©Copyright 2024

Xiaofeng Xiang

Defect Modeling and Device Optimization in Chalcogenide Photovoltaics from First Principles to AI-assisted Design

Xiaofeng Xiang

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

Doctor of Philosophy

University of Washington

2024

Reading Committee:

Scott T. Dunham, Chair

Ting Cao

Hugh W. Hillhouse

Program Authorized to Offer Degree:
Molecular Engineering

University of Washington

Abstract

Defect Modeling and Device Optimization in Chalcogenide Photovoltaics from First Principles to AI-assisted Design

Xiaofeng Xiang

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

This dissertation presents a comprehensive framework for defect modeling and device optimization in chalcogenide photovoltaics, focusing on Cu(In,Ga)Se_2 (CIGS) and CdSeTe materials. By integrating first principles density functional theory (DFT), continuum modeling, and AI-driven methods, a robust predictive model for electronic performance and optimization of thin-film solar cells has been developed.

Detailed analyses of defect formation and diffusion mechanisms were conducted using DFT calculations, enabling precise control over defect profiles to optimize material properties and enhance device performance. Continuum modeling and device modeling were employed to bridge atomic-scale defect properties with macroscopic device behavior, successfully predicting the impact of defect distributions on carrier lifetimes, providing insights into manufacturing process impacts and solar cell performance optimization. The developed predictive Technology Computer Aided Design (TCAD) model offers a detailed depiction of defect behaviors and their impact on device performance, serving as a powerful tool for guiding experimental efforts and advancing the development of high-efficiency thin-film solar cells.

The integration of AI-driven methods positively influences TCAD simulations by employing machine learning (ML) for defect property prediction, thereby accelerating technology advancement. A framework powered by ML and DFT has been developed for predicting and screening functional

impurities in semiconductors. This framework can be further implemented into the TCAD model to explore new dopants in solar cells and other electronic devices.

TABLE OF CONTENTS

Page

List of Figures . . . iv

List of Tables . . . viii

Chapter 1: Introduction . . . 1

Chapter 2: Background . . . 3
- 2.1 Chalcogenide Photovoltaics . . . 3
  - 2.1.1 CIGS . . . 3
  - 2.1.2 CdSeTe . . . 3
- 2.2 Methodology . . . 4
  - 2.2.1 Density Functional Theory (DFT) . . . 4
    - 2.2.1.1 Defect Formation Energy . . . 5
    - 2.2.1.2 Migration Barrier . . . 6
    - 2.2.1.3 Defect-assisted Recombination . . . 6
  - 2.2.2 Continuum Modeling . . . 9
    - 2.2.2.1 Process Simulation . . . 9
    - 2.2.2.2 Device Simulation . . . 11
  - 2.2.3 AI-Driven Methods . . . 12
    - 2.2.3.1 Training Data . . . 13
    - 2.2.3.2 Basic Machine Learning Algorithms . . . 14
    - 2.2.3.3 Cross-Validation . . . 15

Chapter 3: Coupled Process and Device Modeling of CIGS Solar Cells . . . 16
- 3.1 Process Simulation in CIGS . . . 16
  - 3.1.1 Defect Diffusion in CIGS . . . 17
  - 3.1.2 Continuum Models for Defect Diffusion . . . 19
  - 3.1.3 Comparison Between Simulation and Experiment . . . 19
- 3.2 Coupling of Process and Device Simulations . . . 20

3.2.1 Cooling Process . . . 25
3.2.2 Solar cell device simulation . . . 26
3.2.3 Simulation Results . . . 30
3.2.3.1 Effect of GGI Gradient on Device Performance . . . 31
3.3 Summary . . . 34

Chapter 4: Point Defects Engineering in CdTe and CdSeTe Alloy . . . 36
4.1 Methods . . . 37
4.1.1 First Principles Calculation Methods . . . 37
4.1.2 Computational Details . . . 37
4.2 Results and Discussion of Point Defects . . . 42
4.2.1 Intrinsic Defects . . . 42
4.2.2 Group V Dopants . . . 51
4.2.2.1 Defect Complex . . . 55
4.2.3 Copper . . . 59
4.3 Continuum Modeling and Optimization of Group-V Doping in CdTe and CdSeTe . . 60
4.3.1 Defect Thermodynamics . . . 61
4.3.2 Process Simulation . . . 64
4.3.3 Results and Discussion . . . 64
4.3.3.1 Grain Boundary Segregation . . . 71
4.4 Summary . . . 74

Chapter 5: Accelerated Defect Property Predictions and Device Design of Cubic Semiconductor Crystals . . . 77
5.1 Machine Learning Framework for Qualitative Screening . . . 79
5.1.1 Semiconductor and Impurity Chemical Space . . . 79
5.1.2 Descriptors . . . 81
5.1.2.1 Feature Selection . . . 81
5.1.3 Training Regression Models . . . 83
5.1.3.1 Linear Regression . . . 84
5.1.3.2 Kernel Ridge Regression . . . 88
5.1.3.3 Gaussian Process Regression . . . 88
5.1.3.4 Neural Network Regression . . . 88
5.1.4 High-Throughput Screening of Dominating Impurities . . . 91
5.2 Deep Learning Models for Quantitative Screening . . . 93
5.2.1 Training Dataset . . . 94

5.2.2 Chemical Environment Encoded Artificial Neural Network 95
5.2.3 Crystal Defect Graph Neural Network 99
5.2.4 ML-assisted Device Design Framework 101
5.2.5 Discussion 105
5.3 Summary 108

Chapter 6: Conclusion 109

Bibliography 111

Appendix A: Sdevice example (CIGS solar cells) 131

Appendix B: ML-assisted Dopant Screening for Device Design (Example of CdTe Solar Cells) 135
B.1 Device Simulation Details 135

Appendix C: Point Defects Engineering in CdTe and CdSeTe Alloy 140
C.1 Hubbard U Values Determination 140
C.2 Energy Cutoff Validation 141
C.3 Se Distribution Dependence in CdSeTe Alloy 142

LIST OF FIGURES

Figure Number Page

2.1 Workflow of Modeling Methods. 5

2.2 Representation of carrier capture at a defect: (a) band diagram and (b) configuration coordinate diagram. The defect in the diagram is a deep acceptor with a negative and a neutral charge state. $\triangle E$ is the ionization energy of the defect, and Q is a properly selected configuration coordinate. In (b), step(1) is the change of the electronic state due to electron-phonon coupling (nonradiative), step(1^*) is the change of the electronic state due to electron-photon coupling (radiative) and step(2) is the vibration relaxation due to phonon-phonon interactions. 7

2.3 Schematic of neural network machine learning algorithms. 13

3.1 Defect transition levels in (a) $CuInSe_2$; (b) $CuGaSe_2$. Dashed lines indicate valence band maximum and conduction band minimum. 18

3.2 Indium tracer diffusion experiment vs. continuum model (a) normal scale non-linear fitting curve (blue); (b) log scale non-linear fitting curve (blue); (c) normal scale continuum model simulated curve (green); (d) log scale continuum model simulated curve (green). Experimental data points are shown in gold scattered points above. . 21

3.3 Configuration coordinate diagram of Cu_{III} (0/-1) (3.3a) and (-1/-2) (3.3b) transition states. Q in X-axis corresponds to the mass-weighted configuration coordinate. $\triangle$Q is the difference in the equilibrium geometries of the initial and final state. $\triangle$E is the energy difference between the initial and final state. 23

3.4 SeM vs. μ_{Se} for GGI=0.15, 0.35 and 0.55 at CGI=0.85; μ_{Se} is Se chemical potential referenced to standard Se chemical potential used in DFT calculation. Temperature for μ_{Se} is 873K. 24

3.5 Deep level defect distribution with different CGI, μ_{Se} and GGI. (a) CGI = 0.60 $\sim$ 0.99, GGI = 0.30, $\mu_{Se} = -1.00$[eV]; (b) $\mu_{Se} = -0.76 \sim -1.20$[eV], CGI = 0.85, GGI = 0.30; (c) GGI = 0.05 $\sim$ 0.75, CGI = 0.85, $\mu_{Se} = -1.00$[eV]. 27

3.6 Complex refractive index of CIGS with different GGI values computed by Eq. (3.4). κ is extinction coefficient while n is refractive index. 28

3.7 GGI curve extracted from GDOES profile in *Jackson et al.*[73] 30

3.8 Deep level defect concentration distribution in CIGS absorber before and after cooling process. 30

3.9 Device schematic and band diagram under AM 1.5G illumination (no bias applied). CIGS layer ranges from 0 to 2.8 μm, but only a portion is shown in the schematic for concise illustration. Red dashed line indicates Fermi level. Upper green solid line denotes conduction band minimum while lower green solid line denotes valence band maximum. 31

3.10 Schematic of GGI profile tuning. Three constraints in the GGI profile are back GGI (Mo interface), front GGI (CdS interface) and the turning point location. All the device simulations are conducted under the same conditions except GGI profile and corresponding defect level change. 34

3.11 Simulation of device efficiency for various GGI gradients with CIGS thickness of 1.4μm. 34

3.12 Simulation of device efficiency for various GGI gradients and corresponding GGI values at two interfaces, mapped for several GGI at CdS interface (0.15, 0.25, 0.35 and 0.40). The turning point location ranges from 0.20 μm (close to Mo layer) to 2.80 μm (CdS interface). GGI at Mo interface ranges from 0.45∼0.73. 35

4.1 Defect formation energies vs inverse supercell size for (a) $As_{Te}^{0,-1}$, (b) $(Cd_{int}+As_{Te})^{+}$ and (c) Cu_{int}^{+2} . The defects are relaxed with fixed lattice constants. "L" on the bottom x-axis denotes the length of corresponding CdTe supercell. The Fermi level is set to the VBM in these plots. 40

4.2 Defect formation energies vs Fermi level of intrinsic defects in CdTe under (a) Cd-rich and (d) Te-rich condition, $CdSe_{0.25}Te_{0.75}$ under (b) Cd-rich and (e) Te-rich condition, $CdSe_{0.50}Te_{0.50}$ under (c) Cd-rich and (f) Te-rich condition. The chemical potential conditions are $\mu_{Cd} + \mu_{Te} = -1.28$ eV and $\mu_{Cd} + \mu_{Se} < 0.15$ eV. 43

4.3 Configuration Coordinate Diagram of V_{Cd} (-1/0) and (-2/-1). Q in X-axis corresponds to the mass-weighted configuration coordinate. Q indicates the configurational coordinate path between equilibrium configurations. $Q = 0$ indicates defects ground state with T_d symmetry. 44

4.4 Schematic of Te_{Cd} C_{3v} defect configuration. c denotes length of the long bond. a denotes length of three short bonds. 45

4.5 Comparison between Defect Neighbor Model predictions and DFT calculations for the Te_{Cd}^{0} defect across Se ratios ranging from 0% to 50% (56 data included). The root mean square error (RMSE) for test data and correlation coefficient (R^2) demonstrate excellent agreement with the data. 47

4.6 Schematic of Te_{int} split defect configuration 49

4.7 Schematic of AX center defect configuration 50

4.8 Schematic of AX center defect band splitting. Left one indicates AX in CdTe and right one indicates AX in $CdSe_{0.25}Te_{0.75}$. $\triangle E_{split}$ is the band splitting energy 51

4.9 New coordinate system for defining atom locations relative to an oriented AX dimer. The coordinates are defined by As (0, 0, 0), the distorted Te (1, 0, 0), and the Cd bonded to both (0.5, 0, 0.5). Additional sites for a Se at (0.5, 0.5, 1) and a Te at (0, 0, 2) are also depicted. Black arrows show the As and Te deformation when the AX center forms. 53

4.10 Defect Vector Model Prediction vs Calculation of (a) Arsenic AX defect and (b) Phosphorus AX defect in CdSeTe alloy (Se ratio ranges from 0% to 50%). 55

4.11 Configuration Coordinate Diagram of As_{Te} (+1_AX/0) and Cd_{int} (+2/+1) in $CdSe_{0.25}Te_{0.75}$. As_{Te}^{+} with formation energy close to the Boltzmann distribution at 300 K is selected as the excited state. As_{Te}^{0} exhibits minimal dependence on the Se/Te arrangement, and the choice of As_{Te}^{0} does not significantly affect the results. In (b), E_{opt} corresponds to the optical transition level. $Q = 0$ indicates interstitial ground state positioned in a octahedra vacant site. $Q \approx 30$ indicates interstitial ground state positioned in a tetrahedral vacant site. Cd_{int}^{+1} is favorable in octahedra site while Cd_{int}^{+2} is favorable in tetrahedral site. 56

4.12 Defect Structure of $(Cd_{int} + As_{Te})^{+}$ in CdTe. 58

4.13 Defect formation energies vs Fermi level of Cu related defects and dominant intrinsic defects in CdTe under (a) Cd-rich and (d) Te-rich condition, $CdSe_{0.25}Te_{0.75}$ under (b) Cd-rich and (e) Te-rich condition, $CdSe_{0.50}Te_{0.50}$ under (c) Cd-rich and (f) Te-rich condition. The chemical potential condition is $\mu_{Cd} + \mu_{Te} = -1.28$ eV, $\mu_{Cd} + \mu_{Se} < 0.15$ eV and $\mu_{Cu} + \mu_{Te} < -0.42$ eV. 60

4.14 Chemical potential range for Group-V dopants. Compounds above indicate competitive phases. 63

4.15 Hole and acceptor densities in As doped CdTe and $CdSe_{0.25}Te_{0.75}$ with μ_{Cd} = -0.80 eV. Initial temperature is 873K. Cooling rate is 2°C/min in (a) and 20°C/min in (b). 66

4.16 Dominant defect densities, carrier density and As activation in As doped CdTe and $CdSe_{0.25}Te_{0.75}$. Initial temperature is 873K. Cooling rate is 20°C/min. 67

4.17 Dominant defect densities, carrier density and As activation in As doped CdTe and $CdSe_{0.25}Te_{0.75}$. Initial temperature is 1173K. Cooling rate is 20°C/min. 68

4.18 Dominant defect densities, carrier density and P activation in P doped CdTe and $CdSe_{0.25}Te_{0.75}$. Initial temperature is 873K. Cooling rate is 20°C/min. 69

4.19 Dominant defect densities, carrier density and P activation in P doped CdTe and $CdSe_{0.25}Te_{0.75}$. Initial temperature is 1173K. Cooling rate is 20°C/min. 70

4.20 Grain boundaries and grain structure of polycrytalline CdSeTe. The width of GBs is 5nm and width of grain is $5\mu m$. Se ratio is chosen to be 25% and μ_{Cd} is -0.8 eV. . 71

4.21 Hole density vs. E_{seg}. Red dotted line indicates baseline model without GBs. 73

5.1 (a) The DFT-ML workflow followed in this work, and the semiconductor-impurity chemical space in terms of (b) the cation (A) and anion (B) choices for group IV, II-VI, and III-V compounds, (c) types of defect sites, and (d) impurity atoms selected from across the periodic table. 80
5.2 Absolute values of coefficient of linear correlation between every unique descriptor and every property; darker boxes imply high correlation whereas white boxes mean there is zero correlation. 82
5.3 Parity plots for best regression models: (a) Random Forest, (b) Gaussian Process, and (c) Neural Network regression. 89
5.4 Parity Plots for the Artificial Neural Network 98
5.5 Schematic of Crystal Defect Graph Neural Network 100
5.6 Schematic of ML-assisted Device Design Framework. DFT is employed to build a comprehensive defect property database including defect formation energy and thermodynamic transition levels. A surrogate neural network model is trained using this dataset to predict defect properties for dopants across the entire chemical space. Next, CCT is used to filter and select dopants with desired properties. Then the defect information of these potential dopants is incorporated into the TCAD model of the specific device of interest, allowing us to gather figures of merit. In the end, we identify promising potential dopants that demonstrate potential for enhancing device performance. Experimental testing and validation are conducted to assess the suitability and efficacy of the selected dopants. 105
C.1 Dependence of lattice parameter a_0 and bandgap on U of different methods: standard DFT (U=0 eV) and GGA+U. The stars represent the HSE06 values. The dashed lines represent the experimental values. 140
C.2 Energy difference ($E_{bulk} - E_{defect}$) of defects (As_{Te}^{-1} and Te_{Cd}^{0}) without correction with energy cutoffs ranging from 400 eV to 550 eV. 141

LIST OF TABLES

Table Number Page

3.1 CIGS Defect Diffusion Mechanism 19

3.2 Defect Reactions in Continuum Model (charge exchange reactions are assumed to be fast for process simulation, so all equations are listed just in terms of holes) 20

3.3 Binding energy (eV) of defect complexes in $CuInSe_2$ and $CuGaSe_2$. Binding energy in pure $CuInSe_2$ and pure $CuGaSe_2$ phases is calculated respectively. Binding energy in $CuIn_{1-x}Ga_xSe_2$ alloy is linearly interpolated according to Ga/In ratio. Charge states of complexes are noted in brackets 24

3.4 Impact of cooling process on deep level defects. [a]Equilibrium concentration was calculated via CCT. CGI, GGI, and μ_{Se} ratio for this sample are 0.85, 0.35, -1.00[eV], respectively. The cooling process is simulated via Sentaurus Process with a cooling rate of -20°C/min. 26

3.5 Critical parameters in CIGS Device Model 29

3.6 CIGS solar cell structure of experiment and simulation 29

3.7 Solar cell device performance of experiment and simulation 31

4.1 LATTICE CONSTANT, BANDGAP FOR CDTE AND CDSE OBTAINED FROM GGA+U HSE06 AND EXPERIMENTS. 39

4.2 Capture cross section of deep level defects in CdTe at 300K 45

4.3 Te^0_{Cd} Defect Neighbor Model: coefficients ΔE and their standard deviations for each feature. For nearest neighbor counting features, the count is specific to Se atoms. The distinction between short bond and long bond is shown in Fig. 4.4. 46

4.4 Arsenic AX Center Defect Vector Model 54

4.5 Phosphorus AX Center Defect Vector Model 54

4.6 Arsenic Defect Formation Energy and Complex Binding Energy (eV) in $CdSe_xTe_{1-x}$. 55

4.7 Phosphorus Defect Formation Energy and Complex Binding Energy (eV) in $CdSe_xTe_{1-x}$. . . 57

4.8 Bond Lengths (Å) and Bond Orders for Nearest Neighbors Surrounding Cd Interstitial in $(Cd_{int} + As_{Te})^+$ and $(Cd_{int} + P_{Te})^+$. Bond lengths are provided outside the brackets, while bond orders are denoted inside the brackets. 57

4.9 CdSeTe Defect Migration Barrier (NEB) 64

4.10 Defect Reactions in Continuum Model: Example of As dopant 65

4.11 Selected defect densities and carrier densities in polycrystalline $CdSe_{0.25}Te_{0.75}$ with varying E_{seg}. Baseline model has no GBs. Initial temperature is 873K. Cooling rate is 20°C/min. Storage time at 300K is 2 weeks. Units of defect/hole density are cm^{-3}. 73

5.1 ML test set prediction RMSE values for transition levels. Lowest prediction errors are highlighted in bold. 86

5.2 ML test set prediction RMSE values for formation energies. Lowest prediction errors are highlighted in bold. 87

5.3 Selected dominating impurities identified by both DFT and ML (GPR), at A-rich chemical potential conditions. 92

5.4 ANN Model Performance 98

5.5 CDGNN Model Performance 100

5.6 ML-assisted Dopant Screening for Device Design (Example of CdTe Solar Cells) . . 106

B.1 Critical parameters in CdTe Device Model 135

B.2 ML-assisted Dopant Screening for Device Design (Example of CdTe Solar Cells, Full Table) 136

C.1 Mean Values and Standard Deviation (σ) of CdSeTe Alloy Bulk Energy E_{bulk} and Lattice Constants a_0. 142

C.2 CdSeTe Alloy Linear Regression Model: coefficients ΔE for each feature. All features are standardized before training. For nearest neighbor counting features, the count is specific to Se atoms. The training error is 0.1 meV while the testing error is 0.6 meV. 142

ACKNOWLEDGMENTS

I am deeply grateful to my professor and committee chair, Dr. Scott Dunham, for his invaluable patience and feedback throughout this journey. My sincere thanks also go to my defense committee Ting Cao, Hugh W. Hillhouse, Shijing Sun, Daniel R. Gamelin, and Lucien N. Brush for their generous provision of knowledge and expertise, which has been instrumental to my success. This research would not have been possible without the generous support of the Department of Energy, NSF MRSEC, UW Clean Energy, and the Hyak supercomputer system at UW.

I would also like to extend my heartfelt thanks to my cohort members and department advisors, especially my office mates, for their research assistance, editing help, and moral support. I am equally grateful to the research assistants and study participants from the university, whose contributions and inspiration have profoundly impacted my work.

Finally, I would like to express my deepest gratitude to my family, especially my parents. Their unwavering belief in me has kept my spirits and motivation high throughout this process.

DEDICATION

To my beloved parents, whose unconditional love and countless sacrifices have been my beacon of strength and motivation throughout this journey.

To my esteemed advisor, Dr. Dunham, for your unwavering guidance and relentless support that have been indispensable in navigating this academic voyage.

To my resilience, for withstanding the stress and countless challenges that were part and parcel of this profound journey.

Chapter 1

INTRODUCTION

The market for thin-film solar cell has grown rapidly in recent years due to predicted future energy trends and the drive toward zero carbon emissions. Traditional photovoltaic technologies, such as Si and GaAs solar cells, demonstrate high efficiencies (20% - 30%), but they require high quality crystalline material, which largely increases the cost of production. Thin-film solar cell, however, have higher defect tolerance, lower temperature coefficient, tunable bandgap, high absorption coefficient, and lower material costs. In addition, they are light and flexible, and thus can be seamlessly integrated into windows, cars, or road surfaces. Two of the leading thin-film solar cell materials are $Cu(In,Ga)Se_2$ (CIGS) and CdSeTe.

Despite tremendous efforts in thin-film solar cell technology in recent years, performance of thin-film solar cell devices are still well below theoretical limits. Defect-assisted Shockley-Read-Hall (SRH) recombination, as well as grain boundary and interface recombination significantly limit carrier lifetime and impact solar cell device performance. A detailed understanding of the coupling of manufacturing processes to device behavior through Technology Computer-Aided Design (TCAD) modeling is essential to fully understand and optimize thin-film solar cells technology.

However, while TCAD modeling has been successful in optimizing silicon-based electronic devices, its application to chalcogenide thin-film solar cells is still limited. This limitation is partly due to the complex nature of intrinsic point defects and extrinsic dopants in compound semiconductors. As a result, many TCAD models for chalcogenide solar cells rely on effective lifetime and carrier density parameters, which fail to capture the many of the connections between device behavior and the fabrication process. In this work, device modeling combined with *ab initio* quantum mechanical calculations has helped to guide device performance improvement through quantifying the influences of material properties and variant device designs.

In chapter 2, some background on chalcogenide photovoltaics and the methodology used in this work are discussed, including Density Functional Theory (DFT), TCAD, and Machine Learning. Some basic theories involving semiconductors, defects and generation-recombination are then introduced.

In chapters 3 and 4, process and device simulations are coupled to develop TCAD model for thin-film solar cells. In general, density functional theory calculations are firstly used to determine the formation energy and diffusion energy barriers of dominant defects in CIGS and CdSeTe. Next, continuum reaction-diffusion models are developed to analyze the redistribution of defects during manufacturing processes. We estimate defect capture cross sections using a first-principles based approach. These cross sections are combined with our calculated defect profiles and trap energy levels to parameterize a Shockley-Read-Hall (SRH) recombination model, which we implement into a device simulator to predict carrier lifetimes and device performance. In that way, a predictive TCAD model is built to predict and optimize the performance of CIGS and CdSeTe solar cells.

In chapter 5, AI-driven methods are applied for point defect formation energies and charge transition level prediction. Descriptors based on tabulated elemental properties, defect coordination environment, and relevant semiconductor properties are used to train ML regression models for the DFT computed data. Kernel ridge, Gaussian process, random forest, and neural network regression models are firstly applied to screen impurities with lower formation energy than dominant native defects in all compounds. Moreover, advanced convolutional neural network model based on defect supercell structure is utilized to augment the basic ML framework to achieve quantitative prediction accuracy.

Chapter 2
BACKGROUND

2.1 Chalcogenide Photovoltaics

2.1.1 CIGS

Photovoltaic devices are essential for renewable clean energy systems. While silicon-based solar modules currently dominate the market, various emerging techniques based on thin-film semiconductors are rapidly developing. Among these technologies, chalcopyrite $Cu(In,Ga)Se_2$ (CIGS) show excellent light conversion efficiency[166]. The world record efficiency of CIGS is 23.6%[160], which is still below theoretical efficiency limit of 33%. Defects in CIGS such as Group III antisites (In_{Cu}, Ga_{Cu}) and copper antisites (Cu_{III}) often possess deep energy levels within the energy band gap which can act as Shockley-Read-Hall (SRH) recombination centers limiting solar cell performance. The distribution of these defects is highly dependent on atomic composition[187]. Therefore, it is crucial to develop predictive models to explore how composition variations in CIGS absorber layers as well as growth and annealing conditions impact defect profiles and thus affect device performance. However, current TCAD tools lack models that couple composition-dependent defect distributions with device simulations.

2.1.2 CdSeTe

CdTe and its alloy, CdSeTe, are widely regarded as promising materials for thin-film solar cells, attributed to their high light absorption coefficients, low fabrication costs, and optimal band gap[108]. Recent advancements have pushed the efficiency of these materials to a remarkable 22.3%[114]. Typically, undoped CdTe films exhibit only weak p-type conductivity. To enhance p-type conductivity, Cu doping has been traditionally employed; however, this can lead to device instability and degradation due to the high diffusivity of Cu [174].

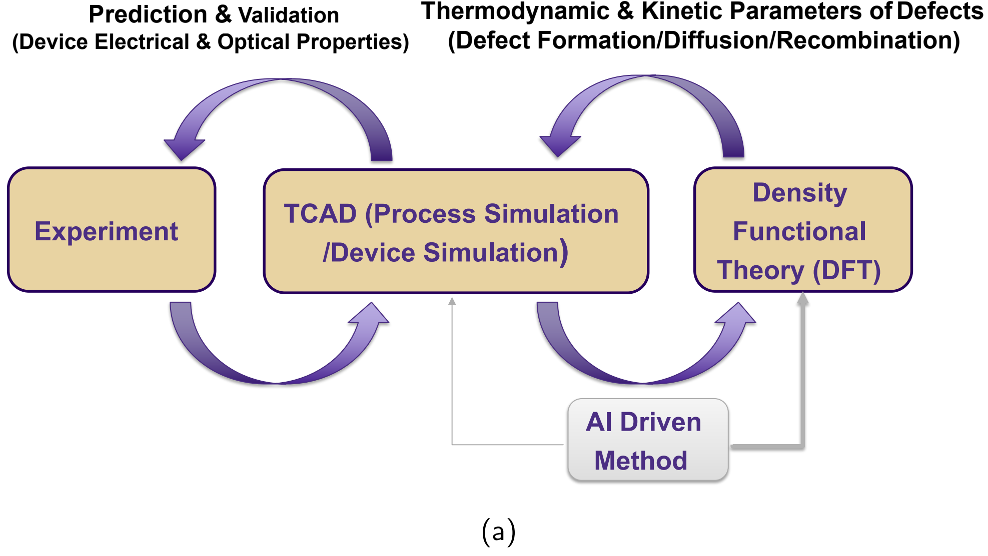
Recent efforts have demonstrated that Group V doping—utilizing elements such as P, As, and Sb—yields higher carrier densities, improved efficiency and enhanced long-term stability [23, 130, 92, 22]. These elements, when substituted for Te, act as effective acceptors [174]. Despite ongoing progress, the defect properties of Group V doped CdSeTe, particularly in alloys, remain poorly understood. Furthermore, the development of a comprehensive TCAD model for CdSeTe is essential to fully understand and optimize these materials. Current TCAD models need to incorporate the effects of Group V doping and accurately simulate the impact of these dopants on device performance.

2.2 Methodology

In this study, we combine multiple modeling methods, including DFT, process simulation, device simulation and AI-driven method. As shown in Fig. 2.1a, thermodynamic and kinetic parameters of defects are obtained from DFT. These parameters are then implemented into Technology computer-aided design (TCAD), which is a branch of electronic design automation that simulates semiconductor fabrication (process simulation) and device operation (device simulation). The outcome from these models is compared to experimental data for validation. If any discrepancy found, the experimental data is used as reference to optimize our models. AI-driven method is also applied here to accelerate/replace time consuming modeling by providing high throughout prediction, especially for DFT computation. This section is a brief overview of these modeling methods.

2.2.1 Density Functional Theory (DFT)

Density Functional Theory provides a computational quantum mechanical modeling method applied generally in physics, chemistry and materials science to study material properties at atomic scale. There are many DFT packages available. In this study, we use plane wave basis code VASP (Vienna Ab-initio Simulation Package)[97, 96]. In this section, we will introduce how to extract critical parameters of defects, which will then be used in TCAD and machine learning.



(a)

Figure 2.1: Workflow of Modeling Methods.

2.2.1.1 Defect Formation Energy

The defect formation energy can be derived from DFT using the supercell method[50] by the following equation,

$$E^f_{k,q} = E_{k,q} - E_{\text{bulk}} - \sum_i n_i \mu_i + q(E_{\text{VBM}} + \mu_{\text{Fermi}}) + E_{\text{corr}} \ ,$$

where $E^f_{k,q}$ is the formation energy of a defect k, $E_{k,q}$ is the total energy of one defect state with charge q, E_{bulk} is the total energy of a bulk unit cell with the same size of the defected cell, and $n_i\mu_i$ is the reference energy of n_i added/removed atoms of element i at given chemical potential μ_i. E_{VBM} is the valence band maximum as computed from electronic structure of bulk supercell. E_{corr} is the charge correction energy using the FNV scheme implemented in PyCDT[21] to account for interaction between periodic images.

2.2.1.2 Migration Barrier

Defect migration barrier is one of the important parameters for defect diffusion. The nudged elastic band (NEB)[63, 62] is a method in VASP for finding saddle points and minimum energy path between initial and final state. The method is used here to identify defect migration path and associated barrier energy.

2.2.1.3 Defect-assisted Recombination

Recombination in semiconductors is a process in which carriers (electrons and holes) annihilate each other. There are mainly three types of recombination: Radiative recombination, Auger recombination and Shockley-Read-Hall (SRH) recombination.

Radiative recombination is also called band-to-band recombination. During radiative recombination, electrons jumping down from the conduction band to the valence band and emit photons. Radiative recombination is efficient in direct bandgap materials such as GaAs. However, in indirect bandgap materials, radiative recombination is less significant because it must be accompanied by a change in momentum of the electron and the generation of phonons.

In Auger recombination, the energy of recombined carriers is given to a third carrier which is then excited to a higher energy level without moving to another energy band. After the excitation, the third carrier loses its excess energy via thermal vibrations. Auger recombination is a nonradiative recombination process. It is most important at high carrier concentrations caused by heavy doping or high level injection under concentrated sunlight.

In this section, we will mainly focus on the third type of recombination: SRH recombination, which is a defect-assisted recombination process. We will introduce how to use DFT calculation to compute capture cross section for each defect. There are two types of SRH recombination: radiative and nonradiative. Configuration coordinate diagram (CCD) is used to closely resemble the actual transition of carriers. Underlying idea of the CCD is the observation that the initial and final states can typically be connected by a one-dimensional collective reaction coordinate. In fact, it is this difference in relaxation that leads to the difference between thermodynamic transition

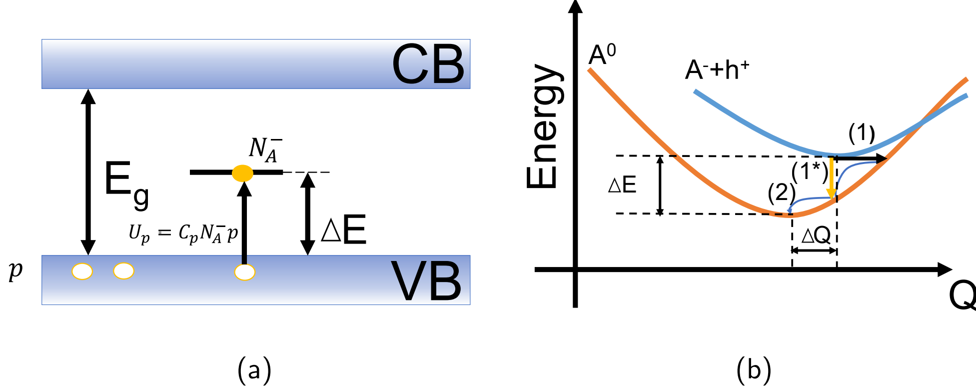


Figure 2.2: Representation of carrier capture at a defect: (a) band diagram and (b) configuration coordinate diagram. The defect in the diagram is a deep acceptor with a negative and a neutral charge state. $\triangle E$ is the ionization energy of the defect, and Q is a properly selected configuration coordinate. In (b), step(1) is the change of the electronic state due to electron-phonon coupling (nonradiative), step(1*) is the change of the electronic state due to electron-photon coupling (radiative) and step(2) is the vibration relaxation due to phonon-phonon interactions.

levels and optical levels. As shown in Fig 2.2, if a carrier capture process follows step(1)+step(2), it would be a nonradiative recombination. On the other hand, step(1*)+step(2) leads to radiative recombination. Q in Fig 2.2 is defined as

$$Q^2 = \sum M \Delta R^2_{\text{displacment}}, \tag{2.1}$$

where M is atomic mass of any element of defects, $R_{\text{displacment}}$ is the displacement vector from its equilibrium position.

The rate of SRH recombination can be estimated by the following equation,

$$R^{\text{SRH}} = \frac{np - n_i^2}{\tau_p \left[n + n_i e^{(E_j - E_i)/k_{\text{B}}T}\right] + \tau_n \left[p + n_i e^{(E_i - E_j)/k_{\text{B}}T}\right]} \tag{2.2}$$

in which, R^{SRH} is the net SRH recombination rate, n and p are electron and hole concentration, n_i is the intrinsic carrier concentration, E_j is the energy level of defect j calculated by DFT in Fig. 3.1, and E_i is the intrinsic Fermi level. The carrier lifetimes, τ_n and τ_p, of electron and holes

are given by

$$\tau_n = \frac{1}{v_{thn} N_t \sigma_n} \qquad \tau_p = \frac{1}{v_{thp} N_t \sigma_p}\ , \tag{2.3}$$

where v_{thn} and v_{thp} are the thermal velocities of electrons and holes, and N_t is the trap density. In Radiative SRH recombination, optical transitions happens due to coupling to the electric field. The capture cross sections σ_n and σ_p of electrons and holes for defects can be estimated by the following expression[39],

$$\sigma_{n/p} * v_{th} = (5.77 \times 10^{-17})(V f n_r \eta_{sp} E_{opt} \frac{|p_{if}|^2}{2m}) \text{ cm}^3 \text{ s}^{-1}, \tag{2.4}$$

where V is the volume of supercell. n_r is the index of refraction, η_{sp} accounts for the spin selection rule ($\eta_{sp} = 1$ for transitions from a spin singlet to a doublet, $\eta_{sp} = 0.5$ for a transition from a doublet to a singlet or from a triplet to a doublet), m is mass of carriers and E_{opt} is the optical transition energy corresponding to energy difference in step(1*) according to Franck-Condon approximation. $\mathbf{P}_{if}$ is the momentum matrix element for perturbed conduction-band or valence-band state ψ_i and defect state ψ_f. It is described by $\mathbf{P}_{if}(Q) = \langle \Psi_i(Q) | \hat{\mathbf{P}} | \Psi_f(Q) \rangle$. This momentum matrix element can be derived from WAVEDERF output file in VASP by setting LOPTICS=True (In fact, the interband dipole matrix elements are presented in the output, but can be easily transferred into momentum matrix elements[58]).

Nonradiative capture cross section are determined by NONRAD[5, 198] based on evaluation of electron-phonon coupling for a special phonon mode in the capture process. The equation is shown below:

$$\sigma_{n/p} * v_{th} = \frac{2\pi}{\hbar} g \sum_m w_m \sum_n \mid \Delta H^{e-ph}_{im;fn} \mid^2 \times \delta(E_{im} - E_{fn}) \text{ cm}^3 \text{ s}^{-1}. \tag{2.5}$$

Here, g is atomic configuration degeneracy. w_m is the thermal occupation of the vibrational state m of the excited state. E_{im} and E_{fn} are the energies of the initial and the final vibrational state. The electron-phonon coupling matrix elements $\Delta H^{e-ph}_{im;fn}$ are evaluated in terms of wavefunction overlaps via the NONRAD code. The delta functions in Eq. 2.5 denotes energy conservation on

the vibronic transitions. Detailed derivation and implementation of this equation can be referred to Ref. [5, 198].

The scaling function f in Eq. (2.4) and Eq. (2.5) has two contributions: the interaction between defect supercell periodic images and the Coulombic interaction between the defect and delocalized carrier is the defect is not neutral (known as Sommerfeld parameter). These two corrections have been considered in NONRAD.

2.2.2 Continuum Modeling

2.2.2.1 Process Simulation

Process simulation (implemented in Sentaurus SProcess[193] via Alagator) aims to model defect redistribution, defect complex formation processes during cooling and annealing process in semiconductor manufacture. The defect diffusion and evolution are described by continuity equations. The model of defects diffusion is based on Fick's first law: the flux of a diffusing specie is linearly dependent to the concentration gradient and electrochemical potential gradient[123],

$$J = -D\nabla C - C\mu\nabla U \ , \tag{2.6}$$

where D is the diffusivity, C is the number density of particles, μ is the mobility of the diffusing species, and U is the local electrochemical potential. The potential is dependent on electric potential and local chemical compositions, so the variation of electric field and composition causes drift of species. Poisson's equation is used to derive the electric potential ψ from the charge density ρ,

$$\nabla^2\psi = \frac{\rho}{\epsilon} \ , \tag{2.7}$$

$$\rho = p - n + \sum_i z_i N_{Di}^+ - \sum_j z_j N_{Aj}^- \ , \tag{2.8}$$

in which p is the hole density, n is the electron density, N_D^+ is the density of defects acting as donors and N_A^- is the density of defects acting as acceptors. z_i is the charge state of donor N_{Di}^+ and z_j

is the charge state of acceptor N_{Aj}^{-}. To incorporate the reactions between defects when computing the defect concentration's change with time, Fick's second law is augmented by a reaction rate term R,

$$\partial_t C_\gamma = -\nabla \cdot J_\gamma - \sum_l \eta_{\gamma l} R_l \ , \tag{2.9}$$

$$R = k_f \left[\prod_{\alpha=1}^{n} C_\alpha - \frac{1}{K} \left(\frac{p}{n_i} \right)^h \prod_{\beta=1}^{m} C_\beta \right] \ , \tag{2.10}$$

$$k_f = 4\pi r \sum_\alpha D_\alpha \ , \tag{2.11}$$

$$k_f = \sigma_{n,p} \upsilon_{th}, \tag{2.12}$$

$$K = \frac{\theta_l}{C_s^{m-n}} e^{-\Delta E_l^f / k_b T} \ , \tag{2.13}$$

where η_γ is the number of species γ consumed by reaction l, and k_f is the forward reaction rate coefficient which is diffusion limited and estimated by a simple kinetic approximation. r is the interaction radius and D_α is the diffusivity of the reactant α. The reaction equilibrium constants K are calculated from the formation energy difference ΔE_l^f of products and reactants. C_s is the concentration of possible defect sites and θ_l is the entropy term. For defect interaction reactions, Eq (2.11) is used to estimate reaction rate. For defect ionization reactions, Eq (2.12) is used, where $\sigma_{n,p}$ is defect electron/hole capture cross section and υ_{th} is thermal velocity. The electron/hole capture cross sections are estimated via the NONRAD method as mentioned in Section 2.2.1.3.

