TEL AVIV UNIVERSITY
The Iby and Aladar Fleischman Faculty of Engineering
The Zandman-Slaner School of Graduate Studies

SIMULATING UNKNOWN & DISORDERED MATERIALS

A thesis submitted toward the degree of
Master of Science in Engineering

by

Nadav Moav

This research was carried out at Tel Aviv University
in the Department of Materials Science and Engineering
Faculty of Engineering
under the supervision of Dr. Lee A. Burton

September 2025

Abstract

This study addresses the computational challenges of modeling disordered materials, such as alloys, by employing the Special Quasirandom Structures (SQS) approach to represent disorder within periodically finite cells. Focusing on the Cu–Mn system, SQS and random substitution methods were used to model maximally disordered configurations across various compositions. Subsequent density functional theory (DFT) calculations revealed only positive formation enthalpies for all configurations, suggesting that the Cu–Mn system is unstable as a homogeneous solid solution, thereby challenging phase diagrams reported in the literature. *Ab initio* exploration using evolutionary algorithms for structure prediction across all compositions further investigated the energy landscape but continued to support a thermodynamically driven decomposition, as no stable phases were identified. Beyond Cu–Mn, these approaches were extended to new materials discovery, generating over one million initial candidates and down-selecting three previously unreported binary semiconductors for potential experimental validation.

Table of Contents

Nomenclature

AFLOW Automatic Flow for Materials Discovery

ATAT Alloy Theoretic Automated Toolkit

BOA Born–Oppenheimer Approximation

DFT Density Functional Theory

EDS Energy Dispersive Spectroscopy

GGA Generalized Gradient Approximation

HPC High Performance Computing

ICSD Inorganic Crystal Structure Database

LDA Local Density Approximation

PAW Projector Augmented Wave

PBE Perdew–Burke–Ernzerhof functional

PyMatGen Python Materials Genomics

SEM Scanning Electron Microscopy

SQS Special Quasirandom Structures

USPEX Universal Structure Predictor: Evolutionary Xtallography

VASP Vienna Ab Initio Simulation Package

XC Exchange-Correlation

List of Figures

List of Tables

1 Introduction

> *One of the continuing scandals in the physical sciences is that it remains in general impossible to predict the structure of even the simplest crystalline solids from a knowledge of their chemical composition.*
>
> —John Maddox, *Nature (1988)*

1.1 Material Innovations and Historical Impact

The impact of newly discovered materials is so profound that entire periods of history are defined by them, such as the Stone Age, Bronze Age, and Iron Age, as illustrated in Figure 1.1. The Bronze Age was marked by the use of bronze, an alloy of copper and tin, which was harder and more durable than tin or copper individually; enabling superior tools, weapons, and craftsmanship. Producing bronze required advanced fabrication techniques and access to both copper and the much rarer tin, fostering long distance trade networks and granting significant power to those who controlled these resources, often sparking conflicts that reshaped political boundaries. The Iron Age was driven by the scarcity of tin and the mastery of high temperature methods for extracting and working iron, which, though initially inferior to bronze, was far more abundant and widely distributed. Iron's availability democratized access to tools and weapons, fueling agricultural advances, enabling larger armies, and transforming economies. Over time, the refinement of iron into steel defined the weapons and armor of the middle ages, with societies possessing the best steel holding significant advantage over others. Even today, this pattern continues, with some referring to the present era as the 'Silicon Age' [1], highlighting the dominance of the silicon transistor in electronics and its central role in driving technological innovation. As in earlier periods, control over such a critical resource now

shapes global power structures, influencing diplomatic relations, industrial policy, and security strategies. Competition for leadership in semiconductor manufacturing and technology has led to trade disputes, technology embargoes, and targeted tariffs between global superpowers, while the concentration of advanced chip production expertise in regions such as Taiwan has made them focal points of geopolitical tension, much like tin-rich areas in the Bronze Age.

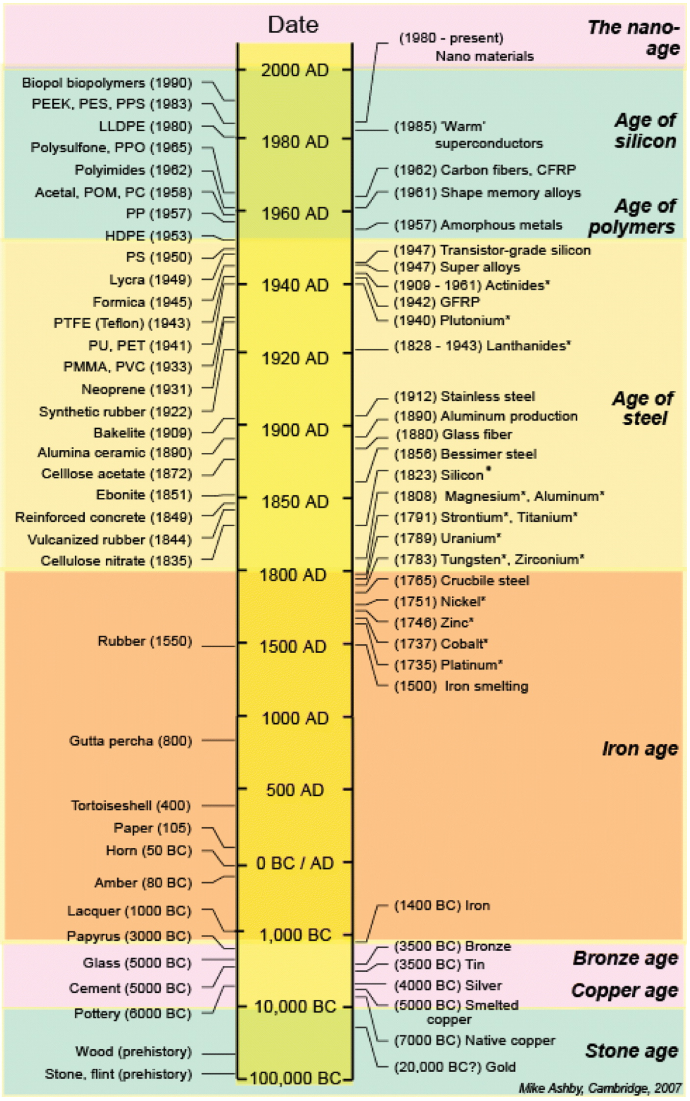

Figure 1.1: Timeline showing major material discoveries and the historical periods with which they are associated , as defined by the primary materials used during those times [2].

1.2 Material Limitations in Modern Technology

Today, many technologies are constrained by the materials properties on which they rely, even when scientists and engineers can identify the ideal attributes for a given application. For example, solar energy is highly desirable as a reliable, clean and abundant energy resource, which can be generated entirely by the judicious combination of solid materials in a solar cell. The performance of solar cells in converting light to electricity is entirely determined by the physical properties of their constituent materials. Decades of research have made it clear what an ideal photovoltaic material should look like in terms of strong optical absorption, thermal stability, good electrical conductivity *etc.* The problem is that in reality, no known material perfectly combines all of these traits. Furthermore, devices require multiple materials to hold multiple ideal properties within the same device at high temperatures without any inter-mixing or long term degradation. Knowing all of these possible attributes and predicting each of these complex relationships at once and ahead of time is impracticle to say the least, leading many laboratories and manufacturers to seek development through trial-and-error approaches or incremental improvements on established methods rather than being theory led. This is known as empirical research, but still requires a deep understanding of the methods and data that have been attempted in the past, in order to effectively target new ways of trialling improvements.

When it comes to materials discovery, a similar historical pattern of development is observed, beginning with an empirical research foundation and gradually incorporating well established theory to guide the work.

1.3 Empirical and Theory-Led Materials Discovery

Historically, the discovery of new materials and elements often occurred by chance, with all conditions aligning perfectly. A notable example is the story of German alchemist Hennig Brand who in 1669 was searching to turn base metals into gold. Brand collected large quantities of human urine, on the basis that the color of urine and gold were similar, allowed it to ferment, boiled it down, and distilled the residue. During this peculiar experiment, he observed a white, waxy substance that glowed in the dark and spontaneously ignited upon exposure to air. Brand had accidentally discovered elemental phosphorus, which is a vital plant nutrient widely used in fertilizers to enhance crop yields [3, 4]. Another well known example is Harry Brearley's

accidental discovery of stainless steel in 1913 during wartime research. While testing iron alloys to create new materials for gun barrels, he discarded a sample that initially failed to meet his expectations. Later, he noticed that, unlike other discarded alloys, this particular sample did not rust. Stainless steel subsequently became a crucial material in modern engineering and manufacturing [5].

Relying on accidental breakthroughs is inherently unreliable, underscoring the need for a well thought out approach, such as through recognition of underlying theoretical patterns. This discussion is often framed as empirical versus theory-led research, where empirical approaches depend on trial and error experimentation, while theory-led methods apply established principles to guide discovery in a more targeted and efficient manner. A known case of this is Dmitri Mendeleev's work in the 19th century. By identifying systematic patterns in elemental properties, later understood as periodic trends governed by atomic number, Mendeleev constructed the periodic table and predicted the existence of elements that had not yet been discovered. He accurately anticipated the properties of germanium, gallium, and scandium, all experimentally confirmed years later [6]. Using theoretical patterns for future predictions relies on access to sufficient, reliable and clean data. Large scale datasets enhance the accuracy of these predictions, allowing researchers to identify patterns, validate models, and make informed decisions in the search for new materials.

1.4 Role of Repositories in Materials Science

The dependence on data driven predictions points to the importance of organizing and preserving scientific information. The need for a standardized method to store and maintain research findings has always been a fundamental part of academic work. Without proper documentation, future researchers cannot follow or build upon previous efforts, requiring the process to start from the beginning. Centralized and accessible repositories of scientific knowledge are therefore essential. Historically, printed volumes have played a major role in compiling critical information. For example, the International Tables for Crystallography, Volume A [7], compile the symmetries of all 230 space groups, ensuring that the intensive work required to classify them is not lost and remains accessible to all.

In recent decades, the digital revolution has transformed how scientific data is stored, accessed, and used. In crystallography, this transformation can be represnted

by the Bilbao Crystallographic Server [8], a widely used online platform that provides the same fundamental symmetry data found in the International Tables, but in a digital, searchable, and interactive format. It also includes powerful tools for visualizing and analyzing symmetry operations, representations, and related crystallographic properties. Additional example is the Inorganic Crystal Structure Database (ICSD) [9, 10, 11], which serves as a digital repository of experimentally determined inorganic crystal structures, enabling rapid access and analysis by the global research community. The database contains over 200,000 entries, each representing a distinct crystallographic structure along with metadata such as chemical composition, oxidation state, space group, unit cell parameters, and bibliographic references. These entries are standardized and searchable, allowing researchers to perform targeted queries based on element combinations, crystal structure types, symmetry, or other criteria.

The digitization of scientific knowledge and the creation of accessible databases have significantly changed the landscape of modern research. By organizing information in standardized, computer readable formats, these resources make it possible to store, share, and analyze large volumes of data efficiently. This, in turn, enables researchers to detect patterns, generate insights, and form new hypotheses that would be difficult or even impossible to uncover manually. As a result, digital databases play a key role in advancing discovery across scientific fields. However, to truly accelerate progress, there is a growing need for tools specifically designed to query and interpret this information effectively. In particular, ensuring compatibility between such databases and computational simulation tools.

1.5 Simulation Driven Databases

Early simulation tools primarily calculated and analyzed the properties of materials that had already been synthesized and characterized, focusing on validating theoretical models and interpreting experimental data rather than guiding new discoveries [12]. Today, a range of computational methods are employed in materials research, most notably density functional theory for electronic structure calculations and molecular dynamics for modeling atomic motion. Widely used software packages such as Quantum Espresso and the Vienna Ab initio Simulation Package (VASP) implement first principles approaches to approximate the solution for the Schrödinger equation describing electron behavior in solids [13]. These tools enable the investigation of materials at the atomic scale without reliance on prior exper-

imental data. Advancements in simulation tools and growing confidence in their accuracy, combined with the rise of high performance computing (HPC), have significantly expanded the role of computers in materials science. No longer limited to post-experimental analysis, computational methods are now widely used to predict material properties before synthesis. This shift has led to the development of toolkits that help process and analyze large volumes of simulation data. One such tool is the Python Materials Genomics (PyMatGen) [14], an open source Python library that facilitates the generation, analysis, and manipulation of materials data. Using such tools, the Materials Project database was created for computing and distributing first-principles properties of inorganic materials [15], with its growth over time, both in the number of entries and in the range of computational methods used for simulations, shown in Figure 1.2. The Automatic Flow for Materials Discovery (AFLOW) is another notable resource with a similar aim, providing standardized data for thousands of compounds [16]. Unlike the ICSD, which is based solely on experimentally determined structures and measured properties, the Materials Project is built entirely on simulation results. This enables the inclusion of a broader set of material properties, including those that are difficult or uncommon to measure experimentally. These databases support data driven research by providing programmatic access, analysis tools, and an expanding repository of computed materials data [17]. Integrating simulation methods into accessible databases, alongside coding tools built on widely used programming languages, has opened vast possibilities for high throughput strategies that accelerate the discovery of promising new material across a broad range of possibilities.

1.6 Modeling Disordered Materials and Predicting Unknown Structures

While repositories like the Materials Project and AFLOW provide extensive data for materials research, they are largely limited to well ordered materials, where each atom occupies a fixed, periodic position within the crystal structure. making it well suited for simulating ideal, stoichiometric compounds, but less applicable for disordered systems such as alloys and solid solutions, where atomic positions are only partially defined and elements are randomly distributed over specific lattice sites. As a result, the direct use of repositories such as the Materials Project for studying disordered materials is limited. However, the database still plays an important role by providing a vast collection of well characterized crystal structures that can be used as starting points, or structural frameworks, for modeling disorder. A structural frame-

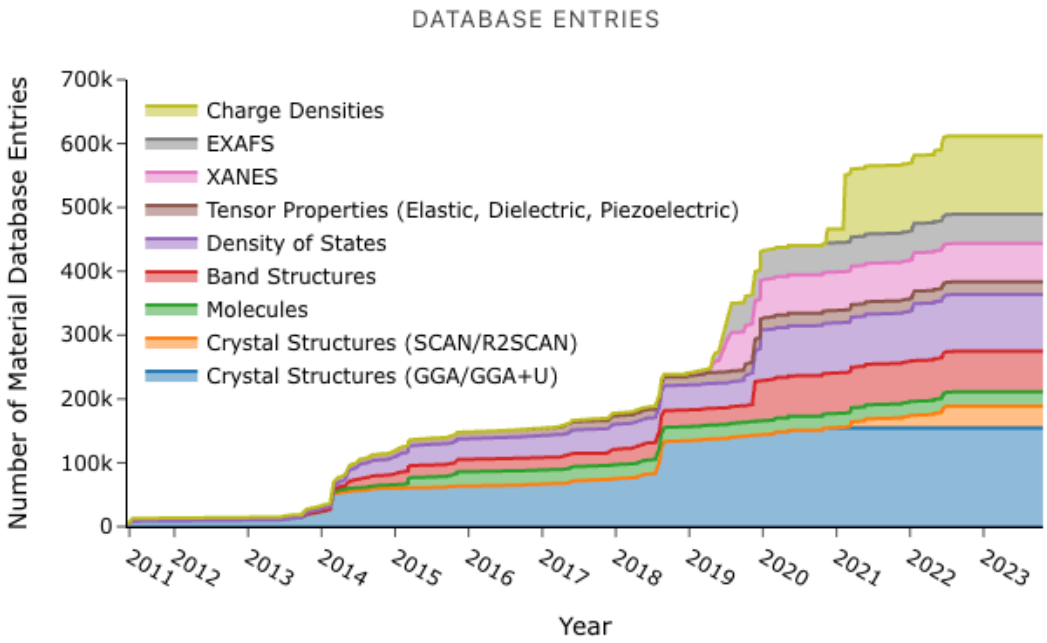

Figure 1.2: Growth of the Materials Project database from 2011 to 2023, showing the increasing number and diversity of material entries generated through different computational methods.

work refers to the lattice geometry of a known compound, which can be adapted to introduce chemical randomness by probabilistically substituting different atoms across equivalent sites. This approach enables the construction of representative disordered models while retaining the essential symmetry and bonding environment of the original structure. Techniques such as the Special Quasirandom Structure method rely on these frameworks to simulate disordered systems within the constraints of first principles calculations. In this way, the Materials Project continues to serve as a foundational resource, even in research focused on materials that fall outside its direct scope. A more detailed explanation of this approach is provided in Section 2.

Modeling disordered materials using structural frameworks depends on prior knowledge of the crystal structure of the alloy in question. However, when no such structural information exists, a fundamental challenge emerges in determining how to begin from scratch to design and simulate materials that are entirely unknown. Overcoming this challenge requires new tools capable of predicting crystal structures from only a chemical formula, incorporating physical principles and computational criteria

to guide the process. These approaches often employ evolutionary algorithms, which iteratively generate, evaluate, and refine candidate structures by mimicking natural selection, gradually converging on energetically favorable configurations. This enables the exploration of a vast space of atomic arrangements, the identification of stable or metastable phases, and the design and study of their properties, laying the groundwork for high throughput computational exploration of materials before they are ever realized experimentally.

1.7 High Throughput Materials Discovery

By leveraging HPC technology capable of handling high throughput calculations, along with structure prediction software, the vast combinatorial landscape of all potentially possible chemical formulas can be explored computationally before proceeding to experimental validation in the lab. When applied correctly and grounded in established chemical and solid state physics principles, computational approaches can be used to determine the stability and properties of new materials based solely on a chemical formula. This methodology lays the foundation for a novel strategies in materials discovery, For example, by using established structural formula types from various material families, such as perovskite (ABX_3), spinel (AB_2X_4), and rocksalt (AX) *etc.* Using these types of formula as templates, researchers can systematically explore chemical substitutions and compositional variations to identify new, stable compounds within these families. This shifts the challenge of discovering new materials into a data driven problem, where a well designed filtration process can be used to determine the most promising candidates for stability. Furthermore, rather than evaluating materials one by one HPC enables the simultaneous analysis of multiple candidates, significantly improving efficiency and accelerating the material discovery process.

1.8 Summary

Combining all of the tools at hand opens the opportunity to take computational materials science from a mostly interpretative tool to a predictive one. As computational power continues to grow, these tools will play an even more critical role in shaping the future of all fields of science, enabling discoveries that were once thought impossible. One of the most striking recent achievements in this regard is the development of AlphaFold, an AI-powered tool that accurately predicts protein folding, a

problem that had remained unsolved for decades. In fact, Cyrus Levinthal famously claimed in 1969 that predicting protein folding would be mathematically impossible [18]. The importance of AlphaFold's breakthrough was recognized with the 2024 Nobel Prize in Chemistry [19], underscoring the growing credibility and influence of computational science tools within the scientific community.

2 Methods

"If you are not completely confused by quantum mechanics, you do not understand it."

—John Archibald Wheeler

In materials science, calculations are generally classified based on whether they use classical or quantum mechanics. Quantum mechanical methods, known as *ab initio* (Latin for "from the start"), are among the most computationally demanding. Unlike classical approaches, they do not rely on empirical parameters or experimental data but instead derive material properties and interactions directly from quantum mechanical principles. This chapter demonstrates the theoretical foundations and computational tools behind *ab initio* calculations, explaining key concepts and introducing the specific software used for performing calculations and simulations for this project.

