A Modern Approach to Doping, Defects, and Disorder - Insights Into the Topological Semimetal Cd_3As_2 Through Advanced Thermodynamic Modeling

by

Chase Brooks

B.S., New Mexico State University, 2019

M.S., University of Colorado Boulder, 2022

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Department of Physics

2025

Committee Members:

Daniel Dessau, Chair

Mark van Schilfgaarde

Victor Gurarie

Gang Cao

Sanghamitra Neogi

Brooks, Chase (Ph.D., Physics)

A Modern Approach to Doping, Defects, and Disorder - Insights Into the Topological Semimetal Cd_3As_2 Through Advanced Thermodynamic Modeling

Thesis directed by Stephan Lany of the National Renewable Energy Laboratory and Professor Daniel Dessau.

Topological semimetals are an intriguing class of solid-state materials with many distinct physical properties, but those properties can only be readily observed or utilized if the Fermi level E_F is in close proximity to the symmetry protected band crossing. In reality, lab-grown materials will suffer from crystallographic defects and unintended doping that can move E_F away from the dispersive crossing point, obscure the unique topological features, and affect carrier transport. Understanding and controlling sources of disorder in these materials is therefore essential to fully harness their potential. Thermodynamic models serve as powerful tools for gaining insights into the formation of such undesirable defects. They can be utilized to determine why certain defects form, how synthesis conditions can be changed to mitigate them, and how extrinsic dopants and alloying can be used to combat them. While many techniques have been developed to simulate disorder in conventional solid-state systems with first-principles calculations, special attention has not historically been given to how techniques need to be updated when applied to topological semimetals. Simultaneously, typical models often include assumptions that certain quantities, such as defect concentrations in a sample after growth or the enthalpy of mixing in an alloy, can be treated as temperature independent. This work strives to address both of these shortcomings by combining a unique treatment of defect behavior in topological semimetals with expanded, broadly applicable models that capture the missing temperature dependence. I apply the revised approaches to the Dirac semimetal Cd_3As_2, which is known to form with an elevated Fermi level, and predict strategies for tuning E_F closer to the Dirac point. These novel methods lay the foundation for future studies of disorder and defect control in other topological semimetals and beyond.

Dedication

To Miles Parker Brooks. May he live in a world where compassion and logic supersede hate and fear.

Acknowledgements

I would first like to thank my advisor, Dr. Stephan Lany, for guiding me through my PhD. Your knowledge and expertise were truly invaluable, and I feel strongly that you have exhibited all the qualities of a great mentor. I would also like to thank Drs. Kirstin Alberi, Mark van Schilfgaarde, and Dan Dessau for their collaboration and for the opportunity to contribute to this project.

My success in graduate school is also largely a result of the friendships and collaborations with my fellow students. Special thanks to Dr. Evan Owen for his incredible knowledge and overt friendliness that got me out of my shell; to Benjamin Marinoff for being my partner for both homework assignments and watching TV; to soon-to-be Dr. Kenny Higginbotham for always destroying me at racquetball while being the most genuine person I know; to Dr. Curtis Peterson and his wife Sam for being the most generous, inviting, and outgoing. Thanks also to the Righteous Crusaders—Bah (Ryan), Brion (Evan), Darryl (Jesse), Garth (Curtis), Nik (Cash), Peri (Sam), and Sungura (Amanda)—and their GIGA quest to destroy the Astral Cage.

Lastly, I would like to thank my amazing family. To my parents Marty and Kelly, I have felt your love and support every day since I left home, even here in Colorado. To my siblings Bailey and Tyler, you are probably two of the funniest human beings on the planet (you're welcome), you are two of my best friends, and you have both greatly reinforced my own knowledge by asking for help with homework (again, you're welcome). To my son Miles, you have made finishing my graduate work and writing this thesis 100 times harder than it otherwise would have been, but I am reminded every morning when I see your smile that it is all worth it. To my wife Dr. Amanda Brooks, you not only kept me grounded and focused on the light at the end of the tunnel... you are that light.

Contents

Chapter

1 Introduction **1**

1.1 Overview of topological materials 1

1.1.1 Weyl semimetals 2

1.1.2 Dirac semimetals 4

1.1.3 Properties of topological semimetals 4

1.1.4 Cadmium Arsenide 6

1.2 Density functional theory 8

1.2.1 DFT formalism 9

1.2.2 Applications beyond zero temperature - thermodynamic modeling 11

1.2.3 Shortcomings 12

1.3 Corrections and modifications 14

1.3.1 Strongly constrained and appropriately normed meta-GGA functional 15

1.3.2 GW theory 16

1.4 Overview of this thesis 18

2 Band energy dependence of defect formation and self-doping in a topological semimetal **21**

2.1 Introduction 21

2.2 Methods 23

2.3 Finite-size effects of defect formation energies 24

2.4 Band energy dependence . . . 28
2.4.1 QSGW band structure and high-resolution density of states . . . 29
2.5 Absolute defect formation energies . . . 29
2.5.1 Limits to chemical potentials . . . 31
2.5.2 Results for intrinsic formation energies . . . 32
2.5.3 Defect density of states and numerical values of formation energies . . . 32
2.6 Defect equilibria . . . 33
2.6.1 Concentration of non-dilute defects . . . 35
2.6.2 Charge balance condition in the thermodynamic model . . . 35
2.6.3 Results for defect equilibria . . . 36
2.6.4 Unique behavior of Fermi level with temperature for a topological semimetal 39
2.7 Conclusions . . . 40

3 Extrinsic doping control from defect theory with site redistribution in a pseudo-equilibrium 42

3.1 Introduction - Constrained Defect Equilibria . . . 43
3.2 Redistribution of Intrinsic Defects . . . 45
3.3 Defect behavior and dopant density of states . . . 46
3.4 Choice of Extrinsic Dopant . . . 48
3.5 Combined Defect and Dopant Redistributions . . . 51
3.6 Cd_3As_2 Extrinsic Doping: Au . . . 52
3.7 Cd_3As_2 Extrinsic Doping: Ge . . . 56
3.8 Discussion and Conclusions . . . 59
3.9 Methods . . . 60
3.9.1 First-principles Total Energy Calculations and Defect Equilibria . . . 60
3.9.2 Constrained Equilibrium and Temperature Dependence of Defect Concentrations with Competing Elements . . . 61
3.9.3 Simulated Diffusion with the Nudged Elastic Band Method . . . 64

4 Heterostructural alloying and phase diagram construction for $(Cd_{1-x}Zn_x)_3As_2$ **67**

4.1 Introduction . . . 67

4.2 Methods . . . 69

4.2.1 Compuational details . . . 69

4.3 First-principles Monte Carlo . . . 70

4.3.1 Probabilities in the chosen Monte Carlo ensemble . . . 72

4.4 Thermodynamic free energy integration . . . 73

4.4.1 Origin of the expression for thermodynamic integration . . . 76

4.5 Phase diagram . . . 77

4.5.1 Additional free energies of mixing . . . 80

4.6 Conclusions . . . 82

5 Final thoughts and outlook **83**

5.1 Further possibilities for Cadmium Arsenide . . . 83

5.2 Applying these models to other systems . . . 85

References **87**

Appendix

A Berry curvature in the vacinity of a Weyl point **98**

B Derivation of the non-dilute concentration of defects in thermal equilibrium **102**

C Derivation of Gibbs-Helmholtz equation **104**

Tables

Table

2.1 Calculated defect formation energies for all Wyckoff positions at reference condition $E_\mathrm{F} = E_\mathrm{DP}$ and $\Delta\mu_\mathrm{Cd} = \Delta\mu_\mathrm{As} = 0$. 33

2.2 Defect equilibria for Cd_3As_2 obtained from first-principles defect formation energies. Given are the concentrations c_D of the intrinsic defects and the resulting net doping c_nd depending on the growth temperature T_g and the Cd- or As-rich condition. . . . 37

Figures

Figure

2.1 (a) The supercell size dependence of the formation energy ΔH of the $\mathrm{Cd_i}$ defect in Cd_3As_2 as a function of inverse linear supercell size $1/L = \Omega^{-1/3}$ obtained from supercell volume Ω. ΔH_0 corresponds to the dilute limit of the formation energy. (b-d) The electronic band structure of Cd_3As_2 with energies measured from the Dirac point energy E_{DP}, calculated with GGA (b), SCAN (c), and quasiparticle self-consistent GW theory (d). All band structures were calculated for the same atomic structure that was relaxed with the SCAN functional. 25

2.2 The supercell size-dependence of the formation energy ΔH of the V_{Cd} defect in Cd_3As_2 as a function of inverse linear supercell size $1/L = \Omega^{-1/3}$ obtained from supercell volume Ω. ΔH_0 corresponds to the dilute limit of the formation energy. Labels indicate the number of atoms in the respective host supercells. 26

2.3 (a) The QSGW band structure zoomed in near the Dirac point. (b) Occupied and (c) unoccupied density of states calculated in QSGW as a function of energy E measured with respect to the Dirac point E_{DP}. Dashed red line shows the numerically fitted model function g_{QSGW}. 30

2.4 Chemical potential stability phase diagram for Cd_3As_2. 31

2.5 Formation energies $\Delta H_{D,q}$ of the intrinsic point defects in Cd_3As_2 as a function of Fermi level E_F in (a) Cd-rich and (b) As-rich conditions. Only the lowest energy Wyckoff positions are shown for V_{As} and V_{Cd}. The GW correction $\Delta E_{DP} = -0.19$ eV and equilibrium Fermi level E_F^{eq} for growth at $T_g = 400$ K and 800 K are indicated. . 33

2.6 Formation energies $\Delta H_{D,q}$ of the intrinsic point defects in Cd_3As_2 as a function of Fermi level E_F in (a) Cd-rich and (b) As-rich conditions. (c) Host and defect density of states (DOS) for point defects in Cd_3As_2 calculated with the SCAN functional as a function of energy E measured with respect to the Dirac point E_{DP}. Only the lowest energy Wyckoff positions are shown. 34

2.7 (a) Temperature dependence of E_F in the presence of defect concentrations determined for three different growth temperatures T_g (circles). (b) Low temperature limit of E_F as a function of T_g. Solid and dashed lines indicate Cd-rich and As-rich conditions, respectively. 37

2.8 The (a) electron and (b) hole concentrations n_e and n_h plotted as a function of temperature T for the three growth temperatures in Fig. 3(a) (indicated as upper limits). Shown for Cd-rich and As-rich conditions. A logarithmic scale is used for hole concentrations. 38

2.9 (a) Fermi-Dirac occupation of electrons corresponding to As-rich growth of Cd_3As_2 at 800 K (blue) and then cooled to 250 K (orange). The temperature dependent Fermi levels E_F are indicated by vertical dashed lines. (b) QSGW density of states data as a function of energy E near the Dirac point E_{DP}. (c)-(d) Model density of states $g_{QSGW}(E)$ weighted with the Fermi-Dirac distribution for electrons (dark gray) and holes (light gray) at the respective temperatures, 800 K (c) and 250 K (d). 41

3.1 Concentrations c_D of intrinsic Cd defects in Cd_3As_2 as a function of temperature T when grown under As-rich conditions at 600 K (dashed) and 800 K (solid). 46

3.2 Host and defect/dopant density of states (DOS) in Cd_3As_2 calculated with the SCAN functional (including spin-orbit coupling) as a function of energy E measured with respect to the Dirac point E_{DP}. Only the lowest energy Wyckoff positions are shown for each defect. 48

3.3 (a) Formation energies $\Delta H_{\mathrm{D},q}$ of donor and acceptor defects for Cd, Au, and Ge, as functions of Fermi level E_{F}. The intersection points between donors and acceptors are utilized in constructing (b) The two-objective dopant selection for lowering the Fermi level (low $E_{\mathrm{F}}^{\mathrm{dc}}$) with high dopant solubility (low ΔH_{D}). Considered are Group I (circles) and IV (squares) dopants under Cd- and As-rich conditions, in comparison to the intrinsic Cd defects (crosses). Au and Ge are chosen as the most promising extrinsic dopants on the Pareto frontier (red line). 50

3.4 Initial defect concentrations $c_{\mathrm{D}}^{\mathrm{g}}$ and net carrier concentration c_{nc} in Cd_3As_2 as functions of Au doping concentration for Cd-rich (dashed) and As-rich (solid) growth at (a) $T_{\mathrm{g}} = 800$ K and (b) $T_{\mathrm{g}} = 400$ K. The defect concentrations after growth are then measured from the initial growth values and plotted as a function of temperature T for As-rich growth at (c) 800 K with an equilibrium concentration of Au (1.95×10^{18} cm^{-3}) and (d) 400 K with 1% Au. 54

3.5 (a) Defect concentrations c_{D} for Ge-doped Cd_3As_2 measured from the indicated initial growth values $c_{\mathrm{D}}^{\mathrm{g}}$. (b) The associated net carrier concentration c_{nc}, both plotted as functions of temperature T. Results are shown for Cd-rich growth at 800 K with an equilibrium concentration of Ge (2.21×10^{19} cm^{-3}). 57

3.6 Low temperature ($T \to 0$) net carrier concentration c_{nc} of Cd_3As_2 grown at $T_{\mathrm{g}} =$ 400 K, plotted as a function of both the Ge doping concentration and Cd chemical potential $\Delta\mu_{\mathrm{Cd}}$. Net n-type doping is colored red and net p-type is colored blue. . . 59

3.7 Nudged Elastic Band supercell energies E plotted as a function of the image index for the simulated diffusion paths of (a) the recombination of a Cd_i and V_{Cd}, (b) Au moving from an interstitial to a neighboring vacant Cd site, and (c) Ge moving from a Cd site to a neighboring As site. Energies are measured with respect to the minimum for each diffusion. 65

4.1 Monte Carlo simulated annealing of $(Cd_{0.5}Zn_{0.5})_3As_2$ for the Cd_3As_2 ground-state crystal structure, averaged over ten independent samples and equilibrated sequentially at (a) 800, (b) 600, and (c) 400 K. Energies are measured with respect to the average of the initial random ($T = \infty$) configurations. Error bars indicate the standard deviation across the independent MC runs at each iteration. 71

4.2 Temperature dependent alloy mixing enthalpy ΔH_m data and quadratic fits for Monte Carlo simulations of (a) space group $I4_1/acd$ (solid) and (b) space group $P4_2/nmc$ (dashed) as functions of composition x. Energies are measured with respect to the respective ground states at the end-point compositions ($x = 0$ and $x = 1$). (c) Discrete MC data and exponential decay function fit for the quadratic bowing parameter Ω of both structures as a function of reciprocal temperature. 74

4.3 (a) Temperature-composition phase diagram for the $(Cd_{1-x}Zn_x)_3As_2$ heterostructural alloy. The miscibility regions for the two structures are colored red and blue, and the metastable region between them is colored white. Corresponding free energies of mixing ΔG_m at (b) $T = 300$ K and (c) $T = 370$ K are shown as a function of composition x. The black dashed line in (a) marks the minimum free energy crossover between the two structures, illustrated by black dots in (b) and (c). Green and orange dots represent points on the binodal lines as governed by common tangent construction. Diamonds indicate inflection points that determine the spinodal lines in (a), shown as solid (dashed) for compositions at which the respective structure is stable (metastable). 79

4.4 Free energies of mixing ΔG_{m} at (a) $T = 400$ K and (c) $T = 500$ K as a function of composition x. The black dots mark the minimum free energy crossover between the two structures. Green and orange dots represent points on the binodal lines as governed by common tangent construction. Diamonds indicate inflection points that determine the spinodal line in Fig. 4.3(a). 81

5.1 The electronic band structure of Cd_3As_2 with energies measured from the Dirac point energy E_{DP}, calculated with quasiparticle self-consistent GW theory (red) and GGA with different values of a nonlocal external potential V applied to all s-like states. All band structures were calculated for the same atomic structure that was relaxed with the SCAN functional. 85

Chapter 1

Introduction

1.1 Overview of topological materials

In typical semiconducting or insulating solid state systems, the valence band is derived from relatively lower energy p-like atomic orbitals, while the unoccupied conduction band is made up of higher energy s-like orbitals with a greater principal quantum number. The difference in energy between the two usually determines the conventional band gap. However, in some materials this typical band ordering can be reversed so the s-like states lie below the p-like states. By contrast to trivial band gap systems with the standard band ordering, there is a unique set of materials, known as topological insulators (TIs), with crystal symmetries that constrain the wavefunctions to enforce the inverted band behavior. They are just one example of a broader class of topological materials that manifest symmetry protected band crossings close to the Fermi level E_{F} [9, 22, 102].

Modern scientific interest in topological materials was ignited by the discovery of TIs such as HgTe, Bi_2Se_3, and Bi_2Te_3. The defining characteristic of TIs is that they are insulating everywhere in the bulk but have conducting, metallic states at their surface [48, 161]. It turns out that everywhere in the bulk BZ of these materials the electronic states have the inverted band behavior described above, and a strong spin-orbit interaction opens an insulating gap between the inverted p-states. Meanwhile, at the surface there is a (minimum of one) band crossing at the Fermi level that originates from the band inversion and gives rise to the conducting behavior at the edges. The crossing itself is found to be protected by the symmetries of the crystal, and it is considered topological in that there is a topological invariant associated with the existence of time reversal symmetry that fundamentally

differentiates the TI from a trivial insulator [63]. As a result, a TI cannot be adiabatically transformed into an ordinary insulator (e.g. by slowly tuning the energy of an inverted band) without passing through an intermediate conducting state.

In another related set of topological materials, crystal symmetries can give rise to band crossings in the bulk electronic states rather than at the surface. Conducting materials that possess these distinctive bulk band crossings are called topological semimetals. In three dimensions (3D) the points in momentum space where the conduction band and valence band are degenerate are classified as either Weyl points or Dirac points depending on if their dispersion relations correspond to the Weyl or Dirac Hamiltonians from particle physics. The primary subject of investigation in the current work is a Dirac semimetal, so I shall briefly summarize the formalism associated with topological semimetals as well as the unique properties they possess that motivate their dedicated study.

1.1.1 Weyl semimetals

As introduced in the previous section, the defining feature of 3D topological semimetals is the existence of symmetry protected crossing points in the bulk band structure where two or more bands are degenerate at the Fermi energy E_{F}. They serve as 3D analogues to graphene, which has a similar topological band crossing in 2D [22]. As with graphene, the band dispersion in the vicinity of the crossing is linear in crystal momentum, but in 3 rather than 2 dimensions. Specifically, for the first type of topological semimetal known as a Weyl semimetal , the band structure in the vicinity of the band crossing (assuming no anisotropies) is described by the Weyl Hamiltonian:

$$H_{\mathrm{Weyl}} = \pm v_{\mathrm{F}}\, \boldsymbol{\sigma} \cdot \boldsymbol{k}. \tag{1.1}$$

Here, v_{F} is the Fermi velocity ($\hbar = 1$) that determines the slope of the bands, $\boldsymbol{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$ are the Pauli matrices, the bands are measured from E_{F}, and the crossing (Weyl point) is chosen to be at momentum $\boldsymbol{k} = 0$. At any given Weyl point, the bands are two-fold degenerate. Several

3D materials, such as TaAs [92] and NbP [159], have been identified as Weyl semimetals and are currently being extensively studied both theoretically and experimentally.

A vital feature of the Weyl Hamiltonian is the nontrivial topology of its solutions. The eigenstates can be distinctly labeled by their chirality, either right-handed or left-handed based on if the projection of $\boldsymbol{\sigma}$ onto $\boldsymbol{k}$ is positive or negative, respectively ($+$ and $-$ in Eq. 1.1). These chiralities can be associated with a topological invariant, namely the flux of the so-called Berry curvature (See Appendix A),

$$\boldsymbol{\Omega}(\boldsymbol{k}) = \pm\frac{1}{2}\frac{\boldsymbol{k}}{k^3}, \tag{1.2}$$

through any surface in momentum space that encloses a Weyl point of that chirality [22]. It should be clear that any closed surface flux integral of $\boldsymbol{\Omega}(\boldsymbol{k})$ can be evaluated in the manner of Gauss's law and will give 2π times the chirality of the Weyl point. This gives rise to the (integer) Chern number, defined as

$$C = \frac{1}{2\pi}\int_{BZ}\boldsymbol{\Omega}(\boldsymbol{k})\cdot d\boldsymbol{S}, \tag{1.3}$$

i.e. $1/(2\pi)$ times the flux integral of $\boldsymbol{\Omega}(\boldsymbol{k})$ over the first BZ. The topologically invariant Chern number is either $+1$ or -1 for a single Weyl point [39].

An important feature of Eq. 1.2 is that it has the same form as the electric field of a point charge but in momentum space rather than real space. This leads to the commonplace expression that Weyl points can be thought of as sources and sinks of Berry curvature. Furthermore, the monopole behavior gives rise to an important prerequisite for the formation of Weyl points. The field lines of $\boldsymbol{\Omega}(\boldsymbol{k})$ must begin and end within the first BZ because of the necessary periodicity [9, 22], so Weyl points must necessarily form in pairs of opposite chirality, and the Chern number must thereby sum to zero for the full Weyl semimetal system [39]. Therefore, the only way to eliminate the topologically protected Weyl points is pairwise, when they exist at the same point in momentum space. This situation corresponds to the next class of 3D topological semimetal: the Dirac semimetal.

1.1.2 Dirac semimetals

When two Weyl points of opposite chirality exist at the same point in momentum space, and the symmetry of the crystal still ensures the crossing of the valence and conduction bands, the result is known as a Dirac semimetal. For such a material, the low-energy excitations near E_F can be described by the 4×4 Dirac Hamiltonian for a massless particle:

$$H_\mathrm{Dirac} = v_\mathrm{F} \begin{pmatrix} \boldsymbol{\sigma} \cdot \boldsymbol{k} & 0 \\ 0 & -\boldsymbol{\sigma} \cdot \boldsymbol{k} \end{pmatrix}, \tag{1.4}$$

where again the band crossing is chosen to be at $\boldsymbol{k} = 0$ and energy is measured from E_F. As implied in the above discussion and in the size of H_Dirac, at the Dirac point the bands (made up of two Weyl nodes) are four-fold degenerate. Materials with bulk band structures that have been found to contain 3D Dirac points at E_F include Na_3Bi [88] and the focus of the current work, Cd_3As_2 [87, 106].

The leading difference between a Dirac and Weyl semimetal phase is in the crystal symmetries of the material. It turns out that both time reversal symmetry and inversion symmetry are necessary for a stable Dirac crossing; breaking either (or both) of them lifts the degeneracy of the Dirac point and creates a pair of Weyl points with opposite chirality. For example, introducing a magnetic field through a tunable Zeeman term in Eq. 1.4 explicitly breaks time reversal symmetry and yields a Weyl semimetal, and the Dirac point is only recovered if the magnetic coupling is set to zero [67].

1.1.3 Properties of topological semimetals

There are many consequences of the topology and symmetry protected band crossings present in Dirac and Weyl semimetals, and they give rise to a number of unique properties that can be experimentally observed. One such feature is the appearance of distinct surface states. Owing to the point source and sink Berry curvature behavior of Weyl points, projections of the region between the two Weyl nodes onto the surface BZ results in edge-localized states called Fermi arcs [9, 22, 91, 154].

These Fermi arcs show up as line segments when mapping the Fermi surface in $\boldsymbol{k}$-space, as opposed to the typical closed loops of a topologically trivial material. Fermi arcs are also often present for Dirac semimetals because, even though the net flux of Berry curvature at a Dirac point is zero, their symmetry constraints can still connect different bands at the surface BZ to give arc-like structures.

Another distinctive feature of topological semimetals that has contributed to their ongoing research is the transport properties that arise from the so-called chiral anomaly. When external electric and magnetic fields are applied in the same direction in a topological semimetal there is a resulting disparity between the number of electrons with opposite chiralities [161]. The imbalance leads to several unique transport properties [112], the most striking of which is a large, negative longitudinal magnetoresistance [91, 112, 160, 161]. The chiral anomaly also gives rise to other universal transport phenomena such as the anomalous Hall effect in magnetic Weyl semimetals [22]. Other special properties that are unique to topological semimetals include ultrahigh electron mobilities [91, 132] and a unique optical conductivity that is linear in frequency [91].

These characteristic properties make topological semimetals very appealing quantum materials for applications to next-generation devices. The combination of their ultrahigh mobilities and low phonon mean free paths make them good candidates for low dissipation electronics, and 3D Dirac and Weyl semimetals can be epitaxially integrated into devices in a way that has proven difficult for graphene [85, 106, 168]. Additionally, topological semimetals have the potential to be utilized in infrared emitters, photodetectors, and optical switches [95, 157, 164, 170], and their capacity for preserving spin currents makes them viable for photovoltaics and spintronic devices [109, 158]. Lastly, topological semimetals are potentially tangentially applicable in quantum computing because they might host Majorana zero modes when paired with s-wave superconductors [82].

There are additional classes of topological semimetals and topological materials not discussed in detail here, such as nodal line semimetals, where the topologically protected band crossing is along a continuous line or loop in $\boldsymbol{k}$-space rather than at discrete points [38, 67]. Rather than go into exhaustive detail on all of the possible materials, I shall simply reiterate that the very interesting and novel physics associated with topological semimetals makes them worthwhile to study

experimentally, theoretically, and computationally. However, investigating these materials leads to an important question that is quite difficult to answer: what happens to these materials in the presence of disorder? When grown in a lab the material will not be a perfect crystal with universally defined symmetries. Instead, the lab-grown material will have crystallographic defects, impurities, and unintended dopants, and it is unclear precisely how these sources of disorder affect the electronic properties of the system and the topology of the band structure. The current work makes several steps toward answering this question for Cd_3As_2, a representative topological semimetal.

1.1.4 Cadmium Arsenide

Just over a decade has elapsed since angle-resolved photoemission spectroscopy experiments confirmed a 3D Dirac crossing in the band structure for Cd_3As_2 [87, 106]. The ground state crystal structure [4] has the inverted band behavior and necessary symmetries (time reversal and inversion) to support the four-fold degenerate linear band crossing as described above. Additionally, Cd_3As_2 has many of the desirable properties associated with topological semimetals described in the previous section. Fermi arc surface states, large linear magnetoresistance, ultrahigh mobilities, and a linear band dispersion have all been reported [83–85, 87, 98, 104, 106].

There are two primary reasons to study Cd_3As_2 with thermodynamic models based on first-principles calculations. First, it is advantageous to have a better understanding of how point defects and unintentional doping affect Cd_3As_2. At the outset of this work it was understood that when grown at or near thermodynamic equilibrium Cd_3As_2 was inherently n-type, with an excess number of electron donating defects (originally thought to be As vacancies) that elevate the Fermi level more than 100 meV above the Dirac point [59, 103, 107, 129]. The elevated E_F limits access to the characteristic properties and unique electron behaviors of Cd_3As_2 as a topological material [32, 103]. That is to say, the system is only really a Dirac semimetal if E_F is in close proximity to the Dirac point [21, 22]. Therefore, the hope is to computationally study the origin of this n-type growth, and use those insights to manipulate the formation of defects during growth and tune E_F closer to the Dirac point.

To further motivate this goal of Fermi level control, it is important to examine how reducing the electron carrier concentration and aligning E_F to the Dirac point enhances its useful properties as a quantum material. Recall that topological semimetals are generally sought after because of their ultrahigh electron mobilities and their capacity for a large linear or negative longitudinal magnetoresistance. It has been observed that samples of Cd_3As_2 with lower carrier concentrations have greater electron mobilities [36, 49]. This is typically attributed to the large Fermi velocity (band dispersion) nearer to the Dirac point and the band structure's protection against quasiparticle backscattering [3, 49]. Simultaneously, increases in electron mobility can give a stronger linear magnetoresistance [3], and indeed both thin-film and bulk samples of Cd_3As_2 are also found to have higher magnetoresistance when the carrier concentration is low [49, 104]. This idea is further supported by experimental observations that the magnetoresistance increases as temperature is decreased [36], because the electron carrier concentration is reduced at lower temperatures as it approaches a zero-temperature limit (see e.g. Fig. 2.8 in Ch. 2). Lastly, it has been proposed that a low carrier density is necessary to experimentally observe the chiral anomaly and a negative longitudinal magnetoresistance in Cd_3As_2 because the chiral anomaly is dependent on the nontrivial Berry curvature that only exists close to the Weyl nodes [84] (where the Dirac point has been split into two Weyl points by the applied magnetic field). In summary, manipulating point defects to reduce the carrier density and subsequently lower E_F to the Dirac point is expected to enhance the highly sought after transport properties associated with Cd_3As_2 being a Dirac semimetal.

The second reason to develop new thermodynamic models for studying Cd_3As_2 is to discover how the existing models for defects and alloys need to be updated for a topological semimetal. There is a well-established framework for using first-principles calculations to analyze the formation of point defects in insulators, semiconductors, and metals (see, e.g., Ref. [37]). However, it is not clear to what degree these tools can be applied to the fundamentally different system of a topological semimetal. For example, if E_F lies in either the conduction or valence bands in a traditional band gap system, the limited width of the Fermi-Dirac distribution at typical temperatures means only excitations of a single carrier need to be considered (electrons in the conduction band or holes in the

valence band). Whereas for a topological semimetal with E_F in the vicinity of the band crossing, simultaneous excitations of both electrons and holes may result in new and non-intuitive behavior. Hence, I make it a point in the current work to expand upon the existing first-principles based methodologies of defect formation and apply the updated theory to Cd_3As_2. To this end, I will first briefly summarize the first-principles computational methods that serve as the basis for constructing the thermodynamic models utilized in the rest of this work.

1.2 Density functional theory

One of the leading modern approaches to exploring the properties of condensed matter systems—both experimentally observed and theoretically predicted—is calculations from first-principles. These first-principles (or *ab initio*) calculations are based on the fundamental principles of quantum mechanics, do not rely on fitting empirical parameters to experiment, and have risen to such prominence because of their relative simplicity to implement combined with their general applicability. Specifically, I am referring to those methods that transform the difficult problem of calculating the ground state of a system of N interacting electrons into a more manageable calculation of single-particle or quasiparticle energies. Two of these many-body approaches, namely density functional theory (DFT) and GW theory, serve as the primary tools for determining the electronic structures and energetic inputs for the thermodynamic models in the current work. I shall therefore briefly summarize in this section the DFT approach, how it is most often implemented, and its challenges and limitations. Then in Sec. 1.3 I will give an overview of some of the corrections and modified theories beyond standard DFT (including GW theory) that address some of its shortcomings.

1.2.1 DFT formalism

The many-body Schrödinger equation for the electrons in a solid state system is given by

$$\hat{H}\left|\Psi\right\rangle = \left[\sum_{i}^{N}\left(-\frac{\hbar^2\nabla_i^2}{2m} + V_{\text{ext}}(\boldsymbol{r}_i)\right) + \frac{1}{2}\sum_{i\neq j}^{N} U(\boldsymbol{r}_i, \boldsymbol{r}_j)\right]\left|\Psi\right\rangle = E\left|\Psi\right\rangle, \tag{1.5}$$

where the first term is the kinetic energy, $V_{\text{ext}}(\boldsymbol{r})$ is the external potential from the (fixed) atomic nuclei, and $U(\boldsymbol{r}_i, \boldsymbol{r}_j)$ is the repulsive Coulomb interaction between the N electrons. The starting point for recasting this complex many-body problem into a set of single-particle equations is in the two theorems by Hohenberg and Kohn [53]. The theorems state first that the energy E in Eq. 1.5 is a unique functional of the electron density ρ, and second that said energy functional gives the correct ground state energy only if the input ρ is the true ground state density. The significance of the Hohenberg-Kohn theorems is that the fundamental variable for solving for the ground state energy can be chosen to be the single-variable, real-valued function $\rho(\boldsymbol{r})$ instead of the more complicated N-variable, complex-valued ground state $\left|\Psi_0\right\rangle$.

The complexity of the ground state wave function is not actually lost when the problem is redefined in terms of the electron density: it is instead relocated to the complicated dependence of E on ρ [39]. Kohn and Sham expressed the total energy functional as [65]

$$E[\rho] \equiv T_0[\rho] + E_{\text{ext}}[\rho] + E_{\text{H}}[\rho] + E_{\text{xc}}[\rho]. \tag{1.6}$$

Here, T_0 is the kinetic energy of a non-interacting electron system with the same density as the system being studied, E_{ext} comes from integrating $V_{\text{ext}}(\boldsymbol{r})\cdot\rho(\boldsymbol{r})$, E_{H} is the Hartree contribution to the electron-electron Coulomb energy, and the unknown exchange-correlation energy functional E_{xc} is simply defined to be the remaining contribution that comes from the electron-electron interactions.

