Modeling Complex Multicomponent Materials with First-Principles Based Statistical Mechanics and Machine Learning

David E. Sommer

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington

2021

Reading Committee:

Scott T. Dunham, Chair

Kai-Mei C. Fu

Anton V. Andreev

Program Authorized to Offer Degree:
Department of Physics

University of Washington

Abstract

Modeling Complex Multicomponent Materials with First-Principles Based Statistical Mechanics and Machine Learning

David E. Sommer

Chair of the Supervisory Committee:
Scott T. Dunham
Department of Electrical Engineering

Recent progress in the engineering of multicomponent, solid-state compounds for opto-electronic applications has entailed an ever expanding range of material chemistries and a rapid increase in material complexity. For example, within the classes of chalcogenide and halide-perovskite semiconductors, fundamental material properties can be effectively tuned by alloying various isovalent chemical species and by the controlled incorporation of dopants. In characterizing these materials at an atomistic level, one has to contend not only with the presence of a multitude of point defects, but also with the potential formation of ordering and instabilities against secondary phases. This poses a fundamental challenge to first-principles modeling that is only exacerbated by an exponentially large configuration space. This work is primarily concerned with the development and application of methods, rooted in statistical mechanics and machine learning, for modeling these complex, multicomponent semiconductors. Much of this manuscript focuses on modeling defects and configurational disorder in specific chalcogenides and halide-perovskites of particular technological relevance, employing well-established multiscale methods such as density functional theory, point defect thermodynamics, statistical learning, cluster expansions and Monte Carlo simulation. The latter portion of this work concerns more recent developments in the fields of deep learning and tensor networks, adapted and applied to material structure-property prediction.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would foremost like to thank Scott Dunham for his consistent support during the completion of this work. He has been an invaluable guide to the world of materials modeling and an indispensable source of deep knowledge across many disciplines. As an academic mentor, he set a high bar for research without which much of this work would be woefully incomplete. I would also like to thank Mari Ostendorf, who, along with Scott, provided the initial impetus for the machine learning research that eventually evolved into the latter part of this work. I owe many thanks to the members of the Clean Energy Institute and the Molecular Engineering Materials Center at the University of Washington, who fostered a rich atmosphere of interdisciplinary collaboration. In particular, I would like to thank Daniel Gamelin for helping to keep much of this research grounded by experimental evidence and for introducing me to the wide world of perovskites. Indeed, much of this work would not have happened without the extensive collaborative efforts from the following current and former members of the Gamelin group: Sidney Creutz, Mike de Siena, Diana Roh, Tyler Milstein, and Kyle Kluherz. Finally, special thanks are owed to current and former members of the Dunham group whom I have had the pleasure to work with over the years: Daniel Mutter, Yu Jin, Shifeng Zhu, Xiaofeng Xiang, Aaron Gehrke, Yijun Tong, and Anna Merkulova.

DEDICATION

For Sarah

Chapter 1

INTRODUCTION

Advances in electronic structure methods and large-scale computing have enabled increasingly more accurate predictions of fundamental properties of condensed matter systems from first principles. At a basic level, these include electronic band structure, ground state energies and lattice structure, interatomic forces and, to a certain extent, vibrational and electronic quasiparticle excitation spectra. Given the widespread availability of state-of-the-art software for first-principles electronic structure calculations, it is now the case that these methods form an essential component of condensed matter and materials science research. Nonetheless, certain properties that implicate the technological application of a material, such as atomic kinetics and phase equilibria, require characterization at length and time scales which remain computationally inaccessible from first principles. To bridge these gaps, significant efforts have been applied to the development of multiscale modeling methods, often parameterized by *ab initio* data. A summary of standard multiscale methods is shown in Figure 1.1.

In the search for new semiconductors for novel optoelectronic applications, there is a persistent need for atomistic modeling methods. For instance, point defects often play a crucial role in dictating the physical properties of a material but are difficult to identify and characterize experimentally. In the limit that the concentrations of relevant defects are dilute, there exist well-established formalisms [72] based on density functional theory (DFT) and many-body perturbation theory (MBPT) for modeling fundamental defect properties, such as formation energies and charge-state transition levels, as well as their thermodynamics and diffusion. In this limit, a point defect is viewed as an isolated perturbation

in an otherwise pristine, crystalline environment (cf. Figure 1.4). Importantly, these dilute approximations remain valid as long as defect-defect interactions are negligible, typical for defect concentrations less than 1 atomic %. This picture breaks down as the degree of point-like disorder increases and the spatial arrangement of various defects and impurities becomes increasingly correlated. In this case, the distinction between the original host and the system of defects becomes ill-defined, and the space of possible configurations grows exponentially. This non-dilute regime is particularly difficult to capture with computationally intensive first-principles methods, where each calculation requires one to explicitly specify an initial arrangement atoms in the system and subsequent optimization is limited to relatively local displacements. However, the expanding range of technologically relevant multicomponent compounds necessitates theoretical methods that can handle these highly disordered systems.

The search for defect-tolerant semiconductors [287, 30, 292, 283, 118] and improvements in synthetic control over their chemical composition and stability has driven a recent expansion in the viable design space of optoelectronic materials. For example, substantial progress has been made in recent years in engineering efficient, low-cost thin-film photovoltaics (PVs) based on absorber layers which are increasingly complex in terms of both atomic structure and chemistry. This is exemplified by the development of multicomponent chalcogenides, which are derived from an elemental zinc-blend lattice by repeatedly replacing subgroups of sites by valence-compensated pairs of ions (Figure 1.2). Absorbers made from $Cu(In,Ga)Se_2$ (CIGS), in particular, have achieved PV device efficiencies exceeding 20% despite significant off-stoichiometries and polycrystalline morphologies. Due to their semi-ionic character, these materials are heavily compensated by the formation of a diverse range of point defects and defect complexes [292, 201, 235]. Understanding how defect formation, clustering, and diffusion controls the stability and electronic properties of these materials continues to be an important task for improving these PV technologies.

Metal halide perovskites (MHPs), with general formula ABX_3 (A = Cs, CH_3N_3, CH_5N_2; B = Sn, Pb; X = Cl, Br, I), constitute another technologically important class of complex,

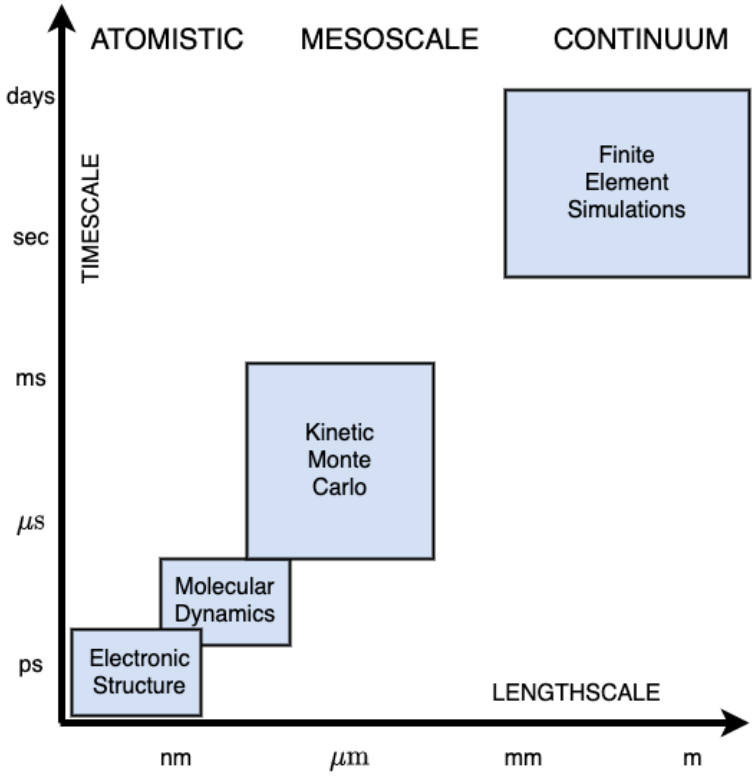


Figure 1.1: Standard multiscale methods for modeling condensed matter organized by typically accessible length- and timescales.

defect-tolerant semiconductors with an ever diversifying range of energy and lighting applications [294, 207, 181, 2, 17, 56, 152]. Binary mixtures of halides in the perovskite lattice are particularly stable, and the high room-temperature mobility of substitutional halides has enabled a substantial degree of synthetic control over the band gap [207, 2]. Doping by impurity ions, such as transition metals and lanthanides, has also emerged as an effective route toward engineering additional optical and electronic behavior and controlling perovskite structural stability [17, 56, 152, 144, 39, 164, 270, 225, 296, 194, 153, 288, 171, 212]. However, even the simplest cases involving undoped, mixed halide perovskites can exhibit complex thermodynamic behavior, such as photoinduced phase separation [107, 16, 27], which can have

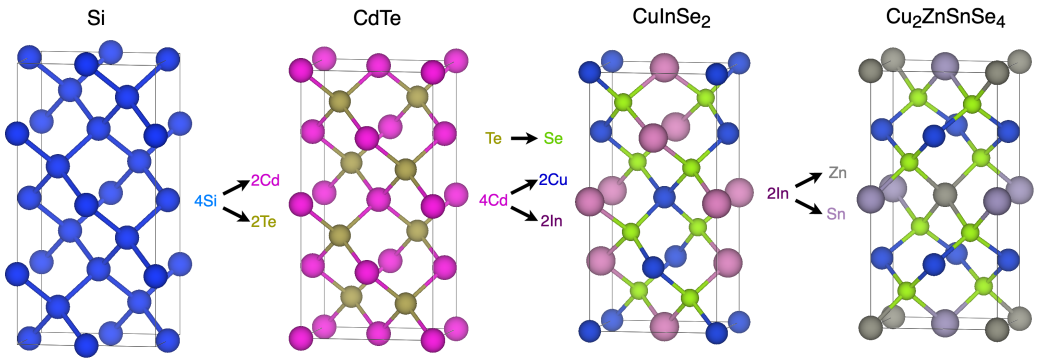


Figure 1.2: Relationship between various semiconductor compounds derived from an elemental zinc-blend structure.

important implications on subsequent device performance [254]. The introduction of dopants further complicates this picture. For instance, Mn-doping-induced phase separation was recently proposed in [225] to explain changes in Mn EPR signal in nanocrystals of $CsPbX_3$ under Cl $\rightarrow$ Br anion exchange (Figure 1.3).

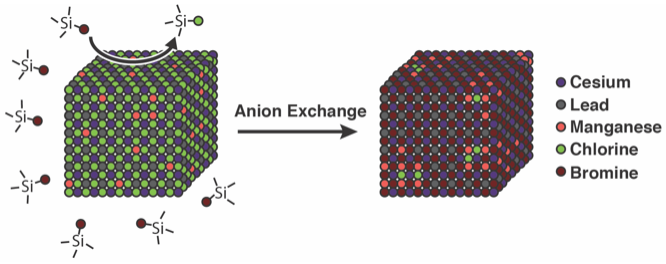


Figure 1.3: Manganese clustering during Cl $\rightarrow$ Br anion exchange in $CsPbCl_3$. (Figure from [225].)

In this work, we study defects and configurational disorder in these systems, combining electronic structure methods with statistical mechanics approaches of increasing complexity. In Chapter 2, we focus on the chalcogenide $CuInSe_2$, using DFT to characterize a myriad of intrinsic defects and defect complexes. With this *ab initio* data, we then construct ther-

modynamic models of defect formation as a function material composition. This allows us to connect non-stoichiometry to defect formation, phase stability, and changes in material conductivity through self-doping. Building on this equilibrium picture, we apply the ladder of multiscale methods shown in Figure 1.1 toward understanding the diffusion of Cu in this material, attempting to resolve certain discrepancies in the experimental literature.

In Chapter 3, we model ordering in mixed halide perovskites and consider potential implications to halide exchange reactions, both in the undoped case and with Mn doping. To do so, we construct effective lattice Hamiltonians in terms of the atomic configurational degrees of freedom, deriving interaction coefficients from DFT using statistical regression. To explore finite-temperature thermodynamics, we sample these effective models using Markov Chain Monte Carlo and connect our observations of phase separation to recent experimental work [225]. We then return to the single halide perovskite $CsPbCl_3$ and characterize the formation of defects and defect complexes involving Yb dopants. We discuss the implications of our findings to recent observations of Yb-mediated solar downconversion / quantum cutting in these materials [171, 212].

The statistical mechanics approaches used in Chapters 2 and 3 rely either on dilute / noninteracting approximations or on the assumption that the DFT ground state structures of various alloy configurations can be mapped unequivocally to the same underlying lattice. The latter assumption is violated in the presence of large lattice distortions or competing structural phases. To move beyond these limitations, we consider, in Chapter 4, the general construction of translationally and rotationally invariant atom-density representations, which we use as inputs to machine learning algorithms. Motivated by the development of tensor network methods in many-body quantum theory, we show how tensor networks can be adapted to machine learning tasks, both supervised and unsupervised, defined on these atom-density bases. We benchmark these methods on ternary alloys of transparent conducting oxides and binary metal alloys, finding that they maintain strong performance when the amount of training data is limited. We then show how these techniques could be applied to the compression of atom-density representations and extended to SO(3)-equivariant cases,

and we discuss promising connections between model expressibility and the entanglement structure encoded in the model weights during the course of training.

Figure 1.4 summarizes some of the fundamental techniques employed in this work for obtaining coarse-grained atomistic models which are rooted in statistical mechanics and machine learning. They are organized in terms of the increasing complexity of the structural and chemical degrees of freedom that they attempt to capture, and the main contributions of each chapter roughly follow the order of structural complexity in the diagram. For example, the results of Chapter 2 rely on a model of intrinsic defect thermodynamics in $CuInSe_2$ coupled via compositional constraints. Much of Chapter 3 goes beyond this relatively isolated point defect picture to account for higher degrees of chemical and configurational disorder. Finally, in Chapter 4, the construction of atomistic representations based on local expansions of the atomic density and the explicit incorporation of fundamental symmetries allows large atomic relaxations relaxations, multiple lattices symmetries, and multiple chemical components to be treated with a single effective model.

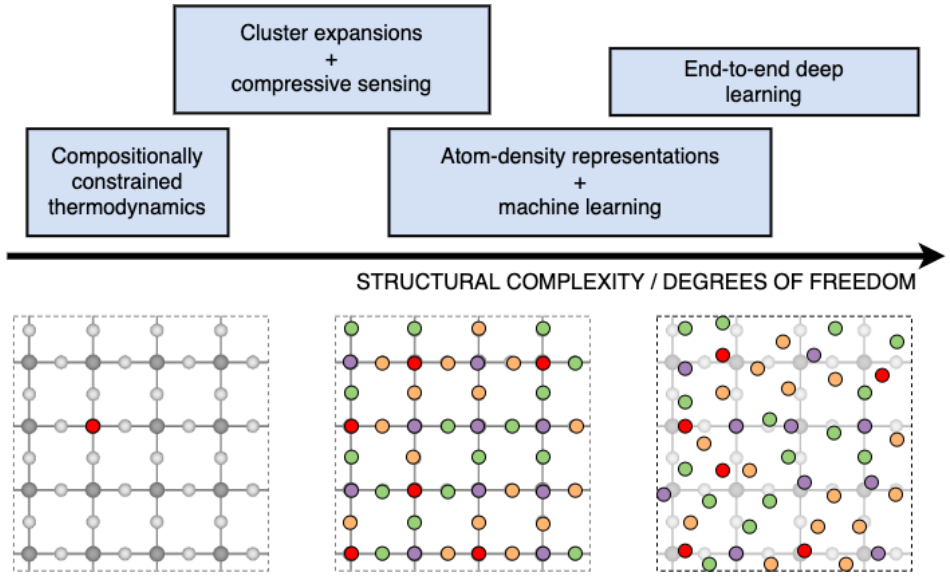


Figure 1.4: Coarse-grained modeling techniques developed and/or applied in this work, organized by applicability to material degrees of freedom with increasing complexity.

Chapter 2

CHALCOPYRITES: THERMODYNAMICS AND KINETICS OF POINT DEFECTS

The chalcopyrite semiconductor $CuInSe_2$ (CIS) is widely recognized as a promising absorber layer for low-cost, high-efficiency thin-film photovoltaics (PVs) due to its intrinsic tolerance to the presence of crystalline disorder. Indeed, much effort has been applied toward characterizing the atomistic defect properties of the material and the development of optimal protocols for device fabrication. For example, it is well-appreciated that CIS is heavily compensated by the formation of a diverse range of point defects and their complexes [201], leading to stable material compositions which can deviate substantially from ideal stoichiometry [224]. This flexibility is often advantageous, as it allows the electronic properties of the material to be effectively controlled by the material synthesis conditions and post-synthetic defect engineering. Indeed, such compositional tuning is facilitated by the typically high mobility of cations diffusing through Cu sublattice [102].

In this chapter, we study the formation of a wide variety of intrinsic point defects and their complexes in CIS. We start from first-principles density functional theory (DFT) calculations, which provide a basic characterization of the energetics of defect formation and their electronic behavior. We identify several defects which introduce shallow acceptor levels into the bandgap and can form easily in the material. Among these, shallow Cu vacancies tend to be abundant, especially in technologically relevant Cu-poor material, and can lead to *p*-type self-doping by pinning the Fermi level near the valence band. Cation antisites also form with relative ease. While In_{Cu} and its complexes formed with Cu vacancies are essentially electronically benign in *p*-type material, the Cu_{In} antisite exhibits a deeper electronic state which may act as a detrimental recombination center. Suppressing or compensating

its formation is likely crucial to improving PV device efficiencies.

Given that stable CIS can often deviate significantly from ideal stoichiometry and this can dictate subsequent device performance [224], it is worthwhile to connect this behavior to material synthesis conditions and to the underlying formation of defects. To do so, we formulate an approximate thermodynamic model parameterized by DFT defect formation energies and study the impact of non-stoichiometry on defect formation and phase stability. Our model reproduces some well-know features of the the experimental In_2Se_3-Cu_2Se pseudobinary phase diagram [83, 102, 45]. In particular, we find that the stable CIS easily supports Cu-deficient compositions due to the formation of Cu vacancies and In_{Cu}, and we connect the location of the $CuIn_5Se_8$ secondary phase boundary to the proliferation of fully compensated $2V_{Cu} - In_{Cu}$ complexes. By contrast, the narrower existence region for Cu-rich compositions is found to be due to the substitution of Cu on the In sublattice and the increasing density of Cu interstitials, which destabilize the CIS chalcopyrite phase. Our analysis further confirms the observed self-doping behavior in non-stoichiometric material, where CIS fabricated under Se-rich atmospheres leads to p-type material [224].

In subsequent sections, we utilize this model of CIS defect thermodynamics to understand the diffusion of Cu in the system. The Cu composition is observed to play an important role in the fabrication of high-quality PV absorber layers [102, 224, 45], and Cu is a particularly mobile species in CIS. However, measurements of the Cu diffusivity vary over several orders of magnitude at room temperature [86, 20, 52, 79, 151, 236, 231, 41, 147], and this is somewhat confounding to the interpretation the underlying diffusion mechanisms. By combining *ab initio* transition state calculations with simple diffusion models, we find that we can capture wide variations in the diffusivity related to the underlying material compositions. These models are further supported by kinetic Monte Carlo and continuum reaction-diffusion simulations. In particular, we clarify the roles of Cu vacancy- and interstitial-mediated mechanisms and account for the orders-of-magnitude differences between tracer and chemical diffusion coefficients.

2.1 Point Defects from First Principles

In many physically relevant cases, the concentration $x_{d,q}$ of a point defect d in charge state q is much lower than the concentration of lattice sites. In this limit, the number of distinct ways to arrange $N_{d,q}$ defects on N_s sites is $W \approx (\theta_{d,q} N_s)^{N_{d,q}}/N_{d,q}!$, where $\theta_{d,q}$ is a degeneracy factor accounting for internal degrees of freedom of the defect (e.g., orientation or spin). W corresponds to the number of configurational microstates of the system and thus gives a contribution to the entropy of

$$S^{(\mathrm{conf})} = k_\mathrm{B} \ln W \approx k_\mathrm{B} \left[N_{d,q} - N_{d,q} \ln(N_{d,q}/N_s) + N_{d,q} \ln \theta_{d,q} \right] \tag{2.1}$$

Let $\Delta E^{(f)}_{d,q}$ be the energy cost of creating a single defect. Then, the configurational free energy per atom of a system of $N_{d,q}$ defects at temperature T with fractional concentration $x_{d,q} = N_{d,q}/N_s$ is given by

$$F = x_{d,q} E^{(f)}_{d,q} - T \frac{S^{(\mathrm{conf})}}{N_s} = x_{d,q} E^{(f)}_{d,q} - k_\mathrm{B} T \left[x_{d,q} - x_{d,q} \ln(x_{d,q}) + x_{d,q} \ln \theta_{d,q} \right] \tag{2.2}$$

By solving for the minimum of the free energy, $\arg\min_{x_{d,q}} F(x_{d,q})$, the equilibrium defect concentration in the dilute limit is given by

$$x_{d,q} = \theta_{d,q} e^{-\Delta E^{(f)}_{d,q}/k_\mathrm{B} T} \ . \tag{2.3}$$

A simple expression for the defect formation energy, $\Delta E^{(f)}_{d,q}$, follows by considering the exchange of lattice atoms with external chemical reservoirs,

$$\Delta E^{(f)}_{d,q} = E_{d,q} - E_\mathrm{bulk} + \sum_\alpha m^\alpha_d \mu_\alpha + q(E_\mathrm{VBM} + E_\mathrm{Fermi}) + E_\mathrm{corr} \ . \tag{2.4}$$

Here, $E_{d,q}$ is the total ground state energy of structure containing the defect, created in a pristine reference structure which would otherwise possess an energy of E_bulk. The integers m^α_d count the number of chemical species α either added ($m^\alpha_d < 0$) or removed ($m^\alpha_d > 0$) from the system in forming the defect, and the energy to exchange these atoms with a chemical

reservoir is characterized by the chemical potentials μ_α. It is often convenient to define the chemical potentials, $\mu_\alpha = \mu_\alpha^0 + \Delta\mu_\alpha$, with respect to a set of standard references, μ_α^0. Deviations with respect to these references, $\Delta\mu_\alpha$, capture the synthesis conditions of the material of interest and satisfy, when the material is thermodynamically stable against the spontaneous formation of secondary compounds C, the following constraints:

$$\Delta \mathrm{H}_{\mathrm{bulk}}^{(f)} = \mathrm{E}_{\mathrm{bulk}}^{(\mathrm{u})} - \sum_{\alpha \in \mathrm{bulk}} n_\alpha \mu_\alpha = \sum_{\alpha \in \mathrm{bulk}} n_\alpha \Delta\mu_\alpha \tag{2.5}$$

$$\Delta \mathrm{H}_{\mathrm{C}}^{(f)} > \sum_{\alpha \in \mathrm{C}} n_\alpha \Delta\mu_\alpha \tag{2.6}$$

The first constraint is provided by the formation enthalpy, $\Delta \mathrm{H}_{\mathrm{bulk}}^{(f)}$, of the reference material with composition $(A_{n_A} B_{n_B} C_{n_C} \cdots)$ and ground state energy per formula unit $\mathrm{E}_{\mathrm{bulk}}^{(\mathrm{u})}$, while the second constraint specifies the stability against secondary phase formation. Since we will consider defects that can become charged, an additional term reflecting the electronic reservoir is required. For semiconductors in the thermodynamic limit, this provided by the equilibrium Fermi level, E_{Fermi}, referenced with respect to the valence band maximum (VBM). When considering solids with multiple types of defects, it is natural to choose the VBM corresponding to the bulk reference system. As defects in semiconductor and insulators tend to introduce electronic levels into the band gap or near the band edges, which subsequently influence the electronic behavior of the system, it is useful to identify them. As charge carriers are exchanged with the electronic states of the host material in changing the charge state of the defect, the transition level between defect charge states q_1 and q_2 can be located by

$$\epsilon_d(q_1/q_2) = \frac{\Delta E_{d,q_1}^{(f)}|_{E_{\mathrm{Fermi}}=0} - \Delta E_{d,q_2}^{(f)}|_{E_{\mathrm{Fermi}}=0}}{q_2 - q_1} \,. \tag{2.7}$$

Each of the terms in the defect formation energy (3.11) can be derived from DFT calculations [72]. In particular, when employing DFT in a plane-wave basis, which enforces periodic boundary conditions on the structure, approximating the dilute limit can require fairly large

simulation cells. Practical limitations on computational resources, however, restrict the size of the structures that one can easily calculate, and a defect with nonzero charge state may spuriously interact with its periodic images via a long-range Coulomb potential. Hence, a finite-size correction term E_{corr} is often included to offset these effects.

Various schemes exist for deriving electrostatic energy corrections, which we briefly summarize in the next section.

2.2 Compositionally Constrained Thermodynamics

The defect formation energies (3.11) depend on a set of chemical potentials, $\mu_\alpha = \mu_\alpha^0 + \Delta\mu_\alpha$, whose deviations $\Delta\mu_\alpha$ with respect to a standard reference μ_α^0 characterizes the synthesis conditions of the material. While these chemical potentials can be constrained by requiring stability against secondary phase formation, their absolute values are nonetheless difficult to choose without extensive experimental data. However, net atomic fractions k_α are readily measurable, and given a set of expected defects in the system, one can calculate deviations from perfect stoichiometry [177], k_α^{p}, by

$$k_\alpha = \frac{k_\alpha^{\text{p}} - \sum_{d,q} x_{d,q} m_d^\alpha}{1 - \sum_{d,q} x_{d,q} \sigma_d} \,, \tag{2.8}$$

where $\sigma_d = \sum_\alpha m_d^\alpha$, and m_d^α is the number of atoms of species α removed in the formation of a single defect, d. Note that, by construction, these atomic fractions satisfy the constraint $\sum_\alpha k_\alpha = 1$. Given a set of atomic fractions and standard reference chemical potentials, one can self-consistently obtain defect concentrations, total chemical potentials and the equilibrium Fermi level by solving the inverse problem [177] implicitly defined by (2.8). By imposing compositional constraints, this method, which we subsequently refer to as compositionally constrained thermodynamics (CCT), amounts to modeling the defect thermodynamics in the canonical ensemble. Note that the number of potential defects $x_{d,q}$ often outnumbers the number of constraints that can be imposed on the system. In the following, we detail how a solvable inverse problem can be defined by making explicit use of the Boltzmann-Gibbs

forms of the defect concentrations (2.3).

The Fermi level dependence of the concentration of defect d in charge state q can be factorized via

$$f_d(E_{\text{Fermi}}) \equiv \sum_q \frac{x_{d,q}}{x_{d,0}} = \sum_q \frac{K_{d,q}}{(n/n_i)^q}, \tag{2.9}$$

where $K_{d,q}$ are coefficients for mass-action relations involving defect charging [177], and n is the free electron concentration, which we normalize with respect to the intrinsic carrier concentration n_i, as is standard [7]. Thus, the total defect concentrations, $x_d \coloneqq \sum_q x_{d,q}$, can be written in terms of the neutral defect concentration, $x_{d,0}$, and the free electron concentration, n. Charge neutrality, $p - n + \sum_{d,q} q x_{d,q} = 0$, where p is the free hole concentration, combined with the equilibrium relationship, $pn = n_i^2$, provides an additional constraint to solve self-consistently for n:

$$\left(\frac{n}{n_i}\right)^2 - 1 = \sum_d x_{d,0} \sum_q q \frac{K_{d,q}}{n_i} \left(\frac{n}{n_i}\right)^{1-q}. \tag{2.10}$$

A reduction in the number of independent defect concentrations, $x_{d,0}$, can be made by considering hypothetical transformations between defect species in the absence of particle exchange with external chemical reservoirs [177]. This yields effective mass-action relations between defect species, allowing us to rewrite equations (2.8) and (2.10) in terms of an independent basis set of defect concentrations. Hence, we obtain a coupled and fully determined set of algebraic constraint equations which are polynomial in a set of basis defect concentrations. The basis defect concentrations can be determined using standard root finding algorithms, and to maintain numerical stability, it is useful to choose defects which typically possess low formation energies and therefore occur at higher concentrations. By inverting the mass-action transformations, we can generate the remaining set of charged defect concentrations. Finally, the Fermi level can be found by self-consistent iteration of the set of solutions until charge neutrality (2.10) is satisfied. We summarize this procedure in Algorithm 1.

Of course, there is a one-to-one correspondence between atomic fractions k_α and chemical

Data: DFT defect formation energies and mass action relations

Input : Atomic fraction, k_α, and temperature, T, initial Fermi level, $E^{(0)}_{\text{Fermi}}$

Output: Defect concentrations, chemical potentials, and self-consistent Fermi level

Initialize compositional constraints;
Factor out Fermi level dependence;
Eliminate defect variables using basis-defect mass action relations;
while $n \leq n_{\text{Fermi loops}}$ **do**
  for $E_{\text{Fermi}} = E^{(0)}_{\text{Fermi}}$ **to** E_{gap} **do**
    Solve polynomial constraint equations, $\{k_\alpha = F(\{x_{d_b,0}\})\}$, in terms of basis defect concentrations, $\{x_{d_b,0}\}$;
    Reconstruct $x_{d,q} = f_{d,q}(E_{\text{Fermi}}) \sum_b K_{d,d_b} x_{d_b,0}$;
    if $p - n + \sum_{d,q} q x_{d,q} < 0$ **then**
      $E^{(0)}_{\text{Fermi}} = E_{\text{Fermi}}$;
      break;
    end
  end
end

Algorithm 1: Summary of the CCT algorithm

potentials μ_α, and in determining defect concentrations one is free to set a combination that spans the chemical space of the material of interest. This results in a mixed canonical / grand canonical treatment of the defect thermodynamics, and for every chemical potential set, the dimensionality of compositionally constrained inverse problem is reduced by 1.

2.3 Thermodynamics of Point Defects and Complexes in $CuInSe_2$

In this section, we study the formation of intrinsic defects in $CuInSe_2$ using the DFT and CCT methods discussed above. In particular, we consider an extensive set of intrinsic defects

[201, 232], namely all cation and anion vacancies (V_{Cu}, V_{In}, V_{Se}) and interstitials (Cu_i, In_i, Se_i), cation antisites (In_{Cu}, Cu_{In}) and typically prevalent defect complexes ($V_{Cu} - In_{Cu}$, $2V_{Cu} - In_{Cu}$, $V_{Cu} - V_{Se}$).

Due to the large number of calculations required to build a set of reference formation energies and the computational demand of *ab initio* methods, we use the efficient Perdew-Burke-Ernzerhof (PBE) [197] generalized gradient approximation (GGA) functional rather than a potentially more accurate, but much more computationally intensive, method such as hybrid Heyd-Scuseria-Ernzerhof (HSE) [103]. HSE provides an improved description of the localized $3d$-orbitals of Cu-containing semiconductor compounds compared to local density functionals [201, 72]. We approximate this behavior by applying a +U correction to the Cu d states following the method of Dudarev, *et al.* [61], with a value of U = +5 eV chosen to reproduce the experimental band gap of CIS [200]. Defect formation energies, defect charge transition levels, and compound formation enthalpies were thus calculated using the VASP simulation package with PAW-PBE GGA functionals [123]. Ground state energy calculations were performed using a 64-atom cell and $2\times2\times2$ Monkhorst-Pack k-point mesh [175], and the intra-ionic Hellmann-Feynman forces were relaxed to below 0.01 eV/Å. For defect formation energies and charge transition levels, we included post-processing corrections for spurious electrostatic interactions following the scheme (FNV) proposed by Freysoldt, Neugebauer, and Van de Walle [74]. Figures 2.1a,c show the resulting defect formation energies and transition levels. The formation energies are shown for moderately Se-rich material, marked as point A on the predicted secondary phase diagram (Figure 2.1b), and the Cu, In and Se chemical potentials are measured with respect to the energies of their low-energy pure elemental phases.

Among the vacancies, we find that V_{Cu} has especially low formation energy and introduces a shallow acceptor level near the valence band. Despite a low In chemical potential, V_{In} possesses a relatively high formation energy and exhibits a series of transitions levels, one moderately shallow and two closer to mid-gap. Given the high Se chemical potential, V_{Se} also shows a high formation energy, but can form in moderate quantities for very Se-poor

material. The isolated Se vacancy shows a shallow (+2/0) negative-U type transition [131] associated with large local relaxation in the surrounding group-III cations. As discussed in previous studies [131, 201], the divacancy $V_{Cu} - V_{Se}$ shows a deep $(+1/-1)$ transition often attributed to the observation of light-induced metastabilities in CIS-based solar cells [132]. As we will verify below with CCT calculations, these divacancy complexes can form in modest amounts in n-type material but are otherwise rare in p-type material, at least in equilibrium [201].

Considering the interstitials, we find that only Cu_i can form in relative abundance and that it acts as singly charged shallow donor. Despite its small concentration relative to Cu vacancies, we will find in Section 2.4 that it can play a significant role in Cu diffusion.

Both the cation antisites possess relatively low formation energies in p-type material, suggesting a propensity for the formation of weakly cation-disordered sublattices. While In_{Cu} introduces a shallow donor level near the conduction band, Cu_{In} exhibits both a somewhat shallow acceptor level as well as a deeper $(-1/-2)$ transition. Given the possibly high concentration of Cu_{In}, this $(-1/-2)$ transition may act as a deleterious deep recombination center. Finally, we find favorable binding between Cu vacancies and In_{Cu} antisites, forming either partially or full compensated complexes, $V_{Cu} - In_{Cu}$ and $2V_{Cu} - In_{Cu}$, consistent with previously reported first-principles studies [292, 201, 189]. It has been argued [292, 148] that the coalescence of these vacancy-antisite complexes is responsible for the observed formation of an ordered defect phase [102] in Cu-depleted CIS.

Our calculations of charge transition levels agree reasonably well with the recent experimental picture laid out by Spindler, *et al.* [235]. In particular, through a combination of photoluminescence spectroscopy (PL) and Hall measurements, they identified three distinct acceptor levels, two donor levels, and a single deep defect state. The acceptor levels were subsequently attributed to V_{Cu}, Cu_{In}, and V_{In}. The identification of V_{Cu} and Cu_{In} are consistent with our calculations, although it remains an open question as to whether V_{In} can occur in sufficiently high enough concentrations to be measured. As our CCT calculations show below (Figure 2.3), this may only be possible in material synthesized under strong

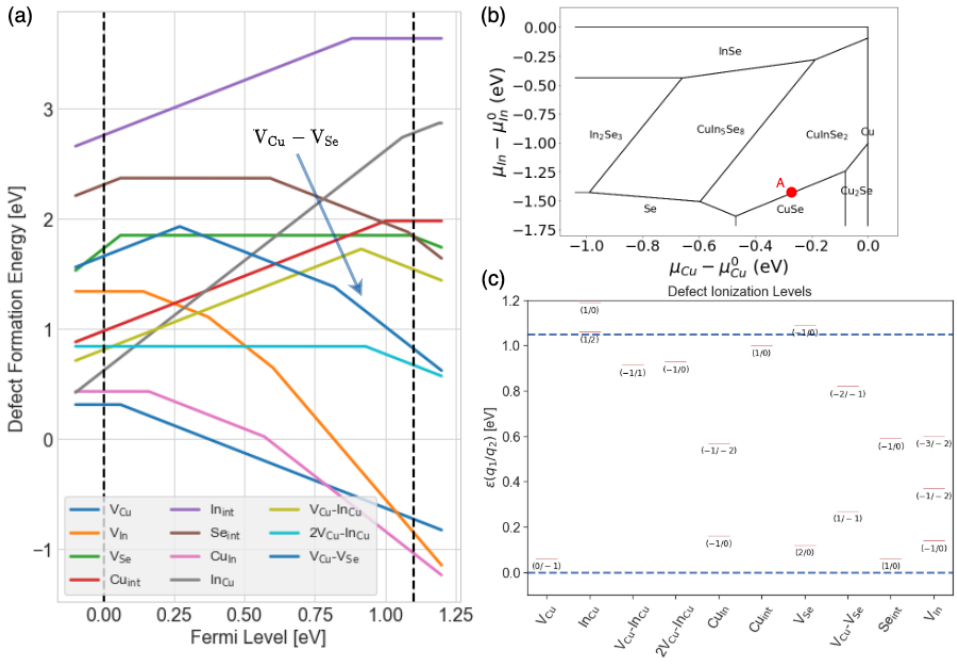


Figure 2.1: (a) Point defect and defect complex formation energies in CIS predicted from DFT using the supercell method, with chemical potentials chosen at point A in the secondary phase diagram (b). Point A reflects moderately Cu-poor, Se-rich synthesis conditions. (c) Corresponding thermodynamic charge transition levels for the intrinsic defects plotted in (a).

Cu- and Se-rich conditions. The existence of a deep (-1/-2) transition level due to Cu_{In} also agrees with their findings. However, their measurement puts this level approximately 0.2 eV closer to the conduction band, although there is some question as to the origin of the broadness of the measured transition. We also find a shallow donor state due to In_{Cu}. While Spindler, *et al.* [235] attribute the presence of a second donor level to $V_{Cu} - V_{Se}$, it is unclear whether they form in detectable quantities, and our calculations put this transition level deeper within the bandgap. Incidentally, this additional donor level may be due to either Cu_i or $nV_{Cu} - In_{Cu}$ complex formation.

The region of thermodynamically (meta)stable CIS spans a broad range of Cu compo-

sitions in the experimental In_2Se_3-Cu_2Se pseudobinary phase diagram [83, 102], extending largely toward the Cu-poor side of ideal stoichiometry and less so into the Cu-rich side, $0.65 \lesssim k_{Cu}/k_{In} \lesssim 1.05$. Furthermore, excess Se supplied during material growth leads to p-type CIS, characterized by a nominal valence ratio $2k_{Se}/(k_{Cu} + 3k_{In}) > 1.0$. Figure 2.2 exhibits this relationship, predicted by CCT calculations at typical synthesis temperature of 700 K, relative to the predicted secondary phase diagram. It is worth noting that the predicted phase boundaries are based on zero-temperature ground state compound formation enthalpies, and therefore variations on the order of k_BT should be expected at finite temperature. This accounts for the degree to which the displayed composition maps extend beyond the ideal, 0K CIS region.

Let us discuss some salient features of Figure 2.2 and how they relate to the underlying formation of intrinsic defects (Figure 2.3). As expected, the largest changes in composition occur near the phase boundaries, signaled by large fluctuations in the relevant particles numbers $\langle(\delta N_\alpha)^2\rangle = k_BT\partial\langle N_\alpha\rangle/\partial\mu_\alpha$. For example, the boundary shared with the Cu-depleted phase $CuIn_5Se_8$ features steep gradients in the cation, k_{Cu}/k_{In}, and anion-cation $k_{Se}/(k_{Cu} + k_{In})$, ratios. This is accompanied by (Figure 2.3) the formation of Cu vacancies, In_{Cu} antisites, and their stabilized complexes $nV_{Cu} - In_{Cu}$ $(n = 1, 2)$. Indeed, upon decreasing Cu composition, CIS is found to become unstable relative to the formation of mixed ordered defect compound (ODC) phases [102], and this often attributed to the aggregation and ordering of large numbers of $nV_{Cu} - In_{Cu}$ complexes [293, 148]. Near the pure Se and Cu_xSe boundaries we find large changes in the nominal valence ratio and the cation ratio. Variations in the self-consistent Fermi level are also shown and found to agree with the model results in [23] and the experimental study in [273], where larger Se chemical potential / partial pressure led to stronger p-type doping. Indeed, approaching the Se phase boundary requires decreasing the Cu chemical potential, leading to a proliferation of Cu vacancies, which act as shallow acceptors. For fixed Se chemical potential, we also find that increasing the [Cu]/[In] ratio decreases the Fermi level [224] due primarily to decreases in the concentration of In_{Cu}, which acts as a shallow donor, and increasing Cu_{In}, which features an acceptor level close to the

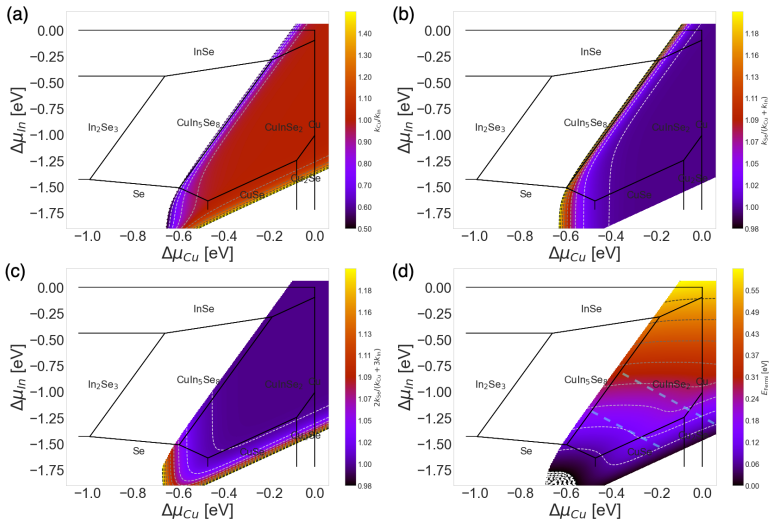


Figure 2.2: Predicted variations of the CIS composition (a-c) and self-consistent Fermi level (d) at 700 K mapped to the secondary phase diagram. Shown are (a) the [Cu]/[In] cation ratio, (b) the [Se]/([Cu]+[In]) anion-cation ratio, and (c) the nominal valence ratio 2[Se]/([Cu]+3[In]). Sparse isocurves (white, dashed curves) are included to highlight the nonlinearity in the mapping between chemical potential and composition space. Dashed, light blue lines in (d) corresponds to levels of constant Se chemical potential. (Reprinted from [234], with the permission of AIP Publishing.)

valence band [235]. While approaching these Cu-rich secondary phases, not only does the In-sublattice become increasingly replaced by Cu, but open interstitial sites become occupied by Cu as well. This saturation of Cu destabilizes the chalcopyrite CIS lattice, allowing for phase separation of Cu_xSe. Indeed, the synthesis of Cu-rich CIS often requires the removal of surface Cu_2Se by wet chemical etching [45].

Cu-poor CIS synthesized under an Se-rich atmosphere has been found to yield high efficiency p-type absorbers for photovoltaic devices [224]. To emulate these conditions, we plot in Figures 2.4a,b the defect concentrations with fixed Se chemical potential as a function of the [Cu]/[In] ratio. The chosen Se chemical potentials correspond to the cuts through the

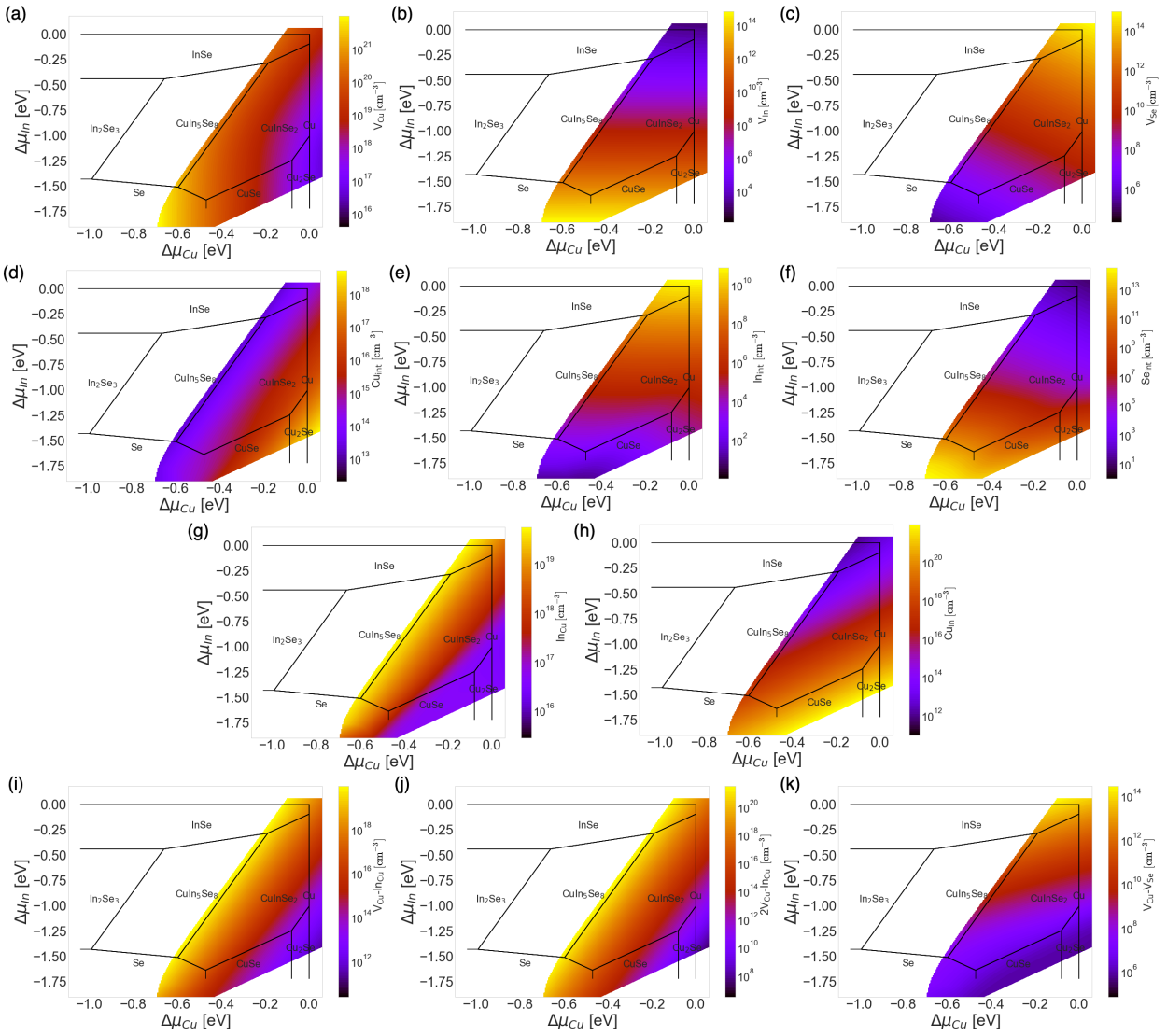


Figure 2.3: Predicted variations in point defect and defect complex concentrations in CIS at 700 K mapped to the secondary phase diagram. Shown are (a-c) vacancies, (d-f) interstitials, (g-h) cation antisites, and (i-k) Cu vacancy complexes.

phase diagram shown in Figure 2.2d, reflecting different degrees of intrinsic p-type doping. In Cu-poor material ($k_{\mathrm{Cu}}/k_{\mathrm{In}} < 1$), $\mathrm{V_{Cu}}$, $\mathrm{In_{Cu}}$, and their complexes are the most abundant defects, while a significant amount of $\mathrm{Cu_{In}}$ is also present. In Cu-rich material ($k_{\mathrm{Cu}}/k_{\mathrm{In}} >$

1), high concentrations of V_{Cu} and In_{Cu} remain, while the concentrations Cu_{In} and, to a lesser extent Cu_i, grow substantially. Indeed, this crossover region is particularly narrow in composition space, while broad in chemical potential space (Figure 2.2a), which implies the existence of large thermodynamic factors (2.15) discussed in Section 2.4. As expected, the concentration of metastable V_{Se} and $V_{Cu} - V_{Se}$ is low at these Se chemical potentials. While Cu-poor compositions reduce the concentration of potential hole traps Cu_{In} [201], they remain present in significant quantities. In the absence of compensated antisite pairs $In_{Cu} - Cu_{In}$, which we have not accounted for, more moderately p-type material synthesized under smaller excess Se shows lower concentrations of Cu_{In}, while maintaining small concentrations of metastable Se vacancies.

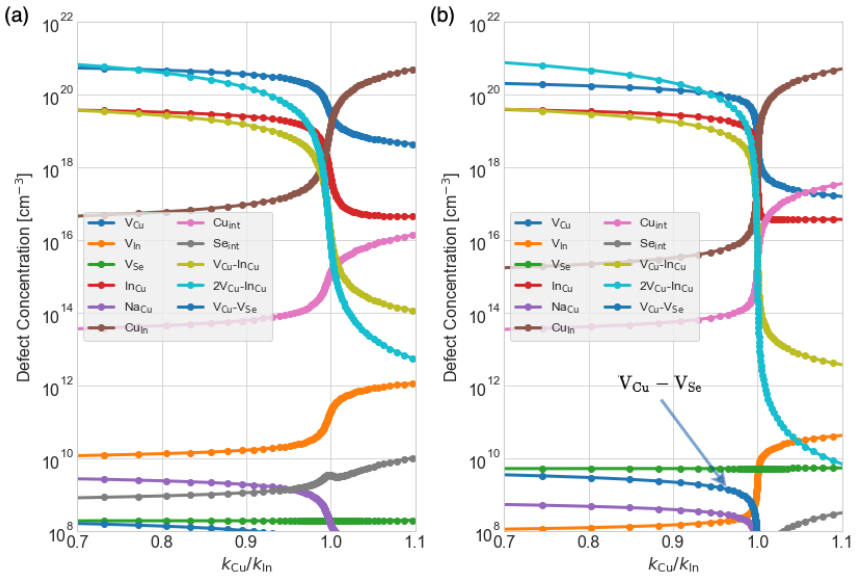


Figure 2.4: [Cu]/[In] composition dependence of point defect and defect complex concentrations in CIS at 700 K along the contours of constant Se chemical potential shown in Figure 2.2d. The fixed Se chemical potentials correspond to different degrees of self-doping, yielding (a) strongly and (b) moderately p-type material.

2.4 Models of Atomic Transport in $CuInSe_2$

The Cu composition is observed to play a crucial role in dictating PV device performance, where it has been found that Cu-poor, Se-rich materials lead to high efficiency p-type absorbers [224]. From an atomistic perspective, Cu-deficient stoichiometries disfavor the formation of Cu_{In} antisites, which can act as deep recombination centers, while promoting the formation of shallow Cu vacancies. The high mobility of Cu vacancies can, in turn, facilitate the diffusion of beneficial alkaline dopants as well as the stabilization of various defect complexes. Indeed, Cu itself is a particularly mobile species and its migration can locally change the self-doping in a CIS absorber layer. This is thought to be essential for p-n junction formation at the CdS-CIS interface and thus on the subsequent quality of the PV device [87, 224, 102].

Previous experimental work has found that Cu diffusion can depend significantly on material composition, with reported [86, 20, 52, 79, 151, 236, 231] values of the Cu diffusivity varying over several orders of magnitude ($10^{-12} - 10^{-8}$ cm^2/s) at elevated temperature ($500 - 700$ K). Recent first-principles calculations [202, 203] support the conclusion [20, 52, 79] that Cu vacancies play a fundamental role in mediating diffusion, while also highlighting the potentially significant contribution of Cu interstitials. In this section, we consider the role of material composition and synthesis conditions in Cu diffusion. Since Cu vacancies are highly abundant in CIS compositions most relevant to PV applications, we first study their impact on Cu diffusion using a combination of thermodynamic calculations and kinetic lattice Monte Carlo simulations parametrized by first-principles calculations [233]. In doing so, we find that while Cu vacancy diffusion accounts for a significant contribution to experimental diffusion measurements using dilute radioactive tracers, it is insufficient to explain much of the wider variations in reported diffusivities.

Thus, in subsequent subsections, we extend this analysis, following ref. [203], by including the contributions of Cu interstitial diffusion and more realistic constraints on the material synthesis conditions. In particular, we model the total Cu diffusivity across the full ther-

modynamic stability range of CIS and numerically quantify the difference between so-called *tracer* and *chemical* diffusion coefficients. In doing so, we disentangle two effects that can contribute to larger measured values of Cu diffusivity. First, in the absence of significant Cu composition gradients and in the limit where the concentration of tracked particles, such as radioactive tracers, is small, the measured Cu diffusivity is largely controlled by the composition. This dictates the relative contributions of vacancy and interstitial mechanisms, and for Cu-rich compositions we find that a sufficient number of rapidly diffusing interstitials dominates the total diffusivity. Conversely, when significant Cu concentration gradients exist, for instance near the interface between CIS phases of disparate compositions or between CIS and some secondary phase, one enters the regime of chemical diffusion. In this case, we find that the inclusion of an extra thermodynamic factor can account for observed diffusivities 2-3 orders of magnitude larger than tracer self-diffusion. This thermodynamic factor arises from the gradient in chemical potential due to the composition gradient that gives an effective drift term that goes beyond standard entropy-driven diffusion. To validate the consistency of this picture, which is fundamentally based on a linear response approximation, we build a minimal continuum model that we then use to simulate a typical experimental Cu diffusion measurement. Strong agreement with available experimental data [79, 147] supports the fundamental role played by local variations in Cu composition in interpreting diffusion measurements.

2.4.1 Methods and Model Selection

The rearrangement of native cations in CIS depends on the availability of atomistic pathways mediating their diffusion. Since the local hopping of atoms in semiconducting crystals tends to be thermally activated, a given hop will depend on a migration energy barrier separating the initial and final configurations [159]. Moreover, the likelihood that such a transition is thermodynamically accessible in the first place is determined by the probability to find the local initial configuration in the system. For defect-mediated diffusion processes, this probability is given by the fractional concentrations of the participating defects and their local

binding energies. The equilibrium concentration of intrinsic defects in complex semiconductor compounds with semi-ionic character such as CIS will generally depend on the underlying composition of the material [232]. Thus, the characterization of diffusion mechanisms in CIS requires a model for the underlying defect concentrations as a function of composition at relevant temperatures.

The contribution of a particular diffusion mechanism mediated by defect d in net charge state q to the total diffusivity, $D^*_{\text{tot}} = \sum_{d,q} D^*_{d,q}$, of a given species, α, can be approximated using the typical Langmuir-Arrhenius form for the tracer diffusion coefficient:

$$D^*_{d,q} = g f \lambda^2 p_{d,q} \nu e^{-\Delta E^{(m)}_{d,q}/k_{\text{B}}T}, \tag{2.11}$$

where λ is the hopping distance, $\Delta E^{(m)}_{d,q}$ is migration energy barrier, f is a correlation factor dependent on the mechanism and the underlying crystalline lattice, and $g = z/2d$ is a geometric factor determined by the nearest-neighbor coordination number z and the spatial dimension d of the diffusion process. The attempt frequency ν characterizes the equilibrium thermal motion around the initial configuration projected onto the reaction coordinates. While typically on the order of the Debye frequency of the lattice, it can be approximated, at the level of quasi-harmonic transition state theory [266, 156], by the normal vibrational modes around the ground state, $\nu^{(G)}_j$, and saddle point, $\nu^{(S)}_j$, configurations,

$$\nu = \frac{\prod_j^{3N} \nu^{(G)}_j}{\prod_j^{3N-1} \nu^{(S)}_j} \ . \tag{2.12}$$

Migration energy barriers were calculated using the climbing-image nudged elastic band (CINEB) method [100, 123, 175] and density functional theory (DFT), while normal vibrational modes were calculated using density functional perturbation theory (DFPT).

Near equilibrium, the probability, $p_{d,q}$, that the defects mediating diffusion in net charge state q are locally available depends on their equilibrium fractional concentrations, $x_{d,q}$. In the dilute limit, these defect concentrations are given by $x_{d,q} = \theta_{d,q} e^{-\Delta E^{(f)}_{d,q}/k_{\text{B}}T}$, where the formation energy, $\Delta E^{(f)}_{d,q}$ can be derived from DFT using the supercell method [72] discussed

above, and $\theta_{d,q}$ is a degeneracy factor accounting for entropic contributions to the defect free energy. In the following sections, defect concentrations have been derived using CCT calculations.

In the absence of concentration gradients, the diffusive motion of individual tracer particles is well-described by the tracer diffusion coefficients (2.11). Formally speaking, these can be derived on the basis of a Green-Kubo formula,

$$D_\alpha^* = \frac{1}{N_\alpha d} \sum_{i=1}^{N_\alpha} \int_0^\infty dt \, \langle \mathbf{v}_i(t) \cdot \mathbf{v}_i(0) \rangle \ , \tag{2.13}$$

where $\langle \cdot \rangle$ denotes an equilibrium ensemble average, and the integrand is the velocity-velocity autocorrelation function over individual particle trajectories [84, 3, 158]. Note that in taking the time integral and assuming time translation invariance, one can recover the conventional random walk formula for the diffusion coefficient in terms of the mean-square displacement $\langle |\mathbf{r}_i|^2 \rangle$. In the case of tracer diffusion, the effects of varying composition on the diffusivity is entirely characterized by the prefactors $p_{d,q}$ in (2.11).