2.1 Ab initio Calculations

Ab initio calculation methods generally are built on the assumption that electrons determine macroscopic properties in the solid state. These electrons have a ground state at which, for a given configuration of atomic positions, they have the lowest internal energy. This ground state configuration will be overwhelmingly prevalent for specific structures and determine a material's intrinsic properties. In *ab initio* codes the ground state is reached through iteration and then, when the ground state is identified, it is considered a "converged" calculation. A calculation is considered physically meaningful only when it has converged.

The steps to achieve convergence are typically consistent across *ab initio* software, and are illustrated in Figure 2.1. The following sections will provide a detailed explanation of the theoretical background behind each step.

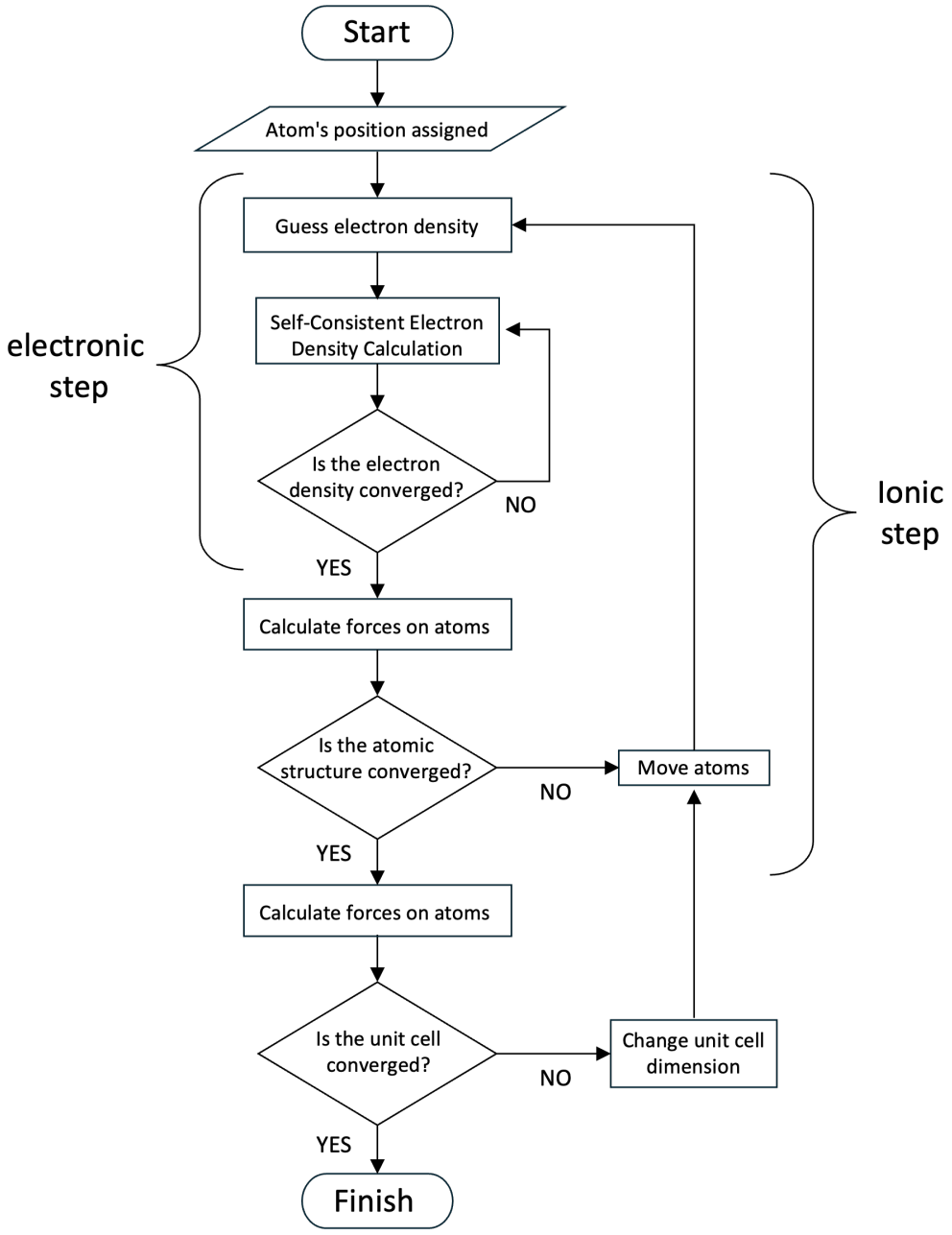

Figure 2.1: Representation of the *ab initio* computational workflow as implemented in software packages.

2.1.1 Electronic Convergence

As shown in Figure 2.1, the first step in an *ab initio* calculation is to define the atomic positions, which determine the number of electrons and protons and thus the nuclear potential. The software then generates an initial guess for the electron density and evaluates the corresponding energy. This is followed by an iterative

self consistent field cycle, where the electron density and total energy are updated until they converge within predefined thresholds. In *ab initio* approaches, nearly all methods begin with an approximate solution to the Schrödinger equation.

2.1.1.1 The Schrödinger Equation

A major breakthrough in quantum mechanics was the realization that particles, such as electrons, exhibit wave like behavior. This led to the concept of the wavefunction, which provides a probabilistic description of their state. In *ab initio* calculations, these wavefunctions are obtained by solving the Schrödinger equation. However, the equation in its full form is too complex to solve for systems with many electrons. As the number of electrons increases, the number of interactions that must be considered grows exponentially, making an analytical solution impossible. This challenge is known as the many body problem [20]. As a result, the Schrödinger equation can only be solved analytically for relatively simple systems, such as a single hydrogen atom or hydrogen like ions with only one electron, including He^{+} and Li^{2+} [21, 22]. To address this challenge, certain assumptions are introduced to obtain approximate solutions to the Schrödinger equation. These assumptions simplify the calculation as much as possible while preserving the essential physics of the system. The following sections will describe the methods and assumptions used to achieve this.

The main objective of the calculations is to determine the ground state of the system,a stationary state that does not change with time and corresponds to the lowest possible energy configuration. Relying on the fact that there is no time dependence, the first approximation is to work with the time independent form of the Schrödinger equation [23], which is written as:

$$\hat{H}\Psi(r_i, r_I) = E\Psi(r_i, r_I) \tag{2.1}$$

$\hat{H}$ is the Hamiltonian operator which describes the total energy of a given system in quantum mechanics. It consists the potential (U) and kinetic (T) energies of all particles in a given system. These energy components include the kinetic energy of the electrons (T_i) and the nuclei (T_I), as well as the potential energy arising from their attraction interactions with each other (U_{iI}) and repulsive interaction between themselves (U_{ij}, U_{IJ}):

$$\hat{H} = T_I + T_i + U_{ij} + U_{IJ} + U_{iI} \tag{2.2}$$

Ψ represents the eigenstates of the Hamiltonian, commonly referred to as the wavefunction. The wavefunction itself does not have a direct physical interpretation. Instead, the square of its magnitude at a given point in space is proportional to the probability of finding a particle at that location. The variables r_i and r_I are the coordinates of the electron and nucleus respectively. E is the eigenvalue of the Hamiltonian and describes the energy of each state in the system.

2.1.1.2 The Born–Oppenheimer Approximation

The Born–Oppenheimer Approximation (BOA) is that nuclei can be considered stationary in space relative to the electrons, with fixed coordinates. The BOA was first introduced by Max Born and J. Robert Oppenheimer in their 1927 paper, *On the Quantum Theory of Molecules* [24], and is now known by their names. The assumption is based on the fact that protons in the nucleus are significantly more massive than the surrounding electrons - approximately 1,860 times heavier [25]. This implies that even the slightest movement of the nucleus will result in an almost immediate response from the electrons. As a result, the system's total wavefunction can be separated into an electronic wavefunction and a nuclear wavefunction. The nuclear wavefunction does influence the electronic wavefunction but remains unchanged during the electronic solution and can be expressed as:

$$\Psi(r_i, r_I) = \Psi_e(r_i) \cdot \Psi_n(r_I) \tag{2.3}$$

If the nucleus is considered stationary, its kinetic energy T_I can be approximated to zero. Furthermore, the nucleus-nucleus (n–n) interaction U_{IJ} becomes a classical Coulomb interaction that can be solved analytically for a given atomic configuration. As a result, the system can be solved with respect to the fixed U_{IJ}. Therefore, the Hamiltonian shown in equation 2.2 can be simplified to the electronic Hamiltonian:

$$\hat{H}_e = T_i + U_{ij} + U_{iI} \tag{2.4}$$

The full expression of Equation 2.4 can be written in Hartree atomic units. In these units, fundamental constants such as the Coulomb constant, the reduced Planck constant, the electron mass, etc. are set to 1, which simplifies quantum mechanical equations by removing these physical constants, resulting in:

$$\hat{H}_e = -\frac{1}{2}\sum_{i=1}^{N_e}\nabla_i^2 + \sum_{i=1}^{N_e}\sum_{j>1}^{N_e}\frac{1}{r_{ij}} - \sum_{i=1}^{N_e}\sum_{I=1}^{N_n}\frac{Z_I}{r_{iI}} \tag{2.5}$$

where ∇^2 is the Laplacian operator, representing the second derivative of the wavefunction with respect to spatial coordinates. N_e and N_n are the numbers of electrons and nuclei, respectively, while Z_I is the charge of nucleus I. The term r_{ij} is a scalar representing the distance between electrons i and j, and r_{iI} is a scalar representing the distance between electron i and nucleus I. Repulsive interactions are represented with a positive sign, while attractive interactions are represented with a negative sign. With this reduced electronic Hamiltonian, the electronic Schrödinger equation takes the form:

$$\hat{H}_e\Psi_e(r_i, r_I) = E_e\Psi_e(r_i, r_I) \tag{2.6}$$

When examining Equation 2.5, even after applying the BOA to simplify the problem, it does not effectively address the many body problem. The presence of multiple electron–electron (e–e) and electron–nucleus (e–n) interactions remains a significant challenge, requiring further approximations to make the problem solvable.

2.1.1.3 The Single Electron Wavefunction

A different approach is required to account forthe vast number of e–e interactions to tackle the many body problem. In 1928, Douglas Rayner Hartree proposed treating electrons as non interacting particles. By doing so, the full electronic wavefunction can be expressed as a product of N_e single electron wavefunctions, known as the Hartree product [26]:

$$\Psi(x_1, x_2, x_3, ..., x_N) = \psi_1(x_1)\cdot\psi_2(x_2)\cdot\psi_3(x_3)...\cdot\psi_N(x_N) \tag{2.7}$$

Where x is a combined expression of both position and spin; s describes the spin of each electron, and r represents the position.

The quantum properties of electrons must be properly represented in the electronic and single electron wavefunction. Electrons are fermions and must obey the Pauli exclusion principle, which states that no two electrons can occupy the same quantum state with the same spin [27]. This principle plays a fundamental role in

shaping the behavior of multi electron systems. To ensure agreement with the Pauli exclusion principle, the wavefunction must be anti symmetric with respect to the exchange of particles. This means that if two electrons swap their spatial or spin coordinates, the wavefunction must reverse its sign. For a system with only two electrons, this condition is expressed as:

$$\Psi(r_1, s_1; r_2, s_2) = -\Psi(r_2, s_2; r_1, s_1) \tag{2.8}$$

or since x accounts for both s and r:

$$\Psi(x_1, x_2) = -\Psi(x_2, x_1) \tag{2.9}$$

John Clarke Slater introduced a different approach to representing the electronic wavefunction using single-electron wavefunctions in 1929. He incorporated exchange interactions by leveraging the intrinsic antisymmetry of the determinant, naturally enforcing the Pauli exclusion principle. A determinant is a mathematical operation applied to a square matrix that produces a single value and notably changes sign when two rows are swapped. This formulation became known as the Slater determinant [28], with the following form:

$$\Psi_{N_e} = \frac{1}{\sqrt{N_e!}} \begin{vmatrix} \psi_1(x_1) & \psi_2(x_1) & \cdots & \psi_{N_e}(x_1) \\ \psi_1(x_2) & \psi_2(x_2) & \cdots & \psi_{N_e}(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_1(x_{N_e}) & \psi_2(x_{N_e}) & \cdots & \psi_{N_e}(x_{N_e}) \end{vmatrix} \tag{2.10}$$

In the Slater determinant, each row represents an electron, and each column corresponds to a different state (spatial and spin). Swapping the coordinates of two electrons mathematically corresponds to swapping two rows in the Slater determinant. This structure ensures that each electron is considered in all possible states rather than being assigned a single fixed state, naturally accounting for superposition. The property of superposition can be observed when reducing the general Slater determinant (equation 2.10) to a single determinant that includes only two electrons for simplicity:

$$\Psi(x_1, x_2) = \frac{1}{\sqrt{2}} \begin{vmatrix} \psi_1(x_1) & \psi_2(x_1) \\ \psi_1(x_2) & \psi_2(x_2) \end{vmatrix} = \frac{1}{\sqrt{2}} \left(\psi_1(x_1)\psi_2(x_2) - \psi_1(x_2)\psi_2(x_1)\right) \tag{2.11}$$

Swapping x_1 and x_2 corresponds to exchanging the two rows in the determinant:

$$\Psi(x_2, x_1) = \frac{1}{\sqrt{2}} \begin{vmatrix} \psi_1(x_2) & \psi_2(x_2) \\ \psi_1(x_1) & \psi_2(x_1) \end{vmatrix} = \frac{1}{\sqrt{2}} \left(\psi_1(x_2)\psi_2(x_1) - \psi_1(x_1)\psi_2(x_2)\right) \tag{2.12}$$

Since swapping two rows in a determinant introduces a negative sign, the result is as shown in equation 2.9

Confirming that the wavefunction is antisymmetric under particle exchange. Such antisymmetry directly enforces the Pauli exclusion principle, ensuring that no two electrons can occupy the same quantum state, meaning they cannot share both the same spatial orbital and the same spin. A direct consequence of this antisymmetry is revealed when two electrons attempt to occupy the same quantum state. If both electrons are in the same state, then two columns in the determinant become identical. For example, if both electrons are in state ψ_1, the determinant takes the form:

$$\Psi(x_1, x_2) = \frac{1}{\sqrt{2}} \begin{vmatrix} \psi_1(x_1) & \psi_1(x_1) \\ \psi_1(x_2) & \psi_1(x_2) \end{vmatrix} \tag{2.13}$$

Since the determinant of a matrix with two identical columns is always zero, this results in:

$$\Psi(x_1, x_2) = 0 \tag{2.14}$$

This result shows that the antisymmetric Slater determinant inherently enforces the Pauli exclusion principle by ensuring the wavefunction vanishes (does not exist) when two electrons occupy the same state. The limitation of this representation is that it does not account for e–e correlation. However, by initially considering non interacting electrons, the single electron wavefunctions can be used in approaches that incorporate both exchange and correlation effects.

2.1.1.4 Density Functional Theory (DFT)

As the name suggests, Density Functional Theory (DFT) is fundamentally based on the concept of electron density. To discuss the Hohenberg-Kohn theorems, it is first necessary to define the electron density. This function is obtained by summing the probability densities of all single electron wavefunctions, yielding a function that describes the probability of finding an electron at a given point in space. It is expressed as:

$$n(\mathbf{r}) = \sum_{i=1}^{N_e} |\psi_i(\mathbf{r})|^2 \tag{2.15}$$

where N_e is the number of electrons in the system, $n(\mathbf{r})$ represents the electron density function, and $|\psi_i(\mathbf{r})|^2$ denotes the probability density associated with the i-th single electron wavefunction.

The primary motivation for working with the electron density function is that, instead of dealing with $3N_e$ spatial coordinates for $\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_{N_e}$, the problem is reduced to only three coordinates. This reformulation addresses the many body problem by describing the interactions between electrons and the electron density cloud, rather than tracking a vast number of individual electron positions.

2.1.1.5 The Hohenberg–Kohn Theorems

In 1964, Walter Kohn and Pierre Hohenberg further demonstrated the usefulness of the electron density by mathematically proving two theorems, published in their paper "Inhomogeneous Electron Gas" [29]. These fundamental results are encapsulated in the two Hohenberg–Kohn theorems:

- Given a system of interacting particles in an external potential, its ground state density $n_0(\mathbf{r})$ uniquely determines the external potential $v(\mathbf{r})$, up to a constant.

- For any external potential $v(\mathbf{r})$, a universal energy functional $E[n]$ can be defined in terms of the density $n(\mathbf{r})$. For a given $v(\mathbf{r})$, the exact ground state energy of the system corresponds to the global minimum of this functional. The density that minimizes the functional is the ground state density $n_0(\mathbf{r})$.

In other words, the first theorem states that the ground state energy E, as derived from Schrödinger's equation, is a unique functional of the electron density $n(\mathbf{r})$. A functional is simply a function whose input is another function, in this case the electron density.

$$E[n(\mathbf{r})] = E \tag{2.16}$$

The second theorem asserts that the electron density that minimizes the energy functional is the exact ground state electron density, which corresponds to the full solution of Schrödinger's equation. To take advantage of these theorems, an appropriate method for their practical implementation must be developed.

2.1.1.6 The Kohn–Sham Method

In 1965, Kohn and Sham approached this problem using the Hohenberg–Kohn theorems and the approximation of non interacting electrons. By considering non interacting electrons, they made it possible to work with single electron wavefunctions rather than the complex many body wavefunction. They formulated an approximation of the total energy functional of the electron density as the sum of a known functional, $E_{\text{known}}[n]$, and the exchange correlation (XC) functional, $E_{\text{XC}}[n]$, as follows [30]:

$$E_{\text{tot}} = E_{\text{known}}[n] + E_{\text{XC}}[n] \tag{2.17}$$

The known functional, $E_{\text{known}}[n]$, consists of the kinetic energy functional of non interacting electrons, $T_s[n]$, the Hartree potential functional, $E_H[n]$, which accounts for the coulomb interactions between electron densities, and the coulomb interactions between nuclei and the electron density. These components are expressed as:

$$E_{\text{known}}[n] = \underbrace{-\frac{1}{2}\sum_{i}^{n}\int \psi_i^*(r)\nabla^2\psi_i(r)\,dr}_{T_i[n]} + \underbrace{\frac{1}{2}\int\int \frac{n(r)n(r')}{|r-r'|}\,dr\,dr'}_{E_H[n]} + \underbrace{\int U_{iI}(r)n(r)\,dr}_{\text{electron–nuclei interaction}} \tag{2.18}$$

The XC functional, which accounts for all unknown interaction terms, including the interacting component of the kinetic energy. Since the kinetic energy term $T_i[n]$

(Equation 2.18) is derived from the non interacting single electron wavefunctions, referred to as Kohn–Sham orbitals.