Performing a functional derivative of Eq. 1.6 with respect to ρ subject to the constraint of a

fixed total number of electrons [61, 65] gives the famous Kohn-Sham equations:

$$[T + V_{\text{ext}}(\boldsymbol{r}) + V_{\text{H}}(\boldsymbol{r}) + V_{\text{xc}}(\boldsymbol{r})]\, \varphi_{n\boldsymbol{k}}(\boldsymbol{r}) = \varepsilon_{n\boldsymbol{k}} \varphi_{n\boldsymbol{k}}(\boldsymbol{r}). \tag{1.7}$$

In these equations, $\varphi_{n\boldsymbol{k}}$ are non-interacting single-particle orbitals with band index n that feel an effective potential $V_{\text{ext}} + V_{\text{H}} + V_{\text{xc}}$, where V_{xc} is the functional derivative of $E_{\text{xc}}[\rho]$. The orbitals can be labeled by BZ momentum $\boldsymbol{k}$ if periodic boundary conditions are applied, because then the charge density and effective potential have the periodicity of the crystal lattice and Bloch's theorem can be applied. The DFT implementation for finding the ground state energy in a periodic system then amounts to choosing a suitable approximation for $E_{\text{xc}}[\rho]$; solving the set of eigenvalue problems in Eq. 1.7 for the orbitals $\varphi_{n\boldsymbol{k}}$, corresponding eigenvalues $\varepsilon_{n\boldsymbol{k}}$, and the electron density given by

$$\rho(\boldsymbol{r}) = \sum_{\text{occ.}} |\varphi_{n\boldsymbol{k}}(\boldsymbol{r})|^2\,; \tag{1.8}$$

and finally evaluating Eq. 1.6 to obtain the total energy.

There are a few key aspects of the approach just described that need to be addressed. First, there is an implied requirement of self-consistency in Eqs. 1.7 and 1.8. The effective potential in Eq. 1.7 is determined for a given density, so it is essential that $\rho(\boldsymbol{r})$ must be found such that the potential it generates yields orbitals that reproduce it via Eq. 1.8 [141]. This is accomplished in a straightforward manner by, e.g., iteratively alternating between solving Eq. 1.7 and evaluating Eq. 1.8. Second, because Bloch's theorem is employed and the $\boldsymbol{k}$-dependence is integrated out in Eq. 1.8, the Kohn-Sham equations can be solved independently on a chosen grid of irreducible points in the first BZ. However, a sufficiently dense number of $\boldsymbol{k}$-points must be chosen to sample enough of the BZ and ensure good convergence. Third, it is important to note the eigenvalues $\varepsilon_{n\boldsymbol{k}}$ in the Kohn-Sham equations are Lagrange multipliers introduced when enforcing the constraint of fixed electron count when minimizing the energy functional [39, 61], so they have a limited physical interpretation. Nevertheless, the eigenvalues can be used to construct a band structure that can

(with important caveats discussed below) predict many electronic properties of the solids being studied.

Different implementations of DFT can be classified by how they choose to approximate the exchange-correlation potential and further subdivided by how they go about solving the Kohn-Sham equations. For approximating E_{xc}, the most rudimentary and straightforward choice is known as the local density approximation (LDA), in which it is chosen to equal the exchange-correlation energy for a uniform, interacting electron gas (jellium). The electron gas E_{xc} is considered "known" in the sense that there is an exact analytic expression for the exchange energy, and the contribution from correlations has been numerically modeled to high accuracy by Monte Carlo methods [23, 40, 119]. The other frequently used approximation to E_{xc} is known as the generalized gradient approximation (GGA) and is estimated by considering an electron gas with a slowly varying density [117]. The GGA functional depends on, in addition to the value of the density, the magnitude of its gradient. It is generally accepted that the additional information contained in the gradient of ρ leads to improved descriptions of the electrons in many solids [117].

As for the different methods of solving the Kohn-Sham equations, they are often identified by their choice of basis set. That is, the Kohn-Sham orbitals $\varphi_{n\boldsymbol{k}}(\boldsymbol{r})$ can be expressed as linear combinations of a chosen basis set of orthonormal functions so that solving Eq. 1.7 amounts to finding the expansion coefficients, and different DFT implementations have different choices for those orthonormal functions. Common choices of basis sets include plane waves, combinations of atomic orbitals, linearized muffin-tin orbitals, and linearized augmented plane waves. Broad overviews of these different basis sets are given in, e.g., Refs. [5, 141].

1.2.2 Applications beyond zero temperature - thermodynamic modeling

There was a historical misconception associated with DFT that because it is used to find the ground state energy of a system, it is not capable of giving insights above zero temperature. However, many finite-temperature effects and properties can be successfully modeled with DFT calculations, as long as the electronic ground state is well described. In general, the DFT formalism

of the previous section can be generalized to finite temperatures by considering, rather than the internal energy, a thermodynamic free energy that includes contributions from different sources of entropy. For example, the formation energy of a point defect in a solid depends on the atomic chemical potentials (see Ch. 2), which include contributions from ideal gas entropy as a function of temperature and partial pressure [74]. Other entropic contributions to the thermodynamic free energy include configurational entropy (see e.g. Sec. 4.4 and Appendix B), electronic entropy [73, 96], and vibrational entropy [12, 37].

Modeling a solid state system at finite temperatures then amounts to determining under what conditions the overall free energy (with relevant contributions from entropy included) is minimized and assessing the resulting physical implications. This gives rise to many applications and thermodynamic models that successfully describe physical phenomena at elevated temperatures, even though the individual first-principles calculations that provide the input data for these models are performed at zero temperature. Examples applicable to the current work include determining the equilibrium concentrations of point defects under certain growth conditions [150, 169] (see Chs. 2 and 3), and Monte Carlo simulations either via the cluster expansion approach [100, 128, 148] or through explicit energy sampling (see Ch. 4).

1.2.3 Shortcomings

The most frequently cited limitation of DFT is the band gap problem for electronic structures it generates; it typically underestimates the band gaps for semiconductors and insulators by as much as 40% [118]. This band gap error is rooted in the construction of the auxiliary single-particle Hamiltonian in Eq. 1.7. While the Lagrange multiplier eigenvalues of this effective Hamiltonian can give meaningful insights into the electronic structure, they cannot be interpreted as true quasiparticle excitation energies. Furthermore, it is important to emphasize that this error is well understood to be an inherent feature of the single-particle energies for a local density functional (independent of any approximations) [40, 41, 45, 46, 118, 136], and beyond-DFT methods (see Sec. 1.3) that incorporate nonlocal exchange effects are necessary to correct it.

Despite the typical emphasis on the DFT band gap error as the most common problem associated with the approach, it is worth noting that there are other practical limitations of the theory. For example, in a true interacting electron system, Pauli exclusion leads to a spatial separation of same-spin electrons and an overall reduction in the electron-electron Coulomb repulsion energy. In Hartree-Fock theory this manifests as a negative exchange energy from considering explicit pairwise swaps between single particle states. As for DFT, it should be included in the exchange-correlation functional E_{xc}. The Hartree term in the Kohn-Sham equations includes an unphysical self-interaction energy (the electron feeling Coulomb repulsion from itself) that should be canceled by the exchange-correlation potential. However, the approximations of this potential in LDA or GGA cannot capture the true (nonlocal) exchange behavior. The remaining uncanceled Hartree energy leads to a spurious delocalization of electron states, called the delocalization error or self-interaction error [99, 119]. Standard DFT implementations like LDA and GGA are also known to underestimate the heat of formation (DFT value is less negative) of semiconducting and insulating compounds, predicting them to be less stable than they actually are [57, 71]. Moreover, standard GGA functionals like that of Ref. [120] can underestimate the strength of atomic bonds and result in lattice constants that are 1-2% too large [47, 69].

Beyond just quantitative errors for calculated band gaps, the single-particle energies for the auxiliary Hamiltonian in standard DFT can also give rise to incorrect qualitative descriptions of band structures. For example, there is a class of materials, called Mott insulators, that are verified experimentally to be insulating but typically predicted by DFT to be metallic. This false characterization of the nature of the band structure is associated with strongly correlated systems with localized, partially filled d- and f-states [7, 27]. In these systems, DFT fails to capture the strong correlations, is unable to properly describe these localized states, and incorrectly produces overlapping valence and conduction bands. This has significant consequences even beyond the electronic structure. For systems with only a quantitative band gap error the ground state energy is expected to still be close to the true value, but if the wrong electronic states are occupied/unoccupied the electron and spin densities and even the predicted total energy can be remarkably inaccurate [7].

Other examples of a qualitatively incorrect band structure relevant to the current work are DFT sometimes incorrectly classifying topological materials as trivial [2, 89, 152].

Regardless of the limitations of DFT summarized here, I would like to emphasize that they by no means undercut its usefulness. The theory is computationally efficinent, broadly applicable, and undoubtedly useful in studying the electronic properties of solid state systems. With that in mind, I will now discuss some of the refined theories that address these shortcomings and further improve DFT's predictive power.

1.3 Corrections and modifications

At present I shall briefly mention some of the modern modifications to standard DFT that are used to address the weaknesses discussed in the previous section, with a particular emphasis on those relevant to the work contained in this thesis. One of the most common expanded DFT methods that addresses the self-interaction error is the use of so-called hybrid functionals [15, 51]. These functionals introduce semi-empirical parameters that mix the approximate exchange energy from DFT with the explicit nonlocal exchange energy from Hartree-Fock (HF) theory. The restraints of the DFT approximation lead to delocalized states and too small band gaps, whereas HF does not include electron correlations, site exchanges between distant electron states are not effectively screened, and band gaps end up overestimated. Therefore, a fractional mixing of the two (typically around 20-25% HF, 75-80% DFT) is understood to give a much improved approximation of the true exchange energy.

There are several other modified density functional theories that serve to produce updated band structures consistent with experiment or otherwise correct the delocalization error. The DFT+U approach includes an intra-atomic Hubbard U repulsion term in the potential for strongly correlated d- or f-electrons [34, 165]. The empirical parameter U can be varied to give band gaps that agree with experiment, but the drawback is the method can no longer be considered predictive from first-principles. Another semi-empirical method employed when considering phase stability and the formation of point defects utilizes fitted elemental-phase reference energies (FERE) to give

much improved heats of formation [71, 144, 156]. Within the FERE approach, energies of elemental reference phases that form binary or ternary compounds are not calculated explicitly with the chosen LDA or GGA functional, but rather are determined by an overdetermined fit based on experimental heats of formation and corresponding DFT energies with good error cancellation [71]. Other modified theories include explicitly including correlation effects beyond standard DFT [24], utilizing the random phase approximation (RPA) to give improved total energies when DFT predicts the wrong ground state structure [114], and intentionally increasing the d-like character of poorly described p-like states to correctly achieve an experimentally observed topological insulating gap [151].

There are many applications in which one or several of the above methods are valuable in improving the effectiveness of DFT calculations, and each of them comes with advantages and disadvantages. Rather than belabor the usefulness of each of them, I will instead concentrate on two specific approaches that were central to the rest of this thesis: the strongly constrained and appropriately normed meta-GGA functional and GW theory.

1.3.1 Strongly constrained and appropriately normed meta-GGA functional

The defining feature of meta-Generalized Gradient Approximation (meta-GGA) DFT functionals is that the approximation to the exchange-correlation potential depends on, in addition to the density ρ and the magnitude of its gradient $|\nabla\rho|$, either the Laplacian of the density $\nabla^2\rho$, the kinetic energy density of the occupied single-particle orbitals $\sum |\nabla\varphi_{n\boldsymbol{k}}|^2/2$, or both [146]. It is understood that these higher order terms in the approximation of V_{xc} can still not fully capture the true nonlocal behavior of electron exchange and correlation, but the additional information they contain can notably improve the approximation. In fact, the meta-GGA exchange-correlation energy can be implemented in such a way that what enters the Kohn-Sham equations is a differential operator rather than an explicitly local multiplicative potential as in standard DFT [105, 146]. It turns out that this refined potential can partially or even nearly correct the systematic errors of standard DFT functionals as described in Sec. 1.2.3 above.

The specific meta-GGA functional employed in this thesis is the strongly constrained and

appropriately normed (SCAN) functional of Ref. [147]. Interest in and utilization of the SCAN functional have grown in recent years because many general features of DFT solutions in solids are improved when using it. For example, SCAN does not have the 1-2% overestimated lattice constants of standard GGA functionals [146], and it no longer systematically underestimates the magnitude of heats of formation and can thereby more accurately predict stability conditions [13, 57]. I will note that the FERE approach described above can also be implemented for the SCAN functional, but it provides less of an improvement than for typical GGA functionals because SCAN does not have pronounced systematic errors in the first place. Finally, SCAN is known to give slight improvements in calculated band gaps over standard GGA functionals [57]. These quantitative improvements make SCAN a promising choice for general solid-state calculations, albeit with the computational cost increased by a factor of 2 to 5.

The reasons to use the SCAN functional when studying Cd_3As_2 are outlined in Ch. 2. In short, in addition to the general improvements just described, SCAN was better able to reproduce the quasiparticle band energies of GW calculations than GGA. It turns out that for a Dirac semimetal, a proper description of point defects requires as precise a description of the band energies near the Dirac point as possible. With that in mind, it is important to provide an overview of GW theory, as it produces band structures that often align most accurately with experiment.

1.3.2 GW theory

Recall that the fundamental origin of the band gap error in DFT is in the interpretation of the single-particle energies in the Kohn-Sham equations (Eq. 1.7) because the effective Hamiltonian is artificially constructed with a local exchange correlation potential. To truly capture the interacting behavior of electrons, it is necessary that they be described in a manner that is explicitly nonlocal. One particularly effective and powerful approach that accounts for said nonlocality and provides more accurate electronic structures than DFT is GW theory. Note that GW is not explicitly a correction to DFT, but rather a distinct formulation for quasiparticle excitations in a solid-state system derived from many-body perturbation theory. However, it is most often implemented by

using the Kohn-Sham single-particle orbitals and eigenvalues as a starting point as will be seen presently.

In many-body perturbation theory, collective electron and hole quasiparticle excitations are described by eigenvalue equations of the form [56]

$$[T + V_{\text{ext}}(\boldsymbol{r}) + V_{\text{H}}(\boldsymbol{r})]\,\varphi_{n\boldsymbol{k}}(\boldsymbol{r}) + \int d\boldsymbol{r}'\;\Sigma(\boldsymbol{r},\boldsymbol{r}',\omega=\varepsilon_{n\boldsymbol{k}})\varphi_{n\boldsymbol{k}}(\boldsymbol{r}') = \varepsilon_{n\boldsymbol{k}}\varphi_{n\boldsymbol{k}}(\boldsymbol{r}), \tag{1.9}$$

where the terms in the effective potential are the same as in DFT but with the exchange-correlation potential replaced by the many-body self-energy $\Sigma(\boldsymbol{r},\boldsymbol{r}',\omega)$. In addition to being explicitly nonlocal (depending on both $\boldsymbol{r}$ and $\boldsymbol{r}'$), the self-energy is also dynamic (depends on frequency ω), and is in principle determined by way of Hedin's Equations [50]. Within the GW approach, Σ is approximated to lowest order in terms of the Green's function G of the effective Hamiltonian and the screened Coulomb potential W (hence the name). The screened Coulomb potential can be evaluated with the random phase approximation by using the irreducible polarizability χ to construct the inverse of the dielectric function ϵ^{-1}. Technical details for determining Σ can be found in, e.g., [50, 56, 138]).

Solving Eq. 1.9 requires having an initial Ansatz for the quasiparticle orbitals $\varphi_{n\boldsymbol{k}}(\boldsymbol{r})$ and eigenvalues $\varepsilon_{n\boldsymbol{k}}$ because they themselves are necessary to construct the self-energy. This is where an initial DFT calculation becomes relevant. The Kohn-Sham orbitals and energies from a standard implementation of DFT are used to create an initial Σ, and then Eq. 1.9 can be solved or updated to varying degrees of self-consistency. For example, single shot GW uses the zeroth order self-energy from DFT and only one iteration to find new eigenvalues while keeping the orbitals unchanged [56], partially self-consistent (or eigenvalue self-consistent) GW uses the calculated eigenvalues at each iteration to update the poles of G and re-evaluate the self-energy operator but still maintains the DFT orbitals, and fully self-consistent GW additionally updates both the effective Hamiltonian and the orbitals until neither change with more iterations.

As it happens, fully self-consistent GW tends to give worse predictions than even single-shot GW [44]. The truncation in the approximation of the self-energy in the self-consistent approach

leads to an underestimated dielectric screening and overestimated band gaps, while the less rigorous approaches benefit from fortuitous error canceling. However, without self-consistency the GW result is often sensitive to the DFT starting point, which is also detrimental. A refined implementation of self-consistent GW, known as quasiparticle self-consistent GW theory (QSGW), modifies the self-energy in Eq. 1.9 with an optimized perturbation that removes much of the systematic errors while still allowing for overall self-consistency [35, 130]. The current work includes both an eigenvalue self-consistent GW calculation based on DFT results using the SCAN functional and QSGW electronic structures for Cd_3As_2, with the latter provided by collaborator Mark van Schilfgaarde.

I will conclude this section with an overview of why GW corrections on top of DFT calculations are so advantageous. It is important to note that, because the self-energy is both nonlocal and requires many unoccupied orbitals when constructing the frequency-dependent dielectric screening, GW calculations are significantly more computationally expensive than standard DFT. The increase in computation time is customarily justifiable though, for the predicted band structures are greatly improved over the alternatives. The inclusion of screening through the dielectric response function makes GW preferable to HF theory, which only considers bare (unscreened) exchange and does not consider electron correlations. Simultaneously, the nonlocal treatment in GW generally mitigates the errors inherent to DFT and gives far better predictions of band gaps [40, 130, 139]. On top of that, because point defect formation energies depend on both the Fermi level and the energies of the bands on an absolute scale (see e.g. Ch. 2), GW band gap corrections also lead to better predictions of defect concentrations [115]. It is for this reason that in this thesis I combine total energies from DFT using the SCAN functional with band corrections from GW to give the best possible description of defects and disorder in Cd_3As_2.

1.4 Overview of this thesis

The current work introduces and applies several thermodynamic models that serve to identify the sources and subsequent ramifications of disorder in a topological semimetal. Through the study of the Dirac semimetal Cd_3As_2 I address many unique questions that differentiate it from more

traditional band gap systems: How does the well-known DFT band gap error manifest in predicted electronic structures involving topological band crossings? When excess charge carriers from point defects occupy the highly dispersive bands near the crossing point, how is the energy of forming those defects impacted? To what extent can one use variable growth conditions, extrinsic doping, and alloying to exert control over the equilibrium Fermi level? At the same time, I revisit and expand on conventional thermodynamic approximations to better describe intermediate temperatures between the zero- and infinite-temperature limits. The remainder of this thesis covers the motivation behind these updated theories, their mathematical and computational implementation, and the resulting predictions when they are applied to Cd_3As_2.

Chapter 2 focuses on adapting the existing methods of simulating point defects in band gap systems so they can be applied to a topological semimetal. It identifies in Cd_3As_2 a strong dependence of defect formation energies on the band energies in the vicinity of the Dirac point. The effect gives rise to a self-consistent equilibrium solution between the defect concentrations, free carriers, and the Fermi level. Modeling the overall effect with a rigid band model requires a very accurate numerical representation of the band structure, which is obtained with QSGW. Lastly, I use the defect equilibrium model to study the origin of the innate elevated Fermi level in Cd_3As_2 and recommend growth conditions that might tune E_F closer to the Dirac point. The predicted values of E_F for typical growth conditions are consistent with the cited experimental values between 80 and 200 meV, and similarly for the carrier concentrations (cf. Fig. 2.7 and Refs. [59, 103, 104, 107, 129]). Our insights into the formation of dominant Cd defects have also helped develop and explain strategies for reducing the carrier carrier concentrations and increasing the magnetoresistance in thin-film grown Cd_3As_2 [104, 123].

Subsequently, chapter 3 gives further insights into defect formation in Cd_3As_2 with the added consideration of using extrinsic dopants to gain even greater control over E_F. Additionally, in this chapter I develop one of the expanded thermodynamic theories that fills the gap between the low- and high-temperature limits. Specifically, I introduce a computational approach for analyzing the temperature dependent redistribution of point defects on a fixed lattice with the stoichiometry of

each element maintained, which corresponds to a sample that is isolated after being grown so it can not exchange atoms with an external reservoir. This "pseudo-equilibrium" approach is motivated by the persistence of mobile defects at lower temperatures well after an initial growth. That is, if there is a lack of significant energy barriers to atomic diffusion, defect atoms can migrate through the material such that individual defect concentrations can drastically change [11, 58, 70, 93, 134], and the pseudo-equilibrium aims to describe that behavior. Within this model I determine which Group I and Group IV dopants have the best potential to overcome the intrinsic self-doping and achieve net-neutral Cd_3As_2 at lower growth temperatures.

Finally, Ch. 4 moves beyond a dilute point defect model and instead considers alloying Cd_3As_2 with Zn. The alloy is predicted to simultaneously undergo both a topological transition from Dirac semimetal to trivial semiconductor and a doping transition from n-type to p-type. In the process of determining the solubility limits and relative stability of the $(Cd_{1-x}Zn_x)_3As_2$ heterostructural alloy, I construct another updated thermodynamic model that more accurately describes solid-state systems at moderate temperatures. Particularly, I use Monte Carlo sampling of the alloy to go beyond the regular solution approximation by way of a temperature dependent enthalpy of mixing that captures short-range order effects. This modification is useful for for an accurate description of alloys with strong short-range order because it can result in a significant reduction in the alloy enthalpies compared to random mixing [25, 60]. I subsequently perform a thermodynamic integration to construct the alloy phase diagram, which exhibits novel features unique to heterostructural alloys. The Monte Carlo equilibrated atomic structures then serve as a basis for studying variations in the electronic structure of the alloy.

Chapter 2

Band energy dependence of defect formation and self-doping in a topological semimetal

This chapter is adapted from Ref. [19] and its Supplemental Material: Copyright 2023 by the American Physical Society; and from Sec. III of Ref. [3], reproduced with the permission of AIP Publishing.

Abstract. Cadmium arsenide (Cd_3As_2) is a prototypical Dirac semimetal that manifests topological properties in a three-dimensional (3D) bulk material. In defect-free Cd_3As_2, the Fermi level E_F lies at a minimum in the density of states at the Dirac point, but experimentally it forms with excess electron carriers and an elevated E_F, thereby masking the topological features. To computationally study the self-doping of Cd_3As_2, we combine density functional theory (DFT) calculations for defect formation energies with quasiparticle self-consistent GW (QSGW) electronic structure calculations. We demonstrate an innate dependence of the point defect formation energies on carrier concentrations and use the QSGW calculated density of states to extrapolate formation energies to arbitrary electron concentrations. This approach allows the quantitative modeling of thermodynamic defect equilibria in topological semimetals and is used to predict how control of growth conditions might be utilized to achieve doping-neutral Cd_3As_2.

2.1 Introduction

Topological semimetals are an intriguing class of novel materials with a significant and sustained scientific interest. They are represented by several different compounds and exhibit a wide variety

of transport and electronic structure phenomena as a consequence of their inherent symmetries, including ultrahigh electron mobility, large linear magnetoresistance, gapless bulk states, and linear band dispersion [55, 85, 101, 132, 140]. However, in real topological semimetal crystals, defects and impurities can cause doping and disorder, thereby constraining the accessibility of these topological features for application in next-generation devices.

Cd_3As_2 is a prototypical three-dimensional Dirac semimetal which is stable in ambient conditions unlike many of its alternatives [85, 87]. The ground state is a centrosymmetric crystal structure derived from an underlying antifluorite lattice with an ordered sublattice of empty cation sites (space group 142, I41/acd) [4]. Defect-free Cd_3As_2 has a Fermi level that lies at a fourfold degenerate Dirac point along Γ-Z in the Brillouin zone (BZ) [4, 29]. Unfortunately, Cd_3As_2 crystal growth tends to induce unintentional n-type doping, so that the Fermi level E_F typically ends up about 100–200 meV above the Dirac point [28, 59, 104, 129].

We investigate computationally the role of intrinsic point defects in the self-doping behavior and examine how growth conditions of Cd_3As_2 might be manipulated to tune E_F closer to the Dirac point and more readily access its topological features. Furthermore, our goal is to determine what computational methods beyond standard supercell approaches are necessary for an accurate representation of defect formation energies in a topological semimetal. Defect equilibrium calculations based on first-principles formation energies have a long history in semiconductors and insulators [169], where they are applied to increasingly complex systems with disorder [110] and metastability [43]. Recent interest in such simulations for topological semimetals like TaAs [20] and Cd_3As_2 in the present work motivates revisiting the issue of finite-size effects in supercell calculations in view of the absence of a band gap and the resulting differences in the screening behavior. We show that the dominant intrinsic point defects in Cd_3As_2 (Cd interstitials and vacancies) can be modeled as fully ionized species in charge balance with electrons and hole carriers, and we present predictions for Fermi level control during synthesis.

2.2 Methods

For total energy calculations, we use density functional theory (DFT) methods and perform supercell calculations of the intrinsic point defects, using the projector augmented wave (PAW) method as implemented in the VASP code [17, 69], and utilizing both the generalized gradient approximation (GGA) of Ref. [120] and the strongly constrained and appropriately normed (SCAN) meta-GGA functional of Ref. [147]. Compared to standard DFT functionals, SCAN generally improves the prediction of lattice constants and avoids the systematic errors in the elemental reference energies [13]. To obtain an accurate electronic structure with high resolution BZ sampling, we perform quasiparticle self-consistent GW (QSGW) calculations with the Questaal code [68, 113, 130]. To obtain the alignment of the GW band energies relative to SCAN, we also performed eigenvalue self-consistent GW calculations [72] in VASP [138]. The final results include spin-orbit coupling in both total-energy and electronic structure calculations.

For all total energy calculations we used a plane wave basis set with a cutoff energy of 380 eV and version 5.4 of the PAW-PBE pseudopotential data sets. The atomic positions of defect supercells were relaxed with an atomic force convergence requirement of 0.01 eV/Å. For studying the supercell size dependence of the defect energies in Fig. 2.1(a), we used the Generalized Gradient Approximation (GGA) of Ref. [120] without spin-orbit coupling. The 80-, 160-, and 320-atom supercells were calculated using 6×6×6, 4×4×4, and 2×2×2 Γ-centered k-meshes respectively. The large band dispersions in Cd_3As_2, especially in the unoccupied states, required the use of these relatively dense k-point samples. Even so, the size dependence of the uncorrected energies (red) in Fig. 1(a) of the main text still shows a noticeably jagged behavior which is attributed to incomplete k-point convergence. However, after subtraction of the band-filling energies or in case of the charged Cd_i defect, these effects disappear because the contribution from the dispersive conduction band states is eliminated.

We further relaxed the atomic structure of the 80-atom primitive cell with the SCAN meta-GGA functional of Ref. [147], including spin-orbit coupling and a 4×4×4 Γ-centered k-mesh. The

relaxation gave lattice constants $a = 6.31$ Å and $c = 12.77$ Å. This structure serves as the basis for subsequent defect and electronic structure calculations. All of the defect total energies were calculated using the 80 atom primitive cell in SCAN with spin-orbit coupling on a 4×4×4 k-mesh. The potential alignment contribution for the charged defects [80] was included. However, the respective image charge corrections were excluded, because they are expected to vanish in the long-distance limit, as discussed below.

2.3 Finite-size effects of defect formation energies

First-principles calculations of point defects are most frequently performed using supercells under periodic boundary conditions. However, this approach can be subject to spurious interactions between the defect and its periodic images. The corrections necessary to extrapolate the formation energies to the dilute limit have been thoroughly investigated for band gap systems [37, 79]. To investigate how these methods can be adopted to a Dirac semimetal like Cd_3As_2, and to define an approach for evaluating the defect formation energies in a defect equilibrium with varying defect and carrier concentrations, we first calculate the supercell size dependence for both electron and hole donating intrinsic defects.

In the supercell approach to defects in a solid, the formation energy of a defect D in a charge state q is given by

$$\Delta H_{\mathrm{D},q}(E_{\mathrm{F}}, \{\mu\}) = [E_{\mathrm{D},q} - E_{\mathrm{h}}] + \sum_{\alpha} n_{\alpha}\mu_{\alpha} + qE_{\mathrm{F}}, \tag{2.1}$$

where $E_{\mathrm{D},q}$ and E_{h} are the total energies of the defect and host supercells, respectively, μ_{α} are the chemical potentials of the atom species α that are removed ($n_{\alpha} = +1$) or added ($n_{\alpha} = -1$) to the host crystal in forming the defect, and E_{F} is the Fermi energy. Here, E_{F} corresponds to the thermodynamic electron chemical potential, e.g., in a defect equilibrium, which is generally different from the Fermi level in the supercell calculation of a specific defect. The Cd_3As_2 primitive cell has 80 atoms with 3 nonequivalent sites for each Cd and As and one empty cation site which we use as interstitial site.

Figure 2.1(a) shows the dependence of ΔH_{D} of the charge-neutral ($q = 0$) $\mathrm{Cd_i}$ interstitial defect on the supercell size between 80 and 320 atoms (GGA level calculation without spin-orbit coupling). Figure 2.2 shows the calculated formation energies for these same supercell sizes for the V_{Cd}. We observe strong finite-size effects for both the donor-type $\mathrm{Cd_i}$ (Fig. 2.1(a)) and the acceptor-type V_{Cd} (Fig. 2.2) defects. This effect is expected because these defects alter the stoichiometry and electron count, but do not introduce localized defect states that would bind the two excess electrons ($\mathrm{Cd_i}$) or holes (V_{Cd}). In this situation, the electrons (holes) introduced by the defect fill the band continuum of the host material and raise (lower) the Fermi energy, in analogy to "shallow" dopants in semiconductors [121]. In contrast to the Cd interstitials and vacancies, the As vacancy V_{As} acts analogous to "deep" dopants in the sense that it introduces a localized defect states close to the Dirac point [104]. The implications of this different behavior for the absolute formation energies will be discussed below.

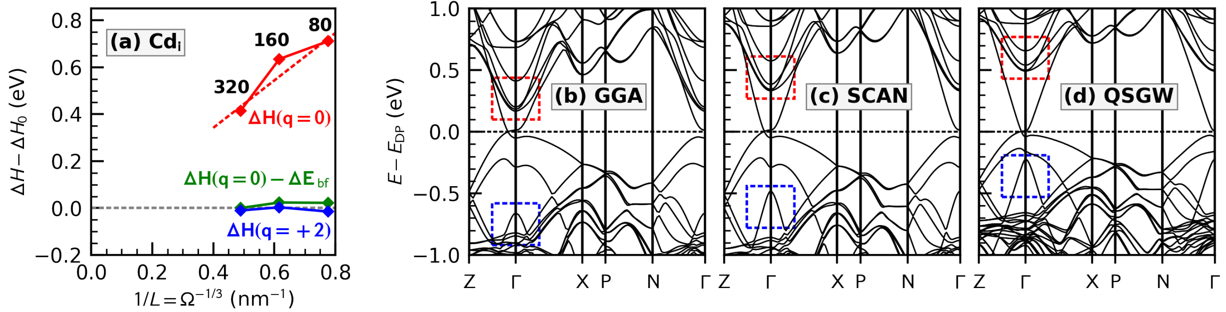


Figure 2.1: (a) The supercell size dependence of the formation energy ΔH of the $\mathrm{Cd_i}$ defect in $\mathrm{Cd_3As_2}$ as a function of inverse linear supercell size $1/L = \Omega^{-1/3}$ obtained from supercell volume Ω. ΔH_0 corresponds to the dilute limit of the formation energy. (b-d) The electronic band structure of $\mathrm{Cd_3As_2}$ with energies measured from the Dirac point energy E_{DP}, calculated with GGA (b), SCAN (c), and quasiparticle self-consistent GW theory (d). All band structures were calculated for the same atomic structure that was relaxed with the SCAN functional.

