Skolkovo Institute of Science and Technology

# COEVOLUTIONARY SEARCH FOR MATERIALS WITH OPTIMAL PROPERTIES IN THE SPACE OF BINARY SYSTEMS

*Doctoral Thesis*

by

ZAHED ALLAHYARI

DOCTORAL PROGRAM IN MATERIALS SCIENCE AND ENGINEERING

Supervisor

Professor Artem R. Oganov

Moscow – 2020

I hereby declare that the work presented in this thesis was carried out by myself at Skolkovo Institute of Science and Technology, Moscow, except where due acknowledgement is made, and has not been submitted for any other degree.

Candidate (Zahed Allahyari)

Supervisor (Prof. Artem R. Oganv)

# DEDICATION

I dedicate this thesis to my dear wife whose support sustained me throughout my graduate studies.

# Abstract

Fast developments in technology and industry require optimal materials for various applications, which highlights the role of materials science in solving its central problem: discovery of optimal materials, having any required properties, in the entire chemical space (i.e., among all possible combinations of elements across the periodic table). Despite the great development of computers and the success of the crystal structure prediction methods such as metadynamics[1] and evolutionary algorithms,[3–5] the problem has remained unsolved because of the high level of complexity. This thesis summarizes our efforts to solve it. We present the coevolutionary optimization algorithm, or Mendelevian search (MendS) method, that we developed to predict optimal materials while aiming for target properties, which was made possible by the implementation of well-designed variation operators. The multi-objective optimization technique, explained in Chapter 3, was used to ensure that the predicted materials are simultaneously optimal in a given property and have a minimal energy. We introduce the idea of organizing the chemical space so that neighboring systems are likely to have similar crystal structures and properties; therefore, the prediction of one optimal material in this space leads to the prediction of other optimal materials. We examined our MendS method by searching for low-energy hard and superhard binary systems. The results of this search are presented in the last chapter.

# Publications

1. Kvashnin, A. G., Oganov, A. R., Samtsevich, A. I., Allahyari, Z. Computational Search for Novel Hard Chromium-Based Materials. J. Phys. Chem. Lett. **8**, 755–764 (2017).
2. Núñez-Valdez, M., Allahyari, Z., Fan, T., and Oganov, A. R. Efficient technique for computational design of thermoelectric materials. Comput. Phys. Commun. **222**, 152–157 (2018).
3. Allahyari, Z. and Oganov, A. R. Multi-Objective Optimization as a Tool for Material Design. In: Andreoni W., Yip S. (eds) Handbook of Materials Modeling. Springer, Cham (2019).
4. Kvashnin, A. G., Allahyari, Z. & Oganov, A. R. Computational discovery of hard and superhard materials. J. Appl. Phys. **126**, 040901 (2019).
5. Allahyari, Z., Oganov, A.R. Coevolutionary search for optimal materials in the space of all possible compounds. npj Comput. Mater., **6**, 55 (2020).

# Acknowledgments

Dear God, thank you for my existence, for my eyes to see, my brain to think, my ears to hear, for the opportunity to grow in a great family and excellent academic environment, for my hands to write this thesis, and for all I have. I am very grateful to have you.

I would like to express my sincere gratitude to my supervisor, Prof. Artem Oganov, who continuously supported my Ph.D. study and related research, for his patience, motivation, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis. I could not imagine a better supervisor and mentor for my Ph.D. study.

I would like to thank the jury members and the chairman of my Ph.D. defense: Prof. Andrey Zhygayevych, Prof. Sergey Levchenko, Prof. Pierre Villars, Prof. Sandro Scandolo, and Prof. Keith Stevenson, and Prof. Xavier Gonze for their kind support and insightful comments.

Deeply appreciating my family, I thank my wife, my parents, my brother, and my sister for supporting me spiritually during my Ph.D. study and my life in general.

Last but not least, I am very grateful to all my friends: Nastya Naumova, Artem Samtsevich, Ivan Kruglov, Dr. Alexander Kvashnin. Vladimir Baturin, Sergey Lepeshkin, Michele Galasso, Pavel Bushlanov, Congwei Xie, Valery Roizen, Dmitry Rybkovsky, Alexander Kvashnin, and many others, who made my time at Skoltech a wonderful experience.

# Personal contribution

The idea of the Mendelevian Search method – defining an organized chemical space in order to perform an unbiased search for optimal materials – was proposed by the author's supervisor, Prof. Artem R. Oganov.

The author implemented all the discussed methods in this thesis, i.e. multi-objective Pareto optimization, an organized Mendelevian chemical space, and the coevolutionary algorithm in the form of a computer program – MendS.

The author implemented multi-objective Pareto optimization in the USPEX code.

The author used the implemented methods and performed all the calculations.

The discussed database in Chapter 4. is created by the author.

All the discussed results in Chapters 3, 4, 5, and 6 are analyzed by the author under the supervision of Prof. Artem R. Oganov.

# Thesis Outline

Chapters 1 and 2 of this thesis introduce the crystal structure prediction.

In Chapter 3, the multi-objective (MO) optimization technique is introduced in the form of the published work:

Allahyari, Z. and Oganov, A. R. Multi-Objective Optimization as a Tool for Material Design. In: Andreoni W., Yip S. (eds) Handbook of Materials Modeling. Springer, Cham (2019).

And a short explanation of the method presents in the following published paper:

Núñez-Valdez, M., Allahyari, Z., Fan, T., and Oganov, A. R. Efficient technique for computational design of thermoelectric materials. Comput. Phys. Commun. **222**, 152–157 (2018).

The nonempirical organization of a chemical space which is discussed in Chapter 4, is about to be submitted to Journal of Physical Chemistry Letters.

In Chapter 5, we introduce our Mendelevian search method in the form of a published paper: Allahyari, Z. & Oganov, A. R. Coevolutionary search for optimal materials in the space of all possible compounds. npj Comput. Mater., **6**, 55 (2020).

We show the results of our search for hard and superhard materials in Chapter 6, which is based on the following papers:

- Kvashnin, A. G., Oganov, A. R., Samtsevich, A. I., Allahyari, Z. Computational Search for Novel Hard Chromium-Based Materials. J. Phys. Chem. Lett. **8**, 755–764 (2017).
- Kvashnin, A. G., Allahyari, Z. & Oganov, A. R. Computational discovery of hard and superhard materials. J. Appl. Phys. **126**, 040901 (2019).
- Allahyari, Z. & Oganov, A. R. Coevolutionary search for optimal materials in the space of all possible compounds. npj Comput. Mater., **6**, 55 (2020).

# Contents

# List of Figures

# List of Tables

# CHAPTER 1.
# INTRODUCTION

Discovery of new materials with specific properties is the main goal of materials science, whose progress has played a significant role in the history of humanity and the development of civilization. From the Stone Age to the Bronze Age to the modern times, novel materials helped the development of societies, with pioneer nations in this field attaining a greater power and influence over other nations. For example, the discovery of iron, copper, and some of their alloys thousands of years ago had a great impact on the technology of that time, and by utilizing these materials in tools and weapons, developed nations spread and increased their influence over other nations. In today's world, materials science is closely tied to the technological and industrial applications. The discovery of steel and silicon are just two examples of new materials that caused revolutions in transportation and communication, bringing about railways and electronic chips.

Until recently, materials discovery was purely empirical. Up to 1912, it was not possible to determine the positions of atoms in a crystal. Later, this limitation was overcome with the development of X-ray diffraction techniques.[5] Nowadays, several diffraction techniques, such as the X-ray diffraction, neutron diffraction, and electron diffraction, can accurately determine the crystal structure of materials. However, despite all the developments, the experimental discovery of materials is limited and even made unaffordable by the high cost of trial and error and the time-consuming procedure of synthesis. On the other hand, the rapid progress in technology and technological race shorten the time for invention and application and increase the demand for new materials.

Unlike experimental studies, the computational materials design opened a way to predict new materials at much lower costs. The development of modern computers and advanced methods, such as density functional theory (DFT), made it possible to accurately predict the properties of materials numerically without going through the expensive and time-consuming procedures of blind experimental study. Although computational studies need to be verified by experiments, this seems to be an acceptable alternative helping to meet the growing demand for new materials.

This thesis summarizes our efforts to develop the advanced Mendelevian search (MendS) method that provides unbiased systematic solution to the central problem of materials science — the discovery of optimal materials with the required properties in the entire chemical space — and its application in the search for hard and superhard materials. The method consists of several submethods. In Chapter 1, some background of the crystal structure and geometry optimization is provided. Chapter 2 introduces the important features and concepts of the evolutionary methodology USPEX for the crystal structure prediction. In Chapter 3, the multi-objective optimization method as a tool for materials design is presented in the form of a published paper. In Chapter 4, we introduce a new model of organizing a chemical space. Chapter 5 focuses on the well-designed coevolutionary algorithm that combines all the previously discussed submethods into a single code, MendS. In Chapter 6, we show the results

of the Mendelevian search for low-energy hard and superhard materials. A brief discussion and conclusion is presented in Chapter 7.

## 1.1. Structure of Materials

The first step toward a computational study of a material is the knowledge of its atomic structure, which enables the precise simulation and then the calculation of numerous properties using state-of-the-art quantum mechanical methods.

In this thesis, we are mainly interested in the atomic structures of crystalline solids. Unlike amorphous materials such as glass, which have no long-range atomic order, crystal structures are periodic in the three-dimensional space and thus have long-range order.

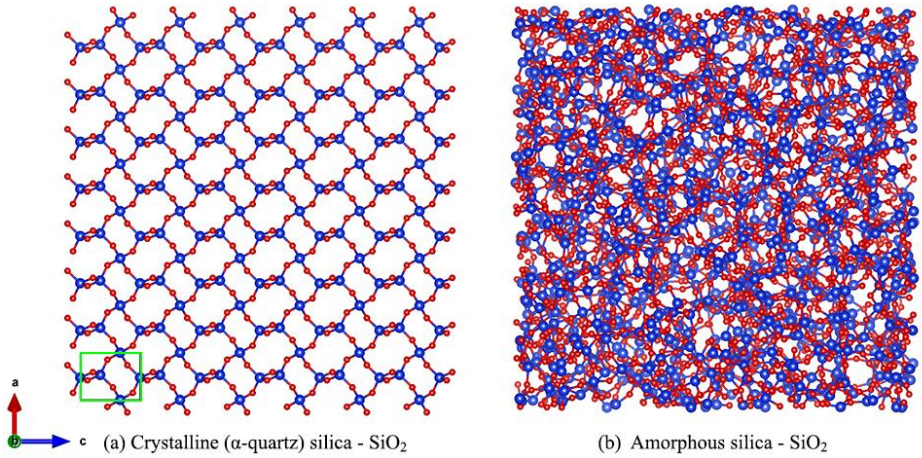



Figure 1. The crystalline and amorphous forms of silica — $SiO_2$.

The smallest unit of this repeated pattern is called a unit cell. The unit cell is, in fact, a parallelepiped defined by six parameters — three lengths of the cell edges (*a*, *b*, *c*) and three angles between them (α, β, γ) — known as the lattice parameters.

Fig. 1 shows structures of a crystalline and a glassy (amorphous) form of silicon dioxide, or silica. A regular arrangement of atoms in the crystalline structure is the result of the repetitive translation of the unit cell (i.e. the green square in Fig. 1a) along its principal axes.

Crystalline materials are classified into seven different crystal systems. These crystal systems are easily identified by the cell parameters – see Table 1. Among these systems, triclinic (anorthic) is the lowest symmetry system with no restrictions on the values of lattice parameters while in other crystal systems, symmetry reduces the number of unique lattice parameters. In all the crystal systems, crystalline lattices are considered to have 'lattice points' on the corners of the unit cell. However, it is possible to generate a lattice with non-trivial centering vectors.

In 1849, Auguste Bravais found that all regular crystals can be described with 14 lattice types for the seven crystal systems. Lattices with lattice points only on the corners are called primitive and are designated with the symbol P. Lattices with additional lattice points (non-primitive lattices) are described as *centered lattice* and regarding the position of additional lattice points in center of the unit cell or center of side/all faces, it is called body-centered (designated symbol – I), based-centered (designated symbol – S or C), face-centered

(designated symbol – F), respectively. There is a rhombohedral-centered lattice specifically for trigonal crystal system, and it is designated with the symbol R – see Table 1.

Table 1. Crystal systems, Bravais lattice types and their point groups.

| Crystal systems | Cell parameters | Bravais Lattice | Point groups |
|---|---|---|---|
| Triclinic | $a \neq b \neq c$; $\alpha \neq \beta \neq \gamma$ | aP | 1, $\bar{1}$ |
| Monoclinic | $a \neq c$; $\alpha = \gamma = 90^\circ$, $\beta \neq 90^\circ$ | mP, mS | 2, *m*, 2/*m* |
| Orthorhombic | $a \neq b \neq c$; $\alpha = \beta = \gamma = 90^\circ$ | oP, oS, oI, oF | 222, *mm*2, *mmm* |
| Tetragonal | $a = b \neq c$; $\alpha = \beta = \gamma = 90^\circ$ | tP, tI | 4, $\bar{4}$, 422, 4/*m*,4*mm*, $\bar{4}$2*m*, 4/*mmm* |
| Hexagonal - hexagonal | $a = b$; $\alpha = \beta = 90^\circ$, $\gamma = 120^\circ$ | hP | 6, $\bar{6}$, 622, 6/*m*, 6*mm*, $\bar{6}$*m*2, 6/*mmm* |
| Hexagonal - rhombohedral | $a = b = c$; $\alpha = \beta = \gamma \neq 90^\circ$ | hR | 3, $\bar{3}$, 32, 3*m*, $\bar{3}$*m* |
| Cubic | $a = b = c$; $\alpha = \beta = \gamma = 90^\circ$ | cP, cI, cF | 23, *m*$\bar{3}$, 432, $\bar{4}$3*m*, *m*$\bar{3}$*m* |

A unit cell reflects full information about the symmetry of crystalline materials. Generally, there are two types of symmetries: the translational symmetry, which is the periodic repetition of a unit cell along its principal axes, and the point symmetry. The point group includes such symmetries as reflection, rotation, inversion, and rotoinversion. In total there are 32 unique crystallographic point groups. These 32 point groups are shown in Table 1.

Also, by combining translational symmetry and point group symmetry, new symmetry operations can be created. For example, combination of proper rotations with translations give rise to operations described as screw axes and combination of reflections (mirror planes) with translations give rise to glide plane operations.

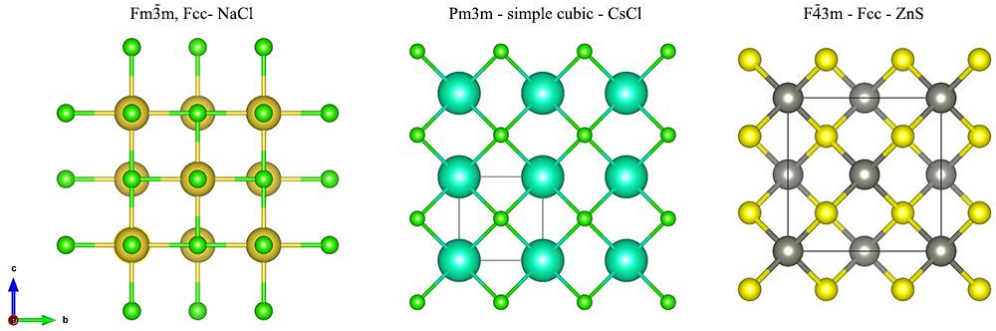



Figure 2. The crystal structure of NaCl, CsCl, and ZnS. All these crystal structures belong to the cubic crystal system, but with different space groups. Both NaCl and ZnS have the face-centered lattice type, whereas the atomic arrangement is octahedral in NaCl and tetrahedral in ZnS, making them belong to different space groups.

Combination of the seven crystal systems with the 14 Bravais lattices, the 32 point groups, screw axes, and glide planes yields to a total of 230 unique space groups in the three dimensional space. Space group is, in fact, a group of symmetry operations that are combined to describe the symmetry of a periodic object (crystal structure) in a three dimensional space. In point groups, all the symmetry elements pass through a single point in the object while in space groups, there is no need for intersect of symmetry elements in a single point. The initial

letter of a space group symbol represents the lattice type (P, C, F, etc.), and followed by the point group symbol in which the rotation and reflection elements are extended to include screw axes and glide planes. The crystal structures of some compounds with cubic crystal system and different space groups are shown in Fig. 2.

## 1.2. Geometry Optimization

Geometry optimization, local optimization, relaxation, and energy minimization are different terms representing the same task: finding the spatial arrangement of atoms in which the net interatomic force on each atom is zero or close to zero and the configuration is located at a stationary point on the potential energy surface (PES). The gradient, or the first derivative of energy with respect to geometry *r*, gives the negative of the force: $F(r) = -\frac{\partial U}{\partial r}$. If the first derivative is equal to zero, the point is located at a minimum, maximum, or transition state (saddle point) of the PES. To distinguish among these points, the second derivative of the energy must be examined. The matrix of the second derivative is called the Hessian matrix. By the diagonalization of the Hessian matrix, eigenvalues and eigenvectors can be calculated. The vibration frequencies are proportional to the square root of the eigenvalues. If all the eigenvalues are positive, the point is located at a minimum on the PES. In case all the eigenvalues are negative, the point is located at a maximum. If there are one or more negative eigenvalues – leading to imaginary frequencies – the point is located at a saddle point on the PES.

There are several techniques of geometry optimization (e.g., the conjugate gradients method,[6] Newton-Raphson method,[7] Broyden (BFGS) algorithm,[8] etc.), whereas the energy function can be determined by solving the quantum mechanical equations and by the density functional theory (DFT) methods, or by parameterized analytical functions.

## 1.3. Energy of a Structure

From a thermodynamic point of view, the most stable crystal structure is that with the lowest Gibbs free energy at given temperature and pressure.

$$G = E + PV - TS, \tag{1.1}$$

where *E* is the internal energy, *P* is the pressure, *V* is the volume, *T* is the temperature, and *S* is the entropy of the system. In different situations, Gibbs free energy can be formulated differently. The Eq. (1.1) only applies to isotropic materials with hydrostatic pressure case. In general, for dealing with solid materials, the *PV* term is replaced by strain-stress tensor as formulated in Eq. (1.2)

$$G = E - \sum_{ij} \varepsilon_{ij}\sigma_{ij} - TS, \tag{1.2}$$

where $\varepsilon$ and $\sigma$ are strain and stress tensors respectively.

Among these terms, the calculation of *PV* and $\varepsilon_{ij}\sigma_{ij}$ is rather straightforward, whereas the calculation of *TS* using the first principles is complicated. In this thesis, all the energy calculations were done at zero pressure and temperature; this also means that the strain-stress

term in Eq. (1.2) is zero. Therefore, the internal energy *E* is the only term to be calculated. This calculation is discussed in the following section.

## 1.4. Density Functional Theory

In quantum physics, particles show a wave behavior, and an expression describing their various properties, equivalent to Newton's second law in classical physics, is the Schrödinger equation. Finding the wave function $\psi$ of an *N*-electron quantum system makes it possible to calculate the energy and many other properties of the system. To do it, the Schrödinger equation must be solved:

$$\hat{H}\psi = E\psi, \tag{1.3}$$

$$\hat{H} = \hat{T}_{\text{nucl}} + \hat{U}_{\text{nucl}} + \hat{T}_{\text{e}} + \hat{U}_{\text{e}} + \hat{V}. \tag{1.4}$$

The Eq. (1.3), is known as time-independent Schrödinger equation. This equation is for systems in the stationary state and does not contain the relativistic effects – for studying the behavior of quantum mechanical systems with relativistic effects, one may use the Klein–Gordon equation. The Born-Oppenheimer approximation separates the degrees of freedom of fast electrons from those of slow ions and neglects the ionic terms. Thus, the Hamiltonian of Schrödinger equation can be written as:

$$\hat{H} = \hat{T}_{\text{e}} + \hat{U}_{\text{e}} + \hat{V} = \sum_{i}^{N}\left(-\frac{\hbar^2}{2m_i}\nabla_i^2\right) + \sum_{i<j}^{N} U(r_i, r_j) + \sum_{i}^{N} V(r_i), \tag{1.5}$$

where

$$U(r_i, r_j) = \frac{e^2}{4\pi\epsilon_0|r_i - r_j|}, \text{and } V(r_i) = \sum_{l}\left(-\frac{e^2 Z_l}{4\pi\epsilon_0|R_l - r_i|}\right).$$

In these equations, $\hat{T}$ describes the kinetic energy of electrons, $\hat{U}$ describes the electron-electron Coulomb repulsion, $r_i$ and $r_j$ are the positions of interacting electrons, $\hat{V}$ describes the electron-ion Coulomb attraction, $R_l$ is the position of nuclei and $Z_l$ is the atomic number of nuclei. The subscripts e and nucl represent the electronic and ionic parts of the operators.

The many-body Schrödinger equation (Eq. 1.3) provides almost exact wave function for studying the behavior of the system, but except for a few simple systems, i.e. hydrogen atom, it is not solvable analytically. There are sophisticated approximations which one can conveniently solve numerically, such as the Hartree–Fock and post-Hartree–Fock methods based on the expansion of wave functions in Slater determinants. The most accurate of these methods are exhaustingly expensive and impractical, especially for large systems.

In 1964, Hohenberg and Kohn (H–K) published and proved two theorems[9] that originated the density functional theory. The theorems stated that:

1. The ground state charge density of a system uniquely determines its potential and all other properties (such as the wave function, energy, etc.).
2. A universal functional of energy $E[n]$ can be defined in terms of the density $n(r)$, which is valid for any external potential $v_{\text{ext}}(r)$. For any particular $v_{\text{ext}}(r)$, the exact

ground state electron density of the system is determined by minimizing this functional.

DFT attracted attention as it provided a way to map the many-body problem onto a single-body problem. It determines the properties of a many-body system using the electron density functional.

In 1965, Kohn and Sham introduced the equation[10] which was a major step toward a quantitative modeling of electronic structures.[11] In the proposed approach, each particle interacts with a density of electrons rather than other particles. This reduces the many-body problem of *N* electrons with 3*N* spatial coordinates to a single-body problem with 3 spatial coordinates, which can be perfectly solved using the computers.

The Kohn–Sham equation is a Schrödinger-like equation that applies on a fictitious system with noninteracting electrons that generates the same electron density as any given real system with interacting electrons. As the electrons in the Kohn–Sham system are noninteracting, the Kohn–Sham wave function is a single Slater determinant constructed from a set of orbitals that are the lowest-energy solutions to the Kohn–Sham equation:

$$\left(-\frac{\hbar^2}{2m}\nabla^2 + v_{\text{eff}}(r)\right)\varphi_i(r) = \varepsilon_i\varphi_i(r), \tag{1.6}$$

where the local effective potential energy acting on a system is

$$v_{\text{eff}}(r) = v_{\text{ext}}(r) + e^2\int\frac{n(r')}{|r-r'|}dr' + \frac{\delta E_{\text{xc}}[n]}{\delta n(r)}, \text{and } n(r) = \sum_i |\varphi_i(r)|^2.$$

The external potential $v_{\text{ext}}(r)$ is the Coulomb attraction from nuclei. All the terms of the Kohn–Sham equation are known except for the exchange-correlation term $E_{\text{xc}}[n]$. The Kohn–Sham equation can be solved self-consistently and the results are expected to be exact if the exchange-correlation term is known. Extremely good results can be produced by making simple approximations. Many efforts have been made to develop the approximations, such as the local density approximation (LDA), generalized gradient approximation (GGA), Meta-GGA, hybrid functional, and others, which nowadays are referred to as the "Jacob's ladder of the DFT".[12]

DFT has been widely used during the past decades and remains one the most popular methods for studying materials computationally. It can accurately describe the ground state properties of a solid-state system (such as lattice parameters, formation energy, etc.) and is therefore a powerful evaluation tool.

### 1.4.1. Exchange and Correlation terms

All the terms in Kohn-Sham equation are known except for the exchange and correlation terms. Exchange is due to the Pauli exclusion principle, which forbids two identical fermions to lie in the same quantum state. Therefore, two electrons with the same quantum numbers, *n*, *l* and *m*, can occupy the same quantum state only if their last quantum number, $m_s$, is different. Meaning that two electrons with parallel spins are not allowed to sit in the same quantum state.

This phenomenon causes an effective repulsion between electrons with similar spins which is named as the *exchange* interaction term.

The *correlation* term is due to impossibility of two molecules to occupy the same place. This effect already exists at the classical level. For example, in an ensemble of interacting classical particles (e.g. billiard balls), there are correlation effects – if a ball is in one place, it is not possible for another ball to be at the same place. The same rule applies to the gas molecules. In fact, this effect is one of the essential ingredients of Van der Waals equation of gases.

Although the form of these terms are unknown there are several methods to approximate their values.

### 1.4.2. Local Density Approximation (LDA)

To estimate the exchange and correlation term, several approximations are examined. The simplest approximation, known as local density approximation, uses the energy of electrons as in a homogenous electron gas (HEG). The exchange-correlation energy, $E_{XC}[n(r)]$, in the LDA is defined as:

$$E_{XC}[n(\vec{\boldsymbol{r}})] = \int \varepsilon_{XC}\big(n(\vec{\boldsymbol{r}})\big)n(\vec{\boldsymbol{r}})d^3r \tag{1.7}$$

where $\varepsilon_{XC}(n(\boldsymbol{r}))$ is the exchange-correlation energy per unit volume of a homogenous electron gas with density n(**r**). The values of $\varepsilon_{XC}$ were calculated by Ceperley et al., using the Quantum Monte Carlo technique,[13] and were parameterized by Perdew et al.[14]

LDA has a tendency to favor more homogeneous systems and overbinds molecules and solids. In weakly bonded systems, these errors are exaggerated and bond lengths are too short.[15] Generally, LDA gives good results for systems with the smoothly varying charge densities – materials consisting of *sp* electrons, with nice covalent bonds, or simple metals.

### 1.4.3. Generalized Gradient Approximation (GGA)

Although, LDA leads to good results (for materials with slowly varying charge density) for some properties (i.e. lattice constants, bulk moduli, equilibrium geometries, and vibrational frequencies),[16] its drastic overbinding, makes it necessary to calculate the derivatives of electronic densities in order to account for the non-homogeneity of the true electron density. Doing this to satisfy known constraints for the exact functional, one arrives at a family of functionals at the level of generalized gradient approximation (GGA) and have the following form:

$$E_{XC}^{GGA}[n(\vec{\boldsymbol{r}})] = \int \varepsilon_{XC}\big(n(\vec{\boldsymbol{r}}), \nabla n(\vec{\boldsymbol{r}})\big)n(\vec{\boldsymbol{r}})d^3r \tag{1.8}$$

Using the GGA, better results for molecular geometries and ground-state energies have been achieved. As it was mentioned earlier, there are meta-GGA methods that are often more accurate than GGA. In their natural form, these methods contain the first and second derivatives of electron density (the Laplacian) – more accurate but more expensive. There are several types

of GGA and meta-GGA methods, because of the form of density functionals and different definition of their parameters. In this thesis, we widely used the GGA-PBE[17] method. The advantage of this method over other GGA functionals is that it is fully non-empirical and has good numerical behavior (avoiding pathologies, albeit rare, of PW91).

# CHAPTER 2.

# Crystal Structure Prediction Using the Evolutionary Algorithm USPEX


## Abstract

*For a long time, it was thought that crystal structures cannot be predicted on the basis of only the chemical composition because even the simplest systems have a huge number of possible configurations. However, many algorithms that solve this problem have been proposed during the past decades, including the metadynamics, minima hopping, evolutionary algorithm, and other techniques that screen the structures in the promising regions of the potential energy surface (PES) and discard those in the unpromising regions. In this chapter, we briefly explain some of the state-of-the-art methods for crystal structure prediction and then present a concept of evolutionary algorithm (EA) as implemented in the USPEX code, which showed high efficiency in the crystal structure prediction.*



**Keywords:** Evolutionary algorithm, potential energy surface, relaxation, global search.


## 2.1. Introduction

Knowing only the crystal structure of a material is enough to calculate its numerous properties using state-of-the-art quantum mechanical methods. However, until recently, it was thought that crystal structures cannot be predicted and the only practical way to obtain them is by using diffraction methods such as the X-ray diffraction, which are limited to the quality of data, especially at high pressures and temperatures. Therefore, finding another way to obtain crystal structures was essential to overcome these limitations. On the other hand, theoretical crystal structure prediction is a good alternative when the experimental studies are limited.

The potential energy surface (PES) has an overwhelming number of local minima (Fig. 3). Fortunately, it is not necessary to study the entire energy landscape to locate the global minimum, and only the promising regions can be explored. There are several methods for doing this, such as simulated annealing,[18,19] basin hopping,[20] minima hopping,[21] and metadynamics.[1,22] Here, we briefly introduce these methods.

### 2.1.1. Simulated annealing

This approach is based on the concept of physical annealing. The Monte Carlo method is used to evolve the initial configuration. The new configuration is generated by a random move of ions, and in each step, the algorithm decides if the move is accepted or not. The simulation starts at a high temperature, where almost all the moves are accepted, and moves are attempted until the system reaches the thermal equilibrium point. This process is continued until the system is frozen. If the annealing process is continuous and carried out very slowly, there is a high chance to reach the global minimum (most stable structure). However, in practice it is

extremely hard to run such calculations with sufficiently slow cooling rate, and results are often far from global minimum.

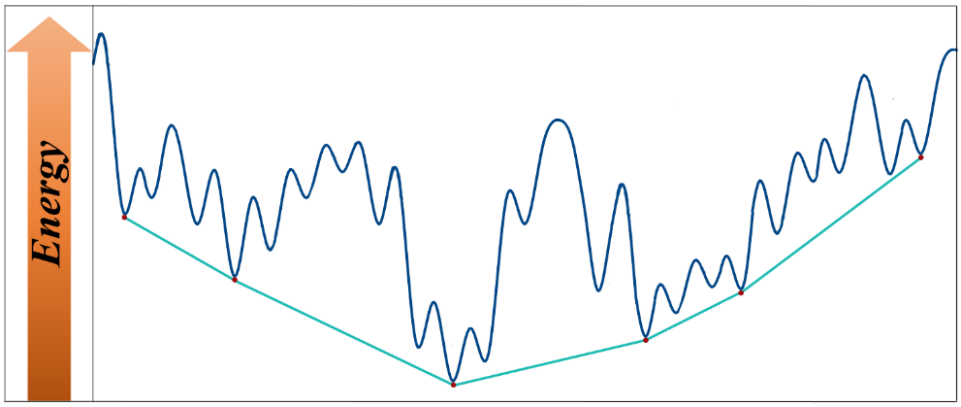



Figure 3. A schematic plot of the potential energy surface.

### 2.1.2. Basin hopping

This algorithm maps the PES onto a set of staircases with plateaus, or basins of attraction, corresponding to the set of configurations that lead to the given minimum after the optimization. For exploring the energy landscape the Monte Carlo simulation was used with a constantly reducing temperature by 0.8. In each step, all coordinates are displaced by a random number in the range of [-1, 1]. The energy of new coordinates is then minimized using conjugated gradients. If the energy of the new configuration is lowered, the move is accepted. Otherwise, the move is only accepted with a probability of $((E_{old} - E_{new})/(k_B T))$.[23] The key to the possible success of the basin hopping algorithm is due to removing the energy barriers of PES and converting the PES into the set of basins of attraction of all local minima. Therefore, it is possible for a system to hop between basins. Basin hopping method was used to predict the structure of Lennard-Jones clusters with 1-110 particles,[20] and for even larger systems.

### 2.1.3. Minima hopping

Minima hopping performs molecular dynamic simulations instead of Monte Carlo as in basin hopping.[23] This method is consisting of (1) an inner part: to perform jumps from one local minimum to another, and (2) an outer part: to accept or reject the new local minimum. The new local minimum is accepted only if the energy difference between the current local minimum and new local minimum is less than a predefined $E_{diff}$. $E_{diff}$ is not a constant and each time that a new local minimum is rejected or accepted, $E_{diff}$ is increased or decreased. Therefore, if the inner part only proposes the moves that go up in energy, such moves will finally be accepted after $E_{diff}$ has been sufficiently increased after many rejections. In the original method, the geometry relaxations were done using the combination of standard steepest descent and conjugate gradient method.[21] Jumping to a new local minimum (results of the inner part) leads to three different cases: (1) geometry relaxation gives back the current local minimum, (2) reaching an already visited local minimum, and (3) finding a new local minimum – the most desirable outcome, because of the chance to explore new configurations. If the first or second case happens, the value of $E_{kinetic}$ is increased to help the algorithm jumping

over high energy barriers and explore new local minima. If the third case happens, $E_{kinetic}$ will be decreased.

Minima hopping recognizes the regions that have been already explored and avoids revisiting these parts of the energy landscape. The method has been successfully applied to study the high pressure structures of $Si_2H_6$,[24] $LiAlH_4$,[25] $Zn(BH_4)_2$,[26] etc.

### 2.1.4. Metadynamics

The metadynamics[22] algorithm adds a positive Gaussian potential to the energy landscape to lower and overcome the barriers.[23] The algorithm starts from a set of equilibrated collective coordinated (i.e. cell parameters) at a given pressure and temperature. Evaluates the pressure tensor, while keeping the cell parameters fixed, during a long enough molecular dynamic relaxation. Then the cell parameters are updated and particle positions are rescaled to fit into the new lattice box. As the initial free-energy well is gradually filled, the lattice undergoes a set of deformations until a transition state is reached and the system enters into the basin of attraction of a new state.[1] Adding positive Gaussian potential energy to fill the wells, makes it easier for a system to jump from one well to another well to explore other regions of the PES. The metadynamics algorithm keeps the history of visited spots and discourages a system from revisiting the same spot and encourages an efficient exploration of the PES.

All the methods discussed above need to have an initial guess of the structure. If a good structure is guessed, in a region close to the global minimum, these methods have the advantage to converge fast into the global stable structure. Otherwise, these methods are not very effective. This way, the problem of global search turns into a local search in a relatively small region of the energy landscape. As a characteristic of the energy landscape, most of the low-energy minima are located in the same regions, and corresponding structures have similar bond lengths and coordination environment of atoms,[2] which makes the final structure similar to the initial one. Therefore, these methods are more effective when the initial structure is close to the global minimum, whereas a good starting point is not always known in advance.