The diffusivity D_α of the defects in above equation is computed based on random walk theory[126], in which defects diffuse via a random walk on the sublattice:

$$D_\alpha = gz\lambda^2 Tf, \tag{2.14}$$

where g is a geometric factor determined by the dimensionality of the crystal lattice, z is the coordination number of the atomic sites in the lattice. λ is the nearest neighbor hopping distance,

which represents the distance between neighboring defect sites which involved in diffusion. T is the jump frequency of the defects, it is denoted as:

$$T = vexp(-\frac{E_m}{kT}), \tag{2.15}$$

in which v is the attempt frequency, which is estimated to be in the order of the Debye frequency $10^{13}s^{-1}$. E_m is the migration barrier of defect diffusion calculated via DFT.

2.2.2.2 Device Simulation

Device simulation (implemented in Sentaurus SDevice[192]) describes carriers behavior across the device and thereby providing optical and electrical characteristics of electronic devices. Device models are based on Poisson's equation (Eq. 2.7) and continuity equations. The former calculates the electrostatic potential in the device. The continuity equations solve the carrier concentration:

$$\frac{\partial n}{\partial t} = \nabla \cdot j_n/q + (G - R) \tag{2.16}$$

$$\frac{\partial p}{\partial t} = \nabla \cdot j_p/q + (G - R) \tag{2.17}$$

where j_n and j_p denote electron and hole current densities, G and R represent the carrier generation and recombination rate, respectively. Drift and diffusion of carriers can be described by:

$$j_n = q\mu_n n\varepsilon + qD_n\frac{\partial n}{\partial x} \tag{2.18}$$

$$j_p = q\mu_p p\varepsilon - qD_n\frac{\partial p}{\partial x} \tag{2.19}$$

where μ_n and μ_p are electron and hole mobility, D_n and D_p are diffusivities of electron and hole, and ε is the electric field derived from Poisson's equation.

Process simulation provides information about doping concentration and defect densities in a device. This information will be coupled into device simulator. Shallow level defects are considered as

dopants via Incomplete Ionization model, while deep level defects are described by Shockley-Read-Hall recombination model to trap carriers. Capture cross sections of these defects are computed via the procedure mentioned in Section 1.1.1. Physical Model Interface is exploited to smoothly combine process simulation with device simulator, and account for the variation in defect transition levels due to material properties.

Another important part for device simulation is optical modeling, which is critical for modeling optoelectronic devices such as solar cells. The optical simulator models light propagation through the device layer and calculates the amount of light absorption and corresponding carrier generation rate G in Eq. 2.16 and Eq. 2.17. Specifically, quantum yield model is used to describe how many of the absorbed photons are converted to generate electron-hole pairs. This model considers the band gap. If the excitation energy is greater than or equal to the bandgap energy E_g, the quantum yield is set to one; otherwise, it is set to zero.

In this work, transfer-matrix-method (TMM)[192] is used to simulate the optical generation of CuInGaSe and CdSeTe solar cells. ElementWise option is turned on to account for the variation of materials properties within the device. The complex refractive index and other parameters for optical modeling come from reported experimental data. Excellent match between experimental data and simulation results are achieved, which establishes the accuracy and validity of this method.

The final output of the device simulation are distribution of electric potentials, carrier densities, and optical and electrical characteristics which are of interest. These data provides understanding of the device behavior and insight for optimization of the device performance. An example SDevice *cmd* file is provided in appendix A.

2.2.3 AI-Driven Methods

Recently, with tremendous material data and powerful supercomputer available, data-driven models, also called machine learning (ML) or artificial intelligence (AI) expand their application in semiconductor industry. In this section, we will discuss critical components of a common ML pipeline in semiconductor material design.

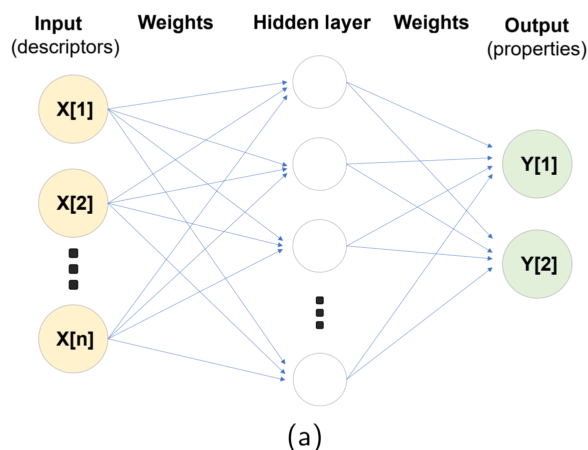


Figure 2.3: Schematic of neural network machine learning algorithms.

2.2.3.1 Training Data

DFT is a widely used source of training data and descriptors among material discovery. Recently, DFT calculation of large-scale materials has become more accessible as computational power has increased. The construction of high-throughput DFT databases now has become a major contributor to material informatics. DFT provides a supplement to experimental data as it is well structured, computationally accessible, and self-consistent. In this work, DFT is used to construct baseline of training data.

The selection of features can have a large impact on ML model performance. Normally, empirical descriptors are typical used in ML for materials discovery. For example, physical features of atoms, such as electronegativity, ionic radius. These features are encoded into a machine-readable format, normally as a vector of descriptors. Including as many features as possible is one route to ensure that no known physics is excluded in a model. However, the existence of redundant and/or meaningless features could deteriorate model performance. Therefore, feature dimensionality reduction methods, such as k-nearest neighbor and principal component analysis, are often used to obtain a reduced set of features which preserve sufficient information.

2.2.3.2 Basic Machine Learning Algorithms

After building a reliable database, it is necessary to select suitable machine learning algorithms. Popular ML algorithms include least squares method, Gaussian process, random forest, and neural network. Each algorithm has its own pros and cons. Here, we introduce several algorithms that are being exploited in our research.

Multiple linear regression (MLR) is a basic least squares method. Given a vector of properly standardized features $X^T = (X_1, X_2, ..., X_p)$, and calculated output vector Y, the matrix of coefficients β corresponding to each feature and output is determined by minimizing the least square error $|Y - X^T\beta|^2$. While MLR yields an unbiased predictor, it is prone to overfitting when several features are highly correlated with the output. To address this issue, we use three shrinkage methods, namely LASSO (least absolute shrinkage and selection operator) regression, ridge regression, and elastic net regression [60], all of which yield a biased predictor but with lower variance, leading to less overfitting compared to standard least square. Ridge regression shrinks the coefficients β by imposing an L_2 penalty whereas LASSO uses an L_1 penalty [139]. Elastic Net is another regularized linear regression technique that combines both L_1 penalty and L_2 penalty. Typically, β shrinkage inside LASSO regression progresses more severely compared to the other two approaches, and some of the coefficient are brought down to 0.

Kernel ridge regression (KRR) is another regression technique which uses the kernel trick to solve a non-linear problem in a linear fashion. The original low-dimensional features are used as input and mapped to a high-dimensional kernel space in which they can be linearly interpreted. There are different possible choices for the kernel function, such as polynomial, radial basis function (RBF), and Laplacian.

Another non-linear regression technique that employs the "kernel trick" is Gaussian process regression (GPR) [213, 177]. GPR uses the kernel and the observation to define a likelihood function on account of the covariance of a prior distribution over the target functions. The prior and likelihood function is assumed to have a Gaussian distribution. Based on Bayes' theorem [161], we get a predictive posterior distribution, from which we can attain a point prediction using its

mean, and an uncertainty value using its variance. A major difference between GPR and KRR is that GPR can internally choose each kernel's hyperparameters by applying gradient-ascent on the marginal likelihood function, while KRR requires a grid or random search using a loss function.

The decision tree method splits source dataset (root node) into several subsets (successor children). RF consists of multiple decision tree models based on randomly bootstrapped subsets of the traning data. Predictions are made according to the collected prediction of the decision trees. The main advantages of RF are robustness to sparse data and easy to perform feature selection.

NN is another type of ML algorithm that imitates biological neurons. A NN model consists of many interconnected logical gates, called perceptrons. A typical network consists of an input layer, hidden layer and output layer as illustrated in Fig. 2.3a. NN is a highly flexible algorithm. If we allow enough nodes in hidden layer, a NN model can represent any function, which is also likely to overfit. Some techniques are generally applied to avoid it, like regularization, early stopping, and use of a better topology.

2.2.3.3 Cross-Validation

The main purpose of machine learning models is prediction; Thus, it is critical to measure the performance of ML models. Cross-validation (CV) is the standard approach for testing the quality of the ML models. In CV, models are trained on a subset of data and then employed to predict values for regression problems or label data for classification problems for a holdout dataset. CV is generally used to estimate an ML model's ability to generalize to unseen data.

The most widely used technique is k-fold, in which the training data are randomly separated into k folds. The ML models are trained on k-1 folds and used to predict the holdout fold over k iterations. It is the most widely used generalization error estimation method. There are also some other techniques, such as leave-p-out cross validation (LPO)[164], Monte Carlo cross validation (MCCV)[222], etc.

Chapter 3

COUPLED PROCESS AND DEVICE MODELING OF CIGS SOLAR CELLS

In this work, we employ a multi-scale modeling framework, combining first principles calculations, continuum models, and device simulation tools to predict CIGS defect distributions and device performance. Density functional theory (DFT) calculations are utilized to determine the formation energies and diffusion energy barriers of defects as well as binding energies of defect complexes. Equilibrium defect concentrations are calculated through compositionally constrained thermodynamics (CCT) calculator[133] for any given composition. After that, continuum models are developed for the defect diffusion and redistribution processes. Finally, the process simulation results are coupled with device simulation in which the deep level defects limit the carrier lifetime and solar cell efficiency via SRH recombination. In the end, a predictive model is built to help interpret experimental data and provide new perspectives for engineering CIGS composition with desirable functionality.

3.1 Process Simulation in CIGS

We used the DFT code VASP with projected augmented wave (PAW) approach[98]. To correct the bandgap error, the Heyd-Scuseria-Ernzerhof (HSE06)[65] is used to calculate defect formation energies and diffusion energy barriers[97, 96]. The defect formation energy can be derived from DFT using the method mentioned in Section 2.2.1.1. Under dilute approximation, defect concentrations per CIGS lattice sites are given by $x_{k,q} = \frac{\theta_{k,q}}{N^{total}} e^{-H^f_{k,q}/k_{\mathrm{B}}T}$, where $\theta_{k,q}$ is the degeneracy factor counting the number of equivalent defect configurations. N^{total} is the total number of lattice sites in a perfect material. The formation entropy is neglected. Since the atomic fractions f_α are experimentally measurable, one can obtain deviations from perfect stoichiometry f^p_α under given defect concentrations in the material using the following relation

$$f_\alpha = \frac{f_\alpha^p - \sum_{k,q} x_{k,q} n_k^\alpha}{1 - \sum_{k,q} x_{k,q}\sigma_{k,q}} \ , \tag{3.1}$$

where $x_{k,q}$ is defect concentrations per CIGS lattice sites, n_k^α is the number of atoms of type α added or removed from the system when one defect k exists. $\sigma_{k,q} = \sum_\alpha n_k^\alpha$ is the atom amount differences in a system with and without the defect k. If atomic fractions and standard reference chemical potentials are provided, one can determine the defect concentration, chemical potentials for each element and the Fermi level by solving the set of equations defined by (3.1). We refer to this method as compositionally constrained thermodynamics (CCT) [133]. The CCT model in this work contains major defects reported in literature[157, 187], namely, vacancies (V_{Cu}, V_{III}, V_{Se}), interstitial (Cu_i, In_i, Ga_i, Se_i), antisites (In_{Cu}, Ga_{Cu}, Cu_{III}), and typical defect complexes (V_{Cu}+In_{Cu}, $2V_{Cu}$+In_{Cu}, V_{Cu}+Ga_{Cu}, $2V_{Cu}$+Ga_{Cu}, V_{Cu}+V_{Se}, V_{Cu}+In_{Cu}+Cu_{III}, V_{Cu}+Ga_{Cu}+Cu_{III}). The transition level $\varepsilon(q_1/q_2)$ of these defects are calculated by

$$\varepsilon(q_1/q_2) = \frac{E^f_{k,q_1}(\mu_{\mathrm{Fermi}} - VBM = 0) - E^f_{k,q_2}(\mu_{\mathrm{Fermi}} - VBM = 0)}{q_2 - q_1} \ . \tag{3.2}$$

These defect transition levels are depicted in Fig. 3.1. Diffusion energy barriers were calculated using the climbing-image nudged elastic band method[63]. Then the continuum models are built in accordance with the DFT results and implemented within the Sentaurus 1/2/3D process simulator via the Alagator scripting language[193]. Through CCT and the continuum models, we can obtain the defect concentrations across the thin film.

3.1.1 Defect Diffusion in CIGS

We have reported in previous work that copper in Cu-poor CIGS diffuses via copper vacancies[185, 186]. However, the diffusion mechanism for group III is different since there are very few group III vacancies in Cu-poor CIGS. We found the group III antisite formation energy to be low, and there is a significant binding energy between group III antisites and copper vacancies (Table 3.3), which are abundant in Cu-poor CIGS. Therefore, it is not difficult for copper vacancies to find

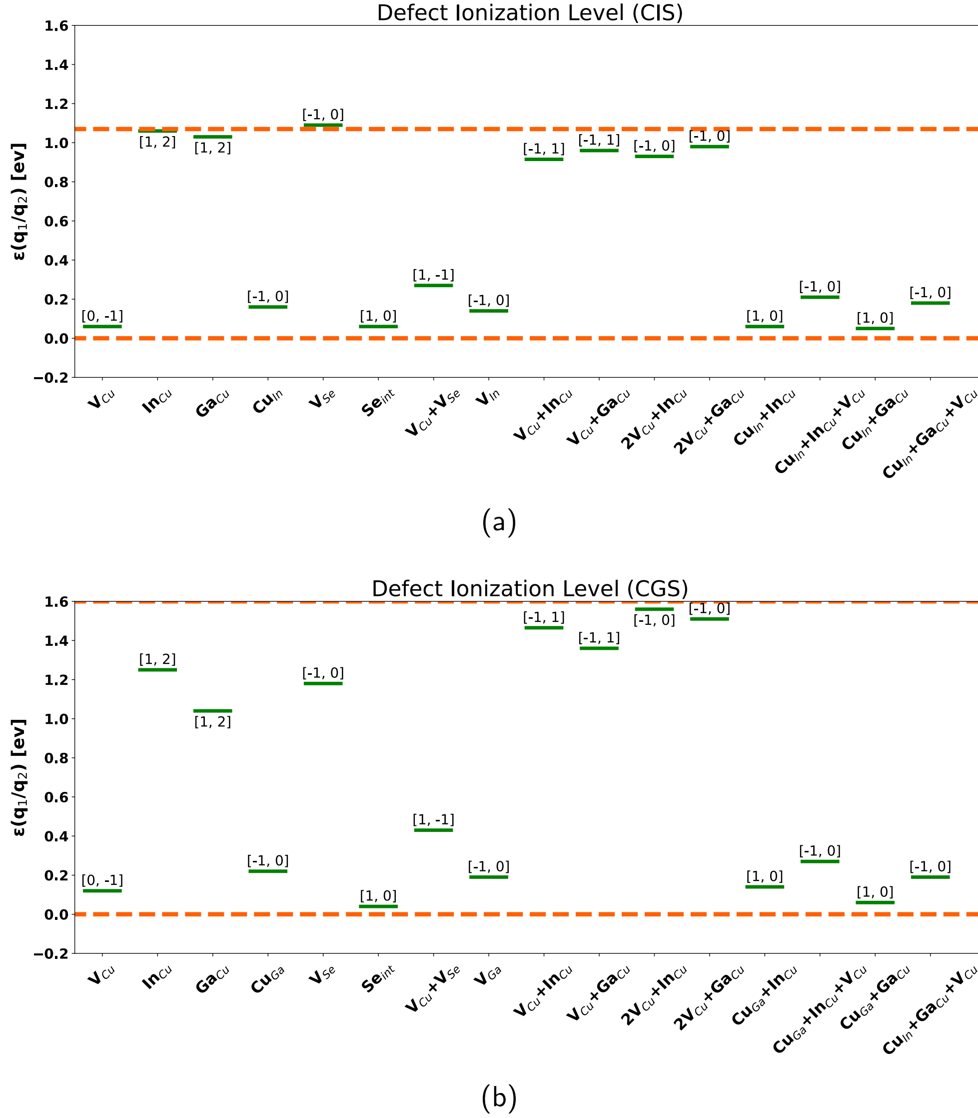


Figure 3.1: Defect transition levels in (a) $CuInSe_2$; (b) $CuGaSe_2$. Dashed lines indicate valence band maximum and conduction band minimum.

group III antisites and form mobile defect complexes in the material. Group III antisites may bind with either one or two copper vacancies due to their respective charges (nominally +2 for group III antisites and -1 for Cu vacancies) and substantial Coulomb attraction. We considered two energetically competitive routes for the long-range diffusion of these vacancy complexes (shown in Table 3.1), involving either (i) a vacancy restricted to nearest-neighbor hops out to a third

Table 3.1: CIGS Defect Diffusion Mechanism

Migration barrier	E_m in eV
$V_{Cu} \rightarrow V_{Cu}$	1.10(-1), 1.12(+0)
$(In_{Cu}+V_{Cu}) \rightarrow (In_{Cu}+V_{Cu})$ [Direct]	1.21(+), 1.26(+0)
$(In_{Cu}+2V_{Cu}) \rightarrow (In_{Cu}+2V_{Cu})$ [via 3NN]	1.24(+0)
$(Ga_{Cu}+V_{Cu}) \rightarrow (Ga_{Cu}+V_{Cu})$ [Direct]	1.21(+), 1.19(+0)
$(Ga_{Cu}+2V_{Cu}) \rightarrow (Ga_{Cu}+2V_{Cu})$ [via 3NN]	1.75(+0)

nearest-neighbor (3NN) site from an antisite, or (ii) a nearest-neighbor vacancy hopping directly to another nearest-neighbor site. The latter "direct" mechanism (ii) requires a second nearest-neighbor hopping distance and effectively bypasses mechanism (i). Based on DFT and CCT[55, 186], we determined that single vacancy complexes mainly diffuse through this direct mechanism, while mechanism (i) is the dominant process for the diffusion of double vacancy complexes.

3.1.2 Continuum Models for Defect Diffusion

The continuum model is based on Section 2.2.2. The formation energy of each defect is different in $CuInSe_2$ and $CuGaSe_2$ as In-Ga composition could affect E_{VBM} in Eq. (2.1). Thus, we assume that the formation energy of one defect can be linearly interpolated by [Ga]/([In] + [Ga]) ratio (GGI) and defect formation enthalpy $H^f_{k,q}$ is also dependent on GGI ratio. C_α is the concentration per unit volume of reactant α, C_β is the concentration per unit volume of product β, n_i is the intrinsic carrier concentration and h is the net number of holes released by the reaction. It is noteworthy that $x_{k,q} = C_{k,q}/C_s$, where C_s is the concentration of possible defect sites.A list of defect reactions is shown in Table 3.2. We assume there are no fluxes and transfer across the interface and thus Neumann boundary condition are used for all defects.

3.1.3 Comparison Between Simulation and Experiment

As we reported previously, the copper diffusion model has good consistency with experiments [186]. To verify the group III (In, Ga) diffusion mechanisms, we have performed a simulation under the

Table 3.2: Defect Reactions in Continuum Model (charge exchange reactions are assumed to be fast for process simulation, so all equations are listed just in terms of holes)

Exchange Reaction	**Complex Formation**	**Ionization Reaction**
$In_{Cu}^{2+} + V_{III}^{2-} \rightarrow V_{Cu}^{0} + In_{III}$	$V_{Cu}^{0} + In_{Cu}^{2+} \rightarrow (In_{Cu}+V_{Cu})^{+} + h^{+}$	$V_{Cu}^{0} \rightarrow V_{Cu}^{-} + h^{+}$
$In_{Cu}^{2+} + V_{III}^{-} \rightarrow V_{Cu}^{0} + In_{III} + h^{+}$	$V_{Cu}^{-} + In_{Cu}^{2+} \rightarrow (In_{Cu}+V_{Cu})^{+}$	$V_{III}^{0} \rightarrow V_{III}^{-} + h^{+}$
$In_{Cu}^{2+} + V_{III}^{0} \rightarrow V_{Cu}^{0} + In_{III} + 2p^{+}$	$V_{Cu}^{0} + (In_{Cu}+V_{Cu})^{+} \rightarrow (In_{Cu}+2V_{Cu})^{0} + h^{+}$	$V_{III}^{-} \rightarrow V_{III}^{2-} + h^{+}$
$V_{Cu}^{0} + Cu_{III}^{-} \rightarrow V_{III}^{-}$	$V_{Cu}^{-} + (In_{Cu}+V_{Cu})^{+} \rightarrow (In_{Cu}+2V_{Cu})^{0}$	$Cu_{III}^{0} \rightarrow Cu_{III}^{-} + h^{+}$
$V_{Cu}^{-} + Cu_{III}^{-} \rightarrow V_{III}^{2-}$	$V_{Cu}^{-} + (In_{Cu}+V_{Cu})^{+} \rightarrow (In_{Cu}+2V_{Cu})^{0} + h^{+}$	$Cu_{III}^{-} \rightarrow Cu_{III}^{2-} + h^{+}$
$Ga_{Cu}^{2+} + V_{III}^{2-} \rightarrow V_{Cu}^{0} + Ga_{III}$	$V_{Cu}^{0} + Ga_{Cu}^{2+} \rightarrow (Ga_{Cu}+V_{Cu})^{+} + h^{+}$	$V_{Se}^{0} \rightarrow V_{Se}^{-} + h^{+}$
$Ga_{Cu}^{2+} + V_{III}^{-} \rightarrow V_{Cu}^{0} + Ga_{III} + h^{+}$	$V_{Cu}^{-} + Ga_{Cu}^{2+} \rightarrow (Ga_{Cu}+V_{Cu})^{+}$	$V_{Se}^{0} + h^{+} \rightarrow V_{Se}^{+}$
$Ga_{Cu}^{2+} + V_{III}^{0} \rightarrow V_{Cu}^{0} + Ga_{III} + 2p^{+}$	$V_{Cu}^{0} + (Ga_{Cu}+V_{Cu})^{+} \rightarrow (Ga_{Cu}+2V_{Cu})^{0} + h^{+}$	$(V_{Cu}+V_{Se})^{+} \rightarrow (V_{Cu}+V_{Se})^{0} + h^{+}$
$(In_{Cu}+V_{Cu})^{+} + Cu_{III}^{-} \rightarrow (In_{Cu}+Cu_{III})^{+} + V_{Cu}^{-}$	$V_{Cu}^{-} + (Ga_{Cu}+V_{Cu})^{+} \rightarrow (Ga_{Cu}+2V_{Cu})^{0}$	$(V_{Cu}+V_{Se})^{-} + h^{+} \rightarrow (V_{Cu}+V_{Se})^{0}$
$(In_{Cu}+V_{Cu})^{+} + Cu_{III}^{-} \rightarrow (In_{Cu}+Cu_{III})^{0} + V_{Cu}^{0}$	$V_{Se}^{0} + V_{Cu}^{-} \rightarrow (V_{Cu}+V_{Se})^{-}$	$In_{Cu}^{2+} \rightarrow In_{Cu}^{+} + h^{+}$
$(In_{Cu}+V_{Cu})^{+} + Cu_{III}^{0} \rightarrow (In_{Cu}+Cu_{III})^{2+} + V_{Cu}^{-}$	$V_{Se}^{0} + V_{Cu}^{0} + h^{+} \rightarrow (V_{Cu}+V_{Se})^{+}$	$In_{Cu}^{+} \rightarrow In_{Cu}^{0} + h^{+}$
$(In_{Cu}+2V_{Cu})^{0} + Cu_{III}^{-} \rightarrow (In_{Cu}+Cu_{III})^{+} + 2V_{Cu}^{-}$	$(In_{Cu}+V_{Cu})^{+} + Cu_{III}^{0} \rightarrow (In_{Cu}+Cu_{III}+V_{Cu})^{+}$	$Ga_{Cu}^{2+} \rightarrow Ga_{Cu}^{+} + h^{+}$
$(In_{Cu}+2V_{Cu})^{0} + Cu_{III}^{-} \rightarrow (In_{Cu}+Cu_{III})^{0} + V_{Cu}^{0} + V_{Cu}^{-}$	$(In_{Cu}+V_{Cu})^{+} + Cu_{III}^{-} \rightarrow (In_{Cu}+Cu_{III}+V_{Cu})^{0}$	$Ga_{Cu}^{+} \rightarrow Ga_{Cu}^{0} + h^{+}$
$(In_{Cu}+2V_{Cu})^{0} + Cu_{III}^{0} \rightarrow (In_{Cu}+Cu_{III})^{2+} + 2V_{Cu}^{-}$	$(In_{Cu}+V_{Cu})^{+} + Cu_{III}^{2-} \rightarrow (In_{Cu}+Cu_{III}+V_{Cu})^{-}$	$(In_{Cu}+Cu_{III})^{2+} \rightarrow (In_{Cu}+Cu_{III})^{+} + h^{+}$
$(In_{Cu}+2V_{Cu})^{0} + Cu_{III}^{0} \rightarrow (In_{Cu}+Cu_{III}+V_{Cu})^{+} + V_{Cu}^{-}$	$(Ga_{Cu}+V_{Cu})^{+} + Cu_{III}^{0} \rightarrow (Ga_{Cu}+Cu_{III}+V_{Cu})^{+}$	$(In_{Cu}+Cu_{III})^{+} \rightarrow (In_{Cu}+Cu_{III})^{0} + h^{+}$
$(In_{Cu}+2V_{Cu})^{0} + Cu_{III}^{-} \rightarrow (In_{Cu}+Cu_{III}+V_{Cu})^{0} + V_{Cu}^{-}$	$(Ga_{Cu}+V_{Cu})^{+} + Cu_{III}^{-} \rightarrow (Ga_{Cu}+Cu_{III}+V_{Cu})^{0}$	$(In_{Cu}+Cu_{III}+V_{Cu})^{+} \rightarrow (In_{Cu}+Cu_{III}+V_{Cu})^{0} + h^{+}$
$(In_{Cu}+2V_{Cu})^{0} + Cu_{III}^{2-} \rightarrow (In_{Cu}+Cu_{III}+V_{Cu})^{-} + V_{Cu}^{-}$	$(Ga_{Cu}+V_{Cu})^{+} + Cu_{III}^{2-} \rightarrow (Ga_{Cu}+Cu_{III}+V_{Cu})^{-}$	$(In_{Cu}+Cu_{III}+V_{Cu})^{0} \rightarrow (In_{Cu}+Cu_{III}+V_{Cu})^{-} + h^{+}$
$(Ga_{Cu}+V_{Cu})^{+} + Cu_{III}^{-} \rightarrow (Ga_{Cu}+Cu_{III})^{+} + V_{Cu}^{-}$		$(Ga_{Cu}+Cu_{III})^{2+} \rightarrow (Ga_{Cu}+Cu_{III})^{+} + h^{+}$
$(Ga_{Cu}+V_{Cu})^{+} + Cu_{III}^{-} \rightarrow (Ga_{Cu}+Cu_{III})^{0} + V_{Cu}^{0}$	**Vacancy-Interstitial Pair Annihilation**	$(Ga_{Cu}+Cu_{III})^{+} \rightarrow (Ga_{Cu}+Cu_{III})^{0} + h^{+}$
$(Ga_{Cu}+V_{Cu})^{+} + Cu_{III}^{0} \rightarrow (Ga_{Cu}+Cu_{III})^{2+} + V_{Cu}^{-}$		$(Ga_{Cu}+Cu_{III}+V_{Cu})^{+} \rightarrow (Ga_{Cu}+Cu_{III}+V_{Cu})^{0} + h^{+}$
$(Ga_{Cu}+2V_{Cu})^{0} + Cu_{III}^{-} \rightarrow (Ga_{Cu}+Cu_{III})^{+} + 2V_{Cu}^{-}$		$(Ga_{Cu}+Cu_{III}+V_{Cu})^{0} \rightarrow (Ga_{Cu}+Cu_{III}+V_{Cu})^{-} + h^{+}$
$(Ga_{Cu}+2V_{Cu})^{0} + Cu_{III}^{-} \rightarrow (Ga_{Cu}+Cu_{III})^{0} + V_{Cu}^{0} + V_{Cu}^{-}$	$V_{Cu}^{-} + Cu_{i}^{+} \rightarrow (Cu_{Cu})$	
$(Ga_{Cu}+2V_{Cu})^{0} + Cu_{III}^{0} \rightarrow (Ga_{Cu}+Cu_{III})^{2+} + 2V_{Cu}^{-}$	$V_{Cu}^{0} + Cu_{i}^{+} \rightarrow (Cu_{Cu}) + h^{+}$	
$(Ga_{Cu}+2V_{Cu})^{0} + Cu_{III}^{0} \rightarrow (Ga_{Cu}+Cu_{III}+V_{Cu})^{+} + V_{Cu}^{-}$		
$(Ga_{Cu}+2V_{Cu})^{0} + Cu_{III}^{-} \rightarrow (Ga_{Cu}+Cu_{III}+V_{Cu})^{0} + V_{Cu}^{-}$		
$(Ga_{Cu}+2V_{Cu})^{0} + Cu_{III}^{2-} \rightarrow (Ga_{Cu}+Cu_{III}+V_{Cu})^{-} + V_{Cu}^{-}$		

conditions reported by Beckers *et al*[12]. They tracked indium tracer diffusion from a thin In isotope layer deposited on the top of CIGS. Experimental data extracted from the paper is plotted in Fig. 3.2a and Fig 3.2b. Since there are few group III vacancies and In diffuses mainly via the copper sublattice, we assume that the initial indium tracer resides on the copper sublattice. Our continuum model simulated curves are plotted in Fig. 3.2c and Fig. 3.2d. Moreover, the authors also tracked tracer diffusion for different temperatures and diffusion times, and we compared our models with their non-linear fitting curves. The simulation results are in good agreement with the experimental measurements and supports the group III diffusion mechanism as proposed previously in Section 3.1.1[55].

3.2 *Coupling of Process and Device Simulations*

The defect distributions obtained from process simulation can be implemented into Sentaurus Device simulator [192]. Specifically, ionization model, which provides an ionization probability model

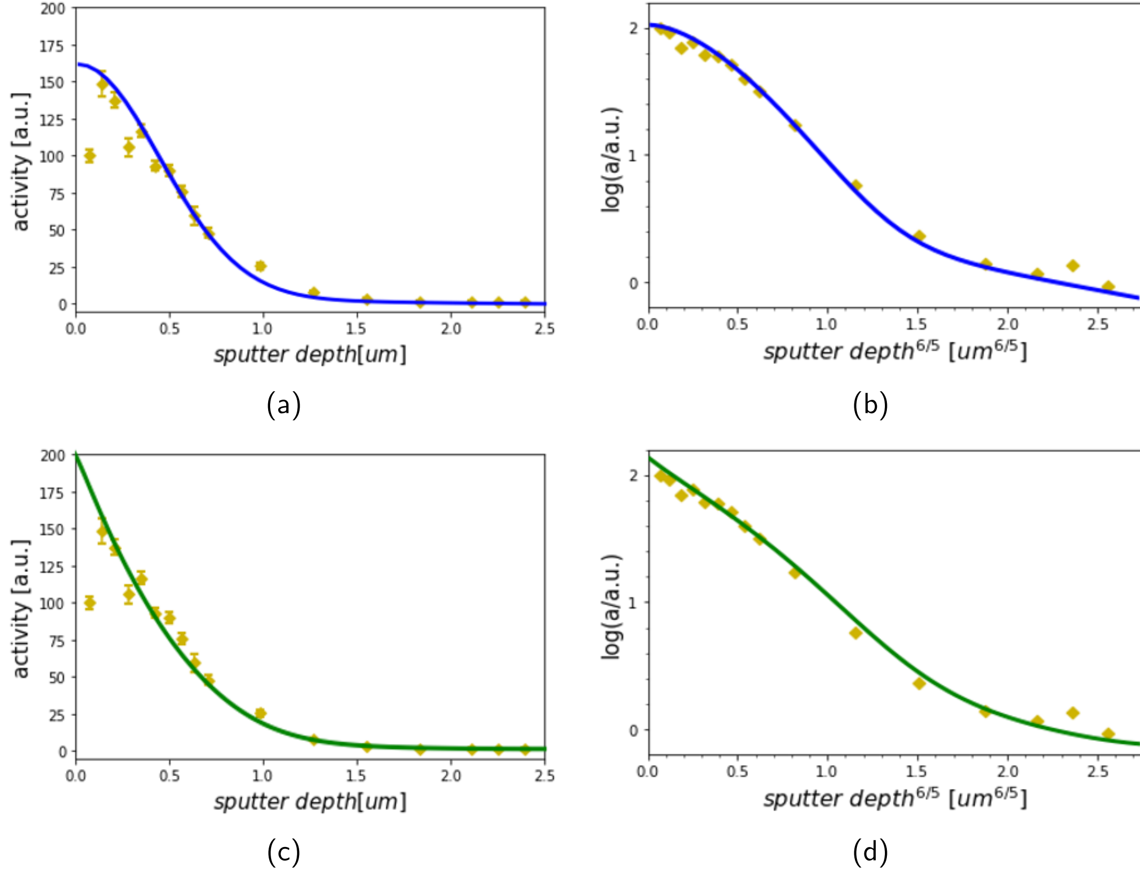


Figure 3.2: Indium tracer diffusion experiment vs. continuum model (a) normal scale non-linear fitting curve (blue); (b) log scale non-linear fitting curve (blue); (c) normal scale continuum model simulated curve (green); (d) log scale continuum model simulated curve (green). Experimental data points are shown in gold scattered points above.

based on activation energy for each species, is used for shallow defects and trap model[192], which explicitly models SRH recombination by taking into account the occupation and space charge storage of defects, is used for deep level defects in device simulator. Here, CCT [187] was applied to derive the equilibrium defect profile. We used GGI ratio, $[\mathrm{Ga}]/([\mathrm{In}]+[\mathrm{Ga}])$ ratio, copper-to-group III (CGI) ratio, $[\mathrm{Cu}]/([\mathrm{In}]+[\mathrm{Ga}])$, and selenium chemical potential, μ_{Se}, as three constraints in CCT. The reason we selected selenium chemical potential rather than valence (SeM) ratio, $(2[\mathrm{Se}]/([\mathrm{Cu}]+3([\mathrm{In}]+[\mathrm{Ga}])))$, is due to the valence ratio of CIGS has a large variance during

three-stage manufacture process. However, the pressure of the chamber, which mainly contributed by selenium vapor, is well controlled in each stage[54, 170], which suggests the selenium chemical potential would be roughly constant in the material if steady state is assumed. In addition, we also found from Fig. 3.4 that μ_{Se} changes rapidly even with tiny variance (10^{-3}) of SeM ratio under typical material composition while the SeM ratio changes only around 0.05% with μ_{Se} from -0.8[eV] to -1.2[eV]. Moreover, state-of-art characterization techniques are not able to detect such minor change of SeM ratio in CIGS layer but the selenium vapor pressure can be roughly fixed as we discussed previously. Thus, we conclude that applying μ_{Se} as third constraint will provide more stable and trustful simulation results. In this work, we used -1.0[eV] as μ_{Se} at 873K with respect to the standard elemental Se reference (the solid trigonal phase of Se with space group $P3_121$) used in our DFT calculations. At this selenium chemical potential, the resulting carrier densities and Fermi level match with experiment.

Bulk SRH recombination has been implemented using the SRH recombination model in the Sentaurus TCAD device simulation software. The SRH recombination rate R^{SRH} is described by Eq. 2.2 as mentioned in Section 2.2.1.3. There are two fundamental steps for carrier capture: the change of the electronic state due to electron-phonon coupling and vibrational relaxation due to phonon-phonon interactions. The electronic transition is much slower and thought to be the rate-limiting step for non-radiative capture. Following Alkauskas *et al*[5], the capture cross section is derived from the perturbation that causes the transitions between initial atomic configuration (Q = 0) and one special phonon mode of the excited state (Q = ΔQ). The configuration coordinate diagrams for a commonly recognized deep level defect $\mathrm{Cu_{III}}$ are given in Fig. 3.3. For $\mathrm{Cu_{III}}$ transition state (0/-1), we obtained approximately $10^{-23} \sim 10^{-21}$ cm^2 for both electron and hole capture cross section. While we got around $10^{-12} \sim 10^{-11}$ cm^2 for hole capture cross section and $10^{-14} \sim 10^{-13}$ cm^2 for electron capture cross section for $\mathrm{Cu_{III}}$ transition state (-1/-2). This suggests that (-1/-2) state is more likely to be deep level transition center while (0/-1) is shallow. This has been further validated by Bader charge analysis[61]: we found that $\mathrm{Cu_{III}}$ transition state (-1/-2) is more localized than transition state (0/-1). We also used NONRAD to examine $\mathrm{Ga_{Cu}}$ in CGS and find

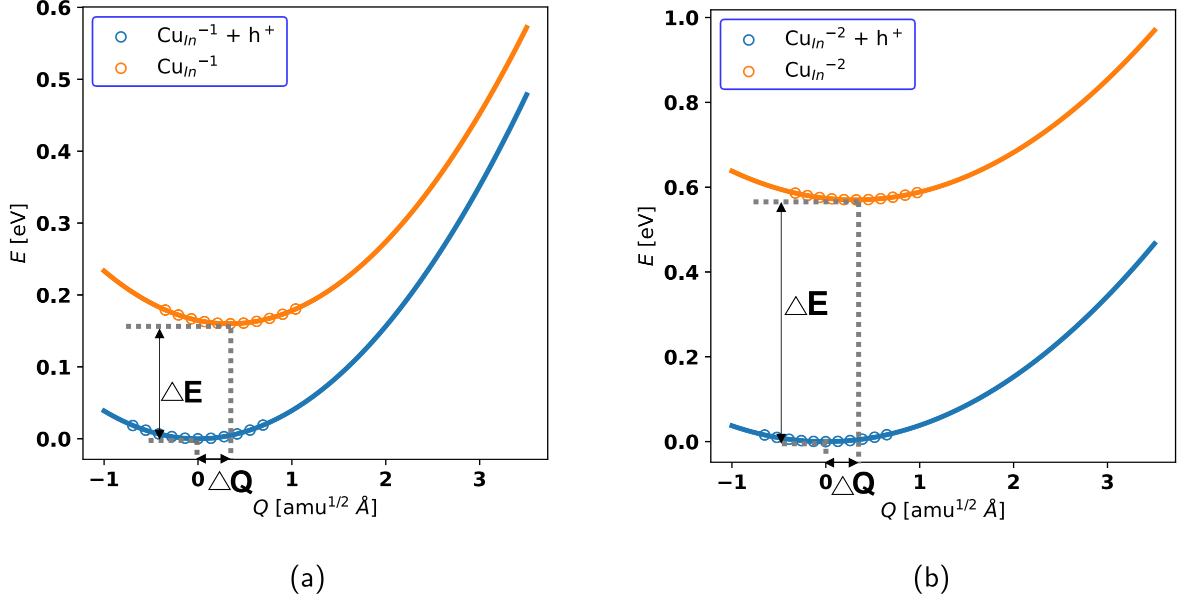


Figure 3.3: Configuration coordinate diagram of Cu_{III} (0/-1) (3.3a) and (-1/-2) (3.3b) transition states. Q in X-axis corresponds to the mass-weighted configuration coordinate. △Q is the difference in the equilibrium geometries of the initial and final state. △E is the energy difference between the initial and final state.

the capture cross section is approximately $10^{-17} \sim 10^{-16}$ cm^2 for holes and $10^{-14} \sim 10^{-13}$ cm^2 for electrons. These values are consistent with experimental measurement[68]. However, the defect tends to be shallow in CIS and have very low capture cross sections. This indicates Ga_{Cu} could be a dominant recombination center in CIGS devices with high GGI. The observation is consistent with reports by other groups[64, 188]. In fact, gallium antisites tend to have large concentration after high temperature film growth process according to CCT thermal equilibrium calculation. However, during cool-down process in the continuum model, they bind with copper vacancies, transforming into shallow defect complexes. This is due to large binding energy between group III antisite and copper vacancy as mentioned in Section II.A, which means cooling and annealing processes during manufacture could lower the deep level defect concentrations and improve the device performance. We also find significant binding energy between gallium antisite and copper

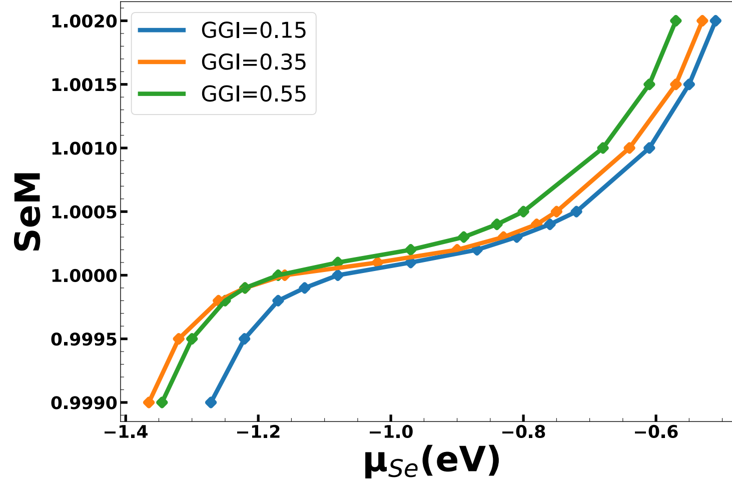


Figure 3.4: SeM vs. μ_{Se} for GGI=0.15, 0.35 and 0.55 at CGI=0.85; μ_{Se} is Se chemical potential referenced to standard Se chemical potential used in DFT calculation. Temperature for μ_{Se} is 873K.

