

DATA-DRIVEN DESCRIPTORS FOR THE THERMOCHEMISTRY OF INORGANIC
CRYSTALLINE SOLIDS

by

CHRISTOPHER JOSEPH BARTEL

B.S., Auburn University, 2014

M.S., University of Colorado, 2016

A thesis submitted to the
Faculty of the Graduate School of the
University of Colorado in partial fulfillment
of the requirement for the degree of
Doctor of Philosophy
Department of Chemical and Biological Engineering
2018

This thesis entitled:
Data-driven Descriptors for the Thermochemistry of Inorganic Crystalline Solids
written by Christopher Joseph Bartel
has been approved for the Department of Chemical and Biological Engineering

Charles B. Musgrave

Alan W. Weimer

Date_____

The final copy of this thesis has been examined by the signatories,
and we find that both the content and the form meet acceptable presentation standards
of scholarly work in the above-mentioned discipline.

Abstract

Bartel, Christopher Joseph (Ph.D., Chemical Engineering)

Data-driven descriptors for the thermochemistry of inorganic crystalline solids

Thesis directed by Professors Charles Musgrave and Alan Weimer

The advancement of society has been historically predicated on the discovery or invention of new materials and in particular, inorganic solid-state materials have had transcendent influence on society – stone for tools, steel for structures, silicon circuits and solar cells, and so forth. Synthesizing a new material is time-consuming, costly, and frustrating for those tasked with the job. The success of solid-state synthesis can be greatly improved if one knows the thermodynamic stability of the material they are trying to make and those they are trying to avoid. This dissertation addresses the prediction of thermodynamic stability for solid-state materials primarily using a branch of quantum chemistry called density functional theory (DFT) and statistical approaches that fall under the umbrella of data analytics and machine learning.

We partitioned the pathways of solid-state decomposition into three types to quantify the success of DFT-based approaches for predicting thermodynamic stability in a high-throughput manner. By comparing with experiment, we find that in general, DFT performs quite well. Importantly, when the decomposition pathway type is elucidated for all known inorganic crystals, we find that the type that DFT performs the best on is the most prevalent, supporting the efficacy of DFT-based stability predictions.

Still, DFT is computationally expensive and not always practical for a given problem. This motivates the use of data analytics to accelerate the prediction of thermodynamic stability using so-called “descriptors”. We applied the SISO (sure independence screening and sparsifying

operator) algorithm to identify a new tolerance factor (descriptor) for predicting the experimentally realized stability of perovskites, which are a class of inorganic solids having significant utility as solar absorbers, catalysts, and capacitors.

This new tolerance factor was applied to identify new double perovskite solar absorbers in the cesium-chloride chemical space. In doing so, we gain insights into the stability of these materials, point out some pitfalls of common high-throughput approaches, and reveal a number of potential all-inorganic solar absorbers which may become active components in high-efficiency solar cells.

Much of the computational materials field is resigned to studying temperature-independent thermodynamics because of the expense of including the effects of vibrational entropy in the solid-state. To address this problem, we again used SISSO, this time to identify a simple descriptor for the Gibbs energy of an arbitrary inorganic crystalline solid. We show how this descriptor can be used for rapid predictions of temperature-dependent stability and thermochemical equilibrium.

As a demonstration of the utility of the Gibbs energy descriptor, we used it to screen for active materials that might be able to mediate the conversion of air, water, and sunlight into ammonia using chemical looping. These results provide a detailed thermodynamic analysis of the involved reactions for this process, highlighting the challenging tradeoff between metal oxide and metal nitride stability that must be met for the process to succeed.

This work helps reveal the lack of exploration of metal nitride compounds relative to their oxide counterparts. We show that the space of ternary metal nitrides that have been synthesized has the potential to double based on DFT-based stability predictions. We also developed quantitative descriptors for the bonding in metal nitrides to help rationalize their stability and highlight opportunities for synthesizing new nitrides with interesting technological properties.

*To my parents
for their unwavering support*

Acknowledgments

It's difficult to comprehensively acknowledge all the people who made this thesis possible within this brief space, as there are so many.

I would first like to acknowledge my thesis advisors, Charles Musgrave and Al Weimer, who, throughout my graduate career, have been beacons of support, encouragement, and advice. I was drawn to CU because of Al and my experience working in his research group as an REU student the summer of my junior year. I'm greatly appreciative of Al's generosity in allowing me to drift away from the experimental work I signed up for upon joining his group and focus my research on computational materials problems, which I now plan to work on for the rest of my career. Charles has also demonstrated a great deal of generosity, encouraging me to work on the problems I find most interesting and to seek out collaborations that expand these efforts. For my advisors' generosity, I am deeply grateful.

I would have never gone down the path of academia if Mark Byrne at Auburn University had not accepted my request to perform undergraduate research in his lab. Dr. Byrne was a great first introduction to how a mentor should be, taking a sincere interest in my experience in his lab and doing whatever he could to increase my chances of going to graduate school. Dr. Byrne encouraged me to apply for an REU which initially brought me to CU to work with Al and directly with Aaron Palumbo, who was another great mentor for my undergraduate research.

My good fortune with mentors has certainly extended to my graduate career. Chris Muhich was a senior graduate student in the Weimer/Musgrave groups when I started at CU and introduced me to the worlds of computational chemistry, maintaining an ALD reactor, and preparing for a life in academia. Chris provided a benchmark for how much one student could accomplish during

graduate school and serves as a constant inspiration. I was fortunate to work with Ann Deml while she was a postdoc at NREL as she got me started using *Python* and showed me the importance of making aesthetically pleasing figures. It's difficult to imagine any of the subsequent chapters without these two principal tools. Aaron Holder has been my *de facto* third advisor for most of my Ph.D. He has been everything I could want in a mentor and collaborator, providing technical hands-on advice, teaching fundamental concepts, brainstorming, and preparing me for life after graduate school. It is impossible to imagine this thesis coming together without his guidance.

Through Ann and Aaron, I was able to regularly interact with Vladan Stevanović and Stephan Lany at NREL, who helped me realize how much I did not yet know about solid-state physics and how much I wanted to study the topic. This collaboration with NREL opened doors to my involvement with the Center for Next Generation Materials by Design, fostering a collaboration with Elisabetta Arca and Andriy Zakutayev at NREL and Wenhao Sun and Gerd Ceder at Berkeley Lab, which comprises much of Chapter 7.

My luck with collaborations extends also outside the U.S. thanks to Chris Sutton from the Fritz-Haber Institute in Berlin who approached me after I gave a talk at AIChE in 2016. He facilitated a trip for me to work with him in Berlin to collaborate on the project that eventually became Chapter 3 of this thesis and helped orient me towards the worlds of data analytics for materials science; and also the world outside the U.S. This collaboration continues today along with another former FHI postdoc, Bryan Goldsmith. Both Chris and Bryan have been irreplaceable mentors during my graduate career and I consider myself truly fortunate that Chris attended my talk that day at AIChE in 2016. I would be remiss to not also gratefully acknowledge my other collaborators at the FHI – Runhai Ouyang who introduced me to the SISSO algorithm used in Chapters 3 and 5, Luca Ghiringhelli, who continues to inform my perspective on machine learning

and data science, and Matthias Scheffler for supporting now two trips from Colorado to Berlin to visit his department. Each these collaborators are co-authors on the work described in Chapter 3, an indispensable component of this thesis.

I’ve also had the privilege of trying to emulate my great mentors by advising five devoted undergraduate students – Austin Drake, Maila Kodas, Jack Rumptz, Matt Jankousky, and Tunkie Saunders. Austin put in countless hours helping me set up and test experimental equipment in the Weimer lab which I’m sure was not easy, given my lack of aptitude for experimental work. While on the subject, I must point out that I’m endlessly grateful to Hans Funke for putting up with my general incompetence in the lab and showing me the correct way to do the few things I managed to accomplish as an experimentalist. Also an experimentalist, Maila showed great independence in taking over the ceramic 3D printer that I no longer had time to work with as my interests drifted towards computation. Jack was my first computational undergrad and started working with me when I was a novice at both *Python* and machine learning, so I’m grateful to him for putting up with my umpteen missteps in both and for informing me about good coding practices. Matt had the unfortunate position of being the first person that I saddled with an independent project. In doing so, he was forced to suffer through the growing pains of me learning how to direct a computational chemistry research project – while the road has been bumpy, I’m optimistic that we’ve both learned a great deal and enjoyed the process. Tunkie joined me late in his undergraduate career, so I’m thankful to him for his ambition in learning what he could of *Python*, data science, and computational chemistry in such a short time. I’m appreciative that he so eagerly took on the role of being my guinea pig for testing out a range of my often not-so-practical ideas for machine learning and crystal structure prediction. While not an undergraduate, I had the privilege of mentoring graduate student, Becky O’Toole, when she started in the Weimer group

my 3rd year. Becky probably enabled this thesis as much as anyone by very quickly changing her role from helping me in the lab to making sure I never stepped foot in the lab again. Without her willingness and capability to take on my experimental work in addition to her other projects, I would not have been able to shift my focus to the work that comprises this thesis.

There are a number of people behind the scenes that facilitate a painless graduate experience. I'm especially grateful to Dom De Vangel, who promptly and effectively answers any and all questions related to the mechanics and paperwork of getting through the Ph.D. program.

In addition to my advisors, I'm appreciative of my Ph.D. committee – Will Medlin, Adam Holewinski, and Ronggui Yang. I had the pleasure of being taught by Will for kinetics and Ronggui for an essential introduction to solid-state physics and nanoscale energy transport. Will also provided an excellent example for how to instruct a course as I assisted him as a TA in my 3rd year. Adam has consistently come up with insightful questions during my committee meetings, providing a new perspective on my hypotheses and results.

This work would not be possible without government funding and I was fortunate to receive a graduate research fellowship from the National Science Foundation and GAANN fellowships from the Department of Education. Thanks is owed to Al, Charles, and Dr. Byrne who assisted me with these applications. This independent funding has helped enable me to work on this diverse array of projects, and for that I'm truly grateful.

My interest in science and engineering stems from the great teachers I've had for those subjects growing up. Ms. Disher and Mr. Traylor especially come to mind for introducing me to calculus and chemistry during high school. Without their captivating lectures and positive attitude about their courses, I'm not sure I would have chosen to study chemical engineering.

Finally, I cannot thank my family and friends enough, for making me believe I'm capable of accomplishing anything and for sharing in all the good times. Although she can't read, shout out to my dog, Keema, for embodying the idea of a man's best friend.

Contents

1	Introduction	1
1.1	Materials by design and density functional theory.....	1
1.2	Predicting thermodynamic stability	3
1.3	Descriptor-based property prediction.....	4
1.4	Thesis scope	5
2	The role of decomposition reactions in assessing first-principles predictions of solid stability.....	8
2.1	Abstract	8
2.2	Introduction	9
2.3	Results	11
2.3.1	Relevant reactions for determining the stability of compounds.....	11
2.3.2	Benchmarking performance on formation enthalpy.....	14
2.3.3	Optimizing elemental reference energies	15
2.3.4	Decomposition reaction analysis.....	17
2.4	Discussion	21
2.5	Methods.....	23
3	New tolerance factor to predict the stability of perovskite oxides and halides	25
3.1	Abstract	25
3.2	Introduction	25
3.3	Results	28
3.3.1	Finding an improved tolerance factor	28

3.3.2	Comparing to calculated perovskite stabilities.....	34
3.3.3	Extension to double perovskite oxides and halides.....	35
3.3.4	Compositional mapping of perovskite stability	38
3.4	Discussion	41
3.5	Methods.....	42
3.5.1	Radii assignment	42
3.5.2	Descriptor selection.....	43
3.5.3	Alternative features	45
3.5.4	Increasing dimensionality	45
3.5.5	Potential for overfitting	46
3.5.6	Alternative radii for more covalent compounds.....	47
3.5.7	Computer packages used.....	47
4	Computational screening for double perovskite solar absorbers: a case-study of cesium chlorides	48
4.1	Abstract	48
4.2	Introduction	48
4.3	Results	51
4.3.1	Stability analysis	51
4.3.2	Structural distortions	53
4.3.3	Candidate solar absorbers.....	58
4.4	Conclusions	59

4.5	Methods	60
5	Physical descriptor for the Gibbs energy of inorganic crystalline solids and temperature-dependent materials chemistry	61
5.1	Abstract	61
5.2	Introduction	61
5.3	Results	63
5.3.1	Trends in the Gibbs energies of compounds and elements	63
5.3.2	Descriptor identification and performance.....	67
5.3.3	Thermochemical reaction equilibria.....	72
5.3.4	Effects of temperature and composition on material stability.....	74
5.4	Discussion	79
5.5	Methods	81
5.5.1	Data retrieval	81
5.5.2	Feature retrieval.....	81
5.5.3	Descriptor identification.....	82
5.5.4	Descriptor sensitivity.....	83
5.5.5	Comparing to quasiharmonic approximation.....	84
5.5.6	Stability analysis	85
5.5.7	Structure considerations	85
5.5.8	Application of the descriptor.....	86
5.5.9	Extension to new materials	86

6	High-throughput equilibrium analysis of active materials for solar thermochemical ammonia synthesis	88
6.1	Abstract	88
6.2	Introduction	89
6.3	Results	91
6.3.1	Formation energies of redox pairs.....	91
6.3.2	Energetics of each reaction	93
6.3.3	Limiting reaction analysis	96
6.3.4	Equilibrium product distributions	99
6.3.5	Shifting equilibrium	101
6.3.6	Non-equilibrium considerations.....	103
6.4	Discussion	104
6.5	Methods.....	105
7	Descriptors for bonding in metal nitrides	108
7.1	Abstract	108
7.2	Introduction	109
7.3	Results	110
7.3.1	Covalent bonding and the kinetics of aluminum nitride hydrolysis.....	110
7.3.2	Redox-mediated stabilization in zinc molybdenum nitrides	113
7.3.3	New stable ternary nitrides.....	118
7.3.4	Experimental synthesis of predicted ternary nitrides	126
7.3.5	Metastable ternary nitrides	128

7.3.6	Thermochemical decomposition into competing binaries	130
7.3.7	Electronic origin of ternary nitride stability	133
7.4	Discussion	138
7.5	Methods	138
7.5.1	First-principles calculations of aluminum nitride hydrolysis kinetics	138
7.5.2	Data-mined structure prediction.....	140
7.5.3	Extracting chemical insights from first-principles calculations.....	141
8	Outlook	145
	References	147

Figures

Figure 2.1 Three unique decomposition reactions	12
Figure 2.2 Prevalence of reactions among known materials	13
Figure 2.3 Experimental vs. theoretical formation enthalpies (Type 1)	17
Figure 2.4 Experimental vs. theoretical decomposition enthalpies	20
Figure 3.1 Perovskite structure and composition.....	27
Figure 3.2 Assessing the performance of the improved tolerance factor, τ	30
Figure 3.3 Map of predicted double perovskite oxides and halides	38
Figure 3.4 The effects of ionic radii and oxidation states on the stability of single and double perovskite oxides and halides	41
Figure 4.1 Cesium chloride double perovskite space	50
Figure 4.2 Stability analysis conducted with the SCAN DFT functional.....	53
Figure 4.3 Structural distortion effects on stability and band gap	55
Figure 4.4 Octahedral distortions of optimized structures.....	57
Figure 4.5 Refined band gap and stability of top candidates.....	59
Figure 5.1 Contributions to the Gibbs energies of compounds	65
Figure 5.2 Descriptor performance	69
Figure 5.3 Benchmarking descriptor against ab initio methods	72
Figure 5.4 High-throughput reaction engineering	74
Figure 5.5 Survey of temperature-dependent (meta)stability	77
Figure 5.6 Composition-dependence of metastability	79
Figure 6.1 Solar thermochemical ammonia synthesis (STAS) reaction scheme	91
Figure 6.2 Temperature-dependent Gibbs energies	93

Figure 6.3 Reaction energies of the STAS cycle	96
Figure 6.4 Volcano-dependence of STAS energetics	99
Figure 6.5 Predicted yields for tungsten nitride hydrolysis	100
Figure 6.6 Reaction-wise equilibrium yields for all pairs.....	101
Figure 6.7 Increasing yields by shifting equilibrium	103
Figure 7.1 Rate-limiting step of AlN hydrolysis	113
Figure 7.2 Bonding analysis in Zn-Mo-N.....	116
Figure 7.3 Bonding for alternative cations in X-Mo-N.....	118
Figure 7.4 Stability map of ternary nitrides	124
Figure 7.5 Experimental confirmation of predicted nitrides.....	128
Figure 7.6 Thermodynamic perspective of ternary nitride stability	132
Figure 7.7 Rationalizing nitride stability through bonding analysis.....	135

1 Introduction

1.1 Materials by design and density functional theory

Throughout history, emergent technologies have been enabled by the discovery and application of new or improved materials. Materials-driven societal advancement dates back at least to the Stone Age, when early hominins made use of naturally occurring rock in the form of tools to obtain food and build shelter. A transformational change occurred with the dawn of metallurgy and the observation that naturally occurring elements, copper and tin, could be mixed (alloyed) at high temperature, to form bronze, which could be much harder than pure copper. This may be considered the first chemical design of materials. Many thousands of years later, computing was enabled by the invention of the integrated circuit based on the semiconductor, silicon.¹

Today, these same silicon chips have enabled a new paradigm for materials design using high-performance computing and quantum chemistry. Supercomputers hosted by governments and companies around the world can collectively perform 10^{18} floating-point operations per second, a million-fold increase in the last ~20 years.² A significant fraction of this supercomputing power has been made available to academic research groups, many of which are focused on the design of improved materials for a variety of applications. In particular, there are thousands of researchers relying on density functional theory (DFT) to calculate the electronic, chemical, and physical properties of materials at the atomistic scale.^{3, 4}

The density functional approximation is that a many-body system of interacting entities can be approximated as a non-interacting system. In the case of quantum chemistry as it's applied

within this work (DFT), the entities are electrons and the many-body problem that is being approximated is the time-independent Schrodinger's equation⁵:

$$\hat{H}\Psi = E\Psi \quad (1.1)$$

where the energy, E , of a quantum system is the expectation value of the Hamiltonian, $\hat{H}$, of the wave function, Ψ . All properties manifested by a system of nuclei and electrons (i.e., materials) – stability, physical structure, electronic structure, reactivity, etc. – are determined by Ψ . Of particular relevance is Ψ such that E is minimized, which provides the ground-state distribution of electron density.

In the commonly imposed Born-Oppenheimer approximation,⁶ the primary challenge in solving Schrodinger's equation is the computation of electron-electron interactions. For a simple example with two protons and two electrons (e.g., H_2), Schrodinger's equation in the Born-Oppenheimer approximation becomes:

$$\left[-\frac{1}{2} \sum_{i=1,2} \nabla_i^2 + \frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|} + \sum_{i=1,2} v_{ext}(\mathbf{r})\right] \Psi(\mathbf{r}_1, \mathbf{r}_2) = E\Psi(\mathbf{r}_1, \mathbf{r}_2) \quad (1.2)$$

where $\mathbf{r}$ is a position vector, $\mathbf{r}_i$ is the position of electron, i , and v_{ext} is an external potential imposed by the nuclei. **Equation 1.2** is complicated by the second term of the Hamiltonian – the electrostatic repulsion – which couples the two electrons and requires the solution of a partial differential equation in six coordinates.⁷ In contrast, The non-interacting Schrodinger's equation has only three coordinates and reduces to the following:

$$\left[-\frac{1}{2} \nabla^2 + v_s(\mathbf{r})\right] \varphi_i(\mathbf{r}) = \epsilon_i \varphi_i(\mathbf{r}) \quad (1.3)$$

where φ is a one-electron wavefunction or orbital and ϵ_i the eigenstate of that orbital. Critical to the degree to which **Equation 1.3** approximates **Equation 1.2** is the choice of v_s which is v_{ext} corrected by some function that helps mimic the electron-electron interactions. In DFT, v_s is a functional of the electron density, $n(\mathbf{r})$, and is designed to approximate the exchange and correlation energies of the electrons in the system. A number of these density functionals have been developed in the ~50 years since the original DFT ansatz was proposed, each of which have certain constraints and associated success and failure cases (e.g., materials or properties), have an inherent tradeoff of computational expense and accuracy, and frequently appear among the most cited papers in the history of science.⁸⁻¹¹ Owing to the availability of high-performance computing resources and the development of DFT codes designed to operate on them,¹²⁻¹⁴ DFT has become a principal tool for scientists of a range of disciplines to understand and design new materials.

1.2 Predicting thermodynamic stability

An essential criterion for the success of a proposed material is that there are some reasonable conditions where the proposed material exists – i.e., it is stable. DFT allows for the optimization of electronic and geometric structure for an initial configuration of ions and electrons. The resulting total energy, E , which has been variationally minimized with respect to the configuration of ions and electrons is a key result of the calculation.¹⁵ With recent advances in supercomputing power, it has become somewhat commonplace for a single research group to calculate E for 10^2 - 10^4 structures using typically the generalized gradient approximation (GGA)⁸ for the density functional. This capability has fortunately been commensurate with the advent of materials databases which openly provide these DFT-calculated structures and E (typically along with other readily obtainable properties).¹⁶⁻¹⁸ Many of these databases support the NOMAD

repository¹⁹ which accepts input and output files for any of these calculations and currently makes publicly available more than 50,000,000 DFT calculations of molecules and crystals.

Public access to this magnitude of DFT calculations has several advantages for the materials community, chief among them the ability to calculate E for a new structure and compare to E for all other structures that are known and could potentially be created from the structure of interest. A convenient and often necessary way to make this comparison is by the convex hull construction, where, the Gibbs free energy, G , (or some approximation, e.g., E) is generated with respect to some handle on which G depends. In practice, this is almost exclusively composition, but in principle, can be any parameter that influences the energy.²⁰ Considering a material with N unique elements, $G(x_1, x_2, \dots x_N)$ is plotted and the N -dimensional convex polyhedra of maximal size that connects some of these points dictates the materials that are stable (vertices) or unstable (non-vertices) with respect to property, x .²¹ For the synthesis of materials, the most convenient property of interest is composition – e.g., a furnace is loaded with a fixed amount of elements A and B (in the binary $N = 2$ example) and allowed to reach equilibrium at fixed temperature, T , and pressure, P . The convex hull in the $G(x_A, x_B)$ space then determines stability. In practice, G is rarely computed due to the expense of doing so from a first-principles approach,²² and G is approximated only by comparing the formation enthalpies that are readily available in the materials databases.

1.3 Descriptor-based property prediction

In addition to enabling large-scale stability analyses, open databases of high-throughput DFT calculations also provide the opportunity to apply statistical approaches to learn from this vast data space.²³ Historically, empirical parameters have been developed in attempts to generalize the observation of some property based on simple fundamental relationships, e.g., predicting crystal structure from the radii of the involved elements.²⁴ These approaches are extremely useful

for directing theorists and experimentalists towards which materials require more investigation, without significant computational expense. These descriptors typically involve the computation of a simple analytical expression, in contrast to DFT calculations that require a complex numerical integration of many differential equations. With the availability of so-called “big-data”, materials scientists are now afforded the opportunity to apply sophisticated statistical algorithms to learn these relationships, whereas historically these descriptors were developed with only intuition and limited empirical observation.

Within supervised learning algorithms, a compound or structure is represented by a matrix of features or properties, $\mathbf{X}$. The design of $\mathbf{X}$ is to best represent the material(s) of interest as it pertains to predicting some property, $\mathbf{y}$. Learning algorithms then identify an optimal relationship between $\mathbf{X}$ and $\mathbf{y}$ such that given new materials, they too can be represented by the same features and used to predict the target property with significant accuracy. The most effective form of representation, $\mathbf{X}$, is an ongoing problem in materials science,²⁵⁻²⁸ and there are countless developments on the side of algorithms to map $\mathbf{X}$ to $\mathbf{y}$, mostly outside the scope of materials scientists. This thesis will not devote too much time to either of these problems, instead demonstrating how existing approaches to representing materials and algorithms to predict both discrete and continuous properties can yield effective tools for predicting the thermochemistry of inorganic crystalline solids.

1.4 Thesis scope

This thesis leverages high-throughput DFT and data analytics to predict, understand, and quantify the stability and chemistry of inorganic crystalline solids. The first study (Chapter 2) focuses on stability predictions within existing DFT approaches. This acts as a tutorial for how stability is determined using the convex hull approach and also benchmarks various

approximations that are made by the high-throughput DFT community, revealing the consequences of these approximation for predicting stability. In Chapter 3, we demonstrate the application of compressed sensing to predicting the experimentally observed stability of the perovskite structure using only the chemical composition as input. This work was inspired by an empirical tolerance factor introduced by Viktor Goldschmidt in 1921 and we show how state-of-the-art statistical learning algorithms can improve quantitatively on the intuition of great solid-state chemists while still providing important chemical insights. We applied our statistically learned tolerance factor to assist in the rapid screening of candidate double perovskite solar absorbers, revealing a number of interesting candidates that are predicted to be stable and have attractive electronic properties for photovoltaic applications (Chapter 4). A persistent approximation in the computational materials community is that the formation enthalpy of solid compounds is a suitable proxy for the Gibbs energy, which actually dictates stability at a given temperature and pressure. In Chapter 5, we again used compressed sensing to learn a simple correction for the formation enthalpy, that reproduces experimental and first-principles Gibbs energies with high accuracy. We demonstrated the ultra-high-throughput application of this descriptor to explore the temperature- and composition-dependence of (meta)stability for all known inorganic solids. This Gibbs energy descriptor provides the unique capability of accurately predicting the thermochemical reaction equilibrium. Leveraging this, we applied the descriptor to pursue redox-active materials to mediate the conversion of air and water into ammonia by way of solar thermochemical ammonia synthesis (Chapter 6). Finally, a combined data-driven and “conventional” solid-state chemistry approach is used to understand bonding and stability in metal nitrides (Chapter 7). In this work, we reveal the driving forces behind the thermodynamic and chemical stability of the ternary nitrides space and identify hundreds of new, stable materials,

seven of which have since been synthesized. In Chapter 8, the key findings of this thesis are summarized and a few notes on future directions are provided.

2 The role of decomposition reactions in assessing first-principles predictions of solid stability

Much of the content in this chapter appears verbatim in the following manuscript:

C.J. Bartel, A.W. Weimer, S. Lany, C.B. Musgrave, A.M. Holder (2018) [Under review] {arXiv:1810.08248}

2.1 Abstract

The performance of density functional theory (DFT) approximations for predicting materials thermodynamics is typically assessed by comparing calculated and experimentally determined enthalpies of formation from elemental phases, ΔH_f . However, a compound competes thermodynamically with both other compounds and their constituent elemental forms, and thus, the enthalpies of the decomposition reactions to these competing phases, ΔH_d , determines thermodynamic stability. We evaluated the phase diagrams for 56,791 compounds to classify decomposition reactions into three types: 1. those that produce elemental phases, 2. those that produce compounds, and 3. those that produce both. This analysis shows that the decomposition into elemental forms is rarely the competing reaction that determines compound stability and that approximately two-thirds of decomposition reactions involve no elemental phases. Using experimentally reported formation enthalpies for 1,012 solid compounds, we assess the accuracy of the generalized gradient approximation (GGA) (PBE) and meta-GGA (SCAN) density functionals for predicting compound stability. For 646 decomposition reactions that are not trivially the formation reaction, PBE (MAD = 70 meV/atom) and SCAN (MAD = 59 meV/atom) perform similarly, and commonly employed correction schemes using fitted elemental reference energies make only a negligible improvement (~ 2 meV/atom). Furthermore, for 231 reactions

involving only compounds (Type 2), the agreement between SCAN, PBE, and experiment is within ~ 35 meV/atom and is thus comparable to the magnitude of experimental uncertainty.

2.2 Introduction

The design and discovery of new materials are being rapidly accelerated by the growing availability of density functional theory (DFT) calculated property data in open materials databases which allow users to systematically retrieve computed results for experimentally known and yet-to-be-realized solid compounds.^{16, 17, 29-32} The primary properties of interest are the optimized structure and corresponding total energy, E , with, for example, $\sim 50,000,000$ compiled structures and energies available *via* the NOMAD repository.¹⁹ Given E for a set of structures, one can routinely obtain the reaction energy, E_{rxn} , to convert between structures. E for a compound is typically compared with E for its constituent elements to obtain the formation enthalpy, ΔH_f , which provides the thermodynamic driving force at zero temperature and pressure for stability of a given structure with respect to its constituent elements:

$$\Delta H_f(A_{\alpha_1}B_{\alpha_2} \dots) = E(A_{\alpha_1}B_{\alpha_2} \dots) - \sum_i \alpha_i E_i \quad (2.1)$$

where E is the calculated total energy of the compound ($A_{\alpha_1}B_{\alpha_2} \dots$), α_i the stoichiometric coefficient of element i in the compound, and E_i the total energy (chemical potential) of element i . ΔH_f computed by **Equation 2.1** is typically compared to ΔH_f obtained experimentally at 298 K with varying degrees of agreement depending on the density functional and compounds (chemistries) under investigation.^{29, 30, 33-38}

However, ΔH_f is rarely the useful quantity for evaluating the stability of a compound. More relevant are the reaction energies for a given compound relative to all other compounds within the

same composition space, where the reaction with the most positive E_{rxn} is the decomposition reaction.^{21, 36, 39} For example, for a given ternary compound, ABC , the relevant space of competing materials includes the elements (A , B , and C), all binary compounds in the A - B , A - C , and B - C spaces, and all ternary compounds in the A - B - C space. The stability of ABC is obtained by comparing the energy of ABC with that of the linear combination of competing compounds with the same average composition – ABC – that minimizes the combined energy of the competing compounds, E_{A-B-C} . The decomposition enthalpy, ΔH_d , is then obtained by:

$$\Delta H_d = E_{rxn} = E_{ABC} - E_{A-B-C}. \quad (2.2)$$

$\Delta H_d > 0$ indicates an endothermic reaction for a given ABC forming from A - B - C ; the sign notation that $\Delta H_d > 0$ indicates instability is chosen to be commensurate with the commonly reported quantity, “energy above the hull”, where ΔH_d also provides the energy with respect to the convex hull but can be positive (for unstable compounds) or negative (for stable compounds). A ternary example was shown for simplicity, but the decomposition reaction and ΔH_d can be obtained for any arbitrary compound comprised of N elements by solving the N -dimensional convex hull problem.

For the high-throughput screening of new materials for a target application, stability against all competing compounds is an essential requirement for determining the viability of a candidate material.³⁹ In this approach, compounds are typically retained for further evaluation (more rigorous calculations or experiments) if $\Delta H_d < \gamma$, where the threshold γ commonly ranges from ~ 20 to ~ 200 meV/atom depending on the priorities of the screening approach and the breadth of materials under evaluation.⁴⁰⁻⁴⁵ The success of high-throughput screening approaches thus depends directly on the accuracy of ΔH_d , which is typically obtained using DFT with routinely employed approximations

to the exchange-correlation energy. Nevertheless, despite the intimate link between stability predictions and ΔH_d , new approaches (e.g., the development of improved density functionals and/or statistical correction schemes) are primarily benchmarked against experimentally obtained ΔH_f . Here, we show that the decomposition reactions that are relevant to stability can be classified into three types, and that the ability of DFT-based approaches to predict ΔH_d for each type relative to experiment is the applicable determinant of the viability of that method for high-throughput predictions of compound stability.

2.3 Results

2.3.1 Relevant reactions for determining the stability of compounds

The decomposition reactions that determine ΔH_d fall into one of three types: Type 1 – a given compound is the only known compound in that composition space, the decomposition products are the elements, and $\Delta H_d = \Delta H_f$ (**Figure 2.1**, left); Type 2 – a given compound is bracketed (on the phase diagram) by compounds and the decomposition products are exclusively these compounds (**Figure 2.1**, center); and Type 3 – a given compound is not the only known compound in the composition space, is not bracketed by compounds and the decomposition products are a combination of compounds and elements (**Figure 2.1**, right). For a given compound, one of these three types of decomposition reactions will be the relevant reaction for evaluating that material’s stability. Notably, these decomposition reactions apply to compounds that are stable (vertices on the convex hull, $\Delta H_d \leq 0$, **Figure 2.1**, top) and unstable (above the convex hull, $\Delta H_d > 0$, **Figure 2.1**, bottom).

As it pertains to thermodynamic control of synthesis, Type 2 reactions are insensitive to adjustments in elemental chemical potentials that are sometimes modulated by sputtering, partial

pressure adjustments, or plasma cracking. Any changes to the elemental energies will affect the decomposition products and the compound of interest proportionally, and therefore, while ΔH_f for all compounds will change, ΔH_d will be fixed. This is in contrast to Type 1 reactions which will become more favorable with increases in the chemical potential of either element. The thermodynamics of Type 3 reactions can be modulated by these synthesis approaches if the elemental form of the species whose chemical potential is being adjusted participates in the decomposition reaction, i.e. the compound must be the nearest (within the convex hull construction) stable, or lowest energy metastable, compound to the elemental chemical potential being adjusted.⁴⁶

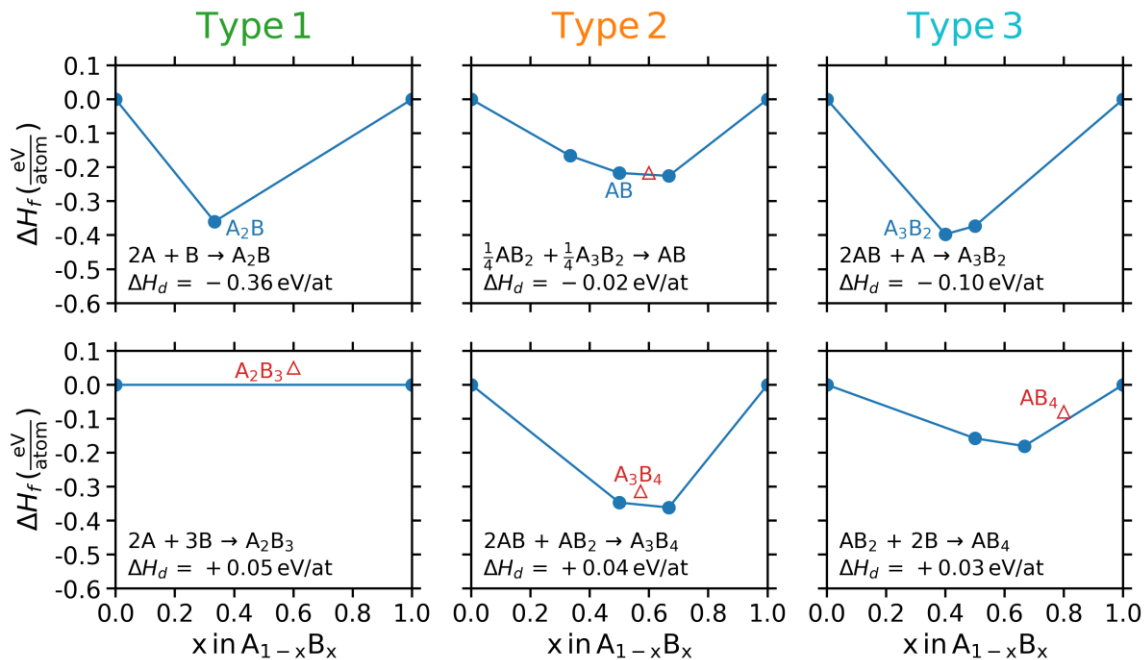


Figure 2.1 Three unique decomposition reactions

A stable (top) and metastable (bottom) example of each reaction type. Left: reaction Type 1 – the decomposition products are the elements; Center: reaction Type 2 – the decomposition products contain no elements; Right: reaction Type 3 – the decomposition products contain elements and compounds. Solid blue circles are breaks in the hull (stable) and open red triangles are above the hull (metastable). In all examples, A and B are arbitrary elements.

The relative prevalence of each decomposition pathway is not yet known, although the phase diagrams of most inorganic crystals can be resolved using open materials databases. At present, the Materials Project¹⁶ provides 56,791 unique inorganic crystalline solid compounds with computed ΔH_f . Using the N -dimensional convex hull construction, we determined ΔH_d and the stability-relevant decomposition reaction for each compound and report the prevalence of each reaction type in **Figure 2.2**. For these 56,791 compounds, Type 2 decompositions are found to be most prevalent (63% of compounds) followed by Type 3 (34%) and Type 1 (3%). Notably, 81% of Type 1 reactions (for which $\Delta H_d = \Delta H_f$) are for binary compounds, which comprise only 13% of the Materials Project, and $< 1\%$ of the non-binary compounds compete for stability exclusively with elements (**Figure 2.2**, right). As the number of unique elements in the compound, N , increases it becomes increasingly probable that other compounds will be present on the phase diagram and the decomposition will therefore be dictated by these compounds.

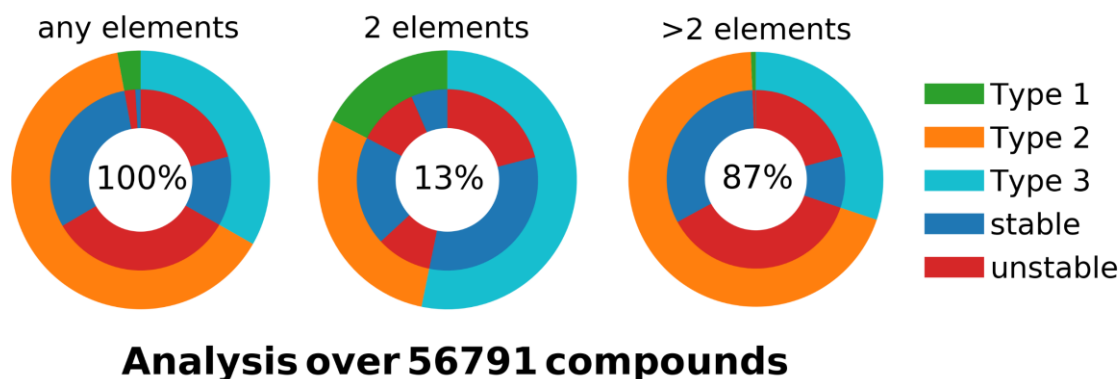


Figure 2.2 Prevalence of reactions among known materials

Partitioning Materials Project data into each of the three decomposition reaction types (outer circle). Then, for each type, partitioning compounds as stable (on the convex hull) and unstable (above the convex hull). Left – the entire database of 56,791 compounds; Center – only binary compounds; Right – only non-binary compounds. The fraction of the Materials Project comprising each circle is shown in the interior of each diagram.

2.3.2 Benchmarking performance on formation enthalpy

The decomposition reactions determining compound stability that are Type 1 are the least prevalent among Materials Project compounds ($\sim 3\%$) suggesting that ΔH_d rarely equals ΔH_f , especially for $N > 2$. Despite this, the primary approach used to benchmark first-principles thermodynamics methods is to compare experimental and computed ΔH_f . We compared experimentally obtained ΔH_f from FactSage⁴⁷ to computed ΔH_f using the generalized gradient approximation (GGA) as formulated by Perdew, Burke, and Ernzerhof (PBE)⁸ and using the strongly constrained and appropriately normed (SCAN)⁹ meta-GGA density functionals for 1,012 compounds spanning 63 elements. Importantly, this reduced space of compounds with experimental thermodynamic data decompose into the full range of Type 1 (37%), 2 (22%), and 3 (41%) reactions, but first we analyze only ΔH_f for all compounds to establish a baseline for subsequent comparison to ΔH_d . On this set of 1,012 compounds, the mean absolute difference (MAD) between experimentally determined ΔH_f (at 298 K)⁴⁷ and calculated ΔH_f , nominally at 0 K and without zero-point energy (ZPE), was found to be 196 meV/atom for PBE and 88 meV/atom for SCAN (**Figure 2.3a**). In addition to a reduction in the magnitude of residuals by $\sim 55\%$, the distribution of residuals is nearly centered about 0 for SCAN in contrast to PBE which consistently underestabilizes compounds relative to their constituent elements (particularly diatomic gases), leading to predictions of ΔH_f that are too positive by ~ 200 meV/atom. Unlike PBE, SCAN has been shown to perform well for a range of diversely bonded systems^{9, 48, 49} and does not suffer from this same systematic error.

The near zero-centered residuals produced by SCAN suggest that no global systematic difference likely exists between the energies predicted by this density functional and those obtained experimentally, and thus, some of the lingering disagreement may arise from deficiencies

in the functional for describing certain types of compounds, e.g. those with transition metals,⁴⁹⁻⁵² and/or be related to correlated noise in experimental measurement. For 228 binary and ternary compounds reported in Ref. ³⁰ (compiled from Ref. ⁵³), the MAD between experimental sources (i.e., Refs. ⁴⁷ and ⁵³) for ΔH_f is 30 meV/atom. This difference agrees well with the scale of chemical accuracy expected for the experimental determination of ΔH_f of ~ 1 kcal/mol⁴⁹ and suggests that the disagreement between experiment and theory should not be lower than ~ 30 meV/atom on average. A potential source of disagreement between experimental and DFT-calculated ΔH_f is the incongruence in temperature, where experimental measurements are performed at 298 K and DFT calculations of ΔH_f are computed at 0 K, and thus neglect heat capacity, as well as usually neglecting ZPE. These contributions are typically assumed to be small based on the results obtained for a limited set of compounds.⁵⁴ This assumption is robustly confirmed here for 647 structures where the vibrational and heat capacity effects on ΔH_f are found to be ~ 7 meV/atom on average at 298 K.

2.3.3 Optimizing elemental reference energies

Various approaches have been developed to improve the PBE prediction of ΔH_f by systematically adjusting the elemental energies, E_i , of some or all elemental phases.^{29, 30, 33-35} In the fitted elemental reference energy scheme, the difference between experimentally measured and calculated ΔH_f is minimized by optimally adjusting E_i by a correction term, $\delta\mu_i$:

$$\Delta H_{f, A\alpha_1 B\alpha_2 \dots} = E_{A\alpha_1 B\alpha_2 \dots} - \sum_i \alpha_i (E_i + \delta\mu_i). \quad (2.3)$$

To quantify the magnitude of errors that can be resolved by adjustments to the elemental reference energies, we applied **Equation 2.3** to ΔH_f computed with PBE and SCAN (**Figure 2.3b**) with all elements considered in this optimization (these approaches are denoted in this

work as PBE+ and SCAN+, respectively). Fitting reference energies for PBE approximately halves the difference between experiment and calculation and centers the residuals (MAD = 100 meV/atom). Because the difference between experiment and SCAN is less systematic, fitting reference energies improves SCAN errors less than it improves PBE, and only reduces the MAD by ~20% (MAD = 68 meV/atom).

While adjusting elemental reference energies is simple and effective in reducing the difference between experimentally determined and calculated ΔH_f , there are a number of limitations to this approach. Because it is a fitting scheme, the optimized $\delta\mu_i$ are sensitive to the set of experimental and calculated data used for fitting and do not necessarily have physical meaning – i.e., $\delta\mu_i$ accounts for the systematic disagreement between a density functional and experimental measurement across different types of materials, yet this can be difficult to interpret. The fitted reference energy scheme, as implemented here, produces a single $\delta\mu_i$ for each element whether a given element appears in the compounds as a cation or anion (e.g., Sb^{3+} or Sb^{3-}). For the majority of the compounds considered in this work, the use a single fitted value is appropriate because elements only appear in the data as either anions or cations. However, if one was interested in studying compounds containing elements that appear as cationic or anionic, statistically resolving a separate $\delta\mu_i$ for cation- and anion-specific use would be more appropriate, as the fitted correction can differ in both magnitude and sign for cations and anions. Additionally, fitted reference energies have only been available for PBE (and SCAN reported in this work), so the calculation of ΔH_f using alternative functionals, which may be better suited for a given problem, would require a re-fitting of reference energies within that functional. These limitations make it advantageous to avoid fitted reference energies for the high-throughput

prediction of stability, particularly if they have negligible effects on the validity of first-principles predictions.

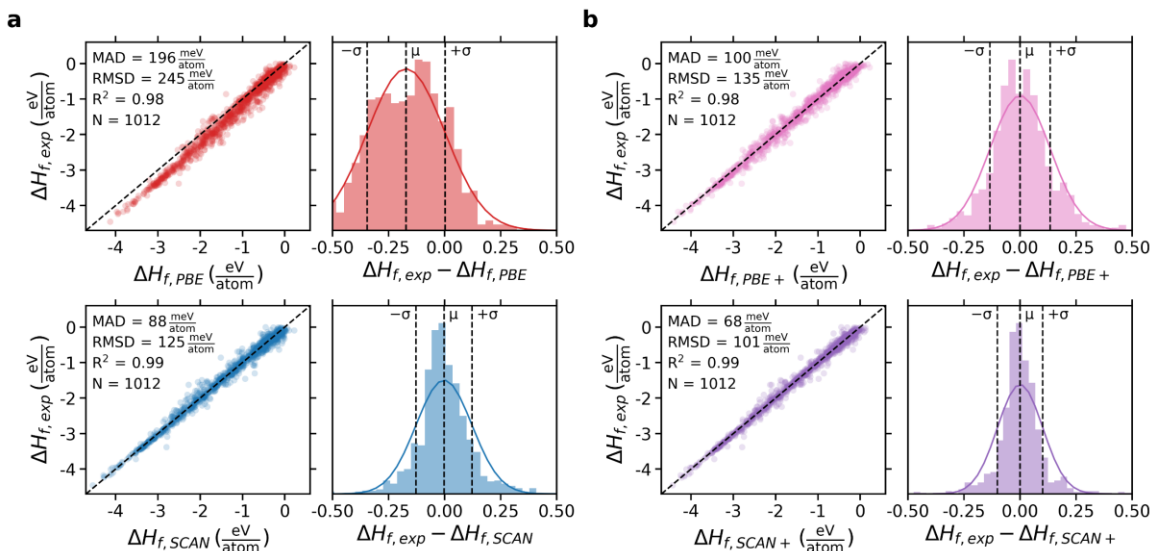


Figure 2.3 Experimental vs. theoretical formation enthalpies (Type 1)

a) A comparison of experimentally measured and DFT-calculated ΔH_f for all 1,012 compounds analyzed (PBE above; SCAN below) showing that SCAN significantly improves the prediction of ΔH_f over PBE. MAD is the mean absolute difference; RMSD is the root-mean-square difference; R^2 is the correlation coefficient; N is the number of compounds shown; μ is the mean difference; σ is the standard deviation. A normal distribution constructed from μ and σ is shown as a solid curve. **b)** For the same compounds, a comparison of PBE and SCAN with experiment using fitted elemental reference energies for the calculation of ΔH_f (PBE+ above; SCAN+ below) showed that for Type 1 reactions fitted elemental reference energies significantly improve the prediction of ΔH_f , especially predictions by PBE.

2.3.4 Decomposition reaction analysis

While the improved construction of the meta-GGA density functional (i.e., SCAN) and the use of fitted reference energies ameliorates errors associated with the insufficient description of the elements and thus improves the prediction of ΔH_f considerably relative to PBE, the effects these approaches have on the prediction of thermodynamic stability – i.e., ΔH_d – have not yet been quantified. We used ΔH_f obtained from experiment, PBE, and SCAN for the 1,012 compounds analyzed in **Figure 2.3** to perform the N -dimensional convex hull analysis to determine the

decomposition reaction and quantify ΔH_d . For 646 compounds which decompose by Type 2 or 3 reactions, the MAD between experimentally measured and DFT-computed ΔH_d is substantially lower than for ΔH_f – ~60% lower for PBE and ~30% lower for SCAN (**Fig. 4**). Notably, the decomposition reaction that results from using experiment, PBE, or SCAN is identical in terms of the competing compounds and their amounts for 89% of the 1,012 compounds evaluated.

