NORTHWESTERN UNIVERSITY

Data-Driven High-Throughput Materials Discovery and Knowledge Extraction

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Materials Science and Engineering

By

Vinay Ishwar Hegde

EVANSTON, ILLINOIS

December 2020

ABSTRACT

Data-Driven High-Throughput Materials Discovery
and Knowledge Extraction

Vinay Ishwar Hegde

Recent advances in high-performance computing have resulted in massive databases of materials properties calculated with techniques such as ab initio density functional theory. In fact, some of the largest of such databases have calculations of nearly all distinct, ordered, experimentally-reported compounds. This thesis discusses the application of data in one such database, the Open Quantum Materials Database (OQMD), along three broad fronts. (a) High-throughput materials discovery: prediction of hitherto unreported ternary oxyfluorides and Heusler-based compounds using techniques such as prototype decoration and cluster expansion to generate novel hypothetical structures, in conjunction with convex hull-based phase stability analysis. (b) Fingerprinting the high-pressure materials genome: exploration of the enthalpy landscape of all materials at high-pressure using a simple linear approximation for compound enthalpy, and ambient-pressure energy and volume data in the OQMD; using the framework to predict novel high-pressure-stable

phases in ambient-immiscible systems. (c) Network representation of all materials: a top-down view of phase diagrams through the lens of complex network theory, unraveling the complete thermodynamic network of all materials, and using it to extract otherwise-intractable knowledge—a quantitative scale of material reactivity, the "nobility index"—and thereby identifying the noblest materials in nature.

Acknowledgements

First and foremost, Chris, thank you for being a fantastic mentor and guide over this journey, and this has been a journey indeed. Thank you for being ever so patient, and for giving me the freedom to take research "detours" throughout grad school. I always marveled at (and continue to marvel at) the keen insights you so readily bring to every discussion, and the great clarity with which you impart it to others. It really is inspiring. Thank you also for Mimi and Glitter—I will miss them.

I have collaborated with several people over the past five years, both within the Wolverton group and outside it. I want to thank all those collaborators for making the experiences truly enjoyable. Thank you for all the wonderful discussions, and bearing with my nitpicking and nagging.

My family—Amma, Ubbi, Eekmaava, Atte—for being understanding through all the near-radio silences. Despite the frequent, "when are you graduating?", "just complete it already!", "how many more years are you going to keep *studying*?!", you have been extremely supportive of me over the years. I would not be where I am today without you. Thanks also, of course, to Sohan, Sadhu, Iggychops, and Anvesh—you light up my heart.

Thanks to all my labmates and friends over the years. Tony, Max, and Shahab—I am grateful for your friendship. Jiangang, Eric, and Koushik—I will miss all the lunchtime banter. Shash and Koushik—is there any substitute for our impromptu Devon runs? Partha, Kritish, and especially Amit—thanks for being tolerable roommates (ha! not

letting you off that easy). Chaiti, Mer, Raj, and Datta—our Whatsapp group really kept me sane. Adithi and Sharan—you made Chicago so much more memorable. Lakumi—your (wireless) companionship steadied me in the darkest of times. I'm deeply indebted to all of you.

I have two more shout-outs: first, the graduate MSE batch of '13. You all made me feel welcome, and made NU feel like home away from home. Second, the administrative staff in the MSE office, and Jeannine and Emily in particular. You have been unbelievably helpful over the years. A heartfelt thanks to you all.

To used Schwinns and The IceFrog,

to Chicago and the winds of winter,

and as always, to you, Amma

Table of Contents

List of Tables

List of Figures

CHAPTER 1

Introduction and motivation

The Materials Genome Initiative (MGI) for Global Competitiveness was launched in 2011 with a vision to at least double the pace of discovery, development, and deployment of advanced materials, at a fraction of current costs [1]. The MGI recognized that a close integration of computational, experimental, and data informatics tools is central to achieving its vision (Figure 1.1). On the topic of materials discovery in specific, the initiative identified the following: (a) the traditional process of materials discovery has been largely based on scientific intuition and Edisonian trial-and-error experiments—often expensive and time-consuming, and (b) state of the art computational tools can

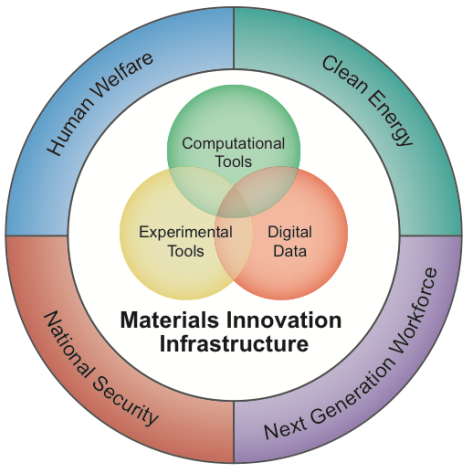

Figure 1.1. An overview of the vision of Materials Genome Initiative (MGI); central to the vision is the close integration of computational, experimental and digital data tools for the acceleration of materials discovery and development. (Source: Ref. [1])

perform virtual experiments *in silico* and greatly aid in guiding, and reducing the need for, physical experimentation. Indeed, computational materials discovery and design has become fairly widespread, across a variety of material types and applications, and is a rapidly burgeoning field.

The tools and techniques employed for computational materials discovery are many and varied, and depend on the class of materials being studied and their properties. Nonetheless, all the current approaches can be broadly classified into the following three emergent paradigms:

A. High-throughput databases and data mining:

This approach is characterized by the combination of computational quantum-mechanical–thermodynamic approaches and techniques rooted in database construction and intelligent mining of data. The practical implementation of computational high-throughput (HT) is highly non-trivial and involves the following connected steps: (i) materials growth: thermodynamic and electronic structure calculations of hundreds of thousands of materials, both currently known *and* hypothetical, with a consistent, transferable set of parameters across all calculations, (ii) materials storage: systematic storage of the previously-calculated materials properties in databases that enable quick and structured retrieval of data, and (iii) materials selection: thermodynamic analysis and/or intelligent mining of the previously-stored data to identify novel materials with the desired properties and/or gain new physical insights into existing materials.

Due to rapid advances in high-performance computing, several large HT computational databases have emerged over the past decade: the Open Quantum Materials Database (OQMD) [19, 20], the Materials Project [21], the AFLOWLIB.org consortium [22], are among the prominent ones. While density functional theory (DFT) is the common workhorse for the atomistic calculation of materials properties in most such HT databases, the two key features differentiating them are the number of materials calculated, and the variety of properties evaluated for each material. For instance, both the OQMD and the Materials Project have calculated most of the currently reported inorganic compounds, ∼50,000 unique ordered ones in total, sourced from a database such as the Inorganic Crystal Structure Database (ICSD) [23]. However, the OQMD has more than 400,000 hypothetical compounds generated by decorating common structural prototypes (see Sec. 3.1.3, and Ref. [24] for details of the decoration of prototypes approach), while the Materials Project has fewer, close to 20,000 hypothetical compounds. On the other hand, the OQMD has calculated properties such as formation energy, phase stability, electronic density of states, magnetic moment, etc., for all the materials in the database, while the Materials Project, in addition to those properties, has more complex quantities calculated for a subset of compounds, e.g., band structures for ∼54,000, elastic tensors for ∼4,700, and piezoelectric tensors for ∼1,000 compounds.

Such HT DFT databases have been used to search for materials for a variety of applications: new battery electrodes [25–27], electrode coatings [28–30], water-splitting materials [31, 32], strengthening precipitates in alloys [24, 33, 34], thermoelectrics [35–37], transparent conducting oxides [38–40], photovoltaics [39, 41], scintillators [42, 43], topological insulators [44, 45], and others.

B. Global structural searches

The HT DFT approach involving involving databases with decorations of known structural prototypes has a limited success rate when the ground state structure is uncommon or previously unknown. More advanced methods for crystal structure prediction exist which allow for a systematic search of the entire potential energy surface of the system. Such structural search methods can be broadly distinguished by the level of restriction imposed on the structural degrees of freedom being optimized: (a) neither the underlying lattice type nor the atomic configuration is known, and (b) the underlying Bravais lattice is known but the atomic configuration is unknown.

Some of the commonly used methods in the former category (both lattice type and atomic configuration unknown) are: (a) Simulated Annealing [46] — a Monte Carlo technique in which a random new state close to the current solution is accepted with a probability of 1 if the new state is lower in energy, and a probability $e^{\frac{-\Delta E}{k_B T}}$, where ΔE is the difference in energy, k_B is the Boltzmann constant, and T is temperature, if its energy is higher. (b) Metadynamics [47] — a steepest-descent-like discrete evolution of the history-dependent Gibbs potential representing the potential energy landscape in order parameters (usually, the lattice vectors) space. (c) Minima Hopping [48] — the random steps of simulated annealing are replaced by local optimization to find the local minimum and molecular dynamics based moves are used to escape it via low-energy saddle points. (d) Random Structure Searching [49] — using random but sensible initial configuration of atoms-in-a-box, and "shaking" such structures so that they "fall" into nearby minima. (e) Evolutionary Algorithms [50, 51] — an initial population of structures, each with a fitness (usually energy or enthalpy), is used to derive a new generation of structures based

on genetic principles of heredity and mutation of parent structures followed by selection, and the entire process is repeated. Such structural methods have been used to successfully predict the cystal structure of a variety of materials, e.g., framework materials [52], benzene [53], polymorphs of $MgSiO_3$ [54], silicon and gold clusters [55, 56], high-pressure phases of silane [49], lithium alloys [49, 57], stable and metastable phases of $CaCO_3$ [51], and many others.

In the latter category of problems where the underlying lattice type is known but the atomic configuration is unknown, perhaps the most common optimization method of choice is cluster expansion [58], a technique based on a generalized Ising-like Hamiltonian, using interaction energies extracted for a set of unique sub-configuration of atoms ("clusters") to identify the ground state structures (see Section 2.3 for a brief formalism of the method as employed in this work). Cluster expansion based methods have been used to predict the ground state structures of II-III spinel oxides [59], vacancy order-disorder in battery electrodes [60], structural order in Heusler alloys [61], ground state of adsorbed species on substrates [62], a variety of intermetallic systems [63–67], among others.

C. Machine learning

This paradigm of materials discovery is characterized by the use of machine learning (ML) techniques to extract knowledge from large materials datasets. Approaches under this umbrella typically have the following three components: (a) source of, preferably curated and consistent, materials properties data, (b) quantitative representation of each material in the dataset (often referred to as "attributes" or "descriptors"), and (c) ML algorithm(s) to be trained on the data. A host of ML algorithms, ranging from linear and

ridged regression, principal component analysis, decision trees, support vectors machines, neural networks, etc., are used depending on the nature of the dataset available and the property under study. However, the distinctive feature of each approach is arguably its second component, i.e., quantitative representation of a material—composed of descriptors based solely on the composition or, in addition, the crystal and electronic structure of a material. A few relevant examples of the application of ML techniques include predicting the crystal structure of binary alloys [68–70], classification of wide band gap AB compounds [71], composition-only models of thermodynamic stability of ternary compounds [72], data-mined structural similarity to predict ternary oxides [73], and models to predict the DFT total and/or formation energies [74, 75].

It should be noted here that the three paradigms mentioned above very effectively complement each other. For example, a high-throughput DFT approach via decoration of known structural prototypes (A., above) can help identify regions in the composition-structure space of a chemical system which merit further detailed investigations using a structural search method (B., above). Further, new structure types discovered by global structure searches can be used as input to a secondary round of HT DFT for a quick sampling of chemistries, and this entire process can be repeated in a constant feedback-loop increasing the extent and efficiency of materials discovery. An example of such a feedback mechanism is discussed in Section 3.2.5.

This rest of this thesis is organized as follows: Chapter 2 briefly discusses relevant methods and formalisms that are referred to throughout the thesis (the more specific methods and computational details are relegated to the corresponding chapters). Chapter 3 demonstrates the use of the Open Quantum Materials Database (OQMD) and

related resources in discovering new materials—ternary oxyfluoride compounds via high-throughput prototype decoration, and Heusler-based compounds via cluster expansions. Chapter 4 presents a method to extend the applicability of high-throughput computational databases to explore the high-pressure phase stability landscape of materials by leveraging already available zero-temperature/zero-pressure data within a simple approximation to compound enthalpy; it also uses the newly-developed method to find stable binary high-pressure phases in ambient-immiscible systems. Lastly, Chapter 5 introduces a completely novel paradigm of marrying complex network theory and material science. It presents a graph-based representation of materials phase diagrams, and uses this top-down approach to extract new knowledge, namely, a quantitative metric for material reactivity in the form of a nobility index. Some relevant supplemental data and crystal structures are made available in the Appendix.

CHAPTER 2

Methods and formalisms

2.1. Density functional theory

Density functional theory (DFT) has become the primary tool for electronic structure calculations of materials. The fundamental tenet of DFT is that *any* property of a system of many interacting particles can be viewed as a functional of the ground state density $n_0(\mathbf{r})$ [76]. The Hamiltonian for any system of interacting nuclei and electrons in an external potential $V_{\text{ext}}(\mathbf{r}_i)$ can be written as

$$\hat{H} = -\frac{\hbar^2}{2m_e}\sum_i \nabla_i^2 - \sum_i V_{\text{ext}}(\mathbf{r}_i) + \frac{1}{2}\sum_{i\neq j}\frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|} + \frac{1}{2}\sum_{I\neq J}\frac{Z_I Z_J e^2}{|\mathbf{R}_I - \mathbf{R}_J|} \tag{2.1}$$

where electrons are denoted by lower case subscripts and nuclei, with charge Z_I, at positions $\mathbf{R}_I$ are denoted by upper case subscripts, and the kinetic energy of nuclei has been ignored according to the adiabatic approximation. A universal functional for the energy $E[n]$ in terms of the density $n(\mathbf{r})$, valid for any external potential $V_{\text{ext}}(\mathbf{r})$ can then be written as

$$E[n] = T[n] + E_{\text{e-e}}[n] + E_{\text{NN}} + \int V_{\text{ext}}(\mathbf{r})\, n(\mathbf{r}) \tag{2.2}$$

where $T[n]$ and $E_{\text{e-e}}[n]$ are the kinetic and potential energy (due to electron-electron interaction) functionals, and E_{NN} is the Coulombic energy of nuclei-nuclei interaction. The proof for the existence of such functionals of the ground state electronic density,

uniquely determined for a given external potential, was originally given by Hohenberg and Kohn [77]. Further, the above problem of an interacting many-body system was made tractable by the Kohn-Sham ansatz [78] which replaces the exact ground state density by the ground state density of an auxiliary system of non-interacting electrons, with all the many-body effects grouped into the exchange-correlation energy functional, $E_{\mathrm{xc}}[n]$. This results in independent-particle equations of the Kohn-Sham auxiliary system for which the ground state density $n(\mathbf{r})$ must be solved for self-consistently with the effective external potential $V_{\mathrm{ext}}(\mathbf{r})$.

This reformulation of DFT is, in principle, exact—the ground state energy and energy of the many-body electron problem can be found by solving the Kohn-Sham equations for independent particles. However, the exact form of the exchange-correlation functional E_{xc} is unknown, and thus the method is a feasible approach to calculating the ground state properties of a system to the extent that an approximate form for $E_{\mathrm{xc}}[n]$ can describe the true exchange-correlation energy. In addition to the approximation for $E_{\mathrm{xc}}[n]$, a practical solution of the Kohn-Sham equations involves several other factors that influence the accuracy of a calculation: the use of pseudopotentials to approximate electron-nuclei interactions, finite size of basis sets, fineness of the grids used for representation of physical quantities, and others.

All DFT calculations reported in this work were performed using the Vienna Ab initio Simulation Package (VASP) [79, 80], with projector augmented wave (PAW) [81, 82] potentials, and the Perdew-Burke-Ernzerhof (PBE) [83] formulation of a generalized gradient approximation (GGA) to the exchange-correlation energy functional. We use a constant energy cutoff of 520 eV for the plane-wave basis set, and Γ-centered k-point meshes with a

density of ∼8,000 k-points per reciprocal atom (KPPRA). All structures were completely relaxed with respect to volume and cell-internal degrees of freedom by minimizing the energy till forces on all atoms were less than 0.01 eV/Å and stresses on the cell were within a few kbar. All elements with a half-filled *d*- or *f*-shell were given an initial magnetic moment of 5 or 7 μ_{B}/atom respectively, in a ferromagnetic spin configuration, and then allowed to relax to self-consistency.

2.2. Calculation of energetic quantities

2.2.1. Formation energy

The formation energy of a compound $(A_a B_b C_c \ldots)$ is defined as

$$\Delta E_f(A_a B_b C_c \ldots) = E(A_a B_b C_c \ldots) - \sum_{i=A,B,\ldots} x_i \, \mu_i \tag{2.3}$$

where $E(A_a B_b C_c \ldots)$ is the DFT total energy of the compound, $x_A = \frac{a}{a+b+\ldots}$ and μ_A are the mole fraction and chemical potential of component A, respectively. For most elements, the chemical potential μ_A is simply the DFT total energy of the respective ground state structure of the element. However, for elements that are solid or liquid at, or exhibit a solid-solid phase transformation below, room temperature, the chemical potentials have been fit to experimental formation energies data (see Ref. [20] for details).

2.2.2. Thermodynamic stability

We calculate the thermodynamic stability of a compound by constructing the so-called convex hull of formation energies of all phases in the respective chemical space. Every phase *on* the convex hull has the following property: its formation energy is lower than

any other phase or linear combination of phases at that composition. Thus, all phases on the convex hull are, by definition, thermodynamically stable (e.g., phases S1, S2, S3, and S4 in in Figure 2.1). Conversely, any phase that does not lie on the convex hull is thermodynamically unstable — there is another phase or combination of phases on the convex hull which is lower in energy. For example, in Figure 2.1, phase U1 is unstable because there exists another phase (S2) at the composition that has a lower formation energy; similarly, phase U2 is unstable because a linear combination of phases S2 and S3 has a lower energy at that composition ("E_{hull}").

A measure of thermodynamic stability of a phase is its distance from the convex hull. In other words, the farther away a phase is from the convex hull, higher is the thermodynamic driving force for it to transform or decompose into another phase or combination of phases. The distance from the convex hull ΔE_{HD} for a phase with formation energy ΔE_f can be calculated as

$$\Delta E_{\mathrm{HD}} = E_{hull} - \Delta E_f \tag{2.4}$$

where E_{hull} is the energy of the convex hull at the composition of the phase (see Figure 2.1 for an illustration). The energy of the convex hull at any composition is given by a linear combination of energies of stable phases. This is thus a linear composition-constrained energy minimization problem [25, 84], and is available as a tool called "grand canonical linear programming" (GCLP) on the OQMD website (`http://oqmd.org/analysis/gclp`). Thus, by definition, the hull distance E_{HD} for a phase on the convex hull (hence thermodynamically stable) is 0, and for all unstable phases is greater than 0.

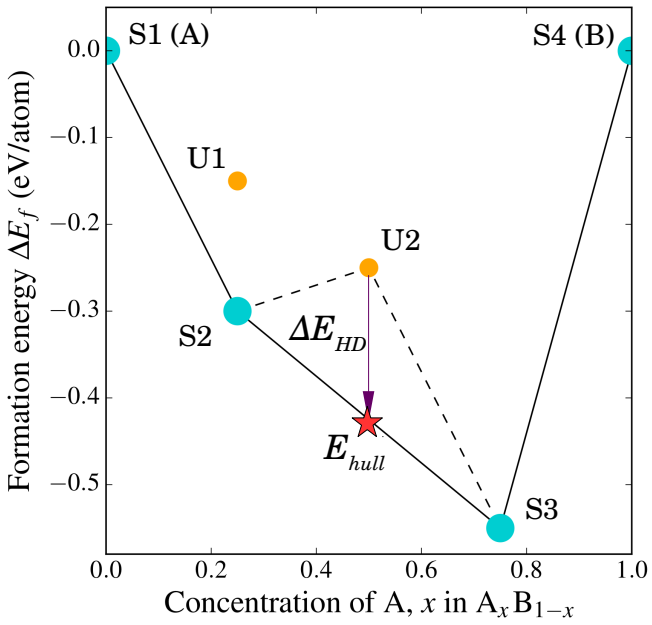

Figure 2.1. A schematic convex hull in the A–B chemical space. Phases S_i lie *on* the convex hull and are thermodynamically stable, i.e., for each phase S_i, there is no other phase or combination of phases at its composition lower in energy. Phases U_i are off the convex hull and thus unstable. For example, the formation energy of phase U2 is higher than that of a linear combination of phases S2 and S3. The distance from the convex hull (ΔE_{HD}) of phase U2 is given by the difference between its formation energy and the energy of the convex hull at its composition (E_{hull}, represented by the crimson star). (Source: Ref. [2])

We note here that the distance from the convex hull of a compound depends on the completeness of the set of phases considered in the construction of the convex hull. Ideally, for calculating the convex hull of a system A–B–C, one would have to investigate all possible compounds that can be formed from elements A, B, and C, which is not feasible. A practical approach is to construct the convex hull with all the currently reported compounds in the A–B–C phase space. For all thermodynamic stability analyses reported in the present work, we have limited our universe of considered phases to those in

the OQMD—almost all unique, ordered compounds reported in the ICSD, and a further ∼400,000 hypothetical compounds based on common structural prototypes.

2.2.3. Significance of calculated thermodynamic stability

A negative value of formation energy ΔE_f of a compound, indicates that at zero temperature, the compound is more stable than its constituent elements. This is a *necessary but far from sufficient* condition for ground state thermodynamic stability. As discussed above, a phase is thermodynamically stable only if its distance from the convex hull $\Delta E_{\text{HD}} = 0$. Further, we assert that the distance of a phase from the convex hull (or simply its "hull distance") ΔE_{HD} is an indicator of the likelihood of its synthesis in experiments.

In a recent study [2], we compiled all the experimentally synthesized XYZ compounds reported in the ICSD (X = Cr/Mn/Fe/Co/Ni/Ru/Rh, Y = Ti/V/Cr/Mn/Fe/Ni, and Z = Al/Ga/In/Si/Ge/Sn/P/As/Sb), if any, and tabulated their formation energies and hull distances as calculated in the OQMD — a total of 110 compounds (with 98 distinct compositions). From a plot of the formation energies and hull distances (Fig. 2.2), we see that the vast majority of the reported compounds that have been experimentally synthesized (green diamonds) lie on or close to the calculated convex hull — 37 compounds are on the convex hull (i.e., a hull distance of 0) and an additional 53 lie relatively close to it (i.e., a hull distance less than about 0.1 eV/atom). Overall, most experimentally reported compounds in the XYZ compositions considered (90 of 99) were found to have a hull distance less than about 0.1 eV/atom. A similar trend of most known compounds being close to the convex hull of formation energies has been documented for other systems (e.g., for half-Heuslers, see Fig. 2.3).

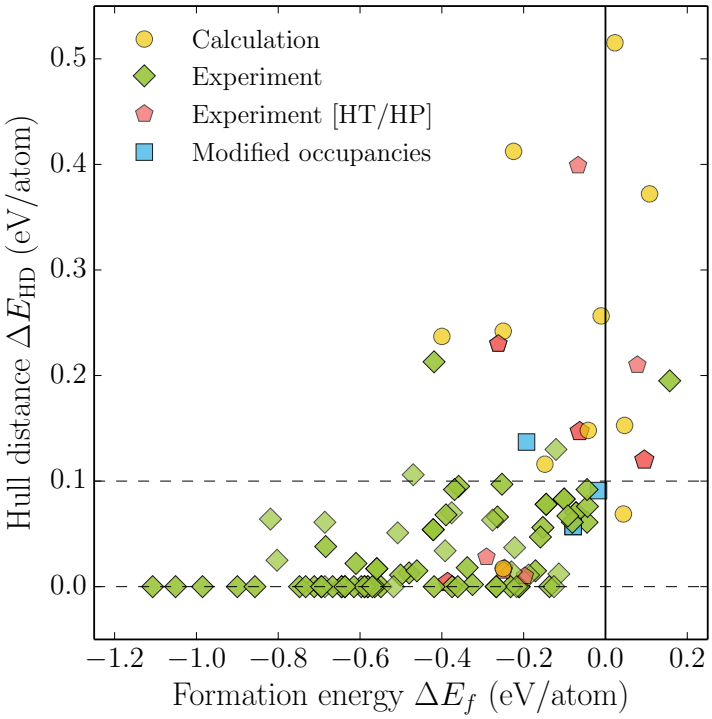

Figure 2.2. DFT-calculated formation energy vs. hull distance of all XYZ compounds reported in the ICSD (X = Cr/Mn/Fe/Co/Ni/Ru/Rh, Y = Ti/V/Cr/Mn/Fe/Ni, and Z = Al/Ga/In/Si/Ge/Sn/P/As/Sb). A hull distance $\Delta E_{\mathrm{HD}} = 0$ indicates a stable ground state compound. (Yellow circles = phases sourced into ICSD from electronic structure calculations, green diamonds = experimentally synthesized phases, red pentagons = phases reported to be stable at high temperature or pressure, light-blue squares = phases with site occupations that differ from the DFT calculation.)

We also note that since we use 0 K DFT energies in our analysis, a phase that is above the convex hull may be either actually metastable or stabilized (i.e., moved on to the convex hull, and thus become experimentally accessible) due to (a) finite temperature contributions to the free energy such as phonons, magnons, configurational entropy, and/or (b) other external conditions such as pressure. Thus, while a phase that is above the convex hull may be experimentally realizable under carefully controlled conditions,

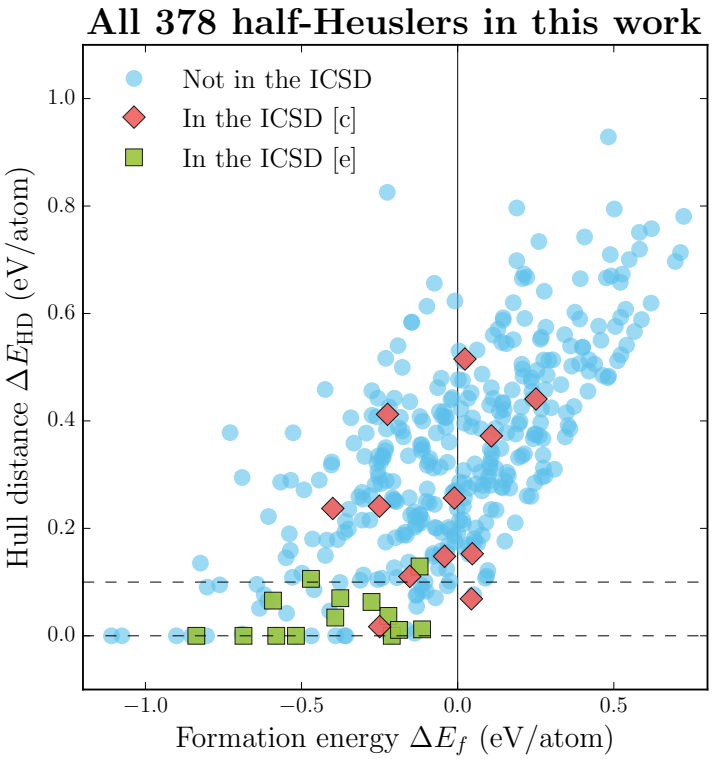

Figure 2.3. DFT-calculated formation energy vs. hull distance of 378 *XYZ* half-Heusler compounds considered in Ma et al. [2]. Almost all the experimentally reported half-Heusler compounds (green squares, "In ICSD [e]") have a hull distance less than 0.1 eV/atom (the window represented by the two horizontal dashed lines); half-Heusler compounds sourced into the ICSD from previous computational work are represented by red diamonds (labeled "In ICSD [c]"); the two compounds in the ICSD that are possibly mischaracterized to be in the AlLiSi half-Heusler structure, RhTiSb and RhTiSn, as represented by yellow pentagons (labeled "Rh(Ti,X) AlLiSi").

and although free energy at the processing temperature may be more relevant in determining phase stability, the 0 K hull distance is still a very good measure available of the *likelihood* of its experimental synthesis.

2.3. Cluster expansion

A cluster expansion for a binary alloy A_xB_{1-x} with N lattice sites can be viewed as a generalized Ising model in which a property $\mathcal{F}$ for any arbitrary configuration $\{\hat{S}_i\}$ is expressed as a linear combination of basis functions defined as the product of occupation variables of the lattice sites:

$$\mathcal{F}\left(\{\hat{S}_i\}\right) = J_0 + \sum_i J_i \hat{S}_i + \frac{1}{2!}\sum_{i\neq j} J_{ij}\hat{S}_i\hat{S}_j + \frac{1}{3!}\sum_{i\neq j\neq k} J_{ijk}\hat{S}_i\hat{S}_j\hat{S}_k + \ldots, \tag{2.5}$$

where the configuration $\{\hat{S}_i\}$ is the vector of all occupation variables ($= \{\hat{S}_1, \hat{S}_2, \ldots, \hat{S}_N\}$), and the occupation variables $\hat{S}_i = +1$ or -1, depending on whether the site i is occupied by an atom of type A or type B. The expression in Eq. (2.5) is essentially exact for any property uniquely determined by the atomic configuration, if all the terms are included in the series, and by extension should hold for a truncated series, if the cluster expansion is well converged. Further, Eq. (2.5) can be rewritten as

$$\mathcal{F}\left(\{\hat{S}_i\}\right) = J_0 + \sum_f D_f J_f \,\overline{\prod}_f \left(\{\hat{S}_i\}\right) \tag{2.6}$$

where f is a set of lattice sites (called a "cluster"). The sum is defined over all clusters f that are *not* equivalent by symmetry, while the product average $\overline{\prod}_f$ is taken over all D_f clusters that are equivalent to f by symmetry. The unknown parameters $\{J_f\}$, called effective cluster interactions (ECI), can then be obtained by fitting them to known values of the property $\mathcal{F}$ for a set of configurations, by the structure inversion method [85].

The transferability of the fit to unknown configurations is measured by a leave-one-out cross-validation (CV) score, defined as:

$$\text{CV score} = \left(\frac{1}{n} \sum_{i=1}^{n} \left(\mathcal{F}_i^{\text{direct}} - \mathcal{F}_{(i)}^{\text{CE}} \right)^2 \right)^{\frac{1}{2}} \tag{2.7}$$

where the $\mathcal{F}_{(i)}^{\text{CE}}$ are predicted by a least-squares fit to $(n-1)$ directly calculated values of the property, excluding $\mathcal{F}_i^{\text{direct}}$. The set of clusters and ECI that minimizes the CV score is termed the optimum cluster expansion. The usefulness of the cluster expansion approach relies on the ECI being rapidly decreasing functions of the number of sites and intersite separation, so that only a few clusters can be included in Eq. (2.6) without a significant loss in accuracy.

All cluster expansions reported in this work were constructed using the implementation of the above methodology in the `maps` utility of the Alloy Theoretic Automated Toolkit (ATAT) [86, 87].

CHAPTER 3

High-throughput materials discovery using database-screening

3.1. Novel ternary oxyfluorides via high-throughput prototype decoration

3.1.1. Background

In comparison to single-anion compounds such as oxides, chalcogenides, fluorides, and nitrides, compounds containing multiple anions (referred to as "mixed-anion" or "heteroanionic") have not been as widely studied [88]. Heteroanion compounds are broadly classified into various families based on the combination of anions present, e.g. oxyhalides, oxychalcogenides, oxypnictides, chalcohalides, and so on (see the leftmost subpanel in Fig. 3.1).

The presence of two anions with different properties such as electronegativity, polarizability, charge, and size, and the various ways in which they can be ordered opens up new ways to engineer the properties of a compound. Similarly, the varying solubility of different elements can be used to fine-tune the doping/defect levels to control the electronic properties in heteroanion compounds. Locally asymmetric coordination can lead to noncentrosymmetric and/or natural heterostructures (schematics in Fig. 3.1).

It is therefore not surprising that many of the known heteroanionic compounds have various compelling properties, making them potential candidates for a wide variety of applications. For example, BiXO (X = Cl, Br, and I) and $LaTiNO_2$ are high-efficiency photocatalysts in the visible light region [89, 90], BiCuSeO and Bi_2SeO_2 show promise for

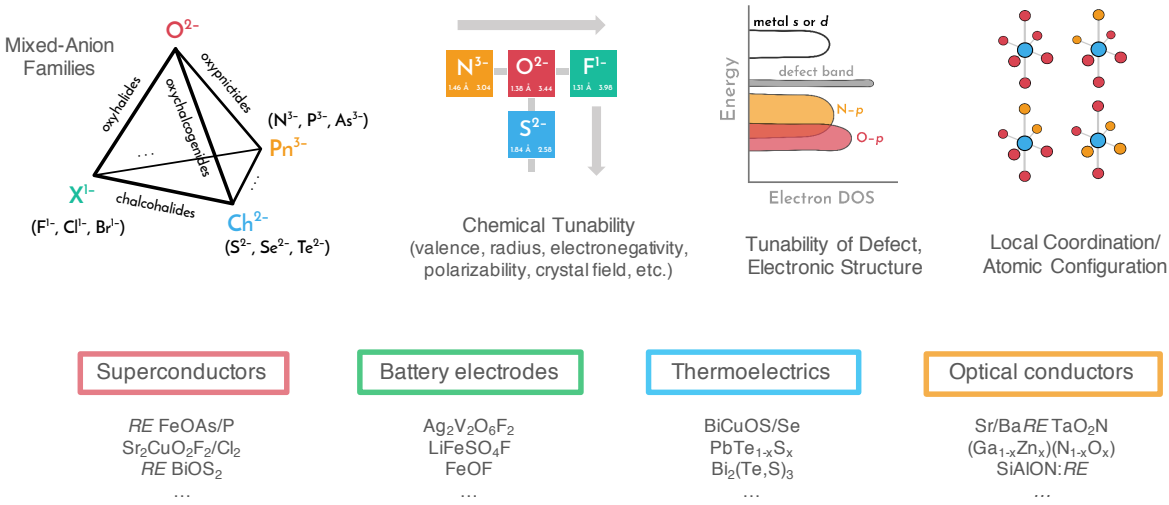

Figure 3.1. The various heteroanion family of compounds, categorized by the combination of anions in them. The diversity in the properties of individual anions provides opportunities to tune those of the heteroanion compound for desired applications.

thermoelectric applications [91, 92], $Bi_4S_3O_4$ and LaFeAsO family show superconductivity [93, 94], LaCuSO is an excellent p-type transparent conductor oxide [95], FeOF [96] is a potential cathode for conversion batteries, and so on.

In this chapter, we focus on ternary oxyfluoride compounds. Oxyfluoride materials have shown improved performance over their oxide and fluoride counterparts in applications such as electrochemical cells and solid-state lighting. They also show a myriad of technologically useful properties such as nonlinear optical behavior and superconductivity. We propose to use the prototype decoration followed by high-throughput DFT calculations to discover novel hitherto unreported single-metal oxyfluorides.

Our approach to discovering novel single-metal oxyfluoride compounds involves the following steps: (a) systematic survey and analysis of compounds reported in the literature to identify common structural prototypes, (b) decorating such structural prototypes

with all the elements in the periodic table on the cationic sites to generate new hypothetical heteroanionic compounds, (c) performing HT DFT calculations of the prototype decorations to determine their formation energy, and (d) calculating the phase stability of the prototype decorations against all the phases to identify new heteroanionic compounds that are likely to be thermodynamically stable.

3.1.2. Survey of the ICSD and data curation

We queried the Inorganic Crystal Structure Database (ICSD) for all single-metal oxyfluorides of the form $M_aO_bF_c$ (where M = one of 66 metals/semi-metals: $2-6s$ block alkali metals, $4-6d$ transition metals, lanthanide elements, those belonging to the boron group (Al, Ga, In, Tl), carbon group (Si, Ge, Sn, Pb), nitrogen group (As, Sb, Bi), and Te). The resulting 165 entries were used to create a curated dataset as follows (schematic summarizing data curation in Fig. 3.2):

(1) All materials in which either O or F is not anionic are discarded.

(2) Any structure exhibiting site disorder, i.e., one or more crystallographic sites statistically occupied with an occupancy factor of less than unity for one or more species, was discarded. 73 of the structures queried from the ICSD had occupancy disorder on one or more sites.

(3) Only unique structures, i.e., structures which were *not* a duplicate of another structure in the dataset at the same composition, were considered. Structural similarity algorithms implemented in the MINT package [97, 98] were used to

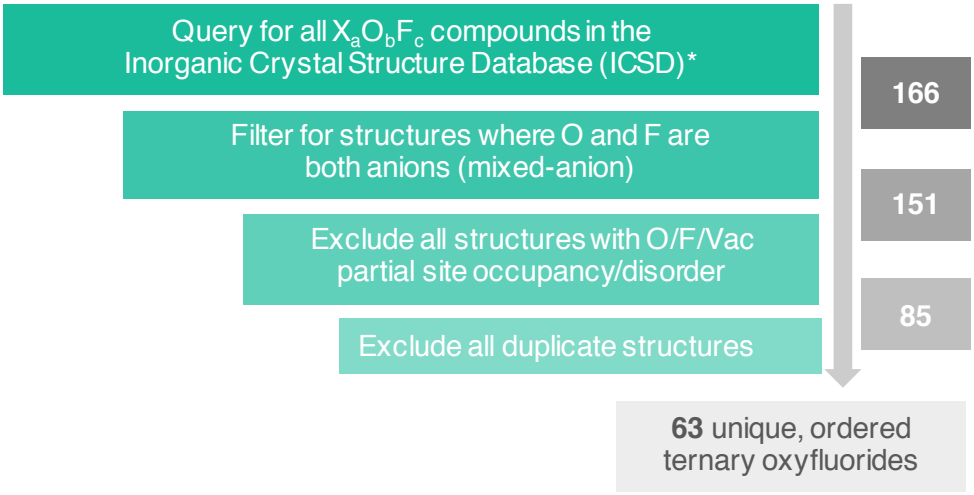

Figure 3.2. Schematic describing the steps in curating the ICSD dataset to retain only distinct, ordered ternary oxyfluorides.

compare structures with a tolerance factor of 0.2. Of all the entries queried from the ICSD, 31 were duplicates, and hence discarded.

The final curated dataset of 63 ordered single-metal oxyfluorides was considered for all subsequent analysis. A distribution of site-ordered $M_aO_bF_c$ compounds in the ICSD is shown in Figure 3.3.

3.1.3. Summary of reported compounds and prototype calculations

The compounds reported in the ICSD occur with ∼24 different stoichiometries, and several structural prototypes (see Table 3.1 for a summary). We identify the more frequently occurring prototypes with simple stoichiometries, grouped according to the oxidation state of the cation (examples of the structures used for prototype decoration in Fig. 3.4):

(a) $M_2O_1F_2$ (M^{2+}): Pb_2OF_2 ($P4_2/ncm$), Ba_2OBr_2 ($Ibam$)

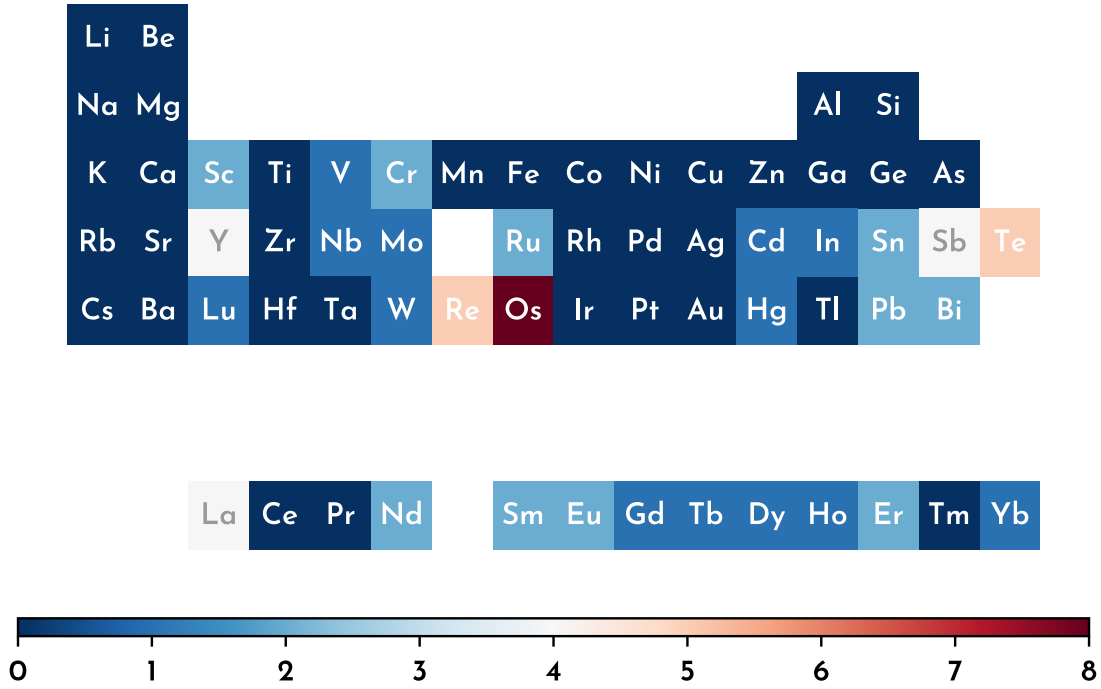

Figure 3.3. Distribution of single-metal oxyfluoride compounds in the Inorganic Crystal Structure Database (ICSD). The background color of each element M represents the number of distinct, ordered compound $M_aO_bF_c$ reported in the ICSD.

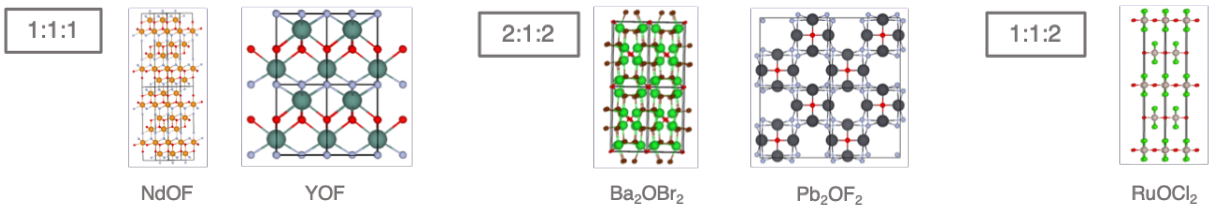

Figure 3.4. Some common prototype oxyhalide structures at simple stoichiometries used for generating hypothetical oxyfluoride compounds via prototype decoration.

(b) $M_1O_1F_1$ (M^{3+}): NdOF ($R3m$), YOF ($P4nmm$), VOCl ($Pmmn$)

(c) $M_1O_1F_2$ (M^{4+}): $RuOCl_2$ ($Immm$), $NbOI_2$ ($C2$)

$M_a O_b F_c$	Number
all entries in the ICSD	165
unique, ordered entries in the ICSD	63
entries from the ICSD calculated in the OQMD	49
compounds on the OQMD convex hull	29
compounds on or < 0.1 eV/atom of the OQMD convex hull	44
unique stoichiometries in the ICSD	24
possible decorations	$24 \times 66 = 1584$
unique (named) structural prototypes in the ICSD	17
possible decorations	$17 \times 66 = 1122$

Table 3.1. A summary of the reports of single-metal oxyfluoride compounds in the Inorganic Crystal Structure Database (ICSD). Number of all possible prototype decorations, based on the unique stoichiometries and structural prototypes in the ICSD, are also listed.