Table 3.3: Binding energy (eV) of defect complexes in $CuInSe_2$ and $CuGaSe_2$. Binding energy in pure $CuInSe_2$ and pure $CuGaSe_2$ phases is calculated respectively. Binding energy in $CuIn_{1-x}Ga_xSe_2$ alloy is linearly interpolated according to Ga/In ratio. Charge states of complexes are noted in brackets

Defect	$CuInSe_2$	$CuGaSe_2$
In_{Cu}+V_{Cu}	0.18 (+1)	0.50 (+1)
Ga_{Cu}+V_{Cu}	0.30 (+1)	0.66 (+1)
In_{Cu}+Cu_{III}	0.89 (0), 0.44 (+1)	0.92 (0), 0.46 (+1)
In_{Cu}+Cu_{III}+V_{Cu}	0.69 (0), 0.46 (+1)	0.72 (0), 0.48 (+1)
Ga_{Cu}+Cu_{III}	0.89 (0), 0.46 (+1)	0.93 (0), 0.46 (+1)
Ga_{Cu}+Cu_{III}+V_{Cu}	0.69 (0), 0.46 (+1)	0.70 (0), 0.47 (+1)

antisite. This may also lead to formation of triplets such as $Ga_{Cu} + V_{Cu} + Cu_{III}$. A summary of defect binding energy is listed in Table 3.3. We have used NONRAD to examine these defects as well. However, the capture cross sections for both electrons and holes are very low ($< 10^{-24}$ cm^2). Thus, we consider these triplet defects, $Ga_{Cu} + V_{Cu} + Cu_{III}$ and $In_{Cu} + V_{Cu} + Cu_{III}$, as delocalized shallow defects. To verify the model, we compare the device simulation with experiment[73] as described in Section 3.2.3.

3.2.1 Cooling Process

Cooling process simulation is conducted in Sentaurus Process by the continuum model we built for defect diffusion. We find that the cooling process diminishes the density of unpaired group III antisites (Ga_{Cu}). This is due to a lowered probability for bound antisite-vacancy complexes to thermally dissociate following cool-down. Similarly, the concentration of unpaired Cu antisites also drops down during the cooling process. The final defect density (and thus carrier lifetime) depends on the post-growth thermal exposure, and, due to the low activation energy for Cu vacancy diffusion, even room temperature storage can impact the density of recombination-active defects in CIGS. A typical example for how cooling process could affect deep level defect concentration is shown in Table 3.4. By applying this model to a range of CGI ratio, GGI ratio and μ_{Se}, we can track how these deep level defects concentration changes with compositions.

As shown in Fig. 3.5a, for a fixed GGI ratio, and μ_{Se}, gallium antisite concentration only changes slightly until CGI is higher than 0.95. Ga_{Cu} drops sharply when CGI is larger than 0.95, which is accompanied by a significant drop of copper vacancy concentration when CGI is close to stoichiometry. Copper antisites increase drastically with CGI>0.95, suggesting a CGI ratio close to stoichiometry could be detrimental to device performance. We find a similar trend in Fig. 3.5b. Changes in the Ga_{Cu} concentration are within an order of magnitude, while copper antisites see large reduction when μ_{Se} is decreasing due to group III vacancy deficiency. However, when the μ_{Se} is lower than -1.10[eV], the valence ratio SeM will be lower than 1.0000 as shown in Fig. 3.4. Under such circumstance, the CIGS becomes n-type, which is generally not applicable for state-of-the-art CIGS solar cells. According to Fig. 3.5c, Ga_{Cu} density decreases a lot especially for high GGI materials. Even though the capture cross section for Ga_{Cu} is large as mentioned before, the low density of Ga_{Cu} makes it impossible to be an effective recombination center. When the GGI is low, there is significant amount of Ga_{Cu}. However, it is considered as shallow defect and has very low capture cross section in the low GGI region, which is also hard to be a deep level defect. On the other hand, Cu_{III} increases faster when GGI goes higher than 0.60. Thus we conclude that Ga_{Cu} has little impact on CIGS solar cell performance but Cu_{III} is at least partially responsible for the

Table 3.4: Impact of cooling process on deep level defects. [a]Equilibrium concentration was calculated via CCT. CGI, GGI, and μ_{Se} ratio for this sample are 0.85, 0.35, -1.00[eV], respectively. The cooling process is simulated via Sentaurus Process with a cooling rate of -20°C/min.

Concentration(cm^{-3}) / Defect Type / Condition	$\mathrm{Ga_{Cu}}$	$\mathrm{Cu_{III}}$
Equilibrium at 873K[a]	2.41e18	1.46e15
Cooling (873K to 300K, ramp = -20°C/min)	1.38e16	9.31e13

efficiency loss in wide-gap CIGS solar cells[64].

From Fig. 3.5, we can infer that controlling atomic ratio of elements in CIGS absorber is critical as it impacts defect concentration, lifetime, and finally device behavior. Also, it is obvious from the three plots that the cooling process could help decrease the deep defect concentrations for both gallium and copper antisites. Once we obtain the defect distribution profile of CIGS absorber layer, it is possible to predict solar cell performance via device simulation.

3.2.2 Solar cell device simulation

We used the above model to simulate solar cell device in [73]. The solar cell structure simulated is equivalent to the literature with some minor changes shown in Table 3.6. Some critical parameters used in device simulation are shown in Table 3.5. For the thickness of MgF_2 and ZnO:Al layer, we used values slightly different than reported. The reason is that the refractive index values of these two layers are not specified in the experimental paper[73]. The values we used here are typical values reported in literature[179, 7, 197, 37, 212], but we found large variances in refractive index values due to disparate synthesis and deposition processes. The function of both layers, however, is to reduce the reflection of photons which can be absorbed at maximum efficiency. Thus, we optimized the thickness of these two layers to maximize carrier generation in the absorber layer as calculated via $d = \frac{\lambda_{max}}{4\eta}$, where d is the thickness, λ_{max} is the maximum absorbed wavelength of CIGS, and η is the refractive index of anti-reflective material at λ_{max}. For the other layers, we keep them as reported. The complex refractive index values of $CuInGaSe_2$ are dependent on the material composition[154]. To accommodate the index value change with GGI across the CIGS

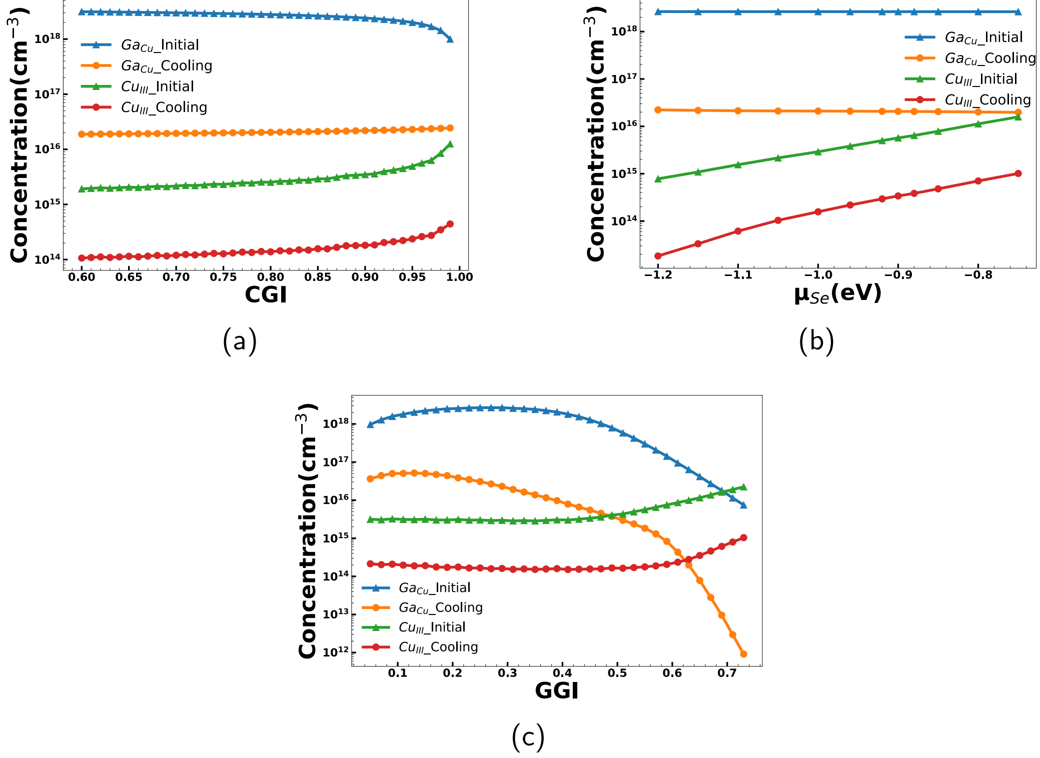


Figure 3.5: Deep level defect distribution with different CGI, μ_{Se} and GGI. (a) CGI $= 0.60 \sim 0.99$, GGI $= 0.30$, $\mu_{Se} = -1.00$[eV]; (b) $\mu_{Se} = -0.76 \sim -1.20$[eV], CGI $= 0.85$, GGI $= 0.30$; (c) GGI $= 0.05 \sim 0.75$, CGI $= 0.85$, $\mu_{Se} = -1.00$[eV].

layer, we used a linear interpretation with a shift in the spectrum according to the band gap. The relation between band gap and GGI are demonstrated in the following equation[190]:

$$E_g^{total} = (1-\alpha)E_g^{\text{CIS}} + \alpha E_g^{\text{CGS}} - b\alpha(1-\alpha) + 1.7 * (f_{\text{Cu}}^p - f_{\text{Cu}}) \; . \tag{3.3}$$

Here α denotes the GGI value, b is the optical bowing parameter which we chose as 0.20 in our work, and $(f_{\text{Cu}}^p - f_{\text{Cu}})$ accounts for the change in band gap with deviations in Cu composition from perfect stoichiometry ($f_{\text{Cu}}^p = 0.25$) [190]. Values for $E_g^{\text{CIS}} = 1.04$[eV] and $E_g^{\text{CGS}} = 1.68$[eV] are applied for all calculations[209, 6].

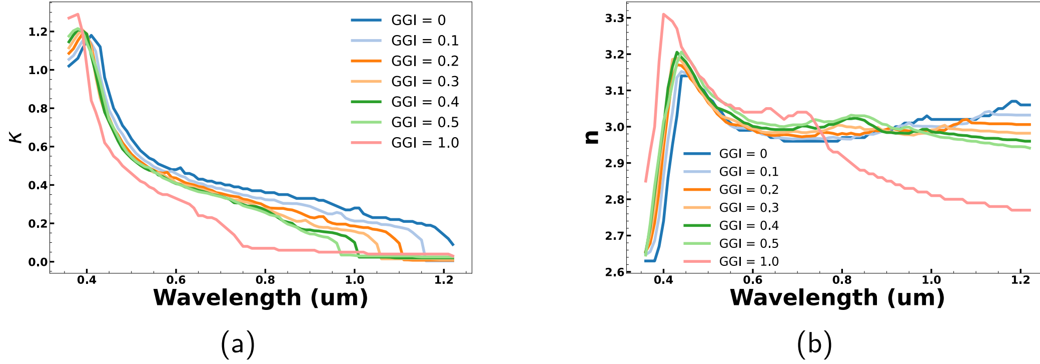


Figure 3.6: Complex refractive index of CIGS with different GGI values computed by Eq. (3.4). κ is extinction coefficient while n is refractive index.

The refractive index n for a specific GGI value can be calculated by

$$n(\alpha, E + \alpha\Delta E_g) = (1 - \alpha)n(0, E) + \alpha n(1, E + \Delta E_g) \ . \tag{3.4}$$

Here E is the energy of photon, ΔE_g is the band gap difference between CIS and CGS, α is GGI ratio while $\alpha\Delta E_g$ in Eq. (3.4) indicates the band gap energy change for a specific GGI versus band gap of CIS. The refractive index n of a photon with energy $E + \alpha\Delta E_g$ for GGI ratio α is linearly interpolated by n of a photon with energy E in CIS and n of a photon with energy $E + \Delta E_g$ in CGS. The extinction coefficient κ follows a similar relationship. Using the relation above, we can obtain complex refractive index values for any given GGI values. When the energy of photon is lower than the band gap, we applied an exponential decay in the extinction coefficient to account for the reduction in absorption intensity. Some cases are shown in Fig. 3.6. This method is used to obtain the complex refractive index for any GGI profile. Subsequently, the Transfer Matrix Method (TMM) is used in Sentaurus Device to calculate the propagation of plane waves through layered media[192] under the AM 1.5 Global spectrum with an integrated power of 1000 $\mathrm{W/m^2}$.

In CIGS layer, we used GDOES profile in the literature to extract GGI profile as illustrated in Fig. 3.7. To set constraints on the composition, a constant CGI value of 0.85 from paper and μ_{Se} (-1.00[eV]) are used in CCT. After that, we obtained distribution of all the defects. Among them,

Table 3.5: Critical parameters in CIGS Device Model

	$Cu(In,Ga)Se_2$	CdS	i-ZnO	ZnO:Al
Bandgap (eV)	Eq. 3.3[190]	2.4	3.3	3.3
Carrier Density (cm^{-3})	varies with composition	1.0 x 10^{19}	1.0 x 10^{19}	1.0 x 10^{19}
Refractive Index	Eq.3.4 [154]	Data [197]	Data [179]	Data [7]
Effective mass e/h	0.08(CIS), 0.13(CGS)/0.35	0.25/0.70 [113]	0.24/0.59 [183]	0.24/0.59 [183]
Mobility e/h (cm^2/V s)	300/30 [100]	100/25 [113]	200/25 [109]	200/25 [109]
Lifetime e/h (ns)	Defect-assisted SRH Model	1.0/1.0	-	-

Table 3.6: CIGS solar cell structure of experiment and simulation

Material	Layer Thickness (Ref. [73])	Layer Thickness (Simulation)
MgF_2	(105 ± 10) nm	100 nm
ZnO:Al	100 nm	55 nm
i-ZnO	50 nm	50 nm
CdS	(50 ± 5) nm	50 nm
$Cu(In,Ga)Se_2$	(2.8 ± 0.2) μm	2.8 μm
Mo	(1.5 ± 0.2) μm	1.5 μm

shallow defects like Cu vacancy were implemented as dopants in the device simulation through ionization model while deep level defects were considered as traps in the trap model[192]. The deep level defects profile are shown in Fig. 3.8. The carrier recombination due to deep level defect is modeled by SRH recombination model. We applied 5×10^{-14} and 10^{-12} for electron and hole capture cross sections of Cu_{III}, respectively , while 5×10^{-14} and 5×10^{-17} for electron and hole capture cross sections of Ga_{Cu}. In addition to recombination due to deep level defects Cu_{III} and Ga_{Cu}, we also include radiative recombination with the radiative recombination rate 1×10^{-10} cm^3/s[211], surface recombination at back and front surface with recombination velocity 1×10^{-7} cm/s[210]. Grain boundaries and other structural defects are accounted for via a uniform background midgap SRH recombination with lifetime of 15ns. The third recombination source is likely to be attributed to V_{Cu}+V_{Se}[69, 188, 77, 70, 71]. However, the detailed examination of this metastable defect is out of the scope of this paper. The effective lifetime in the device simulation is between $1-20$ ns for typical material compositions, which is within the range of reported data in literature [77, 206].

In addition, we found in literature that there is a thin n-type layer located at CIGS/CdS

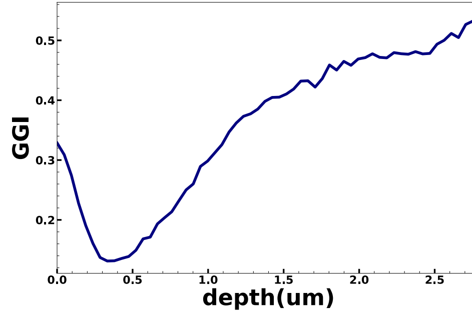


Figure 3.7: GGI curve extracted from GDOES profile in *Jackson et al.*[73]

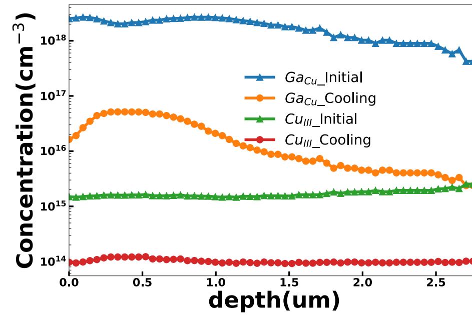


Figure 3.8: Deep level defect concentration distribution in CIGS absorber before and after cooling process.

interface, which may be due to Cd diffusion into CIGS and occupies copper sites[105, 2, 142, 218]. Other people argue that there is a layer with high resistance between CdS and CIGS[142, 75] possibly due to ordered defect compound. Accordingly, we added n-type doping at interface from $0.00 - 0.01$ μm in the device structure at Fig. 3.9, which results in a low carrier density region thus high resistance layer. The additional surface doping provides a smooth transition from heavily p-type CIGS to n-type CdS, which facilitates electron transfer and boosts device performance.

3.2.3 Simulation Results

As shown in Table 3.7, four typical parameters which are used to characterize solar cells, short circuit current ($J_{\rm sc}$), open circuit voltage ($V_{\rm oc}$), fill factor, and efficiency (η%) are all in good agreement

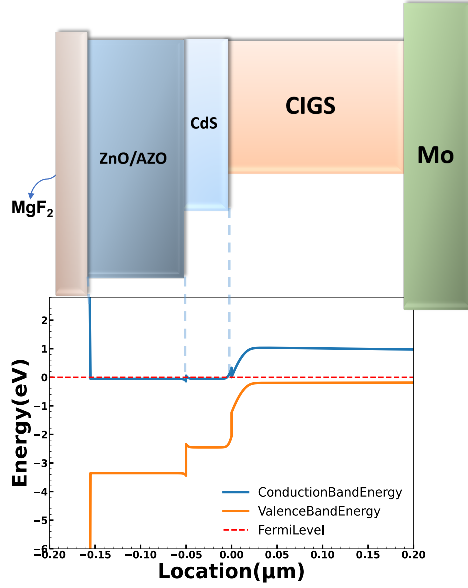


Figure 3.9: Device schematic and band diagram under AM 1.5G illumination (no bias applied). CIGS layer ranges from 0 to 2.8 μm, but only a portion is shown in the schematic for concise illustration. Red dashed line indicates Fermi level. Upper green solid line denotes conduction band minimum while lower green solid line denotes valence band maximum.

Table 3.7: Solar cell device performance of experiment and simulation

	J_{sc} (mA)	V_{oc} (V)	η%	Fill Factor
Experiment (*Jackson, P. et al*)	34.7	0.7113	19.0	77.1
Device Simulation	34.7	0.7146	19.2	77.6

with experiment values. This confirms the validity of our simulation model. To determine the most critical deep level recombination center, we turned off the SRH recombination of Ga_{Cu}, and there is little effect on the solar cell efficiency. The reason is discussed previously in Section III.A. Thus, we conclude that Cu_{III} is the only deep level bulk defect in the model that significantly impacts the solar cell performance.

3.2.3.1 Effect of GGI Gradient on Device Performance

The GGI gradient directly impacts the band gap variation in CIGS layer, which essentially changes the conduction band position. A typical three-stage process will generate a GGI profile which

has a minimum notch close to the CdS interface. Then GGI increases towards both CdS and Mo interface, which is known as double gradient or V-shaped profile[140].

Generally, the common consensus is that double gradient profile is or could be beneficial to the device performance[53]. However, it is difficult to judge if gains of device efficiency are due to GGI grading or solely from three-stage deposition process, as the latter is believed to increase grain size as compared to one-stage process[171, 33]. Generally, the back gradient toward Mo side is thought to be beneficial. Based on electro-optical modeling[110], an increased GGI ratio towards Mo contact creates a quasi-electric field that repels minority carriers (electrons) from the high recombination region at back contact. In addition, high energy photons that have not been absorbed at front region can be further absorbed at back region because of the advantageous absorption coefficient for high energy photons according to Fig. 3.6. Thus the effect of higher GGI toward Mo layer is considered as positive. However, the role of increased GGI towards the CdS is not clear. Actually, a theoretical study suggests that the increased GGI in front part does not lead to performance gain[110], and, under certain conditions, a front grading has been found to be detrimental to device performance[195]. In fact, CIGS solar cells in experiment appear to have higher performance when the minimum GGI position is close to CdS interface[29]. Also, people found a nearly flat GGI gradient toward CdS interface shows great efficiency[52]. In our model, we find the notch in GGI profile generates electron traps due to formation of quantum well which increase non-radiative recombination. Even though the lower band gap at the notch facilitates the absorption of low energy photons, in our simulations it cannot totally compensate the loss due to additional SRH recombination. Thus, we conclude that the front Ga gradient towards CdS is only a consequence of three-stage process but it does not lift the device efficiency per se. The increase in grain size from the three-stage process as compared to the one-stage process and the back GGI gradient would likely be the reasons of higher solar cell efficiency[171]. Therefore, we believe the optimal GGI profile should have a gradient toward Mo layer and flat GGI toward CdS interface.

To explore the effect of GGI gradient and find the optimal GGI profile, we systematically investigated a wide range of possible GGI profiles to create a contour maps of device efficiency

using our device simulation model as shown in Fig. 3.10. We used three constraints to determine GGI profile: GGI at back Mo interface, GGI at CdS interface, and the turning point location. The turning point location indicates where GGI transitions from flat to graded.

The simulation results are shown in Fig. 3.12. One key finding of our simulation is that the optimal GGI at CdS interface is 0.35∼0.40, at which the offset between CdS and CIGS has been optimized. In an optical GGI profile, the turning point location should be very close to CdS interface, which is consistent with report from experiment[53]. Also, Yang *et al.*[228] found that when the back GGI is increased from around 0.5 to 0.65∼0.70 with front GGI of approximately 0.35, the efficiency gain is around 0.3%∼0.4%. This is in line with our simulation results and it suggests our model can provide quantitatively comparable results with experiment. Also, we did not find a loss in device performance when further increasing GGI at back region, but the performance gain did slow down when back GGI hits 0.65. We believe this is due to the Cu_{III} defect density drastically increasing when GGI>0.60, as shown in Fig. 3.5c. Thus, controlling the back GGI over a reasonable range should be beneficial to device performance. Generally, CIGS solar cell has a large tolerance for GGI profiles for which the device performance is nearly optimal, which allows flexibility during manufacture process.

Additionally, we cut the CIGS layer thickness in half to see how device efficiency would differ in Fig. 3.11. Surprisingly, the efficiency of solar cell can still be maintained at high level (∼21.6%) if the GGI profile is optimized. Compared to the device with 2.8μm CIGS, the optimal back GGI is lower (0.63∼0.73) and the turning point location is further from CdS interface. The reason is that there is less effective light absorption area for CIGS with thickness of 1.4μm. Thus, the absorption coefficient of thin film becomes more critical. Lower GGI region is more favored than higher GGI region at front region as the former has more optical absorption and generation. On the other hand, the back region in thin CIGS samples is close to high generation and recombination region. As a result, the impact of deep level defect Cu_{III} would be more significant. Consequently, a lower GGI profile is favored at both front and back region of 1.4μm CIGS layer.

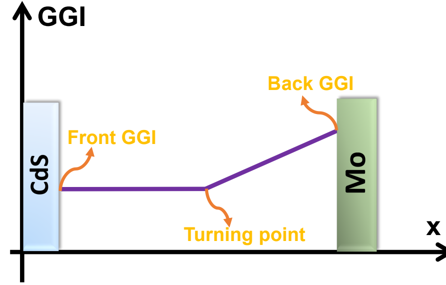


Figure 3.10: Schematic of GGI profile tuning. Three constraints in the GGI profile are back GGI (Mo interface), front GGI (CdS interface) and the turning point location. All the device simulations are conducted under the same conditions except GGI profile and corresponding defect level change.

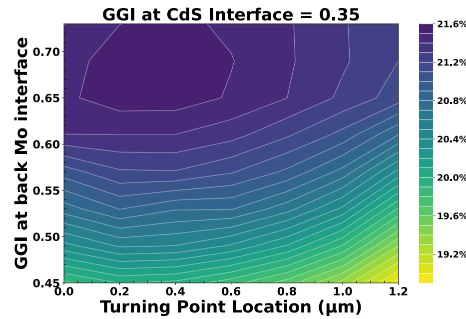


Figure 3.11: Simulation of device efficiency for various GGI gradients with CIGS thickness of 1.4μm.

3.3 Summary

We have demonstrated coupled process and device modeling for CIGS solar cells with excellent match to experimentally-observed device performance. A detailed set of continuum models for defect, ion, impurity and complex evolution using parameters calculated via DFT has been implemented in 1/2/3D process/device simulation tool Sentaurus. Continuum models based on DFT parameters successfully predict Cu diffusion and In tracer diffusion in CIGS. The effect of GGI gradient is also explored in this work. We found that a high GGI value at back contact (Mo) would reduce the recombination but the slope toward the front side (CdS) will possibly lead to minority carrier trap region and increase the non-radiative recombination. Also, the increased amount of deep level traps with increasing GGI would hinder the device performance gains by raising the back

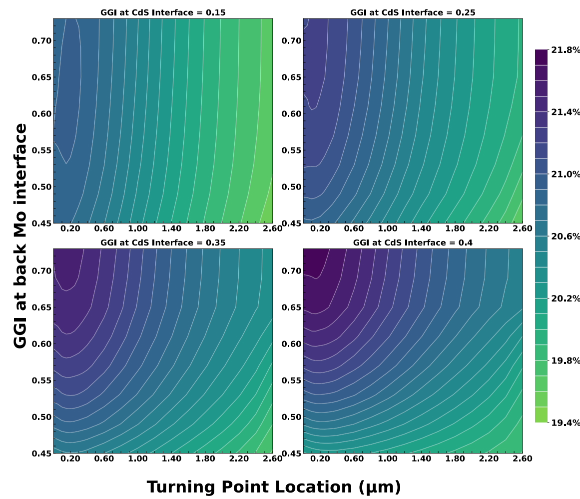


Figure 3.12: Simulation of device efficiency for various GGI gradients and corresponding GGI values at two interfaces, mapped for several GGI at CdS interface (0.15, 0.25, 0.35 and 0.40). The turning point location ranges from 0.20 μm (close to Mo layer) to 2.80 μm (CdS interface). GGI at Mo interface ranges from 0.45∼0.73.

GGI gradient. Besides, we notice there is a large range where the CIGS solar cell performance can reach optimal level which allows flexibility in CIGS manufacturing. Additionally, the CIGS solar cell efficiency can be maintained at a high level even if the thickness of the absorber layer has been halved. This suggests less materials have to be used for an almost equivalently efficient solar cell module.

Chapter 4

POINT DEFECTS ENGINEERING IN CDTE AND CDSETE ALLOY

Cadmium telluride (CdTe) and its alloy with selenium (CdSeTe) have been extensively studied due to their high potential for use in optoelectronic devices, including solar cells and radiation detectors. Compared to CdTe, the addition of Se to the alloy has been shown to improve the material's electronic properties, such as its carrier lifetime and carrier mobility, allowing tuning of bandgap, making it a promising candidate for high-efficiency solar cell applications [48]. The performance of these devices, however, can be significantly impacted by the presence of point defects, which can affect the material's electrical and optical properties. To better understand the effects of defects on the properties of CdSeTe, it is necessary to investigate the formation and recombination of defects and defect complexes in this material. In this work, we use density functional theory (DFT) calculations to investigate the properties of intrinsic defects, copper and group V dopants in CdSeTe, with a focus on the properties of these point defects and their impact on the material's electronic properties.We consider intrinsic defects that have been widely reported and are considered to be significant, such as V_{Cd}[226, 94, 79, 182, 16], Te_{Cd}[223, 149], V_{Te}[223, 40, 152, 127], Cd_{Te}[223, 40, 152], Cd_{int}[223], Te_{int}[94, 223, 89, 40], Se_{Cd}[178], $V_{Cd} + Te_{Cd}$[93] and $Te_{int} + Te_{Cd}$[93]. For copper in CdTeSe, we consider four dominant extrinsic defects reported in the literature, including Cu_{Cd}[227, 38], Cu_{int}[227], $Cu_{Cd} + Cd_{int}$[227] and $Cu_{Cd} + Cu_{int}$[227]. And for group V dopants, in addition to the commonly recognized AX defects and substitutional defects As_{Te}[89, 225], we also consider two types of defects complexes that has not been thoroughly studied, including $(Cd_{int} + As_{Te})^{+}$ and $(V_{Te} + As_{Te})^{+}$. The choice of these two defects is based on the potential Coulombic interaction between As_{Te}^{-1} and Cd_{int}^{2+} or V_{Te}^{2+}.

4.1 Methods

4.1.1 First Principles Calculation Methods

Most DFT calculations apply the generalized gradient approximation (GGA) or local density approximation (LDA) as the exchange-correlation functional. However, it is widely known that DFT with LDA or GGA is likely to underestimate the bandgap of semiconductors. For CdTe, GGA gives us a bandgap of 0.68 eV, which is much smaller than experimental values ($\approx$ 1.5 eV). This can be explained by the overestimation of the delocalization of Cd-4*d* electrons, which lifts the valence electron energies (the Te-5*p* valence band) [215]. There are many approaches to correct the exchange-correlation functional, such as self-interaction correction (SIC) calculations[128], and hybrid functional of Heyd, Scuseria and Ernzerhof (HSE)[79, 226]. However, these methods are computationally expensive, which makes them impractical for study of defects in alloys due to the massive number of possible configurations. DFT with coulomb self-interaction potentials (GGA+U) is another method that has been frequently used to correct the calculated bandgaps [215]. This method combines Hubbard-like model for a portion of states in the system with Coulomb self-interaction potentials (U) to select bands for correction. Non-integer or double occupations of states are described by introducing of two parameters: (1) U, which reflects the intensity of the on-site Coulomb interaction, and (2) J, which adjusts the intensity of the exchange interaction. Typically, for simplicity, an effective parameter $U_{eff} = U - J$ is used. This effective parameter U_{eff} is typically referred to as U.

4.1.2 Computational Details

Our investigation reveals that setting U=12.2 eV[216] for Cd-4*d* orbitals in CdTe closely matches the experimental lattice constant and bandgap, showcasing the effectiveness of the GGA+U method. In comparison, while the HSE method accurately reproduces the experimental bandgap, it tends to substantially overestimate lattice constants, indicating a larger error margin compared to the GGA+U approach.

Applying the optimized U parameter to zinc blende and wurtzite structures of CdSe, we observe

a surprisingly good agreement with experimental data, surpassing the performance of the HSE06 hybrid method. Specifically, the bandgap and lattice parameters of zinc blende CdSe align closely with experimental results, as detailed in Table 4.1. For the wurtzite structure, although GGA+U slightly underestimates lattice parameters, the bandgap predictions remain excellent.

This consistent accuracy across both CdTe and CdSe can be attributed to the similar *d*-*s* coupling inherent to Cd-Te and Cd-Se bonds, with both Te and Se being group VI elements. The GGA+U method effectively lowers the Cd 4*d* bands, enhancing *d*-*s* coupling. With U=12.2 eV, it appears to provide a suitable correction for both materials. The details of U value determination are provided in Appendix C.1.

Leveraging this optimized U value, we extend our study to the full composition range of the CdSeTe alloy. A notable advantage of the GGA+U method is its lower computational cost compared to HSE06, facilitating the application to supercells of the CdSeTe alloy with varied Se/Te arrangements and defect calculations. This approach also allows for the extension to larger supercells at a manageable computational expense when some defect calculations yield less reliable results due to finite cell-size errors[25].

Structural optimizations and energy calculations were carried out using the VASP code. For the GGA+U calculations, we employed the Perdew, Burke, and Ernzerhof (PBE) exchange-correlation functional. Additionally, the HSE06 functional with default parameters was utilized for comparative analysis. The calculations were performed with a plane wave cutoff energy of 450 eV for the wave functions. The chosen energy cutoff value is validated in Appendix C.2. Spin-polarized calculations are performed by setting ISPIN to 2. The Brillouin zones for all structures under investigation were sampled using Γ-centered k-point grids. Specifically, a 2×2×2 k-point grid was employed for both 64-atom and 216-atom supercells, while a 1×1×1 grid was applied to the 512-atom supercell.

In the study of CdSeTe alloy, we focus on compositions of $CdSe_{0.25}Te_{0.75}$ and $CdSe_{0.50}Te_{0.50}$. This selection is informed by the fact that the cubic phase of $CdSe_{x}Te_{1-x}$ is stable for $0 \leq x \leq 0.45$[175]. In contrast, the hexagonal phase (wurtzite structure) of $CdSe_{x}Te_{1-x}$ emerges for $0.55 \leq x \leq 1$, which is not considered to be photoactive[159, 175]. Moreover, in most state-of-the-art

CdSeTe technologies, the Se ratio typically remains below 40%[102, 48, 99]. Consequently, we will not extend our study to higher Se ratios.

Utilizing a supercell program[148], we generate 30 configurations for each alloy composition, prioritizing those with the highest frequency of occurrence. The configuration selected as the reference structure is the one whose formation energy is closest to the Boltzmann distribution average of the formation energies for all configurations sharing the same alloy ratio. This reference supercell then serves as the basis for generating defect supercells. In Appendix C.3, we demonstrate that the Se-Se interaction, energy difference, and lattice constant difference in CdSeTe alloys with the same Se ratio but different Se arrangements are all small. Therefore, the selection of one reference structure will not significantly affect the overall results. In addressing defects within the alloy, we examine multiple configurations that exhibit varying Se/Te arrangement environments. We found some defects, such as Te_{Cd}, As_{Te} AX defects, show a significant dependence on the local environment. However, most defects with T_d symmetry do not exhibit strong dependence on local arrangements. For all these defects, we opt for the Boltzmann distribution average at 873 K as a representative formation energy values across at least four configurations for each point defect. 873K is a typical annealing temperature for CdTe deposition process[173, 4].

Table 4.1: LATTICE CONSTANT, BANDGAP FOR CDTE AND CDSE OBTAINED FROM GGA+U HSE06 AND EXPERIMENTS.

Method	GGA+U		HSE06		Experiment[101]	
	Bandgap (eV)	a_0 (**Å**)	**Bandgap (eV)**	a_0 (**Å**)	**Bandgap (eV)**	a_0 (**Å**)
CdTe (Zinc Blende)	1.50	6.46	1.50	6.58	1.50	6.48
CdSe (Zinc Blende)	1.72	5.95	-	-	1.71	5.98
CdSe (Wurtzite)	1.79	$a = b = 4.21$ $c = 6.86$	1.68	$a = b = 4.30$ $c = 7.01$	1.80	$a = b = 4.30$ $c = 7.02$

The defect formation energy can be derived from DFT using the supercell method[50] by the Equation 2.1, E_{corr} is the charge correction energy to account for interaction between periodic images. The conventional FNV correction method[51] exhibits a tendency for overestimating or underestimating the charge correction, especially when dealing with partially or fully delocalized defects[86]. In order to address this issue and enhance the accuracy of the correction process, a strategic modification is implemented. Instead of directly applying the FNV correction to these

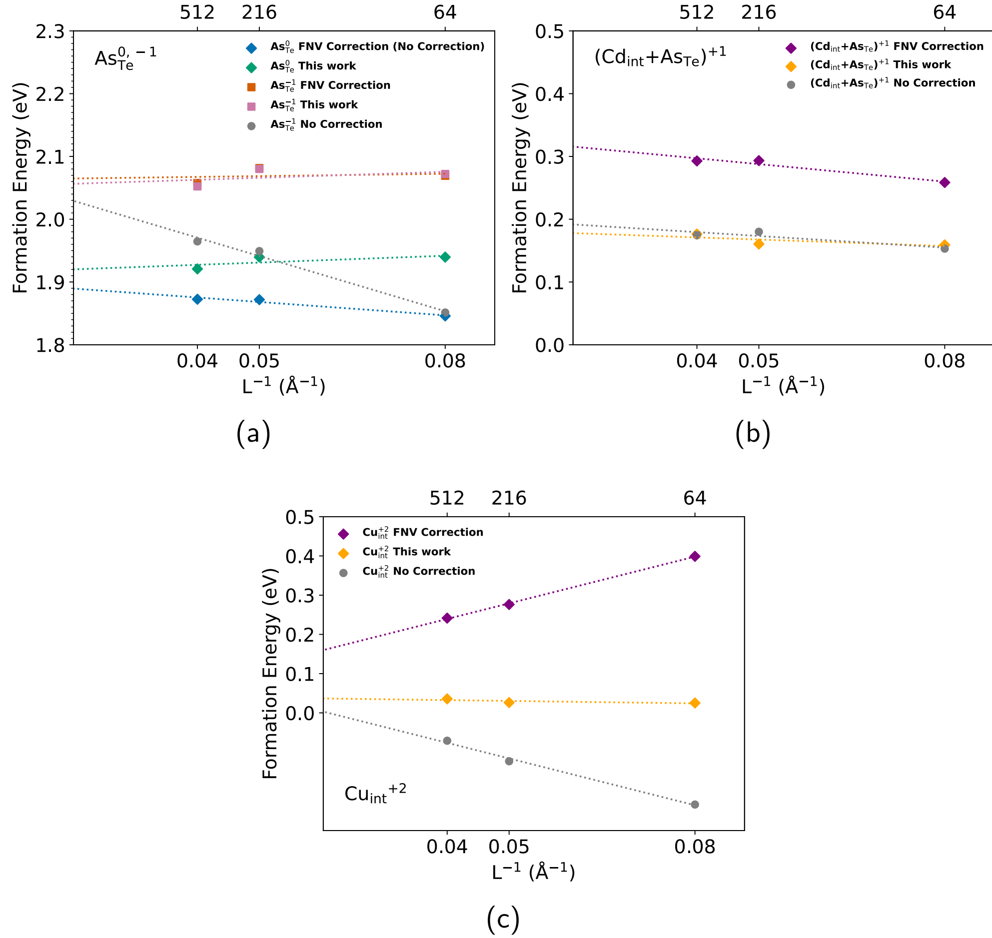


Figure 4.1: Defect formation energies vs inverse supercell size for (a) $As_{Te}^{0,-1}$, (b) $(Cd_{int} + As_{Te})^{+}$ and (c) Cu_{int}^{+2} . The defects are relaxed with fixed lattice constants. "L" on the bottom x-axis denotes the length of corresponding CdTe supercell. The Fermi level is set to the VBM in these plots.

complex defects, we choose to initiate the correction process by employing the density derived electrostatic and chemical (DDEC) method to determine the net atomic charges (NACs)[194, 106, 121]. This preliminary step serves the purpose of precisely ascertaining the charge localization within the system. Following the DDEC analysis, the q_{NAC} of the defects are determined. To find out the unscreened localized charge used in FNV correction, a scaling factor k is introduced. This scaling factor k is carefully calibrated to reintroduce the screening effect that may be lost during the

charge localization process. Subsequently, q_{NAC}, scaled by the aforementioned factor k, is treated as the input localized charge q_{loc} for the FNV charge correction method,

$$q_{\mathrm{loc}} = k * q_{\mathrm{NAC}}. \tag{4.1}$$

The factor k can be determined using one of two approaches. The first approach involves selecting defects that are fully localized; in this scenario, the nominal charge of the defects is equal to q_{loc}, facilitating straightforward determination of k. However, this method may not always be feasible due to the prerequisite knowledge required about defect localization. Alternatively, the second approach entails adjusting the supercell size and iteratively testing k values to achieve convergence in the defect formation energies across varying supercell dimensions. For both strategies, it is advisable to select defects with substantial q_{NAC} values, as this can lead to significant correction values. Such a selection ensures the derived k values are both reasonable and effective.