Among these methods, metadynamics has reduced the dimensionality of the problem to a small number of order parameters. The approach is often successful and efficient but having its failures. In this method, the lattice parameters are usually presented as a six-dimensional order parameter; its choice, however, is not always clear.[2]

In the rest of this chapter, first, we briefly mention a few methods that their effectiveness is not influenced by the choice of the initial population, such as random sampling,[27] particle swarm optimization [28] (PSO), and evolutionary algorithm [29] (EA). Then we discuss the concept and features of the evolutionary algorithm implemented in the USPEX [2,4,30] (Universal Structure Predictor: Evolutionary Xtallography) code which was extensively used in the prediction of new materials in our work.

### 2.1.5. Random sampling

This method starts by distributing (sampling) a large number of randomly generated structures all over the PES. These structures are then optimized to their local minimum and the energetically best structure is the natural choice of structure searching. However, the method

becomes efficient when these structures are not generated fully randomly. By choosing stoichiometry, imposing symmetry, using chemical ideas, shaking, etc., Pickard and Needs implemented an efficient random structure search algorithm in the AIRSS (*ab initio* random structure searching) code.[27]

### 2.1.6. PSO

The particle swarm optimization [28] (PSO) algorithm is inspired by the social behavior of birds flocking, and designed to solve the multidimensional optimization problems. The PSO algorithm starts with a number of random structures (random initial population similar to genetic algorithm), while a randomized velocity is assigned to each of these structures. Each structure keep track of its current velocity and location before and after the local optimization (relaxation), in addition to the location of the lowest energy structure in the population. Based on this information, the new velocity is calculated for each structure. Then the new location can be calculated based on the new velocity and previous location before the local optimization.

Velocity plays an important role in determination of the direction and speed of the structural movements and gives a superior ability to the PSO algorithm to overcome the large barriers of the energy landscape.[23] We can mention CALYPSO [31] (Crystal structure AnaLYsis by Particle Swarm Optimization) as one of the successful crystal structure prediction packages based on PSO algorithm.

### 2.1.7. Evolutionary algorithm

These algorithms are inspired by Darwin's evolutionary theory, and are based on the concept of "survival of the fittest" in nature. The evolutionary algorithm does not need any initial guess of the structures or the order parameters. They are fully nonlocal and suitable for multidimensional problems. EA is more efficient when the energy landscape has an overall shape that is expected for chemical systems (Fig. 3).

Mimicking the evolution in nature, an evolutionary algorithm tries to generate better and better structures in iterations (generations) by selecting the fittest structures as parents and applying to them variation operators such as heredity and mutation. The EA starts with a number of random structures (initial population). These structures are locally optimized and more chance are given to the best structures – those with lower energy among the population – to produce new structures and transfer their favorable properties to their children.

One of the important advantages of an evolutionary algorithm is its capability to predict several low-energy metastable structures (local minima) in addition to those with the global minimum energy. However, it must be mentioned that neither the evolutionary algorithm nor any other algorithm in practice can guarantee finding the global minimum.

## 2.2. Evolutionary Algorithm USPEX

USPEX is a crystal structure prediction code, based on the evolutionary algorithm, that demonstrated high efficiency,[32–34] especially in searching for atomic structures of materials with up to 100 atoms in the unit cell. In addition to 3D bulk materials, USPEX can search for optimal structures of surfaces, 2D crystals, polymers, nanoparticles, and proteins. During an

evolutionary search, USPEX optimizes materials with respect to a fitness function. This fitness can be any property or a combination of properties such as enthalpy, hardness, magnetization, elastic constants, and others, as explained in Chapter 3. USPEX is interfaced with a variety of the available DFT, molecular dynamics, and semi-empirical codes that are used to relax the structures and obtain their energies and other properties.

The algorithm of USPEX is shown in Fig. 4. In the first generation, several structures (initial population) are produced using random operators (e.g., random symmetry and random topology). These structures are then relaxed using the quantum mechanical calculations or other methods, their energies are obtained and used for ranking the relaxed structures – the structures are ranked based on the so-called fitness which can be energy, or any other property of a material that is defined by user. In the population of the fittest structures, only those with the highest rank, usually 70%, have a chance to be selected as parents. Then, variation operators such as mutation or heredity are applied to these parents, producing the structures of the next generation. The loop continues until the predefined number of generations is reached.

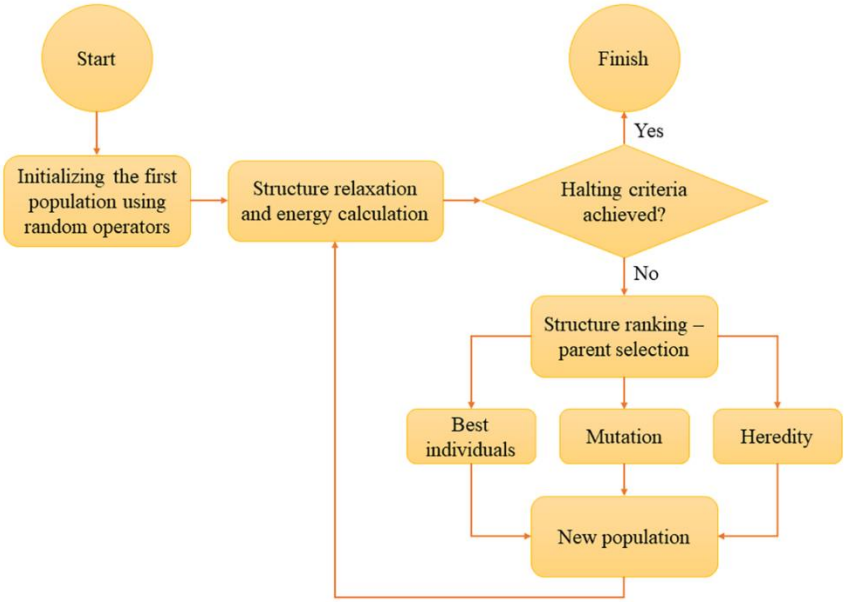



Figure 4. The algorithm of USPEX.

### 2.2.1. Selection Method

Selecting good parents is very important for the success of the algorithm. In this work, the roulette wheel selection method[35] was used and each system was assigned the weight function $w_r > 0$ ($r = 1, 2, \ldots, N$) derived from its fitness rank. To reasonably increase the chance of selection for best-ranked systems, a quadratic dependence to the fitness rank is assigned. The probability of selection of each system with the fitness rank $r$ is defined as:

$$P_r = \frac{w_r}{n_r \times \sum w_r}, r = 1, 2, \ldots, N \quad (2.1)$$

$$w_r = (N + 1 - r)^2 \quad (2.2)$$

where $N$ is the last fitness ranking number, which is equal to the number of all systems in the simple ranking or to the number of the Pareto fronts in the multi-objective Pareto ranking (Chapter 3); $n_r$ is the number of systems having the same fitness ranking $r$.

### 2.2.2. Variation Operators

Variation operators are at the heart of the evolutionary algorithm. They are used for generating new structures and must be designed carefully. The following variation operators are provided in USPEX:

**Heredity** produces new structures by cutting and pasting slices of parent structures. These slices are selected based on their degree of order.

**Softmutation** obtains a new structure by moving an atom or atoms of a parent structure along the eigenvectors of the softest modes because low-frequency eigenmodes correspond to the directions of low curvature of the energy surface.[4]

**Permutation** produces a new structure by switching two or more atoms in a parent structure.

**Transmutation** generates new structures by changing the chemical identity of a randomly selected atom to another chemical species in the system – it is used in the variable-composition mode.

**Rotational mutation** creates new structures by mutating orientations of randomly selected molecules – only for molecular crystal structure prediction.

**Lattice mutation** finds new structures by distorting the lattice of some known structures.[2]

**Spin mutation** randomly reinitializes the magnetic moments of a parent structure in a way that at maximum, one symmetry element of multiplicity two is broken.

### 2.2.3. Random Structure Generators

A fully random initialization of structures is not a good option, especially for large systems. Therefore, two random generators — random symmetry and random topology — are implemented in the code. The initial population is generated randomly; to preserve its diversity, random generation is used for obtaining a fraction of the subsequent populations.

**Random symmetry:** using 230 space groups, a well-ordered and chemically reasonable structure is generated by randomly placing an atom in a general Wyckoff position and multiplying it using the space group symmetry operations. If two or more atoms are placed closer to each other than a predefined threshold, they are merged into one atom put in a special Wyckoff position by averaging their coordinates.[4]

**Random topology:** structures are generated using the previously known topologies of crystal structures from the databases.[36]

### 2.2.4. Fingerprint Function

If a structure is allowed to be generated over and over again, especially if it is energetically favorable, the algorithm will be stuck at some local minima, making it very difficult to produce radically new structures. To avoid this inefficiency, the USPEX code provides the fingerprint

function that distinguishes between structures (e.g., using the bond lengths, angles, symmetry, stoichiometry, energy difference, etc.) and discards the identical ones.[37]

### 2.2.5. Degree of Order

The order parameter quantifies the degree of symmetry of the environment of a given atom; it is also possible to define the average degree of order of a fragment of a structure.[4] This is especially important when selecting the less defective part of the parent structure for heredity and the more defective part for mutation.

### 2.2.6. Perturbation

The relaxation does not break the symmetry, it can instead increase it to the supergroup.[4] To make the symmetry breaking possible during the relaxation, a small random displacement of the atomic position is applied to structures.

# CHAPTER 3.
# Multi-objective Optimization as a Tool for Materials Design


## Abstract

*In this chapter, we explain the concepts of Pareto optimality and Pareto dominance and use them in solving the multi-objective (MO) optimization problems. We discuss several different MO optimization methods and show how the MO optimization can be used for designing new materials. A simple Pareto-based MO optimization method is examined on several practical case studies to assess its efficiency in optimizing the double-objective problems.*



**Keywords:** Domination, Pareto dominance, hardness, convex hull


## 3.1. Introduction

Computational optimization is essential for solving a wide range of scientific, technological, and engineering problems. Generally, these problems can be classified into single-objective and multi-objective (MO) ones. Single-objective problems are those with only one objective to be optimized. In MO problems, several objectives, often conflicting, are optimized simultaneously, which makes these optimizations more complex than single-objective ones. An improvement in one objective usually leads to a degradation of the other ones. MO problems have a set of optimal solutions called the Pareto front, and the task of MO optimization methods is to find it. A decision maker can then select the preferable solutions from the set taking into account various considerations (e.g., the cost, expected ease of synthesis, etc.). Optimizing the energy alone results in a stable, usually (but not always) easy to synthesize phase which, however, does not necessarily possess the best properties. If only the physical properties are optimized, the theoretical limit for a property of interest can be achieved, but the energy of the predicted phase may be so high that the chances of its synthesis become negligible. The MO optimization of the energy and a target property is the best known way to find compounds of practical interest possessing attractive properties and a high chance, albeit not a guarantee, of being synthesizable. In this chapter, we show how several MO optimization methods can be used in solving the central problem of computational materials design: the discovery of new materials optimal in multiple properties. A simple Pareto-based MO optimization method is applied to several practical cases, demonstrating its efficiency with double-objective problems.

## 3.2. What Is the Pareto Front?

The goal of MO optimization methods is to find a set of optimal solutions with a trade-off between the objectives. This set of solutions is called the Pareto front. We use the Pareto front concept to rank the solutions in the population and give more weight to the best solutions (for more details see section 3.4.). Solution $S_1$ is said to be Pareto dominant compared to solution $S_2$ if it is at least as good as solution $S_2$ in every objective and better than solution $S_2$ in at least

one objective. A solution that is not dominated by any other solution is called a nondominated solution. A schematic representation of solutions on the Ashby plot[38] (a two-dimensional plot of any two objectives where each objective is plotted along one of the axes) is shown in Fig. 5. The goal is to find a set of solutions that minimize objectives *X* and *Y* in the best way. Solution *C* dominates solution *D* because it has a lower (or better) value of both objectives than solution *D*. Likewise, solutions *A* and *B* dominate solution *E*. Solutions *A*, *B*, and *C* do not dominate each other because solution *A* is better than solutions *B* and *C* in one objective but worse than them in the other objective. Solutions *A*, *B*, and *C* are nondominated solutions. The Pareto front is a set of nondominated solutions.

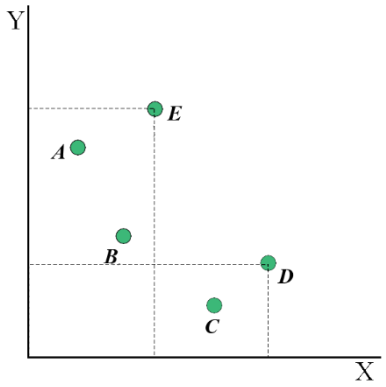



Figure 5. Illustration of the dominance concept.

In mathematical terms, for a set of solutions $\{S_n\}$ $(n = 1, \dots, N)$, with each solution having $F_m$ $(m = 1, \dots, M)$ objectives, the MO optimization can be formulated as follows:

$S = [S_1, S_2, \dots, S_N]$; # A set of solutions
$F = [F_1, F_2, \dots, F_M]$; # A set of objectives
$counter = 0$;
$P_{counter} = \emptyset$;
*while length*$(\{S\}) \neq 0$ # While there are solutions in the population
    $S = (S \cup P_{counter}) - (S \cap P_{counter})$ # Removing the nondominated solutions from the population
    *for* $(i,j = 1 : N,$ *and* $i \neq j)$
        *if* $F_k(S_i) \leq F_k(S_j)$ & $F_k(S_i) < F_k(S_n)$, $(k = 1,\dots, M, S_n \subseteq S \;\&\; S_n \neq \emptyset)$
            $S_i >> S_j$ ; # $S_i$ dominates $S_j$
        *end*
    *end*
    *for* $(i,j = 1 : N,$ *and* $i \neq j)$
        *if* $S_i >> S_j$
            $S^* = S_i$ ; # $S_i$ a nondominated solution
        *end*
    *end*
    $counter = counter + 1$;
    $\{P_{counter}\} = S^*$; # Copying the nondominated solutions to $\{P\}$
*end*

Solution $S^*$ that is not dominated by any other solution is called Pareto optimal. $\{P_t\}$ $(t = 1, \ldots, T,$ and $T \leq N)$ is a set of nondominated solutions, or the Pareto front. After the first Pareto front $\{P_1\}$ is found, it is removed from the main population and a search for nondominated solutions is conducted in the remaining population, yielding the second Pareto front $\{P_2\}$. The process is repeated until all solutions are classified in different Pareto fronts. In the worst situation, when each Pareto front includes only one solution, the number of Pareto fronts $T$ can be equal to the number of sets of solutions $N$. This usually happens when the objectives are not conflicting (i.e., in the case of linear relation between all objectives).

In a MO optimization, if one solution can be found which is more optimal than all other solutions in every objective, it is called the utopian solution. It always exists for single objective, whereas MO problems do not necessarily have a utopian solution.

## 3.3. Different MO Methods

In general, in an efficient Pareto-based MO optimization method, the distance to the optimal front should be minimized and the diversity of the generated solutions should be maximized. The latter is useful when more than two objectives are optimized. In this section, we discuss a simple Pareto ranking method, which is sufficient for a double-objective optimization, and some more complex MO optimization methods.

### 3.3.1. Layer Classification (a Simple Pareto Ranking)

The set of nondominated solutions, which form the Pareto front, is determined using the dominance concept: it is assigned the highest rank as having the highest probability of selection in the current population. Then it is temporarily removed, and the next Pareto front is found for the remaining population. This procedure is repeated until all solutions are classified into different Pareto fronts (Fig. 6). There is no priority in selecting the structures within the same Pareto front, which can be done using any binary method, like a tournament selection. The solutions of the first Pareto front have the highest probability of selection, followed by the second Pareto front, and so on. This method, which we call the layer classification Pareto technique, is used in some of the other methods introduced in this chapter. In some cases, however, it fails, particularly when the number of objectives increases, and therefore all of the following methods use additional techniques to obtain reliable results.

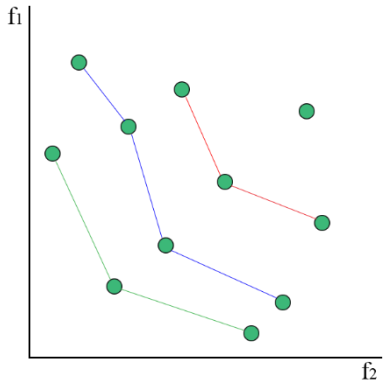



Figure 6. Layer classification of solutions in different Pareto fronts for minimizing two objectives.

### 3.3.2. Vector Evaluated Genetic Algorithm (VEGA)[39]

This method is not based on Pareto optimality techniques. It works as a typical genetic algorithm with one difference in the parent selection method. Let us consider an *m*-objective problem with maximum *N* number of best solutions that are allowed to be selected as parents in each generation. At the end of each generation, the population is ranked for each single objective and *N/m* of the best ranked solutions are added to a hypothetical archive. If a solution is optimal in *n* objectives ($n \leq m$), it is *n* times more likely to be selected than a solution optimal in only one objective. However, this method is susceptible to the shape of the Pareto front and fails to generate solutions that do not necessarily excel in one objective but are optimal in the sense of the Pareto front.[40]

### 3.3.3. Nondominated Sorting Genetic Algorithm (NSGA)

In the proposed NSGA,[41] the layer classification method is used and a sharing function is then applied to the solutions of each Pareto front so that some priority is given to the solutions in the less crowded regions. This sharing function is totally dependent on the choice of the user-predefined sharing parameter $\delta_{\text{share}}$ that denotes the largest value of distance metric within which two solutions share each other's fitness.[42] To overcome the user dependencies and other failures of this fitness sharing method,[41] in NSGA-II it was replaced with the "crowded comparison approach"[42] which also gives some priority to the solutions placed in the less crowded regions of each Pareto front but has no user-dependent parameters.

### 3.3.4. Pareto Envelope-Based Selection Algorithm (PESA)

In this method, only nondominated, or Pareto front, solutions are permitted to be ranked and selected. After the Pareto front is determined, a square (generally, 'hyperbox' in higher dimensional objective space) is defined as a result of finding the closest distance between two neighbor solutions for each objective. In a hyperbox that has the maximum number of solutions, each solution gets the minimum isolation value, and vice versa (Fig. 7). The probability of selecting these solutions is proportional to their isolation value. The goal is to give a higher probability of selection to solutions in the isolated regions than those in the crowded regions. The selection of individuals on the basis of the probability of selection given to each of them is called the individual-based selection. In PESA-II, it was replaced with the region-based selection, in which the isolation value is given to hyperboxes instead of solutions. As a result, hyperboxes get selected instead of solutions. The hyperbox that contains fewer solutions is considered more isolated than the one with more solutions, and hyperboxes with higher isolation values are preferred. After the hyperbox is selected, one of its solutions is picked randomly. This method, more efficient than the individual-based selection, is preferred by the author. More details on the distribution of probabilities and the selection mechanism are available in the original paper.[43]

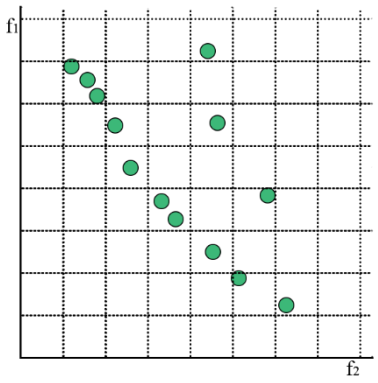



Figure 7. In PESA, hyperboxes are defined to help select the solutions.

### 3.3.5. Strength Pareto Evolutionary Algorithm (SPEA)

The first version of SPEA was published in 1999.[44] The method appeared to have serious problems in special situations and was claimed to be unsuccessful[42,43,45] in finding the optimal solution. In SPEA2 (2001),[45] the selective fitness assignment method was completely changed and a more solid algorithm was proposed. Here we discuss the fitness assignment in both methods and briefly mention the main failures of SPEA1.

In SPEA1,[44] the population is divided into two groups: external, composed of nondominated solutions, and internal, composed of dominated solutions. The optimal solutions can be selected from both groups, hence the fitness is assigned to all population members. The fitness is evaluated using a predefined strength function. For a nondominated solution, the strength is equal to the number of other solutions it dominates divided by the size of the internal group *N*. For a dominated solution, the strength is obtained by summing up the strengths of nondominated solutions that dominate it. The selective fitness is equal to the strength of the solution if the solution is a member of the external group. Otherwise, the selective fitness is equal to the strength of the solution plus one to make sure that nondominated solutions are always preferred to dominated ones (Fig. 8a). However, the method fails when many solutions are dominated by the same nondominated individuals and therefore have the same fitness. For example, in the particular case when only one nondominated solution exists, SPEA1 behaves like a random search algorithm. In addition, SPEA1 is mostly unsuccessful in the convergence and quick convergence with the optimal solutions.

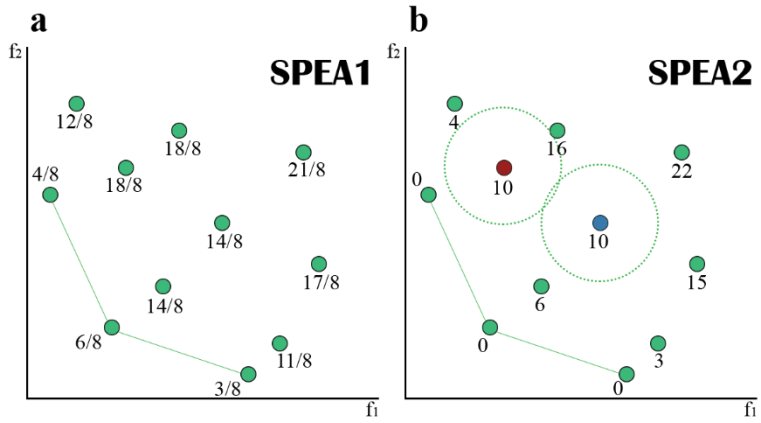



Figure 8. Different fitness assignment in (a) SPEA1 and (b) SPEA2 for the same distribution of solutions. According to the density estimation technique used in SPEA2, because the solution shown in blue lies in the less crowded region, it has a higher chance of being selected than the one shown in red.

To eliminate these deficiencies, completely different fitness assignment was proposed in SPEA2.[45] To improve the convergence to the best optimal solutions, an archive with a fixed size is defined, and only the solutions belonging to it are allowed to be selected. If the number of nondominated solutions equals the size of the archive, only these solutions are selected. If it exceeds the size of the archive, nondominated solutions from the dense regions (those containing many nondominated solutions close to each other) are eliminated from the archive. If the size of the archive is larger than the number of nondominated solutions, the free places are filled by dominated solutions according to their fitness. If many solutions have the same fitness, the density estimation technique is used. More details on the density estimation technique used in SPEA2 are available in the original paper.[45]

The strength in SPEA2 is computed for each solution, either dominated or nondominated, as the number of other solutions it dominates. The raw fitness is obtained by summing up the strengths of all solutions that dominate the target solution (therefore, the raw fitness is zero for all nondominated solutions). If the raw fitness is the same for two or more solutions, the fitness assigned to each solution is calculated as the sum of the raw fitness and the density of the region. In both SPEA1 and SPEA2, the fitness is to be minimized. Although SPEA2 was claimed to be very effective in the MO optimization, sometimes it fails to give a proper ranking to nondominated solutions. For example, in Fig. 8b, the solution shown in blue belongs to the third Pareto front but, because it is placed in the less crowded region, it has a better ranking than the one shown in red, belonging to the second Pareto front.

## 3.4. Combining the MO Optimization with USPEX for Materials Design

We performed the MO evolutionary search for materials optimal in hardness and stability (the energy above the convex hull) for $Fe_xB_y$, $Mo_xB_y$, and $Mo_xN_y$ binary systems using the USPEX code.[2,4] In these calculations, the simple Pareto ranking method (layer classification Pareto technique) was used, without any further procedures such as the fitness sharing, crowded comparison, clustering, or any other methods taking into account the density of solutions in the objective space. These additional procedures strengthen the MO Pareto-based methods when three or more objectives are under study. Although they can also be useful when dealing with

two objectives, we chose to keep it simple as long as it works well (this method was efficiently used in the prediction of new hard[46] and thermoelectric[47] materials).

In this method, at the end of each evolutionary generation, the fitness derived from the Pareto rank is assigned to each structure. The highest rank, or the lowest fitness, is assigned to the structures forming the first Pareto front, and these structures have the highest probability of being selected as parents. The structures of the next generation are created by applying heredity and mutation operators to these parents. The combination of the MO optimization and evolutionary search is expected be very efficient. The method always tries to fill the blank spaces in the Ashby plot in which structures with a combination of several optimal properties could be placed. These spaces are filled by creating new structures produced by the optimum parent structures selected from the same region of the Ashby plot. The blank spaces in the Ashby plot get filled with the good structures because only the optimum structures are selected from an unlimited number of those that can be created, whereas the bad structures are discarded from the population instantly (this method is also useful when a concave Pareto front is obtained). Finding all the optimum structures in these regions may take a long time and even then is not guaranteed. However, useful (even if approximate) Pareto fronts are obtained using this well-defined method which intelligently produces optimal solutions even when the time and computational power are limited.

The method of MO optimization discussed here is effective for two dimensions. Its efficiency decreases dramatically with an increase in the number of objectives, or properties to be studied for each material, because of an increase in the number of optimal structures. The dimensionality of the Pareto front is proportional to the number of properties that we wish to optimize. For example, when the dimensionality increases from two to three, the Pareto front, which was a one-dimensional line, becomes a two-dimensional plane. Similarly, for four properties, the Pareto front becomes a 3D hyperplane, and so on. The number of optimal structures placed on the Pareto front increases and the intelligent evolutionary algorithm becomes useless with a random parent selection.

Several examples of crystal structure prediction using the evolutionary methodology USPEX in combination with the simple Pareto ranking (layer classification) method are presented here. The goal of these calculations is to find hard and stable structures in target systems using the MO optimization method and assess the efficiency of this method and the evolutionary crystal structure prediction. In these examples, all the ab initio calculations were done using the PBE exchange-correlation functional[17] and the projector augmented-wave (PAW) method[48,49] as implemented in the VASP code[50] with the plane-wave basis set cutoff of 550 eV and the Brillouin zone sampling using a grid with a spacing of $2\pi \times 0.06$ Å$^{-1}$. The evolutionary calculations for each binary system had the initial population size of 120, the subsequent population size of 60 structures, and were run for 50 generations. The hardness was computed using the Lyakhov–Oganov model.[51]

### 3.4.1. Example 1: $Mo_xN_y$

Stable compounds of the $Mo_xN_y$ system were previously studied[52] using the single-objective optimization mode of USPEX, and the obtained results can be used to evaluate the efficiency of the simplest MO optimization method implemented in USPEX.

In the previous work,[52] MoN, $MoN_2$, and $Mo_4N_3$ were the stable compounds under zero pressure. Also, two very low-energy metastable $Mo_2N$ structures and one metastable $Mo_3N_2$ structure have been reported.[52] In this research, we found four stable compounds of this system: MoN, $MoN_2$, $Mo_2N$, and $Mo_3N_2$. The reason for this difference is that in this research, the stable $Mo_4N_3$ structure was not found, and, as a consequence, metastable structures of $Mo_2N$ and $Mo_3N_2$ appeared on the convex hull (Fig. 9a). It worth mentioning that since evolutionary algorithms do not guarantee finding the global minimum, failing to find a stable structure (global minimum energy) may occur in any SO or MO evolutionary algorithm, however, generally, the possibility of losing a global minimum in a SO evolutionary search should be less than a MO search. All other stable and metastable compounds reported in the previous work were found in this research. Moreover, a new metastable compound $Mo_5N_2$ with the energy of 2 meV/atom above the convex hull, which has never been reported before, was discovered in our calculations. We also found another structure of $Mo_2N$ with *Cmcm* space group which is a little more stable than the computationally predicted $I4_1/amd$ structure[52] of this compound. The stable $P6_3/mmc$-$MoN_2$ structure found in this work is in perfect agreement with the computationally found stable structure of $MoN_2$[52] and is lower in energy by 0.8 eV/atom, and therefore thermodynamically more stable, than the experimentally reported $R\bar{3}m$ structure (later shown by Yu et al. (2016) to be incorrect).

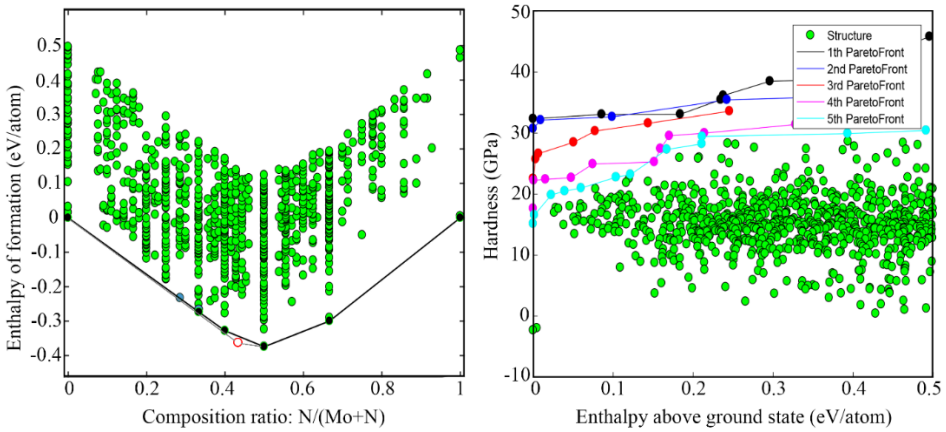



Figure 9. Convex hull diagram (left) and Ashby plot of hardness and energy above the convex hull (right) for the Mo–N system. The stable $Mo_4N_3$ compound reported in the literature is shown by a hollow red circle. The low-energy metastable $Mo_5N_2$ and *Cmcm* structure of $Mo_2N$ are shown in blue.

In this example, the MO optimization not only discovered several unknown low-energy metastable structures, but also successfully reported all the already studied stable and metastable structures of the $Mo_xN_y$ system except one, $Mo_4N_3$. This indicates that if two properties are optimized, the MO optimization method is almost as efficient as the single-objective optimization.

The Ashby plot of hardness and instability of the $Mo_xN_y$ system is shown in Fig. 9b. Several hard and stable or low-energy metastable structures were found in this system. Although most of the structures are in the hardness range of 10–20 GPa and energy range of 0.1–0.5 eV/atom above the convex hull, the algorithm tried to focus on the regions where optimal structures in both properties are located.

### 3.4.2. Example 2: $Fe_xB_y$

Using the same method to study the $Fe_xB_y$ system, in this work we found two stable binary compounds, FeB and $FeB_3$, and several metastable compounds: $FeB_2$, $FeB_3$, $FeB_4$, $Fe_2B$, $Fe_3B_5$, and $Fe_3B_{10}$. Some of these compounds have been reported to be stable or metastable in the literature,[53,54] whereas some were discovered in this work. The differences between our results and the findings reported in other works are discussed below.

In the literature, FeB, $FeB_2$, and $Fe_2B$ are reported to be stable, whereas our calculations showed $Fe_2B$ and $FeB_2$ to be metastable by 3 meV/atom and 42 meV/atom, respectively. Although both these compounds were found in this research, their lowest-energy structures have not yet been discovered. Two structures of $FeB_3$ were found in our calculations. Most likely, this compound would not appear on the convex hull if the stable structure of $FeB_2$ were found. Nevertheless, $FeB_3$ is not reported either as stable or metastable in the literature. $FeB_4$ was found to be metastable in this work, in agreement with the earlier studies.[53,54] $Fe_4B$ is one of the reported metastable compounds that were not detected in our calculations. In this work, several simultaneously hard and stable or low-energy metastable phases were detected for the $Fe_xB_y$ system, and some of them are superhard (Fig. 10b).

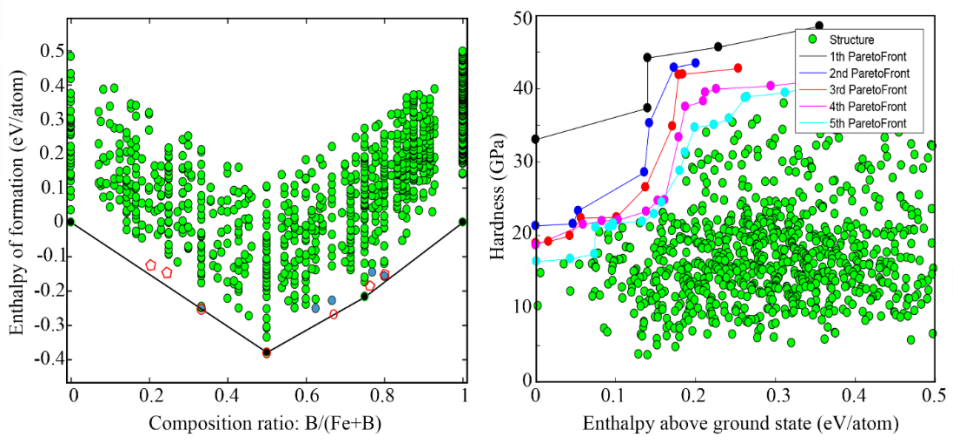



Figure 10. Convex hull diagram (left) and Ashby plot of hardness and energy above the convex hull (right) for the Fe–B system. The stable and metastable compounds reported in the literature are shown by hollow red circles and polygons. The low-energy metastable structures found in this work are shown in blue.