To obtain the wavefunctions that minimize the total energy functional (equation 2.17), the functional is differentiated with respect to the electron density, leading to the Kohn–Sham potential:

$$v_{\mathrm{KS}}(\mathbf{r}) = v_H(\mathbf{r}) + v_n(\mathbf{r}) + v_{\mathrm{XC}}(\mathbf{r}) \tag{2.19}$$

which results in the Kohn-Sham equation:

$$\hat{h}_{KS}\psi_i = \left(-\frac{1}{2}\nabla_i^2 - v_{\mathrm{KS}}(\mathbf{r})\right)\psi_i(\mathbf{r}) = \varepsilon_i\psi_i(\mathbf{r}) \tag{2.20}$$

Here, v_{KS} is the Kohn–Sham potential, and ε_i represents the energy corresponding to the Kohn–Sham orbital. Expanding this expression explicitly:

$$\left(-\frac{1}{2}\nabla^2 + \int \frac{n(\mathbf{r})}{|\mathbf{r}-\mathbf{r}'|}\,d^3r + v_n(\mathbf{r}) + \frac{\delta E_{\mathrm{XC}}[n]}{\delta n(\mathbf{r})}\right)\psi_i = \epsilon_i\psi_i. \tag{2.21}$$

With all these elements in place, the final step is to solve the equations iteratively until convergence is achieved, resulting in a self consistent electron density corresponding to the system's ground state. This procedure follows these steps:

1. Define an initial trial electron density, $n(\mathbf{r})$.

2. Solve the Kohn–Sham equation (equation 2.20) using the trial electron density to determine the single electron Kohn-Sham orbitals, $\psi_i(\mathbf{r})$.

3. Compute the electron density from the obtained Kohn-Sham orbitals, $n_{\mathrm{KS}}(\mathbf{r})$.

4. Compare the calculated electron density, $n_{\mathrm{KS}}(\mathbf{r})$, with the initial trial density, $n(\mathbf{r})$. If the difference is within a specified tolerance, the process converges, yielding the ground state electron density. Otherwise, update the trial electron density and repeat from step 2.

Despite this well defined procedure, the XC potential v_{XC} remains unknown and must be approximated to effectively carry out the iterative process and predict the accurate energy of a given electronic system.

2.1.1.7 Exchange Correlation Functionals

The challenge of finding a suitable approximation for the exchange correlation functional E_{XC} serves as a bottleneck in fully leveraging the capabilities of the DFT method. Although it contributes only a small fraction of the total energy, it plays a crucial role in determining material properties [31].

One of the first and simplest approaches to approximating E_{XC} is the Local Density Approximation (LDA). This method takes advantage of the fact that E_{XC} can be calculated exactly for a homogenies electron gas system [32, 33]. To use this knowledge the LDA, as its name suggests, computes the exchange correlation functional locally at each spatial coordinate $\mathbf{r}$. That is, the XC energy is determined at every point in space as if the electron density at that location belonged to an imaginary homogenies electron gas system [34, 35, 36]. E_{XC} is calculated using LDA as:

$$E_{XC}^{LDA}[n] = \int n(\mathbf{r})\epsilon_{XC}^{hom}(\mathbf{r})dr \tag{2.22}$$

LDA has been successful in predicting structures and electronic properties with reasonable accuracy [37]. However, it encounters limitations in systems with significant electron density fluctuations or strong electronic correlations, such as those involving d or f electrons [38]. In these cases, LDA may fail to accurately capture the properties of strongly correlated systems. Moreover, is known to overestimate binding energies between atoms, leading to inaccuracies in predicting ground state structures and associated properties [39].

LDA has limitations that require improvements to enhance calculation reliability. A key advancement is incorporating semilocal density information. As the name suggests, the Generalized Gradient Approximation (GGA) extends LDA by considering not only the local electron density but also its gradient, leading to a more accurate XC functional.

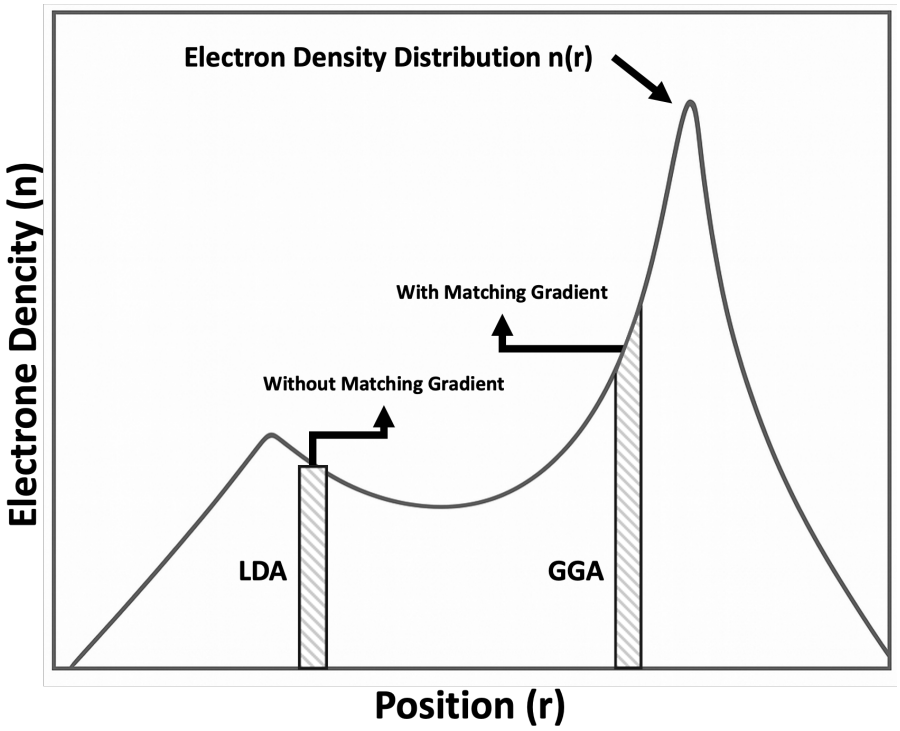

Figure 2.2: Representation of LDA and GGA approaches to electron density, highlighting the improved accuracy of GGA through the incorporation of density gradients. Adapted from [40].

$$E_{XC}^{GGA} = \int n(\mathbf{r})\epsilon_{XC}^{GGA}(n(\mathbf{r}), \nabla n(r))dr \tag{2.23}$$

The schematic in Figure 2.2 illustrates the difference between the two methods. LDA assumes a constant electron density around a given position, while GGA other provides a more accurate representation by incorporating variations in the actual density distribution. For non metallic systems, such as ZnO and MgO, studies have shown that GGA provides a more accurate representation of structural properties compared to LDA. [41]. By incorporating density gradients, GGA corrects these tendencies, resulting in better agreement with experimental bond lengths, cohesive energies, and overall structural predictions [42]. As there are many ways the information from the gradient of the electron density can be included in a GGA functional, there are numerous distinct functionals such as Perdew–Zunger, self-interaction correction and Perdew–Wang functional [37, 43]. The method that was chosen for this work is the Perdew–Burke–Ernzerhof (PBE) functional [44].

With the XC functional established, all essential terms for the total energy func-

tional are defined, enabling the convergence of electron density in an *ab initio* calculation base on the DFT method.

2.1.2 Ionic Convergence

As part of the overall *ab initio* calculation process shown in Figure 2.1, the electronic convergence discussed in the previous section is embedded within a larger iterative loop known as the ionic convergence (or ionic step). According to the Born-Oppenheimer approximation, each nuclei is treated as a fixed particle in space. This means that electronic convergence is determined with respect to a specific, fixed arrangement of the ions.

To find the ground state of a system, the forces acting on each ion are calculated after each electronic convergence. These forces are then used to displace the ions along the direction of the force vectors, guiding them toward more favorable positions before the next electronic iterative convergence calculation. The forces act as a guideline for refining the positions of the ions, taking advantage of the relationship between force and energy gradient. Atoms move toward positions that minimize the system's total energy, driving the system toward equilibrium, where the sum of forces on all atoms approaches zero. This iterative process continues until the forces acting on each ion meet a predefined criterion, indicating that the system has reached a stable configuration and its minimum energy, corresponding to the ground state.

2.1.2.1 Hellman–Feynmann Theorem

The forces on the atoms result from the combined contributions of the electrons and nuclei. The Hellmann–Feynman theorem is compatible with computational calculations and implemented as follows [45, 46].

Much like in classical physics, where force is defined as the negative gradient of the potential energy, the theorem states that the force on a specific atom (F_I) in a quantum system is given by the expectation value of the partial derivative of the Hamiltonian with respect to the atom's position $\mathbf{r}_I$. The expectation value represents the weighted average value of a physical quantity in a given quantum state Ψ. Since the Hamiltonian ($\hat{H}$) is the total energy operator, its expectation value corresponds to the weighted average of the energy E.

$$F_I = -\frac{\partial E}{\partial \mathbf{r}_I} \tag{2.24}$$

The only terms that depend on $\mathbf{r}_I$ are the n–n (U_{IJ}) and e–e (U_{Ii}) interaction potentials. With some approximations, this leads to the following expression for the force:

$$F_I = -\frac{\partial U_{IJ}}{\partial \mathbf{r}_I} - \int \frac{\partial U_{Ii}}{\partial \mathbf{r}_I} n(\mathbf{r})\, d\mathbf{r} \tag{2.25}$$

With this approach, after an electronic iteration is completed, the forces can be calculated by performing simple derivative operations.

2.1.3 Periodicity of Solids

Considering all the methods that were discussed, one common challenge that arises in materials science in particular is the large number of atoms in a bulk material, given the magnitude of Avogadro's number (10^{23}). In solids, this issue can be addressed by utilizing their inherent property of a repeating periodic arrangement of atoms. Assuming a solid has a perfect structure, it can be treated as a repeating unit cell with periodic boundary conditions. One approach for implementing this in *ab initio* methods is the use of plane wave expansion.

2.1.3.1 Plane Wave Representation

In most solid materials, the atomic positions are arranged in a periodic pattern, meaning that any property dependent on position $\mathbf{r}$ will also be periodic. Specifically, this applies to the electron density n and the potential energy U, both of which depend on position and remain unchanged under translation by the position vector $\mathbf{R}$, as shown by the following equations:

$$\mathbf{R} = n_1\mathbf{a} + n_2\mathbf{b} + n_3\mathbf{c} \tag{2.26}$$

$$U(\mathbf{r}) = U(\mathbf{r} + \mathbf{R}) \tag{2.27}$$

$$n(\mathbf{r}) = n(\mathbf{r} + \mathbf{R}) \tag{2.28}$$

The position vector is expressed as a linear combination of the basis vectors (a, b, c), which are three independent vectors that define the unit cell of a crystal

lattice. These basis vectors determine the geometry of the lattice. Additionally, the position vector includes integer multiples of these basis vectors (n_i). The set of all possible combinations of these integer multiples and unit cell vectors generates the entire crystal lattice, which is periodically repeated in space.

In 1929, Felix Bloch introduced an approach to implement the periodicity of the lattice. He realized that the electrons in a solid have the same form as a propagating free electron with the plane wave form as [47]:

$$\Psi(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}} \tag{2.29}$$

This differs only by the crystal lattice's periodic modulation C, which is associated with the atomic positions:

$$\Psi(\mathbf{r}) = Ce^{i\mathbf{k}\cdot\mathbf{r}} \tag{2.30}$$

Here, $\mathbf{k}$ is the wave vector, and the crystal momentum is $p = \hbar\mathbf{k}$. By mapping these plane waves onto the periodic pattern of the solid crystal lattice, Bloch was able to create a quasi-periodic plane wavefunction with a periodic cell potential $u_k(\mathbf{r})$, which has the same periodicity as the lattice potential. This resulted in the modified wavefunction:

$$\Psi(\mathbf{r}) = u_k(\mathbf{r})e^{i\mathbf{k}\cdot\mathbf{r}} \tag{2.31}$$

Due to the periodicity of $u_k(\mathbf{r})$, it satisfies the condition:

$$u_k(\mathbf{r} + \mathbf{R}) = u_k(\mathbf{r}) \tag{2.32}$$

As a result, the wavefunction can be rewritten as:

$$\Psi(\mathbf{r} + \mathbf{R}) = u_k(\mathbf{r} + \mathbf{R})e^{i\mathbf{k}\cdot(\mathbf{r}+\mathbf{R})} = e^{i\mathbf{k}\cdot\mathbf{R}}\Psi(\mathbf{r}) \tag{2.33}$$

This expression shows that when the wavefunction is translated by a lattice vector $\mathbf{R}$, it changes only by a phase factor $e^{i\mathbf{k}\cdot\mathbf{R}}$, rather than altering its overall shape. This fundamental property of Bloch wavefunctions captures the underlying periodicity of electrons in a crystal.

In this way, all relevant physical quantities, including wavefunctions, potential energy, electron densities, and total energy, can be represented in a periodic form using plane waves. This leads to the final stage of the *ab initio* calculation procedure, as shown in Figure 2.1, where the periodic boundaries of the model solid can be adjusted to provide greater degrees of freedom in the geometric relaxation process.

2.1.3.2 Pseudo Potential Approximation

The periodicity associated with solids can be leveraged to reduce the computational cost of *ab initio* calculations. Using a plane wave basis allows simulations to efficiently incorporate the periodic nature of the solid lattice. However, there is no single method to implement this approximation, and different software packages adopt distinct approaches.

Some software accounts for all the electrons in the system. However, as mentioned in Section 2.1.1.3, the more particles included in the calculation, the more computationally complex it becomes. In particular, when using plane waves, this approach is highly expensive and impractical. The challenge arises because the wavefunctions of electrons near the nucleus exhibit rapid oscillations due to the strong Coulomb attraction between the nucleus and these electrons. To capture these sharp features accurately, plane waves with very short wavelengths and high momentum are required. This demand increases the number of plane waves in the basis set, directly leading to a significant rise in computational cost.

To address this issue, atoms are modeled by distinguishing between "core" and "valence" electrons, using a defined cutoff radius (r_c). Core electrons, which reside within r_c, are treated as bound, while valence electrons, located beyond r_c, are considered free. Since core electrons contribute less to the chemical bonding and most observable material properties, their wavefunctions do not need to be computed in full detail. Instead, they are represented by pseudopotentials seen in Figure 2.3, which encapsulate the long range interactions of the core while smoothing the strongly diverging potential in the core region. This process generates a pseudo wavefunction that, within the core region, exhibits fewer oscillations, thereby significantly reducing the number of required plane waves and lowering computational costs. In the valence region, however, the pseudo wavefunction remains unchanged and accurately follows the behavior of the all electron wavefunction.

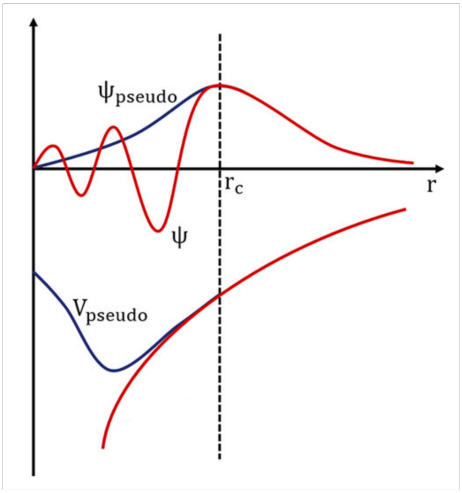

Figure 2.3: Representation comparing the all electron wavefunction and atomic potential with the pseudo wavefunction and pseudopotential, highlighting the cutoff region r_c [48].

Plane wave expansion is not the only method used for the pseudopotential approximation. Some *ab initio* methods employ local pseudopotentials, which are radially symmetric functions that approximate the potential experienced by valence electrons. These pseudopotentials depend only on the radial distance r from the nucleus, meaning they are identical for all angular components. While this approach can be simpler to implement and computationally less expensive, it is generally less accurate for complex systems.

In contrast, the Vienna Ab Initio Simulation Package (VASP) , used in this project, implements projector augmented wave (PAW) potentials, combining the advantages of both plane wave expansion with pseudopotentials and a more precise treatment of core electrons. Originally introduced by Blöchl, the PAW method improves the treatment of core electrons in electronic structure calculations by ensuring the true wavefunction is preserved in the core region [49]. In this approach, the core region wavefunction Ψ_{core} is reconstructed from the pseudo wavefunction using augmentation projectors. The complete all electron wavefunction Ψ_{PAW} is then derived using the following relation:

$$\Psi_{\text{PAW}} = \Psi_{\text{inter}} + \Psi_{\text{core}} - \Psi_{\text{net}} \tag{2.34}$$

Here, Ψ_{inter} represents the valence contribution and is expressed using a plane wave expansion. After incorporating the core contribution, the overlapping compo-

nent Ψ_{net} is subtracted to prevent double counting in the final wavefunction. This approach combines the computational efficiency of pseudopotentials with the accuracy of the PAW method, striking a balance between practicability and precise electronic structure representation.

2.2 The Vienna Ab Initio Simulation Package

In this work, all DFT calculations were performed using the Vienna Ab Initio Simulation Package (VASP) [50, 51, 52, 53]. It is a powerful and widely used tool in computational materials science for electronic and atomic structure calculations. VASP has become a central tool in academic research, as evidenced by its numerous citations. Furthermore, its predictions show a strong correlation with experimental data, highlighting the reliability of its results [54]. What sets VASP apart from other DFT based codes is its precision and efficiency, achieved by employing the PAW method alongside Plane Wave Pseudopotential techniques to solve the Schrödinger equation for periodic systems, as discussed in Section 2.1.3 [55, 56].

2.2.1 VASP Input Files

For each VASP calculation, four different input files are needed, each serving a specific purpose in setting up the calculation. These files define the atomic structure, computational settings, and convergence criteria to ensure accurate results.

INCAR File

The INCAR file serves as the primary control file in VASP, defining the computational parameters for each calculation. It determines how the calculation is carried out by specifying various settings, convergence criteria, and algorithm choices, ensuring that VASP employs the most suitable methods for the system under study.

This file contains a series of entries, each controlling a specific aspect of the calculation. These include settings for electronic relaxation, as described in Section 2.1.1 (e.g. `EDIFF`, `NELM`), ionic relaxation, as discussed in Section 2.1.2 (e.g. `IBRION`, `NSW`, `EDIFFG`), and plane wave expansion energy cutoffs, as covered in Section 2.1.3.1 (`ENCUT`). Adjusting these values allows for the optimization of accuracy, efficiency, and stability in the calculation.

If certain parameters are not explicitly defined, VASP automatically assigns default values based on its internal presets. While these defaults are intended to provide reasonably accurate results, they may not always be optimal for every system. Therefore, careful selection and refinement of INCAR settings are essential to ensure reliable and meaningful outcomes.

POSCAR File

The POSCAR file defines the overall geometry of the system, specifying the lattice base vectors, atomic positions (in Cartesian or reduced coordinates), space group, and number of species. This structural information is essential for the calculation, making the POSCAR file the starting point for all VASP based calculations. Creating a POSCAR file is typically the first step in setting up a VASP calculation, as shown at the beginning of Figure 2.1. It can be manually written using crystallographic data, generated with structure building software, or obtained from online resources. Since VASP calculations are performed based on the structure defined in the POSCAR file, any inaccuracies can directly impact the results.

KPOINTS File

The KPOINTS file in VASP defines the k-point sampling used to discretize the Brillouin zone. In crystallography, there is a real-space lattice and a reciprocal-space lattice, with the latter representing the inverse of real space. In reciprocal space, the first Brillouin zone is the region around a lattice point that is closer to that point than to any other. It possesses special properties that are exploited in electronic structure calculations, as it is the smallest region in reciprocal space that contains all unique wavevectors (represented by **k**-points) needed to describe the periodic system. Proper selection of k-points ensures convergence in electronic relaxation processes, influencing both accuracy and computational efficiency. The KPOINTS file defines the subdivisions along the reciprocal lattice vectors to determine the k-point density, creating a three-dimensional grid mesh of all points within the Brillouin zone, and may include optional shifts in the mesh to optimize accuracy. It also specifies the k-point scheme, indicating whether the k-points are manually defined or automatically generated, and the type of mesh used, such as the Γ-centered grid, which is symmetrically centered around the Γ-point $(0, 0, 0)$. The quality of the k-point sampling directly impacts the results of the calculation. The optimal choice of k-points

is influenced by the system's size, symmetry, and the required level of accuracy. The number of points in each direction is typically chosen to be approximately inversely proportional to the corresponding unit cell dimensions, meaning that larger unit cells require fewer k-points, while smaller unit cells need denser k-point meshes. This relation ensures efficient sampling of the Brillouin zone while keeping computational costs manageable.

