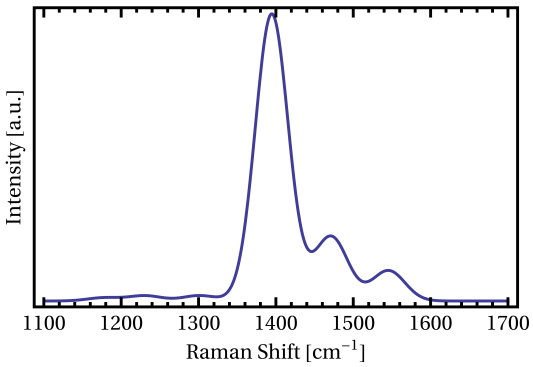
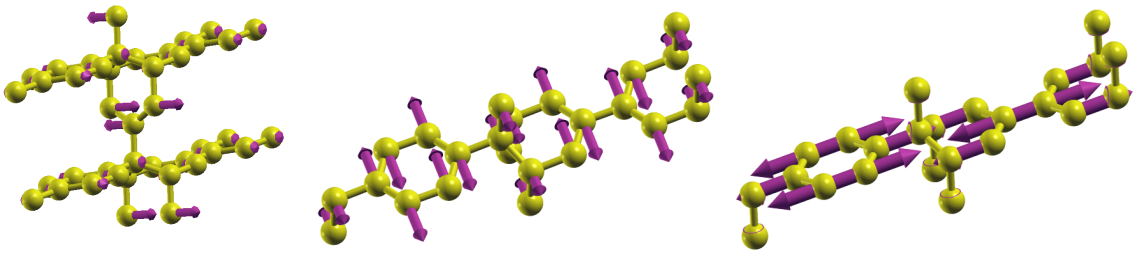


Fig. 6.7.: Calculated Raman spectrum of crossed graphene; the range was chosen to show the relevant peaks. The most pronounced mode lies at 1390 cm^{-1}.

ESPRESSO which gives relative Raman intensities of the single modes. The predicted Raman spectrum (figure 6.7) was obtained by summing up gaussian functions with a FWHM (full width at half maximum) of 50 cm^{-1} and the correspondent areas. As can be seen, there are three major Raman modes the intensities of which dominate the spectrum. Also, the spectrum differs significantly from the ones of diamond (one sharp peak at 1332 cm^{-1}) and graphite (two broad bands at approximately 1360 cm^{-1} and 1580 cm^{-1}) [142]. This is mainly due to the fact that the most prominent peak at 1390 cm^{-1} stems from a structural feature present in neither of the natural phases: By visualising the eigenvectors of the respective modes the peaks can be associated with the atomic vibrations that cause them. Figure 6.8 depicts the atomic motions of the three modes possessing the highest relative intensities. The G band at 1550 cm^{-1} present in all graphitic materials is identified to be a result of the vibration of the C-C pairs in the graphene plane. The afore mentioned pronounced peak at 1390 cm^{-1} originates from a new structural feature: The vibration of the C atom connecting the graphene planes.



(a) (b) (c)

Fig. 6.8.: Vibration patterns of the three dominating peaks: (a) lies at 1390 cm^{-1} and is associated with the motion of the zigzag lines of carbon atoms (C1) in between the graphene layers. (b) can be found at 1470 cm^{-1} while (c) at 1550 cm^{-1} corresponds to the G vibration known from graphite/graphene.

Electronic structure The interest in graphene is mainly founded by its outstanding electronical properties [115] and crossed graphene can be expected to share many of them. The band structure and resulting density of states (DOS) were calculated by QUANTUMESPRESSO and WIEN2k yielding once more similar results. Figure 6.9 shows the results of the WIEN2k calculations. Extensive symmetry considerations led to the insight that the band crossings seen at selected points in the Brillouin zone are 'real' ones rather than valence and conduction bands just getting very close. Moreover, similar to the famous Dirac cones [115] of graphene the band crossings are linear though in a much smaller energy range. The DOS reveals the origin of the responsible states: As expected, the electronic states near the Fermi level are entirely π ones which do not participate in the strong bonds connecting the carbon atoms. Thus, they are far less localised and may fill states near the Fermi level.

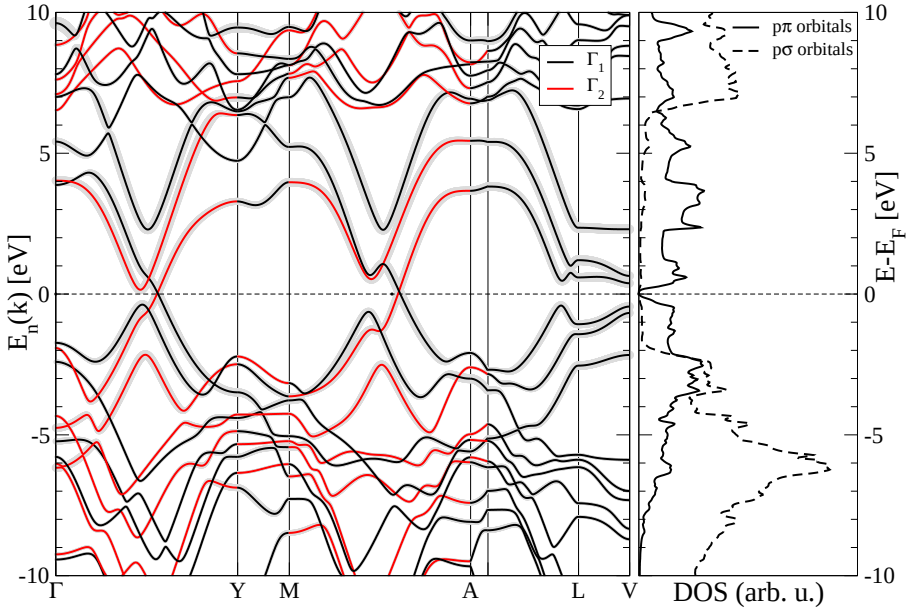


Fig. 6.9.: Band structure of crossed graphene. Bands with different symmetries differ by colour thus showing 'real' band crossings. The thickness of the grey shades visualises the π-character of the shown bands. In addition, the projected DOS of the p-states (pσ and pπ) demonstrates the importance of the pπ electrons for the observed zero band gap.

The crossing points found in the band structure lie at the approximate reciprocal points $(0, 0, 0.555)$ and $(0, 0.5, 0.397)$[2]. For further examination of the linear dispersion, the vicinity of the crossing points has been sampled more thoroughly in three dimensions. As can be seen from figure 6.10, the linearity of the crossings is strongly dependent on the perspective: While having clearly linear energy dispersion in one direction, another view shows parabolic bands. Also, the linear dispersion is not as

[2] Furthermore, symmetry reveals that these points are restricted to the reciprocal $(0, 0, z)$ and $(0, 1/2, z)$ axes as well as the $(x, y, 0)$ and $(x, y, 1)$ planes.

great as in graphene where it holds for about 1 eV around the Fermi-energy E_F. This is mostly due to the vicinity of the adjacent layers.

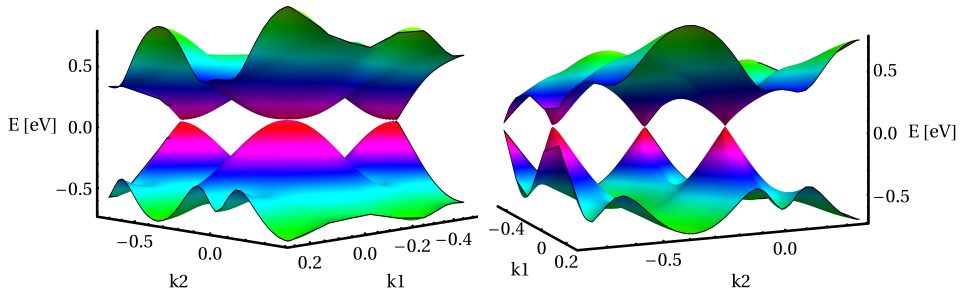


Fig. 6.10.: 3D bands of crossed graphene near the crossing point at (0, 0, 0.555) from different perspectives.

Contradicting the interpretation above, Gradhand and coworkers [143] argue that band crossings are 'infinitely unlikely' in nonmagnetic crystals with inversion symmetry which applies for crossed graphene. All crossings that occur are to be considered accidental and should be removed when including spin-orbit coupling. Further research in this direction would probably solve this contradiction.

6.1.3. Conclusion

The crossed graphene example shows that EVO successfully finds new and probably interesting crystal structures but that significant additional work in crystal structure prediction lies in refining the structure and the actual determination of stability and properties. In particular, crossed graphene was verified to be stable regarding phonons and elastic constants which also confirm the great anisotropy present in the structure. For possible identification a Raman spectrum was calculated exhibiting a pronounced mode that raises from a unique structural feature. The electronic structure can be seen to be derived from graphene/graphite reproducing also linear band crossings at the Fermi energy. The predicted phase is expected to bridge the gap between carbon nanostructures (i.e. similarities to nanotubes) and bulk material.

6.2. Stable phase in the germanium nitrofluoride system

A second example for crystal structure prediction of bulk materials with EVO is germanium nitrofluoride. Respecting the number of valence electrons of the elements yields the formula GeNF. It opens new ground considering that no structure composed of the three elements can be found in the inorganic crystal structure database (ICSD).

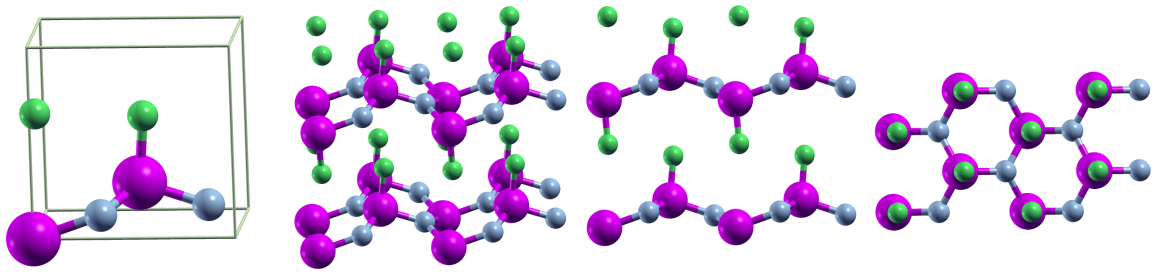
Atom	Wyckoff site	x	y	z
Ge	2a	0.000000	0.607584	0.000000
N	2a	0.000000	0.487977	0.337002
F	2a	0.000000	0.454731	0.025245

Tab. 6.2.: Structural data of GeNF in the orthorhombic space group (31) P $mn2/1$, $a = 3.108$ Å, $b = 5.040$ Å, $c = 5.184$ Å. All sites are fully occupied.