In this study, we select Te_{Cd}^{+} as the reference defect due to its advantageous properties. Specifically, the Kohn-Sham level of Te_{Cd}^{+} is located within the band gap, and it exhibits a large capture cross section for both the (0/+1) and (+1/+2) transitions, as discussed in Sec. 4.2.1 and corroborated by existing literature[94]. We firstly apply first approach to determine universal k value of 3.5 for CdTe, $\mathrm{CdSe_{0.25}Te_{0.75}}$, and $\mathrm{CdSe_{0.50}Te_{0.50}}$. This k value is then verified by the second approach by increasing supercell size.

Figure 4.1 showcases the effectiveness of our charge correction method in overcoming the limitations associated with the FNV correction approach. Our method notably addresses several critical issues:

- Non-zero NACs in neutral defects, such as As_{Te}^{0} illustrated in Fig. 4.1a, result in image interactions that necessitate charge correction, a scenario not accommodated by FNV correction. This issue has been observed in other work as well [145]. It occurs when a carrier fills states with delocalized characteristics, such as states close to the CBM or VBM. Consequently, the free carrier charges do not fully screen the nuclear charge of point defects, leaving a long-range Madelung interaction between the point defects even in neutral supercells. Our method's ap-

plication of charge correction to these defects is supported by faster convergence, as evident in Fig. 4.1a. For instance, the transition level for As_{Te} (0/-1) shifts from 0.22 eV without charge correction to 0.13 eV with it, more closely aligning with experimental ($\approx$ 94 meV) [135] and hydrogenic model predictions ($\approx$ 113 meV) [176]. Similarly, the transition level for P_{Te} (0/-1) improves from 0.16 eV to 0.11 eV, nearing experimental values ($\approx$ 87 meV) [135].

- For charged defects that are not fully localized, our method accurately accounts for the discrepancy between localized charge q_{loc} and nominal charge, avoiding the overcorrection observed with FNV correction. This accuracy is validated by the consistent formation energy results for Cu_{int} across supercells of 64, 216, and 512 atoms, as shown in Fig. 4.1c.

- Additionally, our approach adeptly determines the localized charge distribution of complex defects, a task for which FNV correction's charge assignment proves ambiguous and results in poor convergence. Our method's ability to assign q_{loc} to specific defect positions leads to improved convergence for complex defects such as $(Cd_{int} + As_{Te})^{+}$, as demonstrated in Fig. 4.1b.

For localized defects, such as As_{Te}^{-1} in Fig. 4.1a, Te_{Cd}^{+1} and Te_{Cd}^{+2}. Our methods show little difference compared to FNV correction. These findings highlight the comprehensive capabilities of our charge correction method in accurately modeling defect properties, surpassing the limitations of existing approaches.

4.2 Results and Discussion of Point Defects

4.2.1 Intrinsic Defects

The cadmium vacancy V_{Cd} is a pivotal intrinsic defect in CdTe and its alloys, with its configuration and transition levels being subjects of significant debate in both theoretical and experimental literature[226]. Our research reveals that the neutral state of V_{Cd} primarily adopts D_{2d} symmetry (dimer structure), featuring two fully occupied degenerate states and one unoccupied state. In contrast, the -1 and -2 charged states are most stable in T_d symmetry. The transition levels of V_{Cd}, as

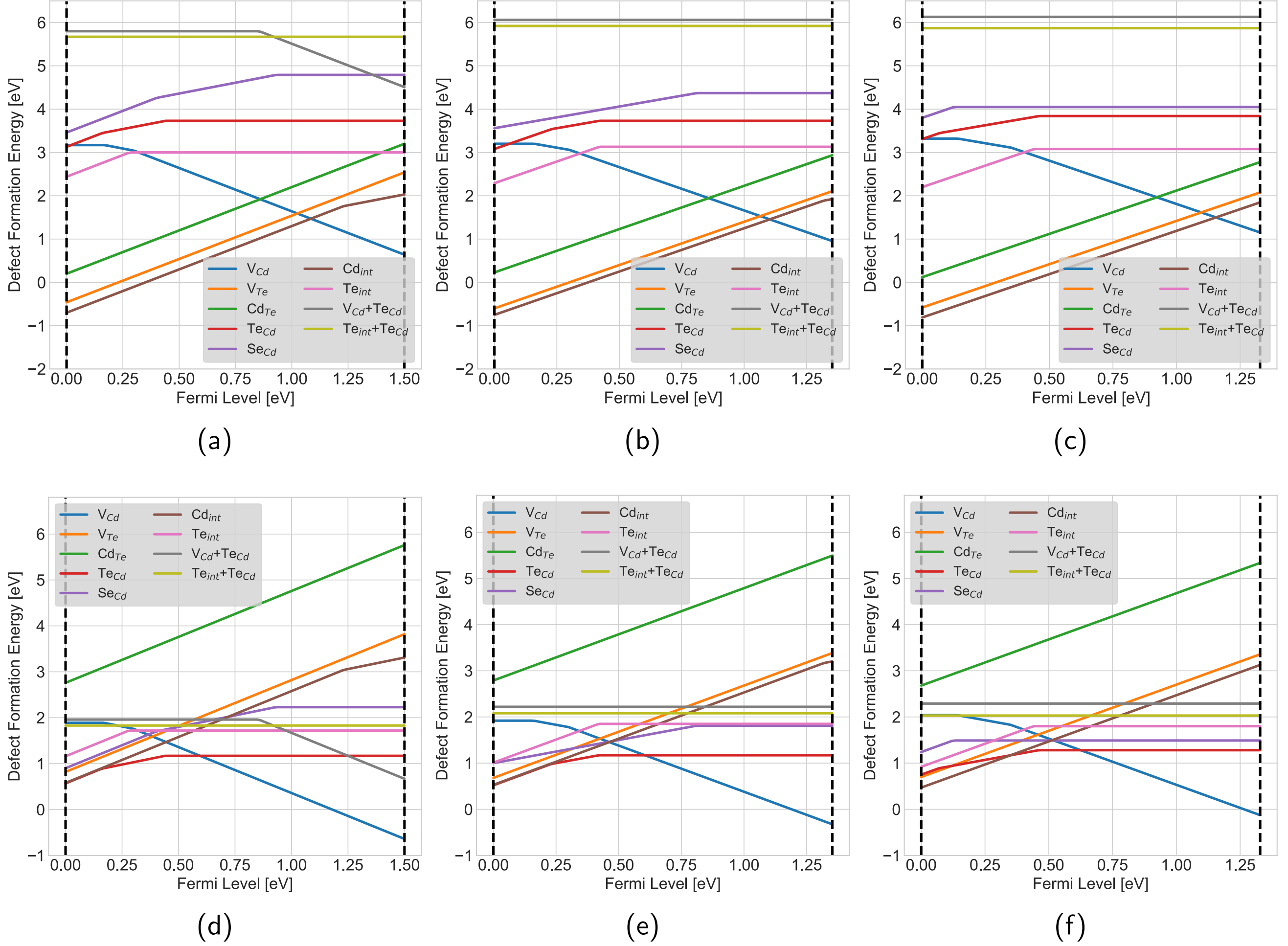


Figure 4.2: Defect formation energies vs Fermi level of intrinsic defects in CdTe under (a) Cd-rich and (d) Te-rich condition, $CdSe_{0.25}Te_{0.75}$ under (b) Cd-rich and (e) Te-rich condition, $CdSe_{0.50}Te_{0.50}$ under (c) Cd-rich and (f) Te-rich condition. The chemical potential conditions are $\mu_{Cd} + \mu_{Te} = -1.28$ eV and $\mu_{Cd} + \mu_{Se} < 0.15$ eV.

established in various studies, range from 0.1 to 0.8 eV above the VBM [42, 13, 208, 28, 40, 24, 167]. In this work, the transition levels for V_{Cd} in CdTe are identified as 0.17 eV for the (-1/0) transition and 0.30 eV for the (-2/-1) transition. For $CdSe_{0.25}Te_{0.75}$, these levels are 0.16 eV and 0.30 eV, respectively, while for $CdSe_{0.50}Te_{0.50}$ they are 0.14 eV and 0.35 eV, respectively. These findings indicate that the transition levels of V_{Cd} exhibit relative consistency in alloys with a selenium ratio below 50%.

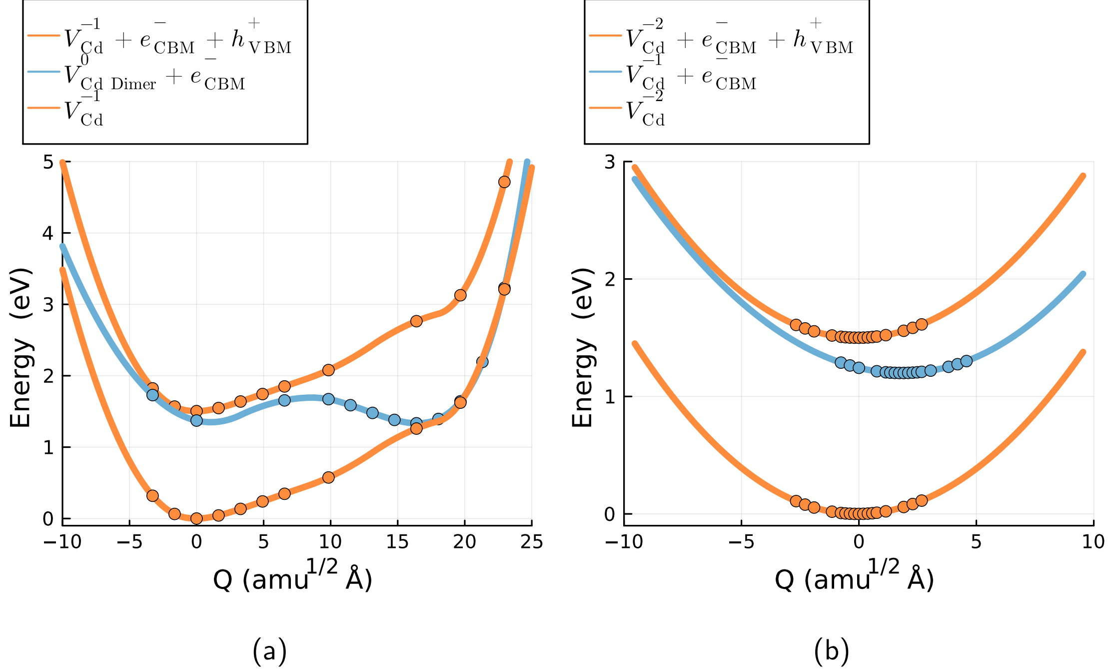


Figure 4.3: Configuration Coordinate Diagram of V_{Cd} (-1/0) and (-2/-1). Q in X-axis corresponds to the mass-weighted configuration coordinate. Q indicates the configurational coordinate path between equilibrium configurations. $Q = 0$ indicates defects ground state with T_d symmetry.

Furthermore, we calculated the defect capture cross section of V_{Cd} in CdTe based on static approximation[153, 50, 191, 5, 84, 83]. The results, presented in Fig. 4.3 and Table 4.2, reveal that the potential energy surface (PES) of the (-2/-1) transition is quasi-harmonic, while the PES of the (-1/0) transition is anharmonic. This discrepancy can be attributed to the proximity of metastable D_{2d} V_{Cd}^{-1} and T_d V_{Cd}^{0} states, which merge with the most stable configurations of these respective charge states—a phenomenon also observed elsewhere in HSE06 calculation of V_{Cd}[79] and defects in GaAs[83]. From the carrier capture cross section calculations, we determined that the (-1/0) transition exhibits strong nonradiative recombination intensity, whereas the (-2/-1) transition demonstrates only weak nonradiative recombination. This observation aligns with previous reports using the HSE06 functional[79, 94].

The second critical intrinsic defect under investigation is Te_{Cd}. The charge states Te_{Cd}^{+} and

Table 4.2: Capture cross section of deep level defects in CdTe at 300K

Defect	Transition Level	Trapping Process	$\sigma(T=300K)[cm^2]$
V_{Cd}	(-2/-1)	hole capture	8.0×10^{-17}
		electron capture	1.9×10^{-20}
V_{Cd}	(-1/0)	hole capture	2.8×10^{-16}
		electron capture	2.1×10^{-15}
Te_{Cd}	(+2/+1)	hole capture	2.1×10^{-18}
		electron capture	1.0×10^{-15}
Te_{Cd}	(+1/0)	hole capture	3.3×10^{-15}
		electron capture	8.1×10^{-18}

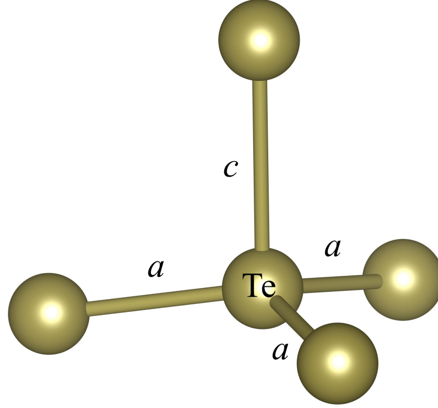


Figure 4.4: Schematic of Te_{Cd} C_{3v} defect configuration. c denotes length of the long bond. a denotes length of three short bonds.

Te^0_{Cd} both exhibit C_{3v} symmetry as shown in Fig. 4.4, while Te^{2+}_{Cd} displays T_d symmetry. A significant dependence on the Se/Te arrangement is observed for Te_{Cd} defects with C_{3v} symmetry. To exemplify this, we considered Te^0_{Cd} and calculated its formation energy for 56 different local Se/Te environments. These 56 Te^0_{Cd} defects were randomly split into an 80% training set and a 20% testing set. We employed a neighbor counting linear regression model, as detailed in Table 4.3, which effectively captures this dependence using eight features. This regression model was trained using a bootstrapping strategy, a resampling technique that involves repeatedly drawing samples from our dataset with replacement. We repeat this process 10 times. The fitting results, illustrated in Fig. 4.5, demonstrate the model's accuracy.

The eight features considered in the model include the Se ratio, the square of Se ratio, the presence of a first nearest neighbor (1NN) Se atom with a long bond, the count of 1NN Se atoms with short bonds, the square of the count of 1NN Se atoms with short bonds, the presence of second nearest neighbors (2NN) with one long and one short bond, 2NN with one long bond, and 2NN with one short bond. The positive coefficients for the Se ratio and Se ratio squared suggest that higher Se ratios make the formation of Te_{Cd}^{0} less favorable. The feature 1NN Se with a long bond indicates whether the elongated bond in the C_{3V} configuration is occupied by Se, with its negative coefficient implying that the presence of Se in this bond promotes C_{3V} defect formation. The features 1NN Se with short bonds and its square capture a quadratic relationship between the formation energy of C_{3V} defects and the occupancy of short bonds by Se, with the minimum of the quadratic function occurring near a Se ratio of 0.8. This indicates a decrease in formation energy as more Se atoms occupy the 1NN short bonds with Se ratio less than 80 %. The final three features account for the 2NN Se/Te atoms and whether they are neighbors to the 1NN with elongated or short bond, or both. The positive coefficient for 2NN (1 Long, 1 Short) and negative coefficients for both 2NN (1 Long) and 2NN (1 Short) reflect the diverse preferences for the C_{3V} defect configuration. This model can be further integrated with Lattice Monte Carlo simulations to elucidate the impact of the local environment on Te_{Cd}. Additionally, we performed calculations for the capture cross section of this defect, with results presented in Table 4.2. These findings validate that Te_{Cd} can create deep-level trap centers, aligning with previous HSE06 calculations and experimental observations[94, 111].

Table 4.3: Te_{Cd}^{0} Defect Neighbor Model: coefficients ΔE and their standard deviations for each feature. For nearest neighbor counting features, the count is specific to Se atoms. The distinction between short bond and long bond is shown in Fig. 4.4.

Feature	ΔE (meV)	Feature	ΔE (meV)
Se ratio	383 ± 52	(1NN Se with short bonds)^2	32 ± 2
(Se ratio)^2	760 ± 70	2NN (1 Long, 1 Short)	38 ± 2
1NN Se with long bond	-74 ± 4	2NN (1 Long)	-44 ± 2
1NN Se with short bonds	-51 ± 8	2NN (1 Short)	-29 ± 2

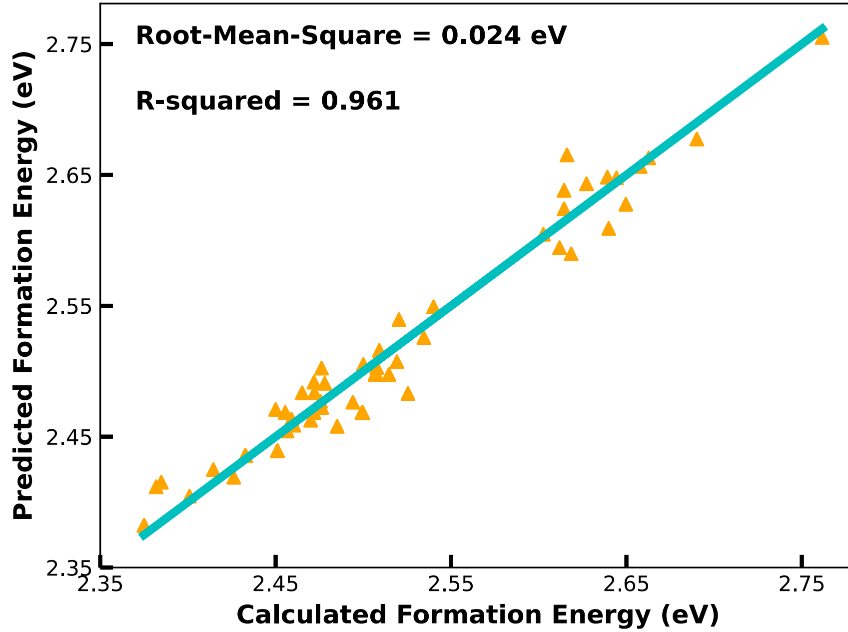


Figure 4.5: Comparison between Defect Neighbor Model predictions and DFT calculations for the Te_{Cd}^{0} defect across Se ratios ranging from 0% to 50% (56 data included). The root mean square error (RMSE) for test data and correlation coefficient (R^2) demonstrate excellent agreement with the data.

The defects V_{Te}, Cd_{int}, and Cd_{Te} are identified as shallow donors, exhibiting several common characteristics. Both V_{Te} and Cd_{Te} display negative-U behavior. The (+2/0) transition states of V_{Te} and Cd_{Te}, along with the (+2/+1) transition state of Cd_{int}, are situated near the conduction band minimum (CBM) or embedded within the conduction band (CB). Consequently, the +2 charge state is predominant for all three defects in p-type CdTe and its alloys. Only minor variations in their formation energies are observed with Se alloying, suggesting that these defects will continue to act as compensating defects in intrinsic CdSeTe alloys. Among these, Cd_{Te} exhibits the highest formation energy under both Cd-rich and Te-rich conditions, indicating that it is unlikely to be the dominant compensating defect. The neutral state of V_{Te} is most stable in a Jahn-Teller distorted structure, consistent with HSE06 calculations[226]. The V_{Te}^{+2} state is most stable in T_d symmetry, while Cd_{int}^{+2} is most stable when positioned in a tetrahedral vacant site surrounded by four group

VI atoms.

Interstitial tellurium Te_{int} is identified as a donor. We used ab-initio molecular dynamics (MD) to run simulated annealing, a technique for finding global energy minimums of Te_{int} without an initial guess structure. Essentially, the system begins at high temperature and is slowly cooled, eventually settling into a low energy configuration. MD was run in the canonical (NVT) ensemble, using the Nose-Hoover thermostat to control the system temperature. The system was advanced with a 6 femtosecond time step for 1000 steps, starting at 1000K and ending at 1K. We found Te_{int} prefers to form split interstitial oriented along the [110] directions. Ref. [112] find identical stable configurations. Some studies suggest that Te_{int} could be a deep-level defect[94, 111], and our findings indicate that the transition level of Te_{int} is indeed located near the mid-gap. However, the formation energy of Te_{int} is not particularly favorable, even under Te-rich conditions. Based on its formation energy and our compositionally constrained thermodynamic (CCT) method[133], we estimate that the density of Te_{int} at 300K is around three orders of magnitude lower than the carrier density and the densities of two other deep-level defects, Te_{Cd} and V_{Cd}, under Te-rich condition. Consequently, the effectiveness of Te_{int} in limiting carrier lifetime is questionable. This aspect will be further explored in our future work.

The defect Te_{Cd} can potentially form complexes with Te_{int} and V_{Cd} due to strong binding energy[93]. We have investigated the formation energies of the $V_{Cd} + Te_{Cd}$ and $Te_{int} + Te_{Cd}$ complexes in Fig. 4.2 and employed the CCT to estimate their densities. However, the results are similar to those for Te_{int}. We determined that the density of these two complexes at 300 K is approximately three orders of magnitude lower than the carrier density. Further investigation is required to ascertain whether these complexes significantly impact the defect distribution.

We have also explored dominant deep level defect formation energy in CdSeTe alloy with varied Se/Te ratio. Dominant defect formation energy in CdTe and CdSeTe for different charge states are depicted in Fig. 4.2. It is clearly observable that deep level defects have higher formation energy in Cd-rich condition. In fact, both experimental observation and HSE06 results validate this observation that as CdTe becomes more Cd-rich, a longer minority carrier lifetime can be achieved[111].

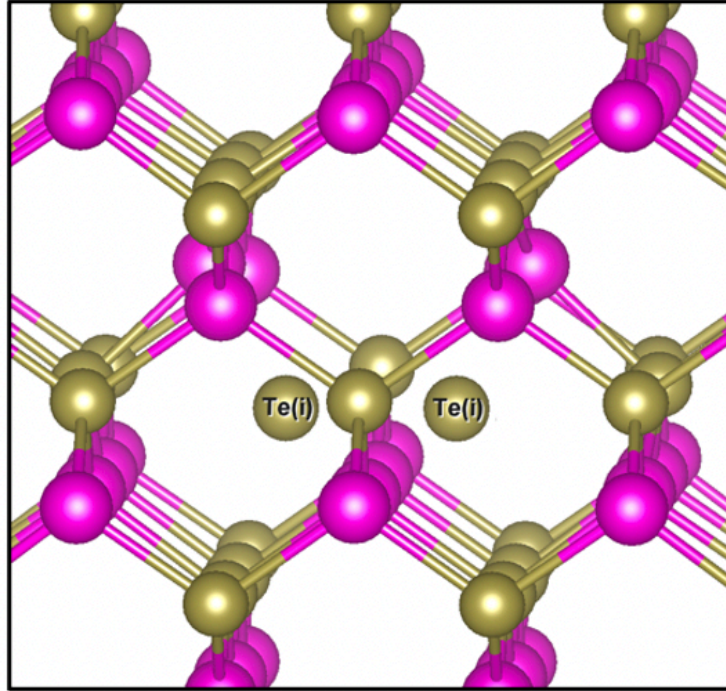


Figure 4.6: Schematic of Te_{int} split defect configuration

This conclusion remains valid in CdSeTe alloy. Moreover, from experiment observation[9, 91], the bulk recombination in CdSeTe alloy is lowered in CdTe. Fig. 4.2 shows that the formation energy of deep level defects (Te_{Cd} and V_{Cd}) slightly increases as we raise the Se ratio. Specifically, The E_f of V_{Cd}^0 is 0.15 eV higher in 50% alloy while E_f of Te_{Cd}^{+2} is 0.18 eV higher. This suggests a lower density of deep level centers in CdSeTe alloy. In addition, Se alloying could possibly deactivate the recombination centres by forming Se-complexes[91]. Thus, it is expected that bulk recombination will be reduced in CdSeTe alloy.

Examination of Fig. 4.2d, Fig. 4.2e, and Fig. 4.2f reveals that the Fermi level E_F is consistently pinned around 0.6 eV across CdTe, $CdSe_{0.25}Te_{0.75}$, $CdSe_{0.50}Te_{0.50}$ compositions for Te-rich condition. Under such condition, the dominant acceptor is V_{Cd} and the dominant compensating donors are Te_{Cd}, Cd_{int}, V_{Te} and Se_{Cd}. For Cd-rich condition, E_F is observed to be pinned between 1.2 eV and 1.3 eV for the same compositions. Under such condition, the dominant acceptor is V_{Cd} and

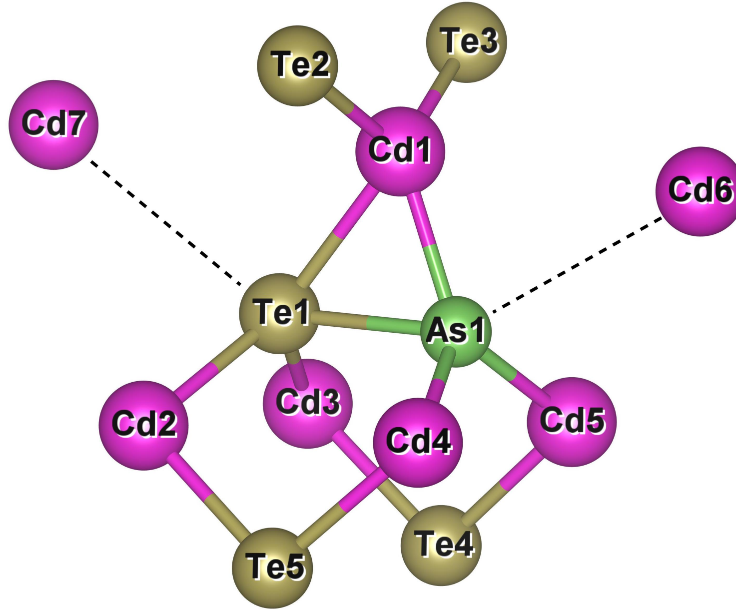


Figure 4.7: Schematic of AX center defect configuration

the dominant compensating donors are Cd_{int}, and V_{Te}. These findings align closely with Yang's first-principles calculations using the HSE06 functional[226], highlighting the intrinsic self-doping limits under equilibrium growth conditions. Interestingly, this limitation may potentially be circumvented by employing rapid cooling from high temperatures[226]. Nonetheless, the current method for calculating defect formation energies in the alloy, utilizing a Boltzmann distribution average at 873 K, simplifies the analysis but might not adequately address the intricacies arising from defects significantly influenced by the Se/Te arrangement. In addition, defect interaction is omitted when the material is cooling down from high temperature. Achieving a more detailed defect profile necessitates the integration of Lattice Monte Carlo simulations and continuum simulations[218, 55, 219]. These advanced simulation techniques are planned for inclusion in our future work, promising a deeper understanding of defect dynamics within the alloy.

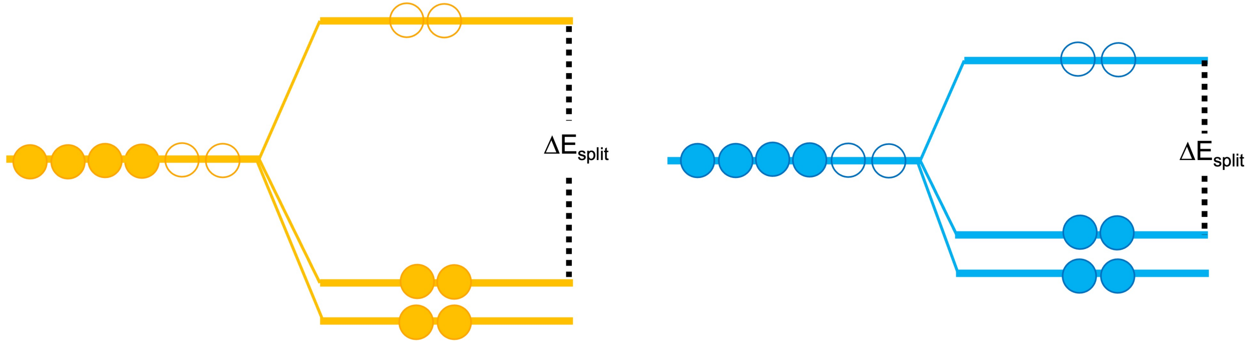


Figure 4.8: Schematic of AX center defect band splitting. Left one indicates AX in CdTe and right one indicates AX in $CdSe_{0.25}Te_{0.75}$. $\triangle E_{split}$ is the band splitting energy

4.2.2 Group V Dopants

Arsenic (As) and phosphorus (P) have been identified as effective dopants in CdTe-based materials, offering potential for enhancing p-type doping in solar cells[225]. However, the behavior of As and P in CdSeTe, such as the primary sources of compensating defects in As- and P-doped CdTe, remains less understood[30].

As anticipated, As can substitute for Te or Se in CdSeTe, acting as an acceptor due to its group V nature, possessing one electron fewer than Te. In this study, we focus on the substitutional defect As_{Te} and P_{Te}, as other related defects, like As/P interstitials, exhibit relatively higher formation energies[225]. Our calculations show that the neutral and -1 charged states of As_{Te} and P_{Te} display T_d symmetry, consistent with previous reports[225]. The calculated (0/-1) transition levels for As_{Te} and P_{Te} are 0.13 eV and 0.18 eV, respectively (Table 4.6, Table 4.7), indicating shallow levels suitable for *p*-type CdTe doping. However, the transition level of As_{Te} shifts toward the CBM, reaching 0.31 eV above the valence band maximum (VBM) when the Se ratio is 50%. Conversely, the transition level of P_{Te} remains close to the VBM even as the Se ratio increases. This suggests that for As-doped CdSeTe, careful control of the Se ratio is necessary to maintain reasonable As doping efficiency.

Furthermore, As and P can form a +1 charged AX center defect, converting it into a donor[225, 89]. As depicted in Fig. 4.7, this asymmetric defect occurs when As/P shifts toward one of the

neighboring Te atoms, causing the triply degenerate T_d state to split into two fully occupied states and one empty state, as illustrated in Fig. 4.8. This process results in the breaking of one bond with a neighboring Cd atom for each of the two atoms. If the band splitting energy ΔE_{split} exceeds the bonding energy E_{bond} with Cd, the AX defect becomes more stable than the tetrahedral As/P substitutional defect. However, our calculations indicate that in CdTe the formation energy of the As AX center is 1.92 eV, similar to that of T_d As_{Te}^{+1} at 1.93 eV. Similarly, the P AX center has a formation energy of 2.04 eV, close to T_d P_{Te}^{+1} at 2.10 eV, suggesting that the AX center does not provide additional stability.

As the Se ratio increases, the T_d +1 charge state of As_{Te} exhibits a formation energy of 1.93 eV in CdTe, 1.89 eV in $CdSe_{0.25}Te_{0.75}$, and 1.84 eV in $CdSe_{0.50}Te_{0.50}$, indicating a slight preference for the T_d +1 defect with higher Se alloying. However, the formation energy of the AX defect significantly decreases with increasing Se ratio in the CdSeTe alloy. For instance, in $CdSe_{0.25}Te_{0.75}$, the formation energy of the AX defect ranges from 1.57 eV to 1.79 eV for different Se/Te arrangements, compared to 1.92 eV in CdTe. This suggests a possible interaction between Se and the AX defect that could lower the formation energy of the AX defect and compensate for p-type doping. As shown in Fig. 4.8, the favorable formation energy of the AX defect is not attributed to an increase in ΔE_{split}, but rather to the nearby Se atom potentially weakening the bond between Cd and As/Te. In fact, our study of a 64-atom supercell of CdSeTe alloy with only one Se atom in the group VI lattice revealed that the lowest formation energy of the AX defect occurred when either the Te4 or Te5 atom was replaced with a Se atom. From Fig. 4.7, it can be inferred that when a smaller Se atom is placed into positions Te4 and Te5, the nearby Cd atoms are compressed in the direction of As1-Te1, strengthening the As1-Te1 bond but weakening the As1-Cd6 and Te1-Cd7 bonds. Moreover, we observed that the As atom tends to move toward the Te atom to form a dimer structure in the alloy, while moving toward the Se atom is not energetically favorable. A similar behavior is observed for P_{Te} as well.

To determine the optimal group VI positions for AX defect formation, we established a new coordinate system in a 64-atom supercell of the CdSeTe alloy. In this system, an As/P atom was

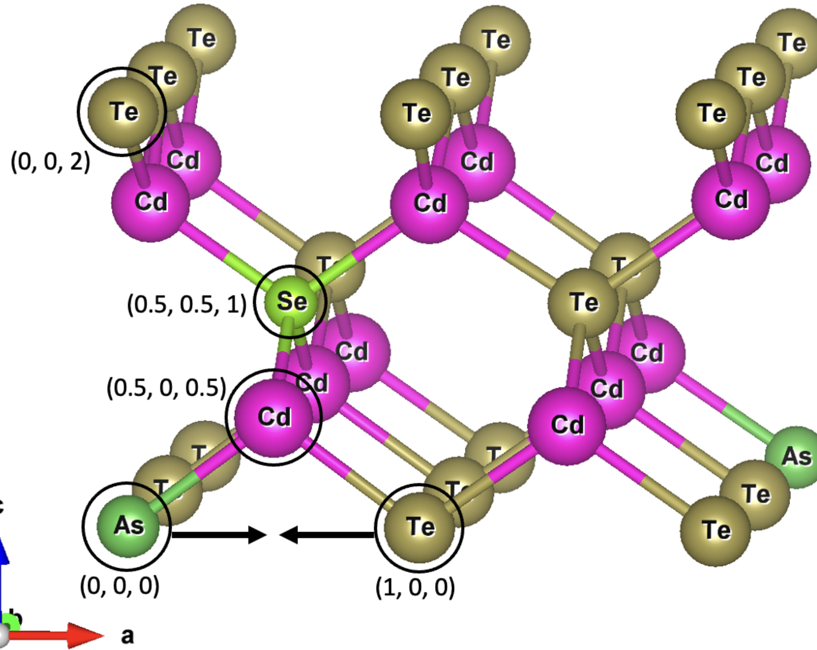


Figure 4.9: New coordinate system for defining atom locations relative to an oriented AX dimer. The coordinates are defined by As (0, 0, 0), the distorted Te (1, 0, 0), and the Cd bonded to both (0.5, 0, 0.5). Additional sites for a Se at (0.5, 0.5, 1) and a Te at (0, 0, 2) are also depicted. Black arrows show the As and Te deformation when the AX center forms.

placed at the origin, and a distorted Te atom was located at (1, 0, 0), as depicted in Fig. 4.9, with all other atoms labeled accordingly. This expanded 128-atom configuration enabled us to systematically vary the number of Se atoms at different vector positions originating from As/P and assess their impact on AX defect formation. As the b-direction vectors are symmetrically equivalent, their sign is considered arbitrary.

We calculated 54 formation energies of AX defects for As in CdSeTe with Se alloy ratios ranging from 0% to 50% and developed a regression model (Table 4.4, Fig. 4.10a) to predict the AX defect formation energy based on Se/Te ordering. A similar model was also constructed for phosphorus defects, as shown in Table 4.5 and Fig. 4.10b. In both Table 4.4 and Table 4.5, some vectors exhibit a positive effect on formation energy, while others have a negative effect. This can be attributed to

Table 4.4: Arsenic AX Center Defect Vector Model

Vector	ΔE (eV)	Vector	ΔE (eV)
$(0.5, \pm 0.5, -1)$	-0.100	$(0, \pm 1, 2)$	-0.015
$(0.5, \pm 0.5, 1)$	-0.030	$(-1, 0, 2)$	0.023
$(-0.5, \pm 0.5, -1)$	0.045	$(1.5, \pm 0.5, 1)$	-0.020

Table 4.5: Phosphorus AX Center Defect Vector Model

Vector	ΔE (eV)	Vector	ΔE (eV)
$(0.5, \pm 0.5, -1)$	-0.110	$(0, \pm 1, 2)$	-0.017
$(0.5, \pm 0.5, 1)$	-0.024	$(-1, 0, 2)$	0.009
$(-0.5, \pm 0.5, -1)$	0.054	$(1.5, \pm 0.5, 1)$	-0.023

the strain effects caused by the formation of AX defects and the substitution of Se for Te, which results from the mismatch in atomic radii. If these two strain effects lead to similar lattice expansion or contraction, it will favor the formation of AX defects.

To further unveil the impact of AX center defects on group V dopability, we explore the transition mechanism between the AX center defect and the neutral As_{Te} substitutional defect within a 25% CdSeTe alloy, we plotted the configuration coordinate diagram (CCD) of As_{Te} (+1_AX/0) in Fig. 4.11a. The barrier (~ 0.4 eV) between the AX and T_d As_{Te}^{+1} states likely limits the occurrence of AX defects, a finding consistent with similar transition barriers reported in CdTe [89]. Although certain alloy arrangements may promote AX defect formation, the overall quantity of AX defects may be limited, due to the transition barrier between the AX defect and the positively charged T_d As_{Te} defect. Recent computational findings [30] suggest that AX centers do not constitute a bottleneck for p-type doping in CdTe and CdSeTe alloys. However, the significance of AX centers on the dopability of group V elements requires further investigation, taking into account both the kinetic and thermodynamic aspects of AX defect formation.

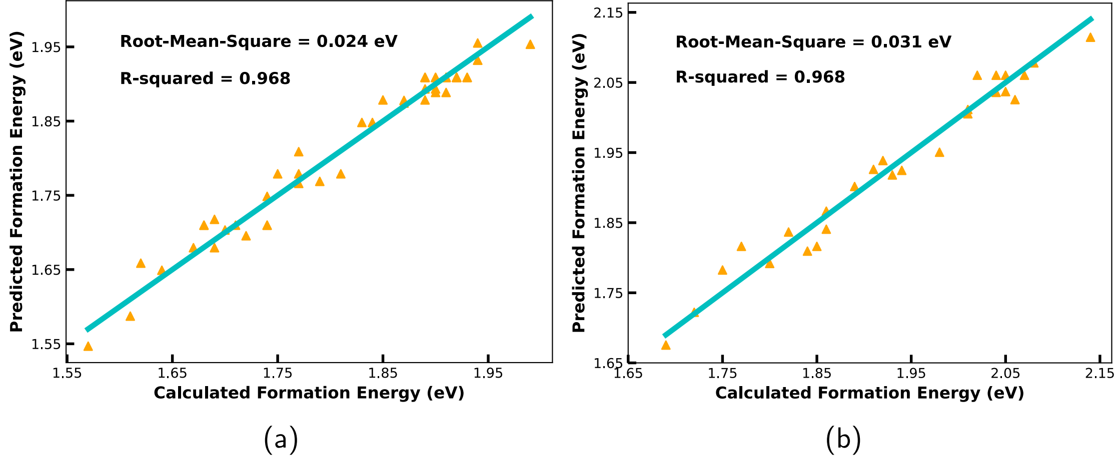


Figure 4.10: Defect Vector Model Prediction vs Calculation of (a) Arsenic AX defect and (b) Phosphorus AX defect in CdSeTe alloy (Se ratio ranges from 0% to 50%).