3.1.4. Assessment of thermodynamic stability of known ternary oxyfluorides

We first investigate the accuracy of using distance to the convex hull to quantify thermodynamic stability of a given compound. We see that the vast majority of experimentally reported oxyfluorides in the ICSD are either predicted to lie on the convex hull or within 0.1 eV/atom of it according to the formation energies calculated in the OQMD (see Fig. 3.5). This gives us confidence in our approach towards finding new stable oxyfluorides using the same hull distance metric.

3.1.5. High-throughput prototype decoration

For each structural prototype, we replace the metal M with all elements in the periodic table that can assume the corresponding oxidation state, i.e., elements that can exist in 3+, 4+, 2+ states for M in the $M_1O_1F_1$, $M_1O_1F_2$, and $M_2O_1F_2$ prototypes respectively. We perform HT DFT calculations of all such decorations of the structural prototypes,

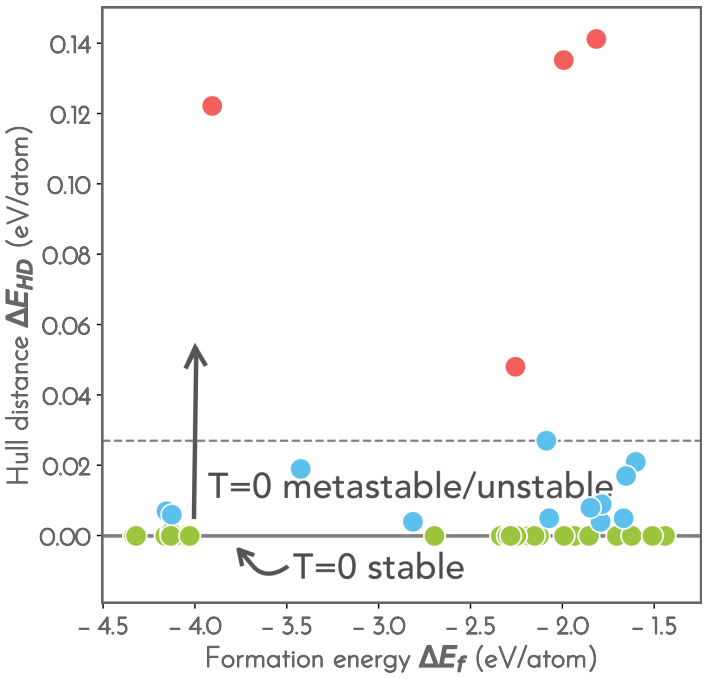

Figure 3.5. Formation energy vs distance from the convex hull for all the ternary oxyfluoride compounds calculated in the OQMD. Most compounds lie on or close to the convex hull.

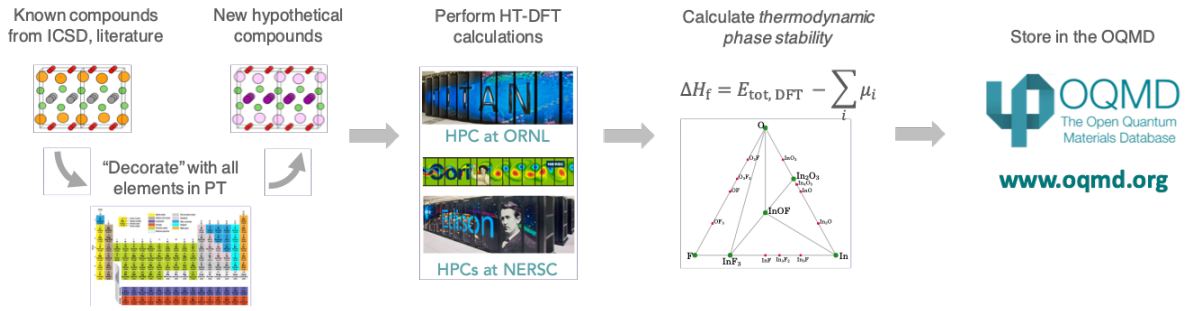

Figure 3.6. A summary of the high-throughput DFT workflow for the prototype decoration paradigm.

spanning a wide range of chemistries, and calculate their stability against all other phases in the OQMD. A summary of this approach is shown as a schematic in Fig. 3.6.

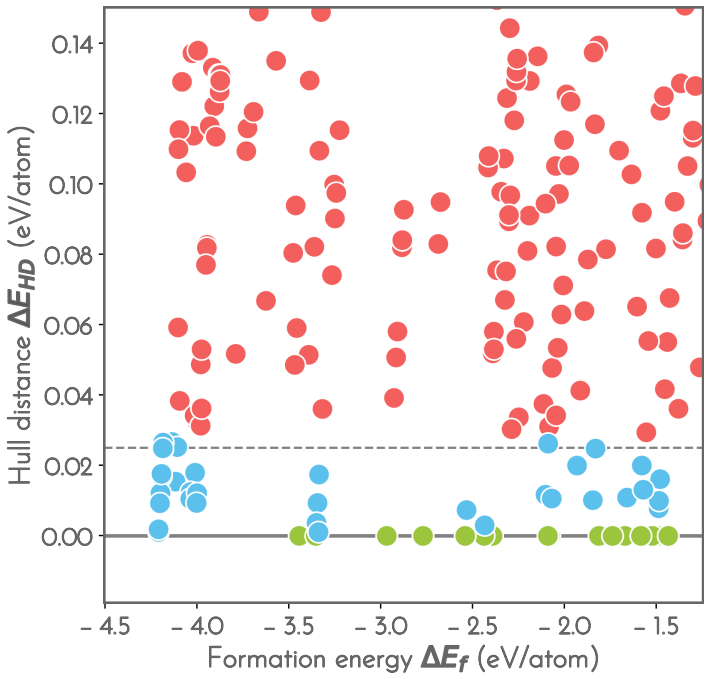

Figure 3.7. Formation energy vs distance from the convex hull for hypothetical ternary oxyfluorides generated by decorating common structural prototypes. There are several new oxyfluorides predicted to be stable (green circles) or nearly so (blue circles).

3.1.6. Predictions of thermodynamically stable ternary oxyfluorides

Our approach predicts a number of novel hitherto unreported oxyfluoride compounds to be thermodynamically stable or nearly so in each of the stoichiometries sampled (see Fig. 3.7). Some of the compounds predicted include CrOF, VOF, NiOF, $RuOF_2$, $MoOF_2$, Fe_2OF_2, and Co_2OF_2. A summary of the compounds predicted to lie on the OQMD convex hull, i.e., to be thermodynamically stable, by our calculations is given in Table 3.2, and schematics of their crystal structure are shown in Figs. 3.8, 3.9 and 3.10.

Cation	Compound	Structural prototype	Space group	E_g (eV)	E_{HD} (eV/atom)	Competing phases
M^{2+}						
	Mn_2OF_2	Ba_2OBr_2	*Ibam*	0.5	−0.142	MnO–MnF_2
	Fe_2OF_2	Ba_2OBr_2	*Ibam*	1.1	−0.169	FeO–FeF_2
	Co_2OF_2	Ba_2OBr_2	*Ibam*	1.7	−0.179	CoO–CoF_2
	Ni_2OF_2	Ba_2OBr_2	*Ibam*	2.2	−0.090	NiO–NiF_2
M^{3+}						
	CrOF	VOCl	*Pmmn*	2.2	−0.110	Cr_2O_3–CrF_3
	VOF	VOCl	*P*41212	1.5	−0.127	V_2O_3–VF_3
	MnOF	VOCl	*P*41212	0.5	−0.072	Mn_2O_3–MnF_3
	CoOF	VOCl	*P*41212	0.0	−0.041	Co_3O_4–CoF_3–O_2
	NiOF	ScOF	*P*21/*c*	2.7	−0.075	NiO_2F_2–O_2
M^{4+}						
	$RuOF_2$	$RuOCl_2$	*Immm*	0.0	−0.007	RuO_2–RuF_4
	$MoOF_2$	$NbOI_2$	*C*2	0.0	−0.026	MoO_2–MoF_6–MoF_3
	$OsOF_2$	$NbOI_2$	*P*2	0.0	−0.012	OsO_2–OsF_6–Os

Table 3.2. A summary of single-metal oxyfluoride compounds predicted to lie on the convex hull of stable phases by our calculations. The columns respectively represent the stable compound predicted, the structural prototype which was decorated, the space group, the calculated DFT band gap, the distance of the compound from the convex hull of *currently known phases*, the lowest energy phase/combination of phases at the composition according to currently known compounds in the OQMD. Only predictions without lanthanide/actinide/radioactive elements are shown.

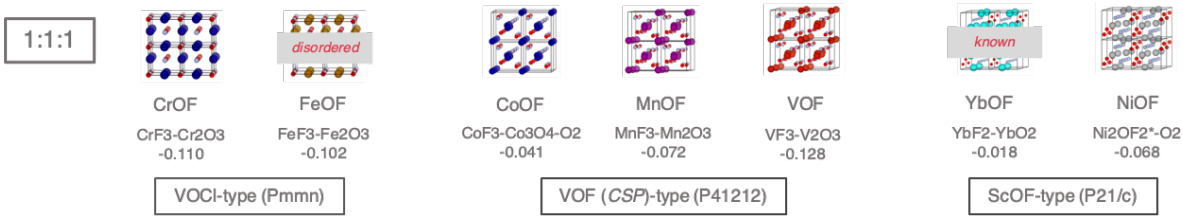

Figure 3.8. Novel, hitherto-unreported ternary oxyfluoride compounds with the 1:1:1 stoichiometry predicted to be stable.

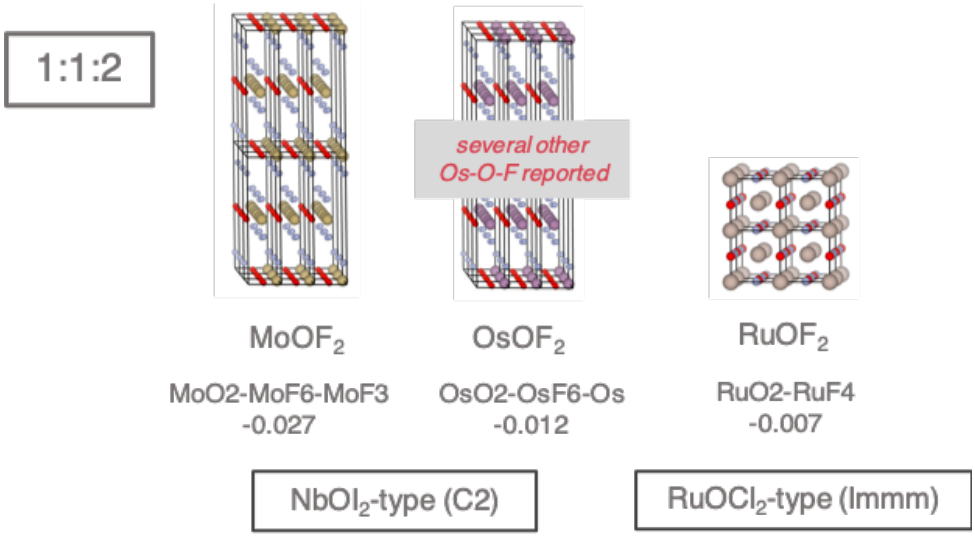

Figure 3.9. Novel, hitherto-unreported ternary oxyfluoride compounds with the 1:1:2 stoichiometry predicted to be stable.

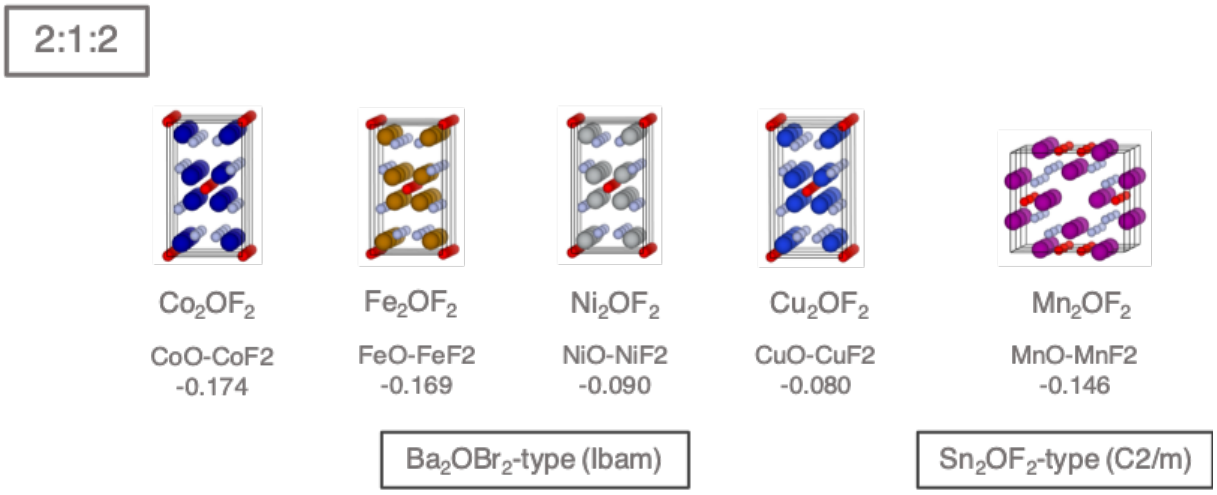

Figure 3.10. Novel, hitherto-unreported ternary oxyfluoride compounds with the 2:1:2 stoichiometry predicted to be stable.

3.2. Novel Heusler-based compounds from cluster expansion

3.2.1. Background

Half-metallic ferromagnets (HMFs) have an energy gap in the minority-spin channel and a finite density of states in the majority-spin channel leading to complete spin polarization at the Fermi level, and thus are attractive for spintronic applications. Many full- and half-Heusler phase compounds are HMFs and are particularly promising due to their relatively high Curie temperatures. A simple screen over all Heusler-related phases in the OQMD yielded, among other candidates, half-Heusler FeTiSb and full-Heusler Fe_2TiSb phases that are both near half-metals and both predicted to on or close to the Fe–Ti–Sb convex hull, making them good candidates for spintronic applications. However, it is known that defects such as vacancies (on the X site of an X_2YZ Heusler phase) can lead to a loss of half-metallic character. While this effect has been reported in Co-, and Mn-based Heusler compounds, it has not been studied in Fe-based Heusler compounds. Moreover, the effect of vacancies on the stability and order of these compounds has not been adequately explored. Here we propose to use the cluster expansion method to study such effects in the Fe–Ti–Sb system.

3.2.2. Constructing the cluster expansion

We calculate the formation energy ΔE_f (per mixing atom) of a compound with composition $\mathrm{Fe}_x\Box_{1-x}\mathrm{Ti}_{0.5}\mathrm{Sb}_{0.5}$ with respect to the end members $\mathrm{FeTi}_{0.5}\mathrm{Sb}_{0.5}$ and $\Box\mathrm{Ti}_{0.5}\mathrm{Sb}_{0.5}$ using

$$\Delta E_f = E\big(\mathrm{Fe}_x\Box_{1-x}\mathrm{Ti}_{0.5}\mathrm{Sb}_{0.5}\big) - \Big\{x\,E\big(\mathrm{FeTi}_{0.5}\mathrm{Sb}_{0.5}\big) + (1-x)\,E\big(\Box\mathrm{Ti}_{0.5}\mathrm{Sb}_{0.5}\big)\Big\} \tag{3.1}$$

where all the required energies are calcuted using DFT. The parameters used for all calculations, and other details are discussed in Section 2.1.

We fit a binary CE to the formation energies calculated using Eq. (3.1) by minimizing the sum of squares of the residuals (see Section 2.3 for details). We generate new structures, use their formation energies in the fit, and add terms to the CE till the cross-validation (CV) score is minimized, and adding further structures results in overfitting. We find a well-converged CE, with a CV score of ∼20 meV/mixing atom by using 66 $Fe_x\square_{1-x}Ti_{0.5}Sb_{0.5}$ structures and their formation energies (Figure 3.11).

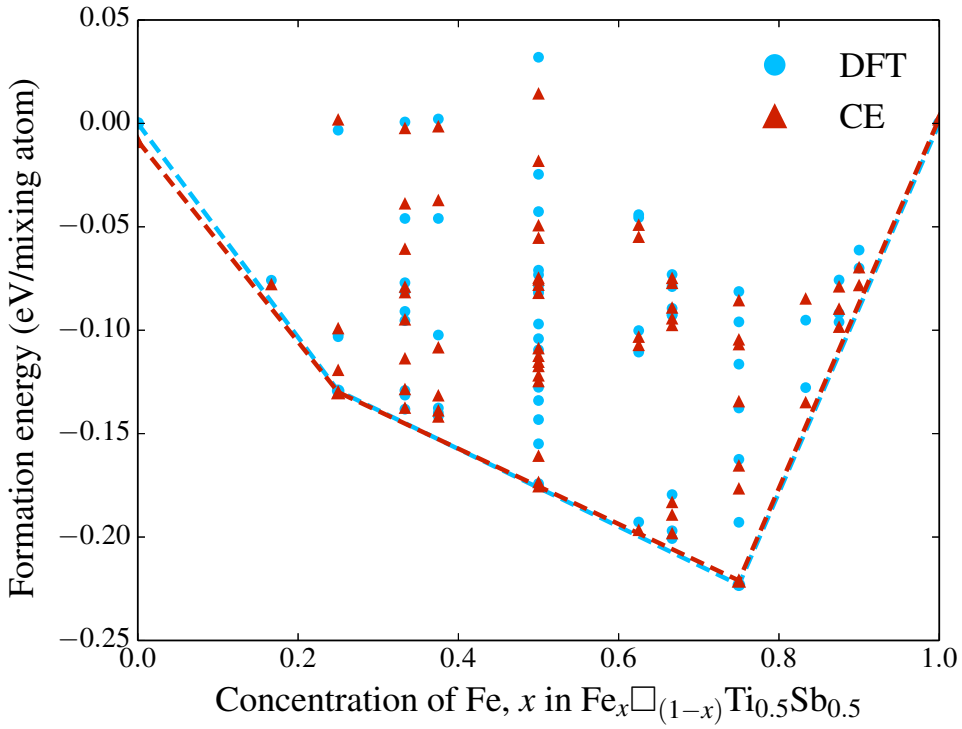

Figure 3.11. (a) Energies of 66 different arrangements of Fe and □ on the Fe sublattice of Fe_2TiSb in the $L2_1$ structure, calculated using DFT, and those predicted by a binary cluster expansion (CE) fit to the DFT energies. The DFT convex hull is captured well by the CE, with a cross-validation error of ∼20 meV/mixing atom.

The effective cluster interactions (ECI) decrease rapidly with increase in the intersite distance for clusters of all types (i.e., two-, three-, and four-point clusters; see Figure 3.12a), implying that excluding larger sized clusters does not result in a significant loss in accuracy. A comparison between the energies predicted by the CE and those calculated by DFT (Figure 3.12b) show excellent agreement, giving us confidence in our approach. All symmetrically unique clusters with corresponding effective cluster interactions (ECI) greater than 0.005 eV/mixing atom in magnitude are listed in Table 3.3, and depicted schematically in Figure 3.13. We note that the cluster with the dominant ECI is a two-point antiferromagnetic cluster ($J_{2,4}$ in Table 3.3) that favors the ordering of Fe and □ in the ⟨111⟩ direction on the Fe sublattice.

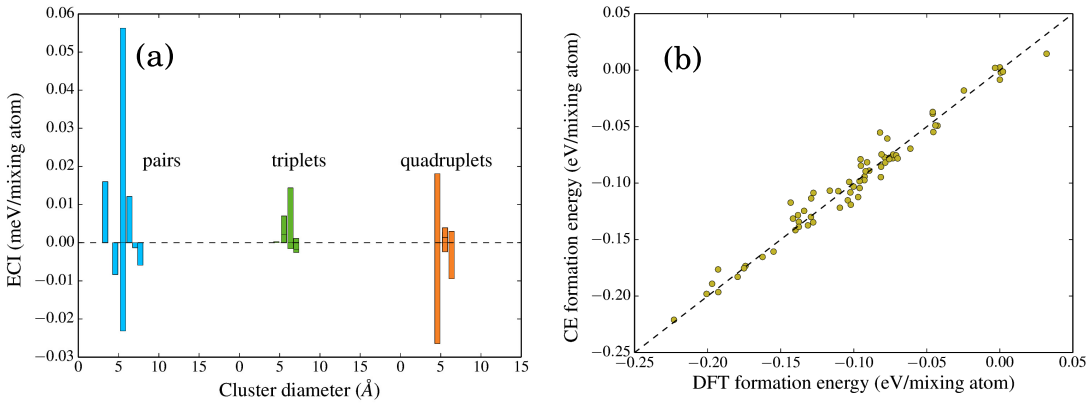

Figure 3.12. (a) Variation of effective cluster interactions (ECI) with increasing intersite distance for various two-, three-, and four-point clusters used in the CE. (b) Comparison of formation energies calculated with DFT with that predicted by the CE; shows excellent agreement.

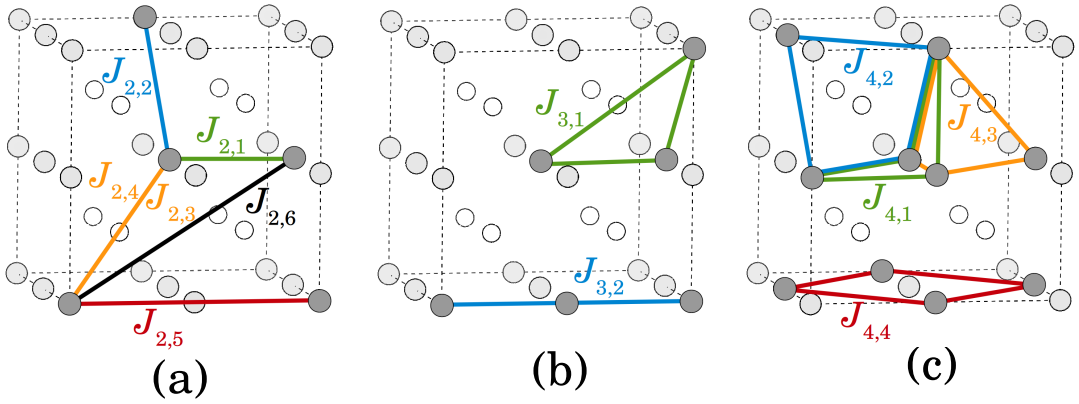

Figure 3.13. All symmetrically unique clusters with corresponding ECI greater than 5 meV/mixing atom (see Table 3.3) included in the binary CE; (a) pairs, (b) triplets, and (c) four-point clusters. $J_{k,m}$ refer to the ECI corresponding to a cluster $f = (k, m)$ with k vertices, and separated by m-th nearest neighbor distance. Light and dark gray filled circles represent sites belonging to the Fe sublattice, whereas unfilled circles represent sites belonging to the Ti/Sb sublattices.

3.2.3. Low-energy crystal structures from cluster expansion

From our cluster expansion, we predict the energy of 66 configurations, a complete enumeration of all possible Fe/□ orderings with 20 atoms or less per cell. Out of this expremely large configuration space, we find that the lowest energy structure corresponds to a composition of $Fe_{0.75}\square_{0.25}Ti_{0.5}Sb_{0.5}$, i.e., between the full-Heusler and half-Heusler compositions, belongs to the trigonal $R3m$ space group, and has a formation energy of -0.223 eV/mixing atom. We also find structures on or within a few meV/mixing atom of the convex hull at other compositions: (a) at the composition of the half-Heusler phase ($Fe_{0.5}\square_{0.5}Ti_{0.5}Sb_{0.5}$), we find two structures in space groups $C1m1$ and $Amm2$ (labeled as such in Figure 3.14), 1 and 2 meV/mixing atom above the convex hull respectively, while the half-Heusler phase ('$C1b$' in Figure 3.14) itself is ∼20 meV/mixing atom above the

Cluster (coordinates)	# Sites	Diameter (Å)	Multiplicity	ECI (eV/mixing atom)
$J_{0,1}$: empty	0	0	1	-0.130
$J_{1,1}$: 000	1	0	2	-0.055
$J_{2,1}$: $\frac{1}{2}\frac{1}{2}\frac{1}{2}$, $0\frac{1}{2}\frac{1}{2}$	2	3.009	6	0.016
$J_{2,2}$: $\frac{1}{2}\frac{1}{2}\frac{1}{2}$, $0\frac{1}{2}1$	2	4.255	12	-0.008
$J_{2,3}$: $\frac{1}{2}\frac{1}{2}\frac{1}{2}$, 101	2	5.212	4	-0.023
$J_{2,4}$: $\frac{1}{2}\frac{1}{2}\frac{1}{2}$, 011	2	5.212	4	0.056
$J_{2,5}$: 000, 010	2	6.018	6	0.012
$J_{2,6}$: 000, $\frac{1}{2}1\frac{1}{2}$	2	7.371	24	-0.006
$J_{3,1}$: $\frac{1}{2}\frac{1}{2}\frac{1}{2}$, $1\frac{1}{2}\frac{1}{2}$, 101	3	5.212	24	0.007
$J_{3,2}$: 000, $0\frac{1}{2}0$, 010	3	6.018	6	0.014
$J_{4,1}$: $\frac{1}{2}\frac{1}{2}\frac{1}{2}$, $0\frac{1}{2}\frac{1}{2}$, $00\frac{1}{2}$, $0\frac{1}{2}1$	4	4.255	8	-0.017
$J_{4,2}$: $\frac{1}{2}\frac{1}{2}\frac{1}{2}$, $\frac{1}{2}01$, $00\frac{1}{2}$, $0\frac{1}{2}1$	4	4.255	2	-0.026
$J_{4,3}$: $\frac{1}{2}\frac{1}{2}\frac{1}{2}$, $0\frac{1}{2}\frac{1}{2}$, $01\frac{1}{2}$, $0\frac{1}{2}1$	4	4.255	8	0.018
$J_{4,4}$: $\frac{1}{2}00$, $1\frac{1}{2}0$, $0\frac{1}{2}0$, $\frac{1}{2}10$	4	6.018	6	-0.009

Table 3.3. List of symmetrically unique clusters included in the binary cluster expansion, the number of sites in each of them, the cluster diameter, multiplicity and the corresponding effective cluster interactions (ECI). All atomic coordinates are in units of the lattice vectors of the Fe_2TiSb $L2_1$ structure.

convex hull, and (b) at a composition of $Fe_{0.25}\square_{0.75}Ti_{0.5}Sb_{0.5}$, we find a structure in the space group $R3m$ (labeled as such in Figure 3.14) to be on the convex hull. We discuss in detail, the common structural motifs in the arrangement of Fe and $\square$ in several competing structures at the composition of the lowest energy structure, i.e., $Fe_{0.75}\square_{0.25}Ti_{0.5}Sb_{0.5}$, below.

We find three tetragonal structures (I.)—(III.), and one trigonal structure (IV.) within 100 meV (per mixing atom) of the CE convex hull:

(I.) In the space group $P\bar{4}m2$ (No. 115 in the International Tables of Crystallography), ∼85 meV/mixing atom above the convex hull. It has an $ABA'B'$-type

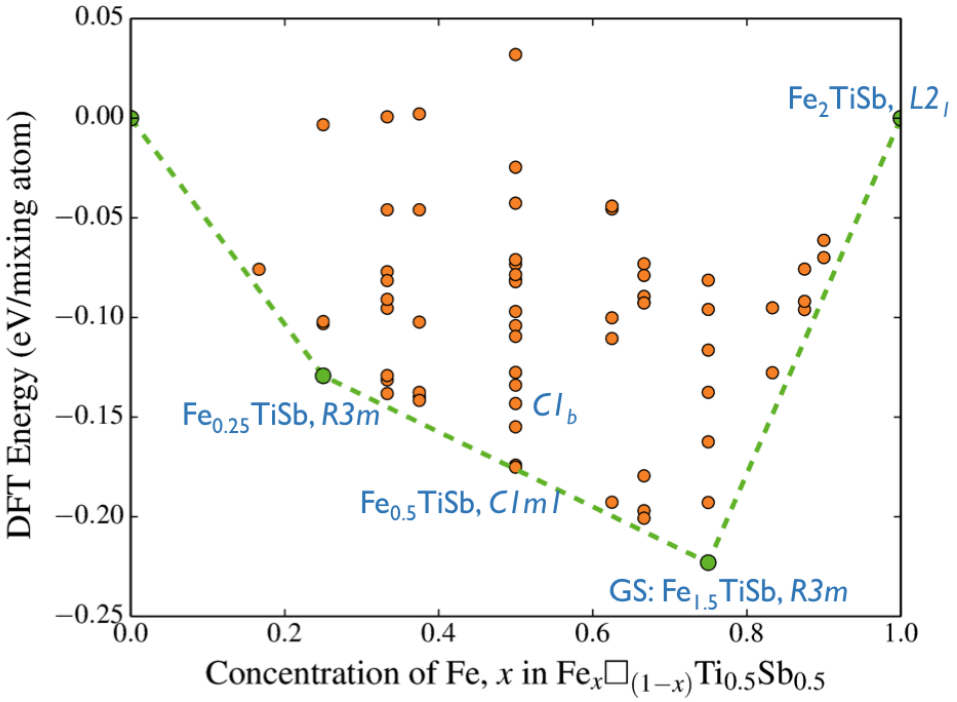

Figure 3.14. Various stable and nearly stable (within ∼100 meV/mixing atom of the convex hull) structures, and their formation energies calculated with DFT.

stacking along the ⟨001⟩ direction, where planes A and A' make up the interpenetrating fcc sublattices of Ti and Sb, plane B has completely occupied Fe sites, and plane B' has Fe and □ occupying alternating sites on the Fe sublattice of the $L2_1$ structure (Figure 3.15a).

(II.) In the space group $P42/nmc$ (No. 137 in the International Tables of Crystallography), ∼61 meV/mixing atom above the convex hull. It has an $AB'A'BAB''A'B$-type stacking along the ⟨001⟩ direction, where planes A and A' make up the interpenetrating fcc sublattices of Ti and Sb, plane B has completely occupied Fe sites, and planes B' and B'' have alternating chains of Fe and □ in the ⟨110⟩ direction on the Fe sublattice of the $L2_1$ structure (Figure 3.15b).

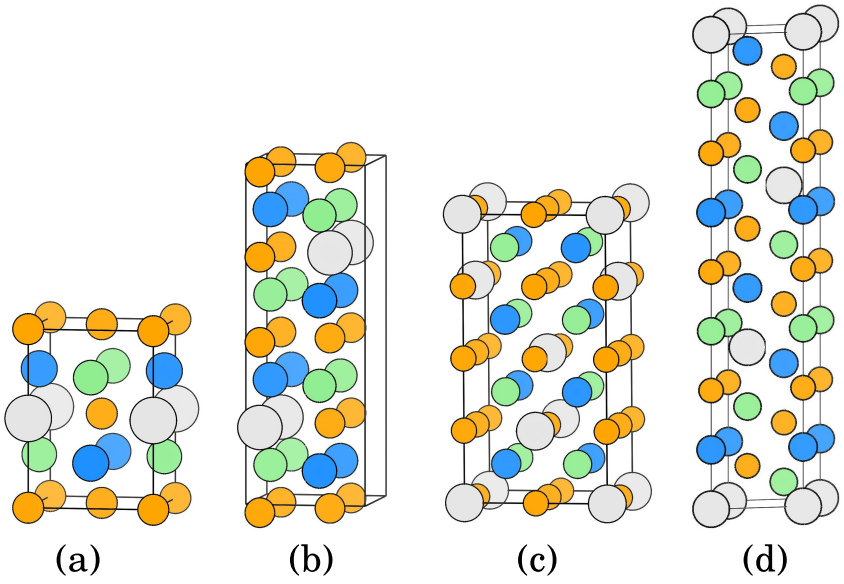

Figure 3.15. Competing structures at a composition of $Fe_{0.75}\square_{0.25}Ti_{0.5}Sb_{0.5}$ belonging to space group (a) $P\bar{4}m2$, (b) $P42/nmc$, (c) $I\bar{4}2d$, (d) $R3m$, that are ∼85, ∼60, ∼30, and 0 (stable) meV/mixing atom above the convex hull respectively. (Fe = orange, □ = light gray, Ti = blue, and Sb = green)

(III.) In the space group $I\bar{4}2d$ (No. 122 in the International Tables of Crystallography), about 30 meV/mixing atom above the convex hull. It has an $AB'A'BABA'B''$-type stacking in the ⟨001⟩ direction. While the planes A and A' make up the interpenetrating fcc sublattices of Ti and Sb, plane B has completely occupied Fe sites, planes B' and B'' have alternating chains of (i) Fe atoms, and (ii) Fe and □ occupying alternate sites, on the Fe sublattice of the $L2_1$ structure but are shifted from each other by half a lattice vector along ⟨100⟩ (Figure 3.15c).

(IV.) In the space group $R3m$ (No. 160 in the International Tables of Crystallography), *on the convex hull* ('$GS, R3m$' in Figure 3.14). It has chains of Fe and □ arranged alternatingly along the ⟨111⟩ direction (Figure 3.15d) on the Fe sublattice of the

$L2_1$ structure, consistent with our previous observations based on the ECI of the CE.

We list all the stable and nearly stable structures at various compositions, their formation energies, and their distance from the convex hull in Table 3.4. In addition, we list the correlation coefficients of the above discussed competing structures and the perfectly disordered structure at composition $x = 0.75$ in Table 3.5.

Composition (x)	Space group	ΔE_f (eV/mixing atom)	ΔE_{HD}	E_g (eV)	M (μ_B)
0.0	$Fm\bar{3}m$	0	0	0	0
0.25	$R3m$	−0.129	0	0.439	0
0.50	$F4\bar{3}m$ ($C1b$)	−0.155	0.021	0	0.92
	$Amm2$	−0.174	0.002	0.08	0
	$C1m1$	−0.175	0.001	0.15	0
0.75	$P\bar{4}m2$	−0.138	0.085	0.23	0
	$P42/nmc$	−0.162	0.060	0.25	0
	$I\bar{4}2d$	−0.193	0.030	0.40	0
	$R3m$	−0.223	0	0.64	0
1.0	$Fm\bar{3}m$ ($L2_1$)	0	0	0.72 ↓	1.04

Table 3.4. List of stable and nearly stable structures at various compositions x, for x in $Fe_x\square_{1-x}Ti_{0.5}Sb_{0.5}$, the space group they belong to, their formation energy ΔE_f as defined in Eq. (3.1), distance from the convex hull ΔE_{HD} ("stability"), band gap E_g (gap present in both spin channels unless specified; ↓ indicates gap only in the minority spin channel), and magnetization M; all properties are calculated with DFT.

	$P\bar{4}m2$	$P42/nmc$	$I\bar{4}2d$	$R3m$	$\mathcal{R}$
$J_{2,1}$	0.000	0.000	0.000	0.000	0.250
$J_{2,2}$	0.333	0.333	0.333	0.500	0.250
$J_{2,3}$	0.000	0.000	0.000	0.000	0.250
$J_{2,4}$	0.000	0.000	0.000	0.000	0.250
$J_{2,5}$	1.000	0.667	0.667	0.000	0.250
$J_{2,6}$	0.333	0.000	0.667	0.500	0.250
$J_{3,1}$	-0.167	-0.167	-0.167	0.000	0.125
$J_{3,2}$	0.500	0.167	0.167	-0.500	0.125
$J_{4,1}$	0.000	0.000	0.000	0.000	0.062
$J_{4,2}$	1.000	1.000	1.000	0.000	0.062
$J_{4,3}$	0.000	0.000	0.000	0.000	0.062
$J_{4,4}$	1.000	0.333	0.333	1.000	0.062

Table 3.5. Correlation coefficients of competing structures at a composition of $Fe_{0.75}\square_{0.25}Ti_{0.5}Sb_{0.5}$ labeled by their space group ($P\bar{4}m2$, $P42/nmc$, $I\bar{4}2d$, and $R3m$), compared with those of the perfectly disordered structure $\mathcal{R}$.

3.2.4. Interplay between ordering of vacancies, thermodynamic stability, and electronic structure

To understand the effect of the ordering of vacancies on the electronic structure of the $Fe_x\square_{1-x}Ti_{0.5}Sb_{0.5}$ compounds, we calculated the density of electronic states in these compounds. First, our calculations show that (a) the Fe_2TiSb full-Heusler compound is a half-metallic ferromagnet with the Fermi energy falling at the edge of the gap in the minority spin channel (Figure 3.16a), (b) the Fe□TiSb half-Heusler is a nearly half-metallic ferromagnet with the Fermi energy falling at the edge of the valence band in the majority spin channel (Figure 3.16b), consistent with previous experimental and theoretical studies [99, 100], and (c) the $\square_2$TiSb compound in the $L2_1$ structure is a non-magnetic metal (Figure 3.16c).

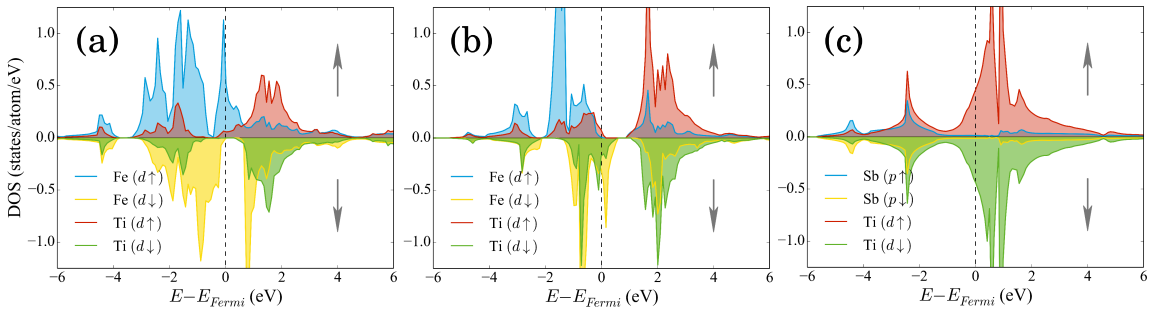

Figure 3.16. Orbital-projected density of electronic states (DOS) of (a) Fe_2TiSb (full-Heusler phase), (b) Fe□TiSb (half-Heusler phase), (c) $□_2TiSb$ in the $L2_1$ structure. ↑ and ↓ represent majority and minority spin channels respectively.

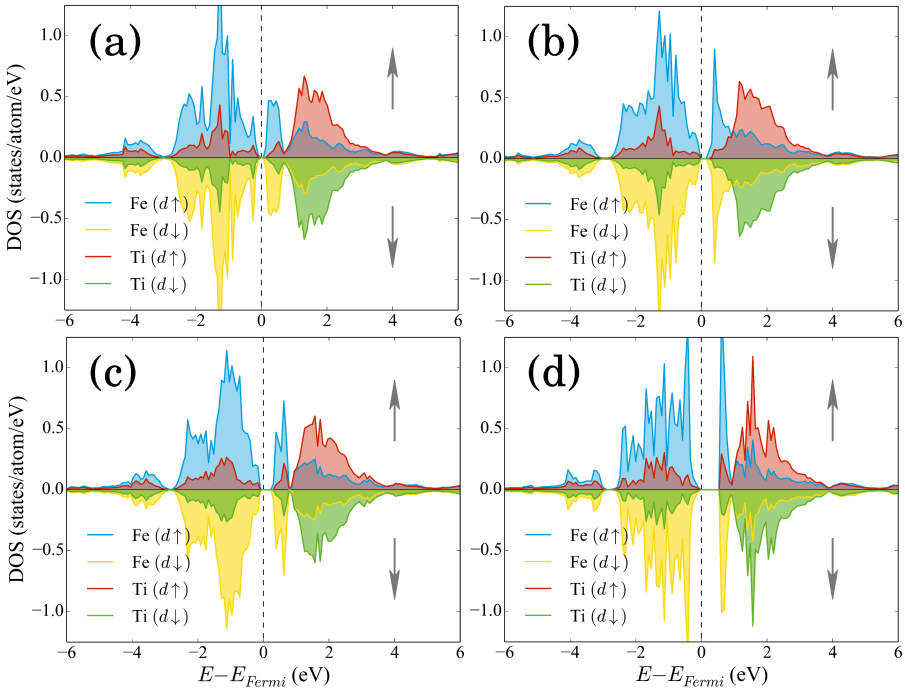

Figure 3.17. Comparison of orbital-projected density of electronic states (DOS) of four competing structures at composition $Fe_{0.75}□_{0.25}Ti_{0.5}Sb_{0.5}$ (corresponding to Figures 3.15a–d). While the distance from the convex hull of the structures decreases (i.e., they become increasingly stable) from (a) to (d), this corresponds to opening of increasingly larger gaps, ↑ and ↓ represent majority and minority spin channels respectively.

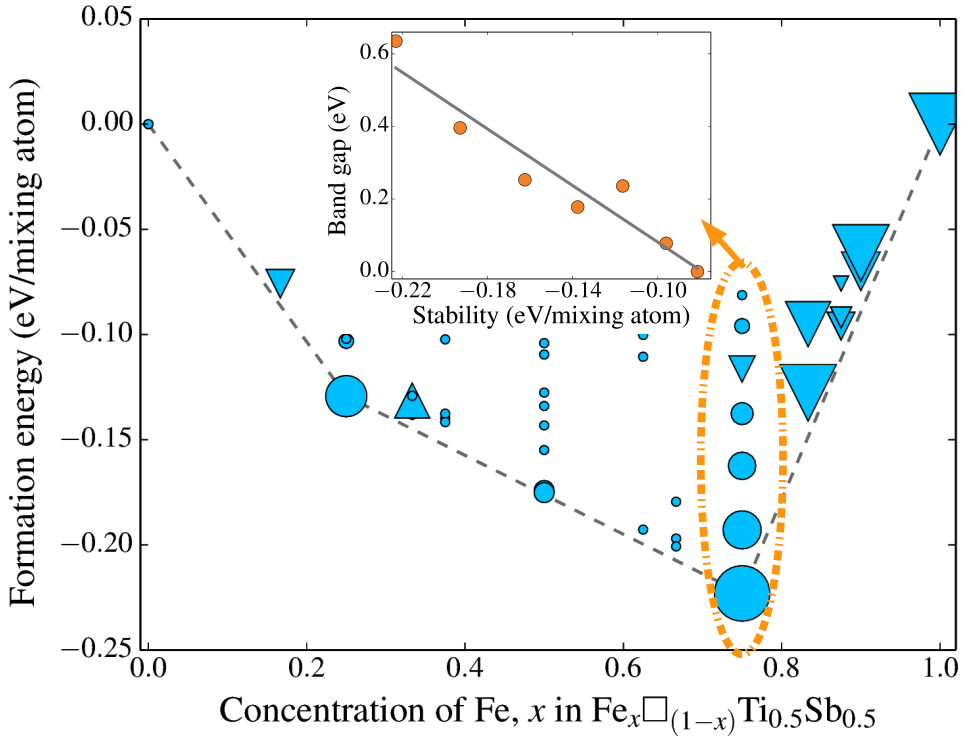

Figure 3.18. Correlation between band gap and stability for structures at various $Fe_x\square_{1-x}Ti_{0.5}Sb_{0.5}$ compositions: size of a marker is directly proportional to the band gap of the corresponding structure (band gaps vary from 0.0 to 0.75 eV). Tiny circles indicate a band gap of zero, bigger circles indicate gaps in both the spin channels, upward and downward facing triangles indicate gaps in the majority and minority spin channels respectively. Structures with a large band gap lie closer to the convex hull than those with a small or no gap. (Inset) The relationship between band gap and stability for structures with composition $x = 0.75$ is almost linear.