In case of the Cd defects, the concentration dependent change of E_{F} causes corresponding "band-filling" energies [121]

$$\Delta E_{\mathrm{bf}} = \sum_{\substack{n,\mathbf{k} \\ \varepsilon_{n,\mathbf{k}} > E_{\mathrm{DP}}}} w_{\mathbf{k}} f_{n,\mathbf{k}} (\varepsilon_{n,\mathbf{k}} - E_{\mathrm{DP}}) \tag{2.2}$$

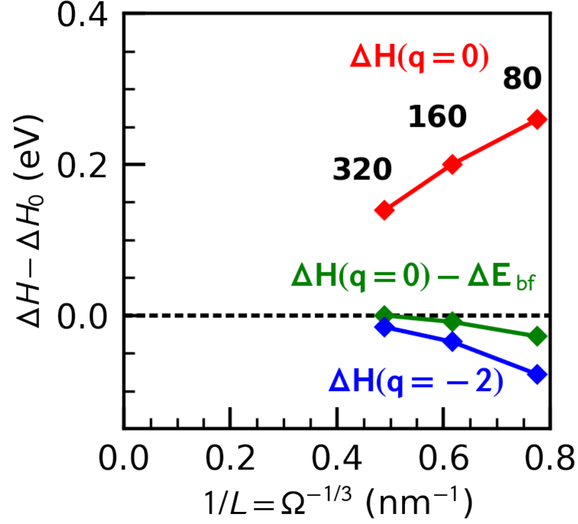


Figure 2.2: The supercell size-dependence of the formation energy ΔH of the V_{Cd} defect in $\mathrm{Cd_3As_2}$ as a function of inverse linear supercell size $1/L = \Omega^{-1/3}$ obtained from supercell volume Ω. ΔH_0 corresponds to the dilute limit of the formation energy. Labels indicate the number of atoms in the respective host supercells.

for electron donating defects ($\mathrm{Cd_i}$). Using the occupancy $f_{n,\mathbf{k}}$ of each band n at k-point $\mathbf{k}$ with BZ weight $w_{\mathbf{k}}$, the sum in Eq. (2.2) and in the analogous equation for acceptors integrates the contributions of the band energies $\varepsilon_{n,\mathbf{k}}$ of occupied states above (donors) or unoccupied states below (acceptors) the Dirac point energy E_{DP}, which plays the role of the reference value for the Fermi level of the undoped material. Here, the potential alignment technique of Ref. [80] is used to obtain consistent energy scales between defect and host cells. The analogous expression to Eq. 2.2 for electron accepting defects (V_{Cd}) is

$$\Delta E_{\mathrm{bf,h}} = \sum_{\substack{n,\mathbf{k} \\ \varepsilon_{n,\mathbf{k}} \leq E_{\mathrm{DP}}}} w_{\mathbf{k}}(1 - f_{n,\mathbf{k}})(E_{\mathrm{DP}} - \varepsilon_{n,\mathbf{k}}), \tag{2.3}$$

where instead we sum over states occupied by holes below the reference level E_{DP}.

Subtracting the band-filling energies ΔE_{bf} from ΔH_{D}, we obtain formation energies that are in good approximation cell-size independent (Figs. 2.1(a) and 2.2). The absolute values of the finite-size energies for the V_{Cd} are less than half of those for the $\mathrm{Cd_i}$, reflecting the higher density of

states of the occupied (valence-band like) states in Cd_3As_2. Also, the uncorrected energies in Fig. 2.2 are less susceptible to k-point convergences than in case of Cd_i.

In semiconductors or insulators, electron donating or accepting defects can form in a charged defect state when they exchange electrons with each other or with the band continuum of the host material. A similar mechanism is expected here as well, but there are important differences arising from the fact that Cd_3As_2 does not have a band gap. While the detailed screening behavior of Dirac semimetals is still subject to active research [26], we expect that defect charges are fully screened in the long-distance limit. As a consequence, there is no bound state analogous to the hydrogenic level in effective-mass theory [155]: the neutral Cd_i^0 essentially constitutes a charged Cd_i^{+2} defect with 2 electrons in the conduction-band like states above E_{DP}. To demonstrate this equivalence, Figs. 2.1(a) and 2.2 also include the size dependence for the charged defects, where the additional carriers are excluded from the calculation.

To remove spurious image charge interactions in case of charged defect supercell calculations, there are additional terms for the formation energies (Eq. 2.1) needed to correct for the electrostatic interaction between the defect and the compensating background [79, 80, 94]. This contribution depends on the dielectric constant and can exceed 1 eV in wide-gap systems [115, 116]. In a topological semimetal, however, it should vanish for sufficiently large cell sizes. Indeed, for the largest 320 atom cell, we find that the formation energies of the band-filling corrected neutral (green) and the charged (blue) defects agree closely within about 0.01 eV (see Fig. 2.1a), for both the Cd_i^{+2} and V_{Cd}^{-2} defects. The more significant energy differences in the smaller cells are likely reflecting residual image charge interactions due to incomplete screening at shorter distances. Thus, from the cell size dependence in Fig. 2.1(a), we can estimate the screening length in Cd_3As_2 as $L \sim 2$ nm. We conclude then that Cd_i and V_{Cd} can be described as fully ionized defects, where the formation energy depends on the Fermi level (see Eq. 2.1). In this model, the contribution of band-filling to $\Delta H_{D,q}$ is encoded in the change of E_F with varying free carrier (electron or hole) concentrations [78].

2.4 Band energy dependence

Since the formation energies of defects at finite concentrations in Cd_3As_2 depend innately on the electronic structure and density of states (DOS) near E_F, it is imperative that a quantitative description of its defects involves an accurate representation of the electron band energies. It is well known that the local approximation of standard DFT functionals leads to a significant underestimation of the band gap in insulating systems [118, 136]. This limitation often has a considerable effect on calculated electronic properties and defect formation energies [79]. Similar problems exist for DFT calculations of topological materials: the approximation to the functional can give rise to quantitative errors in band dispersions [33, 163], but it can also lead to an incorrect description of the overall band ordering [2, 152]. DFT electronic structure errors must therefore be addressed when considering the Cd_3As_2 Dirac semimetal band structure.

We performed QSGW calculations of Cd_3As_2 in the 80-atom primitive cell to obtain an accurate model of the electronic structure. These QSGW results are shown alongside the band structures for both the GGA and SCAN functionals in Fig. 2.1(b)-(d). There are several key features that distinguish the QSGW electronic structure from standard DFT. First, the Dirac point lies closer to the center of the Brillouin zone, and the band dispersion nearby is greater (a close-up of the band structure around the Dirac point is shown below). Second, the unoccupied s-like bands above the Dirac point (highlighted in red) are moved upward. Third, the QSGW calculation exhibits a strong upward shift of an occupied band below E_F (highlighted in blue). This band has a dominant As-p like character throughout the BZ but becomes s-like at Γ, revealing its origin from the topological band inversion.

These modifications to the electronic structure are analogous to GW corrections of DFT band energies in semiconductors, which generally increase the energy separation of unoccupied s-like and occupied p-like states [10, 72, 130]. It is essential these corrections be included in the defect model for Cd_3As_2 because the band energies near the Dirac point affect the formation energies of defects. We find that the SCAN functional partially accounts for each of the above features, so that the

SCAN band structure lies somewhere in between GGA and QSGW (see Fig. 2.1). To obtain the most accurate defect model, we will combine below the defect formation energies from supercell calculation in SCAN, with the DOS obtained from QSGW with high-resolution BZ integration. This strategy is inspired by a similar approach that proved successful in semiconductors and insulators [115].

2.4.1 QSGW band structure and high-resolution density of states

The quasiparticle self-consistent GW calculations [68] were performed using the Questaal code [113]. To make the potential self-consistent, we used a 3×3×3 k-mesh in the construction of the polarizability and self-energy, and a 4×4×4 k-mesh for the one-body part. A 4×4×4 mesh (for GW) and 6×6×6 mesh (for the one-body part) was also used to confirm the system was converged in k. The basis set consisted of atom-centered smooth Hankel functions, described in Ref. [113], distributed as follows. As consisted of *spdfspd* orbitals with the $3d$ partial wave included in the valence as a local orbital. Cd had *spdfspd* orbitals with the $5d$ partial wave included as a local orbital. The basis was enlarged with "floating orbitals" centered on 48 interstitial sites. All told, the Hamiltonian had a rank of 2832. The G cutoff for the coulomb interaction in the mixed product basis was set at 4.4 Ry, and the G cutoff for the one-particle states generating G and W was 5.2 Ry. For energy integration the frequency mesh spacing was 0.04 Ry at low frequency. Figure 2.3(a) shows the same band structure data as Fig. 2.1(d), but focusing on the vicinity of the Dirac point. The total DOS, shown in Fig. 2.3(b) and (c), was obtained from a Brillouin-zone integration on a very fine 24×24×48 k-mesh.

2.5 Absolute defect formation energies

In addition to an improved description of the band structure (Fig. 2.1) and the associated electronic density of states, GW calculations also allow us to determine the change in absolute band energies relative to the underlying DFT functional. In semiconductors and insulators, the corresponding conduction and valence band edge shifts have been shown to dramatically improve the

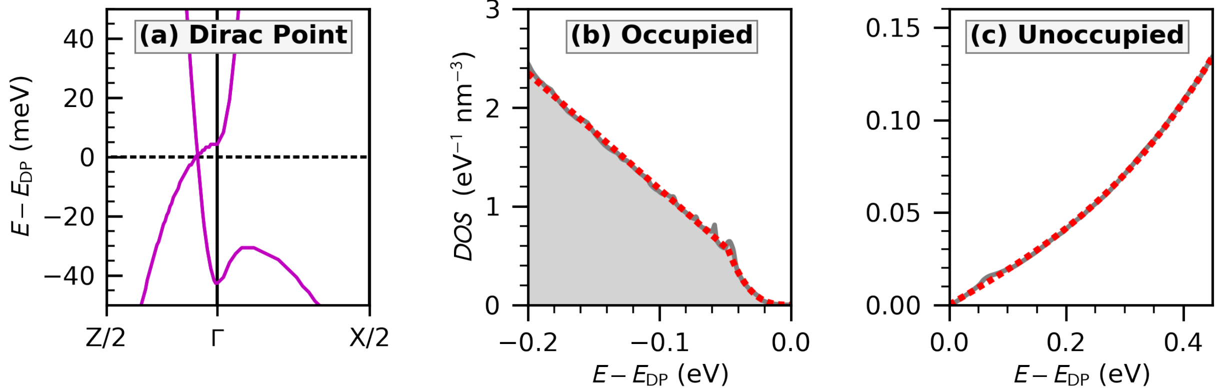


Figure 2.3: (a) The QSGW band structure zoomed in near the Dirac point. (b) Occupied and (c) unoccupied density of states calculated in QSGW as a function of energy E measured with respect to the Dirac point E_{DP}. Dashed red line shows the numerically fitted model function g_{QSGW}.

formation energies of fully ionized defects calculated in DFT [115]. Similar corrections are necessary for Cd_3As_2: The DFT supercell calculations yield the formation energies for the charged defects using the Dirac point E_{DP} as a reference energy for the Fermi level. In so far the GW calculation alters E_{DP} on an absolute scale, the Fermi level must follow suit to maintain the correct charge balance, thereby affecting the formation energies according to Eq. (2.1). To determine the shift of E_{DP}, we performed GW calculations of Cd_3As_2 in which the wavefunctions and hence the electron density and electrostatic potential, of the DFT-SCAN functional are maintained. That is, we performed an eigenvalue self-consistent GW calculation with VASP similar as in previous work on insulators [72, 115], but using the SCAN functional and including spin-orbit coupling. In this type of calculation, the wavefunctions, electron density, and electrostatic potential of the underlying SCAN solution are kept fixed, thereby providing GW band energies that are directly comparable to those in SCAN. To reduce the computing intensity, we combined the results of a single-shot GW calculation on a 2×2×2 k-mesh with a full eigenvalue self-consistent solution at the Γ point. While this relatively coarse k-mesh does not include the Dirac point explicitly, its shift can be approximated by the energy change of the highest occupied states in the GW calculation, because the occupied valence band like states below the Fermi level have much less dispersion than those above E_{F} [see Fig. 2.1(b)-(d)].

The resulting shift of the reference energy is determined as $\Delta E_{\rm DP} = -0.19$ eV, which we include in the calculation of the formation energies via Eq. (2.1).

2.5.1 Limits to chemical potentials

The defect energies depend on synthesis conditions through the chemical potentials $\mu_\alpha = \mu_\alpha^0 + \Delta\mu_\alpha$ of Cd and As, observing $\Delta H_{\rm f} = 3\Delta\mu_{\rm Cd} + 2\Delta\mu_{\rm As} = -0.59$ eV/fu, where μ_α^0 are the energies of the elemental solids and $\Delta H_{\rm f}$ is the calculated formation enthalpy of Cd_3As_2 per formula unit (fu). Considering CdAs and $CdAs_2$ as potential competing phases, we find that the Cd-rich limit is given by phase coexistence with Cd metal ($\Delta\mu_{\rm Cd} = 0$, $\Delta\mu_{\rm As} = -0.30$ eV) and the As-rich limit by coexistence with $CdAs_2$ ($\Delta\mu_{\rm Cd} = \Delta\mu_{\rm As} = -0.12$ eV), whereas CdAs lies slightly above the convex hull. Figure 2.4 illustrates a phase diagram for the stable compounds formed by Cd and As, and shows this stable range of chemical potentials for Cd and As.

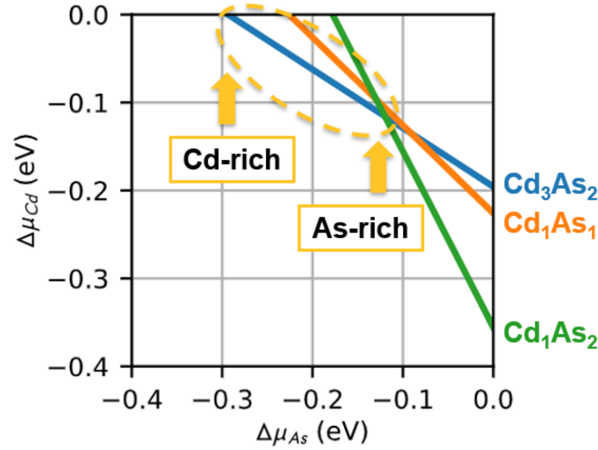


Figure 2.4: Chemical potential stability phase diagram for Cd_3As_2.

To future students, stability phase diagrams like Fig. 2.4 can be constructed in the following way. First, perform total energy calculations for all of the possible competing phases, including the elemental reference phases. Second, normalize the energies of all phases by dividing out the number of formula units in the unit cell. Third, calculate the formation enthalpy of each phase by subtracting the (normalized) elemental energies multiplied by the stoichiometry coefficients from

the (normalized) energy of that phase, e.g. $\Delta H_{\mathrm{f}} = E_{\mathrm{Cd_3As_2}} - 3E_{\mathrm{Cd}} - 2E_{\mathrm{As}}$. Finally, the formation enthalpy must equal the combined change in chemical potentials, e.g. $\Delta H_{\mathrm{f}} = 3\Delta\mu_{\mathrm{Cd}} + 2\Delta\mu_{\mathrm{As}}$, but those equations can be rearranged to express one chemical potential in terms of the others, e.g. $\Delta\mu_{\mathrm{Cd}} = \frac{1}{3}\Delta H_{\mathrm{f}} - \frac{2}{3}\Delta\mu_{\mathrm{As}}$. This allows you to plot lines (higher dimensional contours when considering more constituent elements) like those in Fig. 2.4, and the phase that minimizes the other chemical potentials when one $\Delta\mu$ is held fixed is the stable phase.

2.5.2 Results for intrinsic formation energies

Figure 2.5 shows the formation energies for the energetically most favorable, fully ionized defects as a function of E_{F}, where we calculated the energies of all elements, compounds, and defects using the SCAN functional with spin-orbit coupling. In contrast to the two Cd defects described above, the As vacancy V_{As} introduces a localized defect state close to the Dirac point [104], which can be fully occupied ($q = -1$), half occupied ($q = 0$), or empty ($q = +1$). The neutral state turns out to be slightly unstable ("negative-U defect" [6]), resulting in a $(+/-)$ charge transition level at $E_{\mathrm{DP}} + 0.15$ eV in the formation energy diagram in Fig. 2.5. Overall, the formation energies of V_{As} are higher than those of $\mathrm{Cd_i}$ and V_{Cd} throughout the range of chemical potentials. Therefore, the net doping will be largely determined by the balance between the Cd interstitial and vacancy defects. The numerical values of the defect energies ΔH_{D} for all nonequivalent sites are given in Sec. 2.5.3. The formation energies for the $\mathrm{As_{Cd}}$ and $\mathrm{Cd_{As}}$ antisite and the $\mathrm{As_i}$ interstitial defects were also calculated for completeness. As seen below, they are sufficiently high to exclude them from the following analysis.

2.5.3 Defect density of states and numerical values of formation energies

For completeness, we include here Fig. 2.6, with formation energies and density of states for the defects considered in the main text as well as the As interstitial and antisite defects $\mathrm{As_{Cd}}$ and $\mathrm{Cd_{As}}$. The $\mathrm{As_i}$ and antisite defects are seen to be considerably higher than the other defects and hence lower in concentration, so they are excluded from the defect equilibria results below. In

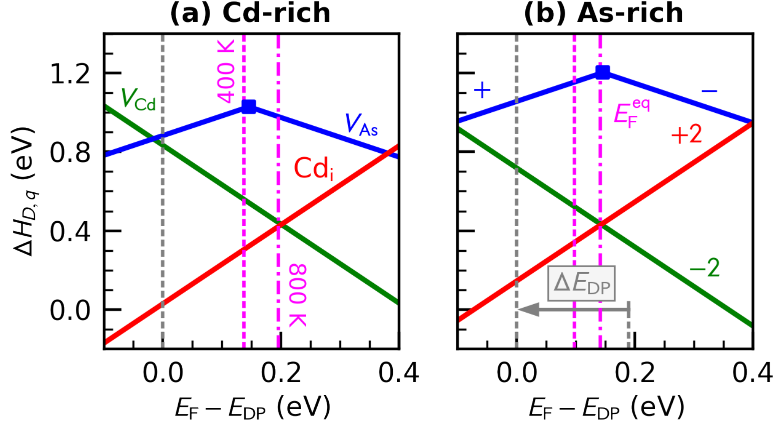


Figure 2.5: Formation energies $\Delta H_{D,q}$ of the intrinsic point defects in Cd_3As_2 as a function of Fermi level E_F in (a) Cd-rich and (b) As-rich conditions. Only the lowest energy Wyckoff positions are shown for V_{As} and V_{Cd}. The GW correction $\Delta E_{DP} = -0.19$ eV and equilibrium Fermi level E_F^{eq} for growth at $T_g = 400$ K and 800 K are indicated.

contrast to V_{Cd} and Cd_i, all other defects exhibit localized defect state close to the Dirac point, causing a sequence of distinct charge states corresponding to the slopes of the formation energy in Fig. 2.6(a). Table 2.1 shows the formation energies calculated via the above procedure at a reference condition $E_F = E_{DP}$ for the Fermi level and $\Delta\mu_{Cd} = \Delta\mu_{As} = 0$ for the chemical potentials. The formation energies for specific values of E_F and the chemical potentials can be evaluated with help of Eq. (2.1).

Table 2.1: Calculated defect formation energies for all Wyckoff positions at reference condition $E_F = E_{DP}$ and $\Delta\mu_{Cd} = \Delta\mu_{As} = 0$.

	Cd_i	$V_{Cd,1}$	$V_{Cd,2}$	$V_{Cd,3}$	$V_{As,1}$	$V_{As,2}$	$V_{As,3}$
$\Delta H_{D,(q=-2)}$ (eV)		0.83	1.17	1.03			
$\Delta H_{D,(q=-1)}$ (eV)					1.61	2.05	1.47
$\Delta H_{D,(q=+1)}$ (eV)					1.20	1.47	1.18
$\Delta H_{D,(q=+2)}$ (eV)	0.03						

2.6 Defect equilibria

It remains an open question how competing point defects collectively impact the properties of experimentally grown Cd_3As_2. We therefore consider a defect equilibrium and investigate the

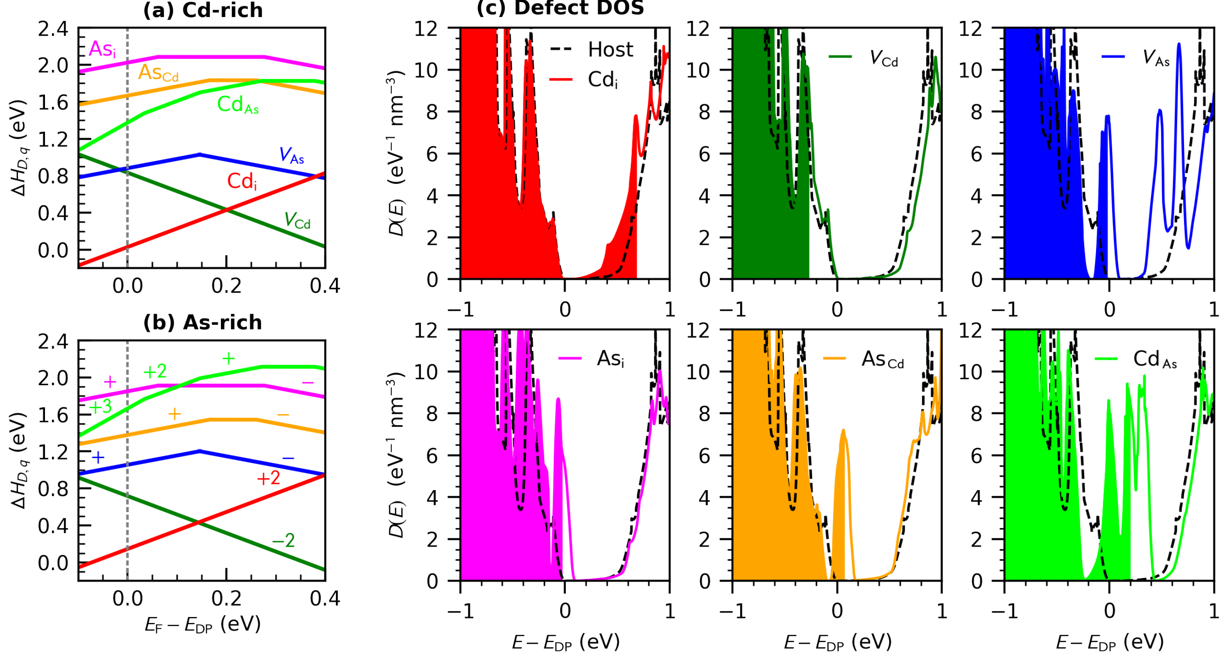


Figure 2.6: Formation energies $\Delta H_{D,q}$ of the intrinsic point defects in Cd_3As_2 as a function of Fermi level E_F in (a) Cd-rich and (b) As-rich conditions. (c) Host and defect density of states (DOS) for point defects in Cd_3As_2 calculated with the SCAN functional as a function of energy E measured with respect to the Dirac point E_{DP}. Only the lowest energy Wyckoff positions are shown.

combined effect of all of the contributing defects simultaneously. For this thermodynamic equilibrium, we assume an initial growth temperature and calculate the associated defect concentrations as a function of the Cd and As chemical potentials, following a procedure similar to Refs. [73, 76]. For the relatively dilute defect concentrations expected here, the defect concentrations in equilibrium at a temperature T are given by

$$c_{\mathrm{D},q}(E_{\mathrm{F}}, \{\mu\}, T) = N_{\mathrm{D}} e^{-\Delta H_{\mathrm{D},q}(E_{\mathrm{F}},\{\mu\})/k_{\mathrm{B}}T}, \tag{2.4}$$

where N_{D} is the concentration of the atomic site where the defect D substitutes, k_{B} is the Boltzmann constant, and $\Delta H_{\mathrm{D},q}(E_{\mathrm{F}}, \{\mu\})$ contains the implicit E_{F} dependence as given by Eq. (2.1) above.

2.6.1 Concentration of non-dilute defects

To future students, Eq. 2.4 is the limiting dilute approximation for the concentration of point defects in a solid. For completeness, I will also give the complete equation for non-dilute defects that is rarely included in the literature but becomes relevant for defect concentrations greater than a few percent. In that regime, the defect concentrations in equilibrium are (see Appendix B)

$$c_{\mathrm{D},q} = N_{\mathrm{D}} \frac{e^{-\Delta H_{\mathrm{D},q}/k_{\mathrm{B}}T}}{1 + e^{-\Delta H_{\mathrm{D},q}/k_{\mathrm{B}}T}}, \tag{2.5}$$

or equivalently

$$\frac{c_{\mathrm{D},q}}{N_{\mathrm{D}} - c_{\mathrm{D},q}} = e^{-\Delta H_{\mathrm{D},q}/k_{\mathrm{B}}T}, \tag{2.6}$$

where each quantity is the same as in Eq. 2.4. The dilute limit corresponds to $\Delta H_{\mathrm{D},q} \gg k_{\mathrm{B}}T$, or equivalently $c_{\mathrm{D},q} \ll N_{\mathrm{D}}$, and applying that limit to Eq. 2.5 or 2.6 gives Eq. 2.4 as it should.

2.6.2 Charge balance condition in the thermodynamic model

The defect equilibrium requires self-consistency between the defect densities $c_{\mathrm{D},q}$ and the Fermi level E_{F} while simultaneously observing overall charge balance between defect charges and the carrier concentrations. To accurately account for the band-filling effects discussed above and to obtain the corresponding relationship between E_{F} and the carrier concentrations, we integrate the DOS $g_{\mathrm{QSGW}}(E)$ from the QSGW calculation, weighted with the Fermi-Dirac distribution, i.e.,

$$n_{\mathrm{e}} = \int_{E_{\mathrm{DP}}}^{\infty} \frac{g_{\mathrm{QSGW}}(E)}{e^{(E-E_F)/k_{\mathrm{B}}T} + 1} dE, \tag{2.7}$$

for electrons and, analogously,

$$n_{\mathrm{h}} = \int_{-\infty}^{E_{\mathrm{DP}}} g_{\mathrm{QSGW}}(E) \left(1 - \frac{1}{e^{(E-E_F)/k_{\mathrm{B}}T} + 1} dE \right), \tag{2.8}$$

for holes. The integration is facilitated by using a model function for $g_{\mathrm{QSGW}}(E)$, discussed presently.

To facilitate the integration of the Fermi-Dirac distribution in eq. 2.7, we fit the DOS from QSGW on either side of the Dirac point with a model function g_{QSGW} using a mix of linear and cubic polynomials for both the occupied and unoccupied states. Figure 2.3 shows the DOS for both the QSGW data and the model function which reproduces the original data very closely. We use this fit for the self-consistent defect equilibrium calculated here. The integration of g_{QSGW} weighted with the Fermi-Dirac distribution is performed numerically with 200 data points per $k_{\mathrm{B}}T$, with an upper bound at least 50 $k_{\mathrm{B}}T$ above E_{F}. The self-consistent solution is found by iteratively updating E_{F} until the overall charge balance is satisfied to within $\Delta q = 10^5\ e/\mathrm{cm}^{-3}$.

2.6.3 Results for defect equilibria

Experimentally, Cd_3As_2 is grown in either bulk or thin film form. Molecular Beam Epitaxy (MBE) has been proven as a valuable tool for growth of high-quality Cd_3As_2 with good control over the elemental chemical potentials [104, 125, 133]. MBE has a growth temperature window of about $T_{\mathrm{g}} \approx 110$ - 250 °C, where the upper limit results from the modest formation enthalpy ΔH_{f} (see above) and the fugacity of the atomic and molecular species. On the other hand, bulk synthesis can reach temperatures above 400 °C [31, 129]. Thus, we consider here a range between 400 and 800 K for the calculations of defect equilibria.

The resulting defect concentrations are given in Table 2.2 and include a sum over all Wyckoff sites for V_{Cd} and V_{As} as well as a sum over the different charge states of V_{As}. As expected from the formation energies, the $\mathrm{Cd_i}$ and V_{Cd} defects are dominant and the concentrations of the As vacancies are several orders of magnitude lower. On account of the charge balance condition, the equilibrium Fermi level $E_{\mathrm{F}}^{\mathrm{eq}}$ at T_{g} lies close to the intersection point between the $\mathrm{Cd_i}$ and V_{Cd} formation energies (Fig. 2.5). Table 2.2 also shows the net doping concentrations $c_{\mathrm{nd}} = 2[c_{\mathrm{D}}(\mathrm{Cd_i}) - c_{\mathrm{D}}(V_{\mathrm{Cd}})]$, indicating that n-type growth ($c_{\mathrm{nd}} > 0$) is favored, as expected, under Cd-rich conditions for the full range of T_{g}. Interestingly, however, Cd_3As_2 becomes net p-type ($c_{\mathrm{nd}} < 0$) at the higher end of growth temperatures under As-rich conditions. This transition suggests the intriguing possibility of doping-neutral growth of Cd_3As_2.

Table 2.2: Defect equilibria for Cd_3As_2 obtained from first-principles defect formation energies. Given are the concentrations c_D of the intrinsic defects and the resulting net doping c_{nd} depending on the growth temperature T_g and the Cd- or As-rich condition.

Cd/As rich	T_g (K)	c_D (cm^{-3}) Cd$_i$	V_{Cd}	V_{As}	c_{nd} (cm^{-3})
Cd	400	1.0×10^{18}	8.2×10^{14}	1.3×10^{9}	2.0×10^{18}
	600	3.0×10^{18}	1.2×10^{18}	3.0×10^{13}	3.7×10^{18}
	800	1.6×10^{19}	1.5×10^{19}	4.6×10^{15}	2.9×10^{18}
As	400	3.5×10^{17}	2.4×10^{15}	2.2×10^{7}	6.9×10^{17}
	600	2.1×10^{18}	1.7×10^{18}	1.0×10^{12}	9.1×10^{17}
	800	1.5×10^{19}	1.6×10^{19}	3.6×10^{14}	-2.7×10^{18}

To elucidate the physical origin of the doping-type inversion, we study in more detail the temperature dependence of the Fermi level, shown in Fig. 2.7(a) for three different growth temperatures (circles). Both the asymmetry of the DOS (see Fig. 2.3 showing a much larger DOS below E_{DP} than above) and the width of the Fermi-Dirac distribution play important roles in determining the position of E_F^{eq}. At higher temperatures, their combined effect is to add positive charge from hole carriers even when E_F lies above E_{DP}. To maintain charge balance, E_F is pushed further upwards, which favors the formation of negatively charged acceptor-type defects (cf. Fig. 2.5). As a result, the net doping c_{nd} can become negative (cf. Table 2.2).

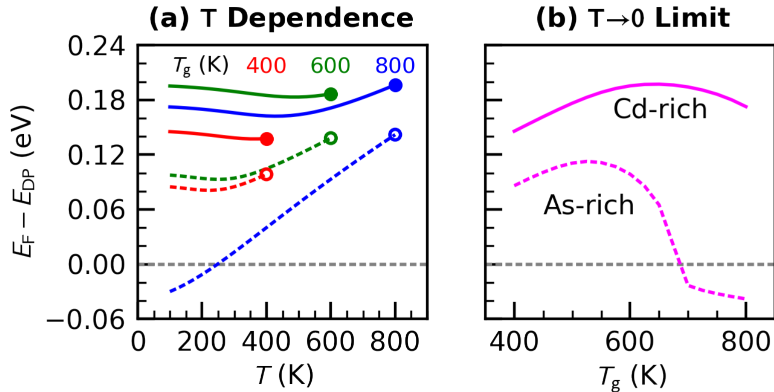


Figure 2.7: (a) Temperature dependence of E_F in the presence of defect concentrations determined for three different growth temperatures T_g (circles). (b) Low temperature limit of E_F as a function of T_g. Solid and dashed lines indicate Cd-rich and As-rich conditions, respectively.

For accessing the topological band structure features of Cd_3As_2, the position of E_F at low temperature matters more than that at T_g. To model the T dependence of the electronic system, we reevaluate the charge-balance condition with the Fermi-Dirac distribution (Eq. 2.7), while keeping the defect and corresponding net doping concentrations from the equilibrium at T_g. (At this point, Cd vacancies and interstitials could recombine if at least one of the species is mobile at the respective temperature. However, since the net doping, E_F, and the electron and hole concentrations are not affected by such recombination events, this effect is not explicitly modeled.) Figure 2.7(a) shows the resulting Fermi level $E_F(T)$, and the associated carrier concentrations $n_e(T)$ and $n_h(T)$ are shown in Fig. 2.8 . These self-consistent carrier concentrations n_e and n_h correspond to the values of E_F in Fig. 2.7 and the defect concentrations in Table 2.2. In the low temperature limit, the dominant carrier type approaches the net doping concentration ($n_e \to c_{nd}$ or $n_h \to -c_{nd}$). In the n-type case, we observe an interesting nonmonotonic behavior of $E_F(T)$. To maintain the charge balance, the Fermi level first decreases with T, reflecting the positive slope of the DOS above E_{DP} (Fig. 2.3). Above a certain temperature that depends on c_{nd}, it starts to increase again, reflecting the high-T behavior and DOS asymmetry described above.