Table 4.6: Arsenic Defect Formation Energy and Complex Binding Energy (eV) in $CdSe_xTe_{1-x}$.

	As_{Te}^{+1} AX	As_{Te}^{+1} T_d	As_{Te}^{0}	As_{Te}^{-1}	(-1/0)	$E_{binding}$ $(Cd_{int} + As_{Te})^+$	$E_{binding}$ $(V_{Te} + As_{Te})^+$
CdTe	1.92	1.93	1.94	2.07	0.13	1.21	0.55
$CdSe_{0.25}Te_{0.75}$	1.57∼1.79	1.89	1.93	2.11	0.18	1.32	0.37
$CdSe_{0.50}Te_{0.50}$	1.70∼1.87	1.84	1.87	2.19	0.31	1.35	0.47

4.2.2.1 Defect Complex

A key unresolved issue in the study of compensating donors in CdSeTe is the possible existence of dominant compensating donors other than the AX center. Chatratin 2023 suggests that V_{Te} and Cd_{Te} could serve as potential compensating donors, although Cd_{int} is excluded from consideration due to its high mobility[30]. However, the likelihood of Cd_{Te} becoming the predominant compensating donor is debatable, given its relatively unfavorable formation energy, especially under Cd-rich conditions where As and P doping prove most effective [135, 220]. Thus, we lean towards V_{Te} and Cd_{int} as the likely dominating compensating defects.

Despite concerns regarding the instability caused by the high mobility of Cd_{int}, we find that Cd_{int}^{+2} can form stable complexes with As_{Te}^{+1} or P_{Te}^{+1} through electrostatic interactions and/or covalent bonding, resulting in $(Cd_{int} + As_{Te})^+$ (Fig. 4.12) or $(Cd_{int} + P_{Te})^+$ complexes. The formation

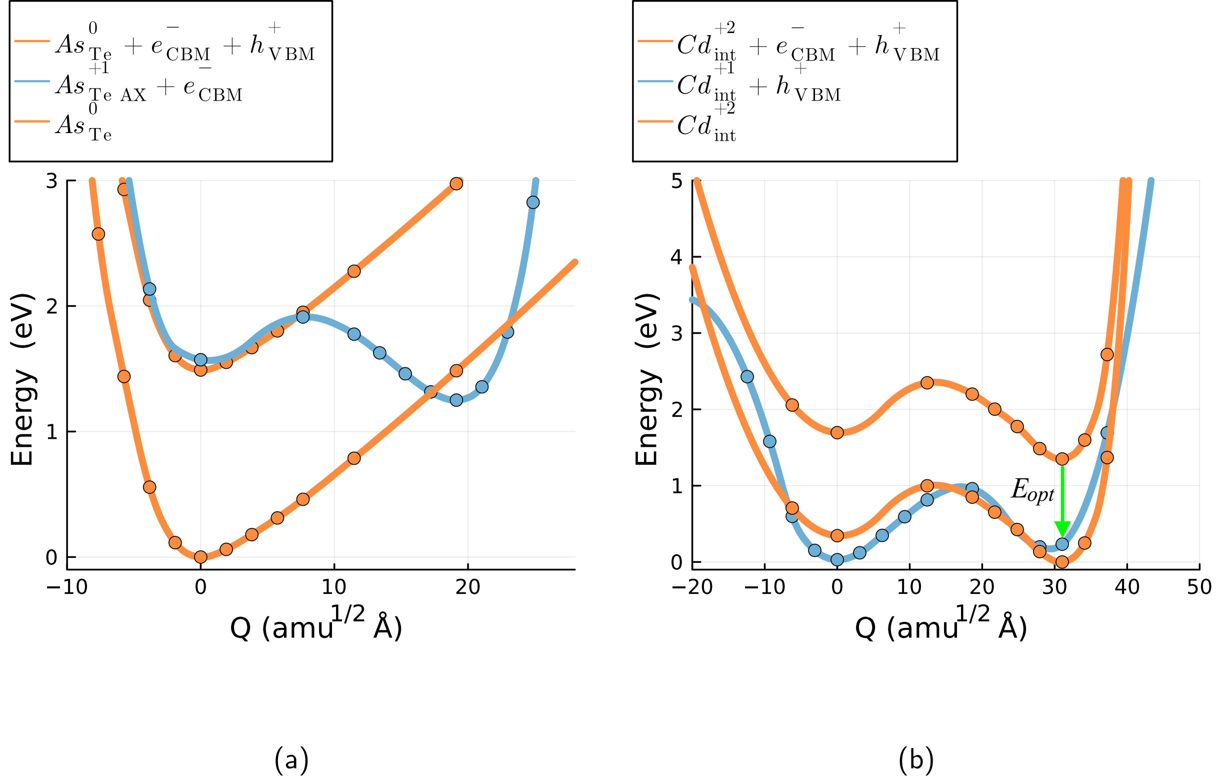


Figure 4.11: Configuration Coordinate Diagram of As_{Te} (+1_AX/0) and Cd_{int} (+2/+1) in $CdSe_{0.25}Te_{0.75}$. As^{+}_{Te} with formation energy close to the Boltzmann distribution at 300 K is selected as the excited state. As^{0}_{Te} exhibits minimal dependence on the Se/Te arrangement, and the choice of As^{0}_{Te} does not significantly affect the results. In (b), E_{opt} corresponds to the optical transition level. $Q = 0$ indicates interstitial ground state positioned in a octahedra vacant site. $Q \approx 30$ indicates interstitial ground state positioned in a tetrahedral vacant site. Cd^{+1}_{int} is favorable in octahedra site while Cd^{+2}_{int} is favorable in tetrahedral site.

energies of these complexes are comparatively low, indicating their potential significance in the CdSeTe system (see Table 4.6, Table 4.7). Additionally, while there is a noticeable binding energy between Group V acceptors and V_{Te}, it is significantly lower than that observed with Cd_{int}, leading us to discount $(V_{Te} + As_{Te})^{+}$ and $(V_{Te} + P_{Te})^{+}$ as the most dominant donor.

We carefully examine the bonds between Cd^{+2}_{int} and As^{+1}_{Te} or P^{+1}_{Te}, as detailed in Table 4.8. This table reports the bond lengths (in angstroms, Å) and bond orders (BO) for the nearest neighbor

Table 4.7: Phosphorus Defect Formation Energy and Complex Binding Energy (eV) in $CdSe_xTe_{1-x}$.

	P_{Te}^{+1} AX	P_{Te}^{+1} T_d	P_{Te}^{0}	P_{Te}^{-1}	(-1/0)	$E_{binding}$ $(Cd_{int}+P_{Te})^{+}$	$E_{binding}$ $(V_{Te}+P_{Te})^{+}$
CdTe	2.04	2.10	1.93	2.04	0.11	1.14	0.35
$CdSe_{0.25}Te_{0.75}$	1.72∼1.94	2.08	2.05	2.09	0.04	1.31	0.31
$CdSe_{0.50}Te_{0.50}$	1.82∼2.01	2.10	2.07	2.13	0.06	1.31	0.38

atoms surrounding Cd interstitials within $(Cd_{int} + As_{Te})^{+}$ and $(Cd_{int} + P_{Te})^{+}$ complexes in both CdTe and CdSeTe alloys. The BO values, estimated via the density derived electrostatic and chemical (DDEC) approach [122], range from 0 (indicative of a purely ionic bond) to 1 (indicative of a purely covalent bond). From the data, it is observed that the bond lengths between Cd-Te, Cd-Se, Cd-As, and Cd-P decrease very slightly as the alloy ratio increases, potentially due to lattice constant contraction. Additionally, the BO values exhibit minimal variation across different alloy ratios, suggesting that the bonding nature of $(Cd_{int}+As_{Te})^{+}$ and $(Cd_{int}+P_{Te})^{+}$ remains relatively consistent regardless of alloy composition. We infer from these metrics that the Cd-Te and Cd-Se bonds approximate a 50% ionic and 50% covalent character, whereas the bonds between Cd-As and Cd-P are predominantly covalent. This predominance of covalent bonding contributes to the strong bonding energies observed, indicating the significance of both Coulomb interactions and covalent bonding in these complexes.

Table 4.8: Bond Lengths (Å) and Bond Orders for Nearest Neighbors Surrounding Cd Interstitial in $(Cd_{int}+As_{Te})^{+}$ and $(Cd_{int}+P_{Te})^{+}$. Bond lengths are provided outside the brackets, while bond orders are denoted inside the brackets.

Defect Type	$(Cd_{int} + As_{Te})^{+}$			$(Cd_{int} + P_{Te})^{+}$		
Material \ Bond	**Cd - Te**	**Cd - Se**	**Cd - As**	**Cd - Te**	**Cd - Se**	**Cd - P**
CdTe	2.84 (0.52)	-	2.57 (0.64)	2.83 (0.52)	-	2.49 (0.66)
$CdSe_{0.25}Te_{0.75}$	2.82 (0.52)	2.65 (0.52)	2.55 (0.65)	2.81 (0.52)	2.64 (0.52)	2.47 (0.67)
$CdSe_{0.50}Te_{0.50}$	2.80 (0.55)	2.64 (0.53)	2.54 (0.69)	2.79 (0.55)	2.63 (0.54)	2.46 (0.70)

Recently, *Kuciauskas et al.*[99] investigated defects in CdSeTe with As doping[99]. Their spectral data revealed a defect with an activation energy of 0.14 eV – 0.22 eV from the CBM in undoped CdSeTe, which changes or disappears after As doping. They speculated that this defect could be

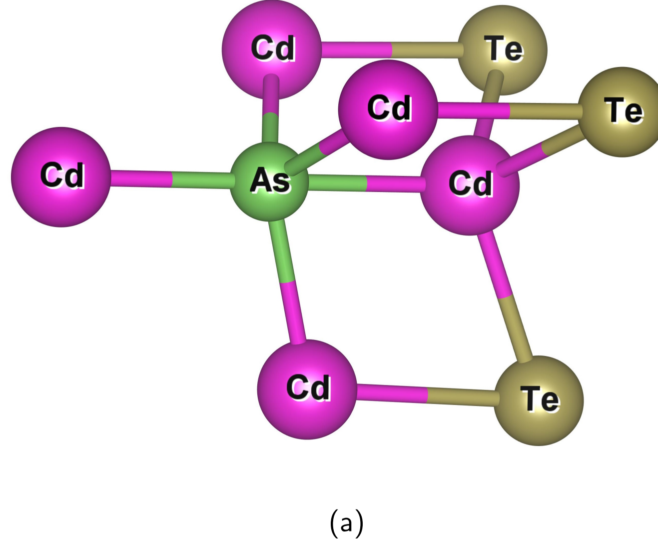


Figure 4.12: Defect Structure of $(Cd_{int} + As_{Te})^{+}$ in CdTe.

V_{Te}. To test this hypothesis, we calculated the formation energy and transition levels of V_{Te} in a 25% CdSeTe alloy. From two typical configurations of V_{Te}, we found that the transition levels are 0.64∼0.66 eV for (0/+1) and 1.38∼1.41 eV for (+1/+2), which do not support the assumption that the defect is V_{Te}.

Furthermore, we investigated the defect-assisted optical transition of Cd_{int} using first-principles methods[39]. Intriguingly, from the CCD of Cd_{int} in $CdSe_{0.25}Te_{0.75}$ shown in Fig. 4.11b, the optical transition level E_{opt} of tetrahedral site interstitial Cd_{int} is approximately 1.12 eV from the VBM, or 0.23 eV to the CBM. This level aligns well with the photoluminescence (PL) signal detected by *Kuciauskas et al.*[99]. The significant binding between Cd_{int} and As_{Te}^{+1} could explain why the PL signal disappears after As doping. These calculations suggest that the change in the PL signal is likely due to Cd_{int} rather than V_{Te}.

4.2.3 Copper

Copper (Cu) is well-known for enhancing carrier density in CdTe solar cells, leading to higher power conversion efficiency[104]. However, the doping instability of Cu and the presence of a large number of compensating defects[8, 92, 32] result in inferior efficiency and stability compared to group V-doped counterparts[104, 92]. To explore the dopability of Cu in CdSeTe, we calculated the formation energies of Cu defects under Cd-rich and Te-rich conditions (Fig. 4.13). Four dominant defects, Cu_{Cd}, Cu_{int}, $Cu_{Cd} + Cd_{int}$ and $Cu_{Cd} + Cu_{int}$, along with the dominant intrinsic defects, are considered in this study.

Our calculations show that the defect transition level of Cu_{Cd} (-1/0) in CdTe is 0.33 eV above the valence band maximum (VBM). Experimental results for this transition level vary between 0.15 and 0.37 eV[9, 56, 27, 169], with our result falling within the upper bound of this range, while HSE06 calculations suggest lower bound values around 0.20 eV[226]. Despite appearing deep in the gap, Cu can still act as an effective acceptor due to its relatively low formation energy, as shown in Fig. 4.13d. In CdTe, Cu_{int} and Cd_{int} emerge as dominant compensating defects, with the Fermi level being pinned around 0.3 eV, a significant improvement over intrinsic CdTe (Fig. 4.2d). However, as the Se ratio increases, the pinned level shifts toward the conduction band minimum (CBM) (Fig. 4.13e, Fig. 4.13f), rendering the system more n-type. This shift suggests a decrease in the doping efficiency of Cu in CdSeTe, which may account for the observed degradation in net acceptor density in Cu-doped devices[92].

In conclusion, we believe that group V dopants hold more promise than Cu in CdSeTe alloys. Apart from the doping efficiency challenges associated with Cu in CdSeTe, Cu implantation in CdTe typically necessitates a Te-rich environment[230, 8], which may lead to the introduction of more deep-level traps, as illustrated in Fig. 4.2 and Fig. 4.13. In contrast, group V dopants generally favor a Cd-rich environment[136, 135], which can enhance carrier lifetime in solar cells.

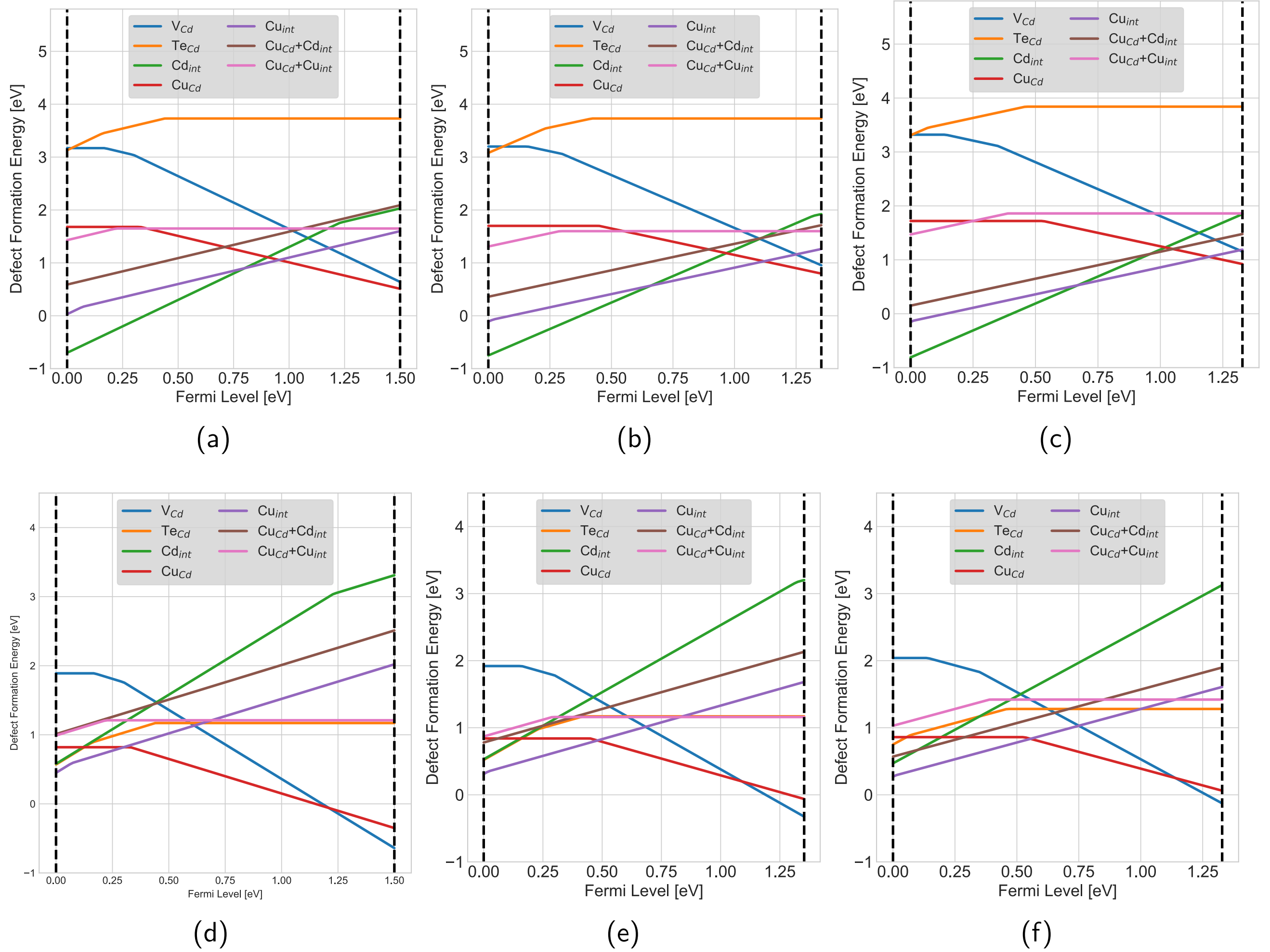


Figure 4.13: Defect formation energies vs Fermi level of Cu related defects and dominant intrinsic defects in CdTe under (a) Cd-rich and (d) Te-rich condition, $CdSe_{0.25}Te_{0.75}$ under (b) Cd-rich and (e) Te-rich condition, $CdSe_{0.50}Te_{0.50}$ under (c) Cd-rich and (f) Te-rich condition. The chemical potential condition is $\mu_{Cd} + \mu_{Te} = -1.28$ eV, $\mu_{Cd} + \mu_{Se} < 0.15$ eV and $\mu_{Cu} + \mu_{Te} < -0.42$ eV.

4.3 Continuum Modeling and Optimization of Group-V Doping in CdTe and CdSeTe

Building on the insights derived from the point defect calculations, we now shift our focus to the continuum model, which allows for a more comprehensive understanding of the defect dynamics and thermodynamics in the CdSeTe system under varying conditions.

4.3.1 Defect Thermodynamics

Compositionally Constraint Thermodynamics (CCT)[133] is a canonical approach to calculate the defect concentrations. Under a dilute approximation, defect concentrations are given by

$$x_{k,q} = \frac{\theta_{k,q}}{N^{total}} e^{-\Delta H^{f}_{k,q}/k_{\mathrm{B}}T}, \tag{4.2}$$

where $\theta_{k,q}$ is the degeneracy factor that counts the number of equivalent defect configurations, $\Delta H^{f}_{k,q}$ is the formation energy of defect k at charge q, which we obtained from KLMC and DFT. The formation entropy is neglected. N^{total} is the total number of lattice sites in a perfect material. One can obtain deviations of atomic fractions f_α from perfect stoichiometry f^{p}_{α} under given defect concentrations in the material using the following relation

$$f_\alpha = \frac{f^{p}_{\alpha} - \sum_{k,q} x_{k,q} n^{\alpha}_{k}}{1 - \sum_{k,q} x_{k,q} \sigma_{k,q}} , \tag{4.3}$$

where n^{α}_{k} is the number of atoms of type α added or removed from the system when one defect k exists. $\sigma_{k,q} = \sum_\alpha n^{\alpha}_{k}$ is the atom amount differences in a system with and without the defect k. If atomic fractions and standard reference chemical potentials are provided, one can determine the defect concentration, chemical potentials for each element and the Fermi level by solving the set of equations defined by (5.6). In this study, the atomic fractions in CdSeTe are hard to determine with sufficient accuracy via characterization, since the material is very close to stoichiometric. Instead, we can use the chemical potentials of Cd, Te, Se and extrinsic dopants and impurities under thermodynamic constraints to determine the defect concentrations and the Fermi level. If the formation energies $H^{(r)}_{k,q}$ are known for given reference values $\mu^{(r)}_{\alpha}$ and $E^{(r)}_{F}$, then $H_{k,q}$ for any values of μ_α and E_F can be expressed using the following equations:

$$\begin{aligned} \Delta H_{k,q}(\mu_\alpha, E_F) = \Delta H^{(r)}_{k,q} + \sum_{\alpha} n^{\alpha}_{i} (\mu_\alpha - \mu^{(r)}_{\alpha}) \\ + q(E_F - E^{(r)}_{F}). \end{aligned} \tag{4.4}$$

Using equations (4.2) and (4.4), we can represent the total defect density for any defect using the density of a single neutral defect:

$$x_k = x_{k,0} \sum_q \frac{\theta_{k,q}}{\theta_{k,0}} e^{-\frac{\Delta H^{(r)}_{k,q} - \Delta H^{(r)}_{k,0} + q(E_F - E^{(r)}_F)}{k_B T}}. \tag{4.5}$$

From (4.2) and (4.5), we can easily obtain

$$\sum_\alpha n^\alpha_k \mu_\alpha = \sum_\alpha n^\alpha_k \mu^{(r)}_\alpha - k_B T \ln \frac{x_{k,0} \times N^{total}}{\theta_{k,0}} - \Delta H^{(r)}_{k,0}. \tag{4.6}$$

By applying equation (4.6) and the charge neutrality equation (4.7), we can determine the equilibrium defect distribution and the Fermi level for any given set of chemical potentials and temperature.

$$n + \sum_k \sum_{q<0} |q| x_{k,q} = p + \sum_k \sum_{q>0} q x_{k,q} \tag{4.7}$$

Here, we need to consider chemical potential constraints for each element. Under thermodynamic equilibrium, CdTe should satisfy

$$\mu_{Cd} + \mu_{Te} = \Delta H_f(CdTe) = -1.28(eV). \tag{4.8}$$

$\Delta H_f(CdTe)$ comes from previous GGA+U calculation[220], and this formation energy value is very close to experimental measurement -1.30(eV)[18]. There are other competing phases when group V is introduced, such as Cd_3As_2, $CdAs_2$, and CdP_2. However, due to the difficulty determining the formation energy using solely GGA or GGA+U[74], we instead use experimental values for these

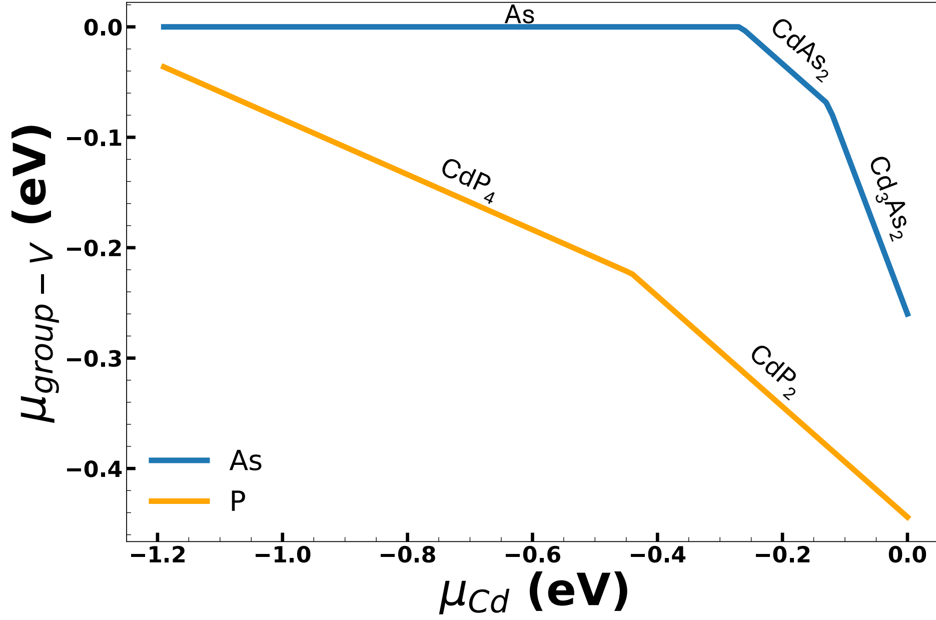


Figure 4.14: Chemical potential range for Group-V dopants. Compounds above indicate competitive phases.

compounds[156, 26, 143, 67, 95]. This gives following constraints:

$$\begin{aligned}
\mu_{Cd} &\leq \mu\,(Metal\,Cd), \mu_{Te} \leq \mu\,(bulk\,Te), \\
\mu_{As} &\leq \mu\,(bulk\,As), \mu_{P} \leq \mu\,(bulk\,P), \\
3\mu_{Cd} + 2\mu_{As} &\leq \Delta H_f(Cd_3As_2) = -0.52(eV), \\
\mu_{Cd} + 2\mu_{As} &\leq \Delta H_f(CdAs_2) = -0.267(eV), \\
3\mu_{Cd} + 2\mu_{P} &\leq \Delta H_f(Cd_3P_2) = -0.79(eV), \\
\mu_{Cd} + 2\mu_{P} &\leq \Delta H_f(CdP_2) = -0.888(eV), \\
\mu_{Cd} + 4\mu_{P} &\leq \Delta H_f(CdP_4) = -1.335(eV).
\end{aligned} \tag{4.9}$$

The resulting chemical potential range for Group-V dopants is depicted in Fig. 4.14.

Table 4.9: CdSeTe Defect Migration Barrier (NEB)

Migration barrier	E_m in eV
V_{Cd}	0.93
V_{Te}	1.20
Cd_i	0.52
Te_i	0.16

4.3.2 Process Simulation

CCT can be applied to extract defect densities in thermodynamic equilibrium[218, 219]. Process simulation described in Section 2.2.2 is then be employed to simulate defect reactions and diffusion processes during cooling and annealing.

We consider V_{Cd}, V_{Te}, Cd_i and Te_i as mobile species. The migration barriers are determine via Nudged Elastic Band (NEB) in VASP[62] (Table 4.9). For defect ionization reactions, Eq (2.12) is used, where $\sigma_{n,p}$ is defect electron/hole capture cross section and v_{th} is thermal velocity. The electron/hole capture cross sections are estimated via the NONRAD method[5, 198].

Table 4.10 presents a comprehensive list of defect reactions for the As-doped CdSeTe solar cell continuum model. We operate under the assumption that there are no fluxes or transfers across the interface; therefore, Neumann boundary conditions are applied to all defects. The defect reactions summarized in Table 4.10 can be categorized into four main types: vacancy-interstitial pair annihilation, complex formation, exchange reactions, and ionization reactions. These reactions are integral to understanding the defect dynamics within the solar cell structure.

4.3.3 Results and Discussion

Continuum models are applied to both As-doped and P-doped CdTe as well as $CdSe_{0.25}Te_{0.75}$ under 873K, 1023K, and 1173K. These results are shown in Fig. 4.16, Fig. 4.18, Fig. 4.17, and Fig. 4.19. We plots domiant defect densities, carrier densities and activation percent in these plot. Here, the

Table 4.10: Defect Reactions in Continuum Model: Example of As dopant

Vacancy-Interstitial Pair Annihilation	**Complex Formation**	**Ionization Reaction**
$V_{Cd}^{2-} + Te_i^{2+} \rightarrow Te_{Cd}^{0}$ $V_{Cd}^{-} + Te_i^{2+} \rightarrow Te_{Cd}^{+}$ $V_{Cd}^{0} + Te_i^{2+} \rightarrow Te_{Cd}^{2+}$ $V_{Cd}^{0} + Te_i^{0} \rightarrow Te_{Cd}^{0}$ $V_{Cd}^{2-} + Cd_i^{2+} \rightarrow (Cd_{Cd})$ $V_{Cd}^{-} + Cd_i^{2+} \rightarrow (Cd_{Cd}) + h^+$ $V_{Te/Se}^{2+} + Te_i^{0} \rightarrow (Te_{Te/Se}) + 2h^+$ $V_{Te/Se}^{2+} + Cd_i^{2+} \rightarrow Cd_{Te/Se}^{2+} + 2h^+$ $V_{Cd}^{2-} + (As_{Te}+Cd_i)^+ \rightarrow As_{Te}^{-}$ $V_{Cd}^{-} + (As_{Te}+Cd_i)^+ \rightarrow As_{Te}^{0}$	$Te_i^{0} + Te_{Cd}^{0} \rightarrow (Te_i+Te_{Cd})^0$ $V_{Cd}^{2-} + Te_{Cd}^{2+} \rightarrow (V_{Cd}+Te_{Cd})^0$ $V_{Cd}^{-} + Te_{Cd}^{+} \rightarrow (V_{Cd}+Te_{Cd})^0$ $V_{Te}^{2+} + As_{Te}^{-} \rightarrow (As_{Te}+V_{Te})^+$ $Cd_i^{2+} + As_{Te}^{-} \rightarrow (As_{Te}+Cd_i)^+$ **Exchange Reaction** $Cd_i^{2+} + Te_{Cd}^{0} \rightarrow Te_i^{2+} + (Cd_{Cd})$	$Te_{Cd}^{+} + h^+ \rightarrow Te_{Cd}^{2+}$ $Te_{Cd}^{0} + h^+ \rightarrow Te_{Cd}^{+}$ $Te_i^{0} + h^+ \rightarrow Te_i^{+}$ $Te_i^{+} + h^+ \rightarrow Te_i^{2+}$ $Se_{Cd}^{+} + h^+ \rightarrow Se_{Cd}^{2+}$ $Se_{Cd}^{0} + h^+ \rightarrow Se_{Cd}^{+}$ $V_{Cd}^{2-} + h^+ \rightarrow V_{Cd}^{-}$ $V_{Cd}^{-} + h^+ \rightarrow V_{Cd}^{0}$ $As_{Te}^{-} + h^+ \rightarrow As_{Te}^{0}$ $As_{Te}^{0} + h^+ \rightarrow As_{Te}^{+}$ (AX)

activation is defined as

$$activation = \frac{Group\,V\,acceptor - Group\,V\,donor}{Total\,Group\,V\,Species}. \tag{4.10}$$

Additionally, we investigated the effects of varying cooling rates and storage durations in the simulations, which showed minimal impact on the final acceptor and carrier densities (Fig. 4.15). This is because certain reactions with high rates, such as ionization reactions, reach equilibrium rapidly at 300K. Other reactions may not reach thermodynamic equilibrium, but their slow rates prevent significant changes even after extended storage periods. This validates the stability of As and P doping.

In Fig.4.16, Fig.4.17, Fig.4.18, and Fig.4.19, it is clear that the AX defect is not the predominant compensating defect in As- or P-doped CdTe. However, due to favorable local arrangements in the CdSeTe alloy, AX defects can become more significant at specific μ_{Cd} levels. Higher initial temperature also amplifies the role of AX defects. Despite this, the complexes $(As_{Te}+Cd_i)^+$ and $(P_{Te}+Cd_i)^+$ tend to be the dominant compensating donors, except at low μ_{Cd} levels. Reducing μ_{Cd} decreases the influence of compensating defects and increases hole density, but at the cost of reducing

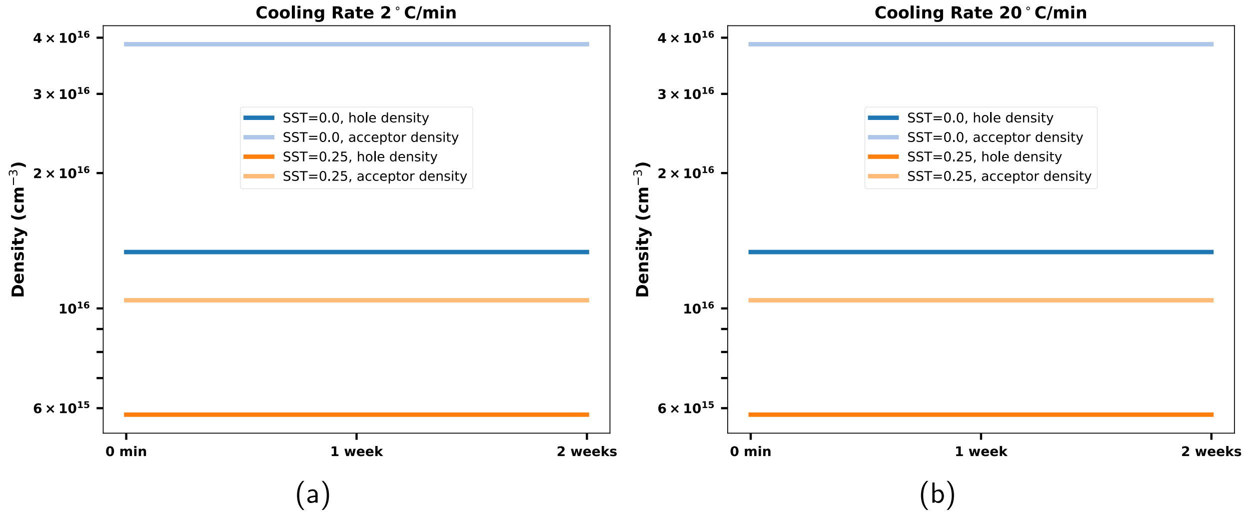


Figure 4.15: Hole and acceptor densities in As doped CdTe and $CdSe_{0.25}Te_{0.75}$ with μ_{Cd} = -0.80 eV. Initial temperature is 873K. Cooling rate is 2°C/min in (a) and 20°C/min in (b).

the solubility of As and P. In addition, in Te-rich conditions, deep-level defects V_{Cd} and Te_{Cd} become significant. Thus, an optimal μ_{Cd} is needed to minimize compensating defects while maintaining Group-V dopant solubility, achieving optimal carrier densities and carrier lifetime. These optimal ranges are approximately -0.80 eV for As-doped CdTe, -0.90 eV for As-doped $CdSe_{0.25}Te_{0.75}$ at 873K, -1.00 eV for As-doped CdTe and $CdSe_{0.25}Te_{0.75}$ at 1173K, -0.70 eV for P-doped CdTe, -0.80 eV for P-doped $CdSe_{0.25}Te_{0.75}$ at 873K, -0.90 eV for P-doped CdTe, and -0.95 eV for P-doped $CdSe_{0.25}Te_{0.75}$ at 1173K. Even though Cd-rich conditions favor Group V doping, overly Cd-rich conditions should be avoided due to the rise in compensating defects, while overly Te-rich conditions must also be avoided due to deep-level defects.

On the other hand, raising the initial temperature from 873K to 1173K results in about a tenfold increase in carrier density for both As- and P-doped CdTe and $CdSe_{0.25}Te_{0.75}$ from $\sim 10^{16} cm^{-3}$ to $\sim 10^{17} cm^{-3}$. The higher doping level is consistent with experimental report[135]. This is because Group V acceptors formed at 1173K are an order of magnitude higher in concentration than those formed at 873K, providing more carriers after cooling and annealing. While compensating donors also increase, $(As_{Te}+Cd_i)^+$ and $(P_{Te}+Cd_i)^+$ complexes are significantly reduced at optimal chemical

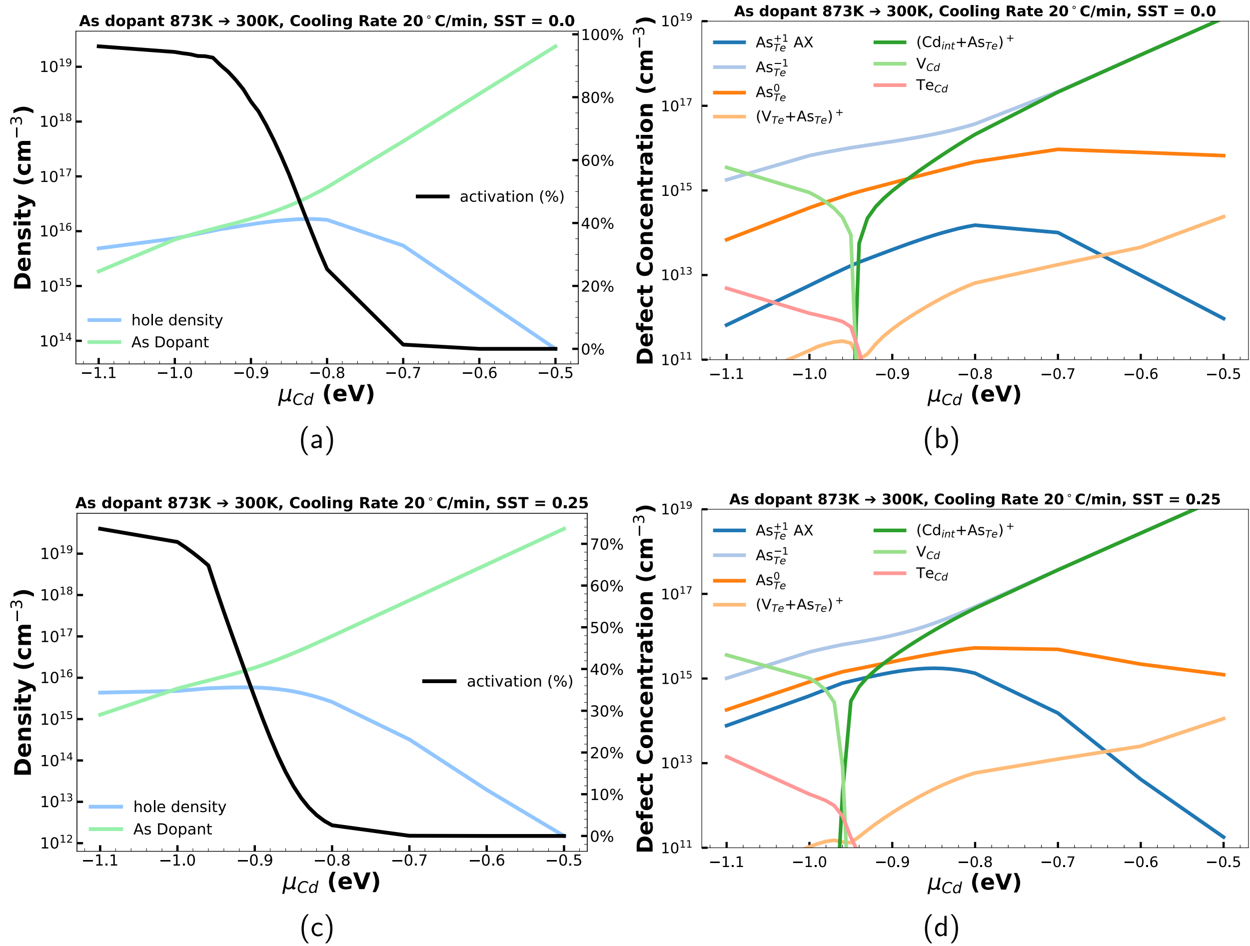


Figure 4.16: Dominant defect densities, carrier density and As activation in As doped CdTe and $CdSe_{0.25}Te_{0.75}$. Initial temperature is 873K. Cooling rate is 20°C/min.

potential conditions, enabling greater carrier densities. However, the AX defect become quite significant especially in CdSeTe alloy at 1173K, as shown in Fig.4.17d and Fig.4.19d. AX defects remain relatively stable across different chemical potentials and can even drive activation rates to negative values, making them a significant issue under these conditions (Fig. 4.17c). A further concern when increasing the initial growth temperature is the overlapping of the Cd chemical potential regions that yield high carrier densities with those that have high levels of deep level defects. This necessitates a careful balance between beneficial and detrimental effects.