For 231 Type 2 decomposition reactions where compounds compete only with compounds and fitted reference energies thus have no influence on ΔH_d , SCAN and PBE are found to perform comparably with MADs of ~35 meV/atom compared with experiment on ΔH_d . This difference approaches the “chemical accuracy” of experimental measurements (~1 kcal/mol), the difference in ΔH_f between experimental sources (30 meV/atom), and the difference found previously for the formation of 135 ternary metal oxides from their constituent binary oxides using an approach based on PBE with a Hubbard U correction fit specifically for transition metal oxides (24 meV/atom).³⁶ Because Type 2 decomposition reactions only involve compounds, computing the decomposition reaction energy using total energies or formation enthalpies is equivalent – therefore the results with (**Fig. 4a**) and without (**Fig. 4b**) fitted reference energies are identical.

Elemental energies are included in the calculation of ΔH_d for compounds that compete thermodynamically with both compounds and elements (Type 3 decomposition reactions). However, for 415 reactions of this type and using either SCAN or PBE we found that the use of fitted reference energies does not significantly affect the agreement with experiment for ΔH_d with improvements of only ~2 meV/atom (**Fig. 4c, d**). For these compounds, SCAN improves upon PBE by ~20% and the MAD between SCAN and experiment (73 meV/atom) falls between those for Type 1 (88 meV/atom) and Type 2 (34 meV/atom) reactions.

The prevalence of each reaction type was quantified for the Materials Project database, with Type 2 reactions accounting for 63% of all decompositions evaluated and this fraction increasing from 29% to 67% to 75% for binary, ternary, and quaternary compounds, respectively. For these cases, our results show that both SCAN and PBE can be expected to yield chemically accurate predictions of ΔH_d , which quantifies the driving force for thermodynamic stability. While on average, SCAN and PBE perform similarly for ΔH_d , this analysis is performed only on ground-state structures within each functional. It was recently shown that SCAN performs significantly better than PBE for structure selection – i.e., identifying the correct polymorph ordering of which crystal structure is the lowest energy at fixed composition.⁴⁹ Here, ~10% of the 2,238 structures optimized were found to have different space groups using PBE and SCAN. Considering only ground-states, the lowest energy PBE and SCAN structures differ for ~11% of the 1,012 unique compositions assessed in this work. While the MAD from experiment for ΔH_d differs by only ~20% between SCAN and PBE, additional advantages when considering structure and properties are likely associated with the use of SCAN for the accurate description of compounds.^{9, 48, 49, 52, 55}

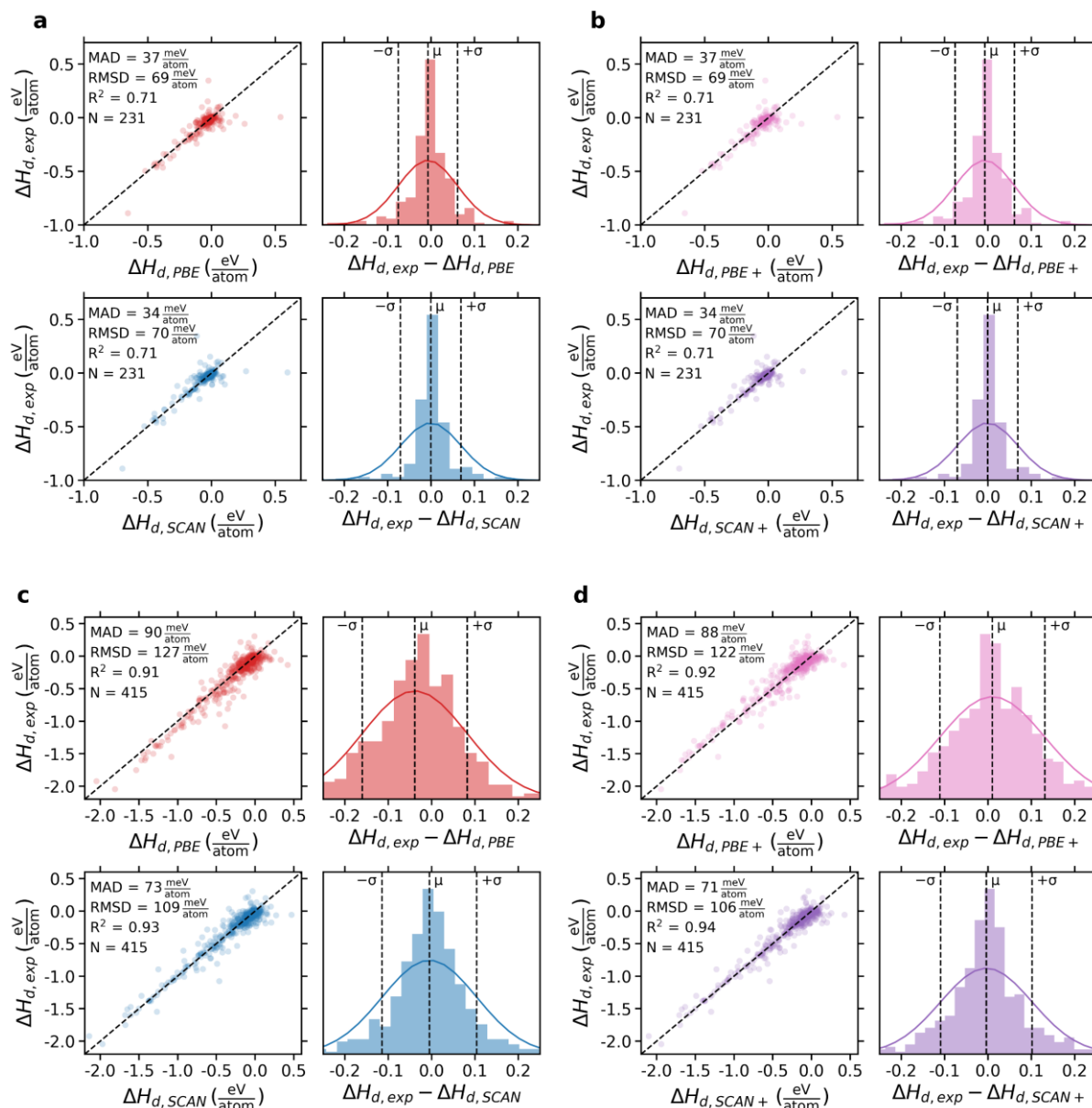


Figure 2.4 Experimental vs. theoretical decomposition enthalpies

a) A comparison of experimentally measured and DFT-calculated ΔH_d (PBE above; SCAN below) for 231 compounds that undergo Type 2 decomposition reactions showing similar performance between PBE and SCAN in predicting ΔH_d . **b)** For the same compounds, a comparison of PBE and SCAN with experiment using fitted elemental reference energies for the calculation of ΔH_d (PBE+ above; SCAN+ below) showing identical results as **(a)** due to a cancellation of elemental energies for these Type 2 decomposition reactions. **c)** A comparison of experimentally measured and DFT-calculated ΔH_d (PBE above; SCAN below) for 415 compounds that undergo Type 3 decomposition showing similar performance between PBE and SCAN in predicting ΔH_d . **d)** For the same compounds, a comparison of PBE and SCAN with experiment using fitted elemental reference energies for the calculation of ΔH_d (PBE+ above; SCAN+ below) showing that adding fitted elemental reference energies does not significantly

improve the prediction of ΔH_d for Type 3 decomposition reactions. Annotations are as described in the **Figure 2.3** caption.

2.4 Discussion

For 1,012 compounds, we show the improvement in computed formation enthalpies, ΔH_f , afforded by fitting elemental reference energies for both GGA (PBE) and meta-GGA (SCAN) density functionals (**Figure 2.3**). However, to accurately predict the stability of materials it is essential to accurately compute the decomposition enthalpy, ΔH_d , which dictates stability with respect to all compounds and elements in a given chemical space. ΔH_d is computed by determining the stoichiometric decomposition reaction with the most positive reaction energy. ΔH_f is only relevant for the stability of compounds that undergo Type 1 decompositions, where the compound only competes with elemental phases and consequently, $\Delta H_d = \Delta H_f$. (**Figure 2.1**). Furthermore, Type 1 decompositions occur for only 17% of binaries and almost never ($< 1\%$) for non-binaries, as shown for the $\sim 60,000$ N -component compounds evaluated (**Figure 2.2**). For this reason, ΔH_f and the agreement between experiment and theory for ΔH_f are rarely relevant to the stability of materials. However, for other applications such as the calculation of defect formation energies, ΔH_f is the relevant materials property and the adjustment of calculated chemical potentials using the fitted elemental reference energy scheme may still have significant utility, especially when using PBE.

The stability of a substantial fraction of compounds, those that undergo Type 2 decompositions, can be determined without any consideration of elemental energies. For these compounds, PBE and SCAN perform similarly and approach the resolution of experimental approaches to determining ΔH_f (~ 30 meV/atom) (**Figure 2.4a**). Importantly, the performance metrics we provide are evaluated over a wide range of compounds and chemistries. For chemical

spaces that are known to be problematic for a given approach (e.g., 3d transition metals for PBE), the error can significantly exceed the average difference reported here.^{49, 52}

While the majority of compounds in the Materials Project compete with Type 2 decomposition reactions, this is not generally known when first evaluating a compound and so high-throughput screening approaches that typically survey a wide range of compounds will likely include analysis of Type 1 and Type 3 decomposition reactions that do require the calculation of elemental energies. Type 1 decompositions, which occur for binary compounds in sparsely explored chemical spaces, will be highly sensitive to the functional and elemental energies and SCAN improves significantly upon PBE for these compounds. Notably, fitting elemental reference energies for PBE still results in larger errors than SCAN and fitting reference energies for SCAN leads to only modest additional improvements. For Type 3 decompositions, which are $\sim 10\times$ more prevalent than Type 1 reactions in Materials Project, SCAN improves upon PBE by $\sim 20\%$ and the use of fitted elemental reference energies has almost no effect (~ 2 meV/atom on average) on either approach (**Figure 2.4c, d**). Interestingly, considering the $\sim 60,000$ compounds in Materials Project (**Figure 2.2**, left), a roughly equal fraction of Type 2 compounds are stable (48%) and unstable, yet only 37% of Type 3 compounds are stable. However, Type 3 compounds are more amenable to non-equilibrium synthesis approaches that allow for increased chemical potentials of the elements and potential access of metastable compounds.⁴⁶

In summary, we've shown that the decomposition reactions that dictate the stability of solid compounds can be divided into three types that depend on the presence of elemental phases in the decomposition reaction. Through a global evaluation of phase diagrams for $\sim 60,000$ compounds in the Materials Project, we quantify the prevalence of these reaction types and show that the formation enthalpy is rarely the quantity of interest for stability predictions ($\sim 3\%$ of Materials

Project compounds). Instead, the decomposition enthalpy, which may or may not include the calculation of elemental phases is the most relevant quantity. Benchmarking the PBE and SCAN density functionals against decomposition enthalpies obtained from experimental data reveals quantitatively and qualitatively different results than benchmarking only against formation enthalpies and in most cases mitigates the need to systematically correct DFT-calculated elemental energies for the assessment of stability. We showed that for 231 reaction energies between compounds, the agreement between SCAN, PBE, and experiment (~ 35 meV/atom) is comparable to the expected noise in experimental measurements. Therefore, because this type of decomposition reaction is predominant in determining solid stability, we show that high-throughput DFT approaches to stability predictions are generally in excellent agreement with experiment. For alternative decomposition reactions that include both compounds and elements or problems that require higher energy resolution such as polymorph energy ordering,^{50, 56} the choice of functional (e.g., SCAN instead of PBE) can have non-negligible effects on stability predictions.

2.5 Methods

Experimental values for ΔH_f were obtained from the FactSage database⁴⁷ for 1,012 compounds as reported at 298 K and 1 atm. For each compound, the NREL Materials Database (NRELMatDB)⁵⁷ was queried for structures matching the composition within 50 meV/atom of the ground-state structure as reported in the database. If a given compound had no calculated structures tabulated in NRELMatDB, the procedure was repeated with the Materials Project database¹⁶. Structures containing potentially magnetic elements were sampled in non-magnetic, two ferromagnetic (high- and low-spin), and up to 16 antiferromagnetic configurations (depending on cell configuration) where the ground-state magnetic configuration was retained for each structure. Sampling was performed using the approach described by NRELMatDB. This process was also

repeated for all 63 elements represented in the dataset with the exception of H₂, N₂, O₂, F₂, and Cl₂ which were calculated as diatomic molecules in a 15×15×15 Å box. After magnetic sampling, 2,238 unique structures were found for the 1,012 compounds and 63 elements. All structures were optimized with PBE and SCAN using the Vienna Ab Initio Simulation Package (VASP)^{12, 58} using the projector augmented wave (PAW) method^{59, 60}, a plane wave energy cutoff of 520 eV, and a Γ -centered Monkhorst-Pack k-point grid with $N = 20|b_i|$ discretizations along each reciprocal lattice vector, b_i . The energy cutoff, k-point density, and related convergence settings were sufficient to achieve total energy convergence of < 5 meV/atom for all calculations. For the calculation of phonons to compute thermal effects, the finite displacement method with 2×2×2 supercells as implemented in PHONOPY²² was used with SCAN and an increased plane wave cutoff of 600 eV and further tightened convergence criteria for total energy convergence of < 1 meV/atom.

3 New tolerance factor to predict the stability of perovskite oxides and halides

Much of the content in this chapter appears verbatim in the following manuscript:

C.J. Bartel, C. Sutton, B.R. Goldsmith, R. Ouyang, C.B. Musgrave, L. Ghiringhelli, M. Scheffler (2018) [Under review] {arXiv:1801.07700}

3.1 Abstract

Predicting the stability of the perovskite structure remains a longstanding challenge for the discovery of new functional materials for many applications including photovoltaics and electrocatalysts. We developed an accurate, physically interpretable, and one-dimensional tolerance factor, τ , that correctly predicts 92% of compounds as perovskite or nonperovskite for an experimental dataset of 576 ABX_3 materials ($X = O^{2-}, F^-, Cl^-, Br^-, I^-$) using a novel data analytics approach based on SISSO (sure independence screening and sparsifying operator). τ is shown to generalize outside the training set for 1,034 experimentally realized single and double perovskites (91% accuracy) and is applied to identify 23,314 new double perovskites ($A_2BB'X_6$) ranked by their probability of being stable as perovskite. This work guides experimentalists and theorists towards which perovskites are most likely to be successfully synthesized and demonstrates an approach to descriptor identification that can be extended to arbitrary applications beyond perovskite stability predictions.

3.2 Introduction

Crystal structure prediction from chemical composition continues as a persistent challenge to accelerated materials discovery.^{24, 61} Most approaches capable of addressing this challenge require several computationally demanding electronic-structure calculations for each material

composition, limiting their use to a small set of materials.⁶²⁻⁶⁵ Alternatively, descriptor-based approaches enable high-throughput screening applications because they provide rapid estimates of material properties.^{66, 67} Notably, the Goldschmidt tolerance factor,⁶⁸ t , has been used extensively to predict the stability of the perovskite structure based only on the chemical formula, ABX_3 , and the ionic radii, r_i , of each ion (A, B, X):

$$t = \frac{r_A + r_X}{\sqrt{2}(r_B + r_X)} \quad (3.1)$$

The perovskite crystal structure, as shown in **Figure 3.1a**, is defined as any ABX_3 compound with a network of corner-sharing BX_6 octahedra surrounding a larger A -site cation ($r_A > r_B$), where the cations, A and B , can span the periodic table and the anion, X , is typically a chalcogen or halogen. Distortions from the cubic structure can arise from size-mismatch of the cations and anion, which results in additional perovskite structures and nonperovskite structures. The B cation can also be replaced by two different ions, resulting in the double perovskite formula, $A_2BB'X_6$ (**Figure 3.1b**). Single and double perovskite materials possess exceptional properties for a variety of applications such as electrocatalysis,⁶⁹ proton conduction,⁷⁰ ferroelectrics⁷¹ (using oxides, $X = O^{2-}$), battery materials⁷² (using fluorides, $X = F^-$), as well as photovoltaics⁷³ and optoelectronics⁷⁴ (using the heavier halides, $X = Cl^-, Br^-, I^-$).

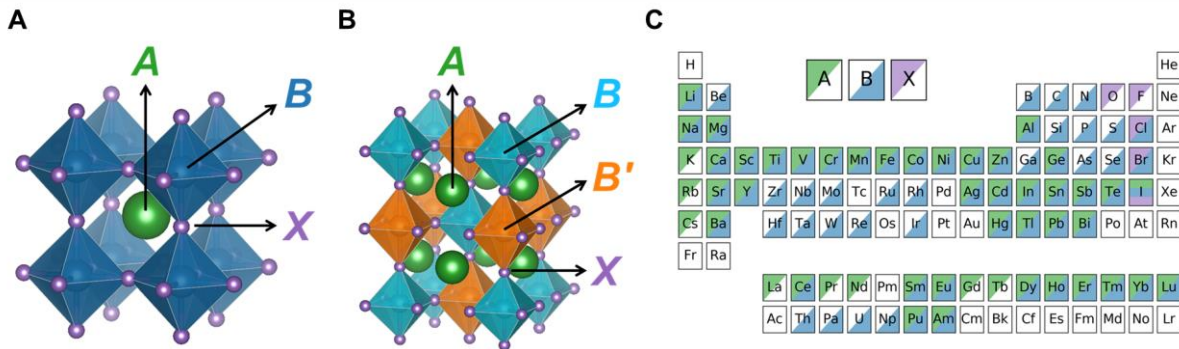


Figure 3.1 Perovskite structure and composition

a) ABX_3 , in the cubic single perovskite structure ($Pm\bar{3}m$), where the A cation is surrounded by a network of corner-sharing BX_6 octahedra. **b)** $A_2BB'X_6$, in the rock salt double perovskite structure ($Fm\bar{3}m$), where the A cations are surrounded by an alternating network of BX_6 and $B'X_6$ octahedra. In this structure, B and B' are indistinguishable. **c)** A map of the elements that occupy the A, B, and/or X sites within the 576 compounds experimentally characterized as perovskite or nonperovskite at ambient conditions and reported in ⁷⁵⁻⁷⁷.

The first step in designing new perovskites for these applications is typically the assessment of stability using t , which has informed the design of perovskites for over 90 years. However, as reported in recent studies, its accuracy is often insufficient.⁷⁸ Considering 576 ABX_3 solids experimentally characterized at ambient conditions and reported in ⁷⁵⁻⁷⁷ (see **Figure 3.1c** for the A, B, and X elements in this set), t correctly distinguishes between perovskite and nonperovskite for only 74% of materials and performs considerably worse for compounds containing heavier halides (chlorides – 51% accuracy, bromides – 56%, and iodides – 33%) than for oxides (83%) and fluorides (83%) (**Figure 3.1a**). This deficiency in generalization to halide perovskites severely limits the applicability of t for materials discovery.

In this work, we present a new tolerance factor, τ , which has the form:

$$\tau = \frac{r_X}{r_B} - n_A \left(n_A - \frac{r_A/r_B}{\ln(r_A/r_B)} \right) \quad (3.2)$$

where n_A is the oxidation state of A , r_i is the ionic radius of ion i , and $r_A > r_B$ by definition. A high overall accuracy of 92% for the experimental set (94% for a randomly chosen test set of 116 compounds) and nearly uniform performance across the five anions evaluated (oxides – 92% accuracy, fluorides – 92%, chlorides – 90%, bromides – 93%, iodides – 91%) is achieved with τ (**Figure 3.2b**). Like t , the prediction of perovskite stability using τ requires only the chemical composition, allowing the tolerance factor to be agnostic to the many structures that are considered perovskite. In addition to predicting *if* a material is stable as perovskite, τ also provides a monotonic estimate of the *probability* that a material is stable in the perovskite structure. The accurate and probabilistic nature of τ as well as its generalizability over a broad range of single and double perovskites allows for new physical insights into the stability of the perovskite structure and the prediction of thousands of new double perovskite oxides and halides, 23,314 of which are provided here and ranked by their probability of being stable in the perovskite structure.

3.3 Results

3.3.1 Finding an improved tolerance factor

One key aspect of the performance of t is how well the sum of ionic radii estimates the interatomic bond distances for a given structure. Shannon’s revised effective ionic radii,⁷⁹ based on a systematic empirical assessment of interatomic distances in nearly 1,000 compounds, are the typical choice for radii because they provide ionic radius as a function of ion, oxidation state, and coordination number for the majority of elements. Most efforts to improve t have focused on refining the input radii^{75, 77, 80, 81} or increasing the dimensionality of the descriptor through two-dimensional structure maps^{76, 82, 83} or high-dimensional machine learned models.⁸⁴⁻⁸⁶ However, all hitherto applied approaches for improving the Goldschmidt tolerance factor are only effective over

a limited range of ABX_3 compositions. Despite its modest classification accuracy, t remains the primary descriptor used by experimentalists and theorists to predict the stability of perovskites.

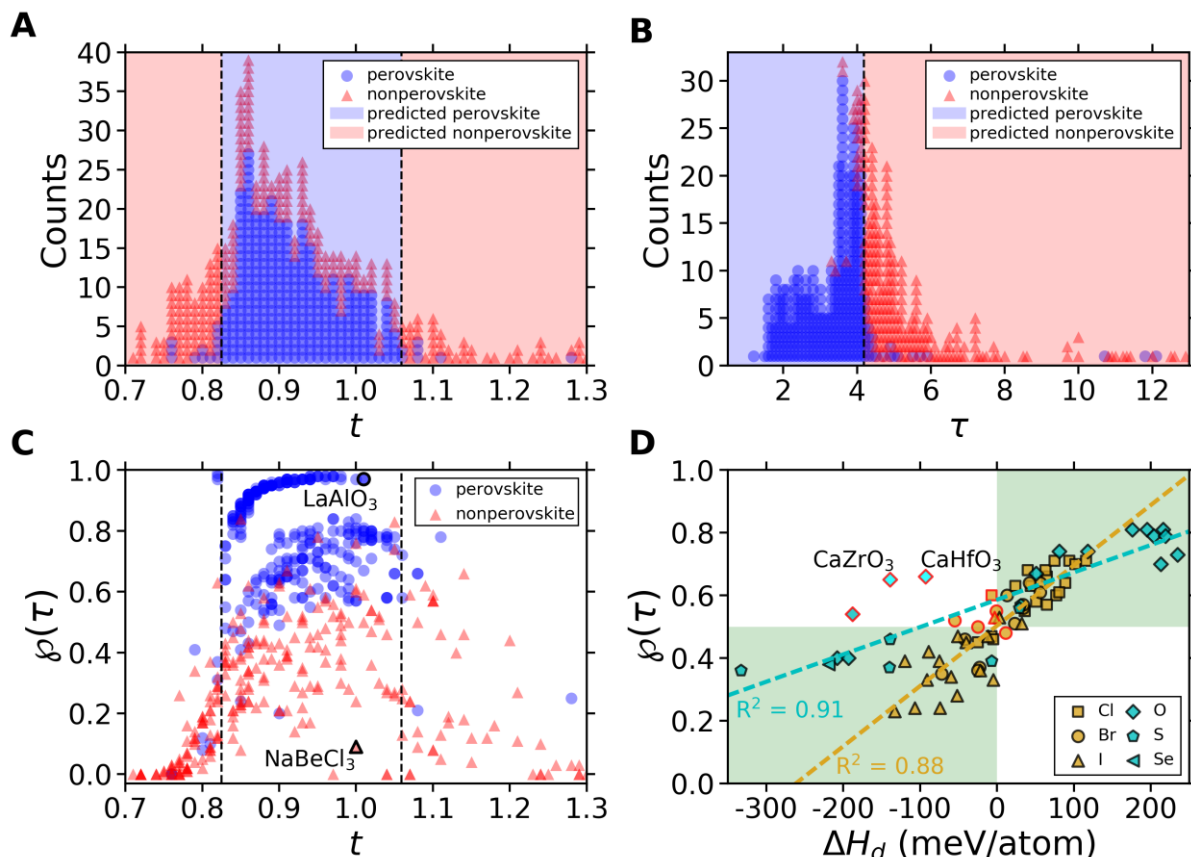


Figure 3.2 Assessing the performance of the improved tolerance factor, τ

a) A decision tree classifier determines that the optimal bounds for perovskite formability using the Goldschmidt tolerance factor, t , are $0.825 < t < 1.059$, which yields a classification accuracy of 74% for 576 experimentally characterized ABX_3 solids. **b)** τ achieves 92% classification accuracy on the set of 576 ABX_3 solids based on perovskite classification for $\tau < 4.18$. The largest value of τ in the experimental set of 576 compounds is 181.5, however, all points with $\tau > 13$ are correctly labeled as nonperovskite and not shown to highlight the decision boundary. The outlying compounds at $\tau > 10$ that are labeled perovskite yet have large τ are PuVO_3 , AmVO_3 , and PuCrO_3 , which may indicate poorly defined radii or incorrect experimental characterization. **c)** A comparison of Platt-scaled classification probabilities, $\phi(\tau)$, versus t . LaAlO_3 and NaBeCl_3 are labeled to highlight the variation in ϕ at nearly constant t . **d)** A comparison between ϕ and the decomposition enthalpy (ΔH_d) for 36 double perovskite halides calculated using density functional theory (DFT) in the $Fm\bar{3}m$ structure in ⁸⁷, and 37 single and double perovskite chalcogenides and halides in the $Pm\bar{3}m$ structure in ⁸⁷. The legend corresponds with the anion, X . Positive decomposition enthalpy ($\Delta H_d > 0$) indicates the structure is stable with respect to decomposition into competing compounds. The green and white shaded regions correspond with agreement and disagreement between the calculated ΔH_d and the classification by τ . Points of disagreement are outlined in red. CaZrO_3 and CaHfO_3 are labeled because they are known to be stable in the perovskite structure, although they are unstable in the cubic structure.^{88, 89} For this reason, the best fit line for the chalcogenides ($X = \text{O}^{2-}$, S^{2-} , Se^{2-}) excludes these two points.

The SISSO (sure independence screening and sparsifying operator) approach⁹⁰ was used to identify an improved tolerance factor for predicting whether a given compound will be stable as perovskite (determined by experimental realization of any structure with corner-sharing BX_6 octahedra at ambient conditions). Of the 576 experimentally characterized ABX_3 solids, 80% were used to train and 20% to test the SISSO-learned descriptor. Several alternative atomic properties were considered as candidate features and, among them, SISSO determined that the best performing descriptor, τ (**Equation 3.2**, **Figure 3.2b**), depends only on oxidation states and Shannon ionic radii. For the set of 576 ABX_3 compositions, τ correctly labels 94% of the perovskites and 89% of the nonperovskites compared with 94% and 49%, respectively, using t . The primary advantage of τ over t is the remarkable reduction in false positives – compounds predicted to be perovskite but are not experimentally identified as stable perovskites – with false positive rates for τ and t of 11% and 51%, respectively. The large decrease in false positive rate (from 51% to 11%) while substantially increasing the overall classification accuracy (from 74% to 92%) demonstrates that τ improves significantly upon t as a reliable tool to guide experimentalists towards which compounds can be synthesized in perovskite structures.

Beyond the improved accuracy, a significant advantage of τ is the monotonic (continuous) dependence of perovskite stability on τ . As τ decreases, the τ -based probability of being perovskite ($\wp$) increases, where perovskites are expected for an empirically determined range of $\tau < 4.18$ (**Figure 3.2b**). Probabilities are obtained using Platt's scaling,⁹¹ where the *binary classification* of perovskite/nonperovskite is transformed into a *continuous probability* estimate of perovskite stability, $\wp(\tau)$, by training a logistic regression model on the τ -derived binary classification. Probabilities cannot similarly be obtained with t because the stability of the perovskite structure does not increase or decrease monotonically with t , where $0.825 < t < 1.059$ results in a

classification as perovskite (this range maximizes the classification accuracy of t on the set of 576 compounds). While $\wp$ is sigmoidal with respect to τ because of the logistic fit, a bell-shaped behavior of $\wp$ with respect to t is observed because of the multiple decision boundaries required for t (**Figure 3.2c**). This relationship leads to an increase in $\wp$ (i.e., probability of perovskite stability using τ) with an increase in t until a value of $t \sim 0.9$. Beyond this range, the probabilities level out or decrease as t increases further.

The disparity between $\wp$ and the assignment by t can be significant, especially in the range where t predicts a stable perovskite ($0.825 < t < 1.059$). A comparison of the perovskite, LaAlO_3 , and the nonperovskite, NaBeCl_3 , illustrates the discrepancy between these two approaches. t incorrectly predicts both compounds to be perovskite ($t = 1.0$) whereas $\wp$ varies from $< 10\%$ for NaBeCl_3 to $> 97\%$ for LaAlO_3 , in agreement with experiment. For NaBeCl_3 , instability in the perovskite structure arises from an insufficiently large Be^{2+} cation on the B -site, which leads to unstable BeCl_6 octahedra. This contribution to perovskite stability is accounted for in the first term of τ (**Equation 3.2**, $r_X/r_B = \mu^{-1}$, where μ is the octahedral factor). Note that μ is the typical choice for a second feature used in combination with t ,^{76, 77, 82} and was recently used to assess the predictive accuracy of Goldschmidt’s “no-rattling” principle. In this analysis, six inequalities dependent upon t and μ were derived and used to predict the formability of single and double perovskites with a reported accuracy of $\sim 80\%$.⁵⁷ Notably, training a decision tree algorithm on the bounds of t and μ that optimally separate perovskite from nonperovskite leads to an 85% classification accuracy for this dataset. In contrast to these two-dimensional descriptors based on (t, μ) , τ incorporates μ as a one-dimensional descriptor yet still achieves a higher accuracy of 92%, demonstrating the capability of the SISSO algorithm to identify a highly accurate tolerance factor comprised of intuitively meaningful parameters.

The nature of geometrical descriptors, such as t or μ , is fundamentally different than data-driven descriptors, such as τ . t and μ are derived from geometric constraints that indicate when the perovskite structure is a possible structure that can form. However, these constraints do not necessarily indicate when the perovskite structure is the ground-state structure and does, in fact, form. For instance, if $t = 1$ and the ionic limit on which t was derived is applicable (the interatomic distances are sums of the ionic radii), these criteria do not suggest perovskite is the ground-state structure, only that the interatomic distances are such that the lattice constants in the A - X and B - X directions can be commensurate with the perovskite structure. The fact that t does not guarantee the formation of the perovskite structure is evident by the high false positive rate (51%) in the region of t where perovskite is expected ($0.825 < t < 1.059$). Similarly, although μ may fall within the range where BX_6 octahedra are expected based on geometric considerations ($0.414 < \mu < 0.732$), the octahedra that form may be edge- or face-sharing and therefore the observed structure is nonperovskite. In this work, SISSO searches a massive space of potential descriptors to identify the one that most successfully detects when a given chemical formula will or will not crystallize in the perovskite structure and because this is the target property, τ emerges as a much more predictive descriptor than t or μ .

Although the classification by τ disagrees with the experimental label for 8% of the 576 compounds, the agreement increases to 99% outside the range $3.31 < \tau < 5.92$ (200 compounds) and 100% outside the range $3.31 < \tau < 12.08$ (152 compounds). The experimental dataset may also be imperfect as compounds can manifest different crystal structures as a function of the synthesis conditions due to, for example, defects in the experimental samples (impurities, vacancies, etc.). These considerations emphasize the usefulness of τ -derived probabilities, in addition to the binary

classification of perovskite/nonperovskite, which address these uncertainties in the experimental data and corresponding classification by τ .

3.3.2 Comparing to calculated perovskite stabilities

The precise and probabilistic nature of τ , as well as its simple functional form – depending only on widely available Shannon radii (and the oxidation states required to determine the radii) – enables the rapid search across composition space for stable perovskite materials. Prior to attempting synthesis, it is common for new materials to be examined using computational approaches, therefore it is useful to compare the predictions from τ with those obtained using density functional theory. The stabilities (decomposition enthalpies, ΔH_d) of 73 single and double perovskite chalcogenides and halides were recently examined with density functional theory utilizing the Perdew-Burke-Ernzerhof⁸ exchange-correlation functional (DFT).^{87, 92} τ is found to agree with the calculated stability for 64 of 73 calculated materials. Significantly, the probabilities that result from classification with τ linearly correlate with ΔH_d , demonstrating the value of the monotonic behavior of τ and $\wp$ (**Figure 3.2d**).

Although τ appears to disagree with these DFT calculations for nine compounds, six disagreements lie near the decision boundaries ($\wp = 0.5$, $\Delta H_d = 0$ meV/atom), suggesting that they cannot be confidently classified as stable or unstable perovskites using τ or DFT calculations of the cubic structure. Of the remaining disagreements, CaZrO_3 and CaHfO_3 reveal the power of τ compared with DFT calculations of the cubic structure as these two oxides are known to be isostructural with the orthorhombic perovskite CaTiO_3 , from which the name perovskite originates.^{88, 89} $\Delta H_d < -90$ meV/atom for these two compounds in the cubic structure, indicating they are nonperovskites. In contrast, τ predicts both compounds to be stable perovskites with ~65% probability, which agrees with experiment. These results show that a key challenge in the

prediction of perovskite stability from quantum chemical calculations is the requirement of a specific structure as an input as there are more than a dozen unique structures classified as perovskite (i.e., those having corner-sharing BX_6 octahedra) and many more which are nonperovskite.

Several recent machine-learned descriptors for perovskite stability have been trained or tested on DFT-calculated stabilities of only the cubic perovskite structure.^{28, 92-94} However, less than 10% of perovskites are observed experimentally in this structure,⁸⁰ leading to an inherent disagreement between the descriptor predictions and experimental observations. Recently, it was shown that of 254 synthesized perovskite oxides (ABO_3), DFT calculations in the Open Quantum Materials Database (OQMD)²⁹ predict only 186 (70%) to be stable or even moderately unstable (within 100 meV/atom of the convex hull).⁸⁶ The discrepancy is likely associated with the difference in energy between the true perovskite ground state and the calculated high-symmetry structure(s). Because τ was trained exclusively on the experimental characterization of ABX_3 compounds, τ is informed by the true ground state (or metastable but observed) structure of each ABX_3 and the potential for these compounds to decompose into any compound(s) in the A - B - X composition space. A principal advantage of τ over many existing descriptors is that its identification and validation were based on experimentally observed stability or instability of a structurally diverse dataset.

3.3.3 Extension to double perovskite oxides and halides

Double perovskites are particularly intriguing as an emerging class of semiconductors that offer a lead-free alternative to traditional perovskite photoabsorbers and increased compositional tunability for enhancing desired properties such as catalytic activity.^{69, 78, 95} Still, the experimentally realized composition space of double perovskites is relatively unexplored

compared with the number of possible A , B , B' , and X combinations that can form $A_2BB'X_6$ compounds. The set of 576 compounds used for training and testing τ are comprised of 49 A cations, 67 B cations, and 5 X anions, from which $> 500,000$ double perovskite formulas, $A_2BB'X_6$, can be constructed. Comparing with the Inorganic Crystal Structure Database (ICSD)^{57, 96} reveals only 918 compounds ($< 0.2\%$) with known crystal structures, 868 of which are perovskite.

Although τ was only trained on ABX_3 compounds, it is readily adaptable to double perovskites because it depends only on composition and not structure. To extend τ to $A_2BB'X_6$ formulas, r_B is approximated as the arithmetic mean of the two B -site radii (r_B , $r_{B'}$). τ correctly classifies 91% of these 918 $A_2BB'X_6$ compounds in the ICSD (compared with 92% on 576 ABX_3 compounds), recovering 806 of 868 known double perovskites. The geometric mean has also been used to approximate the radius of a site with two ions.⁹⁷ We find this has little effect on classification with τ as 91% of the 918 $A_2BB'X_6$ compounds are also correctly classified using the geometric mean for r_B , and the classification label differs for only 14 of 918 compounds using the arithmetic or geometric mean. Although τ was identified using 460 ABX_3 compounds, the agreement with experiment on these compounds (92%) is comparable to that on 1,034 compounds (91%) that span ABX_3 (116 compounds) and $A_2BB'X_6$ (918 compounds) formulas and were completely excluded from the development of τ (i.e., test set compounds). This result indicates significant generalizability to predicting experimental realization for single and double perovskites that are yet to be discovered. With τ thoroughly validated as being predictive of experimental stability, the space of yet-undiscovered double perovskites was explored to identify 23,314 charge-balanced double perovskites that τ predicts to be stable at ambient conditions (of $> 500,000$ candidates). Importantly, there are many thousands of additional compounds with substitutions on

the A and/or X sites – $AA'BB'(XX')_6$ – that are expected to be similarly rich in yet-undiscovered perovskite compounds.

Two particularly attractive classes of materials within this set of $A_2BB'X_6$ compounds are double perovskites with $A = \text{Cs}^+$, $X = \text{Cl}^-$ and $A = \text{La}^{3+}$, $X = \text{O}^{2-}$ which have garnered significant interest in a number of applications including photovoltaics, electrocatalysis, and ferroelectricity. The ICSD contains 45 compounds (42 perovskites) with the formula $\text{Cs}_2BB'\text{Cl}_6$, 43 of which are correctly classified as perovskite or nonperovskite by τ . From the high-throughput analysis using τ , we predict an additional 420 perovskites to be stable with 164 having at least the probability of perovskite formation as the recently synthesized perovskite, $\text{Cs}_2\text{AgBiCl}_6$ ($\wp = 69.6\%$).⁹⁸ A map of perovskite probabilities for charge-balanced $\text{Cs}_2BB'\text{Cl}_6$ compounds is shown in **Figure 3.3** (lower triangle). Within this set of 164 probable perovskites, there is significant opportunity to synthesize double perovskite chlorides that contain $3d$ transition metals substituted on one or both B -sites as 83 new compounds of this type are predicted to be stable as perovskite with high probability.

While double perovskite oxides have been explored extensively for a number of applications, the small radius and favorable charge of O^{2-} yields a massive design space for the discovery of new compounds. For $\text{La}_2BB'\text{O}_6$, $\sim 63\%$ of candidate compositions are found to be charge-balanced compared with only $\sim 24\%$ of candidate $\text{Cs}_2BB'\text{Cl}_6$ compounds. The ICSD contains 85 $\text{La}_2BB'\text{O}_6$ compounds, *all of which* are predicted to be perovskite by τ in agreement with experiment. We predict an additional 1,128 perovskites to be discoverable in this space, with a remarkable 990 having $\wp \geq 85\%$ (**Figure 3.3**, upper triangle). All 128 ABX_3 compounds in the experimental set that meet this threshold are experimentally realized as perovskite, suggesting there is ample opportunity for perovskite discovery in lanthanum oxides.

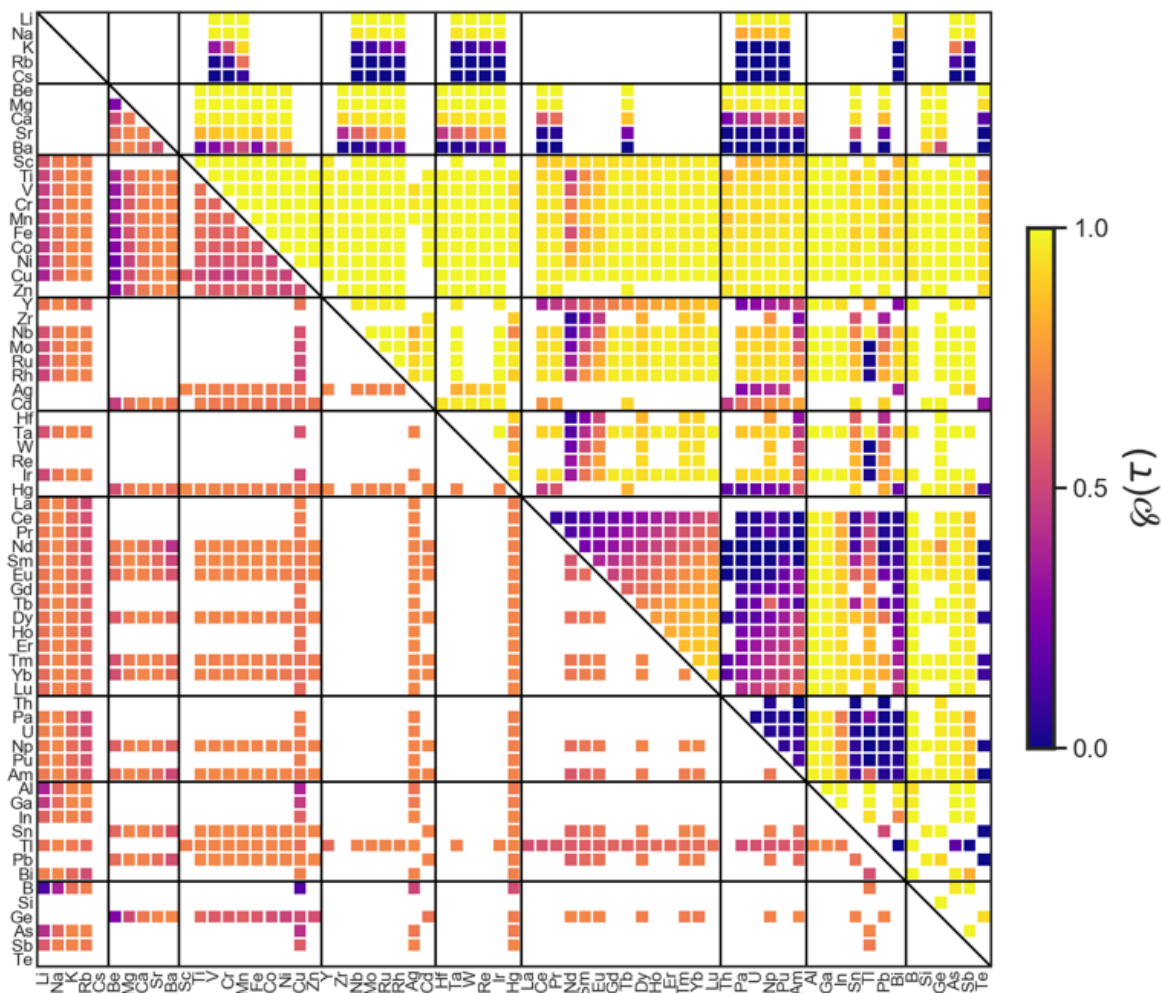


Figure 3.3 Map of predicted double perovskite oxides and halides

Lower triangle: the probability of forming a stable perovskite with the formula $\text{Cs}_2\text{BB}'\text{Cl}_6$ as predicted by τ . *Upper triangle:* the probability of forming a stable perovskite with the formula $\text{La}_2\text{BB}'\text{O}_6$ as predicted by τ . White spaces indicate B/B' combinations that do not result in charge-balanced compounds with $r_A > r_B$. The colors indicate the Platt-scaled classification probabilities of $\phi(\tau)$, with higher ϕ indicating a higher probability of forming a stable perovskite. B/B' sites are restricted to ions that are labeled as B sites in the experimental set of 576 ABX_3 compounds.

3.3.4 Compositional mapping of perovskite stability

In addition to enabling the rapid exploration of stoichiometric perovskite compositions, τ provides the probability of perovskite stability for an arbitrary combination of n_A , r_A , r_B , and r_X , which is shown in **Figure 3.4**. For each grouping shown in **Figure 3.4**, experimentally realized

perovskites and nonperovskites are shown as single points to compare with the range of values in the predictions made from τ . Doping at various concentrations presents a nearly infinite number of $A_{1-x}A'_xB_{1-y}B'_y(X_{1-z}X'_z)_3$ compositions that allows for the tuning of technologically useful properties. τ suggests the size and concentration of dopants on the A , B , or X sites that likely lead to improved stability in the perovskite structure. Conversely, compounds that lie in the high-probability region are likely amenable to ionic substitutions that decrease the probability of forming a perovskite, but may improve a desired property for another application. For example, LaCoO_3 , with $\wp = 98.9\%$, should accommodate reasonable ionic substitutions (i.e., A -sites of comparable size to La or B -sites of comparable size to Co), and was recently shown to have enhanced oxygen exchange capacity and nitric oxide oxidation kinetics with stable substitutions of Sr on the A -site.⁹⁹

The probability maps in **Figure 3.4** arise from the functional form of τ (**Equation 3.2**) and provide insights into the stability of the perovskite structure as the size of each ion is varied. The perovskite structure requires that the A and B cations occupy distinct sites in the ABX_3 lattice, with A 12-fold and B 6-fold coordinated by X . When r_A and r_B are too similar, nonperovskite lattices that have similarly coordinated A and B sites, such as cubic bixbyite, become preferred over the perovskite structure. Based on the construct of τ , as $r_A/r_B \rightarrow 1$, $\wp \rightarrow 0$, which arises from the $+x/\ln(x)$ ($x = r_A/r_B$) term, where $\lim_{x \rightarrow 1} \frac{x}{\ln(x)} = +\infty$ and larger values of τ lead to lower probabilities of forming perovskites. When $r_A = r_B$, τ is undefined, yet compounds where A and B have identical radii are rare and not expected to adopt perovskite structures ($t = 0.71$).

The octahedral term in τ (r_X/r_B) also manifests itself in the probability maps, particularly in the lower bound on r_B where perovskites are expected as r_X is varied. As r_X increases, r_B must similarly increase to enable the formation of stable BX_6 octahedra. This effect is noticeable when separately comparing compounds containing Cl^- (left), Br^- (center), and I^- (right) (bottom row of

Figure 3.4), where the range of allowed cation radii decreases as the anion radius increases. For $r_B \ll r_X$, r_X/r_B becomes large, which increases τ and therefore decreases the probability of stability in the perovskite structure. This accounts for the inability of small B -site ions to sufficiently separate X anions in BX_6 octahedra, where geometric arguments suggest that B is sufficiently large to form BX_6 octahedra only for $r_B/r_X > 0.414$. Because the cation radii ratios significantly affect the probability of perovskite, as discussed in the context of $x/\ln(x)$, r_X also has a significant indirect effect on the lower bound of r_A , which increases as r_X increases.

The role of n_A in τ is more difficult to parse, but its placement dictates two effects on stability – as A is more oxidized (increasing n_A), $-n_A^2$ increases the probability of forming the perovskite structure, but n_A also magnifies the effect of the $x/\ln(x)$ term, increasing the importance of the cation radii ratio. Notably, $n_A = 1$ for most halides and some oxides (245 of the 576 compounds in our set) and in these cases, $\tau = \frac{r_X}{r_B} + \frac{r_A/r_B}{\ln(r_A/r_B)} - 1$ for all combinations of A , B , and X and n_A plays no role as the composition is varied.

This analysis illustrates how data-driven approaches can be used to not only maximize the predictive accuracy of new descriptors, but can also be leveraged to understand the actuating mechanisms of a target property – in this case, perovskite stability. This attribute distinguishes τ from other descriptors for perovskite stability that have emerged in recent years. For instance, three recent works have shown that the experimental formability of perovskite oxides and halides can be separately predicted with high accuracy using kernel support vector machines,⁸⁵ gradient boosted decision trees,⁸⁴ or a random forest of decision trees.⁸⁶ While these approaches can yield highly accurate models, the resulting descriptors are not documented analytically, and therefore, the mechanism by which they make the perovskite/nonperovskite classification is opaque.