### 3.4.3. Example 3: $Mo_xB_y$

For this system, MoB and $MoB_2$ were found to be stable, in full agreement with the previous studies.[55,56] Many metastable $Mo_xB_y$ compounds have been reported both theoretically[57] and experimentally:[55] $MoB_3$, $MoB_4$, $Mo_2B$, $Mo_2B_5$, $Mo_3B_2$, and $Mo_5B_3$. In this search, we found eight low-energy metastable compounds: $Mo_2B$, $MoB_3$, $MoB_5$, $Mo_3B_5$, $Mo_3B_2$, $Mo_4B$, $Mo_5B_2$, and $Mo_2B_3$. Three of them, $Mo_2B$, $Mo_3B_2$, and $MoB_3$, have already been reported as metastable. Five previously unknown compounds discovered in this work — $MoB_5$, $Mo_2B_3$, $Mo_3B_5$, $Mo_4B$, and $Mo_5B_2$ — have their energies lying above the convex hull by 55 meV/atom, 29 meV/atom, 57 meV/atom, 23 meV/atom, and 21 meV/atom, respectively. Although three reported metastable compounds, $MoB_4$, $Mo_2B_5$, and $Mo_5B_3$, were not found in

our calculations, the MO optimization acted successfully, discovering five new low-energy metastable compounds of the $Mo_xB_y$ system, which indicates the efficiency of the method in crystal structure prediction when several properties are to be optimized (see Fig. 11).

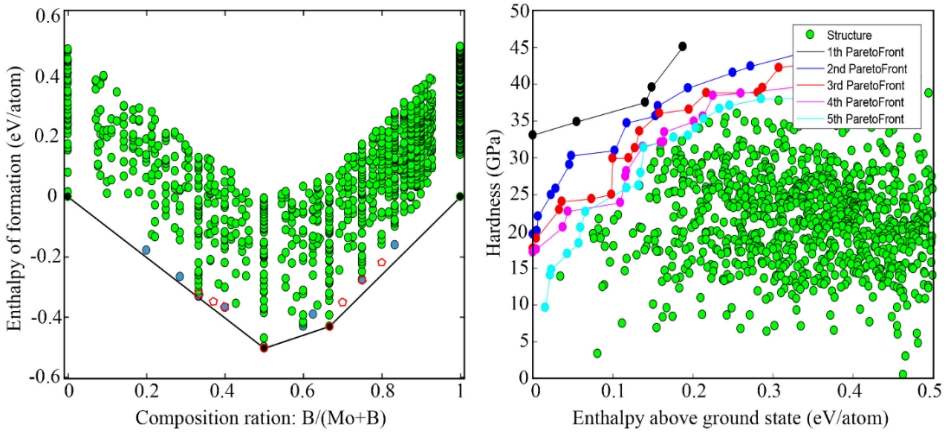



Figure 11. Convex hull diagram (left) and Ashby plot of hardness and energy above the convex hull (right) of the Mo–B system. The metastable compounds reported in previous studies are shown by hollow red polygons, the compounds discovered in this work are shown in blue.

The stable binary compounds with the highest Lyakhov–Oganov hardness for each of the above examples are $MoN_2$ (32.6 GPa), $FeB_3$ (20 GPa), and $MoB_2$ (19.6 GPa). These phases have $P6_3/mmc$, $P2_1/m$, and $R\bar{3}m$ space groups, respectively (Fig. 12). Whereas the Lyakhov–Oganov model of hardness is convenient for screening, being very cheap and numerically robust, more accurate values are predicted using Chen's model.[58] The latter, however, requires the calculation of the elastic tensor and is therefore much more computationally expensive. Therefore, after the initial screening using Lyakhov-Oganov model, the final hardnesses were computed using Chen's model. In Fig. 10b and 11b, the hardest stable compound (with a hardness of 33.6 GPa) corresponds to pure α-boron.

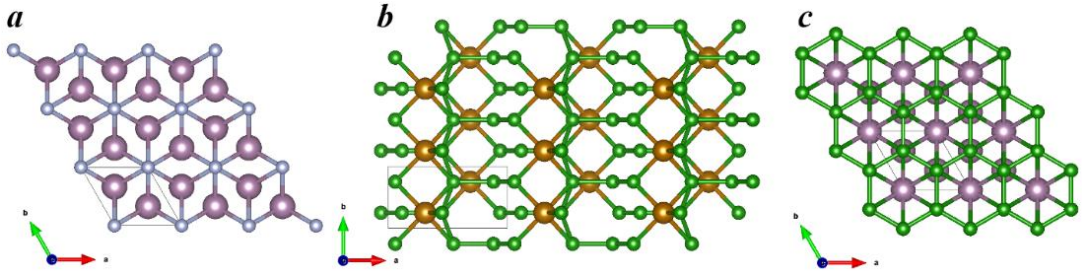



Figure 12. Structures of stable binary compounds with the highest Lyakhov–Oganov hardness among the systems studied in this work: (a) $P6_3/mmc$-$MoN_2$ (Chen's hardness 22.3 GPa), (b) $P2_1/m$-$FeB_3$ (Chen's hardness 30.2 GPa), (c) $R\bar{3}m$-$MoB_2$ (Chen's hardness 28.5 GPa).

## 3.5. Conclusion

In this chapter, some of the most prominent multi-objective optimization methods were discussed, and the layer classification Pareto technique in combination with USPEX was used in the search for simultaneously hard and stable materials. The most desirable and favorable outcome of materials design is the discovery of new materials with simultaneously optimal

multiple properties because, in our examples, unstable hard materials are as useless as stable soft materials. We need to have materials that are optimal in all required properties, and the MO optimization method can be used as a verified tool for attaining this goal. The combination of the evolutionary search and MO Pareto ranking provides a powerful means for the discovery of such new materials.

# CHAPTER 4.

# Nonempirical Definition of the Mendeleev Numbers: Organizing the Chemical Space


## Abstract

*Organizing a chemical space so that elements with similar properties would take neighboring places in a sequence can help to predict new materials. In this paper, we propose a universal method of generating such a one-dimensional sequence of elements, i.e. at arbitrary pressure, which could be used to create a well-structured chemical space of materials and facilitate the prediction of new materials. This work clarifies the physical meaning of Mendeleev numbers, which was earlier tabulated by Pettifor. We compare our proposed sequence of elements with alternative sequences formed by different Mendeleev numbers using the data for hardness, magnetization, enthalpy of formation, and atomization energy. For an unbiased evaluation of the MNs, we compare clustering rates obtained with each system of MNs.*



**Keywords:** Mendeleev number, chemical scale, Pettifor map


## 4.1. Introduction

Vast amounts of information about the physical properties and crystal structures of materials have been produced and need to be organized in a clear way to facilitate insight. Even for known materials many properties remain unexplored, and a clear organization of data similar to Mendeleev's Periodic Table would help to estimate these properties *a priori* and uncover those regions of the chemical space that deserve a deeper study. To solve this challenging problem, it is necessary to construct a coherent chemical space, basically a coordinate system, in which materials with similar properties are closely related and likely to be placed in neighboring regions. This way, prediction of one material would lead to predictions of other materials with similar or perhaps even better properties.

This idea of a chemical space can be explained on a simple example of a set of colored pencils, in which the pencils are put in an order so that the color variation between the adjacent pencils is minimal (Fig. 13). In this example, the pencils represent the elements of the Periodic Table while the colors represent their properties. A combination of two different colors can be considered a binary system in which fractions of colors represent the composition (stoichiometry), while the resulting color shows the properties of the system. A two-dimensional color map, built in such a way, represents a chemical space where binary systems with similar properties are located close to each other, which is the direct result of a suitable one-dimensional arrangement of the elements.

A similar idea of "structure map" was explored in 1984 by Pettifor,[59] who suggested that a well-structured chemical space can be derived by changing the sequence of the elements in the Periodic Table.[59] He proposed a chemical scale that determines the "distance" between the elements on a one-dimensional axis and a Mendeleev number (MN) — an integer showing the

position of an element in the sequence.[60] Pettifor claimed that binary compounds with the same structure type occupy the same region in a two-dimensional map plotted using the MNs (the Pettifor map). He evaluated the chemical scale by presenting a map clearly separating 34 different structure types of 574 binary *AB* compounds (Fig. 14a).[59] Later, Pettifor showed that the MN approach also works for other $A_xB_y$ compounds.[60] Although Pettifor derived the chemical scale and Mendeleev number empirically and based his assessment on only several hundred binary compounds, his study provided a phenomenally successful ordering of the elements confirmed in many later works.[61,62] In this work we denote Pettifor's MN as $MN_P$. We expect that a nonempirical method of finding the MNs would perform even better.

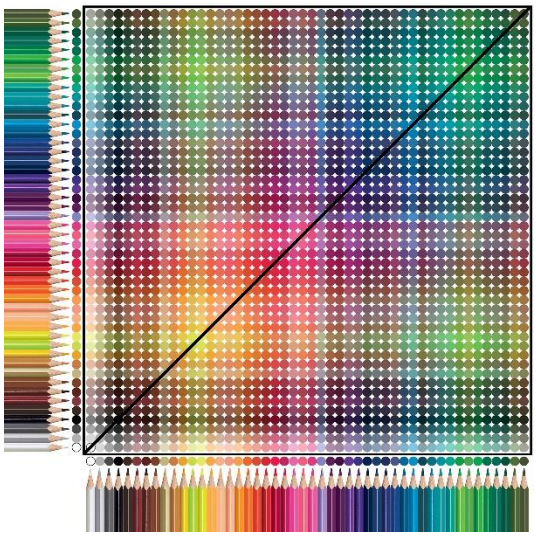


Figure 13. A colored pencil diagram demonstrating the idea of chemical space.

Earlier, in 1929, Goldschmidt tried to find a systematic relationship between the chemical composition and crystal structures of materials. His goal, in particular, was to find how a crystal structure (the geometric arrangement of atoms in a crystal) depends on the chemical composition. The result of his work, known as Goldschmidt's law of crystal chemistry, states that the crystal structure is determined by stoichiometry, atomic size, and polarizability of atoms/ions.[63] In 1955, Goldschmidt's law was modified by Ringwood when he added the electronegativity as another important parameter determining the crystal structure.[64] Based on this premise, we define the chemical scale and MN from these atomic properties.

In 2008, Villars et al. propose a different enumeration of the elements (called periodic number – PN), emphasizing the role of valence electrons.[65] In contrast to the atomic number (AN), PN depends in details on the underlying Periodic Table of the elements.

In 2016, Glawe et al. proposed another sequence of elements (modified MN – in this work we show by $MN_m$) based on their similarity, defining elements A and B to be similar if they crystallize in the same structure type when combined with other elements of the Periodic Table. For example, the alkali metals (Li, Na, K, etc.), forming the rocksalt crystal structure when mixed with Cl, are similar by this definition.[62] Applying this definition and using the available crystal structures in the Inorganic Crystal Structure Database (ICSD),[66] the degrees of similarity of each element with respect to other elements were calculated. Based on these data, the best sequence of elements was optimized using a genetic algorithm, so that similar elements occupy neighboring places in this arrangement.

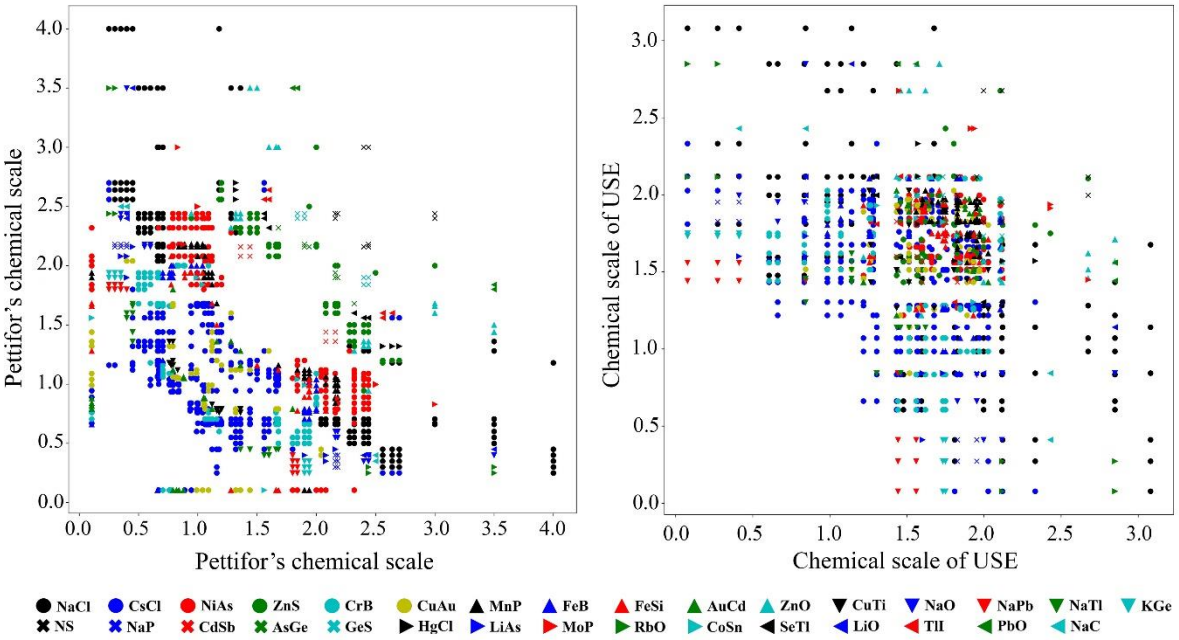



Figure 14. Structure maps of 521 binary *AB* compounds using Pettifor's chemical scale and our redefined chemical scale.

However, defining the MNs with the help of databases has its drawbacks. The first and most important one is that the calculations of the MNs in this case are property-dependent. The quality of the results is lowered because all the structures in the ICSD were taken into account, including theoretical and experimental, stable and metastable at the same time. Also, note that for some elements the data in the ICSD are insufficient.

In this paper, we present a simple, physically meaningful, fully nonempirical universal method of defining the MNs and obtaining the universal sequence of elements (USE). We then compare different MNs using our own theoretical database, which contains about 500,000 crystal structures.

As the chemistry of the elements and materials changes under pressure, so will the MNs. The proposed universal method makes it possible to define the MNs of the elements by their electronegativity and atomic radius at any pressure. In Section 3, we use these properties to compute MNs of a number of elements at high pressures (50 GPa, 200 GPa, and 500 GPa).

## 4.2. Computational Methods

Unlike Pettifor, who derived his $MN_P$ empirically, we offer a nonempirical (and, therefore, more universal) definition. The most important chemical properties of an atom are the radius $R_a$, electronegativity $\chi$, polarizability $\alpha$, and valence $v$. We disregarded the polarizability in favor of the electronegativity because they are strongly correlated.[67] For simplicity, we also excluded the valence, which is not constant for many elements. Thus, we only consider the electronegativity and atomic radius to define the MNs and obtain the USE (Table 2).

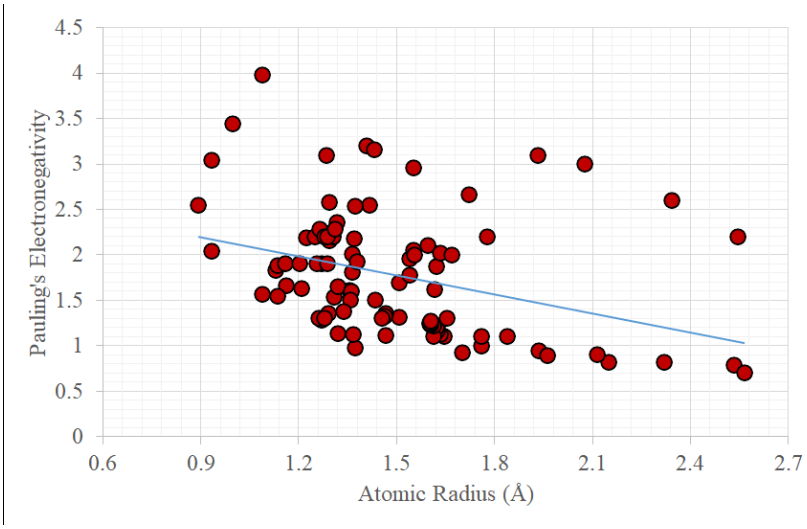



Figure 15. Electronegativities and atomic radii of the elements. The regression line is shown in blue.

We used the Pauling scale for the electronegativity $\chi$.[68] For each element there are many values of atomic radius depending on the bonding type (ionic, covalent, metallic, and van der Waals), oxidation state, and coordination number. The problem is that we need to use values obtained in a consistent way for all elements, and such values were not available. In this work, the atomic radius $R_a$ is defined as half the shortest interatomic distance in the relaxed simple cubic structure of an element. A significant correlation between Pauling's electronegativity $\chi$ and atomic radius $R_a$ (Fig. 15) means that one of them or better some combination of the two can be used as a single parameter approximately characterizing the chemistry of an element. To find an approximate combination of these two parameters into one, the regression line in the space of $\chi$ and $R_a$ was computed and all the elements were projected onto it (Fig. 15). The zero value on this scale was assigned to the projection of the first element (the one having a large atomic radius and low electronegativity) onto the regression line, while the coordinates of other elements on the line were defined as the distance of their projections from zero – these are defined as the chemical scale. The Mendeleev number, USE, was defined as the sequential number of the projected element on the regression line (see Table 3).

Table 2. Electronegativities and atomic radii of the elements used for obtaining the universal sequence of elements (USE).

| Element | Atomic radius $R_a$ (Å) | Pauling Electronegativity (χ) | Element | Atomic radius $R_a$ (Å) | Pauling Electronegativity (χ) |
|---|---|---|---|---|---|
| H | 0.727 | 2.2 | In | 1.541 | 1.78 |
| He | 1.286 | 3.1 | Sn | 1.541 | 1.96 |
| Li | 1.374 | 0.98 | Sb | 1.553 | 2.05 |
| Be | 1.090 | 1.57 | Te | 1.596 | 2.1 |
| B | 0.933 | 2.04 | I | 1.721 | 2.66 |
| C | 0.891 | 2.55 | Xe | 2.344 | 2.6 |
| N | 0.932 | 3.04 | Cs | 2.535 | 0.79 |
| O | 0.997 | 3.44 | Ba | 1.962 | 0.89 |
| F | 1.089 | 3.98 | La | 1.647 | 1.1 |
| Ne | 1.409 | 3.2 | Ce | 1.467 | 1.12 |
| Na | 1.701 | 0.93 | Pr | 1.367 | 1.13 |
| Mg | 1.508 | 1.31 | Nd | 1.320 | 1.14 |
| Al | 1.355 | 1.61 | Pm | 1.635 | 1.13 |
| Si | 1.269 | 1.9 | Sm | 1.626 | 1.17 |
| P | 1.223 | 2.19 | Eu | 1.620 | 1.2 |
| S | 1.293 | 2.58 | Gd | 1.623 | 1.2 |
| Cl | 1.431 | 3.16 | Tb | 1.613 | 1.1 |
| Ar | 1.933 | 3.1 | Dy | 1.613 | 1.22 |
| K | 2.151 | 0.82 | Ho | 1.604 | 1.23 |
| Ca | 1.761 | 1 | Er | 1.602 | 1.24 |
| Sc | 1.466 | 1.36 | Tm | 1.602 | 1.25 |
| Ti | 1.308 | 1.54 | Yb | 1.759 | 1.1 |
| V | 1.209 | 1.63 | Lu | 1.605 | 1.27 |
| Cr | 1.162 | 1.66 | Hf | 1.454 | 1.3 |
| Mn | 1.136 | 1.55 | Ta | 1.358 | 1.5 |
| Fe | 1.131 | 1.83 | W | 1.316 | 2.36 |
| Co | 1.137 | 1.88 | Re | 1.287 | 1.9 |
| Ni | 1.160 | 1.91 | Os | 1.278 | 2.2 |
| Cu | 1.203 | 1.9 | Ir | 1.288 | 2.2 |
| Zn | 1.320 | 1.65 | Pt | 1.311 | 2.28 |
| Ga | 1.365 | 1.81 | Au | 1.374 | 2.54 |
| Ge | 1.365 | 2.01 | Hg | 1.556 | 2 |
| As | 1.369 | 2.18 | Tl | 1.617 | 1.62 |
| Se | 1.418 | 2.55 | Pb | 1.622 | 1.87 |
| Br | 1.551 | 2.96 | Bi | 1.635 | 2.02 |
| Kr | 2.077 | 3 | Po | 1.670 | 2 |
| Rb | 2.319 | 0.82 | At | 1.777 | 2.2 |
| Sr | 1.935 | 0.95 | Rn | 2.544 | 2.2 |
| Y | 1.625 | 1.22 | Fr | 2.567 | 0.7 |
| Zr | 1.463 | 1.33 | Ra | 2.114 | 0.9 |
| Nb | 1.362 | 1.6 | Ac | 1.838 | 1.1 |
| Mo | 1.294 | 2.16 | Th | 1.655 | 1.3 |
| Tc | 1.257 | 1.9 | Pa | 1.436 | 1.5 |
| Ru | 1.249 | 2.2 | U | 1.339 | 1.38 |
| Rh | 1.264 | 2.28 | Np | 1.291 | 1.36 |
| Pd | 1.306 | 2.2 | Pu | 1.271 | 1.28 |
| Ag | 1.379 | 1.93 | Am | 1.261 | 1.3 |
| Cd | 1.509 | 1.69 | Cm | 1.279 | 1.3 |

## 4.3. Results and Discussion

In a well-ordered sequence of elements, the atoms with similar properties are close to each other. Therefore, in the two-dimensional chemical space based on such sequence, the properties of neighboring binary systems should exhibit a close relation. On this premise, we evaluate different MNs: atomic number (AN), Villars' periodic number [65] (PN), Pettifor's Mendeleev number [60] ($MN_P$), modified Mendeleev number [62] ($MN_m$), and Mendeleev number in this work, the universal sequence of elements, (USE). These MNs are shown in Table 3.

To examine different MNs, a database containing about 500,000 theoretical and experimental crystal structures of unary and binary compounds was compiled. These structures were relaxed using density functional theory within the generalized gradient approximation (DFT-GGA) and the database was set up so as to contain neither duplicates nor very unstable structures (whose energy is more than 0.5 eV/atom above the convex hull). Some crystal structures in the database were imported from other online databases, such as ICSD [66] and COD,[69] while the majority came from the previous calculations based on the evolutionary algorithm USPEX.[2–4]

The database contains the crystal structure information for 1,591 binary and 80 unary systems – excluding Ar, Xe, Ce, Nd, Pm, Sm, Eu, Gd, Tb, Dy, Ho, Er, Tm, Yb, Lu, Rn, Bk, Cf, Es, Fm, Md, No, Lr, Rf, and Db. Of these, only 446 systems have the magnetic information that are obtained in several multi-objective evolutionary searches for low-energy and highly magnetized phases, as implemented in the USPEX algorithm.[70] The hardness of all crystal structures in this database was computed using the Lyakhov-Oganov model.[51] The database is fully consistent because all crystal structures were relaxed and their energies computed in the same settings using the density functional theory with the projector-augmented wave method (PAW) and PBE [17] functional as implemented in the VASP code.[48,50] To compare the performance of different MNs for binary systems, the 2D maps of various properties were plotted, among them the hardness (representing the mechanical properties), magnetization (electronic properties), enthalpy of formation, and atomization energy (thermodynamic and chemical properties).

For hardness and magnetization, the representative structure of each binary system is a structure with the energy less than 0.1 eV/atom above the convex hull, having the highest hardness or magnetization, respectively. For generating the chemical spaces of the enthalpy of formation and atomization energy, the representative structure of each binary system is a structure with the lowest enthalpy of formation or lowest atomization energy, respectively. In all cases, no restrictions on stoichiometries of studied structures were imposed. The generated chemical spaces of hardness, magnetization, enthalpy of formation, and atomization energy using different MNs are shown in Fig. 16, 17, 18, and 19, respectively.

Table 3 . The universal sequence of elements (USE), coordinates of the elements on the regression line – chemical scale (CS), atomic number (AN), periodic number [65] (PN), Pettifor's Mendeleev number [60] ($MN_P$), modified MN [62] ($MN_m$).

| # | USE | CS | AN | PN | $MN_P$ | $MN_m$ | # | USE | CS | AN | PN | $MN_P$ | $MN_m$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Fr | 0 | H | Li | He | He | 51 | Bi | 1.517 | Sb | Re | V | V |
| 2 | Cs | 0.077 | He | Na | Ne | Ne | 52 | Sn | 1.56 | Te | Fe | W | Cr |
| 3 | Rb | 0.272 | Li | K | Ar | Ar | 53 | Zn | 1.566 | I | Ru | Mo | Mo |
| 4 | K | 0.411 | Be | Rb | Kr | Kr | 54 | Hg | 1.571 | Xe | Os | Cr | W |
| 5 | Ra | 0.486 | B | Cs | Xe | Xe | 55 | Te | 1.594 | Cs | Co | Tc | Re |
| 6 | Ba | 0.606 | C | Fr | Rn | Rn | 56 | Sb | 1.601 | Ba | Rh | Re | Tc |
| 7 | Sr | 0.662 | N | Ca | Fr | Fr | 57 | Ga | 1.62 | La | Ir | Mn | Os |
| 8 | Ac | 0.827 | O | Sr | Cs | Cs | 58 | V | 1.646 | Ce | Ni | Fe | Ru |
| 9 | Ca | 0.834 | F | Ba | Rb | Rb | 59 | Mn | 1.661 | Pr | Pd | Os | Ir |
| 10 | Na | 0.843 | Ne | Ra | K | K | 60 | Ag | 1.676 | Nd | Pt | Ru | Rh |
| 11 | Rn | 0.871 | Na | Sc | Na | Na | 61 | Cr | 1.702 | Pm | Cu | Co | Pt |
| 12 | Yb | 0.892 | Mg | Y | Li | Li | 62 | Be | 1.71 | Sm | Ag | Ir | Pd |
| 13 | La | 0.984 | Al | La | Ra | Ra | 63 | Kr | 1.71 | Eu | Au | Rh | Au |
| 14 | Pm | 1.011 | Si | Ac | Ba | Ba | 64 | Ge | 1.733 | Gd | Be | Ni | Ag |
| 15 | Tb | 1.012 | P | Ce | Sr | Sr | 65 | Re | 1.735 | Tb | Mg | Pt | Cu |
| 16 | Sm | 1.041 | S | Th | Ca | Ca | 66 | Si | 1.75 | Dy | Zn | Pd | Ni |
| 17 | Gd | 1.061 | Cl | Pr | Yb | Eu | 67 | Tc | 1.76 | Ho | Cd | Au | Co |
| 18 | Eu | 1.063 | Ar | Pa | Eu | Yb | 68 | Cu | 1.804 | Er | Hg | Ag | Fe |
| 19 | Y | 1.071 | K | Nd | Y | Lu | 69 | I | 1.81 | Tm | B | Cu | Mn |
| 20 | Dy | 1.081 | Ca | U | Sc | Tm | 70 | Fe | 1.824 | Yb | Al | Mg | Mg |
| 21 | Th | 1.091 | Sc | Pm | Lu | Y | 71 | As | 1.827 | Lu | Ga | Hg | Zn |
| 22 | Ho | 1.094 | Ti | Np | Tm | Er | 72 | Ni | 1.845 | Hf | In | Cd | Cd |
| 23 | Er | 1.101 | V | Sm | Er | Ho | 73 | Co | 1.847 | Ta | Tl | Zn | Hg |
| 24 | Tm | 1.107 | Cr | Pu | Ho | Dy | 74 | Mo | 1.877 | W | C | Be | Be |
| 25 | Lu | 1.116 | Mn | Eu | Dy | Tb | 75 | Ar | 1.885 | Re | Si | Tl | Al |
| 26 | Li | 1.141 | Fe | Am | Tb | Gd | 76 | Pd | 1.89 | Os | Ge | In | Ga |
| 27 | Ce | 1.144 | Co | Gd | Gd | Sm | 77 | Ir | 1.905 | Ir | Sn | Al | In |
| 28 | Mg | 1.218 | Ni | Cm | Sm | Pm | 78 | Os | 1.913 | Pt | Pb | Ga | Tl |
| 29 | Pr | 1.232 | Cu | Tb | Pm | Nd | 79 | Pt | 1.931 | Au | N | Pb | Pb |
| 30 | Hf | 1.257 | Zn | Bk | Nd | Pr | 80 | Ru | 1.937 | Hg | P | Sn | Sn |
| 31 | Xe | 1.263 | Ga | Dy | Pr | Ce | 81 | P | 1.953 | Tl | As | Ge | Ge |
| 32 | Zr | 1.266 | Ge | Cf | Ce | La | 82 | Rh | 1.97 | Pb | Sb | Si | Si |
| 33 | Nd | 1.276 | As | Ho | La | Ac | 83 | W | 1.973 | Bi | Bi | B | B |
| 34 | Sc | 1.281 | Se | Es | Fm | Th | 84 | Se | 1.997 | Po | O | Bi | C |
| 35 | Tl | 1.304 | Br | Er | Es | Pa | 85 | Au | 2.027 | At | S | Sb | N |
| 36 | Pa | 1.385 | Kr | Fm | Cf | U | 86 | B | 2.106 | Rn | Se | As | P |
| 37 | Pu | 1.396 | Rb | Tm | Bk | Np | 87 | S | 2.116 | Fr | Te | P | As |
| 38 | U | 1.397 | Sr | Yb | Cm | Pu | 88 | Br | 2.12 | Ra | Po | Po | Sb |
| 39 | Cm | 1.401 | Y | Lu | Am | Am | 89 | Cl | 2.332 | Ac | H | Te | Bi |
| 40 | Am | 1.416 | Zr | Ti | Pu | Cm | 90 | H | 2.366 | Th | F | Se | Po |
| 41 | Np | 1.425 | Nb | Zr | Np | Bk | 91 | Ne | 2.373 | Pa | Cl | S | Te |
| 42 | Cd | 1.433 | Mo | Hf | U | Cf | 92 | He | 2.418 | U | Br | C | Se |
| 43 | Pb | 1.442 | Tc | V | Pa | Es | 93 | C | 2.43 | Np | I | At | S |
| 44 | Ta | 1.449 | Ru | Nb | Th | Fm | 94 | N | 2.675 | Pu | At | I | O |
| 45 | In | 1.458 | Rh | Ta | Ac | Sc | 95 | O | 2.849 | Am | He | Br | At |
| 46 | Po | 1.477 | Pd | Cr | Zr | Zr | 96 | F | 3.08 | Cm | Ne | Cl | I |
| 47 | At | 1.502 | Ag | Mo | Hf | Hf | 97 | | | Bk | Ar | N | Br |
| 48 | Nb | 1.503 | Cd | W | Ti | Ti | 98 | | | Cf | Kr | O | Cl |
| 49 | Ti | 1.513 | In | Mn | Nb | Ta | 99 | | | Es | Xe | F | F |
| 50 | Al | 1.514 | Sn | Tc | Ta | Nb | 100 | | | Fm | Rn | H | H |

### 4.3.1. Evaluation of a chemical space

In a correctly defined chemical space, closely located materials should have similar properties. The most promising materials will then be clustered in one or few "islands" in this space. To predict new materials, it could be sufficient to explore these islands instead of the entire chemical space. The fewer these islands are, the easier it would be to locate and explore them for promising materials. A chemical space containing many small islands is less amenable for the prediction of materials than the one with fewer big islands. Therefore, for evaluating each chemical space, it is useful to find these islands and calculate the number of (similar) materials they cover.

For doing this, we used the idea of the clustering algorithm proposed by Rodriguez and Laio [71] and applied it to clustering regions of the chemical space on the basis of their similarity. In this simple method, each cluster is defined by a cluster center and a number of similar data points around it. For finding the cluster centers, two quantities are to be calculated for each data point *i*: its local density $\rho_i$, and its distance $\delta_i$ from the nearest point with a higher density. In the original method, $\rho_i$ is equal to the number of points that are closer than $d_c$ to the point *i* (we call these points: local neighbors), where $d_c$ is a cutoff radius. Also, $\delta_i$ for the point with the highest density is equal to its distance from the furthest data point. This way, the cluster centers are those points with high value of both $\rho$ and $\delta$. Clearly, the point with the highest density $\rho_i$, is always a cluster center.

In our modified method, we only consider the point with the highest density as a cluster center, and therefore, there is no need for calculation of $\delta_i$. Then, we remove the cluster center and all its local neighbors from the dataset, we calculate $\rho_i$ again for the remaining data points, and find a new cluster center. We continue this loop until all the data points are assigned to a cluster. The points with zero local density $\rho_i$ are isolated points. In our method, $\rho_i$ is equal to the number of points that are closer than $d_c$ to the point *i*, and their property difference to the point *i* is less than $d_p$, where $d_p$ is a property difference cutoff. We need to clarify that data points closer than $d_c$ to a local neighbor – neighbors of the local neighbors – with property difference less than $d_p$ from the cluster center, are also included in the cluster and considered as local neighbors of the cluster center, but these points are not included in calculation of local density $\rho_i$ in the first place.

The number of clusters (i.e. islands) that cover all binary systems in the chemical spaces of the MNs, is a good quantitative evaluation of these MNs. The lower the number of clusters, the better-clustered the chemical space. However, as cutoff values, i.e. $d_c$ and $d_p$, are increased, the number of clusters decreases (see Fig. 20).

For finding the cluster centers, the constant cutoff radius $d_c$ equal to 5 blocks was used – clusters expand by including neighbors of the local neighbors as mentioned above, and $d_c$ is only used to bound the neighborhood area of each system. To see how number of clusters in different MNs changes with respect to $d_p$, we let this value changes as shown in Fig. 20 and Fig. 21.

Another quantitative evaluation of the MNs is the number of systems that are covered by clusters. For this purpose, we define an imaginary "ideal MN" ($MN_{ideal}$) for each property, which clusters all the materials in a minimum number of clusters ($N_{min}$) in the target chemical

space. $N_{min}$ can be easily calculated by having the property range of distributed systems in a chemical space (as shown in the color bar of Fig. 16, 17, 18, and 19) and $d_p$ (maximum property difference between a cluster member and the cluster center) – the range of this value with regard to the change of the $d_p$ is shown in Table 4. Therefore, our second evaluation criterion is the fraction of all systems that are covered by the first (biggest) $N_{min}$ clusters – the results of this evaluation are shown in Fig. 21. These two evaluations provide an insight into the clustering rate of different MNs.