Long-wavelength fluctuations in the particle density, due either to an external perturbation or to local thermal fluctuations, introduce finite concentration gradients into the system. The relaxation of the system back toward equilibrium is then controlled by the collective diffusion of an ensemble of particles, characterized by the chemical diffusion coefficient $D_{\alpha,\text{chem}}$. Again, this transport coefficient can be formally defined in the Green-Kubo linear response formalism [158, 71],

$$D_{\alpha,\text{chem}} = \frac{\Theta_\alpha}{N_\alpha d} \int_0^\infty dt \left\langle \sum_{i=1}^{N_\alpha} \mathbf{v}_i(t) \cdot \sum_{j=1}^{N_\alpha} \mathbf{v}_j(0) \right\rangle \ , \tag{2.14}$$

where the velocity-velocity correlation function captures the time correlation of the total flux of particles of species α. This correlation function can be related to the mean-square displacement of the center of mass coordinate of the ensemble of particles, $\left\langle \left| \sum_{i=1}^{N_\alpha} \mathbf{r}_i \right|^2 \right\rangle / N_\alpha$. The so-called thermodynamic factor,

$$\Theta_{\alpha\alpha} = \left(\frac{\langle (\delta N_\alpha)^2 \rangle}{\langle N_\alpha \rangle} \right)^{-1} = \frac{\partial (\mu_\alpha / k_\mathrm{B} T)}{\partial \ln k_\alpha} \ , \tag{2.15}$$

accounts for the chemical non-ideality of the system [5], and amounts to an additional thermodynamic driving force due to the change in chemical potential with changes in composition. Its presence in (2.14) follows fundamentally from the fluctuation-dissipation theorem under the assumption that local detailed balance is satisfied [84, 3, 71, 158, 71]. We further note that this assumption is valid when Fick's law continues to hold [3], that is, in the linear response regime. In comparing these expressions for tracer (2.13) and chemical (2.14) diffusion coefficients, the terms related to cross-correlations between distinct pairs ($i \neq j$) of particle trajectories in chemical diffusion typically lead to additional factors of $O(1)$ for self-diffusion [84, 5]. This is the case for vacancy-mediated Cu diffusion, as we verify below using KLMC [233]. Neglecting these cross-correlations then leads to the following approximation for chemical diffusion,

$$D_{\alpha,\mathrm{chem}} = \Theta_{\alpha\alpha} D_J \approx \Theta_{\alpha\alpha} D^*_\alpha \ , \tag{2.16}$$

originally due to Reed and Ehrlich [210]. Thus, D^*_α approximately equals $D_{\alpha,\mathrm{chem}}$ when interactions between diffusing particles can be ignored ($\Theta_\alpha = 1$), applicable for instance to the case of radioactive tracers in the dilute limit. The so-called *jump* diffusion coefficient [84], D_J, accounts not only for the mean-squared displacement (MSD), $\langle (\Delta r_i^{(\alpha)})^2 \rangle$, of N_α individual particles of species α, but also the cross-correlation between distinct particle trajectories,

$$D_J = \lim_{t \to \infty} \frac{1}{2dt} \left\langle \frac{1}{N_\alpha} \left(\sum_{i=1}^{N_\alpha} \Delta \mathbf{r}_i^{(\alpha)}(t) \right)^2 \right\rangle \ . \tag{2.17}$$

Here, $\Delta \mathbf{r}_i^{(\alpha)}(t)$ is the net displacement of particle i of type α in time t, and d in the spatial dimension of the corresponding random walk. The origin of these terms can be understood on the basis of the phenomenological treatment of irreversible thermodynamics [5, 159]. In

this framework, one starts with the assumption of a linear response relationship between the flux / current of particles J_α and a set of generalized forces X_α,

$$J_\alpha = \sum_\beta L_{\alpha\beta} X_\beta \ , \tag{2.18}$$

characterized by a set of phenomenological Onsager transport coefficients $L_{\alpha\beta}$. For the solid-state diffusion processes considered in this work, the generalized forces correspond to gradients in the chemical potentials, $X_\alpha = -\nabla\mu_\alpha$, but could also include terms due to electric-field induced drift, $\sim q\vec{E}$, or gradients in temperature, $\sim \nabla \ln(T)$. Let $c_\alpha \coloneqq k_\alpha/\Omega$ denote the concentration of species α, where Ω is the volume per lattice site. Then the relationship between the generalized diffusion matrix, $D_{\alpha\beta}$, and the matrix of Onsager coefficients is determined by the change of basis between chemical potentials μ_α and concentrations c_α in the fluxes (2.18),

$$D_{\alpha\beta} = \sum_\mu L_{\alpha\mu}\Theta_{\mu\beta} \tag{2.19}$$

The matrix of thermodynamic factors, $\Theta_{\alpha\beta} = \partial\mu_\alpha/\partial k_\mathrm{B}T \ln k_\beta$, acts as the Jacobian of this transformation, and the diagonal elements of the diffusion matrix are the self-diffusion coefficients. Green-Kubo relations starting from current-current correlation functions yield the dependence of the transport coefficient on individual particle displacements [84, 5]. For the Onsager coefficients, this gives the Einstein-Smoluchowski relations

$$L_{\alpha\alpha} = \frac{1}{2dVk_\mathrm{B}T}\left(\sum_i \frac{\langle(\Delta\mathbf{r}_i^{(\alpha)})^2\rangle}{t} + \sum_{\substack{i,j \\ i\neq j}} \frac{\langle\Delta\mathbf{r}_i^{(\alpha)}\cdot\Delta\mathbf{r}_j^{(\alpha)}\rangle}{t}\right) \tag{2.20}$$

for self-diffusion, and

$$L_{\alpha\beta} = \frac{1}{2dVk_\mathrm{B}T}\left(\sum_{i,j} \frac{\langle\Delta\mathbf{r}_i^{(\alpha)}\cdot\Delta\mathbf{r}_j^{(\alpha)}\rangle}{t}\right) \tag{2.21}$$

for the diffusion of species α driven by a chemical potential gradient in species β. It is worth noting that two terms in the diagonal Onsager coefficient (2.20) correspond, respectively, to the self-correlation between single particle trajectories and the 2-particle cross-correlation between independent trajectories. Taken together, as in (2.17), these terms capture the stochastic motion of the center of mass of the ensemble of α-type particles. In the dilute limit, both the cross-correlations and thermodynamic factor can be neglected, giving the tracer diffusion cofficient

$$D_\alpha^* = \frac{1}{2dN_\alpha t} \sum_{i=1}^{N_\alpha} \langle (\Delta \mathbf{r}_i^{(\alpha)})^2 \rangle \ . \tag{2.22}$$

Due to the ensemble average, the particle index n in (2.22) can be dropped. Subsequently rewriting the net displacement in terms of the discrete n_J steps in a given time interval, $\Delta \mathbf{r} = \sum_{t=1}^{n_J} \mathbf{r}_t$, and factoring out the uncorrelated part of the tracer diffusivity, $D^* = f D_0^*$, yields the correlation factor

$$f = 1 + \frac{2 \sum_{s=1}^{n_J - 1} \sum_{t=1}^{n_J - s} \langle \mathbf{r}_{t+s} \cdot \mathbf{r}_t \rangle}{\sum_{t=1}^{n_J} \langle r_t^2 \rangle} \ . \tag{2.23}$$

When all hops are the same length λ, the denominator, $\sum_{t=1}^{n_J} \langle r_t^2 \rangle$, can be reexpressed in terms of the average number of hops per unit time, $\Gamma = z p_{d,q} \nu / t$, which gives the uncorrelated part of the diffusivity, $D_0^* = \Gamma \lambda^2 / 2d$. The correlation factor takes on values between 0 and 1, and the extent to which it deviates from unity provides a measure of the correlation between successive displacements in a single particle's trajectory.

One virtue of the Einstein-Smoluchowski relations (2.17, 2.20, 2.21, 2.22) is that they enable the calculation of transport coefficients via numerically sampling the trajectories of diffusing particles. For solid-state diffusion, where dynamic transitions between system configurations are captured by an enumerable set of potential hops of lattice atoms with rates ω_i, the time evolution can be simulated using kinetic Monte Carlo algorithms. The basic procedure is summarized in Algorithm 2, while certain modifications (see e.g., [37]) must be included to accelerate the simulation of systems that possess deep local minima

due, for example, to small energy barriers or otherwise highly separated time scales.

Input: Set of rate constants, $\{\omega_i\}$

while $t < t_{\max}$ **do**
  Determine all possible processes i, occurring n_i times;
  Calculate total rate, $\omega_{\text{tot}} = \sum_i n_i \omega_i$;
  Generate random numbers ρ_1, ρ_2, ρ_3;
  Find process class k s.t. $\sum_{i=1}^{k} n_i \omega_i < \rho_1 \omega_{\text{tot}} < \sum_{i=1}^{k+1} n_i \omega_i$;
  Randomly select process of class k, $\text{int}(n_k \rho_2)$;
  Increment simulation time, $t \to t - \frac{\ln(\rho_3)}{\omega_{\text{tot}}}$;
end

Algorithm 2: N-fold way algorithm (rejection-free kinetic Monte Carlo)

In this way, the simulation explicitly advances the physical time of the system, allowing one to sample its dynamics in a rejection-free way. For our purposes, initial rate constants are determined by the attempt frequencies and activation energies for a given set of processes, which can be subsequently updated over the course of the simulation to account for local interactions between defects and native lattice atoms. In general, MSD and 2-particle correlation statistics are evaluated in terms of the following time-averaging,

$$\overline{\delta_t^2(\Delta t)} = \frac{1}{t - \Delta t} \int_0^{t-\Delta t} dt' \left[\mathbf{r}(t' + \Delta t) - \mathbf{r}(t')\right]^2 \ , \tag{2.24}$$

where the lag time Δt defines the width of the window moved across the time-series $\mathbf{r}(t)$. In order to obtain good statistics, both the lag time and the simulation time t must be sufficiently long, with $\Delta t \ll t$. For ergodic processes [169], this long-time limit recovers the ensemble average,

$$\langle \mathbf{r}^2(\Delta t) \rangle = \lim_{t \to \infty} \overline{\delta_t^2(\Delta t)} \ . \tag{2.25}$$

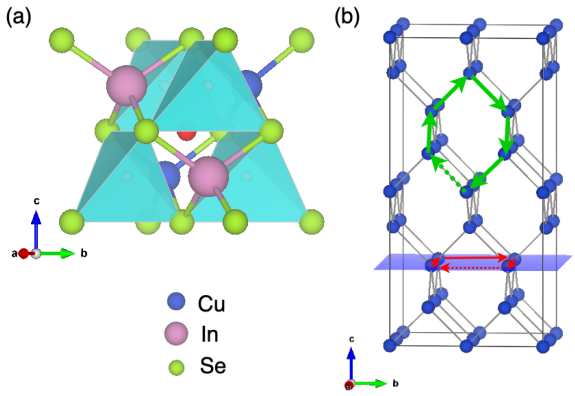


Figure 2.5: (a) Local arrangement of lowest-energy Cu interstitial sites in chalcopyrite CIS (center of blue tetrahedra). Direct Cu interstitial diffusion proceeds by local hopping through the saddle point (red sphere). (b) 1NN (green arrows) and 2NN (red arrows) hops associated with vacancy-mediated diffusion on the Cu sublattice. Closed paths show the minimum number of hops leading to net displacement of a Cu tracer, leading to correlation factors below unity [5, 159]. Note that strictly 2NN vacancy diffusion is restricted to the *ab* planes (blue) normal to the *c* axis. (Reprinted from [234], with the permission of AIP Publishing.)

We have considered the following diffusion pathways:

$$\mathrm{V_{Cu}} \rightarrow \mathrm{V_{Cu}}, \tag{2.26}$$

$$\mathrm{Cu}_i \rightarrow \mathrm{Cu}_i, \tag{2.27}$$

$$\mathrm{Cu_{In}} + \mathrm{V_{In}} \rightarrow \mathrm{V_{In}} + \mathrm{Cu_{In}}, \tag{2.28}$$

$$\mathrm{Cu}_i \rightarrow (\mathrm{Cu} - \mathrm{Cu})_{\mathrm{Cu}} \rightarrow \mathrm{Cu}_i. \tag{2.29}$$

Based on equation (2.11) and CCT calculations, we have determined that Cu diffuses primarily via exchange with Cu vacancies (2.26) and via interstitial sites (2.27). $\mathrm{Cu_{In}}$ antisite diffusion via In vacancies (2.28) can be safely excluded from our model because the probability $(x_{\mathrm{Cu_{In}}} x_{\mathrm{V_{In}}})/(k_{\mathrm{Cu}} k_{\mathrm{In}})$ is irrelevant compared to (2.26) and (2.27) over the compositions

considered, and local binding between the defects is unfavorable due to repulsive Coulomb interactions. For simplicity, we have also excluded the indirect interstitialcy diffusion pathway (2.29). Since this pathway has been found to have a migration barrier similar to direct interstitial diffusion [189, 203], it can be effectively incorporated into our model for direct interstitial diffusion by modifying the local coordination number, z, and the correlation factor[5], f. This amounts to a correction to the diffusivity by a factor of $O(1)$.

Values for the parameters in equation (2.11) for V_{Cu}-mediated and direct interstitial diffusion are listed in Table 2.1. Since V_{Cu} acts as a shallow acceptor, both the neutral and (1-) charge states occur in significant quantities in p-type, Cu-poor material and are furthermore found to have similar migration barriers. Both charge states are therefore included in the vacancy model. We previously found [233] that V_{Cu} hops between second nearest neighbor sites (2NN) on the Cu sublattice possess migration barriers comparable with those of the conventional 1NN processes. Hence, these 2NN processes are also considered. However, it is worth noting that the small anisotropy between c and a, b lattice vectors in the conventional chalcopyrite unit cell of CIS reduces the effective dimensionality of the 2NN processes. While the local 1NN sublattice environment is tetrahedrally coordinated, similar to a zinc-blend lattice, the relative 2NN connectivity is not, likewise, face-center cubic. Instead, 2NN sites are connected in parallel 2D square lattices which form planes along the c axis, and hence strict diffusion via 2NN hopping is quasi-two-dimensional. The correlation factors shown in Table 2.1 reflect this local connectivity [159]. Since Cu interstitials act as shallow donors, and it is difficult to efficiently n-type dope CIS, we consider only the (1+) charge state.

2.4.2 Composition Dependence of Vacancy-Mediated Cu Diffusion

In this section, we develop approximate analytic models for vacancy-mediated Cu diffusion in undoped $CuInSe_2$ as a function of composition. Previously [232], we numerically analyzed the composition dependence of intrinsic defect concentrations based on atomistic data derived from DFT. This section builds on those results by using the composition dependence of the Cu vacancy concentration as a basis for modeling of Cu diffusion. To assess the consistency

Mechanism	$p_{i,q}$	λ	z (d)	f	ν	$\Delta E_{i,q}^{(m)}$ (q)
V_{Cu} (1NN)	$\frac{x_{V_{Cu},q}}{k_{Cu}}$	4.17	4 (3)	0.50	18.8	1.10 (0), 1.12 (1-)
V_{Cu} (2NN)	$\frac{x_{V_{Cu},q}}{k_{Cu}}$	5.88	4 (2)	0.47	34.7	1.20 (0), 1.21 (1-)
Cu_i	$\frac{x_{Cu_i,q}}{k_{Cu}}$	4.18	12 (3)	1.0	1.1	0.26 (1+)

Table 2.1: Parameters (2.11) for vacancy- and interstitial-mediated Cu diffusion in net charge state q. (Reprinted from [234], with the permission of AIP Publishing.)

of our models, we numerically calculate the Cu diffusivity using kinetic lattice Monte Carlo (KLMC) and Onsager transport tensors [134, 251].

An Analytic Model

In modeling the equilibrium concentrations of native point defect as a function of composition in undoped $CuInSe_2$ [232, 177], we have found it useful to parameterize the composition in terms of the cation ratio, [Cu]/[In], and the local valence ratio, 2[Se]/([Cu] + 3[In]). These are respectively denoted

$$\gamma = k_{Cu}/k_{In} \ , \tag{2.30}$$

$$\eta = 2k_{Se}/\left(k_{Cu} + 3k_{In}\right) \ , \tag{2.31}$$

where k_α is the atomic fraction of species α. For compositions of greatest technological interest (Se-rich, p-type material), we found that the high concentration of Cu-vacancies would provide a thermodynamically favored pathway to mediate Cu diffusion [232]. A relatively lower migration barrier for vacancies confined to hop on the Cu sublattice compared to other vacancy transitions on the full cation sublattice supports a simplified model wherein Cu diffusion is restricted to the Cu sublattice [232, 189].

Thus, a simple approximation for the diffusivity of Cu follows by treating diffusion as a correlated random walk involving a single type of vacancy jump of fixed distance λ [5],

giving

$$D^*_{\mathrm{Cu}}(\gamma,\eta) \approx \frac{\lambda^2}{6}\Gamma(\gamma,\eta)\ , \tag{2.32}$$

where z is the lattice coordination number, and the effective jump frequency $\Gamma(\gamma,\eta)$ depends on the composition. For vacancy-mediated diffusion with migration barrier $\Delta E^{(m)}$, the composition dependence is relegated to the fractional concentration of Cu-vacancies, $x_{\mathrm{V_{Cu}}}(\gamma,\eta)$:

$$\Gamma(\gamma,\eta) = \Xi(\gamma,\eta) z f \nu e^{-\beta\Delta E^{(m)}} \frac{x_{\mathrm{V_{Cu}}}(\gamma,\eta)}{k_{\mathrm{Cu}}}\ . \tag{2.33}$$

Here, the fractional concentration $x_{\mathrm{V_{Cu}}}$ is defined as the number of Cu vacancies divided by the total number of lattice sites in defect-free, stoichiometric $CuInSe_2$, and $\beta = 1/k_{\mathrm{B}}T$. The self-diffusion correlation factor, in the limit of dilute vacancy concentration, is denoted by f, and Ξ is a correction that accounts for the dependence of the correlation between successive hops on the vacancy concentration. It is defined explicitly in equation (2.49) and motivated further in the surrounding text. ν is the attempt frequency, and the term $x_{\mathrm{V_{Cu}}}/k_{\mathrm{Cu}}$ is the probability that a Cu site is vacant, where k_{Cu} is the Cu atomic fraction.

To account for the composition-dependence of the Cu-vacancy concentration, we follow a general thermodynamic prescription [177] by solving a coupled series of constraint equations, which are polynomial in a set of basis defect concentrations. This was done numerically in previous work [232], and we use the foresight of those results to reduce the full set of concentration constraints, retaining only the dominant terms. This allows us to obtain simpler analytic expressions.

Generally, the atomic fraction of species α, in the presence of defects d with fractional concentration $x_{d,q}$, can be expressed as in (2.8). Combined with the constraint $\sum_\alpha k_\alpha = 1$, the composition in non-stoichiometric $CuInSe_2$ can be rewritten in terms of the ratios (2.30) and (2.31) as

$$\frac{1-\gamma}{4} = \sum_{d,q} x_{d,q} \left[m_d^{\mathrm{Cu}} - \gamma m_d^{\mathrm{In}} \right] \ , \tag{2.34}$$

$$\eta - 1 = \sum_{d,q} x_{d,q} \left[\eta \left(m_d^{\mathrm{Cu}} + 3m_d^{\mathrm{In}} \right) - 2m_d^{\mathrm{Se}} \right] \ . \tag{2.35}$$

Since $\mathrm{V_{Cu}}$ and $\mathrm{In_{Cu}}$ are dominant defect species for the compositions considered, they are used as the basis defects in the subsequent CCT calculations [232]. After dropping negligible defect concentrations according to the numerical results [232] and retaining only dominant terms in the ensuing constraint polynomials (2.34), (2.35), and (2.10), we can obtain closed-form expressions for $x_{\mathrm{V_{Cu}},0}$, $x_{\mathrm{In_{Cu}},0}$, and n. An analytic approximation for total Cu-vacancy concentration, $x_{\mathrm{V_{Cu}}}$, is then reconstructed using equation (2.9).

For Se-rich p-type material, $\gamma \leq 1$ and $\eta \geq 1$. Qualitatively, either decreasing γ or increasing η tends to drive the material to have a larger equilibrium concentration of vacant Cu sites [232]. Previously [232], we identified two regions in this composition space ($\gamma \leq 1, \eta \geq 1$) with distinct characteristic defects and Fermi level behavior. The boundary between these regions, demarcating the transition between moderately p-type and strongly p-type material, is approximately given by $\gamma = \gamma_c(\eta) \equiv -3.75\eta + 4.75$.

For moderately p-type Se-rich material, $\gamma < \gamma_c(\eta)$, and the dominant defect species, in addition to $\mathrm{V_{Cu}}$, are $\mathrm{In_{Cu}}$, $\mathrm{V_{Cu}}$+$\mathrm{In_{Cu}}$, and $2\mathrm{V_{Cu}}$+$\mathrm{In_{Cu}}$ [232]. The Cu-vacancy concentration in this limit, which we will refer to as model 1, is approximated by

$$x_{\mathrm{V_{Cu}}}^{(1)}(\gamma, \eta) \approx (1 - \eta^{-1} + R_1)(1 + \Lambda) \ , \tag{2.36}$$

where

$$\Lambda = \frac{1 - \frac{\eta}{4}(3+\gamma)(1-8R_1)}{(3+\gamma)(1+R_2)[\eta(1+R_1)-1]} \tag{2.37}$$

is a subleading contribution accounting for Fermi level effects. R_1 and R_2 account for the dissociation of vacancy-antisite complexes at high temperatures in slightly off-stoichiometric material [232]:

$$R_1 = \frac{K_{\mathrm{V_{Cu},-1}} K_{\mathrm{In_{Cu},+2}}}{\kappa_2 K_{\mathrm{V_{Cu}+In_{Cu},+1}}} = \frac{1}{16} e^{\beta E_1^{(\mathrm{b})}} \ , \tag{2.38}$$

and

$$R_2 = \frac{\kappa_1 K_{\mathrm{In_{Cu},+2}}}{\left(\kappa_2 K_{\mathrm{V_{Cu}+In_{Cu},+1}}\right)^2} = \frac{3}{8} e^{\beta E_2^{(\mathrm{b})}} \ . \tag{2.39}$$

The terms $E_{1,2}^{\mathrm{b}}$ in equations (2.38) and (2.39) are binding energies of charged defects:

$$E_1^{(\mathrm{b})} = \Delta E^{(f)}_{(\mathrm{V_{Cu}+In_{Cu}}),+1} - \Delta E^{(f)}_{\mathrm{V_{Cu},-1}} - \Delta E^{(f)}_{\mathrm{In_{Cu},+2}} \ , \tag{2.40}$$

$$E_2^{(\mathrm{b})} = 2\Delta E^{(f)}_{(\mathrm{V_{Cu}+In_{Cu}}),+1} - \Delta E^{(f)}_{(\mathrm{2V_{Cu}+In_{Cu}}),0} - \Delta E^{(f)}_{\mathrm{In_{Cu},+2}} \ . \tag{2.41}$$

The defect formation energies $\Delta E^{(f)}_{d,q}$ for defect d in charge state q can be computed with *ab initio* methods [72], and it is worth noting that the specific combination appearing in $E_{1,2}^{(\mathrm{b})}$ leads to cancellation of chemical potential and Fermi level factors, consistent with our general thermodynamic formalism [177].

The factors κ_i and $K_{d,q}$ defining R_1 and R_2 are coefficients for mass-action relations involving complex formation and defect charging, respectively:

$$\kappa_1 = \frac{x_{\mathrm{2V_{Cu}+In_{Cu},0}}}{\left(x_{\mathrm{V_{Cu},0}}\right)^2 x_{\mathrm{In_{Cu},0}}} \ , \ \kappa_2 = \frac{x_{\mathrm{V_{Cu}+In_{Cu},0}}}{x_{\mathrm{V_{Cu},0}} x_{\mathrm{In_{Cu},0}}} \ , \tag{2.42}$$

$$K_{d,q} = \frac{x_{d,q}}{x_{d,0}} \left(\frac{n}{n_i}\right)^q \ . \tag{2.43}$$

For strongly p-type, Se-rich material, $\gamma > \gamma_c(\eta)$, and along with $\mathrm{V_{Cu}}$, the dominant defect species are $\mathrm{In_{Cu}}$, $\mathrm{V_{In}}$, and $\mathrm{Cu_{In}}$ [232]. As the Fermi level is effectively constant for these compositions [232], the vacancy concentration acquires a much simpler form, which we will refer to as model 2:

$$x^{(2)}_{\mathrm{V_{Cu}}}(\gamma, \eta) \approx \frac{1}{2} - \frac{1 + 2\gamma}{\eta(3 + \gamma)} \ . \tag{2.44}$$

Numerical Benchmarks

In light of the scattered values of diffusion coefficients for Cu measured experimentally, we performed numerical simulations to establish the consistency of the models presented above. When available, we also provide points of experimental data for comparison [20, 79, 236, 245, 231, 151].

In this work, two methods were used to numerically compute the self-diffusion coefficient. First, the mean-squared displacement over time,

$$D^* = \frac{1}{6Nt} \sum_{i=1}^{N} \langle r_i^2 \rangle \ , \tag{2.45}$$

for N mobile Cu atoms was computed using KLMC simulations based on the open source KMCLib code [134]. Each simulation sampled transitions involving both the Cu and In sublattices with migration barriers given in Ref.[189], using cells containing $\sim 10^4$-10^5 cation sites with periodic boundary conditions. Sites were occupied randomly according to the numerical equilibrium concentrations [232], and an initial equilibration of 10^3 steps was run before sampling the mean-squared displacement. Typical runs consisted of $\sim 10^5$ steps, which produced reliably converged values of D^*. Nearest-neighbor binding between Cu-vacancies and $\mathrm{In_{Cu}}$-antisites was included but was not found to significantly affect the results.

Our second method used the open source Onsager code [251] to compute normalized[1] Onsager transport tensors[2], $\tilde{L}_{\alpha\beta}$, from lattice Green's function solutions to the master equation. The tracer diffusivity was calculated via

$$D^*_{\alpha'} = \left\langle \tilde{L}_{\alpha'\alpha'} \right\rangle_{\mathbf{r}} \frac{x_{\mathrm{V}_\alpha}}{k_\alpha} \tag{2.46}$$

for mobile species α and fractional vacancy concentration x_{V_α}. A primed species index (e.g., α') is used to emphasize that this species is a tracer. The brackets $\langle \ldots \rangle_{\mathbf{r}}$ denote a spatial

[1]Briefly, the Onsager terms $\tilde{L}_{\alpha\beta}$ are normalized to have components with units of diffusivity[251], i.e., [area]/[time]. Explicitly, $\tilde{L}_{\alpha\beta} = L_{\alpha\beta} x_\alpha x_\mathrm{V} / (k_\mathrm{B} T \Omega)$, where Ω is the volume per atom, $\beta \in \{\alpha, \mathrm{V}\}$, and $L_{\alpha\beta}$ has typical units [5, 159] of [length]^5[energy]/[time].

[2]In general, for fixed species indices α, β, $\tilde{L}_{\alpha\beta}$ is a 3×3 spatial matrix.

average over the matrix components of $\tilde{L}_{\alpha\beta}$. Use of this method is justified as the vacancy concentration is sufficiently dilute (on the order of ∼1 at. %) for the compositions of interest. We will see below (equation (2.49)) that only a multiplicative correction of order unity is required to account for higher vacancy concentrations.

The vacancy-mediated self-diffusion correlation factor, f, for the Cu sublattice, which is symmetric under the space group $I4_1/amd$ (a subgroup of the $I\bar{4}2d$ space group for the full lattice), is not commonly reported and is difficult to determine analytically. Having at hand numerical methods for simulating tracer diffusion has the added benefit of allowing us to calculate f directly, and motivates us to include an additional factor (2.49) discussed in Section II.A. In KLMC, f can be calculated by

$$f = \frac{1}{N\lambda^2}\sum_{i=1}^{N}\langle r_i^2\rangle = \frac{1+\langle\cos\theta\rangle}{1-\langle\cos\theta\rangle}\ , \tag{2.47}$$

where the first equality holds for a single type of vacancy jump at fixed distance λ, while the second equality assumes that f is dominated by the correlation of consecutive jumps separated by angle θ [5, 159]. Using the Onsager tensors, $\tilde{L}_{\alpha\beta}$, the correlation factor can be computed as

$$f = -\left\langle \tilde{L}_{\alpha'\alpha'}(\tilde{L}_{\alpha'\mathrm{V}})^{-1}\right\rangle_{\mathbf{r}}\ , \tag{2.48}$$

where $\tilde{L}_{\alpha'\mathrm{V}}$ is the off-diagonal contribution of a vacancy chemical potential gradient to the flux of mobile species α [251, 5, 159].

Fixing a representative value of the local valence ratio, η, following Ref. [79], Figure 2.6 compares our model of the diffusivity against numerical calculations limited to 1NN transitions for varying the cation ratio, γ. In this case, the self-diffusion correlation factor was found to be $f = 0.5$. This is not surprising given the local tetrahedral bonding (and thus coordination number $z = 4$) between Cu atoms in the 1NN shell. While the Cu sublattice does not possess the exact symmetry of a diamond lattice, the correlation factors for the two lattices are the same [5]. The local hopping distance is the distance between 1NN Cu sites,

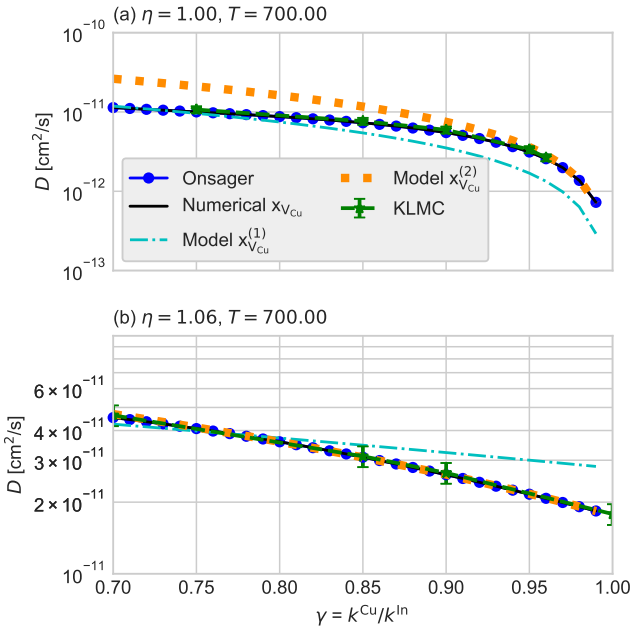


Figure 2.6: Comparison of γ-dependence of analytic and numerical diffusivity models limited to 1nn transitions for (a) $\eta = 1.00$ and (b) $\eta = 1.06$ at temperature 700 K. Note that the diffusivities based on numerical calculation of $x_{\mathrm{V_{Cu}}}$, KLMC, and Onsager tensors overlap for all values of γ. Error bars for KLMC diffusivities are present in (a), but are suppressed by the scale of the plot. The y-axis scales logarithmically in both (a) and (b). (Reprinted from [233], with the permission of AIP Publishing.)

$\lambda \approx 4.17$ Å.

In comparing the tracer diffusivity extracted from the Onsager tensors and the KLMC simulations, we found that our KLMC simulations predicted a diffusivity that was uniformly higher by a factor of order unity. This discrepancy, Ξ, accounts for higher vacancy concentration in off-stoichiometric compositions [178, 243, 242, 5]. Indeed, as calculated (2.47), (2.48), the tracer correlation factors are defined in the limit that the vacancy concentration is very dilute, e.g., where a single vacancy is present on the Cu sublattice. In the opposite limit, where the Cu sublattice contains a single Cu atom, diffusive hopping will be uncor-

related, i.e., $f \to 1$. Thus, for sufficiently defective compositions, the correlation factor acquires a dependence on the vacancy concentration. This dependence is well-approximated by [178, 243, 5]

$$\Xi\left(x_{\mathrm{V_{Cu}}}\right) = \left[f + \frac{\left(1 - x_{\mathrm{V_{Cu}}}/k_{\mathrm{Cu}}\right)\left(1 - f\right)}{\left(1 + x_{\mathrm{V_{Cu}}}/k_{\mathrm{Cu}}\right)}\right]^{-1} , \tag{2.49}$$

where f is the correlation factor in the dilute limit. Hence, the inclusion of Ξf in equation (2.33) accounts for all contributions from correlated hopping in our analytic model. Similarly, replacing $\tilde{L}_{\alpha'\alpha'} \to \Xi\tilde{L}_{\alpha'\alpha'}$ in equation (2.46) improves the agreement between our models and KLMC.

The results of each calculation with appropriate corrections are shown in Figure 2.6. The diffusivities derived from Onsager tensors and KLMC agree within statistical error for all sampled values of the cation ratio. When the numerical results [232] for the vacancy concentration are used in (2.33), we find that our model reproduces the simulated diffusivity within the same statistical error. The general decrease of the self-diffusion coefficient as $\gamma \to 1^-$ is expected, as a higher fraction of Cu sites will be occupied by Cu atoms. Comparing this against (2.32) using the analytic approximations (2.36) and (2.44), we find that model 2 (2.44) matches the numerical results for $\gamma \gtrsim \gamma_c(\eta)$, while agreement with model 1 (2.36) improves for greater deviations from stoichiometry, $\gamma \lesssim \gamma_c(\eta)$. Indeed, the discrepancy between the numerical results and model 1 in Figure 2.6a for $\gamma \to \gamma_c(\eta)^-$ underscores the complexity of the defect thermodynamics in moderately p-type $CuInSe_2$ [232]. Model 1 (2.36) therefore accounts for the leading order behavior for $\gamma < \gamma_c(\eta)$, while higher order corrections are difficult to write down in closed form.

Diffusion Limited to 1NN and 2NN Vacancy Transitions

The migration barrier for transitions between second nearest neighbor (2NN) Cu-sites is only slightly higher than the barrier for 1NN transitions. The inclusion of the 2NN shell introduces four additional sites within the kinetic range of a vacancy. We found that an

effective model for the tracer diffusivity described by the average migration barrier between 1NN and 2NN transitions, as well as the average hopping distance and attempt frequencies, is a good approximation to direct KLMC simulation given that one properly accounts for the effective correlation factor. Accounting for an additional four neighbors at 2NN sites yields an effective coordination number $z = 8$ in (2.32), and from (2.48) the effective tracer correlation factor is found to be $f = 0.71$. This is slightly less than the correlation factor for a bcc lattice ($f_{\text{bcc}} = 0.73$), which is reasonable given the effective coordination number.

Figure 2.7 shows the thermally-activated nature of the tracer diffusivity, as well as the effect of increasing η in the range reported by Gartsman, *et al.* [79], and Becker and Wagner [20]. In this case, γ is fixed according to the values reported by Gartsman, *et al.* [79], and both 1NN and 2NN vacancy have been included. We see that for compositions close to stoichiometry, $\eta = 1$, the predicted diffusivity is close to the values reported by Becker and Wagner [20], who studied stoichiometric samples. Their data is labeled "PRB-83" in Figure 2.7. For these near-stoichiometric compositions, the activation energy is dominated by the migration barrier, but has a correction of ∼20 meV due to the temperature dependence of the Cu vacancy concentration. As the composition deviates further from stoichiometry (increasing η), the diffusivity moves closer to the value reported by Gartsman, *et al.* [79] for non-stoichiometric samples, labeled "JAP-97 (Vol.)" in Figure 2.7. The increase in the diffusivity for increasing $\eta > 1$ is again expected, since structural disorder, and hence the availability of vacant Cu sites, will increase as the material becomes more Se-rich (Figure 2.3a). However, we find that for this model to come within an order of magnitude of the diffusivity reported by Gartsman, *et al.* [79] requires the compositional parameter η to lie outside of a reasonable processing window.

While these tracer diffusion models agree reasonably well with the results in [20, 79], they do not account for other experimental values of Cu diffusivity reported in the literature [52, 151, 236, 231]. Indeed, each of these groups report very rapid Cu diffusion, several orders of magnitude above our model predictions. Lubomirsky, Gartsman, and Cahen have argued [147] that diffusion measurement methods based on transient ion drift [151], potentiostatic

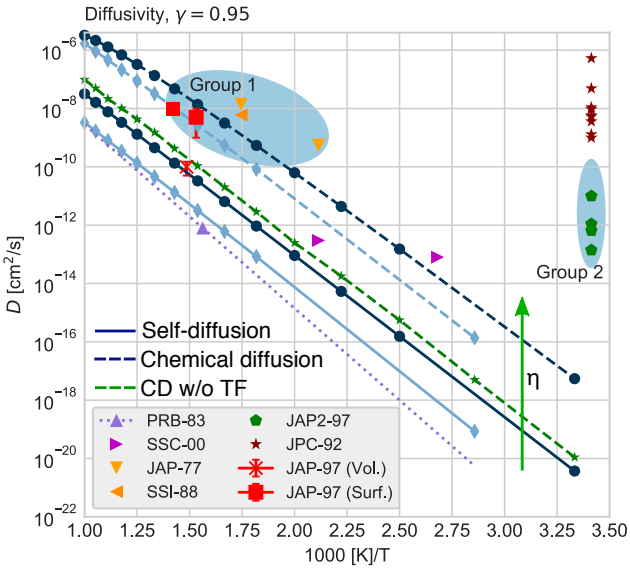


Figure 2.7: Effects of varying local valence ratio η on inverse-temperature dependence of vacancy-mediated self-diffusion models (solid, blue lines) with both 1NN and 2NN transitions. Experimental diffusivities reported in PRB-83 [20], SSC-00 [236], JAP-77 [245], SSI-88 [231], JPC-92 [52], JAP-97 (Vol./Surf.) [79], and JAP2-97 [151] included for comparison. Error bars for JAP-97 (Vol.) estimated based on reported variations in measured composition [79]. "Group 1" and "Group 2" refer to the categorizations given by Lubomirsky, Gartsman, and Cahen [147]. Inclusion of collective center of mass motion (2.17) derived from KLMC leads to small increases in the diffusion coefficient (dashed, green line). Improved agreement with "Group 1" (dashed, blue lines) requires the introduction of both interstitial diffusion and a Cu thermodynamic factor.

current decay [52], and p-n junction motion [231, 245] can exhibit systematic errors due to space charge effects, i.e., internal and external electric fields, leading to overestimation of diffusion coefficients. Gartsman, *et al.* [79], have also tracked in-diffusion from a Cu layer deposited on the sample surface. Their values, labelled "JAP-97 (Surf.)," are plotted in Figure 2.7. Earlier experiments [231, 245] involving p-n junction motion also report similar

values for the Cu diffusion coefficient. These results are summarized in the companion paper [147], where the relevant diffusivities are referred to as "Group 1" in that work. Lubomirsky, Gartsman, and Cahen [147] argue that measuring either multiple samples or a sufficiently large p-n junction depth, as was done in [231, 245], respectively, minimizes the typical errors incurred by space-charge effects. Hence one should expect values for the diffusivity similar to those obtained from diffusion from a constant boundary source [79, 147]. These "Group 1" Cu diffusivities are plotted in Figure 2.7. It is worth noting that Gartsman, *et al.* [79], attribute the approximately two orders of magnitude enhancement of the diffusion coefficient to the presence of a large thermodynamic factor, $\Theta \approx 100$.

Sufficiently Cu-rich compositions possess lower concentrations of Cu vacancies, and thus if diffusion into a Cu-poor bulk material is dominated by a vacancy mechanism, this will give rise to a non-negligible vacancy flux for which we must account. This so-called *vacancy wind* effect [5, 159] should result in enhanced interdiffusion between Cu-rich and Cu-poor regions. However, based on KLMC simulation, we find that simply including additional contributions to the diffusivity arising from the collective center of mass motion (2.17) is insufficient to account for the discrepancies (Figure 2.7). Instead, in the next section, we show that both the inclusion of Cu interstitial diffusion and a large thermodynamic factor are necessary to account for the larger chemical diffusion coefficients. The addition of interstitial diffusion also reconciles the unreasonably large value of the valence ratio otherwise required for strict vacancy diffusion to reproduce the tracer diffusivity measured by Gartsman, *et al.* [79].

2.4.3 Cu Tracer and Chemical Self-Diffusion

We found above that additional terms related to cross-correlations between pairs of distinct particle trajectories [5] amounted to an $O(1)$ rescaling of the self-diffusion coefficient [84]. In contrast, for materials such as CIS which can exhibit large deviations from ideal stoichiometry, the thermodynamic factor provides the dominant correction to tracer diffusion. In modeling the dependence of the chemical potentials on the chemical compositions in the CCT formalism, we can compute thermodynamic factors by simply taking numerical

gradients. In doing so, the derived thermodynamic factors account for the highly nontrivial relationship between the formation of intrinsic defects in CIS and its thermodynamic stability.

Figures 2.8b,c exhibit the variation in the chemical and tracer Cu diffusivities across the stable thermodynamic range of CIS. The data is shown at 700 K, chosen to be within the range of both typical synthesis temperatures and reported temperatures for experimental Cu diffusion measurements [79, 151, 20, 245]. We find that the tracer diffusion coefficient ranges over several orders of magnitude, $10^{-12} - 10^{-7}$ cm^2/s, where the largest values occur for sufficiently Cu-poor and Cu-rich compositions, separated by a local minimum that spans the range of Se chemical potentials. This composition dependence is shown in Figure 2.8a corresponding to a cut across the phase diagram with fixed Se chemical potential, chosen to reflect typical synthesis conditions for p-type material with fixed Se partial pressure. We also show the individual contributions from vacancy- and interstitial-mediated mechanisms to the total diffusivity. As expected, the vacancy mechanisms dominate the diffusion in Cu-poor material, the interstitial mechanism becomes dominant for Cu-rich material, and the crossover between the two regimes results in the local minimum in the total diffusivity. We find that the value at this local minimum agrees with the value measured by Gartsman, *et al.*, [79] for the bulk diffusion of dilute Cu tracers. Furthermore, we find that maximal tracer diffusivities can be achieved either at the Cu_2Se-pure Cu secondary phase boundaries or for sufficiently excess Se chemical potential.

The chemical diffusion coefficient, on the other hand, takes on typical values on the order of $10^{-9} - 10^{-8}$ cm^2/s throughout most of the stable CIS region. The diffusivity reaches its lowest values along the Cu-depleted phase boundary and in the coexistence region with Se and CuSe. The higher chemical diffusion coefficients obtained in the stable CIS region are due to the relatively large changes in chemical potential required to move the material to either Cu-rich or Cu-poor off-stoichiometries. Hence, in the presence of perturbations to the equilibrium Cu compositions, there is a strong thermodynamic driving force back toward equilibrium. For Cu-rich material with an appreciably higher concentration of rapidly

diffusing interstitials, this leads to a significantly larger chemical diffusion coefficient. This agrees with the findings of Gartsman, *et al.*, [79] for the inward diffusion of Cu following the surface deposition of Cu metal. In this case, we find that the total chemical diffusion within the CIS layer is controlled by the local susceptibility of the Cu composition to the chemical potential gradient (Figure 2.8a), and moreover that both interstitial- and vacancy-mediated mechanisms are important. We can rationalize this situation by considering the following microscopic picture. Cu deposited on the surface of moderately Cu-poor, Se-rich CIS yields a Cu-rich interface region that favors the formation of Cu interstitials [102]. The low migration barriers of these interstitials enable their rapid diffusion into the bulk CIS. Far away from this interface region, CIS remains Cu-poor, and the associated high number of Cu vacancies eventually limits the inward diffusion of interstitials via Frenkel pair recombination. In the following section, we validate the fundamental kinetics in this picture by constructing and simulating an approximate continuum model.

The temperature dependence of the total tracer and chemical diffusivities are well-described by an Arrhenius function,

$$D_{\text{tot}}(T) = D_0^{(\text{eff})} e^{-\Delta E_{\text{act}}^{(\text{eff})}/k_{\text{B}}T}, \tag{2.50}$$

where the effective activation energy, $\Delta E_{\text{act}}^{(\text{eff})}$, and diffusion prefactor, $D_0^{(\text{eff})}$, are determined by fits to our numerical data. Figure 2.8d shows the cation ratio dependence of these effective parameters along the same constant Se chemical potential contour considered previously. The effective activation energy for tracer diffusion is within the range of values, 1.25 ± 0.10 eV, reported by Becker and Wagner [20] across the chemical compositions considered. However, it is important to note that while this activation energy is closer to the vacancy-mediated migration barriers, a vacancy mechanism does not completely explain the result. Indeed, as discussed above, diffusion is dominated in Cu-rich CIS by an interstitial mechanism. The crossover between vacancy- and interstitial-dominated regimes is apparent in the onset of nonlinearity in the effective activation energies and prefactors as $k_{\text{Cu}}/k_{\text{In}} \to 1$. As expected, the inclusion of a thermodynamic factor yields Arrhenius fits to the chemical diffusion coef-

ficient that result in effective activation energies lower than their tracer counterparts. This agrees, at least qualitatively, with the lower experimental activation energies, $\approx 0.55 - 0.75$ eV, reported in refs. [41, 147] for chemical diffusion, although it appears that these values were based on averaging across samples with different compositions. Nonetheless, for tracer and chemical diffusion, we find that the effective activation energies are lowered as the material becomes increasingly Cu-rich, signaling faster diffusion with the introduction of more Cu interstitials.

2.4.4 Inward-Diffusion from a Cu-Rich Boundary

To directly study the influence of a nonequilibrium distribution of Cu defects on Cu diffusion, we simulated the in-diffusion of Cu tracers from a layer of Cu metal deposited on the CIS surface and compared the predicted depth profiles to those reported by Gartsman, *et al.* [79]. To that end, we define a simple, continuum reaction-diffusion model for Cu consisting of vacancy- and interstitial-mediated diffusion and Frenkel pair recombination / generation:

$$\begin{aligned}
\partial_t C_{\mathrm{V_{Cu}}} =& \left(D^{(1)}_{\mathrm{V_{Cu}}} + D^{(2)}_{\mathrm{V_{Cu}}} \right) \nabla^2 C_{\mathrm{V_{Cu}}} \\
& - R_{\mathrm{F}} - R_{\mathrm{F}^\dagger}, && (2.51)\\
\partial_t C_{\mathrm{Cu}_i} =& D_{\mathrm{Cu}_i} \nabla^2 C_{\mathrm{Cu}_i} - R_{\mathrm{F}}, && (2.52)\\
\partial_t C_{\mathrm{Cu}_i^\dagger} =& D_{\mathrm{Cu}_i} \nabla^2 C_{\mathrm{Cu}_i^\dagger} - R_{\mathrm{F}^\dagger}, && (2.53)\\
\partial_t C_{\mathrm{Cu}^\dagger_{\mathrm{Cu}}} =& \left(f^{(1)} D^{(1)}_{\mathrm{V_{Cu}}} + f^{(2)} D^{(2)}_{\mathrm{V_{Cu}}} \right) \frac{C_{\mathrm{V_{Cu}}}}{C_{\mathrm{s}}} \\
& \times \nabla^2 C_{\mathrm{Cu}^\dagger_{\mathrm{Cu}}} + R_{\mathrm{F}^\dagger}, && (2.54)
\end{aligned}$$

where a '†' on a chemical component denotes a tracer species, C_i is the per volume concentration of vacancies, interstitials or lattice Cu, and C_{s} is the concentration of Cu sublattice sites. The superscripts (1) and (2) refer to the 1NN and 2NN vacancy-mediated diffusion mechanisms, where $f^{(1)}$ and $f^{(2)}$ are the respective correlation factors. Since the barriers for 1NN and 2NN vacancy-mediated diffusion are nearly equal (see Table 2.1), we treat these

coefficients as independent from the charge state of the vacancy. The Frenkel pair reaction rates R_{F} and $R_{\mathrm{F}^\dagger}$ are defined for non-tracers and tracers, respectively, as

$$R_{\mathrm{F}} = k_f \left(C_{\mathrm{V_{Cu}}} C_{\mathrm{Cu}_i} - C^*_{\mathrm{V_{Cu}}} C^*_{\mathrm{Cu}_i} \right), \tag{2.55}$$

$$R_{\mathrm{F}^\dagger} = k_f \left(C_{\mathrm{V_{Cu}}} C_{\mathrm{Cu}_i^\dagger} - C_{\mathrm{Cu}^\dagger_{\mathrm{Cu}}} \frac{C^*_{\mathrm{V_{Cu}}} C^*_{\mathrm{Cu}_i}}{C^*_{\mathrm{Cu_{Cu}}}} \right), \tag{2.56}$$

$$k_f = 4\pi\lambda_{IV} \left(D^{(1)}_{\mathrm{V_{Cu}}} + D^{(2)}_{\mathrm{V_{Cu}}} + D_{\mathrm{Cu}_i} \right). \tag{2.57}$$

Since the energy barrier for Frenkel pair recombination ($\mathrm{V_{Cu}} + \mathrm{Cu}_i \rightarrow \mathrm{Cu_{Cu}}$) is low relative to the migration barriers, $\Delta E_{IV} = 0.13$ eV, the forward-reaction rate, k_f, is diffusion-limited. The capture cross-section for Frenkel pair recombination is determined by the nearest-neighbor distance ($\lambda_{IV} \approx 2.95$ Å) between interstitial and Cu sublattice sites. Bulk equilibrium concentrations, C^*, were determined by CCT calculations using compositions within the range reported by Gartsman, *et al.* [79] The simulations were carried out using the finite-element solvers in Synopsys's Sentaurus Process.

To model the presence of Cu metal at the CIS surface, we imposed simple Dirichlet boundary conditions, where surface defect equilibrium concentrations were determined by selecting the material composition at which the corresponding chemical potentials were in equilibrium with Cu metal and secondary phase Cu_2Se. The surface concentration of lattice Cu tracers is a free parameter in our model, which we set according to the initial dose estimated from the depth profiles reported by Gartsman, *et al.* [79] We note that, in formulating this reaction-diffusion model (2.54), we have assumed that the depletion of Cu-tracers at the boundary is negligible on the relevant time scales we wish to simulate, which is consistent with the large reported dose of surface Cu [79].

Figure 2.9a shows the simulated diffusion profiles for the in-flow of Cu tracers under a 10 minute anneal at 650 and 700 K. We find excellent agreement between our model and the reported data at 703 K, while our model does not account for the long tail observed experimentally at 653 K. Nonetheless, we find that the rapid in-diffusion of Cu, with a diffu-

sivity two orders of magnitude greater than the reported bulk diffusivity, can be accounted for by the injection of Cu interstitials from secondary Cu-rich phases at the CIS surface. This is further demonstrated in Figure 2.9b, where variations in the boundary Cu interstitial concentration are found to control the subsequent diffusion depth under 10 minute anneal at 700 K.

2.5 Summary

In this chapter, we have characterized the formation of intrinsic point defects in CIS, starting from first principles calculations. Among the myriad of possible defects, we identified Cu vacancies and interstitials as well as cation antisites to be the most significant across a broad range of processing conditions. With the exception of Cu_{In}, these dominant defect species only introduce shallow levels into the bandgap, consistent with the observation that the performance of CIS-based optoelectronic devices typically exhibit a high tolerance to crystalline disorder. By modeling equilibrium defect thermodynamics, we connected this behavior to the phase stability of the material to off-stoichiometric compositions. We demonstrated, in particular, how the Se chemical potential controls the degree of self-doping and that compensated $2V_{Cu} - In_{Cu}$ complexes form in abundance near the Cu-deficient secondary phase boundary.

We then extended this analysis to the study of Cu diffusion, modeling the extent to which differences in material composition can account for the large variations in reported experimental diffusion coefficients. Gartsman, *et al.* [79], Chernyak, *et al.* [41], and Lubomirsky, *et al.* [147] noted a consistent two orders of magnitude difference between the diffusion of Cu measured by tracking bulk Cu tracers versus measurements based on Cu in-diffusion and p-n junction motion. Absent space charge effects, they attributed this difference to the presence of a large thermodynamic factor on the order of $\Theta \approx 100$. The results presented in this work support this argument. Indeed, by employing a model of intrinsic defect thermodynamics in CIS coupled via compositional constraints, we have directly calculated thermodynamic factors of this magnitude in moderately Cu-poor, Se-rich p-type material and explored their

impact on the predicted diffusion coefficients. Our predictions for both tracer and chemical coefficients were shown to be in close agreement with the experimental measurements reported by Gartsman, *et al.* [79], and this further validates the outsized effects that thermodynamic driving forces can have on chemical diffusion in heavily compensated materials like CIS. Moreover, we have related this linear response picture to the underlying correlated random walks executed by individual Cu atoms, hopping either through vacant Cu-sublattice sites or through interstitial pathways. A relatively simple continuum model based on this atomistic picture was able to approximately reproduce the experimental diffusion profiles [79] for the in-flow of Cu into Cu-poor CIS from a Cu-rich interface, demonstrating how Cu interstitials, in particular, can lead to enhanced Cu diffusion. Finally, by mapping out the behavior of the tracer and chemical diffusion coefficients across the CIS secondary phase diagram, we have predicted the extent to which either variations in the composition or the material synthesis conditions can account for the large spread of reported Cu diffusivities measured at elevated temperatures.

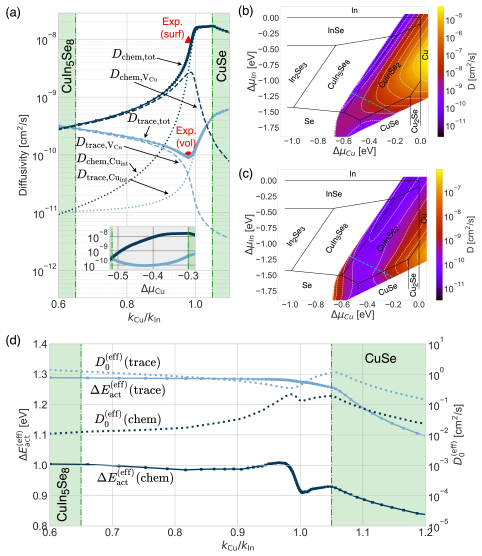


Figure 2.8: (a) Dependence of Cu tracer and chemical diffusion coefficients on the [Cu]/[In] ratio for fixed Se chemical potential at 700 K. The Se chemical potential is chosen to reflect excess Se supply during material synthesis. The diffusivities in (a) are plotted along the cut (blue, dashed line) through the phase diagram, shown in (b-c), and the corresponding locations of the predicted secondary phase boundaries are shown in green. The individual contributions from vacancy- and interstitial-mediated mechanisms to the total diffusivity are shown for both tracer and chemical diffusion. The red data points labeled "Exp. (vol)" and "Exp. (surf)" are the measured diffusivities reported by Gartsman, *et al.*, [79] for radiotracer volume diffusion and inward Cu diffusion following CIS surface deposition of Cu metal, respectively. The [Cu]/[In] positions for these points were determined by the reported stoichiometries of their single crystal CIS samples [79]. The inset to (a) shows the Cu chemical potential dependence of the total tracer and chemical diffusion coefficients along the same cut in (b-c). The full dependence of the chemical (b) and tracer (c) diffusion coefficients on the Cu and In chemical potentials in stable CIS is also shown. Sparse isocurves (white, dashed curves) are included to highlight the nonlinearity of the two diffusivities in the chemical potential space. (d) Effective activation energies and diffusivity prefactors from Arrhenius fits (2.50) to the temperature dependence of the chemical and tracer diffusion coefficients, along the same cut in the phase diagram (b,c) with constant Se chemical potential. (Reprinted from [234], with the permission of AIP Publishing.)

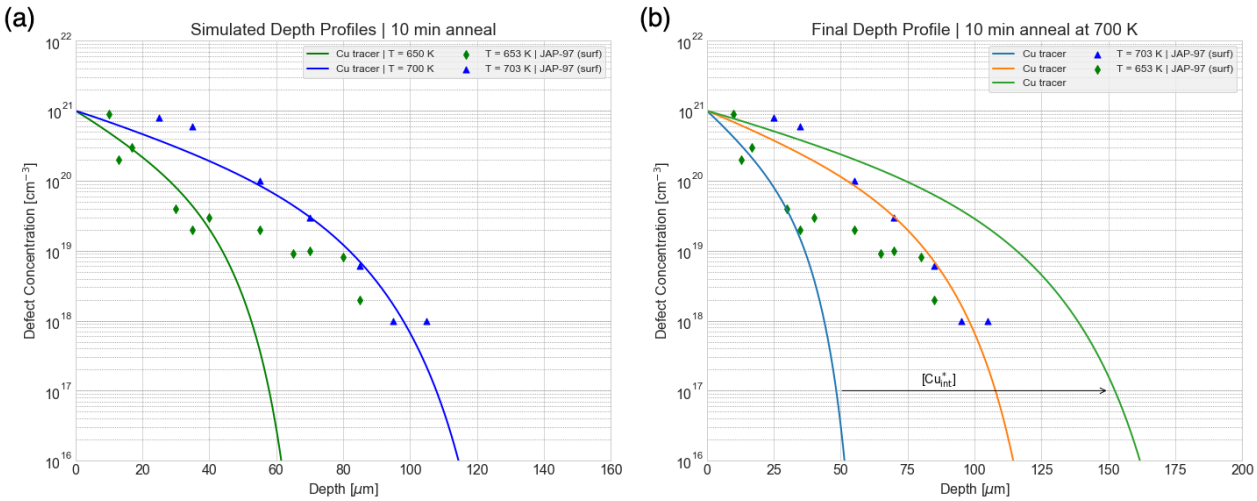


Figure 2.9: Simulated depth profiles for the in-diffusion of Cu tracers from a surface layer of Cu metal into Cu-poor, *p*-type CIS compared to experimental values reported in JAP-97 [79]. (Reprinted from [234], with the permission of AIP Publishing.)

Chapter 3

HALIDE PEROVSKITES: ORDER-DISORDER MODELS AND DOPANT ENGINEERING

Halide perovskites have garnered massive research interest over the past several years due to a combination of favorable optical and electronic properties and high defect tolerance. Nanocrystals (NCs) and thin-films of all-inorganic, metal-halide perovskites $CsPbX_3$ (X = Cl, Br, I), in particular, possess a fortunate combination of broadband absorption, narrow emission bandwidths, high photoluminescence quantum yields (PLQYs), and band-gap tunability across the near-infrared (NIR), visible and ultraviolet (UV) by compositional alloying [207, 181, 2]. Doping by impurity ions has also emerged as a potentially effective route toward engineering additional optical and electronic behavior and controlling perovskite structural stability [17, 56, 152].