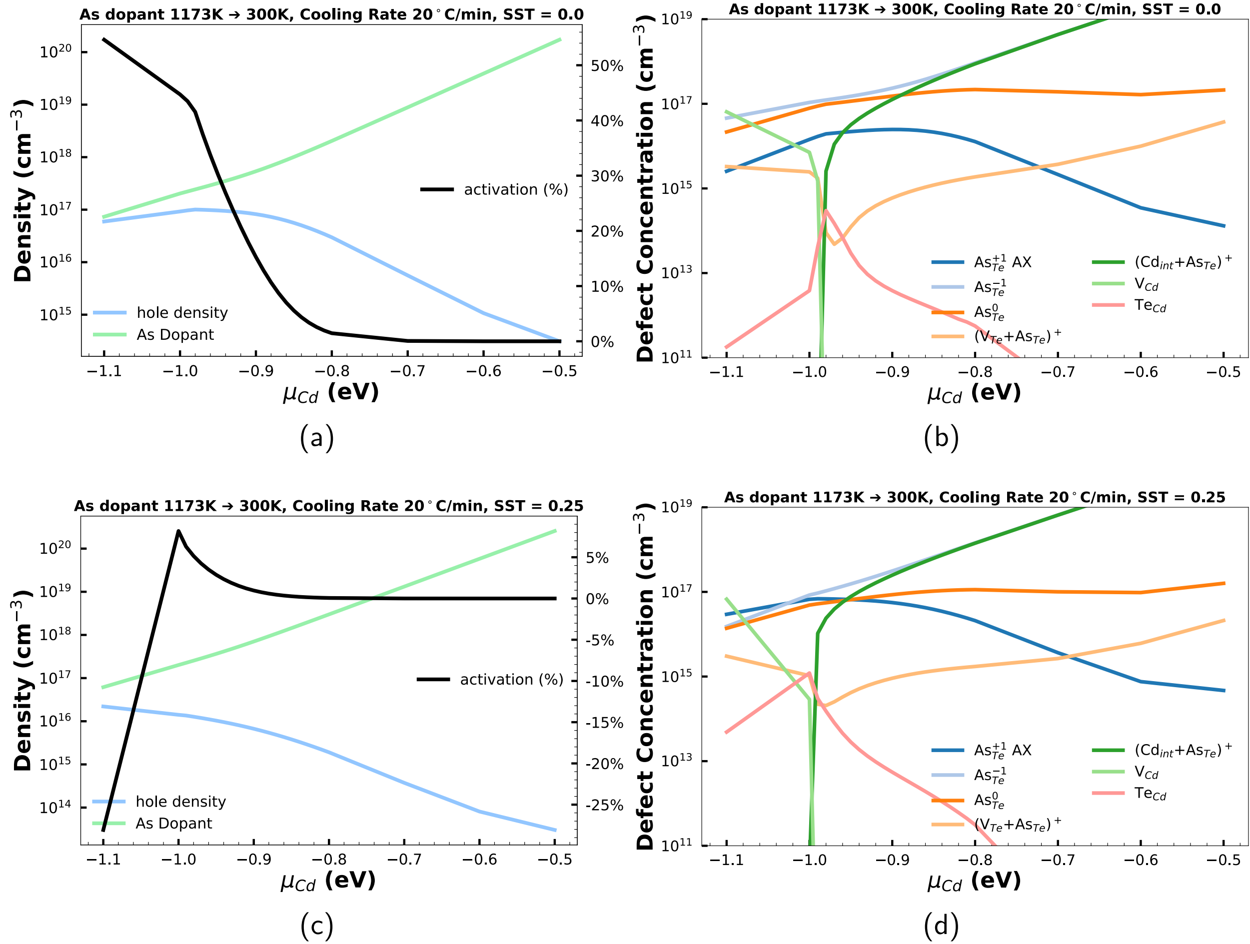


Figure 4.17: Dominant defect densities, carrier density and As activation in As doped CdTe and $CdSe_{0.25}Te_{0.75}$. Initial temperature is 1173K. Cooling rate is 20°C/min.

Further analysis of Figs. 4.16, 4.18, 4.17, and 4.19 reveals that optimal hole densities achieved by doping with P are slightly higher than those in As-doped CdTe, despite P's lower solubility due to the chemical potential constraints shown in Fig 4.14. This pattern holds true in $CdSe_{0.25}Te_{0.75}$ as well. This can be attributed to P's higher activation efficiency. This higher efficiency results from a significant reduction of the $(P_{Te}+Cd_i)^+$ complex at higher μ_{cd}. The lower chemical potential limits for P increase the formation energy of this defect complex, leading to a more pronounced reduction of this complex donor at higher μ_{cd} compared to As, which exhibits higher solubility. A similar

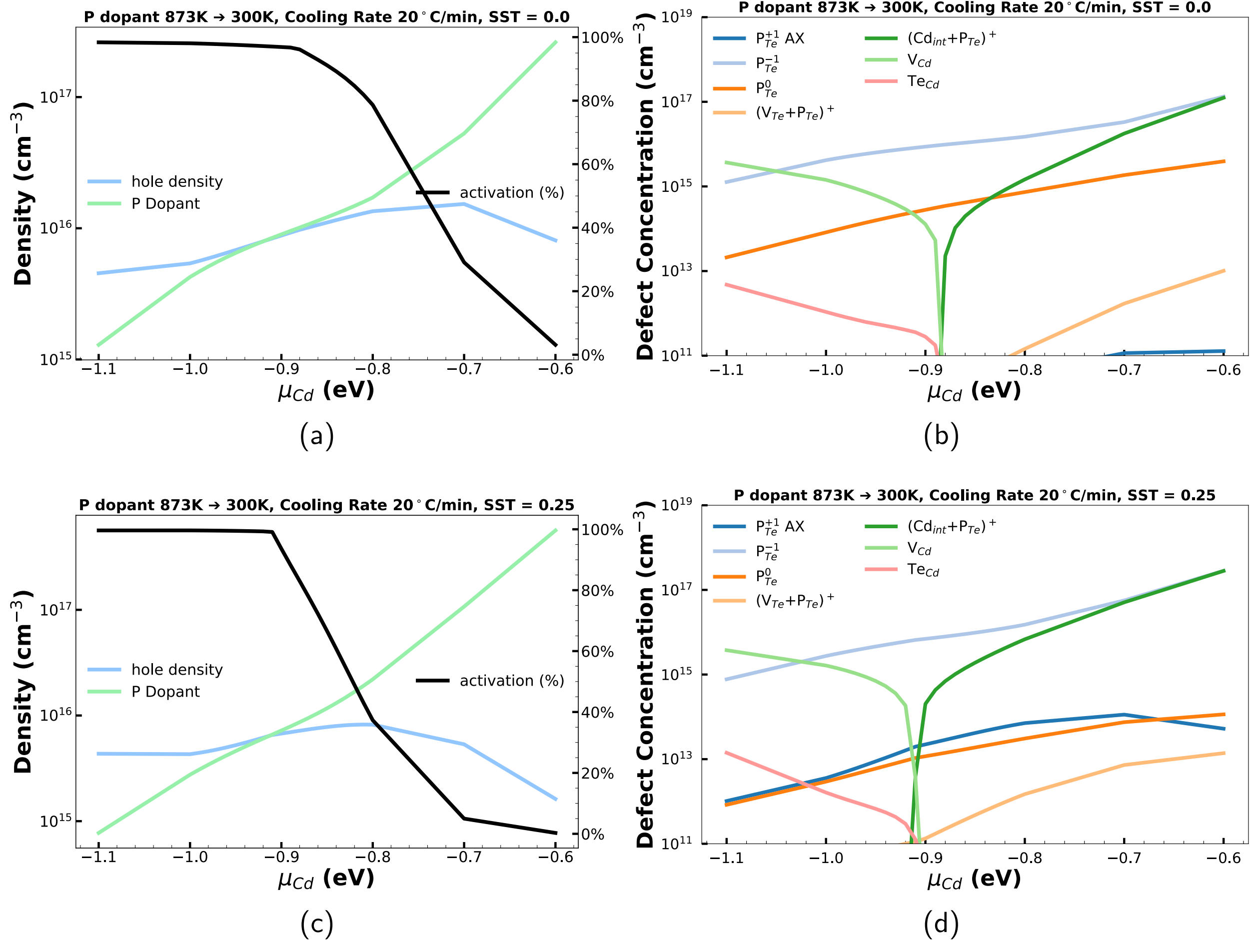


Figure 4.18: Dominant defect densities, carrier density and P activation in P doped CdTe and $CdSe_{0.25}Te_{0.75}$. Initial temperature is 873K. Cooling rate is 20°C/min.

trend is observed with a Se ratio at 25%. Notably, the AX defect is less favorable in P-doped CdTe and CdSeTe, rendering it less detrimental in these materials.

In comparison to previous experimental observations, Nagaoka et al. (2020) [135] fabricated Cd-rich CdTe single crystals doped with As at an initial temperature of 1173K, achieving a hole density close to $10^{17}, \mathrm{cm}^{-3}$ with a 15% activation efficiency at 300K. This result is remarkably consistent with our simulations shown in Fig.4.17a. For P-doped CdTe, the hole density is slightly higher than that of As-doped CdTe, as confirmed in Fig.4.19a. Several reports using As-doped polycrystalline

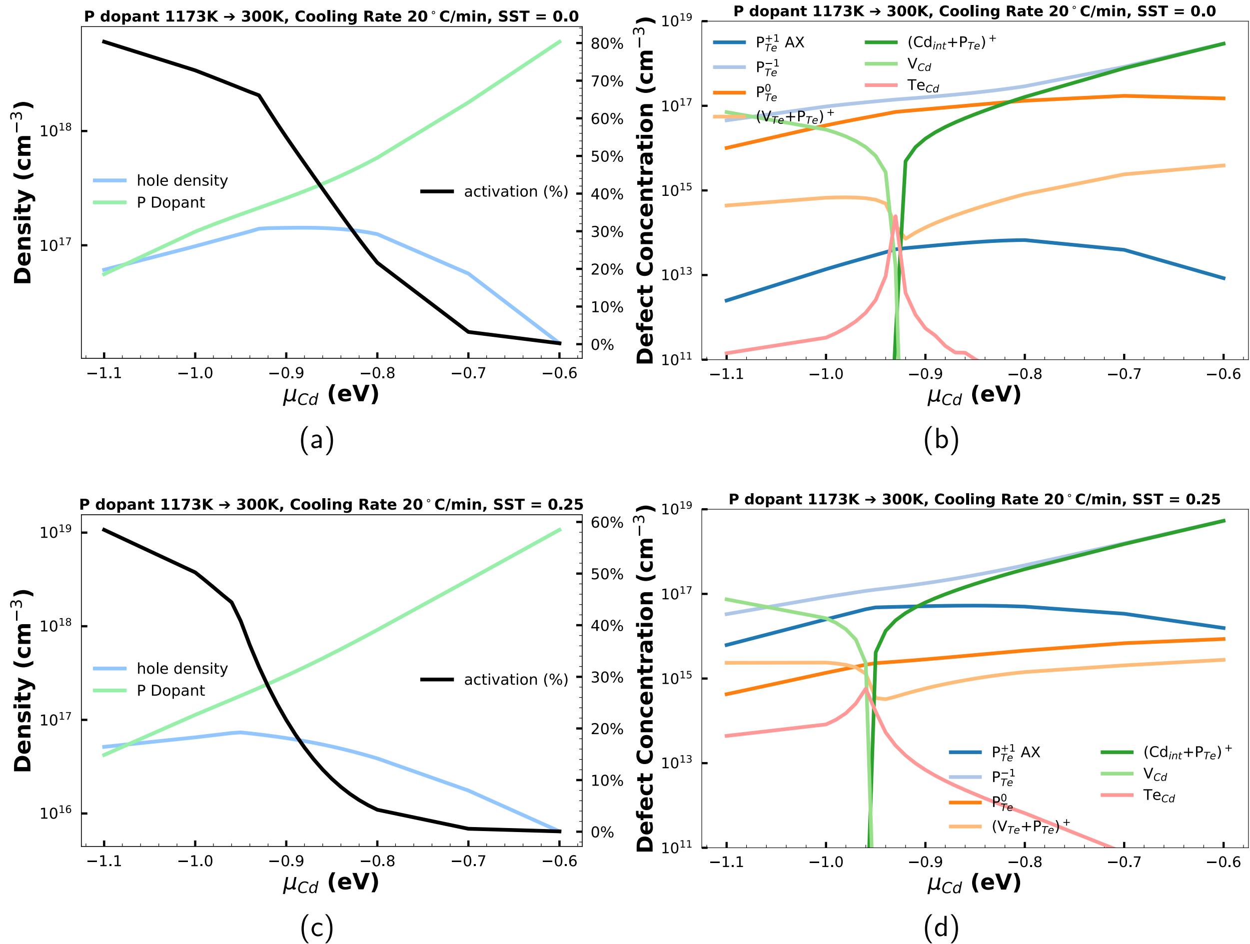


Figure 4.19: Dominant defect densities, carrier density and P activation in P doped CdTe and $CdSe_{0.25}Te_{0.75}$. Initial temperature is 1173K. Cooling rate is 20°C/min.

CdTe or CdSeTe show an As activation ratio ranging from 0% to 6% [78, 114, 1, 104]. Despite this low activation ratio, the reported hole densities are on the order of $10^{16}\,\text{cm}^{-3}$, with solar cell efficiencies near or above 20%. This suggests that a low activation ratio isn't a critical issue for achieving excellent solar cell performance. As illustrated in Fig.4.16, Fig.4.18, Fig.4.17, and Fig.4.19, an activation ratio below 10% can still yield good carrier density, which is either close to or at the optimal point. A high activation ratio isn't always advantageous, as it comes at the expense of Group V solubility in CdTe and CdSeTe, potentially reducing overall carrier density.

However, if a higher carrier density or activation ratio is needed, we propose a possible approach involving adjusting the Cd chemical potential towards the optimal point. This could be achieved experimentally by varying the Cd content during the growth phase.

4.3.3.1 Grain Boundary Segregation

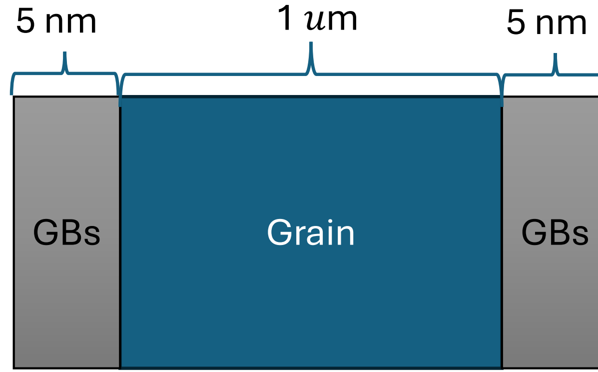


Figure 4.20: Grain boundaries and grain structure of polycrytalline CdSeTe. The width of GBs is 5nm and width of grain is $5\mu m$. Se ratio is chosen to be 25% and μ_{Cd} is -0.8 eV.

Today, most state-of-the-art thin-film photovoltaics, including CdSeTe, $CuInGaSe_2$, and perovskite materials, are polycrystalline and contain grain boundaries (GBs) [57] There is evidence suggesting that the two-dimensional (2D) compound $CdCl_2$ forms at interfaces and GBs in CdTe, reducing surface energy and eliminating surface trap states [125, 180]. Here, we explore whether grain boundary segregation affects the dopability of Group V species. Given the existence of $CdCl_2$ in GBs, GBs or interfaces can act as sinks for Cd interstitials due to the large binding energy[180] between interface structures. This can lead to Cd accumulation in these regions during cooling and annealing, which may subsequently influence the number of complexes and carriers. To investigate this phenomenon, we constructed a 2D model (Fig.4.20) to simulate the effect of grain boundary segregation through process simulation.

If the total dopant fluxes at the interfaces between the grain boundary and the neighboring

layer are balanced, the rate of segregation G is given by[115, 193]

$$G = k_{Transfer}(C_{grain} - \frac{C_{gb}}{f_{gb}k_{Segregation}}), \tag{4.11}$$

where C_{gb} is the defect concentration in grain boundary, C_{grain} is the concentration of defect in grain interior, $k_{Transfer}$ is the transfer rate, and $k_{Segregation}$ is the segregation rate of defect in the grain boundary. f_{gb} is the fraction of unfilled boundary sites:

$$f_{gb} = 1 - \frac{C_{gb}}{GBMaxDensity}, \tag{4.12}$$

where $GBMaxDensity$ represents the density of available sites on grain boundaries. We use diffusivity of Cd_i in Table 4.9 as $k_{Transfer}$, and

$$GBMaxDensity = 5 \times 10^{14}\, cm^{-2} \tag{4.13}$$

$$k_{Segregation} = \frac{GBMaxDensity}{GMaxConc} exp(\Delta E_{seg}/k_b T), \tag{4.14}$$

where E_{seg} is the intersitial-GB segregation energy, $GMaxConc$ represents the density of lattice sites inside grain. Here, we explore serveral different scenarios where E_{seg} is set from 0.5 eV to 2.5 eV. We have also examined varied storage time. However, there is minimal change after extended storage periods. The reason is similar to what we observed in Fig. 4.15. Generally, the transfer rate in room temperature is too slow to make a significant change.

From the results in Table 4.11, it can be seen when the GBs act as a sink of Cd_i, there is an accumulation of Cd_i^{GBs}. The amount of Cd_i^{GBs} depends on the E_{seg}. At growth temperature 873K, only when E_{seg} is larger than binding energy $E_{binding}$ of $(As_{Te}+Cd_i)^+$ in Table 4.6, the amount of Cd_i^{GBs} can exceed Cd_i^{Grain}. After cooling and annealing, we can observe that the concentration of Cd_i^{Grain} is significantly dropped. It could be either due to the formation of defect complexes or Cd_i^{Grain} is being annihilated via vacancy-interstitial pair annihilation (Table 4.10). However, even if the Cd_i^{Grain} is significantly lower than Cd_i^{GBs} at 300K, only a small portion of Cd_i^{GBs} is reversely

Table 4.11: Selected defect densities and carrier densities in polycrystalline $CdSe_{0.25}Te_{0.75}$ with varying E_{seg}. Baseline model has no GBs. Initial temperature is 873K. Cooling rate is 20°C/min. Storage time at 300K is 2 weeks. Units of defect/hole density are cm^{-3}.

Condition	$E_{seg}\,(eV)$	T (K)	hole density	Cd_i^{grain}	Cd_i^{GBs}	$(As_{Te}+Cd_i)^+$
Baseline, no GBs	-	300	2.60×10^{15}	1.79×10^{-1}	-	4.56×10^{16}
$CdSe_{0.25}Te_{0.75}$ with GBs	0.5	873	1.08×10^{16}	6.62×10^{13}	5.10×10^{8}	4.56×10^{16}
		300	2.60×10^{15}	1.23×10^{3}	4.61×10^{-1}	4.56×10^{16}
$CdSe_{0.25}Te_{0.75}$ with GBs	1.0	873	1.08×10^{16}	6.62×10^{13}	3.93×10^{11}	4.56×10^{16}
		300	2.60×10^{15}	1.79×10^{-1}	2.84×10^{10}	4.56×10^{16}
$CdSe_{0.25}Te_{0.75}$ with GBs	2.0	873	1.50×10^{16}	6.09×10^{13}	2.14×10^{17}	4.35×10^{16}
		300	3.47×10^{15}	1.71×10^{-1}	2.33×10^{17}	4.33×10^{16}

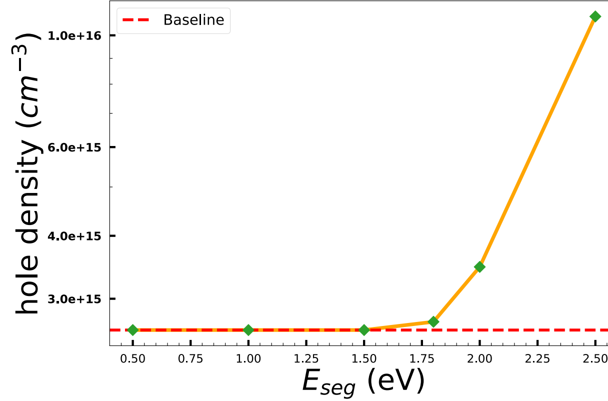


Figure 4.21: Hole density vs. E_{seg}. Red dotted line indicates baseline model without GBs.

transported into grain region when E_{seg} is smaller than $E_{binding}$. This is due to the slow transport rate at 300K. On the other hand, when E_{seg} is greater than $E_{binding}$, we see a slight increase of Cd_i^{GBs} at 300K due to the dissociation of $(As_{Te}+Cd_i)^+$. This will raise the carrier density. When we further increase E_{seg} to be 2.5 eV, we can observe from Fig. 4.21 that the carrier density is increased by around 3 times.

From the results shown in Table 4.11, it is evident that when grain boundaries (GBs) act as a sink for Cd_i, there is an accumulation of Cd_i^{GBs}. The extent of this accumulation depends on the segregation energy E_{seg}. At a growth temperature of 873K, the concentration of Cd_i^{GBs} surpasses that in the grain only when E_{seg} exceeds the binding energy $E_{binding}$ of $(As_{Te}+Cd_i)^+$, as

listed in Table 4.6. Following cooling and annealing, a significant reduction in Cd_i^{Grain} is observed. This reduction may result from the formation of defect complexes or the annihilation of Cd_i^{Grain} through vacancy-interstitial pair reactions, as detailed in Table 4.10. At 300K, a portion of Cd_i^{GBs} is transported back into the grain region when E_{seg} is less than $E_{binding}$. In this scenario, GBs serve as a reservoir for Cd_i, but the impact on carrier density is minimal because the amount of Cd_i^{GBs} accumulated at GBs is not significant. Conversely, if E_{seg} is greater than $E_{binding}$, there is a substantial accumulation of Cd_i^{GBs} at the GBs and a slight increase at 300K due to the dissociation of $(As_{Te}+Cd_i)^+$, which enhances the carrier density. Increasing E_{seg} to 2.5 eV, as shown in Fig. 4.21, leads to a tripling of carrier density compared to baseline model.

Based on these findings, we conclude that GBs can significantly influence carrier densities by accumulating Cd_i within the GBs and reducing the presence of compensating defects. Nevertheless, further first principles and experimental research are required to verify whether GBs act as a sink in CdSeTe and the strength of this segregation. Nonetheless, this study presents a novel approach to boosting carrier density by tailoring GBs to serve as an effective sink for compensating defects or as a component of such defects.

4.4 Summary

In this study, we have undertaken a comprehensive investigation of intrinsic defects, copper, and group V dopants in CdSeTe alloys using DFT calculations. Our tailored charge correction approach demonstrates its advantage over traditional FNV correction methods, particularly when applying to defects with delocalized charge or multiple charge centers. Systematic evaluations reveal that our refined methodology not only improves convergence characteristics but also delivers results that are more consistent with experimental data. Specifically, when analyzing defects such as As_{Te} and P_{Te}, the transition levels calculated by our method show close agreement with experimental findings, underscoring the enhanced accuracy and reliability of our proposed approach.

Within the CdSeTe framework, the formation energy of point defects is noticeably influenced by the local spatial arrangement of Se and Te atoms. This phenomenon is notably evident in instances such as Te_{Cd} (+1, 0) and As_{Te} (+1), where the interplay of Se and Te positions exerts a

pronounced effect. Our investigation extends beyond the mere identification of point defects to a deeper exploration of their intricate relationships with the local atomic environment.

Our study provides insights into the effects of arsenic and phosphorus defects on the electrical properties of CdSeTe alloy. We found that AX formation is more favorable in the alloy due to the interaction of selenium and group V species, which could potentially deteriorate the dopability of As/P in CdSeTe alloy. Our findings could help guide the development of new doping strategies for the fabrication of more efficient solar cells based on CdSeTe alloy.

Overall, we believe that group V dopants hold more promise than Cu in CdSeTe alloys, due to the doping efficiency challenges associated with Cu in CdSeTe and the potential introduction of more deep-level traps in a Te-rich environment. In contrast, group V dopants generally favor a Cd-rich environment, which can enhance carrier lifetime in solar cells.

Furthermore, by employing the continuum models, we investigated the dominant defect types and their effects on doping efficiency and carrier densities. Our analysis reveals that the AX center defect is not the predominant compensating defect in As- or P-doped CdTe. Instead, the complexes $(As_{Te}+Cd_i)^+$ and $(P_{Te}+Cd_i)^+$ tend to dominate, except in low-Cd chemical potential conditions. However, the AX defect becomes significant at particular chemical potentials, particularly in the CdSeTe alloy, where higher initial temperatures amplify its prominence.

This research identifies an optimal range for chemical potential, which balances minimizing compensating defects and maximizing dopant solubility. While Cd-rich conditions generally favor Group V doping, extremely Cd-rich or Te-rich conditions result in higher concentrations of compensating and deep-level defects, respectively. Higher initial temperatures significantly increase carrier densities by fostering the formation of Group V acceptors and reducing compensating complexes under optimal chemical potential conditions.

Our simulations are consistent with experimental observations that show low activation ratios do not severely impact carrier densities or solar cell performance. The improved doping efficacy can be achieved through precise chemical potential adjustment. Grain boundaries functioning as sinks for cadmium interstitials can potentially reduce compensating defects and enhancing carrier

densities. However, additional experimental and theoretical research is required to quantify these effects.

Overall, this comprehensive study elucidates the dynamics of As and P doping in CdTe and CdSeTe, revealing practical doping strategies and providing critical insights into the optimization of solar cell materials. The findings contribute significantly to understanding the challenges of Group V doping and pave the way for enhancing thin-film solar cell performance.

Chapter 5

ACCELERATED DEFECT PROPERTY PREDICTIONS AND DEVICE DESIGN OF CUBIC SEMICONDUCTOR CRYSTALS

Compositional manipulation of semiconductors is one of the primary methods utilized to obtain optimal properties [59, 144, 141, 85, 172, 117]. Apart from alloying, the primary means for compositional control of semiconductor properties is the introduction of dopants or impurities, i.e. guest atoms at a cation, anion, or interstitial sites. Such impurities, even in a very dilute concentration, can potentially cause major changes in the electronic structure and physical properties of the material [162, 93, 229, 153]. The complete understanding of a semiconductor's optoelectronic behavior requires estimating the formation energies of point defects, whether accidental or intentionally introduced [117, 119, 118].

While $\sim$ 90% of solar cells still rely on crystalline Si as the absorber, related group IV semiconductors such as SiC, II-VI semiconductors such as CdTe, III-V semiconductors such as GaAs, and various derivative compounds are all viable as photovoltaic (PV) materials and are currently in use in single terminal as well as tandem solar cells [144, 184, 107, 205, 137, 196]. Many of these compounds have also been used in transistors, photodiodes, lasers, and qubits or quantum sensors. The chemical space of binary group IV, III-V, and II-VI semiconductors contains compounds that exist in the cubic zinc blende (ZB) or wurtzite crystal structures and show systematic trends in lattice constants, electronic band gaps, optical absorption coefficients, and defect properties [168]. Alloying in these spaces has frequently been used for tuning properties and performance, with some prominent examples including the use of CdSeTe in solar cells [17, 41] and AlGaAs in LEDs [10, 204].

Although the structure and optoelectronic properties of binary, ternary and even quaternary compounds in the group IV, III-V, and II-VI semiconductor space have been widely studied both computationally and experimentally [144, 141, 119, 168], a comprehensive understanding of the

formation likelihood and electronic levels of point defects and impurities is missing. A look at functional atomic defects in semiconductors reveals that the energy levels created inside the band gap can (a) reduce PV efficiency via nonradiative recombination of charge carriers, (b) enable sub-gap absorption or emission if the levels are partially filled or if they have low photoionization energies, and (c) enable quantum computing, quantum sensing, and quantum communication via their nuclear or electronic spins. A universal prediction framework for impurity behavior in known and novel semiconductor spaces is thus paramount [119]. Given such a framework for group IV, III-V, and II-VI semiconductors, it would be possible to perform high-throughput screening of impurity atoms from across the periodic table in terms of their energetics relative to dominant native defects (such as vacancies and interstitials), nature of equilibrium conductivity, and location of energy levels with respect to band edges.

Predictive machine learning (ML) models, trained from existing or freshly generated data, act as surrogates for DFT calculations by providing statistical estimates of the desired properties[119]. The burgeoning field of materials informatics has led to many successes, with some of the most notable contributions resulting from the combination of first principle computations and machine learning. ML applied on DFT data has seen the development of predictive and design tools [82, 207, 3], the discovery of novel materials for batteries, capacitors, solar cells, and thermoelectrics [36, 116, 81, 47, 72], and the efficient exploration of extremely large chemical spaces [129, 201]. Indeed, ML has been instrumental in accelerating the prediction of properties related to point defects and dopants in materials. This includes predicting vacancy formation and substitutional energies of oxides using regression algorithms applied on DFT data [203, 34, 35, 181], machine learning formation energies, transition levels, and migration energies of point defects in known semiconductors and alloys [199, 49], and predicting dopability of semiconductors [131]. An extension of these studies in terms of semiconductor and impurity chemical spaces as well as ML techniques can pave the path towards a universal framework for impurity prediction and design.

5.1 *Machine Learning Framework for Qualitative Screening*

In this work, we consider atomic impurities from across the periodic table, in a chemical space of binary group IV, III-V, and II-VI semiconductors in the ZB structure, and use the DFT+ML methodology to predict their complete charge, chemical potential, and Fermi level dependent formation energies. We perform high-throughput DFT computations on impurity atoms simulated at the cation, anion, and different interstitial sites in several selected compounds in the group IV (e.g. Si, SiC, GeC, etc.), III-V (e.g. BSb, GaAs, InP, etc.), and II-VI (e.g. ZnSe, CdS, etc.) chemical space, and use descriptors encoding information about the semiconductor, the impurity atom, and the defect site coordination environment as input to train ML models that predict the neutral state formation energy and 6 types of charge transition levels for any possible impurity. We used sure independence screening and sparsifying operator (SISSO) for feature selection and K-nearest neighbors (KNN) approach for outlier detection, followed by regression techniques such as random forest, Gaussian process, and neural network to yield the predictive models.

5.1.1 Semiconductor and Impurity Chemical Space

The chemical space considered in this work has been pictured in Fig 5.1 in terms of the semiconductor compounds (b), possible defect sites (c), and impurity atoms (d). We include AB semiconductors (with A broadly defined as the cation and B the anion) belonging to groups II-VI, III-V and IV-IV, leading to 8 II-VI compounds (CdO, CdS, CdSe, CdTe, ZnO, ZnS, ZnSe, and ZnTe), 16 III-V compounds (BN, BP, BAs, BSb, AlN, AlP, AlAs, AlSb, GaN, GaP, GaAs, GaSb, InN, InP, InAs, and InSb), and 10 group IV compounds (C, Si, Ge, Sn, SiC, GeC, SnC, SiGe, SiSn, and GeSn). The resulting 34 compounds are modeled in the cubic ZB structure, with A atoms occupying an FCC lattice and B atoms occupying the tetrahedral sites.

In any AB compound in the ZB structure, defects or impurities could be found at the A-site, B-site, or several possible symmetrically-inequivalent interstitial sites. Fig 5.1 also shows the defect sites considered in this work, namely the A and B sites and 3 types of interstitial sites: the A-site interstitial (with 4 neighboring A atoms), the B-site interstitial (with 4 neighboring B atoms), and

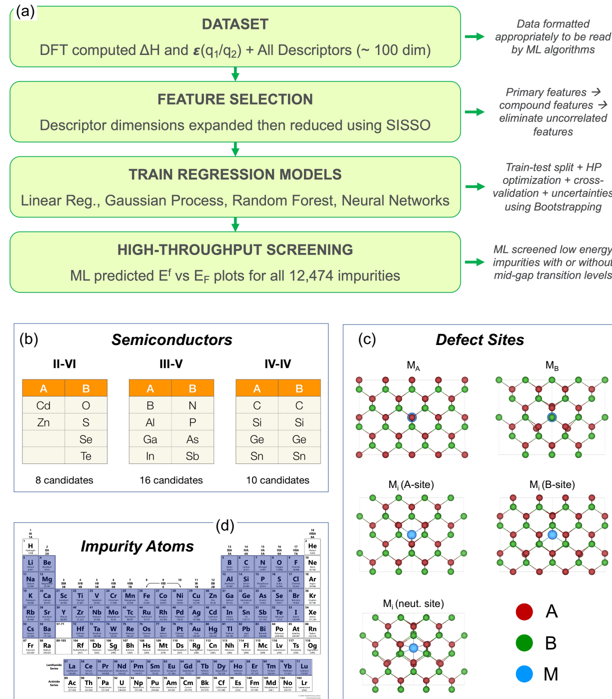


Figure 5.1: (a) The DFT-ML workflow followed in this work, and the semiconductor-impurity chemical space in terms of (b) the cation (A) and anion (B) choices for group IV, II-VI, and III-V compounds, (c) types of defect sites, and (d) impurity atoms selected from across the periodic table.

neutral site interstitial (with 3 neighboring A and B atoms each). The 5 defect sites are considered in the 30 binary compounds while in the remaining 4 elemental systems (C, Si, Ge, Sn), 3 defect sites are considered (A-site, A-site interstitial and neutral site interstitial). In terms of impurity atoms, we consider nearly all elements from periods II to VI as well as all lanthanides, leading to a total of 77 species, as pictured in Fig 5.1. The total number of possible impurities in this chemical space can thus be estimated as: $77 \times 5 \times 30 + 77 \times 3 \times 4 = 12474$. Out of these 12474 data points, about 10% are considered for DFT computations to determine their neutral state formation energies and charge transition levels; ML models trained on these data based on the properties of the semiconductor compound, defect site coordination, and impurity atom lead to generalized predictions applicable to all the data points.

5.1.2 Descriptors

Aside from generating the computational data, the need for *domain expertise* is most evident in creating appropriate descriptors or sets of features that can uniquely represent every point in the dataset. In the semiconductor+impurity chemical space used in this work, we can uniquely identify every data point using the identity of the semiconductor, the identity of the impurity atom, and the defect site it occupies. Thus, we define descriptors for any impurity M at any site S in any compound AB by combining the following three levels of information:

1. AB_{prop}: Available computed or experimental properties of the semiconductor AB, namely the formation energy, lattice constant, band gap, and the electronic and ionic dielectric constants; this leads to 5 dimensions.

2. $Elem_{prop}$: Tabulated elemental properties of the impurity M as well as species A and B, such as ionic radius, ionization energy, electronegativity, etc.; this leads to 81 dimensions.

3. CM_{prop}: Quantifying the chemical coordination environment around the defect site S in terms of A and B neighbors, using the Coulomb Matrix definition [46]; this leads to 8 dimensions.

The complete list of descriptors can be found on the x-axis of Fig 5.2, which shows the Pearson coefficient of linear correlation between the properties of interest, ΔH and $\epsilon(q_1/q_2)$, and each of the descriptors.

5.1.2.1 Feature Selection

The primary feature set of 94 dimensions are all assumed to be relevant to describe the targeted predictors, that is, the impurity transition levels and formation energies. To better explore the nonlinear relationships that may exist between these descriptors dimensions and the properties, we used the sure independence screening and sparsifying operator (SISSO) [150] method to perform feature engineering. First, a set of operators, namely '+, -, *, /, exp, log, ˆ(-1), ˆ2, ˆ3, sqrt, cbrt', are implemented recursively for feature space expansion. The total feature size goes from $\sim 10^2$

to $\sim 10^5$ after two iterations. Next, sure independence screening [44] is used to screen all features from 0D feature space (no iteration) to 1D feature space (one iteration) to 2D feature space (two iterations) using a linear correlation metric, leaving behind only highly correlated features. Finally, a sparsifying operation [45] is applied to filter down the feature space to 80–150 for each output.

The detection of outliers in a dataset helps identify candidates with unusual properties which may either be erroneous or lead to lower accuracy when used to train ML models. K-nearest neighbors (KNN) is a method commonly applied to identify the outliers in a dataset based on a feature space [87]. KNN assigns classes to data points based on the most common assignment of its k nearest neighbors; any point that is surrounded by points belonging to a different class is denoted an outlier. Here, the DFT computed transition levels and formation energy values were used as a combined input to a KNN framework. Using KNN, we selected 10% of the data with the highest outlier scores to be removed.

Figure 5.2: Absolute values of coefficient of linear correlation between every unique descriptor and every property; darker boxes imply high correlation whereas white boxes mean there is zero correlation.

5.1.3 Training Regression Models

In the following sections, we will discuss the optimization and performance of various regression models trained on the computational data, starting with linear regression and moving on to different nonlinear regression techniques, namely random forests, Gaussian processes, kernel ridge, and neural network regression. Common to every method used in this work is the way training-test split, cross-validation, hyperparameter optimization, and error evaluation was performed. A separate model was trained for each of the 8 outputs, namely the 6 types of transition levels and 2 formation energies. Five-fold cross validation was implemented for each model because of a strong dependence of the prediction ability on the exact points chosen for training. Cross-validation helps reduce the reported bias and variance, and is important for avoiding *overfitting*. Various important hyperparameters were optimized for each regression technique; for instance, for neural networks, they include the number of hidden layers, the numbers of nodes in each layer, the dropout rate, etc. All regression models were trained using functions in the Python machine learning library scikit-learn [200].

The metric for evaluating model performance was chosen to be the prediction RMSE. Each of the 5 folds was treated as a validation set over multiple training cycles, and the prediction RMSE for each fold was averaged over the number of folds. This leads to an effective 80-20 training-test split in the dataset, and an effective "test prediction error" is obtained for every data point, providing an unbiased prediction that reveals the true predictive power of the trained model. The optimal set of hyperparameters is chosen such that the cross-validation error is minimized; we ultimately report training and test errors for every model, but optimization is based on the validation error, such that the actual test set in each iteration remains unseen by the model during the training process. Further, the standard deviation in predictions over the multiple training cycles is defined as the uncertainty for each predicted point, providing an error bar that accompanies every prediction. Results are presented as parity plots of ML predictions vs. DFT-calculated properties, with reported RMSE values in eV, and plots between uncertainties and errors in every data point; the predictions are also visualized in terms of semiconductor type. The test prediction RMSE

values listed in Tables 5.1 and 5.2, divided in terms of property, ML technique used, and type of data point.

5.1.3.1 Linear Regression

Fig 5.2 shows the Pearson coefficient of linear correlation between various (primary) descriptor dimensions and the properties of interest. We see that many of the features are 50–70% correlated with the properties, showing a certain degree of linear relationship. We further investigated the correlation coefficients for the ten best SISSO-based compound features (essentially, complex functions of combinations of original features). The highest correlated features reveal the specific descriptors or combinations thereof that could best predict the defect formation energy and charge transition levels. We notice that atomic radii and ionization energy differences (between defect atom M and A/B atoms) are most important for $\epsilon(+3/+2)$ while valence and electronegativity differences dominate for $\epsilon(+2/+1)$; coefficients for both remain small, ~ 0.35. Highest correlations are between 0.5 and 0.55 for $\epsilon(+1/0)$ and $\epsilon(0/\text{-}1)$, with ionization energy and atomic radii differences dominating in both. $\epsilon(\text{-}1/\text{-}2)$ and $\epsilon(\text{-}2/\text{-}3)$ show even higher correlations of between 0.7 and 0.75, with descriptors such as Mendeleev number, covalent radii and ICSD volume of M/A/B atoms being most important. Finally, we find correlation maximums of ~ 0.45 for ΔH (A-rich) and ΔH (B-rich), determined primarily by the semiconductor lattice constant, ionization potential, boiling point, heat of fusion/vaporization, and specific heat capacity. Overall, these correlations reveal that the relative electronegativities, ionization energies and radii of elements are important in placing defect energy levels relative to band edges, while the size of the lattice and heat of fusion may determine how likely it is for the defect atom to exist at site in consideration.