As mentioned earlier, only 1591 binary and 80 unary systems are studied in our database which is about a half of the total binary and unary systems that can be created from the combination of 80 elements – totally 3240 systems can be created. Of these, hardness, enthalpy of formation, and atomization energy are presented for almost all the studied systems (about 50% of total systems) while magnetization was computed only in 446 systems (about 14% of total systems). The amount of missing information can influence the correct clustering of the chemical space – for example, when a cluster cannot expand because of the lack of data points around it, and not because of the existence of dissimilar systems around it. To solve this problem, we assigned a value to the property of each missing system by cubic interpolation of its neighbors' property in the scale of each MN. Then the property of the missing system is calculated as the average of its values in different MN scales – in the spirit of the *committee voting* approach. We evaluated our *committee voting* approach, by removing materials with explicitly calculated properties in our database, and predicting their properties using *committee voting*. On average, the error (difference between the predicted and calculated values) of the predicted values are: 3.24 GPa for hardness, 0.014 $\mu_\beta/Å^3$ for magnetization, 0.175 eV/atom for the enthalpy of formation, and 0.48 eV/atom for atomization energy – between 3.5% (for enthalpy of formation) to 7% (for magnetization) of the property ranges.

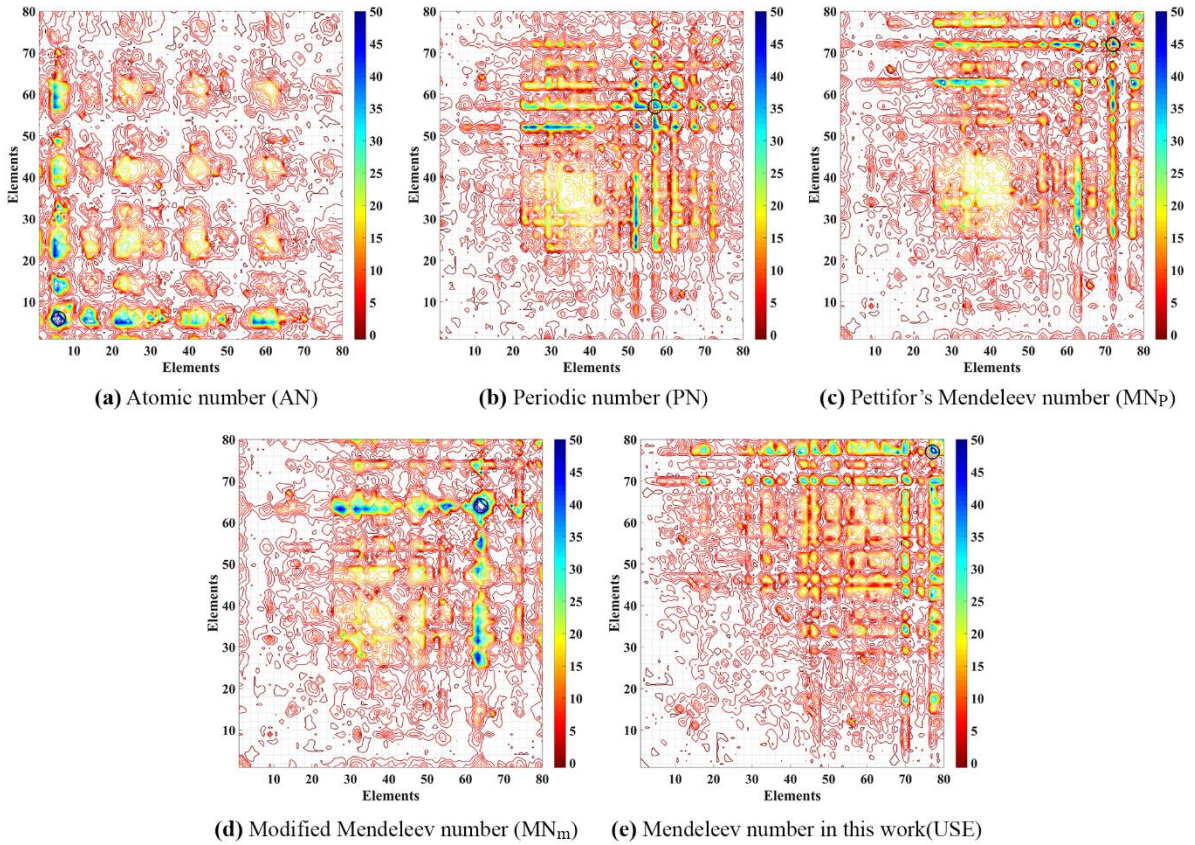
In the following, we discuss different MNs by calculating their clustering rate and visualizing their 2D maps (Pettifor maps) of the hardness, magnetization, enthalpy of formation, and atomization energy.

#### 4.3.2. Hardness

The hardest structure with the energy less than 0.1 eV/atom above the convex hull is the representative structure of a binary system. To get a more accurate map of hardness, the hardnesses of these representative structures were calculated using the Mazhnik-Oganov model [72] of hardness. Then the hardness of the missed systems is calculated using the committee voting method (see Fig. 16) – in Fig. 16, we assigned a hardness value of 50 GPa to materials harder than 50 GPa. In clustering the hardness maps, we only include materials harder than 5 GPa because the majority of materials are soft (with hardness less than 5 GPa) which is not interesting for us and reduces the difference of the clustering rates of different MNs.

Hardest materials are usually compounds of carbon, boron, and nitrogen with each other or with other elements. When these three elements sit in neighboring places (i.e. in AN and $MN_m$), a number of big islands are produced depending on the arrangement of other elements. Despite that, if other similar elements are placed far from each other, they form several distant islands that are not clustered together (see Fig. 16a). Table 4 shows the number of clusters to

cover all binary systems (harder than 5 GPa) in the hardness maps of the MNs. These results are shown in more details in Fig. 20. The maximum number of clusters (islands), in small $d_p$, is found for AN which was expected due to the splotchy hardness map it produced. In the $N_{min}$ biggest clusters, AN covers fewer binary systems than all other MNs in different range of $d_p$ (see Fig. 21). The highest clustering rate is calculated for USE that clusters regions of hardness map in lower number of clusters, than other MNs, in whole range of $d_p$ (Fig. 20). Fig.21 shows that USE covers 78% to 96% (for different $d_p$) of all materials harder than 5 GPa in its biggest $N_{min}$ clusters. Better clustering of materials with similar hardness by USE was expected even by visualizing the produced Pettifor maps of hardness. USE has significantly reduced the size of the regions containing materials harder than 15 GPa – exploring about a quarter of the chemical space is enough to predict almost all the hard materials – that also places soft materials in each other's vicinity. Reducing the size of promising regions of the chemical space is important, especially, when doing an automatic and systematic search for materials with optimal properties.[73]




**(a)** Atomic number (AN) **(b)** Periodic number (PN) **(c)** Pettifor's Mendeleev number ($MN_P$)

**(d)** Modified Mendeleev number ($MN_m$) **(e)** Mendeleev number in this work(USE)

Figure 16. 2D maps of the hardness (GPa) obtained using Mazhnik-Oganov model [72] of hardness of binary systems, plotted in various MNs. The representative for each binary system is the phase with the highest hardness in our database. The material with the highest hardness is shown by black hollow circle.

Table 4. Clustering rate based on: (a) the number of clusters for different MNs in comparison to the minimum number of clusters , $N_{min}$, in a imaginary ideal MN ($MN_{ideal}$), and (b) fraction of binary systems that are covered by the first (biggest) $N_{min}$ clusters in different MNs. The clustering rates are calculated based on the change of the $d_p$.

| | | **(a)** Number of clusters to cover all binary systems | | | | | | **(b)** Fraction of binary systems that are covered by the first $N_{min}$ clusters | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $d_p$ | $N_{min}$ | AN | PN | $MN_P$ | $MN_m$ | USE | AN | PN | $MN_P$ | $MN_m$ | USE |
| **Hardness** | 1.5 | 15 | 98 | 87 | 74 | 80 | 65 | 0.59 | 0.68 | 0.71 | 0.7 | 0.78 |
| | 2.5 | 9 | 62 | 53 | 50 | 55 | 44 | 0.72 | 0.77 | 0.8 | 0.79 | 0.83 |
| | 3.5 | 7 | 43 | 39 | 39 | 51 | 27 | 0.79 | 0.83 | 0.85 | 0.79 | 0.89 |
| | 4.5 | 5 | 32 | 37 | 28 | 43 | 24 | 0.8 | 0.84 | 0.88 | 0.82 | 0.9 |
| | 5.5 | 5 | 29 | 33 | 26 | 34 | 23 | 0.83 | 0.87 | 0.91 | 0.85 | 0.91 |
| | 6.5 | 4 | 21 | 35 | 20 | 26 | 14 | 0.89 | 0.87 | 0.93 | 0.89 | 0.96 |
| **Magnetization** | 0.005 | 18 | 70 | 48 | 60 | 63 | 64 | 0.83 | 0.88 | 0.84 | 0.85 | 0.84 |
| | 0.01 | 9 | 36 | 30 | 30 | 39 | 35 | 0.89 | 0.91 | 0.92 | 0.9 | 0.92 |
| | 0.02 | 5 | 20 | 15 | 12 | 15 | 14 | 0.94 | 0.96 | 0.97 | 0.97 | 0.97 |
| | 0.03 | 3 | 12 | 10 | 11 | 8 | 9 | 0.95 | 0.98 | 0.97 | 0.98 | 0.97 |
| | 0.04 | 3 | 11 | 8 | 10 | 8 | 9 | 0.95 | 0.99 | 0.98 | 0.98 | 0.98 |
| | 0.05 | 2 | 11 | 5 | 8 | 6 | 6 | 0.96 | 0.99 | 0.99 | 0.98 | 0.99 |
| **Enthalpy of formation** | 0.05 | 50 | 240 | 197 | 196 | 182 | 193 | 0.75 | 0.78 | 0.79 | 0.79 | 0.79 |
| | 0.1 | 25 | 136 | 95 | 95 | 87 | 100 | 0.83 | 0.9 | 0.9 | 0.89 | 0.88 |
| | 0.2 | 13 | 76 | 54 | 47 | 59 | 60 | 0.88 | 0.93 | 0.95 | 0.93 | 0.93 |
| | 0.3 | 9 | 59 | 37 | 39 | 39 | 46 | 0.93 | 0.96 | 0.95 | 0.95 | 0.95 |
| | 0.4 | 7 | 35 | 28 | 24 | 31 | 29 | 0.94 | 0.97 | 0.98 | 0.96 | 0.97 |
| | 0.5 | 5 | 24 | 21 | 21 | 21 | 24 | 0.96 | 0.97 | 0.96 | 0.98 | 0.96 |
| **Atomization energy** | 0.1 | 44 | 439 | 325 | 321 | 350 | 440 | 0.33 | 0.46 | 0.51 | 0.44 | 0.35 |
| | 0.2 | 22 | 194 | 141 | 158 | 164 | 190 | 0.53 | 0.76 | 0.71 | 0.63 | 0.64 |
| | 0.3 | 15 | 130 | 96 | 80 | 89 | 94 | 0.69 | 0.8 | 0.79 | 0.77 | 0.84 |
| | 0.4 | 11 | 110 | 54 | 52 | 63 | 84 | 0.71 | 0.86 | 0.87 | 0.8 | 0.86 |
| | 0.5 | 9 | 76 | 49 | 50 | 57 | 70 | 0.76 | 0.89 | 0.87 | 0.8 | 0.88 |
| | 0.6 | 8 | 65 | 37 | 33 | 47 | 47 | 0.82 | 0.9 | 0.89 | 0.88 | 0.89 |

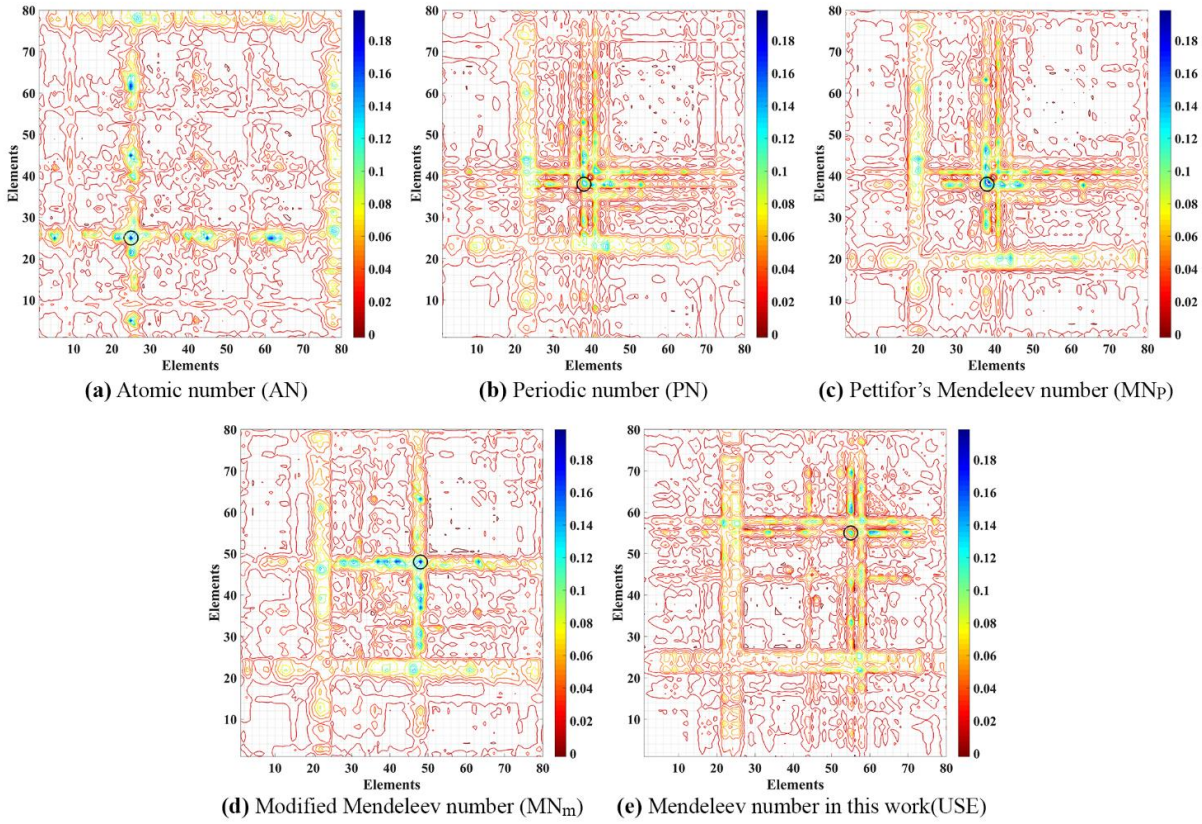
### 4.3.3. Magnetization

As mentioned before, our database contains magnetic information for only 14% of the total binary systems and the magnetic information is assigned to the majority of the systems (about 86%) using committee voting method. This, obviously, increases the clustering rate for all the MNs. The range of magnetization is from zero for nonmagnetic and antiferromagnetic materials to 0.198 $\mu_\beta/Å^3$ corresponding to the magnetization of iron. Among all elements that we included, and their compounds, Fe has the highest magnetization. This result is correct. In reality, Gd has a slightly higher magnetization, but lanthanoids were not included for technical reasons (problems with available pseudopotentials, and with convergence). For evaluating magnetization maps of different MNs, we disregarded materials with magnetization less than 0.02 $\mu_\beta/Å^3$ (see Table 4) – this helps us to better distinguish performance of different MNs.

Looking at Fig. 17, at the first glance, it seems that AN provides a slightly better map with clear separation of materials with similar magnetization. Although that might be true for promising regions, a closer look to Fig. 17a shows that AN clusters other regions of the chemical space, with lower magnetization, less efficiently by following its periodic pattern –

see Table 4, Fig. 20, and Fig. 21. The clustering rate for all the MNs are very high as could be expected – small number of clusters that quickly approaches the $N_{min}$ (minimum number of clusters that is required by an ideal MN), and high coverage of binary systems (from 83% for small $d_p$, to 99% for bigger $d_p$) in the first $N_{min}$ clusters.

In Fig. 17, the main two islands of materials with high magnetization, correspond to the compounds of some transition metals such as Fe, Co, Ni, and some actinoids such as Pu (lanthanoids also form highly magnetic phases, but as we mentioned above, were excluded for technical reasons) – this can be clearly seen in the magnetization map of AN (Fig. 17a).




**(a)** Atomic number (AN) **(b)** Periodic number (PN) **(c)** Pettifor's Mendeleev number ($MN_P$)

**(d)** Modified Mendeleev number ($MN_m$) **(e)** Mendeleev number in this work(USE)

Figure 17. 2D maps of magnetization (in the unit of $\mu_B \cdot Å^{-3}$) of binary systems, plotted in various MNs. The representative for each binary system is the phase with the highest magnetization in our database. The material with the highest magnetization is shown by black hollow circle.

### 4.3.4. Enthalpy of formation

Pettifor maps of the enthalpy of formation produced by different MNs are shown in Fig. 18. The plots were made taking in each binary system the $A_xB_y$ compound with the lowest enthalpy of formation in the database. Unlike hardness and magnetization map, in the maps of the enthalpy of formation, binary systems with lower values of the enthalpy of formation are more favorable (depicted with red color).

Looking at Fig. 18, one can see that PN, $MN_P$, $MN_m$, and USE, have produced similar maps of the enthalpy of formation. In all these maps, promising materials (with more exothermic chemical reactions – shown in orange and red) are gathered in a small region, right bottom – left top, of the map. This means that very dissimilar elements often form stable compounds. The lowest enthalpy of formation was found for $ThF_4$ (−4.11 eV/atom), followed

by $AcF_3$ (−4.09 eV/atom), $CaF_2$ (−3.92 eV/atom) and $ZrF_4$ (−3.62 eV/atom). Other notable values include $Th_4O_7$ (−3.61 eV/atom), $Y_2O_3$ (−3.48 eV/atom), $Al_2Ta$ (−3.18 eV/atom), $Al_2O_3$ (−2.95 eV/atom), CaO (−2.95 eV/atom), $SiO_2$ (−2.79 eV/atom), $Al_5Ge_2$ (−2.44 eV/atom). Note that fluorides and oxides are the most exothermic compounds, which is easy to understand, since F and O have the highest electronegativities. Materials that are shown in yellow color (with an enthalpy of formation between −2 and −2.5 eV/atom) can be found mostly around the promising regions (shown in red). Fig. 20 shows that among these MNs, $MN_m$ provides a slightly better map, while the performance of all MNs (except for AN) are similarly good. Fig. 18 suggests that USE performs better in clustering promising regions (shown in yellow, orange and red) by condensing them in a smaller area. On the other hand, AN produced a periodic map which is inefficient for clustering compounds with similar enthalpies of formation. As expected, our clustering evaluations show that AN clusters regions of the chemical space less efficiently than other MNs (see Fig. 20, and Fig. 21).

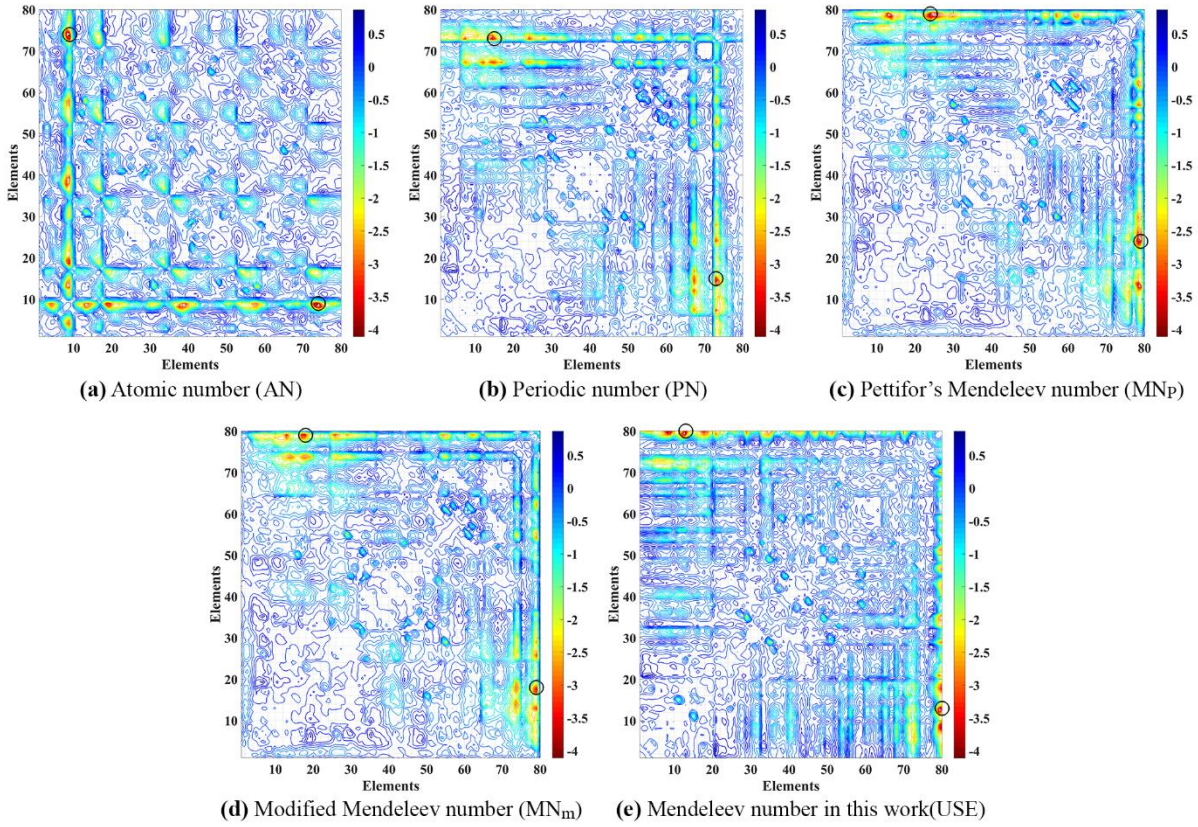



Figure 18. 2D maps of the enthalpy of formation (eV/atom) of binary systems, plotted in various MNs. The representative for each binary system is a structure with the lowest enthalpy of formation in our database. The material with the lowest enthalpy of formation on the map is shown by black hollow circle.

### 4.3.5. Atomization energy

Figure. 19 shows maps of the atomization energy produced by different MNs. Similar to the enthalpy of formation, lower values of atomization energy are preferred. More negative values of atomization energy (shown in orange and red) means that more energy is required to break all bonds in the crystal. For this property, we took into account the spin-polarization energies of atoms, to take into account that ground states of isolated atoms of most elements

are spin-polarized. Among the elements, tungsten has the lowest atomization energy equal to −8.51 eV/atom, while among binary compounds the lowest value is achieved in Ta-C (−8.78 eV/atom for $Ta_6C_5$ and −8.79 eV/atom for $Ta_2C$). Atomization energy measures the total strength of bonding in the solid, and is correlated with the melting temperature. Indeed, tungsten has the highest melting temperature among elements (3695 K), while among binary compounds, HfC and TaC have the highest melting temperatures above 4000 K.[74] For Hf-C, our calculations show atomization energy equal to −8.16 eV/atom. The atomization energy of some representative solids such as graphite, BN (zinc-blende phase), silica ($SiO_2$), and NaCl are −7.98 eV/atom, −7.01 eV/atom, −6.52 eV/atom, and −3.16 eV/atom respectively, which are very close to the values from experiment.

Similar to other properties, i.e. the enthalpy of formation and hardness, AN produces a map with a periodic pattern (Fig. 19a), which means clustering materials with similar properties in many small islands instead of few big islands. Looking at the atomization energy maps in the space of MNs in Fig. 19 and their clustering evaluations in Fig. 20, and Fig. 21, it is clear that PN and $MN_P$ and $MN_m$ do better job by smoothly clustering materials with similar atomization energy, while clustering rates for USE, and AN are progressively lowered. However, by increasing the $d_p$, number of clusters in AN and USE quickly approaches to the number of clusters in PN, $MN_P$, and $MN_m$, while Fig. 21 shows that number of covered systems by minimum number of clusters ($N_{min}$) for AN is less than all other MNs in all range of $d_p$.

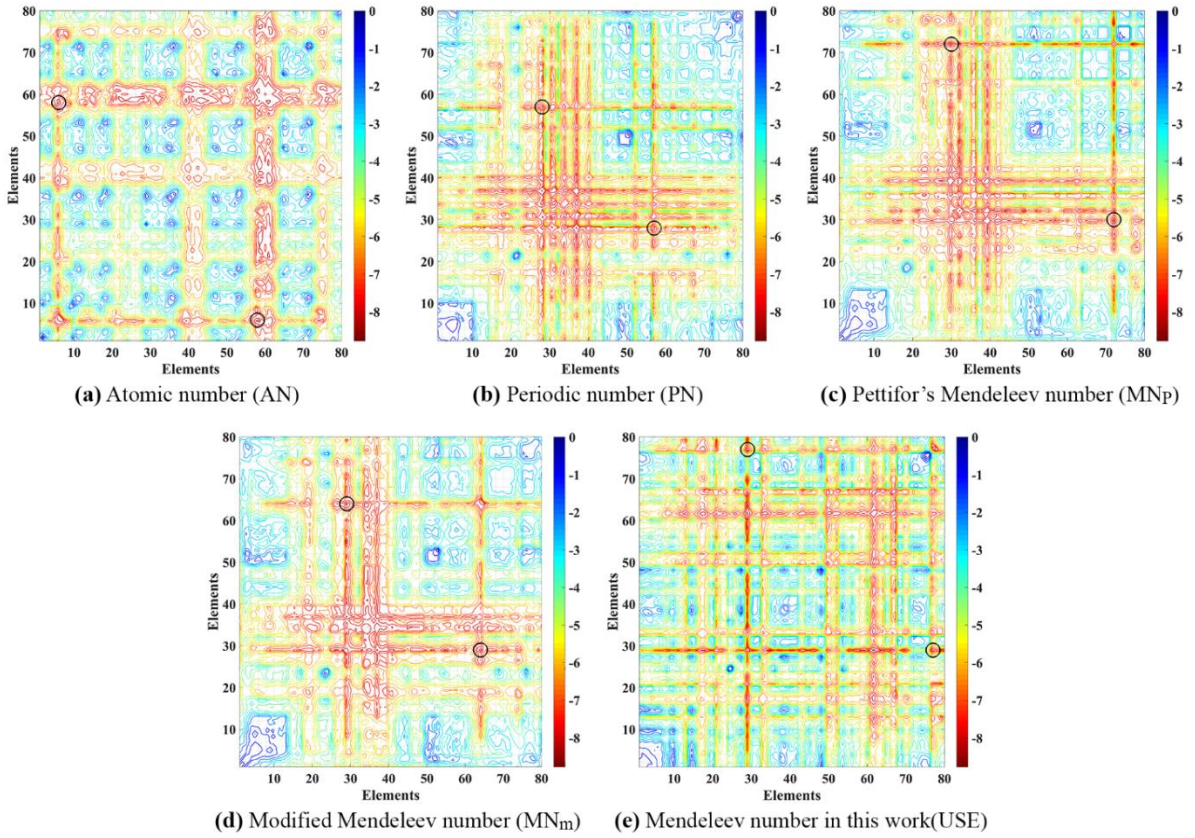



Figure 19. 2D maps of the atomization energy (eV/atom) of binary systems, plotted in various MNs. The representative for each binary system is a structure with the lowest atomization energy in our database. The material with the lowest atomization energy on the map is shown by black hollow circle.

In a nutshell, except for AN which provides a patchy periodic chemical space, other MNs provide a convenient well-structured chemical space for the properties on which we did tests – hardness (representing the mechanical properties), magnetization (electronic properties), enthalpy of formation and atomization energy (thermochemical properties). Among them, USE, with a simple definition from the most important elemental properties (i.e. atomic radius and electronegativity), generates an overall best clustering in the chemical space (see Table 4, and Fig. 20) with clearer separation of regions that contain materials with similar properties. Such well-organized chemical space facilitates the prediction of new materials by exploring the promising regions at the expense of unpromising ones.[73]

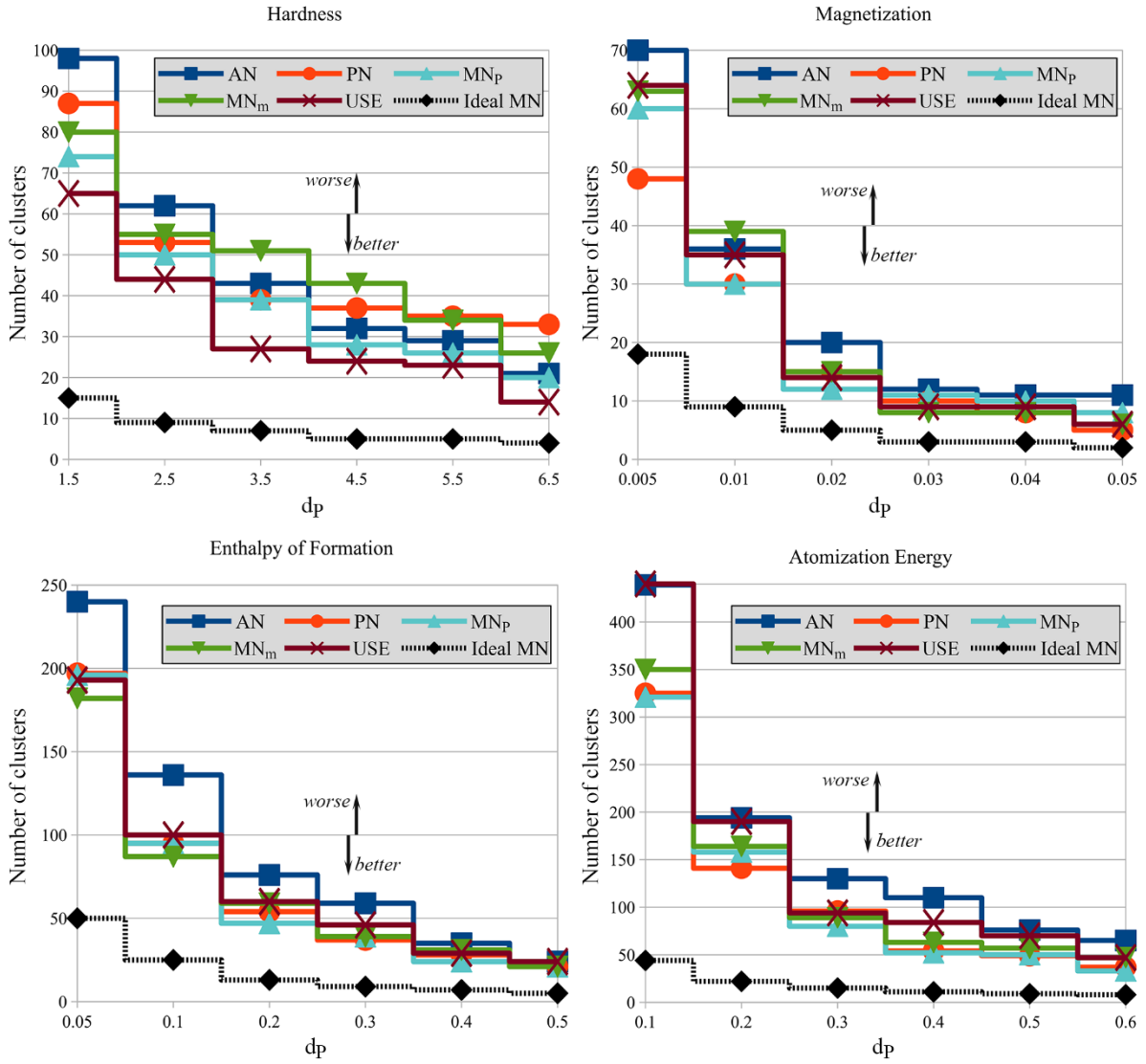



Figure 20. Number of clusters vs. property difference cutoff ($d_P$) for different Mendeleev numbers – in comparison to a hypothetical ideal MN – for the hardness, magnetization, enthalpy of formation, and atomization energy.

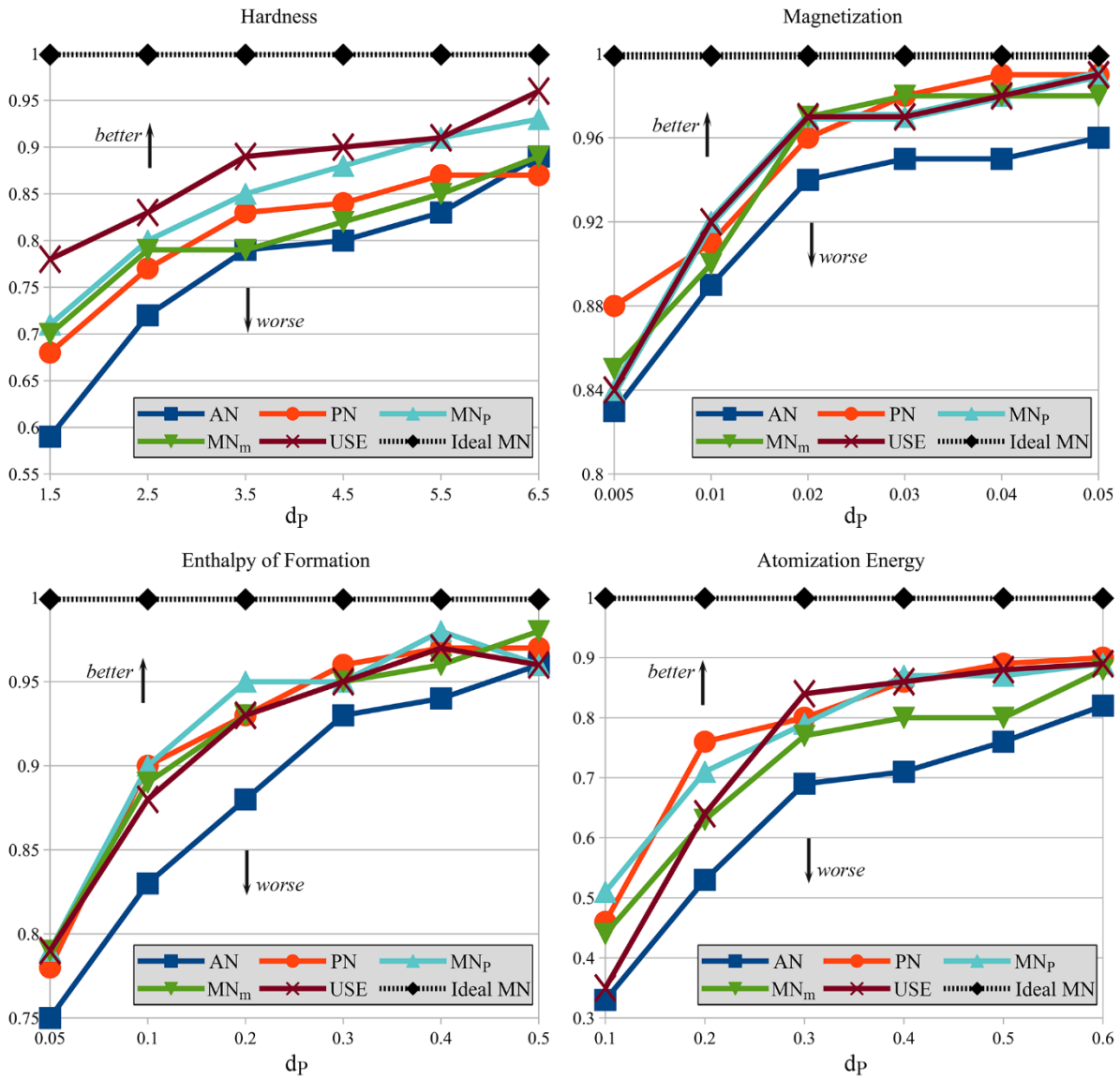



Figure 21. Fraction of binary systems that are covered by a minimum number of clusters as required in an ideal MN to cover all the binary systems for different $d_P$.