In this chapter, we undertake the study of halide exchange reactions and the subsequent formation of ordered alloy phases in $CsPbX_3$, as well as the interactions of certain dopants of high technological relevance with intrinsic disorder. First, we will consider the problem of characterizing the low-energy configurations of mixed halide perovskites, focusing on $CsPb(Br_xI_{1-x})_3$. When the halide composition x is far enough away from the compositional endpoints, $x = 0, 1$, the number of possible configurations that one has to consider grows substantially. It thus becomes computationally demanding to explore the full configuration space using first-principles methods, and one must resort to constructing an effective model. We do this by way of the so-called *cluster expansion* (CE) method, outlined in Section 3.1, which amounts to a decomposition of the total energy of a given alloy configuration in terms of generalized Ising-type interactions. We parameterize these CE models by fitting the interaction parameters to a relatively small set of DFT calculations. Having a sufficiently

accurate model of the energetics then allows us to efficiently sample the configuration space of the alloy using Markov chain Monte Carlo (MCMC) methods. In doing so we characterize the phase stability of halide mixing at finite temperature, as well as the thermodynamics of halide exchange with a given chemical reservoir. With these modeling efforts, in Section 3.2 we make contact with experimental data on halide exchange in $CsPbX_3$ nanocrystals mediated by an idealized precursor molecule [49] and discuss the potential role of persistent long-range ordering in bulk, single crystals.

Building on this methodology, we then consider the case of dopant interactions in mixed halide perovskites. In particular, in Section 3.3 we study the case of B-site doping by Mn in $CsPb(Cl_xBr_{1-x})_3$. Using Metropolis-Hastings MCMC to sample a CE model in the canonical ensemble, we observe room-temperature phase separation into Mn- and Cl-rich domains with increasing Br concentrations, largely consistent with recent experimental evidence [225]. We find that this behavior at intermediate compositions can be rationalized by the difference between Mn-Cl and Mn-Br bond energies: as Cl becomes diluted under ion exchange with Br, the remaining Cl in the system tends to saturate the local Mn environment. Surprisingly, the formation of local Mn-rich phases persists in the pure halide compositional endpoints, suggesting that the formation of Mn clusters is further driven by spinodal decomposition. Based on an analysis of short-range Warren-Cowley type order parameters and partial Bader charges, we argue that this is consistent with a simple model for Coulomb energy gain in the B-site sublattice.

Finally, in Section 3.4 we return to the simpler case of a single halide perovskite $CsPbCl_3$ and study the formation of defect clusters upon doping with Yb. Indeed, exceptionally high experimental photoluminescence quantum yields attributed to highly efficient quantum cutting have recently been observed in ytterbium-doped inorganic metal-halide perovskites such as Yb:$CsPb(X{=}Cl,Br)_3$. Combined with strong, tunable, broadband absorption in the visible spectrum, these materials show great promise for applications in solar downconverter technologies. Much subsequent work has been dedicated to uncovering the fundamental mechanisms behind Yb-mediated quantum cutting, and an accumulation of experimental

evidence has shown that a particular speciation of Yb, believed to be a fully compensated $(2Yb_{Pb} + V_{Pb})^0$ defect complex, dominates this process. Thus, we investigate Yb defect formation in single-crystal $CsPbCl_3$, and in particular the feasibility of forming $(2Yb_{Pb} + V_{Pb})^0$ defect complexes, using first-principles electronic structure calculations. A simple thermodynamic model based on defect formation energies, binding energies, and charge transition levels provides some insight into the distribution of Yb_{Pb} substitutionals and Pb vacancies, underscoring the range of material compositions and synthesis conditions over which locally bound configurations of $(2Yb_{Pb} + V_{Pb})^0$ defect complexes become significant. We complement this analysis with additional calculations of structural and electronic properties and discuss the consistency of these results with the observed onset of quantum cutting.

3.1 Modeling Large Configuration Spaces in Multicomponent Semiconductors

Modeling multicomponent solids in which two or more species can mix on a given sublattice presents a fundamental challenge to first-principles methods. For example, the macroscopic properties of the material of interest often depend on the typical low-energy configurations of the alloyed components, and identifying them requires searching through a combinatorial configuration space that grows exponentially with the number of active lattice sites. Even while the number of possible configurations can be reduced by equivalence under the symmetry operations of the lattice space group, the number of *ab initio* calculations required to sufficiently sample this configuration space is still at best computationally inefficient, and at worst infeasible. An alternative route is to construct a so-called *surrogate model*, which retains the accuracy of first-principles calculations but requires fewer computational resources. The construction of a surrogate model is based fundamentally on the idea of coarse graining the original *ab initio* degrees of freedom, such as the electron density and relaxed ionic positions, using fitted coefficients to implicitly account for their dependence on the alloy configuration space. This procedure fits within the broader class of supervised learning problems, where parameters of a given model ansatz are optimized to reproduce a target property based on a collection of training examples.

The cluster expansion (CE) method [215] is a well-established technique for reducing the computational complexity of this problem by efficiently encoding the chemical degrees of freedom on a fixed lattice [213, 214, 183, 182, 135, 218, 208, 262, 43, 259, 255, 257, 258, 256]. It consists of expanding some computable material property (for example, formation energies or band gaps) in a basis of local many-body interactions. When the relevant on-site degrees of freedom are discrete, as is the case for the chemical occupancy of a lattice site, a CE can be viewed as a generalized Ising model Hamiltonian. An illustration of lower-order clusters in a CE basis is shown in Figure 3.1. It consists of a systematic enumeration of symmetrically equivalent pairs, triplets, quadruplets, and higher-order clusters of lattice sites. This expansion amounts to a discrete, lattice Fourier transform [214] and is, in principle, exact when all possible clusters are included. In many cases, the physically relevant interactions between chemical species are sufficiently local, and one can, in practice, truncate both the spatial range and many-body order of the interaction terms. This leads to a computationally efficient model which maps the atomic structure to the property of interest.

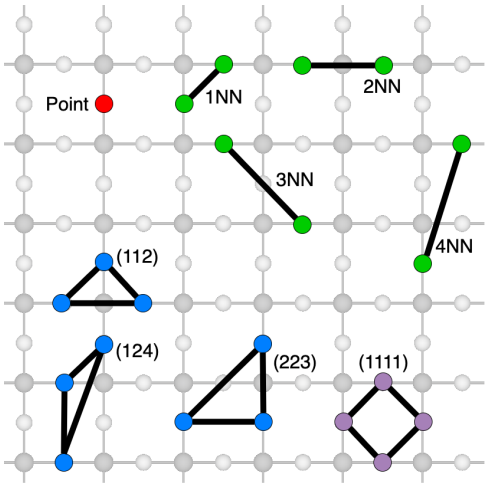


Figure 3.1: Example of symmetrically distinct multisite clusters. Clusters are restricted to a sublattice consisting of edge-centered sites relative to a 2D square lattice. (Figure adapted from [149])

The configuration of various chemical components on a lattice of N sites can be represented by a vector $\sigma = (\sigma_1, \sigma_2, \ldots, \sigma_N)$, where if M_i chemical components can occupy lattice site i, then admissible values of the site occupation variable σ_i are $\sigma_i \in \{0, 1, \ldots M_i - 1\}$. Similarly, a cluster of sites can be represented by a vector $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_N)$ that spans all sites in the lattice, where $\alpha_i = 0$ if site i is not in the cluster, and $\alpha_i \in \{1, \ldots, M_i - 1\}$ otherwise. Then, orthogonal, single-site functions that faithfully represent the occupation of lattice sites in a cluster can be defined in the basis of Chebyshev polynomials [256, 214]:

$$\phi_{\alpha_i, M_i}(\sigma_i) = \begin{cases} 1 & \text{if } \alpha_i = 0 \\ -\cos\left(2\pi \frac{\sigma_i}{M_i} \lceil \frac{\alpha_i}{2} \rceil\right) & \text{if } \alpha_i > 0 \text{ and odd} \\ -\sin\left(2\pi \frac{\sigma_i}{M_i} \lceil \frac{\alpha_i}{2} \rceil\right) & \text{if } \alpha_i > 0 \text{ and even} \end{cases}$$

Here, $\lceil \ldots \rceil$ denotes the ceiling function, and for the binary case ($M_i = 2$), the single site functions reduce to Ising pseudospins, ± 1. Let $E(\sigma)$ denote the material property of interest. In the following, we take $E(\sigma)$ to represent the ground-state energy of the given alloy configuration σ, but a CE can, in principle, be applied to any computable scalar property. A cluster expansion of $E(\sigma)$ can be defined in terms of a complete, orthonormal basis of cluster functions as

$$E(\sigma) = E_0 + \sum_{\alpha} g_\alpha J_\alpha \overline{\Gamma}_\alpha(\sigma) \ , \tag{3.1}$$

where the cluster function

$$\Gamma_\alpha(\sigma) = \prod_i \phi_{\alpha_i, M_i}(\sigma_i) \ , \tag{3.2}$$

is a product of single-site functions in cluster α, and the overline denotes an average over clusters equivalent under the symmetry operations of the (sub)lattice space group:

$$\overline{\Gamma}_\alpha(\sigma) = \frac{1}{N g_\alpha} \sum_{\alpha' \in \mathrm{Orb}(\alpha)} \Gamma_{\alpha'} \ . \tag{3.3}$$

Hence, the sum in (3.1) is carried out over symmetrically distinct clusters, and each representative cluster α carries a multiplicity factor $g_\alpha \coloneqq |\mathrm{Orb}(\alpha)|$ defined as the number of clusters in its group orbit, $\mathrm{Orb}(\alpha)$. The coefficients of the expansion J_α are commonly called effective cluster interactions (ECIs) and capture the coarse-grained energetics of the system.

Being linear in the cluster function basis, the CE in (3.1) defines a linear regression problem for determining the ECI, $Y = XJ$, where Y_i is energy the ith structure in a given set, J_j is the ECI of the jth cluster, and the ij element of the design matrix, X_{ij}, is the jth cluster function of structure i. The optimal ECI, $\hat{J}$, for a DFT surrogate model are determined by the general least-squares solution,

$$\hat{J} = \underset{J \in \mathbb{R}^n}{\arg\min} \, \|XJ - Y\|_2^2 + f_r(J) \ , \tag{3.4}$$

where the energies Y for a training set of structures are calculated from first-principles. The ℓ_p-norm is defined by $\|A\|_p \coloneqq \left(\sum_i |A_i|^p\right)^{1/p}$, and regularization function $f_r(J)$ can be used to reduce overfitting by constraining the values of the ECI [98]. We will follow the compressed sensing paradigm [183, 182, 135] and choose a regularization function $f_r(J) = \lambda\|J\|_1$ corresponding to the least absolute shrinkage and selection operator (LASSO). The hyperparameter λ acts a penalty on the ℓ_1-norm of the expansion coefficients, enforcing sparsity in the number of basis terms with nonzero ECI. The choice of a sparse regularization scheme allows the most relevant clusters to be selected based on the available training data and has the added benefit of yielding a simpler model that is faster to compute.

To construct a reliable surrogate model, the structures included in the training set should sufficiently span the configuration space and be varied enough to capture the relevant physics of the system. Assuming periodic boundary conditions, each unique configuration can be viewed as a specific labeling of a superlattice S derived from a fundamental parent lattice L, where S is a discrete subgroup of L. Hence, a given configuration of chemical species constitutes a labeling of the finite quotient group $G = L/S$. A canonical construction of superlattices and their labelings entails the enumeration of Hermite and Smith normal forms relative to the parent lattice [95]. This construction forms the basis of the Hart-

Forcade algorithms [95, 96, 97], which provide a systematic way of generating candidate configurations for an *ab initio* training data set. Indeed, since the CE amounts to a local decomposition of the lattice degrees of freedom, labeled structures included in the training set need not possess the same unit cell. They must, however, be derived from the same parent lattice.

Computational constraints limit the number of first-principles calculations that can be practically carried out, while favoring structures with smaller unit cells. Having generated a pool of candidate structures up to a certain volume, it becomes imperative to select a representative subset to calculate. While various structure selection methods have been proposed in the literature [183, 182, 218, 257], we have found in practice that a good initial subset of structures for subsequent DFT calculations can be chosen by minimizing the condition number of the design matrix, $\kappa(X)$. We do this by applying simulated annealing to a large pool of candidate structures, accepting a swap of structures between the selected subset and the larger candidate pool if the change in condition number satisfies the Metropolis-Hastings criterion, $p = \min\left(1, e^{-\beta\Delta\kappa(X)}\right)$, where β is a fictitious inverse temperature varied according to an annealing schedule.

In the development of an accurate energy model, it is particularly important to include minimal energy configurations in model training. It is frequently the case that lower energy structures possess smaller unit cells, reflecting a strong degree of long-range ordering under periodic boundary conditions. Since smaller structures incur smaller computational costs, they are often retained in either a training or validation set. Once an initial CE model is fitted, it can be stochastically sampled, either by simulated annealing or (grand-)canonical Monte Carlo, for additional low energy configurations. For small enough simulation cells, these predicted low-energy structures can be saved and their energies subsequently validated with DFT. This active learning cycle, where trial models are used to update the training data, can be repeated until no new feasible ground states are found. Once a final CE model has been determined, thermodynamic observables, including expected values of short-range and long-range order parameters, can be calculated using MCMC methods. We discuss these

methods below in the study of alloyed halide perovskites. A summary of this model-building procedure is shown in Figure 3.2.

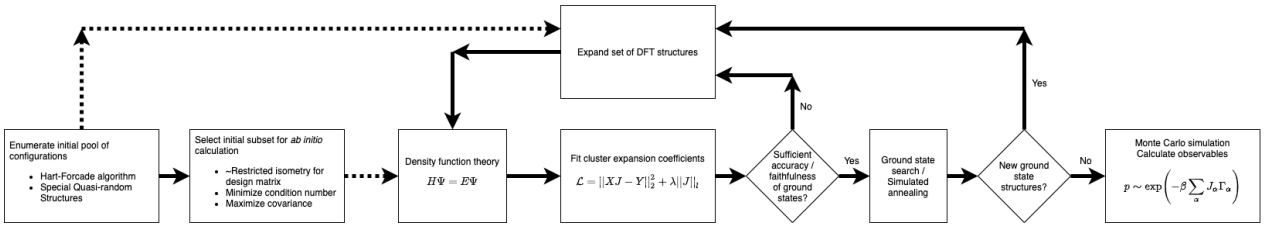


Figure 3.2: Computational workflow for training a cluster expansion (CE) model from first-principles calculations and subsequently predicting thermodynamic observables.

3.2 Thermodynamics of Halide Exchange

Post-synthetic ion exchange has recently emerged as an effective method for tuning the composition of halide perovskites while preserving their lattice structure [113]. This is particularly important for the tuning of optical properties in $CsPbX_3$ (X = Cl, Br, I), which can be otherwise unstable against structural transformations to optically inert allotropes. The high mobility of halide anions, in particular, allows for the facile engineering of mixed halide perovskites spanning the full compositional range [207, 181, 2]. In this section, we study the ordering thermodynamics of mixed halide perovskites, focusing specifically on $CsPb(Br_xI_{1-x})_3$. To do so, we proceed as described above, building a cluster expansion model of the alloy energetics and using it to run MCMC simulations.

To model the energetics of halide mixing, we fit a CE to the DFT mixing / formation energy defined with respect to the compositional endpoints ($x = 0, 1$), $CsPbI_3$ and $CsPbBr_3$:

$$\Delta H_{\text{mix}} = E\left[\text{CsPb}(\text{Br}_x\text{I}_{1-x})_3\right] - xE\left[\text{CsPbBr}_3\right] - (1-x)E\left[\text{CsPbI}_3\right] \tag{3.5}$$

Here, $E[c]$ is the fully relaxed (both electronic and ionic degrees of freedom) ground state DFT energy per halide for the given compound c. The mixing energy characterizes the (zero-

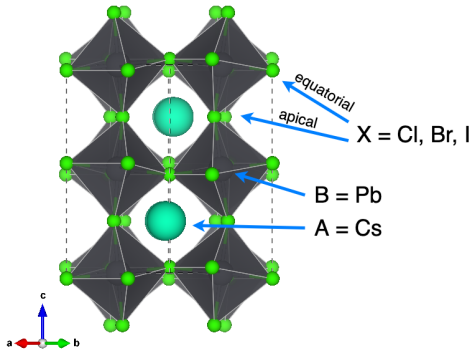


Figure 3.3: The orthorhombic *Pnma* perovskite phase, with general chemical formula ABX_3. The undersized ionic radius of A = Cs within the cuboctahedral void leads to an instability toward Pb-X octahedral tilting. This octahedral tilting creates to symmetrically distinct apical and equatorial halide (X) sites.

temperature) stability of the alloy configuration relative to the pure halide phases. ECIs were determined from DFT calculations of the mixing energies of 150 symmetrically distinct $CsPb(Br,I)_3$ structures in the low-energy, orthorhombic *Pnma* perovskite phase (sometimes called the γ-phase; see Figure 3.3). We used LASSO regression, tuning the ℓ_1 regularization hyperparameter by minimizing the mean square error (MSE) cross-validation (CV) score over 10 training-validation folds [98]. Our final fitted CE was found to have a root-mean square error (RMSE) CV score of ~ 10 meV/atom. The parity plot of CE-predicted energies versus DFT calculations is shown in Figure 3.4.

Figure 3.5 shows the convex hull defined by the mixing energy. We find that our CE model accurately reproduces the DFT data across the entire compositional range. We note, in particular, the existence of low-energy structures at $x = 1/3$ and $x = 1/6$. These correspond to the configurations shown in Figure 3.6, which possess long-ranged layered ordering along the [001] axis. For $x = 1/3$, the minority Br species occupy "apical" halide sites, which in the orthogonal *Pnma* phase are symmetrically distinct from the "equatorial" sites occupied by the majority I (see also Figure 3.3). Similar layered ordering was found by Yin, *et al.*

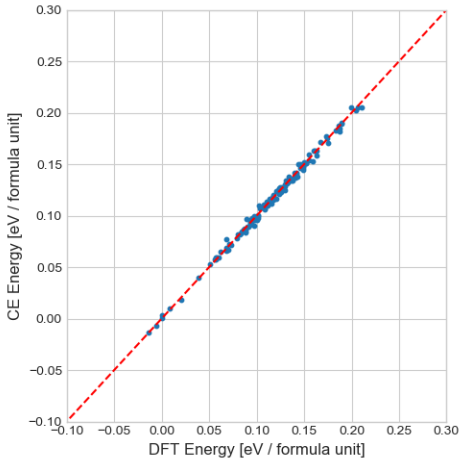


Figure 3.4: Comparison of the predicted formation energies of various configurations of halide binaries $CsPb(I_{1-x}Br_x)_3$ between DFT and a cluster expansion (CE) surrogate model.

[284], in the tetragonal ($I4/mcm$) β-phase. Its stability was rationalized on the grounds of the exchange of partial charges between halides, where the minimum energy configuration for fixed composition $x = 1/3$ maximizes the number of nearest I neighbors to each Br [284, 154]. This form of locally alternating chemical ordering still applies in the orthorhombic γ-phase. We have found, moreover, that these layered configurations continue to lie along the convex hull in the high-temperature cubic ($Pm\overline{3}m$) α-phase. The reduced symmetry defined by the octahedral tilting in the γ-phase thus introduces a uniaxial bias to the direction of the ordering. The selection of the [001] direction, in particular, is likely due to the minimization of lattice strain. Lattice strain also dictates the asymmetry in the mixing energy with respect to the composition, where it is easier to mix the smaller halide (Br) into the material with larger lattice constant ($CsPbI_3$) rather than the reverse.

Semi-grand canonical MCMC at 300 K shows that this asymmetry persists in the expectation value of the mixing energy (Figure 3.5). Following [18], we define a normalized long-range order (LRO) parameter $\zeta/\zeta_{\max}$, where η measures the difference in I occupation between equatorial (e) and apical (a) sites,

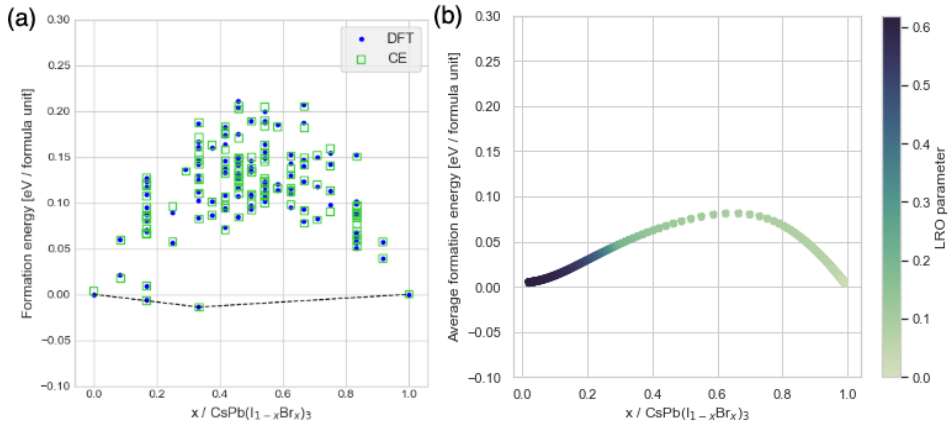
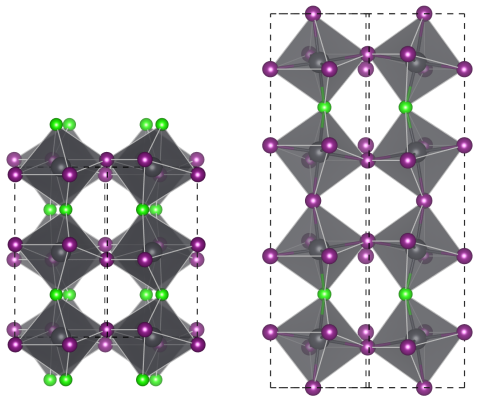


Figure 3.5: (a) The energies of various binary halide configurations in $CsPb(I_{1-x}Br_x)_3$ predicted by DFT and a cluster expansion (CE) surrogate model. The convex hull is marked by a dashed, black line, and the low-energy configurations at $x = 1/3$ and $x = 1/6$ are shown in Figure 3.6. The asymmetry in the 0 K DFT ground state formation energies with respect to the halide composition x is mainained (b) in their room-temperature ensemble average, calculated from the CE model and semi-grand canonical Metropolis Monte Carlo. The halide layering that occupies the convex hull in (a) and characterized by the long-range order (LRO) parameter, $\zeta/\zeta_{\max}$, persists at room temperature.

$$\zeta = \frac{1 + \langle \phi_e \rangle}{2} - \frac{1 + \langle \phi_a \rangle}{2} = \frac{1}{2} \left(\langle \phi_e \rangle - \langle \phi_a \rangle \right) \ , \tag{3.6}$$

and the brackets $\langle \cdots \rangle$ denote a thermodynamic average. Note that, in this case, our single site basis functions ϕ_i reduce to Ising pseudospin variables, where $\phi_i = 1$ for site i occupied by I and $\phi_i = -1$ if occupied by Br. For $x = 1/3$, the parameter $\zeta \leq 1$ is saturated at unity in the completely ordered configuration shown in Figure 3.6. Since the maximum value of η depends on the overall halide composition, we normalize it by its maximum possible value, $\zeta_{\max}$, for every x. We thus find in Figure 3.5 that some degree of layered ordering persists at 300 K, and that the degree of ordering is strongly correlated with the asymmetry in the average mixing energy.



(a) (b)

Figure 3.6: Low-energy ordered binary halide configurations identified by DFT for (a) $x = 1/3$ and (b) $x = 1/6$ in $CsPb(I_{1-x}Br_x)_3$. In the orthorhombic *Pnma* perovskite phase, Pb-X octahedral tilting leads to symmetrically distinct apical and equatorial halide (X) sites, which for $x = 1/3$ are occupied by Br (green) and I (purple) respectively.

The use of common reagents such as alkylammonium halides and surfactant-containing metal halide salt solutions for post-synthetic anion exchange in $CsPbX_3$ is often beset by the underconversion of halide composition and the the formation of secondary metal halide phases [252, 11]. To circumvent some of these challenges, recent experimental work [49] has demonstrated the effectiveness of trimethylsilyl halide (TMSX, X = Cl, Br, I) reagents in mediating anion exchange. Among key advantages to TMSX for anion exchange are favorable bond-dissociation chemistry enabling near-stoichiometric incorporation of heavier halides into lighter halide perovskites and an effective absence of detrimental side reactions. Their favorable thermodynamics, essentially that of an ideal solution, also allows for a simplified theoretical treatment. To that end, we employ our CE model of halide mixing and semi-grand canonical MCMC simulation to study post-synthetic halide exchange mediated by TMSBr and TMSI. In particular, we examine the potential implications of persistent halide ordering in bulk material and compare our results to experimental data on converting nanocrystals

(NCs) from $CsPbBr_3$ to $CsPbI_3$ [49].

Figure 3.7 shows the predicted halide exchange curve at 300 K as a function of fractional TMSI concentration. To alleviate finite-size effects arising from periodic boundary conditions, our data was generated by MCMC in $20 \times 20 \times 20$ supercells of the fundamental 20-atom *Pnma* perovskite unit cell. To model exchange with external chemical reservoirs of Br and I, simulations were conducted over a fine grid of chemical potential differences, $\Delta\mu = \mu_{\mathrm{I}} - \mu_{\mathrm{Br}}$, and single-site flips of the halide identity were accepted with probability p determined by the Metropolis criterion $p = \min\left(1, e^{-\beta(\Delta E + \sum_\alpha \mu_\alpha \Delta N_\alpha)}\right)$. Since the number of sites in the simulation are conserved, this amounts to a treatment of the halide alloy thermodynamics in the semi-grand canonical ensemble. Statistics were gathered over a minimum of 1000 sweeps through the anion sublattice, until the expectation value of the equilibrium halide composition was converged to a precision of 10^{-3} within a 95% confidence interval. Mapping the halide chemical potential to the fractional reagent concentration was accomplished by setting the chemical potential references of the system to those of pure TMSX and assuming an ideal solution, $\mu_\alpha - \mu_\alpha^0 = k_{\mathrm{B}} T \ln(x_\alpha)$.

For Br compositions $x > 0.5$, we observe (Figure 3.7) decent agreement between our model prediction and experimental anion exchange data in 15 nm NCs. The larger discrepancy with data from 6 nm NCs may be due to the increasing relevance of surface effects with larger surface-to-volume ratio. Beyond $x = 0.5$, our model predicts more rapid Br $\rightarrow$ I exchange with increasing reagent concentration. Indeed, this appears to be driven by the formation of lower-energy alloy configurations with increased halide layering, characterized by the $\zeta/\zeta_{\mathrm{max}}$ order parameter, as the Br composition becomes increasingly diluted. The disagreement between our model predictions in bulk material and the experimental NC data again points to the potentially non-trivial role that the NC surface plays during anion exchange. Although we do not do account for it here, it would be interesting to determine how the energetics of halide ordering becomes modified in the presence of finite boundaries. One possibility may follow from the observation [207, 106] that the NC surface can stabilize the cubic perovskite phase at room temperature. While we have found similar low-energy

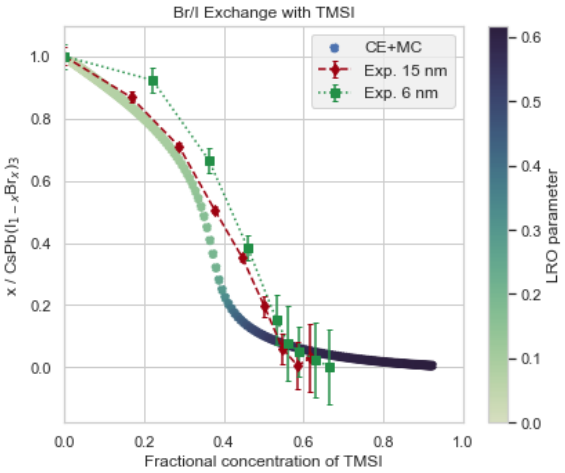


Figure 3.7: Relative Br fraction x as a function of the fractional TMSI reagent concentration at room temperature predicted from a cluster expansion model (CE) and Monte Carlo (MC) simulation, compared to available experimental halide exchange data in $CsPb(I_{1-x}Br_x)_3$ nanocrystals [49]. Facile removal of Br for $x < 0.5$ in bulk crystals is associated with the presence of energetically favorable long-range halide layering, characterized by the long-range order (LRO) parameter, $\zeta/\zeta_{\max}$.

layered configurations in cubic $CsPbX_3$, their DFT mixing energies are not as low as in the orthorhombic *Pnma* phase. This suggests that the degree of layered ordering persisting at room temperature may be lower than as predicted above. Alternatively, halide exchange is fundamentally a nonequilibrium process, and one may reasonably expect [106] a significant activation energy associated with expanding the lattice in incorporating additional I anions. Nonetheless, we find that our model predicts that the exchange reaction can be driven to completion, consistent, at least qualitatively, with experimental observations in both NCs and polycrystalline thin-films of $CsPbX_3$ [49, 207, 106].

Given the importance of local halide morphology on the optical properties of mixed halide perovskites [107, 16, 229, 27, 106], it is worth further scrutinizing the finite-temperature miscibility of mixed-halide solid solutions. Indeed, as the temperature is lowered below 200 K in

our semi-grand canonical MCMC simulations, we observed a discontinuity in the exchange curve, signifying the presence of a miscibility gap (i.e., a demixing transition). Subsequent low-temperature canonical MCMC simulations at fixed composition ($x = 0.5$) verified the existence of phase separation between a relatively pure Br phase and a mixed, I-rich halide phase. While standard Metropolis MCMC combined with thermodynamic integration [128] can, in principle, be used to locate the critical temperature associated with this transition from 2-phase coexistence to a solid solution phase, high rejection probabilities, particularly near the critical point, lead to extended simulation times and poorer quality statistics. This is due, in large part, to the presence of competing many-body interactions in our CE Hamiltonian and concomitant short-range ordering effects. To alleviate these issues, we instead used Wang-Landau (WL) sampling [269, 268, 129, 295, 24], which amounts to a random walk in energy state space, to estimate the microcanoncical density of states, $g(E)$, of the system. Once the density of states is known, the partition function follows easily, $Z = \sum_E g(E)e^{-\beta E}$, along with the thermodynamic observables, $\langle O \rangle = \sum_E Og(E)e^{-\beta E}/Z$, that can be calculated from it. The critical temperature can be determined by finding a local maximum in the specific heat,

$$C(T) = \frac{\langle E^2 \rangle - \langle E \rangle^2}{k_{\mathrm{B}} T^2} \ , \tag{3.7}$$

or a local minimum in the fourth-order Binder cumulant,

$$V_4 = 1 - \frac{\langle E^4 \rangle}{3\langle E^2 \rangle^2} \ , \tag{3.8}$$

subject to finite-size scaling corrections [128]. Figures 3.8b,c show the temperature dependence of the specific heat and the Binder cumulant predicted by WL Monte Carlo using $4 \times 4 \times 4$ supercells of the 20-atom orthorhombic perovskite primitive cell. The sharp divergences observed in these quantities indicate that the cell is large enough to suppress finite-size effects. We find an associated critical temperature at $T_c = 173$ K, in line with previous theoretical efforts [284, 27, 18]. Importantly, our model predicts that Br and I anions remain

miscible well below room temperature, consistent with the experimentally observed ability for halide exchange to effectively interpolate the band gap between $CsPbBr_3$ and $CsPbI_3$.

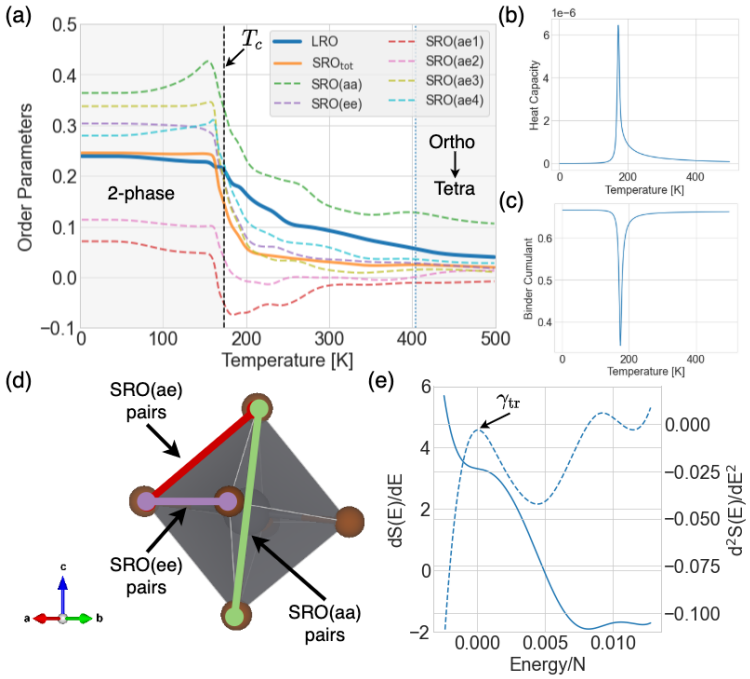


Figure 3.8: (a) Long-range (LRO) and short-range order (SRO) parameters from Wang-Landau Monte Carlo. Lattice site pairs defining the various SRO parameters with respect to the Pb-halide octahedron are shown in (d). Divergences in (b) heat capacity and (c) Binder cumulant of the energy function indicate the presence of a phase transition at the critical temperature $T_c = 173$ K. Phase separation occurs below T_c (2-phase), and the approximate temperature at which a structural phase transition between orthogonal and tetragonal perovskite phases occurs is also marked in (a). The presence of a phase transition T_c can also be inferred from the inflection point γ_{tr} in the microcanonical (Boltzmann) entropy [217].

To characterize halide ordering at finite temperatures, we compute the expected values of the LRO parameter, ζ/ζ_{max}, as well as Warren-Cowley short-range order (SRO) parameters [55],

$$\eta_{ij} = \frac{\langle\overline{\phi_i\phi_j}\rangle - \langle\overline{\phi_i}\rangle\langle\overline{\phi_j}\rangle}{1 - \langle\overline{\phi_i}\rangle\langle\overline{\phi_j}\rangle} \ , \tag{3.9}$$

defined in terms of 2-point correlation functions, $\overline{\Gamma}_{ij}$, from the CE basis (3.2). Note that η_{ij} is normalized such that $-1 \leq \eta_{ij} \leq 1$, and for fixed pairwise distance d_{ij}, its sign indicates whether the alloyed species prefer short-range AA, BB clustering ($\eta_{ij} > 0$) or mixed AB ordering ($\eta_{ij} < 0$). For random pairwise correlations, the SRO parameter vanishes identically ($\eta_{ij} = 0$). Figure 3.8a shows the LRO parameter and a subset of SRO parameters averaged across the WL MC trajectory. We again confirm that the LRO parameter, characterizing whether I atoms preferentially segregate to equatorial sites, decays slowly as function of temperature above the miscibility gap. We also compare SRO parameters within a 3 Å cutoff for which the pairs of sites ij are located either both on apical sites (aa) or equatorial sites (ee), or otherwise connect between these distinct halide sublattices (ae) (see Figure 3.8d). We observe positive and persistent apical-apical correlations, indicative of the tendency for identical halide species to occupy these local pairs of sites. If SRO at first nearest-neighbor (1NN) halide sites is dictated by the quasi-1D halide layering, as in Figure 3.6a, we should also expect negative apical-equatorial SRO parameters. We find, however, that only 1 out the 4 symmetrically distinct 1NN apical-equatorial bonds exhibits this behavior, while another quickly reaches the disordered limit with increasing temperature. The remaining 2 apical-equatorial bonds instead exhibit a tendency towards local clustering. Taken together, this suggests that some degree of local frustration is present around the Pb-X octahedra, complicating the mean-field picture presented in [284].

3.3 Spinodal Decomposition of Mn Dopants During Halide Exchange

Beyond post-synthetic halide exchange, the introduction of certain impurities, such as divalent transition metals or trivalent lanthanides, can provide additional levers for tuning the photophysical properties of metal halide perovskites. For instance, nanocrystals of Mn^{2+}-doped $CsPb(Cl_{1-x}Br_x)_3$, where Mn substitutes for the B-site Pb cations, have shown promise

in lighting and energy-conversion applications via simultaneous manipulation of exciton and Mn^{2+} emission [144, 39, 164, 270]. Recent experimental work [225] used electron paramagnetic resonance (EPR) and photoluminescence (PL) spectroscopies to track the evolution of Mn^{2+} magnetic ordering as well Mn^{2+}- and excitonic PL intensities over the course of Cl $\rightarrow$ Br anion exchange. Surprisingly, while Mn^{2+}-PL was maintained over the full range halide composition, EPR signals disappeared for increasing Br incorporation. Subsequent analysis of Mn^{2+} hyperfine splittings revealed that essentially all EPR-active Mn^{2+} remained exclusively Cl-coordinated during anion exchange. These experimental data are summarized in Figure 3.9 [225].

To explain this behavior, it was hypothesized [225] that the formation of nearest-neighbor $Mn_{Pb} - Mn_{Pb}$ antiferromagnetic dimers, which would be EPR silent yet still PL active, becomes increasingly likely as the Cl concentration is diluted during anion exchange. Indeed, this reasoning is bolstered by the large difference in Mn-Cl and Mn-Br bond dissociation energies and the high mobility of halides and Mn in the perovskite lattice, leading to the proposed mechanism illustrated in Figure 3.10. In this picture, as Br replaces Cl in the system, it becomes energetically favorable for Mn to saturate the remaining Cl bonds. As the Pb-sublattice is approximately simple cubic, with halide sites sitting on the sublattice edges (Figure 3.10b), this leads to the formation, on average, of $Mn_{Pb} - Mn_{Pb}$ dimers with Cl-occupied edges, interacting via antiferromagnetic superexchange [88, 122].

To explore the thermodynamics of Mn clustering, we modeled atom distributions in Mn^{2+}:$CsPb(Cl_{1-x}Br_x)_3$ using a CE energy model and Metropolis-Hastings MCMC simulations in the canonical ensemble. ECIs were determined from a training set of DFT ground state energies consisting of 200 distinct atomic configurations. These were used in canonical MCMC simulations to sample equilibrium alloy configurations at room temperature for various Pb/Mn and Cl/Br compositions. Figure 3.11 shows local snapshots of the equilibrium distributions of Mn_{Pb}, Cl, and Br ions across the full halide composition series. For mixed halide compositions ($0 < x < 1$), our model clearly exhibits increasing Mn clustering as $x \rightarrow 1$. For $x = 4/5$, we observe that Mn clustering is accompanied by Cl clustering.

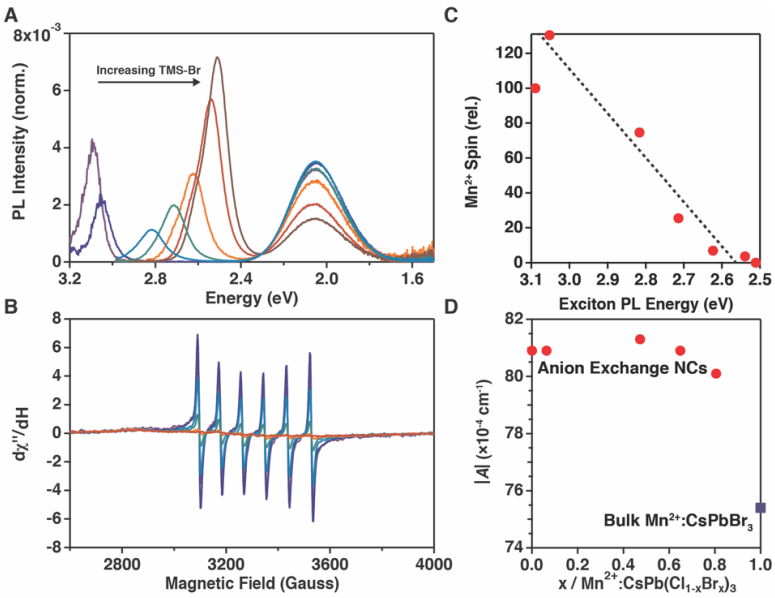


Figure 3.9: (A) PL spectra of 1.1% Mn^{2+}:$CsPbCl_3$ NCs in the EPR tube during the course of an anion exchange, note that there is always Mn^{2+} PL seen centered at 610 nm at every point in the anion exchange. The PL spectra are each normalized to their total integrated PL intensity. Each spectrum was taken at the same NC concentration, and the NCs were never removed from the EPR tube over the entire experiment. (B) X-band EPR spectra collected during the course of the same anion exchange reaction as described in panel A. (C) Scatter plot of the relative Mn^{2+} spin, determined from double integration of the EPR spectra, vs the exciton PL peak energy from (A). The dashed line is a guide to the eye. (D) Summary of hyperfine splitting constants, $|A|$ for Mn^{2+}:$CsPb(Cl_{1-x}Br_x)_3$ NCs during anion exchange, taken from the spectra collected in (B) and for Mn^{2+}:$CsPbBr_3$ bulk powder. Bromide concentrations are estimated from the known dependence of the excitonic PL energy on x. (Reprinted (adapted) with permission from [225]. Copyright 2019 American Chemical Society.)

Surprisingly, at the pure halide endpoints, we also observe a strong degree of Mn-rich phase separation. We proceed with a more thorough analysis of these observations below.

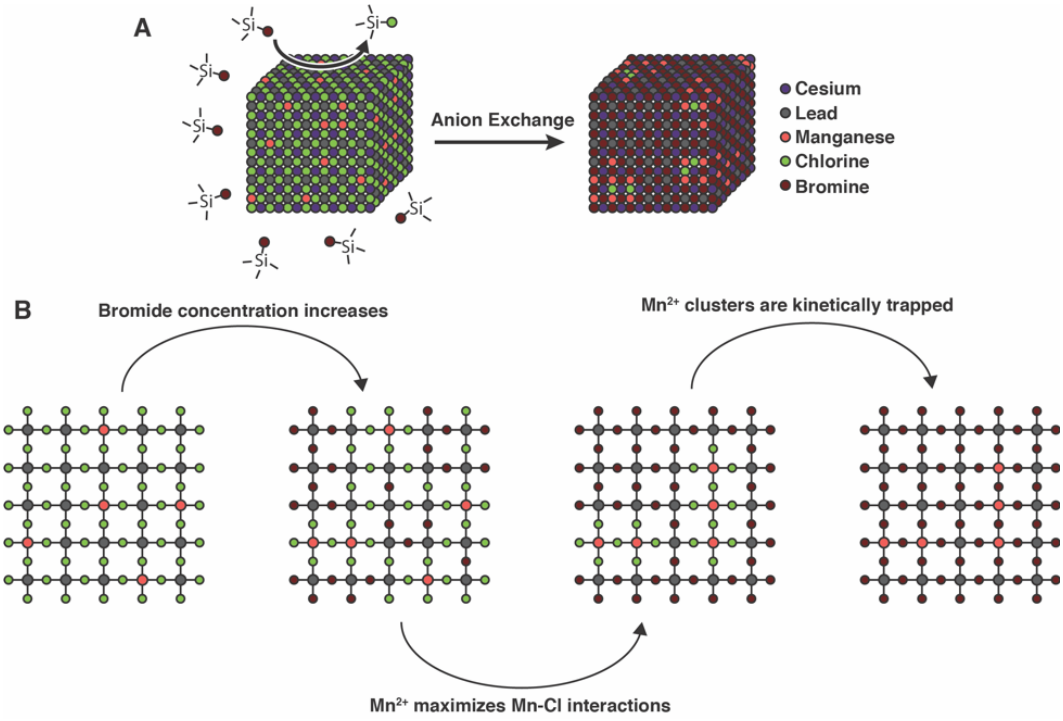


Figure 3.10: Proposed manganese clustering mechanism during anion exchange in Mn^{2+}:$CsPbCl_3$ NCs. (Reprinted (adapted) with permission from [225]. Copyright 2019 American Chemical Society.)

We quantify clustering predicted by the MCMC simulations in terms of Warren-Cowley-type short-range order (SRO) parameters,

$$\eta_i^{(\alpha|\beta)} = \frac{\langle p_i(\alpha|\beta)\rangle - c_\alpha}{1 - c_\alpha} \tag{3.10}$$

where $p_i(\alpha|\beta)$ is the conditional probability that any α species has a β species at a lattice distance $i \in \{1\text{NN}, 2\text{NN}, 3\text{NN}, \dots\}$, and iNN is the ith nearest neighbor shell. Because we do not consider anti-site formation in this work, in the case where $\alpha = \beta$, iNN denotes the ith nearest neighbor shell on the α sublattice. The fractional composition of the alloy species α is given by c_α, and $\langle \dots \rangle$ denotes a thermodynamic average with respect to the canonical ensemble. The SRO parameters measure the degree of pairwise α-β clustering. For nonzero

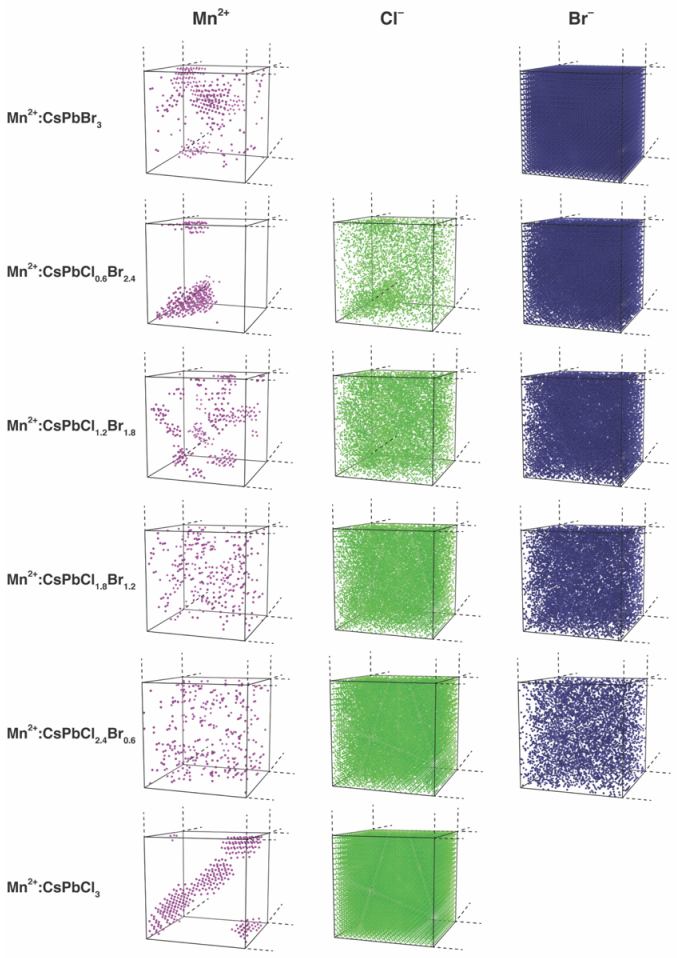


Figure 3.11: Room-temperature equilibrium atomic distributions in $CsPb(Cl_{1-x}Br_x)_3$ with 4% Mn^{2+} doping predicted by canonical MCMC simulation. Local distributions for Mn_{Pb}, Cl, and Br ions are shown at specific halide concentrations. (Reprinted (adapted) with permission from [225]. Copyright 2019 American Chemical Society.)

concentrations of species α and β, the SRO parameters vanish identically for a completely disordered state, which in principle occurs only in the limit of infinite temperature. The lower bound for the SRO parameter is given by $-c_\alpha/(1-c_\alpha)$. A value $0 < \eta_i^{(\alpha|\beta)} \leq 1$ indicates a tendency toward ion segregation, and conversely, $-c_\alpha/(1-c_\alpha) \leq \eta_i^{(\alpha|\beta)} < 0$ indicates a tendency toward local mixing with species distinct from β at the ith nearest neighbor.

Figure 3.12 plots $Mn_{Pb} - Mn_{Pb}$ SRO parameters for 1st, 2nd, and 3rd NN positions as a function of the halide composition parameter, x. Interestingly, the pure and dilute chloride compositions both show spinodal decomposition involving the phase separation of ordered Mn-rich regions, characterized by $Mn_{Pb} - Mn_{Pb}$ pairs occupying 2NN sites. The decomposition of an initially randomly mixed configuration into coexisting phases suggests a tendency for each Mn_{Pb} to maximize its number of Pb nearest neighbors in the absence of significant halide mixing. The stability of the phase separated configuration is consistent with a simple model of Coulomb energy gain [284, 154]. In this picture, there is a transfer of partial charge, δ, between Mn^{2+} and Pb^{2+} nearest neighbors due to their different electronegativities and the high ionicity of the lattice. This transfer leads to nominal oxidation states of $Pb^{(2-\delta)+}$ and $Mn^{(2+\delta)+}$, and it lowers the Coulomb energy by an amount proportional to $-\delta^2/2r$, where $2r$ is the distance between the Pb^{2+} and Mn^{2+} ions. The configuration that maximizes the number of 1NN Mn^{2+}-Pb^{2+} pairs maximizes this Coulomb energy gain. The optimal configuration thus corresponds to $Mn_{Pb} - Mn_{Pb}$ pairs ordered on 2NN sites. Using DFT and Bader charge analysis [244], we find that the average charge transferred between 1NN Mn^{2+} and Pb^{2+} is $\delta = 0.21$ and 0.19 in the cases of $CsPbMnCl_3$ and $CsPbMnBr_3$, respectively. This 2NN ordering may also reduce the lattice strain associated with substituting Pb^{2+} with the smaller Mn^{2+}, leading to further stabilization relative to the random alloy. Because superexchange is only a short-range interaction, superexchange coupling between 2NN Mn^{2+} ions is negligible and 2NN ordering therefore cannot explain the experimental clustering concluded from the loss of EPR intensity in Figure 3.9. To explain this observation, 1NN $Mn_{Pb} - Mn_{Pb}$ clustering is required.

At $x = 0$, the MC simulations predict an absence of 1NN $Mn_{Pb} - Mn_{Pb}$ clustering. Figure 3.12 shows a rapid increase in 1NN clustering with Br addition in the dilute-bromide limit (small x), followed by non-linear and generally increasing 1NN clustering with increasing Br concentration. Indeed, clustering at all the considered $Mn_{Pb} - Mn_{Pb}$ separations (1NN, 2NN, 3NN) generally increases with increasing x, consistent with Mn segregation into enriched crystalline domains as Br is added to the lattice. Our modeling further predicts that Cl will

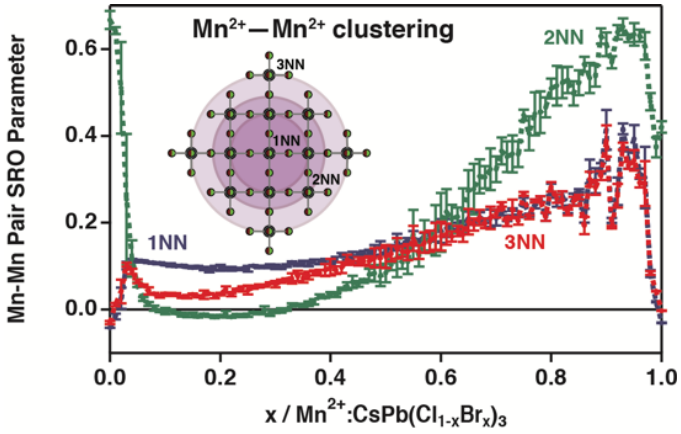


Figure 3.12: Short-range order (SRO) parameters for 1NN, 2NN, and 3NN $\mathrm{Mn_{Pb}} - \mathrm{Mn_{Pb}}$ pairs in 4% $\mathrm{Mn^{2+}}$:$\mathrm{CsPb(Cl_{1-x}Br_x)_3}$ plotted vs the halide composition x, as predicted by a cluster expansion energy model and canonical MCMC methods. Spinodal decomposition is observed at both $x = 0$ and $x = 1$, and $\mathrm{Mn_{Pb}} - \mathrm{Mn_{Pb}}$ 1NN clustering is observed when $x > 0$. The SRO oscillations near $x = 1$ are due to fluctuations between competing low-energy cluster configurations. The inset shows the 1NN, 2NN, and 3NN sites around a central cation, with chloride/bromide ions depicted in green/brown and lead/manganese ions depicted in gray. (Reprinted (adapted) with permission from [225]. Copyright 2019 American Chemical Society.)

also cluster with increasing x, and specifically that the relative configurations of Mn and Cl ions are themselves correlated. For example, Figure 3.13 plots 1NN and 2NN $\mathrm{Mn_{Pb}}$-Cl and $\mathrm{Mn_{Pb}}$-Br SRO parameters as a function of x. These results show that $\mathrm{Mn_{Pb}}$-Cl correlations increase and $\mathrm{Mn_{Pb}}$-Br correlations decrease with increasing x. Although the results generally show anticorrelation for $\mathrm{Mn_{Pb}}$-Br pairs, there is in fact an initial positive correlation at low bromide concentrations that may result from a slight energy stabilization associated with relieving lattice strain by having a smaller Mn dopant next to a larger bromide ion.

We further verify (Figure 3.14) that the Cl-Cl clustering observed in Br-rich compositions is due to the presence of Mn. Indeed, in the absence of Mn, mixed (Cl/Br) halide perovskites

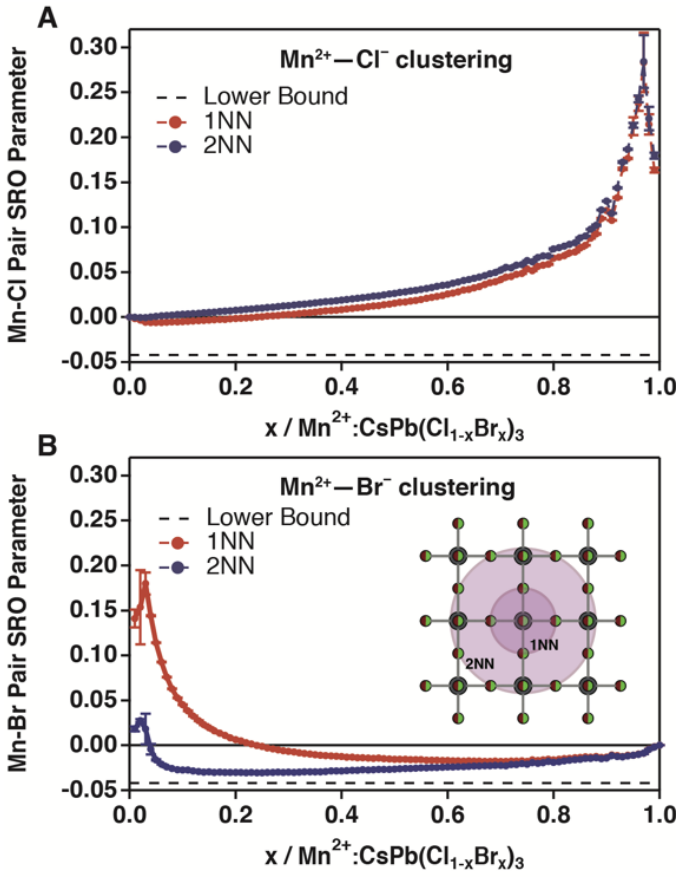


Figure 3.13: Short-range order (SRO) parameters $\eta_i^{(\alpha|\beta)}$ for (A) Mn_{Pb}-Cl and (B) Mn_{Pb}-Br pairs in 4% Mn^{2+}:$CsPb(Cl_{1-x}Br_x)_3$, plotted vs the halide composition x, as predicted by canonical MCMC simulations. Mn_{Pb}-Cl clustering becomes more pronounced as more Br is incorporated. The inset shows the 1NN and 2NN sites around a central cation, with chloride/bromide ions depicted in green/brown and lead/manganese ions depicted in gray. (Reprinted (adapted) with permission from [225]. Copyright 2019 American Chemical Society.)

form miscible solid solution at room temperature, similar to the (Br/I) system discussed in the previous section. As more Mn is added to the mixed (Cl/Br) system (Figure 3.15), it tends to saturate its halide environment with Cl, while driving away Br. In the pure

$CsPbCl_3$, we note (Figure 3.16) that the formation of Mn-rich domains with 2NN ordering occurs almost immediately for dilute Mn concentrations. On the other hand, in $CsPbBr_3$, this demixing transition requires Mn concentrations above $\sim 2\%$.

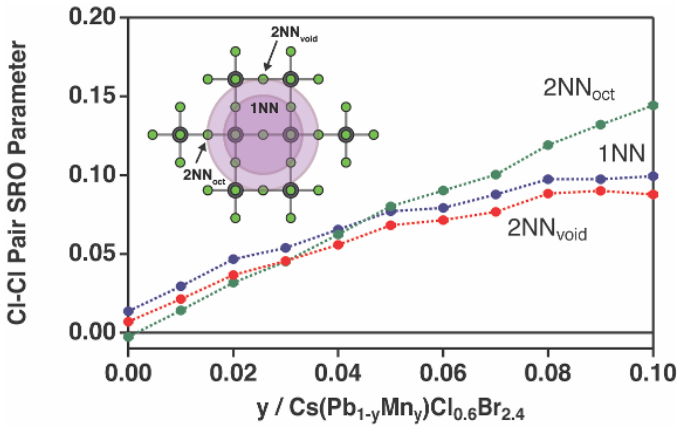


Figure 3.14: Short-range order (SRO) parameters, $\eta_i^{(\mathrm{Cl}|\mathrm{Cl})}$, for Cl pairs under varying Mn^{2+} doping as predicted by canonical MCMC simulation with 20% Cl (cf. Figure 3.11). The lower bound for $\eta_i^{(\mathrm{Cl}|\mathrm{Cl})}$ is not shown for the sake of scale. Chloride clustering is found to increase for increasing Mn^{2+} incorporation. Pairing on 2NN (oct) sites in the same $[BX_6]$ octahedron (B = Mn, Pb) becomes dominant past 5% Mn^{2+}, compared to 2NN (void) sites on distinct octahedra. The inset shows the 1NN and two different 2NN sites with chloride ions in green and lead/manganese ions in gray. (Reprinted (adapted) with permission from [225]. Copyright 2019 American Chemical Society.)