There have been some studies concerning possible modifications of the isoelectronic silicon nitrofluoride [144, 145]. They report on research relying on intuition to derive candidate crystal structures or adapting them from chemically similar compounds by 'geomimetism'. Also, stability was only judged by comparing energies of local minima in the energy landscape obtained by structural relaxation and choosing the lowest one.

EVO provides an elegant way to search the energy landscape without any chemical or crystallographic intuition thus paving paths to new phases. Since for GeNF there are no known phases, every resulting structure can be regarded as new. For the candidate search EVO was started with six atoms (two of each kind) per unit cell at zero pressure and no further restrictions. The resulting structures all looked similar and were energetically comparable hence one of the structures was chosen for further processing. After thorough relaxation a symmetry recognition resulted in an orthorhombic structure. See figure 6.11 and table 6.2 for the crystal structure. Since the cell transformation of EVO was used in this search, the symmetry recognition just introduced minor changes to the structure (i.e. origin shift and small adjustments) to find the orthorhombic space group P $mn2/1$ (nr. 31 in [132]).

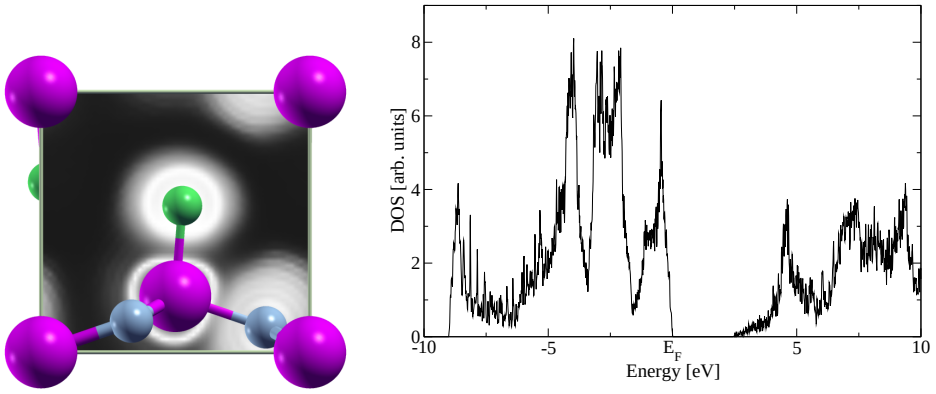
Fluorine introduces a strong ionic character to the compound which is confirmed by inspecting the structure: There are buckled layers of germanium bonded to three nitrogen and vice versa. Each germanium forms a bond to a fluorine which can be regarded to ionically connect the planes by alternating interaction of negatively charged fluorine with the germanium of the opposite layer (see figure 6.11).



(a) (b) (c) (d)

Fig. 6.11.: Structure of GeNF where Ge is purple, N grey and F green. (a) one cell including unit cell boundaries; (b) showing a supercell and (c) and (d) the side and top view of it, respectively.

The bonding behaviour is further investigated by the electron localisation function (ELF). ELFs strive to describe chemical bonds in a graphical way yet some chemical understanding is needed to interpret them. In the case of GeNF, the ELF follows chemical principles (figure 6.12(a)): Fluorine exhibits a strong spherical localisation around the atom corresponding to the strong ionic character of fluorine. The high electronegativity difference between Ge and F leads to high localisation at fluorine. Germanium on the other hand 'looses' its valence electrons. The diffuse localisation around the right nitrogen in figure 6.12(a) accounts for the free electron pair. Altogether, the ELF supports chemical intuition regarding the bonding behaviour of GeNF. For completeness, figure 6.12(b) shows the electron DOS of the predicted structure. The calculated gap is about 2.5 eV wide.



(a) (b)

Fig. 6.12.: (a) 2 dimensional electron localisation function for GeNF. The chosen view visualises the plane of Ge, N (right one) and F in the middle of the cell. ELF values are ranging from 1.0 (white) to 0.03 (black). (b) shows the electron DOS of GeNF. There is a gap of about 2.5 eV.

Furthermore, the stability of GeNF has been investigated using the same methods as in the case of crossed graphene: energy-volume curve, elastic tensor and phonons. The energy-volume curve in figure 6.13 shows a clear but extended minimum where fitting an EOS leads to a bulk modulus of about 20 GPa. This is very low compared to the bulk modulus ($B_{V/R} = 143\text{GPa}$) obtained from the elastic tensor:

$$C_{ij} = \begin{pmatrix} 296 & 42 & 80 & 0 & 0 & 0 \\ & 359 & 18 & 0 & 0 & 0 \\ & & 358 & 0 & 0 & 0 \\ & & & 42 & 0 & 0 \\ & & & & 134 & 0 \\ & & & & & 64 \end{pmatrix} \text{GPa}$$

The derived shear moduli are 82 and 106 GPa according to Reuss and Voigt, respectively. A possible explanation for the discrepancy in the bulk moduli can be the very

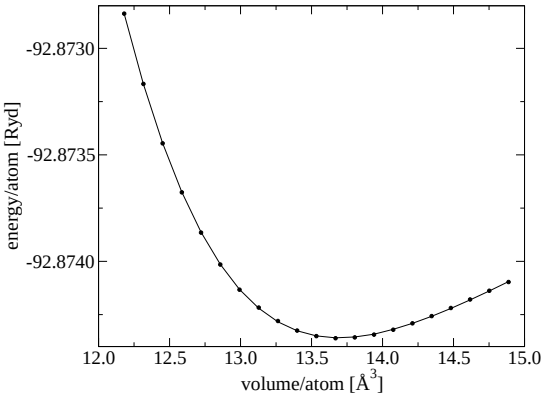
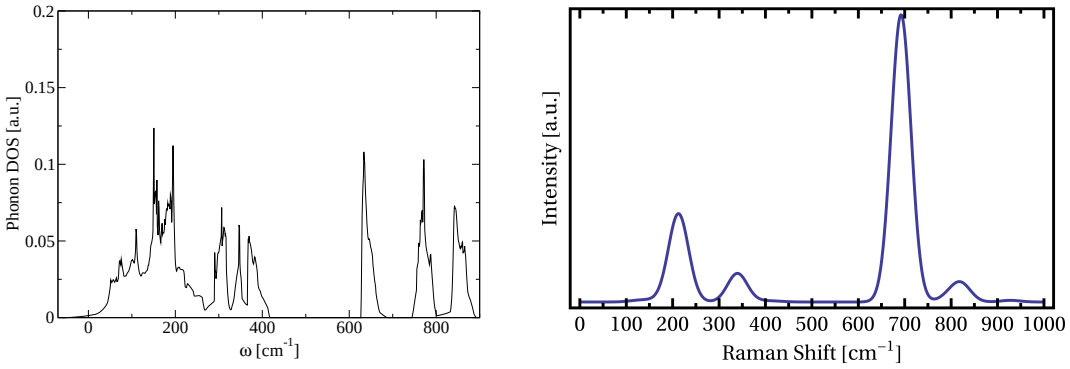


Fig. 6.13.: Energy-volume curve of GeNF.

flat energy-volume curve (figure 6.13) which can lead to large deviations depending on small changes in the parameters of the calculation. However, the clear minimum supports stability.

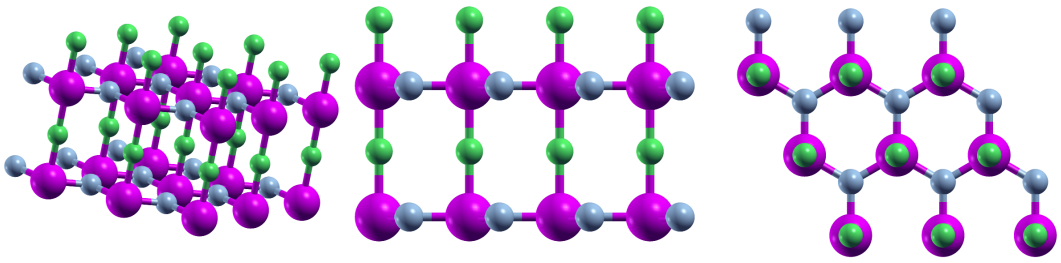
Phonon calculations using DFPT have been carried out as well. The two goals pursued were the confirmation of stability as well as the prediction of a Raman spectrum for possible identification. Figure 6.14(a) shows the phonon density of states; the small values in the negative range (i.e. imaginary phonons) are indicators for either an instability of the structure or insufficient accuracy of the calculations. First further investigations gave some hints that support the assumption of insufficient accuracy. Yet, more research on this purpose is necessary.

The predicted Raman spectrum (figure 6.14(b)) shows four disctint peaks and clearly features possible identification of the structure. Along with X-ray diffraction patterns that can easily be calculated from the structure itself this may lead to detection of the phase.



(a) (b)

Fig. 6.14.: Phonon density of states (a) and Raman spectrum (b) of GeNF calculated using DFPT as implemented in QUANTUMESPRESSO. The spectrum has been simulated using gaussian functions with a FWHM of 50 cm^{-1} which were scaled with the according Raman cross sections.



(a) (b) (c)

Fig. 6.15.: High-pressure phase of GeNF found at 40 GPa. The colours of the elements are chosen in accordance with the figures above.

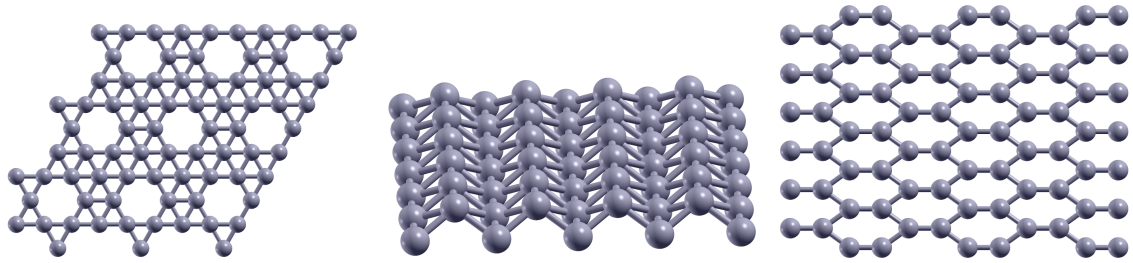
So far, evolutionary algorithms proved highly useful in prediction and solving of high-pressure phases [78, 79, 101, 98]. This may be due to the fact that experiments at pressures of several tens to even hundreds of GPa are though viable yet hard to control. On the other hand, computer simulations easily deal with very high pressures since it is possible to apply a hydrostatic pressure to the unit cell in a variable-cell relaxation. Therefore, also the high-pressure energy landscape of GeNF was explored. There is no need to change any parameter of EVO or the calculations apart from setting a target pressure in the input file of QUANTUMESPRESSO.