### 4.3.6. A well-defined chemical space at high pressures

The chemistry of the elements and compounds changes with pressure. The discussed MNs are either fixed (AN, PN, and $MN_P$) or obtained by optimizing some evaluation function based on big data ($MN_m$), and adapting these MNs to high pressures is either impossible or requires large amounts of data, huge efforts, and vast computational resources. The USE is the only Mendeleev number that was constructed on a fundamental basis, using the most important elemental properties — electronegativity and atomic radius, and as these properties change under pressure, so will the USE.

The atomic radius of an element can be defined (as we defined throughout this work) and calculated as half the shortest interatomic distance in the relaxed simple cubic structure of that element under pressure. The electronegativity of many elements has been calculated at various pressures.[75,76] Using these data, the USE was obtained at various pressures (Table 5). This can

help to predict new materials at arbitrary pressure, only by having a number of relevant data on other systems and plotting them onto the well-organized map produced by the USE.

Table 5. The USE at high pressures.

| # | 50 GPa | 200 GPa | 500 GPa | # | 50 GPa | 200 GPa | 500 GPa |
|---|---|---|---|---|---|---|---|
| 1 | Xe | Cs | Ba | 36 | As | Ir | Ru |
| 2 | Cs | Ba | Cs | 37 | Ge | Nb | Ca |
| 3 | Ba | Po | Bi | 38 | Re | As | As |
| 4 | Po | Bi | Pb | 39 | Ga | Se | Hf |
| 5 | Bi | Pb | Po | 40 | Pt | Pd | Al |
| 6 | Sr | Xe | Sn | 41 | Ti | Sc | Se |
| 7 | Pb | Tl | Tl | 42 | Os | Br | Rh |
| 8 | I | Sn | Xe | 43 | Ir | Ru | Zn |
| 9 | Tl | Sb | Sb | 44 | Tc | Al | Cu |
| 10 | Y | Te | In | 45 | Pd | Na | Sc |
| 11 | Rb | In | Te | 46 | Ru | Ar | Na |
| 12 | Te | I | Hg | 47 | Rh | Rh | Br |
| 13 | Sb | Rb | Rb | 48 | Al | Zn | Cr |
| 14 | Kr | Hg | Cd | 49 | Cl | Ti | Si |
| 15 | Ca | Sr | Au | 50 | V | Cu | Ar |
| 16 | Sn | Y | I | 51 | Zn | V | Nb |
| 17 | In | Cd | Sr | 52 | S | Si | Mn |
| 18 | Lu | Lu | Ag | 53 | Si | Cr | V |
| 19 | Hg | Au | Y | 54 | Cr | Mn | Fe |
| 20 | Hf | Kr | Lu | 55 | P | P | Ni |
| 21 | K | Ta | Zr | 56 | Cu | Fe | Ti |
| 22 | Nb | Ag | W | 57 | Mn | S | P |
| 23 | Sc | Zr | Ta | 58 | Fe | Cl | Co |
| 24 | Cd | W | Re | 59 | Li | Ni | S |
| 25 | Br | Hf | Pt | 60 | Co | Co | Cl |
| 26 | Ta | Re | Os | 61 | Ni | Li | Li |
| 27 | Zr | Pt | Mo | 62 | Ne | Be | Be |
| 28 | Ar | Mo | Kr | 63 | Be | B | B |
| 29 | Na | Os | K | 64 | F | Ne | Ne |
| 30 | Au | K | Ga | 65 | O | C | C |
| 31 | W | Ge | Ge | 66 | B | N | N |
| 32 | S | Ga | Tc | 67 | N | O | O |
| 33 | Mg | Mg | Ir | 68 | C | F | F |
| 34 | Ag | Ca | Pd | 69 | He | He | He |
| 35 | Mo | Tc | Mg | 70 | H | H | H |

## 4.4. Conclusions

Having a well-defined sequence of the elements (Mendeleev numbers, or MNs), where similar elements take neighboring places, one can produce an organized map of properties for binary or more complex systems that leads to the prediction of new materials by having information on their neighboring systems. We defined a simple, physically meaningful, and universal way to order the elements. In this work, we studied our MN (USE), in addition to a number of previously known MNs such as atomic number (AN), Villars' periodic number [65] (PN), Pettifor's Mendeleev number [60] ($MN_P$), modified Mendeleev number [62] ($MN_m$), using provided data on binary systems from our and other online databases, such as ICSD [66] and COD.[69] Two-dimensional maps of the hardness, magnetization, enthalpy of formation, and atomization energy were plotted using the provided data in the space of MNs and it turned out that most of these sequences (except for AN) indeed work well at clustering materials with similar

properties. The evaluation of the MNs showed the overall best clustering rate of the chemical spaces produced by USE for target spaces, i.e. hardness, magnetization, and enthalpy of formation. Also, unlike other MNs, USE can be defined at any arbitrary pressure, which is a step forward for the prediction of materials under pressure. Importantly, our work clarifies the physical meaning of the Mendeleev number (previously defined empirically): it is a collapsed one-number representation of the important atomic properties (such as atomic radius, electronegativity, polarizability, and valence).

# CHAPTER 5.

# Coevolutionary Search for Optimal Materials in the Space of All Possible Compounds


## Abstract

*Over the past decade, evolutionary algorithms, data mining, and other methods showed great success in solving the main problem of theoretical crystallography: finding the stable structure for a given chemical composition. Here we develop a method that addresses the central problem of computational materials science: the prediction of materials that possess the best combination of target properties among all possible combinations of all elements. This nonempirical method combines our coevolutionary approach with the carefully restructured "Mendelevian" chemical space, energy filtering, and Pareto optimization to ensure that the predicted materials have optimal properties and a high chance to be synthesizable. The first calculations, presented here and discussed in Chapter 6, illustrate the power of this approach.*



**Keywords:** Coevolutionary algorithm; evolutionary algorithm; hardness; Mendeleev numbers; multi-objective optimization.


## 5.1. Introduction

Discovery of materials with optimal properties (e.g., the highest hardness, the lowest dielectric permittivity, etc.) or an optimal combination of properties (e.g., the highest hardness and fracture toughness) is the central problem of materials science. Until recently, experiment was the only possible method in materials discovery, with all limitations and expense of trial-and-error approach, but the ongoing revolution in theoretical and computational materials science begins to change the situation. Using the quantum mechanical calculations, it is now routine to predict many properties when the crystal structure is known. In 2003, Curtarolo demonstrated data mining method for materials discovery[77] by screening crystal structure databases, which can include known or hypothetical structures, via ab initio calculations. At the same time, major progress was seen in fully nonempirical crystal structure prediction. Metadynamics[1] and evolutionary algorithms[2–4] have convinced the community that crystal structures are predictable.

Despite the success of these and other methods, a major problem remains unsolved: the prediction of a material with optimal properties among all possible compounds. 4950 binary systems, 161,700 ternary systems, 3,921,225 quaternary systems, and an exponentially growing number of higher-complexity systems can be created by 100 best-studied elements in the periodic table. In each system, a very large number of compounds and, technically, an infinite number of crystal structures can be constructed computationally, and the direct screening of such a multitude is impractical. Only about 72% of binary, 16% of ternary, 0.6% of quaternary, and less than 0.05% of more complex systems have ever been studied experimentally,[78] and even in those systems that have been studied, new compounds are being

discovered.[32,79,80] Studying all these systems, one by one, using global optimization methods is unrealistic. Data mining is a more practical approach, but the above statistics shows that the existing databases are significantly incomplete even for binary systems, whereas using this method for ternary and more complex systems would be inefficient. Besides, data mining cannot find fundamentally new crystal structures. When searching for materials optimal in more than one property, these limitations of both approaches become even greater. We present a new method implemented in our code, MendS (Mendelevian Search), and show its application to discover hard, superhard, and magnetic materials.

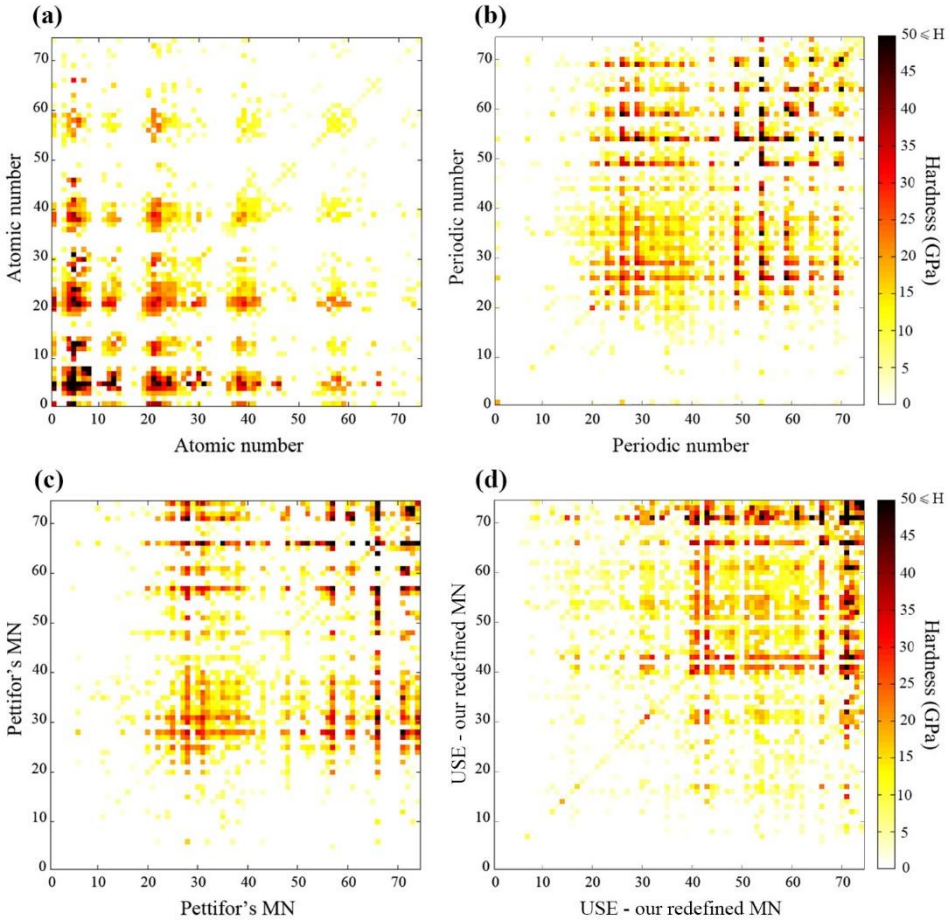



Figure 22. Pettifor maps showing the distribution of hardness in binary systems, using (a) atomic numbers, (b) Villars' Periodic number, (c) Pettifor's MN, and (d) MN obtained in this work. Noble gases were excluded because of their almost complete inability to form stable compounds at normal conditions. Rare earths and elements heavier than Pu were excluded because of the problems of the DFT calculations. In total, we consider 74 elements that can be combined into 2,775 possible binary systems. Each pixel is a binary system, the color encodes the highest hardness in each system.

## 5.2. Mendelevian Space

Global optimization methods are effective only when applied to property landscapes that have an overall organization, e.g., a landscape with a small number of funnels. Discovering materials with optimal properties, i.e. performing a complex global optimization in the chemical and structural space, requires a rational organization of the chemical space that puts

compounds with similar properties close to each other. If this space is created by ordering the elements by their atomic numbers, we observe a periodic patchy pattern (Fig. 22a), unsuitable for global optimization.

As we mentioned in Chapter 4., Pettifor suggested a new quantity, the so-called "chemical scale," that arranges the elements in a sequence so that similar elements are placed near each other, and compounds of these elements also display similar properties.[59] This way, structure maps[60] with well-defined regions of similar crystal structures or properties can be drawn. In a thus ordered chemical space, evolutionary algorithms should be extremely effective: they can zoom in on the promising regions at the expense of unpromising ones.

In Chapter 4., we explained our method in redefining the chemical scale and Mendeleev number (USE), based on the most important chemical properties of the atom — size $R$ and electronegativity $\chi$ (Pauling electronegativity) — the combination of which can be used as a single parameter succinctly characterizing the chemistry of the element. However, we need to emphasize that the chemical scale and MN are only used in this method for visualizing the results (the choice of MN for plotting such a Pettifor map is up to the user), while in our global coevolutionary algorithm, each atom is represented by both its size $R$ and electronegativity $\chi$ to increase the accuracy. In this work, the atomic radius $R$ is defined as half the shortest interatomic distance in the relaxed (for most elements hypothetical) simple cubic structure of an element – see Table 2.

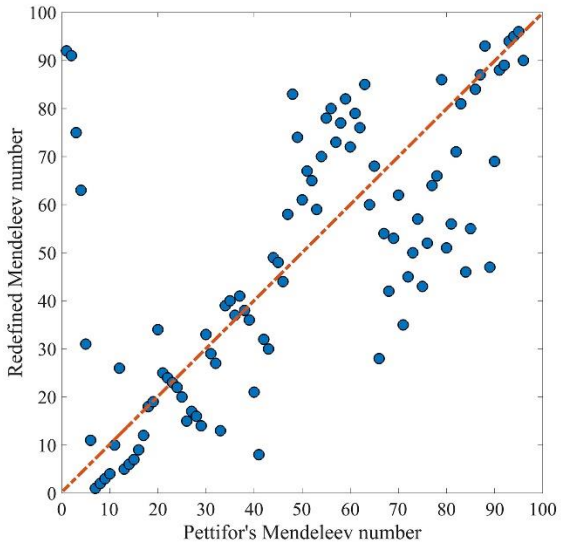



Figure 23. Correlation between the Mendeleev numbers defined in this work and those proposed by Pettifor.

Fig. 23 shows the overall linear correlation between the Mendeleev numbers redefined in this work and those proposed by Pettifor. Carefully defined Mendeleev numbers should lead to a strong clustering in the chemical space, where neighboring systems have similar properties. The results of our searches for hard binary compounds using the MN suggested by Pettifor and our redefined MN are shown on the Pettifor maps (Fig. 22b, c). Satisfyingly, our redefined MNs result in a better-organized chemical space with a clearer separation of regions containing binary systems with similar hardness. In fact, if our MNs (which are the sequences of projected elements on their regression line in the space of crudely correlated atomic radius and

electronegativity) generate a good 2D map, with clear grouping of similar chemical systems (e.g., Na-Cl, K-Cl, Ca-Cl, Na-Br systems are located nearby), then a much better grouping is expected in the space of the initial two parameters $R$ and $\chi$, and it is in this space where variation operators of our method are defined (Fig. 24a, b). Also it worth mentioning that sizes and electronegativities of the atoms change under pressure – and using standard definitions of the Mendeleev number (such as AN, or Pettifor's MN) will not work well. Our recipe, however, is universal and only requires atomic sizes and electronegativities at the pressure of interest. In this paper, we illustrate our method by binary systems, although more complex, at least ternary, systems are also tractable.

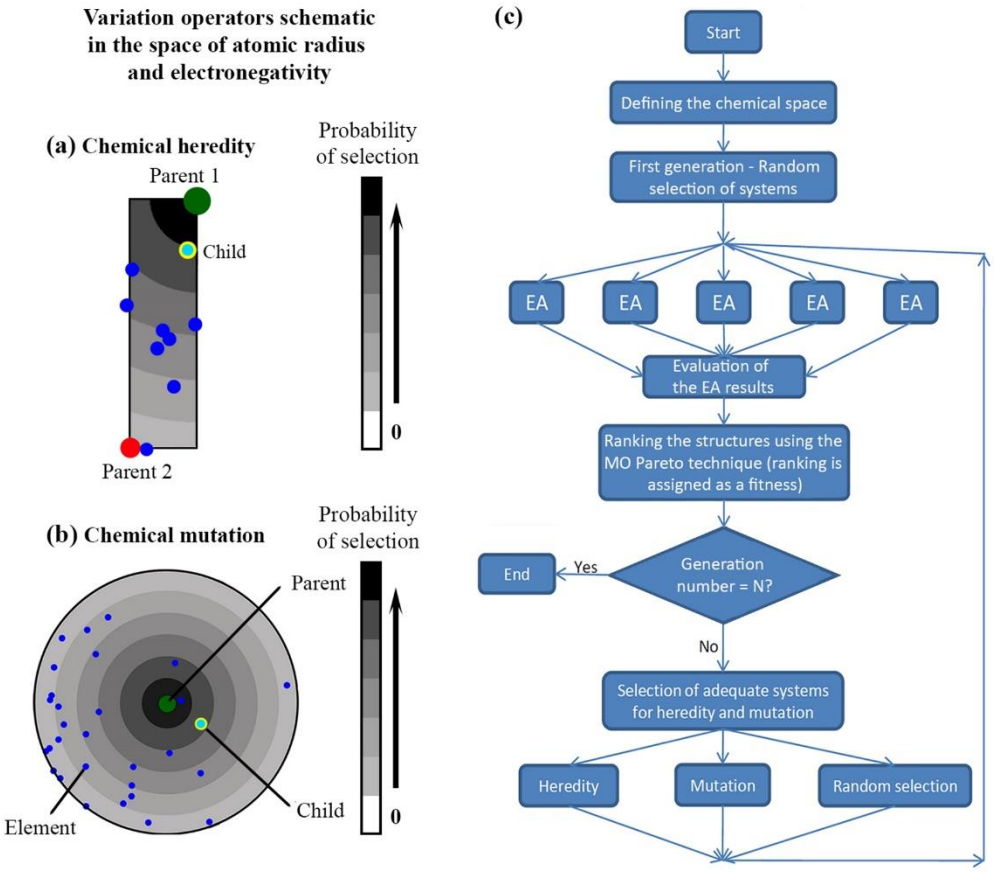



Figure 24. MendS algorithm, (a) Scheme showing how the chemical heredity and (b) chemical mutation create new compositions. The probability, displayed in shades of gray, is given to each possible child according to its distance from the fitter parent (dark green point). (c) Flowchart of the coevolutionary algorithm used in MendS (EA — evolutionary algorithm, MO — multi-objective).

## 5.3. Method

The whole process can be described as a joint evolution (or coevolution) of evolutionary runs, each of which deals with an individual binary variable-composition system. Having defined the chemical space, we initialize the calculations by randomly selecting a small number of systems from the entire chemical space for the first MendS generation. These systems are then optimized by the evolutionary algorithm USPEX[2–4] in its variable-composition mode,[81] searching for compounds and structures with optimal properties, in our example, we

simultaneously maximized hardness and stability, after which MendS jointly analyses the results. Removing the identical structures using the fingerprint method,[82] jointly evaluating all systems, refining and preparing the dataset, and discarding the structures that are unstable by more than 1.0 eV/atom, MendS ranks all systems of the current generation and selects the fittest (usually 70% of them) as potential parents for new systems. Applying the variation operators, such as mutation and heredity, to these parent systems yields offspring systems for the next coevolutionary generation. Additionally, some systems are generated randomly to preserve the chemical diversity of the population. This process is continued until the number of coevolutionary generations reaches the maximum predefined by the user (Fig. 24c). The underlying ab initio structure relaxations and energy calculations were performed using the density functional theory with the projector augmented-wave method (PAW) and GGA-PBE functional as implemented in the VASP code.[48,50] The details on the input parameters of MendS, USPEX, and VASP are given in Appendix 1.

### 5.3.1. Defining the Fitness: Multi-objective (Pareto) Optimization

The combination of multi-objective Pareto optimization with evolutionary algorithm works efficiently, especially when optimizing two-objective problems (see Chapter 3 of this thesis). Here we performed searches that simultaneously optimized (1) stability, measured as the distance above the convex hull, and (2) hardness, computed using the Lyakhov–Oganov model.[51]

The multi-objective optimization usually yields a set of materials, with a trade-off between their properties, that form the first Pareto front. Similarly, 2nd, 3rd, …, *n*th Pareto fronts can be defined (Fig. 25). In our method, the Pareto rank[47,70] is used as a fitness (see Chapter 3).

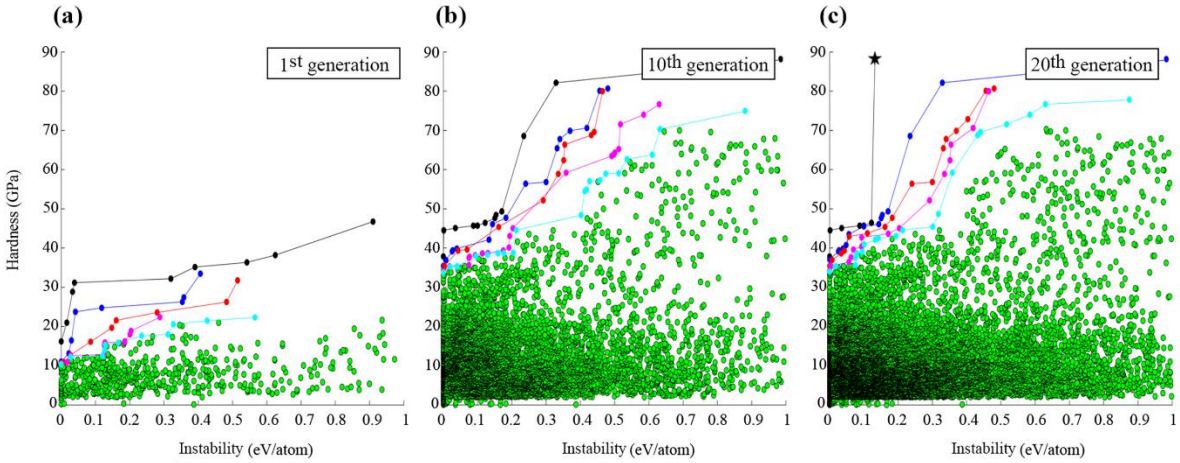



Figure 25. The MendS results of the simultaneous hardness and stability optimization of all unary and binary compounds: (a) 1st generation, (b) 10th generation, (c) 20th generation. The first five Pareto fronts are shown, green points representing all sampled structures. The instability of each compound is defined using Maxwell's convex hull construction. Diamond, the hardest material, is indicated by a star.

### 5.3.2. Variation Operators in a Chemical Space

The variation operators are very important for an efficient sampling of a chemical space using the previously sampled compositions and structures. These operators ensure that different populations not only compete, but also learn from each other. In an efficient algorithm, the chemical space is defined by just one number for each element — the Mendeleev number, or chemical scale. We use this for plotting the Pettifor maps, but within the algorithm, each element is described by two numbers — electronegativity $\chi$ and radius $R$, rescaled to be between 0 and 1 — and it is this space where the variation operators act. Three variation operators are defined in the chemical space: chemical heredity, reactive heredity, and chemical mutation.

Chemical heredity replaces elements in parent systems with new elements so that their electronegativities and atomic radii lie between those of their parents (Fig. 24a). Thus, the regions of the chemical space between the parents are explored:

$$\mathrm{AB} + \mathrm{CD} \rightarrow \mathrm{XY}, \tag{5.1}$$

where X is between A and C or A and D, which is selected randomly, and Y is between the other two elements (B and D or B and C).

Reactive heredity creates an offspring by taking combinations of the elements from parents. For example, if the parents are A–B and C–D, their child is one of the A–C, A–D, B–C, and B–D systems.

Chemical mutation randomly chooses one of the elements of a parent and substitutes it with another element in its vicinity in the space of $\chi$ and $R$ (Fig. 24b).

In both chemical mutation and chemical heredity, the probability is assigned to all elements:

$$P_i = \frac{e^{-\alpha x_i}}{\sum e^{-\alpha x_i}}, \qquad i = 1, 2, \ldots, \tag{5.2}$$

where $x$ is the distance of element $i$ from the parent element, and $\alpha$ is a constant (here we use $\alpha = 1.5$). In the case of chemical heredity, the probability gives a higher weight to the fitter parent, shown by a dark green point in Fig. 24a, b.

The result of applying these variation operators is shown in Fig. 26: the promising regions of the chemical space are sampled more thoroughly at the expense of the unpromising regions. When a new system is produced from parent systems, it inherits from them a set of optimal crystal structures which are added to the first generation, greatly enhancing the learning power of the method.

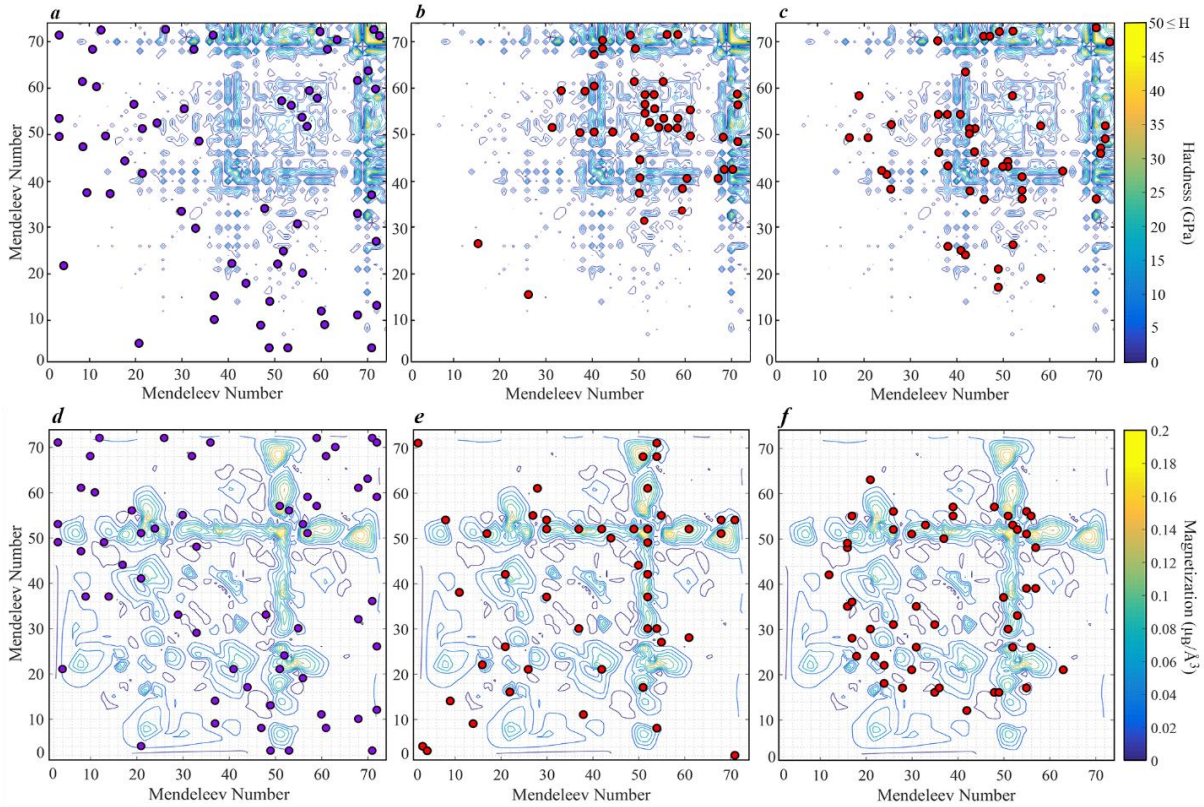



Figure 26. Systems selected (a, d) randomly in the 1st generation, and using all variation operators in (b, e) the 5th and (c, f) 10th generations in searching for (a–c) hard and (d–f) magnetic materials. Randomly generated systems are shown by violet circles.

After finishing the coevolutionary simulation, we took the most promising systems identified in it and performed longer evolutionary runs for each of them, calculating the final hardness using the Chen–Niu model,[58] and fracture toughness — using the Niu–Niu–Oganov model.[83] We reported some of these results in a separate paper on the Cr–B, Cr–C, and Cr–N systems,[46] and our study of the W–B system[84] was inspired by the present finding of promising properties in the Mo–B system. The new results are discussed in Chapter 6.

## 5.4. Mendelevian Search for Hard and Superhard Materials

The Pareto optimization[70] of hardness and stability was performed over all possible structures (with up to 12 atoms in the primitive cell) and compositions of binary compounds of 74 elements: all elements excluding the noble gases, rare earth elements, and elements heavier than Pu. In this work, 600 systems were computed in 20 MendS generations from a total of 2775 binary and unary systems that can be composed of 74 elements (i.e., only about one fifth of all possible systems were sampled). The efficiency of this method in finding optimal materials is shown in Fig. 25. In this fast calculation, numerous stable and metastable hard and superhard materials were detected in a single run. Carbon (diamond and other allotropes) and boron, known to be the only superhard elements, were both found. In addition, not only our calculation revealed several new binary compounds, but also many already known hard systems, or system that were reported to be potentially hard during decades, were found in a single run. Among them $B_xC_y$,[85] $C_xN_y$,[86,87] $B_xN_y$,[88,89] $B_xO_y$,[85,90,91] $Re_xB_y$,[92,93] $W_xB_y$,[94] $Si_xC_y$,[95–98] $W_xC_y$,[96–98] $Al_xO_y$,[96–98] $Ti_xC_y$,[98] $Si_xN_y$,[98] $Ti_xN_y$,[98] $Be_xO_y$,[98] $Ru_xO_y$,[99,100] $Os_xO_y$,[101] $Rh_xB_y$,[102]

$Ir_xB_y$,[102] $Os_xB_y$,[103–105] and $Ru_xB_y$.[103–105] Several binary systems with a very promising hardness were also found in the calculation: $Mo_xB_y$,[57,92] $Tc_xB_y$,[106,107] $B_xP_y$,[108] $V_xB_y$,[109–111] $Fe_xB_y$,[53,112] $Mn_xB_y$,[113,114] and $Mn_xH_y$. They are discussed in Chapter 6. All the systems studied during the calculation are listed in Appendix 2 (Table 7).

All known binary superhard systems were found in a short coevolutionary run, which demonstrates the power of our method. It can be applied to other types of materials. The results of this search is thoroughly discussed in Chapter 6.

## 5.5. Mendelevian Search for Magnetic Materials

In addition to the Mendelevian search for stable/metastable hard and superhard materials, we performed another Mendelevian search for materials with maximum magnetization and stability to examine the power and efficiency of the method in fast and accurate determination of materials with target properties. We performed this calculation over all possible structures (with up to 12 atoms in the primitive cell) and compositions limited to the binary compounds of 74 elements (i.e. all elements excluding the noble gases, rare earth elements, and elements heavier than Pu). In this calculation, well-known ferromagnets iron, cobalt, nickel, and several magnetic materials made from the combination of these and other promising elements were detected before the 6th generation. The magnetization of each structure was found in the spin-polarized calculations using the GGA–PBE functional[17] as implemented in the VASP code.[48,50] More details on the structure relaxation and input parameters can be found in Appendix 1. The chemical space in the Mendelevian search for magnetic materials is shown in Fig. 26d–f, which was formed after calculating the magnetization of 450 binary systems over 15 generations. In this plot, the materials with high magnetization are clustered together. Fig. 26d–f shows how the (co)evolutionary optimization discovered all the promising regions at the expense of the unpromising ones. This calculation has found that among all substances, bcc-Fe has the highest magnetization at zero Kelvin. The system selection in the 5th and 10th generations is shown in Fig. 26e and f.

## 5.6. Conclusion

We developed a method for predicting materials having one or more optimal target properties. The method, based on the suitably defined chemical space, powerful evolutionary algorithm, and multi-objective Pareto optimization technique, was examined by searching for low-energy hard and superhard materials. Well-known superhard systems — diamond, boron allotropes, and the B–N system — were found in a single calculation together with other hard systems, both known and unknown. The results of this search are discussed in Chapter 6. A similar chemical map was produced for magnetic materials; well-known magnetic systems such as Ni, Co, Fe were found within a few generations. The examples of hard materials and ferromagnets found using this method show its power and efficiency, which can be used to search for optimal materials with any combination of properties at arbitrary conditions. As the first step in the prediction of novel materials possessing desired properties, this fully nonempirical method to a large extent solves the central problem of computational materials science.

# CHAPTER 6.
# Computational Prediction of Hard and Superhard Materials


## Abstract

*Superhard materials are a class of materials with unique mechanical properties that are widely used in many industrial applications. In this chapter, we present our results of the Mendelevian search in which several transition metal borides were studied and many new hard and superhard phases were predicted. Several low-energy phases of the Mn–H system were predicted to be unexpectedly hard.*



**Keywords:** Hardness; fracture toughness; elastic properties; Mendelevian search; multi-objective optimization.


## 6.1. Introduction

Modern technology requires new materials displaying unique combinations of mechanical, electronic, and other properties to replace the traditional materials widely used in the industry. We developed a new method,[73] presented in Chapter 5, that makes it possible to predict the optimal materials on the basis of all possible combinations of all the elements from the Periodic Table. This nonempirical method combines our coevolutionary approach with a carefully restructured "Mendelevian" chemical space (Chapter 4), energy filtering, and Pareto optimization[70] (Chapter 3) of target properties and stability to ensure that the predicted materials have optimal properties and a high chance to be synthesizable.