Overall, these results support the two key conclusions drawn from the experimental observations [225], namely that 1NN Mn clustering occurs with increasing x in Mn^{2+}:$CsPb(Cl_{1-x}Br_x)_3$ NCs (Figure 3.9c), and that Mn_{Pb} remains coordinated by Cl despite the addition of Br to the lattice (Figure3.9d). This situation is an example of spinodal decomposition, in which the free energy of an unstable solid solution is reduced by separating into coexisting phases. Beyond supporting the experimental conclusion of Br-induced Mn clustering in Mn^{2+}:$CsPb(Cl_{1-x}Br_x)_3$ NCs, our MCMC simulations reveal additional insights not de-

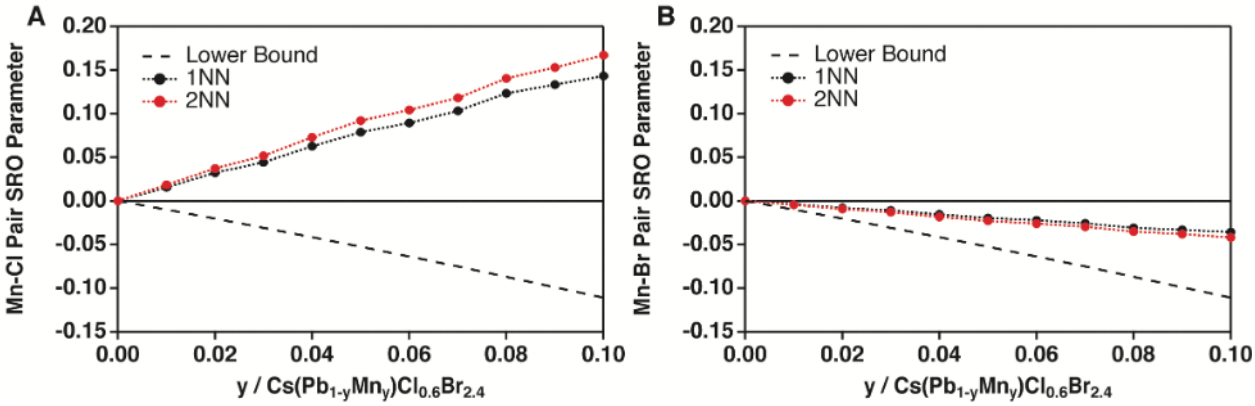


Figure 3.15: Short-range order (SRO) parameters $\eta_i^{(X|\mathrm{Mn})}$ for Cl-$\mathrm{Mn_{Pb}}$ pairs (A) and Br-$\mathrm{Mn_{Pb}}$ pairs (B) under varying $\mathrm{Mn^{2+}}$ doping as predicted by canonical MCMC simulation with 20% Cl. These results show that Cl-$\mathrm{Mn_{Pb}}$ pairs are correlated and that Br-$\mathrm{Mn_{Pb}}$ pairs are anticorrelated suggesting that $\mathrm{Mn_{Pb}}$ has a propensity to cluster into regions of high chloride content. (Reprinted (adapted) with permission from [225]. Copyright 2019 American Chemical Society.)

tected experimentally. In particular, our modeling predicts that $\mathrm{Mn_{Pb}} - \mathrm{Mn_{Pb}}$ ordering is already thermodynamically favored even at the composition endpoints of $\mathrm{Mn^{2+}}$:$\mathrm{CsPbCl_3}$ and $\mathrm{Mn^{2+}}$:$\mathrm{CsPbBr_3}$. The clustering in these endpoint compositions has a strong preference for ordering at 2NN sites, driven by Coulomb interactions and potentially strain minimization. This 2NN ordering cannot, however, explain the Br-induced loss of EPR intensity. Indeed, these conclusions are based on thermodynamics and should be generalizable to $\mathrm{Mn^{2+}}$:$\mathrm{CsPb(Cl_{1-x}Br_x)_3}$ bulk and NCs prepared by different routes. Still, a kinetic factor is also apparent in the observation that the 1NN $\mathrm{Mn_{Pb}} - \mathrm{Mn_{Pb}}$ clustering induced by partial $\mathrm{Cl} \rightarrow \mathrm{Br}$ anion exchange is not lost again upon complete conversion to $\mathrm{CsPbBr_3}$ [225], where computations suggest the 2NN ordering would be more stable. It may, however, be possible that the formation of 1NN ordered Mn in $\mathrm{CsPbBr_3}$ is metastable.

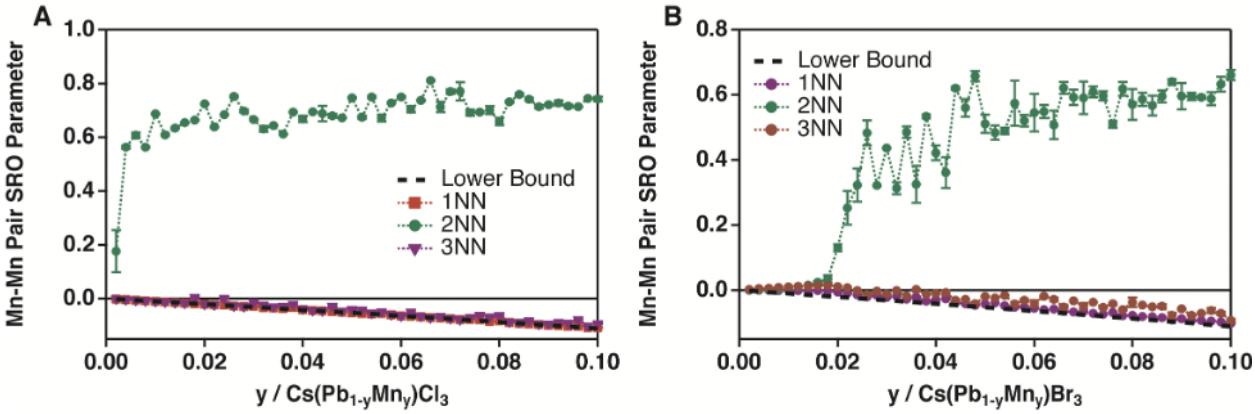


Figure 3.16: Short-range order (SRO) parameters, $\eta_i^{(\mathrm{Mn|Mn})}$, for Mn^{2+} pairs under varying Mn^{2+} doping in pure $CsPbCl_3$ (A) and pure $CsPbBr_3$ (B) as predicted by canonical MCMC simulation. Mn-rich domains featuring 2NN SRO are found for very dilute Mn^{2+} concentration in $CsPbCl_3$ and beyond approximately 1.75% Mn^{2+} concentration in $CsPbBr_3$. (Reprinted (adapted) with permission from [225]. Copyright 2019 American Chemical Society.)

3.3.1 Additional Computational Details

The ECIs were determined from density functional theory (DFT) calculations of the energies of 200 symmetrically distinct structures. To enhance model sparsity and optimize bias-variance tradeoff, cluster selection and ECI fitting were carried out using a Least Absolute Shrinkage and Selection Operator (LASSO) regression combined with 10-fold cross-validation to determine the regularization parameter [182]. The CE model was found to have a root-mean squared error (RMSE) CV score of 10 meV/atom. Monte Carlo (MC) simulations using the CE energy model and the Metropolis-Hastings algorithm with Kawasaki exchange were used to sample room-temperature, equilibrium alloy configurations in the canonical ensemble for various fixed Pb/Mn and Cl/Br compositions. Each MC run was initialized with a random configuration of alloy components, and a minimum of 1000 MC passes were performed to statistically sample the energy and short-range order observables. Each MC

pass is defined as the number of MC trial moves equal to the number of alloyed sites in the simulation cell, where each simulation cell consisted of 32000 alloyed sites. The CASM software [57, 208, 262] and Scikit-learn libraries [195] were used in automating the first-principles calculations, developing the cluster expansion, and performing MC.

DFT calculations based on the all-electron projector-augmented wave (PAW) method were performed using the Vienna Ab Initio Simulation Package (VASP) [125, 124, 126, 28]. The conjugate gradient algorithm was used for structural optimization, and the volume, shape, and atomic positions were relaxed until interatomic Hellmann-Feynman forces were less than 0.01 eV/Å. To reduce total computation time, the structure was optimized using the Perdew-Burke-Ernzerhof (PBE) exchange-correlation functional in the generalized gradient approximation (GGA) [197]. In the pseudopotentials, 9 electrons of Cs ($5s^2 5p^6 6s^1$), 4 electrons of Pb ($6s^2 6p^2$), 7 electrons of Cl ($3s^2 3p^5$), 7 electrons of Br ($4s^2 4p^5$), and 7 electrons of Mn ($3d^5 4s^2$) were treated as valence electrons. The plane-wave basis cut-off was set to 520 eV, and a Gamma-centered k-point mesh was used for Brillouin-zone integration [175]. A k-point density of 5000 k-points per reciprocal number of atoms, which amounts to a $10 \times 10 \times 10$ k-point mesh for a 5-atom, cubic unit cell, was found to be well-converged for various simulation cell sizes.

3.4 Yb Dopants in $CsPbCl_3$

Among the myriad possible dopants in $CsPbX_3$, trivalent rare-earth ions (RE^{3+}, e.g., Sc^{3+}, Y^{3+}, Ce^{3+}, Sm^{3+}, Eu^{3+}, Tb^{3+}, Dy^{3+}, Er^{3+}, and Yb^{3+}) stand out in their ability to provide sharp f-f emission features in the visible and NIR. Surprisingly high PLQYs approaching 200% have recently been achieved in Yb-doped $CsPb(Cl_{1-x},Br_x)_3$ (Yb:$CsPb(Cl_{1-x},Br_x)_3$) owing to efficient quantum cutting [194, 296, 171, 127]. Combined with their strong, broadband absorption in the visible spectrum, these materials are uniquely positioned to act as downconverters for solar technologies. Therefore, it is of particular interest to understand the fundamental mechanisms behind such efficient sensitization of Yb luminescence. Indeed, transient-absorption (TA) spectroscopy in Yb:$CsPbCl_3$ NCs has shown rapid depopulation

of photogenerated excitons on the order of a few picoseconds associated with Yb doping [171], in contrast to much slower energy transfer to Yb observed in other quantum cutting materials [194, 264, 40, 278, 260, 63, 163, 141]. The NIR PLQY was also found to increase with increasingly Cs-rich synthesis conditions [171, 127]. Photoluminescence (PL) measurements of $CsPbCl_3$ similarly doped with optically inert La revealed the existence of a shallow, RE^{3+} dopant-induced defect state. Taken together, these observations led Milstein, *et al.*, to conclude that efficient quantum cutting could be mediated by a charge-compensated defect complex $(2Yb_{Pb} + V_{Pb})^0$ involving two Yb_{Pb}^{1+} substitutionals and a Pb vacancy, V_{Pb}^{2-}. In this picture, a shallow V_{Pb}^{2-} defect rapidly localizes a photogenerated charge carrier, while the spatial proximity of the pair of Yb_{Pb}^{1+} and large exciton Bohr radius allows for simultaneous electronic coupling with both Yb ions, which serve to split the energy of the localized excitation and reemit a pair of NIR photons.

Subsequent work by Kroupa, *et al.*, on Yb:$CsPbCl_3$ thin films verified that quantum cutting is not fundamentally tied to the large surface-to-volume ratio of the NCs, but is rather an intrinsic, bulk-like property of the material [127]. Using time-resolved PL (TRPL), Roh, *et al.*, found direct evidence for a shallow dopant-induced defect state acting as an intermediary for quantum cutting in both NCs and single crystals (SCs) of Yb:$CsPbCl_3$ [212]. Furthermore, while the authors inferred the presence of multiple Yb species in both NC and SC samples, the same species was found to dominate quantum cutting in both forms. These studies are largely consistent with the proposal by Milstein, *et al.*, for the primary role of $(2Yb_{Pb} + V_{Pb})^0$ defect complexes in mediating quantum cutting, but precise atomistic details of the defect structure remain unknown.

In this work, we further investigate defect formation in single-crystal Yb:$CsPbCl_3$ using first-principles electronic structure calculations. We are interested, in particular, in the feasibility of forming $(2Yb_{Pb} + V_{Pb})^0$ defect complexes and the extent to which the formation of this complex competes with other speciations of Yb dopants as a function of material composition. First, we identify the low energy motifs of $Yb_{Pb} + V_{Pb}$ complexes using *ab initio* defect formation energies and binding energies. We provide additional struc-

tural characterization by analyzing Yb defect-induced lattice strain and perturbations to local bonding, discussing their consistency with the experimental studies mentioned above [171, 127, 212]. To understand the distribution of Yb defect speciation under experimentally relevant synthesis conditions, we construct a minimal, self-consistent thermodynamic model for the concentrations of Yb defects as a function of material composition, constrained by the phase stability of $CsPbCl_3$ and parameterized by formation energies. Since variations in material composition are necessarily accompanied by the formation of intrinsic defects, this model also accounts for the formation of intrinsic vacancies, which prior computational studies have identified as the low formation energy defects in $CsPbCl_3$ [114].

Although we identify a motif of $(2Yb_{Pb} + V_{Pb})^0$ with a specific orientation in the perovskite lattice as having the lowest energy, we find that several configurations of the Yb defect complex are likely to be present at finite temperature. Of specific relevance to quantum cutting is the likelihood of finding two Yb_{Pb}^{1+} within a critical radius of a V_{Pb}^{2-}. Our thermodynamic model shows that the fraction of Yb_{Pb}^{1+} in locally bound $(2Yb_{Pb} + V_{Pb})^0$ complexes generally increases both for increasingly Yb-rich synthesis conditions and for increasing the relative Cs:Pb composition. The increasing incorporation of Yb is also found to decrease the concentration of Cl vacancies, which may otherwise act as non-radiative recombination centers. Furthermore, we find that a significant fraction of partially dissociated $Yb_{Pb} + V_{Pb}$ complexes persist even at high Yb doping, which may provide some explanation for the observation of a specific Yb defect speciation dominating quantum cutting despite the presence of multiple Yb species [212].

Finally, we investigate the role of shallow V_{Pb}^{2-} in mediating charge carrier localization prior to quantum cutting, employing a family of effective Hamiltonians to predict bounds on the exciton localization energy and commenting on the potential relevance of large polaron formation. In particular, we find that the inclusion of polaronic effects can, depending on the choice of model, yield good agreement with the experimental PL signatures for RE^{3+} dopant-induced shallow trap emission reported by Milstein, *et al.* [171]. However, our polaron-exciton model choice also predicts weaker exciton binding energies compared to

reported experimental values. We comment on potential sources of this disagreement and argue for our model's consistency with the observation of lower PLQYs and the absence of exciton PL in SC samples reported by Roh, *et al.* [212].

3.4.1 Formation Energies and Charge Transition Levels

In the following, we consider defects in the low-energy, orthorhombic *Pnma* perovskite phase of $CsPbCl_3$ [143, 212]. Formation energies, E_f, of dilute point defects and point defect complexes were calculated from plane-wave, Kohn-Sham density functional theory (DFT) using the supercell approach [121, 125, 73], where

$$E_f(d,q) = E(d,q) - E_{\text{bulk}} + \sum_j n_j(\mu_j^0 + \Delta\mu_j) + q(E_{\text{VBM}} + E_{\text{Fermi}}) + E_{\text{corr}} \ . \tag{3.11}$$

Here, $E(d,q)$ is the total energy of the relaxed supercell containing defect d in charge state q and E_{bulk} is the total energy of the perfect supercell. The integer n_j counts the number of chemical species j removed from ($n_j > 0$) or added to ($n_j < 0$) the system in forming the defect. The synthesis conditions are characterized by the chemical potentials $\Delta\mu_j$ with respect to a set of standard references μ_j^0. Analogously, the Fermi level E_{Fermi} is referenced to the valence band maximum (VBM). A finite-size correction term E_{corr} has been added to account for spurious Coulomb interactions between the point defect and its periodic images [74]. Since the defect complexes considered in this work span multiple Pb-sublattice sites, all defect calculations were performed on a fairly large supercell containing 360 atoms (a $3 \times 3 \times 2$ supercell of the 20 atom *Pnma* unit cell). Assuming stability of $CsPbCl_3$ against secondary phase formation provides additional constraints on the set of primary chemical potentials,

$$x\Delta\mu_{\text{Cs}} + y\Delta\mu_{\text{Pb}} + z\Delta\mu_{\text{Cl}} \leq \Delta H_f(\text{Cs}_x\text{Pb}_y\text{Cl}_z) \ , \tag{3.12}$$

where $\Delta H_f(Cs_xPb_yCl_z)$ are the formation enthalpies of the relevant primary and secondary phases, and the equality is strictly satisfied for the primary phase $CsPbCl_3$, $x = y = 1$ and $z = 3$. The corresponding phase stability diagram (Figure 3.17) shows a narrow existence range for $CsPbCl_3$.

Previous first-principles studies of native point defects in $CsPbX_3$ perovskites (X = Cl, Br, I) have found low formation energies for vacancies over a wide set of processing conditions [114]. Likewise, ample experimental evidence points to the incorporation of Yb via B-site cation substitution [296, 194, 153, 288, 171, 212], so Yb incorporation is only assumed to occur by substituting Pb sites. The Yb chemical potential is chosen to be consistent with experimentally observed Yb composition, [Yb]/([Yb]+[Pb]) [171]. Furthermore, it has been observed that the facile incorporation of Yb occurs under moderately Pb-poor and Cs-rich conditions. Since the Cs chemical potential is tightly constrained by phase stability (Figure 3.17), we choose Pb chemical potentials which are far below the pure Pb secondary phase boundary and consistent with observed Yb compositions and nominal [Cs]:[Pb] ratios [171].

From preliminary calculations of a variety of charge-neutral defect complexes involving

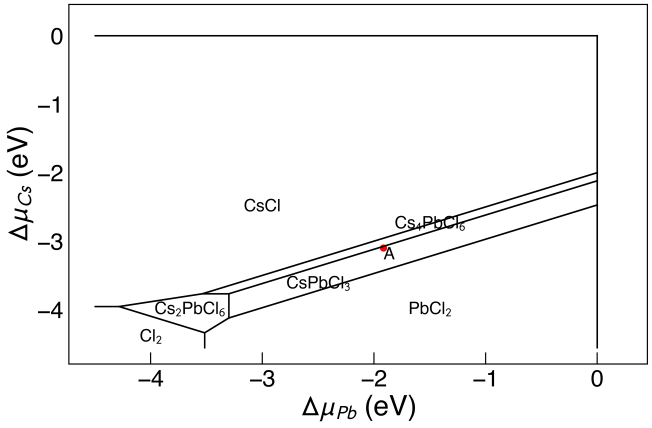


Figure 3.17: Predicted phase stability diagram for $CsPbCl_3$. Chemical potentials chosen at point A reflect moderately Pb-poor, Cs-rich synthesis conditions used in calculating defect formation energies shown in Figure 3.18.

Yb_{Pb}^{1+} compensated by either cation vacancies V_{Cs}^{1-} and V_{Pb}^{2-} or halide interstitials Cl_i^{1-}, we found that $(2Yb_{Pb} + V_{Pb})^0$ consistently had the lowest formation energy for synthesis conditions within the bounds of phase stability. In particular, we considered $(2Yb_{Pb} + 2Cl_i)^0$, $(2Yb_{Pb} + 2V_{Cs})^0$, and $(2Yb_{Pb} + V_{Cs} + Cl_i)^0$ as alternative fully-compensated complexes to $(2Yb_{Pb} + V_{Pb})^0$. For the range of chemical potential consistent with facile Yb incorporation, the formation energy of $(2Yb_{Pb} + V_{Pb})^0$ was lower than the other fully compensated complexes by at least 1 eV. This provides additional justification for treating vacancies as the dominate intrinsic defects in our models, and thus the following analyses focus on complexes involving only Yb_{Pb}^{1+} and V_{Pb}^{2-}.

The computed formation energies of vacancies, Yb_{Pb}, and Yb_{Pb}+V_{Pb} complexes are shown in Figure 3.18 for one choice of moderately Pb-poor, Cs-rich synthesis conditions (point A in Figure 3.17). Table 3.1 lists the corresponding thermodynamic charge transition

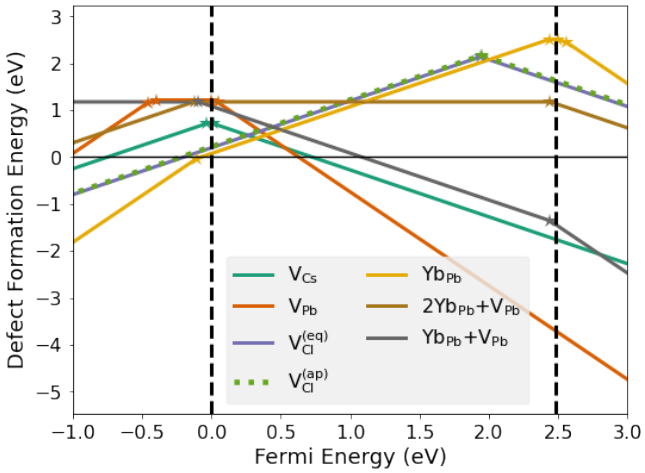


Figure 3.18: DFT formation energies of vacancies, Yb_{Pb} substitutionals, and vacancy-substitutional complexes in moderately Pb-poor $CsPbCl_3$. Chemical potentials correspond to point A in Figure 3.17. The DFT band gap, $E_{gap} = 2.48$ eV, is demarcated by dashed, vertical lines. The difference in formation energy between Cl vacancies at symmetrically distinct apical (ap) and equatorial (eq) is very small.

levels,

$$\varepsilon(q_1|q_2) = \frac{E_f(d,q_1)|_{E_{\text{Fermi}}=0} - E_f(d,q_2)|_{E_{\text{Fermi}}=0}}{q_2 - q_1} \ . \tag{3.13}$$

It is worth emphasizing that the thermodynamic charge transition levels are independent of the chemical potentials. Furthermore, since the ground state of the defect in a given charge state includes geometry relaxation in defining the formation energy, Eq. (3.13) implicitly assumes the defect geometry has sufficient time to equilibrate upon changing the charge state.

We find that the cation vacancies V_{Cs} and V_{Pb} introduce shallow acceptor levels near the VBM. Conversely, an isolated Yb_{Pb} acts as a shallow donor, with a series of thermodynamic charge transition levels near the conduction band minimum (CBM), each separated by ≈ 60 meV (Table 3.1). Interestingly, $2Yb_{Pb} + V_{Pb}$ and $Yb_{Pb} + V_{Pb}$ complexes create, respectively, very deep $(0|-1)$ and $(-1|-2)$ acceptor levels near the CBM. Thus, Yb_{Pb} substitutionals, whether bound or unbound to Pb vacancies, introduce single-particle, Kohn-Sham states resonant within the conduction band. These defect-localized states (DLS) are typically associated with the existence of an effective-mass-like, perturbed-host state slightly below the conduction band; the latter effective-mass state is generally not observable in supercell calculations due to finite size constraints [73, 291, 133]. For isolated Yb_{Pb}, this picture is consistent with its shallow donor character. Adding a Pb vacancy introduces coupling to the valence band via Cl dangling bonds [118]. For fully compensated $2Yb_{Pb} + V_{Pb}$ complexes, this stabilizes the neutral charge state throughout most of the band gap. It is possible that the isolated $(0|-1)$ acceptor state close to the CBM corresponds to the Yb dopant-induced intermediate state proposed as essential to sensitizing efficient quantum cutting [171, 212]. However, it is difficult to resolve the accuracy of DFT formation energies and transition levels below ≈ 50 meV [73]; the presence of this transition level may simply reflect the filling of bulk conduction band states.

We find very little difference in formation energies between Cl vacancies at symmetrically distinct apical (ap) and equatorial (eq) Cl sites (Figure 3.19). These vacancies form

somewhat deep amphoteric defect states with negative-U behavior [1] and may act as non-radiative recombination centers [47, 286]. This negative-U behavior is accompanied by large relaxation in the distance between neighboring Pb in the [$PbCl_5$] octahedra sharing a V_{Cl}^{1-} at their adjacent corner. The resulting formation of a contracted Pb-Pb dimer and the bonding character of Pb p-orbitals shifts the V_{Cl}^{1-} DLS into the band gap [223, 118]. It would be interesting to determine whether Cl vacancies can exhibit metastable behavior under photoexcitation [131, 132], particularly in light of the observation made by Roh, *et al.*, of slow quantum cutting due to metastable trapping states and lower PLQYs in SC samples [212]. However, it is worth noting that due to the stability constraints on $CsPbCl_3$ (Figure 3.17), increasingly Pb-poor synthesis conditions approach Cl-rich secondary phase boundaries and would thus disfavor the formation of Cl vacancies. For the conditions depicted in Figure 3.18, compensation between positively charged V_{Cl}^{1+} and Yb_{Pb}^{1+} and negatively charged V_{Pb}^{2-} and V_{Cs}^{1-} pins the Fermi level at about 0.4 eV above the VBM.

To gain some insight into whether Yb_{Pb}+V_{Pb} complexes can form in sufficient concentrations to contribute to quantum cutting and high PLQYs, we tabulate (Table 3.2) binding

[1] From the thermodynamic charge transition levels, $\varepsilon(+1/0)$ and $\varepsilon(0/-1)$, the effective U parameters is approximately $U_{\text{eff}} = -0.41$ and -0.54 eV for $V_{Cl}^{(ap)}$ and $V_{Cl}^{(eq)}$, respectively.

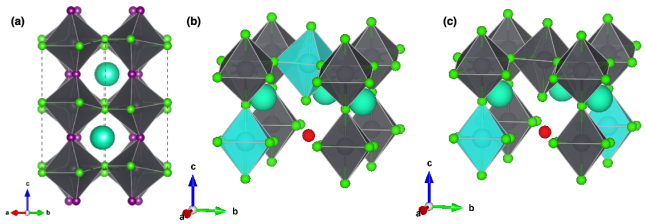


Figure 3.19: (a) Symmetrically distinct apical (purple) and equatorial (green) Cl sites associated with Pb-Cl octahedral (grey) tilting in the 20-atom unit cell of *Pnma* perovskite phase of $CsPbCl_3$. (b) Orthogonal OT $[110]^{\perp}$ and (c) collinear CL [110] motifs of $(2Yb_{Pb} + V_{Pb})^0$ complex, where Yb-Cl octahedra (teal) are first nearest neighbors to a vacant Pb site (red).

energies for $(\mathrm{Yb_{Pb}} + \mathrm{V_{Pb}})^{1-}$ and $(2\mathrm{Yb_{Pb}} + \mathrm{V_{Pb}})^{0}$ complexes in symmetrically distinct configurations. We define the binding energy as the difference between the formation energy of the complex, c, and the sum of the formation energies of the individual, isolated point defects, d,

$$E_{\mathrm{bind}} = E_f(c, q_c) - \sum_{d \in c} E_f(d, q_d) \ . \tag{3.14}$$

Note that this is in contrast to the other commonly used definition, where the energy of the complex is referenced to a single point defect and a complementary sub-complex. Among $(\mathrm{Yb_{Pb}} + \mathrm{V_{Pb}})^{1-}$ complexes, the [001] orientation is marginally energetically favorable. In contrast, among the collinear (CL) $(2\mathrm{Yb_{Pb}} + \mathrm{V_{Pb}})^{0}$ configurations, the [110] orientation is favorable. The lowest energy $(2\mathrm{Yb_{Pb}} + \mathrm{V_{Pb}})^{0}$ configuration is found for an orthogonal (OT) motif, where the vector perpendicular to the plane spanned by the complex lies in the [110] direction, denoted by $[110]^{\perp}$ (*cf.* Table 3.2). Restricting to the minimum energy $(\mathrm{Yb_{Pb}} + \mathrm{V_{Pb}})^{1-}$ [001] complex, we find that about 0.13 eV is gained upon adding another

Table 3.1: Thermodynamic charge transition levels relative to VBM. For reference, the DFT band gap is $E_{\mathrm{gap}} = 2.48$ eV, and intraband states within 0.1 eV to band edges are also included.

Defect	$\varepsilon(q_1\|q_2)$ [eV]
$2\mathrm{Yb_{Pb}} + \mathrm{V_{Pb}}$	2.438 $(0\|-1)$
$\mathrm{Yb_{Pb}} + \mathrm{V_{Pb}}$	-0.093 $(0\|-1)$, 2.440 $(-1\|-2)$
$\mathrm{Yb_{Pb}}$	2.435 $(1\|0)$, 2.497 $(0\|-1)$, 2.559 $(-1\|-2)$
$\mathrm{V_{Pb}}$	0.005 $(-1\|0)$, 0.045 $(-1\|-2)$
$\mathrm{V_{Cs}}$	-0.038 $(1\|0)$, 0.013 $(0\|-1)$
$\mathrm{V_{Cl}^{(ap)}}$	1.945 $(+1\|-1)$
$\mathrm{V_{Cl}^{(eq)}}$	1.939 $(+1\|-1)$

Yb_{Pb}^{1+} to form a $(2Yb_{Pb} + V_{Pb})^0$ OT $[110]^{\perp}$ complex. That the OT motifs are marginally more stable than CL motifs is contrary to what one would expect based on simple electrostatic interactions between the constituent point defects, suggesting that lattice strain among the different configurations plays a nontrivial role. A similar claim was made by Li, *et al.*, for $(2Yb_{Pb} + V_{Pb})^0$ complexes in the cubic perovskite phase of $CsPbCl_3$, who argued that the number of neighboring, distorted Pb-Cl octahedra differentiate OT and CL motifs [139]. The magnitudes of the complex binding energies relative to k_BT indicate relative favorability for the formation of all complex motifs at room temperature. We consider simple thermodynamic models of the defect distributions based on the formation energies in Section 3.4.3.

3.4.2 Structural Characterization

The formation of V_{Pb}^{2-}, Yb_{Pb}^{1+}, and their complexes is accompanied by a certain amount of lattice strain. To quantify the degree of strain induced by defect formation, we calculate the deformation matrix $\mathbf{F}$ of the pristine bulk lattice $\mathbf{L}_0$ which produces the relaxed, defective supercell lattice $\mathbf{L}_1$, where

$$\mathbf{L}_1 = \mathbf{F}\mathbf{L}_0 \ . \tag{3.15}$$

A generalized, one-parameter family of strain tensors can be defined by [247]

$$\mathbf{E}(\mu) = \frac{(\mathbf{F}^{\top}\mathbf{F})^{\mu} - \mathbf{I}}{2\mu} \ , \tag{3.16}$$

where the typical Green-Lagrange strain is given by $\mu = 1$. In Table 3.2, we have tabulated the symmetry-adapted components, e_i, Equations (3.17-3.22), of the Green-Lagrange strain tensor for V_{Pb}^{2-}, Yb_{Pb}^{1+}, and their complexes [247]:

$$e_1 = (E_{xx} + E_{yy} + E_{zz})/\sqrt{3} \tag{3.17}$$

$$e_2 = (E_{xx} - E_{yy})/\sqrt{2} \tag{3.18}$$

$$e_3 = (2E_{zz} - E_{xx} - E_{yy})/\sqrt{6} \tag{3.19}$$

$$e_4 = \sqrt{2}E_{yz} \tag{3.20}$$

$$e_5 = \sqrt{2}E_{xz} \tag{3.21}$$

$$e_6 = \sqrt{2}E_{xy} \tag{3.22}$$

The first component of the Green-Lagrange strain, e_1, characterizes purely volumetric contraction or expansion of the lattice. Volume preserving, deviatoric strains are spanned by e_2 and e_3, while shear strains are given by e_4, e_5, and e_6. We note that the magnitudes of the strain components (Table 3.2) are small enough that they are well-approximated by the linearized strain, $\mathbf{E} \approx \frac{1}{2}\left(\mathbf{F} + \mathbf{F}^\top\right) - \mathbf{I}$.

We find that the largest volume changes occur for the isolated defects $\mathrm{V}_{\mathrm{Pb}}^{2-}$ and $\mathrm{Yb}_{\mathrm{Pb}}^{1+}$, where the lattice expands in forming the vacancy and contracts for the substitutional. Remarkably, we find no significant volume change for the $(\mathrm{Yb}_{\mathrm{Pb}} + \mathrm{V}_{\mathrm{Pb}})^{1-}$ complexes, suggesting that the volumetric strains of the constituent point defects essentially compensate each other. This may provide some explanation for the conspicuous absence of differences in lattice parameters in XRD data between undoped and Yb-doped $CsPbX_3$ NCs and SCs [171, 127, 212, 64, 48, 44, 170]. To further investigate this observation, we invoke a simple thermodynamic model for the defect-perturbed lattice volume V_{def} as a function of defect concentrations $c_{d,q}$,

$$V_{\mathrm{def}} = V_0 + \sum_{d,q} c_{d,q}\nu_{d,q} + n_h\nu_h + n_e\nu_e \ , \tag{3.23}$$

where $\nu_{d,q} = p_{d,q}V_0/B_0$ is the partial volume of defect d in charge state q, B_0 is the bulk modulus (≈ 23.0 GPa), and V_0 is the bulk volume [267]. An effective defect pressure is defined

by the change in the DFT total energy of the defect supercell during volume relaxation [267, 112],

$$p_{d,q} = -\frac{\Delta\left(E_{d,q}(V) - E_{\text{bulk}}(V)\right)}{\Delta V} \ , \tag{3.24}$$

where the difference, $\Delta(\cdot) = (\cdot)\big|_{V=V_{\text{def}}} - (\cdot)\big|_{V=V_0}$, is evaluated at the relaxed, V_{def}, and unrelaxed, V_0, DFT cell volumes. To account for volume changes associated with the introduction of free charge carriers consistent with overall charge neutrality (3.27), we augment Equation (3.23) by the partial volumes $\nu_{e,h}$ of free electrons and holes introduced into bulk $CsPbCl_3$, with respective concentrations $n_{e,h}$.

Using the model for defect concentrations (3.26), discussed in Section 3.4.3, Figure 3.20 shows that the predicted contraction of the effective lattice constant $a_{\text{def}} = V_{\text{def}}^{1/3}$ is less than 0.1 Å for Yb incorporation within the experimentally relevant range. For less than 2% Yb incorporation, the lattice constant is reduced by less than 0.1% relative to the unperturbed value $a_0 = V_0^{1/3}$, limited primarily by the compensation in volumetric strain between V_{Pb}^{2-} and Yb_{Pb}^{1+}. For higher Yb incorporation, the change in the lattice constant becomes dominated by the formation of $(2Yb_{Pb} + V_{Pb})^0$ complexes, but the overall change relative to the bulk is less than 1.8%. Moreover, extensive charge compensation between V_{Pb}^{2-}, Yb_{Pb}^{1+}, and $(Yb_{Pb} + V_{Pb})^{1-}$ also limits the contribution of free charge carriers to the volume deformation, which only reach concentrations on the order of 10^{10}-10^{12} cm^{-3} under the experimental synthesis conditions considered here (*cf.* Section 3.4.3).

To some extent, the differences in binding energies among the configurations of the $(2Yb_{Pb} + V_{Pb})^0$ complexes can be understood from the local distortion of Pb-Cl bond lengths near the Yb_{Pb}^{1+} and the distribution of Green-Lagrange strain components. In the pristine *Pnma* perovskite phase, the average predicted Pb-Cl bond length is 2.89 Å, while a single Yb_{Pb}^{1+} dopant forms contracted 2.62 Å bonds with surrounding Cl, leading to increased Pb-Cl bond lengths at the shared corners of neighboring Pb-Cl octahedra. That the pairing of two Yb_{Pb}^{1+} with a V_{Pb}^{2-} is no longer compensated in volumetric strain (Table 3.2) compared to $(Yb_{Pb} + V_{Pb})^{1-}$ complexes follows from the additional set of contracted Yb_{Pb}-Cl bonds. We

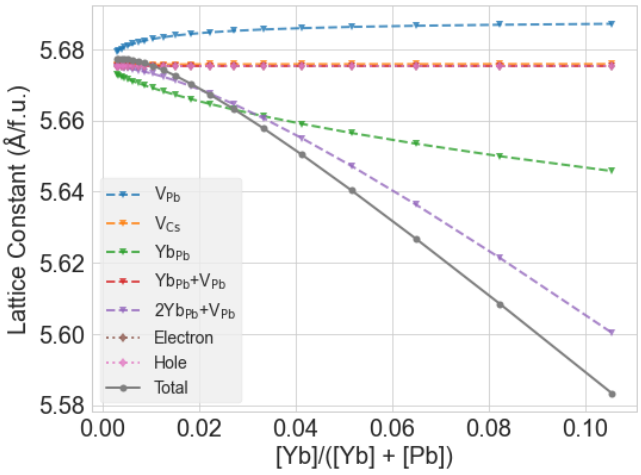


Figure 3.20: Total change in pseudocubic lattice constant $a_{\text{def}} = V_{\text{def}}^{1/3}$ per formula unit (f.u.) due to Yb doping and defect formation, where V_{def} is given by (3.23). Individual contributions to a_{def} from each defect and free electrons and holes are also shown.

also note that among the $(2\text{Yb}_{\text{Pb}} + \text{V}_{\text{Pb}})^0$ motifs, the lower-energy OT configurations induce smaller amounts of xz- and yz-shear strain. Finally, let us mention that the presence of a neighboring Pb vacancy reduces the Yb_{Pb} orthorhombic site symmetry, from O_{h} for isolated $\text{Yb}_{\text{Pb}}^{1+}$, to $D_{2\text{h}}$ (Table 3.2). This is largely consistent with the spectroscopic analysis reported by Roh, *et al.* [212]. It is interesting to note that one Yb in the lowest energy configuration $(2\text{Yb}_{\text{Pb}} + \text{V}_{\text{Pb}})^0$ OT $[110]^{\perp}$ exhibits an additional reduction to reflection site symmetry, C_{s}.

To investigate finite temperature effects on bonding, we performed room-temperature molecular dynamics (MD) simulations of the pristine and defective 360-site supercells and extracted the reduced pair distribution function (rPDF), $G(r)$:

$$G(r) = 4\pi r \langle n \rangle \left(g(r) - 1 \right) \ , \tag{3.25}$$

where $\langle n \rangle$ is the average number density of atoms, $g(r) = \sum_{j \neq k} \langle \delta \left(r - r_{jk} \right) \rangle / \left(\langle n^2 \rangle \Omega \right)$ is the pair distribution function (PDF). The brackets, $\langle \dots \rangle$, denote an ensemble average over the MD trajectories, $r_{j,k}$ is the distance between atoms j and k, and Ω is the supercell volume.

Table 3.2: Binding energies E_{bind} of vacancy-substitutional defect complexes, local [$YbCl_6$] symmetry, relaxed defect volume V_{def} from DFT relative to bulk volume V_0, and symmetry-adapted components e_i of the Green-Lagrange strain tensor with respect to relaxed, bulk $CsPbCl_3$ lattice. The largest absolute values of the strain components in the defect set are highlighted in bold. Lattice orientations of the defect complexes are specified by $[\cdots]$, where $[\cdots]^{\perp}$ denotes the vector perpendicular to the plane spanned by an orthogonal (OT) complex.

Defect Complex	E_{bind} [eV]	Yb Site Sym.	V_{def}/V_0	e_1 [10^{-3}]	e_2 [10^{-3}]	e_3 [10^{-3}]	e_4 [10^{-3}]	e_5 [10^{-3}]	e_6 [10^{-3}]
$(2Yb_{Pb} + V_{Pb})^0$ CL [110]	-0.308	D_{2h}	0.994	-3.323	-1.670	-6.490	**-0.503**	0.027	0.328
$(2Yb_{Pb} + V_{Pb})^0$ CL [001]	-0.278	O_h	0.994	-3.199	-2.631	-4.966	0.087	**-0.347**	**-2.096**
$(2Yb_{Pb} + V_{Pb})^0$ OT $[110]^{\perp}$	-0.340	D_{2h}, C_s	0.994	-3.172	-2.006	-6.004	-0.314	-0.149	1.174
$(2Yb_{Pb} + V_{Pb})^0$ OT $[001]^{\perp}$	-0.315	D_{2h}	0.994	-3.214	-1.873	-6.758	0.016	-0.232	0.388
$(Yb_{Pb} + V_{Pb})^{1-}$ [110]	-0.190	D_{2h}	1.000	-0.231	-2.136	-6.900	-0.294	0.001	0.787
$(Yb_{Pb} + V_{Pb})^{1-}$ [001]	-0.207	D_{2h}	1.000	-0.156	-2.617	-6.179	0.084	-0.060	-1.432
$Yb_{Pb}{}^{1+}$	–	O_h	**0.992**	**-4.626**	**-3.446**	-5.501	0.034	-0.126	-0.370
$V_{Pb}{}^{2-}$	–	–	**1.005**	2.916	-2.714	**-6.951**	0.029	0.068	1.160

In Figure 3.21, we compare the simulated rPDF for the ensemble of $(2Yb_{Pb} + V_{Pb})^0$ complexes against the bulk reference system and supercells containing either a V_{Pb}^{2-} or a Yb_{Pb}^{1+}. We find that the broadening observed in the rPDF of the $(2Yb_{Pb} + V_{Pb})^0$ system is primarily due to the presence of the Pb vacancy. This can be qualitatively understood by invoking a harmonic approximation for the vibrational modes of the lattice: a vacant Pb site is expected to locally reduce the curvature of the Born-Oppenheimer energy landscape, which introduces lower frequency modes and increases the variance of the pair distribution function. Conversely, sharper features in the rPDF of isolated Yb_{Pb}^{1+} are likely due to the shortened Yb-Cl bonds and induced lattice contraction. Interestingly, rPDF features involving Cs provide the clearest distinctions among the four systems. Previous computational works have identified the key role played by displacements of A-site cations in ABX_3 halide perovskites in stabilizing the orthorhombic *Pnma* phase [19, 285]. Moreover, the low-temperature in-

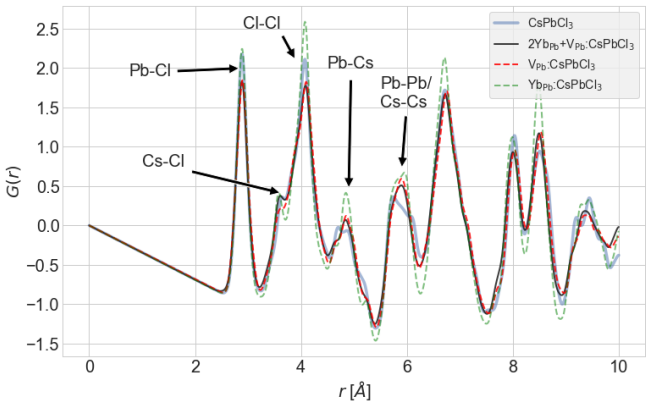


Figure 3.21: Reduced pair distribution functions from room temperature MD simulations of 360-site defect and bulk supercells. Supercells containing an isolated Yb_{Pb} and a compensated $(2Yb_{Pb} + V_{Pb})^0$ complex correspond to Yb fractions, [Yb]/([Yb] + [Pb]), of 1.4% and 2.8%, respectively. The first several features are labeled by the dominant pairs of lattice sites involved.

stability of the cubic perovskite structure to $[BX_6]$ octahedral tilting is often attributed to the presence of an undersized A-site cation; Bechtel and Van der Ven, as well as Young and Rondinelli, found that the tilting mode toward *Pnma* symmetry minimized unfavorable electrostatic interactions and increased A-X covalency by allowing A-site cations to displace from their ideal cubic positions [19, 285]. The presence of Yb_{Pb} substitutionals and Pb vacancies provides additional strain-induced distortions to the volumes of the cuboctahedral voids containing Cs, thereby perturbing the local energy surfaces associated with Cs rattling modes. Whether doping by Yb alters structural phase transitions in halide perovskites, as has been argued for Bi^{3+} and Ni^{2+} doping [112, 152], is a question we leave for potential future work.

3.4.3 Defect Complex Formation

In this section, we discuss a simple model for the concentration of Yb defects and defect complexes based on the DFT data summarized in Section 3.4.1. Again, we take as our starting point the set of native vacancies, Yb_{Pb}, and Yb_{Pb}+V_{Pb} complexes. For a given choice of chemical potentials and temperature, equilibrium defect concentrations are estimated by the standard Arrhenius relation in the dilute approximation [73],

$$c_{d,q}(\Delta\mu_i, E_{\text{Fermi}}, T) = g \exp\left(-E_f(q,d)/k_{\text{B}}T\right) \ , \tag{3.26}$$

where the formation energy $E_f(q,d)$ is given by Equation (3.11) and g is a degeneracy factor accounting for internal and configurational degrees of freedom of the defect. The Fermi level is determined self-consistently by enforcing overall charge neutrality,

$$\sum_{d,q} q c_{d,q} + n_h - n_e = 0 \ , \tag{3.27}$$

where n_h and n_e are the concentrations of free hole and electrons, respectively. The free carrier concentration can be obtained either by a Fermi integral over the electronic density of states or in an isotropic, parabolic band approximation with electron and hole effective masses $m^*_{e,h}$ (Table 3.3).

As discussed in Section 3.4.1, given the tight stability constraints on the Cs chemical potential, we choose chemical potentials of Yb and Pb to be consistent with empirical Yb compositions and nominal Cs:Pb ratios, focusing specifically on moderately Pb-poor synthesis conditions. Figure 3.22 shows the defect concentrations and self-consistent Fermi level for increasing Yb chemical potential, where the Cs and Pb chemical potentials remain fixed at point A in the phase stability diagram (Figure 3.17). The formation of Yb^{1+}_{Pb} defects leads to charge compensation by negative cation vacancies V^{2-}_{Pb} and V^{1-}_{Cs}. A reduction in V^{1+}_{Cl} and the increasing concentration of (partially) compensated $(Yb_{Pb} + V_{Pb})^{1-}$ and $(2Yb_{Pb} + V_{Pb})^{0}$ complexes further leads to a slight increase in the Fermi level for sufficient Yb incorporation. It is worth noting that previous experimental work found that doping by trivalent lanthanide

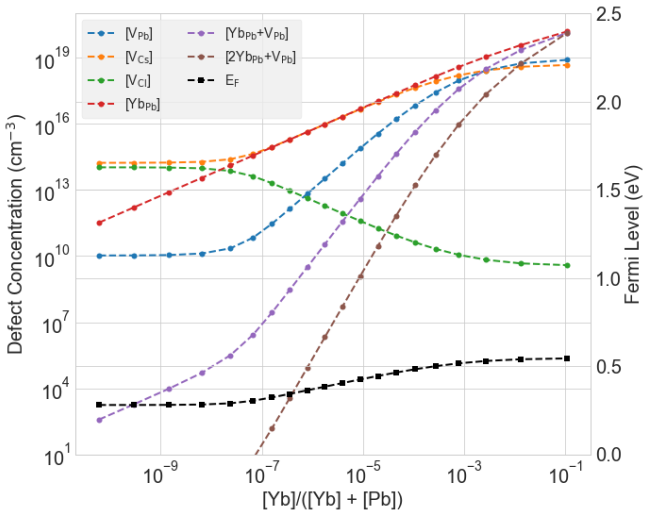


Figure 3.22: Room temperature defect concentrations and self-consistent Fermi level for increasing Yb incorporation with fixed Cs and Pb chemical potentials.

cations suppressed the formation of Cl vacancies, which agrees with our model prediction [271].

Since Milstein, *et al.*, found that NIR PLQY increased with increasingly Cs-rich synthesis conditions [171], we also assess the extent to which controlling the relative chemical potentials of Cs and Pb influences the formation of Yb complexes. Figure 3.23 shows the estimated distribution of various spatial configurations of Yb_{Pb}^{1+} relative to V_{Pb}^{2-} (see below) for increasing [Cs]/[Pb] with fixed, relatively favorable Yb and Cl chemical potentials. We found that the effect of increasingly Cs-rich, Pb-poor synthesis conditions within the bounds set by phase stability is essentially identical to controlling the Yb chemical potential. Hence, Figure 3.23 also shows the equivalent change in Yb fraction, [Yb]/([Yb] + [Pb]). In particular, decreasing the Pb chemical potential increases both the concentration of Pb vacancies and Yb_{Pb}, and increasing the Cs chemical potential disfavors the formation of Cs vacancies. The ensuing charge compensation between Yb_{Pb}^{1+} and V_{Pb}^{2-} and their increasing concentrations lead to the formation of $(2Yb_{Pb} + V_{Pb})^0$ complexes.

The binding energy calculations of Section 3.4.1 (Table 3.2) indicate that several motifs of the $(2\mathrm{Yb_{Pb}} + \mathrm{V_{Pb}})^0$ complex should be expected at room temperature; entropic considerations suggest that an ensemble of partially dissociated configurations of $\mathrm{Yb_{Pb}^{1+}}$ and $\mathrm{V_{Pb}^{2-}}$ should also be present. Furthermore, while the static dielectric constant, ϵ_s, of $CsPbCl_3$ is relatively large, the Debye length scale is on the order of the effective exciton Bohr radius, a_B^*, for the self-consistently determined free carrier concentrations. Hence, some degree of screened Coulomb attraction between $\mathrm{Yb_{Pb}^{1+}}$ and $\mathrm{V_{Pb}^{2-}}$ should additionally bias the distribution of Yb toward partially bound configurations within a_B^*, which may be important for quantum cutting. To characterize the spatial distribution of Yb relative to Pb vacancies, we assume pairwise interactions between point defects in the form of a screened Coulomb potential $\phi(r) \propto \exp(-r/L_\mathrm{D})/(\epsilon_s r)$. The Debye length was determined by the self-consistent free carrier concentrations $L_\mathrm{D} = \sqrt{\epsilon_s \epsilon_0 k_\mathrm{B} T/(e^2(n_e + n_h))}$, and the static dielectric constant was calculated using density functional perturbation theory (Table 3.3). It is important to note that while this model introduces errors in reproducing the DFT binding energies of local motifs (Table 3.2) on the order of $k_\mathrm{B}T$ at room temperature, it is within the expected accuracy of DFT formation energies [73]. The probability of finding $\mathrm{Yb_{Pb}^{1+}}$ within a given distance of a Pb vacancy is determined by sampling the $\mathrm{Yb_{Pb}}$-$\mathrm{V_{Pb}}$ pair distribution function using Metropolis Monte Carlo with a Pb vacancy fixed at the origin. For simplicity and computational efficiency, we maintain a semidilute approximation by populating the simulation with two $\mathrm{Yb_{Pb}^{1+}}$, while scaling the simulation volume with the total $2\mathrm{Yb_{Pb}} + \mathrm{V_{Pb}}$ concentration [2].

Fast rise times observed by Roh, *et al.*, in TRPL measurements support the hypothesis that efficient quantum cutting in Yb:$CsPbCl_3$ depends on short-range, Dexter-type exchange mediating energy transfer between dopant-induced $\mathrm{V_{Pb}^{2-}}$ and nearby $\mathrm{Yb_{Pb}^{1+}}$ [212, 171, 58]. Since such an exchange mechanism is typically short-range, configurations of $\mathrm{V_{Pb}^{2-}}$ and two $\mathrm{Yb_{Pb}^{1+}}$ in which each $\mathrm{Yb_{Pb}^{1+}}$ is within the third nearest neighbor (3nn) shell on the Pb-sublattice of a

[2] Spurious self-interactions between periodic images are avoided by tracking only pairwise interactions between distinct defects, which is computationally tractable in the semidilute limit considered.

V_{Pb}^{2-} are likely more relevant for quantum cutting. These configurations are labeled "local" in Figure 3.23. "Dissociated" configurations are defined as having at least one Yb_{Pb}^{1+} beyond the local (3nn) cutoff set by local exchange but still within the Wannier-Mott exciton Bohr radius (Table 3.3). These configurations may play some role in localizing a photogenerated exciton, but are unlikely to significantly contribute to quantum cutting. Finally, configurations in which at least one Yb_{Pb}^{1+} is beyond the dissociated cutoff a_B^* from a V_{Pb}^{2-} are considered "isolated." In Section 3.4.4, we verify the consistency of this choice of cutoffs with variational calculations of the radius of the exciton bound to the defect (Table 3.3).

For very dilute ($< 1\%$) Yb incorporation, Yb_{Pb}^{1+} are predominantly isolated, but small increases in Yb fraction lead to rapid formation of $(Yb_{Pb} + V_{Pb})^{1-}$ dimers, which have significant binding energy. Additional long range Coulomb attraction and configurational entropy also increase the population of dissociated $(Yb_{Pb} + V_{Pb})_{1nn}^{1-} + Yb_{Pb}^{1+}$ configurations, which become dominant as increasing concentration decreases the average contact distance between defects. At 2% Yb composition, local $2Yb_{Pb} + V_{Pb}$ configurations constitute about 20% of the Yb defect species, where approximately half of those are fully bound at 1nn. This supports the suggestion by Roh, *et al.*, that a specific Yb defect motif accounts for the majority of quantum-cutting PL intensity, despite not being the dominant speciation [212]. Fully delineating, based on first principles, the extent to which different local Yb configurations relative to a Pb vacancy influence local energy transfer is beyond the scope of this work, although we comment further on this in Section 3.4.4.

Beyond 2% Yb composition, the growth in local $2Yb_{Pb} + V_{Pb}$ configurations becomes increasingly linear. Similar, approximately linear behavior for the onset of large PLQYs in Yb:$CsPbCl_3$ was observed by Milstein, *et al.*, and Kroupa, *et al.* [171, 127]. In modeling the PL kinetics of Yb-doped $CsPb(Cl_{1-x}Br_x)_3$, Erickson, *et al.*, found that an effective quantum-cutting rate proportional to the number of Yb-dimers in a specific pairing better reproduced the experimentally observed Yb-concentration dependence in the PLQY compared to a rate proportional to all possible pairs of Yb [64], supporting the hypothesis that the propensity for pairs of Yb to participate in quantum cutting depends on their relative spatial arrangement.

Implicit in their argument was the locality of these Yb-dimers, although they did not model it explicitly. More importantly, they experimentally verified that increasing Yb concentration provided an effective route toward reducing PL saturation at higher photoexcitation rates by increasing the relative fraction of singly-excited Yb-dimers in steady-state [64]. Our model for the spatial distribution of Yb defects accounts for the underlying material synthesis conditions. For small, but increasing Yb incorporation, the fraction of local configurations grows more rapidly than at higher Yb incorporation. This initial nonlinear behavior is due not only to the low probability of finding local Yb pairs for very dilute Yb concentrations, but also depends on the initial scarcity of compensating V_{Pb}^{2-} (*cf.* Figure 3.22). This may provide some explanation for the observed disappearance of excitonic PL in NCs noted by Milstein, *et al.*, as a function of Yb doping, although competing exciton recombination processes may eliminate this behavior in SCs [171, 212]. Further increasing the Yb fraction leads to most Yb_{Pb}^{1+} being within an exciton Bohr radius of a Pb vacancy. Beyond 3% Yb composition, the relative proportion of dissociated $2Yb_{Pb} + V_{Pb}$ configurations decreases in favor of an increasing proportion of local configurations. Importantly, within the experimentally relevant range of Yb incorporation, the growth in local configurations is monotonic, which is consistent with the mechanism proposed by Erickson, *et al.*, for reducing PL saturation by increasing Yb concentration [64].

3.4.4 Exciton Binding and Defect Localization

The quantum cutting mechanism proposed by Milstein, *et al.*, relies not only on the spatial proximity of two Yb_{Pb}^{1+} to a Pb vacancy, but also on the capacity for this doping-induced defect state to bind a photogenerated exciton [171, 127, 212]. In this picture, the Pb vacancy is thought to play the primary role in the initial energy capture before subsequent energy transfer to the nearby Yb_{Pb}^{1+} pair. The shallow nature of this defect state is particularly important, as it facilitates energy localization on the picosecond time scale and is unlikely to decrease the exciton Bohr radius beyond the scale required for simultaneous electronic coupling to both Yb_{Pb}^{1+}.

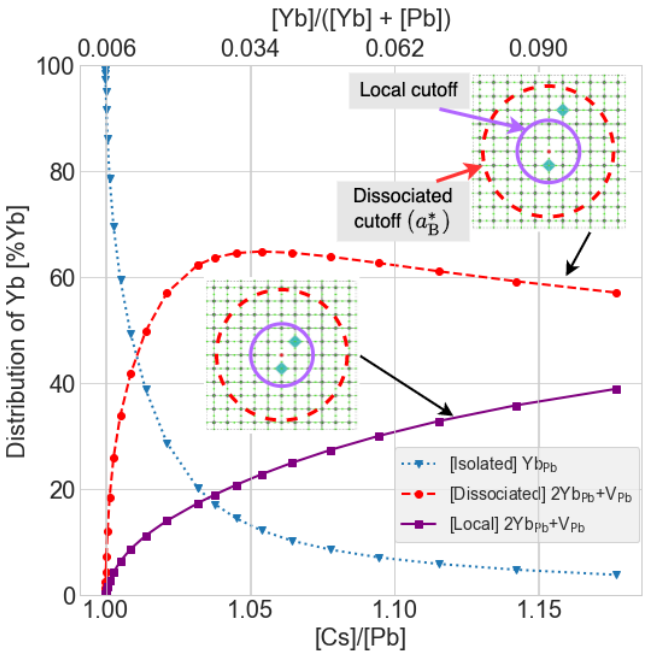


Figure 3.23: Room temperature, spatial distribution of Yb_{Pb}^{1+} relative to V_{Pb}^{2-} for fixed Yb and Cl chemical potentials and increasing relative Cs:Pb composition, [Cs]/[Pb]. The equivalent change in Yb fraction, [Yb]/([Yb] + [Pb]), is shown on the top axis. Local configurations (solid lines) of the $2Yb_{Pb} + V_{Pb}$ complex are defined by having both Yb_{Pb}^{1+} within an approximate length scale for Dexter-type exchange (≈ 10 Å) to the V_{Pb}^{2-}. Dissociated configurations (dashed lines) are defined by having both Yb_{Pb}^{1+} within the effective exciton Bohr radius, a_B^*, to the V_{Pb}^{2-}. Isolated defects (dotted lines) are defined by having at least one Yb_{Pb}^{1+} beyond a_B^*.