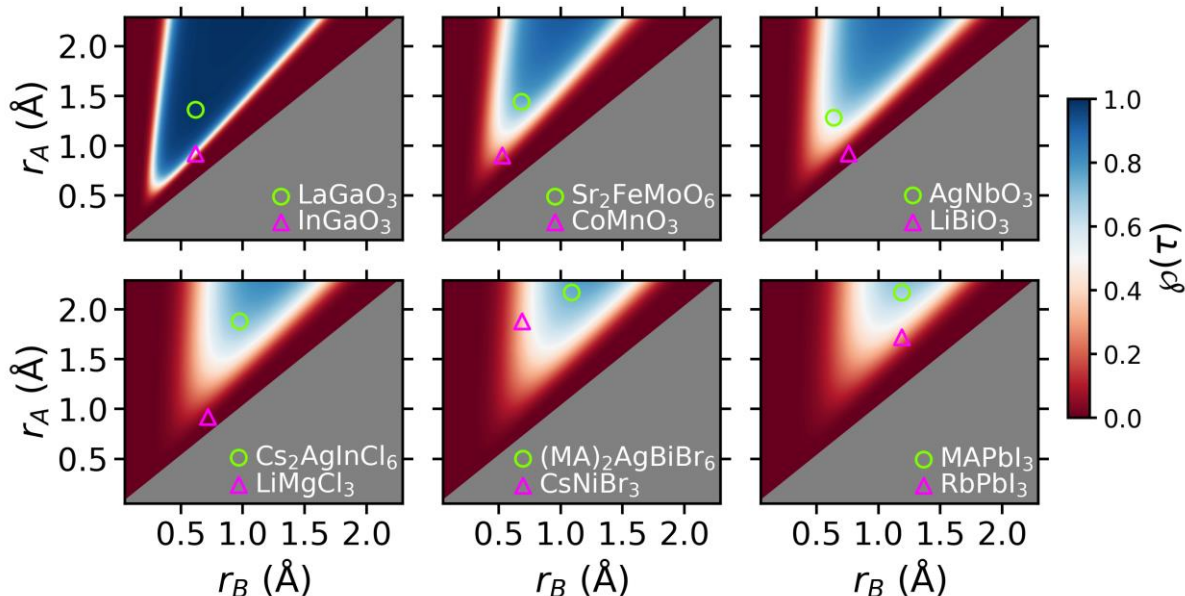


Figure 3.4 The effects of ionic radii and oxidation states on the stability of single and double perovskite oxides and halides

Top row: $X = O^{2-}$ (left to right $n_A = 3^+, 2^+, 1^+$). *Bottom row:* $n_A = 1^+$ (left to right $X = Cl^-, Br^-, I^-$). The experimentally realized perovskites $LaGaO_3$, Sr_2FeMoO_6 , $AgNbO_3$, $Cs_2AgInCl_6$, $(MA)_2AgBiBr_6$, and $MAPbI_3$ are shown as open circles in the corresponding plot, which are all predicted to be stable by τ . The experimentally realized nonperovskites $InGaO_3$, $CoMnO_3$, $LiBiO_3$, $LiMgCl_3$, $CsNiBr_3$, and $RbPbI_3$ are shown as open triangles and predicted to be unstable in the perovskite structure by τ . The organic molecule, methylammonium (MA), is shown in the last two panels. While $(MA)_2AgBiBr_6$ and $MAPbI_3$ are correctly classified with τ , only inorganic cations were used for descriptor identification; therefore, $r_A = 1.88 \text{ \AA}$ (Cs^+) is the largest cation considered. The gray region where $r_B > r_A$ is not classified because when this occurs, A becomes B and vice versa based on our selection rule $r_A > r_B$.

3.4 Discussion

We report a new tolerance factor, τ , that enables the prediction of experimentally observed perovskite stability significantly better than the widely used Goldschmidt tolerance factor, t , and the two-dimensional structure map using t and the octahedral factor, μ . For 576 ABX_3 and 918 $A_2BB'X_6$ compounds, the prediction by τ agrees with the experimentally observed stability for > 90% of compounds, with > 1,000 of these compounds reserved for testing generalizability (prediction accuracy). The deficiency of t arises from its functional form and not the input features

as the calculation of τ requires the same inputs as t (composition, oxidation states, and Shannon ionic radii). Thus, τ enables a superior prediction of perovskite stability with negligible computational cost. The monotonic and one-dimensional nature of τ allows for the determination of perovskite probability as a continuous function of the radii and oxidation states of A , B , and X . These probabilities are shown to linearly correlate with DFT-computed decomposition enthalpies and help clarify how chemical substitutions at each of the sites modulate the tendency for perovskite formation. Using τ , we predict the probability of double perovskite formation for thousands of unexplored compounds, resulting in a library of stable perovskites ordered by their likelihood of forming perovskites. Due to the simplicity and accuracy of τ , we expect its use to accelerate the discovery and design of state-of-the-art perovskite materials for applications ranging from photovoltaics to electrocatalysis.

3.5 Methods

3.5.1 Radii assignment

To develop a descriptor that takes as input the chemical composition and outputs a prediction of perovskite stability, the features that comprise the descriptor must also be based only on composition. Yet it is not known *a priori* which cation will occupy the A -site or B -site given only a chemical composition, $CC'X_3$ (C and C' being cations). To determine which cation is A or B , a list of allowed oxidation states (based on Shannon's radii⁷⁹) is defined for each cation. All pairs of oxidation states for C and C' that charge-balance X_3 are considered. If more than one charge-balanced pair exists, a single pair is chosen based on the electronegativity ratio of the two cations ($\chi_C/\chi_{C'}$). If $0.9 < \chi_C/\chi_{C'} < 1.1$, the pair that minimizes $|n_C - n_{C'}|$ is chosen, where n_C is the oxidation state for C . Otherwise, the pair that maximizes $|n_C - n_{C'}|$ is chosen. With the oxidation

states of C and C' assigned, the values of the Shannon radii for the cations occupying the A and B sites are chosen to be closest to the coordination number of twelve and six, which are consistent with the coordination environments of the A and B cations in the perovskite structure. Finally, the radii of the C and C' cations are compared and the larger cation is assigned as the A -site cation. This strategy reproduces the assignment of the A and B cations for 100% of 313 experimentally labeled perovskites.

3.5.2 Descriptor selection

For the selection of τ , the oxidation states (n_A, n_B, n_X), ionic radii (r_A, r_B, r_X), and radii ratios ($r_A/r_B, r_A/r_X, r_B/r_X$) comprise the primary features, Φ_0 , where Φ_n refers to the descriptor-space with n iterations of complexity as defined in ⁹⁰. For example, Φ_1 refers to the primary features (Φ_0) together with one iteration of algebraic/functional operations applied to each feature in Φ_0 . Φ_2 then refers to the application of algebraic/functional operations to all potential descriptors in Φ_1 , and so forth. Note that Φ_m contains all potential descriptors within $\Phi_{n < m}$ with a filter to remove redundant potential descriptors. For the discovery of τ , complexity up to Φ_3 is considered, yielding $\sim 3 \times 10^9$ potential descriptors. An alternative would be to exclude the radii ratios from Φ_0 and construct potential descriptors with complexity up to Φ_4 . However, given the minimal $\Phi_0 = [n_A, n_B, n_X, r_A, r_B, r_X]$, there are $\sim 1 \times 10^8$ potential descriptors in Φ_3 , so $\sim 1 \times 10^{16}$ potential descriptors would be expected in Φ_4 (based on $\sim 1 \times 10^2$ being present in Φ_1 and $\sim 1 \times 10^4$ in Φ_2), and this number is impractical to screen using available computing resources.

The dataset of 576 ABX_3 compositions was partitioned randomly into an 80% training set for identifying candidate descriptors and a 20% test set for analyzing the predictive ability of each descriptor. The top 100,000 potential descriptors most applicable to the perovskite classification problem were identified using one iteration of SISSO with a subspace size of 100,000. Each

descriptor in the set of $\sim 3 \times 10^9$ was ranked according to domain overlap, as described in Ouyang *et al.*⁹⁰ To identify a decision boundary for classification, a decision tree classifier with a max depth of two was fit to the top 100,000 candidate descriptors ranked based on domain overlap. Domain overlap (and not decision tree performance) is used as the SISO ranking metric because of the significantly lower computational expense associated with applying this metric. Notably, τ was the 14,467th highest ranked descriptor by SISO using the domain overlap metric and, as such, this defines the minimum subspace required to identify τ using this approach. Without evaluating a decision tree model for each descriptor in the set of $\sim 3 \times 10^9$ potential descriptors, we cannot be certain that a subspace size of 100,000 is sufficient to find the best descriptor. However, the identification of τ within a subspace as small as 15,000 suggests that a subspace size of 100,000 is sufficiently large to efficiently screen the much larger descriptor space. We have also conducted a test on this primary feature space ($\Phi_0 = [n_A, n_B, n_X, r_A, r_B, r_X, r_A/r_B, r_A/r_X, r_B/r_X]$) with a subspace size of 500,000. Even after increasing the subspace size by 5 $\times$, τ remains the highest performing descriptor (classification accuracy of 92% on the 576 compound set). An important distinction between the SISO approach described here and by Ouyang *et al.* in⁹⁰ is the choice of sparsifying operator (SO). In this work, domain overlap is used to *rank* the features in SISO, but a decision tree with max depth of two is used as the SO (instead of domain overlap) to *identify* the best descriptor of those selected by SISO. This alternative SO is used to decrease the leverage of individual data points as the experimental labeling of perovskite/nonperovskite is prone to some ambiguity based on synthesis conditions, defects, and other experimental considerations.

The benefit of including the radii ratios in Φ_0 is made clear by comparing the performance of τ to the best descriptor obtained using the minimal primary feature space with $\Phi_0 = [n_A, n_B, n_X,$

$r_A, r_B, r_X]$. Repeating the procedure used to identify τ yields a Φ_3 with $\sim 1 \times 10^8$ potential descriptors.

The best 1D descriptor was found to be $\frac{r_B}{n_X(r_A - r_B)} + \frac{r_B}{r_A} - \frac{r_X}{r_B}$ with classification accuracy of 89%.

3.5.3 Alternative features

We also consider the effects of including properties outside of those required to compute t or τ . Beginning with $\Phi_0 = [n_A, n_B, n_X, r_A, r_B, r_X, r_{cov,A}, r_{cov,B}, r_{cov,X}, IE_A, IE_B, IE_X, \chi_A, \chi_B, \chi_X]$, where $r_{cov,i}$ is the empirical covalent radius of neutral element i , IE_i is the empirical first ionization energy of neutral element i , and χ_i is the Pauling electronegativity of element i , all taken from WebElements (webelements.com), an aggregation of a number of references which are available within. Repeating the procedure used to identify τ results in $\sim 6 \times 10^{10}$ potential descriptors in Φ_3 .

The best performing 1D descriptor was found to be $\frac{r_A/r_B - \sqrt{\chi_X}}{r_{cov,X}/r_B - r_{cov,A}/r_{cov,X}}$ with classification accuracy of 90%, lower than τ which makes use of only the oxidation states and ionic radii, and only slightly higher than the accuracy of the descriptor obtained using the minimal feature set.

3.5.4 Increasing dimensionality

To assess the performance of descriptors with increased dimensionality, following the approach to higher dimensional descriptor identification using SISSO described in ⁹⁰, the residuals from classification by τ (those misclassified by the decision tree, **Figure 3.2b**) are used as the target property in the search for a second dimension to include with τ . From the same set of $\sim 3 \times 10^9$ potential descriptors constructed to identify τ , the 100,000 1D descriptors that best classify the 41 training set compounds misclassified by τ are identified based on domain overlap. Each of these 100,000 descriptors are paired with τ and the performance of each 2D descriptor was assessed using a decision tree with max depth of two. The best performing 2D descriptor was found to be

$\left(\tau, \frac{|r_A r_X / r_B^2 - n_B r_A / r_B|}{|r_A r_B / r_X^2 - r_A / r_B + n_B|}\right)$ with a classification accuracy of 95% on the 576 compound set.

Improvements are expected to diminish as the dimensionality increases further due to the iterative nature of SISSO and the higher order residuals used for subspace selection. Although the second dimension leads to slightly improved classification performance on the experimental set compared with τ , the simplicity and monotonicity of τ , which enables physical interpretation and the extraction of meaningful probabilities, support its selection instead of the more complex 2D descriptor. The benefits and capabilities of having a meaningfully probabilistic one-dimensional tolerance factor, such as τ , are described in detail within the main text.

3.5.5 Potential for overfitting

The SISSO algorithm as implemented here *selects* τ from a space of $\sim 3 \times 10^9$ candidate descriptors and the only parameter that is *fit* is the value of τ that defines the decision boundary for classification as perovskite or nonperovskite, $\tau = 4.18$. This decision boundary is optimized to maximize the classification accuracy on the training set of 460 compounds. The SISSO selection is done out of billions of candidates, but these functions are a discrete set, i.e., a basis in a large dimensional space (the number of training points is the dimension of the space), which is not densely covered by the basis functions. Therefore, the selection of only one function, τ , cannot overfit the data. However, if some physical mechanism determining the stability of perovskites is not represented in the training set, it might be missed by the learned formula (here, τ) and therefore the generalizability of the model would be hampered. However, the 94% accuracy achieved by τ on the excluded set of 116 compounds shows that τ can generalize outside of the training data.

3.5.6 Alternative radii for more covalent compounds

Ionic radii are required inputs for τ (and t) and although the Shannon effective ionic radii are ubiquitous in solid state materials research, a new set of B^{2+} radii were recently proposed for 18 cations to account for how their effective cationic radii vary as a function of increased covalency with the heavier halides.⁷⁷ These revised radii apply to 129 of the 576 experimentally characterized compounds compiled in this dataset (62% of halides). Employing these revised radii results in a 5% decrease in the accuracy of τ to 86% for these 129 compounds compared to a classification accuracy of 91% using the Shannon radii for these same compounds. The application of τ using Shannon radii for presumably covalent compounds is further validated by noting that τ correctly classifies 37/40 compounds which contain Sn or Pb and achieves an accuracy of 91% for 141 compounds with $X = \text{Cl}^-$, Br^- , or I^- . In addition to the higher accuracy achieved by τ when using Shannon radii, we note that the Shannon radii are more comprehensive than these revised radii in⁷⁷, applying to more ions, oxidation states, and coordination environments and are thus recommended for the calculation of τ .

3.5.7 Computer packages used

SISSO was performed using *Fortran 90*. Platt’s scaling⁹¹ was used to extract classification probabilities for τ by fitting a logistic regression model on the decision tree classifications using 3-fold cross-validation. Decision tree fitting and Platt scaling were performed within the *Python* package, *scikit-learn*. Data visualizations were generated within the *Python* packages *matplotlib* and *seaborn*. An implementation of τ is available at github.com/CJBartel/perovskite-stability.

4 Computational screening for double perovskite solar absorbers: a case-study of cesium chlorides

Much of the content in this chapter appears verbatim in the following manuscript:

C.J. Bartel, C. Sutton, B.R. Goldsmith, A.W. Weimer, A.M. Holder, C.B. Musgrave (2018) [In Preparation]

4.1 Abstract

Double perovskite halides, $A_2BB'X_6$, composed of cations A , B and B' and halide X have emerged as lead-free candidates for hybrid perovskite solar absorbers, benefiting from their ability to mimic the effects of lead in lead-containing perovskites through cationic substitutions on the octahedrally coordinated B -site. In this work, we explore the stability and optical properties of 352 potential cesium chloride double perovskites ($Cs_2BB'Cl_6$) using density functional theory. From the results of this analysis, we identify 22 new cesium chloride perovskites that our calculations predict to be stable and have electronic properties suitable for solar absorption. We also emphasize that consideration of decomposition into likely precursors and distortions from the commonly assumed rock salt structure are critical to correctly predicting double perovskite stability and electronic properties.

4.2 Introduction

Double perovskites ($A_2BB'X_6$) have emerged as high-performance materials for a number of applications including electrocatalysts,⁷⁰ ferroelectrics¹⁰⁰ and as lead-free and all-inorganic alternatives to hybrid organic-inorganic lead-halide based solar absorbers.^{87, 98, 101-103} The design and discovery of functional materials is being rapidly accelerated by high-throughput screening approaches based on density functional theory (DFT).^{104, 105} The number of potential compositions

that could crystallize into the double perovskite structure, however, is immense – considering 50 realistic elements that could occupy the *A*- or *B*-site combined with only the four halogens on the *X*-site results in 78,400 $A_2BB'X_6$ formulas. Furthermore, it is not known *a priori* which of these materials will be stable and if so, what crystal structures they will adopt. The large space of possible formulas and structures precludes the use of high-throughput DFT to screen double perovskites for stability and consequently for their suitability for a specific application based on their properties in their stable form. This dilemma encouraged us to explore the use of a descriptor-based approach to predict stability and to thus considerably reduce the space of viable candidate materials to be analyzed using DFT to a more manageable size. Recently, we applied the compressed sensing-based algorithm, SISSO (sure independence screening and sparsifying operator),⁹⁰ to identify a simple descriptor, τ , that requires only the chemical formula as input and correctly classifies over 90% of known single and double perovskite oxides and halides as perovskite or nonperovskite when compared with experiment.¹⁰⁶

$$\tau = \frac{r_X}{r_B} - n_A \left(n_A - \frac{r_A/r_B}{\ln(r_A/r_B)} \right) \quad (4.1)$$

In this work, we computationally prototyped 352 $A_2BB'X_6$ materials with Cs on the *A*-site and Cl on the *X*-site ($Cs_2BB'Cl_6$). Recent synthesis and characterization of several interesting double perovskites in this chemical space, e.g., $Cs_2AgInCl_6$,¹⁰¹ $Cs_2AgBiCl_6$,⁹⁸ $Cs_2AgTlCl_6$,¹⁰³ and $Cs_2AgSbCl_6$ ¹⁰⁷ allows us to analyze our results in the context of experimental observations. Notably, all of these synthesized cesium chloride double perovskites contain Ag, yet τ indicates that 321 *B/B'* combinations not involving Ag have a probability of forming a stable perovskite > 0.5, in addition to 31 more with Ag (**Figure 4.1**). To assess the viability of these 352 τ -stable materials to absorb light, we determined the stability of each compound with respect to structural

distortions and decomposition into likely precursor phases using the recently developed meta-GGA exchange-correlation functional SCAN⁹ and the electronic structure of each using the hybrid HSE exchange-correlation functional with 25% exact exchange (HSE06).¹⁰⁸ Our results reveal the necessity of considering structural distortions away from the ideal cubic perovskite structure in order to accurately predict stability and optical properties. Screening based on these additional considerations predicts that significant opportunities exist for experimentalists to synthesize not yet realized all-inorganic cesium chloride double perovskite solar absorbers with promising properties.

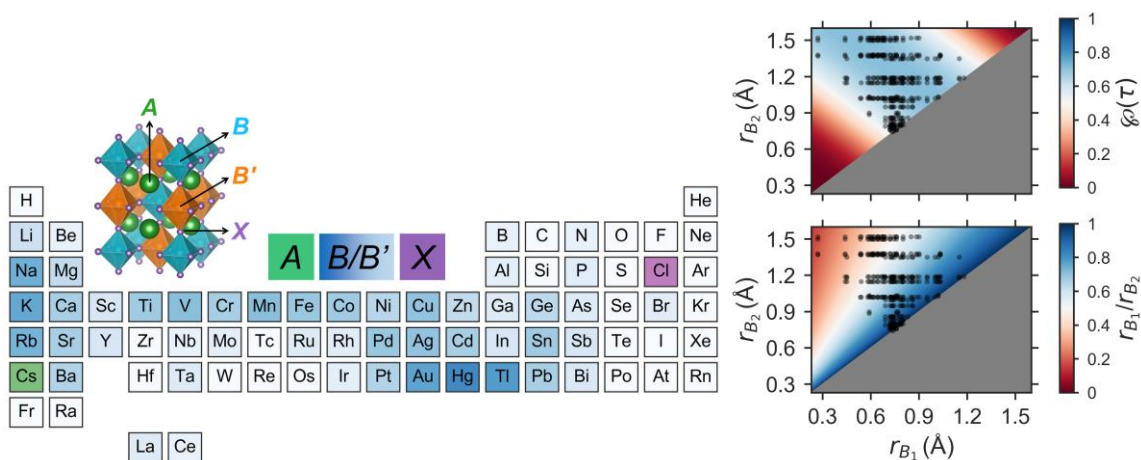


Figure 4.1 Cesium chloride double perovskite space

Left – Rock salt double perovskite crystal structure along with the elements that occupy the A-, B-, and X-sites of double perovskites in this study. The intensity of the blue shading that highlights the considered B/B' sites indicates their prevalence within the dataset (e.g., Hg appears most often – 36 times). Note: B/B' sites are equivalent in the rock salt arrangement. Right – Mapping the space of compounds explored in this work in terms of the B-site radii, r , the tau-derived probability, ϕ , and the B-site cation radii mismatch. Points indicate compositions studied within this work.

4.3 Results

4.3.1 Stability analysis

Prior computational efforts to identify double perovskite solar absorbers have either neglected the calculation of stability with respect to decomposition into competing phases¹⁰⁹⁻¹¹¹ or only focused on a small (~ 10 -100) set of materials.^{87, 112-114} As a first approximation for stability prediction, it is essential to perform a convex hull analysis with respect to all possible (or at least, likely) competing compounds.³⁹ A calculation of the band gap energy using higher levels of theory could then be performed for only the identified (meta)stable compounds, greatly reducing the computational cost of this analysis. In this work, we calculated the decomposition enthalpy, ΔH_d , of 352 $\text{Cs}_2\text{BB}'\text{Cl}_6$ compounds with respect to all competing Cl-containing compounds in the Materials Project database.¹⁶ That is, we compared the calculated formation enthalpy, ΔH_f , of each $\text{Cs}_2\text{BB}'\text{Cl}_6$ structure to the linear combination of calculated ΔH_f 's for all available structures in the Cs, B, B', Cl, Cs-Cl, B-Cl, B'-Cl, Cs-B-Cl, Cs-B'-Cl, and Cs-B-B'-Cl competing chemical spaces that have the same average formula – $\text{Cs}_2\text{BB}'\text{Cl}_6$ – and the most negative combined enthalpy, $\Delta H_{f,\text{competing}}$. ΔH_d was then obtained from:

$$\Delta H_d = \Delta H_{f,\text{Cs}_2\text{BB}'\text{Cl}_6} - \Delta H_{f,\text{competing}} \quad (4.2)$$

A comparison of ΔH_d to ΔH_f is shown in **Figure 4.2**. Out of the 352 $\text{Cs}_2\text{BB}'\text{Cl}_6$ compounds analyzed, 115 are calculated to be stable with respect to competing compounds ($\Delta H_d < 0$, blue), and 152 are calculated to lie within 0.05 eV/atom above the convex hull ($\Delta H_d < +0.05$ eV/atom, purple), which are defined here as potentially metastable³⁷ or stable within the typical errors associated with the SCAN functional.⁵² This analysis predicts that the 13 known $\text{Cs}_2\text{BB}'\text{Cl}_6$

compounds in our dataset that are also reported in the ICSD to have $\Delta H_d < 0$, thus correctly predicting that they should exist as stable perovskites. Furthermore, our analysis also predicts the four recently synthesized Ag-containing double perovskites – $\text{Cs}_2\text{AgInCl}_6$, $\text{Cs}_2\text{AgBiCl}_6$, $\text{Cs}_2\text{AgSbCl}_6$, and $\text{Cs}_2\text{AgTlCl}_6$ – to be stable ($\Delta H_d < 0$).

Each of the compounds shown in **Figure 4.2** was predicted to be perovskite by τ yet 23% of the compounds considered are calculated to be unstable using SCAN ($\Delta H_d > +0.05$ eV/atom). Extending the tolerance from 0.05 eV/atom to 0.1 eV/atom reduces the discrepancy to only 7%. The disagreement between the two approaches is likely due in part to the dependence of τ on only three radii terms – r_A , $r_{B,\text{eff}}$, and r_X – but there being four relevant radii in double perovskites. Applying τ for a stability analysis of double perovskites requires the averaging of r_B and $r_{B'}$ into an effective cationic radius ($r_{B,\text{eff}}$), which becomes a significant approximation as these two B -site radii differ considerably from one another (**Figure 4.1**). It should also be noted that τ is probabilistic and indicates that ~65% of the 352 materials analyzed should be experimentally realizable as perovskite. This prediction generally agrees with the calculations based on ΔH_d , while we note that there is also some inherent disagreement between DFT-calculated stabilities and experimentally realized synthesis.

Comparing ΔH_d to ΔH_f (**Figure 4.2**) reveals the critical effect of not only optimizing the perovskite structure within DFT but also computing the phase diagram with respect to competing compounds (e.g., precursor phases). This comparison demonstrates that ΔH_d has effectively no correlation with ΔH_f and that both strongly and weakly negative ΔH_f compounds span the range of stable, metastable, to unstable when compared with competing compounds using ΔH_d . Therefore, the comparison of a given material’s enthalpy only to the enthalpy of its constituent

elements (ΔH_f) and not to all competing materials (ΔH_d) drastically overestimates the stability of that material.

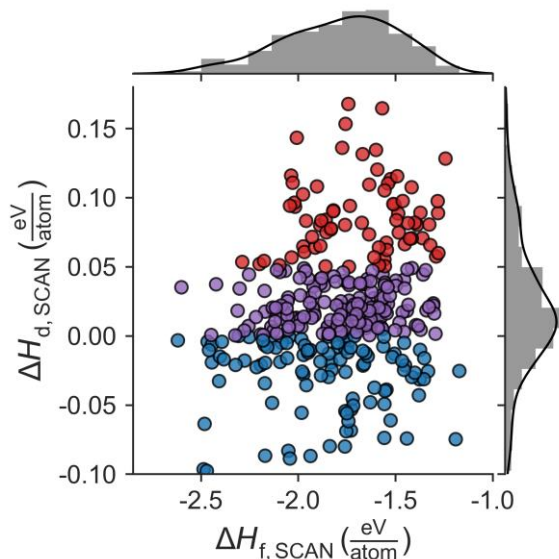


Figure 4.2 Stability analysis conducted with the SCAN DFT functional

Comparing decomposition enthalpy with formation enthalpy for 352 $\text{Cs}_2\text{BB}'\text{Cl}_6$ compounds. Stability is computed with respect to Cl-containing compounds. Blue indicates stable ($\Delta H_d < 0$), purple indicates metastable ($0 \leq \Delta H_d < 0.05$) and red indicates unstable ($\Delta H_d \geq 0.05$ eV/atom).

4.3.2 Structural distortions

In addition to inadvisably using ΔH_f rather than ΔH_d as the metric for materials stability, another approximation that is often used for high-throughput screening is the presumption of crystal symmetry of a preassigned spacegroup.⁸⁷ For instance, the high-symmetry rock salt double perovskite structure might be optimized in a symmetric way to accelerate the DFT determination of structure. This symmetry-constrained optimization precludes the ability of the ions to relax out of their preassigned Wyckoff positions, thus constraining the geometry to the ideal cubic double perovskite symmetry. Although this strategy may seem suitable for searching for double perovskite solar absorbers, it has not been thoroughly benchmarked. Here, we compare the symmetry-constrained structures (*rs*) to those obtained by subjecting the initially symmetric structure to

random atomic displacements of 0.1 Å (*opt*). This symmetry-breaking operation to yield more stable structures has considerable influence not only on the calculated stability, but also on the band gap (**Figure 4.3**).

The stability analyses based on energies calculated for either *rs* or *opt* structures results in a consistent determination for the majority of materials as being stable, metastable, and unstable (**Figure 4.3a**, on or near the diagonal line). This suggests that most structures do not distort significantly when their symmetry is allowed to break or if their structures do distort, the associated energy changes are small (**Figure 4.3a**). We do find, however, that 32 symmetry-constrained structures that fell outside our allowed tolerance for metastability ($\Delta H_{d,rs} > 0.05$) move into this range ($\Delta H_{d,opt} < 0.05$) upon distortion to the fully optimized structure and that in fact 7 of these structures become thermodynamically stable ($\Delta H_{d,opt} < 0$) upon distortion. Therefore, the common stability analysis based on symmetry constrained structures *rs* incorrectly reduces the space of potentially synthesizable $\text{Cs}_2\text{BB}'\text{Cl}_6$ by ~15%.

The critical property of interest for identifying potential solar absorber materials is the band gap, E_g . In **Figure 4.3b** we show a comparison of E_g resulting from the *rs* and *opt* approaches, where 50 structures that are metallic ($E_g = 0$) in their high-symmetry cubic form (*rs*) become semiconducting or insulating ($E_g > 0$) when fully optimized, with E_g becoming as large as 3.4 eV for the optimized structure. For screening purposes, we might consider materials that are stable or metastable and have $0.2 < E_g < 2$ eV in SCAN. This criteria yields 127 structures using the more rigorous *opt* approach, while only 91 of these would have been discovered using the less expensive *rs* approach. These results highlight the necessity of breaking symmetry and performing a full structural and electronic optimization when screening materials for various applications where the critical property of interest is sensitive to the crystal structure.

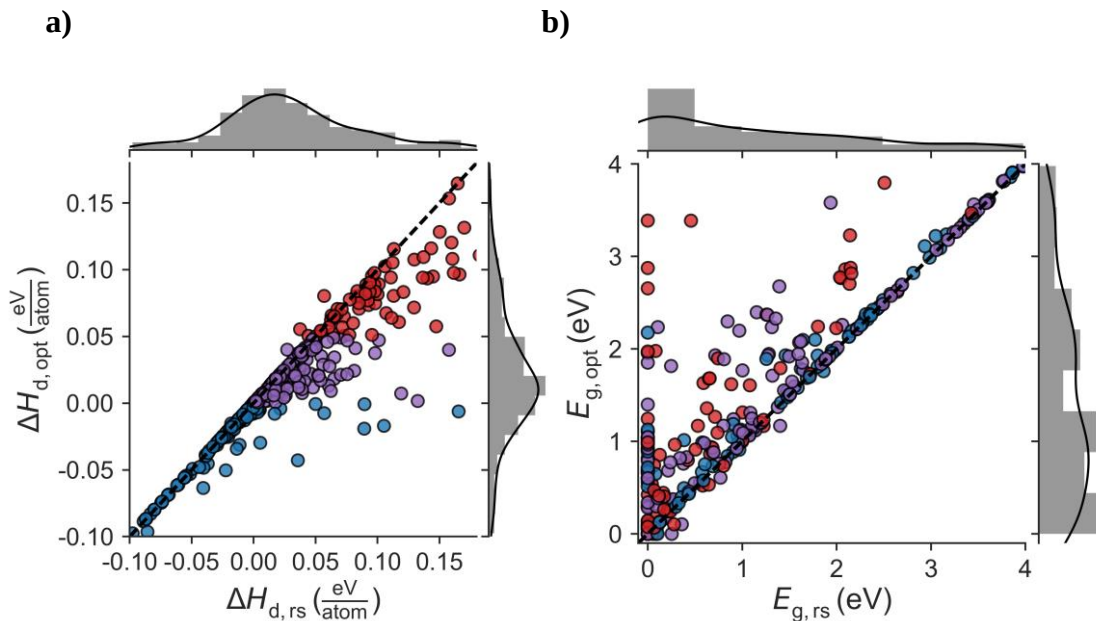


Figure 4.3 Structural distortion effects on stability and band gap

a) Comparison of the SCAN-computed decomposition enthalpy of the fully optimized structure, *opt*, to that of the preassigned rock salt structure, *rs*, and **b)** the band gaps of the *opt* and *rs* structures. The coloring scheme is the same as in **Figure 4.2**.

Lattice distortions away from the ideal *rs* structure have significant effects on stability predictions and fundamentally change the nature of the electronic band structure, yet little is known about the detailed effects of these distortions. To delve further into these effects we explored the tetragonality of the lower symmetry *opt* structures. In perovskites, elongation of the BX_6 octahedra in one direction (*c*) relative to another (*a*) is termed a tetragonal distortion from the ideal ratio of $c/a = 1$ (**Figure 4.4a**). Comparing the difference in calculated E_g between *opt* and *rs* ($\Delta E_{g,opt}$), shows that tetragonality does play some role in controlling the electronic structure (**Figure 4.4b**). As one example, this analysis highlights the distortion of $Cs_2CuInCl_6$, which maintains only slightly elongated $CuCl_6$ and $InCl_6$ octahedra in the *rs* structure (**Figure 4.4c**) but adopts a nonperovskite structure in which Cu is no longer octahedrally coordinated, but instead adopts $CuCl_3$ tetrahedral coordination when perturbations are applied to obtain the *opt* structure (**Figure**

4.4d). The tetragonality ratio, c/a , for CuCl_6 changes from 1.07 to 1.56 upon relaxation to the fully optimized structure, indicating the extreme elongation of the Cu-Cl bonds in the c direction. Although this distortion results in only a 0.007 eV/atom lowering of the energy, it changes the nature of the material from metallic in rs to semiconducting in opt with a gap of $E_g = 1$ eV (using SCAN; in HSE, E_g increases from 1 eV to 2.2 eV). Notably, a large number of structures exist with nearly the ideal ratio for tetragonality but which exhibit a wide variation in E_g between opt and rs , further indicating the important role structure plays in dictating a perovskites electronic properties that should be addressed by more rigorous structural and electronic characterization studies.

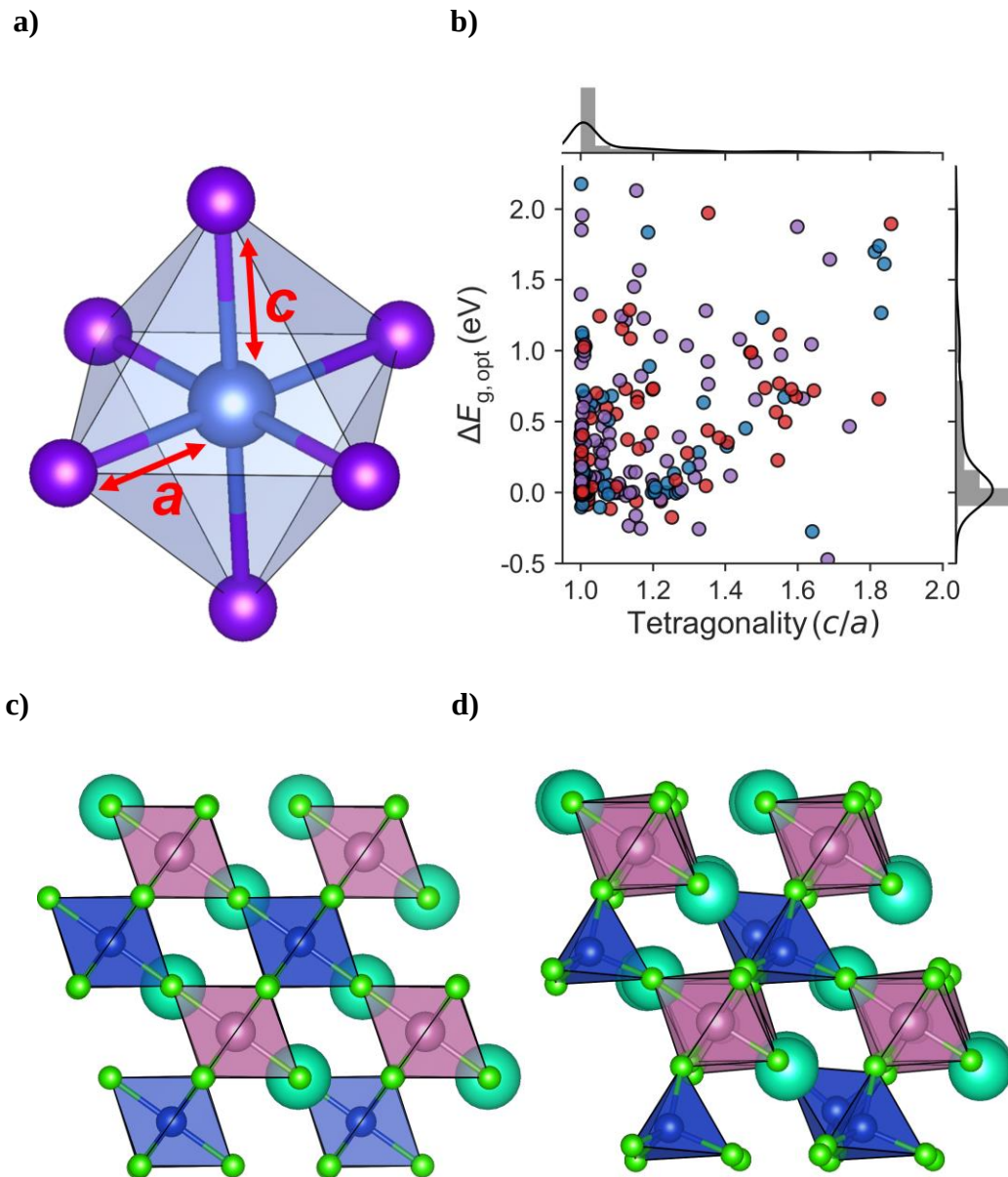


Figure 4.4 Octahedral distortions of optimized structures

a) BX_6 octahedral unit where the blue B -site is 6-fold coordinated by X . In the undistorted case rs , the four $B-X$ bonds are in-plane in the a direction and have the same length while the two $B-X$ bonds are out of this plane (c). If the in-plane and out-of-plane bonds are all the same length, the tetragonality (c/a) is 1. **b)** A comparison of the difference in band gap between opt and rs structures to the tetragonality in the opt structure. **c)** The rs geometry and **d)** the opt geometry of $Cs_2CuInCl_6$. In, Cu, Cs and Cl are shown as pink, blue, teal, and green spheres, respectively.

4.3.3 Candidate solar absorbers

The focus of this effort was to identify potential solar absorbers, making accurate calculation of the magnitude and nature of E_g paramount – only structures that absorb light in the visible spectrum and which have direct or nearly direct gaps are suitable for efficient solar absorption. Our overall goal is to rapidly identify promising solar absorbers using computation to recommend materials for experimental synthesis and characterization while recommending a minimum number of false positives as practically possible. Consequently, because SCAN underestimates E_g , we used HSE06 to calculate band gaps because it predicts gaps that agree well with experiment.¹¹⁵ Because DFT calculations using the HSE06 hybrid functional are computationally demanding, we restricted our analysis to a subset of materials that satisfy the following criteria: 1) $\Delta H_{d,SCAN} < 0.05$ eV/atom, 2) $E_{g,SCAN} < 2$ eV, and 3) direct gap within 0.5 eV of the indirect gap (with SCAN). This screening strategy reduces the set of 352 materials to only viable 99 candidates, 22 of which are identified to have $0.5 < E_g < 2$ eV using HSE – AgAu, AgBr, AgCu, AuCu, AuHg, AuMn, CeHg, CrPb, CuNa, CuRb, HgMn, HgPt, HgTl, KMn, KNi, KPd, MnNa, MnRb, NaNb, NaNi, NaPd, and NiRb (BB' pairs) (**Figure 4.5**). These materials have a high probability of being synthesized and of efficiently absorbing light in the visible spectrum. Furthermore, several are also comprised of only non-toxic earth-abundant elements, and none have thus far been reported in the literature. All-inorganic double perovskite halides are attractive compared to their hybrid counterparts due to improved stability of the inorganic cations. However, the inorganic materials typically have wider band gaps than are preferred for solar absorption.¹⁰¹ Here, we show a number of low-bandgap (< 2 eV) all-inorganic double perovskites that are expected to be stable. These materials span the range of band gaps that are suitable for top- (~ 1.8 eV) or bottom-cells (~ 1.2 eV) in tandem solar cell architectures.¹¹⁶

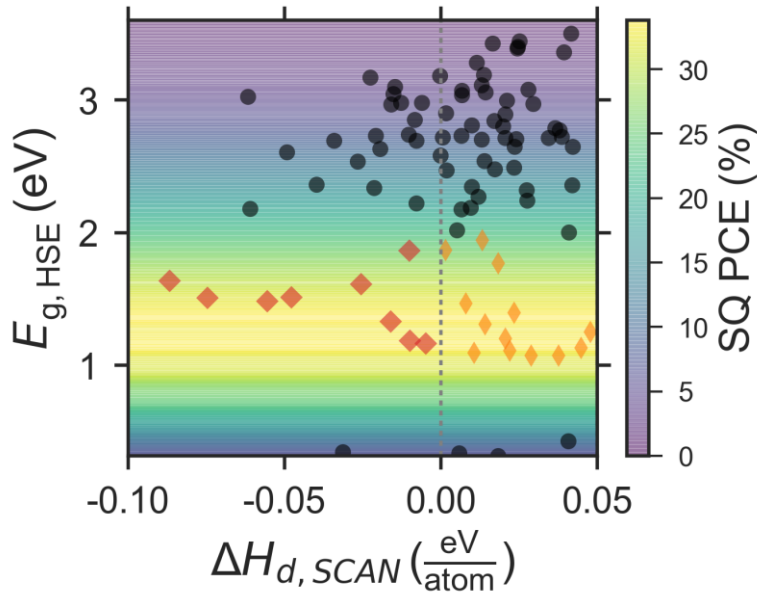


Figure 4.5 Refined band gap and stability of top candidates

Comparing the HSE band gap to the SCAN-computed stability with respect to the Shockley-Queisser limit for power conversion efficiency (SQ PCE)¹¹⁷.

4.4 Conclusions

In this work, we computed the stability of 352 cesium chloride double perovskites with respect to decomposition into potential decomposition products using the SCAN functional. This set was selected rationally using the recently introduced tolerance factor for perovskite stability, τ . We generally found good agreement between the SCAN-computed stability and τ , corroborating the ability of this descriptor to correctly analyze stability at a minimal computational cost while also providing some confidence that SCAN predicts experimental stability well, given that τ was trained on experimentally observed structures. We also tested the commonly used approximation of pre-assigning the perovskite (double perovskite in this case) structure to be rock salt and found that this assumption led to inaccurate predictions of the stability and band gaps of the less symmetric but more stable structures. By randomly perturbing the rock salt structure, we showed

that substantial quantitative differences in stability and optical properties arose between the cubic and less symmetric structures, some of which are attributable to an increased distortion of the octahedral units of the perovskite structure. Finally, we computed the band gaps of 99 potentially viable candidate solar absorbers using the more accurate HSE06 hybrid functional and report 22 new double perovskite chlorides with band gaps between 0.5 and 2 eV and SCAN-calculated decomposition enthalpies < 0.05 eV/atom.

4.5 Methods

Considering 1,228 $\text{Cs}_2\text{BB}'\text{Cl}_6$ formulas resulting from 48 *B*-site cations, τ indicates that 352 have a high probability (> 0.5) of forming stable perovskites. Each of these 352 structures were first optimized within the preassigned rock salt symmetry using the SCAN DFT functional and the projector-augmented wave method, as implemented in VASP.^{12, 58, 59} The resulting geometry was then expanded into a $1\times 1\times 2$ supercell and subjected to random atomic displacements of 0.1 Å using pymatgen.¹¹⁸ This structure was then re-optimized using SCAN and the stability was computed with respect to all relevant decomposition products in the Materials Project database,¹⁶ each of which was also reoptimized with SCAN. 99 materials with $\Delta H_d < 0.05$ eV/atom, $E_{g,\text{SCAN}} < 2$ eV, and having a direct gap within 0.5 eV of the indirect gap were also computed using the HSE06 functional at the fixed SCAN geometry. All structures were optimized with a plane wave energy cutoff of 520 eV and a Γ -centered Monkhorst-Pack k-point grid with $N = 20|b_i|$ discretizations along each reciprocal lattice vector, b_i . The energy cutoff, k-point density, and related convergence settings were sufficient to achieve total energy convergence of < 5 meV/atom for all calculations.

5 Physical descriptor for the Gibbs energy of inorganic crystalline solids and temperature-dependent materials chemistry

Much of the content in this chapter appears verbatim in the following manuscript:

C.J. Bartel, S.L. Millican, A.M. Deml, J.R. Rumpitz, W. Tumas, A.W. Weimer, S. Lany, V. Stevanović, C.B. Musgrave, A.M. Holder, *Nature Communications*, 9, 4168 (2018)

5.1 Abstract

The Gibbs energy, G , determines the equilibrium conditions of chemical reactions and materials stability. Despite this fundamental and ubiquitous role, G has been tabulated for only a small fraction of known inorganic compounds, impeding a comprehensive perspective on the effects of temperature and composition on materials stability and synthesizability. Here, we use the SISO (sure independence screening and sparsifying operator) approach to identify a simple and accurate descriptor to predict G for stoichiometric inorganic compounds with ~ 50 meV atom⁻¹ (~ 1 kcal mol⁻¹) resolution, and with minimal computational cost, for temperatures ranging from 300-1800 K. We then apply this descriptor to $\sim 30,000$ known materials curated from the Inorganic Crystal Structure Database (ICSD). Using the resulting predicted thermochemical data, we generate thousands of temperature-dependent phase diagrams to provide insights into the effects of temperature and composition on materials synthesizability and stability and to establish the temperature-dependent scale of metastability for inorganic compounds.

5.2 Introduction

The progression of technology throughout history has been preceded by the discovery and development of new materials.¹¹⁹ While the number of possible materials and the variety of their

properties is virtually limitless, discovery of new compounds with superior properties that are also stable (or persistently metastable) and synthesizable is a tremendous undertaking that remains as an ongoing challenge to the materials science community.¹²⁰⁻¹²³ The leading paradigm in this effort is the use of first-principles computational methods, such as density functional theory (DFT), and materials informatics to rapidly populate, augment and analyze computational materials databases and screen candidate materials for target properties.^{104, 105} However, despite the exploding growth of these databases with the number of compiled entries currently exceeding 50 million,¹⁹ only a small fraction of realized or potential materials have known Gibbs energies of formation, $\Delta G_f(T)$, which is critical for predicting the synthesizability and stability of materials at conditions of interest for numerous applications which operate at elevated temperature including thermoelectrics,¹²⁴ ceramic fuel cells,¹²⁵ solar thermochemical redox processes,¹²⁶ and CO₂ capture.¹²⁷

Experimental approaches for obtaining $\Delta G_f(T)$ are demanding, and the number of researchers using calorimetry to determine $\Delta G_f(T)$ is significantly smaller than those focused on the discovery and synthesis of new materials. *Ab initio* computational approaches for determining $\Delta G_f(T)$, which involve calculating the vibrational contribution to $G(T)$ as a function of volume,¹²⁸ have benefited from recent advances that reduce their computational cost.^{129, 130} However, despite these advances, calculating the vibrational entropy of phonons quantum mechanically is still computationally demanding, with computed $G(T)$ available for fewer than 200 compounds in the Phonon database at Kyoto University (PhononDB).¹³¹ Highly populated and widely used materials databases currently tabulate 0 or 298 K enthalpies of formation, ΔH_f , which neglect the effects of temperature and entropy on stability. As a result, the growth of computational materials databases has far outpaced the tabulation of measured or computed $\Delta G_f(T)$ of materials, precluding

researchers from obtaining a comprehensive understanding of the stability of inorganic compounds.

The use of machine learning and data analytics to accelerate materials design and discovery through descriptor-based property prediction is becoming a standard approach in materials science,^{27, 106, 132-137} however, these techniques have not previously been used to predict the Gibbs energies of inorganic crystalline solids. Techniques based on symbolic regression have also shown that fundamental physics can be algorithmically obtained from experimental and computed data in the form of optimized analytical expressions of intrinsic properties (features).¹³⁸⁻¹⁴⁰ In this work, we apply a recently developed statistical learning approach, SISSO (sure independence screening and sparsifying operator)¹⁴¹, to search a massive ($\sim 10^{10}$) space of mathematical expressions and identify a descriptor for experimentally obtained $G(T)$ that for the first time enables $\Delta G_f(T)$ to be readily obtained from high-throughput DFT calculations of a single structure (i.e., a single unit cell volume). The descriptor is identified using experimental data⁴⁷ for 262 solid compounds and tested on a randomly chosen excluded set of 47 compounds with measured $G(T)$ and 131 compounds with first-principles computed¹³¹ $G(T)$. We then apply this descriptor to $\sim 30,000$ unique crystalline solids tabulated in the Inorganic Crystal Structure Database (ICSD) to generate the most comprehensive thermochemical data of inorganic materials to date.