GeNF served as a candidate for the search for a high-pressure phase and at pressures of 40 GPa the structure shown in figure 6.15 emerged. Interestingly, there is no fundamental change in the crystal structure. The layers of germanium and nitrogen (see also figure 6.11) flatten and move together so that the coordination of germanium becomes trigonal bipyramidal. However, the top views of the high-pressure and normal-pressure phase (figures 6.15(c) and 6.11(d)) do not differ significantly implying that the bonds within the those planes are rather rigid. The main part of the volume change is due to the approach of the planes.

Of course, a thorough judgement of the stability of high-pressure GeNF would require all the calculations done for the zero-pressure structure. Yet, it was shown that the search for a high-pressure conformation is easily done with EVO.

In summary, the presented phase of germanium nitrofluoride is another example for successful crystal structure prediction using EVO. It was shown that it is also possible to search not only for new phases in known systems but also for new compounds. With GeNF a material was found on which realisation there are no reports. Yet, experimental efforts may be initiated by the presented results.

The unbiased search by EVO and the broad considerations on stability and properties distinguish the present section from works on the isoelectronic SiNF [144, 145]. Also, was pointed that gaining reliable results requires special care and is not in all cases straightforward.



(a) (b) (c)

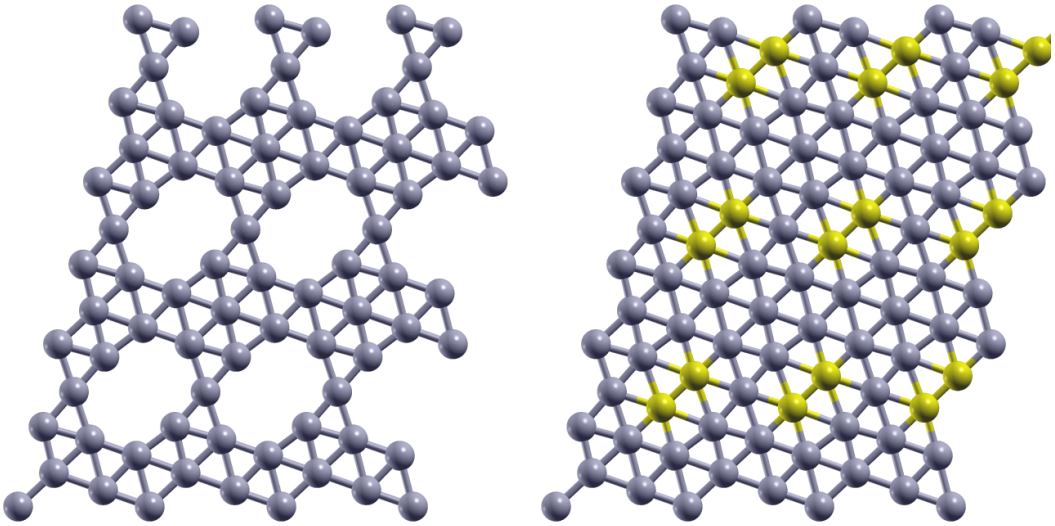
Fig. 6.16.: Boron sheet configurations mainly considered in literature: α-sheet (a), buckled triangular sheet (b) and distorted hexagonal one (c). Since a) and c) are flat, a top view was chosen.

6.3. Boron sheets – using EVOs special feature

As explained in section 4.4 a unique feature of EVO is the ability to search for layered structures in the framework of 3-dimensional periodic boundary conditions. The finding of carbon nanotubes that can be constructed from graphene [146] inspired the search for boron nanotubes and sheet structures. Though there is no experimental evidence for such sheets a set of different structures were proposed theoretically over the years and mainly comparative research was done.

Those boron sheets have been constructed based on graphene, using chemical understanding or relying on Boustanis 'Aufbau principle' [147]. The latter was derived from calculations of diverse boron clusters and states that the inclusion of hexagonal and pentagonal pyramidal units is highly favourable. Several sheets have been predicted and analysed with respect to their electronic structure and energetic stability [148, 149, 150, 151, 152, 153]. Lau and coworkers [151] conducted comparing research on six different sheet configurations: Of course, the graphene-like boron sheet and a flat triangular one (i.e. hexagonal 'holes' filled with one atom each) were considered but found to be unstable because calculations yielded imaginary phonons. In the main focus were the three sheets in figure 6.16 of which the α-sheet [152, 154] was found to be most stable concerning binding energies. More recently, the binding conditions were analysed more thoroughly and the 'hexagon holes' were determined to host excess electrons from the filled hexagons [155]. However, stability has not been proved certainly: For instance, a full phonon dispersion has been published for neither of the sheets. Moreover, there is so far no experimental evidence of their existence. Nevertheless, there are theoretical efforts to construct boron nanotubes from the sheets, see for example [156, 157]. Actual synthesis of single-walled boron nanotubes has been reported in 2004 [158] but since then no confirmation of their existence was published.

From the theoretical point of view, the impartial search for boron sheets may provide new insight into this field and yield new structures that are not that easily derived from graphene. This is the original reason EVO was enabled to search for sheet structures



(a) (b)

Fig. 6.17.: Boron sheet found by EVO: a) shows the actual structure while b) depicts the sheet when filling the hole with two atoms yielding a nearly perfect triangular structure.

as well. The changes and adaptions to the code are explained in section 4.4 and will not be covered here. Though, some results of EVO shall be presented with the special focus on a found structure containing an octagon hole.

The adapted evolutionary algorithm was started using different numbers of boron atoms per unit cell ranging from 2 to 12. This includes the numbers of atoms per unit cell of the of the sheets presented in figure 6.16 namely 8 for the α-sheet, 2 for the buckled triangular one and 4 for the distorted hexagonal sheet. A layer distance of 10 Å ensures vanishing interactions between the layers. It has to be stressed that the layer distance may not be too large since this prolongs the calculations while having no influence on the calculated energies. To get sheet instead of bulk structures the z-coordinates of atoms were restricted to lie between 0 and 0.1 thus taking up only 10 % of the cells in c-direction. By specifying this, there is no need of the sheets to be flat. There were no further constraints on the search.

Actually, two out of the three already predicted sheets were found during the search. However, far more interesting is a boron sheet found in a run started with 8 atoms per unit cell, see figure 6.17(a). It is build from boron triangles/filled hexagons and contains irregular octagon holes which can be built by removing a boron dimer. Filling the octagons (figure 6.17(b)) results in a nearly perfect triangular boron sheet. This is an analogy to the famous α-sheet which can be envisioned as a perfect triangular sheet with an atom removed. Moreover, the energy of the newly predicted sheet is only slightly higher (by 0.02 Ryd per atom) than that of the α-sheet meaning that it is energetically competitive. By thoroughly relaxing the sheet it was also verfied that it occupies a minimum in the energy landscape.

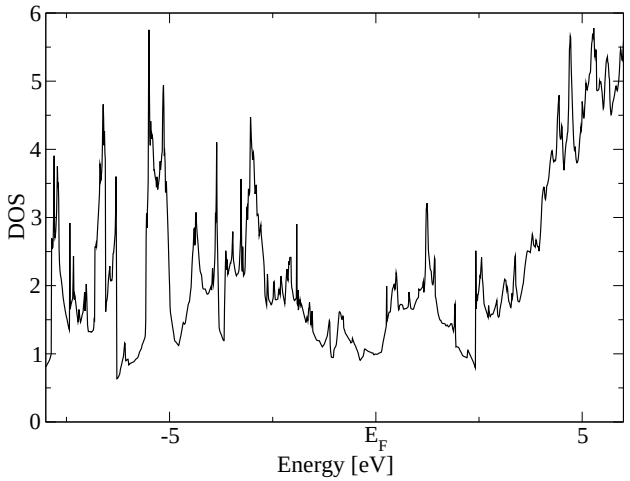
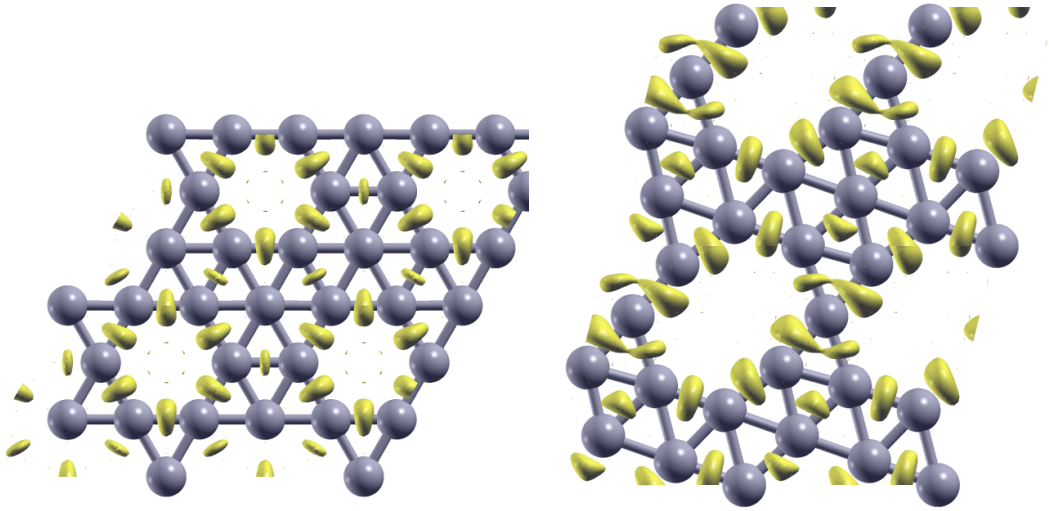


Fig. 6.18.: Electronic DOS of evo-sheet. Since there is no gap, the sheet can be considered to be metallic.

To get access to the electronic structure of the evo-sheet calculations of the electron DOS and the ELF were done. Figure 6.18 shows the electron density of states of the new sheet. It is metallic since there is no gap. This corresponds to the other boron sheets that have shown to be metallic by Özdoğan and coworkers [153]. However, since there are no special features in the DOS, the look at the electron localisation functions is more interesting. For comparison and illustration the 3-dimensional ELFs with an isovalue of 0.8 of the evo-sheet as well as the α-sheet are depicted in figure 6.19. According to [155] the hexagon holes in the α-sheet are necessary to take up excess electrons from the filled hexagons. The authors analyse the n-center 2-electron bonds (where $n >= 2$) of several fragments of the actual sheet[3]. In particular, they find a 7c-2e bond in each filled hexagon while the triangles at edges of the hexagon holes host a 3c-2e bonds each. Such detailled investigations were not done for the new sheet but ELF calculations reveal a similar behaviour for the evo- and α-sheet: The octagon holes host extra electrons. Naturally, as the evo-sheet breaks the all-hexagon structure, its ELF is not as symmetric as the one of the α-sheet.