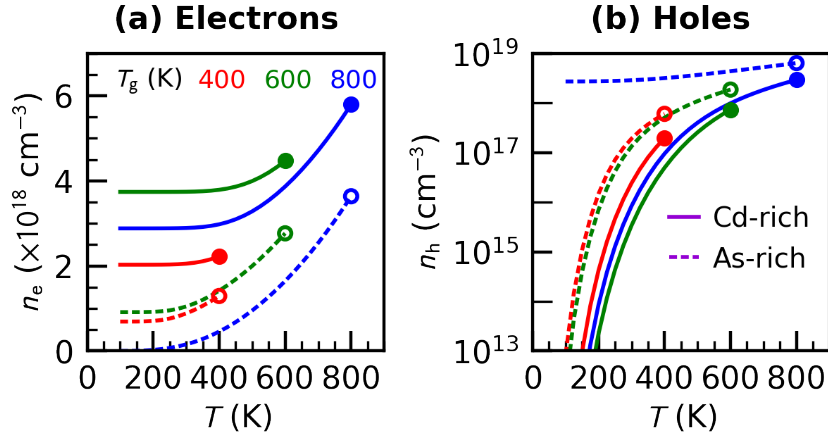


Figure 2.8: The (a) electron and (b) hole concentrations n_e and n_h plotted as a function of temperature T for the three growth temperatures in Fig. 3(a) (indicated as upper limits). Shown for Cd-rich and As-rich conditions. A logarithmic scale is used for hole concentrations.

Figure 2.7(b) summarizes the $E_F(T)$ modeling by showing the low-T limit of E_F as a function

of T_g. The zero-temperature limit of E_F shown in Fig. 3(b) was found using a bisection method for determining E_F given a net doping concentration c_{nd}, the model function g_{QSGW}, and a step function for the Fermi-Dirac distribution. While Cd-rich growth results in n-type doping with $E_F(0) > E_{DP}$ for the full range of T_g, a conversion to net p-type doping with $E_F(0) < E_{DP}$ occurs around 700 K under the As-rich condition. Thus, our modeling results suggest that doping-neutral Cd_3As_2 could be achieved through control of the growth process. There are, however, two potential experimental challenges: Maintaining sufficiently As-rich conditions may require considerable As overpressures at higher temperatures, and rapid thermal processing techniques may be needed to quench in the defect equilibrium at these conditions.

2.6.4 Unique behavior of Fermi level with temperature for a topological semimetal

While the prospect of controlling net doping through growth is quite compelling, the behavior of $E_F(T)$ for a topological semimetal is non-intuitive and requires careful consideration. Several features of the graph of $E_F(T)$ for both n-doped and p-doped Cd_3As_2, and even the origin of p-type doping itself, can be better understood by studying the charge balance condition, Eq. (2.7). The asymmetry of the DOS below/above the Dirac point and the width of the Fermi-Dirac distribution together determine the overall behavior of E_F. The effect of maintaining charge balance with varying temperature is illustrated in Fig. 2.9, which graphs the Fermi-Dirac distribution (Fig. 2.9a), the QSGW model DOS (Fig. 2.9b), and the weighted areas for electrons and holes corresponding to two points on the dashed blue line of Fig. 2.7, i.e., 800 K (Fig. 2.9c) and 250 K (Fig. 2.9d).

At higher temperatures, the Fermi-Dirac distribution is wide enough to capture the much larger density of states below the Dirac point (see Fig. 2.9b) and introduce large numbers of hole carriers even if E_F is above the Dirac point. This means that raising the temperature requires E_F to be pushed upwards to maintain charge balance, and an elevated E_F favors the formation of negatively charged acceptor-type defects (namely the V_{Cd} in Cd_3As_2). The result of growth at this high temperature then, is that acceptors can outweigh donors, so there is net p-type doping (excess holes) even though E_F is above the Dirac point (Fig. 2.9c). Then, as the temperature is decreased,

the difference between electrons and holes is maintained by lowering E_F (Fig. 2.9d), and only in the low-temperature limit does E_F move below the Dirac point.

Conversely, there is an interesting non-monotonic behavior of $E_\mathrm{F}(T)$ in the case of n-type doping (cf. 2.7). This is more easily understood by starting from the low-temperature limit. At zero temperature E_F lies above the Dirac point because of excess electrons from defects. Then, to maintain charge balance, as T increases (but is still relatively low), the positive slope of the DOS means unoccupied DOS is larger than occupied, and charge balance is maintained by E_F decreasing. Finally, once the temperature is large enough that $k_\mathrm{B}T$ is comparable to $E_\mathrm{F} - E_\mathrm{DP}$, which depends on net doping, then the Fermi-Dirac distribution captures the large DOS below the Dirac point and charge balance is maintained by increasing E_F.

2.7 Conclusions

Building upon existing methods for semiconductors and insulators, we defined a computational approach for evaluating first-principles defect equilibria in Cd_3As_2. Several conclusions are more broadly relevant for topological semimetals: due to electronic screening, there are no bound states akin to hydrogenic dopant levels. Nevertheless, defects can have well defined charge states depending on the position of the Fermi level in the band continuum. However, it is particularly important to distinguish electron occupation of defect states from occupation of the band continuum. The latter cases do not represent legitimate charge states of the defect and cause a spurious cell size and concentration dependence of the formation energy. Finally, the overall charge balance condition depends strongly on the shape of the density of states near the Fermi level, often requiring an accurate electronic structure method beyond DFT. In Cd_3As_2, mutually compensating Cd interstitials and vacancies are the dominant intrinsic defects, with the former being the source of unintentional n-type doping, whereas the concentration of As vacancies remains much lower. Neither of the Cd defects cause localized states close to the Dirac point, thereby allowing the Fermi level to penetrate the band continuum. Due to the absence of a band gap and the pronounced asymmetry in the DOS for electron and hole carriers, the defect equilibria exhibit strong and nonmonotonic temperature

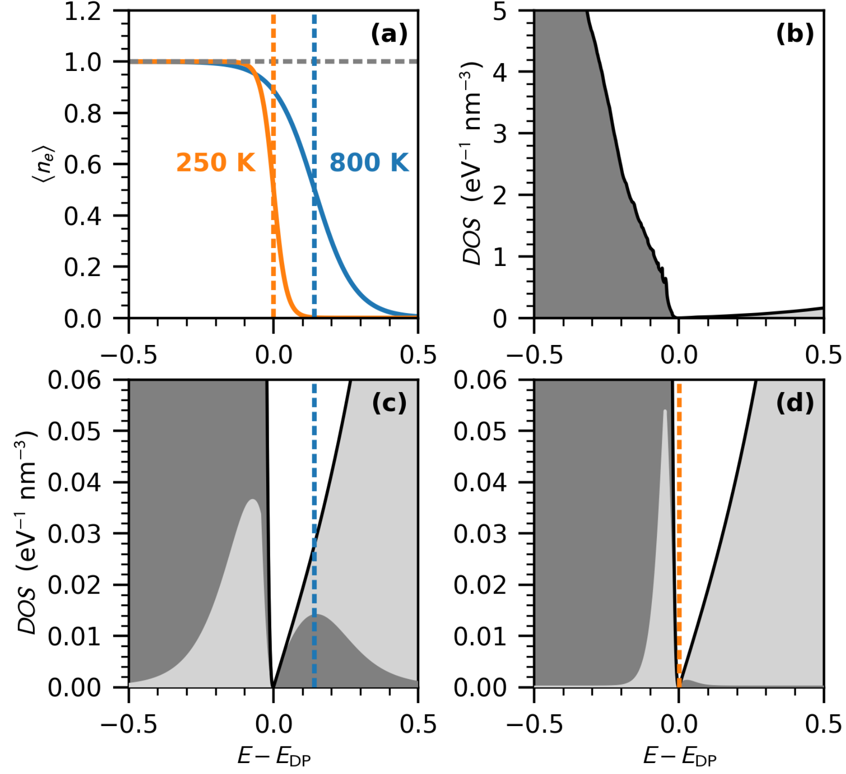


Figure 2.9: (a) Fermi-Dirac occupation of electrons corresponding to As-rich growth of Cd_3As_2 at 800 K (blue) and then cooled to 250 K (orange). The temperature dependent Fermi levels E_F are indicated by vertical dashed lines. (b) QSGW density of states data as a function of energy E near the Dirac point E_{DP}. (c)-(d) Model density of states $g_{QSGW}(E)$ weighted with the Fermi-Dirac distribution for electrons (dark gray) and holes (light gray) at the respective temperatures, 800 K (c) and 250 K (d).

dependencies. The modeling leads to a prediction how doping-neutral Cd_3As_2 could be grown, such that the Fermi level approaches the Dirac point in the low-T limit.

Chapter 3

Extrinsic doping control from defect theory with site redistribution in a pseudo-equilibrium

This chapter is adapted from a manuscript that is soon to be submitted for peer review (working title: "Pseudo-equilibrium theory for extrinsic doping control of the topological semimetal Cd_3As_2"); and from Sec. II of Ref. [3], reproduced with the permission of AIP Publishing.

Abstract. The standard approach for studying defect concentrations in solid-state systems using first principles formation energies is to evaluate a thermodynamic equilibrium at a model growth temperature. For lower (room or operating) temperatures, these defect concentrations are then "frozen in", and only the Fermi level E_F is re-equilibrated. However, this assumption overlooks the possibility of short-range diffusion and defect redistribution, even when the overall composition determined by the growth step remains fixed. To model this redistribution, we developed an approach to solve for the non-equilibrium chemical potentials as a function of temperature while maintaining the elemental stoichiometry. We then apply this approach to the Dirac semimetal Cd_3As_2 in the presence of external dopants. Previous work identified Cd interstitials and vacancies as the primary defects in self-doped Cd_3As_2, with excess interstitials leading to undesirable n-type doping that limits access to the Dirac point under most growth conditions. Here, we utilize the expanded defect equilibrium model to study extrinsic doping with the goal of introducing additional electron accepting defects (Group I on Cd and Group IV on As) and lowering E_F.

3.1 Introduction - Constrained Defect Equilibria

The ground state crystal structure of Cd_3As_2 is derived from an antifluorite lattice with an ordered sublattice of empty cation sites (space group $I4_1/acd$) [4, 29]. The 80-atom primitive cell has 3 nonequivalent Wyckoff sites for both Cd and As and 1 for the empty cation site. We performed total energy calculations of intrinsic defects and dopant substitutions for each unique site in this 80-atom supercell using Density Functional Theory (DFT) and the Projector Augmented Wave (PAW) method as implemented in the VASP code [17, 69]. Specifically, we utilized the Strongly Constrained and Appropriately Normed (SCAN) meta-Generalized Gradient Approximation (meta-GGA) functional [147] with spin-orbit coupling and an eigenvalue self-consistent GW correction to the band energies [72, 138]. The SCAN total energies were then combined with a high resolution Quasiparticle Self-consistent GW (QSGW) [68, 130] calculation of the density of states in Questaal [113] to construct a self-consistent defect equilibria model. Further details are discussed in the Methods section and in Ref. [19].

As a starting point, we maintain the basic assumptions of defects in a solid within the supercell method. That is, the formation energy of a defect D in charge state q is given by

$$\Delta H_{\mathrm{D},q}(E_{\mathrm{F}}, \{\mu\}) = [E_{\mathrm{D},q} - E_{\mathrm{h}}] + \sum_{\alpha} n_{\alpha}\mu_{\alpha} + qE_{\mathrm{F}}, \tag{3.1}$$

where $E_{\mathrm{D},q}$ and E_{h} are total energies of the defect and host supercells, respectively, μ_{α} are the chemical potentials of atom species α that are removed ($n_{\alpha} = +1$) or added ($n_{\alpha} = -1$) to the host crystal in forming the defect, and E_{F} is the Fermi energy. Furthermore, the dilute concentration $c_{\mathrm{D},q}$ of each defect in equilibrium at temperature T is given by

$$c_{\mathrm{D},q}(E_{\mathrm{F}}, \{\mu\}, T) = N_{\mathrm{D}} e^{-\Delta H_{\mathrm{D},q}(E_{\mathrm{F}}, \{\mu\})/k_{\mathrm{B}}T}, \tag{3.2}$$

where N_{D} is the atomic site concentration, k_{B} is the Boltzmann constant, and $\Delta H_{\mathrm{D},q}(E_{\mathrm{F}}, \{\mu\})$ is given by Eq. (3.1). An overall thermodynamic equilibrium is determined by a self-consistent solution

between the concentrations of all defects, E_F, and overall charge balance between charged defects and free carriers (discussed in further detail in Ref. [19]). For unconstrained equilibrium growth, the chemical potentials $\mu_\alpha = \mu_\alpha^0 + \Delta\mu_\alpha$ are associated with thermodynamic exchange between the sample and the associated elemental reservoirs, and they are bounded by the solubility limits of competing phases. In Cd_3As_2, the Cd-rich limit is defined by phase coexistence with Cd metal and the As-rich limit by coexistence with $CdAs_2$ [19]. The limiting values for extrinsic dopants are then subsequently determined by the phases they form with Cd and As, e.g. Au_3Cd_5 for Au [66] or $CdGeAs_2$ for Ge [1].

For temperatures below the chosen growth temperature, defects are traditionally assumed to have fixed concentrations equal to the equilibrium values and only the chemical potential for electrons (E_F) is re-equilibrated [19, 43]. This constraint is enforced through the use of artificial chemical potentials that effectively "freeze in" the defects to what they were during growth. The underlying assumption is that the material is cooled on a fast enough time scale that defects do not have an opportunity to undergo short-range diffusion. However, diffusion and defect recombination events can realistically occur as low as room temperature, where materials are typically stored. The purpose of our expanded thermodynamic defect model is to determine to what extent individual defect concentrations can change as an isolated material is cooled after growth.

The experimental situation we aim to model is a sample that after undergoing thermodynamic equilibration at a chosen growth temperature T_g is then cut off so it can no longer readily exchange atoms with an external reservoir, e.g. an elemental source or the atmosphere. In this scenario atoms can still diffuse through the material so the concentrations of individual defects can change and equilibrate at lower temperatures, but the combined defect stoichiometry (total amount that is not present in the defect-free system) of each element is held fixed to the growth temperature value. This constrained equilibrium system allows us to study the redistribution and recombination of defects at intermediate temperatures between T_g and the lower limit determined by diffusion energy barriers. The current work demonstrates how to incorporate this constraint within the existing framework for defect equilibria. Namely, at the growth step we add the defect and dopant concentrations to

determine the initial defect stoichiometries s_{alpha} of each element. At lower temperatures we then define an objective function that is the absolute difference in the stoichiometries from the growth values. We dynamically update the chemical potentials to minimize this objective function and thereby constrain the system to keep the total amounts of each element fixed. Implementing this constraint becomes more mathematically complex as more atomic species are considered (see below). Therefore, we consider first a constrained equilibrium solution for intrinsic defects in $\mathrm{Cd_3As_2}$, and then subsequently we study how the results change with the introduction of extrinsic dopants.

3.2 Redistribution of Intrinsic Defects

We previously found [19] the dominant defects that determine net doping in $\mathrm{Cd_3As_2}$ are the Cd interstitial $\mathrm{Cd_i}$ (occupying the ordered empty sites in the base crystal structure) and Cd vacancy V_{Cd}, and to a lesser extent the As vacancy V_{As}. Under a majority of growth conditions, the $\mathrm{Cd_i}$ (2 electron donor) outnumbers the V_{Cd} (2 electron acceptor), so the net carrier concentration $c_{\mathrm{nc}} = n_{\mathrm{e}} - n_{\mathrm{h}} \approx 2[c_{\mathrm{D}}(\mathrm{Cd_i}) - c_{\mathrm{D}}(V_{\mathrm{Cd}})]$ involves excess electrons and the system is net n-type. However, with As-rich conditions and a sufficiently high growth temperature the vacancies can surpass interstitials and feasibly give net-neutral or p-type doping. To compare both electron doped and hole doped $\mathrm{Cd_3As_2}$, then, we consider the recombination of intrinsic defects after As-rich growth at $T_{\mathrm{g}} = 600$ K (excess $\mathrm{Cd_i}$) and $T_{\mathrm{g}} = 800$ K (excess V_{Cd}).

If the self-doped system is isolated and cooled after growth, the defect stoichiometries of Cd and As are expected to be maintained by the number of V_{As} being constant and the $\mathrm{Cd_i}$ and V_{Cd} defects recombining at lower temperatures. A constrained equilibrium solution for the intrinsic defects then involves holding the number of As vacancies fixed and finding the chemical potential for Cd such that the difference between interstitials and vacancies is unchanged. We accomplished the latter task of determining the constrained $\Delta\mu_{\mathrm{Cd}}$ at lower temperatures by means of the bisection method. The results of this constrained solution are plotted in Fig. 3.1, and as expected the Cd stoichiometry is maintained by interstitials and vacancies recombining at lower temperatures. As temperature is decreased, the concentrations of both defects decrease significantly, and the defect

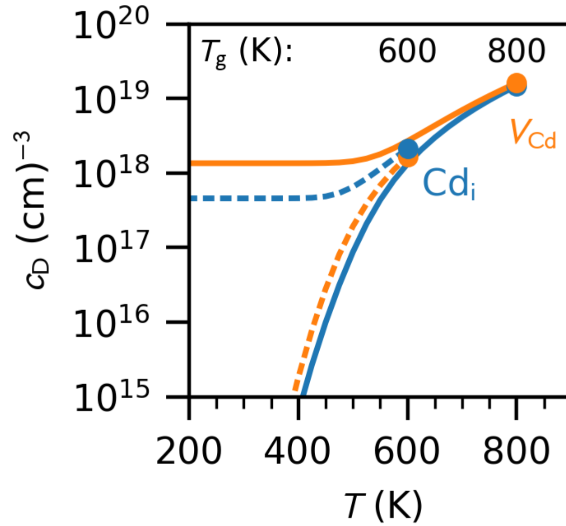


Figure 3.1: Concentrations c_D of intrinsic Cd defects in Cd_3As_2 as a function of temperature T when grown under As-rich conditions at 600 K (dashed) and 800 K (solid).

that was in excess moves toward a limiting value of the initial difference between the two. Because the Cd_i has a +2 charge state and the V_{Cd} a 2- charge state, the net doping (electrons minus holes) is unaffected by their recombination and therefore does not change as a function of temperature. However, these recombination events result in an overall reduction in the defect density, which can be largely beneficial for e.g. improving the electron mobility or other transport properties of the system.

3.3 Defect behavior and dopant density of states

Despite the absence of a band gap, electrically active point defects (either intrinsic defects or dopants or impurities) in topological semimetals can be categorized into one of the two types of behavior, resembling "deep" and "shallow" levels in semiconductors [77]. The first kind are those that introduce localized defect states close to the energy E_{DP} of the Dirac point, within a range of accessible Fermi levels E_F. These defects can form in different charge states depending on the occupation of the state as a function of E_F and, therefore, exhibit a tendency to pin the Fermi level. The other class of defects are those that do not introduce a localized state near E_F, and, instead,

release the carriers (electrons or holes for donor- or acceptor-like defects, respectively) into the band continuum, thereby also shifting the Fermi level.

Figure 3.2 shows the density of states (DOS) for select intrinsic and extrinsic defects in Cd_3As_2 and how they compare to the 80-atom defect-free primitive cell. The shape of the DOS illustrates the two distinct defect behaviors described above. The Cd interstitial Cd_i, Cd vacancy V_{Cd}, and the substitution of tin on arsenic Sn_{As} all introduce delocalized electrons or hole carriers, raising or lowering the supercell Fermi level while the DOS of the Cd_3As_2 host is only moderately perturbed by the presence of the defect (especially when considering the high defect concentration implied by the limited supercell size). Thus, the supercell DOS exhibits a metallic behavior with E_F residing in the band continuum in these cases. In contrast, the V_{As} and Sn_{Cd} defects exemplify the case of localized defect states in the vicinity of the Dirac point. Specifically, V_{As} exhibits amphoteric behavior where a doubly degenerate defect state can be either occupied or unoccupied, corresponding to the negative V_{As}^{-1} As charge state with a defect-induced spike in the DOS at -0.1 eV and to the positive V_{As}^{+1} with a defect state at 0.2 eV, respectively (see Fig. 3.2). Similarly, the neutral Sn_{Cd} causes a defect-localized state just above E_{DP}.

Due to the possibility of occupation changes, the defects with localized states give rise to charge transition levels much like in the case of semiconductors and insulators, except that they occur inside the band continuum instead of within the band gap. This means that the preferred charge state can change as E_F varies with the presence of other electrically active defects and with the changes in temperature. In semiconductors, the defects without localized states are, nevertheless, able to bind carriers in a "shallow" effective mass-like level [155], where the defect state essentially represents an envelope function of the "perturbed host state" [77]. Due to the absence of a band gap and the corresponding screening of the electrostatic potential, this type of bound state does not exist in a semimetal. Instead, defects of this kind, including Cd_i, V_{Cd}, and Sn_{As}, are more appropriately viewed as charged defects being compensated by continuum carriers (or other defects of opposite charge).

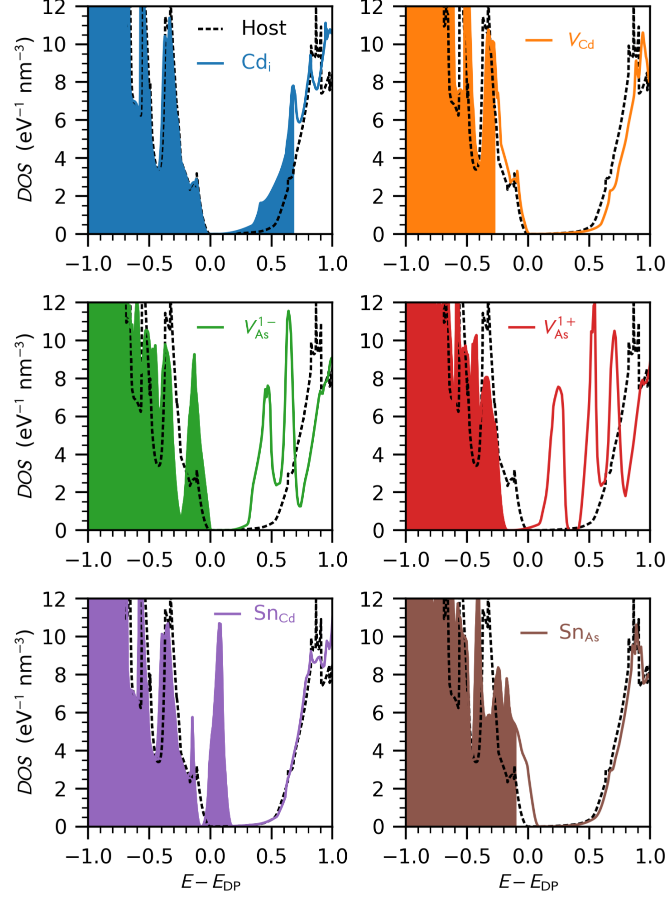


Figure 3.2: Host and defect/dopant density of states (DOS) in Cd_3As_2 calculated with the SCAN functional (including spin-orbit coupling) as a function of energy E measured with respect to the Dirac point E_{DP}. Only the lowest energy Wyckoff positions are shown for each defect.

3.4 Choice of Extrinsic Dopant

With the goal of expanding the constrained equilibrium model to Cd_3As_2 in the presence of external dopants, the question remains of what dopants should be studied. Because the intrinsic defects favor n-type doping under most conditions, we aim to lower E_F closer to the Dirac point by choosing a dopant that favors the formation of electron accepting defects. We therefore consider Group I substitution on Cd and Group IV substitution on As (both with a 1- charge state). However, each potential dopant also introduces compensating electron donating defects. Namely, Group I can occupy the ordered empty sites of the primitive cell similar to the Cd interstitial but with a +1

charge, and Group IV can substitute on Cd with a +2, +1, or 0 charge state (see Ref. [3] and Fig. 3.2 above). This natural compensation behavior requires that we examine the competition between donor and acceptor defects for each prospective dopant to determine if it can feasibly be used to tune E_{F}.

The competition between positive (donor) and negative (acceptor) defects serves as the driving force in determining the equilibrium value of the Fermi level. The behavior comes from the linear dependence of formation energy on charge state multiplied with E_{F} (see Eq. 3.1) and applies very generally beyond just topological semimetals. A low value of E_{F} will favor the formation of additional donors that maintain charge balance by introducing free electrons and raising E_{F}, while a value that is too high will form more acceptors to lower it. Therefore, the point where the formation energies (as functions of E_{F}) of donor and acceptor defects intersect serves as an attractor for the equilibrium Fermi level. This implies that we can plot ΔH_{D} vs E_{F} for the donor and acceptor defects of each possible dopant, and the E_{F} where they intersect is the value the dopant drives the system toward in equilibrium. We refer to this intersection point as the defect (intrinsic) or dopant (extrinsic) compensation point and denote the value where it occurs as $E_{\mathrm{F}}^{\mathrm{dc}}$. Plots of the donor and acceptor formation energies and the resulting compensation points for the intrinsic Cd defects as well as for Au and Ge are shown in Fig. 3.3(a).

For Cd_3As_2, we aim to choose a dopant where the compensation point is at a lower E_{F} (closer to the Dirac point) than the Cd defects so the system will more readily allow neutral or *p*-type doping. The defect compensation point for the $\mathrm{Cd_i}$ and V_{Cd} is 0.14 eV and 0.20 eV above the Dirac point going for As-rich to Cd-rich growth respectively. Despite E_{F}^{dc} always lying over 100 meV above the Dirac point, *p*-type doping is still predicted for growth at a high enough temperature [3, 19]. However, in the current work we seek strategies for achieving neutral doping at lower growth temperatures (discussed in further detail below), so it is still essential that E_{F}^{dc} for the chosen dopants is lower than for the Cd defects.

In addition to requiring that the chosen dopant drives E_{F} closer to the Dirac point, we must also consider the dopant's solubility. This can be accomplished by further studying each dopant

compensation point. Specifically, we aim for dopants where the value of ΔH_{D} at the compensation point is low enough so that a reasonable number can form to compete with the Cd defects but without precipitating a competing phase. To summarize, by constructing a scatter plot of the dopant compensation points for each potential dopant on a graph of ΔH_{D} vs E_{F}, we can analyze both the potential impact on net doping (x-axis) and the solubility (y-axis) simultaneously.

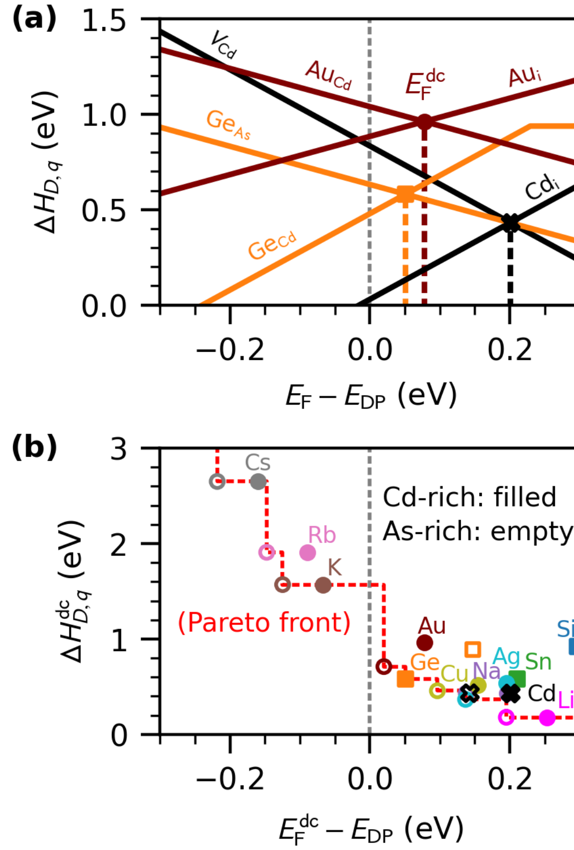


Figure 3.3: (a) Formation energies $\Delta H_{\mathrm{D},q}$ of donor and acceptor defects for Cd, Au, and Ge, as functions of Fermi level E_{F}. The intersection points between donors and acceptors are utilized in constructing (b) The two-objective dopant selection for lowering the Fermi level (low $E_{\mathrm{F}}^{\mathrm{dc}}$) with high dopant solubility (low ΔH_{D}). Considered are Group I (circles) and IV (squares) dopants under Cd- and As-rich conditions, in comparison to the intrinsic Cd defects (crosses). Au and Ge are chosen as the most promising extrinsic dopants on the Pareto frontier (red line).

Figure 3.3(b) shows the dopant compensation points between the primary donor and acceptor defects in Cd_3As_2 for Group I and Group IV dopants as well as for the intrinsic Cd defects. Because

acceptor doping for Group I involves substitution on Cd, those elements favor p-type doping under As-rich growth conditions much like the self-doping from Cd defects, whereas the Group IV elements have improved acceptor doping with Cd-rich growth. To determine which dopants should be considered for further study, we also construct a Pareto front (dashed red line) in Fig. 3.3(b). With the intended dual objective of the compensation point be both low in E_{F} and low in ΔH_{D}, dopants above or to the right of the Pareto front for both Cd-rich and As-rich growth can be disregarded. Therefore, with the requirement that the chosen dopants lie to the left of the Cd defects and because of the limited solubility of the heavy Group IA elements (K, Rb, and Cs), we choose Au from Group I and Ge from Group IV to study with the expanded defect equilibrium model.

3.5 Combined Defect and Dopant Redistributions

When external dopants are introduced, the effects of defect recombinations become more elaborate than for a single element, and constraining the chemical potentials to maintain the initial defect stoichiometries s_α becomes increasingly more complex for each additional element (α) considered. The increased complexity stems from the non-intuitive behavior of varying any given $\Delta\mu$ and how it impacts the defect concentrations. For example, in the case of Au-doped Cd_3As_2 changing $\Delta\mu_{\mathrm{Cd}}$ directly changes the formation energy of the $\mathrm{Cd_i}$, V_{Cd}, and $\mathrm{Au_{Cd}}$ through Eq. 3.1, but the resulting changes to the carrier concentrations and E_{F} then indirectly change the concentration of defects like the $\mathrm{Au_i}$ that do not explicitly depend on $\Delta\mu_{\mathrm{Cd}}$. This implicit dependence of defect concentrations (and combined defect stoichiometries) on the chemical potentials can be exceedingly complex as the results of changing one $\Delta\mu$ greatly depends on both its starting value and the current value of the others. Therefore, a different approach than the bisection method utilized for the Cd defects is necessary to extend the constrained equilibrium model to a system with several conflicting defect elements.

In the current work we employ an iterative Newton-like root-finding algorithm to simultaneously constrain all chemical potentials for the doped system. This is accomplished by defining a vector objective function to be minimized: the absolute difference between the growth stoichiometries and

the updated stoichiometries at lower temperatures (one component per element). At each iteration we update the estimated solution for the constrained chemical potentials based on the magnitude of the objective function and the numerical derivatives of s_α with respect to each $\Delta\mu$. Further details are given in the Methods section. We apply this approach to Cd_3As_2 doped with Au and Ge, which require constraining 2 and 3 chemical potentials respectively. As a result we are able to study, in addition to their effect on net doping, how Au and Ge impact the temperature dependent recombinations of the Cd defects when compared to the intrinsic defect case.

We also consider a modified constrained equilibrium where the initial (growth) dopant concentrations exceed the thermodynamic miscibility limit (i.e. supersaturated doping [43]. In simple cases where the dopant can assume only one lattice site, one could fix the respective concentration and solve the other defect concentrations as in the equilibrium case. However, dopants can often assume different configurations (e.g., Au_{Cd} and Au_i) whose ratios vary with temperature and growth conditions. Thus, we instead solve for a non-equilibrium chemical potential (e.g., $\Delta\mu_{Au}$) that results in the desired total atomic concentration (e.g., sum of Au_{Cd} + Au_i). That is, we choose equilibrium values for the chemical potentials of Cd and As, then we vary the chemical potential of the dopant via the bisection method to find the value that gives a growth concentration equal to a chosen target value, in analogy with the constant-N approach of Ref. [108]. After this supersaturated solution is found, we subsequently enforce the constrained equilibrium by updating the concentrations of all the defects to maintain the initial stoichiometries using the Newton method.