5.3 Results

5.3.1 Trends in the Gibbs energies of compounds and elements

Despite the variations of composition and structure exhibited by different inorganic crystalline compounds, $G(T)$ behaves remarkably similarly over a wide range of materials (**Figure 5.1a**). This similarity prompts the hypothesis that although the underlying physical phenomena

that give rise to $G(T)$ are complex to describe individually, a physically motivated descriptor could be predictive. The origin of the similar behavior of $G(T)$ can be understood from well-known thermodynamic relations, specifically that $\left(\frac{\partial G}{\partial T}\right)_P = -S \leq 0$ for mechanically stable compounds and that $G(T)$ must have negative concavity: $\left(\frac{\partial^2 G}{\partial T^2}\right)_P = -\left(\frac{\partial S}{\partial T}\right)_P = -\frac{C_P}{T} \leq 0$. Indeed, the negative first and second derivatives of experimental Gibbs energies as a function of temperature persist across the composition space of a diverse set of mechanically stable stoichiometric solid compounds (**Figure 5.1a**). We reference the Gibbs energy, G , with respect to the formation enthalpy at 298 K, ΔH_f , because ΔH_f is readily obtained using existing high throughput computational methods – DFT total energy calculations and a suitable correction for the elemental phases:^{29, 30, 33, 35}

$$G^\delta(T) = G(T) - \Delta H_f(298 \text{ K}) \quad (5.1)$$

As expected, the temperature- and material-dependence of the enthalpic contribution to the Gibbs energy, G^δ , is small relative to the entropic contribution (TS). If the standard state formation enthalpy, ΔH_f , is known, the temperature-dependence of the enthalpy is reliably predicted with a simple linear fit for the 309 solid compounds considered in this work. This is assumed implicitly when the quasiharmonic approximation²² of the phonon free energy is used to obtain $G(T)$, but is quantified here across a broad composition and temperature space.

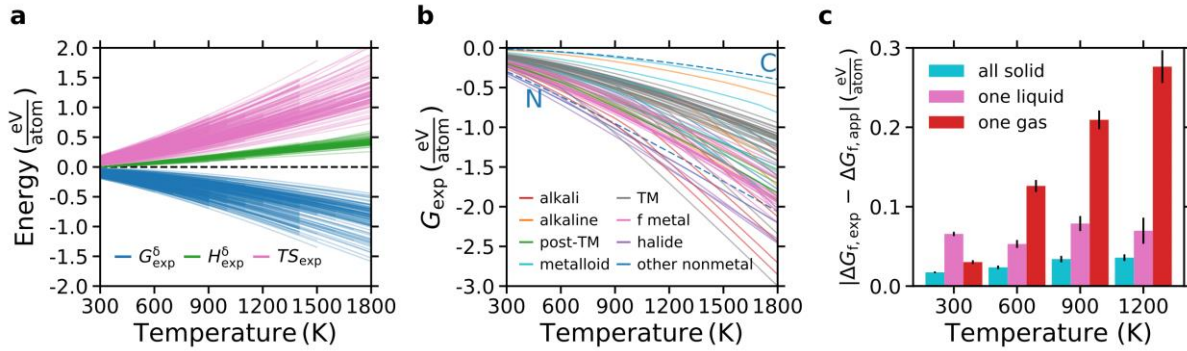


Figure 5.1 Contributions to the Gibbs energies of compounds

a) Experimentally obtained thermodynamic functions of 309 inorganic crystalline solid compounds obtained from FactSage. G^δ is defined in **Equation 5.1**. H^δ is the temperature-dependence of the enthalpy normalized to be zero at 298 K, S is the absolute entropy, and T is temperature. The subscript, *exp*, indicates the quantity is obtained from experimental data. **b)** Experimentally determined absolute Gibbs energies of 83 elements obtained from FactSage. G_C (“C”) and G_N (“N”) are dashed and labeled as they are mentioned in the text. The subscript, *exp*, indicates the quantity is obtained from experimental data. **c)** Mean absolute error in assuming a cancellation of solid vibrational entropy between the compound and the elements comprising it. $\Delta G_f(T)$ is defined in **Equation 5.3**. The subscript, *app*, stands for approximation and $\Delta G_{f,app}(T)$ is defined in **Equation 5.4**.

In addition to the thermodynamic quantities ΔH_f and $G^\delta(T)$, the chemical potentials of the elements $G_i(T)$ also play a critical role in the Gibbs formation energy, $\Delta G_f(T)$, and thus the temperature-dependent stability of a given compound:

$$\Delta G_f(T) = \Delta H_f(298 \text{ K}) + G^\delta(T) - \sum_{i=1}^N \alpha_i G_i(T) \quad (5.2)$$

where N is the number of elements in the compound, α_i is the stoichiometric weight of element i and G_i is the absolute Gibbs energy of element i . While even at low temperatures the differences in G_i between elements can be substantial (e.g., $G_C - G_N = 0.28 \text{ eV atom}^{-1}$ at 300 K), at higher temperatures, differences in G_i of $> 1 \text{ eV atom}^{-1}$ can result between solid and gaseous elements (e.g., $G_C - G_N = 1.12 \text{ eV atom}^{-1}$ at 1200 K, **Figure 5.1b**). In contrast to the elemental Gibbs energies, G_i , which are tabulated and thus require no computation or experiment to obtain, the

Gibbs energies of solid compounds, G^δ , are rarely tabulated and computationally demanding to calculate. Furthermore, assuming that all temperature-dependent effects can be captured by only including the elemental Gibbs energies and neglecting those of the solid compound results in an incomplete cancellation of errors and consequently inaccurate $\Delta G_f(T)$.

The temperature-dependence of the thermodynamic properties of solids have often been assumed to be negligible relative to that of gaseous species.¹⁴² That is, the Gibbs energy is generally assumed to be primarily entropic and principally due to vibrations such that the temperature-dependence of the formation energies of solids is negligible. We examined this assumption for hundreds of solid compounds by comparing the difference between the experimental $\Delta G_f(T)$ and the approximate $\Delta G_f(T)$ that results from assuming negligible temperature dependence of the solid phase:

$$\Delta G_{f,\text{app}}(T) = \Delta H_f(298 \text{ K}) - \sum_{i=1}^N \alpha_i G_{i,\text{gas}}(T) \quad (5.3)$$

Given a binary solid AB , if A and B are both solid at a given temperature, this assumption holds reasonably well and ΔH_f predicts $\Delta G_f(T)$ relatively accurately, e.g. with mean absolute errors of $\sim 50 \text{ meV atom}^{-1}$ at 900 K (**Figure 5.1c**). However, if either A or B are liquid at a given temperature, this error grows to $\sim 100 \text{ meV atom}^{-1}$ at 900 K. Even more alarming is the error produced by this approximation if either A or B are gaseous at T , as is the case for oxides, nitrides, halides, etc. with mean absolute errors for $\Delta G_f(T)$ of $\sim 200 \text{ meV atom}^{-1}$ at 900 K. In this approximation, the chemical potential, $G_i(T)$, of the gaseous element and the formation enthalpy, ΔH_f , of the solid compound are taken from experiment and thus the larger error arises entirely from the missing quantity $G^\delta(T)$. The larger error that arises when an element is a gas or liquid,

but not a solid, is due to the incomplete cancellation of the solid vibrational entropy of the elemental forms and the solid compound. That is, the distribution of phonon frequencies in the crystalline compound of A and B produce vibrational entropy S_{AB} and if A and B are elemental solids, they too have solid vibrational entropies S_A and S_B where from **Figure 5.1c**, we can presume in general: $S_{AB} \approx S_A + S_B$. However, when, for example, A is a diatomic gas, the magnitude of the frequencies of the molecular vibrations of A are significantly larger and the incomplete cancellation of the vibrational entropy of AB and B leads to significant error as temperature increases.

5.3.2 Descriptor identification and performance

Because ΔH_f and $G_i(T)$ are readily obtained from tabulated calculated or experimental results, it is the lack of tabulated $G^\delta(T)$ which prevents the tabulation of $\Delta G_f(T)$ in computational materials databases (**Equation 5.2**). The SISSO (sure independence screening and sparsifying operator) approach¹⁴¹ was used to identify the following descriptor for $G^\delta(T)$:

$$G_{\text{SISSO}}^\delta(T) \left[\frac{\text{eV}}{\text{atom}} \right] \quad (5.4)$$

$$= (-2.48 * 10^{-4} * \ln(V) - 8.94 * 10^{-5} m V^{-1}) T + 0.181$$

$$* \ln(T) - 0.882$$

where V is the calculated atomic volume ($\text{\AA}^3 \text{ atom}^{-1}$), m is the reduced atomic mass (amu), and T is the temperature (K). SISSO efficiently selects this descriptor from a space of $\sim 3 \times 10^{10}$ candidate three-dimensional descriptors, where the dimensionality is defined as the number of fit coefficients (excluding the intercept). A training set of 262 compounds with 2,991 (T, G^δ) points was randomly selected from 309 inorganic crystalline solid compounds with experimentally measured $G^\delta(T)$

(**Figure 5.1a**) and was used for descriptor identification. The remaining 47 compounds with 558 (T, G^δ) points were reserved for testing. The descriptor performs comparably on the training and test sets with mean absolute deviations between the descriptor and experiment of $< 50 \text{ meV atom}^{-1}$ on both sets (**Figure 5.2**). Notably, there is some T -dependence on the magnitude of residuals, with larger deviations as T (and therefore the magnitude of G^δ) increases. There are three plausible reasons for this: 1) the magnitude of G^δ being predicted increases so at fixed relative error, the magnitude of the residuals is larger, 2) the number of compounds with measured $G^\delta(T)$ decreases as T increases, and 3) the physics dictating G^δ at high T are more complex due to e.g., significant anharmonic vibrational effects that are less accurately captured by the simple model of **Equation 5.4**. Approximately 1/3 of the compounds considered have measured $G^\delta(1800 \text{ K})$ and the mean absolute deviation (MAD) between G^δ_{SISSO} and G^δ_{exp} is found to increase from 53 meV atom^{-1} to 92 meV atom^{-1} from 1000 to 1800 K on the 47 compound test set. However, the relative MAD actually decreases from 14% to 11% over this same range on the test set, supporting reason (1) as a primary driver for the increasing residuals at elevated temperature.

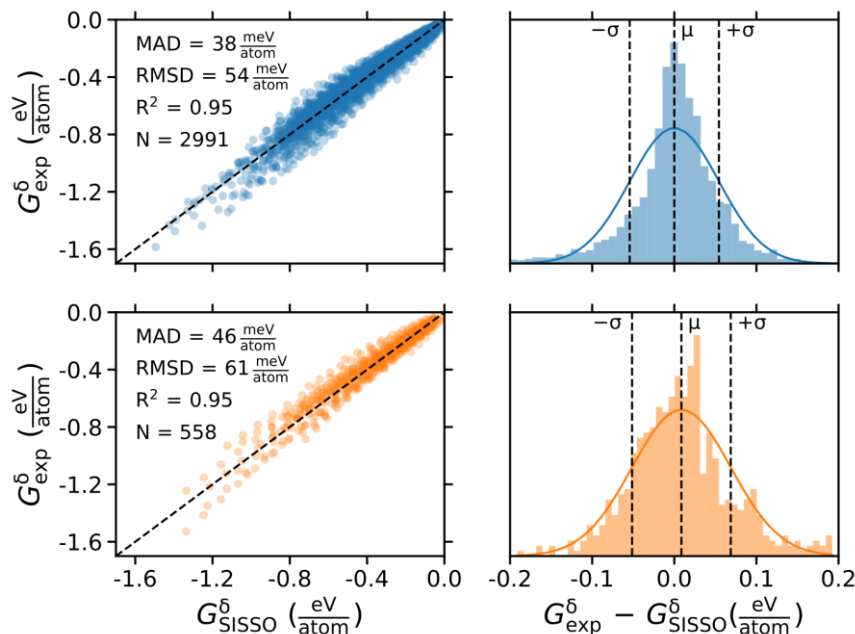


Figure 5.2 Descriptor performance

Performance of the SISSO-learned descriptor (**Equation 5.4**) on the training (top) and test sets (bottom). MAD is the mean absolute deviation, RMSD the root mean square deviation, N the number of points shown, μ the mean deviation and σ the standard deviation. The curved lines are normal distributions constructed from μ and σ .

While a number of elemental and calculated properties were considered as inputs, it is notable that SISSO selects a descriptor dependent on only three quantities – temperature, atomic mass, and (calculated) atomic volume. The identification of these properties agrees well with intuition regarding the properties that most significantly affect the magnitude of vibrational entropy and free energy.^{143, 144} The phonon frequencies in a solid compound, ω , are proportional to the force constant of the vibrational mode, k , and the reduced mass, m , of the vibrating atoms of the mode, with $\omega \sim \sqrt{k/m}$ in the harmonic oscillator approximation. As a mode's stiffness increases or its reduced mass decreases, its vibrational frequency increases, leading to a decrease in vibrational entropy and more positive Gibbs energies. This relationship is also apparent in the descriptor for $G^\delta(T)$, where m is included directly and V appears as a surrogate for k (larger atomic volumes being associated with less stiff bonds or lower k). At constant m and V , increasing

temperature decreases G^δ when $-2.48 * 10^{-4} * \ln(V) - 8.94 * 10^{-5} mV^{-1} \leq 0.181 \ln(T)/T$. This condition is uniformly satisfied for all 309 compounds in the training and test sets from 300 to 1800 K, reflecting the expectation of the negative temperature-dependence of the Gibbs energy from fundamental thermodynamic expressions – e.g., $G = H - TS$. With V and T fixed, increases in m result in more negative Gibbs energies, agreeing with the behavior of a harmonic oscillator for which ω depends inversely on mass and G^δ depends inversely on ω . Finally, with m and T fixed, the descriptor (**Equation 5.4**) indicates that G^δ becomes more negative for larger V (for $V > 1 \text{ \AA}^3 \text{ atom}^{-1}$, i.e. all solid systems), in agreement with V acting as a surrogate for the bond stiffness in the expression for the frequencies of a harmonic oscillator. Importantly, V is the only structural parameter in **Equation 5.4** and therefore, at fixed composition (chemical formula), G^δ varies between structures (i.e., polymorphs) only as V varies and $G^\delta(V)$ dictates that less dense structures of the same composition will have more negative G^δ . Therefore, the prediction of polymorphic phase transitions is beyond the scope of this descriptor.

The quasiharmonic approximation (QHA) is commonly applied as an ab initio method for approximating G (in practice, G^δ).¹²⁸ This approach typically requires a number of DFT calculations because the Helmholtz energy, including the electronic ground state energy and the free harmonic vibrational energy, must be calculated as a function of volume (typically over a range of 10 or more volumes). Because of the high computational cost associated with QHA calculations, the number of structures with calculated G is about 4 orders of magnitude less than the number of structures with calculated formation enthalpies, ΔH_f . As an additional test set for the SISSO-learned descriptor for G^δ , we compare our predictions to 131 compounds with tabulated G^δ in the PhononDB set which are not also in the experimental set compiled from FactSage used for training and testing the descriptor (**Figure 5.3**, top). For these compounds, the descriptor agrees

well with the ab initio values calculated using QHA, with a mean absolute deviation of 60 meV atom⁻¹. Notably, there is a nearly systematic underestimation of QHA-calculated G^δ by the descriptor with $G^\delta_{\text{QHA}} > G^\delta_{\text{SISSO}}$ for 98% of (T, G^δ) points in this set. Comparing QHA to experiment for an additional 37 compounds with experimentally measured G^δ available in FactSage reveals a similar systematic deviation with $G^\delta_{\text{QHA}} > G^\delta_{\text{exp}}$ for 94% of points (**Figure 5.3**, middle). A number of factors likely contribute to the systematic offset between QHA and experiment including the approximations associated with the calculation (e.g., DFT functional and approximation to anharmonic vibrations), the neglect of additional contributions to the Gibbs energy including configurational and electronic entropy, and potential impurities or defects in the experimentally measured samples. It is notable that the deviation between G^δ_{QHA} and G^δ_{exp} is mostly systematic ($R^2 \sim 0.97$), so stability predictions based on convex hull phase diagrams constructed using ab initio G^δ_{QHA} should benefit from a fortuitous cancellation of errors, leading to even lower errors in practice than the already small deviation of 41 meV atom⁻¹ on average. Remarkably, for the same set of 37 compounds, our descriptor has lower mean absolute deviation from experiment than QHA (**Figure 5.3**, bottom) but does not exhibit this systematic underestimation of the magnitude of G^δ owing to its exclusive use of experimentally measured data for descriptor selection. While this magnitude of deviation for G^δ between experiment and prediction (using either QHA or the SISSO-learned descriptor) has been quoted as chemical accuracy (~ 1 kcal mol⁻¹) in the context of ΔH_f ,⁴⁹ it is important to note that temperature-dependent predictions of stability using Gibbs formation energies, $\Delta G_f(T)$, will be affected by errors in both $G^\delta(T)$ and the temperature-independent ΔH_f .

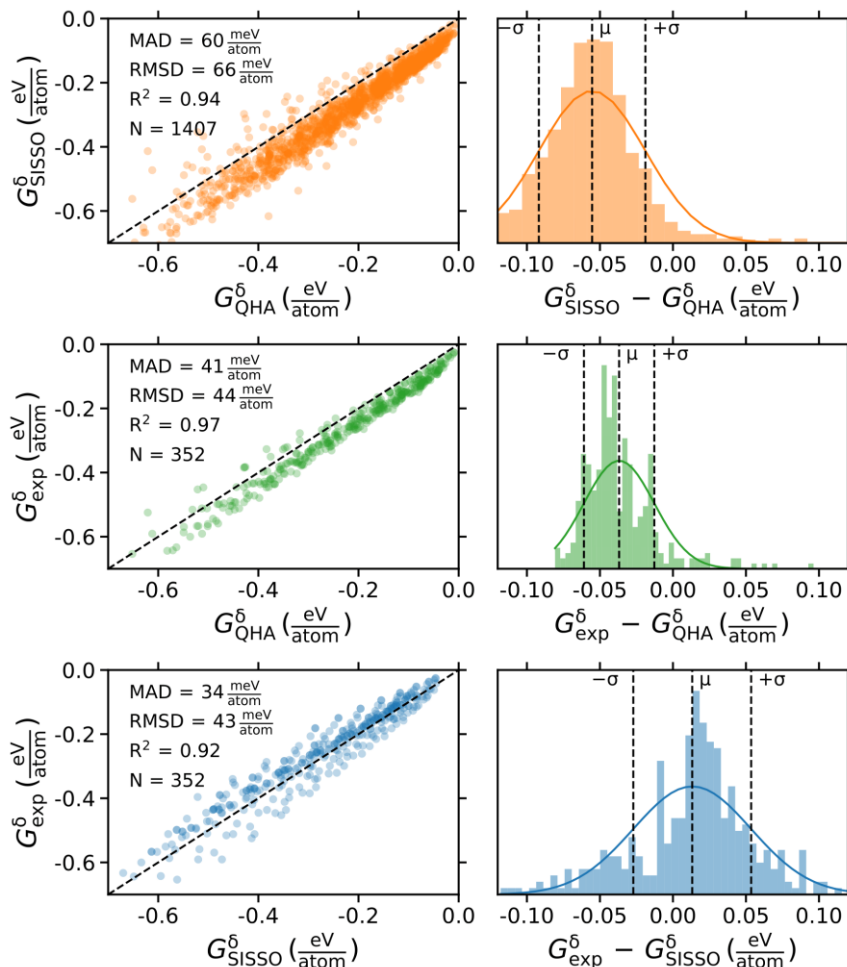


Figure 5.3 Benchmarking descriptor against ab initio methods

Comparing the SISSO-learned descriptor to QHA for 131 compounds not in the experimental dataset used to train or test the descriptor (top), comparing QHA to experiment for 37 compounds which appear in both FactSage and PhononDB (middle), comparing the SISSO-learned descriptor to experiment of these same 37 compounds (bottom). The annotation within each figure is provided in the **Figure 5.2** caption.

5.3.3 Thermochemical reaction equilibria

We combine our high-throughput model for the prediction of $G^\delta(T)$ with tabulated and readily-available DFT calculated ΔH_f and experimental Gibbs energies for the elements, $G_i(T)$ into **Equation 5.2** to enable the rapid prediction of $\Delta G_f(T)$ from a single DFT total energy calculation. Thus, reaction energetics, thermochemical equilibrium product distributions, and temperature-dependent compound stability can be assessed for the millions of structures currently compiled in

materials databases. This unprecedented ability to rapidly predict reaction equilibria for reactions involving solid compounds is illustrated in **Figure 5.4** for a small set of example reactions. In **Figure 5.4a**, the Gibbs energy of reaction, $\Delta G_{\text{rxn}}(T)$, which dictates the equilibrium spontaneity of any reaction event, is demonstrated for: the decomposition of SnSe,¹⁴⁵ solar thermochemical hydrogen generation by the Zn/ZnO redox cycle,¹⁴⁶ the carbothermal reduction of NiO to Ni,¹⁴⁷ the oxidation of MoS₂,¹⁴⁸ and the corrosion of CrN by water.¹⁴⁹ In each case, ΔG_{rxn} computed from the SISSO-learned descriptor for $G^\delta(T)$ agrees both qualitatively and quantitatively with ΔG_{rxn} resulting from the experimental values for $G^\delta(T)$.

As a more sophisticated demonstration, **Figure 5.4b** shows the equilibrium product distribution based on Gibbs energy minimization for the hydrolysis of Mo₂N to MoO₂ in the context of solar thermochemical ammonia synthesis.¹⁵⁰ In this analysis, Mo₂N and H₂O are placed in a theoretical chamber at 1 atm fixed pressure and allowed to reach thermodynamic equilibrium with a set of allowed products – MoO₂, Mo, NH₃, H₂, and N₂ – where the equilibrium product distribution at each temperature is that which minimizes the combined Gibbs formation energy of all species in the chamber. Even for this relatively complex system, the predicted product distribution based on the descriptor for $G^\delta(T)$ agrees both qualitatively and quantitatively with the product distribution calculated from the experimental $G^\delta(T)$. While this capability is demonstrated here to illustrate the utility of the identified descriptor for a few example reactive systems, this procedure is readily amenable for predicting reaction equilibria and product distributions in a high-throughput manner with numerous reacting species for a wide range of solid-state reactions. The accuracy of the descriptor-predicted reaction energies for new systems will be dependent not only on the effectiveness of $G^\delta_{\text{SISSO}}(T)$ to approximate $G^\delta_{\text{exp}}(T)$ but also on the extent to which DFT-

predicted ΔH_f agrees with experiment as both parameters are required to obtain $\Delta G_f(T)$ (Equation 5.2) and therefore $\Delta G_{\text{rxn}}(T)$.

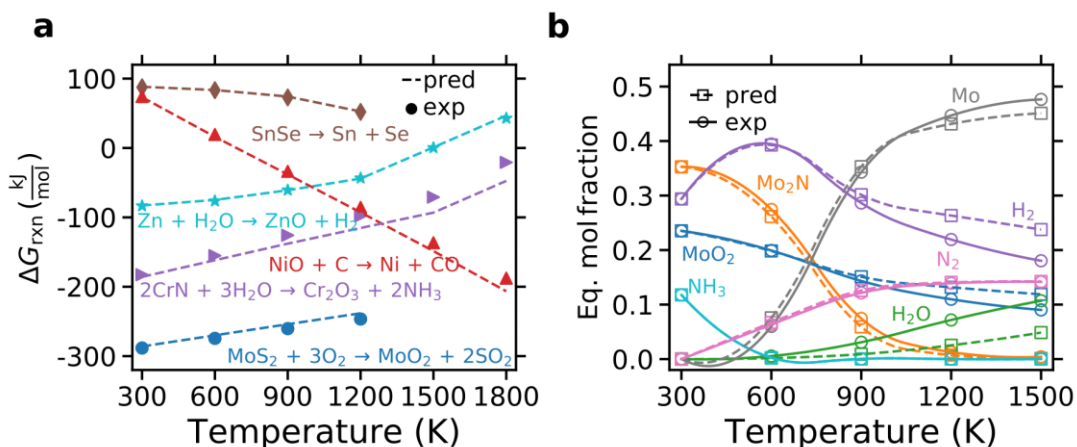


Figure 5.4 High-throughput reaction engineering

a) A comparison of experimental reaction energetics (labels) to those predicted using the machine-learned descriptor for $G^\delta(T)$ (dashed curves). **b)** Reaction product distribution between MoO_2 , Mo_2N , N_2 , H_2 , H_2O , and NH_3 based on Gibbs energy minimization subject to molar conservation and fixed pressure of 1 atm. In both figures, *pred* applies the SISSO-learned descriptor to $G^\delta(T)$ of the solid phases and experimental data for all other components.

5.3.4 Effects of temperature and composition on material stability

Beyond the investigation of solid-state reaction equilibria for a few example systems, we have also used the descriptor for $G^\delta(T)$ to compute phase diagrams to obtain broad insights into the temperature-dependent stability and metastability of thousands of known stoichiometric compounds. In particular, in the convex hull construction, formation energies, ΔG_f , are plotted as a function of composition, and joined to produce the convex object of largest area. If ΔG_f of a composition lies above the convex hull, the composition is thermodynamically metastable and the vertical distance from the hull quantifies the magnitude of metastability of the compound, where larger distances indicate a greater thermodynamic driving force for decomposition of the metastable phase into stable phases. For the first time, temperature can be incorporated as a third

axis in a high-throughput manner using $G^{\delta}_{\text{SISSO}}$ to produce $\Delta G_f(T)$ and assess the stability of compounds.

The Materials Project tabulates calculated structures for 29,525 compositions which also have reported ICSD numbers, suggesting that they have been realized experimentally.¹⁶ Previous efforts to analyze temperature-independent metastability used ΔH_f as a surrogate for formation energy to predict that ~50% of all ICSD structures are metastable at 0 K.¹⁵¹ We predict that ~34% of ICSD compositions are metastable in the absence of temperature effects – i.e., also using ΔH_f . An important distinction between structures and compositions is that if a given composition has more than one known structure, all structures except the ground state at a given set of thermodynamic conditions are, by definition, metastable under those conditions. As such, in our analysis, we consider all structures of the 29,525 compositions, but only report statistics for the ground state structures at each temperature (**Figure 5.5, Figure 5.6**).

The fraction of compositions that are thermodynamically metastable remains nearly constant up to ~900 K where the competing effects of the elemental phases (**Figure 5.1b**) lead to increasing compound destabilization with temperature (**Figure 5.5a**). The fraction of compounds which move onto and off of the convex hull with temperature are also quantified relative to those that are predicted to be metastable and stable at 0 K. If a given composition exhibits no stable structures at 0 K (i.e., ~34% of the ICSD), it is unlikely that any of these structures become thermodynamic ground states at higher temperatures. In fact, only 1,602 of the 10,001 0 K metastable compositions are found to be stabilized when temperature is increased up to 1800 K. For the 1,602 compounds which are 0 K metastable but that come onto the hull to become stable at elevated temperature, the magnitude of their 0 K metastability is quantified in **Figure 5.5b**. In general, compounds must lie very near to the hull at 0 K to have a chance of thermal stabilization

at $T > 0$ K. Even for compounds which become thermodynamic ground states at 1200 K, we find their metastabilities at 0 K to be typically $< 15 \text{ meV atom}^{-1}$ and thus thermal stabilization is often not the active mechanism in the high temperature synthesis of solid compounds.

It is well known that metastable structures are often accessed experimentally, as indicated by the significant fraction of ICSD structures which are realized, but predicted to be metastable across this wide temperature range. A number of routes exist for accessing metastable structures, such as non-equilibrium synthesis conditions and alloying. In these cases, the magnitude of the metastability of these non-equilibrium structures indicates the driving force to convert to one or more stable phases, which is a critical consideration in materials processing and successful application of the material at operating conditions. Given the pool of metastable compositions in the ICSD, a Gaussian kernel density estimate is constructed based on the magnitude of metastability, ΔG_d , and evaluated as a function of temperature (**Figure 5.5c**) and composition (**Figure 5.6**). At 0 K, 54% of metastable (but synthesized) compounds are $> 25 \text{ meV atom}^{-1}$ above the convex hull, 39% are $> 50 \text{ meV atom}^{-1}$, and 26% are $> 100 \text{ meV atom}^{-1}$ above the hull. These results provide some quantification for the false negative rate that is incurred by the $\sim 25\text{-}100 \text{ meV atom}^{-1}$ heuristic error bars of materials screening approaches where compounds are typically allowed to survive stability screening if they are thermodynamically stable or within $\sim 25\text{-}100 \text{ meV atom}^{-1}$ of metastability.¹⁵¹⁻¹⁵⁴ This range has been justifiably augmented in some cases, for example, in the search for novel 2D materials, which are by definition metastable, where the range has been expanded to e.g., $150 \text{ meV atom}^{-1}$.¹⁵⁵ Recent work has also shown that the 0 K energy of amorphous phases can provide an upper bound on the metastability of compounds that can be synthesized.¹⁵³ At low temperatures, the distribution of metastability is mostly constant with a median metastability of 43 meV atom^{-1} at 900 K, suggesting that increasing the temperature from

room temperature to 900 K results in only a small thermodynamic penalty of ~ 20 meV atom⁻¹. Above this temperature, many competing elemental phases undergo phase changes, leading to destabilization of compounds and a median metastability of 113 meV atom⁻¹ at 1800 K. This provides rationale for the viability of high temperature solid-state synthesis approaches where increasing the temperature enables atomic rearrangements to overcome kinetic barriers while maintaining the desired structure as a thermodynamically accessible metastable state.

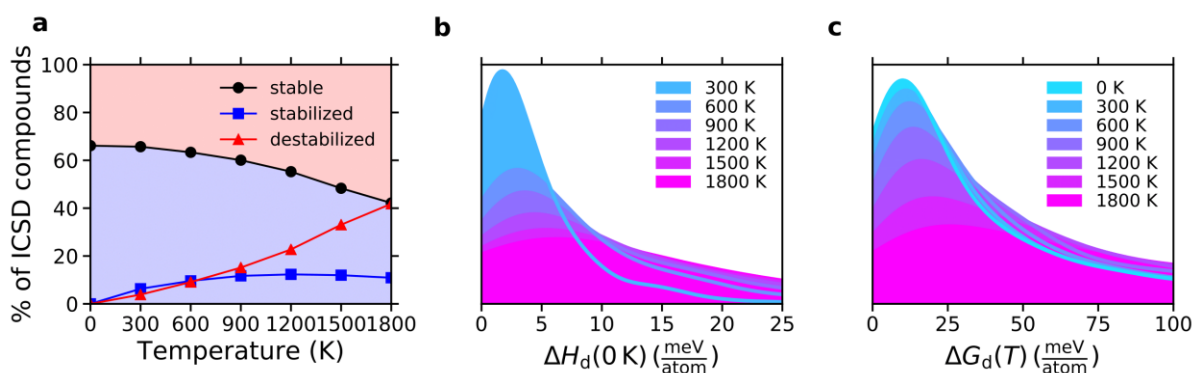


Figure 5.5 Survey of temperature-dependent (meta)stability

a) Fraction of ICSD compositions which are thermodynamic ground states (black), fraction of 0 K metastable compositions which are stable at T (blue), fraction of 0 K stable compositions which are metastable at T (red), **b)** Gaussian kernel density estimate of the 0 K decomposition enthalpy for ICSD compositions which are thermodynamically metastable at 0 K but stable at T , **c)** Gaussian kernel density estimate of the Gibbs decomposition energy at T for metastable compounds at each T .

In addition to the temperature-dependence of metastability, accessible compound metastability is also composition-dependent, as shown in **Figure 5.6**. At 0 K, compounds comprised of most elements have a similar distribution of metastabilities to the overall distribution shown in **Figure 5.5c**, with a few notable exceptions, particularly compounds containing carbon or nitrogen. For carbides and nitrides, the median metastabilities at 0 K are 144 meV atom⁻¹ and 109 meV atom⁻¹, more than five times the median metastability of all other compounds in the ICSD at 0 K (20 meV atom⁻¹). This prevalence of enhanced accessibility of metastable states was

previously recognized for nitrides at 0 K and attributed to high cohesive energy which enables metastable configurations to persist.^{151, 156} The consequences of the high cohesive energies of these materials is low self-diffusion coefficients or high barriers to atomic rearrangement resulting from the tendency of the not-so-electronegative anions, carbon and nitrogen, to form mixed covalent/ionic bonds with electropositive and weakly electronegative elements across the periodic table.

Despite the similar metastability behavior of carbides and nitrides at low temperature, we find that temperature has a dramatically different effect on these two classes of compounds, with nitrides rapidly destabilizing by moving away from the hull and broadening their metastability distribution relative to carbides. The increases in median metastability for carbides and nitrides from 0 to 1800 K are 144 meV atom⁻¹ and 231 meV atom⁻¹, respectively. This can be attributed to the tendency for entropy to stabilize gaseous elemental nitrogen (i.e., N₂) with temperature much more rapidly than solid elemental carbon (i.e., graphite). This creates the considerable high temperature metastability difference that likely plays a critical role in enabling the synthesis of metastable carbides from amorphous precursors, where the lower thermodynamic driving force for phase separation of carbides at high temperature enables the persistence of higher energy amorphous precursor phases and increased thermal energy required to activate crystallization kinetics. The remarkable metastabilities exhibited by carbides and nitrides relative to other classes of materials provide chemical design principles for hindering atomic rearrangements and point towards these underexplored spaces for the discovery of highly metastable materials which are likely synthesizable.

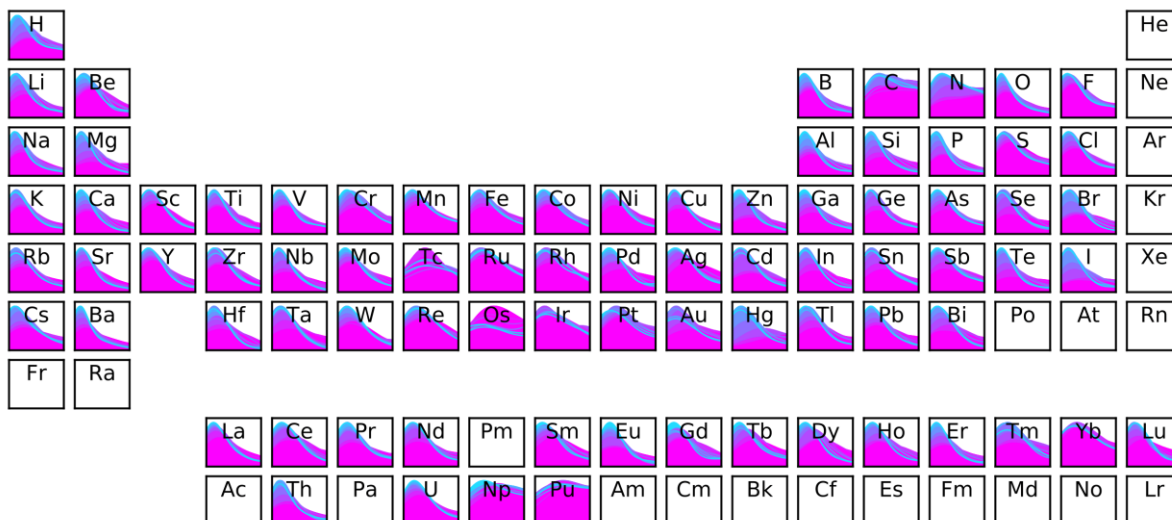


Figure 5.6 Composition-dependence of metastability

Elemental partitioning of the results shown in **Figure 5.5c** by compounds containing element *X*. Each x-axis spans 0 to 100 meV atom⁻¹ and the colors align with the legend as shown in **Figure 5.5c**.

5.4 Discussion

Open materials databases are populated with millions of DFT-calculated total energies and formation enthalpies which have been used extensively for the design and discovery of new materials. However, critically lacking from these databases is the effect of temperature on the thermodynamics of these materials. To address this challenge, we have developed a simple and accurate descriptor for the Gibbs energy of inorganic crystalline solids, $G^\delta(T)$, using the SISSO approach. This low dimensional and physically interpretable descriptor reveals the main drivers for $G^\delta(T)$ to be the mass of the elements which comprise the compound and the volume those atoms occupy in the material, agreeing well with the expectation from fundamental physical expressions and prior work quantifying the magnitude of vibrational entropy in solids. Remarkably, using only these parameters and temperature, the Gibbs energy can be predicted with accuracy comparable to the ab initio QHA approach up to at least 1800 K. Our descriptor for $G^\delta(T)$ can be readily applied to any of the more than one million structures with tabulated DFT total

energy, enabling the high-throughput prediction of temperature-dependent thermodynamics across a wide range of compositions and temperatures.

Utilizing this descriptor, we demonstrate the accurate prediction of reaction energetics for a number of solid-state reactions, including a reaction network of several competing reactions in the context of thermochemical ammonia synthesis. This demonstrates how the descriptor can be incorporated with existing materials databases and tabulated thermochemical data for non-solids to predict the equilibrium products for an arbitrary reaction as a function of temperature. By applying the descriptor to ICSD compounds in the Materials Project database, we obtain the first comprehensive look at materials stability, providing a quantitative determination of how narrowly nature and inorganic synthesis have explored far-from-equilibrium materials and providing guidance for compositional considerations in realizing new metastable materials. While thermodynamic stability is the primary criterion used in high-throughput computational screening of materials to predict the likelihood of a given material being synthesizable, the interplay of thermodynamics with several other criteria, such as kinetics and non-equilibrium process conditions or starting precursors, exhibit a stronger influence over the synthesizability of materials, and currently, there is not a universal and well-defined metric for synthesizability.^{37, 121, 153, 157-159} Importantly, the ~ 50 meV atom⁻¹ resolution in predicting $G^\delta(T)$ achieved by our descriptor exceeds the accuracy of the computational methods that currently predict and populate ΔH_f in materials databases. Therefore, when combining $G^\delta(T)$ with ΔH_f to determine the Gibbs formation energy, $\Delta G_f(T)$, errors in these approaches will be additive, emphasizing the need for new or beyond-DFT methods to calculate ΔH_f when extremely high accuracy is required for a given application. However, there are many examples where DFT-computed ΔH_f was used successfully to realize

new materials¹⁶⁰⁻¹⁶² and the incorporation of temperature effects using the SISSO-learned descriptor for $G^\delta(T)$ should only enhance these efforts.

5.5 Methods

5.5.1 Data retrieval

Gibbs energies were extracted from the FactSage⁴⁷ experimentally determined thermochemical database for 309 solid compounds and from the PhononDB¹³¹ ab initio calculated thermochemical database for 131 additional solid compounds (12 hydrides, 26 carbides, 31 nitrides, 104 oxides, 43 fluorides, 26 phosphides, 47 sulfides, 36 chlorides, 17 arsenides, 30 selenides, 40 bromides, 18 antimonides, 26 tellurides, 34 iodides; 313 binary compounds, 126 ternary compounds, and 1 quaternary compound) and 83 elements. Compound data was extracted only at temperatures where the 298 K solid structure persists as reported in FactSage. Elemental data was obtained for the phase (solid crystal structure, liquid, or gas) with the minimum Gibbs energy at a given temperature. Because the 298 K enthalpy of formation, ΔH_f , is well-predicted for compounds using high-throughput DFT along with appropriate corrections^{29, 30, 33, 35} and readily available for millions of structures in existing materials databases, the Gibbs energy was referenced with respect to ΔH_f (**Equation 5.1**).

5.5.2 Feature retrieval

Nine primary features were considered for this work – five tabulated elemental properties (electron affinity, first ionization energy, covalent radius, Pauling electronegativity, and atomic mass) extracted from pymatgen¹¹⁸ and WebElements⁸²; two calculated properties (atomic volume and band gap) extracted from the Materials Project database; one experimental property (ΔH_f), and temperature. The five tabulated elemental properties were formulated into compound-specific

properties using each of three transformations. For elemental feature, x , we define three forms of averaging – the stoichiometrically weighted mean (avg), the stoichiometrically weighted harmonic mean, akin to the reduced mass (red), and the stoichiometrically weighted mean difference (diff):

$$x_{\text{avg}} = \frac{1}{\sum_{i=1}^N \alpha_i} \sum_i^N \alpha_i x_i \quad (5.5)$$

$$x_{\text{red}} = \frac{1}{(N-1) \sum_{i=1}^N \alpha_i} \sum_{i \neq j}^N (\alpha_i + \alpha_j) \frac{x_i x_j}{x_i + x_j} \quad (5.6)$$

$$x_{\text{diff}} = \frac{1}{(N-1) \sum_{i=1}^N \alpha_i} \sum_{i \neq j}^N (\alpha_i + \alpha_j) |x_i - x_j| \quad (5.7)$$

where when considering a compound, $A_a B_b C_c$, we define α as the vector of coefficients $[a, b, c]$ and N as the length of α . For example, for CaTiO_3 , $\alpha = [1, 1, 3]$ and $N = 3$.

5.5.3 Descriptor identification

The SISSO approach¹⁴¹ was applied to identify the descriptor for G^δ shown in **Equation 5.3** using 262 of the 309 compounds from FactSage with experimentally measured G^δ . To identify this descriptor an initial feature-space, Φ_0 , included 19 features – the five tabulated elemental properties mapped onto each of the three functional forms (**Equations 5.5 – 5.7**), along with the linear forms of atomic volume, band gap, formation enthalpy, and temperature. Two iterations of descriptor construction were performed using an operator space of $[+, -, |-, *, /, \exp, \ln, ^{-1}, ^2, ^3, ^{0.5}]$. Candidate descriptors were constructed by iteratively applying these operators to Φ_0 while conserving the units of constructed features. The first iteration of descriptor construction yielded a space, Φ_1 , with ~600 candidate descriptors and the second iteration a space, Φ_2 , of ~600,000 candidate descriptors. SISSO was then performed on Φ_2 with a subspace size of 2,000 and three

descriptor identification iterations, thereby producing the three-dimensional (3D) descriptor (i.e., three fit coefficients not including the intercept) in **Equation 5.4**. In the first iteration, sure independence screening (SIS) was used to select the 2,000 descriptors S_{1D} from Φ_2 having the highest correlation with G^δ . Within S_{1D} , ℓ_0 -norm regularized minimization, $SO(\ell_0)$, was used to identify the best 1D descriptor. This 1D descriptor is then used to predict the training set and the array of residuals, R_1 , is generated from this prediction. Now with R_1 as the target property (instead of G^δ), SIS identifies a new subspace S_{2D} of 2,000 additional descriptors. $SO(\ell_0)$ then selects the best-performing 2D descriptor from $S_{1D} \cup S_{2D}$ and R_2 is generated as the residuals using this 2D descriptor to predict the training set. This procedure is repeated a third time to yield the 3D descriptor shown in **Equation 5.4**. Therefore, this descriptor is selected among a space of $\binom{6000}{3}$ or $\sim 3 \times 10^{10}$ candidate 3D descriptors.

Importantly, all aspects of the SISO selection algorithm were performed on the training set of 262 compounds with experimentally measured Gibbs energies, leaving an excluded test set of 47 compounds with experimentally measured Gibbs energies in reserve to evaluate the predictive quality of the selected descriptor (Figure 5.2). An additional 131 compounds with QHA-calculated $G^\delta(T)$ not present in the training or test sets were also compared with the SISO-learned $G^\delta(T)$ (**Figure 5.3**).

5.5.4 Descriptor sensitivity

While the random splitting of the experimental set into training and test sets was performed only once, comparing the relevant properties for each set reveals that they are statistically similar, suggesting the model and SISO process would yield similar results for an arbitrary random split of the experimental set. To assess the robustness of the model on diverse training and test sets, we

repeated the random split of the experimental set 1,000 times and evaluate the performance of **Equation 5.4** on each set. The MAD spans 37-42 meV atom⁻¹ on the 85% training set and 26-54 meV atom⁻¹ on the 15% test set, demonstrating that the reported 38 meV atom⁻¹ for training and 46 meV atom⁻¹ for testing (**Figure 5.2**) are not outliers. As an added demonstration, the random split of the experimental set and subsequent SISSO selection process was repeated 12 times. In 10/12 runs, the descriptor shown in **Equation 5.4** appears in the top 3,000 of $\sim 3 \times 10^{10}$ models evaluated (top $\sim 0.00001\%$) in terms of root mean square deviation (RMSD) on the training set. Notably, there are many cases where very slight deviations of **Equation 5.4** also appear in the top models – e.g., replacing $\ln(T)$ with T or $T^{0.5}$. To validate the significance of the three features that comprise the descriptor – temperature, reduced mass, and atomic volume – we assess what fraction of the top 3,000 models contain these features for each of the 12 random train/test splits. Temperature is found to occur in 100% of the top models for each of the 12 random splits. Reduced mass and atomic volume each appear in $\sim 86\%$ of the top 3,000 models on average over the 12 random splits. This analysis was conducted on only the very best models (top $\sim 0.00001\%$) and reveals the significance of these three properties in predicting G^δ to be robust to the random split of the experimental data used to train and test the descriptor. Notably, the first term in **Equation 5.4**, $T\ln(V)$, appears as the feature with the highest correlation with G^δ in all of the 12 random train/test splits.

5.5.5 Comparing to quasiharmonic approximation

QHA-calculated $G(T)$ was extracted from the 2015 version of PhononDB¹³¹ for all compounds with calculated thermal properties. Because a number of approximations are used to calculate ΔH_f from DFT calculations, to isolate the temperature-dependent Gibbs energy for comparison to our descriptor, $G^\delta_{\text{QHA}}(T)$ was calculated as $G^\delta(T) = G(T) - G(0 \text{ K})$.

5.5.6 Stability analysis

For the generation of **Figure 5.5** and **Figure 5.6**, all 34,556 entries (structures) in the Materials Project which have reported formation energies and ICSD numbers were retrieved. For each entry, the temperature-dependent formation energy was calculated as follows:

$$\Delta G_{f,\text{pred}}(T) = \begin{cases} \Delta H_{f,\text{MP}}, & T = 0 \text{ K} \\ \Delta H_{f,\text{MP}} + G_{\text{SISSO}}^{\delta}(T) - \sum_i^N \alpha_i G_{i,\text{exp}}(T), & T \neq 0 \text{ K} \end{cases} \quad (5.8)$$

FactSage elemental energies were used as $G_{i,\text{exp}}$. For all entries, $\Delta G_{f,\text{pred}}(T)$ was evaluated at 0, 300, 600, 900, 1200, 1500 and 1800 K. To avoid overweighting the analysis to compounds which have many polymorphs, the lowest (most negative) $\Delta G_{f,\text{pred}}(T)$ was retained for the analysis at each temperature and for each unique composition (chemical formula). This resulted in 29,525 unique compositions from 34,556 structures with ICSD numbers and reported formation energies in Materials Project. To avoid potentially spurious entries in the ICSD, only the lowest 90% of metastable compositions (with respect to the Gibbs decomposition energy, ΔG_d) were considered. Python was used to construct all possible convex hull phase diagrams and quantify ΔG_d .

5.5.7 Structure considerations

For training, we used 0 K ground-state structures (and magnetic configurations) reported in Materials Project. From this calculation result, we retrieved the volume (per atom) that is then used at all temperatures to generate $G^{\delta}(T)$ as shown in **Equation 5.4**. For a given composition, one could compute $G^{\delta}(T)$ for any number of structural or magnetic configurations and compare the $G(T)$ that results. For the purposes of training and testing, we consider only the calculated ground-

state because this is likely the approach that would be used in practice for the application of the model to new materials which have available calculated but not experimental data.