Superhard materials are a class of materials with unique mechanical properties that are widely used in many industrial applications, like mining, defense, and space industries.[115] Materials with the Vickers hardness greater than 40 GPa are called superhard. The hardest material known to date is diamond, with the Vickers hardness of 60 to 120 GPa.[96,116] Among the noncarbon superhard materials, the cubic boron nitride displays the Vickers hardness of ~60 GPa.[89,96,117] Materials having unique mechanical properties include some borides, carbides, and nitrides of transition metals, such as chromium,[46,118] rhenium,[92,93,119] molybdenum,[56,57] tungsten,[84,94,120,121], and so forth. Some of these carbides (WC) and nitrides (TiN) are widely used in manufacturing and mining, for example, in the drilling equipment.

Diamond and the cubic BN are high-pressure phases. In this work, we pay more attention to hard and superhard transition metal borides which form the majority of our findings in the Mendelevian search calculations. Many of these are stable at ambient conditions and are expected to be easy to synthesize.

Crystal structure prediction methods develop rapidly[122–124] and can already be used to discover novel superhard materials. To facilitate this, simple and reliable ways of computing the hardness and fracture toughness are necessary.

## 6.2. Empirical Models of Hardness

Hardness is a complicated property of materials that cannot be evaluated directly from the atomistic simulations because it usually includes many nonlinear and mesoscopic effects. However, several empirical models make it possible to estimate the hardness from the atomic-scale properties. Chen's model[58] is based on the assumption that the size of an indentation in the hardness test is correlated with the shear modulus of the material *G*, whereas the width of an imprint is inversely proportional to the square of the bulk modulus *B*, and therefore the proportionality for the Vickers hardness $H_V$ is:[58]

$$H_V \propto G(G/B)^2. \tag{6.1}$$

The analysis of the experimental data for many materials[58] has led to the empirical formula for the Vickers hardness:

$$H_V = 2(k^2G)^{0.585} - 3, \tag{6.2}$$

where $k = G/B$ is the Pugh ratio, *G* is the shear modulus, and *B* is the bulk modulus. $H_V$, *B*, and *G* are expressed in GPa. The test calculations of the Vickers hardness using Chen's model agree well with the reference experimental data:[58] the calculated value for diamond is 98 (experimental value ~96),[116] for TiN — 22.6 (20.5),[125] for c–BN — 56.9 (~55).[116,126]

Although this model is reliable, the calculation of the elastic constants of materials on a large scale is computationally expensive.

The Lyakhov–Oganov model,[51] which computes the hardness from bond hardnesses, is more convenient for high-throughput searches: it is numerically stable, usually reliable, and can be used at no significant cost, requiring as an input only the crystal structure and chemical composition.

## 6.3. A Simple Model of Fracture Toughness from First Principles

For industrial applications, the fracture toughness, along with hardness, also plays a key role. Unfortunately, the hardest materials — covalent crystals — are usually brittle, whereas the materials with the highest fracture toughness are metals whose hardness ranges from low to medium.

The fracture toughness can be calculated using the empirical model from Ref.[83]:

$$K_{IC} = \alpha V^{\frac{1}{6}} G \left(\frac{B}{G}\right)^{\frac{1}{2}}, \tag{6.3}$$

where α is the enhancement factor accounting for the degree of metallicity, *V* is the volume per atom, *G* and *B* are the shear and bulk moduli, respectively.[83] For insulators, semiconductors, transition metal carbides, nitrides, borides, and hydrides, α = 1.[83] The calculated values of fracture toughness of diamond, WC, TiN, and c–BN are close to those measured experimentally: 6.33 MPa·m$^{0.5}$ for diamond (experimental value 4–7 MPa·m$^{0.5}$),[127,128] 5.37 MPa·m$^{0.5}$ for WC (experimental value 5–8 MPa·m$^{0.5}$),[129] 3.9 MPa·m$^{0.5}$ for TiN (experimental value 4–5 MPa·m$^{0.5}$),[129] and 5.41 MPa·m$^{0.5}$ for c–BN (experimental value 2–5 MPa·m$^{0.5}$).[126]

## 6.4. Results of the Mendelevian Search for Hard and Superhard Binary systems

We performed a Mendelevian search for hard and superhard binary systems (Chapter 5). The multi-objective Pareto technique[70] (Chapter 3) and energy filtering, which discarded the structures whose energy was above the convex hull by at least 0.5 eV per atom, were used to ensure that the evolutionary algorithm generates hard phases that have a low energy. In this calculation, 600 binary systems, or about one-fifth of all the systems that can be composed of 74 elements, were studied in 20 MendS generations. Impressively, all the hard unary and binary materials reported in the literature and materials claimed to be potentially hard were found in this calculation, among them diamond and its polytype lonsdaleite as the hardest and boron allotropes as the second hardest elemental phases, were found correctly in our Mendelevian search. $B_xC_y$,[85] $C_xN_y$,[86,87] $B_xN_y$,[88,89] $B_xO_y$,[85,90,91] $Re_xB_y$,[92,93] $W_xB_y$,[94] $Si_xC_y$,[95–98] $W_xC_y$,[96–98] $Al_xO_y$,[96–98] $Ti_xC_y$,[98] $Si_xN_y$,[98] $Ti_xN_y$,[98] $Be_xO_y$,[98] $Ru_xO_y$,[99,100] $Os_xO_y$,[101] $Rh_xB_y$,[102] $Ir_xB_y$,[102] $Os_xB_y$,[103–105] and $Ru_xB_y$.[103–105] are some of the examples of binary systems reported to be hard in the literature and found by us in this single calculation – the list of studied systems is shown in Appendix 2.

The huge size of the compositional space (2775 systems, each with about $10^2$ possible compositions (for limited number of atoms – up to 12 atoms in the primitive cell), each composition having an astronomically large number of possible structures) makes it necessary to reduce the number of generations or population size, or both, to shorten the time of calculations. With reduced computational settings, structures and compositions found may be approximate and may need to be refined using a precise evolutionary calculation for each of these systems. Table 6 presents the results for some promising systems that were further studied using the evolutionary calculations. Of these, some transition metal borides are predicted to be hard; $Mo_xB_y$[56,57] and $Mn_xB_y$[114] have been reported as hard materials; a potential of $Tc_xB_y$,[130] $Fe_xB_y$,[131] and $V_xB_y$[110] to be hard has been discussed. Several previously unknown hard structures more stable than the already reported ones were predicted in these systems (Table 6). Completely new hard systems, $S_xB_y$ and $B_xP_y$, were revealed in the calculations, and the $Mn_xH_y$ system was unexpectedly discovered to contain very hard phases. It must be noted that Table 6 only contains phases, usually, harder than 30 GPa. For instance, in each of the predicted systems, a number of low-energy or stable phases were predicted, but due to the low value of hardness were not included in Table 6. The convex hull plots, are based on both the simulation and the experimental phases. However, we declare that not all the experimentally known phases are included in the convex hull plots.

### 6.4.1. Mo–B

Our calculations detected several simultaneously hard and low-energy structures of the $Mo_xB_y$ system (Table 6, Fig. 27), of which only the stable $R\bar{3}m$ structure of $MoB_2$ was studied before experimentally. The hardness of this structure reported both experimentally (24.2 GPa)[132] and theoretically (33.1 GPa)[56] is in close agreement with the value calculated in this work (28.5 GPa). $MoB_3$ and $MoB_4$ were widely studied before,[56,57] with several low-energy and hard structures reported for these systems (i.e., $R\bar{3}m$-$MoB_3$ having the hardness of 31.8 GPa,[57] $P6_3/mmc$-$MoB_3$ — 37.3 GPa,[56] and much softer $P6_3/mmc$-$MoB_4$ — 8.2 GPa[56]). In

this work, new low-energy structures with high hardness were discovered for these systems (Table 6).

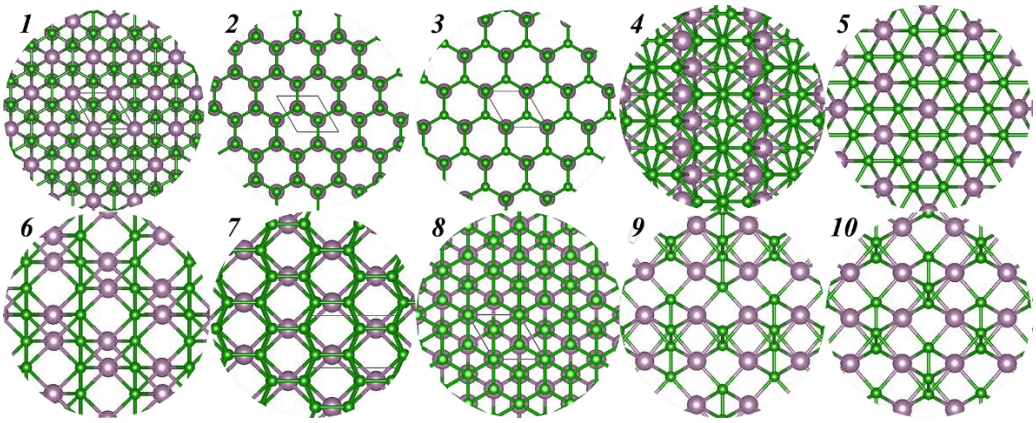



Figure 27. Crystal structures of the Mo–B phases found using the evolutionary calculations: (1) $R\bar{3}m$-$MoB_2$, (2) $P\bar{3}m1$-$MoB_3$, (3) $P\bar{6}m2$-$MoB_5$, (4) $A2/m$-$MoB_3$, (5) $P6_3/mmc$-$MoB_3$, (6) $R\bar{3}m$-$MoB_3$, (7) $Pmmn$-$MoB_4$, (8) $R3m$-$MoB_8$, (9) $Cmcm$-$Mo_2B_3$, (10) $Imm2$-$Mo_2B_3$.

### 6.4.2. Mn–B

We propose several new low-energy and simultaneously hard compounds of the $Mn_xB_y$ system (Table 6). A previous study[113] showed $P2_1/c$-$MnB_4$ as stable and having a very high hardness, obtained both computationally (40.5 GPa)[113] and experimentally (34.6–37.4 GPa),[133] and $C2/m$-$MnB_4$ was claimed to be the second low-energy structure with the energy difference of 18 meV/atom. This research confirms the stability of $P2_1/c$-$MnB_4$. However, we discovered another $MnB_4$ structure, having *Pnnm* space group, whose energy is between the energies of two previously proposed phases of $MnB_4$ (Table 6). In this work, it was found that the ferromagnetic phase of *Pnnm*-$MnB_4$ is more stable than the nonmagnetic one and has the hardness of 40.7 GPa. The crystal structures of studied Mn–B phases are shown in Fig. 28.

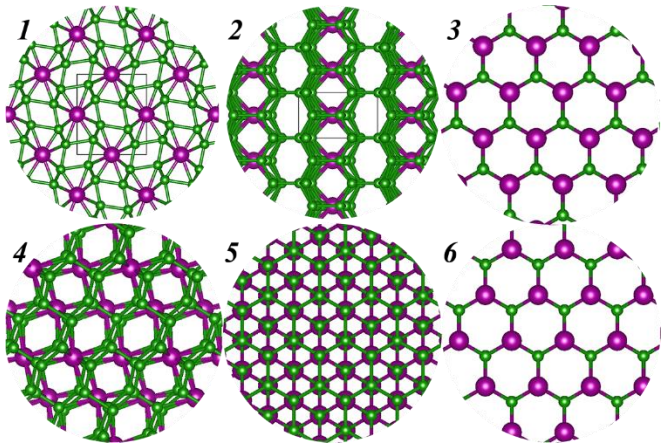



Figure 28. Crystal structures of the Mn–B phases found using the evolutionary calculations: (1) *Pnnm*-$MnB_4$, (2) *Pm*-$MnB_{13}$, (3) $P\bar{6}m2$-$MnB_3$, (4) $P2_1/c$-$MnB_4$, (5) $R\bar{3}m$-$MnB_4$, (6) $P\bar{6}m2$-$MnB_5$.

### 6.4.3. Tc–B

Because of the radioactivity of technetium, the $Tc_xB_y$ system has not been studied experimentally. However, the computational studies of this system started recently.[106,107,130,134] In 2015, $P\bar{3}m1$-TcB was predicted to be energetically more favorable than previously discussed *Cmcm* and WC-type structures.[106] The reported hardness for this structure is 30.3 GPa,[106] which is very close to the hardness of $P\bar{3}m1$-TcB predicted in this work (31 GPa). This structure is positioned 13 meV/atom above the convex hull because of the discovery of other stable compounds (e.g., $Tc_3B_5$) in this work. $P\bar{6}m2$-$TcB_3$ with the predicted hardness of 27.2 GPa was discovered as a stable structure at zero pressure. This structure was also detected in other works[107,134] and was claimed to be synthesizable under pressures above 4 GPa.[107] In addition, we discovered $P\bar{3}m1$-$TcB_3$, another low-energy structure lying 3 meV/atom above the convex hull and having a hardness of 33.1 GPa. $P\bar{6}m2$-$Tc_3B_5$, a compound stable at zero pressure with a hardness of 30.6 GPa, was discovered for the first time in this work. Several low-energy metastable phases of the $Tc_xB_y$ system obtained in this research, with a hardness in the range of 30–36 GPa, are presented in Table 6. Their crystal structures are shown in Fig. 29.

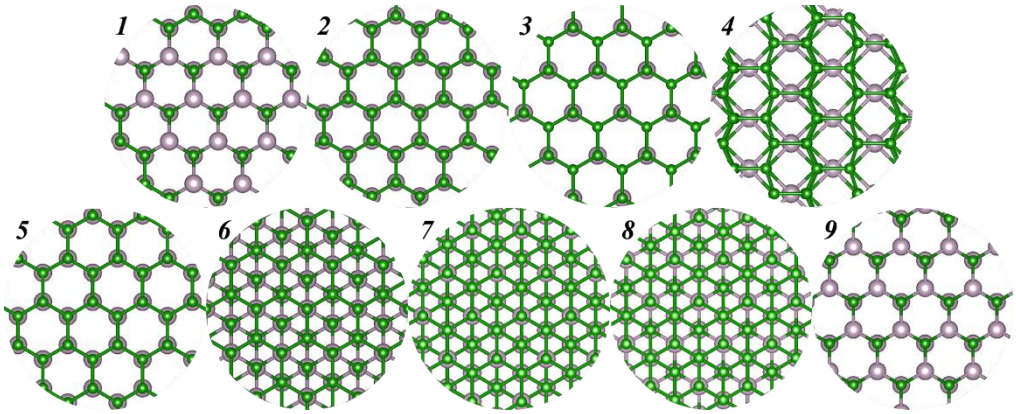



Figure 29. Crystal structures of the Tc–B phases found using the evolutionary calculations: (1) $P\bar{3}m1$-TcB, (2) $P\bar{3}m1$-$TcB_3$, (3) $P\bar{6}m2$-$TcB_3$, (4) $P2_1/m$-$TcB_4$, (5) $P6_3/mmc$-$TcB_4$, (6) $R\bar{3}m$-$TcB_4$, (7) $R3m$-$TcB_7$, (8) $R3m$-$TcB_8$, (9) $P\bar{6}m2$-$Tc_3B_5$.

### 6.4.4. V–B

Many efforts have been focused on searching for low-energy phases of $V_xB_y$ and studying their electrical and mechanical properties. As a result, several low-energy hard and superhard phases have been predicted.[109,110] Nevertheless, the experimental data only exists for the well-known hexagonal $VB_2$ ($AlB_2$-type) with $P6/mmm$ space group,[111] shown in Fig. 30(2). In addition to previously studied *Cmcm*-VB, *Immm*-$V_3B_4$, and $P6/mmm$-$VB_2$,[109] which were also found in our calculations, several boron-rich low-energy phases having very high hardness were discovered (Table 6). Their calculated hardness is very close to or above 40 GPa: 39.7 GPa for $VB_7$, 40 GPa for $VB_5$, and 44.5 GPa for $VB_{12}$. The energy of a newfound extremely hard phase $P\bar{4}m2$-$V_3B_4$ is 6 meV/atom lower than that of the previously known *Immm* structure. The crystal structures of these phases are shown in Fig. 30.

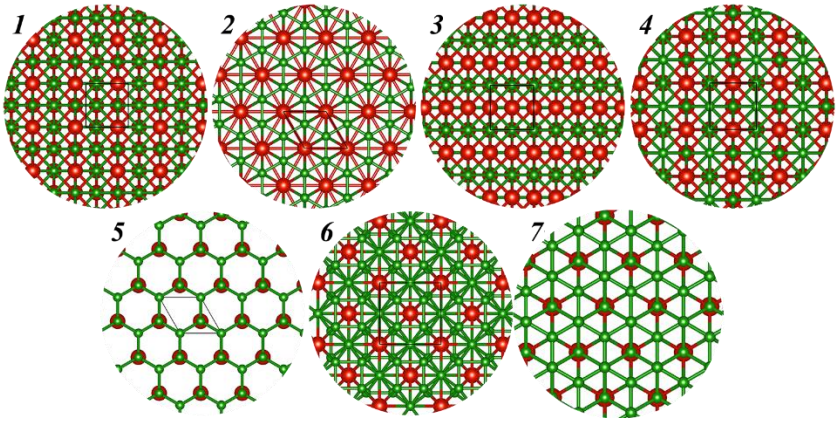



Figure 30. Crystal structures of the V–B phases found using the evolutionary calculations: (1) *Cmcm*-VB, (2) *P*6/*mmm*-$VB_2$, (3) *Immm*-$V_3B_4$, (4) $P\bar{4}m2$-$V_3B_4$, (5) $P\bar{6}m2$-$VB_5$, (6) *I*4/*mmm*-$VB_{12}$, (7) *P*3*m*1-$VB_7$.

### 6.4.5. Fe–B

The studies of the $Fe_xB_y$ system were mostly dedicated to the $FeB_2$ and $FeB_4$ phases,[131,135,136] whereas several works explored different $Fe_xB_y$ compounds.[53,112] The reported stable phases are $Fe_2B$, FeB, and $FeB_2$. In this work we detected another stable phase, $P2_1/m$-$FeB_3$, having a hardness of 30.2 GPa, which to our knowledge has never been reported theoretically or experimentally. The orthorhombic *Pnnm*-$FeB_4$, whose energy is 2 meV/atom above the convex hull (Table 6), has been synthesized at pressures above 8 GPa, and its hardness has been reported to be 62(5) GPa,[137] which has encouraged many computational studies of this structure. However, such a high value of hardness has not been confirmed,[112,131,136,138] and several independent works reported the Vickers hardness varying in the range of 24–29 GPa.[112,131,136,138] The predicted crystal structures of the Fe–B phases are shown in Fig. 31.

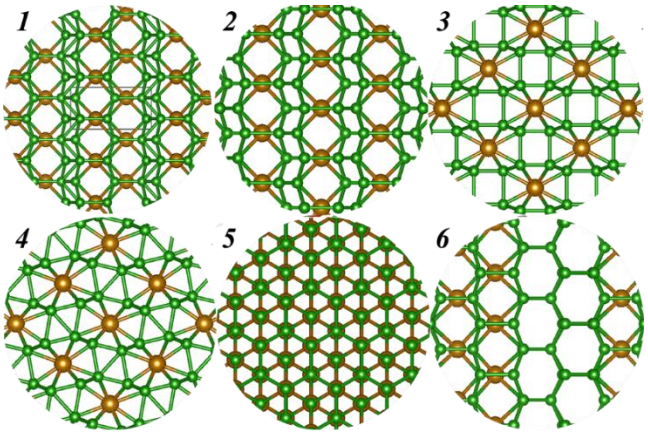



Figure 31. Crystal structures of the Fe–B phases found using the evolutionary calculations: (1) $P2_1/m$-$FeB_3$, (2) *A*2/*m*-$FeB_4$, (3) *Immm*-$FeB_4$, (4) *Pnnm*-$FeB_4$, (5) $R\bar{3}m$-$FeB_4$, (6) *Pm*-$Fe_2B_{11}$.

Table 6. The predicted Vickers hardness ($H_v$), fracture toughness ($K_{1C}$) and enthalpy above the convex hull of selected materials found using MendS. Theoretical values from previous works are shown in parentheses, experimental values are in brackets. The values of hardness for superhard materials are highlighted in bold. The hardness was computed using the Chen-Niu model,[58] the fracture toughness — using the Niu-Niu-Oganov model.[83] Ref: a [95], b [96], c [57], d [56], e [132], f [113], g [133], h [139], i [140], j [106], k [107], m [109], n [111], p [138], q [137].

| | Compounds | $H_v$ (GPa) | $K_{1C}$ (MPa·m$^{1/2}$) | Instability (eV/atom) | Space group | | Compounds | $H_v$ (GPa) | $K_{1C}$ (MPa·m$^{1/2}$) | Instability (eV/atom) | Space group |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Carbon | diamond | **92.7** (**93.6**)[b] [**96**][b] | 6.33 | 0.13 | $Fd\bar{3}m$ | Boron | α-boron | 38.9 (39)[h] [27-34][i] | 2.87 | 0 | $R\bar{3}m$ |
| | lonsdaleite | **93.6** | 6.36 | 0.139 | $P6_3/mmc$ | | B | **44.8** | 3.29 | 0.136 | $Cmc2_1$ |
| B-S | $B_4S_3$ | 30.5 | 1.83 | 0.102 | $Cmcm$ | B-N | BN | **63.4** (**64.5**)[b] [**66**][b] | 5.1 | 0.075 | $F\bar{4}3m$ |
| Mo-B | $MoB_2$ | 28.5 (33.1)[d] [24.2][e] | 3.76 | 0 | $R\bar{3}m$ | Tc-B | TcB | 31 (30.3)[j] | 3.83 | 0.013 | $P\bar{3}m1$ |
| | $MoB_3$ | 35.3 | 3.74 | 0.035 | $P\bar{3}m1$ | | $TcB_3$ | 27.2 (29)[k] | 3.6 | 0 | $P\bar{6}m2$ |
| | $MoB_3$ | 32.2 | 3.63 | 0.077 | $A2/m$ | | $TcB_3$ | 33.1 | 3.79 | 0.003 | $P\bar{3}m1$ |
| | $MoB_3$ | 35.3 (37.3)[d] | 3.63 | 0.017 | $P6_3/mmc$ | | $TcB_4$ | 31.8 | 3.56 | 0.069 | $P2_1/m$ |
| | $MoB_3$ | 33.1 (31.8)[c] | 3.57 | 0.011 | $R\bar{3}m$ | | $TcB_4$ | 30.2 | 3.54 | 0.069 | $R\bar{3}m$ |
| | $MoB_4$ | 35.4 | 3.57 | 0.099 | $Pmmn$ | | $TcB_4$ | 30 (32)[k] | 3.57 | 0.027 | $P6_3/mmc$ |
| | $MoB_5$ | 35.7 | 3.62 | 0.054 | $P\bar{6}m2$ | | $TcB_7$ | 35.9 | 3.35 | 0.084 | $R3m$ |
| | $MoB_8$ | 36.6 | 3.24 | 0.118 | $R3m$ | | $TcB_8$ | 33.9 | 3.3 | 0.113 | $R3m$ |
| | $Mo_2B_3$ | 32.2 | 3.95 | 0.029 | $Imm2$ | | $Tc_3B_5$ | 30.6 | 3.87 | 0 | $P\bar{6}m2$ |
| | $Mo_2B_3$ | 30.4 | 3.87 | 0.043 | $Cmcm$ | | | | | | |
| Si-C | SiC | 33.3 (33.1)[a] [34][b] | 2.94 | 0 | $F\bar{4}3m$ | B-P | BP | 37.2 (31.2)[b] [33][b] | 2.46 | 0 | $F\bar{4}3m$ |
| | SiC | 33.1 | 2.94 | 0.001 | $R3m$ | | $B_6P$ | **41.1** | 2.87 | 0 | $R\bar{3}m$ |
| V-B | VB | 39.1 (38.3)[m] | 3.66 | 0 | $Cmcm$ | Mn-H | MnH | 29.5 | 3.2 | 0 | $P6_3/mmc$ |
| | $VB_2$ | 37.3 (39.5)[m] [27.2][n] | 3.75 | 0 | $P6/mmm$ | | MnH | 27.9 | 3.14 | 0.013 | $R\bar{3}m$ |
| | $VB_5$ | **40** | 3.36 | 0.158 | $P\bar{6}m2$ | | MnH | 26.3 | 3.07 | 0.044 | $Fm\bar{3}m$ |
| | $VB_7$ | 39.7 | 3.19 | 0.143 | $P3m1$ | | $Mn_3H_2$ | 26.8 | 3.22 | 0.017 | $R32$ |
| | $VB_{12}$ | **44.5** | 3.34 | 0.125 | $I4/mmm$ | | $Mn_3H_2$ | 27 | 3.26 | 0.019 | $P6_3/mcm$ |
| | $V_3B_4$ | 37.8 | 3.74 | 0 | $P\bar{4}m2$ | | $Mn_4H_3$ | 27.6 | 3.23 | 0.002 | $P2/m$ |
| | $V_3B_4$ | 35.9 (38.2)[m] | 3.7 | 0.006 | $Immm$ | | $Mn_6H_5$ | 27.3 | 3.17 | 0.011 | $A2/m$ |
| Mn-B | $MnB_3$ | 32.2 | 3.5 | 0.029 | $P\bar{6}m2$ | Fe-B | $FeB_3$ | 30.2 | 3.32 | 0 | $P2_1/m$ |
| | $MnB_4$[†] | **40.7** | 3.65 | 0.009 | $Pnnm$ | | $FeB_4$ | 35.7 | 3.06 | 0.021 | $Immm$ |
| | $MnB_4$ | 38.2 | 3.56 | 0.1 | $R\bar{3}m$ | | $FeB_4$[‡] | 32 | 3.31 | 0.039 | $R\bar{3}m$ |
| | $MnB_4$ | 38.1 (**40.5**)[f] [37.4][g] | 3.76 | 0 | $P2_1/c$ | | $FeB_4$ | **42.7** | 3.31 | 0.063 | $A2/m$ |
| | $MnB_5$ | 32.7 | 3.38 | 0.097 | $P\bar{6}m2$ | | $FeB_4$ | 28.6 (28.4)[p] [**62**][q] | 3.32 | 0.002 | $Pnnm$ |
| | $MnB_{13}$ | **40.4** | 2.9 | 0.181 | $Pm$ | | $Fe_2B_{11}$ | 33.8 | 3.37 | 0.081 | $Pm$ |

[†‡] For these phases we found that ferromagnetic solutions are more stable than non-magnetic. Elastic constant were computed assuming these are ferromagnetic structures, the energy difference between the ferromagnetic and non-magnetic solutions for † and ‡ is 0.037 (eV/transition-metal) and 0.092 (eV/transition-metal) and magnetization is equal to 0.016 and 0.034 $\mu_B \cdot Å^{-3}$, respectively.

### 6.4.6. B–P and Si–C

The cubic boron phosphide BP with the zinc blende structure is a well-known compound in the $B_xP_y$ system (Fig. 32(3)); its hardness was considered to be roughly the same as that of SiC.[108] In our calculations, the hardnesses of SiC and BP were found to be 33 GPa and 37 GPa, respectively. In addition, another stable compound in this system, $B_6P$, was discovered and predicted to be superhard. The computed hardness of $B_6P$ exceeds 41 GPa. This structure contains the α-rhombohedral boron-like clusters which are surrounded by phosphorus atoms, and is different from the experimentally known icosahedra $B_{12}P_2$. For the SiC system, a polytype of β-SiC having *R*3*m* space group was found, with nearly the same hardness and the energy 1 meV/atom higher than that of the known β-SiC with a diamond structure.

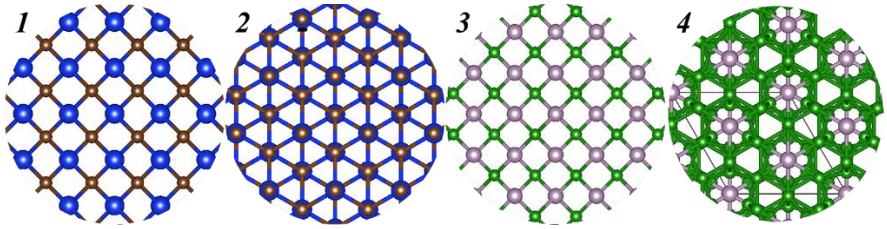



Figure 32. Crystal structures of the Si–C and B–P phases found using the evolutionary calculations: (1) $F\bar{4}3m$-SiC, (2) *R*3*m*-SiC, (3) $F\bar{4}3m$-BP, (4) $R\bar{3}m$-$B_6P$.

### 6.4.7. Mn–H and B–S

Several very hard phases were found for $Mn_xH_y$, an unexpected entry in the list of hard systems (Table 6). All these phases are nonmagnetic, highly symmetric (Fig. 33), and energetically favorable, being positioned either on the convex hull or close to it. Their hardness reaches 30 GPa. Two thermodynamically stable – located on the convex hull – compounds, $P\bar{3}m1$-$Mn_2H$ and $P6_3/mmc$-MnH, were discovered in this system, with the computed hardness of 21.5 GPa and 29.5 GPa, respectively. Only the structures having a hardness above 26 GPa are shown in Table 6.

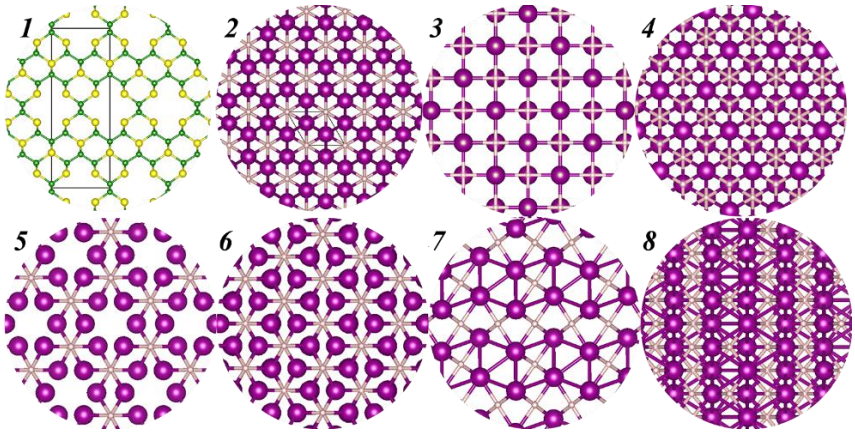



Figure 33. Crystal structures of $B_4S_3$ and the Mn–H phases found using the evolutionary calculations: (1) *Cmcm*-$B_4S_3$, (2) $P6_3/mmc$-MnH, (3) $Fm\bar{3}m$-MnH, (4) $R\bar{3}m$-MnH, (5) $P6_3/mcm$-$Mn_3H_2$, (6) *R*32-$Mn_3H_2$, (7) *P*2/*m*-$Mn_4H_3$, (8) *A*2/*m*-$Mn_6H_5$.

Generally, the $B_xS_y$ system is not hard, but metastable boron sulfides are potentially hard. We found a low-energy metastable phase *Cmcm*-$B_4S_3$ (Fig. 33(1)) with a hardness exceeding 30 GPa, which can stimulate future studies of this system.

### 6.4.8. Cr–C and Cr–B

Chromium-based materials reveal attractive mechanical properties. In the Cr–C system, only three stable carbides, $Cr_{23}C_6$, $Cr_3C_2$, and $Cr_7C_3$,[141–144] are known from experiments. Two metastable chromium carbides, CrC and $Cr_3C$, have also been synthesized.[145–147] In the Cr–B system, six different phases are known from experiments: $Cr_2B$, $Cr_5B_3$, CrB, $Cr_3B_4$, $CrB_2$, and $CrB_4$;[118,148–150] their mechanical properties have been examined theoretically.[46,148,151] Our recent study of chromium carbides[46] using the evolutionary algorithm USPEX[2,4,81] combined with the Pareto optimization of the enthalpy of formation and hardness[46] led to the prediction of a new phase, *Pmn*2-$Cr_2C$, in addition to the already known phases.[152] This *Pmn*2-$Cr_2C$ phase is anticipated to have a shear modulus of 292 GPa, the highest among all chromium carbides, and the Vickers hardness of 27 GPa.[46] The highest Vickers hardness among the chromium-based materials, achieved in $CrB_4$, has been reported to be in the range of 29–44 GPa,[46,153] whereas our calculations yield 48 GPa.

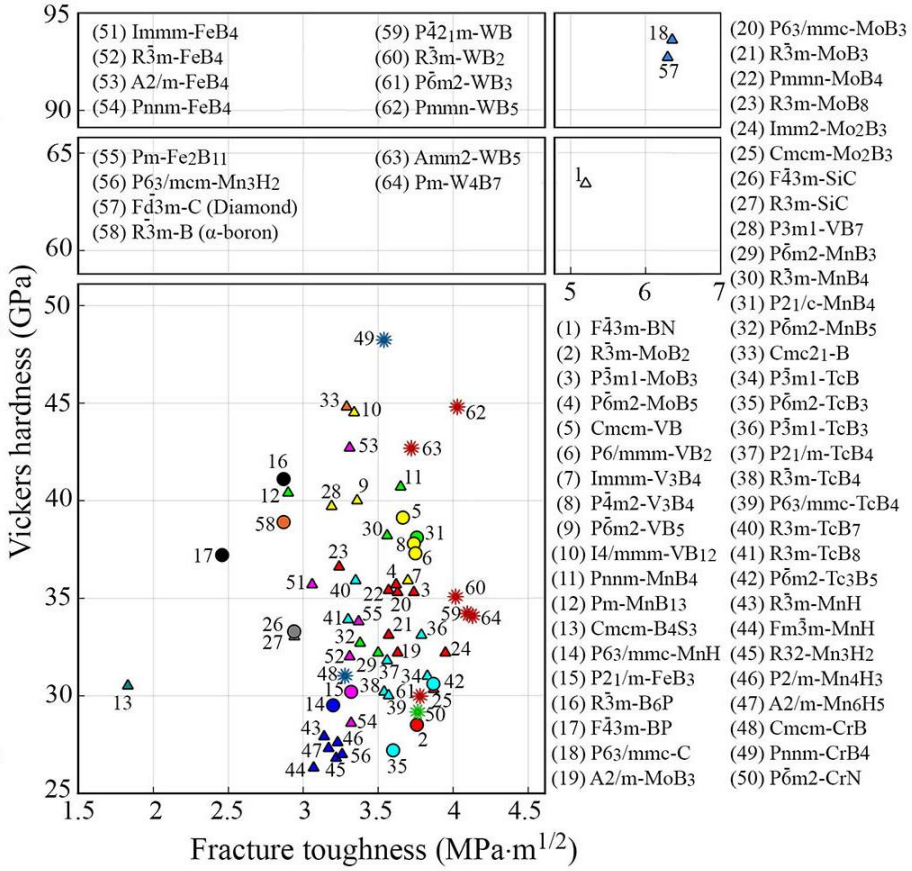



Figure 34. 2D plot of the Vickers hardness vs. fracture toughness. Stable hard compounds from the previous works[24, 25] are shown as "suns"; stable and metastable compounds found in this work are represented by circles and triangles, respectively.