To explore linear relationships further, we chose multiple linear regression (MLR) as the method to train the first predictive models. And then LASSO, ridge regression, and elastic net regression are being applied to regularize linear regression. We see from the plots and from Tables 5.1 and 5.2 that there is a marginal improvement in prediction going from MLR to LASSO, ridge, or elastic net regression. The presented results are using the SISSO features as they provide better predictions

than using the primary feature set, which is presumably due to the non-linear nature of the input-output relationship. We further find that there is a strong dependence of the prediction error on semiconductor type and impurity site. We observe these effects in non-linear models as well, which will be discussed in subsequent sections. Such prediction error differences can be attributed to the imbalanced distribution of training data; the DFT datasets are biased towards II-VI semiconductors and interstitial site impurities. In general, we get better performance on II-VI or interstitial data points since the models we trained work better on majority groups.

Random Forest Regression

The improvement in linear regression model performances upon going from using the primary features to the SISSO-based features shows the importance of interpreting non-linear relationships between the features and properties. However, non-linearity is still limited by the set of operators used in the SISSO method; to further explore this effect, we adopted a popular non-linear regression algorithm known as random forest regression (RFR). RFR is an ensemble measurement method that fits a designated number of classifying decision trees such that each tree is fit on a different randomized sub-sample of the data set, chosen through bootstrapping. During the construction of any tree, the best split for each node is found based on some number of input features. Averaging over all the trees in the forest can be performed in several ways, and in this work, the model combines the results of the trees by averaging their probabilistic prediction, which improves prediction accuracy and can help control overfitting [19].

Hyperparameter tuning focuses on the five most important features in the RFR model, namely, the number of trees in the forest, the maximum depth of each tree, the number of features to consider when looking for the best node split, the minimum number of samples required to split an internal node, and the minimum number of samples required to be at a leaf node. For each of the eight outputs, Bayesian optimization was performed [14] using a function set to minimize both the test RMSE and the difference between the training and test RMSE to balance the bias-variance trade off in the model. Parity plots for the optimized models for all 8 properties are shown in Fig

Table 5.1: ML test set prediction RMSE values for transition levels. Lowest prediction errors are highlighted in bold.

Property	ML Method	II-VI Error (eV)	III-V Error (eV)	IV-IV Error (eV)	Total Error (eV)
ϵ(+3/+2)	MLR	0.35	0.37	0.34	0.35
ϵ(+3/+2)	Ridge	0.35	0.35	0.32	0.34
ϵ(+3/+2)	LASSO	0.36	0.36	0.32	0.35
ϵ(+3/+2)	Elastic Net	0.35	0.35	0.32	0.34
ϵ(+3/+2)	RFR	0.36	**0.31**	0.35	0.34
ϵ(+3/+2)	KRR	0.33	0.37	0.31	0.33
ϵ(+3/+2)	GPR	0.32	0.36	0.32	0.33
ϵ(+3/+2)	**NN**	**0.29**	0.36	**0.29**	**0.31**
ϵ(+2/+1)	MLR	0.42	0.46	0.46	0.44
ϵ(+2/+1)	Ridge	0.42	0.43	0.45	0.43
ϵ(+2/+1)	LASSO	0.43	0.44	0.45	0.44
ϵ(+2/+1)	Elastic Net	0.42	0.43	0.45	0.43
ϵ(+2/+1)	RFR	0.39	0.36	0.40	0.38
ϵ(+2/+1)	KRR	0.33	0.38	0.40	0.36
ϵ(+2/+1)	GPR	0.32	0.38	0.41	0.36
ϵ(+2/+1)	**NN**	**0.29**	**0.35**	**0.38**	**0.33**
ϵ(+1/0)	MLR	0.40	0.39	0.43	0.40
ϵ(+1/0)	Ridge	0.40	0.38	0.42	0.40
ϵ(+1/0)	LASSO	0.41	0.39	0.43	0.41
ϵ(+1/0)	Elastic Net	0.40	0.38	0.42	0.40
ϵ(+1/0)	RFR	0.38	0.36	0.39	0.38
ϵ(+1/0)	KRR	0.31	0.34	0.38	0.33
ϵ(+1/0)	**GPR**	**0.29**	0.32	0.38	**0.32**
ϵ(+1/0)	NN	0.29	**0.31**	**0.37**	0.32
ϵ(0/-1)	MLR	0.37	0.42	0.34	0.38
ϵ(0/-1)	Ridge	0.37	0.40	0.34	0.37
ϵ(0/-1)	LASSO	0.37	0.40	0.34	0.37
ϵ(0/-1)	Elastic Net	0.37	0.40	0.34	0.37
ϵ(0/-1)	RFR	0.37	0.33	0.35	0.35
ϵ(0/-1)	KRR	0.32	0.36	0.32	0.33
ϵ(0/-1)	GPR	0.31	0.34	0.32	0.32
ϵ(0/-1)	**NN**	**0.28**	**0.33**	**0.31**	**0.30**
ϵ(-1/-2)	MLR	0.33	0.38	0.30	0.33
ϵ(-1/-2)	Ridge	0.32	0.37	0.29	0.32
ϵ(-1/-2)	LASSO	0.32	0.37	0.29	0.33
ϵ(-1/-2)	Elastic Net	0.32	0.37	0.29	0.33
ϵ(-1/-2)	RFR	0.34	0.35	0.27	0.33
ϵ(-1/-2)	KRR	0.29	0.32	**0.27**	0.29
ϵ(-1/-2)	GPR	0.29	0.31	0.28	0.29
ϵ(-1/-2)	**NN**	**0.26**	**0.29**	0.28	**0.27**
ϵ(-2/-3)	MLR	0.27	0.26	0.22	0.26
ϵ(-2/-3)	Ridge	0.27	0.26	0.22	0.25
ϵ(-2/-3)	LASSO	0.27	0.26	0.22	0.25
ϵ(-2/-3)	Elastic Net	0.27	0.26	0.22	0.25
ϵ(-2/-3)	RFR	**0.24**	0.28	0.27	0.25
ϵ(-2/-3)	KRR	0.26	0.24	0.21	0.24
ϵ(-2/-3)	GPR	0.25	0.24	0.21	0.24
ϵ(-2/-3)	**NN**	0.25	**0.22**	**0.22**	**0.24**

Table 5.2: ML test set prediction RMSE values for formation energies. Lowest prediction errors are highlighted in bold.

Property	ML Method	II-VI Error (eV)	III-V Error (eV)	IV-IV Error (eV)	Total Error (eV)
ΔH (A-rich)	MLR	0.85	1.57	1.81	1.16
ΔH (A-rich)	Ridge	0.85	1.54	1.78	1.14
ΔH (A-rich)	LASSO	0.88	1.55	1.79	1.16
ΔH (A-rich)	Elastic Net	0.85	1.53	1.78	1.14
ΔH (A-rich)	RFR	1.05	1.03	1.20	1.07
ΔH (A-rich)	**KRR**	0.62	1.35	**1.32**	**0.89**
ΔH (A-rich)	GPR	**0.59**	1.33	1.71	0.96
ΔH (A-rich)	NN	0.62	**1.30**	1.40	0.89
ΔH (B-rich)	MLR	1.04	1.82	1.81	1.31
ΔH (B-rich)	Ridge	1.04	1.73	1.77	1.29
ΔH (B-rich)	LASSO	1.08	1.74	1.80	1.32
ΔH (B-rich)	Elastic Net	1.05	1.72	1.77	1.28
ΔH (B-rich)	RFR	1.09	**1.25**	1.52	1.18
ΔH (B-rich)	KRR	**0.77**	1.52	1.45	1.03
ΔH (B-rich)	GPR	0.82	1.52	1.70	1.11
ΔH (B-rich)	**NN**	0.81	1.34	**1.44**	**1.01**

5.3(a). Looking at the error values listed in Tables 5.1 and 5.2, there is a general improvement in all the transition level prediction RMSEs from between 0.3 and 0.45 eV for the linear models to between 0.25 to 0.38 eV for RFR, and the formation energy RMSEs drop from 1.2 eV or higher to between 1.07 and 1.16 eV.

As observed in the linear regression results, we found that RFR was able to predict formation energies of II-VI semiconductor impurities more accurately than III-V or IV-IV, owing to the larger portion of II-VI semiconductor points in the training data set. Interestingly, the transition levels showed much less of a dependence on semiconductor type; the difference could be due to the larger range of values in the formation energy data versus the transitional levels. We found that the points the model predicted most inaccurately for formation energies are relative outliers as predicted by KNN, and of those points, III-V and IV-IV semiconductor types make up a larger portion than in the data set as a whole. When analyzing the prediction results by site of impurity defect, it was once again seen that interstitials are predicted slightly better than substitutionals, once again owing to the predominance of the former in the dataset.

5.1.3.2 Kernel Ridge Regression

Kernel ridge regression (KRR) is a regression technique which uses the kernel trick to solve a non-linear problem in a linear fashion. The original low-dimensional features are used as input and mapped to a high-dimensional kernel space in which they can be linearly interpreted. In this work, we use different possible choices for the kernel function, namely polynomial, radial basis function (RBF), and Laplacian. For hyperparameter optimization, we applied the grid search method to search a dense space for the best combination of kernel choice and different parameters in the kernel, separately for each output.

The prediction performances for the eight outputs arelisted in Tables 5.1 and 5.2. KRR shows a marked improvement in formation energy prediction and slight improvements in transition level predictions compared to RFR. The improvement is heavily owed to significant lowering of errors for impurities in the II-VI compounds. We find the KRR RMSE for ΔH (A-rich) to be 0.89 eV and for ΔH (B-rich) to be 1.03 eV, while the RMSE values for the 6 transition levels range between 0.25 and 0.35 eV.

5.1.3.3 Gaussian Process Regression

Subsequently, GPR is used for model training. It can be seen from the GPR parity plots in Fig 5.3(b) and from Tables 5.1 and 5.2 that the prediction RMSE values are very similar to those obtained with KRR. The formation energy errors are between 0.96 and 1.1 eV while the transition level errors range from 0.24 to 0.36 eV. This can be explained by the flexibility of the GPR models which likely causes overfitting when dealing with a small dataset and high dimensional features [132].

5.1.3.4 Neural Network Regression

Finally, we used neural networks (NN) to train regression models and compared the results with nonlinear regression models from RFR, KRR, and GPR. The Keras functional API model was used to build a deep feed forward NN to machine learn a multi-output regression [15]. A sequential model trained to predict the six transition levels and two formation energies was found to be time

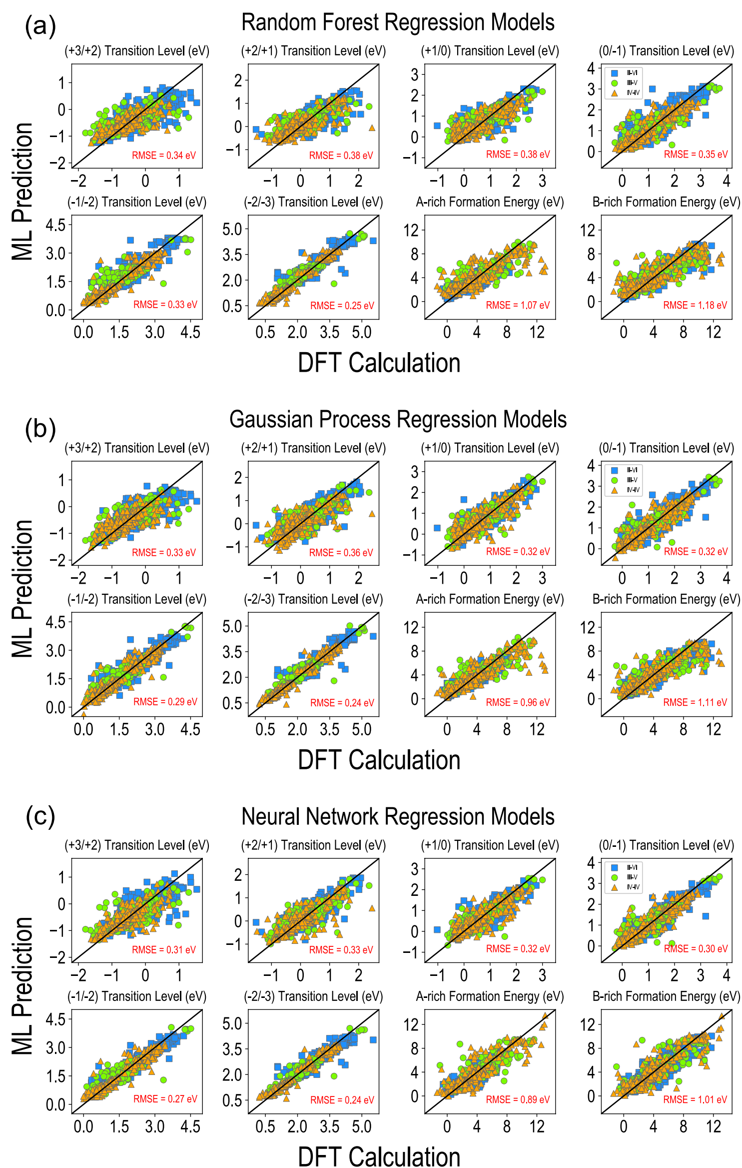


Figure 5.3: Parity plots for best regression models: (a) Random Forest, (b) Gaussian Process, and (c) Neural Network regression.

consuming and lacked the ability to predict multiple outputs at once effectively. Further, a grid search employed to explore the number of hidden layers, number of neurons, learning rate, epochs, batch size, optimizers, and activation functions was found to be inefficient. Separate models were thus trained for each property using the SISSO generated descriptors, and scikit-optimize (skopt) was utilized for Bayesian hyperparameter optimization. To overcome an overfitting problem arising from minimizing only the test RMSE, the optimization function was revised to also include the difference in train and test RMSE.

Each NN architecture contains two to three dense neuron layers, through which the input is concatenated before returning the output through the final layer. The number of neurons in each dense layer varies with the input dimensions for each specific property or output. Kernel and activity regularizers were also integrated in each dense layer to prevent overfitting. The "relu" activation function was ultimately used for each dense layer, beating out sigmoid, softmax, softplus, tanh, and selu functions [163], while the Adam optimizer was selected over SGD, RMSEprop, Adadelta, and Adagrad [31]. NN model training involved 10-repeated 5-fold cross-validation, where the mean and standard deviation of prediction of every data point were used as the predicted value and uncertainty value, respectively. Parity plots for the best models thus obtained are presented in Fig 5.3(c)

It can be seen from the parity plots and Tables 5.1 and 5.2 that NN predictions for both transition levels and formation energies are similar to KRR and GPR. Transition level RMSE values are seen to range from 0.24 to 0.33 eV while the formation energy RMSEs are between 0.9 and 1 eV. A possible disadvantage of the NN models comes from the larger uncertainty values seen in general compared to other methods, while the absolute error values are similar to other methods. This is an effect of the stronger dependence of NN model prediction on the hyperparameter choice, leading to a larger standard deviation in prediction; this is expected to affect NN predictions over the entire chemical space. We further note that standard deviations over 10-folds may not be sufficient to converge the uncertainties, but we use an ensemble of 10 predictions here to save on training time and keep estimates consistent across different ML models.

5.1.4 High-Throughput Screening of Dominating Impurities

The detailed machine learning analysis presented in this work reveals that multiple nonlinear regression techniques can be trained to make predictions of impurity transition levels and formation energies with errors that are within 10% of the range of values across the dataset.

We performed high-throughput prediction of the complete formation energies of the entire dataset of 12,474 impurities, using the best NN, GPR, KRR, and RFR models. It is important to note here that a significant amount of time is saved by replacing full DFT calculations with almost instantaneous ML predictions. On average, any one point defect in a 64 atom supercell simulated in the neutral state requires $\sim$ 500 core hours, while 6 charged state calculations require a further $\sim$ 2000 core hours (running on 8 Intel Broadwell XEON E5-2695 nodes with 36 cores each). For the DFT datasets of $\sim$ 1500 neutral state formation energies and $\sim$ 1000 charge transition levels of 6 types, this translates to $\sim$ 2.75 million core hours. For the entire dataset of 12,474 impurities, $\sim$ 32 million core hours would be required for complete DFT optimization and prediction of all defect properties. On the other hand, every ML model takes a matter of minutes to train and make predictions over the entire chemical space. Thus, based on computations using 1/10th of the total computing time required, we can make reasonable predictions for all the data points.

In Table 5.3, we listed several impurities deemed to be dominating from both DFT and ML (GPR used as example here), along with their stable charge states, the corresponding dominating native defect, the type of shift induced in the equilibrium E_F, and whether mid-gap energy levels are created. For example, it can be seen that Ti at the Al site in AlAs creates a stable +1 charged donor type defect, and along with a -3 charged As vacancy acceptor, makes the conductivity more n-type and creates a transition level in the band gap. Similarly, a Be interstitial defect in Si induces a p-type shift in conductivity. Lists of dominating impurities with or without mid-gap energy levels were thus generated for all compounds.

Table 5.3: Selected dominating impurities identified by both DFT and ML (GPR), at A-rich chemical potential conditions.

Semiconductor	**Impurity**	**Shift in Eqm. E_F**	**Dominating Defects**	**Mid-gap Level?**
CdS	In_{Cd}	n-type	In_{Cd}, q = 1 and V_{Cd}, q = -2	Y
CdS	I_S	n-type	I_S, q = 1 and V_{Cd}, q = -3	Y
CdS	Ti_i	p-type	Ti_i, q = 2 and V_S, q = -1	Y
CdSe	Cu_{Cd}	p-type	Cu_{Cd}, q = -1 and Cd_i, q = 2	Y
CdSe	F_i	p-type	F_i, q = -1 and V_{Se}, q = 2	N
CdSe	Ni_i	p-type	Ni_i, q = -1 and V_{Se}, q = 2	Y
CdTe	Bi_{Cd}	n-type	Bi_{Cd}, q = 1 and V_{Cd}, q = -2	Y
CdTe	As_{Te}	p-type	As_{Te}, q = -1 and V_{Te}, q = 2	Y
CdTe	Na_i	n-type	Na_i, q = 1 and V_{Cd}, q = -2	N
ZnS	Li_i	n-type	Li_i, q = 1 and V_{Zn}, q = -2	N
ZnS	Ti_i	n-type	Ti_i, q = 1 and V_{Zn}, q = -2	Y
ZnSe	Al_{Zn}	n-type	Al_{Zn}, q = 1 and V_{Zn}, q = -2	Y
ZnSe	Br_{Se}	n-type	Br_{Se}, q = 1 and Zn_{Se}, q = -1	Y
ZnTe	Cr_i	n-type	Cr_i, q = 1 and V_{Te}, q = -2	N
ZnTe	Mn_i	n-type	Mn_i, q = 1 and Zn_{Te}, q = -2	Y
AlN	Se_N	p-type	Se_N, q = -1 and V_N, q = 1	Y
AlP	Hf_{Al}	n-type	Hf_{Al}, q = 1 and Al_P, q = -1	Y
AlP	Cr_i	n-type	Cr_i, q = 1 and V_{Al}, q = -2	Y
AlAs	Ti_{Al}	n-type	Ti_{Al}, q = 1 and V_{As}, q = -3	Y
GaN	Tl_{Ga}	p-type	Tl_{Ga}, q = -1 and V_N, q = 1	Y
GaN	P_N	p-type	P_N, q = -2 and V_N, q = 1	Y
GaP	Ni_{Ga}	p-type	Ni_{Ga}, q = -1 and Ga_i, q = 2	Y
GaP	Li_i	n-type	Li_i, q = 1 and Ga_P, q = -2	Y
GaAs	Sc_i	n-type	Sc_i, q = 3 and Ga_{As}, q = -2	Y
GaSb	Al_{Ga}	n-type	Al_{Ga}, q = 1 and V_{Ga}, q = -2	Y
InN	Zr_i	n-type	Zr_i, q = 2 and V_N, q = -1	Y
InP	Cu_i	n-type	Cu_i, q = 1 and In_P, q = -2	Y
InAs	Ca_{In}	p-type	Ca_{In}, q = -1 and In_{As}, q = 2	N
Si	Ti_{Si}	p-type	Ti_{Si}, q = -1 and Si_i, q = 2	Y
Si	Be_i	n-type	Be_i, q = 1 and V_{Si}, q = -3	Y
SiC	V_{Si}	n-type	V_{Si}, q = 1 and V_C, q = -2	Y
SiC	Cr_i	p-type	Cr_i, q = -1 and V_C, q = 1	Y
SnC	As_{Sn}	n-type	As_{Sn}, q = 1 and V_C, q = -2	N
SnC	Cr_{Sn}	p-type	Cr_{Sn}, q = -1 and V_C, q = 2	N

5.2 Deep Learning Models for Quantitative Screening

In our previous work (Section 5.1), we introduced a machine learning (ML) framework for predicting defect formation energies and charge transition levels in a diverse range of cubic phase semiconductors. We evaluated the performance of various ML models and identified the best-performing model, which enabled efficient screening of deep level defects potentially induced by extrinsic dopants. Our screening accuracy exceeded 95%; however, the limitations of the ML framework prevent quantitative studies due to errors that are much larger than the thermal energy at relevant temperatures. This results in carrier density prediction errors exceeding 10^5 in some cases, rendering it unsuitable for an accurate high throughput prediction model. For rigorous quantitative analysis, we must constrain the error to less than one order of magnitude, necessitating a prediction error within around 0.2 eV [50].

In this current work[217], we aim to enhance our model using deep learning techniques. Firstly, we incorporate the chemical environment of defects using the Smooth Overlap of Atomic Positions (SOAP) representation[11] into our existing artificial neural network (ANN) model. We compare this model with a fine-tuned Crystal Graph Convolutional Neural Network (CGCNN) model[221] for defect predictions. Our findings reveal that both models significantly improve the prediction accuracy of formation energies by around 50%. Among them, the chemical environment encoded ANN model outperforms both formation energy and transition level predictions. Specifically, the prediction errors for II-IV semiconductors are all lower than 0.21 eV, making it suitable for quantitative high-throughput screening.

Moreover, we seamlessly integrate our deep learning model with CCT[185] and TCAD technology[192] to investigate potential dopants for enhancing semiconductor device performance. This collaborative approach harnesses the predictive capabilities of our model and TCAD to identify optimal dopants, thereby optimizing device functionality and enhancing overall performance. To validate the efficacy of this design framework, we conduct tests on CdTe solar cell devices and compare to experimental data. Within this framework, we predict four dominant figures of merit: efficiency (η), open circuit voltage (Voc), short-circuit current (Jsc), and fill factor (FF) for solar cells with

various potential dopants in the chemical space. These predictions enable us to determine the most promising dopants that have the potential to significantly impact solar cell performance.

5.2.1 Training Dataset

Our dataset encompassed 1,640 doped semiconductors, incorporating AB-type compounds where 'A' denotes the cation and 'B' signifies the anion, across groups II-VI, III-V, and IV-IV. This categorization yielded 8 II-VI compounds (such as CdO, CdS, CdSe, CdTe, ZnO, ZnS, ZnSe, ZnTe), 16 III-V compounds (including BN, BP, BAs, BSb, AlN, AlP, AlAs, AlSb, GaN, GaP, GaAs, GaSb, InN, InP, InAs, InSb), and 10 group IV compounds (such as C, Si, Ge, Sn, and binary combinations like SiC, GeC, SnC, SiGe, SiSn, GeSn). These compounds, totaling 34, were modeled adopting the cubic Zinc Blende structure, characterized by A atoms forming a face-centered cubic (FCC) lattice and B atoms occupying tetrahedral sites.

Within any given AB compound crystallized in the Zinc Blende structure, potential defects could emerge at the A-site, B-site, or multiple, symmetrically distinct interstitial positions. The present investigation examines five defect locations: the A-site, B-site, A-site tetrahedral interstitial (surrounded by four A atoms), B-site tetrahedral interstitial (surrounded by four B atoms), and the neutral site hexagonal interstitial (equidistant from three A and three B atoms). For the binary compounds, all five defect sites are considered, whereas for the four elemental systems (C, Si, Ge, Sn), three defect sites are analyzed (A-site, A-site interstitial, and neutral site interstitial).

In terms of the doping elements, we take into account a wide spectrum from periods I to VI, along with all lanthanides, culminating in 77 unique species. As a result, the overall count of potential impurities within this chemical structure is 12474. Of these, around 10% have been calculated via DFT to obtain neutral state formation energies under both A-rich and B-rich conditions and six charge transition levels (ϵ(-1/0), ϵ(-2/-1), ϵ(-3/-2), ϵ(+1/0), ϵ(+2/+1) and ϵ(+3/+2)). Following the application of PCA for the removal of outliers, our training dataset includes formation energies for 1,476 compounds and transition levels across 1,076 configurations. The data for this study are obtained from our previous research[120].

Throughout the research process, the physical and chemical descriptors gathered and encoded from our previous work [120] were added to the training data set to be used by the neural networks. Such descriptors were based upon intrinsic properties from a variety of atomic impurities across the periodic table, inherent properties of cubic-structured binary IV-IV, III-V, and II-VI semiconductors, and the Coulomb Matrix which mimics the electrostatic interaction around defect sites.

Furthermore, we collected six inner averages of SOAP power spectrum descriptors [66] with a single radial basis function, degree 1 spherical harmonics and a cutoff distance of 6 Å. These descriptors accurately describe the invariant atomic positions within each semiconductor system by considering the permutations of gathering atomic pairs from the atomic species that compose the defect supercell structure. The input defect structures for ML models are pristine semiconductor supercells with introduced defects, but without any structural relaxation. We experimented with increasing the number of radial basis functions and degrees of spherical harmonics in SOAP, but we observed minimal improvement in neural network models, prompting us to stick with the mentioned configuration.

5.2.2 Chemical Environment Encoded Artificial Neural Network

We constructed our chemical environment encoded artificial neural network (ANN) once all categorical and continuous features from the training dataset were normalized and applied with one-hot encoding. The algorithm was trained on the descriptors mentioned previously to learn and optimize the features to predict defect formation energies and transition levels. Each NN architecture contains two to three dense neuron layers, through which the input is concatenated before returning the output through the final layer.

To determine the regression network's effectiveness in predicting each of the eight outputs (two types of formation energies and six different transition levels), we utilized a five-fold Stratified K-Fold cross-validation sampling on the population of semiconductors in the input dataset based on the type of semiconductor and defect atom type. Such sampling and validation ensure that the

predictive results from the network are limited in bias and variance, reducing overfitting. Hyperparameter optimization was performed via Bayesian optimization to minimize the cross-validation error for the average of each of the five splits; such parameters include learning rate, learning rate decay for the Adam optimizer, the standard deviation for the Gaussian noise distribution, and the kernel regularizer.

From each of the five splits from the Stratified K-Fold cross-validation, our baseline measurement used and calculated to determine the network's performance was the root-mean-square error (RMSE) compared to the predictions and actual DFT values. All training and testing RMSEs for each fold were then averaged. This procedure was repeated for another five runs or trials, in which each run's uncertainties were averaged together.

This procedure gives us a practical testing error obtained for every input data point, providing us with the accurate predictive of the network on each of the eight output values. The optimal set of hyperparameters is chosen to minimize the cross-validation error; we ultimately report training and test errors for every model, but optimization is based on the validation error, such that the actual test set in each iteration remains unseen by the model during the training process.

To better emphasize our testing RMSEs in principle, we crafted parity plots (shown in Fig. 5.4) that compare the benchmark DFT-computed properties to the predicted values from the neural network. As such, we present and visualize the testing results on the A-rich and B-rich formation energies and the six charge transition levels, labeled with the reported average testing RMSEs in eV from Table 5.4 and their standard deviations, for each binary semiconductor type. For each semiconductor group, labeled with different colors for emphasis, all testing data points were cataloged from every individual five Stratified KFold method splits across all five runs.

It can be observed that the integration of descriptors and SOAP power spectrum features significantly enhanced the predictive performance. This improvement is quantified through the average RMSEs obtained from five independent runs, which are systematically tabulated for various output categories as delineated in Table 5.4. Notably, the RMSEs for formation energies exhibited an over 60% enhancement, while the improvements for charge transition levels, specifically ϵ(-1/0),

ϵ(-2/-1), ϵ(+1/0), ϵ(+2/+1), and ϵ(+3/+2), approached nearly 50% relative to the precedent model benchmarks presented in Table 5.1 and Table 5.2. The performance metrics for ϵ(-3/-2) exhibited only marginal advancements. These outcomes substantiate the significant impact of the SOAP features in refining the model's accuracy.

In the "Training Error" column of Table 5.4, we recorded low RMSE values for all output properties across various semiconductor groups, suggesting that the model effectively assimilates the descriptor data from the input and reliably predicts these outputs. Nevertheless, a notable discrepancy between training and testing predictions raises the concern of potential overfitting. To assess the robustness of our models, we determined the standard deviation through a five-fold cross-validation process. As depicted in Fig. 5.4, the standard deviation hovers at approximately half the magnitude of the prediction error, a variance deemed acceptable by prevailing reports in literature[214, 158]. Although a larger regularization term might bridge the test-train performance gap, such a modification was deemed unnecessary. The regularization parameter has been finely tuned via Bayesian optimization, and further increases might introduce undue bias. Moreover, the model's predictions align closely with experimental data in subsequent evaluations of dopant efficacy in devices, reinforcing our confidence in its exceptional performance.

In terms of the network's performance on different outputs, it displays a minor improvement in predicting formation energies of doped semiconductor systems that are A-rich compared to B-rich ones. In addition, it is also shown that group II-VI binary group semiconductors have the lowest RMSEs and have most of their data points in line with the diagonal of the parity plot for both formation energy types compared with the III-V and IV-IV groups.

For each binary semiconductor type in each column of Table 5.4, the RMSE values relatively change regardless of the charge transition level type being measured (with the notable exception of ε(-1/-2) and ε(-2/-3) for IV-IV binary group semiconductors). For instance, the RMSEs for II-VI semiconductors in each individual transition level remain roughly between 0.10 eV to 0.13 eV, while for III-V semiconductors, the errors remain between 0.16 eV to 0.20 eV. Additionally, it is noted that for the binary group II-VI semiconductors, the network again predicts their transition levels

effectively compared to their counterparts, as observed from the low error values and having the majority of their data points remain close to the diagonal, which show reduced variance from the actual DFT values.

Table 5.4: ANN Model Performance

Predictors	Training Error (eV)	Testing Error (eV)	II-VI Error (eV)	III-V Error (eV)	IV-IV Error (eV)
ΔH (A-rich)	0.032	0.31	0.21	0.44	0.46
ΔH (B-rich)	0.036	0.36	0.24	0.56	0.51
ϵ(-1/0)	0.018	0.16	0.11	0.18	0.22
ϵ(-2/-1)	0.024	0.18	0.10	0.18	0.30
ϵ(-3/-2)	0.042	0.24	0.10	0.16	0.47
ϵ(+1/0)	0.012	0.16	0.12	0.20	0.21
ϵ(+2/+1)	0.010	0.15	0.12	0.20	0.20
ϵ(+3/+2)	0.010	0.14	0.13	0.17	0.15

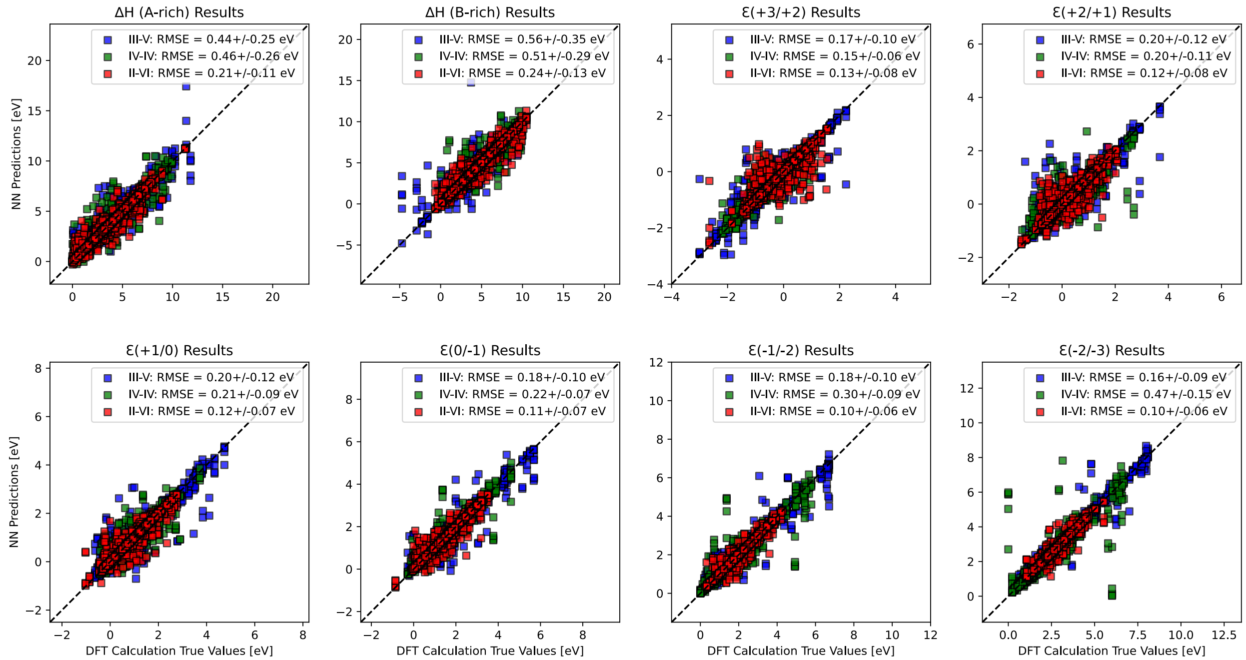


Figure 5.4: Parity Plots for the Artificial Neural Network

5.2.3 Crystal Defect Graph Neural Network

Our CDGNN follows closely with the original Crystal Graph Neural Network (CGCNN) framework of Ref. [221]. We create a surrogate model F_{CDGNN} parameterized by weight $\mathbf{W}$, which has the form:

$$\hat{y}_d = F_{CDGNN}(G_d, \mathbf{u}_{elem}, \mathbf{u}_{struc}; \mathbf{W}), \tag{5.1}$$

where G_d refers to the graph of unrelaxed defect structures, $\mathbf{u}_{elem}$ denotes the elemental properties of atoms comprising the defect structures, and $\mathbf{u}_{struc}$ refers to compound properties (e.g., band gap, compound formation enthalpy, and compound lattice constant) as well as defect structure properties (e.g., Coulomb Matrix, SOAP descriptors). The function F_{CDGNN}, parameterized by weights $\mathbf{W}$, effectively maps a defected crystal G_d along with its corresponding features to the target property $\hat{y}_d$. In this study, $\hat{y}_d$ are defect formation energy and transition levels.

As depicted in Fig. 5.5, the defect elemental properties and bonding information are initially embedded in the feature vector $\mathbf{v}_i$. The convolutional layers then iteratively update $\mathbf{v}_i$ by incorporating information from surrounding atoms and bonds, employing a nonlinear graph convolution function (Eq. (5.2)). After T convolutions, the CDGNN learns a comprehensive feature vector $\mathbf{v}_d$ that represents the unrelaxed defect crystal structure through pooling all nodes in the structure (Eq. (5.3)).

Next, the CDGNN incorporates the compound and defect structure features, such as supercell lattice parameter, bandgap, and SOAP features, by concatenating them with the graph-embedded nodes (Eq. (5.4)). As shown in Table 5.5, we observe an improvement of around 10% in the Root Mean Square Errors (RMSE) after encoding the structure features. This improvement is achieved through several hidden layers, leading to the final output. Upon implementing the graph of defects and their corresponding features into the model, we observed a significant improvement in RMSEs of formation energy predictions compared to our previous models, achieving an RMSE of approximately 0.5 eV (Table .5.5) using the same train-test splitting method from ANN model. However, it is worth noting that no substantial improvement was observed in the transition level predictions. This may be due to the CDGNN model including a large number of features, such as

bonding and element properties of all atoms, which requires a sufficient amount of data for training. To improve the CDGNN model's performance, future work will focus on either incorporating a feature pruning method or increasing the amount of training data

$$\mathbf{v}_i^{t+1} = Conv(\mathbf{v}_i^t, \mathbf{v}_j^t, \mathbf{u}_{(i,j)_q}), \quad \mathbf{u}_{(i,j)_q} \in G_d \tag{5.2}$$

$$\mathbf{v}_d = Pool(\mathbf{v}_0^{(T)}, \mathbf{v}_1^{(T)}, ..., \mathbf{v}_N^{(T)}) \tag{5.3}$$

$$\mathbf{v}_d^{'} = \mathbf{v}_d \oplus \mathbf{u}_{struc} \tag{5.4}$$

Table 5.5: CDGNN Model Performance

Predictors	Training Error (eV)	Testing Error (eV)
ΔH (A-rich) w/o SOAP	0.103	0.589
ΔH (A-rich)	0.117	0.537
ϵ(-1/0)	0.176	0.463
ϵ(-2/-1)	0.420	0.488
ϵ(-3/-2)	0.299	0.391
ϵ(+1/0)	0.148	0.391
ϵ(+2/+1)	0.169	0.406
ϵ(+3/+2)	0.146	0.350

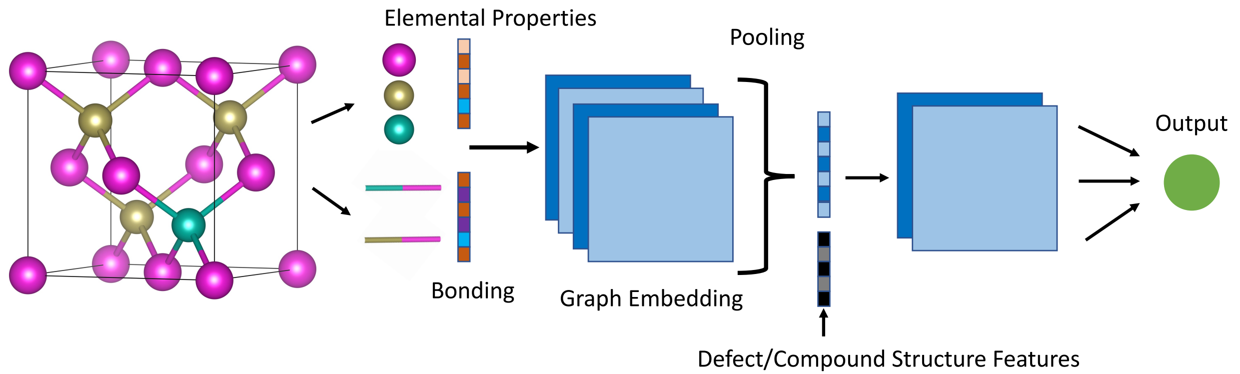


Figure 5.5: Schematic of Crystal Defect Graph Neural Network

5.2.4 ML-assisted Device Design Framework

Most state-of-the-art machine learning models for defects focus on predicting the electronic properties of individual defects. However, in complex real-world systems, such as transistors and solar cells, multiple defects associated with a dopant coexist and collectively influence device performance. Consequently, a comprehensive evaluation framework for dopants is essential to effectively screen and identify optimal dopants. In response to this need, we propose a novel ML-assisted device design framework (Fig. 5.6) that facilitates quantitative and efficient screening and exploration of dopants in semiconductor devices. This innovative framework streamlines the process of material discovery and device design, accelerating the exploration of new dopants for advanced semiconductor devices.