Originally, one aim was not only finding new candidate structures but also verifying or disproving the stability assumption of the above mentioned boron sheet structures by means of the phonon dispersion. So far, the only examination of phonons in boron sheets was done by Lau and Pandey in 2008 [151] yet no phonon dispersion was published. We did extensive calculations on the phonon dispersion of the buckled triangular sheet since it has the smallest unit cell of the ones predicted so far. When calculating a phonon dispersion via DFPT with determination of the force constants and thus indirectly getting phonon frequencies at arbitrary q-points (cf. 2.1.3), some branch showed imaginary phonons. Of course, this would negate the stability assumption. However,

[3]The special tool in use is not available for extended structures.



(a) (b)

Fig. 6.19.: Electron localisation function with an isovalue of 0.8 of the two sheets containing holes: a) α-sheet and b) evo-sheet.

having a closer look at some of those q-points and calculating them directly with DFPT let imaginary phonons vanish. This contradiction made it impossible to judge the stability of the boron sheets regarding phonons throughout the Brillouin Zone. In order to solve this, further examination and probably a large amount of CPU time would be necessary.

To sum up, it was shown that EVO is also helpful in discovering new sheet structures in 3-dimensional periodic boundary conditions. It may introduce unknown and maybe unexpected features while finding already predicted sheets at the same time. Interestingly, the evo-sheet shows a similar electronic structure as the famous α-sheet. Yet, it was not possible to gain meaningful and unquestionable phonon dispersions which could be a topic of further research. Moreover, there is still room for more boron sheet structures and further search with EVO, e.g. using bigger unit cells, might give valuable results.

6.4. Challenges of the determination of stability and properties of candidate structures

The presented chapter showed some examples of successful search for new structures using EVO in different systems. However, it stressed that the determination of candidate structures is just a minor part of the whole process of crystal structure prediction. It was pointed out that there are various examinations that need to be done to confirm a structures stability apart from pure energetic considerations. Problems that can oc-

cur during this process were not concealed since this is by far not uncommon and may happen in many other situations. Yet, the efforts are justified by the great amount of information and insight that can be gained.

7. Conclusion

Computational materials design has evolved to be able to predict new crystal structures. For finding novelties, candidate structures have to be proposed in an unbiased way. There are several methods to deal with this task computationally. The approach of evolutionary algorithms has been chosen here which has matured to a very successful technique in the past years. An evolution strategy for crystal structure search was developed and released as the program EVO. It is available under the terms of the GNU GPL license and currently dependent only on software that is without charge and available for many platforms (Python and QUANTUMESPRESSO or GULP) so that it can be used by everyone.

For crystal structure search, the optimisation has to deal with a highly dimensional search space and an objective function that is costly to compute. Consequently, the algorithm has to minimise the number of evaluations of the objective function. Several key features of EVO turned out to be effective to reach this goal. First, using physical reasoning in the constraints of the candidate unit cells considerably confines the search space by avoiding evaluations of 'impossible' structures. Second, the conjunction of cell transformation and similarity test causes the discovery of different crystal structures within one search by increasing diversity of parents and offspring. This is successfully demonstrated in the searches exploring the carbon and silicon dioxide energy landscapes (see chapter 5). Repeated tests all yield the respective global minima and several local ones. The approach is also suitable for the search of supercells of the corresponding structures where the global minima graphite and α-quartz are reliably detected even though the search space has significantly higher dimension. The tests also confirm the reproducability of the results obtained with EVO.

A third key feature unique to EVO proved to be very valuable: Several dynamic termination criteria automatically end the search when it is considered to be unable find new favourable structures. This relieves the user from subjectively judging the progress of the search to determine an endpoint. As has been shown in section 5 there is no rule for the maximal number of generations the algorithm needs to converge. Thus, this static criterion used in other implementations is infeasible since the needed number of generations depends not only on the size of the system but also heavily on the actual energy landscape (see section 5.2). In addition, information is gained by analysing the termination criterion that applies in the regarded search. It is a measure for the quality of the search run: Just reaching a defined number of generations makes

no statement if the search got caught in a minimum, needs more generations or has been exhaustive. EVO independently rates the progress of the search and terminates when no further progress can be expected.

A further implication came to the fore when analysing the tests on the SiO_2 system with both DFT (section 5.2) and pair potentials (section 5.3) and enters more speculative grounds: The application of pair potentials seems to constrict the energy landscape compared to the case when the fitness is evaluated with DFT. Despite all parameters apart from the evaluation method being equal, all searches using GULP terminated before reaching the maximal number of generations. EVO regards the search to have been exhaustive. On the other hand, the searches with DFT evaluations did not terminate early. This implies that the energy landscape is changed by different evaluation methods which has to be taken into account especially when searching for unknown systems. Further investigation is necessary to validate this finding.

EVO has also been applied to search for several new structures. For crystal structure prediction the search procedure has to be complemented with an evaluation of the stability and various properties of the structures. This was successfully applied to predict a new carbon modification consisting of crossed graphene planes. These are connected by tetrahedrally bonded carbon on the crossing lines to form nearly rectangular pores. Since it combines the bonding characteristics of both, crossed graphene can be considered as a hybrid diamond-graphite like material. Intensive research regarding its properties reveals a strong anisotropy as well as analogies to graphene, especially when looking at mechanical and electronic properties. Furthermore, the challenging and costly task of the whole prediction process was covered by confirming the stability of the found structure.

In addition to crossed graphene, two other structures found with EVO were carefully examined with respect to their stability and properties. It was discovered that there is a stable phase in the germanium nitrofluoride system (GeNF) showing the applicability of the algorithm to ternary systems. Moreover, the introduction of pressure into the search procedure was demonstrated to be convenient in the GeNF case. The search for layered structures is a unique feature of EVO and boron sheets were chosen as an example for it. In this field, several structures had already been proposed using chemical understanding. Our search could find some of the known sheets along with a new one that reveals a previously unknown structural feature. Considering all the results of chapter 6, it can be stated that EVO enables to conveniently search for crystal structures. However, the searches need to be complemented by intensive further investigations that yield detailled insight into the stability and properties of the found structures.

Summarising all these findings, with EVO a practical tool for crystal structure search has been developed. EVO performs reliably for systems of various sizes and compositions. The automatic termination criterion provides a lot of additional information.

The complexity of the whole process of crystal structure prediction was demonstrated for several structures found with EVO. A thorough evaluation provides insights into the stability and properties of the found structures.

A. Appendix

A.1. Notation

N	number of atoms per unit cell
$\hat{H}$	Hamilton operator
E	energy
$\psi(\vec{r})$	wave function
$n(\vec{r})$	density
V	volume
σ	stress tensor
P	pressure
T	temperature
S	entropy
k_B	Boltzmann konstant
$I = (cell, atoms, age, fitness)$	individual
$P = (I_1, \ldots, I_\mu)$	population
μ	number of individuals in the population
λ	number of offspring
$\|x\|$	vector norm
$\{-1, 0, +1\}^{3\times3}$	3 × 3 matrix consisting of -1,0,1
$N(0, \sigma)$	normal distribution with expectation 0 and standard deviation σ
$\mathrm{diag}(x)$	vector of diagonal elements of matrix
$\mathrm{sort}(x)$	sort vector in ascending order
$\|x\|_\infty := max(\|x_1\|, \ldots, \|x_n\|)$	maximum norm
$x[0:n]$	first n elements of a given vector
$\mathbb{1}$	identity matrix
$\lceil x \rceil$	round x up to the next higher integer

A.2. Download and requirements

A.2.1. Requirements

Python interpreter version >= 2.6 is needed as well as a numpy version. (Available as a package of your Linux distribution or at `http://www.python.org` and `http://www.numpy.scipy.org`.)

For using EVO it is mandatory to have one of the following programs for electronic structure calculations installed and tested:

- QUANTUMESPRESSO (`http://www.quantum-espresso.org`)
- GULP (`http://projects.ivec.org/gulp`)

They are needed for the evaluation and rating of the crystal structures that are generated by EVO and are used for structure relaxation if specified. Moreover, it is strongly recommended to gain some experience using GULP/QUANTUMESPRESSO to be able to set proper input parameters for the calculations and to have the expertise in judging the results of EVO obtained using the particular program.

Moreover, it is recommended to install XCRYSDEN (available at `http://www.xcrysden.org`) which is an easy to use crystal visualisation program and is used in the postprocessing tool `visualise.py`.

All of the required programs are free of charge (in general or for academic use).

A.2.2. Download

EVO can be downloaded from the CPC library (`http://www.cpc.cs.qub.ac.uk/`) and has to be unpacked using the command:

```
tar xzf evo.tar.gz
```

A directory `evo` will be created that presently contains three directories that contain all needed files.

`python`	python files for EVO, code of the program
`doc`	documentation, description of EVO options
`test`	tests for EVO

EVO requires no installation because it is called using the python interpreter.

A.3. Usage of EVO

A.3.1. Starting EVO

Assuming that the file `lil_evo.py` is an executable (if not, just type `chmod +x lil_evo.py`) it can be easily called from the `evo/python` directory in the command line:

```
./lil_evo.py PATH_TO_SPECFILE
```

After a program abort EVO may be restarted from the previous complete generation:

```
./lil_evo.py PATH_TO_SPECFILE restart
```

A.3.2. Input files

The so called `specfile` contains all information that is needed to start EVO. More information on the `specfile` can be found in subsection A.3.2. Furthermore, two template files needed for the execution:

- a template input file for the electronic structure program used
- a template job file that is a simple shell script when doing local calculations or, if you want to use a queueing system, a jobfile

each containing several flags that are replaced by EVO. Examples of those files can be found in the `evo/test/`-directory. The specific flags that can be replaced by EVO are explained below.

Specfile

There are a lot of options that may be specified by the user in the `specfile`. To enable EVO an easy reading of the `specfile` some lines have to be present. Thus, a minimal input would look like, where the `atom_list` is needed to know what to do and `prog` is required for correct writing and reading of the input and output files, respectively:

```
CELL_SPECS
atom_list = Al
\
EVO_SPECS
\
COMP_SPECS
prog = pwscf
\
CALC_SPECS
```

The *_SPECS and the backslashes have to be present in the file yet some of the subsections don't need options. Every line starting with '#' is considered to be a comment and not read, blank lines are ignored as well. An overview of all options can be found in subsection A.3.5. Most of the parameters are optional and reasonable default values are set. Even for minimal input all used parameters are printed out in the header of the `parentfile`.