3.6 Cd_3As_2 Extrinsic Doping: Au

As discussed in Sec. 3.2 above, there is a narrow range of growth parameters with the potential to achieve net p-type doped Cd_3As_2. Specifically, As-rich conditions and a temperature above 700 K were needed, at which point the number of V_{Cd} (free holes) surpassed the number of Cd_i (free electrons). In the current work we aim to utilize the chosen extrinsic dopants to achieve net-neutral doping over a wider range of growth conditions, i.e. lower temperatures and chemical potentials closer to the Cd-rich limit. Expanding the accessibility of neutral doping to lower temperatures

is beneficial for two reasons. First, lower growth temperatures in general give rise to fewer overall defects, which can allow for improved electron mobilities and other electronic properties. Second, expanded control over net doping enhances the viability of experimental growth methods that benefit from lower temperatures, such as molecular beam epitaxy (MBE). Thin film growth with MBE has proved very valuable in the modern study of Cd_3As_2 [104, 124, 125, 133] but has an upper limit of $T_g \approx 500$ K owing to the volatility of the constituent atomic and molecular species, so it is sensible to explore strategies for Fermi level control at or below that temperature. Furthermore, the deposition control afforded by MBE allows for the possibility of supersaturated doping not feasible for higher temperature growth methods, which is why we also consider supersaturated initial growth here.

Because the primary Group I defects are substitution on Cd and occupation of the interstitial site and they do not reasonably substitute on As, the constrained equilibrium model for Au-doped Cd_3As_2 only involves constraining two chemical potentials: Cd and Au. The defect stoichiometry of As (s_{As}) can still be maintained by keeping the concentration of V_{As} held fixed as with the intrinsic defects discussed above, and to maintain the Cd and Au stoichiometries only the concentrations of defects involving them need to be dynamically updated to keep the sums

$$s_{\mathrm{Cd}} = c_{\mathrm{Cd_i}} - c_{V_{\mathrm{Cd}}} - c_{\mathrm{Au_{Cd}}}, \tag{3.3}$$

and

$$s_{\mathrm{Au}} = c_{\mathrm{Au_i}} + c_{\mathrm{Au_{Cd}}}, \tag{3.4}$$

equal to their growth values.

As a starting point before analyzing the defect redistributions, we plot both the initial defect concentrations $c_{\mathrm{D}}^{\mathrm{g}}$ and the net carrier concentration $c_{\mathrm{nc}} = n_{\mathrm{e}} - n_{\mathrm{h}}$ of Cd_3As_2 as functions of Au concentration for growth at 800 K and 400 K in Figs. 3.4(a) and (b) respectively. As discussed above, Au can substitute on both Cd and empty sites with a variable ratio, so the Au concentration is measured as a percent occupation of the total available atomic sites. For a topological semimetal such as Cd_3As_2, recall that we aim to tune the net carrier concentration close to zero so E_{F} approached

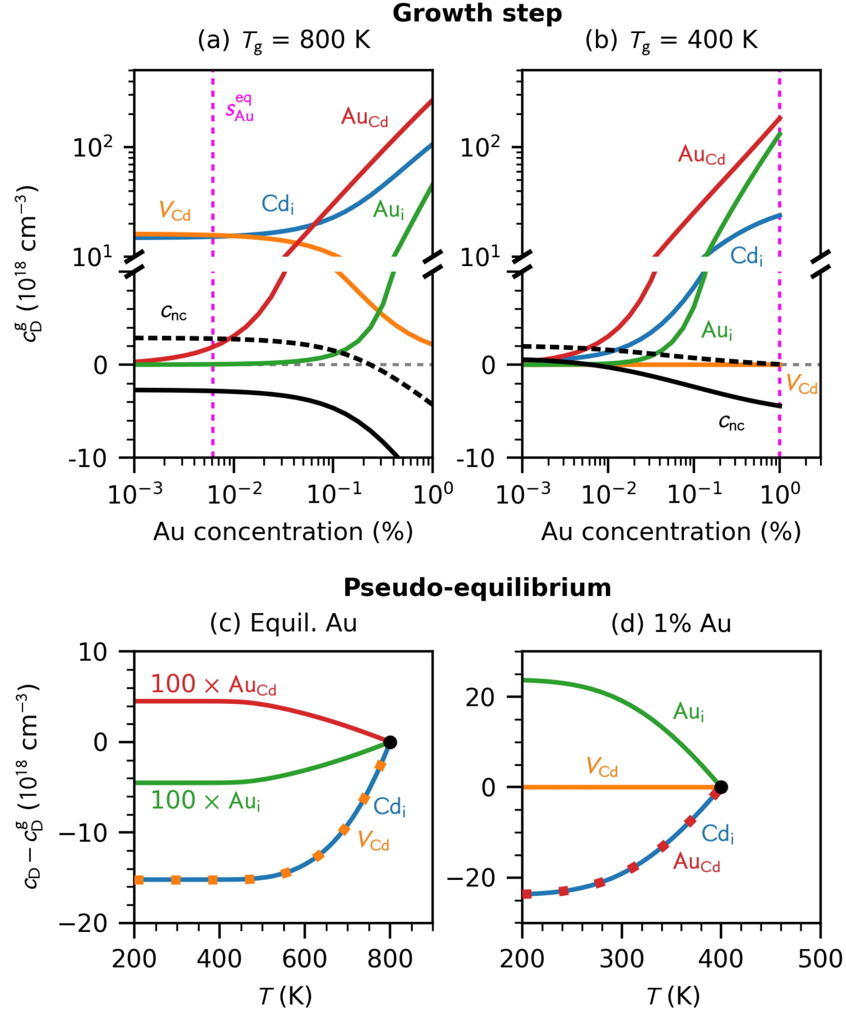


Figure 3.4: Initial defect concentrations $c_{\mathrm{D}}^{\mathrm{g}}$ and net carrier concentration c_{nc} in $\mathrm{Cd_3As_2}$ as functions of Au doping concentration for Cd-rich (dashed) and As-rich (solid) growth at (a) T_{g} = 800 K and (b) T_{g} = 400 K. The defect concentrations after growth are then measured from the initial growth values and plotted as a function of temperature T for As-rich growth at (c) 800 K with an equilibrium concentration of Au (1.95×10^{18} cm^{-3}) and (d) 400 K with 1% Au.

the Dirac point at low temperature. Because both the intrinsic (V_{Cd}) and extrinsic ($\mathrm{Au_{Cd}}$) acceptor defects favor As-rich conditions, c_{nc} is guaranteed to be lower for As-rich growth (solid line) when compared to Cd-rich (dashed). Furthermore, as more Au is introduced at either temperature the increased number of $\mathrm{Au_{Cd}}$ acceptors gives rise to greater number of compensating $\mathrm{Cd_i}$ and $\mathrm{Au_i}$ donors. The net effect is an overall decrease in c_{nc} with added Au, as suggested by the Au defects lying to the left of the Cd defects in Fig. 3.3 and therefore favoring a lower E_{F}. With this understanding of

the initial growth concentrations serving as a baseline, we then study the results of a subsequent constrained equilibrium in two regimes: one dominated by Cd defects and the other by Au.

Figure 3.4(c) shows the constrained concentrations of the Cd and Au defects for thermodynamic As-rich growth at 800 K—the same growth conditions that resulted in net p-type self-doped Cd_3As_2. The equilibrium concentration of Au is limited to 1.95×10^{18} cm^{-3} (0.0062% of Cd and interstitial sites) at this relatively high temperature. We find that the recombination of the Cd defects is largely unchanged compared to the intrinsic case (Fig. 3.1), and the redistribution of Au with decreasing temperature ($\mathrm{Au_i} \rightarrow \mathrm{Au_{Cd}}$) happens on a scale smaller by two orders of magnitude. This suggests that concentrations of Au above the miscibility limit are necessary to significantly alter the doping behavior resulting from the intrinsic Cd defects. However, metastable doping is far less practical at higher temperatures, so we further study supersaturated doping at the lower growth temperature to coincide with thin film growth as discussed above.

The results for a constrained equilibrium starting from As-rich growth at 400 K with 1% Au (3.14×10^{20} cm^{-3}) doping are shown in Fig. 3.4(d). In contrast to the equilibrium results at 800 K, we find that the primary defect redistribution as temperature decreases from 400 K is Cd moving from the interstitial sites to take the place of $\mathrm{Au_{Cd}}$ and the displaced Au moving from the Cd to the interstitial sites. The distinct behavior in the two regimes is understood in the following way. In Fig. 3.4(c), the dominant acceptor defect is the V_{Cd} and there is a resulting net cation deficiency. At the initial temperature of 800 K many Cd and Au atoms are excited to the energetically unfavorable interstitial site to reduce entropy through increased cation disorder. Then, after cooling under the compositional constraint (Eqs. 3.3 and 3.4), both Cd and Au relax to the excess Cd sites to reduce the energy. Whereas in Fig. 3.4(d), the dominant acceptor defect is the $\mathrm{Au_{Cd}}$ and there are very few V_{Cd} (about 3×10^{13} cm^{-3}) so there is a net excess of cations. Upon cooling from 400 K in this regime, energy is reduced by $\mathrm{Au_{Cd}}$ being replaced by the more favorable Cd and being expelled to the interstitial site. For intermediate Au concentrations there would be competition between the V_{Cd} dominated and $\mathrm{Au_{Cd}}$ dominated regimes.

3.7 Cd_3As_2 Extrinsic Doping: Ge

As discussed above in Sec. 3.4, Group IV elements like Ge can reasonably substitute on both As and Cd sites in Cd_3As_2, where it acts as a single acceptor and double donor, respectively (the $\mathrm{Ge_i}$ is not included because it is substantially higher in formation energy). This implies that in an isolated sample the concentration of As vacancies can no longer be held fixed, and the chemical potentials for all 3 elements (Cd, As, and Ge) need to be simultaneously constrained to maintain their stoichiometries. That is, we utilize the same strategy for modeling the temperature dependence of defect concentrations after growth, but now the sums we hold constant are

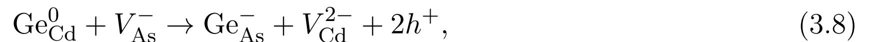
$$s_{\mathrm{Cd}} = c_{\mathrm{Cd_i}} - c_{V_{\mathrm{Cd}}} - c_{\mathrm{Ge_{Cd}}}, \tag{3.5}$$

$$s_{\mathrm{As}} = -c_{V_{\mathrm{As}}} - c_{\mathrm{Ge_{As}}}, \tag{3.6}$$

and

$$s_{\mathrm{Ge}} = c_{\mathrm{Ge_{Cd}}} + c_{\mathrm{Ge_{As}}}. \tag{3.7}$$

There are two additional ways Ge interacts with c_{nc} in Cd_3As_2 that distinguish it from the above results for Au. First, the sought after electron accepting defect is substitution on As rather than Cd, so it favors Cd-rich growth instead of As-rich. Consequently, when trying to achieve doping balance there will be competition between the intrinsic V_{Cd} and extrinsic $\mathrm{Ge_{As}}$ acceptors depending on the value for the Cd/As chemical potentials that was not relevant to Au doping. Second, the compensating $\mathrm{Ge_{Cd}}$ can have one of three different charge states. The fraction of $\mathrm{Ge_{Cd}}$ that has a given charge state changes with E_{F} and T according to Eq. (3.2), in contrast to the case of Au doping where each defect in Eqs. 3.3 and 3.4 only have one charge state. The net carrier concentration can therefore also change with temperature as a result of defects moving between atomic sites or changing their charge state. For example, extra holes can be introduced as a Ge atom moves from a Cd to As site,



$$\mathrm{Ge_{Cd}^{0}} + V_{\mathrm{As}}^{-} \rightarrow \mathrm{Ge_{As}^{-}} + V_{\mathrm{Cd}}^{2-} + 2h^{+}, \tag{3.8}$$

or additional electrons can be introduced by an on-site charge state transition for a $\mathrm{Ge_{Cd}}$,

$$\mathrm{Ge}^0_{\mathrm{Cd}} \rightarrow \mathrm{Ge}^{2+}_{\mathrm{Cd}} + 2e^-. \tag{3.9}$$

The combination of these two processes and their competition gives rise to a nontrivial temperature dependence of c_{nc} that will be explored presently.

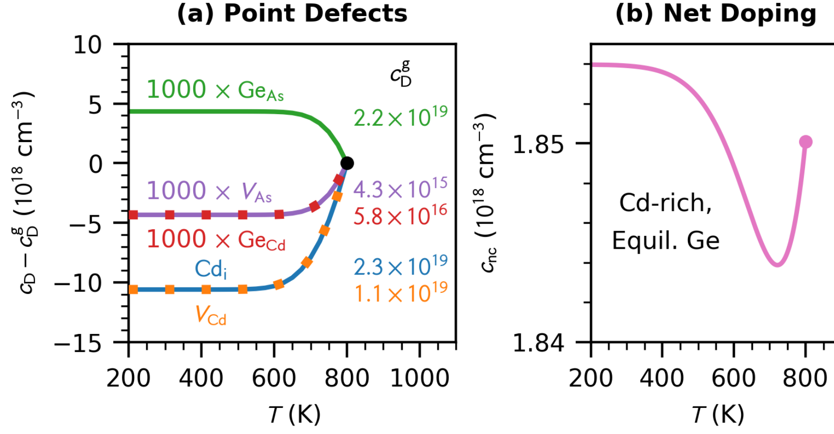


Figure 3.5: (a) Defect concentrations c_{D} for Ge-doped $\mathrm{Cd_3As_2}$ measured from the indicated initial growth values $c^{\mathrm{g}}_{\mathrm{D}}$. (b) The associated net carrier concentration c_{nc}, both plotted as functions of temperature T. Results are shown for Cd-rich growth at 800 K with an equilibrium concentration of Ge (2.21×10^{19} cm^{-3}).

As above for Au, we study Ge equilibrium solubility at the higher temperature $T_{\mathrm{g}} = 800$ K, and then we consider supersaturated doping at a lower temperature where we treat the total Ge concentration as a variable. Figure 3.5 illustrates the redistribution of Ge-doped $\mathrm{Cd_3As_2}$ defects and the temperature dependent net carrier concentration for Cd-rich growth at 800 K. Similar to the results for thermodynamic Au doping, the recombination of the Cd defects is largely unaffected for an equilibrium concentration of Ge (2.21×10^{19} cm^{-3}, 0.056% of Cd and As sites). In fact, the Ge defect concentrations change on a three orders of magnitude smaller scale to the Cd defects despite the initial $\mathrm{Ge_{As}}$ concentration being very similar to the $\mathrm{Cd_i}$ and V_{Cd}.

The net carrier concentration in Fig. 3.5(b), as just discussed, has a nontrivial temperature dependence. In fact, we find c_{nc} varies non-monotonically with temperature. At higher temperatures

(around 800 $\rightarrow$ 700 K) $c_{\rm nc}$ decreases as you cool down because of Ge moving from Cd to As sites (e.g. Eq. 3.8). Then, at lower temperatures (<700 K), a majority of the remaining $\rm Ge_{Cd}$ that did not move to As sites undergo the on-site transition from neutral to charge +2 (Eq. 3.9), so $c_{\rm nc}$ increases. These competing effects from have little overall impact on net doping (see the scale of the vertical axis in Fig. 3.5b), but in other systems beside Cd_3As_2 defects with variable charge states could *substantially* alter net doping from high to low temperatures.

We also examine supersaturated Ge doping concentrations that are potentially accessible by non-equilirbium growth methods in the same manner as for Au. This situation involves lower temperatures and much larger dopant concentrations, but we find again that the redistribution of Ge and changes in net carrier density are not significant. Introducing more Ge increases the concentrations of $\rm Ge_{As}$ and $\rm Ge_{Cd}$, but it does not affect the preexisting small concentration of $V_{\rm As}$. The redistribution of Ge from Cd to As is then greatly restricted because there are so few empty As sites to fill, thereby making the effect minuscule compared to that of the intrinsic Cd defects. This is an example of the general feature of fixed stoichiometry on a finite lattice that extends beyond Cd_3As_2 as introduced in Sec. 3.1: the degree to which defect concentrations change as a function of temperature is bounded by the lowest concentration defect. That is, if a given defect has a restrictively small concentration during growth, redistribution events that involve that defect are limited in number to be less than or equal to that concentration.

As mentioned above, the $V_{\rm Cd}$ and $\rm Ge_{As}$ favor growth at opposing values of the Cd and As chemical potentials. The competition between intrinsic and extrinsic acceptor defects benefiting from As-rich and Cd-rich conditions respectively suggests we should consider the full range of the chemical potentials and not just the endpoints. Therefore, we plot the net carrier concentration as a simultaneous function of both the Cd/As chemical potentials and the Ge concentration, and the resulting contour plot is shown in Fig. 3.6. The carrier concentrations are measured in the low temperature limit ($T \rightarrow 0$). The bottom edge of the plot corresponds to the system being dominated by the intrinsic defects. In that regime $c_{\rm nc}$ is reduced for more As-rich growth, but as found previously there is no potential for tuning $c_{\rm nc}$ to 0 so $E_{\rm F}$ lies at the Dirac point at low T. By

contrast, the region dominated by Ge doping (above around 0.1%) has the behavior inverted, and c_{nc} slopes downward when moving from As-rich to Cd-rich. We find it is indeed feasible to achieve net neutral doping with Ge doping, and in the limiting case of Cd-rich growth p-type doping is found for concentrations above 0.2% Ge.

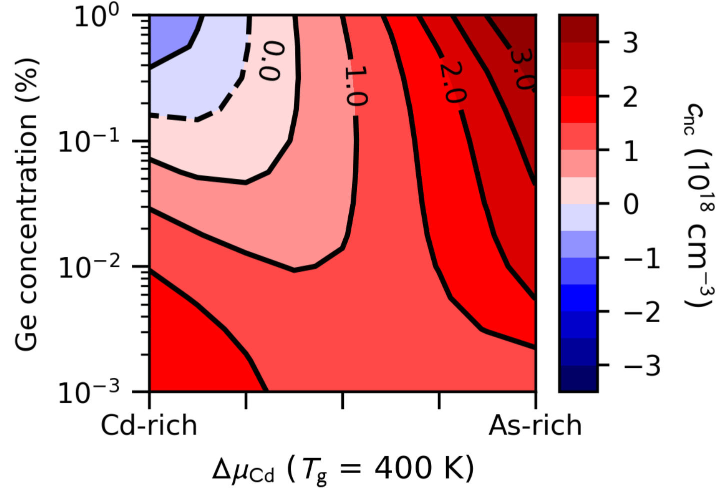


Figure 3.6: Low temperature ($T \to 0$) net carrier concentration c_{nc} of Cd_3As_2 grown at $T_{\mathrm{g}} = 400$ K, plotted as a function of both the Ge doping concentration and Cd chemical potential $\Delta\mu_{\mathrm{Cd}}$. Net n-type doping is colored red and net p-type is colored blue.

3.8 Discussion and Conclusions

We have developed here a new, generalizable approach to first-principles calculations of defect equilibria in solid state systems. By constraining the chemical potentials of all elements associated with the point defects being considered, we are able to model defect redistributions and recombinations at intermediate temperatures via a pseudo-equilibrium. The method can be applied outside of topological semimetals to any material where short-range diffusion is expected to persist at storage or operating temperatures after initial growth. At the same time, we have demonstrated that the intersection between donor and acceptor defects on a plot of formation energy vs. E_{F} (referred to here as the defect or dopant compensation point) is very useful in determining which extrinsic dopants are good candidates for Fermi level control. That is, the dopant compensation

point simultaneously provides information about the value of E_F the dopant drives the system toward and the dopant's solubility. This dual optimization can be utilized to select ideal extrinsic candidates in other systems with undesirable self-doping.

After applying these strategies to the Dirac semimetal Cd_3As_2, we find that both Au and Ge have the potential to overcome the self-doping of Cd defects and attain net-neutral or *p*-type doping at lower growth temperatures, with Au preferring As-rich growth and Ge preferring Cd-rich. Less than 0.01% Au or less than 0.2% Ge (in their preferred regimes) is predicted to tune the low-temperature E_F close to the Dirac point. For Au doping, the redistribution of defects was contingent on which acceptor defect was dominant during growth: $Au_i \rightarrow Au_{Cd}$ for excess V_{Cd} and $Au_{Cd} \rightarrow Au_i$ for excess Au_{Cd}. Ge doping required the extra consideration of defects with variable charge states, so the net carrier concentration (in addition to the defect concentrations) can vary with temperature. The changes in c_{nc} are mostly inconsequential in the case of Cd_3As_2, but the combination of defect charge transitions and redistributions could have a significant effect in other systems.

3.9 Methods

3.9.1 First-principles Total Energy Calculations and Defect Equilibria

In this work, the total energies of host and defect supercells of Cd_3As_2 were calculated using Density Functional Theory (DFT) and the Projector Augmented Wave (PAW) method as implemented in the VASP code [17, 69]. All of the first-principles calculations used a plane wave basis set with cutoff energy 380 eV, version 5.4 of the PAW-PBE pseudopotentials, and the atomic positions of each defect supercell were relaxed with an atomic force convergence criterion of 0.01 eV/Å. We used the Strongly Constrained and Appropriately Normed (SCAN) meta-Generalized Gradient Approximation (meta-GGA) functional [147] with the inclusion of spin-orbit coupling and a 4×4×4 Γ-centered k-mesh for accurate structure prediction [162]. The charged defect formation energies were calculated from total energies of the 80-atom primitive cell (space group $I4_1/acd$),

and we included corrections from the potential alignment between the host and defect cells [80]. Furthermore, we included a correction to the energy of the Dirac point, E_{DP}, based on an eigenvalue self-consistent GW calculation [72, 138].

The primitive cell has 3 distinct Wyckoff sites for Cd and 3 for As, so we considered defect substitutions on each unique site, and those with the lowest energy are shown in Fig. 3.3. The Cd-rich and As-rich limits for thermodynamic growth are set by the enthalpy of formation for Cd_3As_2, $\Delta H_{\text{f}} = 3\Delta\mu_{\text{Cd}} + 2\Delta\mu_{\text{As}} = -0.59$ eV/formula unit, as well as the competing solid-state phases for Cd and As. Specifically, phase coexistence with Cd metal gives the Cd-rich limit, $\Delta\mu_{\text{Cd}} = 0$ and $\Delta\mu_{\text{As}} = -0.30$ eV, and the As-rich limit $\Delta\mu_{\text{Cd}} = \Delta\mu_{\text{As}} = -0.12$ eV comes from coexistence with CdAs_2.

The calculations for defect equilibria were performed in the same manner as in previous work. For any given choice of chemical potentials, we find a self-consistent solution for the defect densities $c_{\text{D}],q}$ and the Fermi level E_{F} with charge balanced between defect charges and free carriers. This requires integrating the density of states (DOS) weighted with the Fermi-Dirac distribution to capture how the carrier concentrations change E_{F} by occupying the band continuum. The integration was facilitated by a model continuous function for the DOS, which was obtained by numerically fitting to a high-resolution Quasiparticle Self-consistent GW (QSGW) calculation performed with the Questaal code [68, 113, 130]. For further details, see the main text and Supplemental of Ref. [19].

3.9.2 Constrained Equilibrium and Temperature Dependence of Defect Concentrations with Competing Elements

The supersaturated doping and constrained equilibrium calculations herein utilize the same self-consistent equilibrium method just described, but the inputs are no longer thermodynamic chemical potentials. Instead, the inputs are artificial chemical potentials that correspond to a target doping concentration or fixed elemental stoichiometries. The complexity in finding the solution for these constrained chemical potentials scales with the number of elements being held fixed. To determine the chemical potential when only one element is being constrained, e.g. when only

considering the recombination of Cd defects or when studying supersaturated doping in otherwise equilibrium growth, we utilize the bisection method. That is, we define a range of possible values of $\Delta\mu_\alpha$, evaluate a defect equilibrium with the midpoint value, update the range so the old midpoint becomes the new lower or upper bound depending on if the element was deficient or in excess, and then iterate until the combined atomic concentration of that element is within a chosen tolerance (fractional error 10^6) from the target value. The process can be repeated for many temperatures below T_g to build up the temperature dependent concentrations of defects involving that element. Constraining the amount of several elements simultaneously, on the other hand, is more involved and requires special consideration.

If the set of atomic chemical potentials $\{\Delta\mu_\alpha\}$ are taken as variable inputs in a defect equilibrium calculation, then any given choice of that input gives as an output the set of defect concentrations $\{c_\mathrm{D}\}$ for all defects that involve any of the atomic species being considered. The output concentrations can then be alternatively expressed in terms of the combined defect stoichiometry for each element by adding and subtracting the contributions from each defect. That is, the defect stoichiometry s_α of species α is defined as

$$s_\alpha = -\sum_{\mathrm{D}(\alpha)} n_\alpha c_{\mathrm{D}(\alpha)}, \tag{3.10}$$

where $\mathrm{D}(\alpha)$ are all defects that involve species α (with concentrations $c_{\mathrm{D}(\alpha)}$), n_α is the same as in Eq. (3.1), and there is also an implied sum over charge state q. The minus sign in Eq. (3.10) is included to reverse the conventional sign of n_α defined above, so the stoichiometry is positive if the atomic species is in excess and negative if it is deficient. For systems such as Cd_3As_2 with several unique Wyckoff positions for a given element, substitutions involving each Wyckoff site can be considered distinct defects with different formation energies and concentrations. In that case, Eq. (3.10) also includes a sum over those nonequivalent symmetry sites. With this definition, the constrained equilibrium system can be modeled by finding the set of $\Delta\mu_\alpha^\mathrm{neq.}$ at each temperature below T_g that gives the same set of s_α as is found during growth. We add the "neq." superscript to

the chemical potentials to indicate they correspond to the constrained system solution rather than a true thermodynamic equilibrium.

The numerical approach employed herein to solve for the constrained equilibrium chemical potentials is a Newton-like root-finding method that iteratively updates the solution with successive linear corrections. We start with a defect equilibrium calculated in the usual way with a model growth temperature and choice for the chemical potentials. We then evaluate the resulting defect stoichiometries for those growth conditions, which we call s^*_α, so the constrained equilibrium condition for an isolated sample at temperature T is $s_\alpha(T) = s^*_\alpha$. If the starting estimate for the chemical potentials gives stoichiometries s^0_α, and the desired change to the stoichiometries is Δs_α such that $s^0_\alpha + \Delta s_\alpha - s^*_\alpha = 0$, then to linear approximation the solution can be updated as

$$\Delta s_\alpha = D_{\alpha\beta}\delta(\Delta\mu^{\text{neq.}}_\beta) = s^*_\alpha - s^0_\alpha, \tag{3.11}$$

where $\delta(\Delta\mu^{\text{neq.}}_\alpha)$ is the correction to the estimate of the chemical potentials and $D_{\alpha\beta} = \partial s_\alpha/\partial(\Delta\mu^{\text{neq.}}_\beta)$ is the (square) Jacobian matrix which can be evaluated numerically. Multiplying Eq. (3.11) by the inverse of $D_{\alpha\beta}$ on the left gives the update to the solution as

$$\delta(\Delta\mu^{\text{neq.}}_\alpha) = (D_{\alpha\beta})^{-1}(s^*_\beta - s^0_\beta). \tag{3.12}$$

The updated chemical potentials can then be used to find a new estimate s_α, and the process can be iterated until the stoichiometries are within a chosen tolerance of s^*_α, chosen here to be $|s_\alpha - s^*_\alpha| \leq 10^{-6}\ s^*_\alpha$.

We have had success with this computational approach, but it is not without challenges. The method makes use of iterative linear corrections to the chemical potentials, but the stoichiometries can have a very nonlinear dependence on those chemical potentials. If E_F was fixed and the element α only had one defect, then s_α would vary exactly exponentially with $\Delta\mu^{\text{neq.}}_\alpha$ (cf. Eqs. 3.1 and 3.2). In that case successive linear corrections work quite consistently, and using the natural log of the

stoichiometries in Eq. (3.12) would converge in one iteration. However, because the stoichiometry instead involves a sum over many competing exponentials, the dependence on chemical potentials becomes very complex and can cause numerical instabilities. Specifically, the chemical potentials can diverge if the initial estimate is far away from the true solution, or they can oscillate around the solution as is common for Newton-like methods. We implement two strategies to manage these instabilities. First, we implement the method at consecutive lower temperatures in steps of 5 K, and we use the solution from the previous temperature as the initial guess for the next. This allows the estimate for the chemical potentials to never start far away from the solution at low temperatures. Second, we multiply the right hand side of Eq. (3.12) by a dynamic damping factor. For each iteration we start the damping factor at a value of 1, then reduce it until the update to all chemical potentials $\delta(\Delta\mu_\alpha^{\mathrm{neq.}})$ is less than 0.01 eV to help prevent the solution from diverging. With the addition of these strategies we were able to consistently implement the Newton method and find the temperature dependent defect concentrations presented herein.

3.9.3 Simulated Diffusion with the Nudged Elastic Band Method

The model for the redistribution and recombination of defects in this article is predicated on the assumption that the defects can undergo short-range diffusion at intermediate or low temperatures. For completeness, we include here simulated diffusion calculations for the defect redistributions observed in the constrained equilibrium results to demonstrate that they are indeed energetically allowed. Specifically, we consider the intrinsic recombination of the $\mathrm{Cd_i}$ and V_{Cd}, a Au atom moving between the interstitial site and a neighboring V_{Cd} site, and Ge moving from a Cd site to a vacant As site, with the lowest energy Wyckoff sites chosen for each defect. We utilized the Nudged Elastic Band (NEB) method [62, 97] to study these diffusion paths in a charge neutral supercell within the framework of our first-principles calculations. For each diffusion path, we created 3 intermediate images with the aid of the Atomic Simulation Environment (ASE) library [81], and then we relaxed the images with the same DFT parameters used for point defects described above. The results for these NEB calculations are shown in Figure 3.7.

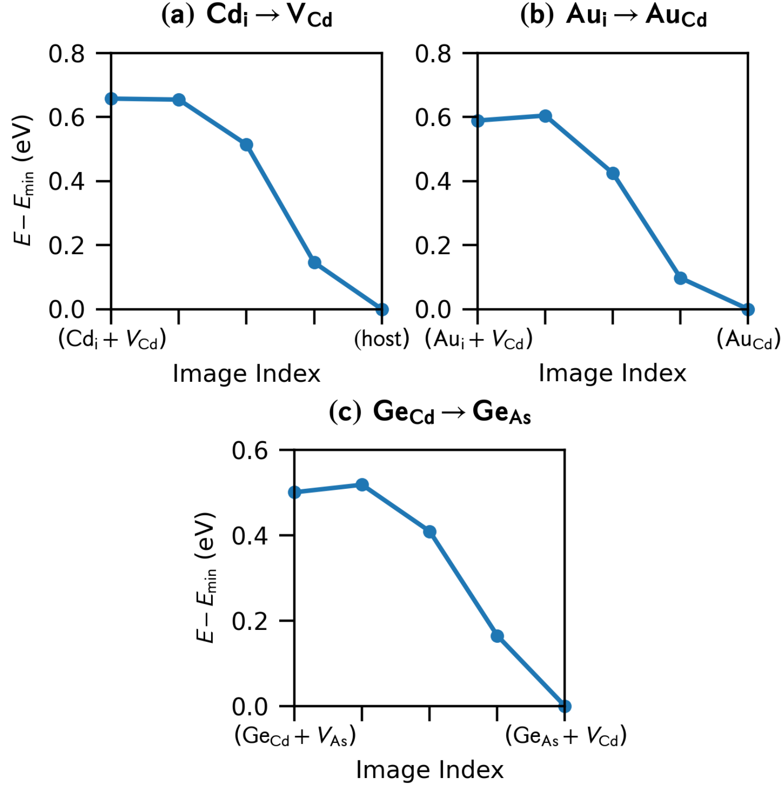


Figure 3.7: Nudged Elastic Band supercell energies E plotted as a function of the image index for the simulated diffusion paths of (a) the recombination of a $\mathrm{Cd_i}$ and V_{Cd}, (b) Au moving from an interstitial to a neighboring vacant Cd site, and (c) Ge moving from a Cd site to a neighboring As site. Energies are measured with respect to the minimum for each diffusion.