5.5.8 Application of the descriptor

To obtain the Gibbs formation energy for a given structure, one must first perform a DFT total energy minimization of the structure. From this, the atomic volume is determined as the volume of the calculated cell divided by the number of atoms in the calculated cell. G^δ can then be computed by **Equation 5.4**. Calculating the Gibbs energy, $G(T)$, using **Equation 5.1** requires the formation enthalpy, ΔH_f , calculated using DFT. If the analysis of interest concerns only one composition (chemical formula), then this is the final step and the relative energies of all structures with this composition can be compared using $G(T)$. If the analysis of interest considers various compositions (e.g., for convex hull stability or thermochemical reaction analysis), the elemental energies must be subtracted to obtain the Gibbs formation energy, $\Delta G_f(T)$ by **Equation 5.2**. Notably, ΔH_f and volumes calculated by DFT are tabulated for many thousands of structures and the elemental $G(T)$ are also tabulated for at least 83 elements. An important point is that users of the descriptor for $G^\delta(T)$ are free to generate ΔH_f and volumes for any number of structural or magnetic configurations for a given composition and compare how $G(T)$ might be sensitive to the changes in structure and magnetism.

5.5.9 Extension to new materials

On the experimental training set of 262 compounds, the mean absolute deviation between experiment and the descriptor is 38 meV atom⁻¹ (**Figure 5.2**). This increases slightly to 46 meV atom⁻¹ (**Figure 5.2**) on the experimental test set and to 60 meV atom⁻¹ on the computed (QHA) test set (**Figure 5.3**). The residuals with respect to experiment are also mostly normally distributed,

suggesting no systematic error in the model. The performance on the test set compounds is a demonstration of validated prediction accuracy or uncertainty on new predictions. These approximate error bars can be expected on additional new predictions to the extent that the sets used for training and testing are comparable to the new materials being predicted. The set we use for training and testing is quite diverse – 83 unique elements, binaries and multinarys, magnetic and nonmagnetic, metallic and insulating, etc. Additionally, the descriptor is relatively simple, having only four fit parameters (including the intercept) and three features (properties) that it depends upon. However, it has not been benchmarked for non-stoichiometric compounds or compounds with defects. For example, one could not expect to obtain the temperature-dependent defect formation energy using our descriptor because this was not benchmarked. Our model is also not capable of predicting the melting point of compounds. $G^\delta(T)$ is for the solid phase and can be obtained even well above a compound's melting point, where the liquid phase has more negative Gibbs energy. As alluded to in the main text, the extension of the descriptor to correctly predict polymorphic phase transitions or temperature-driven magnetic transitions is not practical because the descriptor depends only on the mass, density, and temperature and the magnitude of the energy change for these transitions is typically smaller than the expected error bars of the descriptor. We report substantial evidence that the descriptor is predictive for stability of compounds relative to one another and for the prediction of thermochemical reaction equilibria over a wide range of stoichiometric solid compounds with a diverse set of chemical and physical properties. Data (via public repository), code, and associated protocols are available in a github repository (github.com/CJBartel/predict-gibbs-energies) corresponding to the implementation and application of the model as described within this work.

6 High-throughput equilibrium analysis of active materials for solar thermochemical ammonia synthesis

Much of the content in this chapter appears verbatim in the following manuscript:

C.J. Bartel, J.R. Rumptz, A.M. Holder, A.W. Weimer, C.B. Musgrave (2018) [In Preparation]

6.1 Abstract

Solar thermochemical ammonia (NH_3) synthesis (STAS) is a potential route to produce NH_3 from air, water, and concentrated sunlight. This process involves the chemical looping of an active redox pair that cycles between a metal nitride and its complementary metal oxide to yield NH_3 . To identify promising candidates for the STAS cycle, we performed a high-throughput thermodynamic screening of 1,148 metal nitride/metal oxide pairs. This screening was based on Gibbs energies of the crystalline metal oxides and nitrides at elevated temperatures, $G(T)$, calculated using a recently introduced descriptor and 0 K DFT formation energies tabulated in the Materials Project database. Using the predicted $G(T)$, we assessed the viability of each of the STAS reactions – hydrolysis of the metal nitride, reduction of the metal oxide, and reformation of the metal nitride. From this reaction energy analysis, we identified a volcano-type dependence of the limiting reaction energy on the formation energies of the nitride and oxide that no longer occurs when considering the direct formation of the nitride from the oxide (i.e., a two- instead of a three-step cycle). For all 1,148 redox pairs analyzed and each of the STAS-relevant reactions, we implemented a Gibbs energy minimization scheme to predict the equilibrium composition and yields of the STAS cycle which reveals new active materials based on B, V, Fe, and Ce that warrant further investigation for their potential to mediate the STAS cycle. This work details a high-throughput approach to assessing the temperature-dependent thermodynamics of a

thermochemical redox process that utilizes the wealth of publicly available temperature-independent thermodynamic data calculated using DFT. This approach is readily adaptable to identifying optimal materials for arbitrary thermochemical processes and guiding the thermochemical synthesis of new compounds.

6.2 Introduction

Ammonia (NH_3) synthesis without CO_2 emissions remains a grand challenge for sustainable food, energy, and fuel production.¹⁶³⁻¹⁶⁵ NH_3 is industrially produced by the Haber-Bosch process which is driven by hydrocarbon reforming into H_2 (and CO_2) and catalytically converting H_2 and atmospheric N_2 into NH_3 at high pressure and moderate temperature. While no CO_2 is directly emitted during the NH_3 synthesis reaction, the large scale of production required to make the Haber-Bosch process economically viable requires a tremendous amount of H_2 input. This scale of H_2 production can only currently be obtained from hydrocarbon reforming, which results in the Haber-Bosch process accounting for 1-2% of global CO_2 emissions.¹⁶⁶ This motivates the search for an NH_3 synthesis process that alleviates the need for H_2 or operates on the scale of renewable H_2 production technologies.^{167, 168}

Solar thermochemical water splitting (STWS) leverages the high temperature (> 1500 K) that can be obtained by concentrating solar radiation to produce H_2 from steam at atmospheric pressure and without the consumption of any active material.¹⁶⁹ This is typically achieved by the chemical looping of an active solid, typically a metal oxide (MO) with significant oxygen exchange capacity such as ceria,¹⁷⁰ perovskite,¹⁷¹ or hercynite¹⁷². In the most viable approach to STWS, at high temperature and low oxygen partial pressure, the oxide is partially reduced, generating oxygen vacancies that can then be filled by steam, yielding H_2 . In recent years, an alternative process termed solar thermochemical ammonia synthesis (STAS) was proposed that

cyclically converts the active material between MO and metal nitride (MN), yielding NH_3 from H_2O and N_2 (**Figure 6.1**).¹⁷³⁻¹⁷⁵ In this process, NH_3 is formed by the hydrolysis of the MN by steam, which also converts the nitride into oxide. The MO is then reduced at high temperature to metal by a (preferably gaseous) reducing agent. The metal is then used to reduce atmospheric N_2 to re-form the metal nitride and restart the STAS cycle, yielding NH_3 with the consumption of H_2O , N_2 , and reducing agent. It is possible to conceive of a STAS process without phase change and mediated by oxygen vacancies, however the ability for an oxygen vacancy to reduce N_2 is unlikely.

Unlike current approaches to STWS, which involve the partial reduction of a single active material, the STAS process studied in this work requires a phase-change from MO to MN, so the active material is this redox pair (MO/MN) and the temperature-dependent thermodynamics of these two compounds dictate the viability of a given redox pair for STAS. Prior efforts to identify active materials for STAS have evaluated ~ 35 redox pairs – this number limited by the number of MNs with experimentally obtained Gibbs formation energies, $\Delta G_f(T)$.¹⁷⁶ Recently, we developed a statistically learned descriptor that enables the prediction of $\Delta G_f(T)$ with high accuracy (~ 50 meV/atom) for inorganic crystalline solids (e.g., MO and MN) when the standard-state formation enthalpy, ΔH_f , is known.¹⁷⁷ While the number of MNs with experimentally obtained ΔH_f is similarly small, DFT-calculated ΔH_f are available for thousands of MNs and MOs in open materials databases, such as the Materials Project.¹⁶ By integrating the Materials Project data with our high-throughput approach to obtain $\Delta G_f(T)$, we evaluated the thermodynamic viability of 1,148 redox pairs made from 354 binary (monometallic) oxides and 197 binary nitrides for the STAS cycle shown in **Figure 6.1**. This provides insight into the thermodynamic and materials challenges associated with each of the STAS reactions and guidance towards which yet-unexplored materials

should be the focus of further computational and experimental efforts. Additionally, we present a generic framework for coupling open materials databases of 0 K DFT calculations with our descriptor for $G(T)$ to rapidly assess the thermochemistry of solid-state reactions. This approach has applications for alternative chemical looping processes and for identifying reaction conditions (temperature, precursors, etc.) under which new materials can be synthesized by solid-state thermochemical reactions.

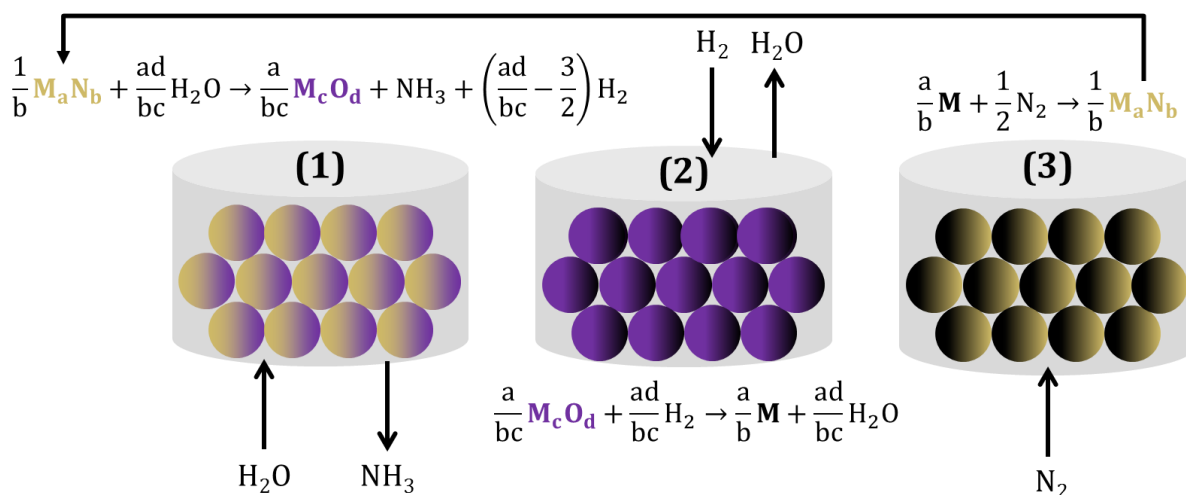


Figure 6.1 Solar thermochemical ammonia synthesis (STAS) reaction scheme

Reactions are shown on a 1 mol NH_3 per cycle basis. In (1), a metal nitride (M_aN_b) is oxidized by steam to yield a metal oxide (M_cO_d), NH_3 , and potentially H_2 (hydrolysis). Note that H_2 can become a necessary reactant for certain MN/MO pairs to achieve the basis of 1 mol NH_3 per cycle. In (2), the metal oxide is reduced by H_2 to yield the metal (M) and steam (reduction). In (3), the metal reduces atmospheric N_2 to yield the metal nitride and restart the STAS loop.

6.3 Results

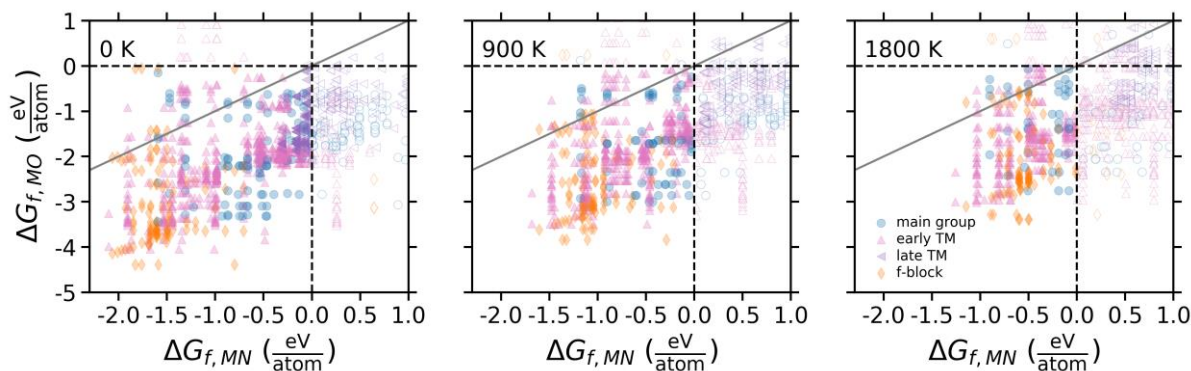
6.3.1 Formation energies of redox pairs

The Gibbs formation energies, $\Delta G_f(T)$, of the oxide and nitride are shown as a function of temperature for all 1,148 pairs considered in this work (**Figure 6.2a**). An essential requirement for viability of a given pair for STAS is that the solid components be formable at the reaction

conditions for each reaction in the STAS cycle where they are formed – i.e., the oxide must be formable at hydrolysis conditions and the nitride must formable at nitridation conditions. $\Delta G_f(T) < 0$ is a necessary (but not sufficient) requirement for stability. Stability is ultimately dictated by the decomposition energy,^{39, 178} but in this case, we show only the formation energy, $\Delta G_f(T)$, as it provides unique insights into the correlated stabilities of oxides and nitrides, which significantly influences the viability of the STAS process. As seen in **Figure 6.2a**, $\Delta G_{f,MO}(T) < \Delta G_{f,MN}(T)$ for the vast majority of pairs (points falling below the gray line). Thus, for most pairs, the stability of the nitride will dictate the stability of the solid compounds throughout the STAS cycle – i.e., if the nitride is stable, then the oxide will likely also be stable. Because cations have similar affinities for oxygen and nitrogen, the stabilities of nitrides and oxides are correlated with highly stable (unstable) metal oxides typically pairing with highly stable (unstable) metal nitrides.

Temperature has a critical effect on which cations can form stable nitrides and oxides (**Figure 6.2b**). At 0 K, 831 pairs spanning 51 cations have a nitride and oxide with $\Delta G_f < 0$. However, this decreases substantially to 566 pairs spanning 41 cations at 900 K and 387 pairs spanning 29 cations at 1800 K. The cations that yield $\Delta G_{f,MO} < 0$ (blue) and $\Delta G_{f,MN} < 0$ (green) are shown for each temperature in **Figure 6.2b**. While most of the periodic table appears potentially viable based on a 0 K analysis, because $\Delta G_{f,MN}(0\text{ K})$ is near zero for most late transition metals, nitrides with these cations have positive $\Delta G_{f,MN}$ even with modest increases in temperature to those more relevant to thermochemical processes (e.g., 900 K). 270 out of 347 oxides (78%) with $\Delta G_{f,MO} < 0$ at 0 K also have $\Delta G_{f,MO} < 0$ at 1800 K in contrast to only 59 out of 131 nitrides (45%) that meet this criteria. This suggests that temperature-swing or the separation of nitrides and oxides during the reaction cycle may be critical to the viability of STAS – i.e., forming the nitride only at low temperature and utilizing higher temperature to form the oxide.

a)



b)

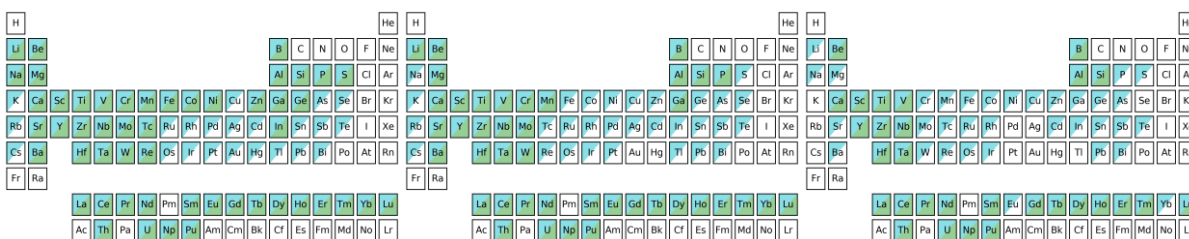


Figure 6.2 Temperature-dependent Gibbs energies

a) Comparing the Gibbs formation energies of oxide (MO) and nitride (MN) for each MN/MO pair in this analysis with temperature increasing from 0 K (left) to 900 K (center) to 1800 K (right). Empty markers correspond with the MO or MN having $\Delta G_f > 0$. The legend for markers by position in the periodic table is provided in the right-most panel. **b)** Indicating which cations have $\Delta G_{f,MO} < 0$ (top triangle, blue) and $\Delta G_{f,MN} < 0$ (bottom triangle, green) for increasing temperature – 0 K (left) to 900 K (center) to 1800 K (right).

6.3.2 Energetics of each reaction

The STAS cycles requires the cyclic conversion of nitride to oxide (hydrolysis), oxide to metal (reduction), and metal to nitride (nitrogen fixation). Each reaction is affected uniquely by the formation energies of oxide and nitride and their dependence on temperature. A thermodynamic assessment of reaction energies for each reaction helps inform the design principles for identifying a viable redox pair for this process (**Figure 6.3**).

The hydrolysis step requires the conversion of water to NH_3 by oxidizing the nitride to oxide. Because $\Delta G_{f,\text{NH}_3} > \Delta G_{f,\text{H}_2\text{O}}$ at all temperatures, this reaction is only spontaneous (reaction energy, $\Delta G_r < 0$) when the conversion of nitride to oxide is thermodynamically downhill by enough to make up for this difference. Low temperature is also preferred because $\text{H}_2\text{O} \rightarrow \text{NH}_3$ has $\Delta G_r = +230$ kJ/mol at 600 K, increasing steadily with temperature to 303 kJ/mol at 1800 K. At 600 K, 628 pairs meet this requirement of the oxide being sufficiently more stable than the nitride such that $\Delta G_{r,\text{hyd}} < 0$ (**Figure 6.3**, top). This is predictable from the results shown in **Figure 6.2a** where the oxide typically has more negative ΔG_f than the corresponding nitride, facilitating the tendency for the nitride to be oxidized by steam and have lattice nitrogen replaced by oxygen. There are two additional challenges to this reaction that have inverse dependence upon temperature. The kinetics of hydrolysis are known to be slow,^{179, 180} and this can be alleviated by increases in temperature. However, NH_3 is thermodynamically favored to decompose into $\text{N}_2 + \text{H}_2$ with decreasing ΔG_r (increased driving force for decomposition) as temperature increases, requiring the fast capture and quenching of the liberated NH_3 .¹⁸¹ While the thermodynamic conversion of nitride to oxide by steam exposure is thermodynamically favorable for ~600 binary redox pairs, these additional challenges will further restrict the number of truly viable pairs for this step.

Once the oxide is formed, the difficult challenge of converting back to the typically less stable nitride must begin. The production of metal (e.g., Mg, Fe, etc.) from oxide ore (e.g., MgO, Fe_2O_3 , etc.) is responsible for some of the oldest high-temperature industrial processes.¹⁸² However, many of these reductions utilize a solid reducing agent which complicates cycling for a chemical looping process such as STAS.^{183, 184} The first redox pair proposed for STAS – $\text{AlN}/\text{Al}_2\text{O}_3$ – requires solid reducing agent (e.g., carbon) to reduce the highly stable Al_2O_3 to Al metal and enable the eventual fixation of atmospheric N_2 to form AlN.¹⁷³ The use of gaseous

reducing agents, such as H_2 , CO , or some combination would be preferred to eliminate the need for reincorporation of a sacrificial reducing agent that must be pelletized with the active materials. In **Figure 6.3** (middle), we show ΔG_r for the reduction of each oxide by H_2 at 1800 K. While H_2 , CO , and CH_4 have comparable reducing power (and therefore similar thermodynamics of reduction), the carbonaceous reducing agents will produce CO_2 and potentially incorporate carbon into the active materials during reduction. Thermodynamically, the use of H_2 as reducing agent provides -147 kJ/mol of reducing power due to the oxidation of H_2 to H_2O . Counterintuitive to these processes being operated typically at high temperature, this reducing power increases (becomes more negative) at lower temperature. However, the increase in $\Delta G_{f,MO}$ more than offsets this such that the highest allowed temperature for a given reactor system is preferred for the oxide reduction step. For the oxide reduction by H_2 to be thermodynamically viable, $\Delta G_{f,MO}$ must be $> \Delta G_{f,H_2O}$ as reflected in **Figure 6.3** with a cutoff of approximately -1.2 eV/atom where oxides with $\Delta G_{f,MO}(1800\text{ K}) < -1.2$ eV/atom have $\Delta G_{r,red}(1800\text{ K}) > 0$. 197 out of 438 oxides (54%) meet this condition considering the high temperature of 1800 K. A complication for the requirement that the oxide have only moderately negative ΔG_f is that the nitride will typically have less negative ΔG_f , frequently becoming > 0 for these weakly stable oxides. Swinging the temperature from high for oxide reduction to low for nitride formation can mitigate this issue because the oxide and nitride are separated during these reactions.

The final reaction in the STAS cycle is the reduction of atmospheric N_2 by the metal in the formation of the metal nitride (**Figure 6.3**, bottom). This reaction is the formation reaction for the metal nitride and is thermodynamically preferred at low temperature due to the consumption of gaseous N_2 (**Figure 6.2**). The 1 mol NH_3 basis used to normalize each pair to one another dictates a non-unity, but still high, correlation between $\Delta G_{r,nit}$ and $\Delta G_{f,MN}$. Because these are nearly the

same property (differing only by the molar coefficients in the reaction), the discussion of $\Delta G_{f,MN}$ in conjunction with **Figure 6.2** also applies to $\Delta G_{r,nit}$ – i.e., viable nitride formation reactions become more and more sparse as temperature is increased with the number of binary nitrides having $\Delta G_{f,MN} < 0$ decreasing from 131 to 59 from 0 K to 1800 K.

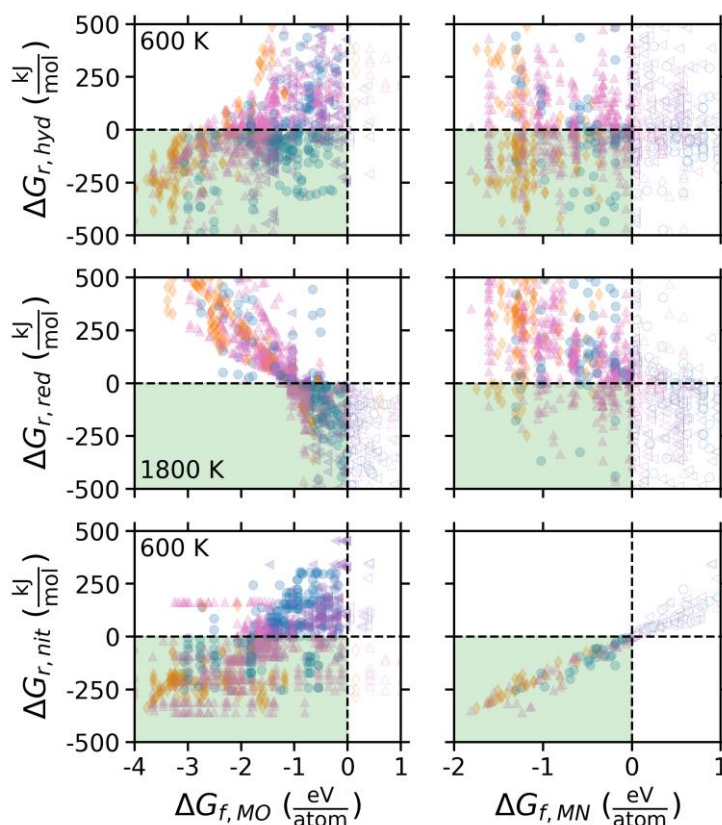


Figure 6.3 Reaction energies of the STAS cycle

Comparing Gibbs reaction energies to Gibbs formation energies for each reaction in the STAS cycle – hydrolysis (*hyd*, top), oxide reduction (*red*, middle), and nitrogen fixation (*nit*, bottom) – for oxides (left) and nitrides (right). Reaction energies are normalized per mol NH_3 per cycle. Temperatures are shown in the left panel of each reaction and chosen to maximize the number of pairs with $\Delta G_r < 0$. The legend is the same as in the right panel of **Figure 6.2a**.

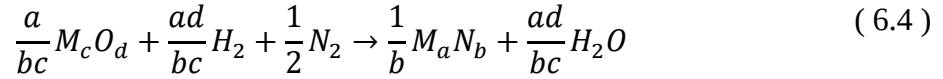
6.3.3 Limiting reaction analysis

Separately analyzing each reaction in the STAS cycle reveals a number of pairs that are viable for each step. However, for the entire STAS cycle to realize equilibrium yields, all reactions must have $\Delta G_r \leq 0$ over some suitable temperature range. This dictates three criteria that must be

met by the redox pair: 1) the conversion of nitride to oxide must be thermodynamically downhill by more than the conversion of H_2O to NH_3 is thermodynamically uphill, 2) H_2O must be thermodynamically downhill relative to the metal oxide ($\Delta G_{f,\text{H}_2\text{O}} < \Delta G_{f,\text{MO}}$), and 3) the formation energy of the nitride must be negative. Because the thermodynamics of oxides and nitrides are correlated for a given cation, it is difficult to find a case where all three of these criteria are met, and a Sabatier-type phenomena emerges where the interactions of cation and anion must be significantly strong to favor nitride formation but not so strong that the oxide cannot be reduced by H_2 . This volcano-type dependence is visualized in **Figure 6.4** (top panel) and pairs based on Mn, Fe, W, Tc, and Yb are the only ones that lie in the favorable region where $\Delta G_{r,\text{lim}} < 0$. Notably, Mn oxide and nitride were recently demonstrated as a promising system for STAS.¹⁸⁵ The volcano plot reveals an apparent optimal formation energy for the oxide and nitride in the redox pair with $\Delta G_{f,\text{MO}}(0\text{ K}) \sim -2\text{ eV/atom}$ and $\Delta G_{f,\text{MN}}(0\text{ K}) \sim -0.5\text{ eV/atom}$ minimizing the limiting ΔG_r for the cycle. Importantly, a number of pairs have $\Delta G_{r,\text{lim}}$ only slightly larger than 0 and these can still be effective active materials for STAS, but they will just yield $< 1\text{ mol NH}_3$ per cycle. For comparison, the maximum yield of H_2 for one vacancy-mediated STWS cycle is the number of oxygen vacancies that can be thermally generated in the oxide, typically much less than 1 mol/mol MO. Additionally, this work is focused only on binary (monometallic) redox pairs as an initial screening, yet binary compounds account for only $\sim 13\%$ of the compounds in the Materials Project database, suggesting that there may be ample opportunity to design redox pairs based on oxides and nitrides with multiple cations.

The volcano-type dependence of the limiting reaction energy on the active material formation energies can be averted by considering a two- instead of three-step cycle where the oxide

is directly converted to nitride by simultaneous exposure to a reducing agent and N₂ (**Figure 6.4**, bottom panel), for example where H₂ is used as the reducing agent:



This approach was utilized previously in STAS with methane as a reducing agent¹⁸⁵ and this reaction is analogous to the ammonolysis of metal oxides which has been used extensively for nitride synthesis.¹⁸⁶ The two-step cycle is the hydrolysis reaction and the reverse of this reaction, replacing NH₃ with H₂ + N₂, so that NH₃ is still produced in each cycle. In this way, the significantly positive $\Delta G_{f,NH_3}$ does not facilitate the transformation of oxide to nitride as in ammonolysis and appears only as a product during hydrolysis. The hydrolysis reaction is rarely thermodynamically challenging because it involves the typically downhill conversion of nitride to oxide. However, the direct formation of nitride from the oxide is the thermodynamically problematic reaction. Because the oxide and nitride are present during both reactions, there is no longer a volcano-type dependence of $\Delta G_{r,lim}$ on ΔG_f and the relative formation energies of the oxide and nitride are more indicative of $\Delta G_{r,lim}$. In addition to altering the thermodynamic considerations of the STAS cycle, the two-step cycle minimizes the number of reaction steps and eliminates the need to handle a pure metal, which may be prone to sintering, melting, or sublimating at elevated temperature. However, only pairs based on Na, Fe, and Tc appear in the viable region where $\Delta G_r < 0$ for both reactions over the temperature range of 600-1800 K.

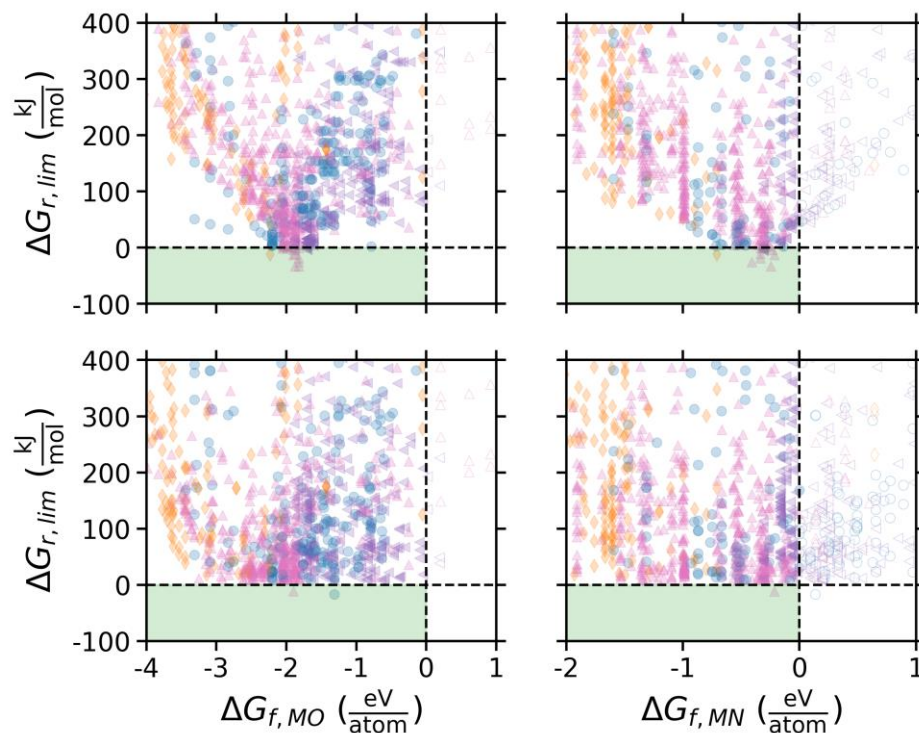


Figure 6.4 Volcano-dependence of STAS energetics

Limiting reaction plot in 3- (top) and 2-step (bottom) cycles over a maximum allowed temperature swing of 600-1800 K. ΔG_f is shown at 0 K to indicate the target formation enthalpies for each reactant that dictates the minimum of the volcano.

6.3.4 Equilibrium product distributions

While instructive, the determination of viable pairs by the limiting reaction energy analysis shown in **Figure 6.4** is not exhaustive because it takes a pairwise approach to each reaction. That is, it considers only each of the reactions as written in **Figure 6.1** when, in reality, there are a number of chemical transformations that could take place for a given set of reactants and reaction conditions. For instance, many of these pairs are complicated by the existence of alternative oxide and nitride phases that are thermodynamically more favorable than the solids in the particular pair. Ascertaining which of the pairs are truly thermodynamically viable for STAS requires an equilibrium analysis by Gibbs energy minimization which determines the molar composition that minimizes the combined free energy of an allowed set of species.¹⁸⁷ In this approach, an initial

feed of reactants is dictated by the stoichiometry of the pair and allowed to reach equilibrium with a given set of species at some temperature under the constraint of molar conservation and with the target of minimizing the Gibbs energy function. Using this approach, it can be determined which oxides and nitrides are favored to form at a given set of reaction conditions. As an example, W_2N_3/W_2O_5 appears thermodynamically viable from the limiting reaction analysis for the three-step cycle, but neither the oxide nor the nitride are the thermodynamically favored oxide or nitride of tungsten (**Figure 6.5**). Instead, the hydrolysis of W_2N_3 yields only WO_3 and not W_2O_5 and the nitridation of W forms WN_2 . This limits the viability of W as a metal reactant because WO_3 is not as readily reduced to W in the presence of H_2 .

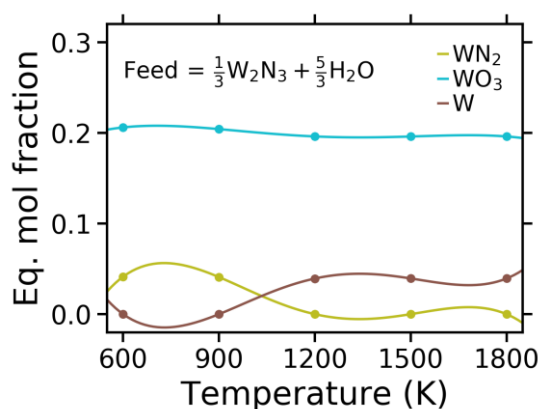


Figure 6.5 Predicted yields for tungsten nitride hydrolysis

Equilibrium product distribution predicted by Gibbs energy minimization for the hydrolysis of WN_2 . The feed considered for each reaction is dictated by the equations shown in **Figure 6.1**. This results in feeds of $1/3$ mol WN_2 + $5/3$ mol H_2O for hydrolysis. Values less than 0 are a result of a spline fit.

Repeating this analysis for all reactions and all pairs produces the yield plots shown in **Figure 6.6** which emphasizes the difficult trade-off that must be achieved between the oxide and nitride in the redox pair. In **Figure 6.6** (left), the two reactions which comprise the two-step cycle are compared – hydrolysis, *hyd*, and the combined reduction of the oxide and formation of the nitride, *redN*. The vast majority of points lie along the x- or y-axis, indicating ~0 yield for one of

the involved reactions. Only 5 of 1,148 pairs have Y_{hyd} and $Y_{\text{redN}} > 0.01$ given a maximum allowed temperature swing of 600 K to 1800 K – BN/B₂O₃, VN/VO₂, Tc₃N/TcO₂, CeN/Ce₇O₁₂, PuN/Pu₂O₃. The pairs based on Tc and Pu are not practical due to their scarcity and radioactivity, leaving B, V, and Ce as cations with significant promise for the two-step cycle based on this thermodynamic analysis.

The three-step cycle requires the careful balance of hydrolysis, reduction and nitride formation, shown in the second through fourth panels of **Figure 6.6**. While each pair-wise combination of reactions show a number of pairs with high yields for both reactions (i.e., points away from the axes), there are only 5 of 1,148 pairs that exhibit Y_{hyd} , Y_{red} , and Y_{nit} all > 0.01 over the same allowed temperature swing of 600-1800 K – P₃N₅/P₂O₅, GaN/Ga₂O₃, Tc₃N/TcO₂, FeN/FeO, and FeN/Fe₃O₄. Tc is again not preferred because of scarcity. P and Ga are problematic for the three-step cycle due to low melting points of their elemental phases (~300 K), leaving Fe as the only metal reactant with significant promise based on this thermodynamic assessment.

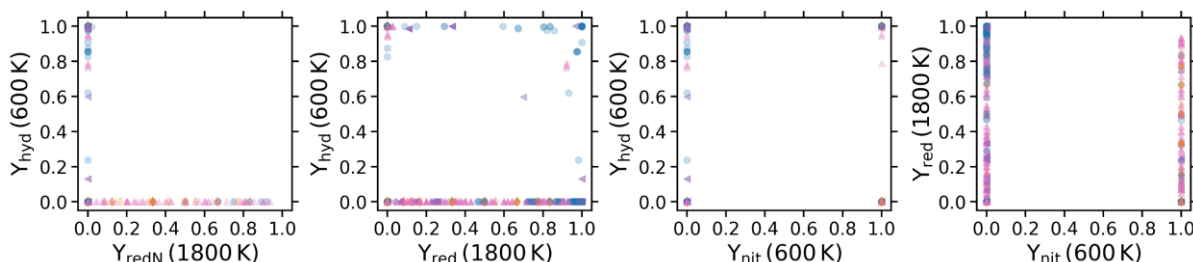


Figure 6.6 Reaction-wise equilibrium yields for all pairs

Yields of each reaction in the two- and three-step cycles. 1st panel – all reactions for the two-step cycle: hydrolysis, *hyd*, and nitride formation from the oxide, *redN*. 2nd panel – hydrolysis and oxide reduction by H₂, *red*. 3rd panel – hydrolysis and nitride formation from the metal, *nit*. 4th panel – reduction and nitride formation from the metal.

6.3.5 Shifting equilibrium

In addition to temperature, the partial pressure of reactant and product species is an important variable for dictating the thermodynamics of each reaction. For STWS, controlling the

partial pressures of H_2O during oxidation (by excess steam feed) and O_2 during reduction (by inert gas sweep) was shown to enable isothermal H_2 production using hercynite as an active material.¹²⁶ For STAS, the reduction (or simultaneous reduction and nitridation) of the metal oxide is the logical step that could benefit from manipulations of partial pressure by considering operation in the regime where the partial pressure of H_2O is decreased relative to the partial pressure of H_2 . This can be achieved by exposing the metal oxide to greater-than-equilibrium amounts of H_2 . Considering $100\times$ excess of H_2 increases the yields of eight pairs to > 0.1 mol NH_3/cycle , three of which were previously identified as having > 0.01 yield at equilibrium reactant amounts – $\text{BN}/\text{B}_2\text{O}_3$, VN/VO_2 , and $\text{FeN}/\text{Fe}_3\text{O}_4$ – and five newly viable pairs – $\text{CrN}/\text{Cr}_2\text{O}_3$, MoN/MoO_2 , WN_2/WO_3 , MnN/MnO , and $\text{Mn}_2\text{N}/\text{MnO}$. Notably, STAS cycles based upon Cr, Mo, and Mn have been previously investigated with varying degrees of success.^{168, 185, 188, 189} For the reverse reaction which occurs during STWS – water splitting over a reduced metal oxide to produce H_2 – similar ratios ($100\times$ the equilibrium amount of H_2O) were used to drive equilibrium towards H_2 production, which led to increases in solar-to-hydrogen efficiencies for the overall cycle.¹⁹⁰ The yet-unexplored pairs based on Fe and V show significant promise as their yields per cycle are increased substantially by considering only $10\times$ (instead of $100\times$) the equilibrium amount of H_2 feed. The effects of excess H_2 for the reduction of Fe_3O_4 to Fe and the formation of VN from VO_2 are shown in **Figure 6.7**. This introduces an additional parameter that can be varied to optimize the thermodynamics of these reactions, but does also introduce a processing challenge associated with separating and heating the excess gas.

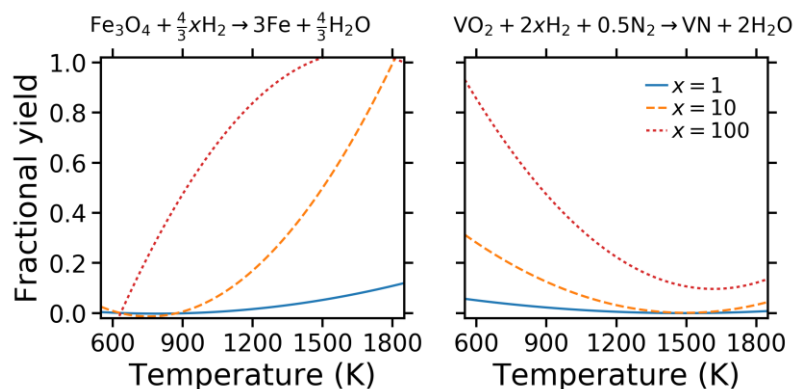


Figure 6.7 Increasing yields by shifting equilibrium

Left – yield of Fe from the reduction of Fe_3O_4 as a function of the number of moles of H_2 feed. Right – yield of VN for the reaction of VO_2 with H_2 and N_2 as a function of the number of H_2 moles in the feed. Values > 1 or < 0 are a result of a spline fit.

6.3.6 Non-equilibrium considerations

Along with thermodynamics, the kinetics of each reaction in STAS are critical to the viability of the NH_3 synthesis process, especially for reactions involving nitrogen. The corrosion of metal nitrides by H_2O is a kinetically limited process for many nitrides below 1200 K, primarily associated with the need to break metal-nitrogen bonds that can have significant covalent character.^{179, 180} Similarly, the reduction of N_2 for metal nitride synthesis requires the activation and cleavage of the triple bonded N_2 molecule, which also poses a significant kinetic challenge. The difficulty in reducing N_2 can result in metals becoming only partially nitridated, limiting the NH_3 yield that can be achieved on a per cycle basis, as this quantity is directly linked with the amount N_2 that can be fixed to the metal. Conversely, the strongly bound metal nitride species can be advantageous compared with what is captured in this equilibrium analysis because of the high metastability of metal nitrides compared with other compounds.^{32, 37, 46, 177} That is, there are likely nitrides that will not appear in a Gibbs energy minimization analysis that are still realized experimentally because of the prevalence of metastable (non-equilibrium) nitride species. While

the presumption of equilibrium is a significant approximation, these results do indicate the challenges associated with each reaction and which active materials show promise to be engineered into viable materials for this process.

6.4 Discussion

Solar thermochemical ammonia synthesis (STAS) is a potential route to producing NH_3 from air, water, and concentrated sunlight. In this work, we utilized the Materials Project database of density functional theory calculations and a recently introduced descriptor for the Gibbs energy of compounds to screen the thermodynamic viability of 1,148 metal nitride/metal oxide redox pairs for this process. Using the resulting thermochemical data, we established thermodynamic rationale for the viability of each reaction in the STAS cycle – hydrolysis of the nitride to oxide (NH_3 synthesis), reduction of the oxide by H_2 , and formation of the metal nitride from atmospheric nitrogen. Taking these reactions together, we identified a volcano-type dependence of the limiting Gibbs energy of reaction over the cycle with respect to the formation energy of the nitride and oxide. By considering an alternative cycle with just two reactions – hydrolysis and the direct formation of the nitride from the oxide – this volcano-type dependence can be removed and the limiting reaction energy depends instead on the relative stabilities of the oxide and nitride. Going beyond reaction energies, we performed a Gibbs energy minimization analysis for all 1,148 pairs and each of the four STAS-relevant reactions to generate a predicted reaction equilibrium product distribution and quantify the yields of each reaction for each pair. This analysis reveals pairs based on B, V, Fe, and Ce as yielding > 0.01 mol NH_3 per cycle given equilibrium feeds for each reaction, none of which have been previously studied for STAS. The effects of shifting equilibrium by feeding excess reactant was also studied. This effect reproduces the viability of previously studied pairs based on Cr, Mo, and Mn, indicates the viability of a yet-unexplored pair based on W, and

enhances the predicted yields of pairs based on V and Fe. Importantly, this work focused exclusively on binary (monometallic) active materials, yet these compounds comprise only ~10% of known materials. Therefore, these results indicate cations that should form the basis of more complex materials (i.e., ternary and quaternary compounds) and establish key thermodynamic considerations for the STAS cycle. The thermodynamic screening approach used here is also readily adaptable to arbitrary thermochemical reactions and can be used to guide the synthesis of new materials or the identification of active materials for other redox processes. For example, the equilibrium yields for an arbitrary solid-state synthesis reaction from solid precursors in a specified atmosphere can be readily predicted using the methods described in this work.

6.5 Methods

Formation enthalpies and structures were retrieved for all binary nitrides and oxides in the Materials Project with < 20 atoms in formula (to avoid including defect structures), excluding azides and peroxides (due to their instability) and compounds with H, C, N, O, F, Cl, Br, I as cations (because they are unlikely to be solid at relevant temperatures).

Gibbs formation energies, ΔG_f were obtained as described in Ref. ¹⁷⁷:

$$\Delta G_f(0\text{ K}) = \Delta H_f(0\text{ K}) \quad (6.5)$$

$$\Delta G_f(T \neq 0\text{ K}) = \Delta H_f(0\text{ K}) + G^\delta(T) - \sum_i^N \alpha_i G_i(T) \quad (6.6)$$

ΔH_f is the Materials Project formation enthalpy, G^δ is the descriptor described in Ref. ¹⁷⁷, α_i is the stoichiometric weight of element α_i in the compound, and G_i is the Gibbs free energy of element, i .

Gibbs reaction energies, ΔG_r were calculated as the stoichiometrically weighted difference between Gibbs formation energies for the products and reactants of each reaction in the STAS cycle:

$$\Delta G_r(T) = \sum_i^{products} \nu_i \Delta G_{f,i}(T) - \sum_i^{reactants} \nu_i \Delta G_{f,i}(T) \quad (6.7)$$

where ν_i is the coefficient of species i that dictates 1 mol NH_3 generated per cycle at equilibrium (Equations 6.1-6.4).

The Gibbs energy minimization approach used to determine equilibrium product distributions and yields was adapted from Ref. ¹⁸⁷. The Gibbs energy of the system, G , is the sum of the product of molar compositions, x_i , and molar Gibbs energies, g_i over N species in equilibrium:

$$G = \sum_i^N x_i g_i \quad (6.8)$$

where the molar Gibbs energies, g_i , are taken to be the Gibbs formation energies, $\Delta G_{f,i}$ modified by the activities, a_i . R is the gas constant and T is the temperature:

$$g_i = \Delta G_{f,i}(T) + RT \ln a_i \quad (6.9)$$

The activity of solid phases is taken to be 1. The activity of gas phases is taken to be the partial pressure, p_i .

$$a_i = \begin{cases} 1 & \text{if solid} \\ p_i & \text{if gas} \end{cases} \quad (6.10)$$

The partial pressure is determined as the product of the number of moles of a certain gaseous species, n_i , and the total pressure, P , divided by the number of moles of gas in the system, n^g . The total pressure, $P = 1$ atm:

$$p_i = \frac{n_i}{n^g} P \quad (6.11)$$

$$n^g = \sum_i^m x_i \text{ for } m \text{ gaseous species} \quad (6.12)$$

Using a molar basis, the Gibbs function thus becomes:

$$\mathbf{G} = \sum_i^m x_i \left[\Delta G_{f,i}(T) + RT \ln \left(\frac{x_i}{\sum_i^m x_i} \right) \right] + \sum_i^{N-m} x_i \Delta G_{f,i}(T) \quad (6.13)$$

for m gaseous species and $N-m$ solid species. $\mathbf{G}$ is then minimized subject to the constraint of molar conservation:

$$\sum_i^N \alpha_{ij} x_i = b_j \quad (6.14)$$

where α_{ij} is the number of moles of element, j , in species i and b_j is the number of moles of element j in the feed. b_j is dictated for each reaction by **Equations 6.1-6.4** which yield 1 mol NH_3 per cycle if ΔG_r for all reactions = 0 (the equilibrium constants are equal to 1).