In Section 3.4.1, we noted the existence of $(1|0)$ and $(0|-1)$ transitions close to the VBM and CBM, respectively, for the $(2Yb_{Pb} + V_{Pb})^0$ complex. To assess the degree of charge carrier localization by the complex, Figure 3.24 shows the difference in DFT electron densities in the plane spanned by the low energy OT $[110]^{\perp}$ and CL [110] configurations for both the $(1|0)$ and $(0|-1)$ charge transitions. To emulate optical excitations, the relaxed atomic positions and lattice parameters from the neutral state are fixed in the (± 1) charge states.

A hole added to the valence band predominantly localizes on Cl and Pb sites, indicative of the fact that the top of valence band of bulk $CsPbCl_3$ is primarily composed of Pb(s)-Cl(p) anti-bonding orbitals [114]. Furthermore, we find that the CL configuration shows marginally stronger charge localization on Cl sites closest to the Pb vacancy, while charge is more evenly distributed across the supercell for the OT configuration. Broadly speaking, the distribution of the added charge carrier reflects the symmetry of the defect configuration. This is particularly apparent in the electron density for the $(0|-1)$ charge transition, where the density distribution of an electron added to the conduction band among Pb p-states (*cf.* Figure 3.25) depends on the configuration of the complex. While the CL configuration shows several Pb sites in chains orthogonal to the complex with somewhat higher electron density, the reduced symmetry of the OT configuration leads to more localization on Pb sites closer to the Yb_{Pb}. Nonetheless, the charge carrier remains relatively delocalized across the supercell, regardless of the defect configuration. Using Bader charge analysis, we have also verified that this charge delocalization persists after allowing for lattice relaxation in the (± 1) charge states [101].

It is worth noting that similar differences in Pb-site electron density were reported by Li, *et al.*, in DFT calculations of OT $(2Yb_{Pb} + V_{Pb})^0$ defect complexes in cubic $CsPbCl_3$ [139]. Following their analysis, in Figure 3.25 we plot the site- and orbital-projected density of states (pDOS) near the conduction band edge for the p-states of Pb sites equidistant from both Yb_{Pb} in the neutral, fully bound defect motifs. Relative to bulk Pb sites, the OT configurations exhibit a higher intensity in the local pDOS near the band edge. The low energy OT $[110]^{\perp}$ configuration, in particular, features a broader band of higher pDOS close to the CBM, which may be attributable to the lower Yb-site symmetry noted in Section 3.4.2. In contrast, the CL configurations show a decreased intensity in their local Pb-site pDOS near the conduction band edge relative to the bulk. This raises the prospect that excited electrons could have a higher probability to populate conduction band states at Pb sites closer to an OT complex [139]. The comparably lower probability associated with CL complexes could provide a mechanism through which only certain local defect complex configurations

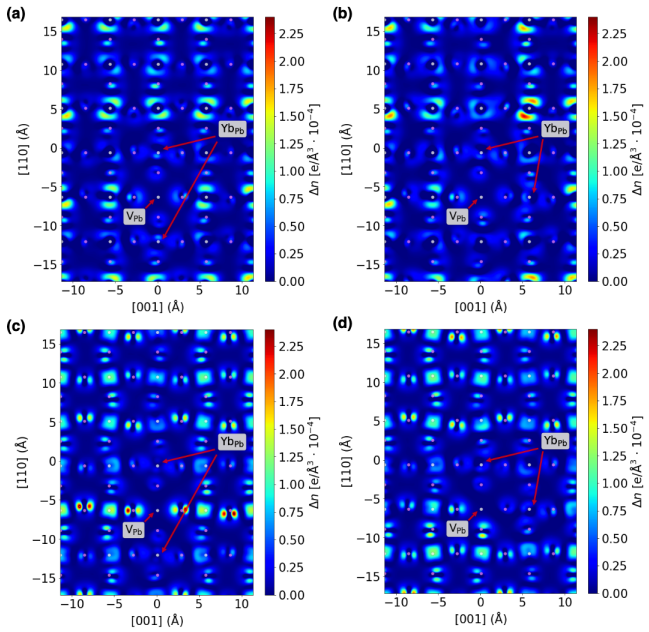


Figure 3.24: Electron density differences for $2Yb_{Pb} + V_{Pb}$ complex in the (a), (c) CL [110] and (b), (d) OT $[110]^{\perp}$ configurations. The top row (a), (b) shows the difference in electron density between the (-1) and neutral charge states associated with the $(0|-1)$ charge transition, while the bottom row (c), (d) shows the difference between the neutral and (+1) charge states associated with the (1|0) charge transition. Density differences are displayed in the plane spanned by the defect complex, and locations of nominal Pb (grey dots) and Cl (purple dots) sites in the undistorted, cubic perovskite phase are provided for spatial reference.

dominate quantum cutting. However, in self-consistently accounting for electronic relaxation in calculating the DFT electron densities, Figure 3.24 reveals a more complicated picture: Pb sites closer to the Yb_{Pb} show somewhat higher electron density, but only marginally so.

We find that the Yb_{Pb} appear to play a secondary role in determining how charge carriers redistribute in the system. Specifically, only a small fraction of charge added to the neutral system ends up on Yb sites or their immediate environment. Moreover, Bader charge analysis reveals that both Yb ions in the neutral complex, regardless of the configuration, have a

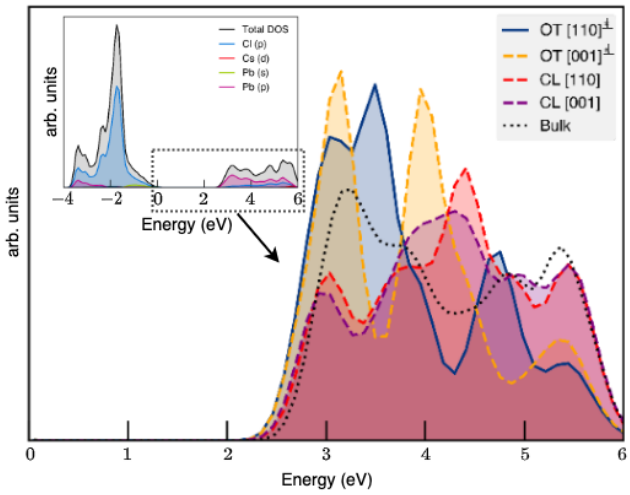


Figure 3.25: Projected density of states near conduction band edge for p-orbitals of Pb sites equidistant from both Yb_{Pb} in $(2Yb_{Pb} + V_{Pb})^0$ defect supercells. (Inset: Total and projected density of states for bulk $CsPbCl_3$.)

nominal (+2) Bader valence [3] due to charge compensation by the Pb vacancy, and this valence state does not appreciably change upon adding a charge carrier (either electron or hole) to the supercell [101]. This is somewhat expected given the shallow nature of the complex and the constituent point defects (Figure 3.18 and Table 3.13), which primarily act as perturbations to the host band states.

Finite-size limits on the DFT supercell make it difficult to definitively resolve spatial features of delocalized charge carriers in shallow defect states in the thermodynamic limit [73]. Moreover, the above analysis is limited by the fact that standard Kohn-Sham DFT is fundamentally a theory of electronic ground states, rather than excited state quasiparticles. Unfortunately, direct excited state calculations of defects from first-principles are often prohibitively computationally expensive. Given that the Pb vacancy acts as a doubly

[3]The Bader valence for a given lattice atom is defined here as the difference between the number of valence electrons in the isolated atom and the Bader (electron) charge within the atom-centered Bader volume.

charged acceptor, while the Yb_{Pb} are only singly charged, the Pb vacancy should provide the dominant contribution to the binding of an exciton. Thus, as an alternative route toward characterizing the role of the defect complex in localizing a photogenerated exciton, we consider an effective model of exciton binding to a V_{Pb}^{2-}, via the following three-body Hamiltonian in the Born-Oppenheimer adiabatic approximation,

$$\hat{H} = -\frac{1}{2}\nabla_1^2 - \frac{1}{2\sigma}\nabla_2^2 - \frac{q_d}{r_1} + \frac{q_d}{r_2} + V(r_{12}) \ , \tag{3.28}$$

with energies $(m_e^* m_0 e^4/(\epsilon_0\epsilon_s\hbar)^2)$ and lengths $(\epsilon_0\epsilon_s\hbar^2/(m_e^* m_0 e^2))$ rescaled to atomic units [241]. The ratio of hole and electron effective masses $\sigma = m_h^*/m_e^*$ is a free parameter in the model. The relative distances of a conduction band electron and a valence band hole to the defect center with nominal charge state $q_d = -2$ are denoted r_1 and r_2, respectively, while the electron and hole effective masses $m_{e,h}^*$ were determined from the DFT band structure. The static dielectric constant ϵ_s and the high-frequency, optical dielectric constant ϵ_∞ of the material were determined using density functional perturbation theory (Table 3.3).

We have implicitly assumed that defect-charge carrier Coulomb interactions are screened by the static dielectric constant ϵ_s, while the electron-hole interaction $V(r_{12})$ is allowed to take a more general form [230]. In particular, the softness and ionic nature of the perovskite lattice and the large difference between the static and high-frequency dielectric constants implies strong Fröhlich coupling $\alpha_{e,h} = (e^2/\hbar\epsilon_*)\sqrt{m_{e,h}^*/2E_{\rm LO}}$ between charge carriers and longitudinal optical (LO) phonons, where $1/\epsilon_* = 1/\epsilon_\infty - 1/\epsilon_s$ and $E_{\rm LO}$ is a characteristic LO excitation energy (see below). The associated formation of large polarons has been previously invoked as a primary source of disagreement between measurements of the exciton binding energy in metal-halide perovskites [12, 166, 167, 219, 173, 209, 29, 253]. To account for the possible relevance of polaron formation, we consider the effective polaron interaction potential derived by Haken (H),

$$V_{\rm H}(r) = -\frac{e^2}{\epsilon_s r} - \frac{e^2}{2\epsilon_* r}\left(\mathrm{e}^{-r/l_h} + \mathrm{e}^{-r/l_e}\right) \ , \tag{3.29}$$

where electron- and hole-polaron radii are given by $l_{e,h} = \sqrt{\hbar^2/2m^*_{e,h}E_{\mathrm{LO}}}$, and the kinetic energy terms in the Hamiltonian (3.28) contain the renormalized effective masses $\tilde{m}^*_{e,h}$ [91, 90, 68]

$$\tilde{m}^*_{e,h} = m^*_{e,h}\left(1 + \frac{\alpha_{e,h}}{6}\right) \ . \tag{3.30}$$

Previous studies have noted that the Haken model often overestimates the exciton binding energy in certain polar semiconductors and insulators [10, 230, 12]. Thus, we also consider the phenomenological model proposed by Pollmann and Büttner (PB),

$$V_{\mathrm{PB}}(r) = -\frac{e^2}{\epsilon_s r} - \frac{e^2}{\epsilon_* r}\left(\frac{m^*_h}{\Delta m^*}\mathrm{e}^{-r/l_h} - \frac{m^*_e}{\Delta m^*}\mathrm{e}^{-r/l_e}\right) \ , \tag{3.31}$$

where $\Delta m^* = m^*_h - m^*_e$, and the bare electron and hole effective masses $m^*_{e,h}$ are used in the kinetic energy terms in the Hamiltonian (3.28) [204]. At short range, both models behave as a Coulomb potential with weaker screening by the optical dielectric constant ϵ_∞, while at long range the interaction is strongly screened by the static dielectric constant ϵ_s. From a strictly mathematical point of view, the Pollmann-Büttner model has the advantage of interpolating between the extreme cases of $\lim_{l_{e,h}\to 0} V(r_{12}) \propto 1/\epsilon_s r_{12}$ and $\lim_{l_{e,h}\to\infty} V(r_{12}) \propto 1/\epsilon_\infty r_{12}$, depending on the length scales $l_{e,h}$.

We take a variational approach to approximating the ground state of the effective Hamiltonian (3.28) using a wave function ansatz $|\Psi\rangle$ expanded in the following s-wave basis $|\psi_i\rangle$:

$$|\Psi\rangle = \sum_i c_i|\psi_i\rangle \tag{3.32}$$

$$|\psi_i\rangle = \exp\left(-\alpha_i r_1 - \beta_i r_2 - \gamma_i r_{12}\right) \tag{3.33}$$

Prior works have shown that this class of exponentially correlated variational ansatz can yield highly accurate ground state predictions on few-body quantum systems with a relatively small number of basis functions [228, 230, 93, 246, 92]. The variational parameters $(\alpha_i, \beta_i, \gamma_i, c_i)$ were optimized according to a two-step procedure discussed in Section 3.4.5.

Table 3.3: Electron and hole effective masses, $m^*_{e,h}$, and static and high-frequency dielectric constants, ϵ_s, ϵ_∞, in cubic and orthorhombic phases of $CsPbCl_3$ from density functional perturbation theory. The exciton binding energies, E_X, and effective Bohr radii, a^*_B, are predicted from the Wannier-Mott model (3.34), (3.35). The variational ground state energy, E_{GS}, of the defect-bound exciton relative to E_X and the expected exciton radius $\langle r_{12}\rangle$ are determined by the same weakly screened electron-hole potential $V(r_{12}) \propto 1/\epsilon_\infty r_{12}$. Approximate phonon screening corrections, ΔE_X, accounting for (neglecting) parabolic electron band dispersion [69] are also shown.

Phase	m^*_h	m^*_e	ϵ_∞	ϵ_s	E_X [meV]	a^*_B [nm]	ΔE_X [meV]	$E_{GS} - E_X$ [meV]	$\langle r_{12}\rangle$ [nm]
$Pm\bar{3}m$ (cubic)	0.17	0.18	4.08	20.56	71.7	2.46	-17.1 (-30.6)	63.7	2.08
$Pnma$ (ortho)	0.25	0.26	4.09	23.67	103.0	1.70	-18.6 (-34.4)	78.9	1.19

The binding energy E_X for a Wannier-Mott exciton in $CsPbCl_3$ provides a natural reference point for the variational ground state energy E_{GS} of the defect-bound exciton. We use a standard hydrogenic model to obtain a baseline estimate for E_X, along with the effective exciton Bohr radius a^*_B,

$$E_X = \frac{m_r}{m_0}\left(\frac{\epsilon_0}{\epsilon_r}\right)^2 E_{Ry} \ , \tag{3.34}$$

$$a^*_B = \frac{m_0}{m_r}\left(\frac{\epsilon_r}{\epsilon_0}\right) a_B \ , \tag{3.35}$$

where the reduced mass of the exciton is $m_r = (1/m^*_e + 1/m^*_h)^{-1}$, and we take the relative dielectric constant ϵ_r to be equal to the high-frequency dielectric constant ϵ_∞ (Table 3.3) [207]. With this choice of dielectric constant along with the effective masses from DFT, we predict an exciton binding energy of ≈ 72 meV which agrees well with the experimental values of 72 ± 3 meV reported by Zhang, *et al.*, [290] and 64 ± 1.5 meV reported by Baranowski, *et al.*, in $CsPbCl_3$ [13, 75, 111]. The associated exciton Bohr radius (2.46 nm) is used in the analysis of Section 3.4.3.

To provide a consistent point of comparison to the variational ground state of the Hamil-

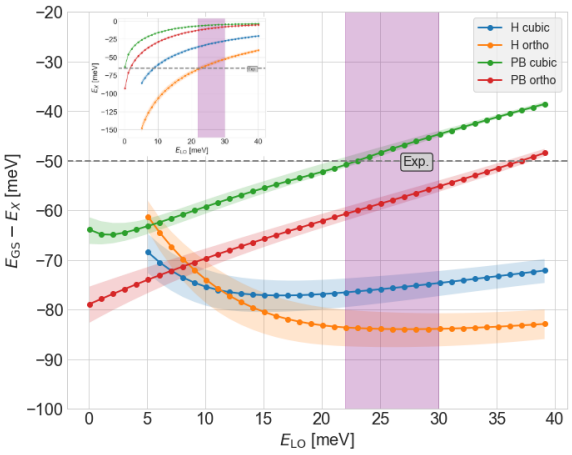


Figure 3.26: Variational ground state energy E_{GS} of polaronic exciton bound to double acceptor A^{2-} with respect to polaronic exciton binding energy E_X (Inset) as a function of characteristic phonon energy E_{LO}. Polaronic effects are incorporated either by the Haken (H) model or the Pollmann-Büttner (PB) model for both the cubic and orthorhombic perovskite phases, with additional uncertainty in the effective mass ratio included. Shaded, vertical region highlights the predicted range of E_{LO} from first-principles calculations for $CsPbCl_3$. Current experimental values for $E_{\mathrm{GS}} - E_X$ from shallow trap emission and for E_X from magneto-optical response (grey, dashed lines) are also included [171, 13].

tonian (3.28) including polaronic effects, we also consider the more general calculation of the exciton binding energy given by the Hamiltonian

$$\hat{H}_X = -\frac{1}{2}\nabla^2 + V(r) \ , \tag{3.36}$$

where the electron-hole interaction potential is again taken to be either the Haken (3.29) or Pollmann-Büttner (3.31) model. The atomic units in the Hamiltonian (3.36) are set by the reduced exciton mass m_r, determined either by the renormalized polaron masses (3.30) for the Haken potential or the bare effective masses for the Pollmann-Büttner potential. We use the following variational ansatz,

$$|\Psi_X\rangle = \frac{\gamma^{3/2}}{\sqrt{\pi}} \mathrm{e}^{-\gamma r} \ , \tag{3.37}$$

to obtain the exciton binding energy $E_X \approx \min_\gamma \langle \Psi_X | \hat{H}_X | \Psi_X \rangle$.

In Figure 3.26, we compare the difference between the ground state energy of an exciton bound to a double acceptor E_{GS} and the exciton binding energy E_X described by either the Haken (3.29) or Pollmann-Büttner (3.31) models and for various choices of physical model parameters. Since the effective masses and static dielectric constant are sensitive to the structure of the $CsPbCl_3$ lattice (Table 3.3), we consider both sets of parameters determined by DFT for the high-temperature cubic $Pm\bar{3}m$ phase and the low-temperature orthorhombic $Pnma$ phase. Furthermore, since the effective band masses derived from DFT may vary based on the choice of semilocal or hybrid functionals and spin-orbit coupling, we allow for an uncertainty of ± 0.05 in the ratio $\sigma = m_h^*/m_e^*$ of effective masses commensurate with variations reported in the literature [219, 209, 165]. In Figure 3.26, we treat the effective LO phonon energy E_{LO} as a free parameter. This is done not only to exhibit the dependence of $E_{GS} - E_X$ on the LO phonon spectra, but also to show the limit $l_{e,h} \to \infty$ of the Pollmann-Büttner model, tantamount to an electron-hole Coulomb interaction weakly screened by the optical dielectric constant ϵ_∞. Of course, this limit is unphysical from the standpoint of the underlying continuum polaron theory [4], which assumes weak to intermediate Fröhlich couplings $\alpha_{e,h} < 6$, but it provides a useful lower bound on the defect-bound exciton ground state as a function of the choice of electron-hole interaction model [91, 90, 68, 204]. Specifically, in terms of formulating the effective Hamiltonian (3.28), this limit corresponds to short-timescale electron-hole interactions against an effectively frozen lattice. We estimate the LO phonon energy $E_{\mathrm{LO}} = \hbar\tilde{\omega}_{\mathrm{LO}} \approx 26$ meV by taking the LO frequency $\tilde{\omega}_{\mathrm{LO}}$ to be the location of the peak of the loss function $\mathrm{Im}(-\epsilon(\omega)^{-1})$ in the mid-infrared [219, 173, 99]. This is close to the values $\approx 26.5 \pm 1.5$ meV found by Carabatos-Nédelec, *et al.*, for $CsPbCl_3$ [35] and ≈ 28 meV found by Sender, *et al.*, for $CH_3NH_3PbCl_3$ [219]. This apparent independence to the

[4] Since the kinetic energy terms in the Haken model contain the renormalized polaron masses (3.30), their curves in Figure 3.26 are cut off before the strong coupling regime.

identity of the A-site cation is to be expected since LO phonons in ABX_3 halide perovskites within this frequency range correspond to Pb-X stretching modes [219, 173, 155]. In the orthorhombic *Pnma* phase, this predicted value of E_{LO} corresponds to dimensionless Fröhlich coupling constants $\alpha_e = 2.37$ and $\alpha_h = 2.30$, while the renormalized, reduced exciton-polaron mass, $\tilde{m}_r = 0.18$, compares well with value of 0.20 ± 0.01 reported by Baranowski, *et al.* [13]. The region of the predicted LO phonon energy E_{LO}, allowing for uncertainty on the order of the optical phonon bandwidth, is included in Figure 3.26.

Milstein, *et al.*, have reported shallow trap emission at $\approx$ 50 meV below the excitonic PL associated with a RE^{3+} dopant-induced defect state [171]. We find that the Pollmann-Büttner model parameterized by the effective masses and dielectric constants in the cubic $Pm\bar{3}m$ phase predicts $E_{GS} - E_X$ to be within 10 meV to this experimental value for most values of the effective LO phonon energy, with particularly good agreement within the *ab initio* range of E_{LO}. When using the heavier effective masses for the orthorhombic *Pnma* phase, the exciton is more tightly bound to the defect by an energy difference $\approx$ 10 meV lower than the cubic phase. However, for both parameter sets, the Pollmann-Büttner model predicts a very weak exciton binding energy $E_X \approx$ 5-15 meV for the *ab initio* range of E_{LO} (Inset to Figure 3.26) compared to the reported experimental values [13, 75, 111, 290].

The Haken model, with parameters set to the orthorhombic phase, shows better agreement to the reported experimental exciton binding energies in this range. Menéndez-Proupin, *et al.*, after correcting a problem in their calculation, made a similar observation that the Haken model better predicted the experimental exciton binding energy in $CH_3NH_3PbI_3$ [166, 167]. Nonetheless, we find that the Haken model, as expected, predicts a stronger defect localization energy with $E_{GS} - E_X$ on the order of 80 meV. Finally, the case where $V(r_{12}) \propto 1/\epsilon_\infty r_{12}$ in the Hamiltonian (3.28) and Wannier-Mott exciton binding (3.34) shows decent agreement to both the doping-induced defect state PL and the experimental exciton binding energy for the cubic phase, while stronger binding is again found for the orthorhombic phase (Table 3.3). In this case, we also calculate the expectation value of the radius $\langle r_{12} \rangle$ for the defect-bound exciton (Table 3.3), verifying for both the cubic and orthorhombic

phases that, while the spatial extent of the exciton is reduced, it remains sufficiently delocalized to simultaneously couple to both Yb_{Pb}^{1+} in a local configuration of the fully compensated complex. As this model possesses the lowest degree of screening among those considered, it provides a characterization of the exciton length scale in the maximally localized case.

In light of the above discussion, let us make some additional comments on interpreting these results. First, to simplify the variational problem defined by (3.28), we have neglected the role of the Yb_{Pb}^{1+} pair in exciton localization. It is reasonable to expect, based on electrostatics, that the addition of Yb_{Pb}^{1+} will increase the defect-bound exciton ground state energy and that the OT configuration of the fully compensated complex should have marginally stronger binding relative to the CL configuration. Thus, our model predictions should provide an approximate lower bound on the ground state energy of the exciton bound to the full defect complex, and this is generally what we find in comparing to the shallow trap emission reported by Milstein, *et al.* [171].

Whether polaronic effects are fundamental to the interpretation of the experimental data remains somewhat of an open question. We found, for instance, that a simple Coulombic electron-hole interaction screened by the optical dielectric constant seems to account for the essential contributions to both the experimental localization energy and the exciton binding energy. However, justification of this model relies on an accurate assessment of the relevant time scales involved in the experimental technique and the exciton lifetime. At very short times ($\lesssim$ 200 fs), the exciton should only be subject to screening by electronic degrees of freedom, while screening by lattice polarization becomes relevant at longer times [165, 249, 42, 167, 13, 76]. Indeed, Miyata, *et al.*, found that large polaron formation occurred on a time scale of 0.7 ps in $CsPbBr_3$ [173]. In contrast, energy transfer from the exciton to the RE^{3+} dopant-induced defect state occurs on a time scale of a few picoseconds, while energy transfer to the Yb pair occurs on the order of a few nanoseconds, suggesting that ionic contributions to the dielectric screening should be considered [171, 127, 212].

It has been argued that transmission measurements based on high-field magneto-optical response, for which the high cyclotron frequency is expected to decouple charge carrier motion

from lattice polarization, can lead to overestimations of the experimental exciton binding energy by factors of 3-4 for methylammonium lead halide perovskites $CH_3NH_3PbX_3$ [76, 172, 249, 32]. Similar behavior can be expected for the class of inorganic lead halide perovskites $CsPbX_3$, as the electronic structure and dielectric screening are largely determined by the $[PbX_3]^{1-}$ sublattice [173, 279]. However, the magneto-optical measurements of the exciton binding energy reported by Baranowski, *et al.*, [13] should be robust against this error since the authors explicitly extrapolate to the zero-field limit, following the earlier studies [172, 279]. Their reported values are further substantiated by the similar values obtained by Zhang, *et al.*, [290] using absorption spectroscopy.

Recent theoretical work by Filip, *et al.*, [69] found that LO phonon screening could account for a significant fraction of the discrepancy between experimental exciton binding energies in orthorhombic $CsPbX_3$ and the overestimated values predicted by the *ab initio* Bethe-Salpeter equation (BSE), which traditionally included only electronic screening effects [29, 253]. Moreover, they established that an isotropic, parabolic band approximation was sufficient to capture the lowest-order phonon screening contribution ΔE_X to the BSE exciton binding energy due to Frölich-type electron-phonon interactions, leading to the analytic expression

$$\Delta E_X = -2E_{\mathrm{LO}} \left(1 - \frac{\epsilon_\infty}{\epsilon_s}\right) \frac{3 + \sqrt{1 + \frac{E_{\mathrm{LO}}}{E_X}}}{\left(1 + \sqrt{1 + \frac{E_{\mathrm{LO}}}{E_X}}\right)^3} \,. \tag{3.38}$$

By ignoring the electronic band dispersion entirely [69], this correction further simplifies to

$$\Delta E_X = -2 \left(1 - \frac{\epsilon_\infty}{\epsilon_s}\right) \frac{E_{\mathrm{X}} E_{\mathrm{LO}}}{E_X + E_{\mathrm{LO}}} \,. \tag{3.39}$$

For the *Pnma* phase (Table 3.3), we find that the correction (3.38) brings the Wannier-Mott value (3.34) for the exciton binding energy to within 10 meV of the value reported by Zhang, *et al.*, [290]. Furthermore, it is interesting to note that the dispersionless correction (3.39) yields close agreement to the exciton binding energy predicted by the Haken model ($E_X^{(\mathrm{H})} \approx 58$ meV). This correction (3.39) is also commensurate with the shift in E_{GS} predicted

by the Haken model relative to the electronically screened $V(r_{12}) \propto 1/\epsilon_\infty r_{12}$ model in (3.28). Thus, it is entirely plausible that phonon screening provides a significant contribution to the defect-bound exciton ground state, as well as to the stability of the free exciton. While we find that there is relatively little difference between the Haken model prediction of $E_{\mathrm{GS}} - E_X$ compared to the electronically screened model, a lower exciton binding energy would imply a higher rate of thermal dissociation and could entail greater competition from alternative pathways for carrier recombination. This may be reasonable given that Roh, *et al.*, did not observe excitonic PL in their SC sample over a broad range of temperatures and suggests that upstream energy losses are reduced in NCs, which exhibited higher PLQYs [212]. Finally, our finding that charge compensation increases the total concentration of $\mathrm{V}_{\mathrm{Pb}}^{2-}$ with increasing Yb incorporation (Section 3.4.3) would then explain why Milstein, *et al.*, do not observe shallow trap emission without RE^{3+} doping [171].

3.4.5 Additional Computational Details

All first-principles calculations in this work were carried out using the Vienna Ab-Initio Simulation Package (VASP) [125, 124]. Plane-wave, Kohn-Sham density functional theory calculations employed the all-electron projector-augmented wave (PAW) method in the generalized gradient approximation (GGA) with the semilocal PBEsol functional [121, 28, 198]. In the pseudopotentials, 9 electrons of Cs ($5s^2 5p^6 6s^1$), 4 electrons of Pb ($6s^2 6p^2$), and 7 electrons of Cl ($3s^2 3p^5$) were treated as valence electrons, while 13 f-state electrons in the [Xe]$4f^{14}6s^2$ valence configuration of Yb were treated as frozen core states. Blocked Davidson iteration was used to optimize the electronic degrees of freedom. A plane-wave basis cut-off was set to 400 eV, and a Gamma-centered k-point mesh was used for Brillouin-zone integration [176]. Only the Gamma-point was included in the 360 site defect supercell calculations ($3 \times 3 \times 2$ supercells of the 20 atom *Pnma* unit cell), while a $4 \times 4 \times 2$ k-point mesh was used for *Pnma* unit cell calculations. The conjugate gradient algorithm was used for structural optimization, and the volume, shape, and atomic positions were relaxed until interatomic Hellmann-Feynman forces were less than 0.01 eV/Å. Because Hellmann-Feynman forces are

derived from the occupied valence states of the system and Pb p-states primarily contribute to the conduction band, we ignored spin-orbit coupling in relaxing atomic structures. However, spin-orbit coupling was included in deriving bulk effective masses and dielectric constants. Molecular dynamics (MD) trajectories were calculated with VASP using the Nosé-Hoover thermostat and NVT ensemble for 100 time steps (2 fs each) at a temperature of 300 K. PDFs were generated by convolving the distribution of pairwise distances with a Gaussian distribution with standard deviation of 0.5 Å and averaging over the MD trajectories.

Corrections to defect formation energies were computed with the help of the PyCDT and Pymatgen programs with suitable modifications [31, 191]. Bulk effective masses were derived from the computed band structures of the cubic $Pm\bar{3}m$ and the orthorhombic $Pnma$ perovskite phases of $CsPbCl_3$ in the parabolic band approximation using weighted least-squares fitting as implemented in the Effmass program [274]. Additional normal mode and phonon spectra analysis utilized the Phonopy program [250]. Visualizations of atomic structures were generated using the VESTA program [174].

Variational ground states to the effective Hamiltonian (3.28) were determined according to the following procedure. The variational optimization problem with wave function anzatz (3.32) can be formulated as a generalized eigenvalue problem,

$$H\mathbf{c} = ES\mathbf{c} \ , \tag{3.40}$$

where $\mathbf{c}$ is the eigenvector of linear variational coefficients c_i in (3.32) associated with the minimal eigenvalue E. Hamiltonian matrix elements H and overlap integrals S,

$$H_{ij} = \langle \psi_i | \hat{H} | \psi_j \rangle \ , \tag{3.41}$$

$$S_{ij} = \langle \psi_i | \psi_j \rangle \ , \tag{3.42}$$

can be computed analytically for the class of exponentially correlated wave functions (3.33) by taking repeated derivatives of the fundamental Calais-Löwdin integral,

$$\Gamma_{0,0,0}(\alpha,\beta,\gamma) = \frac{2}{(\alpha+\beta)(\alpha+\gamma)(\beta+\gamma)} \ , \tag{3.43}$$

with respect to the nonlinear variational coefficients (α,β,γ) [33, 92]. An integral of order $l,m,n \geq 0$, defined as

$$\begin{aligned}&\Gamma_{l,m,n}(\alpha,\beta,\gamma) \\ &\quad = \frac{1}{8\pi^2}\int d\mathbf{r}_1 d\mathbf{r}_2 r_1^{l-1} r_2^{m-1} r_{12}^{n-1} e^{-\alpha r_1 - \beta r_2 - \gamma r_{12}} \ ,\end{aligned} \tag{3.44}$$

is thus given by

$$\Gamma_{l,m,n}(\alpha,\beta,\gamma) = 2l!m!n! \sum_{l'=0}^{l}\sum_{m'=0}^{m}\sum_{n'=0}^{n} \frac{\binom{m-m'+l'}{l'}\binom{l-l'+n'}{n'}\binom{n-n'+m'}{m'}}{(\alpha+\beta)^{m-m'+l'+1}(\alpha+\gamma)^{l-l'+n'+1}(\alpha+\beta)^{n-n'+m'+1}} \ . \tag{3.45}$$

The optimization of linear and nonlinear variational parameters were handled self-consistently by the repeated iteration of two fundamental steps. Initially, nonlinear variational parameters were chosen by quasirandom tempering within the finite intervals $\alpha_i^{(0)} \in [A_1, A_2]$, $\beta_i^{(0)} \in [B_1, B_2]$ and $\gamma_i^{(0)} \in [G_1, G_2]$,

$$\alpha_i^{(0)} = (A_2 - A_1)\left\{\frac{1}{2}i(i+1)\sqrt{p_1}\right\} + A_1 \tag{3.46}$$

$$\beta_i^{(0)} = (B_2 - B_1)\left\{\frac{1}{2}i(i+1)\sqrt{p_2}\right\} + B_1 \tag{3.47}$$

$$\gamma_i^{(0)} = (G_2 - G_1)\left\{\frac{1}{2}i(i+1)\sqrt{p_3}\right\} + G_1 \ , \tag{3.48}$$

where p_j are distinct prime numbers [246, 230, 240]. For a given set of nonlinear parameters, the linear coefficients c_i were then determined by solution to the generalized eigenvalue problem (3.40). For a given solution c_i, the nonlinear parameters $(\alpha_i, \beta_i, \gamma_i)$ were subsequently refined using the Multi-Level Single-Linkage (MLSL) global optimization algorithm with a

series of multi-start local optimizations using the Bound Optimization BY Quadratic Approximation (BOBYQA) algorithm, as implemented in the NLopt package [116, 117, 205, 115]. The finite bounds (A_i, B_i, G_i) on the nonlinear parameters were determined by finite grid search. Variational wave functions with up to 15 basis terms were found give sufficiently converged results.

3.5 Summary

In this chapter, we used coarse-grained lattice Hamiltonians, trained on DFT data, and Monte Carlo simulation to study halide ordering in mixed halide perovskites. For binary halide alloys where the constitutive halides occupy neighboring rows in the periodic table, we have located miscibility gaps well below room temperature. Nonetheless, we found that some degree of layered halide ordering persists at room temperature, which leads to particularly facile Br $\rightarrow$ I halide exchange when the Br fraction is less than 50 %. We then considered the effects of Mn-doping on Cl/Br mixing, where Monte Carlo simulation revealed room temperature phase separation into Mn- and Cl-rich domains. This was, furthermore, found to be consistent with experimental observations of Mn EPR-signal quenching under Cl $\rightarrow$ Br exchange [225]. These results suggest that even isovalent, substitutional cationic impurities interact nontrivially with the surrounding perovskite matrix atoms and may readily order at room temperature. Indeed, the observation of spinodal decomposition of Mn-ordered phases in pure halide material suggests a nontrivial role is played by the transfer of partial charge between naively electrostatically identical ions. Moreover, the apparent anion-dependence of cation distributions in halide binaries has potential ramifications for the effective tuning of the electronic and photophysical properties of doped perovskites.

The fortuitous energy matching between the band gap of $CsPb(Cl_{1-x}Br_x)_3$ and f-f transitions in Yb enables the surprisingly efficient sensitization of Yb luminescence via quantum cutting [170]. A steady accumulation of experimental evidence points to the role played by $(2Yb_{Pb} + V_{Pb})^0$ defect complexes in facilitating these quantum cutting processes, where a shallow doping-induced defect state localizes a photogenerated charge carrier and the spa-

tial proximity of pairs of Yb allows for simultaneous energy splitting and NIR emission of pairs of photons [171, 127, 212, 64, 48, 44, 170]. However, atomistic details of the defect structure, its expected concentration and its coupling to the surrounding perovskite material have remained largely undetermined. To that end, we have attempted to resolve some of these details using density functional theory and various effective models for defect formation and charge carrier localization. We found, for instance, that $Yb_{Pb} + V_{Pb}$ complexes and their constituent point defects introduce only shallow defect levels into the electronic structure of the material. While Cl vacancies form deeper amphoteric defect states, their predicted concentration is limited under the experimentally relevant Cs-rich, moderately Pb-poor synthesis conditions. Increasing Yb incorporation under these conditions was also found to increase the concentration of Pb vacancies and $Yb_{Pb} + V_{Pb}$ complexes via charge compensation. Together, these results are consistent with the hypothesis that high PLQYs in Yb:$CsPbCl_3$ are driven by $(2Yb_{Pb} + V_{Pb})^0$ defect-mediated quantum cutting [171].

We observed that various configurations of the $(2Yb_{Pb} + V_{Pb})^0$ defect complex possess favorable binding energy, where orthogonal motifs are marginally more stable than collinear ones. Screened electrostatic interactions account for most of the complex binding energy, while patterns of local bonding and induced strain provide additional contributions to the small energetic differences between local motifs. We have further characterized, on the basis of a self-consistent thermodynamic model of defect concentrations, how both charge and strain compensation between V_{Pb}^{2-} and Yb_{Pb}^{1+} limits the change in lattice volume under Yb-doping, which may account for the lack of appreciable shifts in XRD data between undoped and Yb-doped $CsPbCl_3$ [171, 127, 212, 48, 44, 170]. Using this thermodynamic model and Monte Carlo sampling, we approximated the room-temperature spatial distribution of Yb_{Pb}^{1+} relative to Pb vacancies as a function of material composition. While the fraction of Yb in local configurations of $2Yb_{Pb} + V_{Pb}$ grows monotonically with either increasing Yb incorporation or increasing Cs:Pb composition, a significant fraction of Yb exists in partially dissociated configurations. Importantly, while these local configurations are conjectured to be the active quantum cutters in the system, TRPL measurements in Yb:$CsPbCl_3$ suggest

that they are not the dominant Yb defect species [212, 64].

Finally, we noted that orthogonal $(2\mathrm{Yb_{Pb}} + \mathrm{V_{Pb}})^0$ defect motifs, compared to collinear motifs, showed marginally stronger Pb-site localization of the DFT electron density associated with the $(0|-1)$ charge transition level. Conversely, collinear motifs exhibited marginally stronger charge localization on Cl sites near the Pb vacancy for the $(1|0)$ transition, which fills a valence band hole. Nonetheless, charge remained essentially delocalized in these supercell calculations, and given the potential relevance of phonon screening and large polaron formation, we resorted to an effective few-body Hamiltonian to scrutinize the role of Pb vacancies in localizing photogenerated charge carriers in the system. Variational ground state calculations predicted shallow defect-bound exciton states with energies lower than the free exciton binding energy in qualitative agreement with the experimentally observed $\mathrm{RE^{3+}}$ dopant-induced shallow trap emission [171, 127, 212]. However, we found notable quantitative disagreements between choices of effective electron-hole interaction with electronic structure parameters determined from DFT. While the inclusion of polaronic effects via the Haken model predicts relatively strong binding to shallow, double acceptor states, it also provides, among the considered models, the best approximation to the experimental exciton binding energy in the low-temperature *Pnma* phase. Reconciling the predicted localization energy of the defect-bound exciton with the value measured from $\mathrm{RE^{3+}}$ dopant-induced shallow trap emission likely requires a full treatment of the entire $2\mathrm{Yb_{Pb}} + \mathrm{V_{Pb}}$ defect complex.

Chapter 4

MACHINE LEARNING TENSOR NETWORKS FOR MATERIALS PROPERTY PREDICTION

Rapid advances in electronic structure methods and computational resources have enabled high-throughput *ab initio* calculations of solids and molecules over broad classes of chemistries and structures. Much of this work has been motivated by a pressing need to expand the tools for engineering new materials with desirable properties, with applications ranging from drug design to optoelectronics, energy storage, and quantum computing. First-principles calculations, in particular, play a crucial role in mapping the atomistic structure and composition of a material to fundamental properties such as ground state energies, potential energy surfaces, band structures and optical excitation spectra. However, a full exploration of materials design space is plagued by various well-known curses of dimensionality. These include the exponential scaling of many-body Hilbert spaces and the associated computational complexity of solving the many-body Schrödinger equation, as well as the combinatorics of decorating finite and periodic lattices by various chemical species. While density functional theory (DFT) and many-body perturbation theory (MBPT) have seen numerous successes in addressing the former, they can still present large computational barriers to address the latter.

In recognition of this challenge, the development and application of machine learning techniques to produce accurate and computationally efficient surrogate models of *ab initio* calculations has become a very active area of research [21, 14, 248, 222, 108, 221, 59, 277, 22, 186, 239, 130]. Broadly speaking, the success of these techniques depends on the identification of a sufficiently descriptive feature space which captures the variance of the data one wishes to model. Much progress has been made recently in identifying and engineering

feature spaces which accurately represent local atomic environments and can be used as inputs to standard machine learning algorithms. Efficient representations are often achieved by constructions that respect physical symmetries, such as invariance to global rotations, translations and permutations of identical chemical species. For the class of atom-centered representations based on expansions of atomic densities, symmetry is incorporated either by directly constructing invariant polynomials, as is the case for atom-centered symmetry functions (ACSFs) [21] and moment tensor potentials (MTPs) [222], or by explicit integration over relevant symmetry groups, as exemplified by the smooth overlap of atomic positions (SOAP) [14] and spectral neighbor analysis potentials (SNAP) [248]. To a large extent, instances in this class of atom-centered representations correspond to different choices of basis sets in an abstract N-body Hilbert space [277], and the hierarchy of N-body features has recently been shown to be organized in the so-called atomic cluster expansion (ACE) framework [59, 60, 9, 150].

It is worth mentioning that an alternative to extensive feature engineering is to employ a so-called *end-to-end* approach, in which inputs to the model are minimally processed, and the relevant feature space is learned by the model architecture during the course of training. Examples of this approach in materials informatics include SchNET and Euclidean neural networks and graph network methods such as Crystal Graph Convolutional Neural Networks (CGCNN), MEGNET, and DIMENET. However, in the absence of prior feature engineering, training accurate deep learning models becomes difficult when the amount of data is limited.

Thus, from a practical standpoint, balancing the bias and complexity of a model between the two extremes of extensive feature engineering and highly flexible model architectures is often influenced by the availability of training data. Strictly end-to-end deep learning can require substantially more data when correlations between features are complex, while naive application of parametrically efficient features can lead to a model which generalizes poorly on new inputs. Of course, the strict distinction between these two extremes can be somewhat arbitrary: the latent space encoded by the hidden layers of a deep learning architecture can be thought of as a kind of renormalized feature space, subject to its own forms of bias

through choices of hyperparameters, regularization schemes and architectural topology. To that end, certain deep model architectures may possess better so-called *inductive bias* [15] over a given dataset, more efficiently prioritizing the search space of solutions when their structure reflects the pattern of correlations in the feature space.

Interesting connections between inductive bias and feature correlations have grown out of recent efforts to introduce techniques from the study of quantum entanglement and strongly correlated quantum many-body systems to the field of machine learning. Specifically, tensor network methods have been successfully applied to supervised and unsupervised learning tasks in computer vision and natural language processing and have provided a basis for analyzing the expressiveness of common deep learning architectures based on their entanglement properties. Indeed, a general argument for the effectiveness of tensor networks in machine learning contexts is that the pattern of entanglement encoded in the network can efficiently represent the pattern of correlation in the feature space of the data.

In this chapter, we argue for the applicability of tensor networks in machine learning structure-property models of materials. This problem is addressed from two directions: we show how both the construction of an input feature space and of an associated machine learning architecture can be formulated in the language of tensor networks. In Section 4.1, a large class of atomic structure representations, corresponding to the SO(3)-invariant tensor basis set of the (smoothed) atomic cluster expansion (ACE), are shown to admit a natural tensor network description. An immediate consequence of this rewriting is that the equivalence classes and the recursive construction of the hierarchy of N-body ACE basis tensors become transparent in the graphical notation of the tensor network. In Section 4.2, we show how common tensor network factorizations for the weights of a machine learning model can be naturally built on top of individual ACE basis tensors. We discuss the relationship between this approach and kernel learning, and in particular how certain tensor networks can realize a form of *alchemical learning* by coupling information between the local atomic structure and the chemical elements within it. We also comment on how the introduction of copying and merging operations in tensor network structures can be used to introduce

higher-order correlations from fixed N-body features, akin to the construction of nonlinear kernels.

In Section 4.3, we benchmark some of the proposals from Sections 4.1 and 4.2 using a widely available dataset of density functional theory calculations of band gaps and formation energies of transparent conducting oxide alloys. For concreteness, we use input features corresponding to the commonly used SOAP power spectrum and study the learning performance of different tensor network factorizations of the model weights on training sets of various sizes. Compared to standard kernel learning methods and fully-connected neural networks, we find that models based on matrix product states (MPS) and matrix product operators (MPO) show strong generalizability and parametric efficiency, with notable performance on small training sets. We subsequently study the entanglement properties of the trained networks and find signatures of high entanglement complexity consistent with the models' strong generalizability. We also provide some evidence that the latent spaces learned by the hidden layers of the networks are able to capture physically relevant structural and chemical information, and we utilize this insight to effectively compress the input basis tensors.

4.1 SO(3)-Invariant Atomic Representations

In this work, we are concerned with the prediction some atomic property V given a configuration $\{\mathbf{r}_i\}$ of N atomic species. For simplicity, we will restrict to the case where V is a scalar, such as the total energy or band gap of the system, but extensions to vectorial and tensorial properties are discussed in Appendix 4.4.4. We begin, following the discussions and derivations in [59, 150], by formulating of a coarse-grained model given by a sum of local terms, $V = V^{(0)} + \sum_i V_i$, where the local property V_i associated with atom i can be approximated by expanding in terms of local n-body interactions $V^{(n)}$:

$$V_i = V^{(1)}(\mathbf{r}_i) + \frac{1}{2}\sum_j V^{(2)}(\mathbf{r}_i, \mathbf{r}_j) + \frac{1}{3!}\sum_{jk} V^{(3)}(\mathbf{r}_i, \mathbf{r}_j, \mathbf{r}_k) + \cdots \tag{4.1}$$

Requiring consistency with fundamental symmetries introduces constraints on the form of the interactions. For instance, translation invariance leads to atom-centered functions of the form $V^{(\nu+1)}(\mathbf{r}_{j_1 i}, \ldots, \mathbf{r}_{j_\nu i})$, where $\mathbf{r}_{ji} = \mathbf{r}_j - \mathbf{r}_i$ can be thought of as an effective bond from the central site i to a site j in its relative environment. Interaction terms can be further decomposed by projecting onto an appropriate ν-order basis set $\Phi_{s_1,s_2,\ldots,s_\nu}$,

$$V^{(\nu+1)}(\mathbf{r}_{j_1 i}, \ldots, \mathbf{r}_{j_\nu i}) = \sum_{\{s\}} J_{s_1,s_2,\ldots,s_\nu} \Phi_{s_1,s_2,\ldots,s_\nu}(\mathbf{r}_{j_1 i}, \ldots, \mathbf{r}_{j_\nu i}) \ , \tag{4.2}$$

with general interaction coefficients $J_{s_1,s_2,\ldots,s_\nu}$. A crucial insight from the recent development of atom-centered descriptors [14, 59, 277, 9] is that an efficient representation of atomic environments can be obtained by a low-rank approximation of the $\Phi_{s_1,s_2,\ldots,s_\nu}$ cluster basis in terms of a single-bond basis $\phi_{s_k}(\mathbf{r}_{j_k i})$,

$$\Phi_{s_1,s_2,\ldots,s_\nu}(\mathbf{r}_{j_1 i}, \ldots, \mathbf{r}_{j_\nu i}) = \prod_{k=1}^{\nu} \phi_{s_k}(\mathbf{r}_{j_k i}) \ , \tag{4.3}$$

combined with a reorganization of the summations in (4.1) and (4.2),

$$\sum_{j_1 \cdots j_\nu} \Phi_{s_1,s_2,\ldots,s_\nu}(\mathbf{r}_{j_1 i}, \ldots, \mathbf{r}_{j_\nu i}) = \prod_{k=1}^{\nu} A_{i,s_k} \ . \tag{4.4}$$

Formally speaking, the atom-centered descriptor $A_{i,s}$ can be understood as the expansion coefficients of the local atomic density σ_i in an abstract atom-centered Hilbert space $|s\rangle \in \mathcal{V}$,

$$|\sigma_i\rangle = \sum_s A_{i,s}|s\rangle \ . \tag{4.5}$$

In practice, the real-space atomic density for a given atomic environment is approximated by a superposition of localized functions h,

$$\langle \alpha \mathbf{r} | \sigma_i \rangle = \sum_{j \neq i} \delta_{\alpha \alpha_j} f_c(r_{ji}) h(\mathbf{r} - \mathbf{r}_{ji}) \ , \tag{4.6}$$

where species α_j occupies site j and $f_c(r_{ji})$ is a smooth cutoff function that restricts the sum to some local environment of atom i. In the original ACE formalism [59], h is chosen to be a delta function, whereas in the SOAP and SNAP formalisms [14], h is chosen to be a Gaussian.

The atom-centered basis (4.5) is not generally invariant under action of the rotation group $\mathcal{G} = \mathrm{SO}(3)$. Rather, according to Maschke's theorem, the atom-centered Hilbert space $\mathcal{V}$ decomposes into a direct sum of irreducible representations ("irreps") of SO(3), $\mathcal{V} \cong \bigoplus_l \mathcal{D}_l \otimes \mathcal{V}_l$, where $\mathcal{V}_l$ is the subspace of the irrep $l = 0, 1, 2, \ldots$, and the degeneracy space $\mathcal{D}_l$ contains additional degrees of freedom (e.g., atomic species α and purely radial components n) which are untouched by the group action. The generic indices, s, of the single-bond basis functions can thus be replaced by the set $(\alpha n l m)$, where $-l \leq m \leq l$ labels components of the irrep subspace $\mathcal{V}_l$, and the angular momentum channels of the real-space single-bond basis correspond to spherical harmonics $Y_l^m(\hat{\mathbf{r}})$,

$$\langle \mathbf{r} | \alpha n l m \rangle = \phi_{\alpha n l m}(\mathbf{r}) | \alpha \rangle = R_{nl}(r) Y_l^m(\hat{\mathbf{r}}) | \alpha \rangle \ . \tag{4.7}$$

The expansion of the local atomic density centered on site i can thus be written as

$$| \sigma_i \rangle = \sum_{\alpha n l m} A_{i, \alpha n l m} | \alpha n l m \rangle \ , \tag{4.8}$$

where the atom-centered descriptors are given by

$$A_{i, \alpha n l m} = \langle \alpha n l m | \sigma_i \rangle \tag{4.9}$$

$$= \int_{\mathbb{R}^3} d\Omega \ R_{nl}(r) Y_l^m(\hat{\mathbf{r}}) \langle \alpha \mathbf{r} | \sigma_i \rangle \ . \tag{4.10}$$

Furthermore, the reorganized $(\nu+1)$-body expansion (4.4) is equivalent to a product state formed by repeated copies of the density $|\sigma_i^{\otimes\nu}\rangle \coloneqq \bigotimes^{\nu} |\sigma_i\rangle$,

$$\prod_{k=1}^{\nu} A_{i,\alpha_k n_k l_k m_k} = \langle \alpha_1 n_1 l_1 m_1 \cdots \alpha_\nu n_\nu l_\nu m_\nu | \sigma_i^{\otimes\nu} \rangle \ . \tag{4.11}$$

The expressions (4.8) and (4.11) suggest that each term in the ν-order series expansion admits a natural description in terms of a tensor network (TN) for the states $|\sigma_i^{\otimes\nu}\rangle$. In the following, we will consider an associated graphical calculus which incorporates the SO(3) representations carried by the tensors, $A_{i,\alpha_k n_k l_k m_k}$, which will be useful in explicitly constructing SO(3)-invariant descriptors. In particular, the formalism presented below closely follows recent treatments of symmetric tensor networks [226, 227, 216, 26] and the earlier, related development of spin networks [196, 187] from angular momentum recoupling theory [1].

4.1.1 Graphical calculus for atomic descriptors

Working with explicit expressions for tensor product spaces can quickly become cumbersome as the order ν increases, and this is particularly true when the tensor space possesses a complex internal structure, as is the case for working with SO(3) symmetry. However, much of the algebraic structure can be encapsulated in a consistent graphical notation, which we utilize in the following. In the standard diagrammatic notation of tensor networks (Appendix 4.4.2), the basic atom-centered descriptor $A_{\alpha nlm}$, as a 4-index tensor, can be represented by a shape with 4 open edges (Figure 4.1). Note that an arrow is added to the edge carrying the irrep space, $\mathcal{V}_l$, of the symmetry group to distinguish it from its dual vector space, $\mathcal{V}_l^*$, and we will subsequently drop the atom site index i for simplicity. Regular representations of SO(3) are given by the unitary Wigner D-matrices, $D^{(l)}_{m'm}(g)$, and the action of $D^{(l)}_{m'm}(g)$ on the atom-centered basis is given by *contraction* with A along the irrep edge, i.e., by

[1] Most of the graphical machinery discussed below carries over more generally to other symmetric tensor categories, of which finite-dimensional representations of SO(3) are an example.

$\sum_m D^{(l)}_{m'm}(g) A_{\alpha n l m}$ (cf. Appendix 4.4.2).

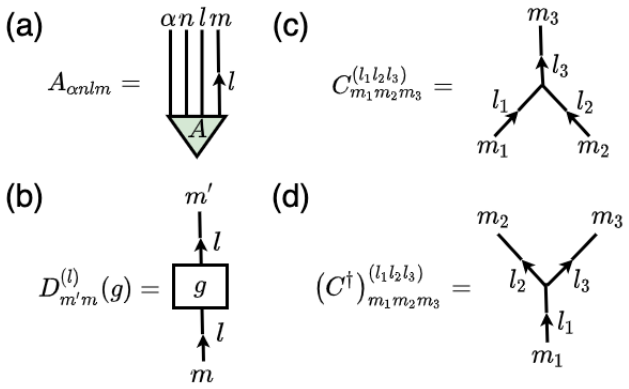


Figure 4.1: Basic graphical components for the construction of invariant descriptors, including (a) atom-centered descriptors (4.10), (b) Wigner D-matrices, and (c,d) Clebsch-Gordan coefficients as fusion and splitting nodes.

A ν-order product state of multiple A-tensors (4.11) is depicted simply by a row of disconnected A's. This ν-order product state is not invariant under three-dimensional rotations, $U_g^{\otimes\nu}|\sigma_i^{\otimes\nu}\rangle \neq |\sigma_i^{\otimes\nu}\rangle$. However, it can be made so by explicitly taking the Haar integral over $g \in \mathrm{SO}(3)$, $|\sigma_i^{\otimes\nu}\rangle_g := \int_g U_g^{\otimes\nu}|\sigma_i^{\otimes\nu}\rangle$, since the homomorphism $U_{g_1}U_{g_2} = U_{g_1 g_2}$ extends to the tensor product space. In tensor components, this symmetrized descriptor is given by

$$\begin{aligned}&\langle \alpha_1 n_1 l_1 m_1 \cdots \alpha_\nu n_\nu l_\nu m_\nu | \sigma_i^{\otimes\nu}\rangle_g \\ &\qquad = \int_{g\in \mathrm{SO}(3)} dg \prod_{k=1}^{\nu} \langle \alpha_k n_k l_k m_k | U_g | \sigma_i \rangle \ , \end{aligned} \tag{4.12}$$

where, upon insertion of resolutions of identity, one obtains a tensor product of Wigner D-matrices, $D^{(l)}_{mm'}(g) = \langle lm|U_g|lm'\rangle$, acting on the A-tensors. An explicit formula for (4.12) can be computed by introducing the intertwiners of the symmetry group, the Clebsch-Gordan (CG) coefficients, into the basic building blocks of the graphical calculus (Figure 4.1). We

outline this derivation in Figure 4.2 for the case $\nu = 3$, but it can be extended by induction to all ν. In particular, repeated application of the CG identities and their equivariance (Figure 4.20) reduces (4.12) to a single Wigner D-matrix contracted with recursively constructed *fusion trees* of CG coefficients. A single, invariant irrep edge must transform as the trivial representation, hence the Haar integral projects the final D-matrix-decorated edge onto the trivial irrep space. We distinguish an edge carrying the trivial representation ($l = 0$) by a dashed line. Because the trivial representation is one-dimensional ($m = 0$) [2], symmetrization yields two disconnected diagrams, which we label Q and B (see below).