POTCAR File

The POTCAR file contains the necessary information for calculating pseudopotentials for each atomic species in the simulation, as explained in Section 2.1.3.2. In addition, VASP employs the PAW method, and as a result, the POTCAR file also specifies the projections for each element. These projections are essential for accurately capturing the electron density in the simulation. The POTCAR file is created by concatenating the pseudopotentials for all elements in the system into a single file. It is critical that the order of the elements in the POTCAR file exactly matches the order in which they appear in the POSCAR file. Any inconsistency between the two files can lead to incorrect results.

For this project, the POTCAR version used is the potpawPBE.64, the latest version from VASP, which is specifically designed for doing calculations with the XC functional PBE method as decribed in Section 2.1.1.7 .

2.3 Universal Structure Predictor: Evolutionary Xtallography

At its core, the Universal Structure Predictor: Evolutionary Xtallography (USPEX) is a crystal structure prediction code that employs an evolutionary algorithm (EA) to achieve high precision and efficiency [57, 58, 59]. This algorithm, inspired by Darwin's theory of evolution and the principle of "survival of the fittest", mimics natural evolutionary processes to refine crystal structures over multiple generations. In each generation, the fittest structures are selected as parents, and evolutionary mechanisms such as heredity and mutation are applied to generate new candidate structures for the next generation, gradually aiming the search toward optimal solutions.

USPEX has the ability to research and discover optimal structures not only for bulk materials but also for surfaces, 2D crystals, polymers, and more. The mate-

rial's structure is optimized (minimize/maximize) throughout the evolutionary process based on a defined set of criteria. These optimization criteria can include any property or combination of properties, such as enthalpy, hardness, magnetization, elastic constants, degree of order, and others. USPEX is also interfaced with a variety of available DFT, molecular dynamics, and semi empirical codes to enhance the optimization process. As mentioned in Section 2.2, in this work VASP was used with USPEX to relax the structures and obtain their energies and other properties.

The basic approach when using USPEX involves a fixed composition, where the elements are predefined by a specific chemical formula. Another powerful feature that USPEX provides is the variable composition method. The purpose of this method is to sample the entire system with different stoichiometry between the elements, meaning that it explores various combinations and proportions of the elements to identify potentially stable structures across a range of chemical compositions and eventually the ground state of the entire system.

The USPEX algorithm is illustrated in Figure 2.4. In the first generation, after determining the chemical formula, an initial population of structures is generated using random symmetry and random topology. Random symmetry ensures well ordered structures by placing atoms in Wyckoff positions based on space group symmetry, while random topology generates structures by modifying known crystal frameworks. These random structures then undergo relaxation through multiple VASP calculations, progressively transitioning from less constrained to more constrained INCAR setups. Their energies are computed and used to rank the relaxed structures based on a user defined optimization criterion. Among the fittest structures, only the highest ranked ones are selected as parents for the next generation. Variation operators, such as mutation and heredity, are then applied to these parents to produce new structures. This iterative process continues until the predefined number of generations is reached.

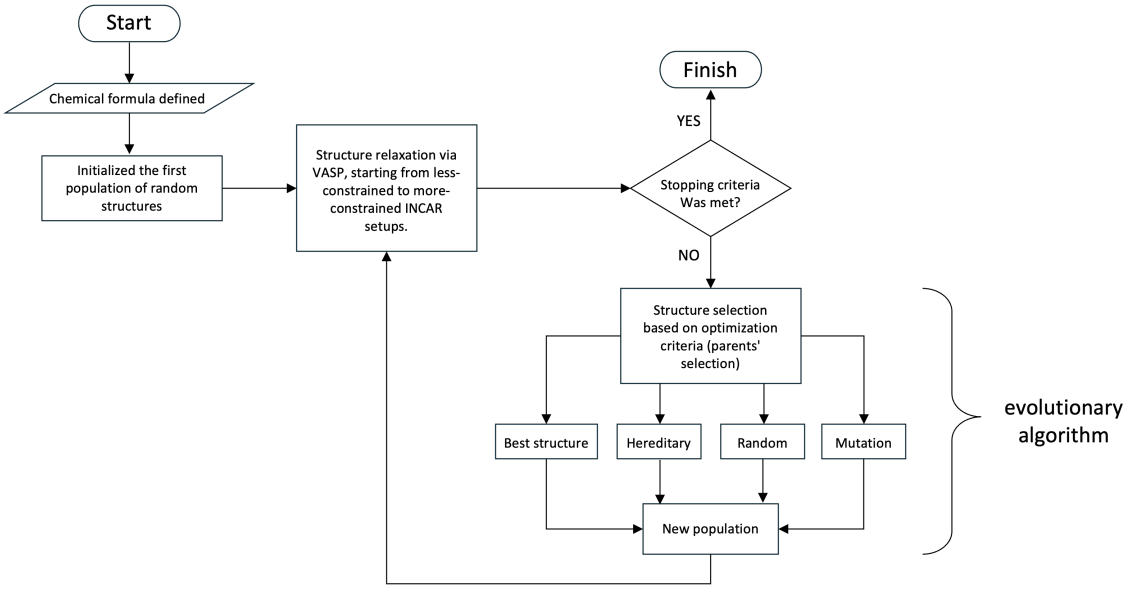

Figure 2.4: Representation of the USPEX algorithm, starting at the top left and proceeding through ordered stages in an iterative loop until the stopping criteria are met.

Each USPEX calculation is configured using a text file called INPUT.txt, which serves as the primary control file for a crystal structure prediction run. It acts as the interface between the user and the USPEX algorithm, specifying interface software, optimization criteria, and computational details. Some of the key parameters defined in the INPUT.txt file include:

- Defines the external computational software (e.g., VASP, Quantum Espresso) used for structure relaxation and energy calculations.
- The number of structures generated per generation.
- The number of top ranked structures selected as the basis for the next generation.
- The distribution of structures among different evolutionary methods.
- The stopping criteria, which can be set as either a fixed number of generations or triggered when the same structure ranks highest for a predefined number of consecutive iterations.

The evolutionary algorithm relies on several key operators that are responsible for generating new structures. USPEX provides the following variation operators:

Heredity creates new structures by combining segments from parent structures, selecting the slices based on their degree of order. The order parameter measures the symmetry of the environment surrounding a given atom, and it is also possible to define the average degree of order for a portion of the structure, selecting the more ordered, less defective parts of the parent structure for heredity, or vice versa, depending on the optimization criteria.

Softmutation generates a new structure by displacing one or more atoms of a parent structure along the eigenvectors of the softest modes. As these low frequency modes correspond to directions of minimal energy curvature, the resulting displacement causes only a small increase in the system's energy, ensuring that the new structure remains close to the original in terms of stability.

Permutation creates a new structure by exchanging the positions of two or more atoms within a parent structure.

Transmutation, used in variable composition mode, generates new structures by changing the chemical identity of a randomly selected atom, replacing it with a different species present in the system.

Lattice mutation creates new structures by modifying the lattice parameters of the parent structure, altering its shape or dimensions while maintaining its overall atomic arrangement.

Another key component of the USPEX algorithm is the Fingerprint Function, which differentiates between structures based on characteristics such as bond lengths, angles, symmetry, stoichiometry, and energy differences, ensuring that duplicate structures are discarded. Without this mechanism, the algorithm could repeatedly generate the same energetically favorable structure, causing it to become stuck in local minima and limiting the exploration of novel configurations that could potentially lead to the global minimum.

The USPEX algorithm offers a highly efficient and versatile approach to crystal structure prediction. Its ability to explore a wide range of structural possibilities, together with its optimization strategies, makes it a valuable tool for discovering stable and novel crystal structures with a high success rate through high throughput methods.

2.4 Special Quasirandom Structures

Alloys, which are a common class of disordered materials, are systems in which atoms occupy lattice sites without long range periodicity in their arrangement. Unlike ordered crystals, where each atomic site is uniquely defined by symmetry, disordered systems exhibit compositional randomness, making the precise atomic configuration indeterminate. Instead, only the overall statistical distribution of atomic species is meaningful. To represent this randomness in crystallographic models, fractional site occupancies are often used. These describe a superposition of atomic species at a given lattice site, weighted by the composition ratio.

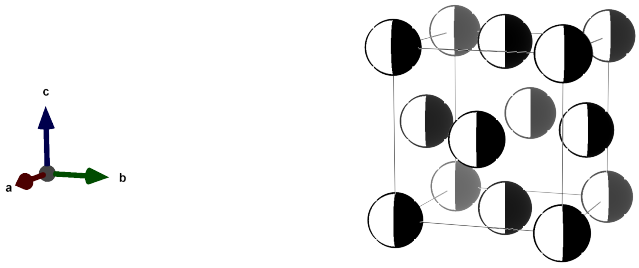

Figure 2.5: Visualization of fractional site occupancies implemented on an FCC crystal structure. Each site represents a 50%-50% statistical mixture of two atomic species, reflecting the composition ratio.

Traditional simulation methods and software packages such as VASP rely on well defined atomic positions, where each atom occupies a specific site with no partial occupancies. Most importantly, the structure must be compatible with periodic boundary conditions. The Special Quasirandom Structures (SQS) method, introduced in 1990 by Alex Zunger *et al.* [60], addresses this limitation by enabling the simulation of disordered materials and alloys through the construction of a finite periodic supercell that statistically approximates a random atomic distribution. In doing so, SQS helps bridge the gap between the requirements of first principles simulations and the inherently non periodic nature of disordered alloys.

2.4.1 Correlation Functions

The SQS method relies on the concept of correlation functions, which quantify the statistical occurrence of specific arrangements of atomic species at given separations

within a lattice. In binary alloys, these include pairwise configurations such as A–A, B–B, and A–B at various neighbor distances, as well as clusters of three or four atoms.

In a perfectly random alloy, the correlation functions are defined as weighted averages that depend solely on the composition and the number of sites in the cluster. For a binary system with composition x and a figure f, defined by its number of sites k (point, pair, triplet, quadruplet, ...), geometry (relative positions), and neighbor shell (first, second, third, ...), the expected value of the correlation function is given by:

$$\langle \Pi_f \rangle = (2x - 1)^k \tag{2.35}$$

At a composition of $x = \frac{1}{2}$, these values vanish for all $k > 0$, indicating the absence of preferential ordering.

For a truly random distribution, the correlation value is independent of the figure's geometry or neighbor shell, it depends only on x and k. However, in SQS, the actual values can differ between shells or geometries, because the limited supercell size cannot reproduce all correlation functions exactly. SQS constructs a finite, periodic supercell in which the configuration the atoms are positioned in it with specific correlation functions that closely match those of the ideal random alloy for the most physically relevant clusters typically involving the first few nearest neighbors. This ensures that the local environment around each atom statistically resembles that of a truly random alloy.

Mathematically, the correlation function for a given cluster type f in a structure σ is defined as:

$$\overline{\Pi}_f(\sigma) = \frac{1}{ND_f} \sum_{l=1}^{ND_f} \prod_{j \in f} S_{l,j} \tag{2.36}$$

where:

- σ is the specific atomic configuration of the lattice, specifying the occupations of all sites.
- $\overline{\Pi}_f(\sigma)$ is the average correlation function for figure f in configuration σ, computed over all symmetry equivalent occurrences of f in the supercell.

- N is the total number of lattice sites.
- D_f is the number of symmetry equivalent clusters of type f per site.
- $S_{l,j}$ is a spin-like occupation variable that can take the value -1 for atom A or $+1$ for atom B in a binary system.
- The product runs over the sites j in cluster f, and the sum runs over all occurrences l of this cluster in the supercell.

By minimizing the deviation of the computed correlation functions from their ideal random target values for a selected set of clusters, the SQS method enables accurate modeling of chemical disorder within small supercells. This makes it possible to apply computational electronic structure methods to disordered alloy systems while preserving the essential statistical features of disorder.

2.4.2 Alloy Theoretic Automated Toolkit

The Alloy Theoretic Automated Toolkit (ATAT) was used as the implementation code for SQS calculations in this work [61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76]. It features an efficient Monte Carlo optimization algorithm that enables the construction of SQS supercells without exhaustively enumerating all possible atomic configurations [77]. Instead, the algorithm performs a randomized search by iteratively swapping atomic species within the supercell and evaluating the resulting correlation functions. It rapidly identify an atomic arrangement that closely matches the target correlation functions of a perfectly random alloy, thereby minimizing the number of steps required to reach the optimal solution. An additional advantage of ATAT is its flexible control over key parameters that directly influence the quality and efficiency of SQS generation. The size of the supercell can be adjusted to match the desired composition, and users can specify the numerical tolerance for how closely the generated structure must reproduce the target correlation functions. These capabilities make ATAT particularly well suited for generating SQS structures quickly and efficiently.

2.5 Summary

The methods discussed in this chapter present the essential theories and computational tools for crystal structure prediction, beginning with the implementation of

density functional theory (DFT) in VASP for structural calculations. The integration of VASP within the USPEX framework, which employs evolutionary algorithms, enables efficient exploration of the materials space for potential discoveries. Additionally, the SQS method, as implemented in ATAT, is used to simulate the complex behavior of disordered structures in alloys and solid solutions. By combining these computational approaches, accurate and reliable predictions of crystal structures can be achieved. In the following chapters, these methods will be applied to the exploration of various materials.

3 Simulating Disordered Materials

Alloys are disordered materials that play a critical role in fields such as electronics, where they are used in high strength and high conductivity copper alloys for electrical wiring [78], and aerospace, where nickel based alloys are employed in jet engine turbine blades [79]. Steel, in particular, play a crucial role in modern infrastructure, serving as a material for building ships, cars, and buildings. Their importance lies in their ability to combine and optimize multiple material properties. To ensure these materials can be reliably used, it is important to evaluate their stability, often through phase diagram analysis.

Cu–Mn alloys are used in applications requiring materials with uniform electrical resistivity over a broad temperature range, which is necessary for obtaining accurate and stable electrical signals in electronic devices. Copper contributes to the low resistivity, while manganese enhances thermal stability. These alloys find applications in mobile electronic devices [80]. However, the long-term reliability of these materials is in question, as prolonged operation at elevated temperatures may lead to degradation, affecting sensing performance. The current understanding of phase stability in the Cu–Mn system relies on phase diagrams that show conflicting results, underscoring the need for a reassessment. In particular, well known phase diagrams from the literature highlight the controversy surrounding the Cu–Mn system, introducing a debate regarding its behavior, as shown in Figure 3.1. Focusing on the 0 – 25 at.% Mn concentration range, there is a clear mismatch: the left phase diagram describes a single phase solid solution, while the right suggests the existence of intermetallic compounds with fixed stoichiometry, such as Cu_5Mn and Cu_3Mn compounds that have never been experimentally observed. To resolve this discrepancy and uncover the system's true behavior, advanced computational tools were employed.

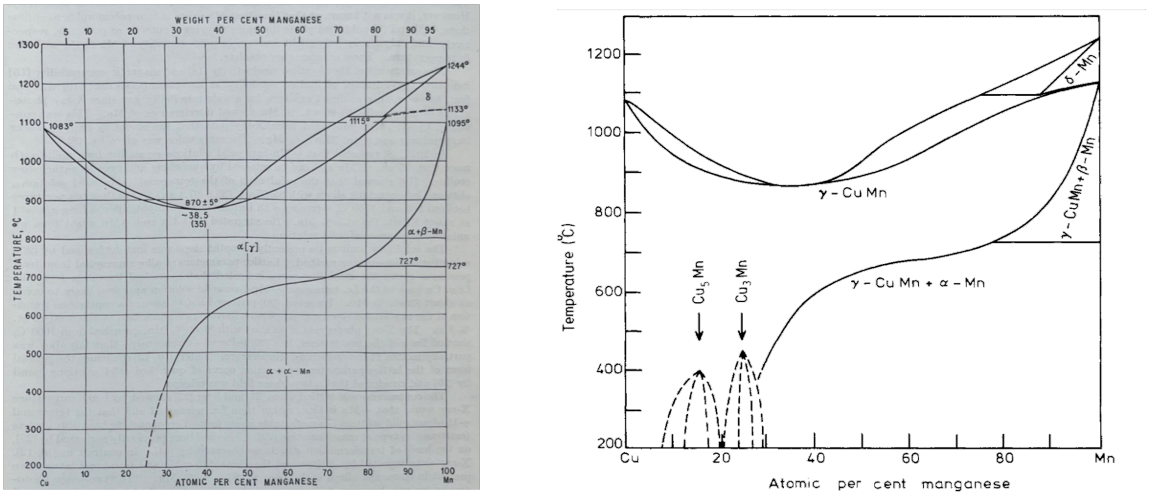

Figure 3.1: Phase diagrams of the Cu–Mn binary system. Adapted from Hansen and Anderko (left) [81], which shows a single phase solid solution in the 0–25 at.% Mn range, and from Cowlam and Shamah (right) [82], which predicts intermetallic compounds such as Cu_5Mn and Cu_3Mn.

3.1 Simulating the Cu–Mn System Using SQS

The Cu–Mn system was simulated and analyzed as a solid solution using the SQS method, implemented through the ATAT software and accessed via PyMatGen. The conventional FCC structure of Cu, obtained from the Materials Project, served as the host framework. This choice is supported by experimental evidence based on X-ray diffraction (XRD) analysis, which shows that Cu–Mn alloys retain the FCC structure under various conditions [83]. The structure was expanded into an 80 atom supercell using a scaling factor of $2 \times 2 \times 5$, and Mn atoms were introduced by replacing Cu atoms. The configuration of atom distribution was refined to maximize disorder, simulating various alloy compositions of the form Cu_xMn_{1-x}, as seen in Figure 3.2.