QUANTUMESPRESSO template file

A typical template file for QUANTUMESPRESSO is shown. Per default, it is expected as `sample.in` in the directory of the `specfile`. All constant input is explicitely specified, everything that needs to be filled by EVO is given as certain flags. The meaning of the flags is given below.

```
&control
calculation = 'vc-relax',
restart_mode = 'from_scratch',
prefix = '$PREFIX',
tstress = .true.,
tprnfor = .true.,
pseudo_dir = '$WORKDIR',
outdir = '/tmp/'
/
&system
ibrav =  0, celldm(1) = 1.88972612477, nat = $NAT, ntyp = $NTYP,
ecutwfc = 40.0,
ecutrho = 640.0,
occupations='smearing',
smearing='mp',
degauss=0.005
/
&electrons
conv_thr = 1.0d-6
/
&ions
/
&cell
cell_factor=4.d0
/
CELL_PARAMETERS
$CELLVECS

ATOMIC_SPECIES
Al 26.982 Al.pbe-n-van.UPF

ATOMIC_POSITIONS crystal
$ATOMS

K_POINTS automatic
$KPOINTS
```

In the following the flags to be replaced are shortly explained:

`$PREFIX`	is prepended to input/output filenames of `pw.x`; EVO chooses a prefix that is unique avoiding conflicts if calculations use the same directory for temporary files
`$NAT`	number of atoms in the unit cell
`$NTYP`	number of types of atoms in the unit cell
`$CELLVECS`	crystal lattice vectors (are generated in Angstrom, make sure that celldm(1) = 1.88972612477 and ibrav = 0)
`$ATOMS`	atomic positions
`$KPOINTS`	k-points for the unit cell; dimensions of the automatically generated grid are determined by EVO for each unit cell

GULP template file

A typical template file for GULP is shown. Per default, it is expected as `sample.in` in the directory of the `specfile`. All constant input is explicitely specified, everything that needs to be filled by EVO is given as certain flags. The meaning of the flags is given below.

```
opti conp
cell
$CELLPARAMS
frac
$ATOMS
library $WORKDIR/clerirosato
```

In the following the flags to be replaced are shortly explained:

`$CELLPARAMS`	crystal lattice parameters
`$ATOMS`	atomic positions
`$WORKDIR`	working directory (where to find GULP library file)

Template job files

In the jobfiles the following flags can be replaced if necessary. More information on the parallelisation of the fitness evaluations can be found in subsection A.3.3.

`$WORKDIR`	working directory (in `sample.queue.job` the queueing system output is directed there)
`$INPUTFILE`	input file for the calculation
`$OUTPUTFILE`	output file for the calculation
`$PREFIX`	is prepended to input/output filenames of `pw.x`; used for cleaning up the temporary directory

K-point sampling for DFT calculations

As k-point sampling is important for accurate results and at the same time dependent on the form of the unit cell, the k-grid is determined for each individual separately. It

depends only on the length of the cell vectors and a given 'kpoint_resolution' k_{res} [77] and the result is rounded to the next higher integer:

$$k_i = \left\lceil \frac{1}{k_{res} * \|vec_i\|} \right\rceil \qquad \forall i \in 1, 2, 3$$

Output files

EVO produces two main output files. As default `out.evo` and `parents.evo` are created in the working directory. If nothing else is specified, this is where the specfile is residing. `out.evo` contains all status and error messages that are produced while EVO is running. It can be used for monitoring the progress of EVO and for debugging (set `verbose=True` in specfile). `parents.evo` contains all information that is needed to restart a calculation or to analyse the results of EVO. The header consists of the options that were used for this specific run of EVO and is followed by the printout of the individuals of all generations. At the end a list of the best individuals of a whole run is printed out.

Apart from that EVO creates the directory `inp_out_evo` in the working directory which contains all information that is needed during the run, e.g. concrete input, output and job files. It is cleaned up after a successful run of EVO and kept if something went wrong.

Postprocessing

For further analysis of the results of EVO two small postprocessing tools available. The first one is used for simple visualisation of a certain generation of crystal structures while the second one can be used to produce an input file for the used electronic structure program of a particular individual. The following list shows the command which invokes the respective tool and a short description. The commands assume that you are in the `evo/python`-directory.

- (python) visualise.py PARENTFILE GENERATION
  The tool produces an input file that is read by XCrysDen reading for it the specified parentfile (path to it) and extracting the wanted generation. The generation is identified by its number. ('best_of_all' is regarded as a number in this context.) The resulting axsf-file to the directory of the PARENTFILE.

- (python) generate_input.py PARENTFILE GENERATION INDIVIDUAL
  The tool fills the data of the INDIVDUAL (number) of the specified GENERATION (number) read from the PARENTFILE (path) in the `sample_input` file. The resulting input file is written to the directory containing the PARENTFILE.

A.3.3. Parallelisation

The fitness evaluations of the individuals are easily parallelised since they are completely independent. There are two possibilities for that: One can use a so-called 'local queue' which applies when working on a single workstation having one or more CPUs. This is feasible for serially running GULP calculations, for example. EVO then distributes the calculations according to the number of available processors. The important options for local GULP calculations that have to be specified in the COMP_SPEC card of `specfile`:

```
prog = gulp
schedule = local
parallel_processes = [NR_OF_FREE/AVAILABLE_PROCESSORS]
```

For big systems and when using DFT (QUANTUMESPRESSO) it is strongly recommended to interface EVO with a compute cluster. In this case, specify `schedule = queue`. Since EVO is working completely by itself it needs to interact with the queueing system installed on your cluster. In detail, the algorithm submits a job, memorises the job id and checks the queue after some period if all memorised jobs have finished. Since the commands to be used and especially the output of them differ according to the queueing system in use, one has to customise a small routine of EVO. `evo/python/queue.py` has to have a structure as shown below since EVO calls the routines of the Queue-Class `submit_job` and `get_jobids`):

```
class Queue()
        def __init__(self):
                pass
        def submit_job(self, jobfile):
                ### instructions to submit job and extract its id ###
                return job_id
        def get_jobids(self):
                ### instructions to get job ids from queueing system ###
                ### pass them over to program as a list ###
                return queued_jobs
```

Two examples (that work on the associated cluster) with lots of comments can be found in `evo/python/theo_queue.py` and `evo/python/adde_queue.py`. It needs just a little experience in Python (or scripting) to adapt the Python file to your needs. (The corresponding output structure can be found in the comments at the end of the file.) By comparing both examples it should be possible also for less experienced users to interface their queueing system. Acutally, `evo/python/queue.py` is the file called by the scheduler. Per default, queue.py is linked to theo_queue.py so that one can have different versions of the queue file in one directory and just do a "ln -s" on the according cluster.

A.3.4. Structure of the code

The structure of EVO is quite easy to understand since the main routine `lil_evo.py` calls automatically all the other parts of the code that are needed for its correct execution. The following list gives an overview on all parts of EVO and their function.

`lil_evo.py`	main program, does everything evolutionary and calls everything else needed for proper execution
`evo_inp_out.py`	responsible for reading `specfile` and writing to and reading all information from `parentfile` (header and generations)
`define_indiv.py`	incorporates the general definition of an individual (crystal structure)
`utils.py`	holds some utility functions called from troughout the program
`handle_except.py`	defines special EVO exceptions
`fit.py`	receives unevaluated individuals and takes care for fitness calculations
`gulp_handler.py`	covers all things that are specific for GULP, writing input, reading output
`pwscf_handler.py`	covers all things that are specific for QuantumESPRESSO, writing input, reading output
`schedule.py`	schedules calculations, either locally or via a queueing system
`queue.py`	interface to queueing system
`adde_queue.py`	example for an interface to another queueing system
`postprocessing.py`	provides functions for postprocessing, utilising WIEN2k's sgroup or XCrysDen
`generate_input.py`	callable, reads individuals from `parentfile` and generate input file for calculation
`visualise.py`	callable, produces XCrysDen input file and opens it

A.3.5. Available options

Section `CELL_SPECS`

Variable	`atom_list`
Type	list of string
Required	yes
Description	list of atom symbols to use in EVO, atomic symbols as present in the Periodic System of Elements first uppercase letter, following lowercase: `'Al'` (not `'AL'` or `'al'`) for aluminium

Variable	`type`
Type	string
Default	`'bulk'`
Description	discriminate between bulk and layered structures: • 'bulk' • 'layer'

Variable	`lengths`
Type	3-list of float ('None' as string allowed)
Default	`None, None, None`
Description	can be specified if one knows cell constants of some material (in Angstrom); also allowed is partial information: lengths = 3.1, None, 4.5

Variable	`angles`
Type	3-list of float ('None' as string allowed)
Default	`None, None, None`
Description	can be specified if one knows cell angles of some material (in Degrees); also allowed is partial information: angles = 90, None, None

Variable	`volume_limits`
Type	2-list of float ('None' as string allowed)
Default	`None, None`
Description	can be specified to limit volume of the generated cell; also allowed is partial information: volume_limits = 105.0, None

Variable	`transform_cell`
Type	bool
Default	`True`
Description	if `True` cells are transformed to shortest possible cell vectors / angles $\sim 90°$ (helps comparability of the cells to identify too similar ones); otherwise specify `False` (*not* recommended).

Variable	`density`
Type	float
Default	`None`
Description	to be specified if one has some hint of possible density of material; density of generated structures may be +/- 20 %; if not specified close-packing is assumed as default density HAS TO BE GIVEN IN g/cm^3

Variable	`layer_height`
Type	float
Default	`None`, if type = layer: `1` (Angstrom)
Description	highest possible height of the generated layer

Variable	`layer_spacing`
Type	float
Default	`None`, if type = layer: `10` (Angstrom)
Description	spacing for layers in Angstrom

Section `EVO_SPECS`

Variable	`my`
Type	integer
Default	`10`
Description	number of individuals per generation

Variable	`lambda`
Type	integer
Default	`30`
Description	number of offspring; 'my' of them are chosen to form the next generation

Variable	`number_of_generations`
Type	integer
Default	`30`
Description	number of generations after that code terminates the calculation

Variable	`number_of_parents`
Type	integer
Default	`2`
Description	number of parent individuals to derive a child

Variable	`maximal_age`
Type	integer
Default	`5`
Description	individuals with superior fitness have to die when having survived this number of generations enhances diversity

Variable	`min_diff_offspr`
Type	integer
Default	`3`
Description	minimal number of dissimilar offspring (if lower, to low diversity, aborting)

Variable	`mutation_std_dev`
Type	float
Default	`0.0`
Description	standard deviation of normal distribution of mutation; if neglecting mutation (e.g. when using full cell relaxation), let this be 0 - otherwise it scales the quantity of the mutation step

Section `CALC_SPECS`

Variable	`kpoint_resolution`
Type	float
Default	`0.1`
Description	resolution of k-grid in reciprocal space (1/(resolution*|lattice_vector|) rounded up)

Section `COMP_SPECS`

Variable	`workdir`
Type	string
Default	folder where specfile is residing
Description	working directory, where put standard output, parent file, etc.