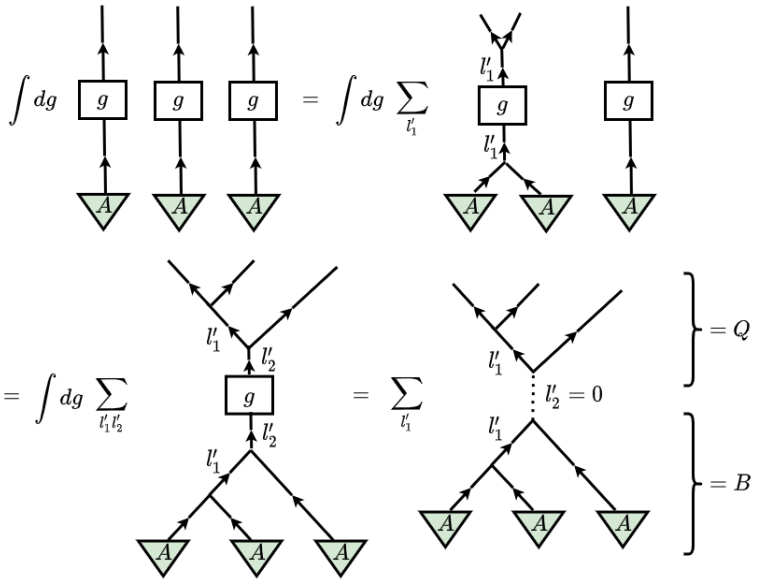


Figure 4.2: The explicit symmetrization (4.12) of a $(\nu+1)$-body product state (4.11) derived from the algebraic rules of SO(3)-recoupling theory. For simplicity, the edges (αnl) are not shown.

A few comments are in order. First, the pattern of contractions encapsulated by the structure of the fusion trees in Figure 4.2 represents an implicit choice of recoupling scheme.

[2] More generally, the trivial representation is the unital object in category $\mathrm{Rep}(\mathcal{G})$ of representations of the group $\mathcal{G}$.

Different choices of recoupling scheme are possible and are related to each other by unitary transformations, i.e., the so-called F-symbols (cf. Appendix 4.4.3). Second, symmetrization of ν-order product states for $\nu > 2$ introduces summations over intermediate angular momenta l', which we represent explicitly, while contractions of irrep edges imply summation over the corresponding magnetic numbers m'. Finally, this direct symmetrization yields the so-called Jucys-Levinson-Vanagas (JLV) theorems [237], also referred to as the generalized Wigner-Eckhart theorem [51, 226, 227, 216]. Accordingly, a symmetric, n-index tensor T decomposes into a tensor product, $T = B \otimes Q$, of a structural tensor Q determined completely by the symmetry group and of a degeneracy tensor B. The structural tensor Q, given by a fusion tree of CG coefficients, is known in other contexts as a spin network [196, 187]. In the spirit of the original Wigner-Eckhart theorem, the degeneracy tensor B is equivalent to the "reduced matrix element" of the decomposition and encapsulates the degrees of freedom not fixed by the symmetry group [51]. This is shown explicitly in Figure 4.4, where the $\{\alpha_k n_k l_k\}$ edges remain open.

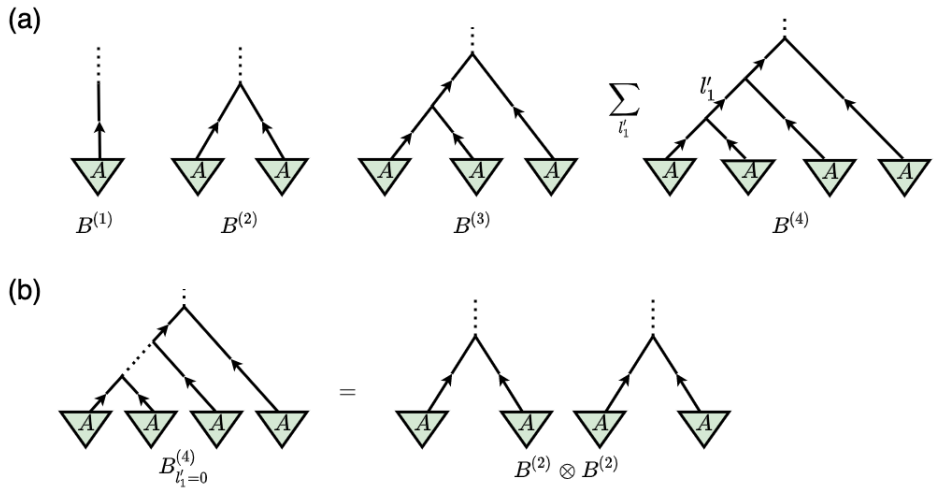


Figure 4.3: (a) The ν-order hierarchy of SO(3)-invariant descriptors in terms of recursively constructed fusion trees. (b) Higher order tensors with trivial intermediate irreps factorize into products of lower order tensors.

The B-tensors constitute the set of SO(3)-invariant descriptors which form the basis of the ACE framework. In the recoupling scheme chosen in Figure 4.2, the ν-order series of invariant tensors is given by the contraction of ν atom-centered tensors A with the appropriate fusion tree (Figure 4.3), and invariance is thus manifested by mapping the tensor product of ν irreps to the trivial representation. It is readily apparent that invariant descriptors with trivial intermediate irreps factorize, e.g., $B^{(4)}_{l'_1=0} = B^{(2)} \otimes B^{(2)}$. Moreover, this construction can be generalized [60] to SO(3)-equivariant descriptors $B^{(\nu)}_L$ by allowing the tensor product of ν irreps to fuse to a non-trivial representation L, where the open L edge of the fusion tree carries the dimension of the irrep, $2L+1$. Higher-order SO(3)-equivariant descriptors, $B^{(\nu+1)}_L$, can be built recursively from lower orders $B^{(\nu)}_{L'} \otimes B^{(1)}_{L''}$ by contracting with the so-called *cup* and *cap* tensors (normalized $2jm$ symbols), which diagrammatically enable the orientation of irrep edges to be reversed. This recursive construction is discussed in more detail in Appendix 4.4.4. In the following sections, we will represent the invariant ACE basis tensors $B^{(\nu)}$, as in Figure 4.4, by suppressing the fusion trees.

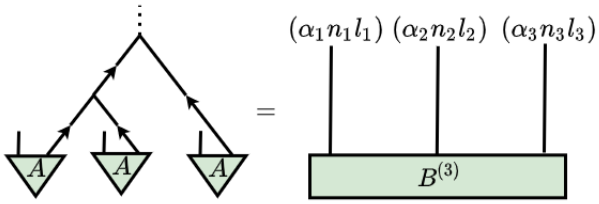


Figure 4.4: Simplified representation of an invariant descriptor with fixed SO(3)-recoupling scheme.

We give explicit formulae [59, 60] for the first few invariant descriptors, while higher order $B^{(\nu)}$ can be constructed via the recursive procedure outlined in Appendix 4.4.4 or simply read off of the corresponding fusion tree:

$$B^{(1)}_{\alpha n} = A_{\alpha n 00} \tag{4.13}$$

$$B^{(2)}_{\substack{\alpha_1\alpha_2\\ n_1 n_2 l}} = \sum_{m=-l}^{l} \frac{(-1)^{l-m}}{\sqrt{d_l}} A_{\alpha_1 n_1 l m} A_{\alpha_2 n_2 l -m} \tag{4.14}$$

$$(B^{(3)})^{\substack{\alpha_1\alpha_2\alpha_3\\ n_1 n_2 n_3\\ l_1 l_2 l_3}} = \sum_{m_1 m_2 m_3} \frac{(-1)^{l_3-m_3}}{\sqrt{d_{l_3}}} C^{(l_1 l_2 l_3)}_{m_1 m_2 -m_3} \times A_{\alpha_1 n_1 l_1 m_1} A_{\alpha_2 n_2 l_2 m_2} A_{\alpha_3 n_3 l_3 m_3} \tag{4.15}$$

Here, $d_l = 2l+1$ is the dimension of the irrep, and the angular momentum coupling in (4.15) can be rewritten in terms of the Wigner $3jm$ symbol, as in [59]. As expected, the $\nu = 1$ term loses all angular information after integrating over the SO(3) rotation group. Hence, accurate expansions of atom-centered properties typically require higher-order terms. In the $\nu = 2, 3$ terms, one finds the SOAP power spectrum and bispectrum, respectively [14, 59]. In a real-space basis, these ν-order invariants can be understood as spherical moments of the bond distribution centered on atom i. For $\nu = 1$, this corresponds to spherically averaging over a single bond $\mathbf{r}_{j_1 i}$, which yields a completely isotropic function depending only on the bond length. The power spectrum, $\nu = 2$, measures the correlation between pairs of bonds ($\mathbf{r}_{j_1 i}$, $\mathbf{r}_{j_2 i}$), where spherical averaging over the local environment results in a function depending on the bond lengths ($r_{j_1 i}$, $r_{j_2 i}$) and the relative angle, $\hat{\mathbf{r}}_{j_1 i} \cdot \hat{\mathbf{r}}_{j_2 i}$, between bond vectors. Similarly, a real-space projection of the bispectrum, $\nu = 3$, depends only on three bond lengths ($r_{j_1 i}$, $r_{j_2 i}$, $r_{j_3 i}$) and the relative angles between each pair of bond vectors.

4.2 Tensor Network Learning

The SO(3)-invariant, atom-centered descriptors $B_i^{(\nu)}$ can be used as input features for a variety of machine-learning methods. In the ACE formalism, this amounts to recasting the expansion (4.1) in the symmetrized basis,

$$
\begin{aligned}
V_i &= \sum_{\nu} w_i^{(\nu)} \cdot B_i^{(\nu)} \\
&= \sum_{\nu} \sum_{\{\alpha_k n_k l_k\}} w_{i\{\alpha_k n_k l_k\}}^{(\nu)} B_{i\{\alpha_k n_k l_k\}}^{(\nu)} \ ,
\end{aligned} \tag{4.16}
$$

where the model weights $w_i^{(\nu)}$ are fully contracted with the free edges, $\{\alpha_k n_k l_k\}$, of the invariant descriptors. As dense tensors, the size of the weights $w_i^{(\nu)}$ scale as $\mathcal{O}(|\alpha|^\nu |n|^\nu |l|^\nu)$ with the number of chemical species $|\alpha|$ and the number of radial $|n|$ and angular momentum $|l|$ channels included in the input descriptors.

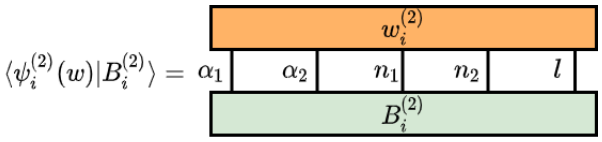


Figure 4.5: ($\nu = 2$)-order term in the ACE expansion as an inner product between the descriptor state, $|B_i^{(\nu)}\rangle$, and a learnable state, $|\psi_i^{(\nu)}(w)\rangle$.

It is clear from (4.16) that each ν-order term can be understood as an inner product, $\langle \psi_i^{(\nu)}(w) | B_i^{(\nu)} \rangle$, between the invariant basis states $|B_i^{(\nu)}\rangle$ (the reduced part of the symmetrized state $|\sigma_i^{\otimes\nu}\rangle_g$) and a state $|\psi_i^{(\nu)}(w)\rangle$ which depends on a set of learnable parameters w (Figure 4.5). We can generalize this construction by considering a tensor network ansatz for $|\psi_i^{(\nu)}(w)\rangle$, where the topology of the network encodes the entanglement structure of the state. Taking inspiration from the application of tensor networks in quantum physics as low-rank approximations of many-body ground states, tensor network factorizations of the model weights will be used to constrain the correlation structure between the tensor elements $\{\alpha_k n_k l_k\}$ of the input descriptor. We will find that this acts as an implicit method to regularize the model fitting [185].

Since there is no inherent geometric relationship between the positions of the descriptor indices, we consider factorizations which preserve their order. To maintain generality, we

will primarily examine factorizations built from matrix product states (MPS),

$$|\psi^{(N)}\rangle = \sum_{s_1,\ldots,s_N} a_1^{s_1} \cdots a_N^{s_N} |s_1 \cdots s_N\rangle \ , \tag{4.17}$$

and matrix product operators (MPO),

$$\hat{T}^{(N)} = \sum_{\substack{s_1,\ldots,s_N \\ s'_1,\ldots,s'_N}} M_1^{s'_1 s_1} \cdots M_N^{s'_N s_N} |s'_1 \cdots s'_N\rangle\langle s_1 \cdots s_N| \tag{4.18}$$

where $a_k^{s_k}$ and $M_k^{s'_k s_k}$ are $\chi_k \times \chi_{k+1}$ matrices, which constitute the learnable model parameters. Note that because the input descriptors and target properties are real-valued, the tensors a_k and M_k will also be real-valued. The internal bond dimensions χ_k are often referred to as virtual dimensions, and the physical indices s_k denote either a descriptor index $\{\alpha_k, n_k, l_k\}$ or a generic, internal vertical bond. MPS and MPO tensor networks are shown in Figure 4.6 for the case of finite virtual boundary conditions, $\chi_1 = \chi_N = 1$. For simplicity, we will set all virtual dimensions to be equal, $\chi_k = \chi$, except for the boundary edges. As discussed further in Section 4.3, the virtual dimensions impose upper bounds on the bipartite entanglement entropies of the state $|\psi^{(\nu)}\rangle$.

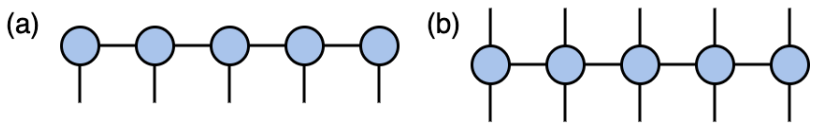


Figure 4.6: Tensor network representations of (a) a matrix product state (MPS) and (b) a matrix product operator (MPO).

It is useful to note that the sequential application of several MPOs to an arbitrary product state $|0\rangle^{\otimes N}$ is equivalent to evolving the state by a finite-depth, variational quantum circuit in the case where each MPO is a unitary operator. If each MPO possesses bond dimension χ, then contraction of n_d MPOs with an arbitrary product state yields an MPS

with bond dimension χ^{n_d} (Figure 4.7a). This can be an efficient way to achieve an MPS with large effective bond dimensions while mitigating the growth in the number of parameters. For the cases studied in this work, the bond dimensions required to achieve accurate models remain relatively small, so we take as our starting ansatz a single MPS whose bond dimensions are treated as hyperparameters. Furthermore, since the MPS tensor elements are real-valued, we optimize them directly, rather than relying on unitary embeddings, using classical (e.g., stochastic gradient descent (SGD)) rather than quantum algorithms. We will return to this tensor network / quantum circuit correspondence in our discussion in Section 4.3. To make use of high-performance automatic differentiation and back propagation algorithms commonly employed in deep learning, the bond dimensions will remain fixed during training. However, certain training methods based on the density matrix renormalization group (DMRG), which adaptively update bond dimensions, would be interesting to compare.

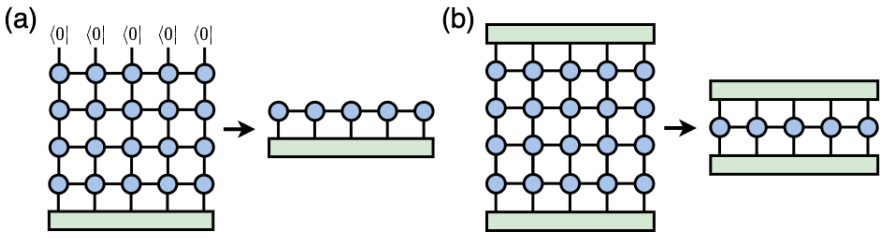


Figure 4.7: The (a) MPS and (b) MPO models used in this work viewed as the contraction of several MPOs. In particular, the upper layer on the LHS of the MPS model (a) defines an embedding of an MPS as an MPO. Learnable model weights are shown in blue, while input descriptors are shown in green.

While we have focused so far on the construction of learnable states $|\psi^{(\nu)}\rangle$, we can alternatively formulate the task of learning scalar contributions to the target property V_i in terms of an expectation value of an MPO, $\langle B_i^{(\nu)}|\hat{T}^{(3\nu)}|B_i^{(\nu)}\rangle$, with respect to the invariant basis states $|B_i^{(\nu)}\rangle$. This can be thought of as a relaxed version of the Born rule, $p(B) = |\langle B|\psi\rangle|^2$, where the pure state density matrix, $|\psi\rangle\langle\psi|$, is replaced by a general mixed state, described by a

matrix product density operator (MPDO), $\hat{\rho}$. Indeed, it was shown in [81] within the context of probabilistic graphical models that locally-purified approximations of MPDOs can represent a larger class of non-negative tensors than an MPS model of equivalent rank. Because we are concerned with the prediction of some target scalar which is not necessarily a probability $p(B)$, we do not enforce non-negativity on the elements of the MPO, $\hat{T}$. Nonetheless, we will provide empirical evidence in Section 4.3 that a similar rank-efficiency relationship may hold for regression tasks when measured against their performance on unseen data. Again, while we could in general train a sequence of MPOs, akin to evolving an invariant basis state back to itself (e.g., similar to the treatment of thermal states as path integrals supported on a compact manifold [66]), we will instead consider a single MPO with adjustable bond dimensions (Figure 4.7b).

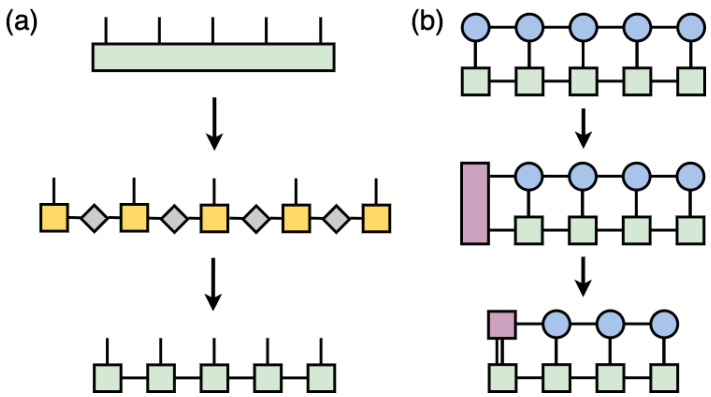


Figure 4.8: (a) A MPS factorization of a dense tensor can be constructed by the sequential application of singular value decompositions between groups of edges. The grey diamonds are matrices of singular values which are subsequently contracted with their neighboring tensors. (b) An efficient contraction order for an MPS model with factorized inputs, where the current contracted tensor is shown in purple.

As the order ν and the dimensions of the individual indices of the dense descriptor tensor increase, its contraction with the variational states $|\psi^{(\nu)}\rangle$ and operators $\hat{T}^{(3\nu)}$ can

become computationally demanding. However, this issue can be alleviated by analogously constructing low-rank, MPS approximations of the states $|B_i^{(\nu)}\rangle$. A standard procedure, shown in Figure 4.8a, is to recursively take singular value decompositions (SVDs) between neighboring pairs of edges, discard a subset of the lowest singular values (SV), and contract the corresponding SV matrices with their neighboring tensors. The virtual bond dimensions of the subsequent MPS is equal to the number of singular values retained at each edge (i.e., the Schmidt rank), which controls the accuracy of the approximation. For small enough bond dimensions, this constitutes a kind of sparsification procedure on the inputs, and we find in practice (cf., Section 4.3.4) that model accuracy can be maintained with surprisingly small bonds. With this MPS decomposition, a more efficient contraction scheme (e.g., Figure 4.8b), can be implemented.

4.2.1 Relationship to other methods

In practice, a balance must be sought between the computational efficiency of the model and the number of input features required to accurately represent the atomic environment. A common simplification is to restrict the expansion (4.16) to many-body descriptors of relatively low order. This is, for instance, the strategy employed in Behler-Parrinello (BP) neural network potentials [21], which use 1-, 2- and 3-body symmetry functions. It is often the case that features derived from order $\nu \leq 3$ expansions are able to differentiate the relevant structural characteristics in a given sample, although certain counterexamples exist [206]. Furthermore, as noted in Section 4.1, higher-order descriptors contain products of lower order, which enables accurate models to be built on a fixed-order descriptor [9, 184]. This is the case for kernel-based models utilizing either the SOAP power spectrum ($\nu = 2$) or bispectrum ($\nu = 3$).

While the expansion (4.16) motivates the systematic introduction of SO(3)-invariant descriptors, one is not limited to a linear model for the prediction of some atomic property V_i. Indeed, kernel methods based on SOAP features and neural networks using ACSFs incorporate general forms of nonlinearity, whether through the choice of covariance kernel

$K(\mathbf{x}, \mathbf{x}')$ in the former, or the structure of the learning architecture in the later. It is worth dwelling on some essential aspects of kernel-based and neural network-based methods, as they will serve as further motivation for the tensor network methods introduced above.

In supervised learning, where the goal is to find a function $f(\mathbf{x})$ which approximates a target quantity $y \approx f(\mathbf{x})$, a nonlinear function on the inputs x_i can be constructed via a mapping φ into a higher-dimensional *feature space*. Such a mapping allows the function f to be formulated as a linear model on the feature space, but the computational complexity of working directly with very large feature vectors $\varphi(\mathbf{x})$ often prohibits their explicit implementation in machine learning tasks. However, since linear models constructed in a feature space can be rewritten in terms of an inner product $\langle\varphi(\mathbf{x})|\varphi(\mathbf{x}')\rangle$ on that space, Mercer's theorem allows one to replace the explicit feature map with a positive, semidefinite kernel function $K(\mathbf{x}, \mathbf{x}') = \langle\varphi(\mathbf{x})|\varphi(\mathbf{x}')\rangle$. This so-called *kernel trick* and the representer theorem leads to a model of the form $f(\mathbf{x}) = \sum_a w_a K(\mathbf{x}, \mathbf{x}_a)$, where the sum runs over a reference set of training instances. Note that the cost to evaluate the model scales linearly with the size of the training set, and thus smaller training sets may generalize poorly to out-of-sample data. Furthermore, the cost for training the model scales like $\mathcal{O}(N^3)$ for N training instances.

An advantage of the expansion (4.8) of the local density in the basis of SO(3) irreps is that it comes equipped with a natural inner product, which induces an inner product on the SO(3)-invariant tensor product states $|\sigma_i^{\otimes\nu}\rangle_g$. Under this induced inner product, the feature space can be identified with the descriptor space. As a tensor network, the associated kernel function is given by contracting the open edges of the SO(3)-invariant descriptors $B_i^{(\nu)}$ and $B'^{(\nu)}_j$ (e.g., Figure 4.10). Note that in kernel-based methods, pairs of local environments in the kernel function come from different structures, which we distinguish using a prime on the descriptors. The (smoothed) overlap $k\left(B_i^{(\nu)}, B'^{(\nu)}_j\right) \coloneqq \langle B_i^{(\nu)}|B'^{(\nu)}_j\rangle$ quantifies the similarity between atomic environments, and raising this kernel to an integer power ζ enhances its sensitivity to the differences between those environments. Up to a normalization of the descriptors, these overlaps yield the class of SOAP kernels upon summing over all sites. Introducing a power ζ to the overlaps is equivalent to taking a ζ-order tensor product of

the invariant basis state, $|B_i^{(\nu)}\rangle^{\otimes\zeta}$, which increases the effective many-body character of the state [277]. We will return to this idea when discussing generalizations of our method below. Note that these $(\nu \cdot \zeta)$-order states do not span the entirety of their SO(3)-invariant tensor product space but instead correspond to states with trivial intermediate irreps (cf. Figure 4.3).

A dense (fully-connected, feed-forward) neural network (NN) model approximates the predictor function $f(\mathbf{x}) = \mathcal{F}\mathcal{L}\cdots\mathcal{F}\mathcal{L}(\mathbf{x})$ by an alternating composition of affine maps $\mathcal{L}(\mathbf{x}) = w\mathbf{x} + \mathbf{b}$ and nonlinear activation functions $\mathcal{F}(\cdot)$ applied element-wise to their input vectors. A dense neural network possesses a relatively simple tensor network representation (Figure 4.9), in which the linear weight matrices w are shown as 2-index tensors while the biases $\mathbf{b}$ and activation functions remain implicit. In the case where the SO(3)-invariant descriptor space is taken to be the input feature space and the activation functions and biases are chosen to be trivial, this neural network architecture provides a tensor network factorization of the model weights $w_i^{(\nu)}$ in the ACE expansion (4.16), where the internal bond dimensions in the tensor network correspond to the number of nodes in the hidden layers of the neural network, and the output (open) bond has dimension 1 when the target property is a scalar. Restoring the biases and nonlinear activation functions thus generalizes the linear maps $w_i^{(\nu)}$. Since the descriptors are treated as input feature vectors with a multi-index $(\alpha_1 n_1 l_1 \cdots)$, the input bond dimension suffers from the same exponential scaling with ν as the original ACE expansion. For deeper neural networks with many hidden nodes, this can lead to an exceptionally large number of model weights and a potential risk of overfitting when the number of training samples is small.

At this point we can draw some connections between the MPO/MPS model in Figure 4.7a and the neural network- and kernel-based methods discussed above. A sequence of n_d MPOs applied to the input descriptors can be viewed as a particular factorization of the weight matrices w in the n_d layers of a dense neural network. For sufficiently small virtual bond dimensions, this can lead to a substantial reduction in the number of model parameters [185, 77]. Hence, the tensor network structure can be understood as a form of

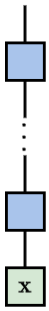


Figure 4.9: A dense neural network as a simple tensor network, where the internal bond dimensions are determined by the number of nodes in each hidden layer, and the layer-wise activation functions and biases are implicit.

regularization on the otherwise dense model weights, where the chosen factorization scheme can enforce a degree of sparsity in the parameters. Moreover, a single MPO applied to an input state $|B_i^{(\nu)}\rangle$ constitutes a mapping from the symmetry-adapted, atom-centered Hilbert space to a potentially lower dimensional space, $|B_i^{(\nu)}\rangle \rightarrow |J_i^{(\nu)}\rangle = \hat{T}^{(3\nu)}|B_i^{(\nu)}\rangle$. Note that since this operator acts on the reduced tensor elements $B^{(\nu)}$ of the SO(3)-invariant subspace, it commutes with the action of the SO(3) rotation group.

The overlap $\langle J_i^{(\nu)}|J'^{(\nu)}_j\rangle$ is equivalent to a generalized kernel function which couples the structure of the atomic environments to the chemical species within them. This constitutes a very general form of low-rank approximation for the "nonfactorizable operators" discussed in [277]. Indeed, this overlap contains the class of so-called "alchemical" kernels built from the SOAP power spectrum (Figure 4.10), which have been shown to improve the accuracy of structural kernel-based methods [53, 276]. Adopting the notation in [277, 276, 53], we see that the alchemical couplings, $\kappa_{\alpha\alpha'}$, can be decomposed in a lower-dimensional elemental basis $|s\rangle$, $\kappa_{\alpha\alpha'} = \sum_s u_{\alpha s} u_{s\alpha'}$. In previous applications of these generalized kernel methods, the values of these couplings were either chosen explicitly, for instance by incorporating physical intuition [53, 6], or learned via additional feature selection steps prior to training the model [276, 110]. Alternatively, by considering general tensor network factorizations of

the the weights $w_i^{(\nu)}$, one can work directly with the atom-centered Hilbert space rather than with distinct pairs of local environments. The above discussion highlights the fact that the choice of network topology imposes a kind of inductive bias on the model which may be used as a way to prioritize a certain kind of solution (e.g., an alchemical one) for the predictor $f(\mathbf{x})$.

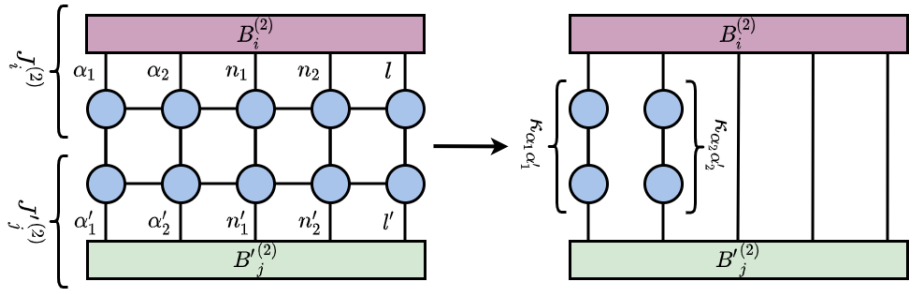


Figure 4.10: A low-rank approximation of a generalized structural kernel can be represented in terms of MPOs. For the appropriate arrangement of physical bonds, this contains the class of alchemical kernels shown at right.

So far, our construction has focused on tensor decompositions of the weights $w_i^{(\nu)}$ as multilinear maps. However, as mentioned above, there may be some advantage to introducing nonlinearity into the model architecture. This could be accomplished analogously to neural networks by introducing element-wise, nonlinear activation functions at every level. Alternatively, following the work of Stoudenmire and Schwab [238], we could construct an explicit higher-dimensional embedding via a feature map φ and work directly with tensor network factorizations of linear weights ω on the embedding space, $f(x) = \omega \cdot \varphi(\mathbf{x})$, without, as with kernel methods, passing to the dual vector space. A convenient embedding is formed by a (unentangled) product state,

$$\varphi(x) = \varphi^{s_1}(x_1) \otimes \varphi^{s_2}(x_2) \otimes \cdots \otimes \varphi^{s_N}(x_N) \ , \tag{4.19}$$

where each element of the input vector $\mathbf{x}$ (i.e., the vectorization and standardization of the

input descriptor $B^{(\nu)}$) is encoded by a local feature map $\varphi^{s_i}(x_i)$. In their original formulation [238], Stoudenmire and Schwab proposed a class of local feature maps in terms of spin-coherent states, $\varphi^{s_i}(x_i) = [\cos(\pi x_i/2), \sin(\pi x_i/2)]^\intercal$. However, we found that in practice a simpler linear mapping $\varphi^{s_i}(x_i) = [1 - x_i, x_i]^\intercal$ performed better when using atom-centered descriptors as input vectors. Because φ maps a d-dimensional input vector to a feature space of dimension 2^d, one is limited to tensor network factorizations of the weights that can be efficiently contracted. As in the original proposal, the simplest choice corresponds to a matrix product state (Figure 4.11). Still, to obtain an accurate model, one is often left with a large number of learnable parameters. Moreover, we have found that TN models built directly on the descriptor space outperform models with an additional feature map, while requiring fewer parameters.

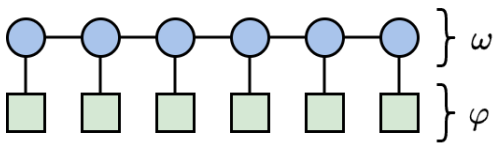


Figure 4.11: Stoudenmire and Schwab MPS model [238], where each input component is explicitly mapped to an unentangled product state, φ.

Another interesting possibility is to mimic the construction of nonlinear structural kernels by forming ζ-order tensor products of the symmetrized descriptors $B^{(\nu)}$:

$$B^{(\nu)} \to B^{(\nu)} \otimes B^{(\nu)} \to \cdots \to \left(B^{(\nu)}\right)^{\otimes\zeta} \tag{4.20}$$

As inputs to a NN or TN model, this construction entails reuse of information in the subsequent network. Information reuse is a common characteristic of modern NN architectures, such as convolutional (CNN) and recurrent (RNN) neural networks, where several trainable filters act on the same input or subsets of input [80]. Recent work [137, 136] has shown that convolutional (CAC) and recurrent (RAC) arithmetic circuits, which share fundamen-

tal architectural components like overlapping filters and pooling with conventional CNNs and RNNs, can be mapped to so-called *generalized* tensor networks [26, 80] by introducing local copy operations into the tensor algebra [3]. It was shown in [137, 136] that a deep CAC with a local pooling scheme maps to a hierarchical, tree-structured tensor network and, for maximally overlapping filters, can support volume-law scaling of the entanglement entropy when modeling the amplitudes of many-body quantum states. Moreover, the entanglement capacity [50] of the CAC tensor network was shown empirically to strongly influence the inductive bias of supervised image classification with a CNN [137].

In a similar vein, the MPO model introduced above can be viewed as a specific choice of learnable architecture acting on $\zeta = 2$ copies of the input descriptor, $T^{(3\nu)}\left(|B^{(\nu)}\rangle^{\otimes 2}\right)$. Since copy operations (4.20) applied to the input descriptors increase their effective many-body character [277], this method provides a potential route toward capturing higher-order correlations with lower order features.

4.3 Numerical Benchmarks

To validate the methods described in Section 4.2, we will focus on a typical learning task encountered in materials property prediction using a widely available dataset (NMD18 [239]) of 3000 transparent conducting oxide (TCO) alloys $(Al_xGa_yIn_z)_2O_3$, a class of wide bandgap materials with high technological applicability in optoelectronic devices. The target property is the DFT mixing enthalpy per cation (referred to as the formation energy in [239]) for a given configuration, referenced to the compositional endpoints of the material,

$$\begin{aligned}\Delta H_{\text{mix}}[(Al_xGa_yIn_z)_2O_3] &= E[(Al_xGa_yIn_z)_2O_3] \\ &- xE[Al_2O_3] - yE[Ga_2O_3] - zE[In_2O_3]\ . \end{aligned} \tag{4.21}$$

Here $x + y + z = 1$, and $E[c]$ is the ground state DFT energy per cation for the given

[3] The general copying of vectors and tensors by a standard tensor network on a finite-sized Hilbert space is forbidden by a no-cloning theorem.

compound c. The mixing enthalpy characterizes the (zero-temperature) stability of the alloy configuration relative to the single cation phases. The underlying crystalline lattices of the alloy configurations span six distinct space groups (C/2m, P$na2_1$, $R\overline{3}c$, P$6_3/mmc$, I$a\overline{3}$, and $Fd\overline{3}m$) and the number of sites in each structure can vary in integer multiples of the number of primitive cell sites. The reference compositional endpoints, however, are fixed to their ground-state lattices, $R\overline{3}c$ for Al_2O_3, C/2m for Ga_2O_3, and I$a\overline{3}$ for In_2O_3. The inclusion of multiple lattice symmetries is difficult to handle with more conventional atomistic modeling methods, such as the standard cluster expansion (CE) [239, 186]. Moreover, allowing for chemically ordered/disordered sublattices presents challenges to deep end-to-end learning methods, where very large datasets are often required to obtain sufficient accuracy [62, 298, 38]. Depending on the application and the computational cost of the first-principles calculations, generating a large dataset can be prohibitively expensive. It is therefore useful to understand how the accuracy of these methods scales with the size of the training dataset.

To further emulate working in a data-constrained environment, we do not assume prior knowledge of the fully relaxed atomic positions and lattice constants. Instead, the input structures retain their ideal lattice positions, while the lattice vectors are simply scaled according to Vegard's law. The target properties, however, are calculated from the relaxed geometries. This constitutes a particularly challenging scenario. Indeed, it was previously found that learning a mapping from unrelaxed structures to ground-state energies is generally more difficult than using relaxed input structures, at least at the level of kernel-based learning [130]. Moreover, the atomic descriptor was not found to be the limiting factor, but rather prediction error was dominated by implicit noise in the underlying set of atomic structures [130].

To construct a global descriptor suitable for the prediction of the global property ΔH_{mix}, we take an average over the local, atom-centered descriptors of the structure,

$$\overline{B}^{(\nu)} = \frac{1}{N} \sum_{i}^{N} B_i^{(\nu)} \ . \tag{4.22}$$

This is a special case of the general prescription of calculating ΔH_{mix} as a sum over (possibly distinct) contributions $H_i^{(\nu)}$ from N local environments, where for (4.22) all local ν-order contributions are contracted with the same weight tensor $w^{(\nu)}$. To simplify the comparison between different machine learning methods and architectures, we choose the SOAP power spectrum $B^{(2)}$ with fixed radial and angular momentum cutoffs ($R_c = 6$ Å, $n_{\text{max}} = 4$, $l_{\text{max}} = 3$) to be the input descriptor. This also allows us to compare our results to those reported in [239, 130], where the former tested the performance of both a deep NN with SOAP feature vectors and Gaussian process regression (GPR) with the SOAP kernel ($\zeta = 2$), and the latter employed kernel ridge regression (KRR) using a Gaussian / radial basis function (RBF) kernel. Thus, in addition to systematically evaluating various tensor network architectures, we provide consistent benchmarks with respect to the performance of GPR with SOAP and RBF kernels, as well as the deep NN architecture described in [239] (see Appendix 4.4.1 for additional details).

4.3.1 Learning curves

We quantified the predictive performance of the MPS and MPO models, as well as GPR and NN models, using stratified k-fold cross-validation. In particular, we considered 10 splits of the NMD18 dataset into a testing set of 600 structures and a remaining pool for training and validation. Of this remaining pool of structures, 10 subsets of a fixed size were chosen with consistent distributions of volume, composition and energies, following [130], from which we constructed 80-20 splits into training and validation sets. Prediction errors were measured on the testing sets, which remain untouched during model training, while validation sets were used for tuning certain hyperparamters (Appendix 4.4.1).

Figure 4.12 compares the average and standard deviations of the root mean square error (RMSE) and the mean absolute error (MAE) of the model predictions over the testing splits, as a function of the training set size. First, while it shows consistently small variance, GPR with an RBF kernel performs the worst among the models considered, likely due to over-localization of the atomic representations in the embedding feature space. By contrast,

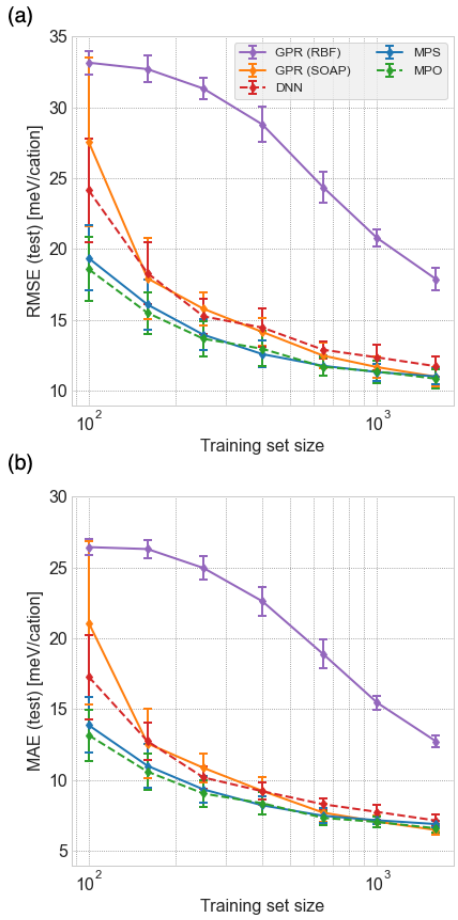


Figure 4.12: Test RMSE (a) and MAE (b) learning curves on the NMD18 dataset. MPS and MPO models are compared to a fully-connected, deep neural network (DNN) and Gaussian process regression (GPR) with RBF and SOAP kernels.

GPR with the SOAP kernel exhibits higher sensitivity to the differences between training structures, improving substantially as the number of reference structures increases. As observed in [239], we find that the DNN model and GPR with SOAP kernel perform similarly across training sets. However, the number of model weights in the DNN architecture remains fixed, while the number of weights increases with the training set size in the kernel methods. Importantly, the DNN is incrementally improvable with the acquisition of new data, while

the kernel-based models must be retrained on the whole dataset. Moreover, for much larger datasets, the computational cost of kernel-based methods becomes prohibitively expensive.

The MPS and MPO models maintain strong generalizability even for very small ($\mathcal{O}(10^2)$ structures) training sets. In this data-constrained regime, they outperform the other models considered, while converging with the DNN and GPR (SOAP) models at the largest training set size. The advantage of the TN models is particularly notable in the RMSE, which is more sensitive to the presence of outliers than the MAE. As with the DNN model, the TN models are systematically improvable with new data without requiring retraining on the full dataset. Unlike the DNN model, they require substantially fewer parameters ($\mathcal{O}(10^3)$ compared to $\mathcal{O}(10^6)$) to achieve similar accuracy. This parametric efficiency may, however, be related to the observed onset of saturation in the model accuracy, particularly for the MAE of the MPS model, as the training set size increases. Yet, similar saturation is also apparent in the DNN model. Moreover, the MPO model marginally outperforms the MPS model, despite possessing a lower virtual bond dimension, and appears less prone to plateaued accuracies for increasing training set sizes. As we show below, the accuracies of the TN models do not suffer from increasing the bond dimensions for fixed datasets, and thus we expect that this saturation can be overcome by simply increasing the number of parameters. It is worth noting that we have additionally tested deeper TN factorizations with stronger entanglement scaling, specifically tree tensor networks (TTN) and the Multi-Scale Entanglement Renormalization Ansatz (MERA) (Figure 4.29), and found that they produce accuracies comparable to the MPS model. They are, however, more expensive to contract. Finally, the computational cost of the TN models scales linearly with the number of samples, as opposed to the quadratic scaling of the kernel-based methods. Thus, we expect the application of these models to be practical across a broad range of dataset sizes.

4.3.2 Entanglement measures

We have seen above that the MPS and MPO models can yield expressive supervised learning models even with a small number of training samples. To what extent does the choice of TN

architecture influence this apparent inductive bias? Since the topology and bond dimensions of the TN constrain the local entanglement structure of the state $|\psi^{(\nu)}\rangle$, it is reasonable to expect that the entanglement in the learned TN parameters captures underlying correlations in the input data. Indeed, similar reasoning has been used to justify a priori the choice of a given TN machine learning architecture by characterizing the spatial scaling of the entanglement entropy [137, 157] or the mutual information [46, 146] for bipartitions of the input features. These studies were based primarily on the analysis of image data, which possess a precise notion of spatial arrangement. Similar analyses could be performed by measuring correlations between individual features x_i or their product state embedding $\varphi(x)$, in the original spirit of Stoudenmire and Schwab [238]. However, since our TN models operate directly on the tensor structure of the input descriptors, the number of available partitions of the tensor indices is small. Hence, an analysis and interpretation of their scaling with subsystem size is limited, except perhaps at large order ν. Instead, in this section, we scrutinize the entanglement learned by the models themselves, conditioned on the target property.

Let us recall that the entanglement entropy is a zero-temperature quantum analogue of the classical Shannon entropy. We will adhere to common practice by taking the entanglement entropy to mean specifically the bipartite, von Neumann entropy,

$$S_{\mathrm{vN}}(\rho_{\mathcal{A}}) \coloneqq -\operatorname{Tr}\left(\rho_{\mathcal{A}} \log \rho_{\mathcal{A}}\right) \ , \tag{4.23}$$

where $\rho_{\mathcal{A}}(|\psi\rangle)$ is the reduced density matrix of a subsystem $\mathcal{A}$ of dimension $d_{\mathcal{A}}$, defined by tracing out the degrees of freedom of a complementary subsystem $\mathcal{B}$ in a pure state $|\psi\rangle$,

$$\rho_{\mathcal{A}}(|\psi\rangle) = \operatorname{Tr}_{\mathcal{B}}\left(|\psi\rangle\langle\psi|\right) \ . \tag{4.24}$$

The entanglement entropy can be computed from the coefficients of a Schmidt decomposition of the pure state $|\psi\rangle$,

$$|\psi\rangle = \sum_{k=0}^{\min(d_{\mathcal{A}},d_{\mathcal{B}})} \lambda_k |\psi_{\mathcal{A}}^k\rangle |\psi_{\mathcal{B}}^k\rangle \ . \tag{4.25}$$

The Schmidt spectra λ_k are the singular values of the of matrix $C_{\mathcal{AB}}$, where $|\psi\rangle = \sum_{\psi_{\mathcal{A}},\psi_{\mathcal{B}}} C_{\mathcal{AB}} |\psi_{\mathcal{A}}\rangle |\psi_{\mathcal{B}}\rangle$. Since the eigenvalues of the reduced density matrix $\rho_{\mathcal{A}}$ are λ_k^2, the entanglement entropy reduces to $S_{\text{vN}}(\rho_{\mathcal{A}}) = \sum_k -\lambda_k^2 \log(\lambda_k^2)$.

We can formulate an analogous entanglement entropy for the MPO models (4.18) by constructing a canonical purification [263, 272]. A vectorization $|T\rangle$ of the MPO $\hat{T}$ follows from the Choi isomorphism [272] (Figure 4.13), from which an effective pure state density matrix can be defined, $\mathcal{Q} = |T\rangle\langle T|$. We note that under this isomorphism, an MPO with physical dimensions d is mapped to an MPS with physical dimension d^2. Equivalent definitions of the entanglement entropy and Schmidt spectra thus follow from (4.23) and (4.24), replacing ρ by $\mathcal{Q}$ and $|\psi\rangle$ by $|T\rangle$.

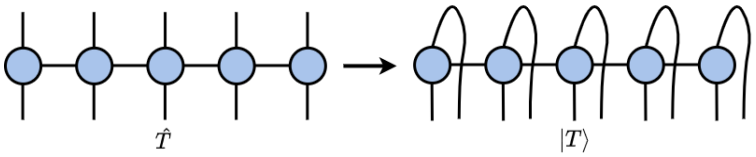


Figure 4.13: The Choi isomorphism constitutes a vectorization $|T\rangle$ for MPO $\hat{T}$.

Figure 4.14 compares average errors and entanglement entropies of the MPS and MPO models, as well as a TTN model (Figure 4.29), as a function of their virtual bond dimensions. The average is taken over 10 training-validation splits with a fixed testing set, and we test two sizes (100 and 1000 structures) of the training set. For the entanglement entropies, we consider two bipartitions of the descriptor indices: (1) a contiguous subsystem $\mathcal{A}$ consisting of chemical degrees of freedom $\{\alpha_1\alpha_2\}$ and $\mathcal{B}$ containing the structural components $\{n_1n_2l\}$, and (2) a noncontiguous subsystem $\mathcal{A}$ consisting of a chemical and radial component $\{\alpha_1 n_1\}$ with $\mathcal{B}$ containing the remaining degrees of freedom. These bipartitions are shown as insets

in Figures 4.14c,d. For comparison, we also plot the errors and entanglement entropies for the corresponding models using a dense ACE tensor (Figure 4.5), and we indicate the bond dimension $\chi \leq \chi_{\text{ct}}$ for which the number of TN model parameters is less than the dense ACE tensor. We will refer to this bond dimension χ_{ct} as the "compression threshold."

For the three TN architectures, we generally find that improved test errors are strongly correlated with larger entanglement enabled by higher bond dimensions. This is particularly true for the smaller training set, which requires larger bond dimensions before reaching minimal errors. The TN models outperform the ACE models once the bond dimension reaches the compression threshold, that is, once the number of parameters exceeds that of the ACE tensor. Again, this improvement is more significant in the small training set than in the larger one. In the larger training set, the test errors of the TN models become competitive with the ACE model well before reaching the compression threshold. This behavior is somewhat counterintuitive in the context of classical statistical learning, where one expects the generalizability of an overparameterized model to degrade by overfitting on small amounts of training data, a consequence of the so-called *bias-variance tradeoff* [98]. However, this phenomena is not uncommon in deep neural networks, where models with trainable parameters greatly exceeding the training sample size nonetheless perform well on unseen data [25, 179, 180].

We observe that the average entanglement entropies converge to values well below the theoretical limits set by the min-cuts in the network [50] and that this convergence closely follows the formation of plateaus in test errors. Similar sub-maximal convergence in the entanglement entropy was also noted in [77] for MPO layers applied to the MNIST dataset, so we expect that this is a general phenomenon when the TN model is complex enough for the learning task. However, the converged entropies in our work appear to be nonuniversal, exhibiting strong dependence on the structure of the network. Indeed, we generally find that networks which can host stronger entanglement entropy scaling with subsystem size relative to others (for instance, logarithmic versus area law scaling between TTNs and MPSs) tend to converge to higher values. Furthermore, for fixed virtual bond dimension, we find that the

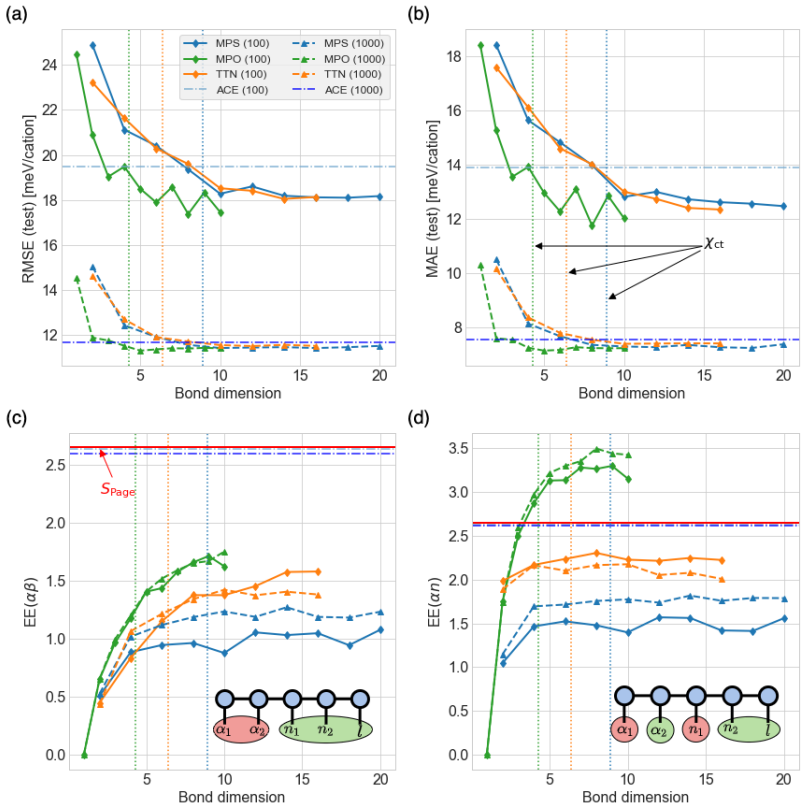


Figure 4.14: Test RMSE (a) and MAE (b) as a function of the virtual bond dimension used in MPS, MPO and TTN models, for training sets containing either 100 or 1000 structures. This is compared to the performance of a model using a dense ACE tensor (dashed, horizontal lines), and the corresponding compression thresholds, χ_{ct}, for each TN model are marked with vertical, dotted lines. The associated entanglement entropies (c,d) are plotted for different bipartitions (insets; partition $\mathcal{A}$ in red, $\mathcal{B}$ in green), and the corresponding Page entropy, S_{Page}, is indicated by a solid, red line. Note that because the MPO model acts on two copies of the input tensor, the effective physical dimensions in each partition (as an MPS under the Choi isomorphism) are larger. The Page entropy associated with this larger Hilbert space is not shown.

MPO model outperforms the MPS model and achieves higher entanglement. This appears consistent with the findings [81] of greater representational efficiency of MPDOs versus MPSs in probabilistic graphical models, although, for this work, it is in the context of regression. Hence, under the same arrangement of virtual bonds, the entanglement entropy seems to provide a consistent measure of the TN model's generalizability.

Nonetheless, comparisons of the entanglement between trained models with distinct virtual bond topologies are difficult to make. For instance, the dense ACE tensors perform somewhat worse than the other TN models, but achieve much higher entanglement entropies. In this case, the entanglement entropies do approach a universal value, that of the Page entropy [192],

$$S_{\text{Page}} = \frac{1 - d_{\mathcal{A}}}{2d_{\mathcal{B}}} + \sum_{k=d_{\mathcal{B}}+1}^{d_{\mathcal{A}}d_{\mathcal{B}}} \frac{1}{k} \tag{4.26}$$

defined as the average entanglement entropy of a pure state randomly drawn from the entire Hilbert space. This would not be entirely surprising for an untrained ACE tensor $w^{(\nu)}$ treated as a length $(|\alpha|^\nu |n|^\nu |l|^\nu)$ vector since the individual elements are initialized according to a standard normal distribution $\mathcal{N}(0, \sigma^2)$. However, it is surprising that the high entanglement of this initial state is essentially preserved under SGD with a mean square error (MSE) loss function, at least up until early stopping. This may provide some explanation for the improved performance of the models with an explicit TN factorization: the entanglement constraints imposed by the network itself help drive the model toward minima in the loss landscape which are characterized by entanglement entropies more finely tuned to the target property.

It has long been recognized that the entanglement entropy alone is insufficient to fully capture the entanglement of general quantum states and that much richer structure can be found in the full entanglement spectrum $\{\lambda_k^2\}$ or its logarithms $\{\xi_k \coloneqq -\log(\lambda_k^2)\}$ [138]. The level spacing statistics between consecutive pairs of eigenvalues have proven useful in characterizing the irreversibility in quantum circuits [36, 220] and emergent entanglement

complexity in many-body quantum dynamics [281, 289, 275, 109]. For entanglement spectra arranged in descending order $\lambda_0^2 \geq \lambda_1^2 \geq \ldots$, let $s_k \coloneqq \lambda_k^2 - \lambda_{k+1}^2$ denote the kth level spacing. To avoid unfolding the spectrum [160], a procedure sensitive to spurious finite size effects, it is common practice [188, 193, 281, 8] to study the ratio of consecutive level spacings $r_k = s_k/s_{k+1}$ or the related quantity $\widetilde{r}_k = \min(r_k, 1/r_k)$, which are independent of the level density of states. It was shown in [36, 220, 281] that either the distribution of spacings $P(s)$ or their ratios $P(r)$, collectively referred to as the *entanglement spectrum statistics* (ESS), provides a measure of a quantum state's entanglement complexity, defined by the existence of an efficient algorithm that completely disentangles the state. Complexly entangled states, for which efficient disentangling algorithms fail, were found to feature ESS with Wigner-Dyson (WD) statistics,

$$P_{\mathrm{WD}}(r) = \frac{1}{Z_\beta} \frac{(r + r^2)^\beta}{(1 + r + r^2)^{1+(3/2)\beta}} \ , \tag{4.27}$$

where Z_β is a normalization factor, and the Dyson index, β, specifies one of three Gaussian random matrix ensembles [8]. On the other hand, states which could be efficiently disentangled possessed Poisson-like ESS, $P_{\mathrm{Poisson}}(r) = 1/(1 + r)^2$. An important feature that distinguishes WD from Poisson level statistics is the presence of level repulsion, $P_{\mathrm{WD}}(r \to 0) \sim r^\beta \to 0$, which reflects universal statistical correlations between adjacent levels [160]. We also note that there appears to be a deep connection between quantum circuits capable of universal computing and ESS [36, 220, 281, 289], in that a universal gate set gives rise to WD statistics. However, a recent study [109] showed that the converse is not true in general, finding that a class of classically simulatable circuits can prepare states with WD ESS.

It is reasonable to expect the above picture to hold, to some extent, for randomly initialized MPS. Indeed, a random MPS, viewed as a unitary embedding [78, 89, 145], forms an approximate 2-design [94]: as a random quantum circuit, the first and second moments approximate those of a Haar distribution. Moreover, Haar-distributed random unitary circuits possess WD ESS corresponding to the Gaussian unitary ensemble (GUE), $\beta = 2$ [289], although a precise relationship between ESS and k-designs remains an open question [109].

Nonetheless, we will explore the ESS of our MPO model, viewed as an MPS under the Choi isomorphism. For simplicity, we consider the entanglement spectrum for the $\{\alpha n\}$ bipartition shown in Figure 4.14d and analyze its evolution under SGD for a fixed training set. To obtain sufficient statistics, we collect data over 100 independent runs of 1000 SGD epochs, and we study models with virtual bond dimensions for which the spectrum is either truncated ($\chi = 12$), marginal ($\chi = 16$), or full-rank ($\chi = 20$).

Figure 4.15 shows both the logarithm of the average spectrum $\langle \xi_k \rangle$ and the ESS at different time steps. First, we find that for all cases the entanglement spectrum deviates from the Marchenko-Pastur (MP) law for Haar-distributed random states sampled from the full Hilbert space [297, 70]. Interestingly, close examination of the truncated ($\chi = 12$) model reveals residual MP structure in the tail of the spectrum, similar to the two-component structure discussed in [280], and this MP tail persists under SGD up to a constant shift from normalization. For all models, upon random initialization of the MPS tensors, the spectrum is relatively flat and thus highly entangled. The largest changes in the spectrum occur at the earliest stages of training, and the spectrum quickly converges to a configuration in which the local level density has decreased. We also find that the largest spectra (low ξ_k) approximately converge to the same values, regardless of the bond dimension, which is consistent with the prior observation that the model retains high accuracy for smaller bonds.