Ensuring that the maximum disordered configurations generated by ATAT conform to one of the core principle of the SQS method, namely, the preservation of the structural framework, was a necessary step prior to further analysis. This validation was carried out by comparing all generated structures to the initial FCC framework and to one another using the `StructureMatcher` class from the `pymatgen.analysis.structure_matcher` module, together with the `FrameworkComparator` method. The comparison was based solely on atomic arrangement and lattice geometry, independent of the chemical species that occupy the sites. The results confirmed that all structures matched within a tightly constrained tolerance, verifying the structural

consistency of the ATAT generated structures presented in Figure 3.2

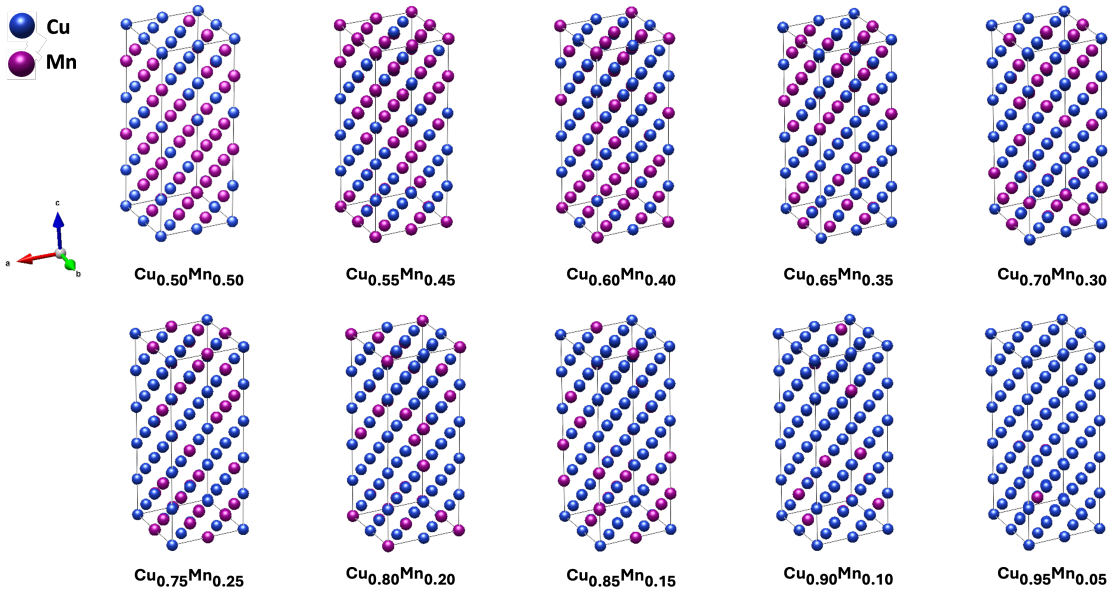

Figure 3.2: Finite supercell representations of Cu–Mn alloy structures generated by ATAT, using a scaling factor of $2 \times 2 \times 5$ applied to the conventional FCC Cu cell. The atomic configurations were optimized to simulate alloy compositions ranging from 5 at.% to 50 at.% Mn (Cu_xMn_{1-x}).

The value of Enthelpy of mixing (ΔH_{mix}) for each composition was computed using VASP with two relaxation steps, based on the SQS maximum disorder configuratin structures. Calculations were performed with respect to the energies of the pure elements and normalized by the Mn concentration. As shown in Figure 3.3, ΔH_{mix} is positive for any Mn concentration, indicating that it is thermodynamically unfavorable for Cu and Mn to mix in the form of a solid solution. The plot also reveals an asymmetric behavior in the ΔH_{mix} values, reaching a maximum of approximately 1.4 eV at around 5 at.% Mn and then gradually decrease as Mn content increases. This behavior suggests that compositions with higher Mn concentrations, such as Cu_3Mn_2, are more likely to form than those with low Mn content. To confirm that this trend is not specific to a particular configuration, additional calculations were performed using ten randomly reassigned atomic species for each composition. These tests yielded the same trend, indicating that the observed behavior does not depend on the specific atomic arrangement, but rather is strongly governed by the overall composition.

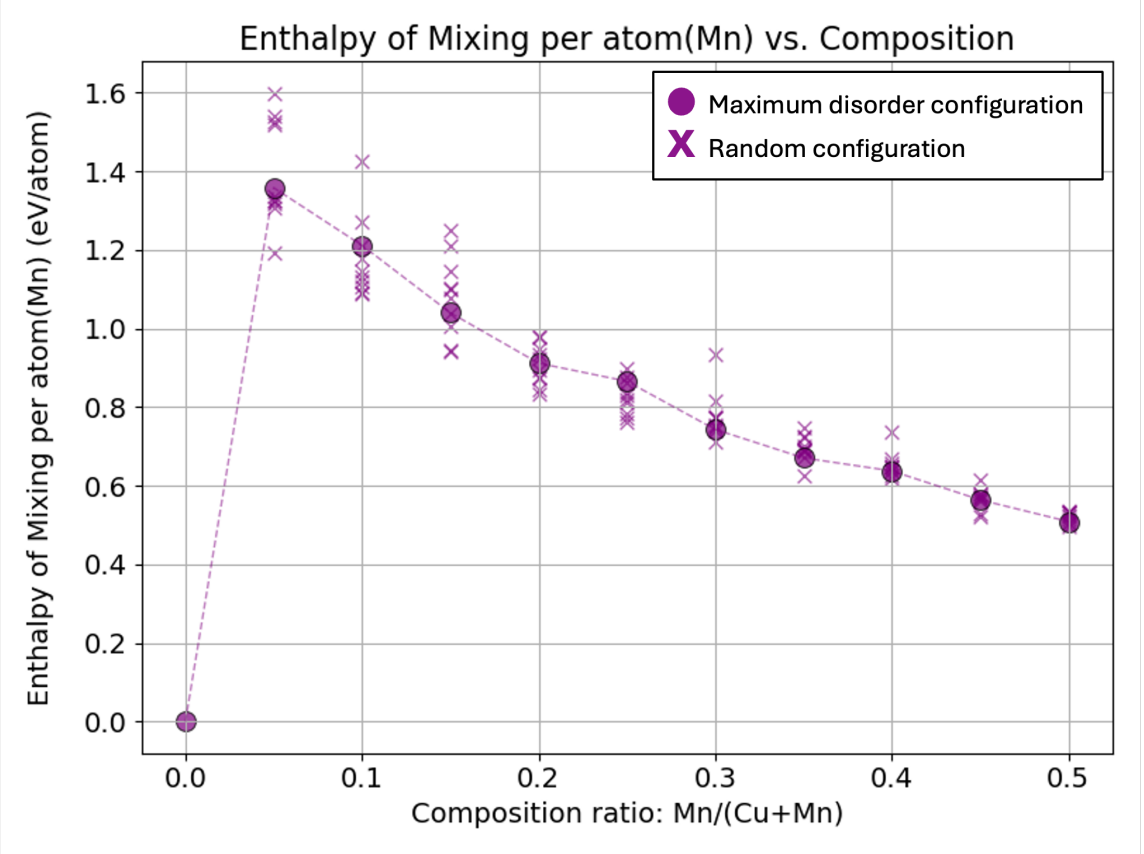

Figure 3.3: Values of enthalpy of mixing per Mn atom with respect to Mn concentration calculated for a simulated Cu–Mn solid solution in the Cu FCC framework, after two relaxation steps. Two sets of calculations are shown: one using maximum disorder configurations (solid circles) and the other using randomly reassigned atomic species (cross markers).

3.2 Structure Prediction of the Cu–Mn System

Fixing the Cu–Mn system to the FCC framework of copper, as done in the SQS based solid solution simulations, may be misleading, as it restricts structural degree of freedom and potentially overlooks more stable configurations. To address this, a variable composition structure prediction approach was employed using USPEX, allowing the atomic arrangement and lattice to evolve freely across the entire compositional range. Assuming sufficient sampling, this approach can reveal compositions that are potentially thermodynamically stable, or at the very least, provide a better understanding of the system's behavior.

Figure 3.4 summarizes the Enthalpy of Formation (ΔH) values calculated for

the Cu–Mn system across a wide compositional range, from 0 to 100 at.% Mn, for structures generated by USPEX in the variable composition structure prediction. similer with the SQS based results, ΔH remains positive for all variable compositions, indicating that mixing is thermodynamically unfavorable even with the additional degree of freedom allowed by this approach. The lowest enthalpy values are observed near the pure elements, which serve as reference points for assessing phase stability across the composition range.

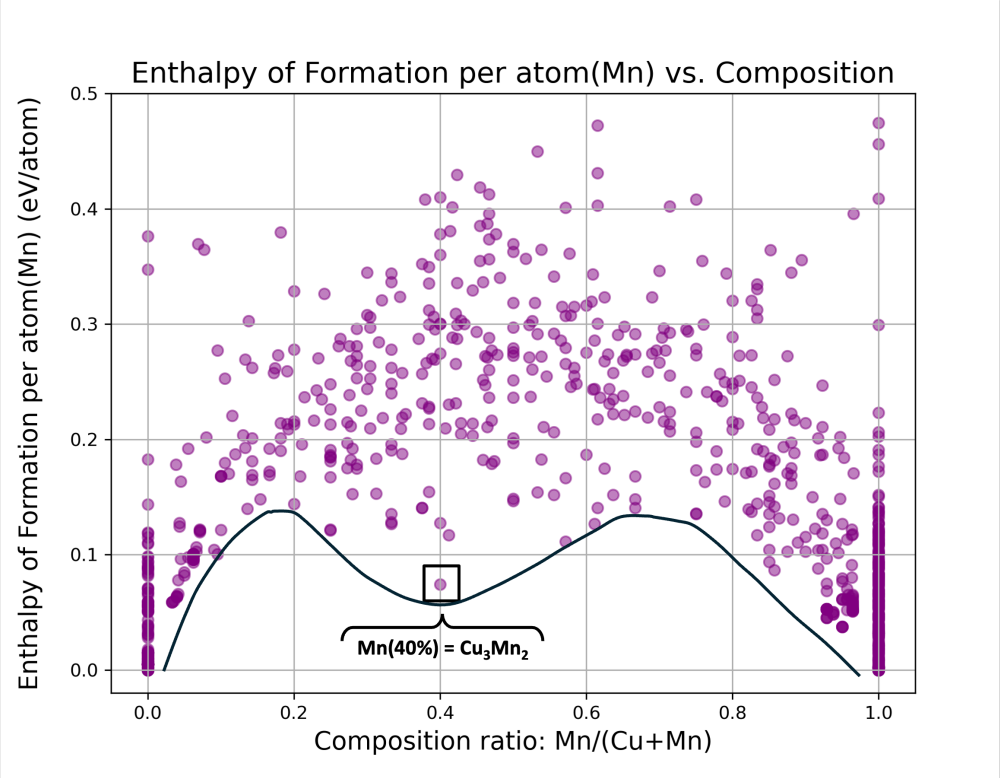

Figure 3.4: Enthalpy of formation (ΔH) per Mn atom as a function of Mn concentration in the Cu–Mn system. Each point represents a distinct structure generated during structure prediction. The variation at each composition reflects the different structures explored in the process of identifying the lowest energy ground state. A solid line shows the overall trend in ΔH, with an unexpected dip observed at Mn/(Cu+Mn) = 0.4, corresponding to the Cu_3Mn_2 composition.

The trend of ΔH observed in Figure 3.4 reveals a noticeable dip at a composition ratio corresponding to 40% Mn, or equivalently the Cu_3Mn_2 stoichiometry. This

relatively stable outlier prompted a fixed composition structure prediction, aimed at probing its potential stability. Following 363 structural relaxations, the predicted configuration, shown on the left side of Figure 3.5(b), yielded an enthalpy of formation of 0.24 eV per Mn atom. This positive value, together with the appearance of the predicted structure, which reflects the periodicity imposed by USPEX, supports the general tendency of the Cu–Mn system to undergo phase separation into compositionally distinct regions.

Manual rearrangement of the atomic configuration was carried out to enhance atomic segregation between Cu and Mn species, while keeping the lattice geometry and atomic positions unchanged. After structural relaxation, the modified configuration, shown on the right side of Figure 3.5(b), achieved a lower ΔH of 0.18 eV per Mn atom. This decrease in enthalpy, resulting from a reduced number of Cu–Mn interfaces, indicates that fewer Cu–Mn interactions provides further thermodynamic evidence that the system favors phase separation.

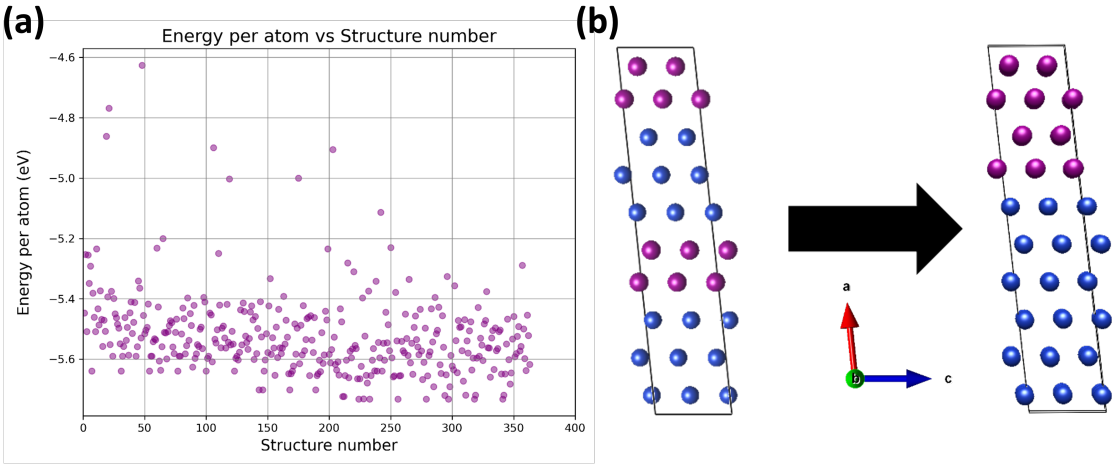

Figure 3.5: Cu_3Mn_2 composition Mn/(Cu + Mn) = 0.4 obtained from fixed composition structure prediction using USPEX. (a) Energy per atom as a function of structure number. (b) Left: lowest energy structure identified after 363 relaxations, exhibiting phase separation. Right: same structure with manually rearranged Cu and Mn atoms to enhance segregation.

The Cu_5Mn composition was also examined, motivated by experimental reports suggesting the possible formation of an intermetallic phase at this atomic ratio, as seen in Figure 3.1 and in additional studies in the literature [84]. Structure prediction calculations, carried out in a manner similar to those for Cu_3Mn_2, showed that the resulting structure also exhibits phase separation with a positive $\Delta H = 0.20$ eV/per

Mn atom. This makes it very clear that the Cu–Mn system exhibits a tendency for the elements to separate, reflecting a strong thermodynamic resistance to mixing, at least in the absence of thermal effects. It is also worth noting that calculations considering magnetic interactions yielded similar results.

3.3 Experimental Validation of Phase Separation in Cu–Mn Alloys

Experimental observations reported in a recent study on the Cu–Mn system reveal a clear tendency toward phase separation into Mn rich and Mn lean regions [83]. As shown in Figure 3.6, scanning electron microscopy (SEM) combined with energy dispersive spectroscopy (EDS) was employed to characterize the decomposed alloys via elemental mapping. The Mn concentration profiles, extracted along an arbitrary line across the sample surface, are presented for both the homogenized and decomposed samples. In the homogenized sample, Mn concentration remains uniform, with only minor fluctuations within the measurement uncertainty. In contrast, the decomposed sample exhibits significant variations in Mn concentration, ranging from 9 to 13 at.%. Notably, the decomposition is compositionally asymmetric, meaning the resulting phases deviate unequally from the initial average composition. This asymmetry is consistent with the asymmetric enthalpy of mixing, as shown in Figure 3.3.

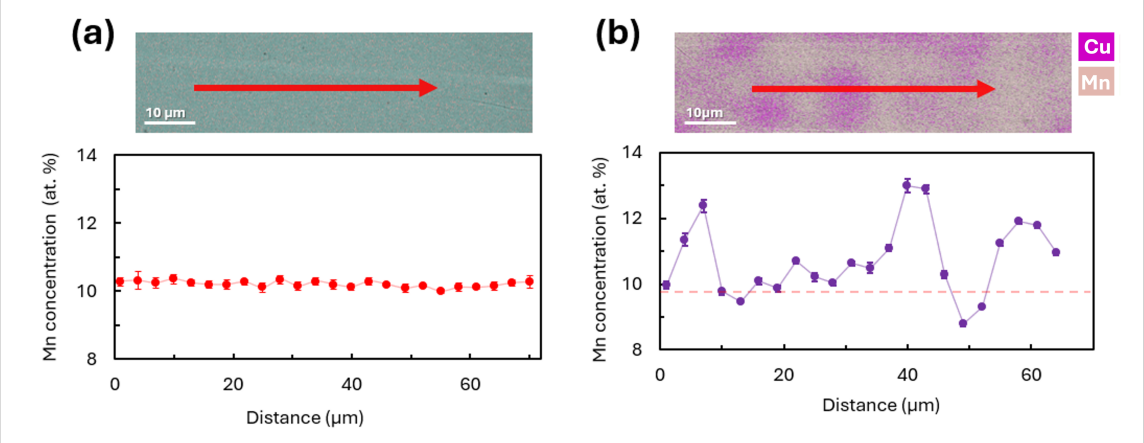

Figure 3.6: Concentration of Mn in the (a) homogenized and (b) decomposed samples. Each value represents the average of five measurements; error bars are standard deviations. Insets in (a) and (b) show SEM-EDS elemental maps along the line scan direction [83].

3.4 Summary

The theoretical results indicate spontaneous decomposition into two phases, one solute-rich and one solute-poor, a process known as spinodal decomposition [85]. This behavior aligns strongly with the presented theoretical predictions of phase decomposition and an unfavorability to mix, which are confidently supported by thousands of calculations showing consistently positive values of ΔH. Given that some industries rely on literature phase diagrams, such as those shown in Figure 3.1, which indicate that the Cu–Mn system should mix and is thermodynamically stable, the presented results underscore the need for a revision. The commonly used Cu–Mn alloys are probably metastable, and it is likely only a matter of time before signs of decomposition begin to emerge in practical applications, if they haven't already.

4 Simulating Unknown Materials

In the previous chapter, the focus was on modeling disordered materials when the underlying crystal structure is already known, using structural frameworks as a starting point. In the case of undiscovered materials, however, no such structural information exists, requiring a different approach that begins with only a chemical formula as input. The following chapter focuses on predicting the structure and stability of such undiscovered inorganic materials. In this work, a central focus is the generation and evaluation of a list of combinations, chemical formulas constructed as part of this study that may correspond to undiscovered, yet potentially stable compounds. This process is carried out using advanced computational tools for structure prediction (USPEX), employing DFT calculations (VASP) to assess stability and guide the search for viable structures. To maintain clarity, the term combinations will be used throughout to refer to these hypothetical candidates, whereas compounds will denote only those already known and documented in established materials databases.

4.1 Electroneutral Combinations for Structure Prediction

By definition a material must have as many electrons within it as are brought by the elements. If the unit cell is periodic, any partition of these electrons between the elements must sum to zero [86]. For example, in sodium chloride, each sodium ion (Na^+) is balanced by a chloride ion (Cl^-), ensuring overall neutrality. This principle is crucial when exploring new materials, as it provides a fundamental guideline for assessing the potential stability of their structures.

4.1.1 Determining Oxidation States from Experimental Databases

The oxidation states of all elements is determined by using the ICSD, one of the largest databases of fully characterized inorganic crystal structures, containing over

200,000 entries. It provides all necessary structural information, atomic coordinates, bond lengths, space groups, and oxidation states. This vast amount of data makes the ICSD a valuable resource for materials science research, promoting the prediction of new inorganic materials. Each entry in the ICSD includes information about the oxidation states of the elements in a compound, listed under the field `_atom_type_oxidation_number`. For example, in the case of magnesium aluminate (Al_2MgO_4), the oxidation states are specified as Al^{+3}, Mg^{+2}, and O^{-2}. Based on these entries, a dataset can be constructed that records all oxidation states observed for each element across the database.

Previous research compiled a dataset that includes all recorded oxidation states in the ICSD and, by ranking them based on their relative commonness, reduced this set to 84 representative oxidation states (see Table A.1 in the Appendix). Constraining oxidation states to these most commonly observed values helps control the combinatorial explosion of possible neutral combinations. This effect is illustrated in Table 4.1 taken from Ref. [87], which quantifies the dramatic growth in the number of charge balanced combinations as the number of ions per compound increases. Notably, these 84 representative states capture over 90 % of all oxidation state assignments in the ICSD (excluding zero states), demonstrating that this reduced set preserves a sufficiently large and representative portion of experimentally reported chemistries for further analysis.