Second, we compare the orbital-projected density of electronic states of stable and nearly stable structures (within 100 meV/mixing atom of the convex hull) at the composition of the lowest energy structure, i.e., $Fe_{0.75}\square_{0.25}Ti_{0.5}Sb_{0.5}$ and find that (a) the top of the valence band and and bottom of the conduction band are predominantly Fe d in character (Figure 3.17), as in the full- and half-Heusler compounds, (b) gaps open in both the spin channels (Figure 3.17a-d) — unlike both the full-Heusler which has a gap in the

minority spin channel, and the half-Heusler which has a gap in the majority spin channel, and (c) this opening of gaps in both the spin channels seems to contribute to the stability of the structure, i.e., the structure with a larger band gap is also the more stable one. We observe that this correlation between band gap and stability is followed across almost all $Fe_x\square_{1-x}Ti_{0.5}Sb_{0.5}$ compositions (Figure 3.18): structures with larger band gaps seem to lie closer to the convex hull than those with small or no band gap, in general. It is also interesting to note that structures with gaps only in one spin channel ('half-metals') are slightly above the convex hull at all compositions (e.g., see the region between $x = 0.75$ and $x = 1.0$ in Figure 3.18). We provide a list of stable or nearly stable structures at various compositions, their band gaps and magnetizations in Table 3.4.

Extending the discussion of thermodynamic stability of a compound from Section 2.2.3 to the context of a cluster expansion, a negative formation energy of a substitutional compound with respect to the end members is a necessary but not sufficient condition for overall phase stability. Accordingly, we construct the convex hull of the Fe–Ti–Sb phase space with all the compounds reported in the ICSD and other hypothetical compounds derived from decoration of common structural prototypes, all energetic quantities queried from the OQMD. The addition of the ground state structures predicted from the cluster expansion alters the 0 K phase diagram—$Fe_{1.5}TiSb$ is now on the convex hull (i.e., predicted to be thermodynamically stable), rendering the half-Heusler FeTiSb unstable (off the convex hull by ∼35 meV/atom) (see Figure 3.19). Our approach and prediction were validated by the recent experimental synthesis of the $Fe_{1.5}TiSb$ compound [101]. In fact, attempts to synthesize the half-Heusler FeTiSb phase via arc-melting resulted

in a two-phase mixture of $Fe_{1.5}TiSb$ and a Fe-deficient phase whose structure was not characterized [101].

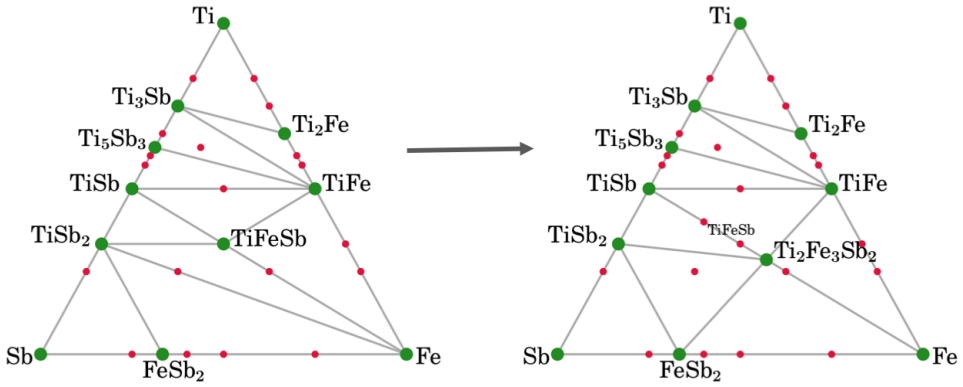

Figure 3.19. The 0 K phase diagram of the Fe–Ti–Sb system constructed using compound energies from the OQMD before (left) and after (right) including the ground state structures from cluster expansion. Green and red circles represent stable and unstable compounds respectively. $Fe_{1.5}TiSb$ is now on the convex hull (stable) rendering the half-Heusler FeTiSb unstable, and altering two neighboring three-phase regions.

3.2.5. Feedback between HT DFT and cluster expansion approaches

The successfully predicted Heusler-based compound $Fe_{1.5}TiSb$ occurs in a novel $R3m$ structure-type based on layering of the full- and half-Heusler primitive cells in the $\langle 111 \rangle$ direction (see Fig. 3.20). This novel structure-type can now be used as input to a prototype decoration + HT DFT approach to rapidly sample other chemistries to discover new stable compounds. We decorate the $X_{1.5}YZ$ $R3m$ prototype with elements from the periodic table such that the total number of electrons per formula unit (e^-/f.u.) is identical to the original $Fe_{1.5}TiSb$ compound (21 e^-/f.u.). We calculate the DFT formation energy of all such hypothetical compounds, a total of 1,470, within the framework of the OQMD and predict their thermodynamic stability against all other phases via construction of convex

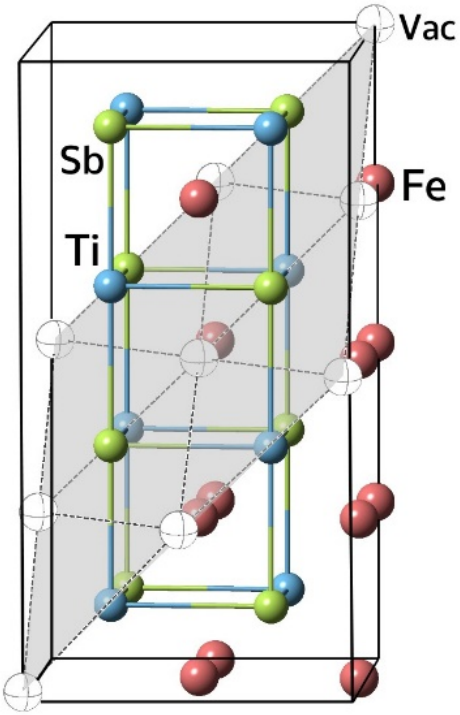

Figure 3.20. The $Fe_{1.5}TiSb$ compound, the "three-quarter Heusler" phase has a crystal structure that is a layering of the half- and full-Heusler primitive unit cells along the $\langle 111 \rangle$ direction.

hulls. Our preliminary non-exhaustive search yields 16 $X_{1.5}YZ$ compounds that lie on the respective convex hull (thermodynamically stable; see Table 3.6 for a partial list), and further 98 compounds that lie within 0.1 eV/atom.

Once interesting compounds have been identified, a more sophisticated structure search, e.g., cluster expansion or crystal structure prediction, can be performed for the system to accurately identify the ground state structures, if different from the $R3m$ prototype. Any such ground states occurring in new structure types can again be fed into a HT DFT loop, in a constant feedback mechanism.

$X_{1.5}YZ$	ΔE_f (eV/atom)	ΔE_{HD}	E_g (eV)
$Ru_{1.5}GeNb$	−0.386	−0.082	0.33
$Ru_{1.5}GeTa$	−0.414	−0.080	0.55
$Fe_{1.5}GeV$	−0.310	−0.065	0.29
$Ru_{1.5}SiTa$	−0.551	−0.042	0.51
$Ru_{1.5}SiNb$	−0.509	−0.037	0.32
$Co_{1.5}TiSn$	−0.450	−0.035	0.00
$Ni_{1.5}MgSn$	−0.348	−0.030	0.00
$Pd_{1.5}SrSn$	−0.680	−0.029	0.00
$Pd_{1.5}MgSn$	−0.675	−0.019	0.00
$Pd_{1.5}BaSn$	−0.650	−0.013	0.00

Table 3.6. $X_{1.5}YZ$ compounds predicted to be stable in the $R3m$ structural prototype. Columns correspond to the composition, formation energy, distance from the convex hull of *currently known phases*, and band gap.

CHAPTER 4

Exploring the high-pressure materials genome

4.1. Background

The laws of thermodynamics dictate that only compounds corresponding to global minima of the Gibbs free energy for a given set of external conditions are viable ground states with infinite lifetimes [102]. For such materials, there always exists a synthetic route that follows an overall exothermic chemical reaction pathway, and all systems at finite temperature will ultimately attain a Boltzmann distribution with a high occupation of the ground state in thermodynamic equilibrium. In practice, however, materials in many industrially relevant applications are metastable, i.e., they have higher energies than the equilibrium ground states. Such metastable phases, or polymorphs, correspond to local minima on the energy landscape and are surrounded by sufficiently high barriers to render them kinetically persistent on a finite time scale [103, 104].

Synthesizing metastable materials essentially requires finding, in some manner, a path in configurational space such that precursors undergo chemical reactions along a downhill trajectory with sufficiently low activation barriers, until the desired product is formed and quenched [105, 106]. A plethora of thermodynamic parameters can be tuned to design such a pathway, including temperature, pressure, electromagnetic fields, compositional variations, choosing specific precursor materials, etc. A special case of this design procedure is to choose a set of thermodynamic parameters such that the desired phase becomes

the thermodynamic ground state at the chosen conditions, where it forms at equilibrium, and can be recovered as a metastable phase at ambient conditions if all transition barriers leading away from it are sufficiently high [107].

This problem of identifying the ground states for a given set of external conditions is commonly tackled in the computational materials discovery community through global optimization of a target fitness function, using advanced crystal structure prediction (CSP) methods [108]. Ideally, this fitness function corresponds to the Gibbs free energy, but it is often approximated by the potential energy (at zero pressure, temperature) or the enthalpy (at zero temperature) or some other biased energy landscape, and is sampled in an unconstrained manner in the configurational space. Many novel materials and their structures have been resolved using CSP at high pressures [109–118], using chemical pressure and thermal degassing [119, 120], as 2-dimensional materials [121–123], or at surfaces and interfaces [124–128]. However, CSP approaches are computationally demanding and their applications are therefore often limited to small subsets of chemical spaces.

On the other hand, data-driven approaches using large materials databases in conjunction with high-throughput (HT) density functional theory (DFT) calculations have become increasingly popular in materials science [21, 129–133]. Such HT databases usually contain DFT-calculated properties such as formation energy, equilibrium volume, and relaxed atomic coordinates for experimentally reported phases available in repositories such as the Inorganic Crystal Structure Database (ICSD) [134]. These datasets are sometimes augmented with hypothetical compounds constructed by decorating common structural prototypes with elements in the periodic table. Subsequent phase stability analysis is often performed to identify stable phases in every chemical space. Although

approaches using such HT-DFT databases are useful for efficient large-scale analysis of energetics across a wide range of chemistries, they lack the power to predict novel materials with unknown crystal structures, and phases beyond ambient conditions since all such databases currently contain only materials properties calculated at zero temperature and zero pressure.

In particular, vast unexplored *pressure-composition* space is becoming increasingly accessible through large-volume press and diamond-anvil cell techniques [135], and improving predictive methods for high-pressure phases is, e.g., relevant to geophysical studies of planetary interiors where there can be numerous polymorphs energetically in close proximity, even in relatively simple compositional systems [136]. Increasingly, many phases created at high pressure can be recovered to ambient conditions (e.g. diamond, silicate perovskite, etc.), where they survive as metastable materials. Further, materials at high pressures show remarkable properties such as exotic stoichiometries and physics, high-temperature superconductivity, high energy density, superhardness, etc. [112]. Thus, a framework to rapidly assess phase stability and predict new materials at high-pressure has the potential for a broad impact ranging from geophysical sciences to uncovering new materials physics.

In this work, we effectively combine big-data in HT-DFT databases with CSP methods to unravel the "high-pressure fingerprint" of all materials and discover novel materials stable at non-ambient pressure conditions. Using the implicitly-available high-pressure information in a HT-DFT database, the Open Quantum Materials Database (OQMD), together with a simple approximation to the formation enthalpy of a compound, we study the effect of pressure on the thermodynamic scale of stability/metastability of inorganic

compounds. Our model correctly predicts most experimentally reported high-pressure elemental and binary phases to become thermodynamically stable at non-ambient pressures. Our statistical analysis of the data in the OQMD shows a large fraction (up to 60%) of ambient-metastable compounds to be thermodynamic ground states at non-zero pressures. Further, we use our model to sample *all* binary intermetallic chemical spaces with no experimentally reported compound in the OQMD (∼1780 chemical spaces) and predict nearly 3800 new compounds to be stable at some finite pressure. Finally, we demonstrate the power of our predictive framework in guiding sophisticated CSP methods by explicitly exploring ten binary-immiscible systems, and discover that our model correctly predicts phase spaces containing novel high-pressure materials, which could be potentially recovered to ambient conditions as metastable compounds.

4.1.1. Linear approximation to enthalpy

The thermodynamic stability of solids determined by the Gibbs free energy $G = E + pV - TS$ (where p and T are pressure and temperature, while E, V and S are the internal energy, volume, and entropy of a phase) is predominantly affected by entropic effects (TS) at ambient pressure since the energy scale of the pV term is rather small: $p\Delta V$ at atmospheric pressure ($p = 0.0001$ GPa) is merely 0.006 meV/atom even for a volume change of $\Delta V = 10$ Å^3/atom. However, this behavior shifts rapidly with pressures in the GPa range, where the effect of pV increases by orders of magnitude compared to entropic contributions. Hence, at zero temperature (neglecting TS), the Gibbs free energy for a given phase reduces to the enthalpy $H = E + pV$. Expanding H as a function of p around

the equilibrium pressure p_0 yields

$$\begin{aligned} H(p) &= H(p_0) + \Delta p\, H'(p_0) + \frac{(\Delta p)^2}{2} H''(p_0) + \cdots \\ &= H(p_0) + \Delta p\, V(p_0) - \frac{(\Delta p)^2}{2} \frac{V(p_0)}{B(p_0)} + \cdots \end{aligned} \tag{4.1}$$

where $\Delta p = (p - p_0)$, and $B = \frac{1}{\beta}$ is the bulk modulus of the phase, where $\beta = -\frac{1}{V}\frac{\partial V}{\partial p}$ is its compressibility. If we neglect all terms higher than second order and consider all phases to be incompressible (i.e., $B(p_0) \to \infty$), for equilibrium pressure $p_0 = 0$, we can approximate the enthalpy of a phase simply as

$$H(p) = E_0 + \Delta p\, V(p_0) \tag{4.2}$$

where E_0 is the internal energy at the equilibrium volume V_0. This approximation is used in CSP approaches to quickly screen large sets of generated candidate structures, and was first mentioned by Pickard and Needs [109]. Conveniently, both E_0 and V_0 are quantities that are also readily available for hundreds of thousands of phases in most HT-DFT materials databases such as the OQMD [129, 130], Materials Project [21], and AFLOWlib [131].

The above linear approximation to enthalpy (henceforth referred to as "LAE") is illustrated in an energy-volume diagram in Fig. 4.1a, where the ground state and two metastable states are each represented by their respective equation of states (EOS) $E(V)$, i.e., their energy as a function of volume, approximated by parabola. The negative slopes of the common tangents connecting the EOS of neighboring phases represent the pressure at which both phases are in equilibrium ("transition pressures", grey dashed lines). With

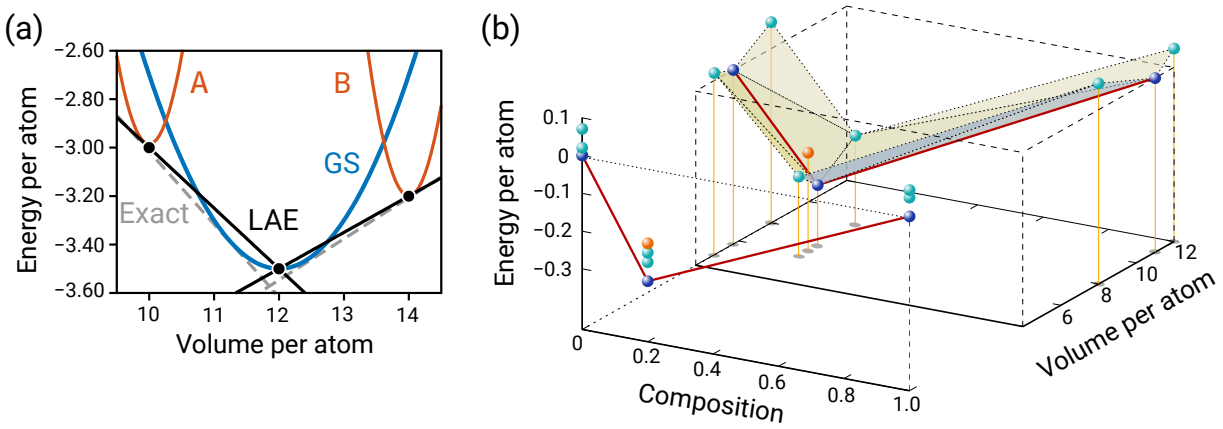

Figure 4.1. (a) A schematic energy-volume (V–E) diagram with three phases (the ground state GS, and two metastable phases A and B) and their EOS, each represented by a parabola. The negative slope of the common tangent to two adjacent EOS (dashed grey line) represents the pressure at which the two phases are in equilibrium. The LAE approximates the common tangent with a line connecting the ambient-condition equilibrium volumes/energies of two adjacent phases (solid black line connecting filled circles). (b) A schematic N–V–E convex hull for a model binary system. Individual phases are represented by spheres, and convex hull boundaries are indicated with solid red and dotted black lines. On the left is the conventional zero-pressure N–E hull, a projection of the extended N–V–E convex hull on the right. Phases that are thermodynamically stable at zero pressure lie on the N–E convex hull (blue spheres). Metastable phases that are stable at some non-ambient pressure lie *above* the N–E hull but *on* the N–V–E convex hull (teal spheres). A phase that is truly unstable at any pressure lies above the N–V–E hull (orange sphere).

our approximation of the bulk moduli $B(p_0) \to \infty$, the EOS curve of each phase would have infinitely large curvature, reducing the parabola to a vertical line originating at the corresponding equilibrium volumes V_0 and energies E_0. Essentially, all information of each phase is then contained in a single point at (V_0, E_0), represented by filled points.

Although the LAE is rather crude, it is reasonably accurate up to pressures in the range of tens, or even hundreds of GPa. As we will show in the rest of this work, the LAE can be used as a powerful tool to enable quick analyses of phase stability of a large

number of materials at non-ambient pressures. Note that we will hereafter use the terms "zero pressure" and "ambient pressure" interchangeably, since the pV contribution to the free energy at atmospheric pressure is insignificant for most inorganic compounds, as discussed earlier. E.g., at one atmosphere the energy contribution of pV in diamond silicon with a volume of $\sim$20 Å^3/atom is merely 0.012 meV/atom, far smaller than the error bars encountered in DFT calculations.

4.1.2. Thermodynamic stability: the convex hull

The thermodynamic stability of a phase at zero temperature can be determined by the construction of the so-called convex hull of all phases in the chemical space. At zero pressure, the convex hull is constructed from the composition and formation energy (composition-energy hull, or simply "N–E convex hull") of all the phases. By definition, a phase *on* the convex hull has a formation energy lower than that of any other phase (or linear combination of phases) at that composition, and is therefore thermodynamically stable. At non-ambient pressures, thermodynamic stability is determined by a convex hull which also takes into account the energy as a function of volume of all phases, given by their respective EOS $E(V)$. The LAE introduced in Section 4.1.1 allows us to simplify the construction of the convex hull by taking into account the ambient volume of each phase, in addition to their composition and formation energy (compositon-volume-energy hull, or simply "N–V–E convex hull"). A phase on the extended N–V–E hull has a formation energy lower than any other phase or combination of phases at that composition *and* volume, and is therefore thermodynamically stable at some pressure. Further, a tie line on

the convex hull represents a two-phase equilibrium, a triangular facet represents a three-phase equilibrium, and so on—a facet with n vertices represents an n-phase equilibrium.

A schematic N–V–E convex hull is shown in Fig. 4.1b. A projection of the extended N–V–E convex hull taking into account only the energy and volumes leads to the N–E hull (indicated by solid red lines). Phases that lie *above* the N–E hull, but *on* the N–V–E hull, are metastable at zero pressure but thermodynamically stable at some finite pressure. For example, in Fig. 4.1b, only two elemental phases and one binary compound (blue spheres) lie on the N–E hull (solid red lines), i.e., are thermodynamically stable at zero pressure, and all other phases are metastable. However, all elemental phases and all binary compounds except one lie on the extended N–V–E hull (teal spheres connected by dotted black lines), i.e., are thermodynamically stable at some non-ambient pressure. Only one phase shown (orange sphere at composition 0.2) is truly unstable at all pressures.

4.1.3. Pressure range of stability

For a system in thermodynamic equilibrium at zero temperature, $dE = -p\,dV + \sum_i \mu_i\,dN_i$, where dE, dV are infinitesimal changes in internal energy E, volume V of the system, respectively, and dN_i is the infinitesimal change in the composition N_i of species i. The equilibrium pressure is thus given by $p = -\left(\frac{\partial E}{\partial V}\right)_{N_i}$, i.e., the derivative of energy with respect to volume at constant composition. Hence, the pressure range of stability of a phase $\mathcal{P}$ with ambient equilibrium volume and energy of V_0 and E_0, respectively, is governed by the phase equilibria at volumes $(V_0 + dV)$ and $(V_0 - dV)$ [1]. In other words,

[1]This procedure is analogous to calculating the range of thermodynamic stability of a compound with respect to the chemical potential of a given species. Since the chemical potential of species i, at p,

the window of pressures $[p_-, p_+]$ where $\mathcal{P}$ is stable is given by

$$p_\pm = -\frac{E_0 - E(V_0 \mp dV)}{dV} \tag{4.3}$$

$E(V_0 \pm dV)$ can be calculated by minimizing the free energy of the system at the target composition *and* volume. Grand canonical linear programming (GCLP) [84] techniques using efficient linear solvers are routinely employed to calculate phase stabilities and equilibrium reaction pathways at 0 K and 0 GPa [2, 25, 31, 137, 138]. In this work, in addition to the average composition of the system being constrained to that of $\mathcal{P}$, the volume is constrained to $V_0 \pm dV$ during energy minimization. Thus, a pressure range of stability can be calculated for every phase that lies on the extended N–V–E convex hull.

As discussed in Sec. 4.1.1, the negative slope of the common tangent to the EOS of two phases is the pressure at which the respective phases coexist, or in other words, one phase transforms into the other under the effect of pressure. In the LAE, the common tangent is reduced to a line connecting the local minima of the two phases (solid black line connecting filled circles in Fig. 4.1a). The LAE introduces errors compared to the real transition pressure, which depend on the overall features of the energy landscape. If we assume that all phases are *compressible* with identical, *finite* bulk moduli, the LAE will consistently lead to an *underestimation* of the magnitude of the transition pressures. In practice, however, high-pressure phases often exhibit shorter, stronger bonds that lead to *higher* bulk moduli. Hence, the LAE would lead to a better agreement with the real transition pressures for phases stable at very high pressures. On the other hand, if the bulk

---

$T = 0$, is given by $\mu_i = \left(\frac{\partial E}{\partial N_i}\right)_{N_{j\neq i}}$, the window of stability $[\mu_i^-, \mu_i^+]$ of the phase can be calculated using $\mu_i^\pm = \frac{E_0 - E(N_i \mp dN_i)}{dN_i}$, i.e., as a perturbation of energy with respect to the composition of the species i.

moduli significantly *decreased* with pressure, the LAE would lead to an *overestimation* of the magnitude of the transition pressures. We also note that transition pressures, based on the above definition, can be positive or negative (e.g., the common tangents connecting the ground state GS with metastable phases A and B, respectively in Fig. 4.1a). A negative pressure can be physically interpreted as a tensile stress, leading to the expansion of a phase toward volumes exceeding its ambient ground state equilibrium volume.

4.2. Results and Discussion

4.2.1. Model assessment

We first evaluate the accuracy of the linear approximation to enthalpy by investigating two elements and five binary systems in detail.

4.2.1.1. Elemental solids. We choose two elements whose high-pressure phase diagrams are among the most complex as well as the most well-studied: silicon and bismuth. Both elements have intricate energy landscapes with several high-pressure allotropes.

a. Silicon

The phase diagram of silicon has been well explored experimentally, partially due to its importance in the semiconductor industry. The ambient ground state is Si-I, which crystallizes in a cubic diamond structure [5]. It transforms around 11 GPa to the Si-II phase, which has a β-Sn structure [6]. This is followed by a transformation to Si-XI with *Imma* symmetry [7] at 13 GPa. Above 16 GPa, Si-V forms in the simple hexagonal structure [8], and at 38 GPa, Si-VI forms in an orthorhombic *Cmcm* structure [9]. The

hexagonal close-packed Si-VII forms above 42 GPa [10], and finally, cubic close-packed Si-X forms at pressures above 78 GPa [11].

We first compute the pressure range of stability of the various silicon allotropes using DFT calculations (see the top bar labeled "DFT" in Fig. 4.2a). For each phase, we calculate the enthalpy explicitly at various pressures at intervals of 2 GPa and 10 GPa in the range of 0–20 GPa and 20–100 GPa, respectively. The transition pressures are then computed by minimizing the interpolated formation enthalpies as a function of pressure. This approach is more accurate than fitting an equation of state (EOS) using energies computed at scaled volumes, since (a) the EOS fit would depend on the functional form (e.g., Vinet, Murnaghan, Birch-Murnaghan, etc.), and (b) the simple scaling of volume assumes isotropic compressibility. We employ above approach for all phase diagrams denoted hereafter with "DFT".

The experimentally reported sequence of formation and transition pressures of high-pressure Si allotropes are well reproduced, with the exception of Si-II, which is effectively degenerate in enthalpy to Si-XI. The discrepancy between experiment and theory for the transition from Si-I to Si-II has been well studied [139, 140], and is attributed to the errors associated with the PBE approximation to the DFT exchange correlation potential.

We then calculate the pressure range of stability of all the allotropes using only the respective equilibrium energies and volumes at 0 GPa, extrapolated linearly as described in Sections 4.1.1–4.1.3 (see the second bar labeled "LAE" in Fig. 4.2a). The agreement between the "DFT" and "LAE" phase diagrams is remarkable: (a) the sequence of the phases is correctly reproduced, with the only exception of Si-X, which the linear approximation model predicts to be unstable even at 100 GPa, and (b) the overall errors in the

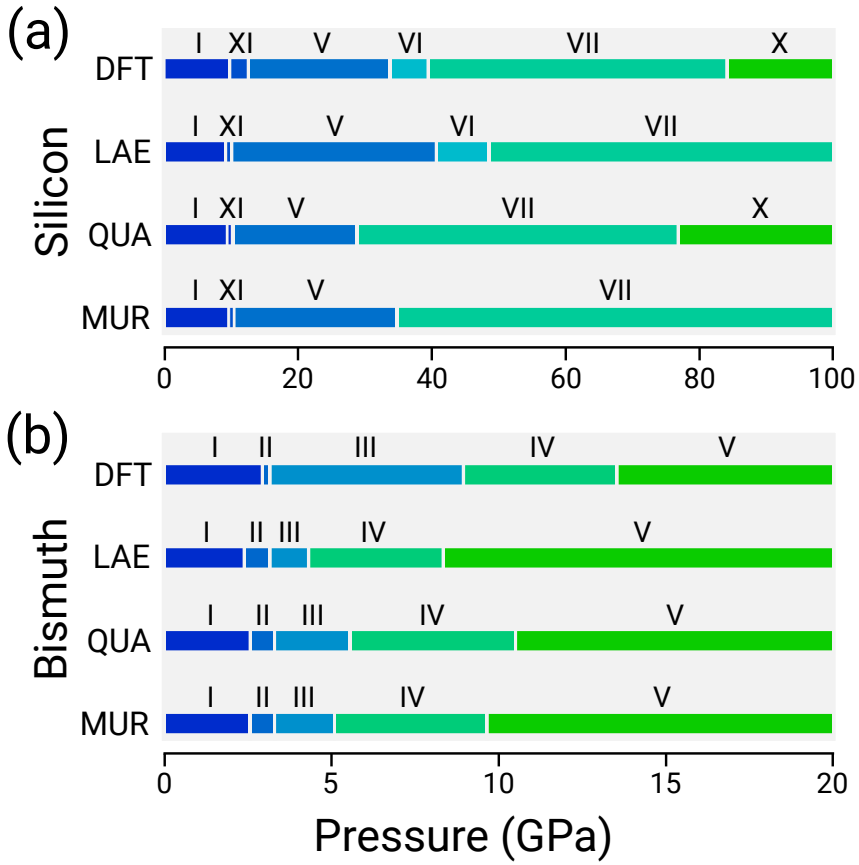

Figure 4.2. The pressure range of stability of high-pressure phases of elemental (a) silicon and (b) bismuth, respectively. Explicitly computed transition pressures using DFT-calculated formation enthalpies are labeled "DFT" (top bar in each panel), and those based on the LAE are denoted with "LAE" (second bar in each panel). In addition, we show transition pressures calculated using the expansion of the enthalpy to second order ("QUA"), and the Murnaghan EOS ("MUR") [3, 4], for comparison. The crystal structures of the silicon and bismuth allotropes were taken from the Refs. [5–11] and [12–18], respectively.

transition pressures predicted by the approximate model are within around 10% of those calculated explicitly.

Additionally, we compare the results from LAE with those from fitted EOS. For this purpose, we scale the lattice vectors of every structure relaxed at 0 GPa by factors

of $x = \{0.95, 0.97, 0.98, 0.99, 1.00, 1.01, 1.02, 1.03, 1.05\}$ and compute the corresponding energies with a single point DFT calculation. The energy-volume data obtained in this way is used to fit two EOS: (a) a quadratic polynomial corresponding to the second order expansion in Equation (4.1) ("QUA", third bar), and (b) the well-known Murnaghan EOS [3, 4] ("MUR", bottom bar). The pressure range of stability calculated using the EOS shows an overall trend and errors comparable to that calculated using the LAE. The quadratic fit fails to predict the stability of Si-VI, while it correctly assigns a pressure range below 100 GPa for Si-X. The Murnaghan EOS performs even worse: it both misses the stability of Si-VI and the onset pressure of Si-X. In other words, using sophisticated EOS does not necessarily improve the predicted phase diagram compared to the LAE, especially given the addition computational cost associated with it [2].

[2]On our HPC infrastructure, the explicit calculation of the enthalpies at different pressures for all relevant Si phases requires ≈ 1120 core hours. This is by far the most expensive, but also most accurate approach, since all degrees of freedom (atomic coordinates and cell parameters) are fully relaxed at relevant pressures. The LAE on the other hand only requires the relaxed cell parameters and energies at 0 GPa, which are computed within ≈ 70 core hours. For fitting the EOS, these relaxed cells are scaled to discrete volumes and the corresponding energies calculated with single-point DFT calculations, a task completed within $\approx 70 + 140 = 210$ core hours (i.e., including the initial relaxation at 0 GPa).

b. Bismuth

At ambient condition, bismuth crystallizes in a rhombohedral Bi-I phase with space group $R\bar{3}m$. It transforms at a pressure of around 2.55 GPa to Bi-II with a $C2/m$ structure [12, 13] and a very narrow range of stability at low temperatures. Upon increasing the pressure, Bi-III forms in a complicated, incommensurate host-guest structure with $P4/ncc$ symmetry [14–16]. A Bi-IV phase with space group $P21/n$ has been reported between 2.4 GPa and 5.3 GPa at temperatures above around 450 K [17]. Finally, the Bi-V bcc phase is observed at pressures above 7.7 GPa [18].

Similar to the case of silicon, we first compute the pressure range of stability of the various bismuth allotropes using enthalpies calculated explicitly at various pressures at intervals of 1 GPa in the range of 0–20 GPa (see the top bar labeled "DFT" in Fig. 4.2b). Although the experimentally reported sequence of allotropes formed is well reproduced, the transition pressures between Bi-III/Bi-IV and Bi-IV/Bi-V are severely overestimated. This behavior has been reported previously by Häussermann et al. [16], and corroborated in our recent work on Cu–Bi intermetallics [141, 142].

The pressure range of stability of all allotropes calculated using the LAE reproduces the correct sequence of formation (see the second bar labeled "LAE" in Fig. 4.2b). However, the agreement between the transition pressures predicted by the approximate model and those calculated explicitly are worse than that for silicon allotropes. We attribute these larger errors to the strong changes in the chemical bonds between the different bismuth phases, especially since ambient Bi-I has a layered structure, in contrast to the high-pressure phases. Hence, our approximation of equal, infinitely large bulk moduli for every phase is perhaps less reasonable for elemental phases of bismuth.

Similar to silicon, more advanced EOS fitted to energy-volume data do not significantly improve the description of the phase diagram. Neither the fit to a quadratic polynomial (QUA) nor the Murnaghan EOS (MUR) provides a much better agreement with the exact DFT transition pressures: the transition pressures between Bi-III/Bi-IV and Bi-IV/Bi-V are in a similar range as the ones from LAE. Overall, these results clearly demonstrate that the LAE can capture the general features that determine high-pressure phase stability at a comparable accuracy to fits of EOS, at a significantly lower computational cost.

4.2.1.2. Binary intermetallics. When compared to pure elements, the high-pressure phase space of binary/higher-order chemical systems have been experimentally relatively unexplored. Including composition and pressure as additional degrees of freedom significantly increases the complexity of the phase space. In this section, we focus on a unique subset of chemistries: intermetallic systems of elements that are not miscible at ambient conditions but form compounds under pressure. Many of these so-called ambient-immiscible systems involve bismuth in combination with other elements. Recently, we investigated three such systems in detail, namely Fe–Bi [143], Cu–Bi [141, 142], and Ni–Bi [144], by performing extensive global structure searches. Here, we use these three systems to further evaluate the performance of the linear approximation to enthalpy.

a. Fe–Bi

Using the minima hopping crystal structure prediction method (MHM), we recently predicted a high-pressure $FeBi_2$ phase with *I4/mcm* symmetry at pressures above 36 GPa [143], which was subsequently experimentally confirmed through evidences found in the *in-situ*

X-ray diffraction pattern at above 30 GPa [111]. We note that the discovery of $FeBi_2$ resulted from extensive MHM structural searches performed at pressures of 0, 50 and 100 GPa. The most promising candidate structures were then relaxed at pressure intervals of 10 GPa to compute enthalpies, which were in turn used to calculate the pressure range of stability of various phases (see top panel in Fig. 4.3a). Besides the $FeBi_2$ *I4/mcm* phase, we find a $FeBi_3$ phase with the *Cmcm* symmetry to be stable in a very small pressure window slightly below 40 GPa. This phase has so far not been observed in experiment.

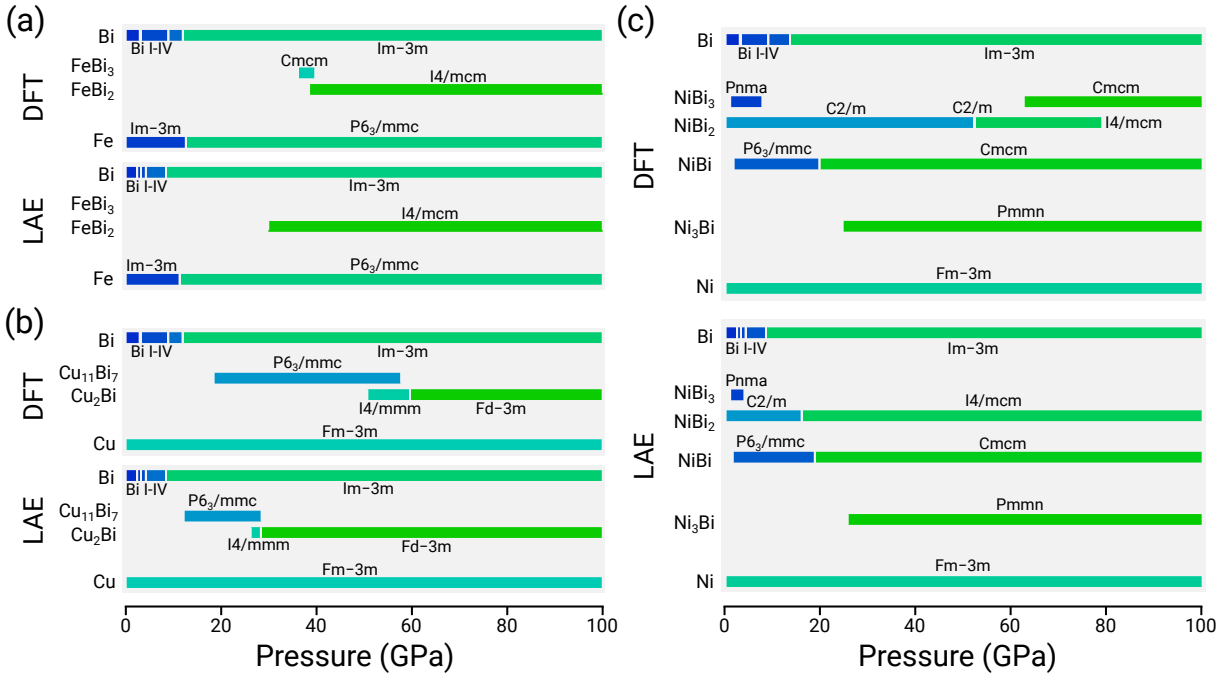

Figure 4.3. Comparison between the explicitly computed phase diagrams with the ones derived from the LAE model for binary systems. The panels (a), (b) and (c) correspond to the Fe–Bi, Cu–Bi and Ni–Bi systems, respectively. Explicitly calculated transition pressures using DFT are denoted with "DFT" (top bar), and results based on the LAE are denoted with "LAE" (bottom bar).

We now compare the pressure range of stability calculated explicitly above against that calculated using the linear approximation to enthalpy, using only the ambient equilibrium

energies and volumes of the phases. The phase diagram predicted by the approximate model (bottom panel in Fig. 4.3a) is qualitatively similar to the exact one: the $FeBi_2$ *I4/mcm* phase becomes stable at comparable pressures. This finding can be conveniently exploited in structural searches: since the MHM samples many low-lying metastable structures at a fixed pressure p_0, one could use the energies and volumes at p_0 of such phases within the LAE to quickly predict if any of the metastable phases become stable at a different pressure $p \neq p_0$. Even for immiscible systems at p_0, potential candidate structures are found if the simulation cells are sufficiently small to prevent phase segregation. This means that a structural search conducted solely at 0 GPa might have been sufficient to uncover the *I4/mcm* structure and correctly predict the experimentally observed $FeBi_2$ phase. The $FeBi_3$ phase, on the other hand, with the narrow pressure window of stability is predicted to be unstable at all pressures by the approximate model; this behavior can be attributed to the approximations inherent to the LAE. The overall good agreement between the exact and approximate phase diagrams is however rather surprising: $FeBi_2$ undergoes a series of magnetic transitions between 0 and 40 GPa, accompanied by abrupt changes in the unit cell volume [143], all of which are neglected in the linear approximation to enthalpy.

b. Cu–Bi

In the ambient-immiscible Cu–Bi system, at least two compounds, with compositions $Cu_{11}Bi_7$ and CuBi, have been recently discovered in diamond anvil cell (DAC) experiments between 3 and 6 GPa [141, 142]. Both phases can be recovered to ambient conditions, and exhibit exciting superconducting and structural properties. For example, CuBi has a

layered structure, rather uncommon for high-pressure phases, and is calculated to have an extremely low energy cost associated with exfoliation from bulk into single sheets [145]. Further, more recent structural searches predict additional, dense Cu_2Bi phases to become thermodynamically accessible at pressures of above 50 GPa [146] (recent, unpublished results show that a Cu_7Bi_2 phase with *C2/m* symmetry might be even lower in enthalpy. However, this structure has not been included in the current phase analysis and will be published separately).

The top panel in Fig. 4.3b shows the pressure range of stability of the various high-pressure Cu–Bi phases computed using explicitly calculated enthalpies for each phase. The CuBi phase is not thermodynamically stable at any pressure at zero temperature, consistent with recent reports of vibrational entropy playing a crucial role in rendering this phase synthesizeable [142]. The $Cu_{11}Bi_7$ phase is thermodynamically accessible at high pressures up to around 60 GPa, when it starts to compete with two dense Cu_2Bi phases [146].

The bottom panel in Fig. 4.3b shows the Cu–Bi phase diagram computed from the LAE, using only the respective equilibrium energy and volume of each phase at 0 GPa. All phases are correctly predicted to be stable by the approximate model, consistent with the exact phase diagram. As expected, the transition pressures predicted by the approximate model are underestimated overall when compared to those calculated explicitly—a trend that is presumably increased due to the significant structural changes in elemental bismuth as a function of pressure (see Section 4.2.1.1). Nonetheless, it is striking that, using the simple linear approximation to enthalpy, we could have correctly predicted *all*

the high-pressure phases in the Cu–Bi system from a structural search only at 0 GPa.

c. Ni–Bi

We tested for the first time the predictive power of our model by investigating the high-pressure phases in the Ni–Bi binary intermetallic system. Two compounds have been experimentally reported at ambient pressures: NiBi in the hexagonal NiAs structure [147], and $NiBi_3$ in the orthogonal $RhBi_3$ structure [148, 149]. Both compounds are superconductors with transition temperatures of 4.25 K and 4.06 K in NiBi [150] and $NiBi_3$ [151, 152], respectively. To generate phase data to be used within the LAE to construct the convex hull and predict transition pressures, we used prototypes from our previous structural searches of the Fe–Bi and Cu–Bi systems, and substituted the Fe/Cu sites with Ni atoms, followed by structural relaxation at ambient pressures. Using this ambient-pressure dataset of energies and volumes, the LAE model predicted stable compounds at high-pressure for the compositions Ni_3Bi and $NiBi_2$. Based on this prediction, we performed a thorough investigation of the Ni–Bi system using MHM simulations at pressures of 10 and 50 GPa, which indeed revealed a number of high-pressure phases.

In particular, our calculations predict new compounds stable at high-pressure at compositions of the previously reported ambient-pressure phases, i.e., NiBi and $NiBi_3$. The hexagonal α-NiBi phase undergoes a structural transition to a TlI-type structure with $Cmcm$ symmetry at pressures above around 20 GPa. Similarly, the orthorhombic $NiBi_3$ phase is thermodynamically unstable above 7.5 GPa, and a $Cmcm$ structure is stable above 62 GPa. Further, we discover additional stable phases at previously unexplored compositions. We find that a $NiBi_2$ phase with $C2/m$ symmetry in the $PdBi_2$ structure

type is in fact thermodynamically stable at ambient pressures, a finding that was reported earlier by Bachhuber *et al.* [153]. At the same composition, a second $C2/m$ phase becomes stable above 52 GPa, over a very small pressure window of less than 1 GPa, followed by a $I4/mcm$ phase, isostructural to $FeBi_2$. Finally, a Ni_3Bi compound with $Pmmn$ symmetry, isostructural to Ni_3Sb in the Cu_3Ti structure type, is predicted to be stable at pressures above 25 GPa.