At initialization, all three cases display WD ESS corresponding to the Gaussian orthogonal ensemble (GOE), $\beta = 1$. This may not be entirely surprising given the above discussion, although in this case the generating random matrix ensemble is real-valued. However, as the state evolves under SGD, the ESS retains a substantial degree of its GOE character. To measure the difference between the GOE ESS, $P_{\mathrm{GOE}}(r)$, and the observed ESS, $P_{\mathrm{obs}}(r)$, we plot the Kullback-Leibler (KL) divergence, $D_{\mathrm{KL}}\left(P_{\mathrm{GOE}} || P_{\mathrm{obs}}\right) = \mathrm{Tr}_r\left[P_{\mathrm{GOE}} \log\left(P_{\mathrm{GOE}}/P_{\mathrm{obs}}\right)\right]$, in the insets to Figure 4.14. First, we note that the higher starting values of D_{KL} in the truncated and marginal spectra relative to the full-rank case are likely due to finite-size effects imposed by the bond dimension. We see that at early times there is rapid growth in the KL divergence followed by a regime in which D_{KL} increases very slowly, accompanied by

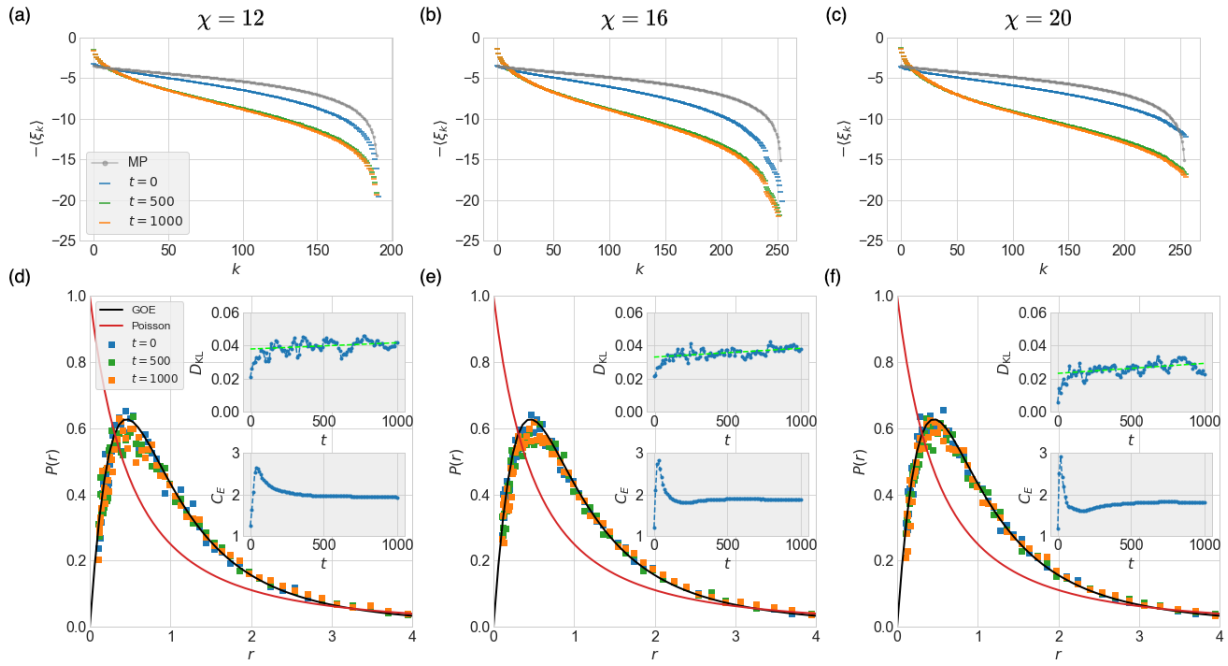


Figure 4.15: (a-c) The average entanglement spectrum of the MPO model as a function of training step, t, compared to the Marchenko-Pastur (MP) law. (d-f) The distribution $P(r)$ of the ratio r of level spacings in the entanglement spectrum (referred to as the entanglement spectrum statistics (ESS) in the main text) closely follows the Gaussian orthogonal ensemble (GOE). Under stochastic gradient descent (SGD), the KL divergence (upper insets) exhibits a short, initial regime of fast growth followed by a slow growth regime. A linear fit to the KL divergence in the slow growth regime (dashed, light blue line) is included to help guide the eye. The crossover between the fast and slow growth in the observed ESS relative to the GOE, coincides with a divergence in the capacity of entanglement, C_E, (lower insets) signaling a phase transition in the entanglement spectrum.

oscillations due to the stochasticity of the optimization. The onset of this slow growth regime prevents the ESS from significantly deviating from the GOE over the course of training. As we noted above, the regime of rapid growth is accompanied by a decrease in the level density,

which essentially stabilizes in the slow growth regime.

Recalling the level repulsion in the GOE, this expansion of the level density resembles the free expansion and re-equilibration of a one-dimensional Coulomb gas following a sudden quench of its confining potential. Specifically, the joint probability distribution of N eigenvalues in the GOE is precisely a Boltzmann-Gibbs factor for a gas of N particles with pairwise, repulsive $-\log(|\lambda_i - \lambda_j|)$ interactions in a quadratic potential [70]. This picture suggests that SGD evolves the entanglement spectrum and the ESS by evolving the confining potential of the joint level distribution. Furthermore, if this potential is driven far enough, we should expect the entanglement spectrum to undergo a phase transition. We can verify that this is the case by calculating the so-called *capacity of entanglement* [282],

$$C_E \coloneqq \langle H_{\text{ent}}^2 \rangle - \langle H_{\text{ent}} \rangle^2 \ , \tag{4.28}$$

where $H_{\text{ent}} \coloneqq -\log \rho_{\mathcal{A}}$ is the *entanglement Hamiltonian* [138, 54]. The capacity of entanglement measures fluctuations in the entanglement spectrum, analogous to the heat capacity in classical statistical mechanics. The insets to Figure 4.14 show the evolution of C_E for the three cases. We observe that the crossover from the fast to slow growth regimes in the KL divergence coincides with a divergence in the capacity of entanglement, signaling a phase transition in the entanglement spectrum. Again, this transition is smoothed out in the truncated and marginal cases due to finite-size effects.

The above discussion provides a mechanism through which the model entanglement adapts to the learning task. Crucially, while the average entanglement spectrum converges to a specific configuration, particularly near its low-ξ_k edge, the ESS maintains universal characteristics of WD statistics, namely level repulsion. Thus, the learned state retains the entanglement complexity from its random initialization. This is consistent with its strong generalizability in that the model is expressive enough to capture complex correlations in the high-dimensional feature space relevant to the target prediction. Moreover, this behavior appears to have close connections to overparameterized neural networks [4, 140] and quantum circuits [275, 119], where SGD applied to sufficiently overparameterized models

finds solutions with small generalization error that are nonetheless close to their random initialization.

4.3.3 Latent space encoding

As discussed in Section 4.2.1, TN factorizations of the learnable states $|\psi^{(\nu)}\rangle$ can be understood as multilinear maps which internally couple structural and chemical degrees of freedom. In this section, we explore this idea further by visualizing the latent tensor product spaces learned by a deeper TN architecture (Figure 4.29) corresponding to a MERA with fixed virtual bond dimension. To do so, we use t-distributed stochastic neighbor embedding (t-SNE), a nonlinear dimensional reduction method that preserves the relative locality between points in the dataset [105, 261, 161]. Again, the MERA network is trained to predict the mixing enthalpy (4.21) on a 1000-structure subset of the full NMD18 dataset, and we employ t-SNE to embed the tensor product spaces mapped by subsequent hidden layers.

Figure 4.16 shows the evolution of the latent space image of each datapoint as function of depth in the MERA network. Note that we apply t-SNE to the entire dataset, not just to the training set. The input average SOAP descriptors, $\overline{B}^{(2)}$, clearly exhibit a high degree of structure, to the extent that t-SNE surprisingly reproduces the ternary phase diagram of the alloy. The first layer of the network reorganizes the data in a hierarchical fashion, grouping structures with the same space group into local clusters. Within each cluster, the separation of structures according to alloy composition is also maintained. As a byproduct, we also see that a significant number of high energy structures in the data occur for space groups I$a\overline{3}$ and $Fd\overline{3}m$ which contain a significant amount of Al. At the final hidden layer, this hierarchical structure has been replaced by a single quasi-1D cluster, with an orientation determined by the mixing enthalpy. This conforms with the intuition provided by a renormalization group interpretation of the network [265, 65, 162, 142], in the sense that lower levels in the network resolve finer details in the structure representations, and these are subsequently coarse grained in mapping to the final target quantities.

4.3.4 Compression of large descriptors

While the expansion (4.16) is mathematically well-defined [9], the exponential scaling of the descriptors $B^{(\nu)}$ with the number of distinct chemical species and the cutoffs of the radial and angular momentum channels can yield exceptionally large feature spaces. This can impose computational limitations on the application of machine learning to atomistic modeling, particularly when the structures in the dataset span compositions with a large number of elements. Thus, recent efforts [276, 110, 6, 82, 67, 221] have been made toward optimizing

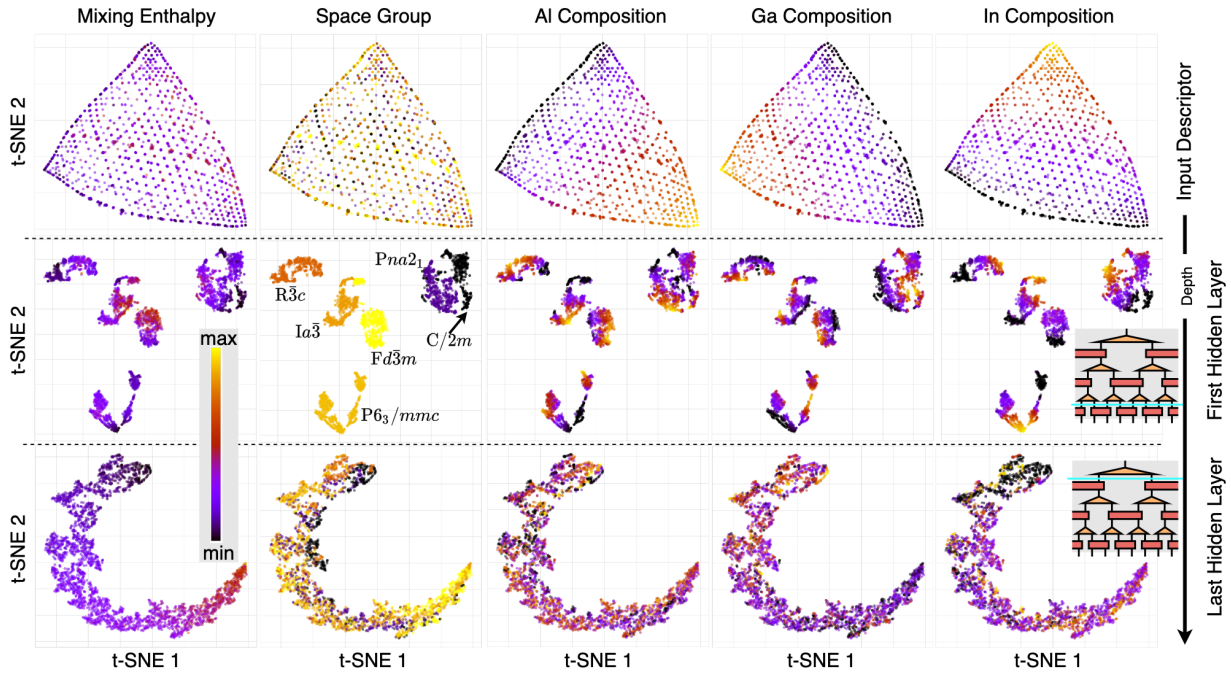


Figure 4.16: Evolution of t-distributed stochastic neighbor embeddings (t-SNE) of the NMD18 dataset with increasing depth in a MERA tensor network model (insets). The regularity of the encoding of chemical composition in the input SOAP descriptor is reorganized by the first hidden layer according to the lattice symmetries of the structures. The features are subsequently coarse-grained by fitting the model to the mixing enthalpy, yielding a quasi-one dimensional encoding ordered by high and low enthalpy.

the efficiency of atomic representations while maintaining their symmetry invariance. This was, for example, a central motivation for the introduction of generalized kernel functions [277], alchemical or otherwise (cf. Section 4.2.1).

As discussed above, one way to alleviate the computational complexity of learning with high dimensional descriptors is to employ TN factorizations of the model architectures, which impose a kind of structured sparsity on the trainable model weights. Similar constructions can be applied to the input descriptors. In this section, we explore this possibility with two methods. First, we validate the MPS factorization scheme (Figure 4.8) discussed in Section 4.2. Second, we implement an autoencoder using MPOs (Figure 4.18). In principle, these two methods can be combined, both in training the autoencoder and in subsequent supervised learning with the encoded input tensors, although we do not test this combination here.

To test these methods, we use the BA10 dataset [186], which consists of 15950 binary alloys and their mixing enthalpies. The binary alloys span composition with 10 different metals and 3 different lattice symmetries, face-centered cubic (fcc), body-centered cubic (bcc) and hexagonal close-packed (hcp), with volumes up to 8 atoms per unit cell. It is worth noting that, like NMD18, the structures in BA10 retain their idealized lattice positions, while the volumes are scaled according to Vegard's law. However, the DFT mixing enthalpies are computed for these unrelaxed structures, as opposed to the relaxed ground states in NMD18. Hence, the prediction accuracy in prior studies [186, 130] was found to be more limited by the descriptor parameters than implicit noise in the dataset. In addition to the large number of distinct chemical species, accurate models based on the SOAP power spectrum, $B^{(2)}$, require a fairly large number of radial ($n_{\max} = 8$) and angular momentum ($l_{\max} = 8$) channels [186, 130]. As in [130], we will consider training-testing splits of 1600 and 1000 structures, respectively, for both the unsupervised task of training the autoencoder and the subsequent supervised task of predicting the mixing enthalpies. For the latter supervised task, we use an MPS model with the encoded tensors as input. Given the exponential scaling of the descriptor dimensions, the input state $|\overline{B}^{(2)}\rangle$ retains only the independent entries of the original descriptor tensor. Furthermore, zero-padding is added to the state $|\overline{B}^{(2)}\rangle$ to allow

for uniform physical dimensions in the subsequent MPS and MPO tensor networks. Note that this does not affect the symmetry invariance of the descriptor, but the physical dimensions of the MPS and MPO architectures in this analysis do not admit the same correspondence with the structural kernels discussed above. Nonetheless, the high accuracies obtained with these models underscores the flexibility in choosing the TN factorization and their broader applicability in machine learning [185, 238, 80, 77, 46, 157, 146].

Figure 4.17 shows the change in the test errors as a function of the maximum bond dimension, χ_{in}, for a MPS factorization of the descriptor state $|\overline{B}^{(2)}\rangle$. In this case, the factorized descriptor is used as input to a supervised MPO model ($\chi = 4$). This is compared to the performance of the unfactorized, dense $\overline{B}^{(2)}$ tensor, and we also plot the corresponding number of input parameters, N_{in}. As expected, retaining large bond dimensions results in small deviations from the baseline error, while the efficiency of contracting the tensor network is substantially improved. The test errors grow as the bond dimension decreases, but surprisingly, this growth is nonmonotonic. Thus, despite reducing the number of input parameters by up to two orders of magnitude, the test errors remain relatively stable, and there are certain smaller bond dimensions that outperform larger ones. The origin of this nonmonotonic behavior is unclear. Naively, one would expect that discarding lower singular values at each virtual bond would introduce noise into the inputs, eventually degrading the model performance. However, it appears that some competing process is present. One possibility is that compressing the input tensors can improve the quality of the loss landscape (i.e., its smoothness or convexity), similar to overparameterization in deep neural networks [4, 140]. If that is the case, then deciding how to balance the input factorization against the TN model architecture would be an interesting open question.

We mentioned previously that an MPO applied to an input state can be viewed as a multilinear map from the atom-centered Hilbert space to a potentially lower dimensional space. Moreover, the topology of certain hierarchical tensor networks, such as TTNs and MERAs, can be viewed as enforcing this kind of dimensional reduction (see Section 4.3.3). For instance, it has recently been recognized that there is a fundamental correspondence

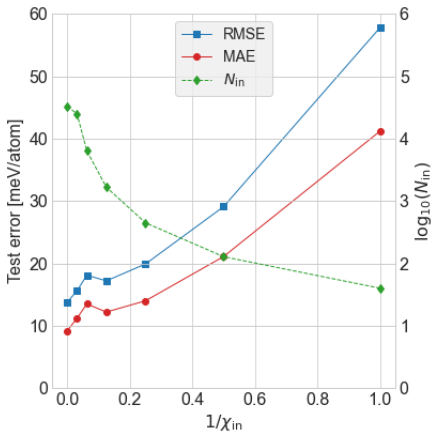


Figure 4.17: The test performance of a supervised MPS model using MPS factorizations of the dense input descriptors on the BA10 dataset. The number of input parameters for a MPS factorization with maximum bond dimension χ_{in} is also plotted. The case where the exact input descriptors are used ($1/\chi_{\text{in}} \to 0$) is included for comparison.

between MERAs and wavelet transformations. We can exploit this general idea to learn approximately faithful compressions of the input descriptors. To do so, we employ MPOs in an autoencoder setup, shown in Figure 4.18, where the output dimensions, d_{latent}, of the MPO tensors are less than the input dimensions, d_{in}. In this setting, the *encoder* MPO maps the input state to a lower dimensional latent space, while the *decoder* MPO attempts to invert this transformation. The autoencoder network is trained by SGD to minimize the MSE between the components of the input state and its reconstruction by the decoder. The trained encoder contracted with the input state (Figure 4.18b) yields a compressed tensor adapted to the data.

In Table 4.1, we record the test reconstruction errors (RMSE and MAE) for autoencoders with different $d_{\text{latent}} < d_{\text{in}} = 8$, as well as the test performance of the encoded inputs in predicting the mixing enthalpies with an MPS model. The bond dimensions of the autoencoder MPOs are fixed ($\chi_{\text{enc}} = \chi_{\text{dec}} = 6$), as are the bond dimensions of the supervised MPS model

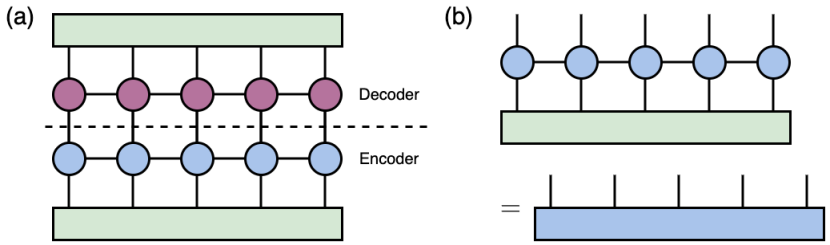


Figure 4.18: (a) An autoencoder based on MPOs used to compress the physical dimensions of large input descriptor tensors. (b) A compressed tensor is given by the contraction of the encoder MPO with the original input.

($\chi_{\text{mps}} = 10$). The compression mediated by the encoder is measured against the original size of the input state, $d^n_{\text{latent}}/d^n_{\text{in}}$. We find that the MPS model performance in predicting the mixing enthalpies remains high even for encoded inputs retaining a few percent of the original number of parameters.

Thus, we have found that both an MPS decomposition and an MPO-based autoencoder are effective methods to reduce the computational complexity imposed by the input descriptors. It is again worth emphasizing the parametric efficiency of these methods compared to more standard deep learning architectures. For example, the deep convolution neural network in the original BA10 work [186] uses $\mathcal{O}(10^8)$ parameters combined with many-body tensor representations (MBTR) [108], which lead to feature vectors with $\mathcal{O}(10^5)$ entries. On the other hand, the high-performing MPS / MPO models possess $\mathcal{O}(10^3)$ parameters with compressed input tensors of similar size.

4.4 Discussion and Summary

In this chapter, we have demonstrated that tensor networks provide both a unifying formalism for a large class of atomic structure representations as well as robust machine learning architectures based on them. Specifically, we detailed how the ACE framework admits a natural description in terms of symmetric tensor networks. This formalism allows for the

Table 4.1: The performance of a 5-site MPO-based autoencoder with virtual bond dimension $\chi_{\text{enc/dec}} = 6$ is measured by the RMSE (MAE) for reconstructing input descriptors in a test set. This test reconstruction error (r-Error) is shown for varying the latent dimensions, d_{latent}, of the encoder, and the subsequent compression of the input $d^5_{\text{latent}}/d^5_{\text{in}}$ is also included. The test RMSE (MAE), denoted s-Error, characterizes the performance of a supervised MPS model ($\chi_{\text{mps}} = 10$) applied to the compressed inputs.

d_{latent}	Compression	r-Error	s-Error [meV/atom]
-	1	-	14.8 (9.9)
6	0.237	0.060 (0.031)	16.3 (11.2)
4	0.031	0.088 (0.047)	19.3 (13.9)
2	9.76×10^{-4}	0.137 (0.070)	103.2 (77.8)

transparent application of SO(3) recoupling theory to the construction of a rotationally invariant basis and generalizes straightforwardly to SO(3)-equivariant cases. By taking seriously the Hilbert space structure of atom-centered descriptors [277], we have shown how tools commonly applied in quantum many-body physics can be repurposed in coarse-grained surrogate models of fundamental material properties.

Indeed, the map between a solid-state or molecular structure and a given property can be written as an inner product between a learnable state and a sufficiently representative descriptor, both of which admit low-rank tensor network factorizations. We have found that, as learnable architectures, MPOs and MPSs exhibit strong generalizability in common learning tasks in material informatics, and their performance is particularly notable when the amount of available data is very limited. We provided two routes toward understanding this apparent inductive bias. First, we showed how learnable tensor networks, viewed as multilinear maps on the atom-centered Hilbert space, can be related to generalized kernel functions. An appropriate choice of network topology thus provides an explicit way of coupling structural and chemical components in the model. Unlike standard structural kernel methods, the

evaluation of these TN models scales more efficiently with the number of structures. This is similar to the application of neural networks, and the MPS models benchmarked in this work can be viewed as parametrically efficient regularizations of a dense neural network. Indeed, the apparent compressibility of the descriptor space and the learning architecture can be used to obtain more efficient atomic representations, and we demonstrated this using both a standard MPS decomposition algorithm as well as a MPO-based autoencoder. The second way to understand the strong generalizability of these TN models is more fundamental and potentially a universal feature of tensor-network machine learning more broadly. In this case, the learned TN states possessed signatures of high entanglement complexity. In particular, the initial WD statistics of the level spacings in the entanglement spectrum were essentially preserved under stochastic gradient descent, while the spectrum itself adapted to the learning task.

The above discussion points to several directions for potential future work. First, general TN factorizations can be used to replace fully connected layers common in deep learning architectures, a program already undertaken in [185, 77]. In the context of materials prediction, adding TN layers in end-to-end architectures like graph neural networks is an effective way to reduce memory overhead and could potentially improve performance on smaller training sets. In certain cases, such as Euclidean neural networks, the representations learned by the network could be used more or less directly as inputs to the methods demonstrated in this work. Similarly, TN layers could be implemented in the Behler-Parrinello neural network framework [21, 22] for modeling high-dimensional potential energy surfaces. In this case, separate tensor networks would replace the neural networks applied to each local environment. On a related note, it would be worth quantifying the performance of these networks on very large datasets and identifying, in particular, whether there is some advantage to using deep versus shallow tensor networks [142].

Let us mention that training TNs with a large number of tensors, as would be the case (4.20) for high orders ν and ζ, using SGD suffers from the presence of barren plateaus in the loss function. For a large MPS, this is intrinsically related to the fact that, under random

initialization, the state constitutes an approximate 2-design for which the expected gradients vanish [145]. It may be possible to overcome this problem by choosing a different initialization scheme [89], or by preconditioning the network with a different algorithm, such as DMRG. Formulating a local loss function should also alleviate the issue [145]. Alternatively, one could utilize common strategies employed in deep learning for regulating gradients. For example, a local tensor network scanned across sub-partitions of an input tensor provides a TN analogue of weight sharing in a convolutional neural network. Residual skip connections may also prove valuable in deeper networks. A related question underlying these approaches is the extent to which they lead to complexly entangled states. If WD statistics in the entanglement spectrum is a fundamental signature of a highly generalizable model, it would be worth characterizing, in general, how the entanglement spectrum in approximate k-designs evolves under stochastic gradient descent.

From a practical standpoint, the parametric efficiency of these methods as well as their high performance on limited training data makes a strong case for their application in large-scale materials simulation and high-throughput screening, particularly when the calculation of target parameters from first principles is computationally demanding. Furthermore, while we have characterized these methods on the important class of atom-density representations, we expect that they are more broadly adaptable to generic material feature spaces [190].

4.4.1 Additional computational details

SOAP power spectra were computed with the DSCRIBE software package [104]. GPR models were trained with the SCIKIT-LEARN library [195] using expectation maximization. The construction and training of deep neural networks and tensor network models were facilitated by the TENSORNETWORK [211] and TENSORFLOW [1] libraries. These models were trained using the Adam optimization algorithm, a variant of stochastic gradient descent that adaptively updates the learning rates based on moments of the gradients [120].

4.4.2 Tensor network graphical notation

An arbitrary state $|\psi\rangle = \sum_{\{s\}} \psi_{s_1 s_2 \cdots s_n} |s_1 s_2 \cdots s_n\rangle$ in a tensor product space $|\psi\rangle \in \mathcal{H}^{\otimes n}$, where each $|s_k\rangle$ forms a basis for a finite-dimensional vector space $\mathcal{H}$, can be described by the tensor $\psi_{s_1 s_2 \cdots s_n}$ formed by its coefficients. In the standard diagrammatic notation of tensor networks, an arbitrary, dense tensor, $\psi_{s_1 s_2 \cdots s_n}$, of order n is represented by a shape or node with n open edges. Summation over a matching pair of tensor indices, otherwise referred to as *contraction*, is represented by a closed edge in the network. The familiar example of matrix multiplication is shown as a tensor network in Figure 4.19, as well as a less trivial network.

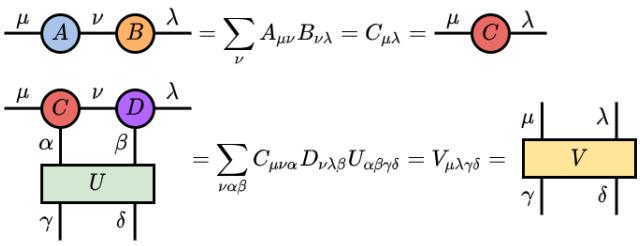


Figure 4.19: Examples of graphical notation for tensor networks. Tensor indices associated with each tensor in the network are represented by open edges. Closed edges between pairs of tensors indicate summation / contraction over the matching tensor indices.

4.4.3 Additional SO(3) recoupling relations

In this appendix, we collect some useful identities in the algebra of SO(3) representations. Recalling the diagrammatics from the main text, each line with an arrow is labelled by an irrep l and spans an associated subspace of dimension $d_l = 2l+1$. Intertwiners between irreps are given by the usual Clebsch-Gordan (CG) coefficients (Figure 4.1c,d), which satisfy the orthogonality relations shown in Figure 4.20a,b. The CG tensors commute with the action of the group (Figure 4.20d for C and analogous for $C^\dagger$); that is, they form a natural (i.e.,

equivariant) transformation. Note that the following three identities are sufficient to derive the generalized Wigner-Eckhart theorem discussed in the main text (Figure 4.2).

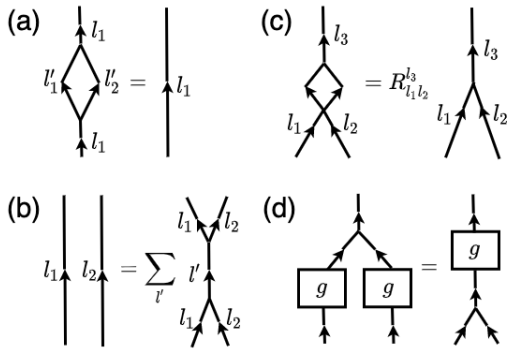


Figure 4.20: Additional recoupling relationships between Clebsch-Gordan tensors and Wigner D-matrices, including (a) orthogonality, (b) completeness, and (d) naturality /equivariance. (c) Braiding between irrep edges is defined in terms of a set of R-symbols, which reduce to sign factors in SO(3).

The recoupling of three or more irreps can be described by the contraction of multiple CG tensors to form a fusion tree, and recoupling schemes represented by different fusion trees can be related to each other via unitary transformations known as F-symbols (Figure 4.21a,b). In SO(3), explicit values for the F-symbols can be derived from the Wigner $6j$ symbols (or alternatively from Racah W coefficients), which are given by the maximally irreducible diagram formed by the contraction of two 4-valent fusion trees. Additionally, pairs of irreps can be swapped (Figure 4.20c) using the so-called R-symbols. In SO(3), they take the explicit values $R^{l_3}_{l_1 l_2} = (-1)^{l_1+l_2-l_3}$ and can be used to constrain to the parity of the state. For instance, an SO(3)-invariant ν-order descriptor possesses inversion symmetry if $\sum_{k=1}^{\nu} l_k$ is even, which corresponds to $\left(\prod_{k=1}^{\nu-2} R^{l'_{k+1}}_{l'_k l_{k+2}}\right) R^{l'_1}_{l_1 l_2} = 1$ for $\nu > 2$ and $l'_{\nu-1} = 0$ in the recoupling scheme used in the main text.

Let us mention that additional coherence relations must be satisfied by the F- and R-symbols, namely the so-called pentagon and hexagon identities, where the pentagon identity

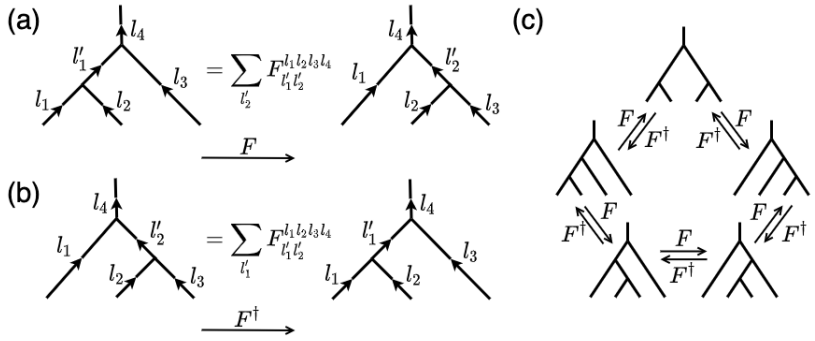


Figure 4.21: Unitary transformations (a,b) known as F-symbols define mappings between recoupling schemes. They must satisfy the consistency condition (c) known as the pentagon identity.

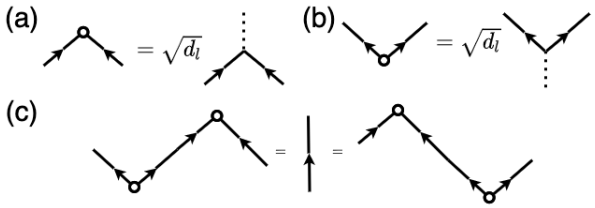


Figure 4.22: (a) Cap and (b) cup tensors defined in terms of Wigner $2jm$ symbols. Diagrammatically, they allow for the reversal of an irrep edge orientation, and satisfy the inversion identity (c).

is shown in Figure 4.21c. Indeed, the algebraic structure of finite-dimensional representations of SO(3) corresponds more generally to a symmetric tensor category. A useful aspect of this construction is the ability to encode nontrivial manipulations of tensors in the local deformations of diagrams. For example, the reversal an irrep arrow can be accomplished by contracting with normalized Wigner $2jm$ symbols, $\epsilon^{(l)}_{mm'} \coloneqq \sqrt{d_l} C^{(ll0)}_{mm'0} = (-1)^{l-m}\delta_{m,-m'}$, where the *cup* and *cap* tensors $\epsilon^{(l)}$, $(\epsilon^{(l)})^\dagger$ are matrix inverses of each other, $\epsilon^{(l)}(\epsilon^{(l)})^\dagger = (\epsilon^{(l)})^\dagger\epsilon^{(l)} = \mathbb{I}^{(l)}$ (Figure 4.22). The cup and cap tensors will be used to recursively construct equivariant descriptors in the following section.

4.4.4 Generalizing to SO(3) equivariants and beyond

In this appendix, we show how the SO(3)-invariant tensor networks in Section 4.1.1 can be generalized to SO(3)-equivariance and applied to the learning of vectorial and tensorial properties. In doing so, we will largely reproduce central results related to the general ACE formalism [60, 9], as well as the so-called λ-SOAP [85] descriptors and the recursive N-body iterative contraction of equivariants (NICE) framework [184].

First, we recall from the main text (Figure 4.2) that ν-order SO(3)-invariant tensors can be explicitly constructed by taking the Haar integral over the action of the group. This can be understood as a projection onto the invariant subspace, $\mathrm{Inv}(\mathcal{V}_{l_1} \otimes \cdots \otimes \mathcal{V}_{l_\nu}) \cong \mathrm{Hom}(\mathcal{V}_{l_1} \otimes \cdots \otimes \mathcal{V}_{l_\nu}, \mathbf{1})$, of the tensor product of ν irrep spaces $\mathcal{V}_{l_k}$. This projection operator,

$$P^{(l_1 \cdots l_\nu)} = \int dg D^{(l_1)}_{m'_1 m_1}(g) \cdots D^{(l_\nu)}_{m'_\nu m_\nu}(g) \tag{4.29}$$

$$= \sum_\iota \overline{\iota_{m'_1 \cdots m'_\nu}} \iota_{m_1 \cdots m_\nu} \tag{4.30}$$

is shown graphically in Figure 4.23, where for $\mathcal{G}$ a semisimple group this projection decomposes, as before, into dual fusion trees (i.e., higher-order intertwiners) factored through the trivial irrep space [187, 199]. Note that this decomposition is a direct consequence of the Peter-Weyl theorem [187]. We denote these general intertwiners by $\iota_{m_1 \cdots m_n}$, and the sum in (4.30) runs over the intermediate irreps, which depend on the chosen recoupling scheme.

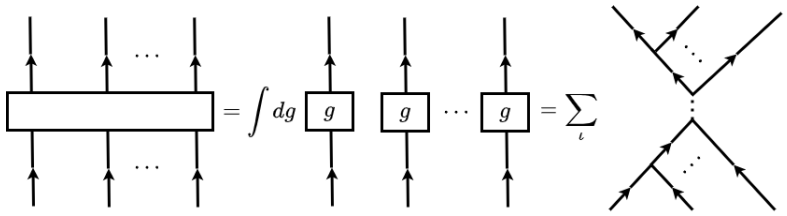


Figure 4.23: General projection operator onto the invariant subspace $\mathrm{Inv}(\mathcal{V}_{l_1} \otimes \cdots \otimes \mathcal{V}_{l_\nu})$ of an arbitrary tensor product space $\mathcal{V}_{l_1} \otimes \cdots \otimes \mathcal{V}_{l_\nu}$.

Under this projection, a ν-order SO(3)-equivariant descriptor, $B_L^{(\nu)}$, transforming as a state with definite angular momentum L can be constructed by symmetrizing over the product of ν atom-centered tensors, $A_{\alpha nlm}$, along with an identity operator carrying the target irrep space L, as shown in Figure 4.24. As before, the descriptor, $B_L^{(\nu)}$, corresponds to the reduced part of the decomposition. By bending upward the open L edge of the corresponding tensor network using cup and cap tensors, one can verify that the tensor product of ν irreps is mapped to the target L irrep, and the fully invariant case from the main text can be recovered by taking $L = 0$.

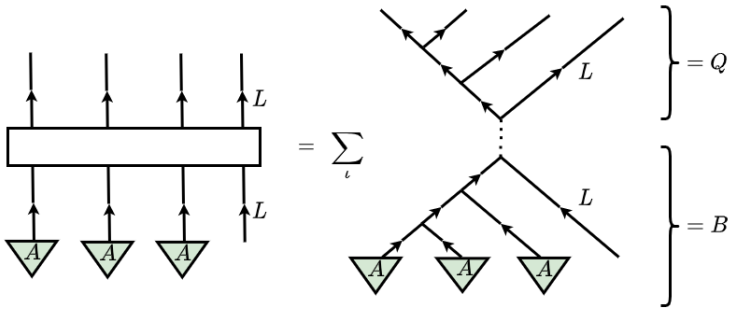


Figure 4.24: Covariant formulation of generalized Wigner-Eckhart theorem in terms of a generalized Haar projector. The reduced element B of this projection yields an equivariant form of the atomic descriptors.

From this picture, it is clear that a $(\nu + 1)$-order descriptor can be obtained by fusing together lower order descriptors. This is shown in our chosen recoupling scheme in Figure 4.25, where an order-1 descriptor, equivalent to the unsymmetrized atom-centered tensor A, is combined with an order-2 descriptor by contraction with the necessary CG, cup and cap tensors. Repeated application of this process yields a recursive formula for higher order descriptors.

The angular momentum L of an equivariant descriptor can be preserved by a tensor network model by acting only on the remaining $\{\alpha_k, n_k, l_k\}$ degrees of freedom. For example, an

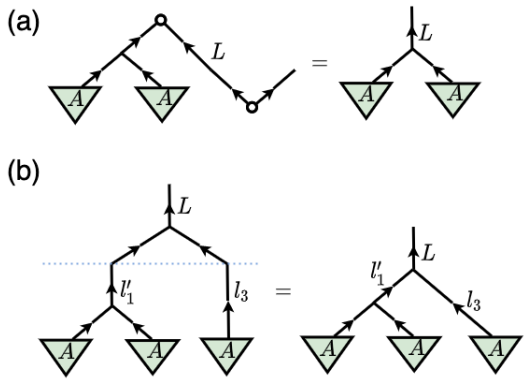


Figure 4.25: Elements of the basic recursive construction of order-($\nu+1$) equivariant descriptors, involving (a) the reorientation of the final momentum irrep edges using cup tensors and (b) the fusion of additional order-1 descriptors with an order-ν descriptor.

MPS-like model which maps to a target quantity $y_{sM}^{(L)}$ living in the space $|s\rangle \otimes |LM\rangle$, where s labels a possible additional degree of freedom, is shown in Figure 4.26a. Naively taking ζ-order tensor products of $L > 0$ equivariant descriptors destroys the angular momentum channel of the original state since products of irreps can be subsequently recoupled. Alternatively, as mentioned in [277], the effective many-body character of an equivariant model can be increased by taking tensor products of the equivariant state, $B_L^{(\nu)}$, with $\zeta - 1$ copies its invariant counterpart, $B_{L=0}^{(\nu)}$. An example $\zeta = 2$ MPO-type model is shown in Figure 4.26b.

A useful aspect of this formalism is that the fusion trees that encode the equivariance of the model can be manipulated purely algebraically and computed independently from the atom-centered A-tensors [216]. As demonstrated in [9, 150] with the ACE formalism, this allows for potential speedups in applications, such as molecular dynamics or Monte Carlo, which constantly update the underlying descriptors. In particular, the fusion trees can be recursively precomputed and contracted with the A-tensors along with the learned model weights at runtime.

Let us point out some implications for the constructions of equivariant structural kernels [85]. Figure 4.27a,b displays the diagrams for kernel functions corresponding to the λ-SOAP

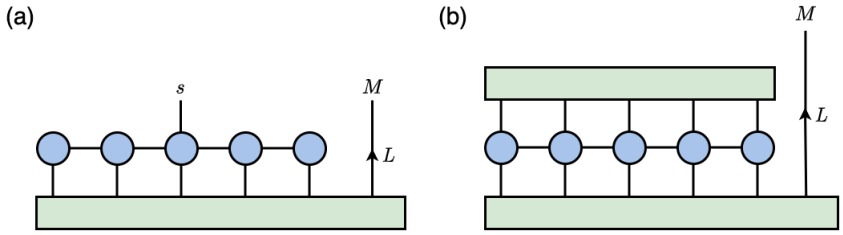


Figure 4.26: Tensor network models which preserve the equivariance of the input descriptor. (a) An example MPS model which allows for additional output degrees of freedom s. (b) An example MPO model which implicitly increases the effective many-body character of the input descriptor by contracting with an equivariant descriptor and its invariant counterpart.

generalization of the power spectrum ($\nu = 2$) and bispectrum ($\nu = 3$), respectively. The pairs of descriptors in these kernels are built from the same recoupling scheme, and for $L = 0$ it can be seen that each kernel function is essentially given by a spin network equivalent to a Wigner $3nj$ symbol [51] weighted by the trace over the remaining $\{\alpha_k, n_k, l_k\}$ degrees of freedom. For either ($L = 0$, $\nu > 3$) or ($L > 0$, $\nu > 2$), there exist topologically distinct kernel functions determined by the application of F-symbols to the pairs of descriptor fusion trees. An example for $\nu = 4$, valid for all L, is shown in Figure 4.27c, and for $L = 0$ the right-hand side is equivalent to the maximally irreducible diagram of a weighted Wigner $6j$ symbol. This raises the interesting question of whether certain nonequivalent inner products for higher ν equivariants play a more privileged role than others, particularly as a basis for machine learning tasks.

The above discussion has focused on the physically important case of global SO(3) covariance. However, this recoupling structure can be further generalized. For example, as described in [60], the inclusion of classical on-site magnetic moments, $\mathbf{m}_i$, either through spin or orbital degrees of freedom can be accommodated by extending the symmetry of the atom-centered descriptors to SO(3) $\otimes$ SO(3). Note that while bond, $\kappa = (nlm)$, and magnetic, $\widetilde{\kappa} = (\widetilde{n}\widetilde{l}\widetilde{m})$, degrees of freedom transform independently in the atom-centered

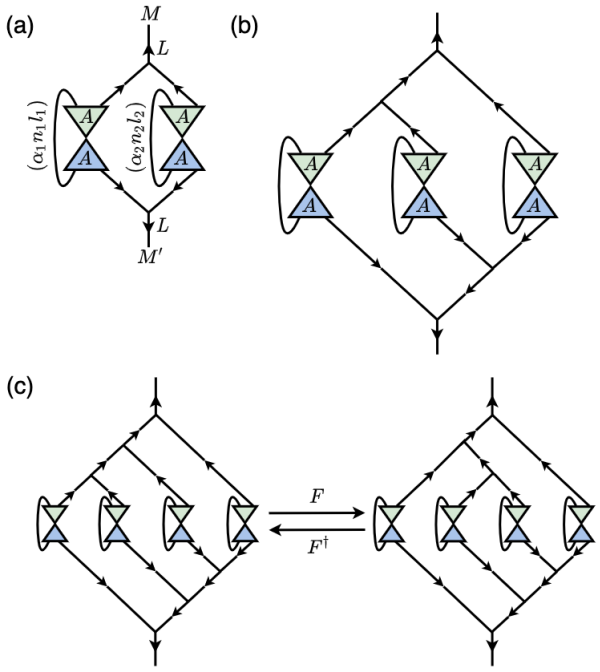


Figure 4.27: Equivariant λ-SOAP generalization of the (a) power spectrum ($\nu = 2$) and (b) bispectrum ($\nu = 3$) structural kernels as tensor networks. (c) Topologically distinct kernel functions exists for either ($L = 0$, $\nu > 3$) or ($L > 0$, $\nu > 2$), determined by the repeated application of F-symbols to the pairs of descriptor fusion trees.

tensors, $A_{\alpha\kappa\tilde{\kappa}}$, they must be recoupled to form a descriptor with definite, total angular momentum. The general construction of such a combined equivariant descriptor is shown in Figure 4.28a,b. Another example is provided by an extension of the SOAP method to the SO(4) group [14], otherwise known as the SNAP method [248], which circumvents the need to explicitly specify a radial basis function, $R_{nl}(r)$, by embedding the radial degree of freedom into the 3-sphere, S^3. This amounts to treating the radial component as an additional polar angle that, along with the two angles inherited from S^2, describes rotations in $\mathbb{R}^4$. The hyperspherical harmonic functions, $U^l_{m\tilde{m}}$, form a complete Fourier basis for functions on S^3. Since, at the level of Lie algebras, there exists an isomorphism SO(4) $\sim$ SO(3) $\otimes$ SO(3), the SO(4) recoupling CG coefficients, H, decompose into products of SO(3) CG coefficients,

$H^{(l_1 l_2 l_3)}_{m_1 \widetilde{m}_1 m_2 \widetilde{m}_2 m_3 \widetilde{m}_3} = C^{(l_1 l_2 l_3)}_{m_1 m_2 m_3} C^{(l_1 l_2 l_3)}_{\widetilde{m}_1 \widetilde{m}_2 \widetilde{m}_3}$ [168, 34, 14]. Recoupling in the SNAP bispectrum using this "parabolic-type" decomposition [168] is shown in Figure 4.28c,d.

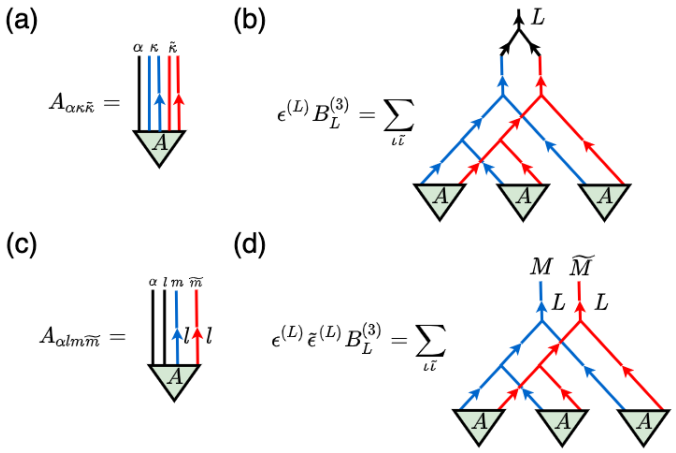


Figure 4.28: Recoupling schemes for (a-b) SO(3) $\otimes$ SO(3) and (c-d) SO(4) equivariant descriptors.

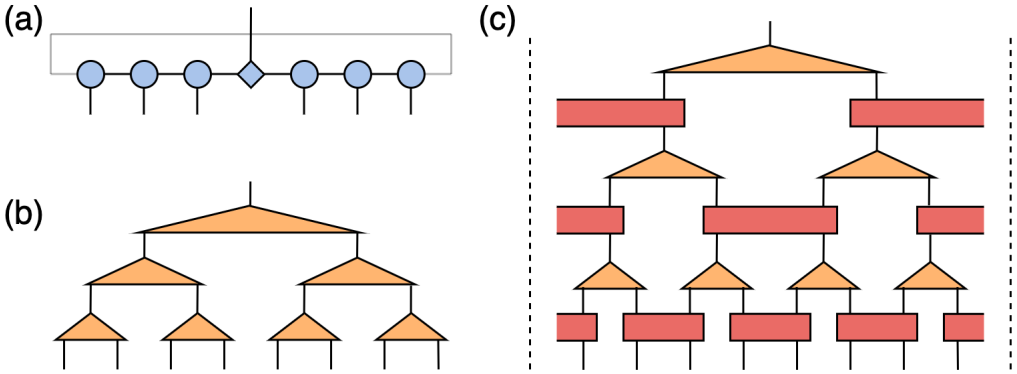


Figure 4.29: Tensor network factorizations corresponding to (a) a matrix product state (MPS), (b) a tree tensor network (TTN), and (c) a Multi-Scale Entanglement Renormalization Ansatz (MERA).

Chapter 5

CONCLUSIONS

In this work, we have utilized first-principles electronic structure methods and multiscale modeling techniques to study the impacts of defect formation and substitutional alloying in technologically important classes of multicomponent semiconductors. An overarching goal of this work was to predict electronic and thermodynamic material properties starting from atomic structure. First, we considered the problem of characterizing point defects and defect complexes in chalcopyrite $CuInSe_2$. Starting from density functional theory calculations, we constructed a thermodynamic model connecting material off-stoichiometry to defect formation. In doing so, we identified the typical defects in the system which impact free carrier concentrations and phase stability, and we tied these properties to material synthesis conditions at finite temperature. We then considered defect-mediated copper diffusion in the system, which likely plays an outsized role in the engineering of efficient absorber layers for photovoltaic device applications. Combining this thermodynamic model, linear response theory, kinetic Monte Carlo and continuum reaction-diffusion simulation, we showed how copper kinetics depends on material composition and how measurements of either tracer or chemical diffusion can lead to large discrepancies in the experimental copper diffusivity. Second, we studied halide alloying and impurity doping in lead halide perovskites. Using Monte Carlo simulation with effective lattice Hamiltonians trained on DFT data, we found that long-range halide ordering persists at room-temperature, well above the predicted miscibility gap, and discussed the potential implications on halide exchange reactions in bulk crystals. We then showed how the introduction of manganese dopants can induce halide phase separation and connected this behavior to experimental observations of EPR signal quenching during halide exchange. Returning to the case of single halide $CsPbCl_3$, we

modeled the energetics of Yb incorporation and the atomistic structure of defect complex formation. We found that a variety of partially charge compensated complexes involving Yb_{Pb} and Pb vacancies form readily at room temperature, and combined with variational methods applied to an effective model of exciton binding to shallow vacancies, we discussed the implications of this complex formation to experimental observation of quantum cutting and enhanced near-IR photoluminescence. Finally, to move beyond the limitations of single lattice models, we constructed machine learning architectures based on tensor networks and adapted to SO(3)-invariant atom-density representations. This allowed us to capture multiple lattice symmetries, atomic displacements, and chemical disorder within a single model. We demonstrated, in particular, that tensor networks act as an effective form of model regularization, allowing for relatively accurate model learning with limited amounts of training data. Moreover, we found promising connections between model expressibility and the entanglement structure encoded in the model weights during the course of training. We then showed how these techniques could be applied to the compression of atom-density representations and extended to SO(3)-equivariant cases.

Chapter 6

Bibliography

[1] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.

[2] Quinten A Akkerman, Valerio D'Innocenzo, Sara Accornero, Alice Scarpellini, Annamaria Petrozza, Mirko Prato, and Liberato Manna. Tuning the optical properties of cesium lead halide perovskite nanocrystals by anion exchange reactions. *Journal of the American Chemical Society*, 137(32):10276–10281, 2015.

[3] T Ala-Nissila, R Ferrando, and SC Ying. Collective and single particle diffusion on surfaces. *Advances in Physics*, 51(3):949–1078, 2002.

[4] Zeyuan Allen-Zhu, Yuanzhi Li, and Zhao Song. A Convergence Theory for Deep Learning via Over-Parameterization. *arXiv*, 2018.

[5] A. R. Allnatt and A. B. Lidiard. *Atomic Transport in Solids*. Cambridge University Press, 1993.

[6] Nongnuch Artrith, Alexander Urban, and Gerbrand Ceder. Efficient and accurate

machine-learning interpolation of atomic energies in compositions with many species. *Phys. Rev. B*, 96:014112, Jul 2017.

[7] N. W. Ashcroft and N. D. Mermin. *Solid State Physics*. Brooks/Cole, 1976.

[8] Y. Y. Atas, E. Bogomolny, O. Giraud, and G. Roux. Distribution of the ratio of consecutive level spacings in random matrix ensembles. *Phys. Rev. Lett.*, 110:084101, Feb 2013.

[9] Markus Bachmayr, Gabor Csanyi, Ralf Drautz, Genevieve Dusson, Simon Etter, Cas van der Oord, and Christoph Ortner. Atomic Cluster Expansion: Completeness, Efficiency and Stability. *arXiv*, 2019.

[10] K.K. Bajaj. Effect of electron-phonon interaction on the binding energy of a wannier exciton in a polarizable medium. *Solid State Communications*, 15(7):1221–1224, 1974.

[11] Subila K Balakrishnan and Prashant V Kamat. Ligand assisted transformation of cubic cspbbr3 nanocrystals into two-dimensional cspb2br5 nanosheets. *Chemistry of Materials*, 30(1):74–78, 2018.

[12] Michal Baranowski and Paulina Plochocka. Excitons in Metal-Halide Perovskites. *Advanced Energy Materials*, 10(26):1903659, 2020.

[13] Michal Baranowski, Paulina Plochocka, Rui Su, Laurent Legrand, Thierry Barisien, Frederick Bernardot, Qihura Xiong, Christophe Testelin, and Maria Chamarro. Exciton binding energy and effective mass of $CsPbCl_3$: a magneto-optical study. *Photonics Research*, 8(10):A50, 2020.

[14] Albert P. Bartók, Risi Kondor, and Gábor Csányi. On representing chemical environments. *Physical Review B*, 87(18):184115, 2013.

[15] Peter W Battaglia, Jessica B Hamrick, Victor Bapst, Alvaro Sanchez-Gonzalez, Vinicius Zambaldi, Mateusz Malinowski, Andrea Tacchetti, David Raposo, Adam San-

toro, Ryan Faulkner, Caglar Gulcehre, Francis Song, Andrew Ballard, Justin Gilmer, George Dahl, Ashish Vaswani, Kelsey Allen, Charles Nash, Victoria Langston, Chris Dyer, Nicolas Heess, Daan Wierstra, Pushmeet Kohli, Matt Botvinick, Oriol Vinyals, Yujia Li, and Razvan Pascanu. Relational inductive biases, deep learning, and graph networks. *arXiv*, 2018.

[16] Rachel E Beal, Daniel J Slotcavage, Tomas Leijtens, Andrea R Bowring, Rebecca A Belisle, William H Nguyen, George F Burkhard, Eric T Hoke, and Michael D McGehee. Cesium lead halide perovskites with improved stability for tandem solar cells. *The journal of physical chemistry letters*, 7(5):746–751, 2016.

[17] Brandon J Beberwyck, Yogesh Surendranath, and A Paul Alivisatos. Cation exchange: a versatile tool for nanomaterials synthesis. *The Journal of Physical Chemistry C*, 117(39):19759–19770, 2013.

[18] Jonathon S Bechtel and Anton Van der Ven. First-principles thermodynamics study of phase stability in inorganic halide perovskite solid solutions. *Physical Review Materials*, 2(4):045401, 2018.

[19] Jonathon S Bechtel and Anton Van der Ven. Octahedral tilting instabilities in inorganic halide perovskites. *Physical Review Materials*, 2(2):025401, 2018.

[20] K. D. Becker and Sigurd Wagner. Temperature-dependent nuclear magnetic resonance in cuinx$_2$ (x = s, se, te) chalcopyrite-structure compounds. *Physical Review B*, 27(9):5240, 1983.

[21] Jörg Behler and Michele Parrinello. Generalized neural-network representation of high-dimensional potential-energy surfaces. *Phys. Rev. Lett.*, 98:146401, Apr 2007.

[22] Jörg Behler. Four generations of high-dimensional neural network potentials. *Chemical Reviews*, 121(16):10037–10072, 2021. PMID: 33779150.

[23] J Bekaert, R Saniz, and B Partoens. Native point defects in cuin1xgaxse2: hybrid density functional calculations predict the origin of p-and n-type conductivity. *Physical Chemistry Chemical Physics*, 16:22299, 2014.

[24] RE Belardinelli and VD Pereyra. Wang-landau algorithm: A theoretical analysis of the saturation of the error. *The Journal of chemical physics*, 127(18):184105, 2007.

[25] Mikhail Belkin, Daniel Hsu, Siyuan Ma, and Soumik Mandal. Reconciling modern machine learning practice and the bias-variance trade-off. *arXiv*, 2018.

[26] Jacob D. Biamonte, Stephen R. Clark, and Dieter Jaksch. Categorical Tensor Network States. *AIP Advances*, 1(4):042172, 2011.

[27] Connor G Bischak, Craig L Hetherington, Hao Wu, Shaul Aloni, D Frank Ogletree, David T Limmer, and Naomi S Ginsberg. Origin of reversible photoinduced phase separation in hybrid perovskites. *Nano letters*, 17(2):1028–1033, 2017.

[28] P. E. Blöchl. Projector augmented-wave method. *Physical Review B*, 50(24):17953–17979, 1994.

[29] Menno Bokdam, Tobias Sander, Alessandro Stroppa, Silvia Picozzi, D. D. Sarma, Cesare Franchini, and Georg Kresse. Role of Polar Phonons in the Photo Excited State of Metal Halide Perovskites. *Scientific Reports*, 6(1):28618, 2016.

[30] Riley E Brandt, Jeremy R Poindexter, Prashun Gorai, Rachel C Kurchin, Robert LZ Hoye, Lea Nienhaus, Mark WB Wilson, J Alexander Polizzotti, Raimundas Sereika, Raimundas Zaltauskas, et al. Searching for "defect-tolerant" photovoltaic materials: combined theoretical and experimental screening. *Chemistry of Materials*, 29(11):4667–4674, 2017.

[31] Danny Broberg, Bharat Medasani, Nils E.R. Zimmermann, Guodong Yu, Andrew Canning, Maciej Haranczyk, Mark Asta, and Geoffroy Hautier. PyCDT: A Python

toolkit for modeling point defects in semiconductors and insulators. *Computer Physics Communications*, 226:165–179, 2018.

[32] H. Büttner and J. Pollmann. Excitons in polar semiconductors. *Physica B+C*, 117:278–283, 1983.

[33] Jean-Louis Calais and Per-Olov Löwdin. A simple method of treating atomic integrals containing functions of r_{12}. *Journal of Molecular Spectroscopy*, 8(1-6):203–211, 1962.

[34] M. A. Caprio, K. D. Sviratcheva, and A. E. McCoy. Racah's method for general subalgebra chains: Coupling coefficients of SO(5) in canonical and physical bases. *Journal of Mathematical Physics*, 51(9):093518, 2010.

[35] C. Carabatos-Nédelec, M. Oussaïd, and K. Nitsch. Raman scattering investigation of cesium plumbochloride, $CsPbCl_3$, phase transitions. *Journal of Raman Spectroscopy*, 34(5):388–393, 2003.

[36] Claudio Chamon, Alioscia Hamma, and Eduardo R. Mucciolo. Emergent irreversibility and entanglement spectrum statistics. *Phys. Rev. Lett.*, 112:240501, Jun 2014.

[37] Abhijit Chatterjee and Arthur F Voter. Accurate acceleration of kinetic monte carlo simulations through the modification of rate constants. *The Journal of chemical physics*, 132(19):194101, 2010.

[38] Chi Chen, Yunxing Zuo, Weike Ye, Xiangguo Li, and Shyue Ping Ong. Learning properties of ordered and disordered materials from multi-fidelity data. *Nature Computational Science*, 1(1):46–53, 2021.

[39] Daqin Chen, Gaoliang Fang, and Xiao Chen. Silica-coated mn-doped cspb (cl/br) 3 inorganic perovskite quantum dots: exciton-to-mn energy transfer and blue-excitable solid-state lighting. *ACS applied materials & interfaces*, 9(46):40477–40487, 2017.

[40] Daqin Chen, Yuansheng Wang, Yunlong Yu, Ping Huang, and Fangyi Weng. Near-infrared quantum cutting in transparent nanostructured glass ceramics. *Opt. Lett.*, 33(16):1884–1886, Aug 2008.