Table 4.1: Number of possible element combinations with the principle of electroneutrality enforced. Showing the combinatorial explosion in the number of neutral combinations as the number of oxidation states considered grows, together with the number of ions per combination [87].

Components	2	3	4	5
Electroneutral combinations for 84 Oxidation States	178	1,913	27,073	268,795
Electroneutral combinations for the 157 Oxidation States	510	9,858	245,551	4,307,895
Electroneutral combinations for the 367 Oxidation State	3,445	179,659	10,016,212	431,839,459

Building on a previously compiled list of common oxidation states, several simplifications were introduced to streamline their application in this work. First, arsenic

(As) can occur either as an anion (-3) or as a cation ($+5$), leading to ambiguity when determining whether a neutral combination contains cationic or anionic arsenic. While this behavior also appears in some other elements (e.g., Ge and P), arsenic stands out for having comparable numbers of entries for both states in the dataset. To address this, arsenic is consistently treated as a cation, as it appears more frequently in its cationic form (4,349 combinations) compared to its anionic form (2,567 combinations). Second, some elements such as Au, Pt, Eu, Sn, Cu, and Fe exhibit multiple common oxidation states, which could introduce ambiguity and make the process more complex. To simplify the approach, only a single oxidation state was assigned to each element, chosen based on the highest number of recorded entries. In most cases, this corresponded to the higher oxidation state. In the case of gold (Au), although the $+1$ state had slightly more entries than the $+3$ state (by 56), the $+3$ oxidation state was chosen due to the small difference and to maintain consistency across elements with multiple oxidation states. The resulting set of adjusted oxidation states is summarized in Table A.1 in the Appendix.

4.1.2 Generating Neutral Combinations

Creating neutral combinations was carried out using a Python script, as shown in Appendix A.1. The script begins by using the reduced common oxidation state list to generate all neutral combinations, disregarding the order of elements in each formula. This ensures that combinations containing the same elements but written differently, such as CO_2 and O_2C, are treated as a single unique combination.

Generating neutral combinations depends on defining the number of ions in each combination, referred to as `n_ions`. To control the size of the resulting list, a maximum number of ions, `n_max`, is specified. This sets the upper limit on how many ions each combination may include, resulting in the generation of all possible neutral combinations that fall within the specified `n_max` range. When `n_ions` is set to seven, the number of possible combinations exceeds that of setting `n_max` to six, making values of seven or higher impractically large for further exploration at this stage of the study. Therefore, in this work, `n_max` was set to six, yielding a large but manageable list of combinations with following count:

Table 4.2: Number of neutral combinations as a function of ion count (`n_ions` = 2-7). The final cell shows the cumulative total number of combinations represented by `n_max`.

Number of Ions (n_ions)	Number of Combinations	Cumulative Sum of all Combinations (n_max)
2	126	126
3	1,363	1,489
4	15,979	17,468
5	144,546	162,014
6	1,169,852	1,331,866
7	8,460,836	9,792,702

4.1.3 Filtering Strategy

Analyzing the number of combinations, as shown in Table 4.2, is a tedious task that might even be considered impossible, which makes the process of reducing the list of combinations especially important. Starting with the straightforward step of discarding all combinations already considered "known" based on the Materials Project database. For example, one of the combinations generated was F_2Ni, according to the Materials Project, this corresponds to nickel fluoride, a well known compound. However, combinations that are not found in the Materials Project after comparing them to the generated list, are considered unexplored. When comparing the generated combinations to entries in the Materials Project database, one might question whether differences in element order, such as HCCr versus CHCr, could lead to missed matches. However, the comparison is performed using the `mpr.summary.search` function from the Materials Project Application Programming Interface (API), which enables structured queries in Python and is insensitive to the order of elements in chemical formulas, ensuring accurate identification of known compounds. Challenges may still arise when working with combinations involving many ions, as certain subgroups can start to be part of the chemical formula according to conventional chemical notation. For instance, the compound $Ca(OH)_2$ is typically written in grouped form, but a generator might output it as CaO_2H_2, causing a mismatch with the database since the grouped representation will not be recognized. This strengthens the reasoning for continuing to focus on combinations with a smaller number of ions at this stage of the work. Applying the comparison

process to the full list of combinations yielded 4,441 matches with known entries in the Materials Project, while 1,327,425 did not appear in the database. These remaining candidates form the basis for the next stages of filtering.

The next stage of filtering focuses on eliminating combinations that pose challenges either due to safety concerns or computational difficulties. This step is based on two primary criteria: the removal of f-block elements and the exclusion of radioactive elements. The first filtering criterion removes all combinations that include f-block elements, which are difficult to model accurately using standard DFT methods due to their complex electronic structures, strong electron correlations, and significant relativistic effects [88]. Properly modeling f-block elements often requires advanced computational techniques, which greatly increase both the complexity and cost of simulation. Excluding them ensures that the combinations left for further evaluation are more suitable for the scope and technical level of this work. The following f-block elements were excluded in this filtering stage:

- **Lanthanides** La, Ce, Pr, Nd, Pm, Sm, Eu, Gd, Tb, Dy, Ho, Er, Tm, Yb, Lu
- **Actinides** Ac, Th, Pa, U, Np, Pu, Am, Cm, Bk, Cf, Es, Fm, Md, No, Lr

The second filter removes all combinations containing radioactive elements due to the risks they pose. These elements emit harmful radiation that can cause cancer, radiation sickness, and genetic mutations, making them hazardous to handle [89]. Their use requires specialized facilities, strict safety protocols, and often results in unpredictable chemical behavior due to their inherent instability. Because of these practical and safety concerns, such combinations are excluded to ensure that the remaining candidates are more feasible for experimental synthesis and application.

> **Filtered radioactive elements:** Ts, Tc, Sg, Rn, Rg, Rf, Ra, Po, Og, Nh, Mt, Mc, Lv, Hs, Fr, Fl, Ds, Db, Cn, Bh, At

The final filtering step continues the approach of maintaining simplicity, focusing on isolating the most straightforward and logically identifiable combinations that remain in the list. In this stage, the focus is directed toward identifying unexplored binary combinations composed of only two distinct elements, with emphasis on potential binary semiconductors, a class of materials known for their wide range of technological applications, including photovoltaics, transistors, and sensors [90].

The search targets combinations involving two elemental species drawn from specific groups in the periodic table that are commonly associated with semiconducting properties, ensuring that the resulting candidates remain both chemically plausible and compositionally simple. Specifically, combinations from the following group pairs were prioritized: (III, V), (II, VI), (I, VII), (IV, VI), (V,VI), and (II, V) [91, 92, 93]. This targeted filtering helps prioritize chemically and functionally meaningful candidates.

- **Group I:** Cu, Ag, Au
- **Group II:** Zn, Cd, Hg
- **Group III:** B, Al, Ga, In, Tl
- **Group IV:** C, Si, Ge, Sn, Pb
- **Group V:** N, P, As, Sb, Bi
- **Group VI:** O, S, Se, Te
- **Group VII:** F, Cl, Br, I

Resulting combinations include three unique binary compounds that satisfy the group pair criteria along with all compositional constraints: $SnTe_2$, $GeTe_2$, and AuI_3. These candidates advance to the next stage of the material discovery process, where their crystal structures will be predicted and their stability examined.

4.2 Predicting and Evaluating Combination Structures

Structure prediction and stability evaluation of the selected candidates are carried out using a series of computational steps, outlined in the following section along with their corresponding results. The fixed composition structure prediction tool in USPEX is used to determine the lowest energy crystal structure for each candidate combination. Once a candidate structure is obtained, it is further relaxed using VASP to calculate its total energy at 0 K. This energy is then compared to the energies of the constituent pure elements to compute the enthalpy of formation.

To assess thermodynamic stability, the calculated enthalpy of formation is combined with those of all experimentally reported compounds within the same chemical system, queried from the Materials Project, to construct a convex hull. A convex hull is a graphical representation that connects the lowest energy phases across all composition ratios within a chemical system. It spans variable composition compounds, allowing one to identify which phases are thermodynamically stable relative to other competing phases. Compounds that lie directly on the convex hull are considered stable, while those above the hull are, at best, metastable or, in some cases, cannot be synthesized under any conditions, and tend to transform into a mixture of lower energy phases [15]. With these computational tools, candidate combinations can be simulated and evaluated to determine whether they are promising enough to justify experimental synthesis efforts in the lab.

4.2.1 $SnTe_2$

$SnTe_2$ is the result of a Group IV–VI combination. The structure prediction process is summarized in Figure 4.1, which shows the energy per atom, calculated using VASP, as a function of structure number. Each point in the figure represents a unique structural configuration generated by USPEX. Over 400 structures were analyzed until the lowest energy structure was identified. This structure remained the minimum energy configuration for eight consecutive generations, which served as the stopping criterion for the USPEX run.

The final predicted structure is shown in Figure 4.2(a). The resulting structure exhibits the symmetry of the Pm space group, which contains relatively few symmetry operations. The optimized lattice parameters are $a = 4.12$ Å, $b = 4.18$ Å, and $c = 6.45$ Å, with angles $\alpha, \beta, \gamma = 90°$. These nearly orthogonal cell angles indicate a weakly distorted monoclinic lattice.

The coordination environment of the tin atoms in this predicted phase consists of distorted $SnTe_6$ octahedra, with an average Sn–Te bond length of approximately 3.15 Å. These octahedra are connected by both corner and edge sharing modes, forming a three dimensional network similar to those observed in experimentally reported tin chalcogenides such as SnS_2 (mp-1170), $SnSe_2$ (mp-665), and SnO_2 (mp-856). However, when considering the bond lengths, there is a notable difference: in most of these experimentally reported structures, the Sn–chalcogen bond lengths typically fall within the range of 2–3 Å(based on the material project). The presence of distorted octahedra in the predicted phase further lowers the symmetry relative

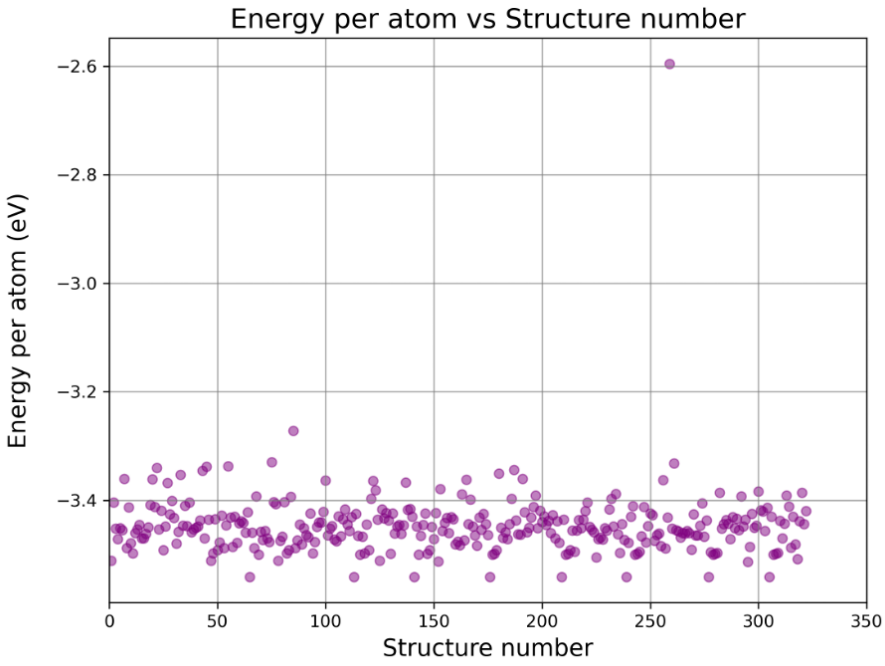

Figure 4.1: Energy per atom as a function of structure number for the $SnTe_2$ combination, obtained from fixed composition structure prediction using USPEX. The minimum energy structure was identified after analyzing over 400 structures.

to these known structures. An interesting observation supporting the plausibility of this predicted structure emerges when using the `StructureMatcher` with the `FrameworkComparator` (available in PyMatGen) to compare it against all experimentally observed entries in the Materials Project. This comparison yields matches within an acceptable loss tolerance, such as EuSnP (mp-19794) and SrSnP (mp-8539), indicating that the predicted structure shares key features with known compounds, thereby lending credibility to its potential as a physically realizable phase. The combination of familiar coordination environments and subtle structural similarities with known compounds. These factors support the structural plausibility of the predicted phase, further reinforced by its negative enthalpy of formation obtained from the convex hull analysis. However, as shown in Figure 4.2(b), the convex hull of the Sn–Te system indicates that $SnTe_2$ does not lie on the hull, being ~ 0.1805 eV per atom (Te) above it. This suggests that the compound is not thermodynamically stable with respect to competing phases and is therefore unlikely to correspond to an experimentally accessible phase.

It is worth noting that $SnTe_2$ has been the subject of previous computational

studies, primarily focusing on a predetermined monolayer phase different from the predicted structure [94, 95]. These studies do not aim to determine the true ground state bulk structure, nor do they provide experimental validation. Moreover, the Materials Project, despite containing numerous hypothetical and metastable entries, does not list $SnTe_2$ at all. This combination of limited computational exploration, lack of experimental evidence, and absence from comprehensive databases reinforces the conclusion that this composition is unlikely to correspond to a practically stable or experimentally accessible bulk phase. At the same time, the existence of such studies highlights that by employing a systematic structure prediction workflow, it is possible to identify promising combination with interesting properties, such as those that motivated these previous investigations.

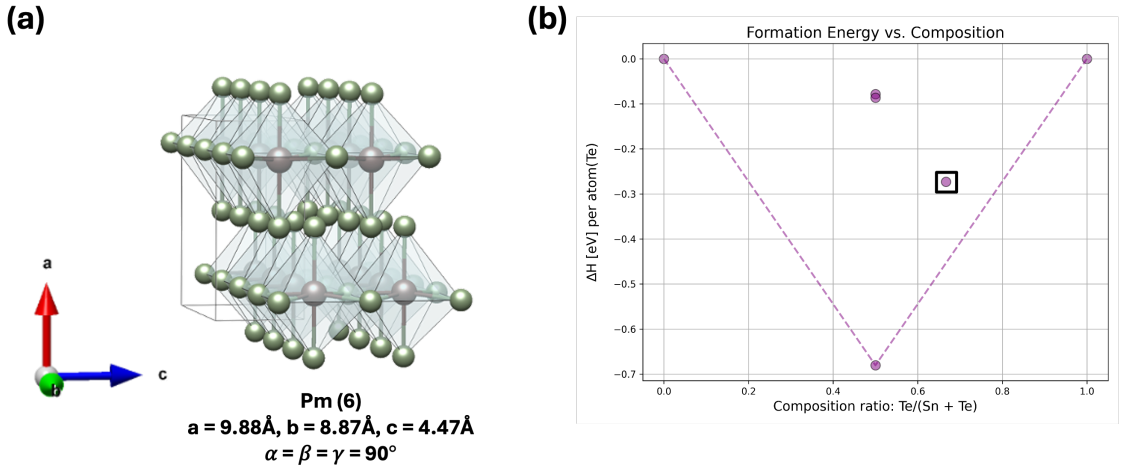

Figure 4.2: Predicted structure and convex hull of the $SnTe_2$ system. (a) The final predicted structure as determined by USPEX. (b) Convex hull of the Sn–Te system showing that $SnTe_2$ (highlighted point) does not lie on the hull.

4.2.2 $GeTe_2$

$GeTe_2$ is another binary combination formed from a Group IV–VI template. The structure prediction process is summarized in Figure 4.3, which shows the energy per atom, calculated using VASP, as a function of structure number. Each point in the figure represents a distinct structural configuration generated by USPEX. Over 400 candidate structures were explored, and the lowest energy configuration was selected based on USPEX's stopping criterion of eight consecutive generations with the same minimum energy structure.

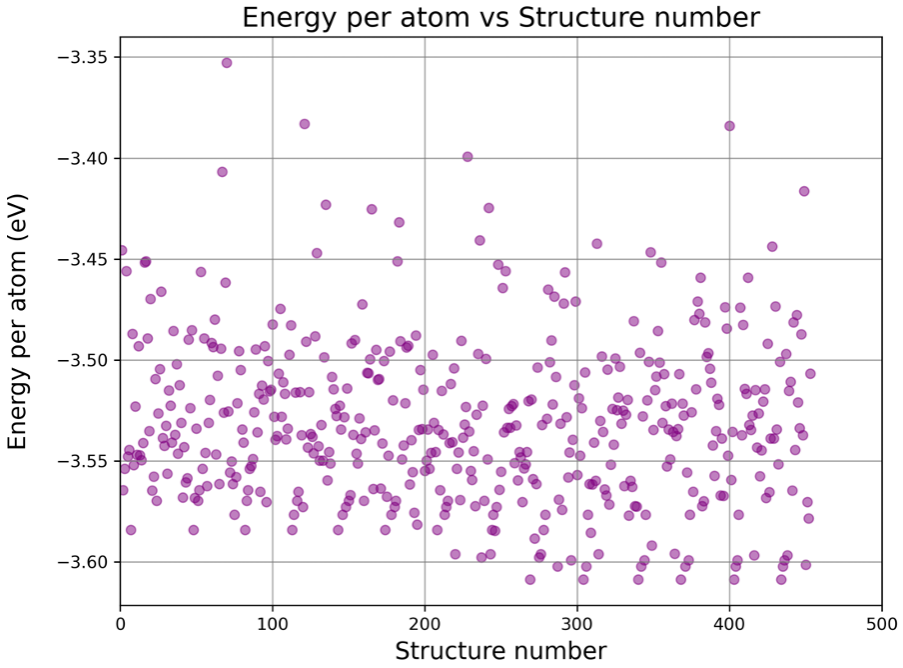

Figure 4.3: Energy per atom as a function of structure number for the $GeTe_2$ combination, obtained from fixed composition structure prediction using USPEX. The minimum energy structure was identified after analyzing over 400 structures.

The final predicted structure is shown in Figure 4.4(a). It crystallizes in the low-symmetry triclinic *P1* space group, reflecting a complex atomic arrangement with minimal symmetry constraints. The optimized lattice parameters are $a = 10.23$ Å, $b = 4.13$ Å, and $c = 11.88$ Å, with angles $\alpha = 89.77°$, $\beta = 112.14°$, and $\gamma = 89.22°$. The coordination environment of the Ge atoms consists of a mixture of $GeTe_4$ tetrahedra and $GeTe_6$ octahedra, forming a pseudo-layered framework. These polyhedra are connected through corner and edge sharing modes, creating an extended network. Average Ge–Te bond lengths in this predicted phase range from approximately 2.7–3.1 Å, which is notably longer than those found in experimentally observed Ge–chalcogenides in the Materials Project (e.g., GeO_2 (mp-470), GeS_2 (mp-572892), $GeSe_2$ (mp-540625)), where typical Ge–chalcogen bond lengths are 1.9–2.9 Å. These known compounds also feature similar networks of tetrahedral and octahedral polyhedra connected by corner and edge sharing modes.

When using the `FrameworkComparator` to compare this predicted phase against all known experimentally observed entries in the Materials Project, no structural matches were found within acceptable tolerances. This absence of matching frame-

works reduces the likelihood of the prediction, as it is unlikely that a simple binary phase would adopt an entirely unobserved structural framework. The lack of a corresponding entry for $GeTe_2$ in the Materials Project further supports the conclusion that this combination is unlikely to be realized.