Variable	`sample_input`
Type	string
Default	`sample.in` in workdir
Description	sample input file containing the needed $FLAGS

Variable	`sample_job`
Type	string
Default	`sample.job` in workdir
Description	sample job file containing the needed $FLAGS

Variable `prog`
Type string
Required yes
Description specify the program to be used for fitness calculation; currently implemented:

- 'pwscf' (v. 4.3.1/4.3.2)
- 'gulp' (v. 3.1/4.0)

Variable `schedule`
Type string
Default `'queue'`
Description type of scheduler used, currently working:

- 'queue'
- 'local'

Variable `parallel_processes`
Type int
Default `1`
Description number of calculation processes to be started when doing local calculations esp. useful for running serial program on a multi-processor machine (for parallelisation)

Variable `save_parents`
Type string
Default `'parents.evo'`
Description file which saves all important data from EVO all options, parents ...

Variable `std_out`
Type string
Default `'out.evo'`
Description records the ongoing of EVO, error messages, etc.

Variable `verbose`
Type bool
Default `False`
Description if True more output is generated, esp. exceptions are printed, useful for debugging etc.

Bibliography

[1] R. M. Martin. *Electronic structure: Basic theory and practical methods.* Cambridge University Press, 2004.

[2] P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Phys. Rev.*, 136: B864–B871, 1964.

[3] W. Kohn and L. J. Sham. Self-consistent equations including exchange and correlation effects. *Phys. Rev.*, 140:A1133–A1138, 1965.

[4] A. Savin, R. Nesper, S. Wengert, and T. F. Fässler. ELF: The electron localization function. *Angew. Chem. Int. Ed.*, 36:1808–1832, 1997.

[5] O. H. Nielsen and R. M. Martin. First-principles calculation of stress. *Phys. Rev. Lett.*, 50:697–700, 1983.

[6] O. H. Nielsen and R. M. Martin. Quantum-mechanical theory of stress and force. *Phys. Rev. B*, 32:3780–3791, 1985.

[7] R. Golesorkhtabar, P. Pavone, J. Spitaler, P. Puschnig, and C. Draxl. ElaStic: A tool for calculating second-order elastic constants from first principles. *Comput. Phys. Commun.*, 184:1861 – 1873, 2013.

[8] M. T. Dove. *Structure and dynamics: an atomic view of materials.* Oxford Univ. Press, 2010.

[9] X. Gonze and C. Lee. Dynamical matrices, Born effective charges, dielectric permittivity tensors, and interatomic force constants from density-functional perturbation theory. *Phys. Rev. B*, 55:10355–10368, 1997.

[10] S. Baroni, S. de Gironcoli, A. Dal Corso, and P. Giannozzi. Phonons and related crystal properties from density-functional perturbation theory. *Rev. Mod. Phys.*, 73:515–562, 2001.

[11] M. Lazzeri and F. Mauri. First-principles calculation of vibrational raman spectra in large systems: Signature of small rings in crystalline SiO_2. *Phys. Rev. Lett.*, 90:036401, 2003.

[12] P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. L. Chiarotti, M. Cococcioni, I. Dabo, A. Dal Corso, S. de Gironcoli, S. Fabris, G. Fratesi, R. Gebauer, U. Gerstmann, C. Gougoussis, A. Kokalj, M. Lazzeri, L. Martin-Samos, N. Marzari, F. Mauri, R. Mazzarello, S. Paolini, A. Pasquarello, L. Paulatto, C. Sbraccia, S. Scandolo, G. Sclauzero, A. P. Seitsonen, A. Smogunov, P. Umari, and R. M. Wentzcovitch. QUANTUMESPRESSO: a modular and open-source software project for quantum simulations of materials. *J. Phys. Condens. Matter*, 21:395502, 2009.

[13] QUANTUMESPRESSO website, 2013. URL `http://www.quantum-espresso.org`.

[14] K. Schwarz and P. Blaha. Solid state calculations using WIEN2k. *Comput. Mater. Sci.*, 28:259 – 273, 2003.

[15] WIEN2K website, 2013. URL `http://www.wien2k.at`.

[16] K. Parlinski, Z.-Q. Li, and Y. Kawazoe. First-principles determination of the soft mode in cubic ZrO_2. *Phys. Rev. Lett.*, 78:4063–4066, 1997.

[17] A. Togo, F. Oba, and I. Tanaka. First-principles calculations of the ferroelastic transition between rutile-type and $CaCl_2$-type SiO_2 at high pressures. *Phys. Rev. B*, 78:134106, 2008.

[18] J. D. Gale and A. L. Rohl. The general utility lattice program (gulp). *Mol. Simul.*, 29:291, 2003.

[19] GULP website, 2013. URL `http://projects.ivec.org/gulp/`.

[20] A. Kokalj. XCrySDen—a new program for displaying crystalline structures and electron densities. *J. Mol. Graphics Modell.*, 17:176 – 179, 1999.

[21] A. Kokalj. Computer graphics and graphical user interfaces as tools in simulations of matter at the atomic scale. *Comput. Mater. Sci.*, 28:155 – 168, 2003.

[22] XCRYSDEN website, 2013. URL `http://www.xcrysden.org/`.

[23] K. Momma and F. Izumi. *VESTA3* for three-dimensional visualization of crystal, volumetric and morphology data. *J. Appl. Crystallogr.*, 44:1272–1276, 2011.

[24] VESTA website, 2013. URL `http://jp-minerals.org/vesta/en/`.

[25] J. Maddox. Crystals from first principles. *Nature*, 335:201, 1988.

[26] S. Baroni, P. Giannozzi, and E. Isaev. Density-functional perturbation theory for quasi-harmonic calculations. *Rev. Mineral. Geochem.*, 71:39–57, 2010.

[27] M. R. Hestenes and E. Stiefel. Methods of conjugate gradients for solving linear systems. *J. Res. Nat. Bur. Stand.*, 49:409–436, 1952.

[28] C. G. Broyden. The convergence of a class of double-rank minimization algorithms 1. general considerations. *IMA J. Appl. Math.*, 6:76–90, 1970.

[29] R. Fletcher. A new approach to variable metric algorithms. *Comput. J.*, 13: 317–322, 1970.

[30] D. Goldfarb. A family of variable-metric methods derived by variational means. *Math. Comput.*, 24:23–26, 1970.

[31] D. F. Shanno. Conditioning of quasi-newton methods for function minimization. *Math. Comput.*, 24:647–656, 1970.

[32] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, editors. *Numerical recipes in FORTRAN.* Cambridge Univ. Press, Cambridge, 2. ed. edition, 1992.

[33] N. Metropolis and S. Ulam. The monte carlo method. *J. Am. Stat. Assoc.*, 44: 335–341, 1949.

[34] A. R. Oganov, editor. *Modern Methods of Crystal Structure Prediction.* Wiley-VCH, 2010.

[35] C. J. Pickard and R. J. Needs. Ab initio random structure searching. *J. Phys. Condens. Matter*, 23:053201, 2011.

[36] C. J. Pickard and R. J. Needs. Structure of phase iii of solid hydrogen. *Nat. Phys.*, 3:473, 2007.

[37] C. J. Pickard and R. J. Needs. High-pressure phases of nitrogen. *Phys. Rev. Lett.*, 102:125702, 2009.

[38] J. Feng, R. G. Hennig, N. W. Ashcroft, and R. Hoffmann. Emergent reduction of electronic state dimensionality in dense ordered Li-Be alloys. *Nature*, 451:445, 2008.

[39] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. Optimization by simulated annealing. *Science*, 220:671–680, 1983.

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

[41] J.C. Schön and M. Jansen. Determination of candidate structures for simple ionic compounds through cell optimisation. *Comput. Mater. Sci.*, 4:43 – 58, 1995.

[42] J. C. Schön and M. Jansen. First step towards planning of syntheses in solid-state chemistry: Determination of promising structure candidates by global optimization. *Angew. Chem. Int. Ed.*, 35:1286–1304, 1996.

[43] J. Pillardy, Y. A. Arnautova, C. Czaplewski, K. D. Gibson, and H. A. Scheraga. Conformation-family monte carlo: A new method for crystal structure prediction. *Proc. Natl. Acad. Sci. U.S.A.*, 98:12351–12356, 2001.

[44] M. Jansen and J. C. Schön. Strukturkandidaten für Alkalimetallnitride. *Z. Anorg. Allg. Chem.*, 624:533–540, 1998.

[45] K. Doll, J. C. Schön, and M. Jansen. Structure prediction based on *ab initio* simulated annealing for boron nitride. *Phys. Rev. B*, 78:144110, 2008.

[46] D. J. Wales and J. P. K. Doye. Global optimization by basin-hopping and the lowest energy structures of Lennard-Jones clusters containing up to 110 atoms. *J. Phys. Chem. A*, 101:5111–5116, 1997.

[47] D. Cvijović and J. Klinowski. Taboo search: An approach to the multiple minima problem. *Science*, 267:664–666, 1995.

[48] S. Goedecker. Minima hopping: An efficient search method for the global minimum of the potential energy surface of complex molecular systems. *J. Chem. Phys.*, 120:9911–9917, 2004.

[49] J. P. K. Doye, M. A. Miller, and D. J. Wales. The double-funnel energy landscape of the 38-atom Lennard-Jones cluster. *J. Chem. Phys.*, 110:6896–6906, 1999.

[50] J. P. K. Doye and D. J. Wales. Global minima for transition metal clusters described by Sutton-Chen potentials. *New J. Chem.*, 22:733–744, 1998.

[51] F. Hoffmann and B. Strodel. Protein structure prediction using global optimization by basin-hopping with NMR shift restraints. *J. Chem. Phys.*, 138:025102, 2013.

[52] S. Goedecker, W. Hellmann, and T. Lenosky. Global minimum determination of the Born-Oppenheimer surface within density functional theory. *Phys. Rev. Lett.*, 95:055501, 2005.

[53] W. Hellmann, R. G. Hennig, S. Goedecker, C. J. Umrigar, B. Delley, and T. Lenosky. Questioning the existence of a unique ground-state structure for Si clusters. *Phys. Rev. B*, 75:085411, 2007.

[54] K. Bao, S. Goedecker, K. Koga, F. Lançon, and A. Neelov. Structure of large gold clusters obtained by global optimization using the minima hopping method. *Phys. Rev. B*, 79:041405, 2009.

[55] S. Roy, S. Goedecker, M. J. Field, and E. Penev. A minima hopping study of all-atom protein folding and structure prediction. *J. Phys. Chem. B*, 113:7315–7321, 2009.