[41] Leonid Chernyak, Konstantin Gartsman, David Cahen, and Oscar M Stafsudd. Electronic effects of ion mobility in semiconductors: Semionic behaviour of cuinse2. *Journal of Physics and Chemistry of Solids*, 56(9):1165–1191, 1995.

[42] Vladimir S Chirvony, Soranyel González-Carrero, Isaac Suárez, Raquel E Galian, Michele Sessolo, Henk J Bolink, Juan P Martínez-Pastor, and Julia Pérez-Prieto. Delayed Luminescence in Lead Halide Perovskite Nanocrystals. *The Journal of Physical Chemistry C*, 121(24):13381–13390, 2017.

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

[44] Theodore A. Cohen, Tyler J. Milstein, Daniel M. Kroupa, J. Devin MacKenzie, Christine K. Luscombe, and Daniel R. Gamelin. Quantum-cutting Yb^{3+}-doped perovskite nanocrystals for monolithic bilayer luminescent solar concentrators. *Journal of Materials Chemistry A*, 7(15):9279–9288, 2019.

[45] Diego Colombara, Hossam Elanzeery, Nicoleta Nicoara, Deepanjan Sharma, Marcel Claro, Torsten Schwarz, Anna Koprek, Max Hilaire Wolter, Michele Melchiorre, Mohit Sood, et al. Chemical instability at chalcogenide surfaces impacts chalcopyrite devices well beyond the surface. *Nature communications*, 11(1):1–14, 2020.

[46] Ian Convy, William Huggins, Haoran Liao, and K Birgitta Whaley. Mutual Information Scaling for Tensor Network Machine Learning. *arXiv*, 2021.

[47] José Coutinho, Vladimir P Markevich, and Anthony R Peaker. Characterisation of negative-U defects in semiconductors. *Journal of Physics: Condensed Matter*, 32(32):323001, 2020.

[48] Matthew J Crane, Daniel M Kroupa, Joo Yeon Roh, Rayne T Anderson, Matthew D Smith, and Daniel R Gamelin. Single-Source Vapor Deposition of Quantum-Cutting Yb^{3+}:$CsPb(Cl_{1-x}Br_x)_3$ and Other Complex Metal-Halide Perovskites. *ACS Applied Energy Materials*, 2(6):4560–4565, 2019.

[49] Sidney E Creutz, Evan N Crites, Michael C Siena, and Daniel R Gamelin. Anion Exchange in Cesium Lead Halide Perovskite Nanocrystals and Thin Films using Trimethylsilyl Halide Reagents. *Chemistry of Materials*, 2018.

[50] Shawn X. Cui, Michael H. Freedman, Or Sattath, Richard Stong, and Greg Minton. Quantum Max-flow/Min-cut. *Journal of Mathematical Physics*, 57(6):062206, 2016.

[51] Predrag Cvitanović. *Group Theory: Birdtracks, Lie's, and Exceptional Groups*. Princeton University Press, New Jersey, 2008.

[52] Geula Dagan, T. F. Ciszek, and David Cahen. Ion migration in chalcopyrite semiconductors. *Journal of Physical Chemistry*, 96(26):11009–11017, 1992.

[53] Sandip De, Albert P. Bartók, Gábor Csányi, and Michele Ceriotti. Comparing molecules and solids across structural and alchemical space. *Physical Chemistry Chemical Physics*, 18(20), 2016.

[54] Jan de Boer, Jarkko Järvelä, and Esko Keski-Vakkuri. Aspects of capacity of entanglement. *Phys. Rev. D*, 99:066012, Mar 2019.

[55] De de Fontaine. The number of independent pair-correlation functions in multicomponent systems. *Journal of Applied Crystallography*, 4(1):15–19, 1971.

[56] Luca De Trizio and Liberato Manna. Forging colloidal nanostructures via cation exchange reactions. *Chemical reviews*, 116(18):10852–10887, 2016.

[57] CASM Developers. Casmcode, v 0.2.1; 2017, doi: 10.5281/zenodo.546148. *Google Scholar There is no corresponding record for this reference.*

[58] D L Dexter. A Theory of Sensitized Luminescence in Solids. *The Journal of Chemical Physics*, 21(5):836–850, 1953.

[59] Ralf Drautz. Atomic cluster expansion for accurate and transferable interatomic potentials. *Physical Review B*, 99(1):014104, 2019.

[60] Ralf Drautz. Atomic cluster expansion of scalar, vectorial, and tensorial properties including magnetism and charge transfer. *Phys. Rev. B*, 102:024104, Jul 2020.

[61] SL Dudarev, GA Botton, SY Savrasov, CJ Humphreys, and AP Sutton. Electron-energy-loss spectra and the structural stability of nickel oxide: An lsda+ u study. *Physical Review B*, 57(3):1505, 1998.

[62] Alexander Dunn, Qi Wang, Alex Ganose, Daniel Dopp, and Anubhav Jain. Benchmarking materials property prediction methods: the matbench test set and automatminer reference algorithm. *npj Computational Materials*, 6(1):1–10, 2020.

[63] J. J. Eilers, D. Biner, J. T. van Wijngaarden, K. Krämer, H.-U. Güdel, and A. Meijerink. Efficient visible to infrared quantum cutting through downconversion with the Er^{3+}-Yb^{3+} couple in $Cs_3Y_2Br_9$. *Applied Physics Letters*, 96(15):151106, 2010.

[64] Christian S Erickson, Matthew J Crane, Tyler J Milstein, and Daniel R Gamelin. Photoluminescence Saturation in Quantum-Cutting Yb^{3+}-Doped $CsPb(Cl_{1-x}Br_x)_3$ Perovskite Nanocrystals: Implications for Solar Downconversion. *The Journal of Physical Chemistry C*, 123(19):12474–12484, 2019.

[65] G. Evenbly and G. Vidal. Class of Highly Entangled Many-Body States that can be Efficiently Simulated. *Physical Review Letters*, 112(24):240502, 2014.

[66] G. Evenbly and G. Vidal. Tensor network renormalization yields the multiscale entanglement renormalization ansatz. *Phys. Rev. Lett.*, 115:200401, Nov 2015.

[67] Felix A Faber, Anders S Christensen, Bing Huang, and O Anatole von Lilienfeld. Alchemical and structural distribution based representation for universal quantum machine learning. *The Journal of Chemical Physics*, 148(24):241717, 2018.

[68] R. P. Feynman. Slow Electrons in a Polar Crystal. *Physical Review*, 97(3):660–665, 1955.

[69] Marina R Filip, Jonah B Haber, and Jeffrey B Neaton. Phonon Screening of Excitons in Semiconductors: Halide Perovskites and Beyond. *arXiv*, 2021.

[70] Peter John Forrester. *Log-Gases and Random Matrices*. Princeton University Press, New Jersey, 2010.

[71] Dieter Forster. *Hydrodynamic Fluctuations, Broken Symmetry, and Correlation Functions*. Frontiers in Physics, 1975.

[72] C Freysoldt, B Grabowski, T Hickel, and J Neugebauer. First-principles calculations for point defects in solids. *Reviews of Modern Physics*, 86:253, 2014.

[73] Christoph Freysoldt, Blazej Grabowski, Tilmann Hickel, Jörg Neugebauer, Georg Kresse, Anderson Janotti, and Chris G. Van de Walle. First-principles calculations for point defects in solids. *Rev. Mod. Phys.*, 86:253–305, Mar 2014.

[74] Christoph Freysoldt, Jörg Neugebauer, and Chris G de Walle. Fully Ab Initio Finite-Size Corrections for Charged-Defect Supercell Calculations. *Physical Review Letters*, 102(1):016402, 2009.

[75] D. Fröhlich, K. Heidrich, H. Künzel, G. Trendel, and J. Treusch. Cesium-trihalogen-plumbates a new class of ionic semiconductors. *Journal of Luminescence*, 18-19:385–388, 1979.

[76] Krzysztof Galkowski, Anatolie Mitioglu, Atsuhiko Miyata, Paulina Plochocka, Oliver Portugall, Giles E. Eperon, Jacob Tse-Wei Wang, Thomas Stergiopoulos, Samuel D.

Stranks, Henry J. Snaith, and Robin J. Nicholas. Determination of the exciton binding energy and effective masses for methylammonium and formamidinium lead tri-halide perovskite semiconductors. *Energy & Environmental Science*, 9(3):962–970, 2016.

[77] Ze-Feng Gao, Song Cheng, Rong-Qiang He, Z. Y. Xie, Hui-Hai Zhao, Zhong-Yi Lu, and Tao Xiang. Compressing deep neural networks by matrix product operators. *Phys. Rev. Research*, 2:023300, Jun 2020.

[78] Silvano Garnerone, Thiago R. de Oliveira, Stephan Haas, and Paolo Zanardi. Statistical properties of random matrix product states. *Phys. Rev. A*, 82:052312, Nov 2010.

[79] Konstantin Gartsman, Leonid Chernyak, Vera Lyahovitskaya, David Cahen, Vladimir Didik, Vitaliy Kozlovsky, Roald Malkovich, Elena Skoryatina, and Valentina Usacheva. Direct evidence for diffusion and electromigration of cu in CuInSe2. *Journal of Applied Physics*, 82(9):4282–4285, 1997.

[80] Ivan Glasser, Nicola Pancotti, and J. Ignacio Cirac. From Probabilistic Graphical Models to Generalized Tensor Networks for Supervised Learning. *IEEE Access*, 8:68169–68182, 2020.

[81] Ivan Glasser, Ryan Sweke, Nicola Pancotti, Jens Eisert, and Ignacio J Cirac. Expressive power of tensor-network factorizations for probabilistic modeling, with applications from hidden Markov models to quantum machine learning. *arXiv*, 2019.

[82] Aldo Glielmo, Claudio Zeni, and Alessandro De Vita. Efficient nonparametric n-body force fields from machine learning. *Physical Review B*, 97(18):184307, 2018.

[83] T Gödecke, Thomas Haalboom, and Frank Ernst. Phase equilibria of cu-in-se. i. stable states and nonequilibrium states of the in {sub 2} se {sub 3}-cu {sub 2} se subsystem. *Zeitschrift für Metallkunde*, 91, 2000.

[84] Rep Gomer. Diffusion of adsorbates on metal surfaces. *Reports on progress in Physics*, 53(7):917, 1990.

[85] Andrea Grisafi, David M Wilkins, Gábor Csányi, and Michele Ceriotti. Symmetry-Adapted Machine-Learning for Tensorial Properties of Atomistic Systems. *Physical Review Letters*, 120(3):036002, 2017.

[86] J. F. Guillemoles, L Kronik, D Cahen, and U Rau. Stability issues of cu(in,ga)se_2-based solar cells. *Journal of Physical Chemistry B*, 104:4849, 2000.

[87] Jean-François Guillemoles, Leeor Kronik, David Cahen, Uwe Rau, Axel Jasenek, and Hans-Werner Schock. Stability issues of cu(in,ga)se2-based solar cells. *The Journal of Physical Chemistry B*, 104(20):4849–4862, 2000.

[88] RP Gupta, MS Seehra, and WE Vehse. Shift of néel temperature and epr linewidth of kmn f 3 with mg doping. *Physical Review B*, 5(1):92, 1972.

[89] Jonas Haferkamp, Christian Bertoni, Ingo Roth, and Jens Eisert. Emergent statistical mechanics from properties of disordered random matrix product states. *arXiv*, 2021.

[90] H. Haken. Die Theorie des Exzitons im festen Körper. *Fortschritte der Physik*, 6(6):271–334, 1958.

[91] Hermann Haken. Zur Quantentheorie des Mehrelektronensystems im schwingenden Gitter. I. *Zeitschrift für Physik*, 146(5):527–554, Oct 1956.

[92] Frank E Harris, Alexei M Frolov, and Vedene H Smith. Singular and nonsingular three-body integrals for exponential wave functions. *The Journal of Chemical Physics*, 121(13):6323–6333, 2004.

[93] Frank E Harris and Vedene H Smith. Highly Compact Wave Functions for He-Like Systems. *The Journal of Physical Chemistry A*, 109(50):11413–11416, 2005.

[94] Aram W. Harrow and Richard A. Low. Random Quantum Circuits are Approximate 2-designs. *Communications in Mathematical Physics*, 291(1):257–302, 2009.

[95] Gus LW Hart and Rodney W Forcade. Algorithm for generating derivative structures. *Physical Review B*, 77(22):224115, 2008.

[96] Gus LW Hart and Rodney W Forcade. Generating derivative structures from multilattices: Algorithm and application to hcp alloys. *Physical Review B*, 80(1):014120, 2009.

[97] Gus LW Hart, Lance J Nelson, and Rodney W Forcade. Generating derivative structures at a fixed concentration. *Computational Materials Science*, 59:101–107, 2012.

[98] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning*. Springer, New York, 2009.

[99] Robert W. Hellwarth and Ivan Biaggio. Mobility of an electron in a multimode polar lattice. *Physical Review B*, 60(1):299–307, 1999.

[100] G Henkelman and B P Uberuaga. A climbing image nudged elastic band method for finding saddle points and minimum energy paths. *Journal of Chemical Physics*, 113(22), 2000.

[101] Graeme Henkelman, Andri Arnaldsson, and Hannes Jónsson. A fast and robust algorithm for Bader decomposition of charge density. *Computational Materials Science*, 36(3):354–360, 2006.

[102] Herberholz, R., Rau, U., Schock, H. W., Haalboom, T., Gödecke, T., Ernst, F., Beilharz, C., Benz, K. W., and Cahen, D. Phase segregation, cu migration and junction formation in cu(in, ga)se2. *Eur. Phys. J. AP*, 6(2):131–139, 1999.

[103] Jochen Heyd, Gustavo E. Scuseria, and Matthias Ernzerhof. Hybrid functionals based on a screened coulomb potential. *The Journal of Chemical Physics*, 118(18):8207–8215, 2003.

[104] Lauri Himanen, Marc O. J. Jäger, Eiaki V. Morooka, Filippo Federici Canova, Yashasvi S. Ranawat, David Z. Gao, Patrick Rinke, and Adam S. Foster. DScribe: Library of descriptors for machine learning in materials science. *Computer Physics Communications*, 247:106949, 2020.

[105] Geoffrey Hinton and Sam T Roweis. Stochastic neighbor embedding. In *NIPS*, volume 15, pages 833–840. Citeseer, 2002.

[106] Jacob B Hoffman, A Lennart Schleper, and Prashant V Kamat. Transformation of sintered cspbbr3 nanocrystals to cubic cspbi3 and gradient cspbbr x i3–x through halide exchange. *Journal of the American Chemical Society*, 138(27):8603–8611, 2016.

[107] Eric T Hoke, Daniel J Slotcavage, Emma R Dohner, Andrea R Bowring, Hemamala I Karunadasa, and Michael D McGehee. Reversible photo-induced trap formation in mixed-halide hybrid perovskites for photovoltaics. *Chemical Science*, 6(1):613–617, 2015.

[108] Haoyan Huo and Matthias Rupp. Unified representation for machine learning of molecules and crystals. *arXiv preprint arXiv:1704.06439*, 13754, 2017.

[109] Jason Iaconis. Quantum state complexity in computationally tractable quantum circuits. *PRX Quantum*, 2:010329, Feb 2021.

[110] Giulio Imbalzano, Andrea Anelli, Daniele Giofré, Sinja Klees, Jörg Behler, and Michele Ceriotti. Automatic selection of atomic fingerprints and reference configurations for machine-learning potentials. *The Journal of Chemical Physics*, 148(24):241730, 2018.

[111] Hisao Ito, Jun'ichirō Nakahara, and Ryumyo Onaka. Magneto-optical Study of the Exciton States in $CsPbCl_3$. *Journal of the Physical Society of Japan*, 47(6):1927–1935, 1979.

[112] Erin Jedlicka, Jian Wang, Joshua Mutch, Young-Kwang Jung, Preston Went, Joseph Mohammed, Mark Ziffer, Rajiv Giridharagopal, Aron Walsh, Jiun-Haw Chu, and

David S Ginger. Bismuth Doping Alters Structural Phase Transitions in Methylammonium Lead Tribromide Single Crystals. *The Journal of Physical Chemistry Letters*, pages 2749–2755, 2021.

[113] Haotian Jiang, Siqi Cui, Yu Chen, and Haizheng Zhong. Ion exchange for halide perovskite: From nanocrystal to bulk materials. *Nano Select*, 2021.

[114] Handong Jin, Elke Debroye, Masoumeh Keshavarz, Ivan G. Scheblykin, Maarten B. J. Roeffaers, Johan Hofkens, and Julian A. Steele. It's a trap! On the nature of localised states and charge trapping in lead halide perovskites. *Materials Horizons*, 7(2):397–410, 2019.

[115] Steven G. Johnson. The NLopt nonlinear-optimization package, 2020.

[116] A. H. G. Rinnooy Kan and G. T. Timmer. Stochastic global optimization methods part I: Clustering methods. *Mathematical Programming*, 39(1):27–56, 1987.

[117] A. H. G. Rinnooy Kan and G. T. Timmer. Stochastic global optimization methods part II: Multi level methods. *Mathematical Programming*, 39(1):57–78, 1987.

[118] Jun Kang and Lin-Wang Wang. High Defect Tolerance in Lead Halide Perovskite CsPbBr3. *The Journal of Physical Chemistry Letters*, 8(2):489–493, 2017.

[119] Bobak Toussi Kiani, Seth Lloyd, and Reevu Maity. Learning Unitaries by Gradient Descent. *arXiv*, 2020.

[120] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[121] W. Kohn and L. J. Sham. Self-Consistent Equations Including Exchange and Correlation Effects. *Physical Review*, 140(4A):A1133–A1138, 1965.

[122] JJ Krebs. Epr determination of the nearest-neighbor exchange constant for mn2+ pairs in kznf3. *Journal of Applied Physics*, 40(3):1137–1139, 1969.

[123] G Kresse and J Furthmüller. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Physical Review B*, 54:11169, 1996.

[124] G. Kresse and J. Furthmüller. Efficiency of ab-initio total energy calculations for metals and semiconductors using a plane-wave basis set. *Computational Materials Science*, 6(1):15–50, 1996.

[125] G Kresse and J Furthmüller. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Physical Review B*, 54(16):11169–11186, 1996.

[126] G. Kresse and D. Joubert. From ultrasoft pseudopotentials to the projector augmented-wave method. *Physical Review B*, 59(3):1758–1775, 1998.

[127] Daniel M Kroupa, Joo Yeon Roh, Tyler J Milstein, Sidney E Creutz, and Daniel R Gamelin. Quantum-Cutting Ytterbium-Doped $CsPb(Cl_{1-x}Br_x)_3$ Perovskite Thin Films with Photoluminescence Quantum Yields over 190%. *ACS Energy Letters*, 3(10):2390–2395, 2018.

[128] David P. Landau and Kurt Binder. *A Guide to Monte Carlo Simulations in Statistical Physics*. Cambridge University Press, 2015.

[129] DP Landau, Shan-Ho Tsai, and M Exler. A new approach to monte carlo simulations in statistical physics: Wang-landau sampling. *American Journal of Physics*, 72(10):1294–1302, 2004.

[130] Marcel F Langer, Alex Goeßmann, and Matthias Rupp. Representations of molecules and materials for interpolation of quantum-mechanical simulations via machine learning. *arXiv*, 2020.

[131] Stephan Lany and Alex Zunger. Anion vacancies as a source of persistent photoconductivity in II-VI and chalcopyrite semiconductors. *Physical Review B*, 72(3):035215, 2005.

[132] Stephan Lany and Alex Zunger. Light- and bias-induced metastabilities in Cu(In,Ga)Se_2 based solar cells caused by the (V_{Se}-V_{Cu}) vacancy complex. *Journal of Applied Physics*, 100(11):113725, 2006.

[133] Stephan Lany and Alex Zunger. Assessment of correction methods for the band-gap problem and for finite-size effects in supercell defect calculations: Case studies for ZnO and GaAs. *Physical Review B*, 78(23):235104, 2008.

[134] Mikael Leetmaa and Natalia V Skorodumova. KMCLib: a general framework for lattice kinetic monte carlo (KMC) simulations. *Computer Physics Communications*, 185(9):2340–2349, 2014.

[135] Zhidong Leong and Teck Leong Tan. Robust cluster expansion of multicomponent systems using structured sparsity. *Physical Review B*, 100(13):134108, 2019.

[136] Yoav Levine, Or Sharir, Nadav Cohen, and Amnon Shashua. Quantum Entanglement in Deep Learning Architectures. *Physical Review Letters*, 122(6):065301, 2019.

[137] Yoav Levine, David Yakira, Nadav Cohen, and Amnon Shashua. Deep Learning and Quantum Entanglement: Fundamental Connections with Implications to Network Design. *arXiv*, 2017.

[138] Hui Li and F. D. M. Haldane. Entanglement spectrum as a generalization of entanglement entropy: Identification of topological order in non-abelian fractional quantum hall effect states. *Phys. Rev. Lett.*, 101:010504, Jul 2008.

[139] Xiyu Li, Sai Duan, Haichun Liu, Guanying Chen, Yi Luo, and Hans Ågren. Mechanism for the Extremely Efficient Sensitization of Yb^{3+} Luminescence in $CsPbCl_3$ Nanocrystals. *The journal of physical chemistry letters*, 10(3):487–492, 2019.

[140] Yuanzhi Li and Yingyu Liang. Learning Overparameterized Neural Networks via Stochastic Gradient Descent on Structured Data. *arXiv*, 2018.

[141] Hang Lin, Daqin Chen, Yunlong Yu, Anping Yang, and Yuansheng Wang. Near-infrared quantum cutting in Ho^{3+}/Yb^{3+} codoped nanostructured glass ceramic. *Opt. Lett.*, 36(6):876–878, Mar 2011.

[142] Henry W Lin, Max Tegmark, and David Rolnick. Why does deep and cheap learning work so well? *Journal of Statistical Physics*, pages 1–25, 2016.

[143] Matthew R Linaburg, Eric T McClure, Jackson D Majher, and Patrick M Woodward. $Cs_{1-x}Rb_xPbCl_3$ and $Cs_{1-x}Rb_xPbBr_3$ Solid Solutions: Understanding Octahedral Tilting in Lead Halide Perovskites. *Chemistry of Materials*, 29(8):3507–3514, 2017.

[144] Wenyong Liu, Qianglu Lin, Hongbo Li, Kaifeng Wu, István Robel, Jeffrey M Pietryga, and Victor I Klimov. Mn2+-doped lead halide perovskite nanocrystals with dual-color emission controlled by halide content. *Journal of the American Chemical Society*, 138(45):14954–14961, 2016.

[145] Zidu Liu, Li-Wei Yu, L M Duan, and Dong-Ling Deng. The Presence and Absence of Barren Plateaus in Tensor-network Based Machine Learning. *arXiv*, 2021.

[146] Sirui Lu, Márton Kanász-Nagy, Ivan Kukuljan, and J Ignacio Cirac. Tensor networks and efficient descriptions of classical data. *arXiv*, 2021.

[147] Igor Lubomirsky, Konstantin Gartsman, and David Cahen. Space charge effects on dopant diffusion coefficient measurements in semiconductors. *Journal of Applied Physics*, 83(9):4678–4682, 1998.

[148] Christian DR Ludwig, Thomas Gruhn, Claudia Felser, and Johannes Windeln. Defect structures in cuinse 2: A combination of monte carlo simulations and density functional theory. *Physical Review B*, 83(17):174112, 2011.

[149] Jonathan Lym, Geun Ho Gu, Yousung Jung, and Dionisios G Vlachos. Lattice convolutional neural network modeling of adsorbate coverage effects. *The Journal of Physical Chemistry C*, 123(31):18951–18959, 2019.

[150] Yury Lysogorskiy, Cas van der Oord, Anton Bochkarev, Sarath Menon, Matteo Rinaldi, Thomas Hammerschmidt, Matous Mrovec, Aidan Thompson, Gábor Csányi, Christoph Ortner, and Ralf Drautz. Performant implementation of the atomic cluster expansion (PACE) and application to copper and silicon. *npj Computational Materials*, 7(1):97, 2021.

[151] Igor Lyubomirsky, M. K. Rabinal, and David Cahen. Room-temperature detection of mobile impurities in compound semiconductors by transient ion drift. *Journal of Applied Physics*, 81(10):6684–6691, 1997.

[152] Ju-Ping Ma, Jia-Kai Chen, Jun Yin, Bin-Bin Zhang, Qing Zhao, Yoshihiro Kuroiwa, Chikako Moriyoshi, Lili Hu, Osman M Bakr, Omar F Mohammed, and Hong-Tao Sun. Doping Induces Structural Phase Transitions in All-Inorganic Lead Halide Perovskite Nanocrystals. *ACS Materials Letters*, 2(4):367–375, 2020.

[153] Ju-Ping Ma, Ya-Meng Chen, Lu-Min Zhang, Shao-Qiang Guo, Jian-Dang Liu, Hong Li, Bang-Jiao Ye, Zhi-Yong Li, Yang Zhou, Bin-Bin Zhang, Osman M. Bakr, Jun-Ying Zhang, and Hong-Tao Sun. Insights into the local structure of dopants, doping efficiency, and luminescence properties of lanthanide-doped $CsPbCl_3$ perovskite nanocrystals. *Journal of Materials Chemistry C*, 7(10):3037–3048, 2019.

[154] Rita Magri, S-H Wei, and Alex Zunger. Ground-state structures and the random-state energy of the madelung lattice. *Physical Review B*, 42(17):11388, 1990.

[155] Arup Mahata, Daniele Meggiolaro, and Filippo De Angelis. From Large to Small Polarons in Lead, Tin, and Mixed Lead–Tin Halide Perovskites. *The Journal of Physical Chemistry Letters*, 10(8):1790–1798, 2019.

[156] M Mantina, Y Wang, R Arroyave, L. Q. Chen, Z. K. Liu, and C Wolverton. First-Principles calculation of Self-Diffusion coefficients. *Physical Review Letter*, 100:215901, 2008.

[157] John Martyn, Guifre Vidal, Chase Roberts, and Stefan Leichenauer. Entanglement and Tensor Networks for Supervised Image Classification. *arXiv*, 2020.

[158] Gene F Mazenko. Statistical mechanical models and surface diffusion. In *Surface Mobilities on Solid Materials*, pages 27–62. Springer, 1983.

[159] H Mehrer. *Diffusion in Solids: Fundamentals, Methods, Materials, Diffusion-Controlled Processes*, volume 155. Springer Series in Solid-State Sciences, 2007.

[160] Madan Lal Mehta. *Random Matrix Theory*. Elsevier, New York, 2004.

[161] Pankaj Mehta, Marin Bukov, Ching-Hao Wang, Alexandre G.R. Day, Clint Richardson, Charles K. Fisher, and David J. Schwab. A high-bias, low-variance introduction to machine learning for physicists. *Physics Reports*, 810:1–124, 2019. A high-bias, low-variance introduction to Machine Learning for physicists.

[162] Pankaj Mehta and David J Schwab. An exact mapping between the Variational Renormalization Group and Deep Learning. *arXiv*, 2014.

[163] Janne-Mieke Meijer, Linda Aarts, Bryan M. van der Ende, Thijs J. H. Vlugt, and Andries Meijerink. Downconversion for solar cells in $YF_3 : Nd^{3+}$, Yb^{3+}. *Phys. Rev. B*, 81:035107, Jan 2010.

[164] Francesco Meinardi, Quinten A Akkerman, Francesco Bruni, Sungwook Park, Michele Mauri, Zhiya Dang, Liberato Manna, and Sergio Brovelli. Doped halide perovskite nanocrystals for reabsorption-free luminescent solar concentrators. *ACS energy letters*, 2(10):2368–2377, 2017.

[165] Sigismund Teunis Alexander George Melissen, Frédéric Labat, Philippe Sautet, and Tangui Le Bahers. Electronic properties of $PbX_3CH_3NH_3$ (X = Cl, Br, I) compounds for photovoltaic and photocatalytic applications. *Physical Chemistry Chemical Physics*, 17(3):2199–2209, 2014.

[166] E. Menéndez-Proupin, Carlos L. Beltrán Ríos, and P. Wahnón. Nonhydrogenic exciton spectrum in perovskite $CH_3NH_3PbI_3$. *Physica Status Solidi (RRL) - Rapid Research Letters*, 9(10):559–563, 2015.

[167] Eduardo Menéndez-Proupin, Carlos L. Beltrán Ríos, Perla Wahnón, and Roel Tempelaar. Erratum to: Nonhydrogenic Exciton Spectrum in Perovskite $CH_3NH_3PbI_3$ (Phys. Status Solidi RRL 2015, 9, 559–563). *Physica Status Solidi (RRL) – Rapid Research Letters*, 13(5):1900075, 2019.

[168] A V Meremianin. Multipole expansions in four-dimensional hyperspherical harmonics. *Journal of Physics A: Mathematical and General*, 39(12):3099, 2006.

[169] Ralf Metzler, Jae-Hyung Jeon, Andrey G Cherstvy, and Eli Barkai. Anomalous diffusion models and their properties: non-stationarity, non-ergodicity, and ageing at the centenary of single particle tracking. *Physical Chemistry Chemical Physics*, 16(44):24128–24164, 2014.

[170] Tyler Milstein, Kyle Kluherz, Daniel Kroupa, Christian S Erickson, James J Yoreo, and Daniel R Gamelin. Anion Exchange and the Quantum-Cutting Energy Threshold in Ytterbium-Doped $CsPb(Cl_{1-x}Br_x)_3$ Perovskite Nanocrystals. *Nano Letters*, 2019.

[171] Tyler J. Milstein, Daniel M. Kroupa, and Daniel R. Gamelin. Picosecond Quantum Cutting Generates Photoluminescence Quantum Yields Over 100% in Ytterbium-Doped $CsPbCl_3$ Nanocrystals. *Nano Letters*, 18(6):3792–3799, 2018.

[172] Atsuhiko Miyata, Anatolie Mitioglu, Paulina Plochocka, Oliver Portugall, Jacob Tse-Wei Wang, Samuel D. Stranks, Henry J. Snaith, and Robin J. Nicholas. Direct measurement of the exciton binding energy and effective masses for charge carriers in organic–inorganic tri-halide perovskites. *Nature Physics*, 11(7):582–587, 2015.

[173] Kiyoshi Miyata, Daniele Meggiolaro, Tuan M Trinh, Prakriti P Joshi, Edoardo

Mosconi, Skyler C Jones, Filippo Angelis, and X-Y Zhu. Large polarons in lead halide perovskites. *Science advances*, 3(8):e1701217, 2017.

[174] K. Momma and F. Izumi. VESTA3 for three-dimensional visualization of crystal, volumetric and morphology data. *Journal of Applied Crystallography*, 44(6):1272–1276, 2011.

[175] H. J. Monkhorst and J. D. Pack. Special points for brillouin-zone integrations. *Physical review B*, 13:5188, 1976.

[176] HJ Monkhorst and JD Pack. Special points for Brillouin-zone integrations. *Physical Review B*, 13(12):5188–5192, 1976.

[177] D Mutter and S. T. Dunham. Calculation of defect concentrations and phase stability in cu_2znsns_4 and $cu_2znsnse_4$ from stoichiometry. *IEEE Journal of Photovoltaics*, 5:1188, 2015.

[178] Kazuo Nakazato and Kazuo Kitahara. Site blocking effect in tracer diffusion on a lattice. *Progress of Theoretical Physics*, 64(6):2261–2264, 1980.

[179] Preetum Nakkiran, Gal Kaplun, Yamini Bansal, Tristan Yang, Boaz Barak, and Ilya Sutskever. Deep Double Descent: Where Bigger Models and More Data Hurt. *arXiv*, 2019.

[180] Preetum Nakkiran, Prayaag Venkat, Sham Kakade, and Tengyu Ma. Optimal Regularization Can Mitigate Double Descent. *arXiv*, 2020.

[181] Georgian Nedelcu, Loredana Protesescu, Sergii Yakunin, Maryna I Bodnarchuk, Matthias J Grotevent, and Maksym V Kovalenko. Fast anion-exchange in highly luminescent nanocrystals of cesium lead halide perovskites (cspbx3, x= cl, br, i). *Nano letters*, 15(8):5635–5640, 2015.

[182] Lance J Nelson, Gus LW Hart, Fei Zhou, Vidvuds Ozoliņš, et al. Compressive sensing as a paradigm for building physics models. *Physical Review B*, 87(3):035125, 2013.

[183] Lance J Nelson, Vidvuds Ozoliņš, C Shane Reese, Fei Zhou, and Gus LW Hart. Cluster expansion made easy with bayesian compressive sensing. *Physical Review B*, 88(15):155105, 2013.

[184] Jigyasa Nigam, Sergey Pozdnyakov, and Michele Ceriotti. Recursive evaluation and iterative contraction of N-body equivariant features. *The Journal of Chemical Physics*, 153(12):121101, 2020.

[185] Alexander Novikov, Dmitry Podoprikhin, Anton Osokin, and Dmitry Vetrov. Tensorizing Neural Networks. *arXiv*, 2015.

[186] Chandramouli Nyshadham, Matthias Rupp, Brayden Bekker, Alexander V. Shapeev, Tim Mueller, Conrad W. Rosenbrock, Gábor Csányi, David W. Wingate, and Gus L. W. Hart. Machine-learned multi-system surrogate models for materials prediction. *npj Computational Materials*, 5(1), 2019.

[187] Robert Oeckl. *Discrete Gauge Theory*. Imperial College Press, London, 2005.

[188] Vadim Oganesyan and David A. Huse. Localization of interacting fermions at high temperature. *Phys. Rev. B*, 75:155111, Apr 2007.

[189] L. E. Oikkonen, M. G. Ganchenkova, A. P. Seitsonen, and R. M. Nieminen. Formation, migration, and clustering of point defects in $cuinse_2$ from first principles. *Journal of Physics: Condensed Matter*, 26(34):345501, 2014.

[190] Berk Onat, Christoph Ortner, and James R Kermode. Sensitivity and dimensionality of atomic environment representations used for machine learning interatomic potentials. *The Journal of Chemical Physics*, 153(14):144106, 2020.

[191] Shyue Ping Ong, William Davidson Richards, Anubhav Jain, Geoffroy Hautier, Michael Kocher, Shreyas Cholia, Dan Gunter, Vincent L. Chevrier, Kristin A. Persson, and Gerbrand Ceder. Python Materials Genomics (pymatgen): A robust, open-source python library for materials analysis. *Computational Materials Science*, 68:314–319, 2013.

[192] Don N. Page. Average entropy of a subsystem. *Phys. Rev. Lett.*, 71:1291–1294, Aug 1993.

[193] Arijeet Pal and David A. Huse. Many-body localization phase transition. *Phys. Rev. B*, 82:174411, Nov 2010.

[194] Gencai Pan, Xue Bai, Dongwen Yang, Xu Chen, Pengtao Jing, Songnan Qu, Lijun Zhang, Donglei Zhou, Jinyang Zhu, Wen Xu, Biao Dong, and Hongwei Song. Doping Lanthanide into Perovskite Nanocrystals: Highly Improved and Expanded Optical Properties. *Nano Letters*, 2017.

[195] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.

[196] Roger Penrose. Angular momentum: an approach to combinatorial space-time. *Quantum Theory and Beyond*, pages 151–180, 1971.

[197] John P. Perdew, Kieron Burke, and Matthias Ernzerhof. Generalized gradient approximation made simple. *Phys. Rev. Lett.*, 77:3865–3868, Oct 1996.

[198] John P. Perdew, Adrienn Ruzsinszky, Gábor I. Csonka, Oleg A. Vydrov, Gustavo E. Scuseria, Lucian A. Constantin, Xiaolan Zhou, and Kieron Burke. Restoring the Density-Gradient Expansion for Exchange in Solids and Surfaces. *Physical Review Letters*, 100(13):136406, 2008.

[199] Alejandro Perez. The Spin-Foam Approach to Quantum Gravity. *Living Reviews in Relativity*, 16(1):3, 2013.

[200] C Persson, Y. J. Zhao, S Lany, and A Zunger. n-type doping of $cuinse_2$ and $cugase_2$. *Physical Review B*, 72:035211, 2005.

[201] J Pohl and K Albe. Intrinsic point defects in $cuinse_2$ and $cugase_2$ as seen via screened-exchange hybrid density functional theory. *Physical Review B*, 87:245203, 2013.

[202] Johan Pohl and Karsten Albe. Thermodynamics and kinetics of the copper vacancy in $cuinse_2$, $cugase_2$, $cuins_2$, and $cugas_2$ from screened-exchange hybrid density functional theory. *Journal of Applied Physics*, 108(2):023509, 2010.

[203] Johan Pohl, Andreas Klein, and Karsten Albe. Role of copper interstitials in CuInSe2: First-principles calculations. *Physical Review B*, 84(12):121201, 2011.

[204] J Pollmann and H Büttner. Effective Hamiltonians and bindings energies of Wannier excitons in polar semiconductors. *Physical Review B*, 16(10):4480–4490, 1977.

[205] M. J. D. Powell. The BOBYQA Algorithm for Bound Constrained Optimization without Derivatives. *Technical Report, Department of Applied Mathematics and Theoretical Physics, University of Cambridge*, 01 2009.

[206] Sergey N. Pozdnyakov, Michael J. Willatt, Albert P. Bartók, Christoph Ortner, Gábor Csányi, and Michele Ceriotti. Incompleteness of atomic structure representations. *Phys. Rev. Lett.*, 125:166001, Oct 2020.

[207] Loredana Protesescu, Sergii Yakunin, Maryna I Bodnarchuk, Franziska Krieg, Riccarda Caputo, Christopher H Hendon, Ruo Yang, Aron Walsh, and Maksym V Kovalenko. Nanocrystals of Cesium Lead Halide Perovskites ($CsPbX_3$, X = Cl, Br, and I): Novel Optoelectronic Materials Showing Bright Emission with Wide Color Gamut. *Nano Letters*, 15(6):3692–6, 2015.

[208] B Puchala and A Van der Ven. Thermodynamics of the zr-o system from first-principles calculations. *Physical review B*, 88(9):094108, 2013.

[209] M. Puppin, S. Polishchuk, N. Colonna, A. Crepaldi, D. N. Dirin, O. Nazarenko, R. De Gennaro, G. Gatti, S. Roth, T. Barillot, L. Poletto, R. P. Xian, L. Rettig, M. Wolf, R. Ernstorfer, M. V. Kovalenko, N. Marzari, M. Grioni, and M. Chergui. Evidence of Large Polarons in Photoemission Band Mapping of the Perovskite Semiconductor $CsPbBr_3$. *Physical Review Letters*, 124(20):206402, 2020.

[210] David A Reed and Gert Ehrlich. Surface diffusion, atomic jump rates and thermodynamics. *Surface Science*, 102(2-3):588–609, 1981.

[211] Chase Roberts, Ashley Milsted, Martin Ganahl, Adam Zalcman, Bruce Fontaine, Yijian Zou, Jack Hidary, Guifre Vidal, and Stefan Leichenauer. Tensornetwork: A library for physics and machine learning, 2019.

[212] Joo Yeon D. Roh, Matthew D. Smith, Matthew J. Crane, Daniel Biner, Tyler J. Milstein, Karl W. Krämer, and Daniel R. Gamelin. Yb^{3+} speciation and energy-transfer dynamics in quantum-cutting Yb^{3+}-doped $CsPbCl_3$ perovskite nanocrystals and single crystals. *Physical Review Materials*, 4(10):105405, 2020.

[213] JM Sanchez. Cluster expansions and the configurational energy of alloys. *Physical review B*, 48(18):14013, 1993.

[214] JM Sanchez. Cluster expansion and the configurational theory of alloys. *Physical Review B*, 81(22):224202, 2010.

[215] Juan M Sanchez, Francois Ducastelle, and Denis Gratias. Generalized cluster description of multicomponent systems. *Physica A: Statistical Mechanics and its Applications*, 128(1-2):334–350, 1984.

[216] Philipp Schmoll, Sukhbinder Singh, Matteo Rizzi, and Román Orús. A programming guide for tensor networks with global su(2) symmetry. *Annals of Physics*, 419:168232, 2020.

[217] Stefan Schnabel, Daniel T Seaton, David P Landau, and Michael Bachmann. Microcanonical entropy inflection points: Key to systematic understanding of transitions in finite systems. *Physical Review E*, 84(1):011127, 2011.

[218] Atsuto Seko, Yukinori Koyama, and Isao Tanaka. Cluster expansion method for multicomponent systems based on optimal selection of structures for density-functional theory calculations. *Physical Review B*, 80(16):165122, 2009.

[219] Michael Sendner, Pabitra K. Nayak, David A. Egger, Sebastian Beck, Christian Müller, Bernd Epding, Wolfgang Kowalsky, Leeor Kronik, Henry J. Snaith, Annemarie Pucci, and Robert Lovrinčić. Optical phonons in methylammonium lead halide perovskites and implications for charge transport. *Materials Horizons*, 3(6):613–620, 2016.

[220] Daniel Shaffer, Claudio Chamon, Alioscia Hamma, and Eduardo R Mucciolo. Irreversibility and entanglement spectrum statistics in quantum circuits. *Journal of Statistical Mechanics: Theory and Experiment*, 2014(12):P12007, dec 2014.

[221] A. Shapeev. Accurate representation of formation energies of crystalline alloys with many components. *Computational Materials Science*, 139:26–30, 2017.

[222] Alexander V Shapeev. Moment tensor potentials: A class of systematically improvable interatomic potentials. *Multiscale Modeling & Simulation*, 14(3):1153–1173, 2016.

[223] Hongliang Shi and Mao-Hua Du. Shallow halogen vacancies in halide optoelectronic materials. *Physical Review B*, 90(17):174103, 2014.

[224] Susanne Siebentritt, Levent Gütay, David Regesch, Yasuhiro Aida, and Valérie Deprédurand. Why do we make cu (in, ga) se2 solar cells non-stoichiometric? *Solar Energy Materials and Solar Cells*, 119:18–25, 2013.

[225] Michael C De Siena, David E Sommer, Sidney E Creutz, Scott T Dunham, and Daniel R Gamelin. Spinodal Decomposition During Anion Exchange in Colloidal Mn 2+ -Doped CsPbX 3 (X = Cl, Br) Perovskite Nanocrystals. *Chemistry of Materials*, 31(18):7711–7722, 2019.

[226] Sukhwinder Singh, Robert N. C. Pfeifer, and Guifré Vidal. Tensor network decompositions in the presence of a global symmetry. *Phys. Rev. A*, 82:050301, Nov 2010.

[227] Sukhwinder Singh and Guifre Vidal. Tensor network states and algorithms in the presence of a global su(2) symmetry. *Phys. Rev. B*, 86:195114, Nov 2012.

[228] T. Skettrup, M. Suffczynski, and W. Gorzkowski. Properties of Excitons Bound to Ionized Donors. *Physical Review B*, 4(2):512–517, 1970.

[229] Daniel J Slotcavage, Hemamala I Karunadasa, and Michael D McGehee. Light-induced phase segregation in halide-perovskite absorbers. *ACS Energy Letters*, 1(6):1199–1205, 2016.

[230] Vedene H. Smith and Piotr Petelenz. Effective electron-hole interaction potentials and the binding energies of exciton-ionized-donor complexes. *Physical Review B*, 17(8):3253–3261, 1978.

[231] D Soltz, G Dagan, and D Cahen. Ionic mobility and electronic junction movement in $cuinse_2$. *Solid State Ionics*, 28:1105, 1988.

[232] D. E. Sommer, D Mutter, and S. T. Dunham. Defects in na-, k-, and cd-doped $cuinse_2$:canonical thermodynamics based on ab initio calculations. *IEEE Journal of Photovoltaics*, 7:2156, 2017.

[233] David E Sommer and Scott T Dunham. Atomistic models of cu diffusion in cuinse2 under variations in composition. *Journal of Applied Physics*, 123(11):115116, 2018.

[234] David E. Sommer and Scott T. Dunham. Understanding copper diffusion in cuinse2 with first-principles based atomistic and continuum models. *Journal of Applied Physics*, 130(23):235701, 2021.

[235] Conrad Spindler, Finn Babbe, Max Hilaire Wolter, Florian Ehré, Korra Santhosh, Pit Hilgert, Florian Werner, and Susanne Siebentritt. Electronic defects in cu (in, ga) s e 2: Towards a comprehensive model. *Physical Review Materials*, 3(9):090302, 2019.

[236] P. E. Stallworth, J.-F. Guillemoles, J Flowers, J Vedel, and S. G. Greenbaum. NMR studies of $cuins_2$ and $cuinse_2$crystals grown by the bridgman method. *Solid State Communications*, 113(9):527–532, 2000.

[237] Geoffrey E. Stedman. *Diagram Techniques in Group Theory*. Cambridge University Press, Cambridge, 1990.

[238] Miles E Stoudenmire and David J Schwab. Supervised Learning with Quantum-Inspired Tensor Networks. *arXiv*, 2016.

[239] Christopher Sutton, Luca M. Ghiringhelli, Takenori Yamamoto, Yury Lysogorskiy, Lars Blumenthal, Thomas Hammerschmidt, Jacek R. Golebiowski, Xiangyue Liu, Angelo Ziletti, and Matthias Scheffler. Crowd-sourcing materials-science challenges with the NOMAD 2018 Kaggle competition. *npj Computational Materials*, 5(1):111, 2019.

[240] Yasuyuki Suzuki and Kalman Varga. *Stochastic Variational Approach to Quantum-Mechanical Few-Body Problems*. Springer, Heidelberg, 1998.

[241] Attila Szabo and Neil S. Ostlund. *Modern Quantum Chemistry*. Dover, New York, 1996.

[242] RA Tahir-Kheli. Correlation factors for atomic diffusion in nondilute multicomponent alloys with arbitrary vacancy concentration. *Physical Review B*, 28(6):3049, 1983.

[243] RA Tahir-Kheli and RJ Elliott. Correlated random walk in lattices: tracer diffusion at general concentration. *Physical Review B*, 27(2):844, 1983.

[244] W Tang, E Sanville, and G Henkelman. A grid-based bader analysis algorithm without lattice bias. *Journal of Physics: Condensed Matter*, 21(8):084204, 2009.

[245] Benjamin Tell and P. M. Bridenbaugh. Photovoltaic properties and junction formation in cuinse$_2$. *Journal of Applied Physics*, 48(6):2477–2480, 1977.

[246] Ajit J. Thakkar and Vedene H. Smith. Compact and accurate integral-transform wave functions. I. The $1S^1$ state of the helium-like ions from H^- through Mg^{10+}. *Physical Review A*, 15(1):1–15, 1977.

[247] John C Thomas and Anton Van der Ven. The exploration of nonlinear elasticity and its efficient parameterization for crystalline materials. *Journal of the Mechanics and Physics of Solids*, 107:76–95, 2017.

[248] A.P. Thompson, L.P. Swiler, C.R. Trott, S.M. Foiles, and G.J. Tucker. Spectral neighbor analysis method for automated generation of quantum-accurate interatomic potentials. *Journal of Computational Physics*, 285:316–330, 2015.

[249] Jenya Tilchin, Dmitry N Dirin, Georgy I Maikov, Aldona Sashchiuk, Maksym V Kovalenko, and Efrat Lifshitz. Hydrogen-like Wannier–Mott Excitons in Single Crystal of Methylammonium Lead Bromide Perovskite. *ACS Nano*, 10(6):6363–6371, 2016.

[250] Atsushi Togo and Isao Tanaka. First principles phonon calculations in materials science. *Scripta Materialia*, 108:1–5, 2015.

[251] Dallas R Trinkle. Automatic numerical evaluation of vacancy-mediated transport for arbitrary crystals: Onsager coefficients in the dilute limit using a green function approach. *Philosophical Magazine*, 97(28):2514–2563, 2017.

[252] Thumu Udayabhaskararao, Lothar Houben, Hagai Cohen, Matan Menahem, Iddo Pinkas, Liat Avram, Tamar Wolf, Ayelet Teitelboim, Michal Leskes, Omer Yaffe, et al. A mechanistic study of phase transformation in perovskite nanocrystals driven by ligand passivation. *Chemistry of Materials*, 30(1):84–93, 2018.

[253] Paolo Umari, Edoardo Mosconi, and Filippo De Angelis. Infrared Dielectric Screening Determines the Low Exciton Binding Energy of Metal-Halide Perovskites. *The Journal of Physical Chemistry Letters*, 9(3):620–627, 2018.

[254] EL Unger, Lukas Kegelmann, K Suchan, D Sörell, Lars Korte, and Susanne Albrecht. Roadmap and roadblocks for the band gap tunability of metal halide perovskites. *Journal of Materials Chemistry A*, 5(23):11401–11409, 2017.

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

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

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

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

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

[260] Bryan M. van der Ende, Linda Aarts, and Andries Meijerink. Near-infrared quantum cutting for photovoltaics. *Advanced Materials*, 21(30):3073–3077, 2009.

[261] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of Machine Learning Research*, 9(86):2579–2605, 2008.

[262] A Van der Ven, JC Thomas, Qingchuan Xu, and J Bhattacharya. Linking the electronic structure of solids to their thermodynamic and kinetic properties. *Mathematics and computers in simulation*, 80(7):1393–1410, 2010.

[263] Evert van Nieuwenburg and Oded Zilberberg. Entanglement spectrum of mixed states. *Phys. Rev. A*, 98:012327, Jul 2018.

[264] P. Vergeer, T. J. H. Vlugt, M. H. F. Kox, M. I. den Hertog, J. P. J. M. van der Eerden, and A. Meijerink. Quantum cutting by cooperative energy transfer in $Yb_xY_{1-x}PO_4 : Tb^{3+}$. *Phys. Rev. B*, 71:014119, Jan 2005.

[265] G. Vidal. Entanglement renormalization. *Phys. Rev. Lett.*, 99:220405, Nov 2007.

[266] G. H. Vineyard. Frequency factors and isotope effects in solid state rate processes. *Journal of Physics and Chemistry of Solids*, 3:121, 1957.

[267] Aron Walsh, Alexey A. Sokol, and C. Richard A. Catlow. Free energy of defect formation: Thermodynamics of anion Frenkel pairs in indium oxide. *Physical Review B*, 83(22):224105, 2011.

[268] Fugao Wang and D. P. Landau. Determining the density of states for classical statistical models: A random walk algorithm to produce a flat histogram. *Phys. Rev. E*, 64:056101, Oct 2001.

[269] Fugao Wang and David P Landau. Efficient, multiple-range random walk algorithm to calculate the density of states. *Physical review letters*, 86(10):2050, 2001.

[270] Qian Wang, Xisheng Zhang, Zhiwen Jin, Jingru Zhang, Zhenfei Gao, Yongfang Li, and Shengzhong Frank Liu. Energy-down-shift cspbcl3: Mn quantum dots for boosting the

efficiency and stability of perovskite solar cells. *ACS Energy Letters*, 2(7):1479–1486, 2017.

[271] Kei Watanabe, Masanori Koshimizu, Takayuki Yanagida, Yutaka Fujimoto, and Keisuke Asai. Luminescence and scintillation properties of La- and La,Ag-doped $CsPbCl_3$ single crystals. *Japanese Journal of Applied Physics*, 55(2S):02BC20, 2016.

[272] Hendrik Weimer, Augustine Kshetrimayum, and Román Orús. Simulation methods for open quantum many-body systems. *Rev. Mod. Phys.*, 93:015008, Mar 2021.

[273] F. Werner, D. Colombara, M. Melchiorre, N. Valle, B. El Adib, C. Spindler, and S. Siebentritt. Doping mechanism in pure CuInSe2. *Journal of Applied Physics*, 119(17):173103, 2016.

[274] Lucy D. Whalley. effmass: An effective mass package, Aug 2018.

[275] Roeland Wiersema, Cunlu Zhou, Yvette de Sereville, Juan Felipe Carrasquilla, Yong Baek Kim, and Henry Yuen. Exploring entanglement and optimization within the hamiltonian variational ansatz. *PRX Quantum*, 1:020319, Dec 2020.

[276] Michael J Willatt, Félix Musil, and Michele Ceriotti. Feature optimization for atomistic machine learning yields a data-driven construction of the periodic table of the elements. *Physical Chemistry Chemical Physics*, 20(47):29661–29668, 2018.

[277] Michael J. Willatt, Félix Musil, and Michele Ceriotti. Atom-density representations for machine learning. *The Journal of Chemical Physics*, 150(15), 2019.

[278] Lechun Xie, Yuhua Wang, and Huijuan Zhang. Near-infrared quantum cutting in YPO_4:Yb^{3+}, Tm^{3+} via cooperative energy transfer. *Applied Physics Letters*, 94(6):061905, 2009.

[279] Z Yang, A Surrente, K Galkowski, A Miyata, O Portugall, R J Sutton, A A Haghighirad, H J Snaith, D K Maude, P Plochocka, and R J Nicholas. Impact of the Halide

Cage on the Electronic Properties of Fully Inorganic Cesium Lead Halide Perovskites. *ACS Energy Letters*, 2(7):1621–1627, 2017.

[280] Zhi-Cheng Yang, Claudio Chamon, Alioscia Hamma, and Eduardo R. Mucciolo. Two-component structure in the entanglement spectrum of highly excited states. *Phys. Rev. Lett.*, 115:267206, Dec 2015.

[281] Zhi-Cheng Yang, Alioscia Hamma, Salvatore M. Giampaolo, Eduardo R. Mucciolo, and Claudio Chamon. Entanglement complexity in quantum many-body dynamics, thermalization, and localization. *Phys. Rev. B*, 96:020408, Jul 2017.

[282] Hong Yao and Xiao-Liang Qi. Entanglement entropy and entanglement spectrum of the kitaev model. *Phys. Rev. Lett.*, 105:080501, Aug 2010.

[283] Wan-Jian Yin, Tingting Shi, and Yanfa Yan. Unusual defect physics in ch3nh3pbi3 perovskite solar cell absorber. *Applied Physics Letters*, 104(6):063903, 2014.

[284] Wan-Jian Yin, Yanfa Yan, and Su-Huai Wei. Anomalous alloy properties in mixed halide perovskites. *The Journal of Physical Chemistry Letters*, 5(21):3625–3631, 2014. PMID: 26278729.

[285] Joshua Young and James M Rondinelli. Octahedral Rotation Preferences in Perovskite Iodides and Bromides. *The Journal of Physical Chemistry Letters*, 7(5):918–922, 2016.

[286] Peter Yu and Manuel Cadona. *Fundamentals of Semiconductors*. Springer, Heidelberg, 2010.

[287] Andriy Zakutayev, Christopher M Caskey, Angela N Fioretti, David S Ginley, Julien Vidal, Vladan Stevanovic, Eric Tea, and Stephan Lany. Defect tolerant semiconductors for solar energy conversion. *The journal of physical chemistry letters*, 5(7):1117–1125, 2014.

[288] Min Zeng, Flavia Artizzu, Jing Liu, Shalini Singh, Federico Locardi, Dimitrije Mara, Zeger Hens, and Rik Van Deun. Boosting the Er^{3+} 1.5 μm Luminescence in $CsPbCl_3$ Perovskite Nanocrystals for Photonic Devices Operating at Telecommunication Wavelengths. *ACS Applied Nano Materials*, 3(5):4699–4707, 2020.

[289] Lei Zhang, Justin A. Reyes, Stefanos Kourtis, Claudio Chamon, Eduardo R. Mucciolo, and Andrei E. Ruckenstein. Nonuniversal entanglement level statistics in projection-driven quantum circuits. *Phys. Rev. B*, 101:235104, Jun 2020.

[290] Qing Zhang, Rui Su, Xinfeng Liu, Jun Xing, Tze Chien Sum, and Qihua Xiong. High-Quality Whispering-Gallery-Mode Lasing from Cesium Lead Halide Perovskite Nanoplatelets. *Advanced Functional Materials*, 26(34):6238–6245, 2016.

[291] S B Zhang. The microscopic origin of the doping limits in semiconductors and wide-gap materials and recent developments in overcoming these limits: a review. *Journal of Physics: Condensed Matter*, 14(34):R881, 2002.

[292] S. B. Zhang, S. H. Wei, A Zunger, and H Katayama-Yoshida. Defect physics of the $cuinse_2$ chalcopyrite semiconductor. *Physical Review B*, 57(16):9642, 1998.

[293] SB Zhang, Su-Huai Wei, Alex Zunger, and H Katayama-Yoshida. Defect physics of the cuinse 2 chalcopyrite semiconductor. *Physical Review B*, 57(16):9642, 1998.

[294] Wei Zhang, Giles E Eperon, and Henry J Snaith. Metal halide perovskites for energy applications. *Nature Energy*, 1(6):1–8, 2016.

[295] Chenggang Zhou and RN Bhatt. Understanding and improving the wang-landau algorithm. *Physical Review E*, 72(2):025701, 2005.

[296] Donglei Zhou, Dali Liu, Gencai Pan, Xu Chen, Dongyu Li, Wen Xu, Xue Bai, and Hongwei Song. Cerium and Ytterbium Codoped Halide Perovskite Quantum Dots: A Novel and Efficient Downconverter for Improving the Performance of Silicon Solar Cells. *Advanced Materials*, 29(42):1704149, 2017.

[297] Marko Žnidarič. Entanglement of random vectors. *Journal of Physics A: Mathematical and Theoretical*, 40(3):F105–F111, dec 2006.

[298] Yunxing Zuo, Chi Chen, Xiangguo Li, Zhi Deng, Yiming Chen, Jörg Behler, Gábor Csányi, Alexander V Shapeev, Aidan P Thompson, Mitchell A Wood, et al. Performance and cost assessment of machine learning interatomic potentials. *The Journal of Physical Chemistry A*, 124(4):731–745, 2020.