Figure 4.4(b) shows the convex hull of the Ge–Te system, where $GeTe_2$ lies above the hull, indicating that the predicted phase is thermodynamically unstable relative to decomposition into competing phases. This observation is consistent with prior work describing $GeTe_2$ as a metastable compound prone to decomposition into GeTe and elemental Te, particularly under high temperature and high pressure conditions [96]. Notably, no stable crystalline bulk form of $GeTe_2$ has been synthesized. Compared to the $SnTe_2$ convex hull in Figure 4.2(b), the Ge–Te hull is noticeably shallower, which can indicate a reduced thermodynamic driving force for mixing of those two elements. Deeper hulls generally reflect a stronger tendency toward phase stability of the system. Taken together, these findings indicate that while Ge–Te compounds are technologically promising, the predicted bulk ground state structure of $GeTe_2$ lacks experimental validation, does not align with known frameworks, and is energetically unfavorable, making it unlikely to be realized in practical applications.

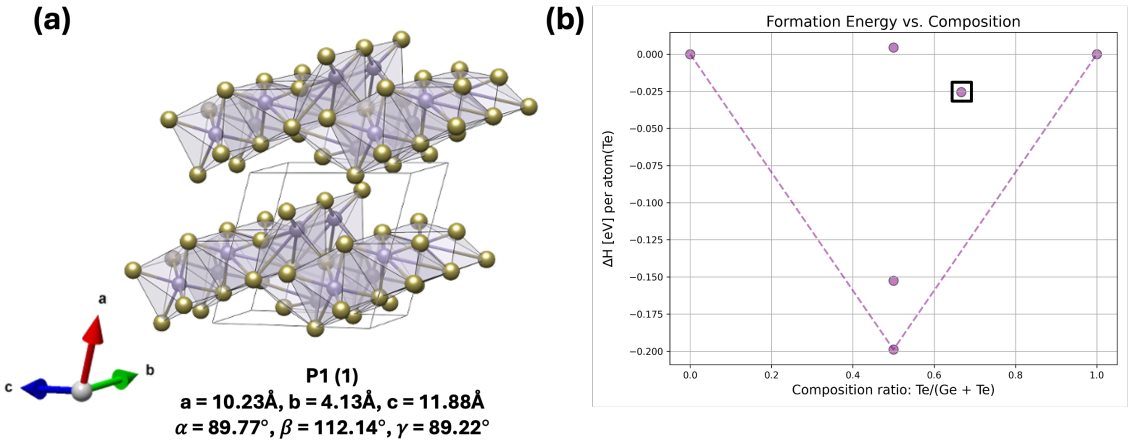

Figure 4.4: Predicted structure and convex hull of the $GeTe_2$ system. (a) The final predicted structure as determined by USPEX. (b) Convex hull of the Ge–Te system showing that $GeTe_2$ (highlighted point) does not lie on the hull.

4.2.3 AuI_3

AuI_3 is a binary combination formed from a Group I–VII template. The structure prediction process is shown in Figure 4.5, which plots the energy per atom as a function of structure number. Each point corresponds to a unique configuration generated during the evolutionary search in USPEX. The minimum energy structure was identified after analyzing over 400 structures, reaching USPEX's stopping criterion of maintining the same lowest enrgy structure for eight consecutive generations.

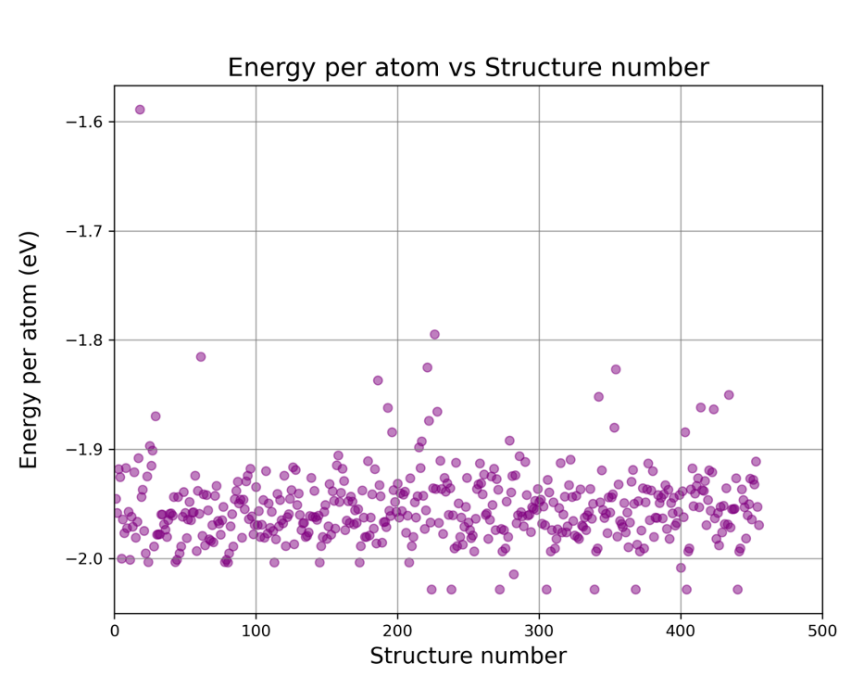

Figure 4.5: Energy per atom as a function of structure number for the AuI_3 combination, obtained from fixed composition structure prediction using USPEX. The minimum energy structure was identified after analyzing over 400 structures.

The final predicted structure is shown in Figure 4.6(a). It crystallizes in the triclinic *P1* space group, presenting a low symmetry configuration. The lattice parameters are $a = 6.78$ Å, $b = 7.04$ Å, and $c = 8.12$ Å, with angles $\alpha = 92.1°$, $\beta = 105.4°$, and $\gamma = 90.6°$. The coordination environment of gold consists of AuI_4 units in a distorted square planar geometry, forming a pseudo layered molecular crystal like framework. The Au–I bond lengths range from 2.63–2.77 Å, which aligns with the average bond lengths observed in other experimentally reported Au–halide trihalides, such as AuF_3 (mp-942), $AuCl_3$ (mp-27647), and $AuBr_3$ (mp-27213), where

typical bond lengths are approximately 2.2–2.6 Å. The last two compounds in particular, show the same molecular crystal like packing and square planar geometry coordination environment of the predicted AuI_3 structure, further supporting the plausibility of this phase.

When using the the `FrameworkComparator` to compare this predicted phase against all known experimentally observed entries in the Materials Project, no structural matches were found within acceptable tolerances. This absence of matching frameworks reduces the reliability of the prediction, as it is unlikely that a simple binary phase would adopt an entirely unobserved structural framework. The lack of a corresponding entry for AuI_3 in the Materials Project further suggests that this compound has not been realized in a stable crystalline form.

Figure 4.6(b) shows the convex hull of the Au–I system, where AuI_3 lies directly on the hull, indicating that the predicted phase is thermodynamically stable with respect to competing phases. However, prior studies have emphasized that AuI_3 is intrinsically unstable in the solid state, tending to decompose into AuI and I_2, which has prevented its successful synthesis as a bulk crystalline solid. Structurally, as mentioned earlier, AuI_3 adopts a molecular crystal–like arrangement in which discrete AuI_4 units are held together by relatively weak intermolecular forces instead of ionic bonding. Such weak interactions can be easily overcome at room temperature, which may explain prior reports of instability even though the phase is stable and lies on the hull. This is because temperature is not accounted for in DFT calculations, where it is effectively set to 0 K. Furthermore, gold was treated from the outset as adopting the +3 oxidation state, rather than the more common +1 state observed in many stable gold halides, which may also contribute to the instability of this phase. Taken together, these findings suggest that while the predicted phase is chemically reasonable and theoretically stable, its experimental validation may remain challenging, illustrating the need for caution when using computational structure prediction to assess whether a given combination is likely to be stable in reality and not only in theory.

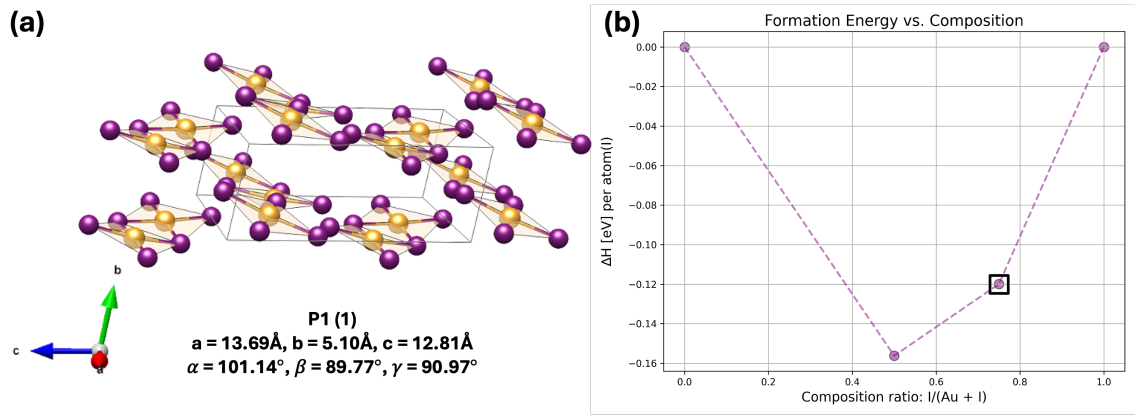

Figure 4.6: Predicted structure and convex hull of the AuI_3 system. (a) The final predicted structure as determined by USPEX. (b) Convex hull of the Au–I system

4.3 Summary

The results presented demonstrate the effectiveness of combining evolutionary structure prediction with DFT calculations for identifying viable chemical combinations. By systematically evaluating the structural and thermodynamic stability of $SnTe_2$, $GeTe_2$, and AuI_3, clear distinctions emerge between combinations with realistic synthesis potential and those that do not. While $SnTe_2$ and $GeTe_2$ were found to be thermodynamically unstable, this still constitutes a positive outcome, as these combinations can now be excluded from an already large list of combinations, allowing the search to focus on other promising options. In contrast, AuI_3 is predicted to be a stable compound, but additional inspections raise concerns about whether it can truly be realized as a bulk crystalline phase. It is also worth noting that the theoretical scope captured in the chosen INCAR settings may not be sufficient to fully describe all relevant effects. Incorporating more advanced treatments could provide greater confidence and more convincing predictions regarding the true stability of these phases.

The limited availability of certain elements likely plays a role in why these compounds have remained unexplored. Tellurium, for example, is a relatively rare element, which limits its use outside of highly targeted applications and may partially explain the lack of experimental efforts toward compounds such as $SnTe_2$ and $GeTe_2$. In contrast, gold and iodine have long been widely used elements, yet their combination as AuI_3 has seen limited exploration. It is possible that such combinations have been investigated but yielded negative results, which often remain unpublished, mak-

ing it difficult to assess their experimental history. Recently, however, the research community has taken steps to address this issue with platforms such as Scientific Reports, PLOS ONE and similar journals encouraging the publication of negative or null results. This suggests that the absence of these compounds from experimental databases is less about oversight and more about fundamental challenges in stabilizing them as bulk crystalline phases.

Taken together, these results highlight the value of data driven, computation guided materials discovery and provide a pathway for directing future experimental efforts toward unexplored, yet potentially synthesizable, materials.

5 Conclusions and Future Work

This project has explored and demonstrated practical approaches for simulating disordered alloys and previously unknown compounds using a range of computational tools that rely heavily on DFT. The primary goal of this work was to assess the viability and usefulness of these approaches, supported by experimental examination where possible. Studying such systems is often limited by the scope of existing databases and tools, which primarily focus on well ordered, experimentally known structures. This creates significant challenges when working with disordered alloys or compounds with unknown crystal structures, where atomic arrangements cannot be easily derived from existing data. Beyond identifying new materials, this work demonstrates that the workflow offers a legitimate and practical means of representing, analyzing, and gaining insight into complex disordered and unknown systems.

For disordered alloys, this work applied the SQS approach to model Cu–Mn solid solutions. This enabled the calculation of mixing enthalpies and provided insight into the system's thermodynamic behavior, including the tendency toward phase separation. These findings were consistent with experimental observations, demonstrating that this framework can reproduce and rationalize known material behavior while offering predictive capability for unexplored compositions. this approach is expected to have an even bigger impact when going to ternary alloys or ever high entropy alloys as the need to keep complex combinations managable becomes ever more challenging

For compounds with unknown crystal structures, the USPEX evolutionary algorithm was used to predict energetically favorable structures for selected chemical combinations, including $SnTe_2$, $GeTe_2$, and AuI_3. Analysis of these predicted structures relative to the convex hull revealed that AuI_3 lies on the hull, indicating its thermodynamic stability, while $SnTe_2$ and $GeTe_2$ occupy near hull positions, sug-

gesting they are accessible metastable phases at best, but to prove their existence further research is needed. These results show that this workflow can meaningfully guide the exploration of compounds that have yet to be synthesized or structurally characterized.

In summary, this thesis establishes a practical computational framework for addressing two key challenges in materials discovery. By combining established techniques with high throughput computational strategies, this work broadens the range of materials that can be investigated computationally and provides a validated pathway for guiding future experimental efforts. For simplicity, and as a demonstration case, the present study focused only on binary systems, which are well documented in the literature and offer limited options for truly novel discoveries. The next step is to extend this workflow to ternary systems, where the combinatorial space is much larger and the potential for uncovering new, technologically relevant materials is greater. Such an expansion will, however, bring additional challenges, including substantially higher computational demands, longer simulation times, and the need for more efficient screening strategies to navigate the large search space. Overcoming these challenges will be essential to realizing the full potential of this approach in the broader context of materials discovery.

References

[1] I. M. Ross, "The foundation of the silicon age," *Bell Labs technical journal*, vol. 2, no. 4, pp. 3–14, 1997.

[2] M. Ashby, "Materials–a brief history," *Philosophical Magazine Letters*, vol. 88, no. 9-10, pp. 749–755, 2008.

[3] F. Krafft, "Phosphorus. from elemental light to chemical element," *Angewandte Chemie International Edition in English*, vol. 8, no. 9, pp. 660–671, 1969.

[4] H. Malhotra, Vandana, S. Sharma, and R. Pandey, "Phosphorus nutrition: plant growth in response to deficiency and excess," in *Plant nutrients and abiotic stress tolerance*, pp. 171–190, Springer, 2018.

[5] A. King, "Materials by accident," *MRS Bulletin*, vol. 25, no. 1, pp. 88–88, 2000.

[6] D. Mendeleev, "The relation between the properties and atomic weights of the elements," *Journal of the Russian Chemical Society*, vol. 1, pp. 60–77, 1869. Originally published in Russian; English translation reprinted in various historical collections.

[7] M. Nespolo, "International tables for crystallography, volume a, space-group symmetry. edited by mois i. aroyo. wiley, 2016. pp. xxi+ 873. price gbp 295.00, eur 354.00 (hardcover). isbn 978-0-470-97423-0.," *Foundations of Crystallography*, vol. 73, no. 3, pp. 274–276, 2017.

[8] M. I. Aroyo, A. Kirov, C. Capillas, J. Perez-Mato, and H. Wondratschek, "Bilbao crystallographic server. ii. representations of crystallographic point groups and space groups," *Foundations of Crystallography*, vol. 62, no. 2, pp. 115–128, 2006.

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

[10] J. K. Stalick and A. D. Mighell, "Crystal data: Version 1.0 database specifications," *Final Report National Bureau of Standards*, 1986.

[11] G. Bergerhoff, R. Hundt, R. Sievers, and I. D. Brown, "The inorganic crystal structure data base," *Journal of Chemical Information and Computer Sciences*, vol. 23, no. 2, pp. 66–69, 1983.

[12] M. L. Cohen and T. Bergstresser, "Band structures and pseudopotential form factors for fourteen semiconductors of the diamond and zinc-blende structures," *Physical Review*, vol. 141, no. 2, p. 789, 1966.

[13] P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, *et al.*, "Quantum espresso: a modular and open-source software project for quantum simulations of materials," *Journal of physics: Condensed matter*, vol. 21, no. 39, p. 395502, 2009.

[14] S. P. Ong, W. D. Richards, A. Jain, G. Hautier, M. Kocher, S. Cholia, D. Gunter, V. L. Chevrier, K. A. Persson, and G. Ceder, "Python materials genomics (pymatgen): A robust, open-source python library for materials analysis," *Computational Materials Science*, vol. 68, pp. 314–319, 2013.

[15] A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, and K. a. Persson, "The Materials Project: A materials genome approach to accelerating materials innovation," *APL Materials*, vol. 1, no. 1, p. 011002, 2013.

[16] S. Curtarolo, W. Setyawan, S. Wang, J. Xue, K. Yang, R. H. Taylor, L. J. Nelson, G. L. Hart, S. Sanvito, M. Buongiorno-Nardelli, *et al.*, "Aflowlib. org: A distributed materials properties repository from high-throughput ab initio calculations," *Computational Materials Science*, vol. 58, pp. 227–235, 2012.

[17] M. K. Horton, P. Huck, R. X. Yang, J. M. Munro, S. Dwaraknath, A. M. Ganose, R. S. Kingsbury, M. Wen, J. X. Shen, T. S. Mathis, *et al.*, "Accelerated data-

driven materials science with the materials project," *Nature Materials*, pp. 1–11, 2025.

[18] C. Levinthal, "How to fold graciously," *Mossbauer spectroscopy in biological systems*, vol. 67, pp. 22–24, 1969.

[19] H. D. Jumper John, "Protein structure predictions to atomic accuracy with alphafold," *Nature methods*, vol. 19, no. 1, pp. 11–12, 2022.

[20] N. H. March, W. H. Young, and S. Sampanthar, *The many-body problem in quantum mechanics*. Courier Corporation, 1995.

[21] D. Griffiths, *Introduction to elementary particles*. John Wiley & Sons, 2020.

[22] L. Grave de Peralta, "Exact quasi-relativistic wavefunctions of hydrogen-like atoms," *Scientific Reports*, vol. 10, no. 1, p. 14925, 2020.

[23] E. Schrödinger, "Quantisierung als eigenwertproblem," *Annalen der physik*, vol. 385, no. 13, pp. 437–490, 1926.

[24] M. Born and R. Oppenheimer, "Zur Quantentheorie der Molekeln," *Annalen der Physik*, vol. 389, no. 20, pp. 457–484, 1927. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/andp.19273892002.

[25] F. Lenz, "The ratio of proton and electron masses," *Physical Review*, vol. 82, no. 4, p. 554, 1951.

[26] D. R. Hartree, "The wave mechanics of an atom with a non-coulomb central field. part i. theory and methods," in *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 24, pp. 89–110, Cambridge university press, 1928.

[27] W. Pauli, "Über den zusammenhang des abschlusses der elektronengruppen im atom mit der komplexstruktur der spektren," *Zeitschrift für Physik*, vol. 31, no. 1, pp. 765–783, 1925.

[28] J. C. Slater, "The theory of complex spectra," *Physical review*, vol. 34, no. 10, p. 1293, 1929.

[29] P. Hohenberg and W. Kohn, "Inhomogeneous electron gas," *Physical Review B*, vol. 136, no. 3B, pp. B864–B871, 1964.

[30] W. Kohn and L. J. Sham, "Self-consistent equations including exchange and correlation effects," *Physical review*, vol. 140, no. 4A, p. A1133, 1965.

[31] J. G. Lee, *Computational materials science: an introduction*. CRC press, 2016.