One of our predictions was very recently verified by compressing NiBi in a DAC. Heating to temperatures above 700°C at pressures above $\approx$ 28 GPa, the hexagonal α-NiBi transforms into β-NiBi in the predicted TlI structure type [144]. The experimental transition pressure is somewhat higher than the calculated value of 20 GPa. This discrepancy could be attributed to the presence of high kinetic reaction barriers in the first-order phase transition, which requires heat to induce the phase change and inevitably leads to calculated transition pressures being lower than those observed in experiment. This hypothesis is supported by detectable evidences of the β-NiBi phase in the XRD pattern upon decompression: the β-NiBi is kinetically persistent as low as 11.62 GPa, hence the equilibrium pressure lies anywhere between 11.62 and 28.3 GPa. In addition, errors inherent to the approximations used in DFT calculations could also explain the difference in the observed and computed transition pressures. The approximations to the exchange correlation potential alone can make a noticeable difference. E.g., the PBE functional predicts that both the experimentally observed NiBi and $NiBi_3$ phases (in their reported structures) are not thermodynamically stable at 0 GPa and 0 K. However, we find that LDA correctly places the two experimental phases on the 0 GPa convex hull, and if we additionally take into account the vibrational entropy contributions to the free energy,

$NiBi_2$ becomes unstable at elevated temperatures. A detailed investigation of the influence of different exchange correlation potentials and temperature effects on the calculated phase stability of Ni–Bi compounds and their properties will be reported elsewhere.

After exploring the high-pressure Ni–Bi system with the MHM, we *a posteriori* compare the phase diagram of Ni–Bi computed using the explicitly calculated enthalpies against that predicted from our LAE model (Fig. 4.3c), and find remarkable agreement. Most phases, and the sequence in which they form under pressure, are correctly predicted by the approximate model. The only exceptions are the $Cmcm$ phase at the $NiBi_3$ composition and the second $C2/m$ compound at the $NiBi_2$ composition at around 50 GPa.

To assess the source of this discrepancy we investigate the convex hull of the Ni–Bi system at 70 GPa, a pressure where the $Cmcm$ phase of $NiBi_3$ and $I4/mcm$ phase of $NiBi_2$ overlap the most (see Fig. 4.4). We see that the $Cmcm$ and $I4/mcm$ phases barely lie on the convex hull of stability, and the energy difference compared to the two-phase mixture of elemental Bi and the $Cmcm$ phase of NiBi is extremely small. This delicate feature of the enthalpy landscape in the Ni–Bi system is hard to capture even with higher order approximations to the enthalpy. As an additional challenge, the $Cmcm$ NiBi phase has a strongly anisotropic compressibility: at 100 GPa, the lattice parameters a, b and c are reduced by factors of 0.813, 0.830 and 0.890, respectively, when compared to those at 0 GPa. Therefore, neither the LAE nor any simple EOS fit can correctly reproduce the exact DFT enthalpy at high pressure—they overestimate the enthalpy, moving the phase off the convex hull. Nevertheless, we recompute the complete phase diagram of the Ni–Bi system by fitting the Murnaghan EOS to the energies at scaled volumes of every relevant

phase (see Fig. 4.5). As expected, the Murnaghan EOS fails to capture the phase stability of *Cmcm* $NiBi_3$, similar to the LAE.

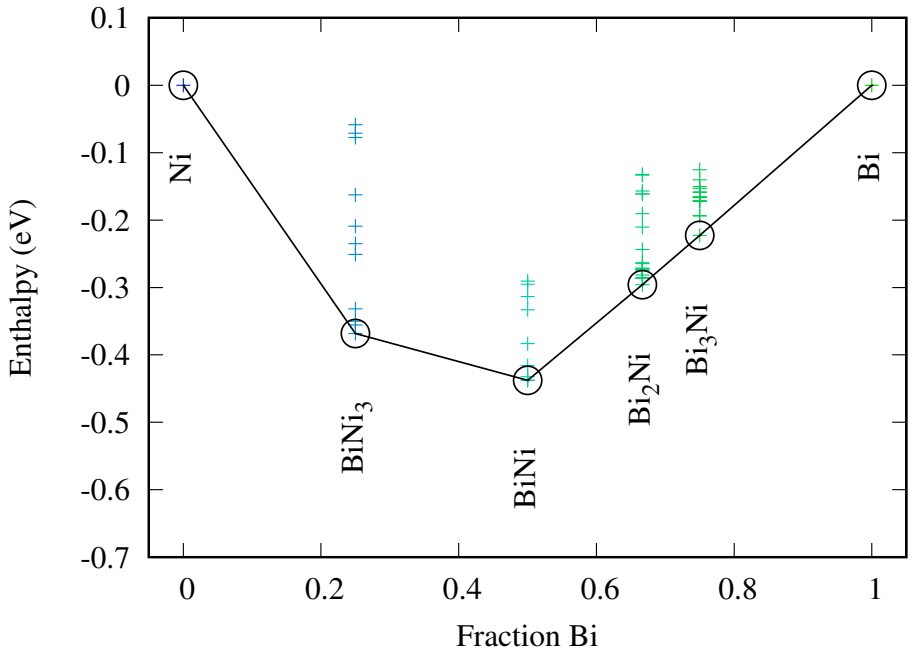

Figure 4.4. The convex hull of stability of the Ni–Bi system at 70 GPa. Note how Bi_3Ni and Bi_2Ni barely lie on the convex hull.

As discussed earlier, the $C2/m$ phase at the $NiBi_2$ composition has a very small pressure range of stability of <1 GPa, so its absence in the phase diagram predicted by the LAE is not surprising. The phase diagram computed using the Murnaghan EOS misses this phase as well. In fact, similar to the *Cmcm* $FeBi_3$ phase that was predicted to be stable in a narrow pressure window of less than 3 GPa but not yet observed experimentally, synthesis of the $NiBi_2$ phase is likely to be challenging, if possible at all. Overall, stability predictions using the LAE and the Murnaghan EOS are very comparable, similar to our findings for the elemental phase diagrams.

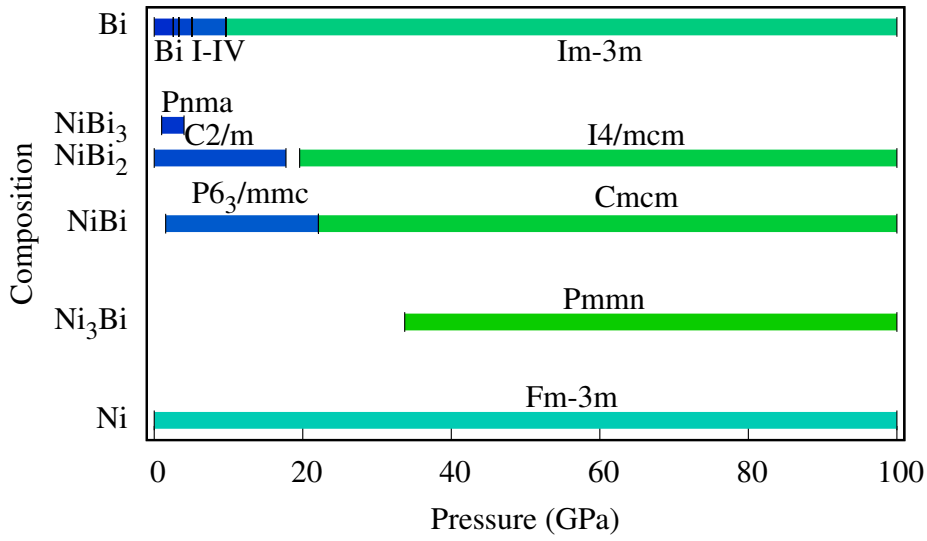

Figure 4.5. The pressure ranges of stability in the Ni–Bi system, computed by fitting the Murnaghan EOS to energy-volume data of every phase. All structures were first relaxed at 0 GPa, then the lattice vectors were scaled with factors of $x = 0.95, 0.97, 0.98, 0.99, 1.00, 1.01, 1.02, 1.03, 1.05$ before performing single point DFT calculations to obtain $E(V)$.

4.2.1.3. Binary oxides. Next, we assess the performance of the LAE in predicting the stability and the transition pressures of phases in two ceramic oxide systems with several experimentally studied high-pressure phases: Zr–O and Ge–O.

a. Zr–O

The phase diagram of Zr–O is complex, where the constituent elements themselves exhibit various thermodynamically stable phases as a function of pressure. Experimentally, oxygen crystallizes in the monoclinic α-phase at ambient pressure and low temperatures [154], transforming into the orthorhombic δ-phase at 3 GPa [155], followed by the monoclinic ϵ-phase at 10 GPa [156, 157], and finally the ζ-phase above 96 GPa through a continuous displacive and isosymmetric transition (ϵ-O is essentially isostructural to ζ-O) [158, 159].

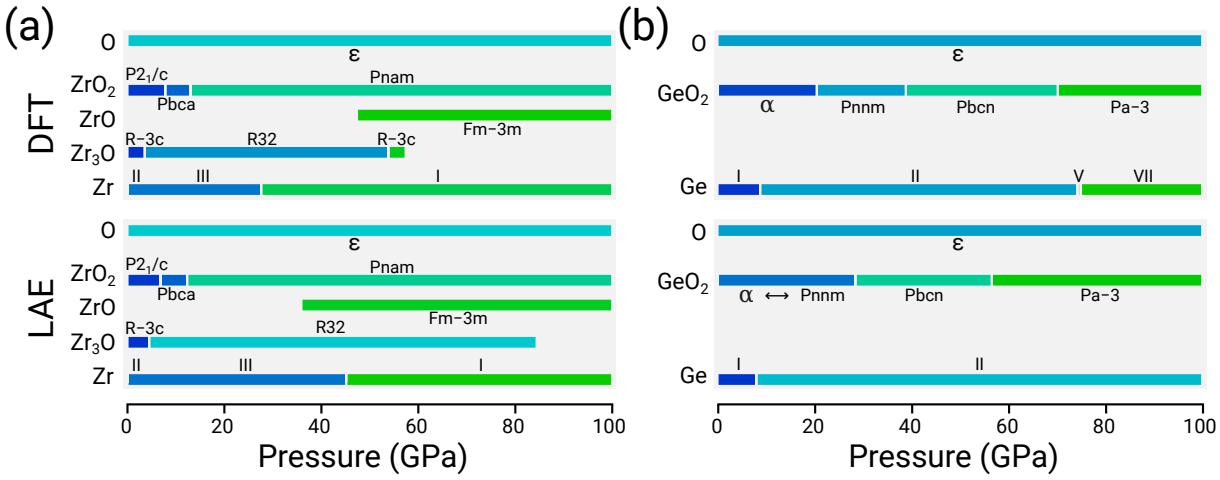

Figure 4.6. Comparison between the explicitly computed phase diagrams and the ones calculated using the LAE model for two oxide systems. The panels (a) and (b) correspond to the Zr–O and the Ge–O systems, respectively. Explicitly calculated transition pressures using DFT are denoted with "DFT " (top bar), and results based on the LAE are denoted with "LAE" (bottom bar). The right arrow denotes the displacive phase transformation.

However, based on our PBE-DFT calculations, only the ϵ-phase (ζ-O) is thermodynamically stable over the whole pressure range as we neglect magnetic ordering which persists in the condensed low-pressure regimes of O_2 (see Fig. 4.6).

At ambient pressure, elemental Zr crystallizes in its low-temperature phase II, which has an hcp structure. At 2.2 GPa, a transformation to phase III with a so-called ω structure of hcp type [160] is observed, and at around 30 GPa Zr adopts a bcc structure (phase I) [161]. The top panel in Fig. 4.6a shows that our exact DFT calculations reproduce these two transitions with reasonable accuracy.

In the binary Zr–O system, ZrO_2 adopts a baddeleyite type structure with $P2_1/c$ symmetry at ambient pressure, which transforms into the orthorhombic *Pbca* and *Pnma* phases at pressures of 3 GPa and 13-30 GPa, respectively [162, 163]. Further, the suboxide

Zr_3O was observed in two structure types with *R-3c* and *R32* symmetries, while the formation of ZrO was reported in the NaCl structure with *Fm-3m* symmetry [164]. The latter phase becomes indeed thermodynamically stable, but only at pressures above ≈ 45 GPa.

A comparison between the exact phase diagram of the Zr–O system with that calculated using the LAE (Fig. 4.6a) shows agreement. The only major discrepancy is found for Zr_3O: from the explicit calculations of enthalpy, the thermodynamically stable low-pressure phase with *R-3c* symmetry loops back and becomes stable again, above ≈ 55 GPa. By definition, such a behavior cannot be predicted within the LAE model.

b. Ge–O

The phase diagram of elemental Ge has been experimentally extensively explored, but the exact sequence and structures of some high-pressure phases have not yet been conclusively determined. At ambient pressure, phase I adopts the diamond structure [165], which transforms to phase II with a β-Sn structure type between 6.7 and 10.5 GPa, depending on the amount of shear stress [166]. Above 74 GPa, phase V forms in the simple hexagonal structure [167], which transforms to an orthorhombic structure with *Cmcm* symmetry (phase VII) around 100 GPa [168], and subsequently to an hcp structure at 160-180 GPa (phase VIII) [169]. Additionally, a series of metastable, intermediate Ge phases have been reported [170]. The formation sequence of the high-pressure phases computed from the exact enthalpies in Fig. 4.6b agrees well with results in the literature [171]. However, the approximate LAE phase diagram incorrectly predicts phase II to be stable up to >100 GPa.

Similar to SiO_2, there are several polymorphs of GeO_2. At ambient pressure, the most stable form is the α-quartz ground state structure [172], which transforms to the rutile-type structure with $P4_2/mnm$ symmetry below 2 GPa [173]. A displacive phase transformation to a $CaCl_2$-type structure with $Pnnm$ symmetry has been reported at $\approx$ 20 GPa, followed by a transition to a α-PbO_2-type structure ($Pbcn$ symmetry) above $\approx$ 33 GPa and a Pyrite-type structure ($Pa\text{-}3$ symmetry) [174, 175].

The top panel in Fig. 4.6b shows that the exact DFT results are in excellent agreement with the experimental observations. The LAE closely reproduces the exact phase diagram, with the exception of the transition from the $P4_2/mnm$ to the $Pnnm$ phase. This is expected as the phase transition is of second order and a DFT relaxation of both structures at 0 GPa results in identical equilibrium energies and volumes.

4.2.2. Large-scale analysis of phase stability at high pressure

4.2.2.1. Elements and binary compounds. The power of our linear enthalpy model lies in its capability to efficiently assess the pressure range of stability of hundreds of thousands of phases. Since the linear approximation requires only equilibrium energies and volumes of phases calculated at ambient pressure, it can be used to leverage the large materials datasets available in HT-DFT databases such as the OQMD [129, 130], Materials Project [21], and AFLOWlib [131]. Here, we present large-scale analysis and statistics of thermodynamic phase stability of materials at high pressure using ambient-pressure phase data calculated in the OQMD.

First, we focus on elemental high-pressure phases, and begin by compiling a "validation-dataset" of experimentally reported high-pressure elemental phases. The crystal structures of many high-pressure phases reported in the Inorganic Crystal Structure Database (ICSD) [134] have been calculated in the OQMD, albeit at ambient pressure. For every element, we filter all entries in the ICSD using the "External Conditions → Pressure" metadata available for each entry. Further, Tonkov et al. [170] compiled a comprehensive list of phase transformations under pressure for nearly 100 elements, on which we rely heavily as a second reference to cross-validate and augment the list of high-pressure phases calculated in the OQMD. Our final compiled dataset contains 132 distinct elemental high-pressure phases. The dataset entire dataset is provided in Table A.1, and a comparison of LAE-predicted vs experimental transition pressures is shown in Fig. 4.7.

For each element in the periodic table, we use the ambient-pressure energy and volume data for *all* ICSD phases (i.e., not limited to high-pressure phases) calculated in the OQMD within the LAE model to predict (a) the number of phases from our validation-dataset that lie on the extended N–V–E convex hull, i.e., the number of phases stable at some finite pressure, and (b) the pressure range of stability of every phase that lies on the N–V–E hull. Fig. 4.8 shows a summary of this analysis in the form of a periodic table: for every element with at least one experimentally reported high-pressure phase, we indicate the number of high-pressure phases in our compiled dataset from the OQMD (bottom-left half) and the number of phases predicted by the linear enthalpy model to lie on the N–V–E hull (top-right half), represented on a color scale. That is, the number of phases reported experimentally and those predicted to be stable at some finite-pressure match exactly whenever the colors in both the left and right segments are identical. This is indeed the

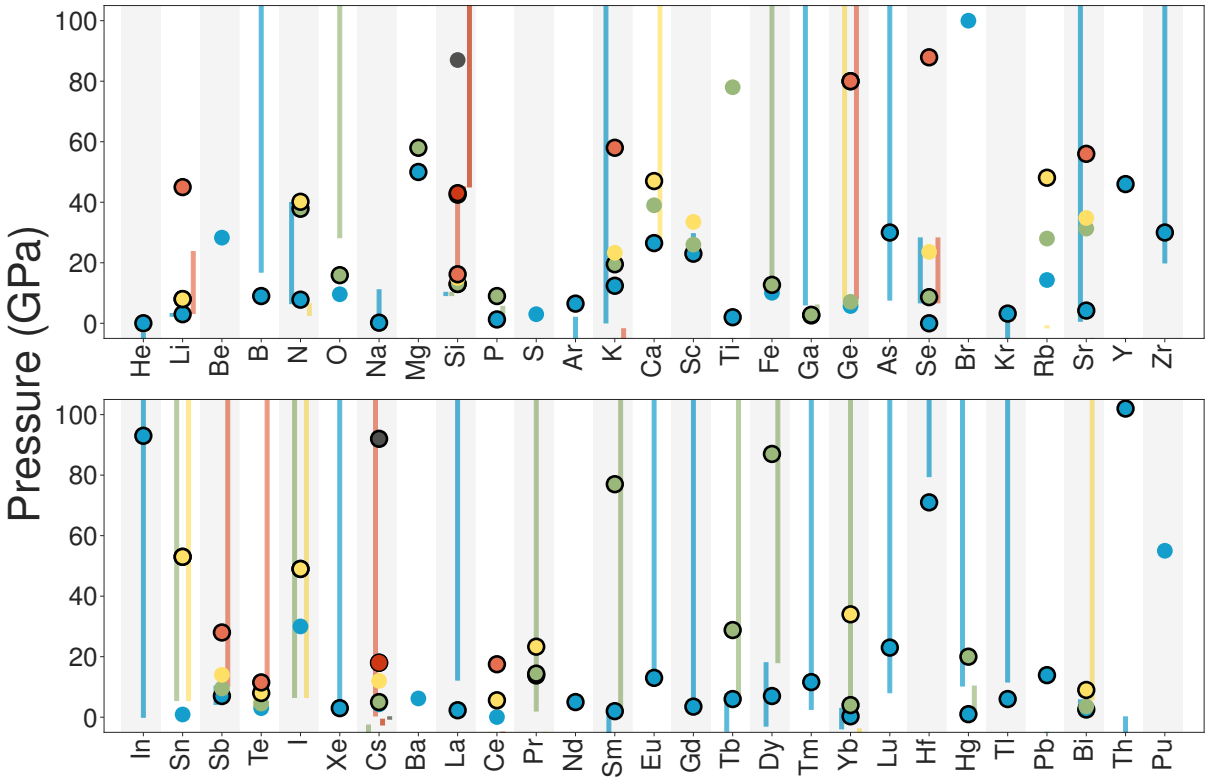

Figure 4.7. Accuracy of the linear enthalpy model in predicting the stability of experimentally reported elemental high-pressure phases. For each element, observed high-pressure phases with transition pressures within 100 GPa are shown as filled circles (different colors are used for different phases). The phases which are correctly predicted by our approximate model to be thermodynamically stable at some finite pressure are indicated by black outlines around the respective circles. The vertical bars represent the pressure range of stability (of the correspondingly colored phase) as predicted by the approximate model.

case for most elements, with a few exceptions. Overall, 75% of all experimentally reported high-pressure phases are predicted to lie on the N–V–E convex hull (see the top panel of Table 4.1). In addition, for around 35% of the phases, the predicted pressure range of stability overlaps with the respective transition pressures reported in experiment. The low success rate in correctly predicting the transition pressure is somewhat expected following

the model validation on Si and Bi in Section 4.2.1. We discuss the possible sources of discrepancy between predictions and experiment toward the end of this Section.

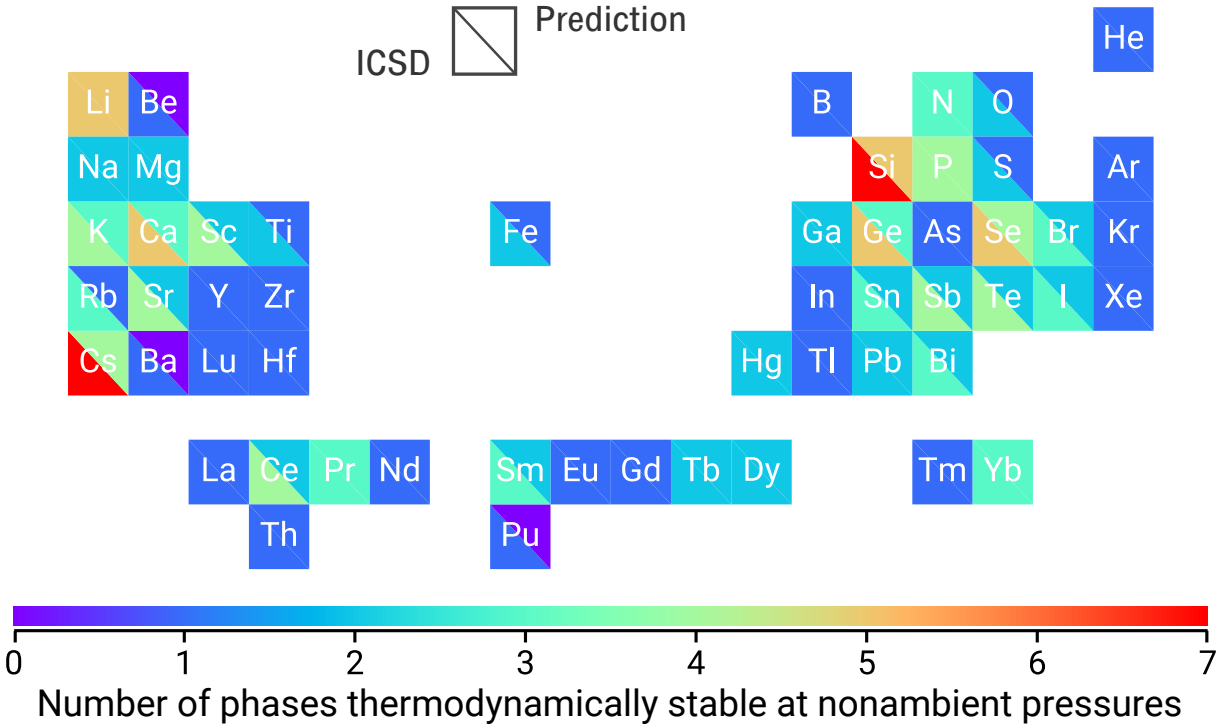

Figure 4.8. Comparison of predictions of high-pressure elemental phases from the LAE model against experiment. For each element, the number of (a) unique phases reported experimentally and (b) predicted by the linear enthalpy model to be thermodynamically stable at non-ambient pressures, are indicated by the color of the bottom-left and top-right halves, respectively, of the corresponding tile in the periodic table. Overall, the model correctly predicts ∼75% of the high-pressure phases in the ICSD to be thermodynamically stable at non-ambient pressures.

Next, we perform a similar large-scale analysis for all experimentally reported binary phases. Using calculations of experimentally reported compounds in the OQMD, curated using pressure-related metadata in the ICSD (in a manner similar to that employed for elemental phases), we compile a dataset of 343 unique binary compounds in total as a validation-dataset (the entire list is available in Table A.2 in the Appendix, along with a

Table 4.1. Accuracy of the linear enthalpy model in predicting the stability (at some finite-pressure) of experimentally reported elemental and binary high-pressure phases.

Elements	
Experimentally reported HP phases	132
Predicted to be stable at finite pressure	97 (75%)
Predicted pressure range of stability matches experiment	45 (35%)
Binaries	
Experimentally reported HP phases	343
Predicted to be stable at finite pressure	273 (80%)
Predicted pressure range of stability matches experiment	125 (35%)

comparison of LAE predictions vs experiment in Fig. 4.9). This number is significantly lower than that expected from a simple combinatorial estimation. For elemental solids, we found in average more than one high-pressure phase per element. If we extend this observation to binaries and assume that every binary system has in average more than one high-pressure phase, the number of potential high-pressure phases considering 90 elements is ${}^{90}C_2 = 4005$. We note that our estimation is very conservative, since binary A–B systems introduce an additional, compositional degree of freedom, which allows multiple high-pressure phases to exist at the same pressure, A_pB_q, as we have seen in Sec. 4.2.1.2. This indicates that the high-pressure phase diagrams of binary systems in general have been relatively underexplored. The linear enthalpy model performs equally well for binary compounds—80% of experimentally reported high-pressure binary phases are predicted to be stable at some finite pressure (see lower panel of Table 4.1). For around 35% of the phases, the predicted pressure range of stability overlaps with the respective transition pressures reported experimentally.

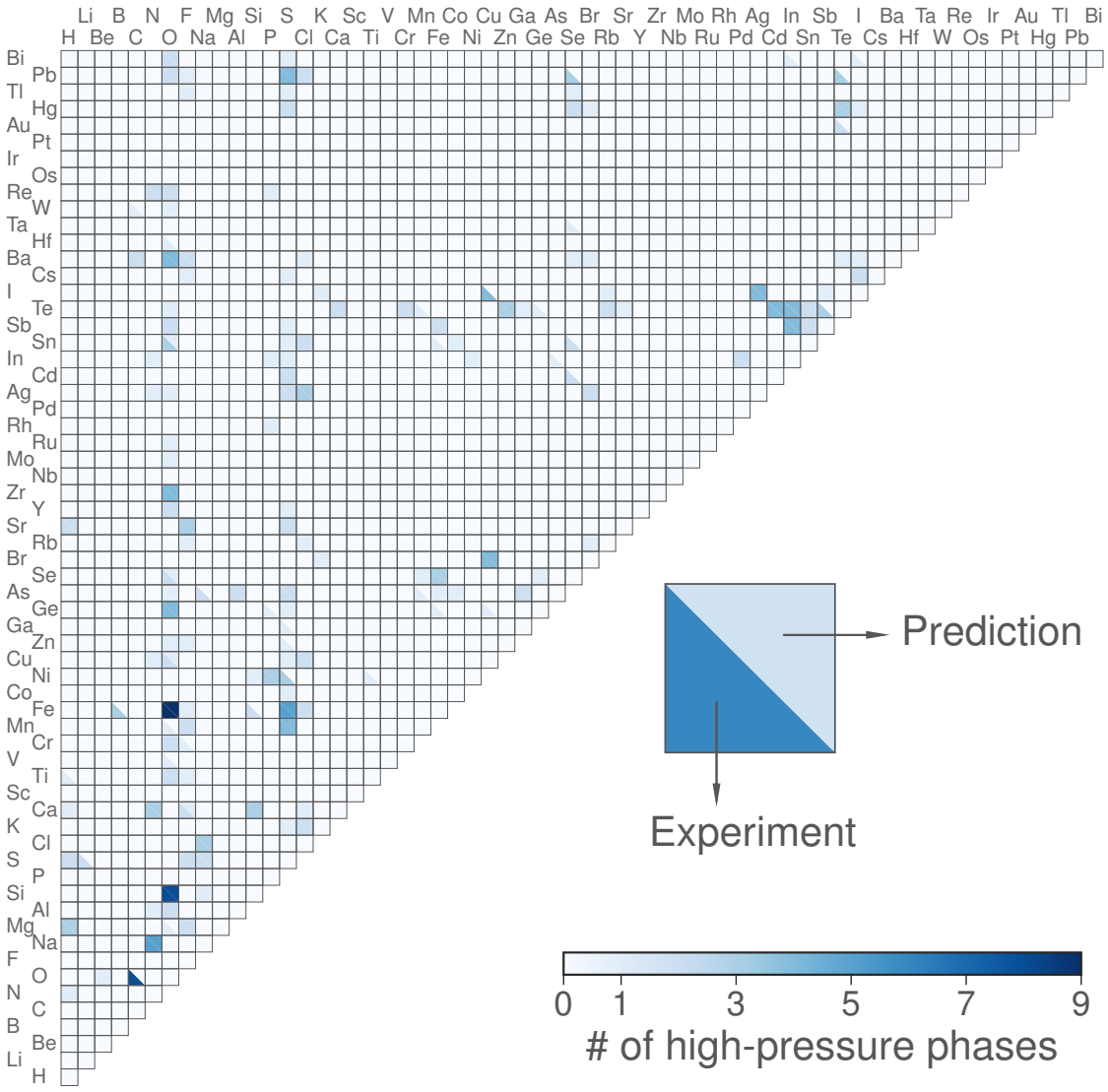

Figure 4.9. High-pressure binary phases reported experimentally compared with predictions using the linear enthalpy model. Each square represents a binary family (e.g., Si-O). The color of the lower left (upper right) triangle indicates the number of observed (predicted) high-pressure phases in that family. The heat map is rather sparse: only a few families have experimentally reported high-pressure phases.

Overall, our "crude" linear enthalpy model performs surprisingly well, with a success rate of 75–80%, in predicting the stability of both elemental and binary high-pressure phases. We identify four potential sources of error that could explain the discrepancy between the number of high-pressure phases reported experimentally and that predicted by our approximate model:

(a) The crystal structure reported experimentally for the phase is erroneous. Resolving the crystal structure, e.g., from in-situ XRD measurements, under high pressure is a difficult and tedious task that can lead to incomplete/incorrect structural characterization. A prominent example is the Bi-III phase, the crystal structure of which was experimentally resolved only after several failed attempts [14]. In fact, Bi-III has an incommensurate host-guest structure and the reported structure is only a representative ordered model with $P4/nnc$ symmetry [16]. A similar incommensurate structure has been reported for phase IV of phosphorus in the pressure range of 107–137 GPa [176].

(b) The high-pressure phase emerges via a phase transition of second order. In this case, the structural relaxations performed using DFT will inevitably transform the high-pressure phase to a lower-pressure structure. Therefore, our linear enthalpy model, which relies on the equilibrium energy E_0 and volume V_0 at ambient pressure of a high-pressure phase, will expectedly not capture its stability.

(c) Errors inherent to DFT calculations and numerical noise, e.g., due to the approximation to the exchange correlation potential, pseudization of core electrons

(which might be important especially at high pressures), unconverged basis sets and sampling meshes, insufficient tolerances during structural relaxations, etc.

(d) Finally, there is the inherent error due to applying a linear approximation to the enthalpy of each phase (i.e., assuming all phases to be perfectly incompressible), which might be unreasonable for some materials at large values of pressure. In particular, the LAE, like all commonly used empirical EOS, assumes isotropic compressibility, which leads to issues with strongly anisotropic phases, as we discuss in detail for the $Cmcm$ phase of $NiBi_3$.

One way to address some of the above limitations in the LAE for practical materials discovery efforts is to take into account phases *within a small distance* ΔH *off* the N–V–E convex hull, in addition to the ones that lie *on* it. For example, an enthalpic tolerance of $\Delta H = 25$ meV/atom can be used to account for the influence of entropic effects at room temperature on thermodynamic phase stability [107]. Phases that are not on the hull, but sufficiently close to it, can then be included in explicit enthalpy calculations to obtain reliable predictions of accurate high-pressure phase diagrams to guide experimental synthesis.

4.2.2.2. All experimentally reported compounds. We now use our linear enthalpy model to analyze the phase stability of all experimentally reported compounds calculated in the OQMD (not limited to high-pressure phases), a total of around 33,000 unique ordered compounds. As earlier, using the equilibrium energy and volume at ambient pressure of each phase in our dataset, we predict the number, and the pressure range

of stability, of all phases that lie on the extended N–V–E convex hull (i.e., presumably thermodynamically stable at some finite-pressure).

First, we find that only around 55% of the 33,000 compounds in our dataset lie on the N–E convex hull, i.e., are thermodynamically stable at ambient pressure conditions, consistent with a previous report on a similar dataset from the OQMD [130]. A recent study by Sun et al. [107] on a dataset of 29,900 experimentally reported compounds calculated in the Materials Project also finds around 50±4% of the phases to be ambient-metastable. In the latter study, it is proposed that the *observed* metastable compounds are generally remnants of thermodynamic conditions where they were once the stable ground states.

We next test this hypothesis of "remnant metastability" by using pressure as a thermodynamic handle and tracking the number of metastable phases that become stable with incremental increase/decrease in pressure, with respect to ambient conditions. Fig. 4.10a shows the fraction of metastable phases as a function of positive (compressive) or negative (tensile) pressure, separated into binary, ternary, quaternary and higher-component systems. We observe a range of trends based on our statistical analysis.

First, the number of metastable phases decreases with incremental application of both positive and negative pressures, relative to 0 GPa. In other words, *a significant fraction of the ambient-metastable phases are in fact thermodynamically stable ground states at non-ambient pressure conditions.* For example, in the case of binary compounds (top left in Fig. 4.10a), the fraction of phases that are metastable at ambient pressure decreases from around 0.45 at 0 GPa to around 0.30 at 100 GPa—33% of the ambient-metastable phases are rendered thermodynamically stable at some pressure $p \in (0, 100]$ GPa. However, in

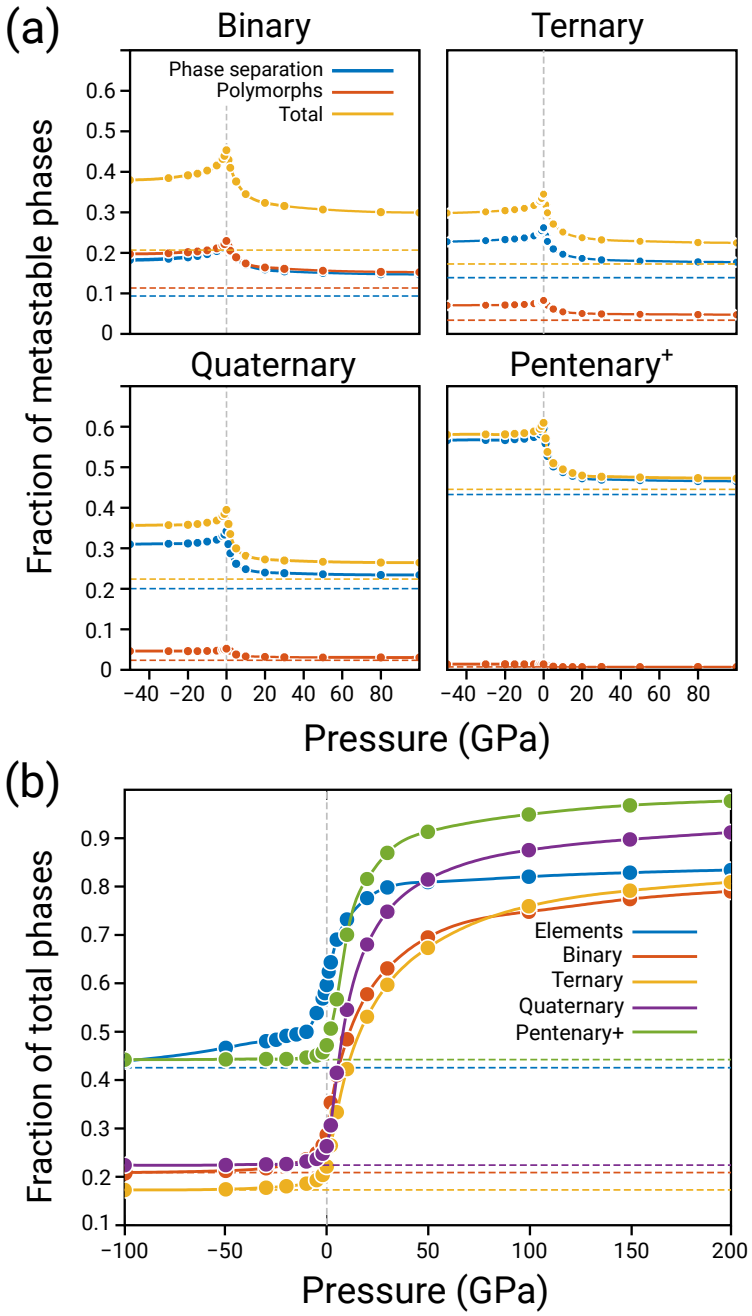

Figure 4.10. (a) Fraction of metastable phases that become thermodynamically stable with incremental increase/decrease in pressure, with respect to 0 GPa. The horizontal dashed lines indicate the fraction of metastable phases that do not lie on the N–V–E convex hull at any pressure. (b) Fraction of ambient-metastable phases that cannot be accessed thermodynamically at any pressure larger than pressure p, equivalent to 1− (fraction of phases that can be accessed at some pressure larger than pressure p).

each case, a sizeable fraction of metastable phases remain metastable at all pressures, i.e., they are not equilibrium ground states at any pressure (represented by horizontal dashed lines in Fig. 4.10a). For example, around 21% of all binary phases remain metastable and cannot be accessed thermodynamically via pressure alone.

Second, the rate of decrease in the number of metastable phases (or increase in the number of metastable phases made stable) with pressure is maximum near zero and decays rapidly toward higher positive/negative pressures. This is most likely due to a bias toward small values of pressure in our compiled set of phases—after all, most compounds reported experimentally are likely observed in near-ambient conditions—but could be also due to a fundamental property of materials, namely, the density of stable ground states as a function of volume/pressure is maximum near zero pressure.

Third, we find considerable differences concerning the "character" of metastability in binary, ternary, and higher order compositional systems. We distinguish two subsets for each n-component dataset (n = 2, 3, 4, $\geq$5)—"polymorphs" and "phase separation"—depending on whether a given phase is metastable with respect to another phase at the same composition or a combination of phases, respectively, at ambient conditions. We note that the higher the number of components present in a metastable compound, the more likely it is to phase-separate rather than transform into a polymorph, in agreement with previous observations [107]. Further, the lower the number of components in a metastable compound, the more likely it is to be stabilized with pressure. Considering the subset of all metastable phases that phase-separate at ambient pressure, 58%, 47%, 42%, and 27% become thermodynamically stable at some finite positive/negative pressure in the case of binary, ternary, quaternary, and higher-component systems, respectively.

Additionally, we observe that the effects of positive and negative pressures on the metastability of phases are *not* symmetric about zero pressure: a much larger portion of ambient-metastable phases become thermodynamically stable under positive (compressive) pressure when compared to negative (tensile) pressure. A difference is perhaps expected considering that the limiting behaviors are very different: large positive pressures favor the formation of close-packed phases before eventual overlap of atomic cores, while the limit of large negative pressures is simply the individual non-interacting atoms of each species in the phase.

Finally, we probe a complementary question: if one were to incrementally tune external conditions from large positive to large negative pressures, how many observed metastable phases $\mathcal{N}$ can be accessed thermodynamically below any given pressure p? We calculate at pressure p, the number of experimentally observed phases from our dataset that *cannot* be thermodynamically accessed at any pressure $> p$. We present this data as a cumulative histogram of the fraction of phases, integrated from pressures p to $+\infty$, separated into elements, binaries, ternaries, etc. in Fig. 4.10b. Hypothetically, if all experimentally reported compounds were thermodynamically stable ground states at some finite pressure, one would expect this cumulative fraction of unstable phases to be 1 and 0 for $p \to \infty$ and $p \to -\infty$, respectively. Consistent with our previous observations, we find that (a) a sizable fraction of the phases do not lie on the extended N–V–E convex hull at all, i.e., they are not ground states under any pressure (represented by horizontal dashed lines in Fig. 4.10b), and (b) the rate at which additional metastable phases can be thermodynamically accessed is maximum near zero pressure (given by the slopes of the curves). In other words, the pressure density of thermodynamic ground states, $\frac{d\mathcal{N}}{dp}$, is maximum near

$p = 0$. Whether this is an artifact of using a dataset of experimentally observed phases or is a fundamental property of matter, needs further analysis, and will be the subject of future work.

4.2.3. Discovery of new high-pressure compounds

So far, we have used the LAE to analyze the phase stability of experimentally reported high-pressure elemental and binary phases, and to probe the accessibility of ambient-metastable phases using pressure as a thermodynamic handle. Now, we go a step further by using the LAE to predict new intermetallic compounds by combining it with CSP methods. For this purpose, we focus on a unique subset of binary systems, namely, the combination of elements that are immiscible at ambient pressures. According to the data we compiled from the OQMD, there currently exist ∼1780 binary systems that do not contain any experimentally observed compounds. Any high-pressure phases that we identify in these systems are therefore true predictions of new materials.

For the dataset to be used for construction of the convex hull and calculation of transition pressures within the LAE, we use ambient-pressure formation energies and volumes of phases calculated in the OQMD. As mentioned in Section 4.4.1, the OQMD contains calculations of more than 450,000 compounds including experimentally reported compounds from the ICSD, and hypothetical compounds generated by decoration of common structural prototypes with all the elements in the periodic table. The Strukturbericht symbols of the prototype structures considered in this section are listed below [130, 177]:

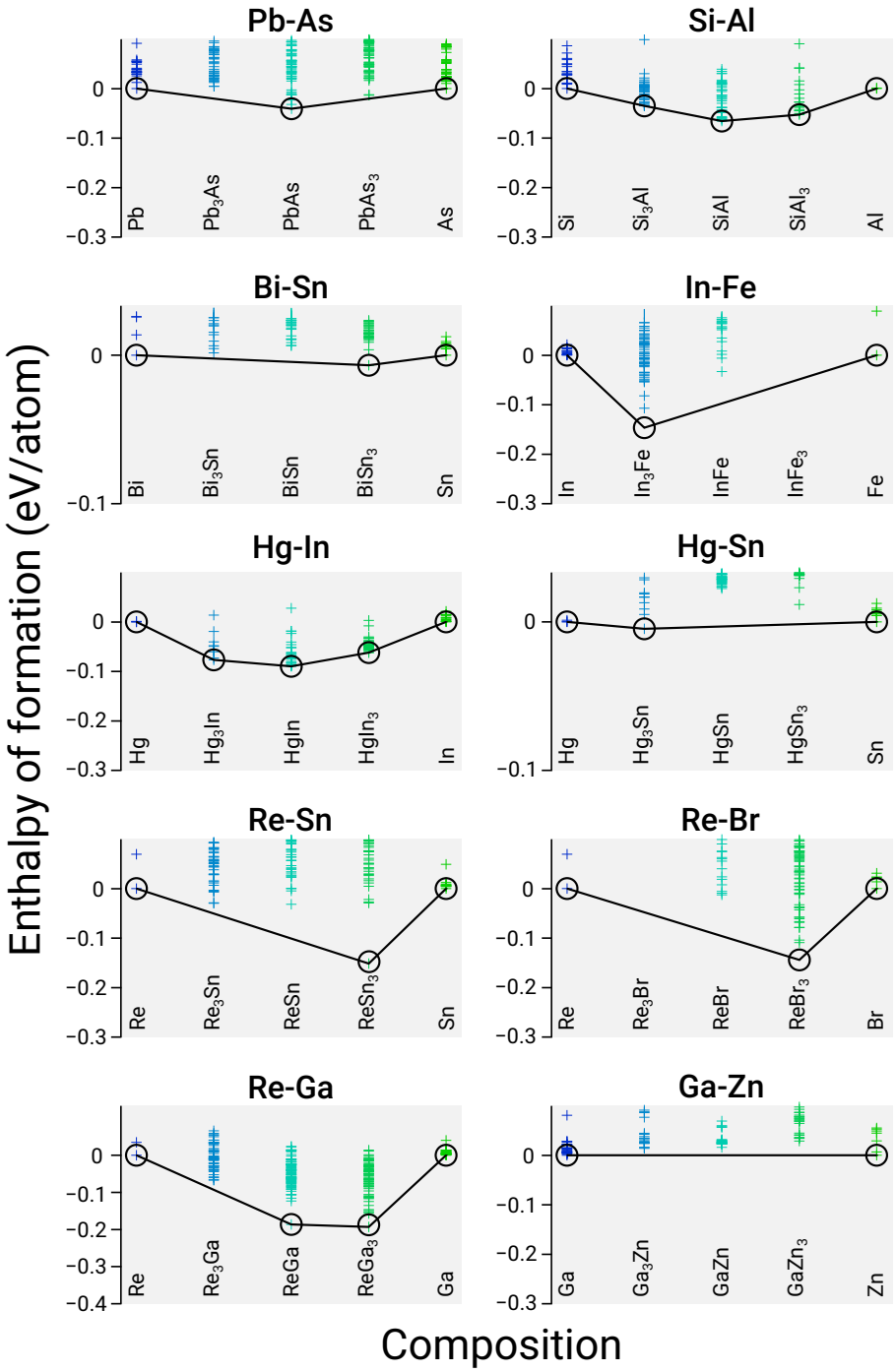

Figure 4.11. The convex hulls of formation enthalpy of ten ambient-immiscible binary systems calculated using structural search at 50 GPa via the MHM. Each cross denotes a phase sampled with the MHM. In all but the Zn-Ga system, we find at least one thermodynamically stable high-pressure phase.