[56] M. Amsler, J. A. Flores-Livas, L. Lehtovaara, F. Balima, S. A. Ghasemi, D. Machon, S. Pailhès, A. Willand, D. Caliste, S. Botti, A. San Miguel, S. Goedecker, and M. A. L. Marques. Crystal structure of cold compressed graphite. *Phys. Rev. Lett.*, 108:065501, 2012.

[57] R. Martoňák, A. Laio, and M. Parrinello. Predicting crystal structures: The Parrinello-Rahman method revisited. *Phys. Rev. Lett.*, 90:075503, 2003.

[58] R. Martoňák, D. Donadio, A. R. Oganov, and M. Parrinello. Crystal structure transformations in SiO_2 from classical and ab initio metadynamics. *Nat. Mater.*, 5:623–626, 2006.

[59] M. Parrinello and A. Rahman. Crystal structure and pair potentials: A molecular-dynamics study. *Phys. Rev. Lett.*, 45:1196–1199, 1980.

[60] R. Martoňák, A. R. Oganov, and C. W. Glass. Crystal structure prediction and simulations of structural transformations: metadynamics and evolutionary algorithms. *Phase Transitions*, 80:277–298, 2007.

[61] S. Curtarolo, D. Morgan, K. Persson, J. Rodgers, and G. Ceder. Predicting crystal structures with data mining of quantum calculations. *Phys. Rev. Lett.*, 91:135503, 2003.

[62] C. C. Fischer, K. J. Tibbetts, D. Morgan, and G. Ceder. Predicting crystal structure by merging data mining with quantum mechanics. *Nat. Mater.*, 5: 641–646, 2006.

[63] R. Eberhart and J. Kennedy. A new optimizer using particle swarm theory. In *Micro Machine and Human Science, 1995. MHS '95., Proceedings of the Sixth International Symposium on*, pages 39–43, 1995.

[64] Y. Wang, J. Lv, L. Zhu, and Y. Ma. Crystal structure prediction via particle-swarm optimization. *Phys. Rev. B*, 82:094116, 2010.

[65] Y. Wang, J. Lv, L. Zhu, and Y. Ma. CALYPSO: A method for crystal structure prediction. *Comput. Phys. Commun.*, 183:2063 – 2070, 2012.

[66] J. Lv, Y. Wang, L. Zhu, and Y. Ma. Predicted novel high-pressure phases of lithium. *Phys. Rev. Lett.*, 106:015503, 2011.

[67] L. Zhu, H. Wang, Y. Wang, J. Lv, Y. Ma, Q. Cui, Y. Ma, and G. Zou. Substitutional alloy of Bi and Te at high pressure. *Phys. Rev. Lett.*, 106:145501, 2011.

[68] C. Darwin and A. Wallace. On the tendency of species to form varieties; and on the perpetuation of varieties and species by natural means of selection. *Journal of the Proceedings of the Linnean Society of London. Zoology*, 3(9):45–62, 1858.

[69] U.S. department of energy genomic science program website, 2013. URL `http://genomicscience.energy.gov`.

[70] K. Pluck. The king, 2013. URL `http://commons.wikimedia.org/wiki/File%3ALion_waiting_in_Namibia.jpg`.

[71] J. H. Holland. *Adaptation in Natural and Artificial Systems: An Introductory Analysis with Applications to Biology, Control and Artificial Intelligence.* MIT Press, Cambridge, MA, USA, 1992. ISBN 0262082136.

[72] H.-P. Schwefel and G. Rudolph. Contemporary evolution strategies. In Federico Morán, Alvaro Moreno, Juan Merelo, and Pablo Chacón, editors, *Advances in Artificial Life*, volume 929 of *Lecture Notes in Computer Science*, pages 891–907. Springer Berlin / Heidelberg, 1995. ISBN 978-3-540-59496-3.

[73] T. Bäck. *Evolutionary Algorithms in Theory and Practice.* Oxford University Press, 1996. ISBN 0-19-509971-0.

[74] V. Nissen. *Einführung in Evolutionäre Algorithmen: Optimierung nach dem Vorbild der Evolution.* Computational intelligence. Vieweg, 1997. ISBN 9783528054991.

[75] S. Bahmann and J. Kortus. EVO — evolutionary algorithm for crystal structure prediction. *Comput. Phys. Commun.*, 184(6):1618 – 1625, 2013.

[76] A. R. Oganov and C. W. Glass. Crystal structure prediction using ab initio evolutionary techniques: Principles and applications. *Journal of Chemical Physics*, 124(24):244704, 2006. ISSN 00219606.

[77] C. W. Glass, A. R. Oganov, and N. Hansen. USPEX — evolutionary crystal structure prediction. *Comput. Phys. Commun.*, 175(11-12):713 – 720, 2006. ISSN 0010-4655.

[78] A. R. Oganov, J. Chen, C. Gatti, Y. Ma, Y. Ma, C. W. Glass, Z.n Liu, T. Yu, O. O. Kurakevych, and V. L. Solozhenko. Ionic high-pressure form of elemental boron. *Nature*, 457:863–867, 2009.

[79] Y. Ma, M. Eremets, A. R. Oganov, Y. Xie, I. Trojan, S. Medvedev, A. O. Lyakhov, M. Valle, and V. Prakapenka. Transparent dense sodium. *Nature*, 458:182–185, 2009.

[80] D. C. Lonie and E. Zurek. XtalOpt: An open-source evolutionary algorithm for crystal structure prediction. *Comput. Phys. Commun.*, 182:372 – 387, 2011.

[81] G. Trimarchi and A. Zunger. Global space-group optimization problem: Finding the stablest crystal structure without constraints. *Phys. Rev. B*, 75:104113, 2007.

[82] S.M. Woodley and C.R.A. Catlow. Structure prediction of titania phases: Implementation of darwinian versus lamarckian concepts in an evolutionary algorithm. *Comput. Mater. Sci.*, 45:84 – 95, 2009.

[83] D. M. Deaven and K. M. Ho. Molecular geometry optimization with a genetic algorithm. *Phys. Rev. Lett.*, 75:288–291, 1995.

[84] T. Bäck, U. Hammel, and H.-P. Schwefel. Evolutionary computation: comments on the history and current state. *IEEE Trans. Evol. Comput.*, 1:3–17, 1997.

[85] M. Valle and A. 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.

[86] A. R. Oganov and M. Valle. How to quantify energy landscapes of solids. *J. Chem. Phys.*, 130:104504, 2009.

[87] M. Valle and A. R. Oganov. Crystal fingerprint space — a novel paradigm for studying crystal-structure sets. *Acta Crystallogr. Sect. A*, 66:507–517, 2010.

[88] E. L. Willighagen, R. Wehrens, P. Verwer, R. de Gelder, and L. M. C. Buydens. Method for the computational comparison of crystal structures. *Acta Crystallogr. Sect. B*, 61:29–36, 2005.

[89] D. C. Lonie and E. Zurek. Identifying duplicate crystal structures: XtalComp, an open-source solution. *Comput. Phys. Commun.*, 183:690 – 697, 2012.

[90] J. A. Chisholm and S. Motherwell. COMPACK: a program for identifying crystal structure similarity using distances. *J. Appl. Crystallogr.*, 38:228–231, 2005.

[91] T. S. Bush, C. R. A. Catlow, and P. D. Battle. Evolutionary programming techniques for predicting inorganic crystal structures. *J. Mater. Chem.*, 5:1269–1272, 1995.

[92] S. M. Woodley, P. D. Battle, J. D. Gale, and R. A. C. Catlow. The prediction of inorganic crystal structures using a genetic algorithm and energy minimisation. *Phys. Chem. Chem. Phys.*, 1:2535–2542, 1999.

[93] N. L. Abraham and M. I. J. Probert. A periodic genetic algorithm with real-space representation for crystal structure and polymorph prediction. *Phys. Rev. B*, 73: 224104, 2006.

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

[95] G. Trimarchi, A. J. Freeman, and A. Zunger. Predicting stable stoichiometries of compounds via evolutionary global space-group optimization. *Phys. Rev. B*, 80:092101, 2009.

[96] X. Zhang, A. Zunger, and G. Trimarchi. Structure prediction and targeted synthesis: A new Na_nN_2 diazenide crystalline structure. *J. Chem. Phys.*, 133:194504, 2010.

[97] 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. *Phys. Rev. Lett.*, 105:217003, 2010.

[98] A. N. Kolmogorov, S. Shah, E. R. Margine, A. K. Kleppe, and A. P. Jephcoat. Pressure-driven evolution of the covalent network in CaB_6. *Phys. Rev. Lett.*, 109: 075501, 2012.

[99] Module for ab initio structure evolution - webpage, 2013. URL `http://bingweb.binghamton.edu/~akolmogo/maise/`.

[100] 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. *Phys. Rev. B*, 83:104106, 2011.

[101] Q. Li, Y. Ma, A. R. Oganov, H. Wang, H. Wang, Y. Xu, T. Cui, H.-K. Mao, and G. Zou. Superhard monoclinic polymorph of carbon. *Phys. Rev. Lett.*, 102: 175506, 2009.

[102] A. R. Oganov and C. W. Glass. Evolutionary crystal structure prediction as a tool in materials design. *J. Phys. Condens. Matter*, 20:064210, 2008.

[103] A. O. Lyakhov, A. R. Oganov, and M. Valle. How to predict very large and complex crystal structures. *Comput. Phys. Commun.*, 181:1623 – 1632, 2010.

[104] A. O. Lyakhov, A. R. Oganov, H. T. Stokes, and Q. Zhu. New developments in evolutionary structure prediction algorithm USPEX. *Comput. Phys. Commun.*, 184:1172 – 1182, 2013.

[105] K. Li, X. Wang, F. Zhang, and D. Xue. Electronegativity identification of novel superhard materials. *Phys. Rev. Lett.*, 100:235504, 2008.

[106] A. R. Oganov, A. O. Lyakhov, and M. Valle. How evolutionary crystal structure prediction works — and why. *Acc. Chem. Res.*, 44:227–237, 2011.

[107] CPC program library, 2013. URL `http://cpc.cs.qub.ac.uk`.

[108] B. Bandow and B. Hartke. Larger water clusters with edges and corners on their way to ice: Structural trends elucidated with an improved parallel evolutionary algorithm. *J. Phys. Chem. A*, 110:5809–5822, 2006.

[109] A. Kuc and G. Seifert. Hexagon-preserving carbon foams: Properties of hypothetical carbon allotropes. *Phys. Rev. B*, 74:214104, 2006.

[110] F. J. Ribeiro, S. G. Louie, M. L. Cohen, and P. Tangney. Structural and electronic properties of carbon in hybrid diamond-graphite structures. *Phys. Rev. B*, 72: 214109, 2005.