In our calculations, some boron hydrides were predicted to be superhard, but they are not included in Table 6 because of their high energy. However, it may be possible to stabilize these hard phases under pressure or by a chemical modification.

Fig. 34 shows the studied materials in the space of hardness and fracture toughness. Diamond and cubic BN possess the best properties but are metastable at normal conditions; among the stable phases, borides of transition metals, especially those from groups VB, VIB, and VIIB stand out: we noted $VB_2$, $V_3B_4$, $MoB_2$, $CrB_4$, $WB_5$, and $MnB_4$ in particular. These and related materials are of high technological interest.

The fact that all known binary superhard systems were found in a short coevolutionary run demonstrates the power of our method, which can be applied to other types of materials.

More details on the structure relaxation and input parameters can be found in Appendix 1.

## 6.5. Conclusion

We discussed several examples of the transition metal boride systems that were computationally predicted and comprehensively studied, some of them potentially promising for practical applications. The new method, Mendelevian search, shows great predictive power in the search for hard and superhard materials. Modern computational techniques enable the prediction and relatively fast development of new materials with enhanced properties, like superhard compounds, that are destined to replace the traditional materials.

# CHAPTER 7.
# Discussion and Conclusions

This work aimed to find an unbiased and nonempirical solution for the main problem of computational materials science – prediction of new materials with optimal properties.

Over the last few decades, with the development of algorithms and computers, computational materials science advanced rapidly. Simulation of materials on computers became possible and using quantum mechanical calculations one could optimize a given crystal structure and compute its different electronic, magnetic, optic, etc., properties. Later, with the implementation of algorithms such as evolutionary algorithms in computers, crystal structure prediction became possible. A huge number of new crystal structures for given materials were predicted in computers and many of them got confirmed by experiment. Computational materials science opened a new way for the design of new materials and helped experiment to save time and reduce the cost of the trial-and-error materials discovery.

Despite all these developments, the central problem of materials science – finding new materials with optimal properties – remained unsolved because of its complexity.

To solve such a challenging problem, we proposed the idea of an organized chemical space (also called Mendelevian space). In Mendelevian space, materials with similar properties are located in neighboring regions. If such space can be created, the prediction of new materials becomes possible by finding and exploring only those regions of the space that are more likely to contain optimal materials for the given properties. This way, prediction of one optimal material leads to the prediction of another optimal material and so on. We used the most important properties of elements, electronegativity and atomic radius, to define such space for elements where similar elements take neighboring places and therefore, compounds of these elements are likely to have similar properties in Mendelevian space.

In the next step, we implemented a coevolutionary method, in which its variation operators (i.e. the central cores of the coevolutionary algorithm) are carefully designed to perform well in such an organized chemical space. Therefore, the outcome of a coevolutionary search is a set of the most optimal materials for the given properties.

Each material, binary system in our work, has a large number of compositions (different stoichiometries), each with an infinite number of crystal structures. The best compounds and crystal structures of each material are predicted using the evolutionary algorithm USPEX. Multi-objective Pareto optimization was implemented in evolutionary algorithm USPEX to ensure that the predicted crystal structures are the best at least in one property, or sufficiently good in multiple properties simultaneously.

Therefore, to solve the aforementioned problem, we implemented, a fully nonempirical, Mendelevian Search (MendS) method that combines our coevolutionary approach with the carefully restructured “Mendelevian” chemical space, energy filtering, Pareto optimization, and evolutionary algorithm crystal structure prediction.

For testing our method, we searched for (1) low-energy hard and superhard binary systems, and (2) low-energy magnetic binary systems. The results of both searches are very promising,

however, we mainly focused on the results of the search for hard and superhard materials in this thesis.

Using MendS, we searched for low-energy and hard crystal structures in 630 binary systems out of 2701 possible systems (about one-fifth of the entire chemical space) in 30 coevolutionary generations. Amazingly, all the hard and superhard binary systems reported in decades, as well as those that were reported to have the potential to be hard were predicted in our search (list of the calculated systems are shown in Appendix 2 – Table 7). In each of these systems, several new phases were predicted, while only a few of the most promising ones are reported in this thesis. Mn-H was predicted to be unexpectedly hard. Well-known superhard systems — diamond, boron allotropes, and the B–N system — were found in a single calculation together with other hard systems, both known and unknown. A similar chemical map was produced for magnetic materials; well-known magnetic systems such as Ni, Co, Fe were found within a few generations. The examples of hard materials and ferromagnets found using this method show its power and efficiency, which can be used to search for optimal materials – with any complexity, at least ternaries – with any combination of properties at arbitrary conditions. As the first step in the prediction of novel materials possessing desired properties, this fully nonempirical method to a large extent solves the central problem of computational materials science.

# APPENDIX 1.

## Input Parameters

The Pareto optimization of hardness and stability was performed over all possible structures with up to 12 atoms in the primitive cell in the space of all possible binary compounds formed by 74 elements (excluding the noble gases, rare earth elements, and elements heavier than Pu). The input parameters of MendS were: population size: 30, 30% chemical heredity, 30% reactive heredity, 20% mutation, and 20% random systems.

For each binary system, the evolutionary optimization was run with the following parameters: initial population size: 100, subsequent population size: 50, number of generations: 6, 40% heredity, 30% mutation, 30% random selection.

The comprehensive multi-objective evolutionary algorithm as implemented in USPEX was run on select promising binary systems with the following input parameters: initial population size: 120, subsequent population size: 60, number of generations: 50, 40% heredity, 20% transmutation, 20% softmutation, 20% random selection.

The underlying ab initio structure relaxations and energy calculations were performed using the density functional theory with the projector augmented-wave method (PAW) as implemented in the VASP code.[48,50] In the spin-polarized calculations, the GGA-PBE functional[17] was implemented. The cutoff energy of 600 eV and a *k*-mesh with the resolution of $0.06 \times \text{Å}^{-1}$ was used in all VASP calculations. In 20 coevolutionary generations, 600 binary systems were studied. This calculation took 3 weeks using 280 cores, a surprisingly minor cost for a full exploration of the chemical space.

# APPENDIX 2.

## List of Studied Systems in the Mendelevian Search for Hard and Superhard Materials

Table 7. Binary systems investigated during the MendS run. The variation operators used for creating the compounds are: R — random selection, CH — chemical heredity, RH — reactive heredity, M — mutation.

| Gen | Compound | Variation operator | 1st parent | 2nd parent | Gen | Compound | Variation operator | 1st parent | 2nd parent |
|---|---|---|---|---|---|---|---|---|---|
| 1 | Rh–B | R | – | – | 2 | Cr–Co | CH | B | Mo |
| 1 | Na–B | R | – | – | 2 | Rh–W | CH | C | Nb |
| 1 | Ca–Po | R | – | – | 2 | Rh–Rh | CH | Mo | C |
| 1 | Zr–Ge | R | – | – | 2 | Os–P | CH | Mo | Pt |
| 1 | H–N | R | – | – | 2 | Al–Si | CH | Mo | Np–Co |
| 1 | Y–N | R | – | – | 2 | V–Cr | CH | Os | Be |
| 1 | Ga–H | R | – | – | 2 | P–P | CH | Mo | Ru |
| 1 | La–Pt | R | – | – | 2 | Ga–S | CH | Nb–N | As |
| 1 | Sn–B | R | – | – | 2 | Re–C | CH | Nb–N | Rh |
| 1 | Ru–N | R | – | – | 2 | Mo–C | RH | C | Mo–Ir |
| 1 | Rb–Be | R | – | – | 2 | Fe–B | RH | Fe | B |
| 1 | W–C | R | – | – | 2 | As–Co | RH | As | Np–Co |
| 1 | Fe–Os | R | – | – | 2 | Al–Os | RH | Os–Ru | Bi–Al |
| 1 | Ti–Pd | R | – | – | 2 | As–Mo | RH | As | Mo |
| 1 | Th–Be | R | – | – | 2 | Co–Rh | RH | Co | Rh |
| 1 | Rb–Mo | R | – | – | 2 | Al–Be | RH | Be | Al–As |
| 1 | Ra–Zr | R | – | – | 2 | Ru–C | RH | Ru | W–C |
| 1 | Ca–Rh | R | – | – | 2 | W–H | RH | W–C | Ga–H |
| 1 | Np–Co | R | – | – | 2 | Pd–W | M | W | – |
| 1 | Rb–H | R | – | – | 2 | Sb–Rh | M | Rh | – |
| 1 | Bi–Al | R | – | – | 2 | Sn–N | M | Ru–N | – |
| 1 | Os–Ru | R | – | – | 2 | Ta–C | M | C | – |
| 1 | Li–Zn | R | – | – | 2 | In–Rh | M | Rh | – |
| 1 | Al–As | R | – | – | 2 | Mn–Be | M | Be | – |
| 1 | Mo–Ir | R | – | – | 2 | Pu–Ge | R | – | – |
| 1 | Zr–Fe | R | – | – | 2 | Li–Ir | R | – | – |
| 1 | Ac–Zn | R | – | – | 2 | Li–Fe | R | – | – |
| 1 | Nb–N | R | – | – | 2 | Ta–Po | R | – | – |
| 1 | Hf–Si | R | – | – | 2 | Sr–P | R | – | – |
| 1 | Pa–Ir | R | – | – | 2 | P–Cl | R | – | – |
| 3 | As–P | CH | C | F–B | 4 | Be–Fe | CH | C | Mn |
| 3 | Fe–Co | CH | C | C–Cr | 4 | Fe–Fe | CH | Cr | B–H |
| 3 | B–H | CH | Fe | C | 4 | Pd–Au | CH | As | Mo–H |
| 3 | Mo–H | CH | W–H | B | 4 | Cu–P | CH | V–Mo | Ru |
| 3 | Cu–Mo | CH | Ru | Co–Rh | 4 | Pb–Nb | CH | Cr | Fe |
| 3 | Ga–Os | CH | B | C | 4 | Ge–Ge | CH | B | Sn |
| 3 | Co–Ni | CH | V | B | 4 | Cr–Cr | CH | Fe | Mn |
| 3 | Cr–Cu | CH | Mn | Os | 4 | Ti–Ru | CH | Cr–Cu | Cr–Rh |
| 3 | Cu–Os | CH | Ir | Co–Rh | 4 | P–Pt | CH | As–P | Mo–H |
| 3 | Mn–B | RH | Mn | B | 4 | As–B | RH | B | As |
| 3 | V–P | RH | V | Os–P | 4 | B–C | RH | C | B–H |
| 3 | Pd–Os | RH | Pd | Al–Os | 4 | Fe–Cl | RH | Fe | Cl |
| 3 | V–Mo | RH | V | Mo–C | 4 | Np–Mn | RH | Mn | Np |

| Gen | Compound | Variation operator | 1st parent | 2nd parent | Gen | Compound | Variation operator | 1st parent | 2nd parent |
|---|---|---|---|---|---|---|---|---|---|
| 3 | Cr–Ir | RH | Cr | Ir | 4 | Ca–C | RH | Ca | C |
| 3 | Sn–C | RH | Sn | C | 4 | Mn–P | RH | Mn | As–P |
| 3 | Cr–Rh | RH | Cr | Co–Rh | 4 | Pb–Mn | RH | Mn | Pb |
| 3 | Mn–Pd | RH | Mn | Pd–W | 4 | V–Ir | RH | V | Cr–Ir |
| 3 | As–C | RH | Mo–C | As–Mo | 4 | P–B | RH | B | P |
| 3 | Pt–Rh | M | Co–Rh | – | 4 | Cd–Mn | M | Mn | – |
| 3 | Se–C | M | Mo–C | – | 4 | Mn–C | M | C | – |
| 3 | Ge–W | M | Pd–W | – | 4 | Ge–Cr | M | Cr | – |
| 3 | Rh–C | M | Ru–C | – | 4 | Ge–C | M | C | – |
| 3 | Se–W | M | W–H | – | 4 | Fe–Rh | M | Fe | – |
| 3 | Mo–Ru | M | Mo | – | 4 | Ti–V | M | V | – |
| 3 | Li–Ta | R | – | – | 4 | V–V | R | – | – |
| 3 | Po–P | R | – | – | 4 | Ru–Rh | R | – | – |
| 3 | Sr–Cl | R | – | – | 4 | Bi–As | R | – | – |
| 3 | Zn–P | R | – | – | 4 | Mn–Rh | R | – | – |
| 3 | Al–Sb | R | – | – | 4 | Nb–C | R | – | – |
| 3 | Pb–V | R | – | – | 4 | Ge–Au | R | – | – |
| 5 | Al–Ru | CH | C | B | 6 | Fe–Se | CH | Cr–C | B–C |
| 5 | Be–Be | CH | Mn | B–C | 6 | Pd–Ir | CH | Fe | C |
| 5 | Fe–Ni | CH | Cr | B–C | 6 | Nb–Ga | CH | Cr–B | Cr–C |
| 5 | As–H | CH | Nb–C | B–H | 6 | Fe–Ru | CH | B | Mn–S |
| 5 | Zn–Cu | CH | V | Mn–Rh | 6 | Tc–Pt | CH | C | Co |
| 5 | Hg–Fe | CH | B | B–C | 6 | Al–P | CH | Mn–Pt | Be–B |
| 5 | Mo–P | CH | Cr | Mn–Rh | 6 | Cu–Cu | CH | P | Co |
| 5 | Si–Cu | CH | Be | Fe | 6 | Ru–H | CH | Mn–S | Cr |
| 5 | Co–Co | CH | Cr | B | 6 | Cu–Co | CH | Fe | Ni |
| 5 | Cr–C | RH | B–C | Ge–Cr | 6 | Mo–B | RH | B–H | Mo |
| 5 | P–H | RH | B–H | P–B | 6 | Cu–H | RH | P–H | Cu |
| 5 | Be–B | RH | B | Be | 6 | Fe–Mo | RH | Fe | Mo |
| 5 | Cr–B | RH | Ge–Cr | Mn–B | 6 | Ni–P | RH | Fe–Ni | Fe–P |
| 5 | Fe–Ir | RH | Fe | V–Ir | 6 | Si–B | RH | Si | B–H |
| 5 | Be–Rh | RH | Mn–Rh | Be | 6 | Cr–P | RH | P | Cr |
| 5 | Mn–Cu | RH | Mn | Cu–P | 6 | Fe–S | RH | Fe–Ni | Mn–S |
| 5 | Fe–P | RH | Fe | Cu–P | 6 | Mn–H | RH | P–H | Mn–Pt |
| 5 | Ir–H | RH | V–Ir | B–H | 6 | Mn–As | RH | Mn | As |
| 5 | Mo–Pd | M | Pd | – | 6 | Ge–Fe | M | Fe–Ir | – |
| 5 | Mn–S | M | Mn–C | – | 6 | Tc–Ni | M | Fe–Ni | – |
| 5 | Pd–Rh | M | Rh | – | 6 | Os–S | M | Mn–S | – |
| 5 | Sc–Nb | M | Nb | – | 6 | H–H | M | B–H | – |
| 5 | V–Ru | M | V–Ir | – | 6 | U–Mn | M | Mn | – |
| 5 | Mn–Pt | M | Mn–Rh | – | 6 | Be–Pt | M | Mn–Pt | – |
| 5 | K–Ti | R | – | – | 6 | Ca–Be | R | – | – |
| 5 | Hg–Ag | R | – | – | 6 | Mo–S | R | – | – |
| 5 | Po–Be | R | – | – | 6 | In–Br | R | – | – |
| 5 | Pu–Sn | R | – | – | 6 | Sc–Co | R | – | – |
| 5 | Cs–O | R | – | – | 6 | Cr–N | R | – | – |
| 5 | Ca–Th | R | – | – | 6 | Ga–V | R | – | – |
| 7 | Ni–B | CH | Cr–N | B–C | 8 | Ni–C | CH | Si–H | B–H |
| 7 | Co–Os | CH | Mo–S | B | 8 | Zn–Zn | CH | Ti | Tc |
| 7 | Os–Pt | CH | Mo–S | B | 8 | Ni–Ni | CH | Co | Fe–Ru |
| 7 | Zn–Fe | CH | Si–B | Cr | 8 | Al–H | CH | Mn–Fe | Cr–N |
| 7 | Ga–Ru | CH | Cr–P | Mn–H | 8 | Ge–Cu | CH | Te | Cr–B |
| 7 | Al–Cr | CH | Ru–H | Mn–H | 8 | Po–Fe | CH | Mn–Fe | C |
| 7 | Ni–Ru | CH | Tc–Pt | B | 8 | Tc–Br | CH | Cr–N | Te |
| 7 | Be–Co | CH | B–H | Mn | 8 | Ti–Ti | CH | Al–Cr | Ti–Cr |
| 7 | Co–Ru | CH | C | Cr–P | 8 | Tc–Ir | CH | Mn–C | Mo |

| Gen | Compound | Variation operator | 1st parent | 2nd parent | Gen | Compound | Variation operator | 1st parent | 2nd parent |
|---|---|---|---|---|---|---|---|---|---|
| 7 | Si–H | RH | Si | Mn–H | 8 | Mn–Tc | RH | Mn | Tc |
| 7 | Tc–P | RH | P | Tc | 8 | Cr–H | RH | Mn–H | Mn–Cr |
| 7 | Cr–Mo | RH | Mo–S | Cr | 8 | Tc–Fe | RH | Tc | Fe |
| 7 | S–B | RH | Mo–S | Si–B | 8 | Os–H | RH | Mn–H | Os–Pt |
| 7 | Tc–B | RH | B–C | Tc | 8 | Cr–Fe | RH | Cr–N | Mn–Fe |
| 7 | Co–B | RH | B–C | Co | 8 | Cr–Os | RH | Os | Mn–Cr |
| 7 | Mn–Cr | RH | Cr–N | Mn–H | 8 | Mn–N | RH | Ti–Mn | Cr–N |
| 7 | Mn–Co | RH | Co | Mn–As | 8 | Ti–Be | RH | Be | Ti–Cr |
| 7 | Mn–Fe | RH | Mn–As | Fe | 8 | Pd–B | RH | B–H | Pd |
| 7 | Fe–H | M | H | – | 8 | Zn–Mo | M | Cr–Mo | – |
| 7 | Tl–Ru | M | Ru–H | – | 8 | B–B | M | Tc–B | – |
| 7 | Ti–Cr | M | Cr | – | 8 | Sb–Ni | M | Ni | – |
| 7 | Te–Mn | M | Mn | – | 8 | V–Tc | M | Tc–B | – |
| 7 | Al–Pt | M | Al–P | – | 8 | Be–Ni | M | Be | – |
| 7 | Ti–Mn | M | Mn–As | – | 8 | Mn–Mo | M | Cr–Mo | – |
| 7 | Ca–Ni | R | – | – | 8 | Ac–Si | R | – | – |
| 7 | Zn–Be | R | – | – | 8 | Hf–Bi | R | – | – |
| 7 | Br–S | R | – | – | 8 | Pa–Si | R | – | – |
| 7 | In–Co | R | – | – | 8 | Pu–Tc | R | – | – |
| 7 | Rb–Sc | R | – | – | 8 | U–Se | R | – | – |
| 7 | Tc–Tc | R | – | – | 8 | Pd–O | R | – | – |
| 9 | P–C | CH | Os–H | B–H | 10 | Pa–P | CH | Rh | B–H |
| 9 | Co–Ir | CH | Mn–C | C | 10 | Si–Fe | CH | Mn | Cr–Ni |
| 9 | Ni–O | CH | Cr–Fe | Mn–N | 10 | Co–P | CH | V–B | B–H |
| 9 | Tc–Cu | CH | Mo | Cr | 10 | In–Ge | CH | Cr–B | Cr–H |
| 9 | Ga–Be | CH | Cr–B | Mn | 10 | V–Ni | CH | Tc–Ru | B |
| 9 | Cu–Ru | CH | Cr–H | Os | 10 | Ge–Ni | CH | Cr | Cr–N |
| 9 | Hg–Ni | CH | C | Cr–B | 10 | Hf–Be | CH | Mn | Cr–H |
| 9 | Cd–Be | CH | C | Mn–H | 10 | Nb–Co | CH | Cr | Si |
| 9 | Cu–Rh | CH | Cr–N | Os | 10 | Np–Cr | CH | Po | Fe |
| 9 | V–B | RH | V | B | 10 | C–O | RH | Ni–O | Tc–C |
| 9 | Ni–N | RH | Cr–N | Ni | 10 | Co–N | RH | Cr–N | Co |
| 9 | Mn–Ni | RH | Ni–C | Mn–H | 10 | Si–Tc | RH | Si | Tc–C |
| 9 | Cr–Ni | RH | Ni–C | Cr–Fe | 10 | Tc–H | RH | B–H | Tc |
| 9 | C–H | RH | B–H | C | 10 | Ga–C | RH | Ga | C |
| 9 | Tc–C | RH | C | Tc | 10 | Nb–V | RH | Nb–Ga | V–B |
| 9 | Ir–N | RH | Ir | Cr–N | 10 | Ga–Ni | RH | Ni | Ga–Be |
| 9 | Tc–Po | RH | Tc | Po | 10 | Be–N | RH | Be | Cr–N |
| 9 | C–N | RH | Mn–N | C | 10 | Po–H | RH | B–H | Po |
| 9 | Zn–Mn | M | Mn | – | 10 | Re–Fe | M | Fe | – |
| 9 | Co–H | M | B–H | – | 10 | Zr–Be | M | Be | – |
| 9 | Cu–C | M | B–C | – | 10 | V–Re | M | V | – |
| 9 | V–Fe | M | Fe | – | 10 | Ga–Tc | M | Ga–Be | – |
| 9 | Tc–Ru | M | Tc–Ir | – | 10 | Cr–W | M | Cr–N | – |
| 9 | Si–C | M | Ni–C | – | 10 | Re–Cu | M | Cu | – |
| 9 | Y–Np | R | – | – | 10 | Sr–U | R | – | – |
| 9 | Li–Sn | R | – | – | 10 | Y–Zr | R | – | – |
| 9 | Sr–I | R | – | – | 10 | Al–Mn | R | – | – |
| 9 | Te–Co | R | – | – | 10 | Sb–Au | R | – | – |
| 9 | Rb–Ba | R | – | – | 10 | Ag–C | R | – | – |
| 9 | Hg–Be | R | – | – | 10 | Tl–Fe | R | – | – |
| 11 | V–Rh | CH | Mn | Al–Mn | 12 | P–N | CH | Br–O | B–H |
| 11 | Mo–Ni | CH | V–Re | Mn–H | 12 | Hg–Os | CH | B–H | B–C |
| 11 | Po–Cu | CH | V–Ni | C | 12 | Be–F | CH | Mn–O | O |
| 11 | Cd–Pt | CH | Ga–C | Np–Cr | 12 | Mo–Mo | CH | P | Si |
| 11 | Al–Co | CH | B | Ga–C | 12 | Mg–Fe | CH | Al–C | Zn–Cr |

| Gen | Compound | Variation operator | 1st parent | 2nd parent | Gen | Compound | Variation operator | 1st parent | 2nd parent |
|---|---|---|---|---|---|---|---|---|---|
| 11 | Mg–In | CH | B–C | Ni | 12 | Re–Ru | CH | Zn–Cr | Si |
| 11 | Zn–Cr | CH | V | Tc–H | 12 | Ta–Be | CH | Cr | Be |
| 11 | C–C | CH | B–H | O | 12 | Co–C | CH | B–H | Cr–O |
| 11 | Co–Mo | CH | Cr–W | Ni | 12 | Be–C | CH | B–H | Be–O |
| 11 | Al–C | RH | B–C | Al–Mn | 12 | Cr–Be | RH | Be | Cr–N |
| 11 | Cr–O | RH | O | Cr | 12 | Zn–V | RH | Zn–Cr | V |
| 11 | V–Si | RH | V | Si | 12 | B–O | RH | Mn–O | Cr–B |
| 11 | Mn–O | RH | Mn | O | 12 | Ga–Mo | RH | Mo | Ga |
| 11 | V–Mn | RH | Al–Mn | V | 12 | N–O | RH | O | Cr–N |
| 11 | V–Be | RH | Zr–Be | V | 12 | Be–Br | RH | Br | V–Be |
| 11 | Fe–N | RH | Fe | Cr–N | 12 | Nb–Cu | RH | Nb–Ga | Cu |
| 11 | Cr–Tc | RH | Tc | Cr–N | 12 | Ga–B | RH | Ga | B–C |
| 11 | Be–O | RH | Be | O | 12 | Zn–O | RH | O | Zn–Cr |
| 11 | Br–O | M | O | – | 12 | Pd–N | M | Fe–N | – |
| 11 | Tl–Mn | M | Mn | – | 12 | Ti–B | M | B–H | – |
| 11 | Pu–N | M | Cr–N | – | 12 | Si–O | M | Cr–O | – |
| 11 | Ge–B | M | B | – | 12 | Nb–Tc | M | Cr–Tc | – |
| 11 | Tc–Co | M | Tc–H | – | 12 | Pa–Be | M | Be | – |
| 11 | Tc–W | M | Tc | – | 12 | Si–S | M | Fe–S | – |
| 11 | Pu–Cu | R | – | – | 12 | Np–Os | R | – | – |
| 11 | Ba–Co | R | – | – | 12 | Te–Ru | R | – | – |
| 11 | Hf–Pa | R | – | – | 12 | Rb–Li | R | – | – |
| 11 | Cu–W | R | – | – | 12 | Y–Nb | R | – | – |
| 11 | Nb–P | R | – | – | 12 | Sc–I | R | – | – |
| 11 | Fr–Cu | R | – | – | 12 | Tl–Pd | R | – | – |
| 13 | N–N | CH | C | O | 14 | V–C | CH | Cr–B | Cr–N |
| 13 | P–Rh | CH | Ti | O | 14 | Ga–Rh | CH | Se | Mn–H |
| 13 | Pu–S | CH | Cr–N | Nb–Tc | 14 | Cu–Ni | CH | Ga | B |
| 13 | Sn–Ru | CH | Cr–B | Ga–Mo | 14 | Zr–Zn | CH | Nb–Mn | Cr |
| 13 | Re–Tc | CH | Ti | Mo | 14 | Pd–Ru | CH | P–Rh | C |
| 13 | Rh–O | CH | Zn–O | F | 14 | Ni–Ir | CH | B–N | P–Rh |
| 13 | Ru–Ru | CH | P | F | 14 | S–N | CH | O | Fe |
| 13 | In–Ru | CH | Si–S | C | 14 | Ti–Mo | CH | Cr | Fe |
| 13 | Au–C | CH | Fe | B–C | 14 | Zn–Tc | CH | Mo | V |
| 13 | P–O | RH | B–O | P | 14 | V–N | RH | Al–V | Cr–N |
| 13 | B–N | RH | P–N | Cr–B | 14 | Nb–Fe | RH | Nb–Mn | Fe–Se |
| 13 | Cr–Se | RH | Cr–Be | Fe–Se | 14 | As–O | RH | As | P–O |
| 13 | Cr–Ru | RH | Cr | Re–Ru | 14 | C–F | RH | C | F |
| 13 | Nb–Mo | RH | Nb–Ga | Ga–Mo | 14 | Nb–H | RH | Nb–Mn | Mn–H |
| 13 | Be–Mo | RH | Mo | Be–C | 14 | Nb–O | RH | Nb–Mn | O |
| 13 | Fe–Br | RH | Fe | Br | 14 | V–O | RH | Rh–O | V |
| 13 | Re–B | RH | Re | B | 14 | Ru–B | RH | Cr–B | Cr–Ru |
| 13 | Si–P | RH | Si | P | 14 | N–F | RH | B–N | F |
| 13 | Al–V | M | V | – | 14 | S–O | M | O | – |
| 13 | Pa–O | M | Si–O | – | 14 | Ta–Mn | M | Mn | – |
| 13 | Nb–Mn | M | Mn | – | 14 | Zn–N | M | B–N | – |
| 13 | Mn–I | M | Mn | – | 14 | Ir–O | M | P–O | – |
| 13 | Re–H | M | Mn–H | – | 14 | Si–N | M | Tc–N | – |
| 13 | Tc–N | M | P–N | – | 14 | Cu–B | M | Cr–B | – |
| 13 | Ac–C | R | – | – | 14 | Ac–Tl | R | – | – |
| 13 | Ga–As | R | – | – | 14 | Sb–I | R | – | – |
| 13 | Ge–As | R | – | – | 14 | Sb–Be | R | – | – |
| 13 | Ag–Ni | R | – | – | 14 | Pa–Tc | R | – | – |
| 13 | Hf–F | R | – | – | 14 | Ge–Si | R | – | – |
| 13 | Ca–Ac | R | – | – | 14 | U–Au | R | – | – |
| 15 | Ni–Br | CH | Si–N | N–F | 16 | Fe–C | CH | Mo–N | B |

| Gen | Compound | Variation operator | 1st parent | 2nd parent | Gen | Compound | Variation operator | 1st parent | 2nd parent |
|---|---|---|---|---|---|---|---|---|---|
| 15 | Ni–F | CH | Si–N | B–N | 16 | Y–Co | CH | Mn–H | Cr |
| 15 | Mo–N | CH | As–O | Ni–P | 16 | Ni–Os | CH | B–N | Si |
| 15 | I–Fe | CH | O | V–N | 16 | U–Co | CH | Cr–N | Cr |
| 15 | Cu–S | CH | Ta–Mn | F | 16 | Re–Au | CH | Mo | V–Ge |
| 15 | Si–Rh | CH | Pd | V–O | 16 | Tc–Mo | CH | B–C | Mo–N |
| 15 | Ag–W | CH | Nb–Fe | Cr–N | 16 | Mn–Au | CH | C | B–N |
| 15 | Pt–Au | CH | C | Ni | 16 | Ni–S | CH | B–N | Fe |
| 15 | Fe–O | CH | Cr–N | F | 16 | Zn–Si | CH | V | Ru |
| 15 | V–F | RH | N–F | V–C | 16 | Cr–Pt | RH | Pt | Cr–B |
| 15 | Ru–O | RH | V–O | Ru–B | 16 | As–Ir | RH | Ir | As |
| 15 | Zr–V | RH | Zr | V–N | 16 | Cr–Si | RH | Cr–N | Si |
| 15 | H–F | RH | Mn–H | N–F | 16 | Si–F | RH | Si | F |
| 15 | Tc–F | RH | F | Tc | 16 | B–F | RH | H–F | B–C |
| 15 | Cr–F | RH | N–F | Cr–N | 16 | Mo–O | RH | O | Mo–N |
| 15 | V–Ge | RH | V | Ge | 16 | Mn–Ru | RH | Mn–H | Ru |
| 15 | As–Ni | RH | As–O | Ni | 16 | As–F | RH | F | Cr–As |
| 15 | Cr–As | RH | As | Cr | 16 | Zr–O | RH | Zr–V | O |
| 15 | Bi–V | M | V–C | – | 16 | V–Ag | M | V | – |
| 15 | Si–Ir | M | Si–N | – | 16 | Nb–Si | M | Si | – |
| 15 | Cd–Ga | M | Nb–Ga | – | 16 | Pu–Ni | M | Ni | – |
| 15 | Mg–C | M | V–C | – | 16 | O–F | M | F | – |
| 15 | Hg–F | M | N–F | – | 16 | Tc–As | M | Cr–As | – |
| 15 | Pu–H | M | Mn–H | – | 16 | Ir–B | M | B–N | – |
| 15 | Sr–Co | R | – | – | 16 | K–Sb | R | – | – |
| 15 | Ag–Mn | R | – | – | 16 | Li–Cu | R | – | – |
| 15 | Ga–At | R | – | – | 16 | In–S | R | – | – |
| 15 | Ti–Cu | R | – | – | 16 | Ac–F | R | – | – |
| 15 | Ag–Pd | R | – | – | 16 | Pu–F | R | – | – |
| 15 | U–Si | R | – | – | 16 | Hg–Te | R | – | – |
| 17 | Ga–O | CH | Fe–C | Cr–N | 18 | Ta–Ru | CH | Si–Mo | Cr–N |
| 17 | Zn–Co | CH | B–H | Cr–B | 18 | Bi–Fe | CH | Cr–N | B–C |
| 17 | Ni–Rh | CH | Si | B–C | 18 | Ag–Fe | CH | C | B–C |
| 17 | Cd–Tc | CH | Cr–Si | Mo | 18 | Re–P | CH | Cr–Pd | Mo |
| 17 | Fe–Pt | CH | O | Cr–N | 18 | Mg–As | CH | B–C | Nb–B |
| 17 | Rh–N | CH | C | Mo–O | 18 | Al–Fe | CH | Mn | Cr |
| 17 | Al–Mo | CH | Tc | Tc–As | 18 | Co–S | CH | B–C | Cr–Au |
| 17 | I–S | CH | Mo–O | B–N | 18 | Zn–Ge | CH | B–N | C |
| 17 | Co–O | CH | Fe–C | F | 18 | Zn–B | CH | Mn–H | Cr |
| 17 | Mn–Si | RH | Mn | Cr–Si | 18 | Tc–Au | RH | Tc | Au |
| 17 | V–H | RH | V | Mn–H | 18 | Al–Rh | RH | Rh | Al–Mo |
| 17 | Mn–F | RH | Mn | O–F | 18 | H–O | RH | O–F | B–H |
| 17 | Ir–C | RH | Fe–C | Ir–B | 18 | Si–Pt | RH | Mn–Si | Pt |
| 17 | Tc–O | RH | O–F | Tc–As | 18 | Nb–Cr | RH | Cr | Nb |
| 17 | Si–Mo | RH | Tc–Mo | Si | 18 | Mo–F | RH | Mo | O–F |
| 17 | As–N | RH | Cr–N | As | 18 | Pd–C | RH | Cr–Pd | B–C |
| 17 | Pt–F | RH | Cr–Pt | F | 18 | Mo–Rh | RH | Mo | Rh |
| 17 | Nb–B | RH | Nb–Si | B–N | 18 | Ga–F | RH | F | Ga–O |
| 17 | Pa–Mn | M | Mn | – | 18 | Hg–Cr | M | Cr | – |
| 17 | Hg–H | M | B–H | – | 18 | Pu–Pt | M | Pt | – |
| 17 | Ag–Tc | M | Tc | – | 18 | Cd–Pd | M | Pd | – |
| 17 | In–C | M | C | – | 18 | Ti–N | M | Cr–N | – |
| 17 | Cr–Au | M | Cr–N | – | 18 | Ag–N | M | B–N | – |
| 17 | Cr–Pd | M | Cr–B | – | 18 | Zr–Mn | M | Mn–H | – |
| 17 | Mg–Be | R | – | – | 18 | Ta–Mo | R | – | – |
| 17 | Ba–Hg | R | – | – | 18 | Hf–Ni | R | – | – |
| 17 | Hg–Ir | R | – | – | 18 | U–Cu | R | – | – |