We find that for all three paths considered, there is a very small or nonexistent energy barrier for the atoms to move from the high energy state to the low energy state. At lower temperatures the $\mathrm{Cd_i}$ will therefore freely and spontaneously diffuse and fill the V_{Cd} to reduce the total energy, and similarly for $\mathrm{Au_i}$ moving to the Cd site and Ge moving from Cd to As. These results coincide the temperature dependent concentrations presented in Figs. 3.1, 3.4(a), and 3.5(a). The inverted behavior of Au in Fig. 3.4(b), as discussed above, is understood to be due to there being a greater energy gain in Cd occupying its home site than Au occupying it. There is a difference of around 50 meV/atom between the recombination of Cd defects and the redistribution of Au, so even though Au prefers to the Cd site, it is an overall energy gain for it to move to an interstitial site and be replaced with a nearby Cd atom.

An aspect of defect mobility after growth not explicitly studied herein is the time scale on which the point defects can move throughout the material to be close enough to interact or recombine with one another. We have found that there is little to no barrier for the atoms in adjacent defects to move between them to reduce the overall energy, but we did not study the extended diffusion processes of defects moving across several unit cells. Such an analysis would require a detailed kinetic model with effective rate constants for all possible diffusion processes and is beyond the scope of the current work.

Chapter 4

Heterostructural alloying and phase diagram construction for $(Cd_{1-x}Zn_x)_3As_2$

This chapter is adapted from Ref. [18] and its Supplemental Material: Copyright 2024 by the American Physical Society.

Abstract. Alloying the topological semimetal Cd_3As_2 with Zn_3As_2 provides a potential route for controlling the electronic properties. We predict the alloy phase diagram from first-principles calculations, considering that both end members have a crystal structure derived from the antifluorite lattice, but with different arrangements of the unoccupied cation sites. To overcome the limitations of the regular solution approximation and to include short-range order effects, we perform Monte Carlo simulations, parameterize the temperature dependence of the mixing enthalpy ΔH_{m}, and perform thermodynamic integration of the free energy. The resulting phase diagram exhibits features that are unique to heterostructural alloy systems and provides computational predictions of solubility limits and composition ranges that are stable against spinodal decomposition.

4.1 Introduction

Alloys formed between compounds with different crystal structures, commonly known as heterostructural alloys, are a rapidly evolving means of advancement in solid-state material synthesis. This unique class of mixtures has many applications, including fine-tuning of material properties [131], fabricating metastable materials at compositions that are otherwise inaccessible [54], and creating negative-pressure polymorphs [142]. One of the most important aspects of studying heterostructural alloys is determining the composition dependence of the alloy mixing enthalpies in

the different possible crystal structures. First-principles density functional theory (DFT) and cluster expansion methods have been very successful in modeling the formation of both isostructural and heterostructural alloys and sampling their configuration space [8, 14, 86, 100, 127, 135, 137, 143, 148]. In this work, we apply first-principles Monte Carlo sampling to study the combined effects of short-range order (SRO) and substitutions on several nonequivalent symmetry sites, and we utilize a thermodynamic integration to fully describe the temperature and composition dependence of alloy free energies.

Cd_3As_2 is a three-dimensional Dirac semimetal with sustained scientific interest. However, intrinsic point defects, recently found to be predominantly Cd interstitials [19], induce n-type doping and an elevated Fermi level after growth under typical synthesis conditions [19, 28, 59, 104, 129]. This unintentional self-doping effect inhibits access to the unique topological band structure features of Cd_3As_2. Conversely, Zn_3As_2 is a trivial semiconductor with a direct band gap of 1.0 eV and typically forms as p-type [149]. The alloy between these two compounds, $(Cd_{1-x}Zn_x)_3As_2$, is then expected to undergo both a crossover from n- to p-type and a topological phase transition from Dirac semimetal to semiconductor with increasing x. Both of these features were observed experimentally [90, 107, 126]. The purpose of the present work is to determine a detailed model for the phase behavior as a function of temperature and composition. Furthermore, the present results also provide representative atomic structure models for future electronic structure and doping studies.

The ground state of Cd_3As_2 is a body-centered tetragonal structure (space group 142, $I4_1/acd$) [4] with three distinct Wyckoff positions for both cations and anions. Zn_3As_2 was observed with this same structure [167] as well as a related primitive tetragonal structure (space group 137, $P4_2/nmc$) [166]. Both structures are obtained by constructing supercells of an anti-fluorite lattice with an ordered set of cation empty sites [4, 29], and the only difference between the two is in the ordered stacking of these empty sites, i.e., they can be viewed as different polytypes. For our choice of DFT functional, namely, the strongly constrained and appropriately normed meta generalized gradient approximation (SCAN meta-GGA) [147], we find $P4_2/nmc$ to be the lower energy structure for Zn_3As_2, so we consider it to be the ground state herein. The $(Cd_{1-x}Zn_x)_3As_2$ alloy was previously

observed in this structure [167], but has also recently been seen to transition from body-centered to primitive tetragonal and then back to body-centered as x is increased [90].

With the backdrop of the currently somewhat inconclusive experimental picture, we here study this heterostructural alloy system by performing first-principles Metropolis Monte Carlo (MC) sampling in both crystal structures. We develop a model that captures temperature-dependent SRO effects in the enthalpy of mixing between the two compounds and then use this model to predict the alloy phase diagram with the crossover of the preferred crystal structure and with the solubility limits as functions of temperature and composition.

4.2 Methods

We obtained DFT total energies of alloy supercells with the projector augmented wave (PAW) method [17] as implemented in the VASP code [69]. While the MC energy sampling made use of a DFT based MC workflow [137] with a standard GGA functional [120], calculations of the final configuration energies used the SCAN functional with spin-orbit coupling, which is known to give improved energy differences between atomic configurations [162] and has been found to more accurately describe Cd_3As_2 [19]. Additional details on DFT calculations are given in Sec. 4.2.1. Common tangent construction between free energy plots was performed using the optimize package in SciPy [153].

4.2.1 Compuational details

For all total energy calculations we used a plane wave basis set with version 5.4 of the PAW-PBE pseudopotential data sets. For the Monte Carlo (MC) sampling and simulated annealing we used the Generalized Gradient Approximation (GGA) of Ref. [120] without spin-orbit coupling, a cutoff energy of 300 eV, and a 2×2×2 Monkhorst-Pack k-mesh. The atomic positions of each proposed MC step were relaxed with a force convergence requirement of 0.01 eV/Å after swapping one pair of Cd and Zn sites. For the calculations of final configurations at each temperature, we used the Strongly Constrained and Appropriately Normed (SCAN) meta-GGA functional [147], including

spin-orbit coupling, a cutoff energy of 380 eV, and a 4×4×4 Γ-centered k-mesh.

4.3 First-principles Monte Carlo

In preliminary calculations we found that the Zn substitution energies vary markedly among the three nonequivalent Cd sites of the Cd_3As_2 ground state structure. Additionally, we observed significant changes in alloy total energies when varying cation configurations for both the Cd_3As_2 and Zn_3As_2 structures, indicating considerable SRO effects. Both of these contributions give rise to non-random distributions of cations within the alloy. Therefore, instead of modeling a random alloy—e.g., with special quasirandom structures—we performed MC simulations which are capable of capturing the temperature-dependent effects of SRO and yielding more realistic alloy configurations. While SRO is expected to be less pronounced for the isovalent substitutions we consider here, compared to, e.g., the exceedingly strong effects in heterovalent alloys [111], we find it is strong enough that it should be taken into account.

As initial configurations for the MC simulations, we constructed 80-atom supercells for both the $I4_1/acd$ and $P4_2/nmc$ structures over the full $(Cd_{1-x}Zn_x)_3As_2$ alloy composition range in steps of $\Delta x = 0.125$. Here, the lattice constants are approximated by Vegard's Law, i.e.

$$a_{\text{alloy}} = (1-x)\, a_{\text{Cd}_3\text{As}_2} + x\, a_{\text{Zn}_3\text{As}_2}, \tag{4.1}$$

and the cation configurations are sampled by 10 explicitly random Cd/Zn occupations for each composition, corresponding to an initial $T = \infty$ infinite temperature limit. For each of the initial configurations, we performed independent MC equilibrations in the canonical NVT ensemble. Specifically, the cation distribution is sampled by mutual swapping of randomly chosen Cd and Zn sites, while using the resulting DFT energy after atomic-force relaxation to accept or reject the trial configuration in accordance with the Boltzmann probability distribution.

To obtain configurations associated with temperatures in the range of interest, we performed simulated annealing [64] by sequentially equilibrating the MC simulations at 800, 600, and 400 K,

i.e., corresponding to Cd_3As_2 growth temperatures accessible by thin film and bulk synthesis as in previous work [19]. Averaging over ten independent runs for each composition allows for the determination of an effective ensemble average of the energy E_x of the alloy at a given temperature, noting that the energy E is virtually identical to the enthalpy H at ambient pressure in the solid state. The present model maintains the sublattice of unoccupied sites of the anti-fluorite lattice in either polytype. Disorder in the distribution of the empty sites is in principle also possible, presumably being responsible for the cubic phase transition at 900 to 1000 K [122], but it is expected to be of minor importance in the lower temperature range of interest here.

The resulting average energies for the simulated annealing of $(Cd_{0.5}Zn_{0.5})_3As_2$ with the Cd_3As_2 crystal structure are shown in Fig. 4.1. Energies are measured with respect to the average of the initial random configurations. Short-range order at finite temperature gives rise to a significant decrease in energy from the fully randomized starting configuration, i.e. around 40 meV per formula unit (f.u.) at 400 K. Furthermore, the systematic energy changes between finite temperatures suggest a smooth variation as a function of temperature. Because SRO clearly has a significant impact on the average energy of the alloys, it is therefore essential that a model of the alloy mixing enthalpy account for this temperature dependence.

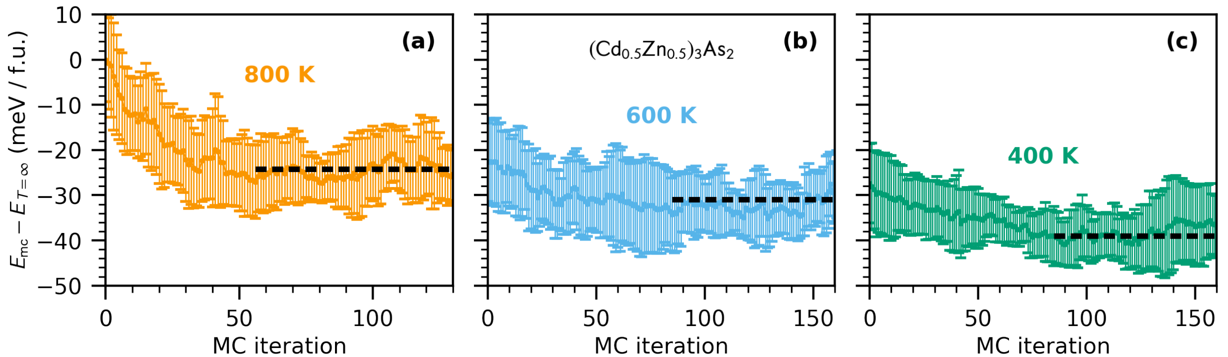


Figure 4.1: Monte Carlo simulated annealing of $(Cd_{0.5}Zn_{0.5})_3As_2$ for the Cd_3As_2 ground-state crystal structure, averaged over ten independent samples and equilibrated sequentially at (a) 800, (b) 600, and (c) 400 K. Energies are measured with respect to the average of the initial random ($T = \infty$) configurations. Error bars indicate the standard deviation across the independent MC runs at each iteration.

4.3.1 Probabilities in the chosen Monte Carlo ensemble

As just described above, the specific ensemble considered for our Monte Carlo simulations of the $(Cd_{1-x}Zn_x)_3As_2$ alloy is the canonical NVT ensemble, because when there is a swap between cation (Cd or Zn) sites, the number of each cation, the cell volume, and the temperature are held fixed. In the NVT ensemble, the probability P of a given microstate (in this case a unique arrangement of cations for fixed number of Cd/Zn) with energy E is given by

$$P = \frac{1}{Z} e^{-E/(k_{\mathrm{B}}T)}, \tag{4.2}$$

where the partition function Z is defined by a sum of the Boltzmann factor over all microstates:

$$Z = \sum_i e^{-E_i/(k_{\mathrm{B}}T)}. \tag{4.3}$$

The enormous number of cation configurations that would need to be calculated to determine Z makes it computationally impractical, so it is far more prudent to perform the MC simulations. Despite not explicitly calculating the partition function, we still make use of Boltzmann statistics by considering the ratio of probabilities between two microstates at each MC step, which does not depend on Z. That is, the ratio of probabilities between two cation configurations is

$$\frac{P_2}{P_1} = \exp\left(\frac{E_1 - E_2}{k_{\mathrm{B}}T}\right), \tag{4.4}$$

which means we can use only the energy difference between the two configurations ($E_1 - E_2$) to model the NVT ensemble. At each MC step we accept or reject potential cation swaps according to Eq. 4.4.

To future students, the process is as follows. Calculate the supercell energy of the current configuration, which is E_1; propose a swap between a randomly chosen pair of Cd and Zn sites, relax the atomic positions (with cell volume fixed), and calculate the total energy of this new possible

configuration, E_2; if $E_2 < E_1$, allow the swap and repeat with the new configuration replacing the old one for the next iteration; if $E_2 > E_1$, calculate the right hand side of Eq. 4.4, generate a random number between 0 and 1 from a uniform distribution, and then compare the two; if the random number is less than the calculated probability, allow the swap and replace the old configuration with the new; otherwise, reject the swap and maintain the old configuration for the next iteration.

4.4 Thermodynamic free energy integration

A common starting point for describing alloy thermodynamics is the regular solution model, where the configurational entropy of mixing is assumed to be equal to the limit of ideal random site occupations, and the mixing enthalpy is considered to be independent of the temperature [52, 145]. The ideal random alloy configurational entropy is given by

$$\Delta S_{\mathrm{m}}(T=\infty) = -k_{\mathrm{B}}[x\ln(x) + (1-x)\ln(1-x)], \tag{4.5}$$

where x and $(1-x)$ are the fractional compositions of the competing atomic species. The enthalpy of mixing is typically modeled as a quadratic function of the alloy composition with a constant bowing parameter, i.e., $\Delta H_{\mathrm{m}} = \Omega x(1-x)$ [145]. We define here a model that incorporates a temperature dependence of the mixing enthalpy through the interaction parameter $\Omega(T)$. Specifically, with the relative polytype energies E_0 and E_1 of the end compounds included, the mixing enthalpy is given by

$$\Delta H_{\mathrm{m}}(T) = \Omega(T) \cdot x(1-x) + (1-x)E_0 + xE_1. \tag{4.6}$$

For each of the two polytypes, we fit bowing parameters $\Omega(T)$ to the mixing enthalpies of the initial random configurations as well as those resulting from the MC equilibrations at the different temperatures. Figure 4.2 shows the raw mixing enthalpy data and the quadratic fits to Eq. (4.6) as a function of composition. Data are shown for infinite temperature (random seed) as well as 800, 600, and 400 K. We tested the possibility of an asymmetric shape of ΔH_{m} by fitting third-order polynomials, but found insufficient statistical significance. Comparing Figs. 4.2(a) and 4.2(b), we

observe similar SRO effects in both structures. The magnitude of Ω at each temperature is large compared to the energy differences between the two polytypes at the endpoints: $E_1 = 5.0$ meV/f.u. for $I4_1/acd$ and $E_0 = 7.3$ meV/f.u. for $P4_2/nmc$. This comparably small energy difference between the two structures for both compounds is attributed to the two structures only being differentiated by the stacking of cation empty sites as discussed above.

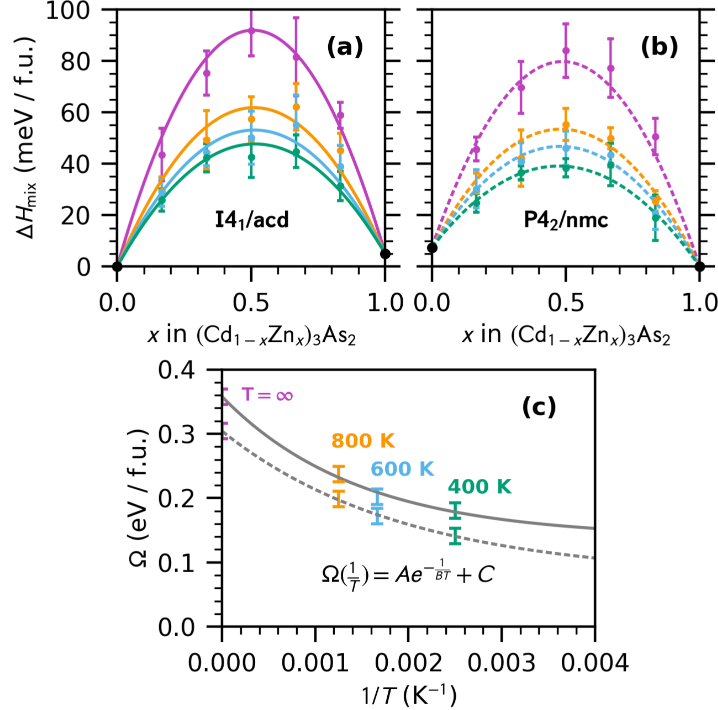


Figure 4.2: Temperature dependent alloy mixing enthalpy ΔH_{m} data and quadratic fits for Monte Carlo simulations of (a) space group $I4_1/acd$ (solid) and (b) space group $P4_2/nmc$ (dashed) as functions of composition x. Energies are measured with respect to the respective ground states at the end-point compositions ($x = 0$ and $x = 1$). (c) Discrete MC data and exponential decay function fit for the quadratic bowing parameter Ω of both structures as a function of reciprocal temperature.

Construction of a heterostructural alloy phase diagram requires a model for the free energy of mixing at all temperatures (see below). To this end, we first determine the T dependence of the mixing enthalpy by extrapolating the discrete set of bowing parameters from the MC simulations. Specifically, we model the temperature dependence of Ω with an exponentially decaying function in

$1/T$, i.e.,

$$\Omega\left(\frac{1}{T}\right) = Ae^{-\frac{1}{BT}} + C, \tag{4.7}$$

where A, B, and C are empirical fit parameters. This functional form for Ω captures an initial value $A + C$ at $T = \infty$ corresponding to the regular solution limit, and then decreases with lower temperatures to a limiting value C as $T \to 0$. The difference $A - C$ captures the maximal energy lowering afforded by SRO while maintaining a long-range disordered state. Figure 4.2(c) shows numerical fits to this exponential decay function for Ω as a function of $1/T$. Substituting this functional form for Ω into Eq. (4.6) yields a temperature-dependent mixing enthalpy that is used for further analysis in the following.

At finite temperatures in a MC simulation—where configurational entropy is not explicitly known—the free energy of mixing $\Delta G_{\mathrm{m}} = \Delta H_{\mathrm{m}} - T\Delta S_{\mathrm{m}}$ is typically obtained via thermodynamic integration starting from either zero or infinite temperature [16, 30]. Either case makes use of a starting point where configurational entropy is known exactly, either $\Delta S_{\mathrm{m}} = 0$ for $T = 0$ or ΔS_{m} given by Eq. (4.5) for $T = \infty$. We employ the infinite temperature starting point because it coincides with the initial random configurations for the simulated annealing. In this framework, the free energy of mixing at finite temperature is given by

$$\Delta G_{\mathrm{m}}(T) = T\left[-\Delta S_{\mathrm{m}}(\infty) + \int_0^{\frac{1}{T}} \Delta H_{\mathrm{m}}\left(\frac{1}{T'}\right) d\left(\frac{1}{T'}\right)\right], \tag{4.8}$$

where the mixing enthalpy $\Delta H_{\mathrm{m}}(1/T)$ is given by substituting Eq. (4.7) into Eq. (4.6). The chosen analytic expression for $\Omega(1/T)$ above makes it possible to evaluate the integral in Eq. (4.8) and obtain the free energies for any composition and temperature not explicitly sampled with MC. Furthermore, the finite-temperature configurational entropy can then be determined by $\Delta S_{\mathrm{m}} = (\Delta H_{\mathrm{m}} - \Delta G_{\mathrm{m}})/T$ if needed.

4.4.1 Origin of the expression for thermodynamic integration

Equation 4.8 originates from a manipulation of an integral involving Gibbs free energy $G = H - TS$ and temperature T. Specifically, we consider the integral

$$\int_T^\infty \left(\frac{\partial\left(G/T'\right)}{\partial T'}\right)_P dT' = \left(\frac{G}{T'}\right)\left(T' = \infty\right) - \left(\frac{G}{T'}\right)\left(T' = T\right), \tag{4.9}$$

where the right hand side comes from the fundamental theorem of calculus. By the Gibbs-Helmholtz equation (derived in Appendix C), the integrand can be rewritten as $-H/T'^2$. Simultaneously, the term at infinite temperature can be simplified as

$$\left(\frac{G}{T'}\right)\left(T' = \infty\right) = \lim_{T' \to \infty}\left(\frac{H}{T'} - S\right) = -S(T' = \infty). \tag{4.10}$$

With the above substitutions, Eq. 4.9 can be rewritten as

$$-\int_T^\infty \frac{H(T')}{T'^2} dT' = -S(T' = \infty) - \frac{G(T)}{T}. \tag{4.11}$$

Then, we can absorb a minus sign into the integration bounds, multiply by T, and solve for $G(T)$ to obtain

$$G(T) = T\left[-S(\infty) - \int_\infty^T \frac{H(T')}{T'^2} dT'\right]. \tag{4.12}$$

Lastly, we can perform a change of variable from T' to $1/T'$ and make the substitution

$$d\left(\frac{1}{T'}\right) = -\frac{1}{T'^2} dT', \tag{4.13}$$

to arrive at

$$G(T) = T\left[-S(\infty) + \int_0^{\frac{1}{T}} H\left(\frac{1}{T'}\right) d\left(\frac{1}{T'}\right)\right], \tag{4.14}$$

which is the same as Eq. 4.8 but for the total Gibbs free energy G instead of the specific free energy of mixing ΔG_{m}.

4.5 Phase diagram

The Gibbs free energy of mixing ΔG_{m} describes the free energy of the alloy relative to the pure end compounds, thereby eliminating any dependence on chemical potentials and associated partial pressures in the gas phase. Further, we did not attempt to evaluate vibrational contributions which are expected to cancel to a large extent, due to the chemical and structural similarity of the end compounds. Thus, having determined $\Delta G_{\mathrm{m}}(x, T)$ via Eq. (4.8) for both polytypes, we can construct a phase diagram containing the composition limits for binodal and spinodal decomposition.

The binodal line defines the global free energy minimum for phase coexistence, thereby describing the thermodynamic solubility limit, and is obtained via common tangent construction. The spinodal line further describes metastable regions that resist small fluctuations in composition, and is determined by the condition $d^2\Delta G_{\mathrm{m}}/dx^2 = 0$. Phase diagram construction for this temperature dependent bowing parameter model follows a similar procedure to the conventional regular solution model [52, 54]. For temperatures within the range of interest, we determine the binodal x points from common tangent construction, considering that the tangents can be either between or within the polytypes. The spinodal x points are determined for each polytype separately as the inflection points of the free energy curve. Repeating the process for each temperature gives a set of (x, T) points that collectively map out the alloy phase diagram.

For the phase diagram construction from the free energy expression, we utilized the SciPy optimize package. The derivative of the free energy $\Delta G'_{\mathrm{m}}(x) = d\Delta G_{\mathrm{m}}/dx$ was evaluated numerically with second order central differences and then fit with a cubic spline interpolation. For the common tangent construction, we used the minimize function to find the values x_1 and x_2 with the smallest absolute difference between the derivatives of the two structures. However, the unique solution that corresponds to a common tangent has the additional constraint that—in addition to the derivatives being equal—the derivatives at those points must equal the slope of the line that connects them, i.e., $\Delta G'_{\mathrm{m},1}(x_1) = \Delta G'_{\mathrm{m},2}(x_2) = [\Delta G_{\mathrm{m},2}(x_2) - \Delta G_{\mathrm{m},1}(x_1)]/(x_2 - x_1)$. Imposing this constraint when minimizing the absolute difference between the derivatives gives the common tangent at a given

temperature such that the solution points x_1 and x_2 then lie on the binodal lines. The spinodal lines were found by solving for the points at which the numerical derivative changes sign.

Previous work on phase diagrams of heterostructural alloys has highlighted the possibility of novel features beyond the regular solution model of isostructural alloys [54], including the decoupling of binodal and spinodal lines, resulting in access to much wider metastable (x, T) ranges where the alloys are stable against spinodal decomposition. The detailed features of heterostructural phase diagrams depend on whether the relation between the crystal structures is displacive or reconstructive, and on the relative magnitude of the end point polymorph energies and the interaction parameters. As discussed in detail in Ref. [54], in the case of a displacive relationship, only the lower-energy structure type of the respective end compounds is dynamically stable, which leads to a single, but possibly strongly non-parabolic mixing enthalpy. In contrast, a reconstructive relationship involves energy barriers in the structure transition, stabilizing the higher-energy polymorph or polytype. In this case, we obtain two distinct free energy curves over the entire composition range.

Figure 4.3(a) shows the temperature-composition phase diagram for $(Cd_{1-x}Zn_x)_3As_2$, calculated from the above-defined free energy model. The relationship between the two polytypes is reconstructive, because the conversion requires overcoming energy barriers associated with the reordering of the unoccupied cation sites. For the two temperatures of 300 K and 370 K, Figs. 4.3(b) and 4.3(c) show the two free energy curves with the corresponding common tangent construction. (Similar figures for 400 and 500 K are shown in Sec. 4.5.1) The structural transition just below $x = 0.2$ is sharp, due to the reconstructive transformation, and weakly temperature dependent. It occurs at Cd-rich compositions, which is a consequence of the fact that interaction parameters $\Omega(T)$ are smaller in the $P4_2/nmc$ (Zn_3As_2) than in the $I4_1/acd$ (Cd_3As_2) structure, even though the relative magnitude of the end point energies E_0 and E_1 would favor a transition in the Zn-rich regime (cf. Fig. 4.2). The structure transition occurs at higher yet comparable Zn compositions than observed in Ref. [90], but the reported change back to $I4_1/acd$ in the Zn-rich regime does not follow from the present model.

The (binodal) miscibility gap narrows quickly around 400 K, but persists up to high tempera-

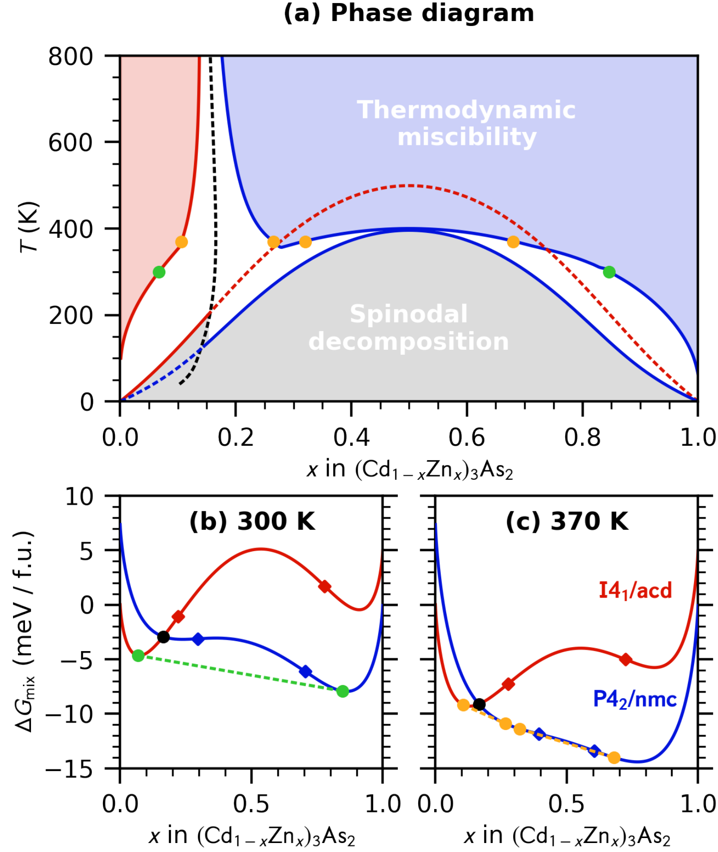


Figure 4.3: (a) Temperature-composition phase diagram for the $(Cd_{1-x}Zn_x)_3As_2$ heterostructural alloy. The miscibility regions for the two structures are colored red and blue, and the metastable region between them is colored white. Corresponding free energies of mixing ΔG_m at (b) $T = 300$ K and (c) $T = 370$ K are shown as a function of composition x. The black dashed line in (a) marks the minimum free energy crossover between the two structures, illustrated by black dots in (b) and (c). Green and orange dots represent points on the binodal lines as governed by common tangent construction. Diamonds indicate inflection points that determine the spinodal lines in (a), shown as solid (dashed) for compositions at which the respective structure is stable (metastable).

tures around the critical composition for structure transition (near $x = 0.2$), forming a characteristic "chimney" structure in the phase diagram [Fig. 4.3(a)]. For temperatures around 370 K, we observe two common tangents [cf. Fig. 4.3(c)], one between the two structures and one within the $P4_2/nmc$ (Zn_3As_2) structure. The possibility of a coexistence of two common tangents is also a unique feature of heterostructural alloys. It was not observed in Ref. [54] because the phase diagrams in that work were dominated by the contribution of the polymorph energies in contrast to the stronger effect of

the interaction parameter in the present case.

In contrast to the binodal lines, the spinodal lines are determined for each structure separately. In the regular solution model, the spinodal lines are exact parabolas, but in our present model they deviate from parabolic, particularly near the endpoints. This deviation is a result of the decreasing magnitude of $\Omega(T)$ with temperature, resulting in a wider range of metastable compositions at low temperatures. Overall, the predicted phase diagram in Fig. 4.3 suggests thermodynamic miscibility for most compositions above about 400 K, which corresponds to the lower limit of common synthesis conditions [104, 125]. The narrow binodal miscibility gap around the structural transition can likely be overcome by thin-film synthesis, e.g., via molecular beam epitaxy [104, 125, 133], where the interatomic mixing can result from the deposition process and does not need to be driven by thermodynamics. Above room temperature, the $(Cd_{1-x}Zn_x)_3As_2$ alloy is predicted to be metastable against spinodal decomposition up to Zn compositions around $x = 0.3$.

The $(Cd_{1-x}Zn_x)_3As_2$ structures generated by the present MC simulations (available in the Supplemental Materials of Ref. [18]) will serve for future investigations, extending upon our previous study in pure Cd_3As_2 [19], where we used a combination of quasi-particle self-consistent GW electronic structure theory [113] and DFT calculations of defect formation energies. Interesting questions include how the Cd/Zn composition and the degree of short-range order affect the topological band structure and how defect-dopant interactions and alloy disorder alter the doping properties [42].

4.5.1 Additional free energies of mixing

For completeness, we include in Fig. 4.4 additional plots of the free energy of mixing that correspond to points in the alloy phase diagram at higher temperatures. For temperatures above the spinodal lines of each structure, the free energy curves are entirely concave up, and there is a narrow composition range for the common tangent between them. At 400 K only space group $P4_2/nmc$ is above the spinodal decomposition limit, and at 500 K both structures are above the limit.

To future students, while the approach described herein is successful in constructing a common tangent between the free energy curves, it is subject to some important limiations. Namely, the

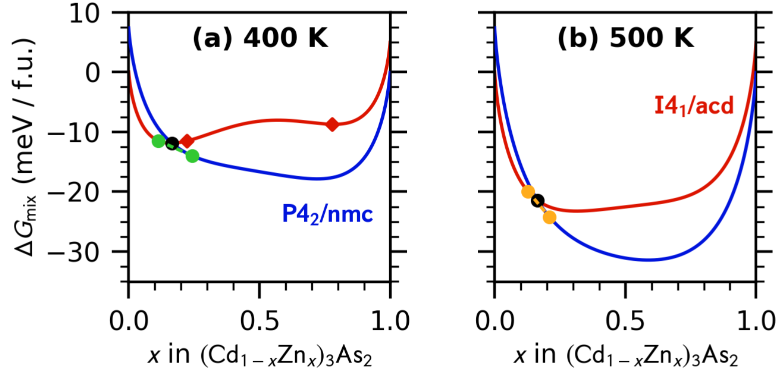


Figure 4.4: Free energies of mixing ΔG_{m} at (a) $T = 400$ K and (c) $T = 500$ K as a function of composition x. The black dots mark the minimum free energy crossover between the two structures. Green and orange dots represent points on the binodal lines as governed by common tangent construction. Diamonds indicate inflection points that determine the spinodal line in Fig. 4.3(a).

solution is strongly dependent on the initial guess for x_1 and x_2 (which would be required by any minimization algorithm). Choosing the initial estimates for these points turns out to be quite complex, and spurious solutions can be obtained if special care is not taken. For example, in temperature ranges where both structures are below their respective spinodal lines and the free energy curves both have pairs of inflection points (e.g. Fig. 4.3b), it is straightforward to use the global minima of the two curves as x_1 and x_2 to obtain the correct solution. However, for temperatures above one or both of the spinodal lines (e.g. Fig. 4.4), choosing the global minima of the free energy curves as the initial guess gives false solutions. To obtain the true common tangents at each temperature as shown in Fig. 4.4, the initial guess unfortunately has to be quite close to the true solution points. These numerical instabilities in finding the solution means that it is very difficult to automate the choice of the initial guess at each temperature, and I was unable to succeed in doing so. Instead, in this regime I manually set the initial guess at a few dozen temperatures by eye, visually inspected each solution to ensure it actually corresponded to a common tangent, then used a cubic spline interpolation to extend the curves over the full range.