(a) elemental prototypes: A1 (fcc), A2 (bcc), A3 (hcp), A3' (α-La), A4 (diamond), A5 (β-Sn), A7 (α-As), A9 (graphite), A10 (α-Hg), A11 (α-Ga), A12 (α-Mn), A13 (β-Mn), A_b (β-U), A_h (α-Po), $C19$ (α-Sm)

(b) binary AB: B1 (NaCl), B2 (CsCl), B3 (zincblende ZnS), B4 (wurtzite ZnS), B19 (AuCd), B_h (WC), $L1_0$ (AuCu), $L1_1$ (CuPt)

(c) binary A_3B: $L1_2$ (Cu_3Au), $D0_{19}$ (Ni_3Sn), $D0_{22}$ (Al_3Ti), $D0_3$ ($AlFe_3$)

We screen for promising chemical systems that contain high-pressure phases in the following manner: for every ambient-immiscible binary system, we use the LAE to predict the thermodynamic phase stability and pressure range of stability of each hypothetical compound in that chemical space. We select systems that contain at least one hypothetical compound predicted to become stable below an arbitrary pressure threshold of 50 GPa. We then rank these systems according to the predicted transition pressures, from lowest to highest, and select 10 of the most promising systems for further investigation. For each system, we further verify that no compound in that chemical space is reported in the ICSD or in phase diagrams available in the ASM Alloy Phase Diagram Database [178]. At each composition where our model predicts a stable high-pressure phase we perform structural searches using the MHM, starting from the respective prototype structure from the OQMD, using simulation (super) cells with up to 10 atoms/cell. Due to the set of binary prototypes currently calculated in the OQMD (see list above), the compositions we sample are limited to A_3B, AB and AB_3. Note that both the system size and the number of sampled compositions are far too low to give accurate predictions of the true high-pressure ground states. The structural searches are merely intended as proof-of-concept,

i.e., to provide a sampling of configurations beyond the limited number of prototype structures.

Fig. 4.11 shows the convex hulls of stability at 50 GPa for the ten selected ambient-immiscible binary systems, namely, As-Pb, Al-Si, Sn-Bi, Fe-In, Hg-In, Hg-Sn, Re-Sn, Re-Br, Re-Ga, and Zn-Ga, after performing structural searches using the MHM. We also provide the approximate LAE convex hulls at the same pressure (extrapolated from the equilibrium volumes and energies of the prototype structures in the OQMD) in Figs. A.1–A.10 in the Appendix. For each binary system, a comparison of the MHM and LAE convex hulls shows a good agreement of the overall features. Note however that the ground state structures found by the MHM calculations are different from the binary prototype structures in the OQMD, often significantly lower in enthalpy.

For all but the Zn-Ga system the MHM searches confirm the existence of at least one new stable high-pressure phase. In fact, in nine out of the ten ambient-immiscible binary systems the LAE correctly predicts that new compounds can be formed at sufficiently high pressures. All thermodynamically stable structures at 50 GPa are provided in the Appendix Section A.3, together with the OQMD ID of the seed structures. Our results show that the LAE is an effective tool to identify "seed" compositions and structures for sophisticated CSP methods. The high-pressure phases predicted here present a number of avenues for experimental synthesis and verification. Overall, the success of the linear enthalpy model in guiding unbiased search techniques based on crystal structure prediction in discovering novel high-pressure phases is remarkable.

4.3. Conclusions

In summary, we present a method that allows an efficient screening for materials that are thermodynamically stable at non-ambient pressures using a simple linear approximation to the formation enthalpy of a phase. Using a generalized convex hull construction, the stability of thousands of compound can be evaluated at a low computational cost based on ambient-pressure data that is currently available in many materials databases without performing any additional DFT calculations. A comparison with higher order EOS fitted to energy-volume data shows that the LAE performs similarly well at significantly lower computational cost.

Through a large-scale analysis of experimentally reported compounds, we show that a large fraction of the observed ambient-metastable phases are in fact thermodynamic ground states at some finite pressure. This results strongly suggests that many phases are likely remnants of high-pressure conditions where they were stable ground states, even if the synthesis occurred at atmospheric pressure—potentially through local pressure fluctuations, chemical pressure, etc. Our method can be readily extended by further generalizing the convex hull construction and taking into account additional thermodynamic degrees of freedom, including temperature or surface areas of finite particles.

Finally, we demonstrate the predictive power of this model when combined with a crystal structure prediction technique by discovering novel high-pressure phases in a set of ambient-immiscible binary intermetallic systems.

4.4. Computational Methods

4.4.1. Calculation of thermodynamic quantities

The equilibrium formation energy and volume data for all the phases considered in our analysis using the LAE were retrieved from the Open Quantum Materials Database (OQMD) [129, 130]. The dataset consists of DFT-calculated properties of over 450,000 compounds which include (a) unique, ordered experimentally reported compounds from the Inorganic Crystal Structure Database (ICSD), and (b) hypothetical compounds generated by the decoration of common structural prototypes with all the elements in the periodic table. Details of the settings used to calculate the equilibrium formation energy and volume of compounds in the OQMD can be found in Ref. [130].

All other DFT calculations reported in this work, i.e., those performed as part of global structure searches, were performed using the Vienna Ab initio Simulation Package (VASP) [179–181]. We use the projector augmented wave (PAW) formalism [182, 183] and the PBE parameterization of the generalized gradient approximation to the exchange correlation functional [184] throughout. For all calculations, we use Γ-centered k-point meshes with about 8000 k-points per reciprocal atom and a plane-wave cutoff energy of 520 eV. All atomic and cell degrees of freedom of a structure are relaxed until the force components on all the atoms are within 0.01 eV/Å, and stresses are within a few kbar.

4.4.2. Structural searches

The minima hopping method (MHM) [185, 186] implements a highly reliable algorithm to explore the low enthalpy phases of a compound at a specific pressure given solely the chemical composition [187–189]. The low lying part of the enthalpy landscape is efficiently

sampled by performing consecutive, short MD escape steps to overcome enthalpy barriers, followed by local geometry optimizations. The Bell-Evans-Polanyi principle is exploited through a feedback mechanism on the MD escape trials, and by aligning the initial MD velocities along soft-mode directions in order to accelerate the search [190, 191]. The MHM has been successfully applied to identify the structure and composition of many materials, also for systems at high pressures [113, 142, 143, 146, 192, 193]. In this work, we performed MHM simulations only at the compositions where a high-pressure phase is predicted to be stable by the linear enthalpy model.

4.4.3. Software implementation

All convex hull constructions in this work were performed using the *Qhull* library [194] with the quickhull algorithm as implemented in the *SciPy* Python package [195]. The GCLP calculations reported in this work were performed using the *Cbc* solver distributed with the *PuLP* Python library [196]. An implementation of the framework described in Sections 4.1.1–4.1.3 has been made available as an open-source Python module [197]. An implementation of the MHM is available through the *Minhocao* package [185, 186].

CHAPTER 5

The Phase Stability Network of All Inorganic Materials

5.1. Background

Several diverse complex systems are modeled as networks of discrete components linked together: man-made systems such as electrical power grids and the world-wide web [198, 199], social systems such as friendship and scientific collaborations [200, 201], and natural systems such as metabolism in a cell and food-webs [202, 203]. Despite significant variation in the nature of individual components and interconnections, many of these networks show striking similarities in their topology [204, 205], often providing new insights into each respective domain of knowledge. For instance, disparate systems such as the world-wide web and metabolic reactions in cellular organisms both have been shown to follow the organizational principles of robust, error-tolerant scale-free networks, with implications for the resilience of the internet and the design of therapeutics [205, 206], respectively.

Recent developments in high-throughput density functional theory (HT-DFT) [132] have resulted in massive computational databases of materials properties [21, 129–131, 207], containing the calculated properties of hundreds of thousands of experimentally reported and hypothetical materials. Such databases have led to new data-driven approaches toward understanding materials. Here we introduce a novel paradigm of viewing materials, and equilibrium phase diagrams in particular, via the lens of complex network theory, i.e. studying the similarities and interactions between materials themselves, in

striking contrast to the traditional bottom-up approaches toward unlocking structure-property relationships in materials [208, 209].

We use the Open Quantum Materials Database (OQMD) [129, 130], a HT-DFT database containing calculations of nearly all crystallographically ordered, structurally unique materials experimentally observed to date (as collected in the Inorganic Crystal Structure Database [210] repository) and a large number of hypothetical materials constructed using commonly occurring structural prototypes—a total of more than half a million materials—to extract the "universal phase stability network" or the "universal T=0 K phase diagram". We accomplish this by using all the phase data in the OQMD within a convex-hull formalism, and identifying *all* thermodynamically stable materials and *all* two-phase equilibria between them. We then represent stable materials as nodes and two-phase equilibria (tie-lines) as edges, thus describing a T=0 K phase diagram as a network encoding thermodynamic stability (illustrated with schematics in Fig. 5.1).

5.2. Results

5.2.1. Overall network connectivity

We find that the phase stability network of all inorganic materials consists of ∼21,300 nodes and is remarkably dense with a total of nearly 41 million edges, and extremely well-connected with ∼3,850 edges per node on average ("mean degree" $\langle k \rangle$). This means that every stable inorganic compound can form a stable two-phase equilibrium with 3,850 other compounds on average. For comparison, $\langle k \rangle$ for other widely-studied networks range from 1.4 (network of email messages) to 113.4 (collaboration network of film actors) [211]. The connectance of the materials network, or the fraction of the maximum possible number

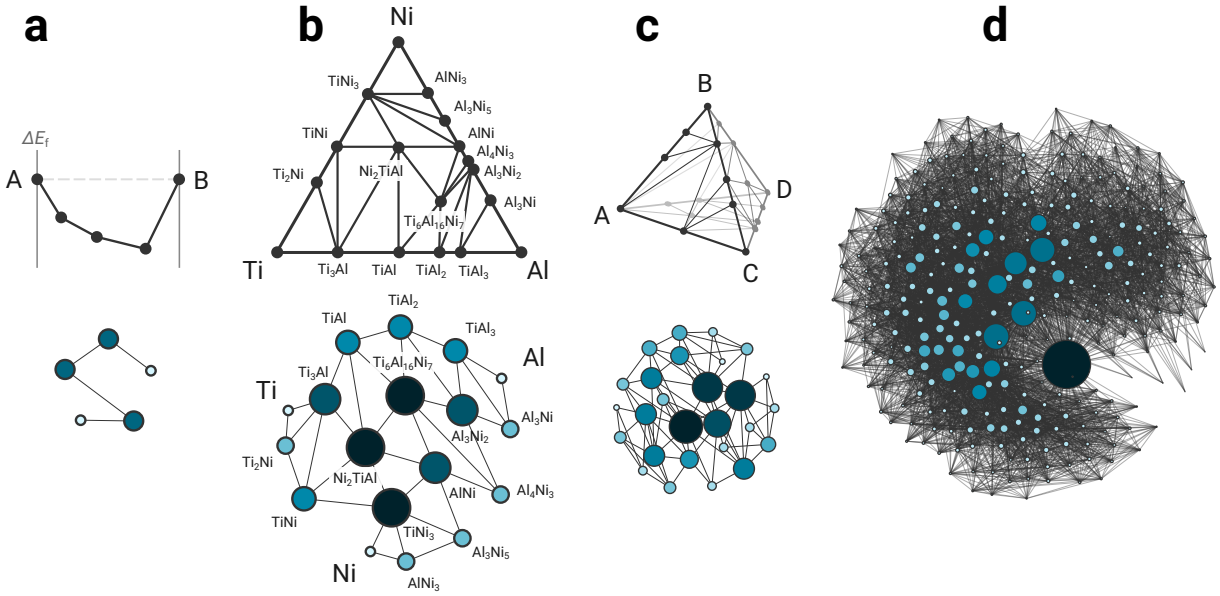

Figure 5.1. Network representation of T=0 K materials phase diagrams. Stable phases and two-phase equilibria (tie-lines) in a phase diagram are represented as nodes and edges, respectively, to create the corresponding network: (**a**) Schematic A-B binary system represented as a typical two-dimensional convex hull of compound formation energies. (**b**) Ti-Ni-Al as an example ternary system, with the T=0 K phase diagram shown as a Gibbs triangle. (**c**) Schematic A-B-C-D quaternary phase diagram shown as a Gibbs tetrahedron. (**d**) The 3*d* transition metal-chalcogen (i.e. 14-dimensional chemical space) materials network. No conventional visual representations exist of phase diagrams at higher than 4 dimensions. Node sizes shown are proportional to node degree.

of edges that are actually present is 0.18. This is an important statistic for the design of "systems of materials", such as electrodes and electrolytes making up batteries [212], or coating materials separating two reactive components [213], where the longevity of the system relies on stable coexistence of such components. Using a lithium-ion intercalation battery as an example "system of materials", a common approach to tackling electrode degradation is to apply protective coatings on electrode particles. In such a battery, the material in the electrode coating should not react with/be consumed by materials in the

electrode *as well as* those in the electrolyte [214, 215]. Thus, the coating–electrode and the coating–electrolyte material pairs must both have tie-lines with each other in order to stably coexist in the system. In other words, both pairs must be neighboring, connected nodes in the materials network.

The degree distribution in the complete phase stability network, specifically the probability $p(k)$ that a material has a tie-line with k other materials in the network follows a lognormal form (Fig. 5.2a; details of the fit in Fig. 5.7 and related discussion in Section 5.4.2). While many widely-studied networks are known to have scale-free power-law degree distributions, lognormal distributions are another member of the "heavy-tail" family, are also relatively common, and behave quite similar to power-laws [216]. In fact, sparsity has been shown to be a necessary condition for the emergence of an exact power-law behavior [217], and densification in sparse, scale-free networks leads to distributions that deviate from a power-law and become closer to lognormal. Thus, the lognormal behavior of the materials network can be understood to result from its extremely dense connectivity, in contrast to the general sparsity of commonly-studied networks.

5.2.2. Network topology

The characteristic path length or mean node-node distance in a network, $\mathcal{L}$, is defined as the number of edges in the shortest path between two nodes, averaged over all pairs of nodes. The longest node-node distance in the network defines its diameter, $\mathcal{L}_{\max}$. The characteristic path length of the materials network $\mathcal{L} = 1.8$, and its diameter $\mathcal{L}_{\max} = 2$. This remarkably short path length indicates that the materials network has "small-world" characteristics [198], i.e. despite its large size, the number of edges that need to be

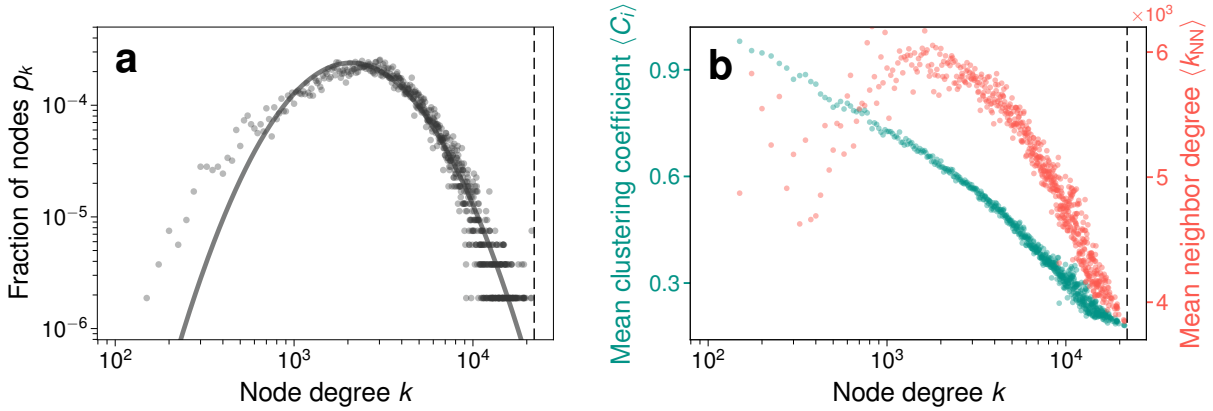

Figure 5.2. Overall structure and topology of the materials network. (**a**) The distribution of node degree in the materials network (grey circles) shows a heavy tail, i.e. a sizeable fraction of materials have tie-lines with nearly all other materials. A lognormal fit is shown as a solid grey line. (**b**) The mean local clustering coefficient $\langle \mathcal{C}_i \rangle$ (green) decreases with node degree k indicating that stable materials form local, high-connected communities. The mean neighbor degree $\langle k_{\mathrm{NN}} \rangle$ (red) also decreases with k, implying a weakly-dissortative network behavior, i.e. materials with a large number of tie-lines connect with those with fewer tie-lines in the network. In both subplots, the vertical dashed line represents the total number of nodes (stable materials) in the network.

traversed from a given node to any other node is relatively small. The extremely small $\mathcal{L}$ for the materials network can be intuitively understood to be a consequence of the almost complete lack of reactivity of noble gases. The non-participation of noble gases in the formation of compounds (and thus having tie-lines with nearly all materials in the network) places an upper bound of 2 on $\mathcal{L}_{\mathrm{max}}$, and since some material pairs already have tie-lines that connect them immediately, the mean path $\mathcal{L}$ is slightly smaller than 2. Even if noble gases are disregarded, the mean path length and diameter of the materials network remain small due to the presence of a few other very-highly connected nodes corresponding to extremely stable and non-reactive materials, e.g. binary halides.

Another metric of interest in a real-world network is transitivity or clustering, quantified by its clustering coefficient, $\mathcal{C}$, which is the probability that two nodes connected to the same third node are themselves connected. In other words, given that there exist stable two-phase equilibria A–C and B–C, what is the probability that A and B can stably coexist? Depending on how the averaging is performed, a global ($\mathcal{C}_g$) or mean local ($\bar{\mathcal{C}}_i$) cluster coefficient of a network can be defined [198, 211]. For the materials network, the clustering coefficients are $\mathcal{C}_g = 0.41$ and $\bar{\mathcal{C}}_i = 0.55$, comparable to other real-world networks, and much higher than random networks of the same density. The mean local clustering coefficient of the materials network decreases with increasing node connectivity (Fig. 5.2b), indicating that stable materials form local highly-connected communities in the network, and such behavior often suggests a hierarchical network structure [218]. The assortativity coefficient or the Pearson correlation coefficient of degree between pairs of connected nodes in the materials network is -0.13, indicating weakly dissortative mixing behavior. This is also confirmed by the distribution of the mean degree of neighbors of a node of degree k being a decreasing function of k (Fig. 5.2a). In other words, materials with a high k (i.e. large number of tie-lines) tend to connect with materials with a lower k (i.e. smaller number of tie-lines). This weakly dissortative behavior of the materials network is similar to that observed in most other technological, information, biological networks, and is likely a virtue of such networks being simple graphs [219].

5.2.3. Hierarchy in the materials network

The mean degree or the average number of tie-lines per material $\langle k \rangle$ decreases with the number of components, $\mathcal{N}$ ($\mathcal{N} = 2$ for binary, $\mathcal{N} = 3$ for ternary, etc. See Fig. 5.3a),

indicating a chemical hierarchy in the materials network. This can be understood to result from an inherent competition for tie-lines that high-$\mathcal{N}$ materials face with low-$\mathcal{N}$ materials in their chemical space, but not vice-versa. In other words, ternary compounds $X_aY_bZ_c$ compete not only with other compounds in the X-Y-Z chemical space but also with binary compounds in the X-Y, Y-Z, Z-X spaces for tie-lines.

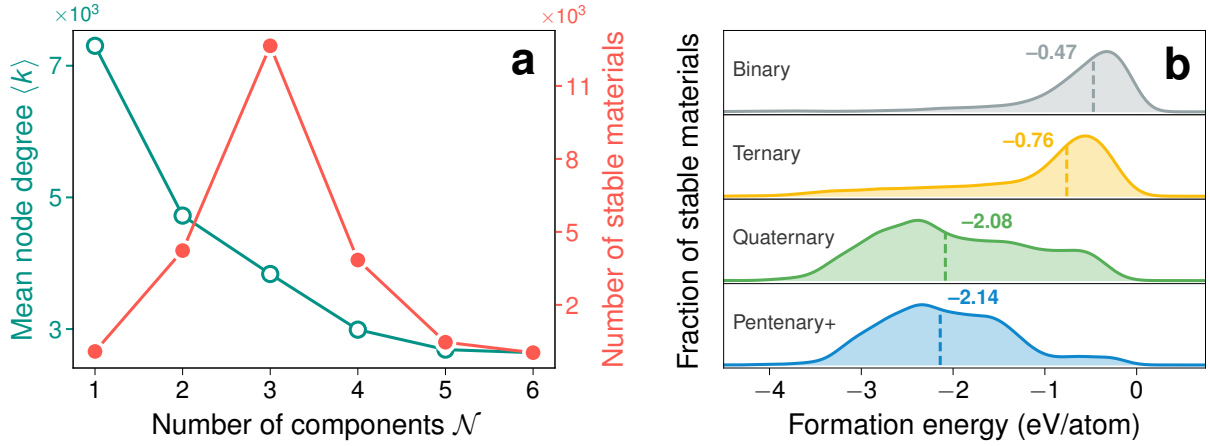

Figure 5.3. Hierarchy in the materials network, and underlying energetic considerations. (**a**) The mean node degree or average number of tie-lines $\langle k \rangle$ (green, open) decreases as a function of number of components $\mathcal{N}$ (i.e. binary, ternary, and so on), which results from high-$\mathcal{N}$ materials having to compete with low-$\mathcal{N}$ materials for stability. The number of known stable $\mathcal{N}$-ary materials (red) itself actually peaks at $\mathcal{N} = 3$ (ternaries). (**b**) Gaussian kernel density estimates of compound formation energies for all stable materials separated by number of components in the material. Dotted vertical lines indicate the respective median of each distribution. High-$\mathcal{N}$ need significantly lower formation energies than low-$\mathcal{N}$ materials to become stable, e.g. -2.08 versus -0.47 eV/atom for quaternary and binary materials, respectively.

We note that this decrease in $\langle k \rangle$ with $\mathcal{N}$ is distinct from the distribution of number of stable $\mathcal{N}$-ary materials itself (Fig. 5.3a), which shows a peak at $\mathcal{N} = 3$. Does this peak in the distribution of stable materials imply the existence of infinite, underexplored space for the discovery of new materials beyond ternaries? The distribution of formation energies

of materials as a function of number of components $\mathcal{N}$ (Fig. 5.3b) reflects the consequence of competition between low- and high-component materials: high-$\mathcal{N}$ compounds appear to need significantly lower formation energies than low-$\mathcal{N}$ ones to become stable. Since there is no obvious underlying reason for the distribution of $T = 0$ K formation energies (with entropic effects neglected) to differ significantly with $\mathcal{N}$, only a few high-$\mathcal{N}$ materials can "survive" as stable phases if the corresponding lower-$\mathcal{N}$ systems already have several stable phases. This is consistent with the recent reports of a "volcano plot" that emerges for stable inorganic ternary nitrides as a function of energetic competition with their corresponding binary nitrides [220], and an increased probability of phase separation with increasing number of components in the system [221]. Widom [222] further argued that the a peak near $\mathcal{N} = 3$ or 4 in such distributions arises from a competition between combinatorial explosion and diminishing volume-to-surface ratio in the composition simplex, as $\mathcal{N}$ increases. Thus, although we do not know of a fundamental law limiting access to thermodynamically stable materials with higher components, a combination of the hierarchy observed in the phase stability network, the distribution of formation energies, and the topology of the convex energy surface all suggest that the scarcity of known high-$\mathcal{N}$ stable materials is not merely a consequence of those chemical spaces being underexplored.

5.2.4. Knowledge extraction: material nobility index

Since the phase stability network practically encompasses all known inorganic crystalline materials as well as a large number of predicted hypothetical materials, the number of tie-lines emerges as a natural metric of nobility of a crystalline material—it is simply

the count of other materials it is determined to have no reactivity against. Thus, while material reactivity or nobility have no standard definitions, a network representation of materials enables us to tackle the chemical nobility of inorganic materials in solid-solid and solid-gas reactions in a completely data-driven fashion, instead of the traditional intuitive or heuristic approaches. Since the number of tie-lines in the materials network is lognormally distributed, we devise a new standard score of material nobility, the "nobility index":

$$\mathcal{Z}_n = \frac{\ln(k) - \mu}{\sigma} \tag{5.1}$$

where k is the node degree or the number of tie-lines a material has, and $\mu = 8.06$ and $\sigma = 0.65$ are the mean and standard deviation of the underlying lognormal distribution. The nobility index is thus agnostic of textbook classifications such as metal, nonmetal, metalloid, ionic, covalent, and so on, and works equally well for any given material. Since the tie-lines in the network are as computed with DFT, the nobilities of materials predicted herewith are only limited by DFT accuracy in estimating relative stabilities of inorganic materials [130, 221, 223].

First, we tackle the reactivity or nobility of elements. Noble gases and fluorine form the bounds of the nobility index (Fig. 5.4), as the noblest and the most reactive, respectively, not only among the elements but in fact among all materials in the network. The most reactive elements following F are P, S, and Cl. Alkali and alkaline earth metals, often considered to be highly reactive metals, are relatively noble in solid-solid and solid-gas reactions, in comparison to early d-block or lanthanide elements, which are, along with

Al, the most reactive metals. The nobility index increases down a group for metals, and increases (decreases) from left-to-right along a row of the periodic table within the d-block (s-block). But what is the noblest metal of them all? Ag emerges as the noblest of all elements after noble gases, followed closely by Hg, Os, Re, W, and Cu, all having more than 14,000 tie-lines. Gold, traditionally considered the noblest element [224], despite being relatively densely connected with 10,000 tie-lines, is less noble in solid-state reactions. Finally, we find that $\mathcal{Z}_n$ is not correlated with other common elemental properties such as electronegativity, atomic radii, melting point, and others [225], indicating that the nobility index encodes new information not readily captured by those properties (see Fig. 5.5 and related discussion below).

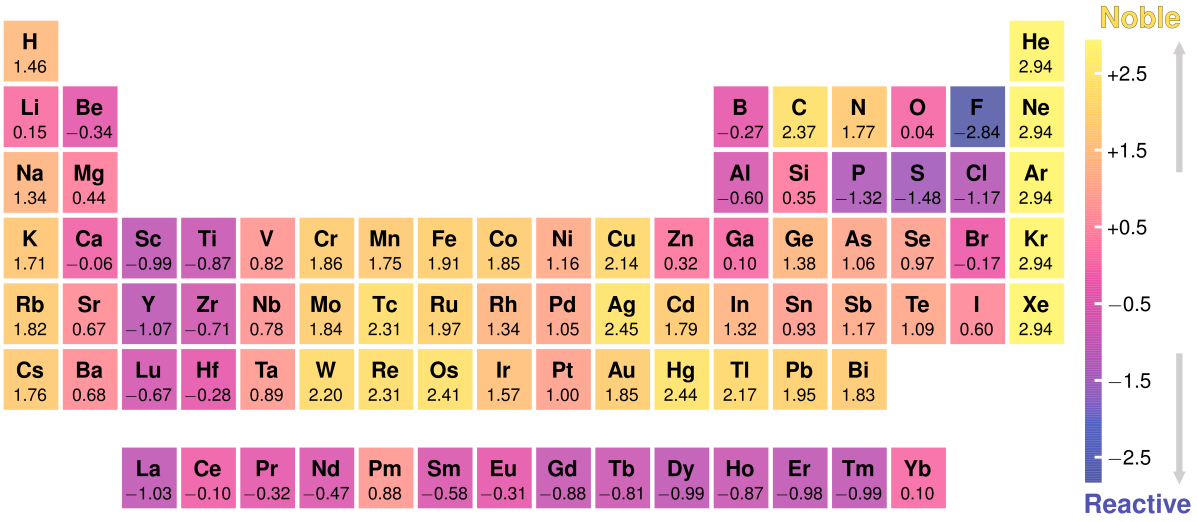

Figure 5.4. Nobility index of all elements. The standard score, $\mathcal{Z}_n$, derived in this work using material connectivity in the phase stability network, as a measure of nobility against solid-solid and solid-gas reactions. Nobility increases up the scale. Numerical values of elemental $\mathcal{Z}_n$ are given below the respective symbols.

Beyond elements, what are the noblest inorganic compounds of all? The compounds at the top of the nobility list are IA/IIA-VIIA compounds such as LiF, NaCl, KCl, CsCl,

KBr, CsBr, KI, RbI, CaF_2, SrF_2, $CsYbF_3$, $RbYbF_3$, and others, their inertness likely due to stability from strong ionic bonding between their constituents. We exclude rareearth- and actinide-containing compounds from the previous analysis of compound nobility in order to account for any shortcomings in the DFT description of f-block elements and compounds containing them.

5.2.5. New information encoded in the "nobility index"

A comparison of the nobility index $\mathcal{Z}_n$ of elements against elemental properties such as electronegativity, boiling point, melting point, atomic volume, etc., as collected by Ward et al. [225] shows little correlation between $\mathcal{Z}_n$ and other properties, with Pearson correlation coefficients close to 0 for most properties (see Fig. 5.5 for a sample comparison set). This indicates that the nobility index defined in this work truly encodes new information about an element/a material not adequately captured by other common properties.

Further, data-driven metrics such as nobility index capture materials knowledge that is not immediately intuitable or is sometimes even counter-intuitive. For instance, intuition derived only from a few elements and some of their compounds may imply that multivalent elements (e.g. transition metals) are likely to have a higher number of tie-lines than monovalent elements (e.g. alkali metals) simply by the virtue of a higher number of compound-forming possibilities. However, data from all materials known so far shows no correlation between number of compounds formed by an element and its total number of tie-lines (i.e. nobility; see Fig. 5.6). In fact, monovalent metals seem to form more compounds on average than their multivalent counterparts!

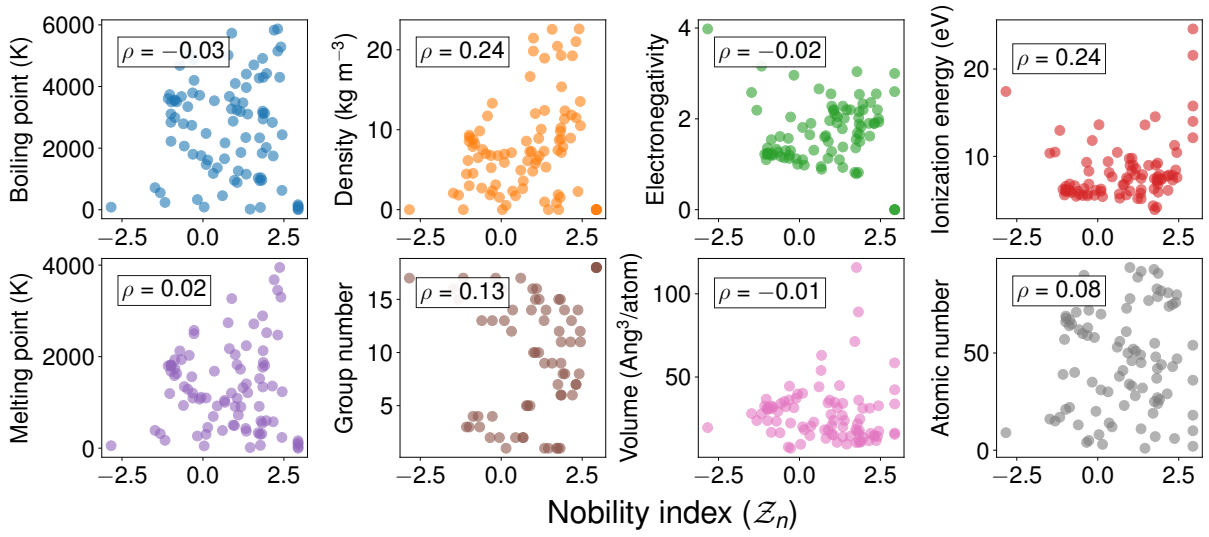

Figure 5.5. Comparison of nobility index versus common elemental properties. There is little to no correlation between the nobility index of an element and any of its properties such as (counterclock-wise from top-left) boiling point, density, electronegativity, first ionization energy, atomic number, atomic volume, group in the periodic table, and melting point. The Pearson correlation coefficient ρ for each comparison is on the top-left of the corresponding panel.

5.3. Discussion

While some of our findings above are in line with chemical intuition, relative nobilities in certain cases, e.g. silver vs gold, deviate from it. This deviation is in part due to the historical context in which these materials have been considered noble or reactive, e.g. whether an element oxidizes or corrodes readily in air, reacts with water and/or certain acids, dissolves in water or electrolytes, and how vigorous such reactions seem. More fundamental approaches to finding descriptors for reactivity go back to electronegativity related concepts, followed by interrelated theories based on perturbation theory, derivatives of electronic energy such as hardness and softness, and others largely developed for molecules [226–228]. In contrast, the nobility index, $\mathcal{Z}_n$, as derived from the tie-lines in

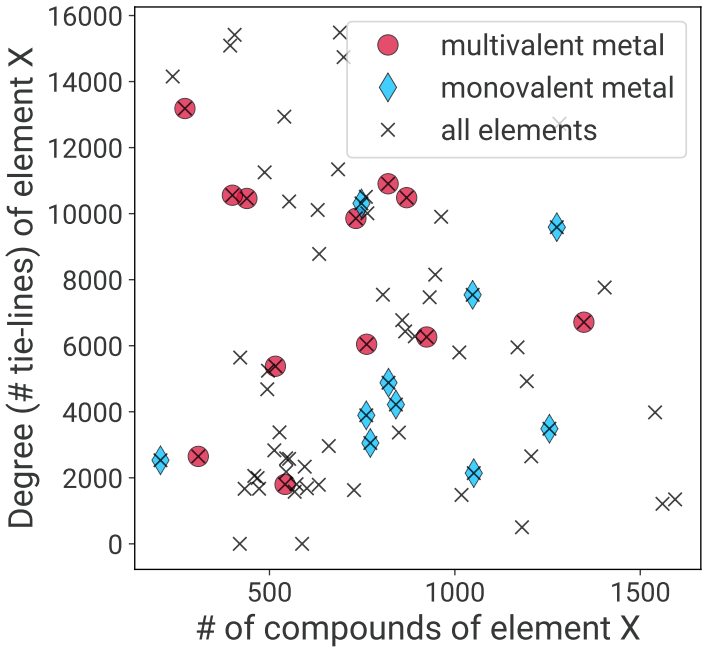

Figure 5.6. Comparison of number of compounds formed by an element versus its node degree. Multivalent metals indicated are all transition metals (Ti, V, Cr, Mn, Fe, Co, Ni, Mo, W, Hf, Pd, Pt), and monovalent metals indicated are mostly alkali/alkaline earth metals (Li, Na, K, Rb, Be, Mg, Ca, Sr, Al, Zn).

the network of all inorganic materials represents a general metric emerging directly from bulk thermodynamic data.

High-throughput experimental and computational techniques are leading to an explosive growth in the size of materials databases. Representation and interpretation of the data at a large-scale, however, remains a challenge. Here we show that tools from complex network theory enable us to access otherwise-difficult-to-extract information from such large datasets. In other words, the emergence of material reactivity from the collective behavior of all materials in the phase stability network serves as a simple, preliminary

example of knowledge extraction out of complex networks of materials. Other similar approaches can be used to discover other hidden knowledge, e.g. analysis of "communities" or "cliques" in the network of all materials can uncover hitherto-unknown relationships between various known materials.

Further, there are various ways our graph theoretic approach to materials data can be used to immediately applied to new materials discovery and design: (a) *direct* techniques, e.g. metrics from network theory such as local clustering and similarity can be used to identify "holes" in the current network—where nodes (i.e. materials) are expected to exist but currently do not, and (b) *indirect* techniques, e.g. using the extracted knowledge or quantities derived from the network as input to other approaches such as in materials informatics. For example, using temporal materials discovery information in combination with thermodynamic phase stability networks can help predict synthesizability [229]. Furthermore, while some of its features resemble other complex networks, the extremely-high connectance and the lognormal degree distribution of the presented phase stability network imply that its underlying generative mechanisms may be unique, and developing generative models for such materials networks can have significant impact on the knowledge discovery of materials in the future.

5.4. Computational Methods

All convex hull constructions were performed using the *Qhull* library [230] as implemented in the *qmpy* (pypi.org/project/qmpy) package. All network analyses were performed using the *graph-tool* [231] and *powerlaw* [232] packages. We describe the approach

used to calculate the universal phase diagram, its network representation and determining the node degree distribution in the network in the following sections.

5.4.1. Calculation of the $T = 0$ K universal phase diagram

The $T = 0$ K phase diagram for a given chemical space is determined by the so-called convex hull construction. A phase is thermodynamically stable iff it lies on (i.e. is a vertex of) the convex hull of $T = 0$ K formation energies of all phases in the chemical space. And phases that are directly connected by a tie-line, i.e., phases that lie on the same facet of the convex hull, are in equilibrium with one another. Determining a binary A-B phase diagram requires constructing a 2-dimensional convex hull of formation energies of all A_xB_y compounds (composition x and formation energy being the two dimensions), a ternary A-B-C phase diagram requires constructing a 3-dimensional convex hull of formation energies of all $A_xB_yC_z$ compounds (compositions x and y, and formation energy being the three dimensions), and so on. The determination of an d-nary phase diagram requires the construction of an d-dimensional convex hull of formation energies of all the N phases in the chemical space.

For low dimensions, i.e. $d = 2$ or 3 (binary or ternary systems), finding the convex hull of N points (total number of phases) has a worst-case time complexity of $\mathcal{O}(N\,log\,N)$. For higher dimensions, standard methods of determining convex hulls such as the Quickhull algorithm, have worst-case time complexities of $\mathcal{O}(N^{\lceil d/2 \rceil})$ [230]. For random data, even the average-case time complexity at higher dimensions scales as $\mathcal{O}(log^{d-1}N)$, i.e. exponentially with d [233]. Such scaling behaviors mean that for moderately large number of points N and dimensions d, finding the convex hull becomes increasingly practically

challenging. For instance, to find the convex hull of all known inorganic materials, even restricting ourselves to experimentally reported compounds in the Open Quantum Materials Database (OQMD), $N \approx 40,000$ and $d = 89$, making the calculation of the convex hull practically impossible with a traditional single-shot approach.

We tackle this challenge of calculating the convex hull at high-dimensions by using a divide-and-conquer approach. While the representational complexity of the convex-hull increases at least exponentially with d, we know from the set of existing materials that not many of them are high-dimensional by themselves. In fact, 99.5% of materials in the OQMD have 4 unique elemental components or fewer. Since the stability of a material is determined only within the chemical subspace of elements that it is made of, we first determine the vertices (i.e. stable materials) of the 89-dimensional convex hull at a reduced computational cost by computing the convex hulls in low-dimensional subspaces for each individual material separately. For instance, to determine if the compound $CaMnO_3$ is stable, it is sufficient to construct the convex hull of all phases $Ca_xMn_yO_z$ in the Ca-Mn-O subspace. This process of constructing convex hulls separately for each unique chemical subspace yields all the vertices of the full convex hull: $\sim 2.1 \times 10^4$ stable materials out of the $> 5.5 \times 10^5$ total materials calculated in the OQMD. Having determined the vertices of the full convex hull, in the second stage, we exhaustively evaluate the existence of a tie-line between any given pair of stable compounds in the OQMD by constructing the convex hull of formation energies in the combined chemical spaces of such candidate pairs, rather than the full 89-dimensional space itself. For example, to determine whether there exists a tie-line between Li_2O and NaCl, we construct the Li-Na-Cl-O convex hull, and find that there indeed exists a Li_2O-NaCl tie-line. In contrast, from a Na-K-F-Cl convex

Phase 1	Phase 2	# Components	Time (s)
Na_2O	KCl	4	∼3
Fe_2S_3	Li_2MnO_4	5	∼6
Li_3PS_4	$SrTiO_3$	6	∼8
$Ba_2Li_3TaN_4$	$LiCoO_2$	6	∼14
$Ba_2Li_3TaN_4$	$NaCoO_2$	7	∼32
$Mn_2Hg_2SF_6$	Li_4CrCoO_6	8	∼34
$Mn_2Hg_2SF_6$	$Ba_2Ca_3Tl_2Cu_4O_{12}$	9	∼65

Table 5.1. Sample compute times for calculating the existence of a tie-line between two phases. The time required is highly dependent on the number of components, i.e. unique elements in the combined chemical space, and further depends on the number of all known compounds in the chemical space. Each calculation was performed on a standard desktop computer utilizing a single core.

hull we find that NaCl and KF, in fact, "react" to form a NaF-KCl two-phase equilibrium. Overall, we construct convex hulls for all ${}^{2.1\times10^4}C_2 \approx 2.3 \times 10^8$ stable phase combinations, and find a total $\sim 41 \times 10^6$ tie-lines.

The computational cost of constructing a convex hull for a unique chemical subspace is expectedly highly dependent on the number of components, and ranges from a few seconds to a few minutes on a standard desktop computer utilizing a single core (some sample times for checking if a tie-line exists between two known materials are provide in Table 5.1). With a conservative estimate of 15–20 seconds per tie-line, the total time required to exhaustively evaluate all possible tie-lines is more than 1 million CPU hours.