[111] T Demuth, Y. Jeanvoine, J. Hafner, and J. G. Ángyán. Polymorphism in silica studied in the local density and generalized-gradient approximations. *J. Phys. Condens. Matter*, 11:3833, 1999.

[112] A. Belsky, M. Hellenbrandt, V. L. Karen, and P. Luksch. New developments in the Inorganic Crystal Structure Database (ICSD): accessibility in support of materials research and design. *Acta Crystallogr. Sect. B*, 58:364–369, 2002.

[113] C. Loose and J. Kortus. Systematic study of the influence of different equations of states on the calculation of elastic properties. *High Pressure Res.* doi: 10.1080/08957959.2013.806657.

[114] The Nobel Prize in physics 2010, Aug 2013. URL `http://www.nobelprize.org/nobel_prizes/physics/laureates/2010/index.html`.

[115] A. K. Geim and K. S. Novoselov. The rise of graphene. *Nat. Mater.*, 6:183–191, 2007.

[116] A. Y. Liu, M. L. Cohen, K. C. Hass, and M. A. Tamor. Structural properties of a three-dimensional all-sp^2 phase of carbon. *Phys. Rev. B*, 43:6742–6745, 1991.

[117] C. Mailhiot and A. K. McMahan. Atmospheric-pressure stability of energetic phases of carbon. *Phys. Rev. B*, 44:11578–11591, 1991.

[118] I. V. Stankevich, M. V. Nikerov, and D. A. Bochvar. The structural chemistry of crystalline carbon: Geometry, stability, and electronic spectrum. *Russ. Chem. Rev.*, 53:640, 1984.

[119] R. Hoffmann, T. Hughbanks, M. Kertesz, and P. H. Bird. Hypothetical metallic allotrope of carbon. *J. Am. Chem. Soc.*, 105:4831–4832, 1983.

[120] R. L. Johnston and R. Hoffmann. Superdense carbon, C_8: supercubane or analog of γ-silicon? *J. Am. Chem. Soc.*, 111:810–819, 1989.

[121] K. M. Merz, R. Hoffmann, and A. T. Balaban. 3,4-connected carbon nets: through-space and through-bond interactions in the solid state. *J. Am. Chem. Soc.*, 109:6742–6751, 1987.

[122] H. R. Karfunkel and T. Dressler. New hypothetical carbon allotropes of remarkable stability estimated by MNDO solid-state SCF computations. *J. Am. Chem. Soc.*, 114:2285–2288, 1992.

[123] K. Umemoto, S. Saito, S. Berber, and D. Tománek. Carbon foam: Spanning the phase space between graphite and diamond. *Phys. Rev. B*, 64:193409, 2001.

[124] L. Wang, B. Liu, H. Li, W. Yang, Y. Ding, S. V. Sinogeikin, Y. Meng, Z. Liu, X. C. Zeng, and W. L. Mao. Long-range ordered carbon clusters: A crystalline material with amorphous building blocks. *Science*, 337:825–828, 2012.

[125] Q. Li, Y. Ma, A. R. Oganov, H. Wang, H. Wang, Y. Xu, T. Cui, H.-K. Mao, and G. Zou. Superhard monoclinic polymorph of carbon. *Phys. Rev. Lett.*, 102: 175506, 2009.

[126] S. E. Boulfelfel, A. R. Oganov, and S. Leoni. Understanding the nature of "superhard graphite". *Sci. Rep.*, 2:471, 2012.

[127] W. L. Mao, H.-K. Mao, P. J. Eng, T. P. Trainor, M. Newville, C.-C. Kao, D. L. Heinz, J. Shu, Y. Meng, and R. J. Hemley. Bonding changes in compressed superhard graphite. *Science*, 302:425–427, 2003.

[128] S. Bahmann, T. Weißbach, and J. Kortus. Crossed graphene: Stability and electronic structure. *Phys. Status Solidi RRL*, 7:639–642, 2013.

[129] R. W. Grosse-Kunstleve. Algorithms for deriving crystallographic space-group information. *Acta Crystallogr. Sect. A*, 55:383–395, 1999.

[130] S. R. Bahn and K. W. Jacobsen. An object-oriented scripting interface to a legacy electronic structure code. *Comput. Sci. Eng.*, 4:56–66, 2002.

[131] B.Z. Yanchitsky and A.N. Timoshevskii. Determination of the space group and unit cell for a periodic solid. *Comput. Phys. Commun.*, 139:235 – 242, 2001.

[132] T. Hahn, U. Shmueli, A. J. C. Wilson, and E. Prince. *International tables for crystallography.* D. Reidel Publishing Company, 2005.

[133] F. Birch. Finite elastic strain of cubic crystals. *Phys. Rev.*, 71:809–824, 1947.

[134] J. F. Nye. *Physical Properties of Crystals: their representation by tensors and matrices.* Clarendon, Oxford, 1957.

[135] J. C. Boettger. All-electron full-potential calculation of the electronic band structure, elastic constants, and equation of state for graphite. *Phys. Rev. B*, 55: 11202–11211, 1997.

[136] W. Voigt. *Lehrbuch der Kristallphysik.* Teubner, Leipzig, 1928.

[137] A. Reuss. Berechnung der Fließgrenze von Mischkristallen auf Grund der Plastizitätsbedingung für Einkristalle. *ZAMM - J. Appl. Math. Mech.*, 9:49–58, 1929.

[138] R. Hill. The elastic behaviour of a crystalline aggregate. *Proc. Phys. Soc. London, Sect. A*, 65:349, 1952.

[139] Blaha, P. et al. *WIEN2k, An Augmented Plane Wave+Local Orbitals Program for Calculating Crystal Properties.* Technische Universität Wien, Austria, 2001. Version 11.

[140] M. S. Dresselhaus, G. Dresselhaus, M. A. Pimenta, and P. C. Eklund. *Analytical applications of Raman spectroscopy*, chapter Raman scattering in carbon materials. Blackwell Science, 1999.

[141] A. C. Ferrari, J. C. Meyer, V. Scardaci, C. Casiraghi, M. Lazzeri, F. Mauri, S. Piscanec, D. Jiang, K. S. Novoselov, S. Roth, and A. K. Geim. Raman spectrum of graphene and graphene layers. *Phys. Rev. Lett.*, 97:187401, 2006.

[142] A. C. Ferrari. Raman spectroscopy of graphene and graphite: Disorder, electron–phonon coupling, doping and nonadiabatic effects. *Solid State Commun.*, 143:47 – 57, 2007.

[143] M. Gradhand, D. V. Fedorov, F. Pientka, P. Zahn, I. Mertig, and B. L. Györffy. First-principle calculations of the berry curvature of bloch states for charge and spin transport of electrons. *J. Phys. Condens. Matter*, 24:213202, 2012.

[144] E. Betranhandy and S. F. Matar. *Ab initio* investigation of the nitrofluoride SiNF. *Phys. Rev. B*, 72:205108, 2005.

[145] E. Betranhandy, G. Demazeau, and S. F. Matar. First principles study of the stability of SiNF. *Comput. Mater. Sci.*, 34:22 – 34, 2005.

[146] N. Hamada, S.-I. Sawada, and A. Oshiyama. New one-dimensional conductors: Graphitic microtubules. *Phys. Rev. Lett.*, 68:1579–1581, 1992.

[147] I. Boustani. Systematic ab initio investigation of bare boron clusters: Determination of the geometry and electronic structures of B_n (n=2-14). *Phys. Rev. B*, 55:16426–16438, 1997.

[148] J. Kunstmann and A. Quandt. Broad boron sheets and boron nanotubes: An *ab initio* study of structural, electronic, and mechanical properties. *Phys. Rev. B*, 74:035413, 2006.

[149] K. C. Lau, R. Pati, R. Pandey, and A. C. Pineda. First-principles study of the stability and electronic properties of sheets and nanotubes of elemental boron. *Chem. Phys. Lett.*, 418:549 – 554, 2006.

[150] K. C. Lau and R. Pandey. Stability and electronic properties of atomistically-engineered 2D boron sheets. *J. Phys. Chem. C*, 111:2906–2912, 2007.

[151] K. C. Lau and R. Pandey. Thermodynamic stability of novel boron sheet configurations. *J. Phys. Chem. B*, 112:10217–10220, 2008.

[152] H. Tang and S. Ismail-Beigi. Novel precursors for boron nanotubes: The competition of two-center and three-center bonding in boron sheets. *Phys. Rev. Lett.*, 99:115501, 2007.

[153] C. Özdoğan, S. Mukhopadhyay, W. Hayami, Z. B. Güvenç, R. Pandey, and I. Boustani. The unusually stable B100 fullerene, structural transitions in boron nanostructures, and a comparative study of α- and γ-boron and sheets. *J. Phys. Chem. C*, 114:4362–4375, 2010.

[154] X. Yang, Y. Ding, and J. Ni. *Ab initio* prediction of stable boron sheets and boron nanotubes: Structure, stability, and electronic properties. *Phys. Rev. B*, 77:041402, 2008.

[155] T. R. Galeev, Q. Chen, J.-C. Guo, H. Bai, C.-Q. Miao, H.-G. Lu, A. P. Sergeeva, S.-D. Li, and A. I. Boldyrev. Deciphering the mystery of hexagon holes in an all-boron graphene α-sheet. *Phys. Chem. Chem. Phys.*, 13:11575–11578, 2011.

[156] J. Kunstmann and A. Quandt. Constricted boron nanotubes. *Chem. Phys. Lett.*, 402:21 – 26, 2005.

[157] H. Tang and S. Ismail-Beigi. First-principles study of boron sheets and nanotubes. *Phys. Rev. B*, 82:115412, 2010.

[158] D. Ciuparu, R. F. Klie, Y. Zhu, and L. Pfefferle. Synthesis of pure boron single-wall nanotubes. *J. Phys. Chem. B*, 108:3967–3969, 2004.

List of Figures

List of Tables

Acknowledgements

At the very end of my thesis, I want to thank several people:

- Special thanks got to Prof. Kortus for providing the interesting topic and for being a constant source of motivation. Thank you for taking the time answering my sometimes trivial and sometimes not so trivial questions and for your seemingly unlimited patience.
- I also want thank Prof. Zurek for being the coreferee of this thesis.
- Furthermore, I want to highlight the great (working) atmosphere within the Institute for Theoretical Physics. It has been awesome to be part of this group. Especially, the room OG 24/25 was the best place for my desk, its inmates (Christian and Andi; Claudi almost lived here, too) have among others offered inspiration and also diversion if needed.
- To my friends and family go thanks for supporting me for all the years.
- Last but not least, I want to thank my husband Helge for encouraging, pushing and helping me throughout the whole time. I doubt this work would have been finished without your support.