4.6 Conclusions

We constructed a heterostructural alloy phase diagram for $(Cd_{1-x}Zn_x)_3As_2$ based on first-principles total energy calculations. Monte Carlo simulated annealing revealed significant short-range order effects, reducing the mixing enthalpy ΔH_m by about a factor of 1/2 compared to the random alloy in the temperature range of interest. The phase diagram reveals characteristic features for alloys with a structural phase transition between the end members, such as the coexistence of two miscibility gaps at a given temperature, and the persistence of a narrow miscibility gap around the structure transition up to high temperatures. The results suggest favorable mixing behavior of $(Cd_{1-x}Zn_x)_3As_2$ alloys, especially in thin film deposition, with stability against spinodal decomposition up to $x = 0.3$.

Chapter 5

Final thoughts and outlook

This thesis has provided many insights into the origins and properties of disorder in Cd_3As_2. There are still open questions, however, particularly with regard to how net doping changes after alloying, e.g. with Zn_3As_2. I would therefore like to briefly discuss some of the future work that could be carried out by subsequent researchers to address these questions. Finally, I will conclude by summarizing the novel techniques and models developed throughout this work and outline how they can be readily applied to other solid-state systems.

5.1 Further possibilities for Cadmium Arsenide

While Ch. 4 addresses the formation of and solubility limits of the $(Cd_{1-x}Zn_x)_3As_2$ alloy, it is still unknown how in detail the electronic structure changes compared to pristine Cd_3As_2. A trivial band gap is expected to open in the Zn-rich limit, but the existence and numerical value of that gap will depend on both the amount of Zn present and the temperature dependent disorder discussed in detail in the previous chapter. Consequently, because the formation of defects is so fundamentally tied to the electronic structure, the net doping will also heavily depend on the amount of Zn and the disorder. A logical next step then is to obtain representative electronic structures for the alloy at different compositions x and temperatures T and incorporate them in an updated defect theory.

There are several challenges associated with construct temperature and alloy composition dependent electronic structures from first-principles calculations. First, the particular numerical values for the band energies depend uniquely on the atomic configuration of the alloy. Therefore,

a representative band structure would require an ensemble average over many configurations at a given composition and temperature. Fortunately, this can be streamlined by using the representative structures that were obtained from MC sampling in Ch. 4 (included as Supplemental Material in Ref. [18], because they were already chosen to represent an equilibrated ensemble average. Averaging the band energies over the MC equilibrated structures should yield a fairly accurate representation of the true electronic structure.

This leads to the second challenge of obtaining the alloy electronic structures: obtaining the representative average at a single temperature and composition requires many individual electronic structure calculations. A computationally efficient approach to each calculation is therefore necessary to perform an electronic structure sampling of this kind. While QSGW provides the most numerically accurate prediction of the band structure for Cd_3As_2, it is impractical to use such an expensive method for the large number of configurations that would be required.

One possible strategy for overcoming this computational bottleneck is through the use of fitted nonlocal external potentials (NLEPs). The NLEP approach is a versatile correction method (see Ref. [75] that can improve DFT band energies in a similar manner to the DFT+U method. It can be differentiated from DFT+U in that the potential introduced to the Kohn-Sham Hamiltonian is applied equally to both occupied and unoccupied orbitals, while the standard Hubbard U potential only acts on occupied states. Different potentials can be applied for each atomic type and for each angular momentum quantum number, and in principle the empirical potentials can be determined so that the DFT band structure largely reproduces that of QSGW. For example, Fig. 5.1 compares the QSGW and DFT band structures of Cd_3As_2, but with several values of the NLEP V applied to the s-like states (equal for Cd and As) in DFT. A value of $V = 0.80$ eV significantly improves the agreement between DFT and QSGW.

To perform the electronic structure sampling necessary for the alloy electronic structure while maintaining the numerical accuracy of QSGW, one could perform a multi-variable least squares fit of the QSGW band structures for pristine Cd_3As_2 and Zn_3As_2. This determines optimal NLEP values for each angular momentum quantum number in Cd, As, and Zn. Then, the electronic structures

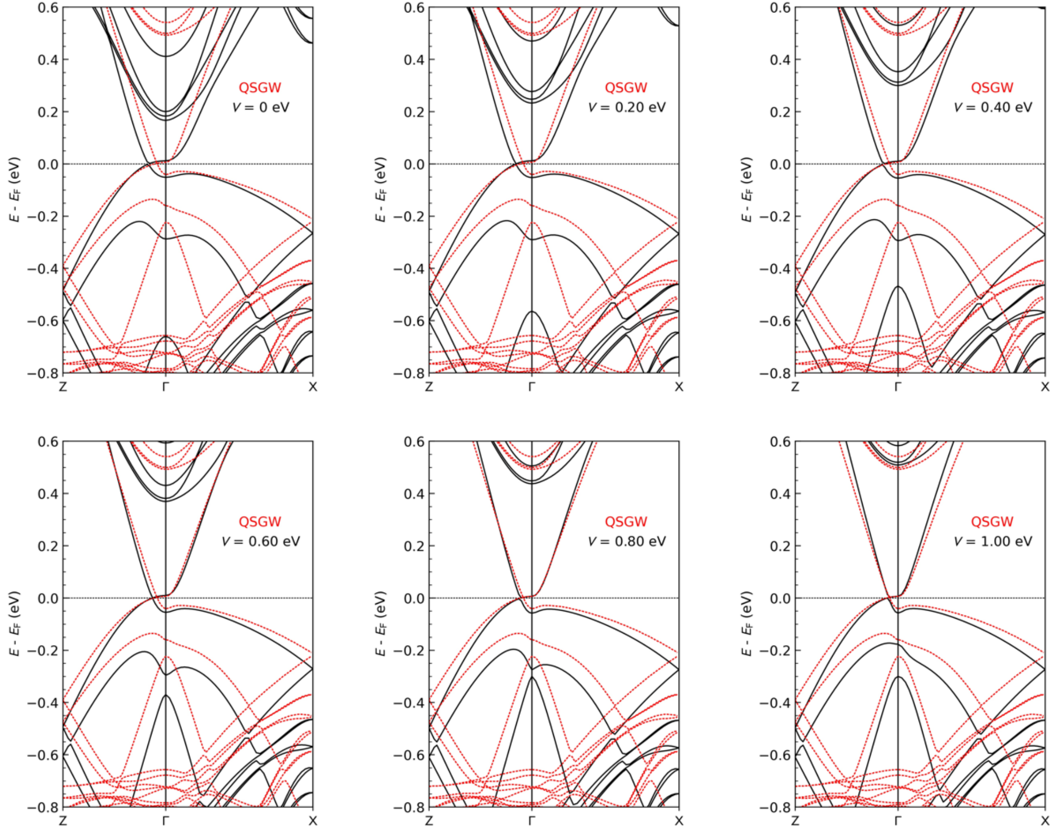


Figure 5.1: The electronic band structure of Cd_3As_2 with energies measured from the Dirac point energy E_{DP}, calculated with quasiparticle self-consistent GW theory (red) and GGA with different values of a nonlocal external potential V applied to all s-like states. All band structures were calculated for the same atomic structure that was relaxed with the SCAN functional.

for the representative alloy structures can be calculated at the DFT level, but with the fidelity of a QSGW calculation. The procedure can be used to construct accurate alloy electronic structures and provides an opportunity to subsequently incorporate the defect pseudo-equilibrium of Ch. 3 to defects and net doping in the $(Cd_{1-x}Zn_x)_3As_2$ alloy.

5.2 Applying these models to other systems

Each of the modernized thermodynamic models developed in this thesis can be adapted and applied to the theoretical study of other material systems. First, the approach of Ch. 2 can be utilized for defect calculations in other topological semimetals. That is, fully screened charged defects

can be used to obtain dilute formation energies without contributions from band-filling; a numerical fit of the DOS near E_F, weighted with the Fermi-Dirac distribution, helps establish the necessary charge balance between donors and acceptors; and a self-consistent defect equilibrium determines net carrier concentrations and the position of the Fermi level. Second, the pseudo-equilibrium model of Ch. 3 can be employed for any system with intrinsic or extrinsic doping where short-range diffusion is expected at operating or storage temperatures. There also remains an open opportunity to more rigorously define this pseudo-equilibrium in terms of a free energy minimization under specific constraints to justify the use of otherwise artificial chemical potentials that maintain the defect stoichiometry in an isolated sample. Lastly, the expanded regular solution approximation with temperature dependent bowing parameter, subsequent thermodynamic integration, and phase diagram construction of Ch. 4 can be easily applied to other heterostructural alloys. The temperature dependence of Ω can be determined through MC sampling with explicit DFT calculations or with a cluster expansion. Together, these models provide a versatile framework for studying disorder and defect-driven properties in a wide range of materials, leading to valuable predictions for practical material design.

References

[1] S. C. Abrahams and J. L. Bernstein, "Piezoelectric nonlinear optic CuGaSe2 and CdGeAs2: crystal structure, chalcopyrite microhardness, and sublattice distortion", The Journal of Chemical Physics **61**, 1140 (1974) (Cited on p. 44).

[2] I. Aguilera, C. Friedrich, and S. Blügel, "Electronic phase transitions of bismuth under strain from relativistic self-consistent GW calculations", Physical Review B **91**, 125129 (2015) (Cited on pp. 14, 28).

[3] K. Alberi, C. Brooks, I. Leahy, and S. Lany, "Tutorial: defects in topological semimetals", Journal of Applied Physics **136** (2024) (Cited on pp. 7, 21, 42, 49).

[4] M. N. Ali, Q. Gibson, S. Jeon, B. B. Zhou, A. Yazdani, and R. J. Cava, "The crystal and electronic structures of Cd_3As_2, the three-dimensional electronic analogue of graphene", Inorganic Chemistry **53**, 4062 (2014) (Cited on pp. 6, 22, 43, 68).

[5] O. K. Andersen, "Linear methods in band theory", Physical Review B **12**, 3060 (1975) (Cited on p. 11).

[6] P. W. Anderson, "Model for the electronic structure of amorphous semiconductors", Physical Review Letters **34**, 953 (1975) (Cited on p. 32).

[7] V. I. Anisimov, J. Zaanen, and O. K. Andersen, "Band theory and Mott insulators: Hubbard U instead of Stoner I", Physical Review B **44**, 943 (1991) (Cited on p. 13).

[8] E. Antillon and M. Ghazisaeidi, "Efficient determination of solid-state phase equilibrium with the multicell Monte Carlo method", Physical Review E **101**, 063306 (2020) (Cited on p. 68).

[9] N. Armitage, E. Mele, and A. Vishwanath, "Weyl and dirac semimetals in three-dimensional solids", Reviews of Modern Physics **90**, 015001 (2018) (Cited on pp. 1, 3, 4).

[10] F. Aryasetiawan and O. Gunnarsson, "The GW method", Reports on Progress in Physics **61**, 237 (1998) (Cited on p. 28).

[11] J. M. Azpiroz, E. Mosconi, J. Bisquert, and F. De Angelis, "Defect migration in methylammonium lead iodide and its role in perovskite solar cell operation", Energy & Environmental Science **8**, 2118 (2015) (Cited on p. 20).

[12] S. Baroni, S. De Gironcoli, A. Dal Corso, and P. Giannozzi, "Phonons and related crystal properties from density-functional perturbation theory", Reviews of modern Physics **73**, 515 (2001) (Cited on p. 12).

[13] C. J. Bartel, A. W. Weimer, S. Lany, C. B. Musgrave, and A. M. Holder, "The role of decomposition reactions in assessing first-principles predictions of solid stability", npj Computational Materials **5**, 4 (2019) (Cited on pp. 16, 23).

[14] S. R. Bauers, A. Holder, W. Sun, C. L. Melamed, R. Woods-Robinson, J. Mangum, J. Perkins, W. Tumas, B. Gorman, A. Tamboli, et al., “Ternary nitride semiconductors in the rocksalt crystal structure”, Proceedings of the National Academy of Sciences **116**, 14829 (2019) (Cited on p. 68).

[15] A. D. Becke, “Density-functional thermochemistry. III. the role of exact exchange”, The Journal of chemical physics **98**, 5648 (1993) (Cited on p. 14).

[16] K. Binder, “Monte Carlo study of entropy for face-centered cubic Ising antiferromagnets”, Zeitschrift für Physik B Condensed Matter **45**, 61 (1981) (Cited on p. 75).

[17] P. E. Blöchl, “Projector augmented-wave method”, Physical Review B **50**, 17953 (1994) (Cited on pp. 23, 43, 60, 69).

[18] C. Brooks and S. Lany, “Heterostructural alloy phase diagram for $(Cd_{1-x}Zn_x)_3As_2$”, Physical Review Materials **8**, L061201 (2024) (Cited on pp. 67, 80, 84).

[19] C. Brooks, M. van Schilfgaarde, D. Pashov, J. N. Nelson, K. Alberi, D. S. Dessau, and S. Lany, “Band energy dependence of defect formation in the topological semimetal Cd_3As_2”, Physical Review B **107**, 224110 (2023) (Cited on pp. 21, 43–45, 49, 61, 68, 69, 71, 80).

[20] J. Buckeridge, D. Jevdokimovs, C. R. A. Catlow, and A. A. Sokol, “Nonstoichiometry and Weyl fermionic behavior in TaAs”, Physical Review B **94**, 180101(R) (2016) (Cited on p. 22).

[21] A. Burkov, “Negative longitudinal magnetoresistance in dirac and weyl metals”, Physical Review B **91**, 245157 (2015) (Cited on p. 6).

[22] A. Burkov, “Topological semimetals”, Nature materials **15**, 1145 (2016) (Cited on pp. 1–6).

[23] D. Ceperley and B. Alder, “Exchange-correlation potential and energy for density-functional calculation”, Phys. Rev. Lett **45**, 567 (1980) (Cited on p. 11).

[24] J. Chan, S. Lany, and A. Zunger, “Electronic correlation in anion *p* orbitals impedes ferromagnetism due to cation vacancies in Zn chalcogenides”, Physical review letters **103**, 016404 (2009) (Cited on p. 15).

[25] X. Chen, Q. Wang, Z. Cheng, M. Zhu, H. Zhou, P. Jiang, L. Zhou, Q. Xue, F. Yuan, J. Zhu, et al., “Direct observation of chemical short-range order in a medium-entropy alloy”, Nature **592**, 712 (2021) (Cited on p. 20).

[26] M. N. Chernodub and M. A. H. Vozmediano, “Direct measurement of a beta function and an indirect check of the Schwinger effect near the boundary in Dirac semimetals”, Physical Review Research **1**, 032002(R) (2019) (Cited on p. 27).

[27] W. Ching, Y.-N. Xu, and K. Wong, “Ground-state and optical properties of Cu_2O and CuO crystals”, Physical Review B **40**, 7684 (1989) (Cited on p. 13).

[28] H. T. Chorsi, S. Yue, P. P. Iyer, M. Goyal, T. Schumann, S. Stemmer, B. Liao, and J. A. Schuller, “Widely tunable optical and thermal properties of Dirac semimetal Cd_3As_2”, Advanced Optical Materials **8**, 1901192 (2020) (Cited on pp. 22, 68).

[29] A. M. Conte, O. Pulci, and F. Bechstedt, “Electronic and optical properties of topological semimetal Cd_3As_2”, Scientific Reports **7**, 45500 (2017) (Cited on pp. 22, 43, 68).

[30] J. J. Cordell, J. Pan, A. C. Tamboli, G. J. Tucker, and S. Lany, “Probing configurational disorder in $ZnGeN_2$ using cluster-based monte carlo”, Physical Review Materials **5**, 024604 (2021) (Cited on p. 75).

[31] I. Crassee, R. Sankar, W.-L. Lee, A. Akrap, and M. Orlita, “3D Dirac semimetal Cd_3As_2: a review of material properties”, Physical Review Materials **2**, 120302 (2018) (Cited on p. 36).

[32] I. Crassee, E. Martino, C. C. Homes, O. Caha, J. Novák, P. Tückmantel, M. Hakl, A. Nateprov, E. Arushanov, Q. D. Gibson, et al., “Nonuniform carrier density in Cd_3As_2 evidenced by optical spectroscopy”, Physical Review B **97**, 125204 (2018) (Cited on p. 6).

[33] I. Di Bernardo, J. Collins, W. Wu, J. Zhou, S. A. Yang, S. Ju, M. T. Edmonds, and M. S. Fuhrer, “Importance of interactions for the band structure of the topological Dirac semimetal Na_3Bi”, Physical Review B **102**, 045124 (2020) (Cited on p. 28).

[34] S. L. Dudarev, G. A. Botton, S. Y. Savrasov, C. Humphreys, and A. P. Sutton, “Electron-energy-loss spectra and the structural stability of nickel oxide: an LSDA+U study”, Physical Review B **57**, 1505 (1998) (Cited on p. 14).

[35] S. V. Faleev, M. Van Schilfgaarde, and T. Kotani, “All-electron self-consistent GW approximation: application to Si, MnO, and NiO”, Physical Review Letters **93**, 126406 (2004) (Cited on p. 18).

[36] J. Feng, Y. Pang, D. Wu, Z. Wang, H. Weng, J. Li, X. Dai, Z. Fang, Y. Shi, and L. Lu, “Large linear magnetoresistance in Dirac semimetal Cd_3As_2 with Fermi surfaces close to the Dirac points”, Physical Review B **92**, 081306 (2015) (Cited on p. 7).

[37] C. Freysoldt, B. Grabowski, T. Hickel, J. Neugebauer, G. Kresse, A. Janotti, and C. G. Van de Walle, “First-principles calculations for point defects in solids”, Reviews of Modern Physics **86**, 253 (2014) (Cited on pp. 7, 12, 24).

[38] H. Gao, J. W. Venderbos, Y. Kim, and A. M. Rappe, “Topological semimetals from first principles”, Annual Review of Materials Research **49**, 153 (2019) (Cited on p. 5).

[39] S. M. Girvin and K. Yang, *Modern condensed matter physics* (Cambridge University Press, 2019) (Cited on pp. 3, 9, 10, 98).

[40] R. W. Godby, M. Schlüter, and L. J. Sham, “Self-energy operators and exchange-correlation potentials in semiconductors”, Physical Review B **37**, 10159 (1988) (Cited on pp. 11, 12, 18).

[41] R. Godby, M. Schlüter, and L. Sham, “Accurate exchange-correlation potential for silicon and its discontinuity on addition of an electron”, Physical review letters **56**, 2415 (1986) (Cited on p. 12).

[42] A. Goyal, M. D. Sanders, R. P. O’Hayre, and S. Lany, “Predicting thermochemical equilibria with interacting defects: $Sr_{1-x}Ce_xMnO_{3-\delta}$ alloys for water splitting”, PRX Energy **3**, 013008 (2024) (Cited on p. 80).

[43] A. Goyal, A. Zakutayev, V. Stevanović, and S. Lany, “Computational fermi level engineering and doping-type conversion of Mg:Ga_2O_3 via three-step synthesis process”, Journal of Applied Physics **129**, 245704 (2021) (Cited on pp. 22, 44, 52).

[44] M. Grumet, P. Liu, M. Kaltak, J. Klimeš, and G. Kresse, “Beyond the quasiparticle approximation: fully self-consistent gw calculations”, Physical Review B **98**, 155143 (2018) (Cited on p. 17).

[45] M. Grüning, A. Marini, and A. Rubio, “Density functionals from many-body perturbation theory: the band gap for semiconductors and insulators”, The Journal of chemical physics **124** (2006) (Cited on p. 12).

[46] M. Grüning, A. Marini, and A. Rubio, “Effect of spatial nonlocality on the density functional band gap”, Physical Review B **74**, 161103(R) (2006) (Cited on p. 12).

[47] P. Haas, F. Tran, and P. Blaha, "Calculation of the lattice constant of solids with semilocal functionals", Physical Review B **79**, 085104 (2009) (Cited on p. 13).

[48] M. Z. Hasan and C. L. Kane, "Colloquium: topological insulators", Reviews of modern physics **82**, 3045 (2010) (Cited on p. 1).

[49] L. He, X. Hong, J. Dong, J. Pan, Z. Zhang, J. Zhang, and S. Li, "Quantum transport evidence for the three-dimensional Dirac semimetal phase in Cd_3As_2", Physical review letters **113**, 246402 (2014) (Cited on p. 7).

[50] L. Hedin and S. Lundqvist, "Effects of electron-electron and electron-phonon interactions on the one-electron states of solids", in *Solid state physics*, Vol. 23, edited by F. Seitz, D. Turnbull, and H. Ehrenreich (Academic Press, 1969), pp. 1–181 (Cited on p. 17).

[51] J. Heyd, G. E. Scuseria, and M. Ernzerhof, "Hybrid functionals based on a screened Coulomb potential", The Journal of chemical physics **118**, 8207 (2003) (Cited on p. 14).

[52] I. Ho and G. B. Stringfellow, "Solid phase immiscibility in GaInN", Applied Physics Letters **69**, 2701 (1996) (Cited on pp. 73, 77).

[53] P. Hohenberg and W. Kohn, "Inhomogeneous electron gas", Physical review **136**, B864 (1964) (Cited on p. 9).

[54] A. M. Holder, S. Siol, P. F. Ndione, H. Peng, A. M. Deml, B. E. Matthews, L. T. Schelhas, M. F. Toney, R. G. Gordon, W. Tumas, J. D. Perkins, D. S. Ginley, B. P. Gorman, J. Tate, A. Zakutayev, and S. Lany, "Novel phase diagram behavior and materials design in heterostructural semiconductor alloys", Science Advances **3**, e1700270 (2017) (Cited on pp. 67, 77–79).

[55] P. Hosur and X. Qi, "Recent developments in transport phenomena in weyl semimetals", Comptes Rendus Physique **14**, 857 (2013) (Cited on p. 22).

[56] M. S. Hybertsen and S. G. Louie, "Electron correlation in semiconductors and insulators: band gaps and quasiparticle energies", Physical Review B **34**, 5390 (1986) (Cited on p. 17).

[57] E. B. Isaacs and C. Wolverton, "Performance of the strongly constrained and appropriately normed density functional for solid-state materials", Phys. Rev. Mater. **2**, 063801 (2018) (Cited on pp. 13, 16).

[58] A. Janotti and C. G. Van de Walle, "Native point defects in ZnO", Physical Review B—Condensed Matter and Materials Physics **76**, 165202 (2007) (Cited on p. 20).

[59] G. S. Jenkins, C. Lane, B. Barbiellini, A. B. Sushkov, R. L. Carey, F. Liu, J. W. Krizan, S. K. Kushwaha, Q. Gibson, T.-R. Chang, H.-T. Jeng, H. Lin, R. J. Cava, A. Bansil, and H. D. Drew, "Three-dimensional Dirac cone carrier dynamics in Na_3Bi and Cd_3As_2", Physical Review B **94**, 085121 (2016) (Cited on pp. 6, 19, 22, 68).

[60] X. Jin, S. Chen, and T. Li, "Coexistence of two types of short-range order in Si-Ge-Sn medium-entropy alloys", Communications Materials **3**, 66 (2022) (Cited on p. 20).

[61] R. O. Jones and O. Gunnarsson, "The density functional formalism, its applications and prospects", Reviews of Modern Physics **61**, 689 (1989) (Cited on p. 10).

[62] H. Jónsson, G. Mills, and K. W. Jacobsen, "Nudged elastic band method for finding minimum energy paths of transitions", in *Classical and quantum dynamics in condensed phase simulations* (World Scientific, 1998), pp. 385–404 (Cited on p. 64).

[63] C. L. Kane and E. J. Mele, “Z_2 topological order and the quantum spin Hall effect”, Physical review letters **95**, 146802 (2005) (Cited on p. 2).

[64] S. Kirkpatrick, C. D. Gelatt Jr, and M. P. Vecchi, “Optimization by simulated annealing”, Science **220**, 671 (1983) (Cited on p. 70).

[65] W. Kohn and L. J. Sham, “Self-consistent equations including exchange and correlation effects”, Physical review **140**, A1133 (1965) (Cited on pp. 9, 10).

[66] B. Koley and P. P. Jana, “Structure and stability of Au_3M_5 (M = Mg, Cd)”, Journal of Solid State Chemistry **274**, 215 (2019) (Cited on p. 44).

[67] M. Koshino and I. F. Hizbullah, “Magnetic susceptibility in three-dimensional nodal semimetals”, Physical Review B **93**, 045201 (2016) (Cited on pp. 4, 5).

[68] T. Kotani, M. van Schilfgaarde, and S. V. Faleev, “Quasiparticle self-consistent GW method: a basis for the independent-particle approximation”, Physical Review B **76**, 165106 (2007) (Cited on pp. 23, 29, 43, 61).

[69] G. Kresse and D. Joubert, “From ultrasoft pseudopotentials to the projector augmented-wave method”, Physical Review B **59**, 1758 (1999) (Cited on pp. 13, 23, 43, 60, 69).

[70] A. Kyrtsos, M. Matsubara, and E. Bellotti, “Migration mechanisms and diffusion barriers of carbon and native point defects in GaN”, Physical Review B **93**, 245201 (2016) (Cited on p. 20).

[71] S. Lany, “Semiconductor thermochemistry in density functional calculations”, Phys. Rev. B **78**, 245207 (2008) (Cited on pp. 13, 15).

[72] S. Lany, “Band-structure calculations for the $3d$ transition metal oxides in GW”, Physical Review B **87**, 085112 (2013) (Cited on pp. 23, 28, 30, 43, 61).

[73] S. Lany, “Communication: the electronic entropy of charged defect formation and its impact on thermochemical redox cycles”, The Journal of Chemical Physics **148**, 071101 (2018) (Cited on pp. 12, 34).

[74] S. Lany, “Chemical potential analysis as an alternative to the van’t Hoff method: hypothetical limits of solar thermochemical hydrogen”, Journal of the American Chemical Society **146**, 14114 (2024) (Cited on p. 12).

[75] S. Lany, H. Raebiger, and A. Zunger, “Magnetic interactions of Cr-Cr and Co-Co impurity pairs in ZnO within a band-gap corrected density functional approach”, Physical Review B—Condensed Matter and Materials Physics **77**, 241201 (2008) (Cited on p. 84).

[76] S. Lany, Y.-J. Zhao, C. Persson, and A. Zunger, “Halogen n-type doping of chalcopyrite semiconductors”, Applied Physics Letters **86**, 042109 (2005) (Cited on p. 34).

[77] S. Lany and A. Zunger, “Anion vacancies as a source of persistent photoconductivity in II-VI and chalcopyrite semiconductors”, Physical Review B—Condensed Matter and Materials Physics **72**, 035215 (2005) (Cited on pp. 46, 47).

[78] S. Lany and A. Zunger, “Dopability, intrinsic conductivity, and nonstoichiometry of transparent conducting oxides”, Physical Review Letters **98**, 045501 (2007) (Cited on p. 27).

[79] S. Lany and A. 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**, 235104 (2008) (Cited on pp. 24, 27, 28).

[80] S. Lany and A. Zunger, "Accurate prediction of defect properties in density functional supercell calculations", Modelling and Simulation in Materials Science and Engineering **17**, 084002 (2009) (Cited on pp. 24, 26, 27, 61).

[81] A. H. Larsen et al., "The atomic simulation environment—a Python library for working with atoms", Journal of Physics: Condensed Matter **29**, 273002 (2017) (Cited on p. 64).

[82] S. R. Lee, P. A. Sharma, A. L. Lima-Sharma, W. Pan, and T. M. Nenoff, "Topological quantum materials for realizing majorana quasiparticles", Chemistry of Materials **31**, 26 (2018) (Cited on p. 5).

[83] C.-Z. Li, L.-X. Wang, H. Liu, J. Wang, Z.-M. Liao, and D.-P. Yu, "Giant negative magnetoresistance induced by the chiral anomaly in individual Cd_3As_2 nanowires", Nature communications **6**, 10137 (2015) (Cited on p. 6).

[84] H. Li, H. He, H.-Z. Lu, H. Zhang, H. Liu, R. Ma, Z. Fan, S.-Q. Shen, and J. Wang, "Negative magnetoresistance in Dirac semimetal Cd_3As_2", Nature communications **7**, 10301 (2016) (Cited on pp. 6, 7).

[85] T. Liang, Q. Gibson, M. N. Ali, M. Liu, R. J. Cava, and N. P. Ong, "Ultrahigh mobility and giant magnetoresistance in the Dirac semimetal Cd_3As_2", Nature Materials **14**, 280 (2015) (Cited on pp. 5, 6, 22).

[86] J. Z. Liu, G. Trimarchi, and A. Zunger, "Strain-minimizing tetrahedral networks of semiconductor alloys", Physical Review Letters **99**, 145501 (2007) (Cited on p. 68).

[87] Z. K. Liu, J. Jiang, B. Zhou, Z. J. Wang, Y. Zhang, H. M. Weng, D. Prabhakaran, S. K. Mo, H. Peng, P. Dudin, et al., "A stable three-dimensional topological Dirac semimetal Cd_3As_2", Nature Materials **13**, 677 (2014) (Cited on pp. 4, 6, 22).

[88] Z. Liu, B. Zhou, Y. Zhang, Z. Wang, H. Weng, D. Prabhakaran, S.-K. Mo, Z. Shen, Z. Fang, X. Dai, et al., "Discovery of a three-dimensional topological dirac semimetal, Na_3Bi", Science **343**, 864 (2014) (Cited on p. 4).

[89] F. Lu, J. Zhao, H. Weng, Z. Fang, and X. Dai, "Correlated topological insulators with mixed valence", Physical review letters **110**, 096401 (2013) (Cited on p. 14).

[90] H. Lu, X. Zhang, Y. Bian, and S. Jia, "Topological phase transition in single crystals of $(Cd_{1-x}Zn_x)_3As_2$", Scientific Reports **7**, 3148 (2017) (Cited on pp. 68, 69, 78).

[91] B. Lv, T. Qian, and H. Ding, "Experimental perspective on three-dimensional topological semimetals", Reviews of Modern Physics **93**, 025002 (2021) (Cited on pp. 4, 5).

[92] B. Lv, H. Weng, B. Fu, X. P. Wang, H. Miao, J. Ma, P. Richard, X. Huang, L. Zhao, G. Chen, et al., "Experimental discovery of weyl semimetal TaAs", Physical Review X **5**, 031013 (2015) (Cited on p. 3).

[93] S. Ma and S. Wang, "Ab initio study of self-diffusion in silicon over a wide temperature range: point defect states and migration mechanisms", Physical Review B—Condensed Matter and Materials Physics **81**, 193203 (2010) (Cited on p. 20).

[94] G. Makov and M. C. Payne, "Periodic boundary conditions in ab initio calculations", Physical Review B **51**, 4014 (1995) (Cited on p. 27).

[95] Y. Meng, C. Zhu, Y. Li, X. Yuan, F. Xiu, Y. Shi, Y. Xu, and F. Wang, "Three-dimensional Dirac semimetal thin-film absorber for broadband pulse generation in the near-infrared", Optics letters **43**, 1503 (2018) (Cited on p. 5).

[96] N. D. Mermin, “Thermal properties of the inhomogeneous electron gas”, Physical Review **137**, A1441 (1965) (Cited on p. 12).

[97] G. Mills, H. Jónsson, and G. K. Schenter, “Reversible work transition state theory: application to dissociative adsorption of hydrogen”, Surface Science **324**, 305 (1995) (Cited on p. 64).

[98] P. J. Moll, N. L. Nair, T. Helm, A. C. Potter, I. Kimchi, A. Vishwanath, and J. G. Analytis, “Transport evidence for Fermi-arc-mediated chirality transfer in the Dirac semimetal Cd_3As_2”, Nature **535**, 266 (2016) (Cited on p. 6).