| Gen | Compound | Variation operator | 1st parent | 2nd parent | Gen | Compound | Variation operator | 1st parent | 2nd parent |
|---|---|---|---|---|---|---|---|---|---|
| 17 | Bi–At | R | – | – | 18 | Ga–Po | R | – | – |
| 17 | Pa–Sb | R | – | – | 18 | Li–Tl | R | – | – |
| 17 | Ru–Br | R | – | – | 18 | Os–Se | R | – | – |
| 19 | Tc–Rh | CH | Pd | Cr–N | 20 | Re–I | CH | Br | Cr–N |
| 19 | S–Cl | CH | Ni | O | 20 | Be–Cu | CH | Mn | Os–B |
| 19 | Th–Rh | CH | C–C | Mo–Rh | 20 | I–Os | CH | Re | Tc–Rh |
| 19 | Ir–Rh | CH | Se | P | 20 | Bi–Ni | CH | Tc–Rh | Os–B |
| 19 | Au–S | CH | Se | C | 20 | Be–S | CH | Mn–H | B–H |
| 19 | Br–C | CH | O–F | B–N | 20 | Ta–Co | CH | Zr–Cr | Zr–Ni |
| 19 | Ti–C | CH | B–H | B–N | 20 | Sn–Al | CH | Zr–Cr | Re–Re |
| 19 | Zr–Nb | CH | Cr | Mn | 20 | Ag–Mo | CH | As | Mn–Re |
| 19 | Rh–H | CH | B–N | Mo | 20 | Cl–H | CH | B | C–C |
| 19 | Mn–Re | RH | Zr–Mn | Re–P | 20 | Zr–C | RH | C | Zr–Ni |
| 19 | Zr–B | RH | Zr–Mn | B | 20 | Os–Rh | RH | Rh | Os–B |
| 19 | Ta–O | RH | O | Ta–Mo | 20 | Rh–Cl | RH | S–Cl | Rh–H |
| 19 | Zr–Ni | RH | Ni | Zr–Mn | 20 | Mn–Os | RH | Os–B | Mn |
| 19 | Zr–Cr | RH | Zr–Mn | Cr–N | 20 | S–F | RH | S | F |
| 19 | As–Rh | RH | As | Rh | 20 | Re–Os | RH | Re | Os–B |
| 19 | Pt–N | RH | Cr–N | Pt | 20 | Si–Os | RH | Os–B | Si |
| 19 | Re–F | RH | Re–P | O–F | 20 | Zr–Rh | RH | Rh | Zr–B |
| 19 | Mn–Se | RH | Se | Mn | 20 | Re–Ir | RH | Re | Ir–Rh |
| 19 | Os–B | M | B | – | 20 | Re–Rh | M | Rh–H | – |
| 19 | Re–Si | M | Re | – | 20 | Pt–B | M | B–N | – |
| 19 | Se–F | M | F | – | 20 | Ti–As | M | Ti–C | – |
| 19 | Bi–Cr | M | Cr–N | – | 20 | Ta–Cr | M | Cr | – |
| 19 | Co–F | M | F | – | 20 | Ti–Si | M | Si | – |
| 19 | Cd–Au | M | Tc–Au | – | 20 | W–B | M | B–C | – |
| 19 | Pu–Ru | R | – | – | 20 | K–Pd | R | – | – |
| 19 | I–N | R | – | – | 20 | At–C | R | – | – |
| 19 | Ac–Be | R | – | – | 20 | At–O | R | – | – |
| 19 | Li–C | R | – | – | 20 | Li–Ti | R | – | – |
| 19 | U–Pt | R | – | – | 20 | Rb–Na | R | – | – |
| 19 | Zn–Os | R | – | – | 20 | Pb–C | R | – | – |

# References


1. Martoňák, R., Laio, A. & Parrinello, M. Predicting Crystal Structures: The Parrinello-Rahman Method Revisited. *Phys. Rev. Lett.* **90**, 075503 (2003).
2. Oganov, A. R. & Glass, C. W. Crystal structure prediction using ab initio evolutionary techniques: principles and applications. *J. Chem. Phys.* **124**, 244704 (2006).
3. Oganov, A. R., Lyakhov, A. O. & Valle, M. How evolutionary crystal structure prediction works--and why. *Acc. Chem. Res.* **44**, 227–37 (2011).
4. Lyakhov, A. O., Oganov, A. R., Stokes, H. T. & Zhu, Q. New developments in evolutionary structure prediction algorithm USPEX. *Comput. Phys. Commun.* **184**, 1172–1182 (2013).
5. Fankuchen, I. Fifty Years of X-ray Diffraction. P. P. Ewald, Ed. Oosthoek's, Utrecht, Netherlands, 1962. ix + 717 pp. Illus. $11.25. *Science (80-. ).* **138**, 1254–1255 (1962).
6. Boland, W. R., Kamgnia, E. R. & Kowalik, J. S. A conjugate-gradient optimization method invariant to nonlinear scaling. *J. Optim. Theory Appl.* **27**, 221–230 (1979).
7. Ypma, T. J. Historical development of the Newton-Raphson method. *SIAM Rev.* **37**, 531–551 (1995).
8. Head, J. D. & Zerner, M. C. A Broyden-Fletcher-Goldfarb-Shanno optimization procedure for molecular geometries. *Chem. Phys. Lett.* **122**, 264–270 (1985).
9. Hohenberg, P. & Kohn, W. Inhomogeneous Electron Gas. *Phys. Rev.* **136**, B864–B871 (1964).
10. Kohn, W. & Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects. *Phys. Rev.* **140**, A1133–A1138 (1965).
11. Zhao, X. Accelerating materials discovery and design: computational study of the structure and properties of materials. *Graduate Theses and Dissertations* (Iowa State University, Digital Repository, 2015). doi:10.31274/etd-180810-3999
12. Hafner, J., Wolverton, C. & Ceder, G. Toward Computational Materials Design: The Impact of Density Functional Theory on Materials Research. *MRS Bull.* **31**, 659–668 (2006).
13. Ceperley, D. M. & Alder, B. J. Ground state of the electron gas by a stochastic method. *Phys. Rev. Lett.* **45**, 566–569 (1980).
14. Perdew, J. P. & Zunger, A. Self-interaction correction to density-functional approximations for many-electron systems. *Phys. Rev. B* **23**, 5048–5079 (1981).
15. Wandelt, K. (Klaus). *Surface and interface science. Volume 4, Solid-solid interfaces and thin films*. (2020).
16. Jones, R. O. & Gunnarsson, O. The density functional formalism, its applications and prospects. *Rev. Mod. Phys.* **61**, 689–746 (1989).
17. Perdew, J. P., Burke, K. & Ernzerhof, M. Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **77**, 3865–3868 (1996).
18. van Laarhoven, P. J. M. & Aarts, E. H. L. *Simulated Annealing: Theory and Applications*. (Springer Netherlands, 1987). doi:10.1007/978-94-015-7744-1
19. Kirkpatrick, S., Gelatt, C. D. & Vecchi, M. P. Optimization by simulated annealing. *Science (80-. ).* **220**, 671–680 (1983).
20. Wales, D. & Doye, J. 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 (1998).

21. Goedecker, S. Minima hopping: An efficient search method for the global minimum of the potential energy surface of complex molecular systems. **120**, 9911–9917 (2004).
22. Laio, A. & Parrinello, M. Escaping free-energy minima. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 12562–12566 (2002).
23. Wang, Y. & Ma, Y. Perspective: Crystal structure prediction at high pressures. *J. Chem. Phys.* **140**, 040901 (2014).
24. Flores-Livas, J. A. *et al.* High-pressure structures of disilane and their superconducting properties. *Phys. Rev. Lett.* **108**, 117004 (2012).
25. Amsler, M. *et al.* Novel structural motifs in low energy phases of LiAlH 4. *Phys. Rev. Lett.* **108**, 205505 (2012).
26. Huan, T. D., Amsler, M., Tuoc, V. N., Willand, A. & Goedecker, S. Low-energy structures of zinc borohydride Zn(BH4)2. *Phys. Rev. B - Condens. Matter Mater. Phys.* **86**, 224110 (2012).
27. Pickard, C. J. & Needs, R. J. Ab initio random structure searching. *J. Phys. Condens. Matter* **23**, 053201 (2011).
28. Eberhart, R. & Kennedy, J. New optimizer using particle swarm theory. in *Proceedings of the International Symposium on Micro Machine and Human Science* 39–43 (IEEE, 1995). doi:10.1109/mhs.1995.494215
29. Goldberg, D. E. & Holland, J. H. Genetic Algorithms and Machine Learning. *Machine Learning* **3**, 95–99 (1988).
30. Glass, C. W., Oganov, A. R. & Hansen, N. USPEX—Evolutionary crystal structure prediction. *Comput. Phys. Commun.* **175**, 713–720 (2006).
31. Wang, Y., Lv, J., Zhu, L. & Ma, Y. Crystal structure prediction via particle-swarm optimization. *Phys. Rev. B - Condens. Matter Mater. Phys.* **82**, 094116 (2010).
32. Zhang, W. *et al.* Unexpected stable stoichiometries of sodium chlorides. *Science* **342**, 1502–5 (2013).
33. Mannix, A. J. *et al.* Synthesis of borophenes: Anisotropic, two-dimensional boron polymorphs. *Science* **350**, 1513–6 (2015).
34. Dong, X. *et al.* A stable compound of helium and sodium at high pressure. *Nat. Chem.* **9**, 440–445 (2017).
35. Lipowski, A. & Lipowska, D. Roulette-wheel selection via stochastic acceptance. *Phys. A Stat. Mech. its Appl.* **391**, 2193–2196 (2012).
36. Bushlanov, P. V., Blatov, V. A. & Oganov, A. R. Topology-based crystal structure generator. *Comput. Phys. Commun.* **236**, 1–7 (2019).
37. Lyakhov, A. O., Oganov, A. R. & Valle, M. How to predict very large and complex crystal structures. *Comput. Phys. Commun.* **181**, 1623–1632 (2010).
38. Ashby, M. F. *Materials selection in mechanical design*. (Butterworth-Heinemann, 2011).
39. Schaffer, J. D. Multiple objective optimization with vector evaluated genetic algorithms. *1st Int. Conf. Genet. Algorithms* 93–100 (1985).
40. Ngatchou, P., Zarei, A. & El-Sharkawi, A. Pareto Multi Objective Optimization. *Proc. 13th Int. Conf. on, Intell. Syst. Appl. to Power Syst.* 84–91 (2005). doi:10.1109/ISAP.2005.1599245
41. Srinivas, N. & Deb, K. Multiobjective optimization using nondominated sorting in

genetic algorithms. *Evol. Comput.* **2**, 221–248 (1994).

42. Deb, K., Pratap, A., Agarwal, S. & Meyarivan, T. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Trans. Evol. Comput.* **6**, 182–197 (2002).
43. Corne, D., Jerram, N., Knowles, J. D., Oates, M. & Martin, J. PESA-II: Region-based Selection in Evolutionary Multiobjective Optimization. *Proc. Genet. Evol. Comput. Conf.* 283–290 (2001). doi:citeulike-article-id:8133801
44. Zitzler, E. & Thiele, L. Multiobjective evolutionary algorithms: a comparative case study and the strength Pareto approach. *IEEE Trans. Evol. Comput.* **3**, 257–271 (1999).
45. Zitzler, E., Laumanns, M. & Thiele, L. SPEA2: Improving the Strength Pareto Evolutionary Algorithm. *Evol. Methods Des. Optim. Control with Appl. to Ind. Probl.* 95–100 (2001). doi:10.1.1.28.7571
46. Kvashnin, A. G., Oganov, A. R., Samtsevich, A. I. & Allahyari, Z. Computational Search for Novel Hard Chromium-Based Materials. *J. Phys. Chem. Lett.* **8**, 755–764 (2017).
47. Núñez-Valdez, M., Allahyari, Z., Fan, T. & Oganov, A. R. A. R. Efficient technique for computational design of thermoelectric materials. *Comput. Phys. Commun.* **222**, 152–157 (2018).
48. Kresse, G. & Joubert, D. From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B* **59**, 1758–1775 (1999).
49. Blöchl, P. E. Projector augmented-wave method. *Phys. Rev. B* **50**, 17953–17979 (1994).
50. Kresse, G. & Furthmüller, J. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Phys. Rev. B* **54**, 11169–11186 (1996).
51. Lyakhov, A. O. & Oganov, A. R. Evolutionary search for superhard materials: Methodology and applications to forms of carbon and TiO2. *Phys. Rev. B* **84**, 092103 (2011).
52. Yu, S. *et al.* Exploring the Real Ground-State Structures of Molybdenum Dinitride. *J. Phys. Chem. C* **120**, 11060–11067 (2016).
53. Kolmogorov, A. N. *et al.* New Superconducting and Semiconducting Fe-B Compounds Predicted with an *Ab Initio* Evolutionary Search. *Phys. Rev. Lett.* **105**, 217003 (2010).
54. Geest, A. G. Van Der & Kolmogorov, A. N. CALPHAD : Computer Coupling of Phase Diagrams and Thermochemistry Stability of 41 metal – boron systems at 0 GPa and 30 GPa from fi rst principles. **46**, 184–204 (2014).
55. Spear, K. E. & Liao, P. K. The B−Mo (Boron-Molybdenum) system. *Bull. Alloy Phase Diagrams* **9**, 457–466 (1988).
56. Liang, Y., Yuan, X., Fu, Z., Li, Y. & Zhong, Z. An unusual variation of stability and hardness in molybdenum borides. *Appl. Phys. Lett.* **101**, 1–6 (2012).
57. Zhang, M., Wang, H. H. H., Wang, H. H. H., Cui, T. & Ma, Y. Structural modifications and mechanical properties of molybdenum borides from first principles. *J. Phys. Chem. C* **114**, 6722–6725 (2010).
58. Chen, X.-Q., Niu, H., Li, D. & Li, Y. Modeling hardness of polycrystalline materials and bulk metallic glasses. *Intermetallics* **19**, 1275–1281 (2011).
59. Pettifor, D. G. A chemical scale for crystal-structure maps. *Solid State Commun.* **51**, 31–34 (1984).
60. Pettifor, D. G. The structures of binary compounds. I. Phenomenological structure

maps. *J. Phys. C Solid State Phys.* **19**, 285–313 (1986).

61. Villars, P. *et al.* Binary, ternary and quaternary compound former/nonformer prediction via Mendeleev number. *J. Alloys Compd.* **317**–**318**, 26–38 (2001).
62. Glawe, H., Sanna, A., Gross, E. K. U. & Marques, M. A. L. The optimal one dimensional periodic table: a modified Pettifor chemical scale from data mining. *New J. Phys.* **18**, 093011 (2016).
63. Goldschmidt, V. M. Crystal structure and chemical constitution. *Trans. Faraday Soc.* **25**, 253 (1929).
64. Ringwood, A. E. The principles governing trace element distribution during magmatic crystallization Part I: The influence of electronegativity. *Geochim. Cosmochim. Acta* **7**, 189–202 (1955).
65. Villars, P., Daams, J., Shikata, Y., Rajan, K. & Iwata, S. A new approach to describe elemental-property parameters. *Chem. Met. Alloy.* **1**, 1–23 (2008).
66. Belsky, A., Hellenbrandt, M., Karen, V. L., Luksch, P. & IUCr. New developments in the Inorganic Crystal Structure Database (ICSD): accessibility in support of materials research and design. *Acta Crystallogr. Sect. B Struct. Sci.* **58**, 364–369 (2002).
67. Nagle, J. K. Atomic polarizability and electronegativity. *J. Am. Chem. Soc.* **112**, 4741–4747 (1990).
68. Pauling, L. The nature of the chemical bond. IV. The energy of single bonds and the relative electronegativity of atoms. *J. Am. Chem. Soc.* **54**, 3570–3582 (1932).
69. Gražulis, S. *et al.* Crystallography Open Database (COD): an open-access collection of crystal structures and platform for world-wide collaboration. *Nucleic Acids Res.* **40**, D420–D427 (2012).
70. Allahyari, Z. & Oganov, A. R. Multi-Objective Optimization as a Tool for Material Design. in *Handbook of Materials Modeling* 1–15 (Springer International Publishing, 2019). doi:10.1007/978-3-319-50257-1_71-1
71. Rodriguez, A. & Laio, A. Clustering by fast search and find of density peaks. *Science (80-. ).* **344**, 1492–1496 (2014).
72. Mazhnik, E. & Oganov, A. R. A model of hardness and fracture toughness of solids. *J. Appl. Phys.* **126**, 125109 (2019).
73. Allahyari, Z. & Oganov, A. R. Coevolutionary search for optimal materials in the space of all possible compounds. *NPJ Comput. Mater.* **6**, 55 (2020).
74. Cedillos-Barraza, O. *et al.* Investigating the highest melting temperature materials: A laser melting study of the TaC-HfC system. *Sci. Rep.* **6**, 1–11 (2016).
75. Rahm, M., Cammi, R., Ashcroft, N. W. & Hoffmann, R. Squeezing All Elements in the Periodic Table: Electron Configuration and Electronegativity of the Atoms under Compression. *J. Am. Chem. Soc.* (2019). doi:10.1021/jacs.9b02634
76. Dong, X. *et al.* How do chemical properties of the atoms change under pressure. (2015). doi:arXiv:1503.00230 [cond-mat.mtrl-sci]
77. Curtarolo, S., Morgan, D., Persson, K., Rodgers, J. & Ceder, G. Predicting Crystal Structures with Data Mining of Quantum Calculations. *Phys. Rev. Lett.* **91**, 135503 (2003).
78. Villars, P. & Iwata, S. Pauling File verifies / reveals 12 principles in materials science supporting four cornerstones given by Nature. *Chem. Met. Alloy.* **6**, 81–108 (2013).
79. Zhu, Q., Oganov, A. R. & Lyakhov, A. O. Novel stable compounds in the Mg–O system under high pressure. *Phys. Chem. Chem. Phys.* **15**, 7696 (2013).

80. Zhu, Q., Oganov, A. R., Salvadó, M. A., Pertierra, P. & Lyakhov, A. O. Denser than diamond: Ab initio search for superdense carbon allotropes. *Phys. Rev. B* **83**, 193410 (2011).
81. Oganov, A. R., Ma, Y., Lyakhov, A. O., Valle, M. & Gatti, C. Evolutionary Crystal Structure Prediction as a Method for the Discovery of Minerals and Materials. *Rev. Mineral. Geochemistry* **71**, 271–298 (2010).
82. Valle, M., Oganov, A. R. & IUCr. Crystal fingerprint space – a novel paradigm for studying crystal-structure sets. *Acta Crystallogr. Sect. A Found. Crystallogr.* **66**, 507–517 (2010).
83. Niu, H., Niu, S. & Oganov, A. R. Simple and accurate model of fracture toughness of solids. *J. Appl. Phys.* **125**, 065105 (2019).
84. Kvashnin, A. G. *et al.* New Tungsten Borides, Their Stability and Outstanding Mechanical Properties. *J. Phys. Chem. Lett.* **9**, 3470–3477 (2018).
85. Haines, J., Léger, J. & Bocquillon, G. Synthesis and Design of Superhard Materials. *Annu. Rev. Mater. Res.* **31**, 1–23 (2001).
86. Liu, A. Y. & Cohen, M. L. Prediction of new low compressibility solids. *Science (80-. ).* **245**, 841–843 (1989).
87. Teter, D. M. & Hemley, R. J. Low-Compressibility Carbon Nitrides. *Science* **271**, 53–55 (1996).
88. He, C. *et al.* Z-BN: a novel superhard boron nitride phase. *Phys. Chem. Chem. Phys.* **14**, 10967 (2012).
89. Li, Y., Hao, J., Liu, H., Lu, S. & Tse, J. S. High-Energy Density and Superhard Nitrogen-Rich B-N Compounds. *Phys. Rev. Lett.* **115**, 105502 (2015).
90. Sasaki, T., Akaishi, M., Yamaoka, S., Fujiki, Y. & Oikawa, T. Simultaneous crystallization of diamond and cubic boron nitride from the graphite relative boron carbide nitride (BC2N) under high pressure/high temperature conditions. *Chem. Mater.* **5**, 695–699 (1993).
91. Hervé Hubert *et al.* High-Pressure, High-Temperature Synthesis and Characterization of Boron Suboxide (B6O). (1998). doi:10.1021/CM970433+
92. Chung, H.-Y. *et al.* Synthesis of ultra-incompressible superhard rhenium diboride at ambient pressure. *Science* **316**, 436–9 (2007).
93. Latini, A. *et al.* Superhard Rhenium Diboride Films: Preparation and Characterization. *Chem. Mater.* **20**, 4507–4511 (2008).
94. Gu, Q., Krauss, G. & Steurer, W. Transition Metal Borides: Superhard versus Ultra-incompressible. *Adv. Mater.* **20**, 3620–3626 (2008).
95. Gao, F. Theoretical model of intrinsic hardness. *Phys. Rev. B* **73**, 132104 (2006).
96. Gao, F. *et al.* Hardness of Covalent Crystals. *Phys. Rev. Lett.* **91**, 015502 (2003).
97. Šimůnek, A. & Vackář, J. Hardness of Covalent and Ionic Crystals: First-Principle Calculations. *Phys. Rev. Lett.* **96**, 085501 (2006).
98. Sung, C.-M. & Sung, M. Carbon nitride and other speculative superhard materials. *Mater. Chem. Phys.* **43**, 1–18 (1996).
99. Leger, J. M., Haines, J. & Blanzat, B. Materials potentially harder than diamond: Quenchable high-pressure phases of transition metal dioxides. *J. Mater. Sci. Lett.* **13**, 1688–1690 (1994).
100. Haines, J. & Léger, J. M. Phase transitions in ruthenium dioxide up to 40 GPa: Mechanism for the rutile-to-fluorite phase transformation and a model for the high-

pressure behavior of stishovite SiO 2. *Phys. Rev. B* **48**, 13344–13350 (1993).

101. Lundin, U. *et al.* Transition-metal dioxides with a bulk modulus comparable to diamond. *Phys. Rev. B* **57**, 4979–4982 (1998).
102. Rau, J. V. & Latini, A. New Hard and Superhard Materials: $RhB_{1.1}$ and $IrB_{1.35}$. *Chem. Mater.* **21**, 1407–1409 (2009).
103. Chung, H.-Y. Y., Weinberger, M. B., Yang, J.-M. M., Tolbert, S. H. & Kaner, R. B. Correlation between hardness and elastic moduli of the ultraincompressible transition metal diborides RuB2, OsB2, and ReB2. *Appl. Phys. Lett.* **92**, 261904 (2008).
104. Robert W. Cumberland *et al.* Osmium Diboride, An Ultra-Incompressible, Hard Material. *J. Am. Chem. Soc.* **127**, 7264–7265 (2005).
105. Hebbache, M., Stuparević, L. & Živković, D. A new superhard material: Osmium diboride OsB2. *Solid State Commun.* **139**, 227–231 (2006).
106. Zhang, G.-T., Bai, T.-T., Yan, H.-Y. & Zhao, Y.-R. New crystal structure and physical properties of TcB from first-principles calculations. *Chinese Phys. B* **24**, 106104 (2015).
107. Miao, X., Xing, W., Meng, F. & Yu, R. Prediction on technetium triboride from first-principles calculations. *Solid State Commun.* **252**, 40–45 (2017).
108. Woo, K., Lee, K. & Kovnir, K. BP: Synthesis and properties of boron phosphide. *Mater. Res. Express* **3**, 074003 (2016).
109. Pan, Y., Lin, Y. H., Guo, J. M. & Wen, M. Correlation between hardness and bond orientation of vanadium borides. *RSC Adv.* **4**, 47377–47382 (2014).
110. Wu, L. *et al.* Unraveling Stable Vanadium Tetraboride and Triboride by First-Principles Computations. *J. Phys. Chem. C* **119**, 21649–21657 (2015).
111. Wang, P. *et al.* Vanadium Diboride ($VB_2$) Synthesized at High Pressure: Elastic, Mechanical, Electronic, and Magnetic Properties and Thermal Stability. *Inorg. Chem.* **57**, 1096–1105 (2018).
112. Li, L.-H., Wang, W.-L., Hu, L. & Wei, B.-B. First-principle calculations of structural, elastic and thermodynamic properties of Fe–B compounds. *Intermetallics* **46**, 211–221 (2014).
113. Niu, H. *et al.* Variable-composition structural optimization and experimental verification of $MnB_3$ and $MnB_4$. *Phys. Chem. Chem. Phys.* **16**, 15866–15873 (2014).
114. Xu, C. *et al.* A first-principles investigation of a new hard multi-layered $MnB_2$ structure. *RSC Adv.* **7**, 10559–10563 (2017).
115. Kvashnin, A. G., Allahyari, Z. & Oganov, A. R. Computational discovery of hard and superhard materials. *J. Appl. Phys.* **126**, 040901 (2019).
116. Andrievski, R. A. Superhard materials based on nanostructured high-melting point compounds: achievements and perspectives. *Int. J. Refract. Met. Hard Mater.* **19**, 447–452 (2001).
117. Solozhenko, V. L., Dub, S. N. & Novikov, N. V. Mechanical properties of cubic BC2N, a new superhard phase. *Diam. Relat. Mater.* **10**, 2228–2231 (2001).
118. Han, L. *et al.* Hardness, elastic, and electronic properties of chromium monoboride. *Appl. Phys. Lett.* **106**, 221902 (2015).
119. Zhong, M. M. *et al.* Phase stability, physical properties, and hardness of transition-metal diborides MB2 (M = Tc, W, Re, and Os): First-principles investigations. *J. Phys. Chem. C* **117**, 10643–10652 (2013).
120. Zhao, E., Meng, J., Ma, Y. & Wu, Z. Phase stability and mechanical properties of

tungsten borides from first principles calculations. *Phys. Chem. Chem. Phys.* **12**, 13158 (2010).

121. Itoh, H., Matsudaira, T., Naka, S., Hamamoto, H. & Obayashi, M. Formation process of tungsten borides by solid state reaction between tungsten and amorphous boron. *J. Mater. Sci.* **22**, 2811–2815 (1987).
122. Oganov, A. R. *Modern methods of crystal structure prediction*. (Wiley-VCH, 2010). doi:10.1002/9783527632831
123. Oganov, A. R., Saleh, G. & Kvashnin, A. G. *Computational Materials Discovery*. (Royal Society of Chemistry, 2018). doi:10.1039/9781788010122
124. Oganov, A. R., Pickard, C. J., Zhu, Q. & Needs, R. J. *Structure prediction drives materials discovery*. *Nature Reviews Materials* **4**, 331–348 (Nature Publishing Group, 2019).
125. Stone, D. S., Yoder, K. B. & Sproul, W. D. Hardness and elastic modulus of TiN based on continuous indentation technique and new correlation. *J. Vac. Sci. Technol. A Vacuum, Surfaces, Film.* **9**, 2543–2547 (1991).
126. Solozhenko, V. L., Kurakevych, O. O. & Le Godec, Y. Creation of Nanostuctures by Extreme Conditions: High-Pressure Synthesis of Ultrahard Nanocrystalline Cubic Boron Nitride. *Adv. Mater.* **24**, 1540–1544 (2012).
127. Dubrovinskaia, N., Sergey Dub, A. & Dubrovinsky, L. Superior Wear Resistance of Aggregated Diamond Nanorods. (2006). doi:10.1021/NL0602084
128. Drory, M. D., Dauskardt, R. H., Kant, A. & Ritchie, R. O. Fracture of synthetic diamond. *J. Appl. Phys.* **78**, 3083–3088 (1995).
129. Munro, R. G., Freiman, S. W. & Baker, T. L. Fracture Toughness Data for Brittle Materials | NIST. *NIST Interagency/Internal Rep. - 6153* (1998).
130. Wu, J. H. & Yang, G. Phase stability and physical properties of technetium borides: A first-principles study. *Comput. Mater. Sci.* **82**, 86–91 (2014).
131. Gou, Y., Fu, Z., Liang, Y., Zhong, Z. & Wang, S. Electronic structures and mechanical properties of iron borides from first principles. *Solid State Commun.* **187**, 28–32 (2014).
132. Okada, S., Atoda, T., Higashi, I. & Takahashi, Y. Preparation of single crystals of MoB2 by the aluminium-flux technique and some of their properties. *J. Mater. Sci.* **22**, 2993–2999 (1987).
133. Gou, H. *et al.* Peierls distortion, magnetism, and high hardness of manganese tetraboride. *Phys. Rev. B* **89**, 064108 (2014).
134. Ying, C., Liu, T., Lin, L., Zhao, E. & Hou, Q. New predicted ground state and high pressure phases of TcB3 and TcB4: First-principles. *Comput. Mater. Sci.* **144**, 154–160 (2018).
135. Harran, I., Wang, H., Chen, Y., Jia, M. & Wu, N. Exploring high-pressure FeB2: Structural and electronic properties predictions. *J. Alloys Compd.* **678**, 109–112 (2016).
136. Li, B., Sun, H. & Chen, C. First-principles calculation of the indentation strength of FeB4. *Phys. Rev. B* **90**, 014106 (2014).
137. Gou, H. *et al.* Discovery of a superhard iron tetraboride superconductor. *Phys. Rev. Lett.* **111**, 1–5 (2013).
138. Zhang, M. *et al.* Hardness of $FeB_4$: Density functional theory investigation. *J. Chem. Phys.* **140**, 174505 (2014).

139. He, C. & Zhong, J. X. Structures, stability, mechanical and electronic properties of α-boron and α*-boron. *AIP Adv.* **3**, 042138 (2013).
140. Veprek, S., Zhang, R. F. & Argon, A. S. Mechanical properties and hardness of boron and boron-rich solids. *J. Superhard Mater.* **33**, 409–420 (2011).
141. Čekada, M., Panjan, P., Maček, M. & Šmíd, P. Comparison of structural and chemical properties of Cr-based hard coatings. *Surf. Coatings Technol.* **151**–**152**, 31–35 (2002).
142. Kok, Y. N. & Hovsepian, P. E. Resistance of nanoscale multilayer C/Cr coatings against environmental attack. *Surf. Coatings Technol.* **201**, 3596–3605 (2006).
143. Cheng, F., Wang, Y. & Yang, T. Microstructure and wear properties of Fe–VC–Cr7C3 composite coating on surface of cast steel. *Mater. Charact.* **59**, 488–492 (2008).
144. Xie, J. Y., Chen, N. X., Teng, L. D. & Seetharaman, S. Atomistic study on the site preference and thermodynamic properties for Cr23−xFexC6. *Acta Mater.* **53**, 5305–5312 (2005).
145. Inoue, A. & Masumoto, T. Formation of nonequilibrium Cr3C carbide in Cr-C binary alloys quenched rapidly from the melt. *Scr. Metall.* **13**, 711–715 (1979).
146. Liu, B. X. & Cheng, X. Y. A metastable Cr carbide of NaCl structure formed by carbon-ion implantation into chromium films. *J. Phys. Condens. Matter* **4**, L265–L268 (1992).
147. Uebing, C., Scheuch, V., Kiskinova, M. & Bonzel, H. P. Segregation of ordered CrN and CrC surface phases on a Fe-15%Cr(100) crystal. *Surf. Sci.* **321**, 89–99 (1994).
148. Niu, H. *et al.* Structure, bonding, and possible superhardness of CrB4. *Phys. Rev. B* **85**, 144116 (2012).
149. Knappschneider, A. *et al.* Possible Superhardness of CrB4. *Inorg. Chem.* **52**, 540–542 (2013).
150. Post, B., Glaser, F. W. & Moskowitz, D. Transition metal diborides. *Acta Metall.* **2**, 20–25 (1954).
151. Zhang, R. F. *et al.* Crystal Field Splitting is Limiting the Stability and Strength of Ultra-incompressible Orthorhombic Transition Metal Tetraborides. *Sci. Rep.* **6**, 23088 (2016).
152. Loubière, S., Laurent, C., Bonino, J. P. & Rousset, A. Elaboration, microstructure and reactivity of Cr3C2 powders of different morphology. *Mater. Res. Bull.* **30**, 1535–1546 (1995).
153. Wang, S. *et al.* Crystal structures, elastic properties, and hardness of high-pressure synthesized CrB2 and CrB4. *J. Superhard Mater.* **36**, 279–287 (2014).