We then represent stable compounds as nodes and tie-lines as edges, thereby generating the "universal phase diagram" as a complete thermodynamic phase stability network of all inorganic materials. We use the *Qhull* library [230] as implemented in the *qmpy* package (pypi.org/project/qmpy) for all the convex hull calculations reported in this work.

5.4.2. Degree distribution of the network of all materials

The probability distribution of node connectivity (number of tie-lines a material has) in the phase stability network of all inorganic materials is heavy-tailed. We examine which of the common heavy-tailed distributions best fit our empirical data. In particular, several well-studied technological, social, and biological networks are thought to have power-law distributions. Is the thermodynamic network of materials similar to other common natural/man-made networks exhibiting power-law behavior or not? To answer this question, we directly compare pairs of heavy-tailed distributions using the method of log likelihood ratios as described in Clauset et al. [234]. For the full materials network, we find that a lognormal distribution ($\mu = 8.06$, $\sigma = 0.65$) is the best fit by far (see Fig. 5.7).

We note that most empirical phenomena obey power-laws only for values greater than some minimum value, i.e. only the tail of the distribution follows a power-law. We investigate if this is indeed the case for the materials network. We find the optimal lower-bound for a power-law behavior, $k_{\min}$, for the materials network as the value that minimizes the Kolmogorov-Smirnov distance between the data and the fit [234]. We find $k_{\min}$ for the materials network = $\sim$5800, and the power-law scaling parameter $\alpha = 4.4$. We note that a $k_{\min}$ of 5800 retains only 17% of the overall materials network (i.e. only 17% of all materials have more than 5800 tie-lines each). Furthermore, even over this tail region of the degree distribution, a lognormal distribution is a better fit (see the inset in Fig. 5.7): the log likelihood ratio $\mathcal{R}$ for power-law versus lognormal is -7.15 with a p-value of 0. In other words, even in the $k_{\min} = 5800$ region (tail) of the materials network, the lognormal distribution fits the data far better than a power-law.

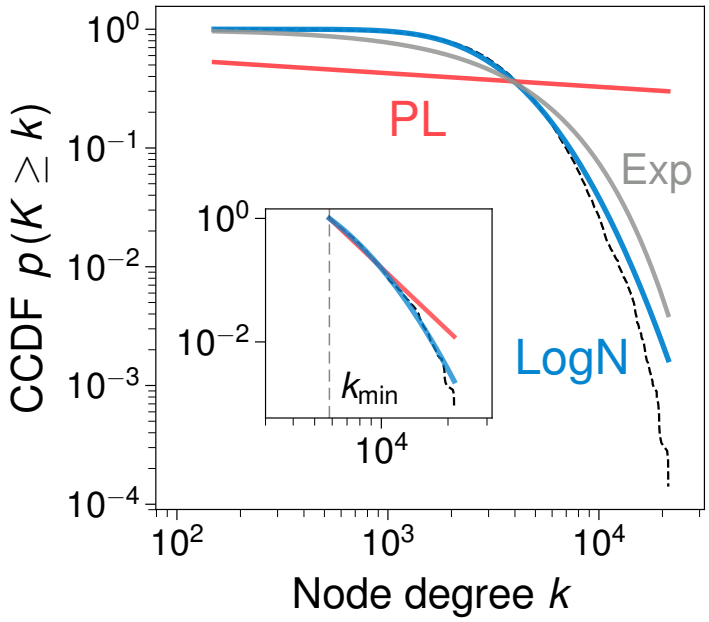

Figure 5.7. Fitting node connectivity data to candidate distributions. The complementary cumulative distribution function of the node degree in the network of all materials is shown as dashed black lines. Power-law (PL), lognormal (LogN), and exponential (Exp) distributions fit to the data are shown as solid red, blue, and grey lines, respectively. The inset shows power-law and lognormal fits to the *tail* of the degree distribution for degree $k > k_{\min} = 5800$.

All analyses of fits of degree distributions mentioned above were performed with the *powerlaw* package [232]. We note that the graph-theoretic analyses reported in this work (e.g. local clustering and centrality meaures) performed with the *graph-tool* package [231], while requiring more than 8 G of memory, take a few hours on a standard desktop utilizing up to 4 cores.

CHAPTER 6

Summary and outlook

In this thesis, I present work in three areas at the fore-front of *data-driven materials science*. I summarize the main conclusions from each area, and list a few themes that I expect to be the natural "next-steps" from the work presented here. I also present scientific tangents from here that, in my opinion, have a potential for high impact:

(I) **Traditional high-throughput computational materials discovery:**

I have used the term "traditional high-throughput" to mean the paradigm of screening computational databases of hundreds of thousands of structures to find stable materials with the properties of interest, and all the related machinery in this approach. I demonstrate the application of such a paradigm to predict new ternary oxyfluorides with simple stoichiometries, and Slater-Pauling semiconductors in between the half- and full-Heusler compositions (the "three-quarter" Heusler compounds). While this technique is conceptually already a few years old, it remains a straightforward way to search for new materials/narrow down the possibilities from an intractable candidate space. I anticipate that instead of decorating structure types with all elements from the periodic table or from conventional intuition, using machine learning or active learning to determine what calculations to perform can significantly increase the speed and efficiency of materials discovery. The logical conclusion of the above line of argument is

that we would be able to calculate the properties only for the materials that are of interest, and those that can be experimentally realized. While the former has been a well-formed problems with promising results from recent materials informatics efforts, the latter case of predicting synthesizability of compounds is currently a hotbed of activity, and even the problem formulation is stil quite amorphous. While being able to accurately determine apriori what materials should be calculated seems like the ideal scenario, I do have a word of caution: since brute-force high-throughput computation still remains the largest source of properties data, biasing the dataset (e.g. calculating only stable or nearly stable compounds) is bound to bias what and how we learn from the materials data in the future. There is significant merit to having negative examples in any dataset we want to learn from, and the easiest way to ensure that is to (at least partly/on a smaller scale) continue calculating enumerated and decorated structures in an unfiltered way, even when such structures do not make physical/chemical sense.

(II) **Fingerprinting the materials genome at high-pressure:**

I have presented how a simple linear approximation to enthalpy can be used in conjunction with density functional data calculated at ideal conditions of zero pressure can be used to quite reliably (and extremely cheaply) assess high-pressure phase stability. With pressures of hundreds of GPa becoming increasingly accessible in laboratories, the method presented here can be used to rapidly identify phases that are stable at high-pressure. I show examples of this approach via predictions of binary ambient-immiscible systems which become compound

forming at high-pressure. These are predictions in the "truest sense"—we predict compounds in systems without a single previously reported compound! Extending the model to include computed bulk modulus data can significantly improve the accuracy of the method presented here without much additional cost. Incorporating such a framework (e.g. using the software tool built here) into open high-throughput databases can be immediately impactful—high-pressure researchers can use the tool as a preliminary screen before investing significant resources into synthesizing a material. A somewhat orthogonal direction of very high potential impact is to devise a framework similar to the one developed here, but for Helmholtz free energy, instead of enthalpy. The challenge there is the lack of large datasets of entropy (analogous to volume for enthalpy). However, the payout for developing such a framework is huge—we would be able to combine the two frameworks to access *the full free energy landscapes* of materials using only calculations performed at zero temperature and zero pressure!

(III) **Complex network theory meets materials science:**

I introduce a completely novel way to look at materials data: viewing ensembles of materials as networks instead of single crystal structures. While there is an inherent loss in information by ignoring the properties of each individual material, there is much to be gained from the properties that are emergent from the interaction between the materials themselves. I show how the "universal phase diagram" of all materials can be naturally represented as a complex network. I also show the emergence of material reactivity from the collective behavior of

all materials in the phase stability network as a simple, preliminary example of knowledge extraction out of networks of materials. Other similar approaches can be used to discover other hidden knowledge, e.g. analysis of "communities" or "cliques" in the network of all materials can uncover hitherto-unknown relationships between various known materials. Further, there are various ways this graph theoretic approach to materials data can be used to immediately applied to new materials discovery and design: (a) *direct* techniques, e.g. metrics from network theory such as local clustering and similarity can be used to identify "holes" in the current network—where nodes (i.e. materials) are expected to exist but currently do not, and (b) *indirect* techniques, e.g. using the extracted knowledge or quantities derived from the network as input to other approaches such as in materials informatics. Furthermore, while some of its features resemble other complex networks, the extremely-high connectance and the lognormal degree distribution of the presented phase stability network imply that its underlying generative mechanisms may be unique, and developing generative models for such materials networks can have significant impact on the knowledge discovery of materials in the future.

References

[1] J. P. Holdren, *et al.*, Materials Genome Initiative for Global Competitiveness, *Tech. rep.*, National Science and Technology Council OSTP, Washington, USA (2011).

[2] J. Ma, V. I. Hegde, K. Munira, Y. Xie, S. Keshavarz, D. T. Mildebrath, C. Wolverton, A. W. Ghosh, W. Butler, Computational investigation of half-heusler compounds for spintronics applications. *Phys. Rev. B* **95**, 024411 (2017).

[3] C. L. Fu, K. M. Ho, First-Principles Calculation of the Equilibrium Ground-State Properties of Transition Metals: Applications to Nb and Mo. *Phys. Rev. B* **28**, 5480–5486 (1983).

[4] R. Evarestov, *Quantum Chemistry of Solids: The LCAO First Principles Treatment of Crystals*, Springer Series in Solid-State Sciences (Springer-Verlag, Berlin Heidelberg, 2007).

[5] B. N. Dutta, Lattice Constants and Thermal Expansion of Silicon up to 900 ° by X-ray Method. *Phys. Stat. Sol. (b)* **2**, 984–987 (1962).

[6] J. C. Jamieson, Crystal Structures at High Pressures of Metallic Modifications of Silicon and Germanium. *Science* **139**, 762–764 (1963).

[7] M. I. McMahon, R. J. Nelmes, New High-Pressure Phase of Si. *Phys. Rev. B* **47**, 8337–8340 (1993).

[8] J. Z. Hu, I. L. Spain, Phases of Silicon at High Pressure. *Solid State Commun.* **51**, 263–266 (1984).

[9] M. Hanfland, U. Schwarz, K. Syassen, K. Takemura, Crystal Structure of the High-Pressure Phase Silicon VI. *Phys. Rev. Lett.* **82**, 1197–1200 (1999).

[10] H. Olijnyk, S. K. Sikka, W. B. Holzapfel, Structural Phase Transitions in Si and Ge Under Pressures up to 50 GPa. *Phys. Lett. A* **103**, 137–140 (1984).

[11] S. J. Duclos, Y. K. Vohra, A. L. Ruoff, Hcp to Fcc Transition in Silicon at 78 GPa and Studies to 100 GPa. *Phys. Rev. Lett.* **58**, 775–777 (1987).

[12] R. M. Brugger, R. B. Bennion, T. G. Worlton, The Crystal Structure of Bismuth-II at 26 kbar. *Phys. Lett. A* **24**, 714–717 (1967).

[13] L. G. Akselrud, M. Hanfland, U. Schwarz, Refinement of the Crystal Structure of Bi-II, at 2.54 GPa. *Z. Kristallogr. - New Cryst. Struct.* **218**, 415–416 (2014).

[14] J. H. Chen, H. Iwasaki, T. Kikegawa, K. Yaoita, K. Tsuji, *AIP Conference Proceedings* (AIP Publishing, 1994), vol. 309, pp. 421–424.

[15] M. I. McMahon, O. Degtyareva, R. J. Nelmes, Ba-IV-Type Incommensurate Crystal Structure in Group-V Metals. *Phys. Rev. Lett.* **85**, 4896–4899 (2000).

[16] U. Häussermann, K. Söderberg, R. Norrestam, Comparative Study of the High-Pressure Behavior of As, Sb, and Bi. *J. Am. Chem. Soc.* **124**, 15359–15367 (2002).
[17] J. H. Chen, H. Iwasaki, T. Kikegawa, Structural Study of the High-Pressure High-Temperature Phase of Bismuth using High Energy Synchrotron Radiation. *J. Phys. Chem. Solids* **58**, 247–255 (1997).
[18] K. Aoki, S. Fujiwara, M. Kusakabe, Stability of the bcc Structure of Bismuth at High Pressure. *J. Phys. Soc. Jpn.* **51**, 3826–3830 (1982).
[19] J. E. Saal, S. Kirklin, M. Aykol, B. Meredig, C. Wolverton, Materials design and discovery with high-throughput density functional theory: The open quantum materials database (oqmd). *JOM* **65**, 1501–1509 (2013).
[20] S. Kirklin, J. E. Saal, B. Meredig, A. Thompson, J. W. Doak, M. Aykol, S. Rühl, C. Wolverton, The open quantum materials database (oqmd): assessing the accuracy of dft formation energies. *npj Comput. Mater.* **1**, 15010 (2015).
[21] A. Jain, S. P. Ong, G. Hautier, W. Chen, W. D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, K. a. Persson, The Materials Project: A materials genome approach to accelerating materials innovation. *APL Mater.* **1**, 011002 (2013).
[22] S. Curtarolo, W. Setyawan, S. Wang, J. Xue, K. Yang, R. H. Taylor, L. J. Nelson, G. L. Hart, S. Sanvito, M. Buongiorno-Nardelli, N. Mingo, O. Levy, Aflowlib.org: A distributed materials properties repository from high-throughput ab initio calculations. *Comput. Mater. Sci.* **58**, 227–235 (2012).
[23] G. Bergerhoff, R. Hundt, R. Sievers, I. D. Brown, The inorganic crystal structure database. *J. Chem. Inf. Comput. Sci.* **23**, 66–69 (1983).
[24] S. Kirklin, J. E. Saal, V. I. Hegde, C. Wolverton, High-throughput computational search for strengthening precipitates in alloys. *Acta Mater.* **102**, 125–135 (2016).
[25] S. Kirklin, B. Meredig, C. Wolverton, High-throughput computational screening of new li-ion battery anode materials. *Adv. Energy Mater.* **3**, 252–262 (2013).
[26] G. Hautier, A. Jain, H. Chen, C. Moore, S. P. Ong, G. Ceder, Novel mixed polyanions lithium-ion battery cathode materials predicted by high-throughput ab initio computations. *J. Mater. Chem.* **21**, 17147–17153 (2011).
[27] H. Chen, G. Hautier, A. Jain, C. Moore, B. Kang, R. Doe, L. Wu, Y. Zhu, Y. Tang, G. Ceder, Carbonophosphates: A new family of cathode materials for li-ion batteries identified computationally. *Chem. Mater.* **24**, 2009-2016 (2012).
[28] M. Aykol, S. Kirklin, C. Wolverton, Thermodynamic aspects of cathode coatings for lithium-ion batteries. *Adv. Energy Mater.* **4**, 1400690 (2014).
[29] D. H. Snydacker, M. Aykol, S. Kirklin, C. Wolverton, Lithium-ion cathode/coating pairs for transition metal containment. *J. Electrochem. Soc.* **163**, A2054–A2064 (2016).
[30] M. Aykol, S. Kim, V. I. Hegde, D. Snydacker, Z. Lu, S. Hao, S. Kirklin, D. Morgan, C. Wolverton, High-throughput computational design of cathode coatings for li-ion batteries. *Nat. Commun.* **7**, 13779 (2016).

[31] A. A. Emery, J. E. Saal, S. Kirklin, V. I. Hegde, C. Wolverton, High-throughput computational screening of perovskites for thermochemical water splitting applications. *Chem. Mater.* **28**, 5621–5634 (2016).

[32] I. E. Castelli, D. D. Landis, K. S. Thygesen, S. Dahl, I. Chorkendorff, T. F. Jaramillo, K. W. Jacobsen, New cubic perovskites for one- and two-photon water splitting using the computational materials repository. *Energy Environ. Sci.* **5**, 9034–9043 (2012).

[33] R. H. Taylor, S. Curtarolo, G. L. W. Hart, Guiding the experimental discovery of magnesium alloys. *Phys. Rev. B* **84**, 084101 (2011).

[34] J. E. Saal, C. Wolverton, Thermodynamic stability of mg-based ternary long-period stacking ordered structures. *Acta Mater.* **68**, 325–338 (2014).

[35] J. He, M. Amsler, Y. Xia, S. S. Naghavi, V. I. Hegde, S. Hao, S. Goedecker, V. Ozoliņš, C. Wolverton, Ultralow thermal conductivity in full heusler semiconductors. *Phys. Rev. Lett.* **117**, 046602 (2016).

[36] S. Wang, Z. Wang, W. Setyawan, N. Mingo, S. Curtarolo, Assessing the thermoelectric properties of sintered compounds via high-throughput ab-initio calculations. *Phys. Rev. X* **1**, 021012 (2011).

[37] R. Gautier, X. Zhang, L. Hu, L. Yu, Y. Lin, T. O. Sunde, D. Chon, K. R. Poeppelmeier, A. Zunger, Prediction and accelerated laboratory discovery of previously unknown 18-electron abx compounds. *Nat. Chem.* **7**, 308–316 (2015).

[38] G. Trimarchi, H. Peng, J. Im, A. J. Freeman, V. Cloet, A. Raw, K. R. Poeppelmeier, K. Biswas, S. Lany, A. Zunger, Using design principles to systematically plan the synthesis of hole-conducting transparent oxides: Cu3vo4 and ag3vo4 as a case study. *Phys. Rev. B* **84**, 165116 (2011).

[39] G. Hautier, A. Miglio, D. Waroquiers, G.-M. Rignanese, X. Gonze, How does chemistry influence electron effective mass in oxides? a high-throughput computational analysis. *Chem. Mater.* **26**, 5447–5458 (2014).

[40] F. Yan, X. Zhang, G. Y. Yonggang, L. Yu, A. Nagaraja, T. O. Mason, A. Zunger, Design and discovery of a novel half-heusler transparent hole conductor made of all-metallic heavy elements. *Nat. Commun.* **6** (2015).

[41] L. Yu, A. Zunger, Identification of potential photovoltaic absorbers based on first-principles spectroscopic screening of materials. *Phys. Rev. Lett.* **108**, 068701 (2012).

[42] W. Setyawan, R. M. Gaume, S. Lam, R. S. Feigelson, S. Curtarolo, High-throughput combinatorial database of electronic band structures for inorganic scintillator materials. *ACS Comb. Sci.* **13**, 382–390 (2011).

[43] C. Ortiz, O. Eriksson, M. Klintenberg, Data mining and accelerated electronic structure theory as a tool in the search for new functional materials. *Comput. Mater. Sci.* **44**, 1042–1049 (2009).

[44] M. Klintenberg, The search for strong topological insulators. *arXiv:1007.4838 [cond–mat.mtrl–sci]* (2010).

[45] K. Yang, W. Setyawan, S. Wang, M. B. Nardelli, S. Curtarolo, A search model for topological insulators with high-throughput robustness descriptors. *Nat. Mater.* **11**,

614–619 (2012).

[46] S. Kirkpatrick, C. D. Gelatt, M. P. Vecchi, *et al.*, Optimization by simulated annealing. *Science* **220**, 671–680 (1983).

[47] R. Martoňák, A. Laio, M. Parrinello, Predicting crystal structures: the parrinello-rahman method revisited. *Phys. Rev. Lett.* **90**, 075503 (2003).

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

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

[50] G. L. Hart, V. Blum, M. J. Walorski, A. Zunger, Evolutionary approach for determining first-principles hamiltonians. *Nat. Mater.* **4**, 391–394 (2005).

[51] A. R. Oganov, C. W. Glass, Crystal structure prediction using ab initio evolutionary techniques: Principles and applications. *J. Chem. Phys.* **124**, 244704 (2006).

[52] M. Deem, J. Newsam, Determination of 4-connected framework crystal structures by simulated annealing. *Nature* **342**, 260–262 (1989).

[53] P. Raiteri, R. Martoňák, M. Parrinello, Exploring polymorphism: the case of benzene. *Angew. Chem. Int. Ed.* **44**, 3769–3773 (2005).

[54] A. R. Oganov, R. Martoňák, A. Laio, P. Raiteri, M. Parrinello, Anisotropy of earth's d" layer and stacking faults in the mgsio3 post-perovskite phase. *Nature* **438**, 1142–1144 (2005).

[55] M. Amsler, S. Goedecker, Crystal structure prediction using the minima hopping method. *J. Chem. Phys.* **133**, 224104 (2010).

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

[57] X. Zhang, A. Zunger, Altered reactivity and the emergence of ionic metal ordered structures in li-cs at high pressures. *Phys. Rev. Lett.* **104**, 245501 (2010).

[58] D. B. Laks, L. Ferreira, S. Froyen, A. Zunger, Efficient cluster expansion for substitutional systems. *Phys. Rev. B* **46**, 12587 (1992).

[59] A. Seko, K. Yuge, F. Oba, A. Kuwabara, I. Tanaka, Prediction of ground-state structures and order-disorder phase transitions in ii-iii spinel oxides: A combined cluster-expansion method and first-principles study. *Phys. Rev. B* **73**, 184117 (2006).

[60] C. Wolverton, A. Zunger, First-principles prediction of vacancy order-disorder and intercalation battery voltages in li x coo 2. *Phys. Rev. Lett.* **81**, 606 (1998).

[61] B. Hülsen, M. Scheffler, P. Kratzer, Thermodynamics of the heusler alloy co 2- x mn 1+ x si: A combined density functional theory and cluster expansion study. *Phys. Rev. B* **79**, 094407 (2009).

[62] M. H. F. Sluiter, Y. Kawazoe, Cluster expansion method for adsorption: Application to hydrogen chemisorption on graphene. *Phys. Rev. B* **68**, 085410 (2003).

[63] Z. Lu, S.-H. Wei, A. Zunger, S. Frota-Pessoa, L. Ferreira, First-principles statistical mechanics of structural stability of intermetallic compounds. *Phys. Rev. B* **44**, 512 (1991).

[64] C. Wolverton, A. Zunger, Comparison of two cluster-expansion methods for the energetics of pd-v alloys. *Phys. Rev. B* **50**, 10548 (1994).

[65] V. Ozoliņš, C. Wolverton, A. Zunger, Cu-au, ag-au, cu-ag, and ni-au intermetallics: First-principles study of temperature-composition phase diagrams and structures. *Phys. Rev. B* **57**, 6427 (1998).

[66] S. V. Barabash, V. Blum, S. Müller, A. Zunger, Prediction of unusual stable ordered structures of au-pd alloys via a first-principles cluster expansion. *Phys. Rev. B* **74**, 035108 (2006).

[67] O. Levy, G. L. Hart, S. Curtarolo, Uncovering compounds by synergy of cluster expansion and high-throughput methods. *J. Am. Chem. Soc.* **132**, 4830–4833 (2010).

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

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

[70] C. S. Kong, W. Luo, S. Arapan, P. Villars, S. Iwata, R. Ahuja, K. Rajan, Information-theoretic approach for the discovery of design rules for crystal chemistry. *J. Chem. Inf. Model.* **52**, 1812–1820 (2012).

[71] P. V. Balachandran, J. Theiler, J. M. Rondinelli, T. Lookman, Materials prediction via classification learning. *Scientific reports* **5**, 13285 (2015).

[72] B. Meredig, A. Agrawal, S. Kirklin, J. E. Saal, J. Doak, A. Thompson, K. Zhang, A. Choudhary, C. Wolverton, Combinatorial screening for new materials in unconstrained composition space with machine learning. *Phys. Rev. B* **89**, 094104 (2014).

[73] G. Hautier, C. C. Fischer, A. Jain, T. Mueller, G. Ceder, Finding nature's missing ternary oxide compounds using machine learning and density functional theory. *Chem. Mater.* **22**, 3762–3767 (2010).

[74] F. Faber, A. Lindmaa, O. A. von Lilienfeld, R. Armiento, Crystal structure representations for machine learning models of formation energies. *Int. J. Quantum Chem.* **115**, 1094–1101 (2015).

[75] A. M. Deml, R. O'Hayre, C. Wolverton, V. Stevanović, Predicting density functional theory total energies and enthalpies of formation of metal-nonmetal compounds by linear regression. *Phys. Rev. B* **93**, 085142 (2016).

[76] R. M. Martin, Electronic structure: basic theory and practical methods. *Cambridge University Press* (2004).

[77] P. Hohenberg, W. Kohn, Inhomogeneous electron gas. *Phys. Rev.* **136**, B864 (1964).

[78] W. Kohn, L. J. Sham, Self-consistent equations including exchange and correlation effects. *Phys. Rev.* **140**, A1133 (1965).

[79] G. Kresse, J. Furthmüller, Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Comput. Mater. Sci.* **6**, 15–50 (1996).

[80] G. Kresse, J. Furthmüller, Efficient iterative schemes for ab-initio total energy calculations using a plane-wave basis set. *Phys. Rev. B* **54**, 11169–11186 (1996).

[81] P. E. Blöchl, Projector augmented-wave method. *Phys. Rev. B* **50**, 17953–17979 (1996).

[82] G. Kresse, D. Joubert, From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B* **59**, 1758–1775 (1999).

[83] J. P. Perdew, K. Burke, M. Ernzerhof, Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **77**, 3865–3868 (1996).

[84] A. R. Akbarzadeh, V. Ozoliņš, C. Wolverton, First-principles determination of multicomponent hydride phase diagrams: Application to the li-mg-n-h system. *Adv. Mater.* **19**, 3233–3239 (2007).

[85] J. W. D. Connolly, A. R. Williams, Density-functional theory applied to phase transformations in transition-metal alloys. *Phys. Rev. B* **27**, 5169-5172 (1983).

[86] A. van de Walle, G. Ceder, Automating first-principles phase diagram calculations. *J. Phase Equilibria* **23**, 348–359 (2002).

[87] A. van de Walle, M. Asta, G. Ceder, The alloy theoretic automated toolkit: A user guide. *Calphad* **26**, 539–553 (2002).

[88] H. Kageyama, K. Hayashi, K. Maeda, J. P. Attfield, Z. Hiroi, J. M. Rondinelli, K. R. Poeppelmeier, Expanding frontiers in materials chemistry and physics with multiple anions. *Nature communications* **9**, 772 (2018).

[89] L. Ye, Y. Su, X. Jin, H. Xie, C. Zhang, Recent advances in biox (x= cl, br and i) photocatalysts: synthesis, modification, facet effects and mechanisms. *Environmental Science: Nano* **1**, 90–112 (2014).

[90] A. Kasahara, K. Nukumizu, T. Takata, J. N. Kondo, M. Hara, H. Kobayashi, K. Domen, Latio2n as a visible-light (< 600 nm)-driven photocatalyst (2). *The Journal of Physical Chemistry B* **107**, 791–797 (2003).

[91] L. Zhao, D. Berardan, Y. Pei, C. Byl, L. Pinsard-Gaudart, N. Dragoe, Bi 1- x sr x cuseo oxyselenides as promising thermoelectric materials. *Applied Physics Letters* **97**, 092118 (2010).

[92] P. Ruleova, C. Drasar, P. Lostak, C.-P. Li, S. Ballikaya, C. Uher, Thermoelectric properties of bi 2 o 2 se. *Materials Chemistry and Physics* **119**, 299–302 (2010).

[93] T. Nomura, S. W. Kim, Y. Kamihara, M. Hirano, P. V. Sushko, K. Kato, M. Takata, A. L. Shluger, H. Hosono, Crystallographic phase transition and high-tc superconductivity in lafeaso: F. *Superconductor Science and Technology* **21**, 125028 (2008).

[94] Y. Mizuguchi, H. Fujihisa, Y. Gotoh, K. Suzuki, H. Usui, K. Kuroki, S. Demura, Y. Takano, H. Izawa, O. Miura, Bis_2-based layered superconductor $bi_4o_4s_3$. *Phys. Rev. B* **86**, 220510 (2012).

[95] K. Ueda, S. Inoue, S. Hirose, H. Kawazoe, H. Hosono, Transparent p-type semiconductor: Lacuos layered oxysulfide. *Applied Physics Letters* **77**, 2701–2703 (2000).
[96] X. Fan, C. Luo, J. Lamb, Y. Zhu, K. Xu, C. Wang, Pedot encapsulated feof nanorod cathodes for high energy lithium-ion batteries. *Nano letters* **15**, 7650–7656 (2015).
[97] K. J. Michel, C. Wolverton, Symmetry building monte carlo-based crystal structure prediction. *Comput. Phys. Commun.* **185**, 1389–1393 (2014).
[98] L. Ward, K. Michel, Materials/mint: Initial release (2016).
[99] H. Luo, G. Liu, F. Meng, J. Li, E. Liu, G. Wu, Half-metallicity in Fe-based heusler alloys Fe_2TiZ (Z = Ga, Ge, As, In, Sn and Sb). *J. Magn. Magn. Mater.* **324**, 3295–3299 (2012).
[100] J. Tobola, J. Pierre, S. Kaprzyk, R. V. Skolozdra, M. A. Kouacou, Crossover from semiconductor to magnetic metal in semi-heusler phases as a function of valence electron concentration. *J. Phys. Condens. Matter* **10**, 1013–1032 (1998).
[101] N. Naghibolashrafi, S. Keshavarz, V. I. Hegde, A. Gupta, W. Butler, J. Romero, K. Munira, P. LeClair, D. Mazumdar, J. Ma, *et al.*, Synthesis and characterization of fe-ti-sb intermetallic compounds: Discovery of a new slater-pauling phase. *Phys. Rev. B* **93**, 104424 (2016).
[102] J. W. Gibbs, On the Equilibrium of Heterogeneous Substances. *Am. J. Sci.* **Series 3 Vol. 16**, 441–458 (1878).
[103] S. Rosenstein, P. P. Lamy, Some Aspects of Polymorphism. *Am. J. Hosp. Pharm.* **26**, 598–601 (1969).
[104] J.-P. Brog, C.-L. Chanez, A. Crochet, K. M. Fromm, Polymorphism, what it is and how to identify it: A Systematic Review. *RSC Adv.* **3**, 16905–16931 (2013).
[105] J. Gopalakrishnan, Chimie Douce Approaches to the Synthesis of Metastable Oxide Materials. *Chem. Mater.* **7**, 1265–1275 (1995).
[106] A. Stein, S. W. Keller, T. E. Mallouk, Turning Down the Heat: Design and Mechanism in Solid-State Synthesis. *Science* **259**, 1558–1564 (1993).
[107] W. Sun, S. T. Dacek, S. P. Ong, G. Hautier, A. Jain, W. D. Richards, A. C. Gamst, K. A. Persson, G. Ceder, The Thermodynamic Scale of Inorganic Crystalline Metastability. *Sci. Adv.* **2**, e1600225 (2016).
[108] A. R. Oganov, *Modern Methods of Crystal Structure Prediction* (Wiley-VCH Verlag GmbH & Co. KGaA, 2010), first edn.
[109] C. J. Pickard, R. J. Needs, Ab Initio Random Structure Searching. *J. Phys.: Condens. Matter* **23**, 053201 (2011).
[110] W. Zhang, A. R. Oganov, A. F. Goncharov, Q. Zhu, S. E. Boulfelfel, A. O. Lyakhov, E. Stavrou, M. Somayazulu, V. B. Prakapenka, Z. Konokova, Unexpected Stable Stoichiometries of Sodium Chlorides. *Science* **342**, 1502–1505 (2013).
[111] J. P. S. Walsh, S. M. Clarke, Y. Meng, S. D. Jacobsen, D. E. Freedman, Discovery of $FeBi_2$. *ACS Cent. Sci.* **2**, 867–871 (2016).
[112] L. Zhang, Y. Wang, J. Lv, Y. Ma, Materials Discovery at High Pressures. *Nat. Rev. Mater.* **2**, 17005 (2017).

[113] J. A. Flores-Livas, M. Amsler, C. Heil, A. Sanna, L. Boeri, G. Profeta, C. Wolverton, S. Goedecker, E. K. U. Gross, Superconductivity in Metastable Phases of Phosphorus-Hydride Compounds Under High Pressure. *Phys. Rev. B* **93**, 020508 (2016).

[114] T. Bi, D. P. Miller, A. Shamp, E. Zurek, Superconducting Phases of Phosphorus Hydride Under Pressure: Stabilization by Mobile Molecular Hydrogen. *Angewandte Chemie* **129**, 10326–10329 (2017).

[115] A. Shamp, T. Terpstra, T. Bi, Z. Falls, P. Avery, E. Zurek, Decomposition Products of Phosphine Under Pressure: PH_2 Stable and Superconducting? *J. Am. Chem. Soc.* **138**, 1884–1892 (2016).

[116] H. Liu, Y. Li, G. Gao, J. S. Tse, I. I. Naumov, Crystal Structure and Superconductivity of PH_3 at High Pressures. *J. Phys. Chem. C* **120**, 3458–3461 (2016).

[117] J. A. Flores-Livas, A. Sanna, E. K. U. Gross, High Temperature Superconductivity in Sulfur and Selenium Hydrides at High Pressure. *Eur. Phys. J. B* **89**, 63 (2016).

[118] Y. Li, J. Hao, H. Liu, Y. Li, Y. Ma, The Metallization and Superconductivity of Dense Hydrogen Sulfide. *The Journal of Chemical Physics* **140**, 174712 (2014).

[119] M. Amsler, S. Botti, M. A. L. Marques, T. J. Lenosky, S. Goedecker, Low-Density Silicon Allotropes for Photovoltaic Applications. *Phys. Rev. B* **92**, 014101 (2015).

[120] R. Rasoulkhani, H. Tahmasbi, S. A. Ghasemi, S. Faraji, S. Rostami, M. Amsler, Energy Landscape of ZnO Clusters and Low-Density Polymorphs. *Phys. Rev. B* **96**, 064108 (2017).

[121] Y. Wang, M. Miao, J. Lv, L. Zhu, K. Yin, H. Liu, Y. Ma, An Effective Structure Prediction Method for Layered Materials Based on 2D Particle Swarm Optimization Algorithm. *J. Chem. Phys.* **137**, 224108 (2012).

[122] X.-F. Zhou, X. Dong, A. R. Oganov, Q. Zhu, Y. Tian, H.-T. Wang, Semimetallic Two-Dimensional Boron Allotrope with Massless Dirac Fermions. *Phys. Rev. Lett.* **112**, 085502 (2014).

[123] H. A. Eivari, S. A. Ghasemi, H. Tahmasbi, S. Rostami, S. Faraji, R. Rasoulkhani, S. Goedecker, M. Amsler, Two-Dimensional Hexagonal Sheet of TiO_2. *Chem. Mater.* **29**, 8594–8603 (2017).

[124] A. L.-S. Chua, N. A. Benedek, L. Chen, M. W. Finnis, A. P. Sutton, A Genetic Algorithm for Predicting the Structures of Interfaces in Multicomponent Systems. *Nat. Mater.* **9**, 418 (2010).

[125] G. Schusteritsch, C. J. Pickard, Predicting Interface Structures: From $SrTiO_3$ to Graphene. *Phys. Rev. B* **90**, 035424 (2014).

[126] S. Lu, Y. Wang, H. Liu, M.-s. Miao, Y. Ma, Self-assembled ultrathin nanotubes on diamond (100) surface. *Nat. Commun.* **5**, 3666 (2014).

[127] X. Zhao, Q. Shu, M. C. Nguyen, Y. Wang, M. Ji, H. Xiang, K.-M. Ho, X. Gong, C.-Z. Wang, Interface Structure Prediction from First-Principles. *J. Phys. Chem. C* **118**, 9524–9530 (2014).

[128] G. Fisicaro, M. Sicher, M. Amsler, S. Saha, L. Genovese, S. Goedecker, Surface Reconstruction of Fluorites in Vacuum and Aqueous Environment. *Phys. Rev. Materials* **1**, 033609 (2017).

[129] J. E. Saal, S. Kirklin, M. Aykol, B. Meredig, C. Wolverton, Materials Design and Discovery with High-Throughput Density Functional Theory: The Open Quantum Materials Database (OQMD). *JOM* **65**, 1501 (2013).

[130] S. Kirklin, J. E. Saal, B. Meredig, A. Thompson, J. W. Doak, M. Aykol, S. Rühl, C. Wolverton, The Open Quantum Materials Database (OQMD): Assessing the Accuracy of DFT Formation Energies. *npj Computational Materials* **1**, 15010 (2015).

[131] S. Curtarolo, W. Setyawan, S. Wang, J. Xue, K. Yang, R. H. Taylor, L. J. Nelson, G. L. W. Hart, S. Sanvito, M. Buongiorno-Nardelli, N. Mingo, O. Levy, AFLOWLIB.ORG: A Distributed Materials Properties Repository from High-Throughput Ab Initio Calculations. *Comput. Mater. Sci.* **58**, 227 (2012).

[132] S. Curtarolo, G. L. W. Hart, M. B. Nardelli, N. Mingo, S. Sanvito, O. Levy, The High-Throughput Highway to Computational Materials Design. *Nat. Mater.* **12**, 191-201 (2013).

[133] C. Toher, *et al.*, The AFLOW Fleet for Materials Discovery. *arXiv:1712.00422 [cond-mat]* (2017). ArXiv: 1712.00422.

[134] Inorganic Crystal Structure Database, https://icsd.fiz-karlsruhe.de. [Online; accessed 20-September-2017].

[135] G. Shen, H. K. Mao, High-Pressure Studies with X-Rays using Diamond Anvil Cells. *Rep. Prog. Phys.* **80**, 016101 (2017).

[136] J. Tsuchiya, First principles prediction of a new high-pressure phase of dense hydrous magnesium silicates in the lower mantle. *Geophys. Res. Lett.* **40**, 4570–4573 (2013).

[137] T. H. Mason, X. Liu, J. Hong, J. Graetz, E. H. Majzoub, First-Principles Study of Novel Conversion Reactions for High-Capacity Li-Ion Battery Anodes in the Li-Mg-B-N-H System. *The Journal of Physical Chemistry C* **115**, 16681–16687 (2011).

[138] D. S. Aidhy, Y. Zhang, C. Wolverton, Prediction of a $Ca(BH_4)(NH_2)$ Quaternary Hydrogen Storage Compound from First-Principles Calculations. *Phys. Rev. B* **84**, 134103 (2011).

[139] J. Sun, R. C. Remsing, Y. Zhang, Z. Sun, A. Ruzsinszky, H. Peng, Z. Yang, A. Paul, U. Waghmare, X. Wu, M. L. Klein, J. P. Perdew, Accurate First-Principles Structures and Energies of Diversely Bonded Systems from an Efficient Density Functional. *Nat. Chem.* **8**, 831–836 (2016).

[140] R. G. Hennig, A. Wadehra, K. P. Driver, W. D. Parker, C. J. Umrigar, J. W. Wilkins, Phase Transformation in Si from Semiconducting Diamond to Metallic β-Sn Phase in QMC and DFT under Hydrostatic and Anisotropic Stress. *Phys. Rev. B* **82**, 014101 (2010).

[141] S. M. Clarke, J. P. S. Walsh, M. Amsler, C. D. Malliakas, T. Yu, S. Goedecker, Y. Wang, C. Wolverton, D. E. Freedman, Discovery of a Superconducting Cu-Bi Intermetallic Compound by High-Pressure Synthesis. *Angew. Chem. Int. Ed.* **55**,

13446–13449 (2016).

[142] S. M. Clarke, M. Amsler, J. P. S. Walsh, T. Yu, Y. Wang, Y. Meng, S. D. Jacobsen, C. Wolverton, D. E. Freedman, Creating Binary Cu–Bi Compounds via High-Pressure Synthesis: A Combined Experimental and Theoretical Study. *Chem. Mater.* **29**, 5276–5285 (2017).

[143] M. Amsler, S. S. Naghavi, C. Wolverton, Prediction of Superconducting Iron–Bismuth Intermetallic Compounds at High Pressure. *Chem. Sci.* **8**, 2226–2234 (2017).

[144] K. M. Powderly, S. M. Clarke, M. Amsler, C. Wolverton, C. D. Malliakas, Y. Meng, S. D. Jacobsen, D. E. Freedman, High-Pressure Discovery of β-NiBi. *Chem. Commun.* **53**, 11241–11244 (2017).

[145] M. Amsler, Z. Yao, C. Wolverton, Cubine, a Quasi Two-Dimensional Copper-Bismuth Nano Sheet. *Chem. Mater.* **29**, 9819–9828 (2017).

[146] M. Amsler, C. Wolverton, Dense Superconducting Phases of Copper-Bismuth at High Pressure. *Phys. Rev. Materials* **1**, 031801 (2017).

[147] G. Haegg, G. Funke, Roentgenalanlyse des Systems Nickel-Wismut. *Z. Phys. Chem., Abt. B* **6**, 272–283 (1929).

[148] V. P. Glagoleva, G. S. Zhdanov, Struktura Sverkhprovodnikov. 7. Rentgenograficheskoe Opredelenie Struktury Bi_3Ni. *Zh. Eksp. Teor. Fiz.* **26**, 337–344 (1954).

[149] M. Ruck, T. Söhnel, Transmissionsoptimierte Einkristallstrukturbestimmung und Elektronische Struktur von Bi_3Ni. *Z. Naturforsch.* **61b**, 785–791 (2006).

[150] N. Alekseevskii, N. Brandt, T. Kostina. *Izv. Akad. Nauk SSSR* **16**, 233 (1952).

[151] N. Alekseevskii, Sverkhprovodimost Soedinenii Vismuta. *Zh. Eksp. Teor. Fiz.* **18**, 101–102 (1948).

[152] T. Herrmannsdörfer, R. Skrotzki, J. Wosnitza, D. Köhler, R. Boldt, M. Ruck, Structure-Induced Coexistence of Ferromagnetic and Superconducting States of Single-Phase Bi_3Ni Seen via Magnetization and Resistance Measurements. *Phys. Rev. B* **83**, 140501 (2011).

[153] F. Bachhuber, J. Rothballer, T. Söhnel, R. Weihrich, Phase Stabilities at a Glance: Stability Diagrams of Nickel Dipnictides. *J. Chem. Phys.* **139**, 214705 (2013).

[154] C. S. Barrett, L. Meyer, J. Wasserman, Antiferromagnetic and Crystal Structures of Alpha–Oxygen. *The Journal of Chemical Physics* **47**, 592–597 (1967).

[155] M. Nicol, K. Syassen, High-Pressure Optical Spectra of Condensed Oxygen. *Phys. Rev. B* **28**, 1201–1206 (1983).

[156] L. F. Lundegaard, G. Weck, M. I. McMahon, S. Desgreniers, P. Loubeyre, Observation of an O_8 Molecular Lattice in the ϵ Phase of Solid Oxygen. *Nature* **443**, 201–204 (2006).

[157] H. Fujihisa, Y. Akahama, H. Kawamura, Y. Ohishi, O. Shimomura, H. Yamawaki, M. Sakashita, Y. Gotoh, S. Takeya, K. Honda, O_8 Cluster Structure of the Epsilon Phase of Solid Oxygen. *Phys. Rev. Lett.* **97**, 085503 (2006).

[158] Y. Ma, A. R. Oganov, C. W. Glass, Structure of the Metallic ζ-Phase of Oxygen and Isosymmetric Nature of the $\epsilon - \zeta$ Phase Transition: Ab Initio Simulations. *Phys. Rev. B* **76**, 064101 (2007).