[99] P. Mori-Sánchez, A. J. Cohen, and W. Yang, “Localization and delocalization errors in density functional theory and implications for band-gap prediction”, Physical review letters **100**, 146401 (2008) (Cited on p. 13).

[100] M. Muzyk, D. Nguyen-Manh, K. J. Kurzydłowski, N. L. Baluc, and S. L. Dudarev, “Phase stability, point defects, and elastic properties of W-V and W-Ta alloys”, Physical Review B **84**, 104115 (2011) (Cited on pp. 12, 68).

[101] N. L. Nair, P. T. Dumitrescu, S. Channa, S. M. Griffin, J. B. Neaton, A. C. Potter, and J. G. Analytis, “Thermodynamic signature of dirac electrons across a possible topological transition in $ZrTe_5$”, Physical Review B **97**, 041111(R) (2018) (Cited on p. 22).

[102] P. Narang, C. A. Garcia, and C. Felser, “The topology of electronic band structures”, Nature Materials **20**, 293 (2021) (Cited on p. 1).

[103] A. Narayanan, M. Watson, S. Blake, N. Bruyant, L. Drigo, Y. Chen, D. Prabhakaran, B. Yan, C. Felser, T. Kong, et al., “Linear magnetoresistance caused by mobility fluctuations in n-doped Cd_3As_2”, Physical Review Letters **114**, 117201 (2015) (Cited on pp. 6, 19).

[104] J. N. Nelson, I. A. Leahy, A. D. Rice, C. Brooks, G. Teeter, M. van Schilfgaarde, S. Lany, B. Fluegel, M. Lee, and K. Alberi, “Direct link between disorder and magnetoresistance in topological semimetals”, Physical Review B **107**, L220206 (2023) (Cited on pp. 6, 7, 19, 22, 25, 32, 36, 53, 68, 80).

[105] R. Neumann, R. H. Nobes, and N. C. Handy, “Exchange functionals and potentials”, Molecular Physics **87**, 1 (1996) (Cited on p. 15).

[106] M. Neupane, S.-Y. Xu, R. Sankar, N. Alidoust, G. Bian, C. Liu, I. Belopolski, T.-R. Chang, H.-T. Jeng, H. Lin, et al., “Observation of a three-dimensional topological Dirac semimetal phase in high-mobility Cd_3As_2”, Nature communications **5**, 3786 (2014) (Cited on pp. 4–6).

[107] S. Nishihaya, M. Uchida, Y. Nakazawa, M. Kriener, Y. Kozuka, Y. Taguchi, and M. Kawasaki, “Gate-tuned quantum Hall states in Dirac semimetal $(Cd_{1-x}Zn_x)_3As_2$”, Science Advances **4**, eaar5668 (2018) (Cited on pp. 6, 19, 68).

[108] T. Ogawa, A. Taguchi, and A. Kuwabara, “An extended computational approach for point-defect equilibria in semiconductor materials”, Npj Computational Materials **8**, 79 (2022) (Cited on p. 52).

[109] G. B. Osterhoudt, L. K. Diebel, M. J. Gray, X. Yang, J. Stanco, X. Huang, B. Shen, N. Ni, P. J. Moll, Y. Ran, et al., “Colossal mid-infrared bulk photovoltaic effect in a type-I Weyl semimetal”, Nature materials **18**, 471 (2019) (Cited on p. 5).

[110] J. Pan, J. Cordell, G. J. Tucker, A. C. Tamboli, A. Zakutayev, and S. Lany, “Interplay between composition, electronic structure, disorder, and doping due to dual sublattice mixing in nonequilibrium synthesis of $ZnSnN_2$:O”, Advanced Materials **31**, 1807406 (2019) (Cited on p. 22).

[111] J. Pan, J. J. Cordell, G. J. Tucker, A. Zakutayev, A. C. Tamboli, and S. Lany, "Perfect short-range ordered alloy with line-compound-like properties in the $ZnSnN_2$: ZnO system", npj Computational Materials **6**, 63 (2020) (Cited on p. 70).

[112] S. Parameswaran, T. Grover, D. Abanin, D. Pesin, and A. Vishwanath, "Probing the chiral anomaly with nonlocal transport in three-dimensional topological semimetals", Physical Review X **4**, 031035 (2014) (Cited on p. 5).

[113] D. Pashov, S. Acharya, W. R. L. Lambrecht, J. Jackson, K. D. Belashchenko, A. Chantis, F. Jamet, and M. van Schilfgaarde, "Questaal: a package of electronic structure methods based on the linear muffin-tin orbital technique", Computer Physics Communications **249**, 107065 (2020) (Cited on pp. 23, 29, 43, 61, 80).

[114] H. Peng and S. Lany, "Polymorphic energy ordering of MgO, ZnO, GaN, and MnO within the random phase approximation", Physical Review B—Condensed Matter and Materials Physics **87**, 174113 (2013) (Cited on p. 15).

[115] H. Peng, D. O. Scanlon, V. Stevanovic, J. Vidal, G. W. Watson, and S. Lany, "Convergence of density and hybrid functional defect calculations for compound semiconductors", Physical Review B **88**, 115201 (2013) (Cited on pp. 18, 27, 29, 30).

[116] H. Peng, D. O. Scanlon, V. Stevanovic, J. Vidal, G. W. Watson, and S. Lany, "Addendum to "Convergence of density and hybrid functional defect calculations for compound semiconductors"", Physical Review B **96**, 239903(E) (2017) (Cited on p. 27).

[117] J. P. Perdew, J. A. Chevary, S. H. Vosko, K. A. Jackson, M. R. Pederson, D. J. Singh, and C. Fiolhais, "Atoms, molecules, solids, and surfaces: applications of the generalized gradient approximation for exchange and correlation", Physical review B **46**, 6671 (1992) (Cited on p. 11).

[118] J. P. Perdew and M. Levy, "Physical content of the exact Kohn-Sham orbital energies: Band gaps and derivative discontinuities", Physical Review Letters **51**, 1884 (1983) (Cited on pp. 12, 28).

[119] J. P. Perdew and A. Zunger, "Self-interaction correction to density-functional approximations for many-electron systems", Physical review B **23**, 5048 (1981) (Cited on pp. 11, 13).

[120] J. P. Perdew, K. Burke, and M. Ernzerhof, "Generalized gradient approximation made simple", Physical Review Letters **77**, 3865 (1996) (Cited on pp. 13, 23, 69).

[121] 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) (Cited on p. 25).

[122] A. Pietraszko and K. Łukaszewicz, "Thermal expansion and phase transitions of Cd_3As_2 and Zn_3As_2", Phys. Stat. Sol. (a) **18**, 723 (1973) (Cited on p. 71).

[123] A. D. Rice, M. Liedke, M. Butterling, E. Hirschmann, A. Wagner, N. M. Haegel, and K. Alberi, "Vacancy complexes in Cd_3As_2", APL Materials **11** (2023) (Cited on p. 19).

[124] A. D. Rice, J. N. Nelson, C. Brooks, S. Lany, and K. Alberi, "Extrinsic n-type doping of Cd_3As_2 thin films", Applied Physics Letters **122** (2023) (Cited on p. 53).

[125] A. D. Rice, K. Park, E. T. Hughes, K. Mukherjee, and K. Alberi, "Defects in Cd_3As_2 epilayers via molecular beam epitaxy and strategies for reducing them", Physical Review Materials **3**, 121201(R) (2019) (Cited on pp. 36, 53, 80).

[126] L. M. Rogers, R. M. Jenkins, and A. J. Crocker, "Transport and optical properties of the $Cd_{3-x}Zn_xAs_2$ alloy system", Journal of Physics D: Applied Physics **4**, 793 (1971) (Cited on p. 68).

[127] C. L. Rom, R. W. Smaha, C. L. Melamed, R. R. Schnepf, K. N. Heinselman, J. S. Mangum, S.-J. Lee, S. Lany, L. T. Schelhas, A. L. Greenaway, et al., "Combinatorial synthesis of cation-disordered manganese tin nitride $MnSnN_2$ thin films with magnetic and semiconducting properties", Chemistry of Materials **35**, 2936 (2023) (Cited on p. 68).

[128] J. M. Sanchez, F. Ducastelle, and D. Gratias, "Generalized cluster description of multicomponent systems", Physica A: Statistical Mechanics and its Applications **128**, 334 (1984) (Cited on p. 12).

[129] R. Sankar, M. Neupane, S.-Y. Xu, C. J. Butler, I. Zeljkovic, I. Panneer Muthuselvam, F.-T. Huang, S.-T. Guo, S. K. Karna, M.-W. Chu, et al., "Large single crystal growth, transport property and spectroscopic characterizations of three-dimensional Dirac semimetal Cd_3As_2", Scientific Reports **5**, 12966 (2015) (Cited on pp. 6, 19, 22, 36, 68).

[130] M. van Schilfgaarde, T. Kotani, and S. Faleev, "Quasiparticle self-consistent GW theory", Physical Review Letters **96**, 226402 (2006) (Cited on pp. 18, 23, 28, 43, 61).

[131] A. Schleife, M. Eisenacher, C. Rödl, F. Fuchs, J. Furthmüller, and F. Bechstedt, "Ab initio description of heterostructural alloys: Thermodynamic and structural properties of $Mg_xZn_{1-x}O$ and $Cd_xZn_{1-x}O$", Physical Review B **81**, 245210 (2010) (Cited on p. 67).

[132] L. M. Schoop, F. Pielnhofer, and B. V. Lotsch, "Chemical principles of topological semimetals", Chemistry of Materials **30**, 3155 (2018) (Cited on pp. 5, 22).

[133] T. Schumann, M. Goyal, H. Kim, and S. Stemmer, "Molecular beam epitaxy of Cd_3As_2 on a III-V substrate", APL Materials **4**, 126110 (2016) (Cited on pp. 36, 53, 80).

[134] E. G. Seebauer and M. C. Kratzer, "Charged point defects in semiconductors", Materials Science and Engineering: R: Reports **55**, 57 (2006) (Cited on p. 20).

[135] A. Seko, K. Shitara, and I. Tanaka, "Efficient determination of alloy ground-state structures", Physical Review B **90**, 174104 (2014) (Cited on p. 68).

[136] L. J. Sham and M. Schlüter, "Density-functional theory of the energy gap", Physical Review Letters **51**, 1888 (1983) (Cited on pp. 12, 28).

[137] A. Sharan and S. Lany, "Computational discovery of stable and metastable ternary oxynitrides", The Journal of Chemical Physics **154** (2021) (Cited on pp. 68, 69).

[138] M. Shishkin and G. Kresse, "Implementation and performance of the frequency-dependent GW method within the PAW framework", Physical Review B **74**, 035101 (2006) (Cited on pp. 17, 23, 43, 61).

[139] M. Shishkin, M. Marsman, and G. Kresse, "Accurate quasiparticle spectra from self-consistent GW calculations with vertex corrections", Physical Review Letters **99**, 246403 (2007) (Cited on p. 18).

[140] B. Singh, B. Ghosh, C. Su, H. Lin, A. Agarwal, and A. Bansil, "Topological hourglass dirac semimetal in the nonpolar phase of Ag_2BiO_3", Physical Review Letters **121**, 226401 (2018) (Cited on p. 22).

[141] D. J. Singh and L. Nordstrom, *Planewaves, pseudopotentials, and the LAPW method* (Springer Science & Business Media, 2006) (Cited on pp. 10, 11).

[142] S. Siol, A. Holder, J. Steffes, L. T. Schelhas, K. H. Stone, L. Garten, J. D. Perkins, P. A. Parilla, M. F. Toney, B. D. Huey, W. Tumas, S. Lany, and A. Zakutayev, “Negative-pressure polymorphs made by heterostructural alloying”, Science Advances **4**, eaaq1442 (2018) (Cited on p. 67).

[143] R. W. Smaha, K. Yazawa, A. G. Norman, J. S. Mangum, H. Guthrey, G. L. Brennecka, A. Zakutayev, S. R. Bauers, P. Gorai, and N. M. Haegel, “Synthesis and calculations of wurtzite $Al_{1-x}Gd_xN$ heterostructural alloys”, Chemistry of Materials **34**, 10639 (2022) (Cited on p. 68).

[144] V. Stevanović, S. Lany, X. Zhang, and A. Zunger, “Correcting density functional theory for accurate predictions of compound enthalpies of formation: fitted elemental-phase reference energies”, Physical Review B **85**, 115104 (2012) (Cited on p. 15).

[145] G. B. Stringfellow, *Organometallic vapor-phase epitaxy: theory and practice* (Elsevier, 1999) (Cited on p. 73).

[146] J. Sun, M. Marsman, G. I. Csonka, A. Ruzsinszky, P. Hao, Y.-S. Kim, G. Kresse, and J. P. Perdew, “Self-consistent meta-generalized gradient approximation within the projector-augmented-wave method”, Phys. Rev. B **84**, 035117 (2011) (Cited on pp. 15, 16).

[147] J. Sun, A. Ruzsinszky, and J. P. Perdew, “Strongly constrained and appropriately normed semilocal density functional”, Physical Review Letters **115**, 036402 (2015) (Cited on pp. 16, 23, 43, 60, 68, 69).

[148] L. K. Teles, J. Furthmüller, L. M. R. Scolfaro, J. R. Leite, and F. Bechstedt, “First-principles calculations of the thermodynamic and structural properties of strained $In_xGa_{1-x}N$ and $Al_xGa_{1-x}N$ alloys”, Physical Review B **62**, 2475 (2000) (Cited on pp. 12, 68).

[149] W. J. Turner, A. S. Fischler, and W. E. Reese, “Physical properties of several II-V semiconductors”, Physical Review **121**, 759 (1961) (Cited on p. 68).

[150] C. G. Van de Walle and J. Neugebauer, “First-principles calculations for defects and impurities: applications to III-nitrides”, Journal of applied physics **95**, 3851 (2004) (Cited on p. 12).

[151] J. Vidal, X. Zhang, V. Stevanović, J.-W. Luo, and A. Zunger, “Large insulating gap in topological insulators induced by negative spin-orbit splitting”, Phys. Rev. B **86**, 075316 (2012) (Cited on p. 15).

[152] J. Vidal, X. Zhang, L. Yu, J.-W. Luo, and A. Zunger, “False-positive and false-negative assignments of topological insulators in density functional theory and hybrids”, Physical Review B **84**, 041109(R) (2011) (Cited on pp. 14, 28).

[153] P. Virtanen, R. Gommers, T. E. Oliphant, M. Haberland, T. Reddy, D. Cournapeau, E. Burovski, P. Peterson, W. Weckesser, J. Bright, S. J. van der Walt, M. Brett, J. Wilson, K. J. Millman, N. Mayorov, A. R. J. Nelson, E. Jones, R. Kern, E. Larson, C. J. Carey, İ. Polat, Y. Feng, E. W. Moore, J. VanderPlas, D. Laxalde, J. Perktold, R. Cimrman, I. Henriksen, E. A. Quintero, C. R. Harris, A. M. Archibald, A. H. Ribeiro, F. Pedregosa, P. van Mulbregt, and SciPy 1.0 Contributors, “SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python”, Nature Methods **17**, 261 (2020) (Cited on p. 69).

[154] X. Wan, A. M. Turner, A. Vishwanath, and S. Y. Savrasov, “Topological semimetal and Fermi-arc surface states in the electronic structure of pyrochlore iridates”, Physical Review B—Condensed Matter and Materials Physics **83**, 205101 (2011) (Cited on p. 4).

[155] H. Wang and A.-B. Chen, "Calculation of shallow donor levels in GaN", Journal of Applied Physics **87**, 7859 (2000) (Cited on pp. 27, 47).

[156] L. Wang, T. Maxisch, and G. Ceder, "Oxidation energies of transition metal oxides within the GGA + U framework", Phys. Rev. B **73**, 195107 (2006) (Cited on p. 15).

[157] Q. Wang, C.-Z. Li, S. Ge, J.-G. Li, W. Lu, J. Lai, X. Liu, J. Ma, D.-P. Yu, Z.-M. Liao, et al., "Ultrafast broadband photodetectors based on three-dimensional Dirac semimetal Cd_3As_2", Nano letters **17**, 834 (2017) (Cited on p. 5).

[158] X. Wang, G. Bian, P. Wang, and T.-C. Chiang, "Dirac semimetal films as spin conductors on topological substrates", Physical Review B **91**, 125103 (2015) (Cited on p. 5).

[159] H. Weng, C. Fang, Z. Fang, B. A. Bernevig, and X. Dai, "Weyl semimetal phase in noncentrosymmetric transition-metal monophosphides", Physical Review X **5**, 011029 (2015) (Cited on p. 3).

[160] J. Xiong, S. K. Kushwaha, T. Liang, J. W. Krizan, M. Hirschberger, W. Wang, R. J. Cava, and N. P. Ong, "Evidence for the chiral anomaly in the dirac semimetal Na_3Bi", Science **350**, 413 (2015) (Cited on p. 5).

[161] B. Yan and C. Felser, "Topological materials: weyl semimetals", Annual Review of Condensed Matter Physics **8**, 337 (2017) (Cited on pp. 1, 5).

[162] J. H. Yang, D. A. Kitchaev, and G. Ceder, "Rationalizing accurate structure prediction in the meta-GGA SCAN functional", Physical Review B **100**, 035132 (2019) (Cited on pp. 60, 69).

[163] L. Yang, J. Deslippe, C.-H. Park, M. L. Cohen, and S. G. Louie, "Excitonic effects on the optical response of graphene and bilayer graphene", Physical Review Letters **103**, 186802 (2009) (Cited on p. 28).

[164] N. Yavarishad, T. Hosseini, E. Kheirandish, C. P. Weber, and N. Kouklin, "Room-temperature self-powered energy photodetector based on optically induced Seebeck effect in Cd_3As_2", Applied Physics Express **10**, 052201 (2017) (Cited on p. 5).

[165] E. R. Ylvisaker, W. E. Pickett, and K. Koepernik, "Anisotropy and magnetism in the LSDA+U method", Physical Review B—Condensed Matter and Materials Physics **79**, 035103 (2009) (Cited on p. 14).

[166] W. Zdanowicz, F. Krolicki, and P. Plenkiewicz, "Preparation and semiconducting properties of pseudobinary solid solutions Zn_3As_2-Zn_3P_2", Acta Physica Polonica A **44**, 447 (1973) (Cited on p. 68).

[167] W. Zdanowicz, K. Lukaszewicz, and W. Trzebiatowski, "Crystal structure of semiconducting system Cd_3As_2-Zn_3As_2", Bull. Acad., Pol. Sci., Ser. Chim **12**, 169 (1964) (Cited on pp. 68, 69).

[168] C. Zhang, T. Zhou, S. Liang, J. Cao, X. Yuan, Y. Liu, Y. Shen, Q. Wang, J. Zhao, Z. Yang, et al., "Unexpected low thermal conductivity and large power factor in Dirac semimetal Cd_3As_2", Chinese Physics B **25**, 017202 (2015) (Cited on p. 5).

[169] S. B. Zhang and J. E. Northrup, "Chemical potential dependence of defect formation energies in GaAs: application to Ga self-diffusion", Physical Review Letters **67**, 2339 (1991) (Cited on pp. 12, 22).

[170] C. Zhu, F. Wang, Y. Meng, X. Yuan, F. Xiu, H. Luo, Y. Wang, J. Li, X. Lv, L. He, et al., "A robust and tuneable mid-infrared optical switch enabled by bulk Dirac fermions", Nature Communications **8**, 14111 (2017) (Cited on p. 5).

Appendix A

Berry curvature in the vacinity of a Weyl point

The following derivation is based on a similar calculation on Pg. 340 of Ref. [39]. For the Weyl Hamiltonian (Eq. 1.1) with positive chirality (the derivation proceeds in the same manner for negative chirality), the two eigenstates (for a specific gauge choice) are

$$|-\rangle = \begin{pmatrix} \sin(\theta/2)e^{-i\phi} \\ -\cos(\theta/2) \end{pmatrix}, \tag{A.1}$$

and

$$|+\rangle = \begin{pmatrix} \cos(\theta/2)e^{-i\phi} \\ \sin(\theta/2) \end{pmatrix}, \tag{A.2}$$

with energies given by $H_{\text{Weyl}} |\pm\rangle = \pm v_{\text{F}} k |\pm\rangle$. Here, the crystal momentum $\boldsymbol{k}$ is given in spherical coordinates, i.e. $\boldsymbol{k} = k(\sin\theta\cos\phi, \sin\theta\sin\phi, \cos\theta)$.

The (vector) Berry connection $\boldsymbol{\mathcal{A}}^n(\boldsymbol{\mathcal{R}})$ of the n-th eigenstate of a Hamiltonian H with a set of D variable parameters labeled by $\boldsymbol{\mathcal{R}} = (\mathcal{R}_1, \mathcal{R}_2, \ldots, \mathcal{R}_D)$, is defined as [39]

$$\boldsymbol{\mathcal{A}}^n(\boldsymbol{\mathcal{R}}) = i \langle n(\boldsymbol{\mathcal{R}})| \frac{\partial}{\partial \boldsymbol{\mathcal{R}}} |n(\boldsymbol{\mathcal{R}})\rangle \,. \tag{A.3}$$

The variable parameters $\boldsymbol{\mathcal{R}}$ in momentum space for the Weyl Hamiltonian are then k, θ, and ϕ. Focusing on the $|-\rangle$ state, the spherical components of the Berry connection are as follows. The

component $\mathcal{A}_k^-$ is trivially 0 because the eigenstates do not depend on k. As for the other components:

$$\begin{aligned}
\mathcal{A}_\theta^- &= i\,\langle -|\partial_\theta|-\rangle \\
&= i\begin{pmatrix}\sin(\theta/2)e^{i\phi} & -\cos(\theta/2)\end{pmatrix}\frac{\partial}{\partial_\theta}\begin{pmatrix}\sin(\theta/2)e^{-i\phi} \\ -\cos(\theta/2)\end{pmatrix} \\
&= \frac{i}{2}\begin{pmatrix}\sin(\theta/2)e^{i\phi} & -\cos(\theta/2)\end{pmatrix}\begin{pmatrix}\cos(\theta/2)e^{-i\phi} \\ \sin(\theta/2)\end{pmatrix} \\
&= \frac{i}{2}(\sin(\theta/2)\cos(\theta/2) - \sin(\theta/2)\cos(\theta/2)) \\
&= 0,
\end{aligned} \tag{A.4}$$

and

$$\begin{aligned}
\mathcal{A}_\phi^- &= i\,\langle -|\partial_\phi|-\rangle \\
&= i\begin{pmatrix}\sin(\theta/2)e^{i\phi} & -\cos(\theta/2)\end{pmatrix}\frac{\partial}{\partial_\phi}\begin{pmatrix}\sin(\theta/2)e^{-i\phi} \\ -\cos(\theta/2)\end{pmatrix} \\
&= i\begin{pmatrix}\sin(\theta/2)e^{i\phi} & -\cos(\theta/2)\end{pmatrix}\begin{pmatrix}-i\,\sin(\theta/2)e^{-i\phi} \\ 0\end{pmatrix} \\
&= \sin^2(\theta/2).
\end{aligned} \tag{A.5}$$

The (gauge independent and anti-symmetric) tensor for the Berry curvature, defined in general as

$$\omega_{\mu\nu}^n(\boldsymbol{\mathcal{R}}) = \partial_{\mathcal{R}_\mu}\mathcal{A}_\nu^n(\boldsymbol{\mathcal{R}}) - \partial_{\mathcal{R}_\nu}\mathcal{A}_\mu^n(\boldsymbol{\mathcal{R}}), \tag{A.6}$$

can then be constructed in spherical coordinates as

$$\begin{aligned}
\omega_{\theta\phi}^- = -\omega_{\phi\theta}^- &= \partial_\theta\mathcal{A}_\phi^- - \partial_\phi\mathcal{A}_\theta^- \\
&= \frac{\partial}{\partial\theta}\sin^2(\theta/2) \\
&= \sin(\theta/2)\cos(\theta/2) \\
&= \frac{\sin(\theta)}{2},
\end{aligned} \tag{A.7}$$

with all other entries equal zero. It can then be transformed back to Cartesian momentum coordinates by way of the Jacobian matrix. For example, the k_x, k_y component can be found as

$$\begin{aligned} \omega^{-}_{k_x k_y} &= \omega^{-}_{\theta\phi} \det \begin{vmatrix} \frac{\partial\theta}{\partial k_x} & \frac{\partial\theta}{\partial k_y} \\ \frac{\partial\phi}{\partial k_x} & \frac{\partial\phi}{\partial k_y} \end{vmatrix} \\ &= \frac{k_z}{2k^3}, \end{aligned} \tag{A.8}$$

where the spherical coordinates that are differentiated are

$$\theta = \cos^{-1}\left(\frac{k_z}{\sqrt{k_x^2 + k_y^2 + k_z^2}}\right), \tag{A.9}$$

and

$$\phi = \tan^{-1}\left(\frac{k_y}{k_x}\right). \tag{A.10}$$

Repeating the above calculation for each Cartesian component of ω^- gives

$$\omega^{-}_{ij} = \frac{1}{2k^3}\begin{pmatrix} 0 & k_z & -k_y \\ -k_z & 0 & k_x \\ k_y & -k_x & 0 \end{pmatrix}. \tag{A.11}$$

Lastly, the tensor can be converted to the pseudovector Berry curvature $\boldsymbol{\Omega}(\boldsymbol{k})$, by contracting it with the Levi-Civita symbol, i.e.

$$\Omega^n_i = \epsilon_{ijk}\omega^n_{jk}, \tag{A.12}$$

so we have

$$\begin{aligned} \Omega^{-}_{k_x} &= \epsilon_{123}\omega^{-}_{k_y k_z} = \frac{k_x}{2k^3}, \\ \Omega^{-}_{k_y} &= \epsilon_{213}\omega^{-}_{k_x k_z} = \frac{k_y}{2k^3}, \\ \Omega^{-}_{k_z} &= \epsilon_{312}\omega^{-}_{k_x k_y} = \frac{k_z}{2k^3}, \end{aligned} \tag{A.13}$$

or for the combined Berry curvature, the point charge expression of Eq. 1.2:

$$\Omega^{-}(\boldsymbol{k}) = \frac{1}{2}\frac{\boldsymbol{k}}{k^3}. \tag{A.14}$$

The derivation is exactly analogus for the $|+\rangle$ state as well as for both eigenstates with negative chirality.

Appendix B

Derivation of the non-dilute concentration of defects in thermal equilibrium

The total Gibbs free energy $G_{\mathrm{def}}^{\mathrm{tot}} = H_{\mathrm{def}}^{\mathrm{tot}} - TS_{\mathrm{def}}^{\mathrm{tot}}$ for an ensemble of various *non-interacting* point defects can be expressed as a sum over contributions from each defect. That is, up to degeneracy factors,

$$G_{\mathrm{def}}^{\mathrm{tot}} = \sum_{\mathrm{D},q} c_{\mathrm{D},q} \Delta H_{\mathrm{D},q} - k_{\mathrm{B}} T \ln \prod_{\mathrm{D},q} W(c_{\mathrm{D},q}, N_{\mathrm{D}}), \tag{B.1}$$

where $c_{\mathrm{D},q}$ and $\Delta H_{\mathrm{D},q}$ are the concentration and enthalpy of formation of the defect, respectively, k_{B} is the Boltzmann constant, T is the temperature, and N_{D} is the atomic site density where each defect D can substitute. The microstate duplicity function $W(c_{\mathrm{D},q}, N_{\mathrm{D}})$ that determines each defect's contribution to the total configurational entropy is simply the number of ways to arrange $c_{\mathrm{D},q}$ defects on N_{D} sites:

$$W(c_{\mathrm{D},q}, N_{\mathrm{D}}) = \frac{N_{\mathrm{D}}!}{c_{\mathrm{D},q}!(N_{\mathrm{D}} - c_{\mathrm{D},q})!}. \tag{B.2}$$

Applying the Stirling Approximation, $\ln \mathrm{N}! \approx N \ln N - N$, to Eq. B.1 then simplifies the total free energy to

$$\begin{aligned} G_{\mathrm{def}}^{\mathrm{tot}} &= \sum_{\mathrm{D},q} c_{\mathrm{D},q} \Delta H_{\mathrm{D},q} - k_{\mathrm{B}} T \sum_{\mathrm{D},q} \ln \left[\frac{N_{\mathrm{D}}!}{c_{\mathrm{D},q}!(N_{\mathrm{D}} - c_{\mathrm{D},q})!} \right] \\ &= \sum_{\mathrm{D},q} c_{\mathrm{D},q} \Delta H_{\mathrm{D},q} - k_{\mathrm{B}} T \sum_{\mathrm{D},q} \left[N_{\mathrm{D}} \ln N_{\mathrm{D}} - c_{\mathrm{D},q} \ln c_{\mathrm{D},q} - (N_{\mathrm{D}} - c_{\mathrm{D},q}) \ln (N_{\mathrm{D}} - c_{\mathrm{D},q}) \right]. \end{aligned} \tag{B.3}$$

The expression for $c_{\mathrm{D}',q'}$ for a specific defect D' with charge state q' in equilibrium is determined by minimizing the free energy with respect to that defect's concentration, and the free energy is

minimal when $\partial G^{\mathrm{tot}}_{\mathrm{def}}/\partial c_{\mathrm{D}',q'} = 0$. Taking the derivative of Eq. B.3 and setting it equal to zero gives

$$0 = \Delta H_{\mathrm{D}',q'} - k_{\mathrm{B}}T\left[-\ln c_{\mathrm{D}',q'} + \ln\left(N_{\mathrm{D}'} - c_{\mathrm{D}',q'}\right)\right], \tag{B.4}$$

which simplifies to

$$-\frac{\Delta H_{\mathrm{D}',q'}}{k_{\mathrm{B}}T} = \ln\left(\frac{c_{\mathrm{D}',q'}}{N_{\mathrm{D}'} - c_{\mathrm{D}',q'}}\right). \tag{B.5}$$

Finally, dropping the unique labels D' and q' and rearranging gives Eq. 2.6,

$$\frac{c_{\mathrm{D},q}}{N_{\mathrm{D}} - c_{\mathrm{D},q}} = e^{-\Delta H_{\mathrm{D},q}/k_{\mathrm{B}}T}, \tag{B.6}$$

which can be solved for $c_{\mathrm{D},q}$ to obtain Eq. 2.5,

$$c_{\mathrm{D},q} = N_{\mathrm{D}}\frac{e^{-\Delta H_{\mathrm{D},q}/k_{\mathrm{B}}T}}{1 + e^{-\Delta H_{\mathrm{D},q}/k_{\mathrm{B}}T}}. \tag{B.7}$$

Appendix C

Derivation of Gibbs-Helmholtz equation

The Gibbs free energy for a closed system (fixed number of particles) at pressure P and temperature T is given by

$$G(P,T) = H - TS = U + PV - TS, \tag{C.1}$$

where the enthalpy $H = U + PV$ is defined in terms of the total energy U, P, and volume V, and the entropy is given by S. The total differential of this expression is

$$dG = dU + PdV + VdP - TdS - SdT, \tag{C.2}$$

where upon substituting the fundamental thermodynamic relation

$$dU = TdS - PdV, \tag{C.3}$$

simplifies to

$$dG = VdP - SdT. \tag{C.4}$$

This allows us to apply the quotient rule to the differential of G/T, and simplify it as

$$\begin{aligned} d\left(\frac{G}{T}\right) &= \frac{TdG - GdT}{T^2} \\ &= \frac{T\left(VdP - SdT\right) - GdT}{T^2} \\ &= \frac{-\left(G + TS\right)dT + TVdP}{T^2} \\ &= \frac{-HdT + TVdP}{T^2} \\ &= -\frac{H}{T^2}dT + \frac{V}{T}dP. \end{aligned} \tag{C.5}$$

Simultaneously, because G is the thermodynamic potential for the natural variables P and T, the same differential can be expressed as

$$d\left(\frac{G}{T}\right) = \left(\frac{\partial\left(G/T\right)}{\partial T}\right)_P dT + \left(\frac{\partial\left(G/T\right)}{\partial P}\right)_T dP. \tag{C.6}$$

Comparing Eqs. C.5 and C.6 gives the Gibbs-Helmholtz Equation,

$$\left(\frac{\partial\left(G/T\right)}{\partial T}\right)_P = -\frac{H}{T^2}, \tag{C.7}$$

and simultaneously the related expression,

$$\left(\frac{\partial\left(G/T\right)}{\partial P}\right)_T = \frac{V}{T}. \tag{C.8}$$