7 Descriptors for bonding in metal nitrides

Much of the content in this chapter appears verbatim in the following manuscripts:

C.J. Bartel, C.L. Muhich, A.W. Weimer, C.B. Musgrave, *ACS Applied Materials & Interfaces*, 8, 28, 18550-18559 (2016)

E. Arca, S. Lany, J.D. Perkins, C.J. Bartel, J. Mangum, W. Sun, A.M. Holder, G. Ceder, B. Gorman, G. Teeter, W. Tumas, A. Zakutayev, *Journal of the American Chemical Society*, 140, 12, 4293-4301 (2018)

W. Sun, C.J. Bartel, E. Arca, S. Bauers, B. Matthews, B. Orvañanos, B. Chen, L. Schelhas, M.F. Toney, W. Tumas, J. Tate, A. Zakutayev, S. Lany, A.M. Holder, G. Ceder (2018) [Under review] {arXiv:1809.09202}

7.1 Abstract

Exploratory synthesis in novel chemical spaces is the essence of solid-state chemistry. However, uncharted chemical spaces can be difficult to navigate, especially when materials synthesis is challenging. Nitrides represent one such space—where stringent synthesis constraints have limited the exploration of this important class of functional materials. Here, we employ a suite of computational materials discovery and informatics tools to construct a large stability map of the inorganic ternary metal nitrides. Our map clusters the ternary nitrides into chemical families with distinct stability and metastability, and highlights hundreds of promising new ternary nitride spaces for experimental investigation—from which we experimentally synthesized 7 novel Zn- and Mg-based ternary nitrides. By extracting the mixed metallicity, ionicity, and covalency of solid-state bonding from the DFT-computed electron density, we reveal the complex interplay between chemistry, composition, and electronic structure in governing large-scale stability trends in ternary nitride materials. In particular, the nature of bonding in aluminum nitride and zinc molybdenum nitrides are connected with the difficulty to hydrolyze the former and the propensity for the latter to undergo redox-mediated control of stability, crystal structure, and optoelectronic

properties and the inductive and reductive effects are shown to separate the space of stable ternary nitrides.

7.2 Introduction

Nitrides are an exciting space for materials design,^{186, 191, 192} as exemplified by state-of-the-art nitride materials for solid-state lighting,^{193, 194} ceramic hard coatings,¹⁹⁵ ammonia-synthesis catalysts,^{196, 197} permanent magnets,¹⁹⁸ superconductors,¹⁹⁹ superinsulators,²⁰⁰ electrides²⁰¹ and more. The nitride (N^{3-}) anion imparts unique electronic and bonding characteristics that are difficult to achieve in other chemical spaces, including hybridization of nitrogen $2p$ states with metal d states for useful optoelectronic and defect-tolerance properties,²⁰² as well as the formation of strong metal-nitrogen bonds leading to structural stability and mechanical stiffness.²⁰³

Despite much promise in the functionality of nitride materials, the nitrides are relatively underexplored, with fewer than 400 unique ternary metal nitrides catalogued in the Inorganic Crystal Structure Database (ICSD) in contrast to over 4,000 ternary metal oxides. The paucity of known nitrides can largely be attributed to the challenging requirements of nitride synthesis. Because the N_2 molecule is so stable, solid-state nitrides generally have small formation energies, decompose at high-temperature, and must be synthesized in oxygen- and water-free atmospheres to achieve high purity.^{186, 192, 204, 205} These stringent synthesis constraints, coupled with the poor intrinsic stabilities of nitrides, impose significant risk on the exploratory synthesis of novel nitride materials.

In our previous data-mining study of crystalline metastability,^{37, 177} we found nitrides to be the most metastable class of chemical compounds—having the largest fraction of metastable phases, as well as the highest average energies above the ground-state phases. The unusual metastability of nitrides can be attributed to the cohesivity afforded by strong metal-nitrogen bonds

in the solid-state, which can kinetically ‘lock-in’ metastable nitride structures. This metastability and cohesivity of the nitrides has interesting implications for the hydrolysis of aluminum nitride for ammonia synthesis, the redox-mediated stabilization of zinc molybdenum nitrides, and the discovery and exploration of the ternary nitride space, each of which are discussed within this chapter.

7.3 Results

7.3.1 Covalent bonding and the kinetics of aluminum nitride hydrolysis

Aluminum nitride (AlN) is a large, direct-band-gap material with a unique array of properties including high thermal conductivity ($319 \text{ W m}^{-1} \text{ K}^{-1}$, theoretical at room temperature),²⁰⁶ low electrical conductivity (resistivity $>10^{13} \Omega \text{ cm}$), high mechanical strength ($>400 \text{ MPa}$), and a thermal expansion coefficient ($4.3 \times 10^{-6} \text{ K}^{-1}$ at room temperature) near that of silicon.²⁰⁷ This remarkable combination of properties makes AlN highly valuable as a packaging material for integrated circuits and multichip modules. However, the propensity for AlN to hydrolyze and produce ammonia (NH_3) precludes the use of water during its processing, mandating the use of organic solvents, such as isopropanol, which increases the cost and complexity of manufacturing AlN substrates.²⁰⁸

AlN has also drawn interest as a metal-nitride redox material for NH_3 production from steam as a renewable alternative to the energy- and fossil-fuel-intensive Haber-Bosch process with significantly lower environmental impact. Numerous research efforts have sought an atmospheric pressure alternative to the Haber-Bosch process, which operates at extremely high pressure (up to 30 MPa)²⁰⁹ and consumes approximately 2% of the world’s energy production in the form of fossil fuels to both produce the H_2 that eventually becomes incorporated into NH_3 and drive the

technologically challenging conversion of H_2 and N_2 to NH_3 . In recent years, a strategy analogous to redox cycling for solar thermochemical water splitting¹⁶⁹ has emerged for the production of NH_3 using solar energy to drive a thermochemical redox cycle where a metal oxide is reduced with hydrogen, carbon monoxide, or solid carbon in the presence of nitrogen to produce a metal nitride that is subsequently hydrolyzed by steam to reform the metal oxide and produce NH_3 .¹⁷³

We have performed ab initio quantum chemical calculations on plausible pathways for the hydrolysis of the $(1\bar{1}00)$ surface of wurtzite AlN. First, we considered six unique H_2O adsorption geometries to identify the lowest energy configuration of dissociatively adsorbed H_2O . H_2O that dissociatively adsorbs to the AlN surface to form $\text{N-H}^* + \text{Al-OH}^*$ is the most stable of those evaluated. Subsequently, reaction paths involving a second adsorbed H_2O were considered for the formation of the NH_2^* intermediate. The NH_2^* intermediate forms via a surface hydroxyl-mediated proton-relay mechanism, which is a surface Grothuss-like proton relay²¹⁰ where adsorbed hydroxyls act to mediate the proton relay. The formation of NH_2^* is accompanied by the dissociation of one of the three Al-N bonds at the NH^* center. In this analysis, we also identify the dissociation of predominantly covalent Al-N bonds as the primary cause of the large barriers that results in the kinetic challenge of liberating nitrogen from AlN by hydrolysis in the case of thermochemical NH_3 generation. From the NH_2^* intermediate, we find a similar Grothuss-like mechanism for NH_3^* formation. During the proton migration through the proton relay and the Al-N bond dissociation step, we predict that a N vacancy is formed, which is subsequently filled by a surface O. In the case of low H_2O coverage, this step is partially concerted where the N vacancy forms transiently as it is concomitantly filled by an O surface atom and NH_3 forms. We elucidate a similar mechanism for the case of additional water present on the AlN surface. In the

case of higher H₂O coverage, the NH₃ generation step involves the formation of an intermediate N vacancy and NH₃*, where the N vacancy is subsequently filled by a surface O atom.

We calculate activation barriers for the overall hydrolysis reaction to be 330 or 359 kJ/mol for minimal water coverage and additional water coverage cases, respectively (**Figure 7.1**). These results are corroborated by previous experimental work that demonstrated that little or no conversion of AlN to NH₃ occurs below 950 °C.²¹¹ In the elucidation of this mechanism and the associated energetics, we present an enhanced understanding of the behavior of AlN and, more broadly, metal nitrides when exposed to water, thereby oxidizing and liberating nitrogen as NH₃. These results have important implications for designing processes that either avoid oxidation by hydrolysis (for example, in the case of processing AlN powder for thermal-management applications) or for designing materials that possess favorable thermodynamics and kinetics for solar thermal redox processes using metal nitrides to efficiently generate carbon-free NH₃. In the search for metal nitrides with fast-oxidation kinetics for carbon-free NH₃ synthesis, it is important to consider the metal–nitrogen bond strength, which can be correlated to the metal–oxygen bond strength of the accompanying oxide. The mechanism we predict for AlN hydrolysis involves nitrogen liberation in the form of NH₃ and large activation barriers associated with the dissociation of mostly covalent Al–N bonds, which explains the slow experimentally observed hydrolysis kinetics. However, a metal nitride with metal–nitrogen bonds with less covalent character will result in lower activation barriers to NH₃ generation and thus improved hydrolysis kinetics.

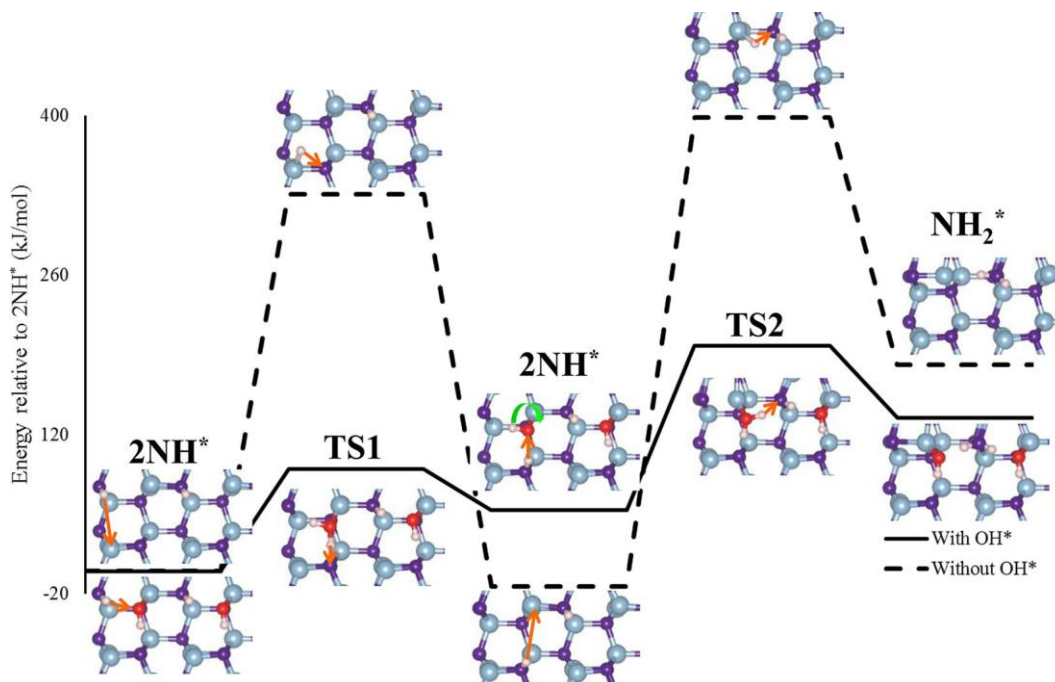


Figure 7.1 Rate-limiting step of AlN hydrolysis

Calculated schematic potential energy surfaces for NH_2^* formation from two neighboring NH^* sites with and without proton-transfer mediation by an OH^* ; NH_2^* formation mechanism where two sequential proton relay steps, each involving two sequential proton transfers, are mediated by the OH^* and NH_2^* formation mechanism in the absence of surface OH^* . Proton transfer mediated by OH^* substantially decreases the activation energy for both the proton “hop” to a vacant N site (TS1) and for the proton transfer to NH^* to form NH_2^* (TS2), demonstrating the ability of OH^* to catalyze proton transfers and, specifically, NH_2^* formation. Light blue is Al, purple is N, red is O, white is H.

7.3.2 Redox-mediated stabilization in zinc molybdenum nitrides

The success of nitrides semiconductors in light-emitting devices has stimulated interest in the design and discovery of novel nitride materials for optoelectronic applications and other functionalities. In addition to their potential application, nitrides are also interesting from the perspective of fundamental science: they are often metastable materials due to the inherent stability of the triple bond in the N_2 molecule and to the strength of the metal–nitrogen bonds.

Of particular interest from both applied and basic perspectives are molybdenum-based nitrides. In solid-state chemistry, Mo-based nitrides can form covalent, ionic, or metallic

compounds, depending on the Mo oxidation state. In ternary molybdenum nitrides, bonding depends on the electronegativity of the second cation, M . Highly electropositive elements (electronegativity $EN \leq 1$) tend to form ionic nitridomolybdates with Mo formally in a +VI oxidation state,²¹² for example, $M_3\text{MoN}_4$ with $M = \text{Ca, Sr, Ba}$ alkaline-earth metals. The electron donation from such electropositive elements stabilizes the $[\text{MoN}_4]^{6-}$ units in monomeric, dimeric, or oligomeric forms. The less electropositive elements ($1.55 \leq EN \leq 2.22$) tend to form more covalent compounds with $M\text{MoN}_2$ stoichiometry,²¹³ where $M = \text{Fe, Co, or Mn}$, and Mo is formally in a +IV oxidation state. Transition-metal elements (Fe, Co, Ni) and Mo can also form $M_3\text{Mo}_3\text{N}$ interstitial nitride compounds, which are metallic.²¹⁴

In other areas of chemistry, the ability of Mo to switch between different oxidation states enables its participation in various catalytic processes, in contrast to other transition-metal elements such as Cr or W.²¹⁵ For example, several organisms use a Mo-containing enzyme for the N_2 fixation process.²¹⁶ Similarly, Mo active sites cycle from +IV to +VI oxidation states in the formate dehydrogenase (FDH) enzyme, relevant to the CO_2 reduction process.²¹⁷ Mo-based nitrides have also been shown to have superior performance in both oxygen reduction reactions (ORRs) and hydrogen evolution reactions (HERs), with the Mo redox flexibility being pivotal for these electrocatalytic processes.^{218, 219}

Here, we report on the discovery of nitrides in the Zn–Mo–N chemical space, whose synthesis was facilitated by the ability of Mo to accommodate different oxidation states and switch between them. Despite the abundance of the elements, the Zn–Mo–N ternary system has not been reported in the literature. *Ab initio* calculations resulted in the prediction of two new compounds with ZnMoN_2 and Zn_3MoN_4 stoichiometries in this chemical space. Experiments show that these two compounds and the metastable alloys at intermediate compositions form in the wurtzite-

derived crystal structure. Retaining the same wurtzite-derived crystal structure across the ZnMoN_2 – Zn_3MoN_4 alloy composition is possible due to the ability of Mo to switch between +VI and +IV oxidation states and the stability of a covalent Zn–N framework. We also show that the optoelectronic properties of Zn–Mo–N vary dramatically as a function of composition: the Zn_3MoN_4 is resistive and has a relatively wide bandgap, whereas the ZnMoN_2 is a highly optically absorbing and electrically conductive material.

In order to get chemical insight into the bonding character of the calculated Zn_3MoN_4 ($\text{Pmn}2_1$) and ZnMoN_2 ($\text{P6}_3\text{mc}$) materials, we performed a Crystal Orbital Hamilton Population (COHP)²²⁰ and density-derived charge analysis of these compounds²²¹ (**Figure 7.2**). In addition, we compared these COHP analysis results to the chemically related Sr_3MoN_4 (Pbca) and MnMoN_2 ($\text{P6}_3/\text{mmc}$) materials (**Figure 7.3**) to understand the role of Zn in the stabilization of these new nitridomolybdates.

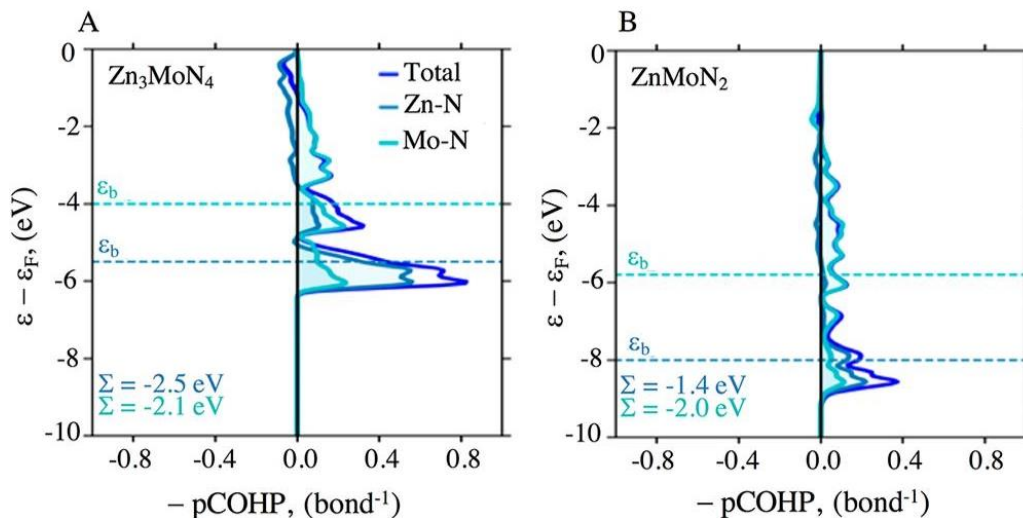


Figure 7.2 Bonding analysis in Zn-Mo-N

Crystal Orbital Hamilton Population (COHP) analysis partitioned by pairwise interactions Zn_3MoN_4 (Pmn₂₁) and ZnMoN_2 (P6_{3mc}). In each plot, bonding and antibonding interactions appear on the right and left side of the vertical black line, respectively. ϵ_b is the center (first moment) of the total bonding distribution, and Σ is the net bond energy.

As shown in **Figure 7.2**, the bonding states are qualitatively similar between the Zn_3MoN_4 in Pmn₂₁ structure and ZnMoN_2 in P6_{3mc} structure, with substantial bonding present for both Mo–N and Zn–N at similar energies. While the covalent bonding is deeper for both Zn–N and Mo–N in ZnMoN_2 compared with Zn_3MoN_4 , the extent of the covalent bonding (net bonding area (Σ)) is greater in Zn_3MoN_4 , primarily due to increased covalency of Zn–N interactions. The layered tetrahedra $[\text{ZnN}_4]/[\text{MoN}_4]$ of Zn_3MoN_4 thus enable increased overall covalency for Zn–N and increased chemical hardness compared with the alternating $[\text{ZnN}_4]$ tetrahedra and $[\text{MoN}_6]$ octahedra of the ZnMoN_2 arrangement. Zn_3MoN_4 also exhibits increased charge transfer between cations and nitrogen anions compared with the metallic ZnMoN_2 , which we quantify by analyzing the ratio of summed bond-orders to density derived electrostatic charges. This is to be expected as the density-derived charge increases by ~ 1 eV for Mo in Zn_3MoN_4 compared with ZnMoN_2 , aligning with the increase in oxidation state from +IV to +VI. The favorable mixed covalency bonding arrangement of the Zn_3MoN_4 structure leads to a significant bandgap (compared with the

electrically conductive ZnMoN_2) and retention of the wurtzite-derived structure even as the cation fraction of Zn, f_{Zn} , is decreased. The structural transition from $\text{Pmn}2_1$ to $\text{P6}_3\text{mc}$ would require significant rearrangement: from $[\text{ZnN}_4]$ and $[\text{MoN}_4]$ tetrahedra in Zn_3MoN_4 ($\text{Pmn}2_1$), to layered $[\text{ZnN}_4]$ tetrahedra and $[\text{MoN}_6]$ octahedra in ZnMoN_2 ($\text{P6}_3\text{mc}$). Thus, as f_{Zn} is decreased from 0.75 to 0.50 and the thermodynamic driving force for forming Mo octahedra is increasing, the metastable wurtzite-derived structure remains favorable for Zn bonding and is kinetically preserved due to the reconstructive barrier toward forming the ordered layers of the ZnMoN_2 structure.

To gain insights into the chemical and structural features of the Zn_3MoN_4 - ZnMoN_2 alloys, and the role of Zn in stabilizing these materials, we analyzed the computationally predicted electronic structure of the end-members and related ternary nitridomolybdates that do not contain Zn. Specifically, we analyzed two isoelectronic structures, Sr_3MoN_4 and Zn_3MoN_4 , for the high Mo oxidation state, and two isoelectronic structures, MnMoN_2 and ZnMoN_2 , for the lower oxidation state, with electronegativity being 0.95, 1.65, and 1.55 for Sr, Zn, and Mn, respectively. The results are shown in **Figure 7.3**. In contrast with Sr_3MoN_4 , Zn_3MoN_4 is found to prefer the $\text{Pmn}2_1$ symmetry to the Pbca symmetry. The $\text{Pmn}2_1$ structure enables Zn_3MoN_4 to gain an additional 7% in net bonding area compared with Pbca , most of which can be attributed to increased Zn-N interactions. Sr_3MoN_4 does not benefit as significantly from added Sr-N interactions in the $\text{Pmn}2_1$ structure and slightly favors the Pbca symmetry within PBE+U. More than 99% of the covalent bonding contribution in Sr_3MoN_4 arises from Mo-N bonding interactions, in contrast to Zn_3MoN_4 , where bonding comprises ~25% Zn-N and ~75% Mo-N in the $\text{Pmn}2_1$ structure (quantified by bonding area below the Fermi energy). The ability for Zn to adopt a low oxidation state (+II) but still participate significantly in M-N bonding alongside a transition metal

(in this case Mo) distinguishes it from the highly ionic bonding of the alkali-earth elements (e.g., Sr) and makes it possible for this element to stabilize the Mo in the +VI oxidation state—despite its electronegativity being substantially higher than that of alkali-earth metals and relatively comparable to that of Mo. In contrast with MnMoN_2 , ZnMoN_2 is found to prefer the $P6_3mc$ symmetry to the $P6_3/mmc$ symmetry. Interestingly, the net covalent bonding area increases significantly for MnMoN_2 going from $P6_3/mmc$ to $P6_3mc$, yet this is not enough to stabilize the $P6_3mc$ structure. This is likely attributed to increased chemical hardness of the Mo-N bonding states for the $P6_3/mmc$ structure as the Mo-N bonding center shifts down ~ 1 eV compared with the $P6_3mc$ structure. Similarly to the Zn_3MoN_4 case, the Zn-N interaction is pivotal in stabilizing the $P6_3mc$ structure, which is not preferred for isoelectronic MnMoN_2 .

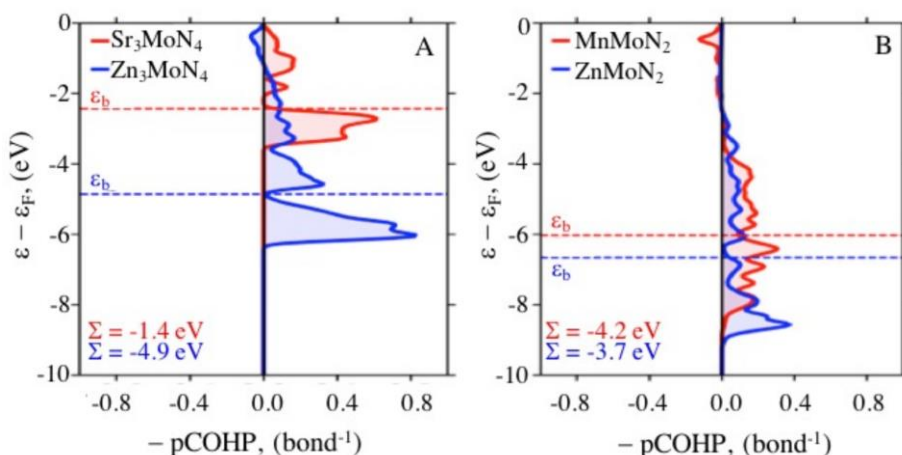


Figure 7.3 Bonding for alternative cations in X-Mo-N

Crystal Orbital Hamilton Population (COHP) analysis partitioned by pairwise interactions for ground-state polymorphs of Zn-based ternary nitridomolybdates introduced in this work (Zn_3MoN_4 – $Pmn2_1$; ZnMoN_2 – $P6_3mc$), compared with the Sr- and Mn-based nitrides of the same stoichiometry. In each plot, bonding and antibonding interactions appear on the right and left side of the vertical black line, respectively.

7.3.3 New stable ternary nitrides

High-throughput computational materials science has emerged as a new paradigm for materials discovery,^{104, 222} helping to guide experimental synthesis efforts across broad and

uncharted chemical spaces. Here, we employ a suite of computational materials discovery^{132, 133, 223, 224} and informatics^{26, 225} tools to survey, visualize, and most importantly, explain stability relationships across the inorganic ternary metal nitrides. Our investigation proceeds in three steps. First, we use crystal structure prediction algorithms to probe the energy landscapes of previously unexplored ternary nitride spaces, surveying novel nitride compounds over 962 M_1 - M_2 -N spaces. We predict a large list of new stable and metastable ternary nitrides, significantly extending the known thermochemical data in this space. Guided by these predictions, we experimentally synthesize 7 new Zn- and Mg-based ternary nitrides, and identify hundreds of promising new ternary nitride systems for further exploratory synthesis.

Much like how Mendeleev's Periodic Table revealed the underlying structure of the elements, an effective visual organization can reveal hidden relationships and chemical families within the ternary metal nitrides. Assisted by unsupervised machine learning algorithms, we next clustered together metals that have a similar propensity to form stable or metastable ternary nitrides. We used these clustered nitride families to construct a large and comprehensive stability map of the inorganic ternary metal nitrides. Not only does our map illustrate broad overarching relationships between nitride chemistry and thermodynamic stability, it further inspires us to rationalize these trends from their underlying chemical origins.^{226, 227} To do so, we extracted from the DFT-computed electron density the mixed metallicity, ionicity, and covalency of solid-state bonding—providing new chemical features to interpret the electronic origins of nitride stability. We show that the nitrogen anion can be surprisingly amphoteric in the solid-state, usually acting as an electron acceptor in nitrogen-rich nitride ceramics, but remarkably, sometimes serving as an electron donor (relative to competing binary nitrides) to stabilize nitrogen-poor metallic nitrides.

Beyond the nitrides, there remain many other unexplored chemical spaces awaiting experimental discovery. Our computational approach here can be further applied to these uncharted chemical spaces, not only to predict and synthesize new compounds, but also to visualize general trends over broad compositional spaces—providing maps and chemical rules to help experimental chemists more rationally navigate exploratory synthesis at the frontier of solid-state chemistry.

In this work, we explore ternary nitrides over a 50×50 M_1 - M_2 -N composition space, where M consists of the 50 most common cations in the known nitrides. These cations broadly sample the periodic table; spanning over the alkali, alkaline earth, transition, precious, and post-transition metals, as well as the main group elements B, C, Si, S and Se. Within this composition matrix, known ternary nitride compounds exist over only 303 M_1 - M_2 -N spaces (~25%). To fill in the missing spaces, we first conducted a high-throughput computational search for novel ternary nitride compounds. Previous computational searches for ternary nitrides have been constrained to either limited composition spaces,^{213, 228, 229} or specific crystal structures.²³⁰⁻²³² Here, we broadly sample over both composition and crystal structure, using a data-mined structure predictor (DMSP)²³³ to perform rational chemical substitutions on the known ternary nitrides, creating unobserved but reasonable novel ternary nitride phases *in silico*.

In a previous work, we trained a DMSP specifically for nitrides discovery, by data-mining which chemical substitutions in the solid-state pnictides are statistically probable.⁴⁶ We found a substitution matrix trained on a pnictides training set to be more predictive for nitride discovery than a substitution matrix trained over all inorganic solids, which otherwise becomes biased towards ionic substitutions that are common in the more thoroughly-explored oxides and chalcogenides. In general, the substitution relationships in oxides are not applicable to nitrides due

to differences in structure-types, elemental coordination, and metal redox chemistry for O^{2-} vs. N^{3-} anions. Here with the pnictides-trained DMSP, we extrapolate the 340 known ternary nitrides (213 stable + 127 metastable) to 6,000 hypothetical ternary nitride structures, sampled over 962 ternary M_1 - M_2 -N spaces.

Using density functional theory (DFT), we computed the formation enthalpies of these DMSP-generated nitrides, which are then used to probe the stability landscapes of unexplored ternary nitride spaces. We evaluate the phase stability of these candidate structures leveraging the tools and precomputed data from the Materials Project database.^{16, 118} **Table 7.1** summarizes the results from this screening. Notably, we predict 203 new stable ternary nitride compounds, nearly doubling the 213 previously-known stable ternary nitrides. These stable ternary nitrides span 277 ternary M_1 - M_2 -N spaces, 92 of which were not previously known to contain any stable ternary compounds. We have made the structures and energies of the newly predicted nitrides freely available on the Materials Project for readers interested in further investigation. We note that a small subset of these stable ternary nitrides have been identified in previous computational searches,^{213, 228-232} which we have reconfirmed here.

Permuting chemistry and crystal structure on the known ternary nitrides offers a computationally efficient probe of formation energies over broad ternary nitride compositions. One limitation of the DMSP is that if the structural prototype of a ground-state nitride has never been observed before, then the DMSP cannot predict it. Nevertheless, because most ternary nitride spaces are unexplored, the prediction of any ternary nitride structure with negative formation energy in an otherwise empty chemical space implies that the true ground-state structures and compositions must be even lower in energy—therefore highlighting that ternary space as a compelling target for further theoretical and experimental investigations.

A large list of predicted compounds is difficult to navigate and does not provide an intuitive picture of the structural form of a chemical space.²³⁴ When Mendeleev constructed the Periodic Table, he produced a conceptual framework to orient our understanding of the relationships and trends between the chemical elements. In higher-order chemical spaces—binaries, ternaries, etc.—these trends become increasingly challenging to extract by hand. Here, we elucidated the structural form of the ternary nitride space using hierarchical agglomeration²³⁵ to cluster together metals with a similar propensity to form either stable or metastable ternary nitrides. In order to capture both large-scale stability trends, as well as local chemical relationships, we built a multi-feature distance metric that considers for each ternary nitride whether it is stable or metastable, its formation energy, and which periodic group the metal lies in. These multiple features represent mixed data-types (nominal, continuous and ordinal, respectively), which we combined into a single distance metric using Gower’s method.²³⁶

The agglomeration algorithm clusters elements hierarchically by minimizing this multi-feature distance matrix. The resulting dendrogram provides a phenotypic representation of the nitride chemical families, grouped by their thermochemical stabilities. With this 1D ordering of metals, we produced a clustered heat map of the inorganic ternary metal nitrides, shown in **Figure 7.4**, colored by the stability of the lowest formation-energy ternary metal nitride in each M_1 - M_2 -N chemical space. Our clustering algorithm parses the ternary nitrides map into distinct regions of stability (blue), metastability against binaries (green), and metastability against elements (red)—highlighting stable ternary nitride spaces that are promising for further exploratory synthesis, and metastable spaces where successful synthesis may require non-equilibrium synthesis routes. An interactive version of the map, with ternary phase diagrams and compound stability information for each M_1 - M_2 -N system, is available at <https://wenhaosun.github.io/TernaryNitridesMap.html>.

Table 7.1. Statistics of the known and predicted ternary metal nitrides, categorized by the thermodynamic stability of ternary M_1 - M_2 -N spaces, and specific ternary $A_xB_yN_z$ phases within those spaces. All spaces are categorized by the H_f of the lowest formation-energy ternary nitride. Metastable phases are categorized by their energy above the convex hull, E_{hull} .

Ternary M_1 - M_2 -N Spaces			
	Previous	New	Portion
Systems with Stable Ternary Nitrides (Blue)	189	92	281 (29%)
- Stable Alkali-Metal-Nitride Systems	124	76	200
- Stable Metal-Metal-Nitride Systems	65	16	81
Metastable vs. Stable Binaries, $H_f < 0$ (Green)	98	322	420 (44%)
Metastable vs. Elements, $H_f > 0$ (Red)	20	241	261 (27%)
Ternary $A_xB_yN_z$ Phases			
	Previous	New	Portion
Stable Ternary Phases	213	203	416
Metastable, $E_{\text{Hull}} < 70$ meV/atom	39	36	75
Metastable, $E_{\text{Hull}} < 200$ meV/atom	85	175	260
Metastable: Stabilizable $\Delta\mu_N < +1$ eV/N	3	92	95

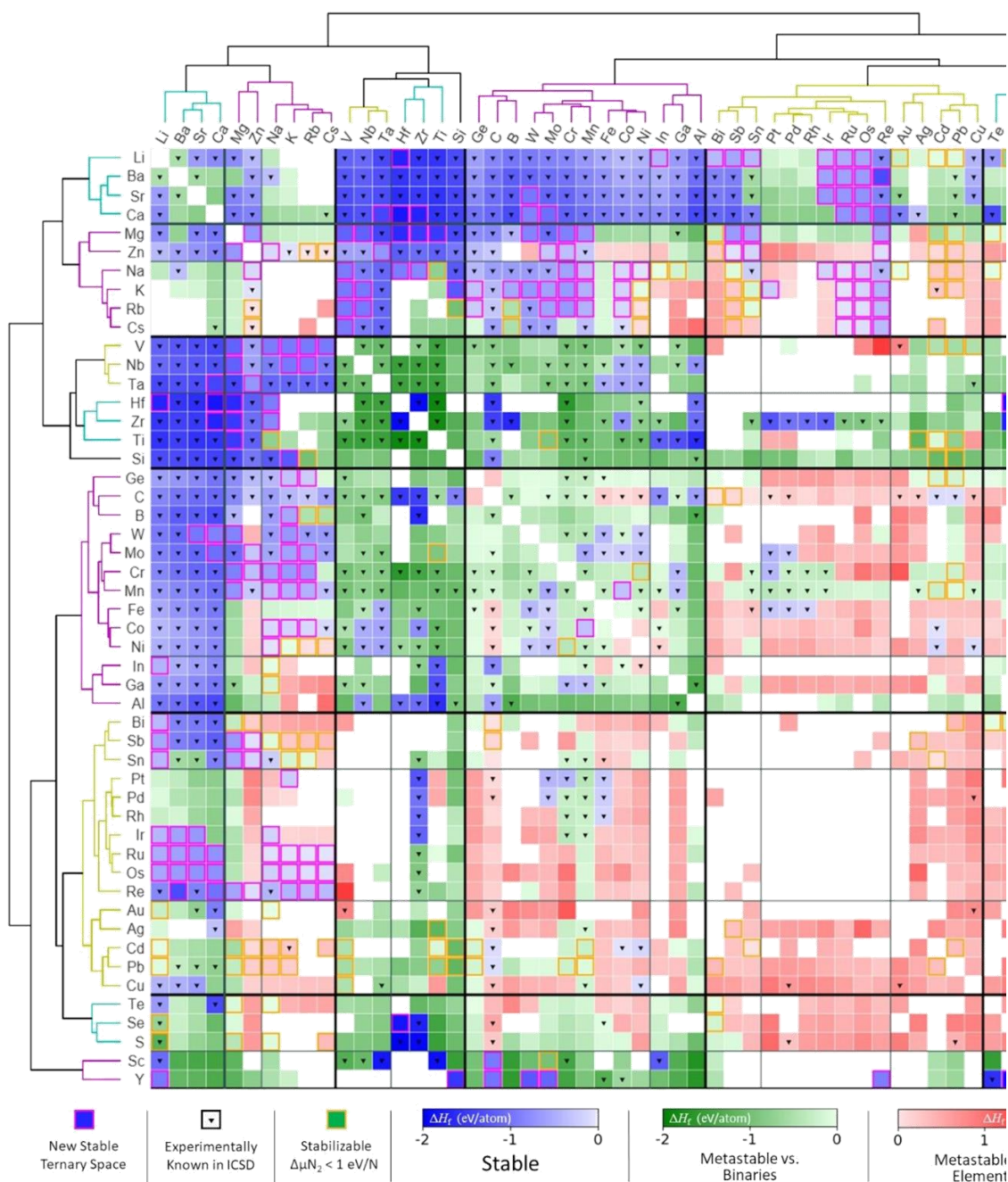


Figure 7.4 Stability map of ternary nitrides

Map of the inorganic ternary metal nitrides, colored by the thermodynamic stability of the lowest formation-energy ternary nitride. Blue: stable ternary nitrides on the convex hull; Green: ternaries with $H_f < 0$ but metastable with respect to binaries; Red: ternaries metastable with respect to elements, $H_f > 0$. Triangles represent ternary nitride systems with entries in the ICSD. White spaces indicate that the DMSP did not find probable chemical substitutions to create a structure in that system. Elements are clustered on multiple features to indicate their propensity to form stable or metastable ternary nitrides. These clustered elements are represented phenotypically by a dendrogram, which parses the ternary nitrides map into regions of distinct stability and

metastability. An interactive version of the ternary nitride map, with phase diagrams and compound stability information for each ternary system, is available at <https://wenhaosun.github.io/TernaryNitridesMap.html>.

Thermodynamically stable ternary nitrides are relatively rare, comprising only a quarter of the map, which is likely a confounding factor in the difficulty of ternary nitride discovery. Alkali and alkaline earth ternary nitrides (Alk-Me-N) represent a majority of the stable ternary nitride spaces ($200/281 = 71\%$), whereas stable non-alkali metal-metal-nitrides (Me-Me-N) are less common, with small islands of stability scattered amongst the mixed transition- and precious-metal nitrides.

The clustered dendrogram distinguishes between three major groups of alkali/alkaline earth elements in their ability to form ternary nitrides. The first group is composed of Li, Ca, Sr, and Ba, which form ternary nitrides with negative H_f with all elements, most of which are thermodynamically stable. The alkali ions Na, K, Rb, and Cs also form stable ternaries with early and first-row transition metals, although they generally react unfavorably ($\Delta H_f > 0$) with precious metals and metalloids. The clustering algorithm places Mg and Zn as intermediate between these two groups. Although one might anticipate Mg to be chemically similar to Ca, Sr and Ba, Mg is less reactive than the other alkaline earth metals—forming ternaries less exothermically and forming fewer stable ternary nitrides overall. For a transition metal, Zn is relatively electropositive, meaning it can react like an alkali when coupled with early transition metals; but when coupled with late- and post-transition metals, Zn-containing ternaries generally have positive formation energy.

Of the 281 spaces with stable ternary nitrides, the 92 indicated on the map by a magenta box do not have any ternary nitride entries in the current ICSD, and therefore represent new theoretical predictions. While most stable ternary nitrides containing Li, Ca, Ba, and Sr have been

synthesized experimentally, ternary nitrides with Na, K, Rb, and Cs have not been as readily explored. This may be because Na, K, Rb, and Cs do not form stable binary nitrides, meaning that their nitride precursors for solid-state synthesis are less conveniently available. Nevertheless, these compositions are promising for further experimental synthesis. Scandium and yttrium form several new stable ternaries, with some of the most negative formation energies on the map. However, the large exothermic formation energies of binary ScN and YN renders most Sc- and Y-based ternary nitrides to be metastable against decomposition, to the extent that the clustering algorithm categorizes Sc and Y as independent chemistries from the rest of the transition metals. Interestingly, we predict the precious metals Ir, Ru, Re, Os to form stable ternary nitrides when synthesized with most alkali and alkaline earth metals, representing new families of stable ternary metal nitrides that await experimental discovery.

7.3.4 Experimental synthesis of predicted ternary nitrides

From our predictions, we identified Zn- and Mg-based ternary nitrides as compelling target spaces for novel materials synthesis. Using magnetron sputtering, we successfully synthesized crystalline nitride thin-films in 7 new ternary nitride spaces: Zn-Mo-N, Zn-W-N, Zn-Sb-N, Mg-Ti-N, Mg-Zr-N, Mg-Hf-N, Mg-Nb-N. Concurrently, we conducted an unconstrained DFT search of ground-state structures and compositions using Kinetically Limited Minimization,¹⁶² with resulting compositions, structures and formation energies tabulated in **Figure 7.5a**. As illustrated in **Figure 7.5b**, these Zn-based ternaries adopt a wurtzite-derived structure, whereas the Mg-based ternaries form in a rock salt-derived structure. **Figure 7.5c** shows experimental synchrotron X-ray diffraction (XRD) patterns of these novel synthesized nitrides together with reference patterns for rock salt (NaCl) and wurtzite (ZnS), adjusted to lattice parameters of $a=4.5 \text{ \AA}$ and $a=3.3 \text{ \AA}/c=5.4 \text{ \AA}$ to approximate the average peak positions in the

experimental patterns. The experimental XRD patterns match the peak positions and intensities of these high-symmetry structures well, with differences in relative intensities arising from both textured growth of the thin films, as well as different scattering powers within the unit cell. Notably, we do not observe peak-splitting relative to the ideal wurtzite or rocksalt structures, suggesting disorder on the cation sites, which is not uncommon for nitrides deposited at low and moderate temperatures.²³⁷ Historically, the rate of discovering new ternary nitride M_1 - M_2 -N spaces has averaged ~ 3.3 per year, as illustrated in **Figure 7.5d**. Our rapid experimental realization of novel nitrides in 7 previously unexplored M_1 - M_2 -N spaces validates the predictions from the map, bolsters confidence in the 85 other predicted spaces with stable nitrides, and highlights the valuable role of computational materials discovery in accelerating exploratory synthesis in novel chemical spaces.

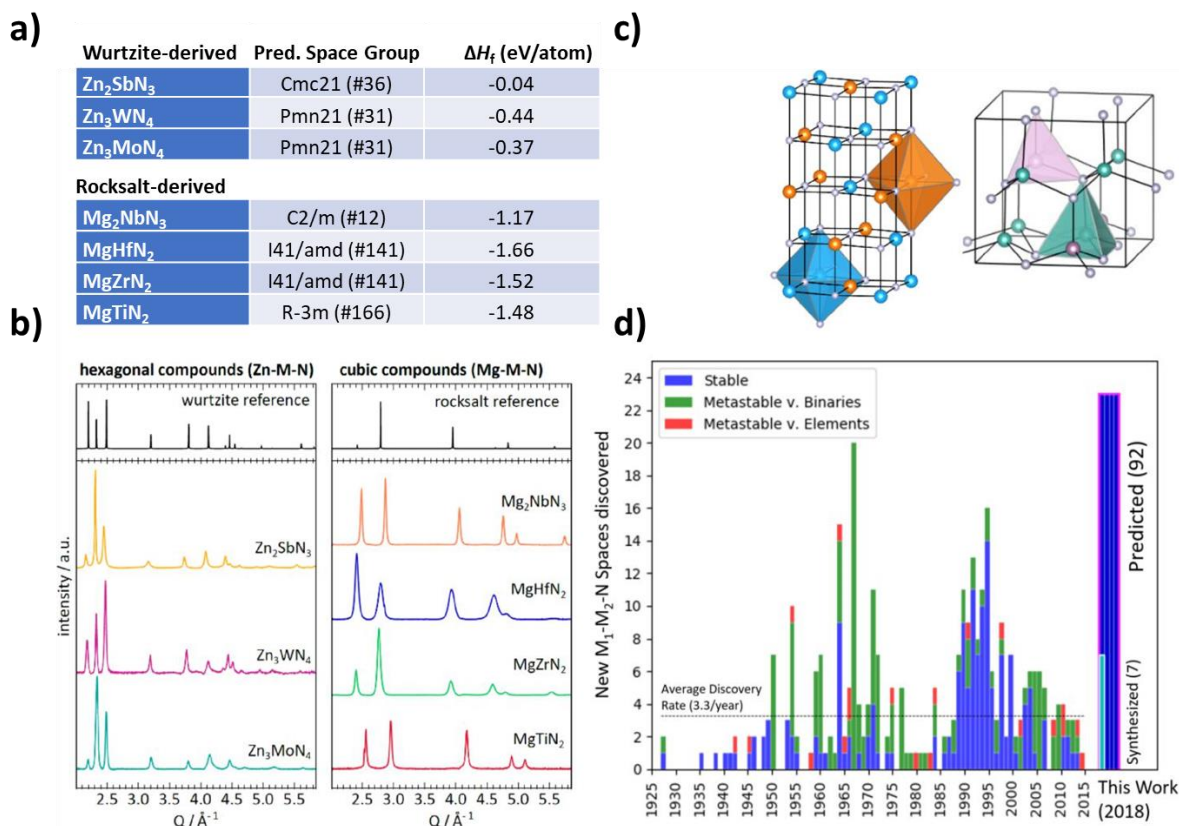


Figure 7.5 Experimental confirmation of predicted nitrides

a.) Seven new Zn- and Mg- based ternary nitrides, with structures predicted by unconstrained structure search, and their corresponding formation energies. **b.)** Mg-based ternary nitrides are deposited in rocksalt-derived structure (left), and Zn-based ternaries in a wurtzite-derived structure (right). **c.)** Synchrotron measured XRD patterns of new Zn- and Mg-based ternary nitrides, shown with reference diffraction patterns for lattice parameter-adjusted wurtzite and rocksalt. Q relates to diffraction angle (θ) and incident wavelength (λ) by $Q = (4\pi/\lambda)\sin(\theta)$ and $\lambda = 0.9744 \text{ \AA}$. **d.)** Discovery histogram for new ternary nitride spaces, based on entries as catalogued in the ICSD.

7.3.5 Metastable ternary nitrides

Ternary nitrides that are metastable against decomposition into binary or elemental phases comprise the majority (71%) of the surveyed spaces. Most of the metastable nitrides with $H_f < 0$ (green) are mixed-transition metal nitrides, whereas mixed precious/post-transition metal nitrides typically have $H_f > 0$ (red). Although metastable nitrides should, in principle, be difficult to synthesize, ternary nitrides have been experimentally realized in 118 of the computed metastable

spaces, shown in **Figure 7.4** by the inverted triangles. In our previous data-mining study of crystalline metastability,^{37, 238} we found nitrides to be the most metastable class of chemical compounds—having the largest fraction of metastable phases, as well as the highest average energies above the ground-state phases. The unusual metastability of nitrides can be attributed to the cohesivity afforded by strong metal-nitrogen bonds in the solid-state, which can kinetically ‘lock-in’ metastable nitride structures.

By formulating rational synthesis strategies to these metastable nitrides, we can expand the design space of functional nitride materials beyond equilibrium phases and compositions. One thermodynamic route to metastable nitrides is via nitrogen precursors that are less strongly bound than triple-bonded N_2 ⁴⁶ such as ammonia,¹⁸⁶ azides,²³⁹ or high-pressure supercritical N_2 .²⁴⁰ As an extreme example, plasma-cracked N_2 can yield atomic N precursors with nitrogen chemical potentials up to $\Delta\mu_N \approx +1$ eV/N.^{241, 242} Thin-film synthesis from these precursors can form remarkably metastable nitrides, such as $SnTi_2N_4$ (metastable by 200 meV/atom),²⁴³ or $ZnMoN_2$ in a wurtzite-derived structure (metastable by 160 meV/atom).¹⁶² In **Figure 7.4**, we use orange boxes to highlight 95 spaces with metastable ternary nitrides predicted to be stabilizable under elevated nitrogen chemical potentials of $\Delta\mu_N < +1$ eV/N.

Metastable ternary nitrides can also be obtained via soft solid-state synthesis routes; for example, delafossite $CuTaN_2$ is metastable by 127 meV/atom, but can be synthesized by ion-exchange metathesis of Cu^+ for Na^+ from the stable $NaTaN_2$ phase.²⁴⁴ Amorphous precursors can also be a route to ternary nitrides that are metastable with respect to phase separation, whereby an atomically homogeneous amorphous precursor with the target ternary nitride composition is gently annealed to a lower energy, but still metastable, target crystalline phase.^{153, 245, 246} On a separate note, decomposition of metastable ternary nitrides can also result in interesting functionality; for

example, the segregation of metastable Si-Ti-N alloys at high temperature results in complex TiN/Si₃N₄ layered heterostructures with superior mechanical properties for tribological applications.²⁴⁷

7.3.6 Thermochemical decomposition into competing binaries

By clustering the ternary nitrides space, we constructed a map that reveals broad overarching relationships between nitride chemistry and thermodynamic stability. We found alkali and alkaline earth ternary nitrides (Alk-Me-N) to comprise the majority of the stable ternary nitrides; whereas non-alkali metal-metal-nitrides (Me-Me-N) were generally found to be metastable with respect to phase separation — albeit with some curious exceptions.