[32] E. Fermi, "Eine statistische methode zur bestimmung einiger eigenschaften des atoms und ihre anwendung auf die theorie des periodischen systems der elemente," *Zeitschrift für Physik*, vol. 48, no. 1, pp. 73–79, 1928.

[33] L. H. Thomas, "The calculation of atomic fields," *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 23, no. 5, pp. 542–548, 1927.

[34] R. G. Parr, "Density functional theory of atoms and molecules," in *Horizons of Quantum Chemistry: Proceedings of the Third International Congress of Quantum Chemistry Held at Kyoto, Japan, October 29-November 3, 1979*, pp. 5–15, Springer, 1989.

[35] P. A. M. Dirac, "Note on exchange phenomena in the thomas atom," *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 26, no. 3, pp. 376–385, 1930.

[36] E. H. Lieb, "Thomas-fermi and related theories of atoms and molecules," *Reviews of Modern Physics*, vol. 53, no. 4, p. 603, 1981.

[37] J. P. Perdew and A. Zunger, "Self-interaction correction to density-functional approximations for many-electron systems," *Physical review B*, vol. 23, no. 10, p. 5048, 1981.

[38] H. Jiang, R. I. Gomez-Abal, P. Rinke, and M. Scheffler, "First-principles modeling of localized d states with the gw@ lda+ u approach," *Physical Review B-Condensed Matter and Materials Physics*, vol. 82, no. 4, p. 045108, 2010.

[39] A. Van de Walle and G. Ceder, "Correcting overbinding in local-density-approximation calculations," *Physical Review B*, vol. 59, no. 23, p. 14992, 1999.

[40] J. G. Lee, *Computational Materials Science: An Introduction.* Boca Raton: CRC Press, 1st ed., 2011.

[41] J. E. Jaffe, J. A. Snyder, Z. Lin, and A. C. Hess, "Lda and gga calculations for high-pressure phase transitions in zno and mgo," *Physical Review B*, vol. 62, no. 3, p. 1660, 2000.

[42] P. Ziesche, S. Kurth, and J. P. Perdew, "Density functionals from lda to gga," *Computational materials science*, vol. 11, no. 2, pp. 122–127, 1998.

[43] J. P. Perdew and Y. Wang, "Accurate and simple analytic representation of the electron-gas correlation energy," *Physical review B*, vol. 45, no. 23, p. 13244, 1992.

[44] J. P. Perdew, K. Burke, and M. Ernzerhof, "Generalized gradient approximation made simple," *Physical review letters*, vol. 77, no. 18, p. 3865, 1996.

[45] H. Hellman, "Einführung in die quantenchemie," *Franz Deuticke, Leipzig*, vol. 285, p. 90, 1937.

[46] R. P. Feynman, "Forces in molecules," *Physical review*, vol. 56, no. 4, p. 340, 1939.

[47] F. Bloch, "Über die quantenmechanik der elektronen in kristallgittern," *Zeitschrift für physik*, vol. 52, no. 7, pp. 555–600, 1929.

[48] R. Islam, K. P. Hazarika, and J. Borah, "Computational techniques for nanostructured materials," in *Handbook of Magnetic Hybrid Nanoalloys and their Nanocomposites*, pp. 1–22, Springer, 2022.

[49] P. E. Blöchl, "Projector augmented-wave method," *Physical review B*, vol. 50, no. 24, p. 17953, 1994.

[50] G. Kresse and J. Hafner, "Ab initio molecular dynamics for liquid metals," *Physical review B*, vol. 47, no. 1, p. 558, 1993.

[51] G. Kresse and J. Hafner, "Ab initio molecular-dynamics simulation of the liquid-metal–amorphous-semiconductor transition in germanium," *Physical Review B*, vol. 49, no. 20, p. 14251, 1994.

[52] G. Kresse and J. Furthmüller, "Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set," *Computational materials science*, vol. 6, no. 1, pp. 15–50, 1996.

[53] G. Kresse and J. Furthmüller, "Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set," *Physical review B*, vol. 54, no. 16, p. 11169, 1996.

[54] Y. Hinuma, T. Hatakeyama, Y. Kumagai, L. A. Burton, H. Sato, Y. Muraba, S. Iimura, H. Hiramatsu, I. Tanaka, H. Hosono, *et al.*, "Discovery of earth-abundant nitride semiconductors by computational screening and high-pressure synthesis," *Nature communications*, vol. 7, no. 1, p. 11962, 2016.

[55] G. Kresse and J. Hafner, "Norm-conserving and ultrasoft pseudopotentials for first-row and transition elements," *Journal of Physics: Condensed Matter*, vol. 6, no. 40, p. 8245, 1994.

[56] G. Kresse and D. Joubert, "From ultrasoft pseudopotentials to the projector augmented-wave method," *Physical review b*, vol. 59, no. 3, p. 1758, 1999.

[57] A. R. Oganov and C. W. Glass, "Crystal structure prediction using ab initio evolutionary techniques: Principles and applications," *The Journal of chemical physics*, vol. 124, no. 24, 2006.

[58] A. R. Oganov, A. O. Lyakhov, and M. Valle, "How evolutionary crystal structure prediction works and why," *Accounts of chemical research*, vol. 44, no. 3, pp. 227–237, 2011.

[59] A. O. Lyakhov, A. R. Oganov, H. T. Stokes, and Q. Zhu, "New developments in evolutionary structure prediction algorithm uspex," *Computer Physics Communications*, vol. 184, no. 4, pp. 1172–1182, 2013.

[60] A. Zunger, S.-H. Wei, L. G. Ferreira, and J. E. Bernard, "Special quasirandom structures," *Physical review letters*, vol. 65, no. 3, p. 353, 1990.

[61] A. van de Walle, "Multicomponent multisublattice alloys, nonconfigurational entropy and other additions to the Alloy Theoretic Automated Toolkit," *Calphad*, vol. 33, pp. 266–278, 2009.

[62] A. van de Walle, M. D. Asta, and G. Ceder, "The Alloy Theoretic Automated Toolkit: A user guide," *Calphad*, vol. 26, pp. 539–553, 2002.

[63] A. van de Walle and G. Ceder, "Automating first-principles phase diagram calculations," *J. Phase Equilib.*, vol. 23, pp. 348–359, 2002.

[64] A. van de Walle and M. D. Asta, "Self-driven lattice-model monte carlo simulations of alloy thermodynamic properties and phase diagrams," *Model. Simul. Mater. Sc.*, vol. 10, p. 521, 2002.

[65] A. van de Walle, P. Tiwary, M. M. de Jong, D. L. Olmsted, M. D. Asta, A. Dick, D. Shin, Y. Wang, L.-Q. Chen, and Z.-K. Liu, "Efficient stochastic generation of special quasirandom structures," *Calphad*, vol. 42, pp. 13–18, 2013.

[66] A. van de Walle, "Methods for first-principles alloy thermodynamics," *JOM - J. Min. Met. Mat. S.*, vol. 65, pp. 1523–1532, 2013.

[67] A. van de Walle, "A complete representation of structure-property relationships in crystals," *Nat. Mater.*, vol. 7, pp. 455–458, 2008.

[68] A. van de Walle and D. Ellis, "First-principles thermodynamics of coherent interfaces in samarium-doped ceria nanoscale superlattices," *Phys. Rev. Lett.*, vol. 98, p. 266101, 2007.

[69] E. Cockayne and A. van de Walle, "Building effective models from scarce but accurate data: Application to an alloy cluster expansion model," *Phys. Rev. B*, vol. 81, p. 012104, 2010.

[70] A. van de Walle, B. G. Chirranjeevi, S. Demers, Q.-J. Hong, A. Kowalski, L. Miljacic, G. S. Pomrehn, and P. Tiwary, "Ab initio calculation of anisotropic interfacial excess free energies," *Phys. Rev. B*, vol. 89, p. 184101, 2014.

[71] A. van de Walle, Q.-J. Hong, S. Kadkhodaei, and R. Sun, "The free energy of mechanically unstable phases," *Nature Commun.*, vol. 6, p. 7559, 2015.

[72] A. van de Walle, S. Kadkhodaei, R. Sun, and Q.-J. Hong, "Epicycle method for elasticity limit calculations," *Phys. Rev. B*, vol. 95, p. 144113, 2017.

[73] R. Sun and A. van de Walle, "Automating impurity-enhanced antiphase boundary energy calculations from ab initio monte carlo," *Calphad*, vol. 53, p. 20, 2016.

[74] A. van de Walle, R. Sun, Q.-J. Hong, and S. Kadkhodaei, "Software tools for high-throughput calphad from first-principles data," *Calphad*, vol. 58, p. 70, 2017.

[75] A. van de Walle, "Reconciling SGTE and ab initio enthalpies of the elements," *Calphad*, vol. 60, p. 1, 2018.

[76] H. Liu and A. van de Walle, "Rapid geometric screening of low-energy surfaces in crystals," *Symmetry*, vol. 14, p. 2067, 2022.

[77] D. P. Kroese and R. Y. Rubinstein, "Monte carlo methods," *Wiley Interdisciplinary Reviews: Computational Statistics*, vol. 4, no. 1, pp. 48–58, 2012.

[78] H. Yang, Z. Ma, C. Lei, L. Meng, Y. Fang, J. Liu, and H. Wang, "High strength and high conductivity cu alloys: A review," *Science China Technological Sciences*, vol. 63, no. 12, pp. 2505–2517, 2020.

[79] G. R. Thellaputta, P. S. Chandra, and C. Rao, "Machinability of nickel based superalloys: a review," *Materials Today: Proceedings*, vol. 4, no. 2, pp. 3712–3721, 2017.

[80] H.-Y. Lee, C.-W. He, Y.-C. Lee, and D.-C. Wu, "A study on the characteristics of cu–mn–dy alloy resistive thin films," *Coatings*, vol. 9, no. 2, p. 118, 2019.

[81] M. Hansen and K. Anderko, *Constitution of Binary Alloys*. New York: McGraw-Hill, 1958.

[82] N. Cowlam and A. M. Shamah, "A search for the ordered phases cu5mn, cu3mn and cumn by neutron diffraction," *Journal of Physics F: Metal Physics*, vol. 10, no. 7, p. 1357, 1980.

[83] S. Sarhan, N. Moav, A. Azulay, D. Levy, L. A. Burton, and H. Bishara, "Spinodal decomposition, phase stability, and electrical resistivity of cu-rich cu-mn alloys," *Acta Materialia*, 2025. Manuscript submitted for publication.

[84] H. Okamoto, "Supplemental literature review of binary phase diagrams: Ag-ni, ag-zr, au-bi, b-ni, co-sb, cu-mn, cu-si, cu-zn, fe-zr, li-sb, mg-pu, and si-zr," *Journal of Phase Equilibria and Diffusion*, vol. 39, pp. 87–100, 2018.

[85] F. Findik, "Improvements in spinodal alloys from past to present," *Materials & Design*, vol. 42, pp. 131–146, 2012.

[86] D. W. Davies, K. T. Butler, A. J. Jackson, A. Morris, J. M. Frost, J. M. Skelton, and A. Walsh, "Computational screening of all stoichiometric inorganic materials," *Chem*, vol. 1, no. 4, pp. 617–627, 2016.

[87] Y. Ding, Y. Kumagai, F. Oba, and L. A. Burton, "Data-mining element charges in inorganic materials," *The Journal of Physical Chemistry Letters*, vol. 11, no. 19, pp. 8264–8267, 2020. PMID: 32852211.

[88] M. Dolg and M. Dolg, *Computational methods in lanthanide and actinide chemistry*. Newark: WILEY, first edition. ed., 2015.

[89] N. Dainiak, J. K. Waselenko, J. O. Armitage, T. J. MacVittie, and A. M. Farese, "The hematologist and radiation casualties," *ASH Education Program Book*, vol. 2003, no. 1, pp. 473–496, 2003.

[90] M. Grundmann, *The physics of semiconductors*, vol. 13. Springer, 2006.

[91] I. Ansara, C. Chatillon, H. Lukas, T. Nishizawa, H. Ohtani, K. Ishida, M. Hillert, B. Sundman, B. Argent, A. Watson, *et al.*, "A binary database for iii–v compound semiconductor systems," *Calphad*, vol. 18, no. 2, pp. 177–222, 1994.

[92] O. Madelung, *Semiconductors: data handbook*. Springer Science & Business Media, 2004.

[93] M. Balkanski and R. F. Wallis, *Semiconductor physics and applications*, vol. 8. Oxford University Press, 2000.

[94] Y. Wang, Z. Gao, and J. Zhou, "Ultralow lattice thermal conductivity and electronic properties of monolayer 1t phase semimetal site2 and snte2," *Physica E: Low-dimensional Systems and Nanostructures*, vol. 108, pp. 53–59, 2019.

[95] L.-H. Qu, J. Yu, Y.-L. Mu, X.-L. Fu, C.-G. Zhong, Y. Min, P.-X. Zhou, J.-M. Zhang, Y.-Q. Zou, and T.-S. Lu, "Strain tunable structural, mechanical and electronic properties of monolayer tin dioxides and dichalcogenides snx2 (xo, s, se, te)," *Materials Research Bulletin*, vol. 119, p. 110533, 2019.

[96] T. Usuki, C. J. Benmore, A. Tverjanovich, S. Bereznev, M. Khomenko, A. Sokolov, D. Fontanari, K. Ohara, M. Bokova, M. Kassem, *et al.*, "Atomic structure and dynamics of unusual and wide-gap phase-change chalcogenides: A gete2 case," *physica status solidi (RRL)–Rapid Research Letters*, vol. 18, no. 10, p. 2300482, 2024.

Appendix A

Table A.1: Representative oxidation states for each element as compiled from previous work based on the ICSD, along with the single oxidation state selected in this study for combinatorial filtering. Oxidation states with multiple entries are shown as combined values separated by slashes.

Element	Oxidation State	Element	Oxidation State	Element	Oxidation State	Element	Oxidation State	Chosen Oxidation State
H	+1	Cr	+3	In	+3	Hf	+4	
Li	+1	Mn	+2	Sb	+3	Ta	+5	
Be	+2	Co	+2	Te	-2	W	+6	
B	+3	Ni	+2	I	-1	Re	+3	
C	-4	Zn	+2	Cs	+1	Os	+4	
N	-3	Ga	+3	Ba	+2	Ir	+4	
O	-2	Ge	+4	La	+3	Hg	+2	
F	-1	Se	-2	Ce	+3	Tl	+1	
Na	+1	Br	-1	Pr	+3	Pb	+2	
Mg	+2	Rb	+1	Nd	+3	Bi	+3	
Al	+3	Sr	+2	Pm	+3	Fe	+2/+3	+3
Si	+4	Y	+3	Sm	+3	Cu	+1/+2	+2
P	+5	Zr	+4	Gd	+3	As	-3/+5	+5
S	-2	Nb	+5	Tb	+3	Sn	+2/+4	+4
Cl	-1	Mo	+6	Dy	+3	Eu	+2/+3	+3
K	+1	Ru	+4	Ho	+3	Pt	+2/+4	+4
Ca	+2	Rh	+3	Er	+3	Au	+1/+3	+3
Sc	+3	Pd	+2	Tm	+3			
Ti	+4	Ag	+1	Yb	+3			
V	+5	Cd	+2	Lu	+4			

```python
#!/usr/bin/env python3
from itertools import combinations_with_replacement
from collections import OrderedDict
from typing import Sequence

def all_electroneutral_combinations(charges: Sequence[tuple], n_ions: int) -> list:
"""
Given a list of (element, charge) pairs, return all combinations of n_ions
that sum to a total charge of zero.
"""
results = []
for tup in combinations_with_replacement(charges, n_ions):
charge_values = [int(item[1]) for item in tup]
if sum(charge_values) == 0:
element_list = [item[0] for item in tup]
results.append(element_list)
return results

def compound_creator(elements: list) -> str:
"""
Build a chemical formula string from a list of elements.
Example: ['Na', 'Cl', 'Cl'] → 'NaCl2'
"""
counts = OrderedDict((el, elements.count(el)) for el in elements)
formula = ""
for el, count in counts.items():
formula += el
if count > 1:
formula += str(count)
return formula

# === USER CONFIGURATION ===
base_dir = "{path_to_project}/" # e.g., "/home/user/project/"
input_filename = "oxidation_states.txt" # file containing element-charge pairs
n_ions = {n_ions} # number of ions in each combination
output_filename = f"Neutral_combinations_{n_ions}_ions.txt"

# === LOAD CHARGES FROM FILE ===
with open(base_dir + input_filename, 'r') as f:
# Example of how oxidation_states.txt should look:
# Na +1
# Cl -1
# Ca +2
# O -2
# Each line: Element <whitespace> Charge
charges = [line.strip().split() for line in f if line.strip()]

# === GENERATE ALL ELECTRONEUTRAL COMBINATIONS ===
neutral_sets = [compound_creator(c) for c in all_electroneutral_combinations(charges, n_ions)]
unique_formulas = sorted(set(neutral_sets))

# === SAVE RESULTS ===
with open(base_dir + output_filename, 'w') as f:
f.writelines(formula + "\n" for formula in unique_formulas)

print(f"Generated {len(unique_formulas)} unique neutral combinations (saved to {output_filename}).")
```

Figure A.1: Python script for generating all electroneutral combinations of ions.

תקציר

מחקר זה עוסק באתגרים החישוביים במידול חומרים חסרי סדר, כגון סגסוגות ותמיסות מוצקות, באמצעות שימוש בגישת המבנים הכמעט אקראיים המיוחדים (SQS) לייצוג אי־סדר בתוך תאים מחזוריים סופיים. המיקוד היה במערכת הנחושת-מנגן (Cu-Mn), תוך שימוש בשיטת SQS ובשיטת החלפה אקראית של מיקומי האטומים ליצירת קונפיגורציות המייצגות אי סדר מקסימלי במגוון תחומי הרכב. חישובי תורת פונקציונל הצפיפות (DFT) הצביעו על אנתלפיות היווצרות חיוביות בלבד בכל הקונפיגורציות, דבר המרמז כי מערכת ה־Cu-Mn אינה יציבה כתמיסה מוצקה הומוגנית, ובכך מערער את אמינותם של דיאגרמות פאזה המופיעות בספרות. חקר *ab initio* באמצעות אלגוריתמים אבולוציוניים לניבוי מבנים בכל תחומי ההרכב בחן את המרחב האנרגטי של המערכת, אך גם בשיטה זו הממצאים תמכו בפירוק מונע תרמודינמית, מאחר שלא נמצאו פאזות יציבות. מעבר למערכת Cu-Mn, הורחב השימוש בגישות אלו לגילוי חומרים חדשים. נוצרו למעלה ממיליון מועמדים ראשוניים, ומתוכם נבחרו שלושה מוליכים למחצה בינאריים שטרם דווחו בספרות, אשר הוצעו כמועמדים לאימות ניסיוני עתידי.

אוניברסיטת תל - אביב

הפקולטה להנדסה ע"ש איבי ואלדר פליישמן
בית הספר לתארים מתקדמים ע"ש זנדמן-סליינר

סימולציה של חומרים בלתי-מוכרים וחסרי סדר מבני

חיבור זה הוגש כעבודת גמר לקראת התואר
"מוסמך אוניברסיטה" במדע והנדסה של חומרים

על - ידי

נדב מואב

העבודה נעשתה במחלקה למדע והנדסה של חומרים
בהנחיית ד"ר לי ברטון

תשרי תשפ״ה