[159] G. Weck, S. Desgreniers, P. Loubeyre, M. Mezouar, Single-Crystal Structural Characterization of the Metallic Phase of Oxygen. *Phys. Rev. Lett.* **102**, 255503 (2009).

[160] S. K. Sikka, Y. K. Vohra, R. Chidambaram, Omega Phase in Materials. *Progress in Materials Science* **27**, 245–310 (1982).

[161] H. Xia, S. J. Duclos, A. L. Ruoff, Y. K. Vohra, New High-Pressure Phase Transition in Zirconium Metal. *Phys. Rev. Lett.* **64**, 204–207 (1990).

[162] S. Desgreniers, K. Lagarec, High-Density ZrO_2 and HfO_2: Crystalline Structures and Equations of State. *Phys. Rev. B* **59**, 8467–8472 (1999).

[163] O. Ohtaka, T. Nagai, T. Yamanaka, T. Yagi, O. Shimomura, *Properties of Earth and Planetary Materials at High Pressure and Temperature* (American Geophysical Union (AGU), 2013), pp. 429–433.

[164] N. Schönberg, The Structure of the Metallic Quaternary Phase ZrTaNO. *Acta Chemica Scandinavica* **8**, 627–629 (1954).

[165] A. S. Cooper, Precise Lattice Constants of Germanium, Aluminum, Gallium Arsenide, Uranium, Sulphur, Quartz and Sapphire. *Acta Cryst, Acta Crystallogr* **15**, 578–582 (1962).

[166] S. B. Qadri, E. F. Skelton, A. W. Webb, High Pressure Studies of Ge Using Synchrotron Radiation. *Journal of Applied Physics* **54**, 3609–3611 (1983).

[167] Y. K. Vohra, K. E. Brister, S. Desgreniers, A. L. Ruoff, K. J. Chang, M. L. Cohen, Phase-Transition Studies of Germanium to 1.25 Mbar. *Phys. Rev. Lett.* **56**, 1944–1947 (1986).

[168] K. Takemura, U. Schwarz, K. Syassen, N. E. Christensen, M. Hanfland, D. L. Novikov, I. Loa, High-Pressure Structures of Ge above 100 GPa. *physica status solidi (b)* **223**, 385–390 (2001).

[169] K. Takemura, U. Schwarz, K. Syassen, M. Hanfland, N. E. Christensen, D. L. Novikov, I. Loa, High-Pressure *Cmca* and hcp Phases of Germanium. *Phys. Rev. B* **62**, R10603–R10606 (2000).

[170] E. Y. Tonkov, E. G. Ponyatovsky, *Phase Transformations of Elements Under High Pressure* (CRC Press, Boca Raton, Fla, 2004), first edn.

[171] A. Mujica, S. Radescu, A. Muoz, R. J. Needs, High-Pressure Phases of Germanium. *J. Phys.: Condens. Matter* **13**, 35 (2001).

[172] J. Haines, O. Cambon, E. Philippot, L. Chapon, S. Hull, A Neutron Diffraction Study of the Thermal Stability of the α-Quartz-Type Structure in Germanium Dioxide. *Journal of Solid State Chemistry* **166**, 434–441 (2002).

[173] N. Brar, H. Schloessin, The Kinetics of the GeO_2 (α-Quartz) $\rightarrow$ (Rutile) Transformation Under High Pressure. *High Temperatures - High Pressures* **13**, 313–320 (1981).

[174] J. Haines, J. M. Lager, C. Chateau, A. S. Pereira, Structural Evolution of Rutile–Type and CaCl–Type Germanium Dioxide at High Pressure. *Phys Chem Min* **27**, 575–582 (2000).

[175] K. Shiraki, T. Tsuchiya, S. Ono, Structural Refinements of High-Pressure Phases in Germanium Dioxide. *Acta Crystallogr., Sect. B: Struct. Crystallogr. Cryst. Chem.* **59**, 701–708 (2003).

[176] H. Fujihisa, Y. Akahama, H. Kawamura, Y. Ohishi, Y. Gotoh, H. Yamawaki, M. Sakashita, S. Takeya, K. Honda, Incommensurate Structure of Phosphorus Phase IV. *Phys. Rev. Lett.* **98**, 175501 (2007).

[177] S. Kirklin, J. E. Saal, V. I. Hegde, C. Wolverton, High-Throughput Computational Search for Strengthening Precipitates in Alloys. *Acta Mater.* **102**, 125–135 (2016).

[178] T. B. Massalski, H. Okamoto, P. Subramanian, L. Kacprzak, W. W. Scott, *Binary Alloy Phase Diagrams*, vol. 1 (American Society for Metals, Metals Park, OH, 1986).

[179] G. Kresse, J. Hafner, Ab Initio Molecular Dynamics for Liquid Metals. *Phys. Rev. B* **47**, 558 (1993).

[180] G. Kresse, J. Furthmüller, Efficiency of Ab-Initio Total Energy Calculations for Metals and Semiconductors Using a Plane-Wave Basis Set. *Comput. Mater. Sci.* **6**, 15–50 (1996).

[181] G. Kresse, J. Furthmüller, Efficient Iterative Schemes for Ab Initio Total-Energy Calculations Using a Plane-Wave Basis Set. *Phys. Rev. B* **54**, 11169 (1996).

[182] P. E. Blöchl, Projector Augmented-Wave Method. *Phys. Rev. B* **50**, 17953 (1994).

[183] G. Kresse, D. Joubert, From ultrasoft pseudopotentials to the projector augmented-wave method. *Phys. Rev. B* **59**, 1758–1775 (1999).

[184] J. P. Perdew, K. Burke, M. Ernzerhof, Generalized Gradient Approximation Made Simple. *Phys. Rev. Lett.* **77**, 3865 (1996).

[185] S. Goedecker, Minima Hopping: An Efficient Search Method for the Global Minimum of the Potential Energy Surface of Complex Molecular Systems. *J. Chem. Phys.* **120**, 9911 (2004).

[186] M. Amsler, S. Goedecker, Crystal Structure Prediction Using the Minima Hopping Method. *J. Chem. Phys.* **133**, 224104 (2010).

[187] M. Amsler, J. A. Flores-Livas, T. D. Huan, S. Botti, M. A. L. Marques, S. Goedecker, Novel Structural Motifs in Low Energy Phases of $LiAlH_4$. *Phys. Rev. Lett.* **108**, 205505 (2012).

[188] T. D. Huan, M. Amsler, V. N. Tuoc, A. Willand, S. Goedecker, Low-Energy Structures of Zinc Borohydride $Zn(BH_4)_2$. *Phys. Rev. B* **86**, 224110 (2012).

[189] T. D. Huan, M. Amsler, R. Sabatini, V. N. Tuoc, N. B. Le, L. M. Woods, N. Marzari, S. Goedecker, Thermodynamic Stability of Alkali-Metal-Zinc Double-Cation Borohydrides at Low Temperatures. *Phys. Rev. B* **88**, 024108 (2013).

[190] S. Roy, S. Goedecker, V. Hellmann, Bell-Evans-Polanyi Principle for Molecular Dynamics Trajectories and its Implications for Global Optimization. *Phys. Rev. E* **77**, 056707 (2008).

[191] M. Sicher, S. Mohr, S. Goedecker, Efficient Moves for Global Geometry Optimization Methods and their Application to Binary Systems. *J. Chem. Phys.* **134**, 044106–044106-7 (2011).
[192] M. Amsler, J. A. Flores-Livas, L. Lehtovaara, F. Balima, S. A. Ghasemi, D. Machon, S. Pailhès, A. Willand, D. Caliste, S. Botti, A. San Miguel, S. Goedecker, M. A. L. Marques, Crystal Structure of Cold Compressed Graphite. *Phys. Rev. Lett.* **108**, 065501 (2012).
[193] J. A. Flores-Livas, M. Amsler, T. J. Lenosky, L. Lehtovaara, S. Botti, M. A. L. Marques, S. Goedecker, High-Pressure Structures of Disilane and their Superconducting Properties. *Phys. Rev. Lett.* **108**, 117004 (2012).
[194] C. B. Barber, D. P. Dobkin, H. Huhdanpaa, The Quickhull Algorithm for Convex Hulls. *ACM Transactions On Mathematical Software* **22**, 469–483 (1996).
[195] E. Jones, T. Oliphant, P. Peterson, *et al.*, SciPy: Open Source Scientific Tools for Python (2001).
[196] S. Mitchell, S. M. Consulting, I. Dunning, PuLP: A Linear Programming Toolkit for Python (2011).
[197] V. I. Hegde, M. Amsler, github.com/wolverton-research-group/nve (2017).
[198] D. J. Watts, S. H. Strogatz, Collective dynamics of 'small-world' networks. *Nature* **393**, 440-442 (1998).
[199] R. Albert, H. Jeong, A.-L. Barabási, Diameter of the world-wide web. *Nature* **401**, 130 (1999).
[200] D. J. Watts, P. S. Dodds, M. E. Newman, Identity and search in social networks. *Science* **296**, 1302–1305 (2002).
[201] M. E. J. Newman, The structure of scientific collaboration networks. *Proc. Natl. Acad. Sci. USA* **98**, 404–409 (2001).
[202] R. J. Williams, N. D. Martinez, Simple rules yield complex food webs. *Nature* **404**, 180 (2000).
[203] R. Guimerà, L. A. Nunes Amaral, Functional cartography of complex metabolic networks. *Nature* **433**, 895–900 (2005).
[204] A.-L. Barabási, R. Albert, Emergence of scaling in random networks. *Science* **286**, 509–512 (1999).
[205] S. H. Strogatz, Exploring complex networks. *Nature* **410**, 268 (2001).
[206] H. Jeong, R. Tombor, R. Albert, Z. N. Oltvai, A.-L. Barabási, The large-scale organization of metabolic networks. *Nature* **407**, 651–654 (2000).
[207] J. Hachmann, R. Olivares-Amaya, S. Atahan-Evrenk, C. Amador-Bedolla, R. S. Sánchez-Carrera, A. Gold-Parker, L. Vogt, A. M. Brockway, A. Aspuru-Guzik, The Harvard clean energy project: Large-scale computational screening and design of organic photovoltaics on the world community grid. *J. Phys. Chem. Lett.* **2**, 2241–2251 (2011).
[208] M. T. Dove, *Structure and dynamics: an atomic view of materials* (Oxford University Press, 2003).

[209] R. Phillips, *Crystals, Defects and Microstructures: Modeling Across Scales* (Cambridge University Press, 2001).
[210] A. Belsky, M. Hellenbrandt, V. L. Karen, P. Luksch, New developments in the Inorganic Crystal Structure Database (ICSD): accessibility in support of materials research and design. *Acta Cryst. B* **58**, 364–369 (2002).
[211] M. E. J. Newman, The Structure and Function of Complex Networks. *SIAM Review* **45**, 167–256 (2003).
[212] S. P. Ong, Y. Mo, W. D. Richards, L. Miara, H. S. Lee, G. Ceder, Phase stability, electrochemical stability and ionic conductivity of the $Li_{10\pm1}MP_2X_{12}$ (M = Ge, Si, Sn, Al or P, and X = O, S or Se) family of superionic conductors. *Energy Environ. Sci.* **6**, 148–156 (2013).
[213] M. Aykol, S. Kim, V. I. Hegde, D. Snydacker, Z. Lu, S. Hao, S. Kirklin, D. Morgan, C. Wolverton, High-throughput computational design of cathode coatings for Li-ion batteries. *Nat. Commun.* **7**, 13779 (2016).
[214] M. Aykol, S. Kim, V. I. Hegde, D. Snydacker, Z. Lu, S. Hao, S. Kirklin, D. Morgan, C. Wolverton, High-throughput computational design of cathode coatings for Li-ion batteries. *Nat. Commun.* **7**, 13779 (2016).
[215] D. H. Snydacker, V. I. Hegde, C. Wolverton, Electrochemically Stable Coating Materials for Li, Na, and Mg Metal Anodes in Durable High Energy Batteries. *J. Electrochem. Soc.* **164**, A3582–A3589 (2017).
[216] M. Mitzenmacher, A brief history of generative models for power law and lognormal distributions. *Internet Math.* **1**, 226–251 (2003).
[217] C. I. Del Genio, T. Gross, K. E. Bassler, All Scale-Free Networks Are Sparse. *Phys. Rev. Lett.* **107**, 178701 (2011).
[218] E. Ravasz, A.-L. Barabási, Hierarchical organization in complex networks. *Phys. Rev. E* **67**, 026112 (2003).
[219] S. Maslov, K. Sneppen, A. Zaliznyak, Detection of topological patterns in complex networks: correlation profile of the internet. *Physica A* **333**, 529–540 (2004).
[220] W. Sun, C. J. Bartel, E. Arca, S. R. Bauers, B. Matthews, B. Orvañanos, B.-R. Chen, M. F. Toney, L. T. Schelhas, W. Tumas, J. Tate, A. Zakutayev, S. Lany, A. M. Holder, G. Ceder, A map of the inorganic ternary metal nitrides. *Nat. Mater.* **18**, 732–739 (2019).
[221] W. Sun, S. T. Dacek, S. P. Ong, G. Hautier, A. Jain, W. D. Richards, A. C. Gamst, K. A. Persson, G. Ceder, The thermodynamic scale of inorganic crystalline metastability. *Sci. Adv.* **2**, e1600225 (2016).
[222] M. Widom, Frequency Estimate for Multicomponent Crystalline Compounds. *J. Stat. Phys.* **167**, 726–734 (2017).
[223] M. Aykol, S. S. Dwaraknath, W. Sun, K. A. Persson, Thermodynamic limit for synthesis of metastable inorganic materials. *Sci. Adv.* **4**, eaaq0148 (2018).
[224] B. Hammer, J. K. Norskov, Why gold is the noblest of all the metals. *Nature* **376**, 238–240 (1995).

[225] L. Ward, A. Agrawal, A. Choudhary, C. Wolverton, A general-purpose machine learning framework for predicting properties of inorganic materials. *npj Comput. Mater.* **2**, 16028 (2016).
[226] G. Klopman, Chemical Reactivity and the Concept of Charge- and Frontier-Controlled Reactions. *J. Am. Chem. Soc.* **90**, 223–234 (1968).
[227] W. Yang, R. G. Parr, Hardness, softness, and the fukui function in the electronic theory of metals and catalysis. *Proc. Natl. Acad. Sci. USA* **82**, 6723–6726 (1985).
[228] H. Chermette, Chemical reactivity indexes in density functional theory. *J. Comput. Chem.* **20**, 129–154 (1999).
[229] M. Aykol, V. I. Hegde, L. Hung, S. Suram, P. Herring, C. Wolverton, J. S. Hummelshøj, Network analysis of synthesizable materials discovery. *Nat. Commun.* **10**, 2018 (2019).
[230] C. B. Barber, D. P. Dobkin, H. Huhdanpaa, The quickhull algorithm for convex hulls. *ACM Trans. Math. Softw.* **22**, 469–483 (1996).
[231] T. P. Peixoto, The graph-tool python library. *figshare* (2014).
[232] J. Alstott, E. Bullmore, D. Plenz, Powerlaw: a Python package for analysis of heavy-tailed distributions. *PLoS ONE* **9**, e85777 (2014).
[233] H. Sartipizadeh, T. L. Vincent, Computing the Approximate Convex Hull in High Dimensions. Preprint at https://arxiv.org/abs/1603.04422 (2016).
[234] A. Clauset, C. R. Shalizi, M. E. J. Newman, Power-Law Distributions in Empirical data. *SIAM Rev.* **51**, 661–703 (2009).

APPENDIX A

Exploring the high-pressure genome: supplemental data

A.1. Experimentally reported high-pressure elemental phases

Table A.1: List of experimentally reported high-pressure elemental phases calculated in the OQMD. For each phase, we report its composition, space group, the ICSD Collection Code, OQMD ID, experimentally reported pressure (p_{ICSD}), pressure range of stability predicted from the linear approximation to enthalpy based model (p_{min}, p_{max}), and formation energy at ambient pressure (ΔE_f).

Composition	Space group	ICSD (Collection	OQMD Code/ID)	p_{ICSD} (GPa)	(p_{min}, p_{max}) (GPa)	ΔE_f (eV/atom)
Ar	P63/mmc	77918	604613	6.5	($-\infty$, 2.0)	0.009
As	$Pm\bar{3}m$	162842	685982	30.0	(7.7, ∞)	0.111
B	Pnnm	165132	630765	9.0	(16.9, ∞)	0.027
Bi	C12/m1	409752	689171	2.5	(2.7, 5.8)	0.042

Table A.1 — Continued

Composition	Space group	ICSD (Collection	OQMD Code/ID)	p_{ICSD} (GPa)	$(p_{\mathrm{min}},\ p_{\mathrm{max}})$ (GPa)	ΔE_f (eV/atom)
	$Im\bar{3}m$	52725	676488	9.0	(5.8, ∞)	0.127
Br	I4/mmm	168174	632558	145.0	(10.3, 21.1)	0.172
	$Fm\bar{3}m$	168175	686736	165.0	(21.1, ∞)	0.638
Ca	$Im\bar{3}m$	52749	592558	26.5	(-22.0, -5.4)	0.015
	Cmmm	168759	632881	47.0	(26.9, ∞)	0.313
	Pn21a	168757	632879	172.0	(1.7, 26.9)	0.015
Ce	Cmcm	601481	594082	5.6	($-\infty$, -24.9)	0.027
	I4/mmm	52844	592229	17.5	(-24.9, -4.9)	0.006
Cs	I41/amdS	109021	620946	5.0	($-\infty$, -2.6)	0.055
	P42/mbc	173928	635511	18.0	(0.4, ∞)	0.004
	Pnma	173929	635512	18.0	(-2.6, -0.7)	0.007
	P63/mmc	57180	676673	92.0	(-0.7, 0.1)	0.000
Dy	$R\bar{3}mH$	629537	593928	7.0	(-3.0, 18.0)	0.000
	Cmmm	157920	626575	87.0	(18.0, ∞)	0.017
Eu	P63/mmc	604034	676509	13.0	(13.4, ∞)	0.028
Fe	P63/mmc	53450	51781	12.7	(12.4, ∞)	0.086
Ga	$I\bar{4}3d$	12173	2814	2.6	(6.1, ∞)	0.026

Table A.1 — Continued

Composition	Space group	ICSD (Collection	OQMD Code/ID)	p_{ICSD} (GPa)	$(p_{\mathrm{min}},\ p_{\mathrm{max}})$ (GPa)	ΔE_f (eV/atom)
	I4/mmm	12174	675527	2.8	(3.0, 6.1)	0.024
Gd	R$\bar{3}$mH	635711	9312	3.5	(1.5, ∞)	0.001
Ge	I41/amdS	53643	676517	80.0	(8.0, 124.6)	0.233
	Imma	246628	676517	80.0	(8.0, 124.6)	0.233
	Cmca	94305	16307	135.0	(124.6, ∞)	0.317
He	P63/mmc	44394	8392	0.0	($-\infty$, -3.3)	0.009
Hf	Im$\bar{3}$m	53023	676496	71.0	(79.5, ∞)	0.180
Hg	I4/mmm	43558	676237	1.0	(10.3, ∞)	0.056
	C12/m1	157861	20117	20.0	(2.5, 10.3)	0.032
I	I4/mmm	109040	19480	49.0	(6.5, ∞)	0.100
	Fmm2	151434	19480	49.0	(6.5, ∞)	0.100
In	Fmmm	57392	676455	93.0	(-0.0, ∞)	0.000
K	Fm$\bar{3}$m	44669	676282	12.4	(0.1, ∞)	0.000
	I4/mcm	157564	100324	19.5	($-\infty$, -7.0)	0.105
	Pnma	165995	21103	58.0	(-7.0, -1.8)	0.035
Kr	P63/mmc	9785	2562	3.2	($-\infty$, 1.8)	0.000
La	Fm$\bar{3}$m	104655	17557	2.3	(12.3, ∞)	0.013

Table A.1 — Continued

Composition	Space group	ICSD (Collection Code/ID)	OQMD	p_{ICSD} (GPa)	$(p_{\min},\ p_{\max})$ (GPa)	ΔE_f (eV/atom)
Li	P63/mmc	44760	676289	3.0	(2.3, 3.2)	0.005
	$Fm\bar{3}m$	57408	13862	8.0	(1.7, 2.3)	0.001
	$R\bar{3}mH$	109011	13862	8.0	(1.7, 2.3)	0.001
	$I\bar{4}3d$	109012	18968	45.0	(3.2, 23.7)	0.027
	P4132	161377	20426	400.0	$(23.7, \infty)$	1.356
Lu	$R\bar{3}mH$	642415	668371	23.0	$(8.1, \infty)$	0.013
Mg	$Fm\bar{3}m$	180453	687441	50.0	$(-\infty, -8.1)$	0.013
	$Im\bar{3}m$	642652	8597	58.0	$(349.5, \infty)$	0.029
N	$R\bar{3}cH$	40936	7386	7.8	(6.5, 39.9)	0.070
	P213	24893	4724	37.9	(174.1, 563.4)	3.461
	P42/mnm	24891	4722	40.1	(2.6, 6.5)	0.010
Na	$Im\bar{3}m$	44757	8534	0.2	(1.7, 11.1)	0.003
	$I\bar{4}3d$	159431	20238	108.0	$(11.1, \infty)$	1.257
Nd	$Fm\bar{3}m$	76591	677940	5.0	$(-127.1, -9.3)$	0.007
O	P63/mmc	164724	20846	15.9	$(28.3, \infty)$	0.098
P	Cmca	36436	4437	1.3	(1.6, 2.0)	0.036
	$R\bar{3}mH$	600019	9540	9.0	(2.0, 5.6)	0.080

Table A.1 — Continued

Composition	Space group	ICSD (Collection	OQMD Code/ID)	p_{ICSD} (GPa)	$(p_{\min},\ p_{\max})$ (GPa)	ΔE_f (eV/atom)
	$Pm\bar{3}m$	600029	20542	124.0	$(5.6, \infty)$	0.132
	Amm2	162244	20542	124.0	$(5.6, \infty)$	0.132
Pb	P63/mmc	77864	682982	13.9	$(-\infty, -25.2)$	0.014
	$Im\bar{3}m$	54314	676681	171.0	$(87.0, \infty)$	0.043
Pr	$R\bar{3}mH$	54338	9789	13.8	$(-115.5, -7.8)$	0.007
	P3121	108682	18882	14.4	$(2.0, \infty)$	0.019
	Cmcm	164283	686415	23.3	$(-\infty, -115.5)$	0.028
Rb	Cmca	109019	18973	48.1	$(-1.5, -0.8)$	0.006
S	$R\bar{3}mH$	57164	676672	206.5	$(17.9, \infty)$	1.111
Sb	$Pm\bar{3}m$	52227	51014	7.0	(4.2, 10.3)	0.043
	$Im\bar{3}m$	108182	18780	28.0	$(10.3, \infty)$	0.228
Sc	I4/mmm	246446	122328	23.0	(27.1, 29.6)	0.212
	P6122	153837	19753	242.0	$(29.6, \infty)$	1.936
Se	$Pm\bar{3}m$	52418	9277	0.0	(6.7, 28.2)	0.166
	P3121	23073	38563	8.6	(0.8, 1.2)	0.004
	$Im\bar{3}m$	659257	684882	150.0	$(28.2, \infty)$	0.578
Si	I41/amdS	109025	18979	13.0	(9.2, 10.2)	0.291

Table A.1 — Continued

Composition	Space group	ICSD (Collection	OQMD Code/ID)	p_{ICSD} (GPa)	$(p_{\mathrm{min}},\ p_{\mathrm{max}})$ (GPa)	ΔE_f (eV/atom)
	I41/amdS	52460	18979	13.0	(9.2, 10.2)	0.291
	P6/mmm	109035	18984	16.2	(10.2, 43.6)	0.308
	Cmca	89414	15748	42.5	(43.6, 45.1)	0.434
	P63/mmc	52459	9288	43.0	(45.1, ∞)	0.508
Sm	P63/mmc	652632	676475	2.0	(-47.9, 3.5)	0.000
	P3121	246657	23315	77.0	(3.5, ∞)	0.011
Sn	Im$\bar{3}$m	52487	92397	53.0	(5.5, ∞)	0.073
	I4/mmm	108748	92397	53.0	(5.5, ∞)	0.073
Sr	Im$\bar{3}$m	52490	8520	4.2	(0.7, ∞)	0.005
	I4/mcm	109026	18980	56.0	($-\infty$, -5.8)	0.143
Tb	R$\bar{3}$mH	652942	9303	6.0	(-4.6, 7.7)	0.658
	P63/mmc	52496	9302	28.8	(7.7, ∞)	0.665
Te	C12/m1	97742	16685	8.0	(3.1, 10.1)	0.045
	R$\bar{3}$mH	52500	676480	11.5	(10.1, ∞)	0.308
Th	I4/mmm	104198	17355	102.0	(-60.8, 0.1)	0.000
Ti	P63/mmc	183409	8079	2.0	($-\infty$, -8.5)	0.014
Tl	Fm$\bar{3}$m	60650	676841	6.0	(11.6, ∞)	0.018

Table A.1 — Continued

Composition	Space group	ICSD (Collection Code/ID)	OQMD (Collection Code/ID)	p_{ICSD} (GPa)	$(p_{\mathrm{min}},\ p_{\mathrm{max}})$ (GPa)	ΔE_f (eV/atom)
Tm	R$\bar{3}$mH	653347	671018	11.6	(2.6, ∞)	0.006
Xe	P63/mmc	9786	675519	3.0	(0.8, ∞)	0.004
Y	Fm$\bar{3}$m	106221	676150	46.0	($-\infty$, -256.3)	0.024
Yb	Fm$\bar{3}$m	43585	8144	0.3	(-3.9, 3.0)	0.000
	Im$\bar{3}$m	43572	8134	4.0	(3.0, ∞)	0.006
	P63/mmc	104205	9319	34.0	($-\infty$, -3.9)	0.007
Zr	Im$\bar{3}$m	169452	9320	30.0	(19.9, ∞)	0.080

A.2. Experimentally reported high-pressure binary phases

Table A.2: List of experimentally reported high-pressure binary phases calculated in the OQMD. For each phase, we report its composition, space group, the ICSD Collection Code, OQMD ID, experimentally reported pressure (p_{ICSD}), pressure range of stability predicted from the linear approximation to enthalpy based model (p_{min}, p_{max}), and formation energy at ambient pressure (ΔE_f).

Composition	Space group	ICSD (Collection	OQMD Code/ID)	p_{ICSD} (GPa)	(p_{min}, p_{max}) (GPa)	ΔE_f (eV/atom)
TeO_2	P212121	26844	5143	2.0	(−57.9, 3.4)	−1.300
CeBi	P4/mmm	58765	10917	11.0	(−65.0, 7.8)	−0.988
FeO	Bbmm	82236	14796	41.0	(−15.3, 92.7)	−1.519
FeO	Pbcn	82236	14796	76.0	(−15.3, 92.7)	−1.519
In_3Pd_2	I4/mmm	59478	11131	0.0	(−40.8, 8.1)	−0.481
ZnTe	Cmcm	31840	6116	15.7	(9.5, 24.7)	−0.176
AgBr	P121/m1	52246	646867	9.3	(−474.7, 4.0)	−0.493
SnTe	Pnma	52489	646930	2.0	(−2.9, 11.1)	−0.332
LaAg	P4/mmm	58288	10639	6.0	(−∞, 57.5)	−0.265
KBr	$Pm\bar{3}m$	61556	11512	2.2	(2.5, ∞)	−1.856

Table A.2 — Continued

Composition	Space group	ICSD (Collection Code/ID)	OQMD	p_{ICSD} (GPa)	$(p_{\mathrm{min}},\ p_{\mathrm{max}})$ (GPa)	ΔE_f (eV/atom)
H_3N	$Fm\bar{3}m$	29321	5723	1.3	(−25.6, 5.2)	−0.374
GdGa	P6/mmm	109240	19090	14.8	(−59.8, 16.2)	−0.622
$CaCl_2$	Pnam	51238	646691	8.0	$(-\infty, \infty)$	−2.636
Sm_3Te_4	$Pm\bar{3}m$	652658	31077	11.0	(−17.4, 24.2)	−1.390
KCl	$Fm\bar{3}m$	18014	3683	1.6	(−12.8, 3.3)	−2.141
KCl	P63/mmc	18014	3683	20.0	(−12.8, 3.3)	−2.141
PbS	Cmcm	62192	11589	3.7	(−0.1, 15.3)	−0.685
RbTe	P63/mmc	73179	13118	1.3	(−8.3, 0.6)	−0.847
MnSe	P63/mmc	76218	13692	9.0	(0.2, 12.7)	−0.355
KI	$Pm\bar{3}m$	22158	4108	2.0	(−18.5, 2.1)	−1.584
$TmTe_2$	P4mm	653106	31101	12.0	(−25.4, 41.7)	−1.074
SrF_2	P63/mmc	168801	21431	44.4	$(-\infty, 49.6)$	−4.001
HgS	$Fm\bar{3}m$	24094	4523	23.4	(−41.5, 2.0)	−0.249
HgS	P3221	24094	4523	0.4	(−41.5, 2.0)	−0.249
BaF_2	$Fm\bar{3}m$	41651	7546	1.5	$(-\infty, 40.5)$	−3.966
MnF_4	$P\bar{4}2m$	62068	11563	3.5	(−168.2, 248.9)	−2.496
Y_2S_3	$Fm\bar{3}m$	67502	12316	12.5	(−9.2, 8.2)	−2.251
SrS	$Pm\bar{3}m$	28900	5624	23.0	(−24.5, 15.4)	−2.295
AgN_3	I4/mcm	27135	5254	13.8	(29.1, 107.9)	0.491

Table A.2 — Continued

Composition	Space group	ICSD (Collection Code/ID)	OQMD	p_{ICSD} (GPa)	$(p_{\mathrm{min}},\ p_{\mathrm{max}})$ (GPa)	ΔE_f (eV/atom)
SiO_2	P41212	100279	87467	0.2	(4.8, 5.3)	−3.039
Bi_4O_7	P31c	51778	646777	0.3	(−7.2, 11.2)	−1.399
In_2Te_5	$Fm\bar{3}m$	1385	1910	7.4	(1.0, 3.0)	−0.219
FeSb	I4/mcm	53535	647220	21.0	(8.5, ∞)	0.009
EuTe	$Pm\bar{3}m$	631697	30151	11.0	($-\infty$, 23.0)	−1.599
EuF_3	$P\bar{3}c1$	95244	16411	8.2	(9.2, 2311.0)	−3.404
ZnTe	P3121	67779	12375	8.9	($-\infty$, −14.0)	−0.466
SnSb	$Fm\bar{3}m$	52302	646878	7.5	(8.5, 74.2)	0.081
$CuBr_2$	$Pa\bar{3}$	409450	25011	7.3	($-\infty$, 16.7)	−0.511
$CuBr_2$	P4/nmmS	409450	25011	5.1	($-\infty$, 16.7)	−0.511
SiO_2	Pbcn	100279	87467	129.0	(4.8, 5.3)	−3.039
CdTe	Pmm2	150941	19425	19.3	(3.3, 3.9)	−0.461
Pb_2O	Pnam	28838	5612	62.7	($-\infty$, −19.2)	−0.546
CdTe	Cmcm	150941	19425	11.7	(3.3, 3.9)	−0.461
SbO_2	C12/c1	31103	42222	9.2	(61.0, ∞)	−1.189
SbO_2	Pna21	31103	42222	9.8	(61.0, ∞)	−1.189
CsI_3	$Pm\bar{3}m$	27252	5285	1.5	(−3.3, 2.9)	−0.969
LaAs	P4/mmm	106265	18199	27.1	(−208.3, 15.1)	−1.570
$BaBr_2$	P1121/a	79892	14425	25.4	(−49.5, −4.0)	−2.432

Table A.2 — Continued

Composition	Space group	ICSD (Collection Code/ID)	OQMD	p_{ICSD} (GPa)	$(p_{\mathrm{min}},\ p_{\mathrm{max}})$ (GPa)	ΔE_f (eV/atom)
Fe_3O_4	$Fd\bar{3}mZ$	65338	11962	11.0	$(-\infty, -65.7)$	−1.033
Fe_3O_4	Cmcm	65338	11962	64.6	$(-\infty, -65.7)$	−1.033
ReO_3	P4/mbm	105546	17897	0.5	(−89.3, 4.1)	−1.901
$CaSi_2$	$P\bar{3}m1$	32006	6152	9.8	(−0.2, 2.3)	−0.390
SnTe	$Pm\bar{3}m$	52489	646930	32.0	(−2.9, 11.1)	−0.332
BaC_2	$R\bar{3}mH$	168408	21391	6.9	$(9.1, \infty)$	0.409
RbCl	$Pm\bar{3}m$	61521	56986	0.8	$(2.0, \infty)$	−2.071
BaC_2	I4/mmm	168408	21391	1.0	$(9.1, \infty)$	0.409
HgTe	$Fm\bar{3}m$	31842	6117	8.2	(5.1, 11.7)	0.014
Cs_2S_3	Pnma	14094	2839	7.4	(−1.0, 29.2)	−0.994
Na_3N	P6/mmm	165991	21099	1.7	(3.9, 5.2)	0.148
HgTe	Cmcm	31842	6117	18.5	(5.1, 11.7)	0.014
Zr_3O	Pnam	27023	5214	24.6	(−16.2, 1.5)	−1.479
ReO_3	$Im\bar{3}$	105546	17897	6.4	(−89.3, 4.1)	−1.901
FeSe	Pbnm	169305	110581	23.0	$(52.7, \infty)$	−0.036
FeS_2	P63mc	10422	2742	5.2	$(2.7, \infty)$	−0.694
SrH_2	Pnma	33160	6231	5.0	$(7.0, \infty)$	−0.495
LaN	$Fm\bar{3}m$	162194	20522	2.6	$(33.5, \infty)$	−0.832
SiO_2	$Fd\bar{3}mZ$	100279	87467	3.0	(4.8, 5.3)	−3.039

Table A.2 — Continued

Composition	Space group	ICSD (Collection Code/ID)	OQMD	p_{ICSD} (GPa)	$(p_{\min},\ p_{\max})$ (GPa)	ΔE_f (eV/atom)
NaCl	$Pm\bar{3}n$	18189	3735	60.0	$(-22.0, 16.5)$	-2.050
InNi	P6/mmm	59434	55518	15.0	$(10.0, \infty)$	-0.115
CdTe	P3121	150941	19425	3.5	(3.3, 3.9)	-0.461
Ga_7Te_{10}	$R\bar{3}mH$	400668	24628	5.5	$(-\infty, 3.1)$	-0.325
FeS_2	$Pa\bar{3}$	10422	2742	3.9	$(2.7, \infty)$	-0.694
Na_4Si_{23}	C12/c1	50796	646594	12.0	(4.0, 12.5)	-0.014
MgH_2	Pbca	155811	19965	13.9	(6.9, 9.4)	-0.140
MgH_2	Pbc21	155811	19965	13.9	(6.9, 9.4)	-0.140
BaO_2	P63/mmc	180398	22213	13.9	$(45.8, \infty)$	-1.980
KCl	$Pm\bar{3}m$	31232	5995	2.1	$(3.2, \infty)$	-2.065
BaO_2	P4/nmmS	180398	22213	18.8	$(45.8, \infty)$	-1.980
Ti_3O_5	P121/c1	20361	3884	34.0	$(-5.5, -2.2)$	-3.164
Tb_2Co_{17}	$Fd\bar{3}mS$	625364	29507	4.1	(11.1, 29.8)	-0.037
CaTe	$Pm\bar{3}m$	41959	7650	39.5	$(-48.3, -46.6)$	-1.626
CdS	PmmnZ	31074	5970	61.0	$(-\infty, 2.0)$	-0.790
FeF_3	$R\bar{3}cH$	29132	5688	1.5	(45.0, 195.1)	-2.339
FeS_2	P121/a1	10422	2742	7.5	$(2.7, \infty)$	-0.694
Y_2O_3	C12/m1	160219	20300	16.2	(6.7, 11.4)	-3.762
As_4S_5	C12/c1	16107	3319	0.8	$(-6.8, -0.6)$	-0.287

Table A.2 — Continued

Composition	Space group	ICSD (Collection Code/ID)	OQMD	p_{ICSD} (GPa)	$(p_{\min},\ p_{\max})$ (GPa)	ΔE_f (eV/atom)
Cr_3Te_4	Pbnm	626888	29597	50.5	(−6.1, −4.9)	−0.189
Sm_2O_3	P$\bar{3}$m1	40475	7283	5.6	(−106.6, 3.9)	−3.701
AgBr	Fm$\bar{3}$m	52246	646867	3.9	(−474.7, 4.0)	−0.493
MnS	Fm$\bar{3}$m	18007	3678	7.2	(0.2, 19.6)	−0.633
SrTe	Pm$\bar{3}$m	602931	656717	19.8	(−33.9, 10.1)	−1.747
$YbGa_2$	P63/mmc	107552	18667	9.9	(−17.5, 0.0)	−0.618
$FeCl_2$	P$\bar{3}$m1	4059	2349	0.6	(−13.2, 1.8)	−0.957
$CuCl_2$	Pa$\bar{3}$	26667	5084	9.2	(−10.7, 49.3)	−0.711
$CuCl_2$	Fm$\bar{3}$m	26667	5084	10.5	(−10.7, 49.3)	−0.711
Fe_3O_4	Pbcm	65338	11962	24.0	(−∞, −65.7)	−1.033
CrO_2	P42/mnm	9423	2509	7.7	(−∞, 96.6)	−2.110
FeS_2	P$\bar{6}$2c	10422	2742	0.0	(2.7, ∞)	−0.694
ReN	P63/mmc	181298	22305	31.0	(2858.7, ∞)	1.071
NaCl	Pm$\bar{3}$m	18189	3735	30.6	(−22.0, 16.5)	−2.050
ReN	P$\bar{6}$m2	181298	22305	13.0	(2858.7, ∞)	1.071
PbF_2	Pnma	14324	2917	1.6	(3.0, ∞)	−2.630
$CeSb_2$	P4/mmm	622083	29184	11.0	(−1.3, 26.8)	−0.820
Na_2S	P63/mmc	56024	10058	9.6	(−23.6, 3.2)	−1.247
Na_2S	Pnma	56024	10058	8.2	(−23.6, 3.2)	−1.247

Table A.2 — Continued

Composition	Space group	ICSD (Collection	OQMD Code/ID)	p_{ICSD} (GPa)	$(p_{\min},\ p_{\max})$ (GPa)	ΔE_f (eV/atom)
FeO	R$\bar{3}$cH	82236	14796	5.2	(−15.3, 92.7)	−1.519
PrN	P4/nmmS	45161	646232	48.0	(−5.1, 62.4)	−1.393
Y_2O_3	Ia$\bar{3}$	160219	20300	0.4	(6.7, 11.4)	−3.762
Fe_3O_4	Cmcm	65338	11962	10.0	$(-\infty, -65.7)$	−1.033
TlF	Pbcm	90993	15947	3.5	(−4.0, 3.0)	−1.922
$SiNi_3$	Pbnm	105343	17803	67.0	(−13.4, 3.7)	−0.465
BaSe	Pm$\bar{3}$m	43655	8173	8.2	(−29.5, 6.5)	−2.052
HgTe	P3221	31842	6117	2.6	(5.1, 11.7)	0.014
SiO_2	P42/mnm	100279	87467	15.0	(4.8, 5.3)	−3.039
LaN	P4/nmmS	162194	20522	39.9	$(33.5, \infty)$	−0.832
$CuBr_2$	Fm$\bar{3}$m	409450	25011	9.3	$(-\infty, 16.7)$	−0.511
NiP_2	Cmc21	22221	4131	5.8	(−1.0, 75.6)	−0.532
WO_3	P121/c1	50728	646576	0.6	(1.4, 2.9)	−2.439
CrO_2	R$\bar{3}$cH	9423	2509	1.9	$(-\infty, 96.6)$	−2.110
$PbCl_2$	P1121/a	15806	3228	25.4	(−20.8, 1.0)	−1.393
$CaSi_2$	R$\bar{3}$mH	32006	6152	5.0	(−0.2, 2.3)	−0.390
CoSn	P6/mmm	55564	9977	5.0	(−4.0, 25.4)	−0.162
In_3Pd_2	Pm$\bar{3}$m	59478	11131	0.0	(−40.8, 8.1)	−0.481
$SnCl_2$	Pnam	81979	14768	4.0	$(-\infty, \infty)$	−1.256

Table A.2 — Continued

Composition	Space group	ICSD (Collection Code/ID)	OQMD	p_{ICSD} (GPa)	$(p_{\min},\ p_{\max})$ (GPa)	ΔE_f (eV/atom)
PbS	Pnma	62192	11589	2.5	(−0.1, 15.3)	−0.685
$SnCl_2$	P1121/a	81979	14768	25.0	$(-\infty, \infty)$	−1.256
SiO_2	C12/c1	100279	87467	20.3	(4.8, 5.3)	−3.039
In_2Te_5	R$\bar{3}$mH	1385	1910	6.4	(1.0, 3.0)	−0.219
$FeCl_2$	R$\bar{3}$mH	4059	2349	0.1	(−13.2, 1.8)	−0.957
GaAs	Imm2	41674	7551	28.1	$(-\infty, 11.9)$	−0.353
GaAs	Pmm2	41674	7551	22.9	$(-\infty, 11.9)$	−0.353
Sb_2S_3	Pnma	22176	4112	3.2	(0.5, 36.3)	−0.429
NdSb	P4/mmm	92168	16038	21.0	$(11.0, \infty)$	−0.939
CsI_3	P4/mmm	27252	5285	55.0	(−3.3, 2.9)	−0.969
SiO_2	P121/c1	100279	87467	4.9	(4.8, 5.3)	−3.039
NiP_2	Pbca	22221	4131	1.3	(−1.0, 75.6)	−0.532
Gd_2O_3	C12/m1	40473	7281	3.5	(−127.9, 4.9)	−3.793
Pr_5Sb_3	P4/mmm	649315	30959	15.0	(−4.9, 49.1)	−0.976
Na_3N	Pm$\bar{3}$m	165991	21099	0.5	(3.9, 5.2)	0.148
Ca_3N_2	C12/m1	162794	20601	15.6	(9.4, 19.0)	−0.843
Ca_3N_2	P$\bar{3}$m1	162794	20601	23.3	(9.4, 19.0)	−0.843
As_4S_5	P121/n1	16107	3319	4.0	(−6.8, −0.6)	−0.287
InP	Fm$\bar{3}$m	41443	646115	10.8	(−65.2, 7.6)	−0.351