We can achieve some insights towards this by considering the geometric requirements of thermodynamic stability. A ternary nitride is stable if it is lower in free-energy than any stoichiometric combination of its competing ternaries, binaries, or elemental constituents. In formation energy versus composition space, this stability requirement manifests geometrically as a convex hull, illustrated for a ternary *A-B-N* space in **Figure 7.6a**. We can therefore rationalize the stability of a ternary nitride from 1) a thermochemical perspective—if a ternary nitride is lower in energy than its competing binary nitride(s), and 2) from a solid-state bonding perspective—how two metals interact electronically within a ternary nitride to raise or lower the bulk lattice energy of the ternary compound.

To quantify the thermochemical propensity of a ternary nitride to decompose into its competing binaries, we first define a feature named the ‘depth of the binary hull’, referring to the lowest energy binary nitride in a binary Me-N space. This binary hull depth, illustrated in **Figure 7.6a** by a black dashed line, serves as a proxy for the strength of the pairwise metal-nitrogen bond in the solid-state. We note that in some binary nitride spaces, the lowest formation-energy binary

nitride has positive formation energy—for example, Cu_3N in the Cu-N hull, indicating that Cu_3N decomposes to $\text{Cu} + \text{N}_2$ under ambient conditions.

Figure 7.6b shows for each element how many stable ternary spaces it forms in, versus the ‘depth’ of the binary hull. A volcano plot emerges, where elements that have either shallow or deep binary nitride hulls tend not to form many stable ternary nitrides, whereas elements that have intermediate binary nitride hull depths (around -0.8 eV/atom) form stable ternary nitrides most readily. From a thermochemical perspective, when the binary hull is deep, there is a greater propensity for a ternary metal nitride to phase-separate into its competing low-energy binary nitrides. On the other hand, a shallow (or positive) binary hull depth indicates intrinsically weak metal-nitrogen bonding; meaning ternary nitride formation is probably unfavorable in the first place. Intermediate binary hull depths indicate favorable metal-nitrogen bonding, but not enough for decomposition of a ternary nitride into its binary constituents, offering a compromise between these two competing effects.

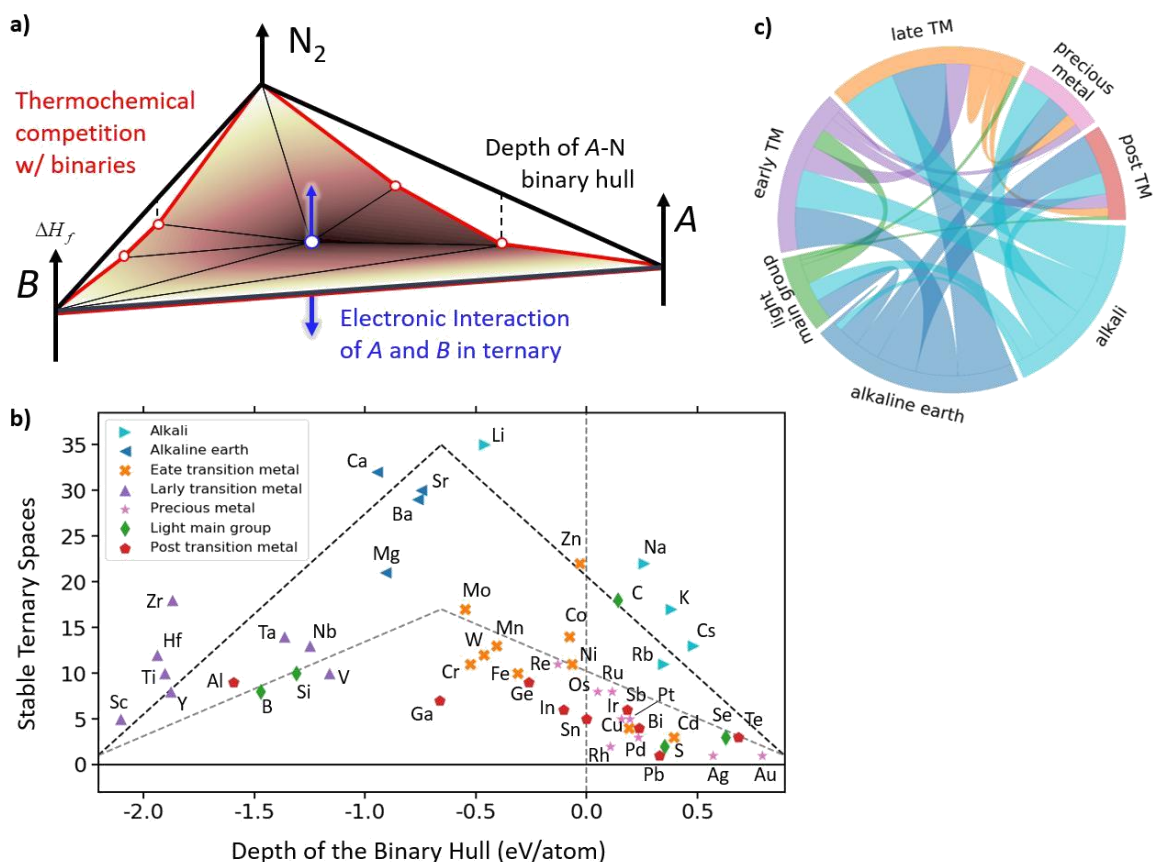


Figure 7.6 Thermodynamic perspective of ternary nitride stability

a.) 3-D projection of a convex hull in a ternary A-B-N space, where the vertical axis is formation energy and the horizontal triangular plane is composition. The stability or metastability of a ternary nitride compound (blue circle) is governed by its propensity to decompose into competing binaries (red lines), as well as by its lattice energy arising from the electronic interaction of two metals in the ternary nitride (blue arrows). **b.)** Scatterplot showing the number of stable ternary systems each metal appears in, plotted against the depth of the binary hull, which corresponds to the formation energy of the lowest-energy binary nitride in the M-N binary space. Eye-guides for the volcano trend provided by dashed lines **c.)** Chord diagram showing the frequency of inter-group and intra-group bonding relationships of stable ternary nitrides.

In **Figure 7.6b**, the alkali and alkaline earth metals stand out on the volcano – forming stable ternaries more readily than other elements with similar binary hull depth. While the volcano plot captures the propensity of individual metals to decompose from a ternary nitride into their corresponding binaries, it does not capture how two metals influence each other electronically within the ternary nitride, which governs the bulk lattice energy of the ternary compound. **Figure 7.6c** displays in a chord diagram which periodic groups the two metals A and B in a stable ternary

come from, where the width of each chord indicates the frequency of stable ternary spaces with that combination. Our chord diagram shows that the two metals in a ternary compound tend to originate from different groups across the periodic table, where one of the elements is often an alkali or alkaline earth metal. This observation is consistent with heuristics arising from Hard-Soft/Acid-Base (HSAB) theory,²⁴⁸ which suggests that ternary nitrides form most readily when the two acids (cations in ternary nitrides) have different HSAB character.²⁰⁵

7.3.7 Electronic origin of ternary nitride stability

Qualitatively, we expect differences in electronegativity between *A*, *B* and N to redistribute the electron density into different bonds, which in the solid-state, may have mixed metallic, ionic and covalent character. Inspired by the role of ‘conceptual DFT’ in rationalizing the reactivity of atoms and molecules,^{226, 227} we constructed new semi-quantitative schemes to extract the nature of solid-state bonding from the DFT-computed electron density. We computed ionic character of each ion from the ratio of the stoichiometrically-normalized Net Atomic Charges (NAC) to the Summed Bond Order (SBO) obtained from the Density Derived Electrostatic and Chemical (DDEC) approach.²²¹ We use Crystal Orbital Hamiltonian Population calculations²²⁰ to decompose the integrated bonding energies of metal-metal interactions (*A-A*, *A-B*, *B-B*) as metallicity,²⁴⁹ and non-metal interactions (*A-N*, *B-N*, *N-N*) as covalency. Using these features, we obtained data-driven insights into how this mixed solid-state bonding character influences ternary nitride stability.

We visualize our results on the classic metallic-ionic-covalent axes of van Arkel triangles²⁵⁰ shown in **Figure 7.7a**, using hexagonal-binned histograms to represent the scatter distribution on these triangles. From **Figure 7.7a**, we see that stable Alk-Me-N ternaries tend to exhibit greater ionicity and metal-nitrogen covalency, whereas stable Me-Me-N ternaries generally

have higher metallicity. This distinction becomes even more apparent when the triangles are further separated by nitrogen-rich and nitrogen-poor nitrides, where this nitrogen excess or deficiency is compositionally referenced against the deepest-hull binary nitrides. For example, the formation of ‘nitrogen-rich’ Ca_2VN_3 from Ca_3N_2 and VN requires excess nitrogen, which is accompanied by the formal oxidation of vanadium from the binary nitride V^{3+}N to the ternary $\text{Ca}_2\text{V}^{5+}\text{N}_3$. On the other hand, formation of ‘nitrogen-poor’ compositions occur by nitrogen release, such as the formation of $\text{Co}_2\text{Mo}_3\text{N}$ from CoN and MoN.

Stable Alk-Me-N are mostly nitrogen rich, whereas most stable Me-Me-N are nitrogen poor. This dichotomy between nitrogen-rich and nitrogen-poor ternary nitrides can largely be rationalized by how electron density redistributes between the nitrogen anion and the more electronegative metal during a reaction from the deepest-hull binaries to a stable ternary nitride. **Figure 7.7b** shows the changes in ionicity of the nitrogen anion, $\Delta\delta_{\text{N}}$, and the more electronegative metal cation, $\Delta\delta_{\text{B}}$, during such a reaction. The formation of nitrogen-rich nitrides typically involves *B*-metal oxidation and nitrogen reduction, whereas the formation of nitrogen-poor nitrides shows the opposite, exhibiting metal reduction and nitrogen oxidation. We emphasize that this nitrogen oxidation and reduction is measured relative to the nitrogen anion from the corresponding binary nitrides, not to the N_2 molecule.

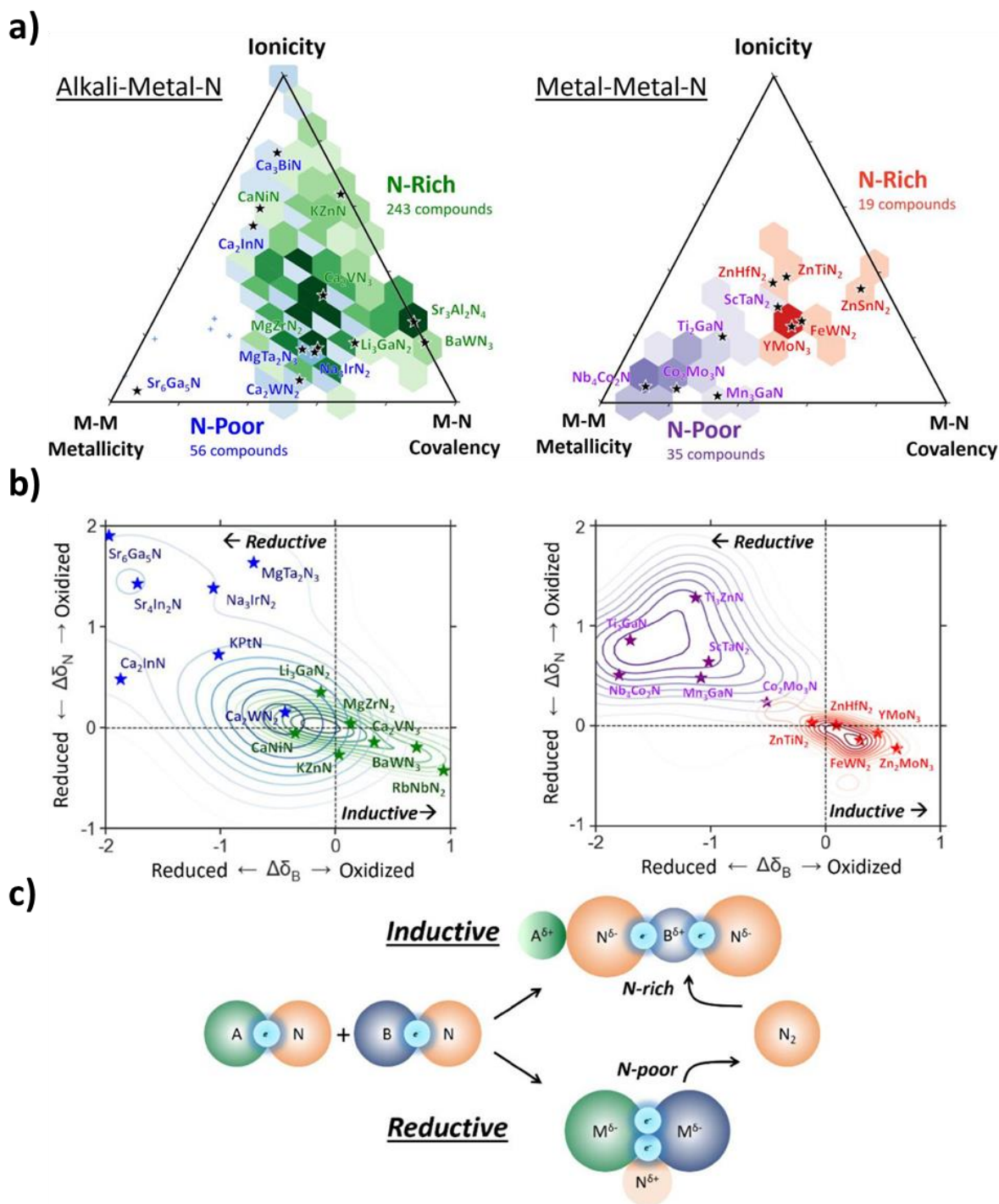


Figure 7.7 Rationalizing nitride stability through bonding analysis

a.) Metallicity, ionicity, and covalency of the stable ternary nitrides; hexagonally binned on van Arkel triangles by the nitrogen-excess or nitrogen-deficiency of the ternary, compositionally-referenced against the deepest-hull binary nitrides. Hexagons plotted for regions with >2 data points only. Color intensity corresponds to number density in each hexagon. Outliers in the Alk-Me-N triangle are shown with small crosses. **b.)** Kernel density distributions of ion oxidation

and reduction between a deepest-hull binary and the stable ternary nitride, for nitrogen (vertical axis) and the more electronegative metal, *B* (horizontal axis). **c.) Inductive effect**, electropositive metal *A* donates electron density to *B*-N covalent bond, oxidizing the more electronegative metal, which can lead to nitrogen-rich nitrides. *Reductive effect*, nitrogen oxidation or nitrogen release provides electrons to Me-Me bonds, reducing the metals and increasing metallicity.

The formation of nitrogen-rich nitrides can be rationalized primarily from the *Inductive Effect*,^{191, 251} where an electropositive metal, *A*, donates electron density to its adjacent nitrogen anion, driving the formation of strong nitrogen covalent bonds with the more electronegative metal, *B*. As illustrated in **Figure 7.7**, this electron donation from *A* generally leads to nitrogen reduction, which in turn oxidizes the metal *B*. Significant oxidation of *B* can be compensated by excess nitrogen— which explains the formation of nitrogen-rich nitrides. An oxidized cation and reduced anion increases the overall ionicity of the $A^{\delta+}[B-N]^{\delta-}$ framework, resulting in nitride ceramics with very negative electrostatic Madelung energies. Because alkali and alkaline metals are so electropositive, the inductive effect drives the strong exothermic formation energies of Alk-Me-N ternaries, rationalizing their predominance within the ternary nitrides map. The inductive effect can also be operative in nitrogen-rich Me-Me-N; most frequently with Zn, which is a relatively electropositive transition metal and can also serve as an electron donor. This fact was previously captured by the hierarchical agglomeration algorithm, which clustered Zn with the other alkali and alkaline earth metals.

For stable nitrogen-poor nitrides, we propose a novel *Reductive Effect*, where remarkably, nitrogen can serve as an electron donor for metal reduction. For some ternary Me-Me-N compositions, Me-Me bonds may be stronger than Me-N bonds. As shown in **Figure 7.7c**, the oxidation or release of electrophilic nitrogen atoms can redistribute electron density back to these strong Me-Me bonds, leading to the reduction of the corresponding metals. The reductive effect can stabilize unusual structures in the nitride chemistry;²⁵² for example, $\text{Co}_2\text{Mo}_3\text{N}$, which exhibits

infinite 1-D chains of covalently-bonded $[\text{Co} - \text{Co}]_{\infty}$ intertwined within an extended Mo-N covalent network. The reductive effect can also be operative in stable nitrogen-poor stoichiometries of Alk-Me-N compounds; for example, in $\text{Sr}_3\text{Ge}_2\text{N}_2$, which features infinite 1-D $[\text{Ge} - \text{Ge}]^{2-\infty}$ chains throughout the otherwise ionic $(\text{Sr}^{2+})_2[\text{GeN}_2]^{4-}$ lattice. The data-mining structure prediction algorithm used in this work operates on ionic substitution, which may not be ideally poised to predict novel nitrogen-poor nitrides due to their ambiguous valence states, suggesting there may still be many reductive effect-stabilized Me-Me-N ternary nitrides awaiting prediction.

Our analysis shows that the nitrogen anion can be fairly amphoteric in the solid-state—usually acting as an electron acceptor under the inductive effect to form ionic/covalent nitrogen-rich nitrides, but sometimes serving as an electron donor in the reductive effect, driving the formation of metallic nitrogen-poor sub-nitrides. The span of electronic structures available to the ternary nitrides offers a rich design space for materials functionality. Incorporating an alkali metal into an otherwise metallic binary nitride can increase charge localization driven by the inductive effect, opening a band gap and thus creating functional semiconducting nitrides suitable for solid-state lighting, photovoltaic energy conversion, piezoelectrics, and more. On the other hand, nitrogen-poor nitrides possess metallic bonding punctuated by charge-localization on nitrogen atoms, which can lead to complex electronic and magnetic structures²⁵³ and may serve as the basis for novel superconductors and topologically-protected quantum materials.^{254, 255} Modifying the nitrogen stoichiometry within a chemical space can be an effective strategy to compositionally tune the electronic structure between the reductive and inductive effect. For example, varying the Zn/Mo ratio in a wurtzite-based Zn-Mo-N compound can modulate the molybdenum oxidation

state from Mo^{4+} to Mo^{6+} , turning conductive ZnMoN_2 into insulating Zn_3MoN_4 , a wide-bandgap semiconductor.¹⁶²

7.4 Discussion

The library of inorganic solids has been dominated by oxides, whose structures and chemistries are often known from mineralogy. Compounds that do not form readily under ambient conditions, such as nitrides, offer a new frontier for materials discovery and design—so long as we have a rational understanding of the factors that drive stability in these relatively unexplored spaces. In this work, we used computational materials discovery and informatics tools to build a large stability map of the ternary metal nitride space. Our objective was not only to predict and synthesize new ternary metal nitrides, but further, to visualize large-scale relationships between nitride chemistry and thermodynamic stability, and to rationalize these trends from their deeper chemical origins. Our map as it stands is necessarily incomplete—it represents a current ‘upper-bound’ on the ternary nitride stability landscape. As new exotic structures and bonding motifs are discovered in the ternary metal nitrides, the procedures in this work can be iteratively re-applied to update and refine our understanding of this extended compositional space. From a broader perspective, our computational approach offers a systematic blueprint for mapping uncharted chemical spaces, providing synthetic chemists guidance in their quest to continuously extend the frontier of solid-state chemistry.

7.5 Methods

7.5.1 First-principles calculations of aluminum nitride hydrolysis kinetics

Plane wave periodic boundary condition DFT calculations were performed using the Vienna Ab-initio Simulation Package.^{12, 58} All calculations employed the Perdew-Burke-Ernzerhof

(PBE) generalized gradient approximation (GGA) exchange and correlation functional⁸ coupled with projector augmented wave (PAW) pseudopotentials.⁶⁰ For these calculations, PAWs were used to explicitly describe the aluminum 3s and 3p; nitrogen 2s and 2p; oxygen 2s and 2p; and hydrogen 1s electrons. A 96 atom, 2×2 supercell of the $(1\bar{1}00)$ surface of wurtzite AlN was used as the model for the reacting surface. This crystal structure was chosen because Schmerler *et al.* showed that the wurtzite phase is the stable structure at the conditions of interest for electrical and NH_3 generation applications.²⁵⁶ The lattice constants computed in this work match those obtained previously by both experiment and calculation to within 1%. Of the low-index facets of wurtzite AlN, Holec *et al.* showed that by far the lowest surface energy facet is the $(1\bar{1}00)$ surface.²⁵⁷ The 96 atom slab consists of six AlN layers and has a thickness of 9.4 Å in a 24.4 Å thick supercell, meaning 15 Å of vacuum space separates the reacting surface from the bottom layer of the adjacent slab. To obtain the model surface, the bulk AlN wurtzite structure was allowed to fully relax, vacuum space was added, and then the top four layers of the slab were allowed to relax while the bottom two layers were frozen to mimic the bulk subsurface. All calculations utilized a 450 eV cut-off energy based on a cut-off energy convergence study over the range of 250-500 eV where a 450 eV cut-off energy was found to produce a total energy within 0.01 eV/supercell of that produced with a 500 eV cut-off energy. The slab was relaxed using various Monkhorst-Pack k-point meshes between $1 \times 1 \times 1$ and $6 \times 6 \times 1$, where the $3 \times 3 \times 1$ k-point grid produced energies within 0.01 eV/supercell of those computed using the finer and more computationally demanding meshes. Bader charge analyses were conducted using software from the Henkelman group.²⁵⁸ All reaction energy barriers were obtained using the climbing-image nudged elastic band (cNEB) method²⁵⁹ or a combination of cNEB and the Dimer method.²⁶⁰ All geometries, including cNEB

images, were relaxed until the magnitude of the maximum force on the optimized ions was less than 0.03 eV/Å.

7.5.2 Data-mined structure prediction

Ternary structures were generated using the Data-Mined Structure-Prediction algorithm (DMSP),²³³ trained specifically for nitride discovery.⁴⁶ Briefly, the substitution training matrix was trained on the Inorganic Crystal Structure Database (ICSD), mapping isostructural compounds and identifying which cations are statistically probable to substitute for one another. The training set for nitride discovery was performed on the set of all pnictides (N + P + As + Sb), which has superior prediction capability to an oxide-containing set. This training set was found via 10-fold cross validation to have 80% probability of recovering known nitride structures. Training of the DMSP algorithm was performed on the ICSD as extracted in October 2015.

Total energies of known and DMSP-suggested nitrides were calculated with density functional theory using VASP, GGA, and PAWs. Plane-wave basis cut-off energies are set to 520 eV. The k-point densities were distributed within the Brillouin zone in a Monkhorst-Pack grid, or on a Gamma-centered grid for hexagonal cells, and used default *k*-point densities in compliance with Materials Project calculation standards,²⁶¹ which were calibrated to achieve total energy convergence of better than 0.5 meV/atom. Each structure is initiated in nonmagnetic, ferromagnetic, and anti-ferromagnetic spin configurations, and the lowest-energy configuration is used for phase stability calculations.

Phase stability calculations are computed from convex hulls, using the phase diagram analysis package in Pymatgen,¹¹⁸ calculated with respect to known nitride phases from the Materials Project obtained using the Materials Project REST API.²⁶² Azides (e.g. NaN₃, WN₁₈)

are removed from the phase diagram when computing phase stability, as they do not typically form during solid-state synthesis techniques. Materials Project data was retrieved in January, 2018.

For the Zn-Mo-N, Zn-W-N, Zn-Sb-N, Mg-Ti-N, Mg-Zr-N, Mg-Hf-N, Mg-Nb-N systems, an unconstrained ground-state search was performed using the “Kinetically Limited Minimization” approach,¹⁶² which does not require prototypical structures from databases. Seed structures are generated from random lattice vectors and atomic positions, subject to geometric constraints to avoid extreme cell shapes, and to observe minimal interatomic distances (2.8 Å for cation-cation and anion-anion pairs, 1.9 Å for cation-anion pairs). For each material, we sampled at least 100 seeds, over the ternary compositions $A_iB_jN_k$ for $ijk = 112, 146, 414, 213, 124, 326, 338, 313$, chosen to accommodate the $(\text{Mg/Zn})^{2+}$, $M^{4+/5+/6+}$, and N^{3-} oxidation states. New trial structures are generated by the random displacement of one atom between 1.0 and 5.0 Å while maintaining the minimal distances. Trial structures are accepted if the total energy is lowered, and the number of trials equals the number of atoms in the unit cell.

7.5.3 Extracting chemical insights from first-principles calculations

The Density Derived Electrostatic and Chemical (DDEC) approach was used to obtain net atomic charges and natural bond orders assigned to each ion in each calculated structure.²²¹ From the DDEC analysis we define the average charge for ion i , δ_i , as the net atomic charge assigned to ion, i (number of electrons) averaged over all ions, i , in the structure, $A_\alpha B_\beta N_\gamma$. The summed bond order for ion i , s_i , was obtained similarly by summing the natural bond orders for all interactions containing i , averaged over all ions, i , in the structure, $A_\alpha B_\beta N_\gamma$. The Crystal Orbital Hamilton Population (COHP) analysis was used to quantify the bonding interactions within each structure and partition these interactions by specific ion-ion pairs using the LOBSTER code.²²⁰ To normalize the comparison of COHPs across a range of structures and compositions, the energy

levels from each PBE calculation were aligned to core levels. Doing so allows for a reasonable comparison of Fermi energies and thus COHP energy depths across the various systems analyzed. To alleviate the effects of varied pseudopotentials across systems, the number of free atom valence electrons was determined for each system using the following equation:

$$N_v(A_\alpha B_\beta N_\gamma) = \alpha N_v(A) + \beta N_v(B) + \gamma N_v(N) \quad (7.1)$$

where N_v is the number of valence (outermost shell) electrons. The minimum energy which contains valence electrons, ϵ_v was then determined for each structure by incrementally decreasing the energy, ϵ , and integrating the density of states (DOS) from ϵ to the Fermi energy, ϵ_F , such that

$$\epsilon_v = \epsilon_v: \int_{\epsilon_v}^{\epsilon_F} DOS(E') dE' = 1 \quad (7.2)$$

where DOS is normalized by N_v and E' is the core-level aligned energy. The magnitude of bonding interactions, Σ , in each structure is then defined as

$$\Sigma = \int_{\epsilon_v}^{\epsilon_F} -COHP(E') E' dE' \quad (7.3)$$

where the COHP is also normalized by N_v .

Using these quantities, we produced the triangle plots shown in **Figure 7.7a**. The ionicity, I , was defined as:

$$I_{A_\alpha B_\beta N_\gamma} = \frac{1}{\alpha + \beta + \gamma} \left(\alpha \frac{\delta_A}{s_A} + \beta \frac{\delta_B}{s_B} + \gamma \frac{\delta_N}{s_N} \right) \quad (7.4)$$

and quantifies the extent of electron transfer in the structure. The metallicity was defined as:

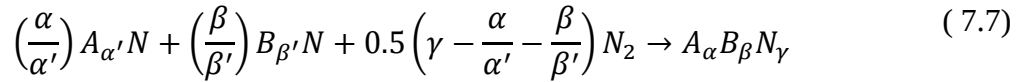
$$M_{A_{\alpha}B_{\beta}N_{\gamma}} = |\Sigma_{A-B} + \Sigma_{A-A} + \Sigma_{B-B}| \quad (7.5)$$

quantifying the net bonding energy of cation-cation interactions. The covalency, C , was defined as:

$$C_{A_{\alpha}B_{\beta}N_{\gamma}} = |\Sigma_{A-N} + \Sigma_{B-N} + \Sigma_{N-N}| \quad (7.6)$$

quantifying the net bonding energy for interactions containing nitrogen. To ensure each quantity (C, I, M) was of the same magnitude, each quantity was divided by the maximum of that quantity within the dataset. In order to plot points on a triangle, the sum of each point, (C, I, M), must equal 1. Therefore, each quantity within each point was normalized by $C+I+M$.

To quantify the extent to which a given ternary was “N-rich” or “N-poor”, we compare the cation/anion ratios in the ternary to the ratios in the deep-hull binaries (A_{α} -N, B_{β} -N) using the assumed formation reaction:



and subsequent condition for being rich or poor in nitrogen:

$$\gamma - \frac{\alpha}{\alpha'} - \frac{\beta}{\beta'} \geq 0 \rightarrow N_{rich}; \gamma - \frac{\alpha}{\alpha'} - \frac{\beta}{\beta'} < 0 \rightarrow N_{poor} \quad (7.8)$$

This reaction was also used to compute the change in charge, $\Delta\delta$, across this formation reaction, as shown in **Figure 7.7b**:

$$\Delta\delta_i = \delta_{i,A\alpha B\beta N\gamma} \quad (7.9)$$

$$- \frac{2}{\frac{\alpha}{\alpha'} + \frac{\beta}{\beta'} + \gamma} \left(\frac{\alpha}{\alpha'} \delta_{i,A\alpha'N} + \frac{\beta}{\beta'} \delta_{i,B\beta'N} + 0.5 \left(\gamma - \frac{\alpha}{\alpha'} - \frac{\beta}{\beta'} \right) \delta_{i,N_2} \right)$$

where δ_{i,N_2} was taken to be 0. A and B are defined as the least and most electronegative cations in the ternary.

8 Outlook

This thesis emphasizes the usefulness of both density functional theory (DFT) and data analytics to rapidly probe thermodynamic stability of inorganic crystalline solids. Our key findings are:

1. High-throughput DFT approaches are making suitable predictions of compound stability, in general, when compared with experiment and the oft-employed correction of fitting elemental reference energies is mostly unnecessary for stability predictions.
2. Statistically motivated algorithms can improve significantly upon empirically derived descriptors. In particular, we show that given the same inputs, the learned descriptor predicts the stability of perovskites with an accuracy of 92% compared with 74% for the empirical standard.
3. The necessity of computing phase diagrams and perturbing structure for high-throughput screening applications is emphasized for the identification of double perovskite solar absorbers. We also present 22 new cesium chloride double perovskites that are stable (or within the error of the methods) with respect to decomposition and have suitable band gaps for solar absorption when calculated using a hybrid density functional.
4. For high-temperature processes, solid-state entropy must be considered to accurately predict the free energy and we present a statistically learned descriptor for this property that compares well with experiment (errors ~ 1 kcal/mol).
5. Solar thermochemical ammonia synthesis as studied in this work is constrained significantly from a thermodynamic perspective. There are promising binary

(monometallic) active materials, but these will likely have to be modified by alloying or the process must be creatively reconceived to enable realization of an efficient process.

6. Metal nitrides are extremely unexplored relative to their potential for identifying interesting materials. We present > 200 ternary metal nitrides that are predicted to be stable with respect to decomposition using DFT. The stability of these materials can be explained by two principal effects – the inductive effect where an electropositive cation enables rich covalent or ionic bonding between a second cation and nitrogen and the reductive effect where nitrogen enables rich metallic bonding between two cations.

While thermodynamic stability is a principal aspect of computationally prototyping a target material, it is not the end of the road. Beyond the enthalpy or free energy of formation (decomposition), which are a primary focus of this work, a number of other factors dictate the successful synthesis of a new material. These other factors include, but are not limited to, kinetics of ion transport, interfacial energy, nucleation, and crystallization. Each of these challenges present opportunities for computational and statistical approaches to bridge the gap between what is predicted on a computer and what is made in the lab, and this should be a chief focus of materials scientists going forward.

References

1. Noyce, R. N. Semiconductor circuit complex having isolation means. 1959.
2. top500.org/statistics/overtime.
3. Kohn, W.; Sham, L. J., Self-Consistent Equations Including Exchange and Correlation Effects. *Physical Review* **1965**, 140, (4A), A1133-A1138.
4. Becke, A. D., Perspective: Fifty years of density-functional theory in chemical physics. *The Journal of Chemical Physics* **2014**, 140, (18), 18A301.
5. Schrödinger, E., An Undulatory Theory of the Mechanics of Atoms and Molecules. *Physical Review* **1926**, 28, (6), 1049-1070.
6. Born, M.; Oppenheimer, R., Zur Quantentheorie der Molekeln. *Annalen der Physik* **1927**, 389, (20), 457-484.
7. Burke, K. The ABC of DFT. (10/9/2018),
8. Perdew, J. P.; Burke, K.; Ernzerhof, M., Generalized Gradient Approximation Made Simple. *Physical Review Letters* **1996**, 77, (18), 3865-3868.
9. Sun, J.; Ruzsinszky, A.; Perdew, J. P., Strongly Constrained and Appropriately Normed Semilocal Density Functional. *Physical Review Letters* **2015**, 115, (3), 036402.
10. Perdew, J. P.; Wang, Y., Accurate and simple analytic representation of the electron-gas correlation energy. *Physical Review B* **1992**, 45, (23), 13244-13249.
11. Becke, A. D., Density-functional exchange-energy approximation with correct asymptotic behavior. *Physical Review A* **1988**, 38, (6), 3098-3100.
12. Kresse, G.; Furthmüller, J., Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Physical Review B* **1996**, 54, (16), 11169-11186.
13. Blum, V.; Gehrke, R.; Hanke, F.; Havu, P.; Havu, V.; Ren, X.; Reuter, K.; Scheffler, M., Ab initio molecular simulations with numeric atom-centered orbitals. *Computer Physics Communications* **2009**, 180, (11), 2175-2196.
14. Torrent, M.; Jollet, F.; Bottin, F.; Zérah, G.; Gonze, X., Implementation of the projector augmented-wave method in the ABINIT code: Application to the study of iron under pressure. *Computational Materials Science* **2008**, 42, (2), 337-351.
15. Ekeland, I., On the variational principle. *Journal of Mathematical Analysis and Applications* **1974**, 47, (2), 324-353.

16. Jain, A.; Ong, S. P.; Hautier, G.; Chen, W.; Richards, W. D.; Dacek, S.; Cholia, S.; Gunter, D.; Skinner, D.; Ceder, G.; Persson, K. A., Commentary: The Materials Project: A materials genome approach to accelerating materials innovation. *APL Materials* **2013**, 1, (1), 011002.
17. Curtarolo, S.; Setyawan, W.; Wang, S.; Xue, J.; Yang, K.; Taylor, R. H.; Nelson, L. J.; Hart, G. L. W.; Sanvito, S.; Buongiorno-Nardelli, M.; Mingo, N.; Levy, O., AFLOWLIB.ORG: A distributed materials properties repository from high-throughput ab initio calculations. *Computational Materials Science* **2012**, 58, 227-235.
18. Saal, J. E.; Kirklin, S.; Aykol, M.; Meredig, B.; Wolverton, C., Materials Design and Discovery with High-Throughput Density Functional Theory: The Open Quantum Materials Database (OQMD). *JOM* **2013**, 65, (11), 1501-1509.
19. Draxl, C.; Scheffler, M., NOMAD: The FAIR Concept for Big-Data-Driven Materials Science. Preprint at <https://arxiv.org/abs/1805.05039> **2018**.
20. Anelli, A.; Engel, E. A.; Pickard, C. J.; Ceriotti, M., A generalized convex hull construction for materials discovery. *arXiv:1803.0932* **2018**.
21. Ong, S. P.; Wang, L.; Kang, B.; Ceder, G., Li-Fe-P-O₂ Phase Diagram from First Principles Calculations. *Chemistry of Materials* **2008**, 20, (5), 1798-1807.
22. Togo, A.; Tanaka, I., First principles phonon calculations in materials science. *Scripta Materialia* **2015**, 108, 1-5.
23. Butler, K. T.; Davies, D. W.; Cartwright, H.; Isayev, O.; Walsh, A., Machine learning for molecular and materials science. *Nature* **2018**, 559, (7715), 547-555.
24. Pauling, L., The principles determining the structure of complex ionic crystals. *J. Am. Chem. Soc.* **1929**, 51, (4), 1010-1026.
25. Bartók, A. P.; Kondor, R.; Csányi, G., On representing chemical environments. *Physical Review B* **2013**, 87, (18), 184115.
26. Isayev, O.; Fourches, D.; Muratov, E. N.; Oses, C.; Rasch, K.; Tropsha, A.; Curtarolo, S., Materials Cartography: Representing and Mining Materials Space Using Structural and Electronic Fingerprints. *Chemistry of Materials* **2015**, 27, (3), 735-743.
27. Isayev, O.; Oses, C.; Toher, C.; Gossett, E.; Curtarolo, S.; Tropsha, A., Universal fragment descriptors for predicting properties of inorganic crystals. *Nature Communications* **2017**, 8.
28. Xie, T.; Grossman, J. C., Crystal Graph Convolutional Neural Networks for an Accurate and Interpretable Prediction of Material Properties. *Physical Review Letters* **2018**, 120, (14), 145301.

29. Kirklin, S.; Saal, J. E.; Meredig, B.; Thompson, A.; Doak, J. W.; Aykol, M.; Rühl, S.; Wolverton, C., The Open Quantum Materials Database (OQMD): assessing the accuracy of DFT formation energies. *npj Computational Materials* **2015**, 1, 15010.
30. Stevanović, V.; Lany, S.; Zhang, X.; Zunger, A., Correcting density functional theory for accurate predictions of compound enthalpies of formation: Fitted elemental-phase reference energies. *Physical Review B* **2012**, 85, (11), 115104.
31. Pizzi, G.; Cepellotti, A.; Sabatini, R.; Marzari, N.; Kozinsky, B., AiiDA: automated interactive infrastructure and database for computational science. *Computational Materials Science* **2016**, 111, 218-230.
32. Sun, W.; Bartel, C. J.; Arca, E.; Bauers, S.; Matthews, B.; Orvananos, B.; Chen, B.-R.; Toney, M. F.; Schelhas, L. T.; Tumas, W.; Tate, J.; Zakutayev, A.; Lany, S.; Holder, A.; Ceder, G., A Map of the Inorganic Ternary Metal Nitrides. *Preprint at <https://arxiv.org/abs/1809.09202>* **2018**.
33. Lany, S., Semiconductor thermochemistry in density functional calculations. *Physical Review B* **2008**, 78, (24), 245207.
34. Wang, L.; Maxisch, T.; Ceder, G., Oxidation energies of transition metal oxides within the $\mathrm{GGA}+\mathrm{U}$ framework. *Physical Review B* **2006**, 73, (19), 195107.
35. Jain, A.; Hautier, G.; Ong, S. P.; Moore, C. J.; Fischer, C. C.; Persson, K. A.; Ceder, G., Formation enthalpies by mixing GGA and GGA+ U calculations. *Physical Review B* **2011**, 84, (4), 045115.
36. Hautier, G.; Ong, S. P.; Jain, A.; Moore, C. J.; Ceder, G., Accuracy of density functional theory in predicting formation energies of ternary oxides from binary oxides and its implication on phase stability. *Physical Review B* **2012**, 85, (15), 155208.
37. Sun, W.; Dacek, S. T.; Ong, S. P.; Hautier, G.; Jain, A.; Richards, W. D.; Gamst, A. C.; Persson, K. A.; Ceder, G., The thermodynamic scale of inorganic crystalline metastability. *Science Advances* **2016**, 2, (11).
38. Pandey, M.; Jacobsen, K. W., Heats of formation of solids with error estimation: The mBEEF functional with and without fitted reference energies. *Physical Review B* **2015**, 91, (23), 235201.
39. Zunger, A., Inverse design in search of materials with target functionalities. *Nature Reviews Chemistry* **2018**, 2, 0121.
40. Ryan, J.; Tam, M.; John, B.; Dane, M., Material Discovery and Design Principles for Stable, High Activity Perovskite Cathodes for Solid Oxide Fuel Cells. *Advanced Energy Materials* **2018**, 8, (11), 1702708.

41. Korbel, S.; Marques, M. A. L.; Botti, S., Stable hybrid organic-inorganic halide perovskites for photovoltaics from ab initio high-throughput calculations. *Journal of Materials Chemistry A* **2018**, 6, (15), 6463-6475.
42. Isaacs, E. B.; Wolverton, C., Inverse Band Structure Design via Materials Database Screening: Application to Square Planar Thermoelectrics. *Chemistry of Materials* **2018**, 30, (5), 1540-1546.
43. Lu, Z.; Ciucci, F., Anti-perovskite cathodes for lithium batteries. *Journal of Materials Chemistry A* **2018**, 6, (12), 5185-5192.
44. Dagdelen, J.; Montoya, J.; de Jong, M.; Persson, K., Computational prediction of new auxetic materials. *Nature Communications* **2017**, 8, (1), 323.
45. Singh, A. K.; Mathew, K.; Zhuang, H. L.; Hennig, R. G., Computational Screening of 2D Materials for Photocatalysis. *The Journal of Physical Chemistry Letters* **2015**, 6, (6), 1087-1098.
46. Sun, W.; Holder, A.; Orvañanos, B.; Arca, E.; Zakutayev, A.; Lany, S.; Ceder, G., Thermodynamic Routes to Novel Metastable Nitrogen-Rich Nitrides. *Chemistry of Materials* **2017**, 29, (16), 6936-6946.
47. Bale, C. W.; Bélisle, E.; Chartrand, P.; Decterov, S. A.; Eriksson, G.; Gheribi, A. E.; Hack, K.; Jung, I. H.; Kang, Y. B.; Melançon, J.; Pelton, A. D.; Petersen, S.; Robelin, C.; Sangster, J.; Spencer, P.; Van Ende, M. A., FactSage thermochemical software and databases, 2010–2016. *Calphad* **2016**, 54, 35-53.
48. Sun, J.; Remsing, R. C.; Zhang, Y.; Sun, Z.; Ruzsinszky, A.; Peng, H.; Yang, Z.; Paul, A.; Waghmare, U.; Wu, X.; Klein, M. L.; Perdew, J. P., Accurate first-principles structures and energies of diversely bonded systems from an efficient density functional. *Nature Chemistry* **2016**, 8, 831.
49. Zhang, Y.; Kitchaev, D. A.; Yang, J.; Chen, T.; Dacek, S. T.; Sarmiento-Pérez, R. A.; Marques, M. A. L.; Peng, H.; Ceder, G.; Perdew, J. P.; Sun, J., Efficient first-principles prediction of solid stability: Towards chemical accuracy. *npj Computational Materials* **2018**, 4, (1), 9.
50. Sai Gautam, G.; Carter, E. A., Evaluating transition metal oxides within DFT-SCAN and SCAN+U frameworks for solar thermochemical applications. *Physical Review Materials* **2018**, 2, (9), 095401.
51. Ekholm, M.; Gambino, D.; Jönsson, H. J. M.; Tasnádi, F.; Alling, B.; Abrikosov, I. A., Assessing the SCAN functional for itinerant electron ferromagnets. *Physical Review B* **2018**, 98, (9), 094413.
52. WebElements.com.

53. Kubaschewski, O.; Alcock, C. B.; Spencer, P. J., *Materials thermochemistry*. 6th ed.; Pergamon Press: Oxford; New York, 1993.
54. Bachmann, K. J.; Hsu, F. S. L.; Thiel, F. A.; Kasper, H. M., Debye temperature and standard entropies and enthalpies of compound semiconductors of the type I-III-VI₂. *Journal of Electronic Materials* **1977**, 6, (4), 431-448.
55. Kitchaev, D. A.; Peng, H.; Liu, Y.; Sun, J.; Perdew, J. P.; Ceder, G., Energetics of MnO_2 polymorphs in density functional theory. *Physical Review B* **2016**, 93, (4), 045132.
56. Kitchaev, D. A.; Peng, H.; Liu, Y.; Sun, J.; Perdew, J. P.; Ceder, G., Energetics of MnO_2 polymorphs in density functional theory. *Physical Review B* **2016**, 93, (4), 045132.
57. Filip, M. R.; Giustino, F., The geometric blueprint of perovskites. *Proceedings of the National Academy of Sciences* **2018**, 115, (21), 5397-5402.
58. Kresse, G.; Furthmüller, J., Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Computational Materials Science* **1996**, 6, (1), 15-50.
59. Kresse, G.; Joubert, D., From ultrasoft pseudopotentials to the projector augmented-wave method. *Physical Review B* **1999**, 59, (3), 1758-1775.
60. Blöchl, P. E., Projector augmented-wave method. *Physical Review B* **1994**, 50, (24), 17953-17979.
61. Woodley, S. M.; Catlow, R., Crystal structure prediction from first principles. *Nat. Mater.* **2008**, 7, 937.
62. Kirkpatrick, S.; Gelatt, C. D.; Vecchi, M. P., Optimization by simulated annealing. *Science* **1983**, 220, (4598), 671-680.
63. Doye, J. P. K.; Wales, D. J., Thermodynamics of global optimization. *Phys. Rev. Lett.* **1998**, 80, (7), 1357-1360.
64. Goedecker, S., Minima hopping: An efficient search method for the global minimum of the potential energy surface of complex molecular systems. *J. Chem. Phys.* **2004**, 120, (21), 9911-9917.
65. Oganov, A. R.; Lyakhov, A. O.; Valle, M., How evolutionary crystal structure prediction works—and why. *Acc. Chem. Res.* **2011**, 44, (3), 227-237.
66. Curtarolo, S.; Hart, G. L. W.; Nardelli, M. B.; Mingo, N.; Sanvito, S.; Levy, O., The high-throughput highway to computational materials design. *Nat. Mater.* **2013**, 12, 191.

67. Ghiringhelli, L. M.; Vybiral, J.; Levchenko, S. V.; Draxl, C.; Scheffler, M., Big data of materials science: Critical role of the descriptor. *Phys. Rev. Lett.* **2015**, 114, (10), 105503.
68. Goldschmidt, V. M., Die Gesetze der Krystallochemie. *Naturwissenschaften* **1926**, 14, (21), 477-485.
69. Hwang, J.; Rao, R. R.; Giordano, L.; Katayama, Y.; Yu, Y.; Shao-Horn, Y., Perovskites in catalysis and electrocatalysis. *Science* **2017**, 358, (6364), 751-756.
70. Duan, C.; Tong, J.; Shang, M.; Nikodemski, S.; Sanders, M.; Ricote, S.; Almonsoori, A.; O'Hayre, R., Readily processed protonic ceramic fuel cells with high performance at low temperatures. *Science* **2015**.
71. Cohen, R. E., Origin of ferroelectricity in perovskite oxides. *Nature* **1992**, 358, 136.
72. Yi, T.; Chen, W.; Cheng, L.; Bayliss, R. D.; Lin, F.; Plews, M. R.; Nordlund, D.; Doeffer, M. M.; Persson, K. A.; Cabana, J., Investigating the intercalation chemistry of alkali ions in fluoride perovskites. *Chem. Mater.* **2017**, 29, (4), 1561-1568.
73. Correa-Baena, J.-P.; Saliba, M.; Buonassisi, T.; Grätzel, M.; Abate, A.; Tress, W.; Hagfeldt, A., Promises and challenges of perovskite solar cells. *Science* **2017**, 358, (6364), 739-744.
74. Kovalenko, M. V.; Protesescu, L.; Bodnarchuk, M. I., Properties and potential optoelectronic applications of lead halide perovskite nanocrystals. *Science* **2017**, 358, (6364), 745-750.
75. Zhang, H.; Li, N.; Li, K.; Xue, D., Structural stability and formability of ABO_3 -type perovskite compounds. *Acta Crystallogr. Sect. B* **2007**, 63, (6), 812-818.
76. Li, C.; Lu, X.; Ding, W.; Feng, L.; Gao, Y.; Guo, Z., Formability of ABX_3 ($X = F, Cl, Br, I$) halide perovskites. *Acta Crystallogr. Sect. B* **2008**, 64, (6), 702-707.
77. Travis, W.; Glover, E. N. K.; Bronstein, H.; Scanlon, D. O.; Palgrave, R. G., On the application of the tolerance factor to inorganic and hybrid halide perovskites: A revised system. *Chem. Sci.* **2016**, 7, (7), 4548-4556.
78. Li, W.; Wang, Z.; Deschler, F.; Gao, S.; Friend, R. H.; Cheetham, A. K., Chemically diverse and multifunctional hybrid organic-inorganic perovskites. *Nat. Rev. Mater.* **2017**, 2, 16099.
79. Shannon, R., Revised effective ionic radii and systematic studies of interatomic distances in halides and chalcogenides. *Acta Crystallographica Section A* **1976**, 32, (5), 751-767.
80. Lufaso, M. W.; Woodward, P. M., Prediction of the crystal structures of perovskites using the software program SPuDS. *Acta Crystallogr. Sect. B* **2001**, 57, (6), 725-738.

81. Kieslich, G.; Sun, S.; Cheetham, A. K., Solid-state principles applied to organic-inorganic perovskites: new tricks for an old dog. *Chemical Science* **2014**, 5, (12), 4712-4715.
82. Li, C.; Soh, K. C. K.; Wu, P., Formability of ABO_3 perovskites. *J. Alloys Compd.* **2004**, 372, (1), 40-48.
83. Becker, M.; Kluner, T.; Wark, M., Formation of hybrid ABX_3 perovskite compounds for solar cell application: first-principles calculations of effective ionic radii and determination of tolerance factors. *Dalton Transactions* **2017**, 46, (11), 3500-3509.
84. Pilania, G.; Balachandran, P. V.; Gubernatis, J. E.; Lookman, T., Classification of ABO_3 perovskite solids: a machine learning study. *Acta Crystallogr. Sect. B* **2015**, 71, (5), 507-513.
85. Pilania, G.; Balachandran, P. V.; Kim, C.; Lookman, T., Finding new perovskite halides via machine learning. *Front. Mater.* **2016**, 3, 19.
86. Balachandran, P. V.; Emery, A. A.; Gubernatis, J. E.; Lookman, T.; Wolverton, C.; Zunger, A., Predictions of new ABO_3 perovskite compounds by combining machine learning and density functional theory. *Physical Review Materials* **2018**, 2, (4), 043802.
87. Zhao, X.-G.; Yang, D.; Sun, Y.; Li, T.; Zhang, L.; Yu, L.; Zunger, A., Cu-In Halide Perovskite Solar Absorbers. *Journal of the American Chemical Society* **2017**, 139, (19), 6718-6725.
88. Helen, D. M., Crystal structure of double oxides of the perovskite type. *Proc. Phys. Soc.* **1946**, 58, (2), 133.
89. Feteira, A.; Sinclair, D. C.; Rajab, K. Z.; Lanagan, M. T., Crystal structure and microwave dielectric properties of alkaline-earth hafnates, AHfO_3 (A=Ba, Sr, Ca). *J. Am. Ceram. Soc.* **2008**, 91, (3), 893-901.
90. Ouyang, R.; Curtarolo, S.; Ahmetcik, E.; Scheffler, M.; Ghiringhelli, L. M., SISSO: A compressed-sensing method for identifying the best low-dimensional descriptor in an immensity of offered candidates. *Physical Review Materials* **2018**, 2, (8), 083802.
91. Platt, J., Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Advances in large margin classifiers* **1999**, 10, (3), 61-74.
92. Sun, Q.; Yin, W.-J., Thermodynamic Stability Trend of Cubic Perovskites. *Journal of the American Chemical Society* **2017**, 139, (42), 14905-14908.
93. Schmidt, J.; Shi, J.; Borlido, P.; Chen, L.; Botti, S.; Marques, M. A. L., Predicting the Thermodynamic Stability of Solids Combining Density Functional Theory and Machine Learning. *Chemistry of Materials* **2017**, 29, (12), 5090-5103.

94. Faber, F. A.; Lindmaa, A.; von Lilienfeld, O. A.; Armiento, R., Machine Learning Energies of 2 Million Elpasolite ABC_2D_6 Crystals. *Physical Review Letters* **2016**, 117, (13), 135502.
95. Kamat, P. V.; Bisquert, J.; Buriak, J., Lead-Free Perovskite Solar Cells. *ACS Energy Letters* **2017**, 2, (4), 904-905.
96. Hellenbrandt, M., The Inorganic Crystal Structure Database (ICSD)—Present and Future. *Crystallography Reviews* **2004**, 10, (1), 17-22.
97. Li, W.; Ionescu, E.; Riedel, R.; Gurlo, A., Can we predict the formability of perovskite oxynitrides from tolerance and octahedral factors? *Journal of Materials Chemistry A* **2013**, 1, (39), 12239-12245.
98. McClure, E. T.; Ball, M. R.; Windl, W.; Woodward, P. M., $\text{Cs}_2\text{AgBiX}_6$ (X = Br, Cl): New Visible Light Absorbing, Lead-Free Halide Perovskite Semiconductors. *Chemistry of Materials* **2016**, 28, (5), 1348-1354.
99. Choi, S. O.; Penninger, M.; Kim, C. H.; Schneider, W. F.; Thompson, L. T., Experimental and Computational Investigation of Effect of Sr on NO Oxidation and Oxygen Exchange for $\text{La}_{1-x}\text{Sr}_x\text{CoO}_3$ Perovskite Catalysts. *ACS Catalysis* **2013**, 3, (12), 2719-2728.
100. Li, Z.; Cho, Y.; Li, X.; Li, X.; Aimi, A.; Inaguma, Y.; Alonso, J. A.; Fernandez-Diaz, M. T.; Yan, J.; Downer, M. C.; Henkelman, G.; Goodenough, J. B.; Zhou, J., New Mechanism for Ferroelectricity in the Perovskite $\text{Ca}_{2-x}\text{Mn}_x\text{Ti}_2\text{O}_6$ Synthesized by Spark Plasma Sintering. *Journal of the American Chemical Society* **2018**, 140, (6), 2214-2220.
101. Volonakis, G.; Haghighirad, A. A.; Milot, R. L.; Sio, W. H.; Filip, M. R.; Wenger, B.; Johnston, M. B.; Herz, L. M.; Snaith, H. J.; Giustino, F., $\text{Cs}_2\text{InAgCl}_6$: A New Lead-Free Halide Double Perovskite with Direct Band Gap. *J Phys Chem Lett* **2017**, 8, (4), 772-778.
102. Greul, E.; Petrus, Michiel L.; Binek, A.; Docampo, P.; Bein, T., Highly stable, phase pure $\text{Cs}_2\text{AgBiBr}_6$ double perovskite thin films for optoelectronic applications. *Journal of Materials Chemistry A* **2017**, 5, (37), 19972-19981.
103. Slavney, A. H.; Leppert, L.; Saldivar Valdes, A.; Bartesaghi, D.; Savenije, T. J.; Neaton, J. B.; Karunadasa, H. I., Small-Band-Gap Halide Double Perovskites. *Angewandte Chemie International Edition* **2018**, 57, (39), 12765-12770.
104. Curtarolo, S.; Hart, G. L.; Nardelli, M. B.; Mingo, N.; Sanvito, S.; Levy, O., The high-throughput highway to computational materials design. *Nature materials* **2013**, 12, (3), 191.
105. Jain, A.; Shin, Y.; Persson, K. A., Computational predictions of energy materials using density functional theory. *Nature Reviews Materials* **2016**, 1, 15004.

106. Bartel, C. J.; Sutton, C.; Goldsmith, B. R.; Ouyang, R.; Musgrave, C. B.; Ghiringhelli, L. M.; Scheffler, M., New Tolerance Factor to Predict the Stability of Perovskite Oxides and Halides. *Preprint at <https://arxiv.org/abs/1801.0770>* **2018**.
107. Deng, W.; Deng, Z.-Y.; He, J.; Wang, M.; Chen, Z.-X.; Wei, S.-H.; Feng, H.-J., Synthesis of Cs₂AgSbCl₆ and improved optoelectronic properties of Cs₂AgSbCl₆/TiO₂ heterostructure driven by the interface effect for lead-free double perovskites solar cells. *Applied Physics Letters* **2017**, 111, (15), 151602.
108. Heyd, J.; Scuseria, G. E.; Ernzerhof, M., Hybrid functionals based on a screened Coulomb potential. *The Journal of Chemical Physics* **2003**, 118, (18), 8207-8215.
109. Takahashi, K.; Takahashi, L.; Miyazato, I.; Tanaka, Y., Searching for Hidden Perovskite Materials for Photovoltaic Systems by Combining Data Science and First Principle Calculations. *ACS Photonics* **2018**, 5, (3), 771-775.
110. Lu, S.; Zhou, Q.; Ouyang, Y.; Guo, Y.; Li, Q.; Wang, J., Accelerated discovery of stable lead-free hybrid organic-inorganic perovskites via machine learning. *Nature Communications* **2018**, 9, (1), 3405.
111. Nakajima, T.; Sawada, K., Discovery of Pb-Free Perovskite Solar Cells via High-Throughput Simulation on the K Computer. *The Journal of Physical Chemistry Letters* **2017**, 8, (19), 4826-4831.
112. Zhao, X.-G.; Yang, J.-H.; Fu, Y.; Yang, D.; Xu, Q.; Yu, L.; Wei, S.-H.; Zhang, L., Design of Lead-Free Inorganic Halide Perovskites for Solar Cells via Cation-Transmutation. *Journal of the American Chemical Society* **2017**, 139, (7), 2630-2638.
113. Filip, M. R.; Liu, X.; Miglio, A.; Hautier, G.; Giustino, F., Phase Diagrams and Stability of Lead-Free Halide Double Perovskites Cs₂BB'X₆: B = Sb and Bi, B' = Cu, Ag, and Au, and X = Cl, Br, and I. *The Journal of Physical Chemistry C* **2018**, 122, (1), 158-170.
114. Jain, A.; Voznyy, O.; Sargent, E. H., High-Throughput Screening of Lead-Free Perovskite-like Materials for Optoelectronic Applications. *The Journal of Physical Chemistry C* **2017**, 121, (13), 7183-7187.
115. Berger, R. F.; Neaton, J. B., Computational design of low-band-gap double perovskites. *Physical Review B* **2012**, 86, (16), 165211.
116. Leijtens, T.; Bush, K. A.; Prasanna, R.; McGehee, M. D., Opportunities and challenges for tandem solar cells using metal halide perovskite semiconductors. *Nature Energy* **2018**, 3, (10), 828-838.
117. Shockley, W.; Queisser, H. J., Detailed Balance Limit of Efficiency of p-n Junction Solar Cells. *Journal of Applied Physics* **1961**, 32, (3), 510-519.

118. Ong, S. P.; Richards, W. D.; Jain, A.; Hautier, G.; Kocher, M.; Cholia, S.; Gunter, D.; Chevrier, V. L.; Persson, K. A.; Ceder, G., Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis. *Computational Materials Science* **2013**, 68, 314-319.
119. Stanley Smith, C., Materials and the Development of Civilization and Science: Empiricism and esthetic selection led to discovery of many properties on which material science is based. *Science* **1965**, 148, 908-17.
120. Hemminger, J. C.; Sarrao, J.; Crabtree, G.; Flemming, G.; Ratner, M., Challenges at the Frontiers of Matter and Energy: Transformative Opportunities for Discovery Science. *USDOE Office of Science* **2015**, 1-78.
121. Holder, A. M., S. Siol, P.F. Ndione, H. Peng, A.M. Deml, B.E. Mathews, L.T. Schelhas, M.F. Toney, R.G. Gordon, W. Tumas, J.D. Perkins, D.S. Ginley, B.P. Gorman, J. Tate, A. Zakutayev, and S. Lany, Novel phase diagram behavior and materials design in heterostructural semiconductor alloys. *Science Advances* **2017**, 3, (6), e1700270.
122. De Yoreo, J., Synthesis Science for Energy Relevant Technology *U.S. Department of Energy Office of Science* **2016**.
123. Siol, S.; Holder, A.; Steffes, J.; Schelhas, L. T.; Stone, K. H.; Garten, L.; Perkins, J. D.; Parilla, P. A.; Toney, M. F.; Huey, B. D.; Tumas, W.; Lany, S.; Zakutayev, A., Negative-pressure polymorphs made by heterostructural alloying. *Science Advances* **2018**, 4, (eaaq1442), 1-7.
124. Zebarjadi, M.; Esfarjani, K.; Dresselhaus, M.; Ren, Z.; Chen, G., Perspectives on thermoelectrics: from fundamentals to device applications. *Energy & Environmental Science* **2012**, 5, (1), 5147-5162.
125. Duan, C.; Tong, J.; Shang, M.; Nikodemski, S.; Sanders, M.; Ricote, S.; Almansoori, A.; O'Hayre, R., Readily processed protonic ceramic fuel cells with high performance at low temperatures. *Science* **2015**, 349, (6254), 1321-1326.
126. Muhich, C. L.; Evanko, B. W.; Weston, K. C.; Lichty, P.; Liang, X.; Martinek, J.; Musgrave, C. B.; Weimer, A. W., Efficient Generation of H₂ by Splitting Water with an Isothermal Redox Cycle. *Science* **2013**, 341, (6145), 540.
127. Dunstan, M. T.; Jain, A.; Liu, W.; Ong, S. P.; Liu, T.; Lee, J.; Persson, K. A.; Scott, S. A.; Dennis, J. S.; Grey, C. P., Large scale computational screening and experimental discovery of novel materials for high temperature CO₂ capture. *Energy & Environmental Science* **2016**, 9, (4), 1346-1360.
128. Stoffel, R. P.; Wessel, C.; Lumey, M.-W.; Dronskowski, R., Ab Initio Thermochemistry of Solid-State Materials. *Angewandte Chemie International Edition* **2010**, 49, (31), 5242-5266.

129. Huang, L.-F.; Lu, X.-Z.; Tennesen, E.; Rondinelli, J. M., An efficient ab-initio quasi-harmonic approach for the thermodynamics of solids. *Computational Materials Science* **2016**, 120, 84-93.
130. Nath, P.; Plata, J. J.; Usanmaz, D.; Al Orabi, R. A. R.; Fornari, M.; Nardelli, M. B.; Toher, C.; Curtarolo, S., High-throughput prediction of finite-temperature properties using the quasi-harmonic approximation. *Computational Materials Science* **2016**, 125, 82-91.
131. Kyoto University Phonon Database; <http://phonondb.mtl.kyoto-u.ac.jp/ph20151124/index.html>. **2015**.
132. Hautier, G.; Fischer, C. C.; Jain, A.; Mueller, T.; Ceder, G., Finding nature's missing ternary oxide compounds using machine learning and density functional theory. *Chemistry of Materials* **2010**, 22, (12), 3762-3767.
133. Meredig, B.; Agrawal, A.; Kirklin, S.; Saal, J. E.; Doak, J.; Thompson, A.; Zhang, K.; Choudhary, A.; Wolverton, C., Combinatorial screening for new materials in unconstrained composition space with machine learning. *Physical Review B* **2014**, 89, (9), 094104.
134. Ghiringhelli, L. M.; Vybiral, J.; Levchenko, S. V.; Draxl, C.; Scheffler, M., Big Data of Materials Science: Critical Role of the Descriptor. *Physical Review Letters* **2015**, 114, (10), 105503.
135. Mueller, T.; Kusne, A. G.; Ramprasad, R., Machine learning in materials science: Recent progress and emerging applications. *Rev. Comput. Chem* **2015**.
136. Nosengo, N.; Ceder, G., Can artificial intelligence create the next wonder material? *Nature* **2016**, 533, (7601), 22-5.
137. Legrain, F.; Carrete, J.; van Roekeghem, A.; Curtarolo, S.; Mingo, N., How Chemical Composition Alone Can Predict Vibrational Free Energies and Entropies of Solids. *Chemistry of Materials* **2017**, 29, (15), 6220-6227.
138. Schmidt, M.; Lipson, H., Distilling Free-Form Natural Laws from Experimental Data. *Science* **2009**, 324, (5923), 81-85.
139. Makarov, D. E.; Metiu, H., Using Genetic Programming To Solve the Schrödinger Equation. *The Journal of Physical Chemistry A* **2000**, 104, (37), 8540-8545.
140. Brunton, S. L.; Proctor, J. L.; Kutz, J. N., Discovering governing equations from data by sparse identification of nonlinear dynamical systems. *Proceedings of the National Academy of Sciences* **2016**, 113, (15), 3932-3937.
141. Ouyang, R.; Curtarolo, S.; Ahmetcik, E.; Scheffler, M.; Ghiringhelli, L. M., SISSO: a compressed-sensing method for identifying the best low-dimensional descriptor in an immensity of offered candidates. *Physical Review Materials* **2018**, 2, (083802), 1-11.

142. Freysoldt, C.; Grabowski, B.; Hickel, T.; Neugebauer, J.; Kresse, G.; Janotti, A.; Van de Walle, C. G., First-principles calculations for point defects in solids. *Reviews of Modern Physics* **2014**, 86, (1), 253-305.
143. Jenkins, H. D. B.; Glasser, L., Standard Absolute Entropy, , Values from Volume or Density. 1. Inorganic Materials. *Inorganic Chemistry* **2003**, 42, (26), 8702-8708.
144. Fultz, B., Vibrational thermodynamics of materials. *Progress in Materials Science* **2010**, 55, (4), 247-352.
145. Zhao, L.-D.; Chang, C.; Tan, G.; Kanatzidis, M. G., SnSe: a remarkable new thermoelectric material. *Energy & Environmental Science* **2016**, 9, (10), 3044-3060.
146. Steinfeld, A., Solar hydrogen production via a two-step water-splitting thermochemical cycle based on Zn/ZnO redox reactions. *International Journal of Hydrogen Energy* **2002**, 27, (6), 611-619.
147. L'vov, B. V., Mechanism of carbothermal reduction of iron, cobalt, nickel and copper oxides. *Thermochimica Acta* **2000**, 360, (2), 109-120.
148. Rodriguez, J. A.; Hrbek, J., Interaction of sulfur with well-defined metal and oxide surfaces: unraveling the mysteries behind catalyst poisoning and desulfurization. *Accounts of Chemical Research* **1999**, 32, (9), 719-728.
149. Bertrand, G.; Mahdjoub, H.; Meunier, C., A study of the corrosion behaviour and protective quality of sputtered chromium nitride coatings. *Surface and Coatings Technology* **2000**, 126, (2), 199-209.
150. Michalsky, R.; Pfromm, P. H.; Steinfeld, A., Rational design of metal nitride redox materials for solar-driven ammonia synthesis. *Interface focus* **2015**, 5, (3), 20140084.
151. Sun, W.; Dacek, S. T.; Ong, S. P.; Hautier, G.; Jain, A.; Richards, W. D.; Gamst, A. C.; Persson, K. A.; Ceder, G., The thermodynamic scale of inorganic crystalline metastability. *Science Advances* **2016**, 2, (11), e1600225.
152. Hautier, G.; Jain, A.; Ong, S. P.; Kang, B.; Moore, C.; Doe, R.; Ceder, G., Phosphates as Lithium-Ion Battery Cathodes: An Evaluation Based on High-Throughput ab Initio Calculations. *Chemistry of Materials* **2011**, 23, (15), 3495-3508.
153. Aykol, M.; Dwarknath, S. S.; Sun, W.; Persson, K. A., Thermodynamic limit for synthesis of metastable inorganic materials. *Science Advances* **2018**, 4, (4).
154. Mueller, T.; Hautier, G.; Jain, A.; Ceder, G., Evaluation of Tavorite-Structured Cathode Materials for Lithium-Ion Batteries Using High-Throughput Computing. *Chemistry of Materials* **2011**, 23, (17), 3854-3862.

155. Ashton, M.; Paul, J.; Sinnott, S. B.; Hennig, R. G., Topology-Scaling Identification of Layered Solids and Stable Exfoliated 2D Materials. *Physical Review Letters* **2017**, 118, (10), 106101.
156. Sun, W.; Holder, A.; Orvananos, B.; Arca, E.; Zakutayev, A.; Lany, S.; Ceder, G., Thermodynamic Routes to Novel Metastable Nitrogen-Rich Nitrides. *Chemistry of Materials* **2017**.
157. Shoemaker, D. P.; Hu, Y.-J.; Chung, D. Y.; Halder, G. J.; Chupas, P. J.; Soderholm, L.; Mitchell, J. F.; Kanatzidis, M. G., In situ studies of a platform for metastable inorganic crystal growth and materials discovery. *Proceedings of the National Academy of Sciences* **2014**, 111, (30), 10922-10927.
158. Gopalakrishnan, J., Chimie Douce Approaches to the Synthesis of Metastable Oxide Materials. *Chemistry of Materials* **1995**, 7, (7), 1265-1275.
159. Aykol, M.; Hegde, V. L.; Suram, S.; Hung, L.; Herring, P.; Wolverton, C.; Hummelshoj, J. S., Network analysis of synthesizable materials discovery. *Preprint available at <https://arxiv.org/abs/1806.05772>* **2018**.
160. Balachandran, P. V.; Kowalski, B.; Sehirlioglu, A.; Lookman, T., Experimental search for high-temperature ferroelectric perovskites guided by two-step machine learning. *Nature Communications* **2018**, 9, 1668.
161. Ren, F.; Ward, L.; Williams, T.; Laws, K. J.; Wolverton, C.; Hatrick-Simpers, J.; Mehta, A., Accelerated discovery of metallic glasses through iteration of machine learning and high-throughput experiments. *Science Advances* **2018**, 4, (4).
162. Arca, E.; Lany, S.; Perkins, J. D.; Bartel, C.; Mangum, J.; Sun, W.; Holder, A.; Ceder, G.; Gorman, B.; Teeter, G.; Tumas, W.; Zakutayev, A., Redox-Mediated Stabilization in Zinc Molybdenum Nitrides. *Journal of the American Chemical Society* **2018**, 140, (12), 4293-4301.
163. Chen, J. G.; Crooks, R. M.; Seefeldt, L. C.; Bren, K. L.; Bullock, R. M.; Darensbourg, M. Y.; Holland, P. L.; Hoffman, B.; Janik, M. J.; Jones, A. K.; Kanatzidis, M. G.; King, P.; Lancaster, K. M.; Lyman, S. V.; Pfromm, P.; Schneider, W. F.; Schrock, R. R., Beyond fossil fuel-driven nitrogen transformations. *Science* **2018**, 360, (6391).
164. Wang, L.; Xia, M.; Wang, H.; Huang, K.; Qian, C.; Maravelias, C. T.; Ozin, G. A., Greening Ammonia toward the Solar Ammonia Refinery. *Joule* **2018**, 2, (6), 1055-1074.
165. Towards sustainable agriculture: Fossil-free ammonia. *Journal of Renewable and Sustainable Energy* **2017**, 9, (3), 034702.
166. Service, R., *Ammonia—a renewable fuel made from sun, air, and water—could power the globe without carbon*. 2018.

167. McEnaney, J. M.; Singh, A. R.; Schwalbe, J. A.; Kibsgaard, J.; Lin, J. C.; Cargnello, M.; Jaramillo, T. F.; Nørskov, J. K., Ammonia synthesis from N₂ and H₂O using a lithium cycling electrification strategy at atmospheric pressure. *Energy & Environmental Science* **2017**, 10, (7), 1621-1630.
168. Michalsky, R.; Avram, A. M.; Peterson, B. A.; Pfromm, P. H.; Peterson, A. A., Chemical looping of metal nitride catalysts: low-pressure ammonia synthesis for energy storage. *Chemical Science* **2015**, 6, (7), 3965-3974.
169. Muhich, C. L.; Ehrhart, B. D.; Al-Shankiti, I.; Ward, B. J.; Musgrave, C. B.; Weimer, A. W., A review and perspective of efficient hydrogen generation via solar thermal water splitting. *Wiley Interdisciplinary Reviews: Energy and Environment* **2016**, 5, (3), 261-287.
170. Chueh, W. C.; Falter, C.; Abbott, M.; Scipio, D.; Furler, P.; Haile, S. M.; Steinfeld, A., High-Flux Solar-Driven Thermochemical Dissociation of CO₂ and H₂O Using Nonstoichiometric Ceria. *Science* **2010**, 330, (6012), 1797-1801.
171. McDaniel, A. H.; Miller, E. C.; Arifin, D.; Ambrosini, A.; Coker, E. N.; O'Hayre, R.; Chueh, W. C.; Tong, J., Sr- and Mn-doped LaAlO₃- δ for solar thermochemical H₂ and CO production. *Energy & Environmental Science* **2013**, 6, (8), 2424-2428.
172. Muhich, C. L.; Ehrhart, B. D.; Witte, V. A.; Miller, S. L.; Coker, E. N.; Musgrave, C. B.; Weimer, A. W., Predicting the solar thermochemical water splitting ability and reaction mechanism of metal oxides: a case study of the hercynite family of water splitting cycles. *Energy & Environmental Science* **2015**, 8, (12), 3687-3699.
173. Gálvez, M. E.; Halmann, M.; Steinfeld, A., Ammonia Production via a Two-Step Al₂O₃/AlN Thermochemical Cycle. 1. Thermodynamic, Environmental, and Economic Analyses. *Industrial & Engineering Chemistry Research* **2007**, 46, (7), 2042-2046.
174. Michalsky, R.; Parman, B. J.; Amanor-Boadu, V.; Pfromm, P. H., Solar thermochemical production of ammonia from water, air and sunlight: Thermodynamic and economic analyses. *Energy* **2012**, 42, (1), 251-260.
175. Gao, Y.; Wu, Y.; Zhang, Q.; Chen, X.; Jiang, G.; Liu, D., N-desorption or NH₃ generation of TiO₂-loaded Al-based nitrogen carrier during chemical looping ammonia generation technology. *International Journal of Hydrogen Energy* **2018**.
176. Michalsky, R.; Pfromm, P. H., Thermodynamics of metal reactants for ammonia synthesis from steam, nitrogen and biomass at atmospheric pressure. *AIChE Journal* **2012**, 58, (10), 3203-3213.
177. Bartel, C. J.; Millican, S. L.; Deml, A. M.; Rumpitz, J. R.; Tumas, W.; Weimer, A. W.; Lany, S.; Stevanović, V.; Musgrave, C. B.; Holder, A. M., Physical descriptor for the Gibbs energy

of inorganic crystalline solids and temperature-dependent materials chemistry. *Nature Communications* **2018**, 9, (1), 4168.

178. Bartel, C. J.; Weimer, A. W.; Lany, S.; Musgrave, C. B.; Holder, A., The role of decomposition reactions in assessing first-principles predictions of solid stability. *arXiv:1810.08248* **2018**.

179. Bartel, C. J.; Muhich, C. L.; Weimer, A. W.; Musgrave, C. B., Aluminum Nitride Hydrolysis Enabled by Hydroxyl-Mediated Surface Proton Hopping. *ACS Applied Materials & Interfaces* **2016**, 8, (28), 18550-18559.

180. Michalsky, R.; Pfromm, P. H., An Ionicity Rationale to Design Solid phase Metal Nitride Reactants for Solar Ammonia Production. *The Journal of Physical Chemistry C* **2012**, 116, (44), 23243-23251.

181. White, A. H.; Melville, W., THE DECOMPOSITION OF AMMONIA AT HIGH TEMPERATURES. *Journal of the American Chemical Society* **1905**, 27, (4), 373-386.

182. Taylor, G. B.; Starkweather, H. W., REDUCTION OF METAL OXIDES BY HYDROGEN. *Journal of the American Chemical Society* **1930**, 52, (6), 2314-2325.

183. Halman, M.; Frei, A.; Steinfeld, A., Magnesium Production by the Pidgeon Process Involving Dolomite Calcination and MgO Silicothermic Reduction: Thermodynamic and Environmental Analyses. *Industrial & Engineering Chemistry Research* **2008**, 47, (7), 2146-2154.

184. Chubukov, B. A.; Palumbo, A. W.; Rowe, S. C.; Wallace, M. A.; Weimer, A. W., Enhancing the Rate of Magnesium Oxide Carbothermal Reduction by Catalysis, Milling, and Vacuum Operation. *Industrial & Engineering Chemistry Research* **2017**, 56, (46), 13602-13609.

185. Heidlage, M. G.; Kezar, E. A.; Snow, K. C.; Pfromm, P. H., Thermochemical Synthesis of Ammonia and Syngas from Natural Gas at Atmospheric Pressure. *Industrial & Engineering Chemistry Research* **2017**, 56, (47), 14014-14024.

186. Elder, S. H.; DiSalvo, F. J.; Topor, L.; Navrotsky, A., Thermodynamics of ternary nitride formation by ammonolysis: application to lithium molybdenum nitride (LiMoN₂), sodium tungsten nitride (Na₃WN₃), and sodium tungsten oxide nitride (Na₃WO₃N). *Chemistry of Materials* **1993**, 5, (10), 1545-1553.

187. Eriksson, G., Thermodynamic Studies of High Temperature Equilibria. *Acta Chemica Scandinavica* **1971**, 25, 2651-2658.

188. Michalsky, R.; Pfromm, P. H., Chromium as reactant for solar thermochemical synthesis of ammonia from steam, nitrogen, and biomass at atmospheric pressure. *Solar Energy* **2011**, 85, (11), 2642-2654.

189. Michalsky, R.; Pfromm, P. H.; Steinfeld, A., Rational design of metal nitride redox materials for solar-driven ammonia synthesis. *Interface Focus* **2015**, 5, (3).
190. Scheffe, J. R.; Weibel, D.; Steinfeld, A., Lanthanum–Strontium–Manganese Perovskites as Redox Materials for Solar Thermochemical Splitting of H₂O and CO₂. *Energy & Fuels* **2013**, 27, (8), 4250-4257.
191. DiSalvo, F. J.; Clarke, S. J., Ternary nitrides: a rapidly growing class of new materials. *Current Opinion in Solid State and Materials Science* **1996**, 1, (2), 241-249.
192. Niewa, R.; DiSalvo, F. J., Recent Developments in Nitride Chemistry. *Chemistry of Materials* **1998**, 10, (10), 2733-2752.
193. Hiroshi, A.; Masahiro, K.; Kazumasa, H.; Isamu, A., P-Type Conduction in Mg-Doped GaN Treated with Low-Energy Electron Beam Irradiation (LEEBI). *Japanese Journal of Applied Physics* **1989**, 28, (12A), L2112.
194. Nakamura, S., The Roles of Structural Imperfections in InGaN-Based Blue Light-Emitting Diodes and Laser Diodes. *Science* **1998**, 281, (5379), 956-961.
195. Knotek, O.; Münz, W. D.; Leyendecker, T., Industrial deposition of binary, ternary, and quaternary nitrides of titanium, zirconium, and aluminum. *Journal of Vacuum Science & Technology A* **1987**, 5, (4), 2173-2179.
196. Zeinalipour-Yazdi, C. D.; Hargreaves, J. S. J.; Laassiri, S.; Catlow, C. R. A., The integration of experiment and computational modelling in heterogeneously catalysed ammonia synthesis over metal nitrides. *Physical Chemistry Chemical Physics* **2018**, 20, (34), 21803-21808.
197. Jacobsen, C. J. H., Novel class of ammonia synthesis catalysts. *Chemical Communications* **2000**, (12), 1057-1058.
198. Sugita, Y.; Mitsuoka, K.; Komuro, M.; Hoshiya, H.; Kozono, Y.; Hanazono, M., Giant magnetic moment and other magnetic properties of epitaxially grown Fe₁₆N₂ single-crystal films (invited). *Journal of Applied Physics* **1991**, 70, (10), 5977-5982.
199. Yamanaka, S.; Kawaji, H.; Hotehama, K.-i.; Ohashi, M., A new layer-structured nitride superconductor. Lithium-intercalated β -zirconium nitride chloride, Li_xZrNCl. *Advanced Materials* **1996**, 8, (9), 771-774.
200. Vinokur, V. M.; Baturina, T. I.; Fistul, M. V.; Mironov, A. Y.; Baklanov, M. R.; Strunk, C., Superinsulator and quantum synchronization. *Nature* **2008**, 452, 613.
201. Lee, K.; Kim, S. W.; Toda, Y.; Matsuishi, S.; Hosono, H., Dicalcium nitride as a two-dimensional electride with an anionic electron layer. *Nature* **2013**, 494, 336.
202. Zakutayev, A., Design of nitride semiconductors for solar energy conversion. *Journal of Materials Chemistry A* **2016**, 4, (18), 6742-6754.

203. Mori-Sánchez, P.; Marqués, M.; Beltrán, A.; Jiang, J. Z.; Gerward, L.; Recio, J. M., Origin of the low compressibility in hard nitride spinels. *Physical Review B* **2003**, 68, (6), 064115.
204. McHale, J. M.; Navrotsky, A.; DiSalvo, F. J., Energetics of Ternary Nitride Formation in the (Li,Ca)–(B,Al)–N System. *Chemistry of Materials* **1999**, 11, (4), 1148-1152.
205. McHale, J. M.; Navrotsky, A.; Kowach, G. R.; Balbarin, V. E.; DiSalvo, F. J., Energetics of Ternary Nitrides: Li–Ca–Zn–N and Ca–Ta–N Systems. *Chemistry of Materials* **1997**, 9, (7), 1538-1546.
206. Slack, G. A.; Tanzilli, R. A.; Pohl, R. O.; Vandersande, J. W., The intrinsic thermal conductivity of AlN. *Journal of Physics and Chemistry of Solids* **1987**, 48, (7), 641-647.
207. Watari, K.; Hwang, H. J.; Toriyama, M.; Kanzaki, S., Effective Sintering Aids for Low-temperature Sintering of AlN Ceramics. *Journal of Materials Research* **1999**, 14, (4), 1409-1417.
208. Weimer, A. W., *Carbide, Nitride and Boride Materials Synthesis and Processing*. Springer Netherlands: 1997.
209. Kirova-Yordanova, Z., Exergy analysis of industrial ammonia synthesis. *Energy* **2004**, 29, (12), 2373-2384.
210. Agmon, N., The Grotthuss mechanism. *Chemical Physics Letters* **1995**, 244, (5), 456-462.
211. Gálvez, M. E.; Frei, A.; Halmann, M.; Steinfeld, A., Ammonia Production via a Two-Step Al₂O₃/AlN Thermochemical Cycle. 2. Kinetic Analysis. *Industrial & Engineering Chemistry Research* **2007**, 46, (7), 2047-2053.
212. Niewa, R.; Jacobs, H., Group V and VI Alkali Nitridometalates: A Growing Class of Compounds with Structures Related to Silicate Chemistry. *Chemical Reviews* **1996**, 96, (6), 2053-2062.
213. Hinuma, Y.; Hatakeyama, T.; Kumagai, Y.; Burton, L. A.; Sato, H.; Muraba, Y.; Iimura, S.; Hiramatsu, H.; Tanaka, I.; Hosono, H.; Oba, F., Discovery of earth-abundant nitride semiconductors by computational screening and high-pressure synthesis. *Nature Communications* **2016**, 7, 11962.
214. Höhn, P.; Kniep, R., Crystal structure of tristrontium [tetranitridomolybdate(VI)], Sr₃[MoN₄]. In *Zeitschrift für Kristallographie - New Crystal Structures*, 2000; Vol. 215, p 325.
215. Burgmayer, S. J. N.; Stiefel, E. I., Molybdenum enzymes, cofactors, and systems: The chemical uniqueness of molybdenum. *Journal of Chemical Education* **1985**, 62, (11), 943.
216. Brown, K. A.; Harris, D. F.; Wilker, M. B.; Rasmussen, A.; Khadka, N.; Hamby, H.; Keable, S.; Dukovic, G.; Peters, J. W.; Seefeldt, L. C.; King, P. W., Light-driven dinitrogen

reduction catalyzed by a CdS:nitrogenase MoFe protein biohybrid. *Science* **2016**, 352, (6284), 448-450.

217. Robinson, W. E.; Bassegoda, A.; Reisner, E.; Hirst, J., Oxidation-State-Dependent Binding Properties of the Active Site in a Mo-Containing Formate Dehydrogenase. *Journal of the American Chemical Society* **2017**, 139, (29), 9927-9936.

218. Perret, N.; Alexander, A.-M.; Hunter, S. M.; Chung, P.; Hargreaves, J. S. J.; Howe, R. F.; Keane, M. A., Synthesis, characterisation and hydrogenation performance of ternary nitride catalysts. *Applied Catalysis A: General* **2014**, 488, 128-137.

219. Zhang, Y.; Ouyang, B.; Xu, J.; Chen, S.; Rawat, R. S.; Fan, H. J., 3D Porous Hierarchical Nickel–Molybdenum Nitrides Synthesized by RF Plasma as Highly Active and Stable Hydrogen-Evolution-Reaction Electrocatalysts. *Advanced Energy Materials* **2016**, 6, (11), 1600221.

220. Maintz, S.; Deringer, V. L.; Tchougréeff, A. L.; Dronskowski, R., LOBSTER: A tool to extract chemical bonding from plane-wave based DFT. *Journal of Computational Chemistry* **2016**, 37, (11), 1030-1035.

221. Manz, T. A.; Limas, N. G., Introducing DDEC6 atomic population analysis: part 1. Charge partitioning theory and methodology. *RSC Advances* **2016**, 6, (53), 47771-47801.

222. Tabor, D. P.; Roch, L. M.; Saikin, S. K.; Kreisbeck, C.; Sheberla, D.; Montoya, J. H.; Dwaraknath, S.; Aykol, M.; Ortiz, C.; Tribukait, H.; Amador-Bedolla, C.; Brabec, C. J.; Maruyama, B.; Persson, K. A.; Aspuru-Guzik, A., Accelerating the discovery of materials for clean energy in the era of smart automation. *Nature Reviews Materials* **2018**, 3, (5), 5-20.

223. Collins, C.; Dyer, M. S.; Pitcher, M. J.; Whitehead, G. F. S.; Zanella, M.; Mandal, P.; Claridge, J. B.; Darling, G. R.; Rosseinsky, M. J., Accelerated discovery of two crystal structure types in a complex inorganic phase field. *Nature* **2017**, 546, 280.

224. Gautier, R.; Zhang, X.; Hu, L.; Yu, L.; Lin, Y.; Sunde, T. O. L.; Chon, D.; Poeppelmeier, K. R.; Zunger, A., Prediction and accelerated laboratory discovery of previously unknown 18-electron ABX compounds. *Nature Chemistry* **2015**, 7, 308.

225. Jain, A.; Hautier, G.; Ong, S. P.; Persson, K., New opportunities for materials informatics: Resources and data mining techniques for uncovering hidden relationships. *Journal of Materials Research* **2016**, 31, (8), 977-994.

226. Parr, R. G.; Yang, W., Density functional approach to the frontier-electron theory of chemical reactivity. *Journal of the American Chemical Society* **1984**, 106, (14), 4049-4050.

227. Hoffmann, R., *How Chemistry and Physics Meet in the Solid State*. 1987; Vol. 26, p 846-878.

228. Tholander, C.; Andersson, C. B. A.; Armiento, R.; Tasnádi, F.; Alling, B., Strong piezoelectric response in stable TiZnN₂, ZrZnN₂, and HfZnN₂ found by ab initio high-throughput approach. *Journal of Applied Physics* **2016**, 120, (22), 225102.
229. Gharavi, M. A.; Armiento, R.; Alling, B.; Eklund, P., Theoretical study of phase stability, crystal and electronic structure of MeMgN₂ (Me = Ti, Zr, Hf) compounds. *Journal of Materials Science* **2018**, 53, (6), 4294-4305.
230. Ching, W.-Y.; Mo, S.-D.; Ouyang, L.; Rulis, P.; Tanaka, I.; Yoshiya, M., *Theoretical Prediction of the Structure and Properties of Cubic Spinel Nitrides*. 2004; Vol. 85, p 75-80.
231. Sarmiento-Pérez, R.; Cerqueira, T. F. T.; Körbel, S.; Botti, S.; Marques, M. A. L., Prediction of Stable Nitride Perovskites. *Chemistry of Materials* **2015**, 27, (17), 5957-5963.
232. Bannikov, V. V.; Shein, I. R.; Ivanovskii, A. L., Elastic properties of antiperovskite-type Ni-rich nitrides MNi₃ (M=Zn, Cd, Mg, Al, Ga, In, Sn, Sb, Pd, Cu, Ag and Pt) as predicted from first-principles calculations. *Physica B: Condensed Matter* **2010**, 405, (22), 4615-4619.
233. Hautier, G.; Fischer, C.; Ehlacher, V.; Jain, A.; Ceder, G., Data Mined Ionic Substitutions for the Discovery of New Compounds. *Inorganic Chemistry* **2011**, 50, (2), 656-663.
234. Kemp, C.; Tenenbaum, J. B., The discovery of structural form. *Proceedings of the National Academy of Sciences* **2008**, 105, (31), 10687-10692.
235. Müllner, D., *Fastcluster: Fast Hierarchical, Agglomerative Clustering Routines for R and Python*. 2013; Vol. 53, p 1-18.
236. Gower, J., *A General Coefficient of Similarity and Some of Its Properties*. 1971; Vol. 27, p 857-871.
237. Lany, S.; Fioretti, A. N.; Zawadzki, P. P.; Schelhas, L. T.; Toberer, E. S.; Zakutayev, A.; Tamboli, A. C., Monte Carlo simulations of disorder in $\text{ZnSn}_{1-x}\text{N}_x$ and the effects on the electronic structure. *Physical Review Materials* **2017**, 1, (3), 035401.
238. Bartel, C. J.; Millican, S. L.; Deml, A. M.; Rumpitz, J. R.; Tumas, W.; Weimer, A. W.; Lany, S.; Stevanović, V.; Musgrave, C. B.; Holder, A. M., Physical descriptor for the Gibbs energy of inorganic crystalline solids and prediction of temperature-dependent materials chemistry. *arXiv preprint arXiv:1805.08155* **2018**.
239. Choi, J.; Gillan, E. G., Solvothermal Metal Azide Decomposition Routes to Nanocrystalline Metastable Nickel, Iron, and Manganese Nitrides. *Inorganic Chemistry* **2009**, 48, (10), 4470-4477.
240. Kroll, P.; Schröter, T.; Peters, M., Prediction of Novel Phases of Tantalum(V) Nitride and Tungsten(VI) Nitride That Can Be Synthesized under High Pressure and High Temperature. *Angewandte Chemie International Edition* **2005**, 44, (27), 4249-4254.

241. Caskey, C. M.; Richards, R. M.; Ginley, D. S.; Zakutayev, A., Thin film synthesis and properties of copper nitride, a metastable semiconductor. *Materials Horizons* **2014**, 1, (4), 424-430.
242. Caskey, C. M.; Holder, A.; Shulda, S.; Christensen, S. T.; Diercks, D.; Schwartz, C. P.; Biagioni, D.; Nordlund, D.; Kukliansky, A.; Natan, A.; Prendergast, D.; Orvananos, B.; Sun, W.; Zhang, X.; Ceder, G.; Ginley, D. S.; Tumas, W.; Perkins, J. D.; Stevanovic, V.; Pylypenko, S.; Lany, S.; Richards, R. M.; Zakutayev, A., Synthesis of a mixed-valent tin nitride and considerations of its possible crystal structures. *The Journal of Chemical Physics* **2016**, 144, (14), 144201.
243. Bikowski, A.; Siol, S.; Gu, J.; Holder, A.; Mangum, J. S.; Gorman, B.; Tumas, W.; Lany, S.; Zakutayev, A., Design of Metastable Tin Titanium Nitride Semiconductor Alloys. *Chemistry of Materials* **2017**, 29, (15), 6511-6517.
244. Yang, M.; Zakutayev, A.; Vidal, J.; Zhang, X.; Ginley, D. S.; DiSalvo, F. J., Strong optical absorption in CuTaN₂ nitride delafossite. *Energy & Environmental Science* **2013**, 6, (10), 2994-2999.
245. Whittaker, M. L.; Sun, W.; DeRocher, K. A.; Jayaraman, S.; Ceder, G.; Joester, D., Structural Basis for Metastability in Amorphous Calcium Barium Carbonate (ACBC). *Advanced Functional Materials* **2018**, 28, (2), 1704202.
246. Kuech, T. F.; Babcock, S. E.; Mawst, L., Growth far from equilibrium: Examples from III-V semiconductors. *Applied Physics Reviews* **2016**, 3, (4), 040801.
247. Vepřek, S.; Reiprich, S., A concept for the design of novel superhard coatings. *Thin Solid Films* **1995**, 268, (1), 64-71.
248. Pearson, R. G., Recent advances in the concept of hard and soft acids and bases. *Journal of Chemical Education* **1987**, 64, (7), 561.
249. Anderson, W. P.; Burdett, J. K.; Czech, P. T., What Is the Metallic Bond? *Journal of the American Chemical Society* **1994**, 116, (19), 8808-8809.
250. Donohue, J., Molecules and crystals in inorganic chemistry. *Journal of Chemical Education* **1950**, 27, (3), 169.
251. Etourneau, J.; Portier, J.; Ménil, F., The role of the inductive effect in solid state chemistry: how the chemist can use it to modify both the structural and the physical properties of the materials. *Journal of Alloys and Compounds* **1992**, 188, 1-7.
252. Bem, D. S.; Gibson, C. P.; zur Loye, H. C., Synthesis of intermetallic nitrides by solid-state precursor reduction. *Chemistry of Materials* **1993**, 5, (4), 397-399.

253. Prior, T. J.; Battle, P. D., Superparamagnetism and metal-site ordering in quaternary nitrides with the η -carbide structure. *Journal of Materials Chemistry* **2004**, 14, (20), 3001-3007.
254. Bing, H.; Cheng, D.; Lihong, Y.; Xiaochao, C.; Linhui, G.; Libin, M.; Youguo, S., CuNNi₃ : a new nitride superconductor with antiperovskite structure. *Superconductor Science and Technology* **2013**, 26, (12), 125015.
255. Zu, L.; Lin, S.; Lin, J.; Yuan, B.; Kan, X.; Tong, P.; Song, W.; Sun, Y., Observation of the Spin-Glass Behavior in Co-Based Antiperovskite Nitride GeNCo₃. *Inorganic Chemistry* **2016**, 55, (18), 9346-9351.
256. Schmerler, S.; Kortus, J., Ab initio study of AlN: Anisotropic thermal expansion, phase diagram, and high-temperature rocksalt to wurtzite phase transition. *Physical Review B* **2014**, 89, (6), 064109.
257. Holec, D.; Mayrhofer, P. H., Surface energies of AlN allotropes from first principles. *Scripta Materialia* **2012**, 67, (9), 760-762.
258. Henkelman, G.; Arnaldsson, A.; Jónsson, H., A fast and robust algorithm for Bader decomposition of charge density. *Computational Materials Science* **2006**, 36, (3), 354-360.
259. Henkelman, G.; Uberuaga, B. P.; Jónsson, H., A climbing image nudged elastic band method for finding saddle points and minimum energy paths. *The Journal of Chemical Physics* **2000**, 113, (22), 9901-9904.
260. Henkelman, G.; Jónsson, H., A dimer method for finding saddle points on high dimensional potential surfaces using only first derivatives. *The Journal of Chemical Physics* **1999**, 111, (15), 7010-7022.
261. Jain, A.; Hautier, G.; Moore, C. J.; Ping Ong, S.; Fischer, C. C.; Mueller, T.; Persson, K. A.; Ceder, G., A high-throughput infrastructure for density functional theory calculations. *Computational Materials Science* **2011**, 50, (8), 2295-2310.
262. Ong, S. P.; Cholia, S.; Jain, A.; Brafman, M.; Gunter, D.; Ceder, G.; Persson, K. A., The Materials Application Programming Interface (API): A simple, flexible and efficient API for materials data based on REpresentational State Transfer (REST) principles. *Computational Materials Science* **2015**, 97, 209-215.