Table A.2 — Continued

Composition	Space group	ICSD (Collection Code/ID)	OQMD	p_{ICSD} (GPa)	$(p_{\mathrm{min}},\ p_{\mathrm{max}})$ (GPa)	ΔE_f (eV/atom)
BaO_2	I4/mmm	180398	22213	6.8	(45.8, ∞)	−1.980
BaO_2	Cmmm	180398	22213	49.4	(45.8, ∞)	−1.980
In_2Te_5	Pnnm	1385	1910	3.8	(1.0, 3.0)	−0.219
Zr_3O	P121/c1	27023	5214	1.5	(−16.2, 1.5)	−1.479
Cu_2O	C12/c1	261853	643065	2.2	(−56785.9, 10.9)	−0.684
As_2O_3	Fd$\bar{3}$mZ	27588	5363	0.5	(1.9, 33.4)	−1.436
Zr_3O	P42/nmcZ	27023	5214	21.4	(−16.2, 1.5)	−1.479
ZnO	P63mc	182360	117131	2.2	($-\infty$, 79.7)	−0.934
ZnTe	P31	67779	12375	11.5	($-\infty$, −14.0)	−0.466
SnO_2	Pa$\bar{3}$	181278	22299	49.0	(16.2, 37.9)	−1.851
SF_6	Im$\bar{3}$m	63360	11802	0.8	(10.5, 170.6)	−2.024
$YbAl_3$	Fd$\bar{3}$mS	58224	10618	2.0	(68.9, ∞)	−0.173
BaTe	Pm$\bar{3}$m	108097	91935	12.3	(4.8, ∞)	−1.623
BaI_2	P1121/a	79893	14426	40.0	(0.1, 1.9)	−1.910
Fe_3O_4	C12/m1	65338	11962	135.0	($-\infty$, −65.7)	−1.033
RbF	Pm$\bar{3}$m	53828	647290	1.2	(−14.8, 5.6)	−2.857
MgF_2	Pnnm	94282	16306	9.4	($-\infty$, 173.8)	−3.724
Co_9S_8	Pa$\bar{3}$	23929	4474	2.2	(−19.2, 24.6)	−0.626
$CuBr_2$	F$\bar{4}$3m	409450	25011	4.8	($-\infty$, 16.7)	−0.511

Table A.2 — Continued

Composition	Space group	ICSD (Collection Code/ID)	OQMD	p_{ICSD} (GPa)	$(p_{\min},\ p_{\max})$ (GPa)	ΔE_f (eV/atom)
SeO_2	Pbam	99464	16878	9.0	(5.6, 22.1)	−0.939
SeO_2	$Pm\bar{3}m$	99464	16878	5.0	(5.6, 22.1)	−0.939
InN	$Fm\bar{3}m$	25677	4825	18.2	(−0.5, 11.3)	−0.019
InSb	Pmm2	10022	2631	4.3	(6.2, 21.5)	0.006
Cr_3Te_4	P63/mmc	626888	29597	10.0	(−6.1, −4.9)	−0.189
Ag_3O_4	$Pn\bar{3}mZ$	202218	638387	0.0	(−6.7, 13.7)	−0.241
CdTe	$Fm\bar{3}m$	150941	19425	3.6	(3.3, 3.9)	−0.461
RuO_2	Pnnm	15071	2972	5.3	(−27.0, 10.6)	−1.266
GeS	$I\bar{4}2d$	1256	1877	5.2	(−0.2, 9.9)	−0.397
AgCl	$Fm\bar{3}m$	56545	53898	3.9	(−∞, 7.7)	−0.576
InSb	Cmcm	10022	2631	5.1	(6.2, 21.5)	0.006
Al_2O_3	$R\bar{3}cH$	161062	20393	6.1	(−∞, ∞)	−2.985
ZnF_2	Pbcn	9169	2482	5.0	(−3059.3, 4.1)	−2.567
SiO_2	P3221	100279	87467	2.3	(4.8, 5.3)	−3.039
BaF_2	Pnam	41651	7546	20.5	(−∞, 40.5)	−3.966
CaH_2	P63/mmc	157943	20130	22.9	(12.1, ∞)	−0.514
Ca_3N_2	$Ia\bar{3}$	162794	20601	0.4	(9.4, 19.0)	−0.843
Tl_6S	$Pm\bar{3}m$	30238	673576	35.5	(−∞, −20.9)	−0.101
InSb	$Fm\bar{3}m$	10022	2631	2.5	(6.2, 21.5)	0.006

Table A.2 — Continued

Composition	Space group	ICSD (Collection Code/ID)	OQMD	p_{ICSD} (GPa)	$(p_{\min},\ p_{\max})$ (GPa)	ΔE_f (eV/atom)
InSb	$Pm\bar{3}m$	10022	2631	35.3	(6.2, 21.5)	0.006
FeSe	Cmma	169305	110581	0.2	(52.7, ∞)	−0.036
MgF_2	$Pa\bar{3}$	94282	16306	16.5	($-\infty$, 173.8)	−3.724
Bi_4O_7	P121/c1	51778	646777	0.1	(−7.2, 11.2)	−1.399
$HgBr_2$	P3	151889	19529	5.4	(2.1, 34.0)	−0.610
HgSe	P3221	56211	10103	1.5	(4.3, 32.0)	−0.025
P_2Rh	P121/c1	42615	7836	3.6	(−7.5, 62.8)	−0.816
PrTe	$Pm\bar{3}m$	29402	5737	9.0	(−86.9, 4.7)	−1.443
Zr_3O	$Fm\bar{3}m$	27023	5214	37.3	(−16.2, 1.5)	−1.479
FeS_2	Pnma	10422	2742	6.3	(2.7, ∞)	−0.694
CeBi	$Pm\bar{3}m$	58765	10917	13.7	(−65.0, 7.8)	−0.988
Pr_4As_3	P4/mmm	44045	8267	45.0	($-\infty$, 28.0)	−1.264
CaF_2	Pnma	28730	5585	9.5	(−92.8, 8.1)	−4.219
$YbGa_2$	P6/mmm	107552	18667	27.0	(−17.5, 0.0)	−0.618
EuS	$I\bar{4}3d$	85577	15258	7.0	($-\infty$, −1.7)	−2.312
AgI	$Fm\bar{3}m$	28230	5478	1.5	(1.6, 3.3)	−0.335
AgI	P121/m1	28230	5478	11.3	(1.6, 3.3)	−0.335
MnS	Pnnm	18007	3678	16.8	(0.2, 19.6)	−0.633
GeSe	$P\bar{4}$	53906	647320	6.0	(1.3, 31.1)	−0.215

Table A.2 — Continued

Composition	Space group	ICSD (Collection Code/ID)	OQMD	p_{ICSD} (GPa)	$(p_{\mathrm{min}},\ p_{\mathrm{max}})$ (GPa)	ΔE_f (eV/atom)
RbTe	$Fm\bar{3}m$	73179	13118	0.5	(−8.3, 0.6)	−0.847
Ti_3O_5	P42/mnm	20361	3884	7.6	(−5.5, −2.2)	−3.164
RbBr	$Pm\bar{3}m$	18017	3686	0.5	(−14.2, 1.5)	−1.928
PrN	$Fm\bar{3}m$	45161	646232	30.0	(−5.1, 62.4)	−1.393
Ag_2S	P121/n1	98452	16747	4.7	(0.3, 20.9)	−0.152
In_2Te_5	$Pm\bar{3}m$	1385	1910	20.0	(1.0, 3.0)	−0.219
YbH_2	P63/mmc	56195	10098	24.5	(−22.2, −5.7)	−0.556
PbS	$Pm\bar{3}m$	62192	11589	25.0	(−0.1, 15.3)	−0.685
SbI_3	$R\bar{3}H$	30906	5946	1.6	(−13.2, −0.5)	−0.484
BiS	Pnma	79515	70066	1.0	(−∞, −1.8)	−0.206
SnS	$P\bar{3}m1$	79374	69959	1.2	(−∞, −0.1)	−0.584
SrF_2	$Fm\bar{3}m$	168801	21431	1.2	(−∞, 49.6)	−4.001
AlN	$Fm\bar{3}m$	163950	20758	20.0	(−∞, 148.5)	−1.394
MnS	$Pa\bar{3}$	18007	3678	4.0	(0.2, 19.6)	−0.633
AgI	$Pm\bar{3}m$	28230	5478	0.3	(1.6, 3.3)	−0.335
MnS	P121/c1	18007	3678	20.0	(0.2, 19.6)	−0.633
FeSe	P63/mmc	169305	110581	1.9	(52.7, ∞)	−0.036
AgCl	Cmcm	56545	53898	1.6	(−∞, 7.7)	−0.576
$BaCl_2$	P1121/a	2190	2091	37.0	(−∞, 39.9)	−2.774

Table A.2 — Continued

Composition	Space group	ICSD (Collection Code/ID)	OQMD	p_{ICSD} (GPa)	$(p_{\mathrm{min}},\ p_{\mathrm{max}})$ (GPa)	ΔE_f (eV/atom)
AgCl	P121/m1	56545	53898	10.0	$(-\infty,\ 7.7)$	−0.576
TbH_2	$Fm\bar{3}m$	56175	10089	0.0	(−33.6, 34.2)	−0.468
GeO_2	P3121	94241	16305	1.6	$(-\infty,\ \infty)$	−1.665
NiP_2	Pnma	22221	4131	16.3	(−1.0, 75.6)	−0.532
Pb_2O	Pbam	28838	5612	13.3	$(-\infty,\ -19.2)$	−0.546
MoO_3	Pbnm	30258	5790	0.1	(12.7, 26.8)	−2.129
CeP_2	$Pm\bar{3}m$	108172	18775	19.0	(0.1, 47.6)	−1.177
GeO_2	Pnnm	9162	2476	29.0	(−0.0, 27.9)	−1.889
Re_3P_4	Pnma	43210	8007	37.0	(−12.3, 84.4)	−0.511
GeO_2	$Pa\bar{3}$	94241	16305	108.0	$(-\infty,\ \infty)$	−1.665
AgI	P4/nS	28230	5478	30.0	(1.6, 3.3)	−0.335
Gd_2O_3	$P\bar{3}m1$	160211	20292	14.7	$(-\infty,\ 71.3)$	−3.742
TiF_3	$R\bar{3}cH$	52160	50979	2.4	(3.2, 159.7)	−3.554
H_2S	I41/acdS	168207	21357	14.0	(5.4, 28.4)	−0.221
Al_2O_3	Pbcn	161062	20393	113.0	$(-\infty,\ \infty)$	−2.985
AlAs	P63/mmc	56967	10263	18.0	(8.0, 234.1)	−0.242
H_2S	P1c1	168207	21357	11.4	(5.4, 28.4)	−0.221
InS	Pmnn	15931	3269	3.6	(−36.9, 3.0)	−0.653
ErN	$Fm\bar{3}m$	167759	21306	5.4	$(121.8,\ \infty)$	−0.929

Table A.2 — Continued

Composition	Space group	ICSD (Collection Code/ID)	OQMD	p_{ICSD} (GPa)	$(p_{\mathrm{min}},\ p_{\mathrm{max}})$ (GPa)	ΔE_f (eV/atom)
SrF_2	Pnam	168801	21431	15.2	$(-\infty,\ 49.6)$	−4.001
La_5Sb_3	P4/mmm	641888	30667	11.0	(−3.2, 0.6)	−1.010
SF_6	C12/m1	63360	11802	3.8	(10.5, 170.6)	−2.024
$FeAs_2$	Pna21	41724	7566	0.0	$(-\infty,\ 33.4)$	−0.269
CuS	P63/mmc	63328	11799	1.8	(−17.5, 9.1)	−0.344
FeSb	Pnnm	53535	647220	5.3	$(8.5,\ \infty)$	0.009
HgI_2	P42/nmcZ	18126	3712	1.1	(−0.5, 0.4)	−0.400
Cu_3N	$Pm\bar{3}m$	25675	4824	0.9	$(-\infty,\ -38.7)$	0.237
K_2S	Pmma	412535	25623	4.4	$(2.8,\ \infty)$	−1.076
CdS	P63mc	620322	29026	2.3	(4.2, 101.5)	−0.658
Na_3N	Pnma	165991	21099	3.9	(3.9, 5.2)	0.148
Na_3N	P63cm	165991	21099	16.6	(3.9, 5.2)	0.148
RbI	$Pm\bar{3}m$	61559	57018	0.5	$(-\infty,\ 10.5)$	−1.551
Na_3N	$Fm\bar{3}m$	165991	21099	35.6	(3.9, 5.2)	0.148
Na_3N	$R\bar{3}cH$	165991	21099	26.1	(3.9, 5.2)	0.148
La_2C_3	$I\bar{4}3d$	26588	5052	5.0	(−1.9, 398.7)	−0.170
$CaSi_2$	I41/amdZ	32006	6152	10.1	(−0.2, 2.3)	−0.390
Tl_5Se_3	$Pm\bar{3}m$	30376	5830	35.7	(−24.0, 21.8)	−0.331
BeO	P63mc	18147	3721	2.2	(−195.4, −11.9)	−2.950

Table A.2 — Continued

Composition	Space group	ICSD (Collection Code/ID)	OQMD	p_{ICSD} (GPa)	$(p_{\min},\ p_{\max})$ (GPa)	ΔE_f (eV/atom)
SnSb	$Pm\bar{3}m$	52302	646878	19.0	(8.5, 74.2)	0.081
SiO_2	$Pa\bar{3}$	100279	87467	268.0	(4.8, 5.3)	−3.039
Fe_3O_4	P121/n1	65338	11962	54.3	$(-\infty, -65.7)$	−1.033
Ag_2S	P212121	98452	16747	6.9	(0.3, 20.9)	−0.152
CaTe	Pbnm	41959	7650	24.7	(−48.3, −46.6)	−1.626
SrH_2	P63/mmc	33160	6231	95.0	$(7.0, \infty)$	−0.495
HgSe	$Fm\bar{3}m$	56211	10103	21.0	(4.3, 32.0)	−0.025
$CoAs_3$	Pnam	9188	2483	2.8	(−16.6, 17.9)	−0.308
$PbCl_2$	Pnam	15806	3228	11.0	(−20.8, 1.0)	−1.393
AlAs	$F\bar{4}3m$	56967	10263	11.0	(8.0, 234.1)	−0.242
$TmTe_2$	P63/mmc	653106	31101	15.0	(−25.4, 41.7)	−1.074
GeO_2	Pbcn	94241	16305	70.7	$(-\infty, \infty)$	−1.665
MgH_2	Pnma	155811	19965	21.9	(6.9, 9.4)	−0.140
CsF	$Pm\bar{3}m$	44288	8354	4.8	(−11.7, 5.7)	−2.833
MnF_4	P42/mnm	62068	11563	1.0	(−168.2, 248.9)	−2.496
SrS	I4/mcm	164144	20780	2.0	$(15.4, \infty)$	−1.996
PbS	$Fm\bar{3}m$	62192	11589	2.1	(−0.1, 15.3)	−0.685
NaCl	$Fm\bar{3}m$	18189	3735	3.4	(−22.0, 16.5)	−2.050

A.3. Novel high-pressure phases in ambient-immiscible systems

A.3.1. As–Pb

Predictions based on OQMD prototypes				
Composition	OQMD	ΔE_f (eV/atom)	$p_{\min}$ (GPa)	$p_{\max}$ (GPa)
As_3Pb	314710	0.392	88.837	∞
$AsPb_3$	312325	0.204	38.082	∞
$AsPb_3$	349738	0.134	9.447	11.080
$AsPb_3$	301783	0.159	11.081	38.083

MHM predicted structure for AsPb at 50 GPa, SPG: Imma (74)			
Unit cell dimensions	a = 3.23238	b = 4.69989	c = 9.20304
	$\alpha = 90.00000$	$\beta = 90.00000$	$\gamma = 90.00000$
Atom	x	y	z
As(1)	0.00000	0.25000	0.64824
Pb(1)	0.00000	0.25000	0.10942

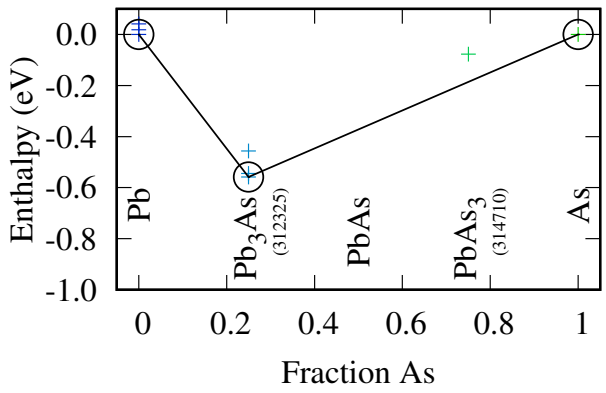

Figure A.1. LAE-predicted convex hull for the As–Pb system at 50 GPa.

A.3.2. Hg–In

Predictions based on OQMD prototypes				
Composition	OQMD	ΔE_f (eV/atom)	$p_{\min}$ (GPa)	$p_{\max}$ (GPa)
$InHg_3$	300764	0.111	7.101	∞
In_3Hg	346334	0.050	3.650	6.289
In_3Hg	299889	0.078	6.288	∞

MHM predicted structure for HgIn, SPG: P4/mmm (123)			
Unit cell dimensions	a = 2.92133	b = 2.92133	c = 4.02749
	$\alpha = 90.00000$	$\beta = 90.00000$	$\gamma = 90.00000$
Atom	x	y	z
Hg(1)	0.50000	0.50000	0.50000
In(1)	0.00000	0.00000	0.00000

MHM predicted structure for $HgIn_3$, SPG: Immm (71)			
Unit cell dimensions	a = 4.33198	b = 4.08785	c = 7.88558
	$\alpha = 90.00000$	$\beta = 90.00000$	$\gamma = 90.00000$
Atom	x	y	z
Hg(1)	0.50000	0.00000	0.50000
In(1)	0.00000	0.00000	0.24986
In(2)	0.00000	0.50000	0.50000

MHM predicted structure for Hg_3In, SPG: I4/mmm (139)			
Unit cell dimensions	a = 4.07722	b = 4.07722	c = 8.18543
	$\alpha = 90.00000$	$\beta = 90.00000$	$\gamma = 90.00000$
Atom	x	y	z
Hg(1)	0.00000	0.00000	0.50000
Hg(2)	0.00000	0.50000	0.25000
In(1)	0.00000	0.00000	0.00000

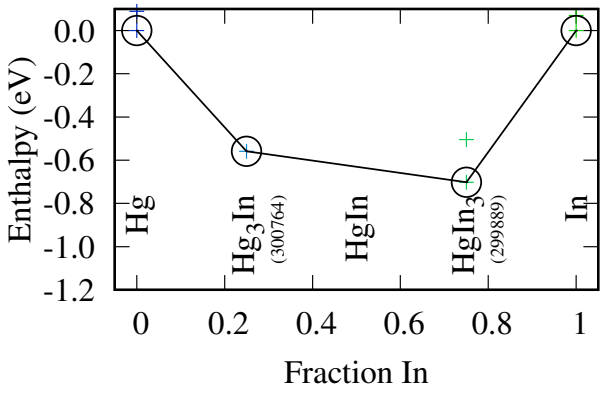

Figure A.2. LAE-predicted convex hull for the Hg–In system at 50 GPa.

A.3.3. Hg–Sn

Predictions based on OQMD prototypes				
Composition	OQMD	ΔE_f (eV/atom)	$p_{\min}$ (GPa)	$p_{\max}$ (GPa)
Sn_3Hg	348253	0.022	0.484	2.997
Sn_3Hg	302858	0.052	2.997	147.627
Sn_3Hg	313400	0.056	147.627	∞
$SnHg_3$	302683	0.147	11.475	∞

MHM predicted structure for Hg_3Sn, SPG: C2/m (12)			
Unit cell dimensions	a = 9.64991	b = 2.89304	c = 5.05054
	α = 90.00000	β = 100.17130	γ = 90.00000
Atom	x	y	z
Hg(1)	0.75186	0.00000	0.25184
Hg(2)	0.00000	0.50000	0.50000
Sn(1)	0.00000	0.00000	0.00000

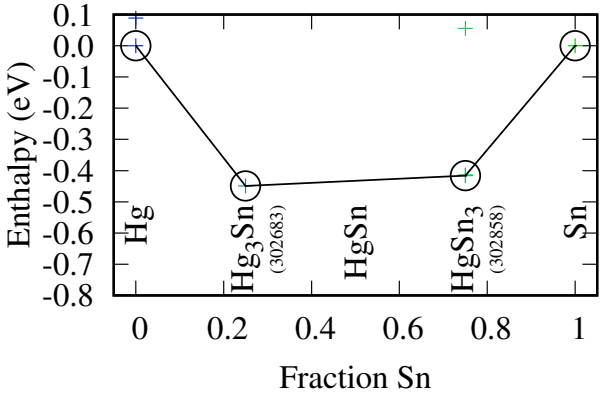

Figure A.3. LAE-predicted convex hull for the Hg–Sn system at 50 GPa.

A.3.4. Fe–In

Predictions based on OQMD prototypes				
Composition	OQMD	ΔE_f (eV/atom)	$p_{\min}$ (GPa)	$p_{\max}$ (GPa)
$InFe_3$	324752	0.099	13.470	23.610
In_3Fe	303668	0.320	138.215	∞
InFe	306831	0.325	20.819	∞

MHM predicted structure for In_3Fe, SPG: Cmcm (63)			
Unit cell dimensions	a = 3.13742	b = 9.40627	c = 8.06256
	α = 90.00000	β = 90.00000	γ = 90.00000
Atom	x	y	z
In(1)	0.00000	0.54711	0.25000
In(2)	0.00000	0.13794	0.54442
Fe(1)	0.00000	0.25703	0.25000

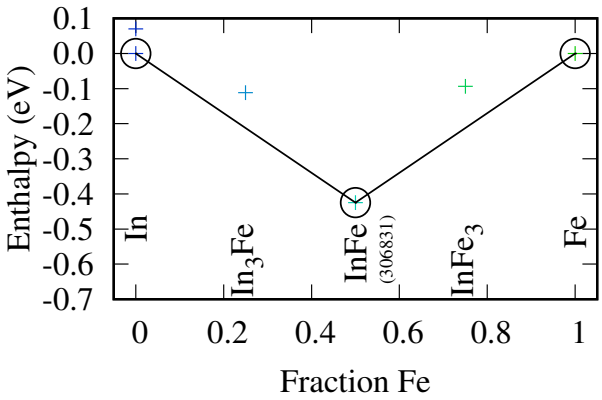

Figure A.4. LAE-predicted convex hull for the In–Fe system at 50 GPa.

A.3.5. Re–Br

Predictions based on OQMD prototypes				
Composition	OQMD	ΔE_f (eV/atom)	$p_{\min}$ (GPa)	$p_{\max}$ (GPa)
$ReBr_3$	298763	1.141	39.970	∞
$ReBr_3$	325023	0.186	9.585	30.275
ReBr	305074	1.615	168.752	∞
ReBr	326158	1.215	45.999	168.774

MHM predicted structure for $ReBr_3$, SPG: C2/m (12)			
Unit cell dimensions	a = 4.94513	b = 9.91552	c = 5.36688
	$\alpha = 90.00000$	$\beta = 102.03314$	$\gamma = 90.00000$
Atom	x	y	z
Re(1)	0.00000	0.69144	0.00000
Br(1)	0.75167	0.00000	0.21595
Br(2)	0.23444	0.14835	0.31348

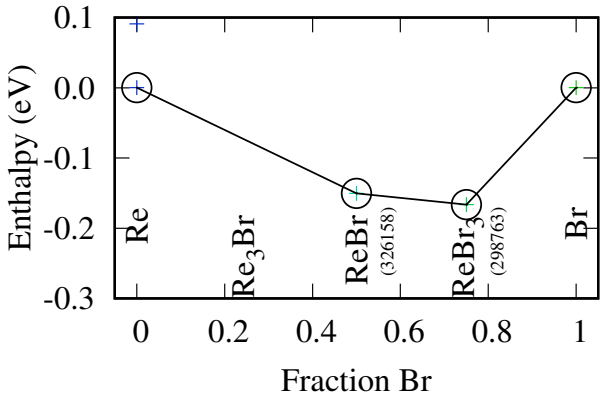

Figure A.5. LAE-predicted convex hull for the Re–Br system at 50 GPa.

A.3.6. Re–Ga

Predictions based on OQMD prototypes				
Composition	OQMD	ΔE_f (eV/atom)	$p_{\min}$ (GPa)	$p_{\max}$ (GPa)
GaRe	327316	0.212	59.768	89.115
GaRe	306232	0.345	89.119	∞
$GaRe_3$	322144	0.205	204.998	∞
Ga_3Re	301060	0.186	11.336	4853.546
Ga_3Re	320470	0.253	4853.549	∞

MHM predicted structure for ReGa, SPG: P4/nmm (129)			
Unit cell dimensions	a = 2.96485	b = 2.96485	c = 5.69732
	$\alpha = 90.00000$	$\beta = 90.00000$	$\gamma = 90.00000$
Atom	x	y	z
Re(1)	0.25000	0.25000	0.86337
Ga(1)	0.25000	0.25000	0.38123

MHM predicted structure for $ReGa_3$, SPG: C2/m (12)			
Unit cell dimensions	a = 8.58183	b = 3.17807	c = 7.81840
	$\alpha = 90.00000$	$\beta = 111.78545$	$\gamma = 90.00000$
Atom	x	y	z
Re(1)	0.63290	0.00000	0.70306
Ga(1)	0.87077	0.00000	0.04567
Ga(2)	0.87745	0.00000	0.55519
Ga(3)	0.37776	0.00000	0.79502

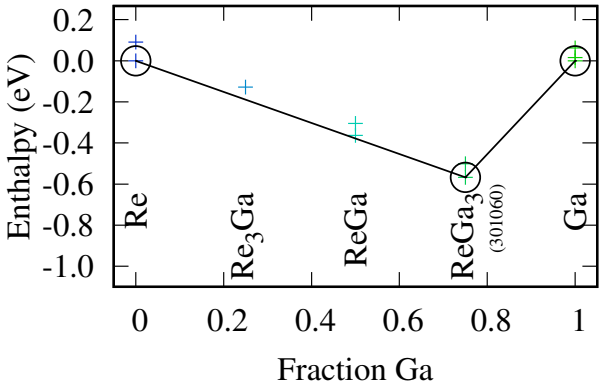

Figure A.6. LAE-predicted convex hull for the Re–Ga system at 50 GPa.

A.3.7. Re–Sn

Predictions based on OQMD prototypes				
Composition	OQMD	ΔE_f (eV/atom)	$p_{\min}$ (GPa)	$p_{\max}$ (GPa)
$ReSn_3$	302922	0.718	45.996	∞
ReSn	304519	0.604	27.398	∞
Re_3Sn	308129	0.546	296.697	∞

MHM predicted structure for $ReSn_3$, SPG: Cmcm (63)			
Unit cell dimensions	a = 3.24850	b = 9.59949	c = 8.56114
	α = 90.00000	β = 90.00000	γ = 90.00000
Atom	x	y	z
Re(1)	0.00000	0.23106	0.25000
Sn(1)	0.00000	0.36635	-0.03189
Sn(2)	0.00000	-0.05211	0.25000

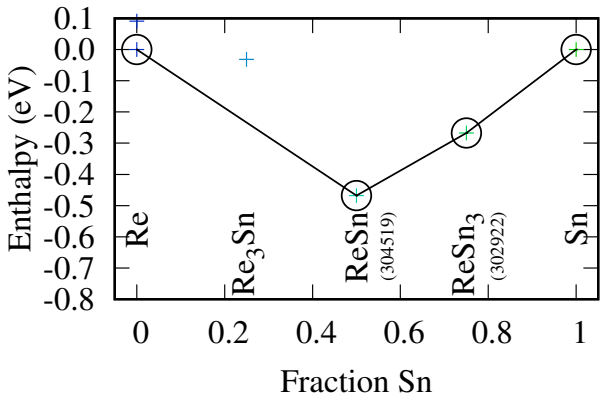

Figure A.7. LAE-predicted convex hull for the Re–Sn system at 50 GPa.

A.3.8. Al–Si

Predictions based on OQMD prototypes				
Composition	OQMD	ΔE_f (eV/atom)	$p_{\min}$ (GPa)	$p_{\max}$ (GPa)
$AlSi_3$	347070	0.367	37.878	∞
AlSi	337976	0.220	29.173	93.707
Al_3Si	301500	0.109	20.866	∞

MHM predicted structure for SiAl, SPG: C2/m (12)			
Unit cell dimensions	a = 16.37599	b = 2.54607	c = 4.31791
	$\alpha = 90.00000$	$\beta = 93.89323$	$\gamma = 90.00000$
Atom	x	y	z
Si(1)	0.06132	0.00000	0.68611
Si(2)	0.18753	0.00000	0.06466
Al(1)	0.56327	0.00000	0.18609
Al(2)	0.31308	0.00000	0.42989

MHM predicted structure for $SiAl_3$, SPG: I4/mmm (139)			
Unit cell dimensions	a = 3.57383	b = 3.57383	c = 14.49798
	$\alpha = 90.00000$	$\beta = 90.00000$	$\gamma = 90.00000$
Atom	x	y	z
Si(1)	0.00000	0.00000	0.87620
Al(1)	0.00000	0.50000	0.00000
Al(2)	0.00000	0.00000	0.62524
Al(3)	0.00000	0.50000	0.25000

MHM predicted structure for Si_3Al, SPG: P4/nmm (129)			
Unit cell dimensions	a = 2.48017	b = 2.48017	c = 14.13708
	$\alpha = 90.00000$	$\beta = 90.00000$	$\gamma = 90.00000$
Atom	x	y	z
Si(1)	0.25000	0.25000	0.06179
Si(2)	0.25000	0.25000	0.56209
Si(3)	0.25000	0.25000	0.81583
Al(1)	0.25000	0.25000	0.31124

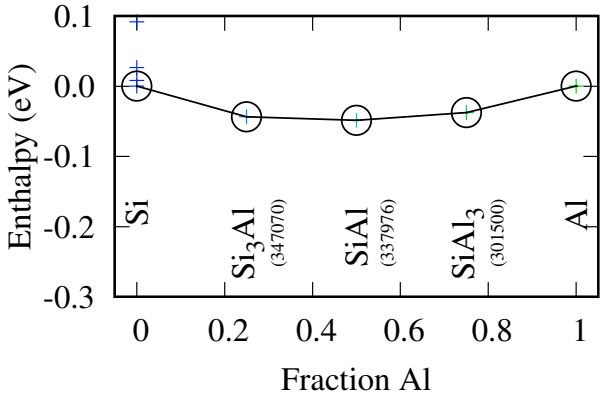

Figure A.8. LAE-predicted convex hull for the Si–Al system at 50 GPa.

A.3.9. Sn–Bi

Predictions based on OQMD prototypes				
Composition	OQMD	ΔE_f (eV/atom)	$p_{\min}$ (GPa)	$p_{\max}$ (GPa)
Sn_3Bi	313053	0.049	4.545	∞
Sn_3Bi	323095	0.010	0.194	4.545
$SnBi_3$	312553	0.113	5.546	∞

MHM predicted structure for Sn_3Bi, SPG: C2/c (15)			
Unit cell dimensions	a = 6.65344	b = 13.55695	c = 4.83451
	α = 90.00000	β = 134.82124	γ = 90.00000
Atom	x	y	z
Sn(1)	0.00000	0.56575	0.25000
Sn(2)	0.00000	0.80932	0.25000
Sn(3)	0.00000	0.31025	0.25000
Bi(1)	0.00000	0.06463	0.25000

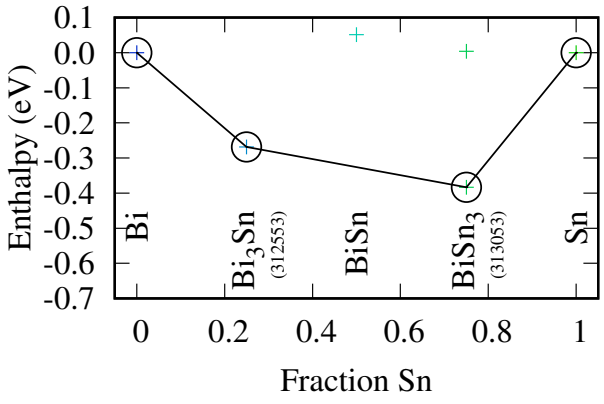

Figure A.9. LAE-predicted convex hull for the Sn–Bi system at 50 GPa.

A.3.10. Zn–Ga

Predictions based on OQMD prototypes				
Composition	OQMD	ΔE_f (eV/atom)	p_{min} (GPa)	p_{max} (GPa)
$ZnGa_3$	343588	0.056	10.162	∞
Zn_3Ga	346113	0.046	15.079	∞

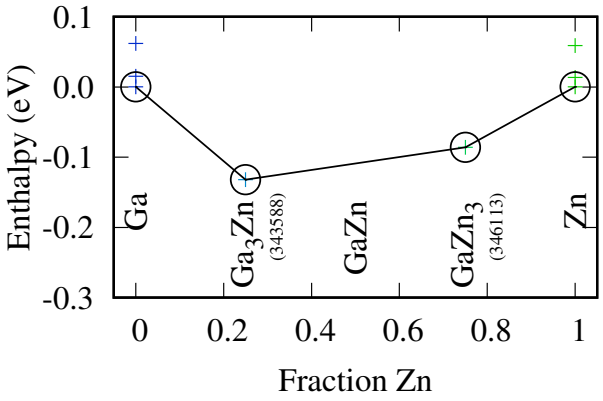

Figure A.10. LAE-predicted convex hull for the Ga–Zn system at 50 GPa.

APPENDIX B

List of Publications

(* = equal contribution)

1. **VI Hegde***, M Aykol*, S Kirklin, C Wolverton, "The Phase Stability Network of All Inorganic Materials", preprint at https://arxiv.org/abs/1808.10869. (Under revision in *Sci. Adv.*)

2. M Amsler*, **VI Hegde***, SD Jacobsen, and C Wolverton, "Exploring the High-Pressure Materials Genome", *Phys. Rev. X* **8**, 041021 (2018). doi:10.1103/PhysRevX.8.041021

3. S Kim*, M Aykol*, **VI Hegde***, Z Lu, S Kirklin, J Croy, MM Thackeray, and C Wolverton, "Materials Design of High-capacity Li-rich Layered-Oxide Electrodes: Li_2MnO_3 and Beyond", *Energy Environ. Sci.* **10**, 2201–2211 (2017). doi:10.1039/C7EE01782K

4. **VI Hegde** and C Wolverton, "Computational discovery of novel ternary oxyfluorides through high-throughput prototype decoration and structure prediction", In preparation.

5. **VI Hegde** and C Wolverton, "Large-scale computational materials discovery using high-throughput prototype decoration: a retrospective analysis of the OQMD", In preparation.

6. **VI Hegde**, Y Xia, and C Wolverton, "qmpy-phon: a Python library for high-throughput compressed sensing lattice dynamics calculations", In preparation.

7. N Naghibolashrafi, **VI Hegde**, S Keshavarz, SS Naghavi, J Ma, A Gupta, P LeClair, WH Butler, C Wolverton, K Munira, D Mazumdar, and AW Ghosh, "Structural and Magnetic Analyses of the $Fe_xCo_{1-x}TiSb$ Alloy System: $Fe_{0.5}Co_{0.5}TiSb$ as a Prototypical Half-Heusler Compound", (Under review in *J. Alloys Compd*)

8. M Aykol, **VI Hegde**, S Suram, L Hung, P Herring, C Wolverton, and JS Hummelshøj, "Network Analysis of Synthesizable Materials Discovery", *Nat. Commun.* **10**, 2018 (2019). doi:10.1038/s41467-019-10030-5

9. Z Yao, **VI Hegde**, A Aspuru-Guzik, and C Wolverton, "Discovery of Calcium-Metal Alloy for Reversible Ca-Ion Batteries", *Adv. Energy Mater.* **9**, 1802994 (2019). doi:10.1002/aenm.201802994

10. M Amsler, L Ward, **VI Hegde**, MG Goesten, Y Xia, and C Wolverton, "Ternary mixed-anion semiconductors with tunable band gaps from machine-learning and crystal structure prediction", *Phys. Rev. Mater.* **3**, 035404 (2019). doi:10.1103/PhysRevMaterials.3.035404

11. M Aykol, S Kim, **VI Hegde**, S Kirklin, and C Wolverton, "Computational Evaluation of New Lithium-3 Garnets for Lithium-ion Battery Applications as Anodes, Cathodes, and Solid-State Electrolytes", *Phys. Rev. Mater.* **3**, 025402 (2019). doi:10.1103/PhysRevMaterials.3.025402

12. DH Snydacker, **VI Hegde**, M Aykol, and C Wolverton, "Computational Discovery of Li-M-O Ion Exchange Materials for Lithium Extraction from Brines", *Chem. Mater.* **30**, 6961–6968 (2018). doi:10.1021/acs.chemmater.7b03509

13. D Wang, M Amsler, **VI Hegde**, JE Saal, A Issa, B-C Zhou, X Zeng, and C Wolverton, "Crystal Structure, Energetics, and Phase Stability of Strengthening Precipitates in Mg Alloys: A First-Principles Study", *Acta. Mater.* **158**, 65–78 (2018). doi:10.1016/j.actamat.2018.07.041

14. J He, SS Naghavi, **VI Hegde**, M Amsler, and C Wolverton, "Designing and Discovering a New Family of Semiconducting Quaternary Heusler Compounds Based on the 18-Electron Rule", *Chem. Mater.* **30**, 4978–4985 (2018). doi:10.1021/acs.chemmater.8b01096

15. S Anand, K Xia, **VI Hegde**, U Aydemir, V Kocevski, T Zhu, C Wolverton, and GJ Snyder, "A Valence Balanced Rule for Discovery of 18-Electron Half-Heuslers with Defects", *Energy Environ. Sci.* **11**, 1480–1488 (2018). doi:10.1039/C8EE00306H

16. S Kim, **VI Hegde**, Z Yao, Z Lu, M Amsler, J He, S Hao, JR Croy, E Lee, MM Thackeray, C Wolverton, "First-Principles Study of Lithium Cobalt Spinel Oxides: Correlating Structure and Electrochemistry", *ACS Appl. Mater. Interfaces* **10**, 13479–13490 (2018). doi:10.1021/acsami.8b00394

17. Z Yao, S Kim, J He, **VI Hegde**, and C Wolverton, "Interplay of Cation and Anion Redox in $Li_4Mn_2O_5$ Cathode Material and Prediction of Improved $Li_4(Mn,M)_2O_5$ Electrodes for Li-ion Batteries", *Sci. Adv.* **4**, eaao6754 (2018). doi:10.1126/sciadv.aao6754

18. DH Snydacker, **VI Hegde**, and C Wolverton, "Electrochemically Stable Coating Materials for Li, Na, Mg Anodes in Durable High Energy Batteries" *J. Electrochem. Soc.* **164**, A3582–A3589 (2017). doi:10.1149/2.0371714jes

19. L Ward, R Liu, A Krishna, **VI Hegde**, A Agrawal, A Choudhary, and C Wolverton, "Including Crystal Structure Attributes in Machine Learning Models via Voronoi Tessellations", *Phys. Rev. B* **96**, 024104 (2017). doi:10.1103/PhysRevB.96.024104

20. SS Naghavi, **VI Hegde**, and C Wolverton, "Diffusion Coefficients of Transition Metals in Fcc Cobalt", *Acta Mater.* **132**, 467–478 (2017). doi:10.1016/j.actamat.2017.04.060

21. J Ma, **VI Hegde**, K Munira, Y Xie, S Keshavarz, DT Mildebrath, C Wolverton, AW Ghosh, WH Butler, "Computational Investigation of Half-Heusler Compounds for Spintronics Applications", (Editors' Suggestion) *Phys. Rev. B* **95**, 024411 (2017). doi:10.1103/PhysRevB.95.024411

22. SS Naghavi, **VI Hegde**, A Saboo, and C Wolverton, "Energetics of Co Alloys and Solute-Vacancy Binding in Fcc Cobalt: A First-Principles Database", *Acta Mater.* **124**, 1–8 (2016). doi:10.1016/j.actamat.2016.10.065

23. M Aykol, S Kim, **VI Hegde**, D Snydacker, Z Lu, S Hao, S Kirklin, D Morgan, and C Wolverton, "High-Throughput Computational Design of Cathode Coatings for Li-ion Batteries", *Nat. Commun.* **7**, 13779 (2016). doi:10.1038/ncomms13779

24. J He, M Amsler, Y Xia, SS Naghavi, **VI Hegde**, S Hao, S Goedecker, V Ozoliņš, and C Wolverton, "Ultralow Thermal Conductivity in Full-Heusler Semiconductors", (Journal Cover) *Phys. Rev. Lett.* **117**, 046602 (2016). doi:10.1103/PhysRevLett.117.046602

25. AA Emery, JE Saal, S Kirklin, **VI Hegde**, and C Wolverton, "High-Throughput Screening of Perovskites for Thermochemical Water Splitting Applications", (Journal Cover) *Chem. Mater.* **28**, 5621–5634 (2016). doi:10.1021/acs.chemmater.6b01182

26. N Naghibolashrafi, S Keshavarz, **VI Hegde**, A Gupta, WH Butler, J Romero, K Munira, P LeClair, D Mazumdar, J Ma, AW Ghosh, and C Wolverton, "Synthesis and Characterization of Fe-Ti-Sb Intermetallics: Discovery of a New Slater-Pauling Phase", *Phys. Rev. B* **93**, 104424 (2016). doi:10.1103/PhysRevB.93.104424

27. S Kirklin, JE Saal, **VI Hegde**, C Wolverton, "High-Throughput Computational Search for Strengthening Precipitates in Alloys", *Acta Mater.* **102**, 125–135 (2016). doi:10.1016/j.actamat.2015.09.016

APPENDIX C

Patents and Invention Disclosures

1. M Aykol, S Kim, S Hao, Z Lu, **VI Hegde**, DH Snydacker, S Kirklin, C Wolverton, "Protective Cathode Coatings for Lithium-Ion Batteries", U.S. Patent No. US 10,374,262 B2, Date of Patent: Aug. 6, 2019.

2. D Snydacker, **VI Hegde**, C Wolverton, "Method for Lithium Extraction via Ion Exchange", U.S. Patent No. US 10,322,950 B2, Date of Patent: Jun. 18, 2019.

3. **VI Hegde**, D Snydacker, C Wolverton, "Protective Anode Coatings for High Energy Batteries", U.S. Patent Appl. Pub. No. US 2019/0088991 A1, Pub. Date: Mar. 21, 2019.

4. C Wolverton, Z Yao, S Kim, J He, **VI Hegde**, "High-energy-density Cathode Materials for Secondary Lithium-ion Batteries", U.S. Provisional Patent No. 62/609,620, Filed Date: Dec. 22, 2017.

APPENDIX D

Software Frameworks and Implementations

*qmpy** Python, Django-based framework that powers the Open Quantum Materials Database (OQMD). github.com/wolverton-research-group/qmpy

kelpie Python package for server-side management of density functional calculations using Vienna Abinitio Simulation Package (VASP). github.com/hegdevinayi/kelpie

nve Fast convex hull and ground state phase stability analysis at non-ambient pressures, using QHull and PuLP libraries. github.com/wolverton-research-group/nve

htdefects Python framework for high-throughput density functional calculations of dilute mixing defects in materials. (Under development; to be open-sourced soon.)

icsd-queryer A Selenium wedriver based data retriever for the Inorganic Crystal Structure Database (ICSD). github.com/hegdevinayi/icsd-queryer

janaf-scraper A Selenium webdriver based data retriever for the NIST-JANAF Thermochemical Tables. github.com/hegdevinayi/janaf-scraper

*qmpy is a collaborative effort. I was the primary developer and maintainer of the package from late-2014 to mid-2019.