In this framework, the ML-assisted device design framework operates as follows: Firstly, we employ DFT to construct a comprehensive defect property database, encompassing defect formation energy and thermodynamic transition levels. Subsequently, this dataset is used to train a surrogate neural network model capable of predicting defect properties for dopants throughout the entire chemical space. Next, we apply compositionally constrained thermodynamics (CCT) [133] to filter and select dopants with desired properties. The defect information from these potential dopants is then incorporated into the TCAD model specific to the device of interest, facilitating the assessment of essential figures of merit. By leveraging this framework, we are able to identify promising potential dopants with the capacity to enhance device performance. To validate the predictions, experimental testing and validation are needed to ensure the suitability and efficacy of the selected dopants for practical device applications.

CCT is a canonical approach to calculate the defect concentrations. Under a dilute approximation, defect concentrations are given by

$$x_{k,q} = \frac{\theta_{k,q}}{N^{total}} e^{-\Delta H^{f}_{k,q}/k_{\mathrm{B}}T}, \tag{5.5}$$

where $\theta_{k,q}$ is the degeneracy factor that counts the number of equivalent defect configurations,

$\Delta H^f_{k,q}$ is the formation energy of defect k at charge q, and N^{total} is the total number of lattice sites in a perfect material. One can obtain deviations of atomic fractions f_α from perfect stoichiometry f^p_α under given defect concentrations in the material using the following relation

$$f_\alpha = \frac{f^p_\alpha - \sum_{k,q} x_{k,q} n^\alpha_k}{1 - \sum_{k,q} x_{k,q} \sigma_{k,q}} \ , \tag{5.6}$$

where n^α_k is the number of atoms of type α added or removed from the system when one defect k exists. $\sigma_{k,q} = \sum_\alpha n^\alpha_k$ is the atom amount differences in a system with and without the defect k. If atomic fractions and standard reference chemical potentials are provided, one can determine the defect concentration, chemical potentials for each element and the Fermi level by solving the set of equations defined by (5.6). We refer to this method as CCT [133].

In this study, the atomic fractions of the semiconductor compound and its impurities are unknown. Instead, we can use the chemical potentials of each element in the semiconductor compound and its potential dopants to determine the defect concentration and the Fermi level. If the formation energies $H^{(r)}_{k,q}$ are known for given reference values $\mu^{(r)}_\alpha$ and $E^{(r)}_F$, then $H_{k,q}$ for any values of μ_α and E_F can be expressed using the following equations:

$$\begin{aligned} \Delta H_{k,q}(\mu_\alpha, E_F) = \Delta H^{(r)}_{k,q} &+ \sum_\alpha n^\alpha_i (\mu_\alpha - \mu^{(r)}_\alpha) \\ &+ q(E_F - E^{(r)}_F). \end{aligned} \tag{5.7}$$

Using equations (5.5) and (5.7), we can represent the total defect density for any defect using the density of a single neutral defect:

$$x_k = x_{k,0} \sum_q \frac{\theta_{k,q}}{\theta_{k,0}} e^{-\frac{\Delta H^{(r)}_{k,q} - \Delta H^{(r)}_{k,0} + q(E_F - E^{(r)}_F)}{k_B T}}. \tag{5.8}$$

From (5.5) and (5.8), we can easily obtain

$$\sum_\alpha n^\alpha_k \mu_\alpha = \sum_\alpha n^\alpha_k \mu^{(r)}_\alpha - k_B T \ln \frac{x_{k,0} \times N^{total}}{\theta_{k,0}} - \Delta H^{(r)}_{k,0}. \tag{5.9}$$

By applying equation (5.9) and the charge neutrality equation (5.10), we can determine the equilibrium defect distribution and the Fermi level for any given set of chemical potentials and temperature.

$$n + \sum_{k} \sum_{q<0} |q| x_{k,q} = p + \sum_{k} \sum_{q>0} q x_{k,q} \tag{5.10}$$

Taking CdTe as an example, a commonly used semiconductor material in thin-film solar cells [20], we employ the above approach to screen dopants' potential impact on device performance. First, we utilize defect formation energies calculated via DFT for intrinsic defects and employ the ANN model to predict formation energies and transition levels for extrinsic defects. With this information, we determine the equilibrium defect concentrations for different charge states at a given temperature via CCT. Considering the typical manufacturing process of CdTe solar cells, which involves high-temperature vapor deposition followed by cooling and annealing at room temperature [165], we select a temperature of 893 K for the equilibrium defect concentration calculations, followed by a quench to room temperature. We assume a "frozen-in" approximation [88], where the total defect concentrations are held constant from the prior equilibrium calculation. Defects are then redistributed among available charge states based on the new Fermi level and temperature.

Once we obtain the defect distribution, we incorporate these defects into a well-constructed CdTe TCAD device model. Here, defects are categorized as either shallow or deep levels. For deep levels (V_{Cd} and Te_{Cd}[223, 79, 224, 90, 220]), we take Shockley-Read-Hall (SRH) recombination into account, while for extrinsic defects, we exclude SRH recombination due to low capture rates. The capture rates of deep levels are computed using the NONRAD method [5, 198]. Further details about the device model are provided in the Supplemental material (Table B.1). In the end, we generate a dopant screening table, as shown in Table 5.6 and Table B.2. For each dopant, we explore two conditions in CdTe: Cd-rich and Te-rich. And the chemical potential for each dopant is the maximum limit of their lowest formation energy binary or ternary compounds formed with Cd or Te. This provides an optimistic or pessimistic estimation based on the dopants' solubility in CdTe. The table includes key characteristics of the defects, such as dominant defects, whether they introduce deep levels (with energy levels within the band gap ± 0.2 eV), the material type after

doping, Fermi level relative to the valence band maximum (VBM) at 300 K, and the majority carrier density predicted by the machine learning framework (negative values indicating electron density and positive values indicating hole density). Experimental reports of majority carrier density at 300 K are also included for comparison. Additionally, the TCAD model provides four key performance metrics for solar cell devices: efficiency (η), open-circuit voltage (Voc), short-circuit current (Jsc), and fill factor (FF). These metrics offer a clear indication of the dopants' potential to enhance device performance.

Using these features, we efficiently screen potential dopants to enhance device performance. Table 5.6 presents a selection of promising dopants for CdTe solar cells, which do not introduce deep-level defects and significantly increases carrier density compared to intrinsic CdTe, suggesting that the formation energies of dominant extrinsic defects of these dopants are more favorable compared to the intrinsic defects under specified conditions. In the first row, intrinsic CdTe is included as a reference. We list several dopants which are able to raise carrier density several orders compared to intrinsic CdTe. Remarkably, this framework exhibits excellent performance, selecting some of the most commonly used dopants such as group V elements (As, P) and Cu. The carrier densities closely align with experimental results, as indicated in Table 5.6. It is noteworthy that, based on predicted carrier density, Cu appears to be a more effective dopant. However, experimental data suggests that Cu doping is less effective when compared to group V species. This phenomenon may be attributed to the instability of copper dopants and the presence of compensating defects [155]. Additionally, it's important to acknowledge that the "frozen-in" approximation might not hold in certain cases, particularly when the interactions between different defects are significant, like vacancy-interstitial pair annihilation and exchange reactions [43]. Also, the strong Coulomb interaction between defects with opposite charges can give rise to defect complexes[218, 219], a consideration not addressed within the current framework.

Additionally, we've also identified experimental evidence supporting the efficacy of fluorine doping in CdTe, which aligns with our results and suggests fluorine as a promising dopant, increasing carrier density to around 10^{17} cm^{-3} [147]. Furthermore, as research groups show growing interest

in n-type CdTe solar cells [151], we have also identified potential dopants in Table B.2. Group VII elements are predicted to be effective for rendering CdTe n-type, potentially achieving carrier densities exceeding 10^{15} cm^{-3}. However, the possibility of introducing deep-level traps by these species may impact device performance.

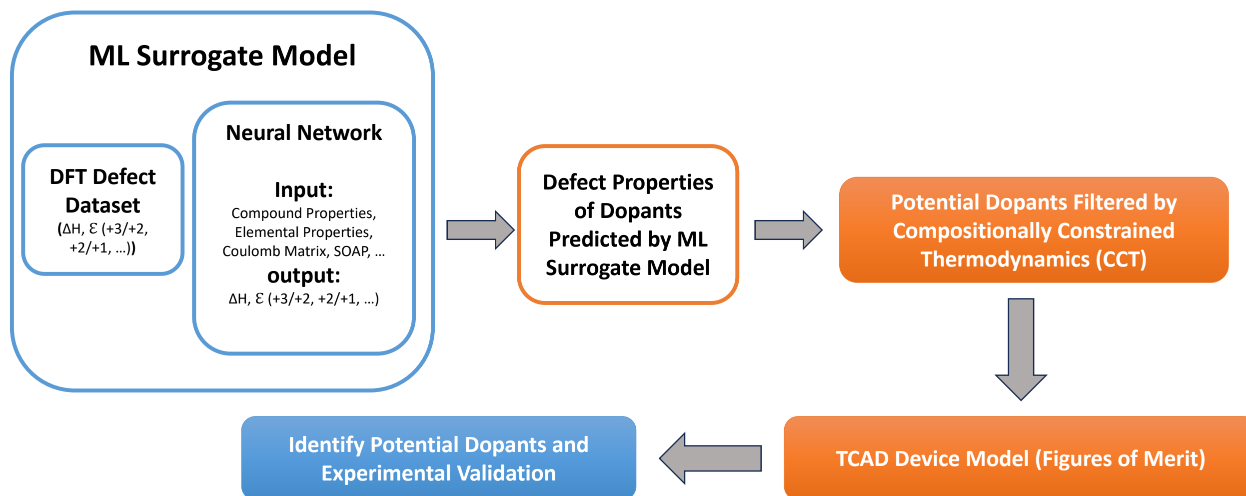


Figure 5.6: Schematic of ML-assisted Device Design Framework. DFT is employed to build a comprehensive defect property database including defect formation energy and thermodynamic transition levels. A surrogate neural network model is trained using this dataset to predict defect properties for dopants across the entire chemical space. Next, CCT is used to filter and select dopants with desired properties. Then the defect information of these potential dopants is incorporated into the TCAD model of the specific device of interest, allowing us to gather figures of merit. In the end, we identify promising potential dopants that demonstrate potential for enhancing device performance. Experimental testing and validation are conducted to assess the suitability and efficacy of the selected dopants.

5.2.5 Discussion

Based on the parity plots of Figure 5.4 and the columns of Tables 5.4 and Tables 5.5 that describe the testing prowess of the models in predicting the two types of formation energies and six types of transition levels, it is observed that the output RMS performance errors for all types of transition levels are lower than the formation energies. A well-known principle that can explain why predicting transition levels, such as reaction barriers or activation energies, is often easier than predicting

Table 5.6: ML-assisted Dopant Screening for Device Design (Example of CdTe Solar Cells)

Dopant	Condition	Dominant Defect	Deep Level?	Type	Fermi Level (eV)	Carrier Density @ 300K (cm^{-3})	Carrier Density (Experiment)	Device Performance			
								η	Voc (V)	Jsc (mA)	FF
-	Te-rich	V_{Cd}, Te_{Cd}	Yes	*p*-type	0.34	7.53e+13	$10^{13} \sim 10^{15}$ [226, 124, 232]	16.59	0.86	22.56	85.69
F	Te-rich	F_int_A	No	*p*-type	0.12	1.98e+17	$10^{14} \sim 10^{17}$ [147]	20.94	1.07	22.31	88.06
N	Cd-rich	N_Te	No	*p*-type	0.25	1.16e+15	$10^{15} \sim 10^{17}$ [146]	18.45	0.85	22.56	87.42
P	Te-rich	P_Te	No	*p*-type	0.20	9.70e+15	$10^{15} \sim 10^{17}$ [135, 124]	19.55	0.90	22.53	87.61
As	Te-rich	As_Te	No	*p*-type	0.24	1.61e+15	$10^{15} \sim 10^{17}$ [135, 134, 1]	18.54	0.94	22.56	87.17
As	Cd-rich	As_Te	No	*p*-type	0.23	2.22e+15	$10^{15} \sim 10^{17}$ [135, 134, 1]	18.55	0.95	22.59	87.44
Cu	Te-rich	Cu_Cd	No	*p*-type	0.14	8.37e+16	$10^{14} \sim 10^{16}$ [80, 103, 155]	20.53	1.05	22.29	88.10

formation energies is that they depend on the relative energies of the given states of the local, isolated process within the system rather than considering the total energies of all particles that composed the entire system.

From the formula describing the calculations for the transition levels within the semiconductor (see Experimental procedures in [120]), such levels can be represented as the energy differences between two charge states. Because these differences depend on the relative positions of the two states in energy space, they can be more accurately predicted than the formation energies, which depend on each particle that makes up the entire structure. Furthermore, it can be shown from the derivation of the transition levels that the chemical potential variable is independent when calculating any defined level compared to the formation energies. This suggests that predicting the formation energies to get the same errors as the transition level requires more features to better incorporate the chemical potential into the network.

Compared to its binary group counterparts, the artificial neural network demonstrates superior prediction accuracy for all output DFT-computed properties in binary group II-VI semiconductors, with smaller errors observed. This enhanced performance can be attributed, first, to the larger portion of II-IV semiconductor data points available for training. However, it is important to consider that the network's underperformance for the IV-IV group can be attributed to the fact that certain crystal compounds within this group may not be stable in their cubic phases (e.g., SiGe). Consequently, this instability leads to unrealistic defect formation energy calculations, expanding the formation energy range, and rendering prediction more challenging. As II-VI binary group semiconductors are the most common types of semiconductor system utilized in thin-film

solar energy technologies, these low error thresholds suggest that the network is a great tool in predicting the development of new dopants in newer semiconductors for such applications.

The improvements in the performance of ANN and CDGNN models due to the utilization of the SOAP descriptors can be attributed to the fact that such features assist in adding additional information regarding the local defect environment, which describes the region where electrons or holes are exchanged from the atomic impurity to the conduction or valence band; therefore, can assist in predicting electronic properties of defect within the system. Using the Gaussian-smeared atomic density around each atom, the descriptors help catalog the overlaps of electron clouds and the interactive effects of chemical bonding. At the same time, these descriptors are invariant to the permutations of atomic indices, meaning that one can fine-tune the model to help consider such invariance from the material structure without the input of a significant amount of training data to learn such cases, allowing for faster computational times and reducing cost for our neural network framework.

The relatively low error of ANN for II-IV group compound prediction makes quantitative screen and prediction possible. The ML-assisted framework for device design presented in this work is a powerful tool that leverages machine learning techniques to identify potential dopants to optimize semiconductor device performance. The framework begins with a comprehensive database of defect properties, including defect formation energies and thermodynamic transition levels, generated using DFT and ANN predictions. However, several challenges and limitations must be considered. Notably, while the ML model is capable of predicting a wide range of properties and can optimize device design, it faces challenges when estimating the deep-level capture rate for defects. Accurate predictions for deep-level defects necessitate theoretical calculations and experimental evidence [231], which the ML model cannot provide at current stage. Additionally, the model's consideration is primarily focused on substitutional defects and interstitials, overlooking other critical point defect types, such as defect complexes [189, 216], which can significantly impact semiconductor devices. To enhance the framework, a more comprehensive defect dataset is required. The other limitation is that the dataset does not include hexagonal phase semiconductors, such as CdSe and GaN. The

framework's performance and applicability are limited to the semiconductor materials present in the available data. Expanding the dataset to include hexagonal phase semiconductors could enhance the framework's versatility.

5.3 Summary

The ML-assisted Device Design Framework presents a powerful approach to streamline the dopant screening process, resulting in significant time and cost savings. Instead of relying solely on exhaustive experimental trials, the framework utilizes machine learning to predict defect properties for various dopants across the chemical space. This predictive capability enables researchers to identify promising potential dopants more efficiently, reducing the need for extensive and resource-intensive experimentation.

By employing the comprehensive defect property database generated through DFT, the framework significantly narrows down the pool of potential dopants with desired properties. This targeted selection process minimizes the need to perform numerous experiments, saving both time and labor.

Moreover, the integration of TCAD simulation further refines the search for dopants, focusing on those that have the most favorable characteristics for device optimization. By combining machine learning predictions with physical models in TCAD simulations, researchers can efficiently gather figures of merit and assess the potential impact of selected dopants on device performance.

The ML-assisted device design framework represents a significant departure from traditional dopant screening methods, providing an innovative, data-driven approach that streamlines the search for superior dopants in semiconductor devices. This transformative framework optimizes resources and expedites the discovery process. Nevertheless, some challenges remain, particularly in accurately predicting deep-level capture rates and addressing the absence of defect complexes, and hexagonal phase semiconductor data. Furthermore, there is room for improvement in achieving quantitative analysis for III-V and IV-IV types, which might require additional data or more refined descriptors. Future research efforts should be directed towards overcoming these limitations while further advancing the application of this framework in semiconductor device design.

Chapter 6

CONCLUSION

This dissertation presents a comprehensive framework for defect modeling and device optimization in chalcogenide photovoltaics, specifically targeting CIGS and CdSeTe materials. By integrating first principles density functional theory (DFT), continuum modeling, and AI-driven methods, we have developed a robust predictive model for the electronic performance and optimization of thin-film solar cells.

One of the key contributions of this work is the detailed analysis of defect formation and diffusion mechanisms in CIGS and CdSeTe. Through DFT calculations, we have accurately determined the formation energies and migration barriers of various intrinsic and extrinsic defects. This fundamental understanding allows for precise control over defect profiles, which is critical for optimizing the material properties and enhancing device performance.

Our continuum modeling approach further bridges the gap between atomic-scale defect properties and macroscopic device behavior. By incorporating defect thermodynamics and kinetics into process and device simulations, we have successfully predicted the impact of defect distributions on carrier lifetimes and solar cell efficiencies. This coupled simulation framework provides valuable insights into the manufacturing process, enabling the fine-tuning of process parameters to achieve optimal device performance.

Additionally, the incorporation of AI-driven methods, including machine learning algorithms, has significantly accelerated the prediction and screening of defect properties. Our machine learning models, trained on extensive DFT datasets, offer rapid and accurate predictions of defect formation energies and charge transition levels. This integration of AI with traditional modeling techniques exemplifies the potential of hybrid approaches in advancing semiconductor research and development.

The predictive TCAD model developed in this work represents a significant advancement in the field of chalcogenide photovoltaics. By providing a detailed and accurate depiction of defect behaviors and their impact on device performance, this framework serves as a powerful tool for guiding experimental efforts and accelerating the development of high-efficiency thin-film solar cells.

In conclusion, this dissertation not only enhances our understanding of defect physics in chalcogenide materials but also establishes a versatile and efficient framework for the optimization of thin-film photovoltaic devices. The methodologies and insights presented here are expected to have far-reaching implications, paving the way for future innovations in photovoltaic technology and beyond.

BIBLIOGRAPHY

[1] Tursun Ablekim, Santosh K Swain, Wan-Jian Yin, Katherine Zaunbrecher, James Burst, Teresa M Barnes, Darius Kuciauskas, Su-Huai Wei, and Kelvin G Lynn. Self-compensation in arsenic doping of cdte. *Scientific reports*, 7(1):4563, 2017.

[2] D Abou-Ras, G Kostorz, A Romeo, D Rudmann, and AN Tiwari. Structural and chemical investigations of CBD-and PVD-CdS buffer layers and interfaces in Cu(In, Ga)Se_2-based thin film solar cells. *Thin Solid Films*, 480:118–123, 2005.

[3] Ankit Agrawal and Alok Choudhary. An online tool for predicting fatigue strength of steel alloys based on ensemble data mining. *International Journal of Fatigue*, 113:389–400, 2018.

[4] DS Albin, Y Yan, and MM Al-Jassim. The effect of oxygen on interface microstructure evolution in cds/cdte solar cells. *Progress in Photovoltaics: Research and Applications*, 10(5):309–322, 2002.

[5] Audrius Alkauskas, Qimin Yan, and Chris G Van de Walle. First-principles theory of nonradiative carrier capture via multiphonon emission. *Physical Review B*, 90(7):075202, 2014.

[6] MI Alonso, K Wakita, J Pascual, M Garriga, and N Yamamoto. Optical functions and electronic structure of $CuInSe_2$, $CuGaSe_2$, $CuInS_2$, and $CuGaS_2$. *Physical Review B*, 63(7):075203, 2001.

[7] Saâd Amara and Mohamed Bouafia. Characterisation of TCO AZO/glass structures by spectroscopic ellipsometry. *International Journal of Nanoparticles 11*, 6(2-3):122–131, 2013.

[8] Elisa Artegiani, Jonathan D Major, Huw Shiel, Vin Dhanak, Claudio Ferrari, and Alessandro Romeo. How the amount of copper influences the formation and stability of defects in cdte solar cells. *Solar Energy Materials and Solar Cells*, 204:110228, 2020.

[9] A Balcioglu, RK Ahrenkiel, and F Hasoon. Deep-level impurities in cdte/cds thin-film solar cells. *Journal of Applied Physics*, 88(12):7175–7178, 2000.

[10] D. Ban, H. Luo, H. C. Liu, Z. R. Wasilewski, A. J. SpringThorpe, R. Glew, and M. Buchanan. Optimized gaas/algaas light-emitting diodes and high efficiency wafer-fused optical upconversion devices. *Journal of Applied Physics*, 96(9):5243–5248, 2004.

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

[12] Thomas Beckers, Lakshmi Nagarajan, and Manfred Martin. Radiotracer diffusion of 114m In in Cu(In, Ga)Se_2 thin films. *Thin Solid Films*, 592:118–123, 2015.

[13] MA Berding. Native defects in cdte. *Physical Review B*, 60(12):8943, 1999.

[14] J. Bergstra, D. Yamins, and D.D. Cox. Making a science of model search: Hyperparameter optimization in hundreds of dimensions for vision architectures. In *Proceedings of the 30th International Conference on Machine Learning*, volume 28, Atlanta, Georgia, USA, 2013. JMLR.

[15] Ekaba Bisong. *Building machine learning and deep learning models on Google cloud platform: A comprehensive guide for beginners*. Apress, 2019.

[16] Koushik Biswas and Mao-Hua Du. What causes high resistivity in cdte. *New Journal of Physics*, 14(6):063020, 2012.

[17] N. Bouarissa, H. Algarni, M. Ajmal Khan, O.A. Al-Hagan, and T.F. Alhuwaymel. Collective effects and optical characteristics of $cdse_x te_{1-x}$. *Optik*, 203:163952, 2020.

[18] Robert F Brebrick. High temperature thermodynamic data for cdte (c). *Journal of phase equilibria and diffusion*, 31:260–269, 2010.

[19] Leo Breiman. Random foresets. *Machine Learning*, 45:5–32, 2001.

[20] J Britt and C Ferekides. Thin-film cds/cdte solar cell with 15.8% efficiency. *Applied physics letters*, 62(22):2851–2852, 1993.

[21] Danny Broberg, Bharat Medasani, Nils ER Zimmermann, Guodong Yu, Andrew Canning, Maciej Haranczyk, Mark Asta, and Geoffroy Hautier. Pycdt: A python toolkit for modeling point defects in semiconductors and insulators. *Computer Physics Communications*, 226:165–179, 2018.

[22] J. M. Burst, J. N. Duenow, D. S. Albin, E. Colegrove, M. O. Reese, J. A. Aguiar, C. S. Jiang, M. K. Patel, M. M. Al-Jassim, D. Kuciauskas, S. Swain, T. Ablekim, K. G. Lynn, and W. K. Metzger. Cdte solar cells with open-circuit voltage breaking the 1 v barrier. *Nature Energy*, 1:16015, 2016.

[23] James M Burst, Stuart B Farrell, David S Albin, Eric Colegrove, Matthew O Reese, Joel N Duenow, Darius Kuciauskas, and Wyatt K Metzger. Carrier density and lifetime for different dopants in single-crystal and polycrystalline cdte. *APL Materials*, 4(11), 2016.

[24] A Castaldini, A Cavallini, B Fraboni, P Fernandez, and J Piqueras. Deep energy levels in cdte and cdznte. *Journal of applied physics*, 83(4):2121–2126, 1998.

[25] CWM Castleton and Susanne Mirbt. Finite-size scaling as a cure for supercell approximation errors in calculations of neutral native defects in inp. *Physical Review B*, 70(19):195202, 2004.

[26] L Červinka and A Hrubỳ. The crystal structure of cdas2. *Acta Crystallographica Section B: Structural Crystallography and Crystal Chemistry*, 26(4):457–458, 1970.

[27] JP Chamonal, E Molva, and JL Pautrat. Identification of cu and ag acceptors in cdte. *Solid State Communications*, 43(11):801–805, 1982.

[28] Yia-Chung Chang, RB James, and JW Davenport. Symmetrized-basis lasto calculations of defects in cdte and znte. *Physical Review B*, 73(3):035211, 2006.

[29] Jakapan Chantana, Taichi Watanabe, Seiki Teraji, and Takashi Minemoto. Influence of minimum position in [Ga]/([Ga]+[In]) profile of Cu(In, Ga)Se_2 on flexible stainless steel substrate on its photovoltaic performances. *Solar Energy Materials and Solar Cells*, 157:750–756, 2016.

[30] Intuon Chatratin, Baoying Dou, Su-Huai Wei, and Anderson Janotti. Doping limits of phosphorus, arsenic, and antimony in cdte. *The Journal of Physical Chemistry Letters*, 14(1):273–278, 2023.

[31] Dami Choi, Christopher J. Shallue, Zachary Nado, Jaehoon Lee, Chris J. Maddison, and George E. Dahl. On empirical comparisons of optimizers for deep learning, 2020.

[32] HC Chou, A Rohatgi, NM Jokerst, EW Thomas, and S Kamra. Copper migration in cdte heterojunction solar cells. *Journal of electronic materials*, 25:1093–1098, 1996.

[33] J de Wild, DG Buldu, T Kohl, G Birant, G Brammertz, M Meuris, J Poortmans, and B Vermang. Intermediate scale bandgap fluctuations in ultrathin Cu(In, Ga)Se_2 absorber layers. *Journal of Applied Physics*, 128(16):163102, 2020.

[34] Ann M. Deml, Aaron M. Holder, Ryan P. O'Hayre, Charles B. Musgrave, and Vladan Stevanović. Intrinsic material properties dictating oxygen vacancy formation energetics in metal oxides. *The Journal of Physical Chemistry Letters*, 6(10):1948–1953, 2015.

[35] Ann M. Deml, Vladan Stevanović, Christopher L. Muhich, Charles B. Musgrave, and Ryan O'Hayre. Oxide enthalpy of formation and band gap energy as accurate descriptors of oxygen vacancy formation energetics. *Energy Environ. Sci.*, 7:1996–2004, 2014.

[36] Hieu A. Doan, Garvit Agarwal, Hai Qian, Michael J. Counihan, Joaquín Rodríguez-López, Jeffrey S. Moore, and Rajeev S. Assary. Quantum chemistry-informed active learning to accelerate the design and discovery of sustainable energy storage materials. *Chemistry of Materials*, 32(15):6338–6346, 2020.

[37] Marilyn J Dodge. Refractive properties of magnesium fluoride. *Applied optics*, 23(12):1980–1985, 1984.

[38] Baoying Dou, Qingde Sun, and Su-Huai Wei. Effects of co-doping in semiconductors: Cdte. *Physical Review B*, 104(24):245202, 2021.

[39] Cyrus E Dreyer, Audrius Alkauskas, John L Lyons, and Chris G Van de Walle. Radiative capture rates at deep defects from electronic structure calculations. *Physical Review B*, 102(8):085305, 2020.

[40] Mao-Hua Du, Hiroyuki Takenaka, and David J Singh. Native defects and oxygen and hydrogen-related defect complexes in cdte: Density functional calculations. *Journal of Applied Physics*, 104(9), 2008.

[41] B.B. Dumre, N.J. Szymanski, V. Adhikari, I. Khatri, D. Gall, and S.V. Khare. Improved optoelectronic properties in $cdse_x te_{1-x}$ through controlled composition and short-range order. *Solar Energy*, 194:742–750, 2019.

[42] P Emanuelsson, P Omling, BK Meyer, M Wienecke, and M Schenk. Identification of the cadmium vacancy in cdte by electron paramagnetic resonance. *Physical Review B*, 47(23):15578, 1993.

[43] Paul Martin Fahey, PB Griffin, and JD Plummer. Point defects and dopant diffusion in silicon. *Reviews of modern physics*, 61(2):289, 1989.

[44] Jianqing Fan and Jinchi Lv. Sure independence screening for ultrahigh dimensional feature space. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(5):849–911, 2008.

[45] Jianqing Fan, Richard Samworth, and Yichao Wu. Ultrahigh dimensional feature selection: beyond the linear model. *Journal of machine learning research*, 10:2013–2038, 2009.

[46] Daniel C. FElton, Zois Boukouvalas, Mark S. Butrico, Mark D. Fuge, and Peter W. Chung. Applying machine learning techniques to predict the properties of energetic materials. *Scientific Reports*, 8:9059, 2018.

[47] Hong-Jian Feng, Kan Wu, and Zun-Yi Deng. Predicting inorganic photovoltaic materials with efficiencies ¿26% via structure-relevant machine learning and density functional calculations. *Cell Reports Physical Science*, 1(9):100179, 2020.

[48] Thomas AM Fiducia, Budhika G Mendis, Kexue Li, Chris RM Grovenor, Amit H Munshi, Kurt Barth, Walajabad S Sampath, Lewis D Wright, Ali Abbas, Jake W Bowers, et al. Understanding the role of selenium in defect passivation for highly efficient selenium-alloyed cadmium telluride solar cells. *Nature Energy*, 4(6):504–511, 2019.

[49] Nathan C. Frey, Deji Akinwande, Deep Jariwala, and Vivek B. Shenoy. Machine learning-enabled design of point defects in 2d materials for quantum and neuromorphic information processing. *ACS Nano*, 14(10):13406–13417, 2020.

[50] Christoph Freysoldt, Blazej Grabowski, Tilmann Hickel, Jörg Neugebauer, Georg Kresse, Anderson Janotti, and Chris G Van de Walle. First-principles calculations for point defects in solids. *Reviews of modern physics*, 86(1):253, 2014.

[51] Christoph Freysoldt, Jörg Neugebauer, and Chris G Van de Walle. Fully ab initio finite-size corrections for charged-defect supercell calculations. *Physical review letters*, 102(1):016402, 2009.

[52] Theresa Magorian Friedlmeier, Philip Jackson, Andreas Bauer, Dimitrios Hariskos, Oliver Kiowski, Richard Menner, Roland Wuerz, and Michael Powalla. High-efficiency Cu(In, Ga)Se_2 solar cells. *Thin Solid Films*, 633:13–17, 2017.

[53] Christopher Frisk, Charlotte Platzer-Björkman, Jörgen Olsson, Piotr Szaniawski, JT Wätjen, Viktor Fjällström, P Salomé, and Marika Edoff. Optimizing Ga-profiles for highly efficient Cu(In, Ga)Se_2 thin film solar cells in simple and complex defect models. *Journal of Physics D: Applied Physics*, 47(48):485104, 2014.

[54] Andrew M Gabor, John R Tuttle, David S Albin, Miguel A Contreras, Rommel Noufi, and Allen M Hermann. High-efficiency $CuIn_xGa_{1-x}Se_2$ solar cells made from $(In_x, Ga_{1-x})_2Se_3$ precursor films. *Applied physics letters*, 65(2):198–200, 1994.

[55] Aaron S Gehrke, David E Sommer, and Scott T Dunham. Atomistic models of in and ga diffusion in cu (in, ga) se2. *Journal of Applied Physics*, 134(11), 2023.

[56] AA Gippius, JR Panossian, and VA Chapnin. Deep-centre ionization energies in cdte determined from electrical and optical measurements. *physica status solidi (a)*, 21(2):753–758, 1974.

[57] Martin A. Green, Ewan D. Dunlop, Masahiro Yoshita, Nikos Kopidakis, Karsten Bothe, Gerald Siefer, and Xiaojing Hao. Solar cell efficiency tables (version 63). *Progress in Photovoltaics: Research and Applications*, 32(1):3–13, 2024.

[58] Baijie Gu, Nai H Kwong, and Rolf Binder. Relation between the interband dipole and momentum matrix elements in semiconductors. *Physical Review B*, 87(12):125301, 2013.

[59] Zhongyuan Guan, Fei Chen, Zhenyang Liu, Peiwen Lv, Mingjun Chen, Mingxuan Guo, Xu Li, Feng Teng, Song Chen, and Aiwei Tang. Compositional engineering of multinary cu–in–zn-based semiconductor nanocrystals for efficient and solution-processed red-emitting quantum-dot light-emitting diodes. *Organic Electronics*, 74:46–51, 2019.

[60] Trevor Hastie, Robert Tibshirani, Jerome H Friedman, and Jerome H Friedman. *The elements of statistical learning: data mining, inference, and prediction*, volume 2. Springer, 2009.

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

[62] Graeme Henkelman and Hannes Jónsson. Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points. *The Journal of chemical physics*, 113(22):9978–9985, 2000.

[63] Graeme Henkelman, Blas P Uberuaga, and Hannes Jónsson. A climbing image nudged elastic band method for finding saddle points and minimum energy paths. *The Journal of chemical physics*, 113(22):9901–9904, 2000.

[64] R Herberholz, V Nadenau, U Rühle, Ch Köble, HW Schock, and B Dimmler. Prospects of wide-gap chalcopyrites for thin film photovoltaic modules. *Solar energy materials and solar cells*, 49(1-4):227–237, 1997.

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

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

[67] J Horn. Crystal structure and absolute configuration of beta-cdp2. *BULLETIN DE L ACADEMIE POLONAISE DES SCIENCES-SERIE DES SCIENCES CHIMIQUES*, 17(2):69, 1969.

[68] Xiaobo Hu, Takeaki Sakurai, Akimasa Yamada, Shogo Ishizuka, Shigeru Niki, and Katsuhiro Akimoto. Investigation of the properties of deep-level defect in Cu(In, Ga)Se_2 thin films by steady-state photocapacitance and time-resolved photoluminescence measurements. *Japanese Journal of Applied Physics*, 54(4S):04DR02, 2015.

[69] M Igalson, A Kubiaczyk, and P Zabierowski. Deep centers and fill factor losses in the CIGS devices. *MRS Online Proceedings Library (OPL)*, 668, 2001.

[70] Magorzata Igalson, Pawe Zabierowski, Daniel Przado, Aleksander Urbaniak, Marika Edoff, and William N Shafarman. Understanding defect-related issues limiting efficiency of CIGS solar cells. *Solar energy materials and solar cells*, 93(8):1290–1295, 2009.

[71] Malgorzata Igalson. Metastable defect distributions in CIGS solar cells and their impact on device efficiency. *MRS Online Proceedings Library (OPL)*, 1012, 2007.

[72] Yuma Iwasaki, Ichiro Takeuchi, Valentin Stanev, Aaron Gilad Kusne, Masahiko Ishida, Akihiro Kirihara, Kazuki Ihara, Ryohto Sawada, Koichi Terashima, Hiroko Someya, Ken-ichi Uchida, Eiji Saitoh, and Shinichi Yorozu. Machine-learning guided discovery of a new thermoelectric material. *Scientific Reports*, 9(1):2751, 2019.

[73] Philip Jackson, Roland Würz, Uwe Rau, Julian Mattheis, Matthias Kurth, Thomas Schlötzer, Gerhard Bilger, and Jürgen H Werner. High quality baseline for high efficiency, Cu(In_{1-x}, Ga_x)Se_2 solar cells. *Progress in Photovoltaics: Research and Applications*, 15(6):507–519, 2007.

[74] Anubhav Jain, Geoffroy Hautier, Shyue Ping Ong, Charles J Moore, Christopher C Fischer, Kristin A Persson, and Gerbrand Ceder. Formation enthalpies by mixing gga and gga+ u calculations. *Physical Review B*, 84(4):045115, 2011.

[75] Seunghwan Ji, Takahiro Hayakawa, Naoki Suyama, Kazuyoshi Nakada, and Akira Yamada. Enhancement of Cu(In, Ga)Se_2 solar cells efficiency by controlling the formation of Cu-deficient layer. *Japanese Journal of Applied Physics*, 59(4):041003, 2020.

[76] Ana Kanevce, Matthew O Reese, TM Barnes, SA Jensen, and WK Metzger. The roles of carrier concentration and interface, bulk, and grain-boundary recombination for 25% efficient cdte solar cells. *Journal of Applied Physics*, 121(21), 2017.

[77] Shankar Karki, Pran Paul, Grace Rajan, Benjamin Belfore, Deewakar Poudel, Angus Rockett, Evgeny Danilov, Felix Castellano, Aaron Arehart, and Sylvain Marsillac. Analysis of recombination mechanisms in RbF-treated CIGS solar cells. *IEEE Journal of Photovoltaics*, 9(1):313–318, 2018.

[78] G Kartopu, O Oklobia, D Turkay, DR Diercks, BP Gorman, Vincent Barrioz, Stephen Campbell, JD Major, MK Al Turkestani, SELÇUK Yerci, et al. Study of thin film poly-crystalline cdte solar cells presenting high acceptor concentrations achieved by in-situ arsenic doping. *Solar Energy Materials and Solar Cells*, 194:259–267, 2019.

[79] Seán R Kavanagh, Aron Walsh, and David O Scanlon. Rapid recombination by cadmium vacancies in cdte. *ACS energy letters*, 6(4):1392–1398, 2021.

[80] Imran S Khan, Vamsi K Evani, Vasilios Palekis, and Chris Ferekides. Effect of stoichiometry on the lifetime and doping concentration of polycrystalline cdte. *IEEE Journal of Photovoltaics*, 7(5):1450–1455, 2017.

[81] Fadoua Khmaissia, Hichem Frigui, Mahendra Sunkara, Jacek Jasinski, Alejandro Martinez Garcia, Tom Pace, and Madhu Menon. Accelerating band gap prediction for solar materials using feature selection and regression techniques. *Computational Materials Science*, 147:304–315, 2018.

[82] Chiho Kim, Anand Chandrasekaran, Tran Doan Huan, Deya Das, and Rampi Ramprasad. Polymer genome: A data-powered polymer informatics platform for property predictions. *The Journal of Physical Chemistry C*, 122(31):17575–17585, 2018.

[83] Sunghyun Kim, Samantha N. Hood, and Aron Walsh. Anharmonic lattice relaxation during nonradiative carrier capture. *Phys. Rev. B*, 100:041202, Jul 2019.

[84] Sunghyun Kim, Ji-Sang Park, Samantha N Hood, and Aron Walsh. Lone-pair effect on carrier capture in Cu_2ZnSnS_4 solar cells. *J. Mater. Chem. A*, 7(6):2686, 2019.

[85] Matthew T. Klug, Anna Osherov, Amir A. Haghighirad, Samuel D. Stranks, Patrick R. Brown, Sai Bai, Jacob T.-W. Wang, Xiangnan Dang, Vladimir Bulovic, Henry J. Snaith, and Angela M. Belcher. Tailoring metal halide perovskites through metal substitution: influence on photovoltaic and material properties. *Energy Environ. Sci.*, 10:236–246, 2017.

[86] Hannu-Pekka Komsa, Tapio T Rantala, and Alfredo Pasquarello. Finite-size supercell correction schemes for charged defect calculations. *Physical Review B*, 86(4):045112, 2012.

[87] Oliver Kramer. *K-Nearest Neighbors*, pages 13–23. Springer Berlin Heidelberg, Berlin, Heidelberg, 2013.

[88] D Krasikov, A Knizhnik, B Potapkin, S Selezneva, and T Sommerer. First-principles-based analysis of the influence of cu on cdte electronic properties. *Thin Solid Films*, 535:322–325, 2013.

[89] D Krasikov and I Sankin. Beyond thermodynamic defect models: A kinetic simulation of arsenic activation in cdte. *Physical Review Materials*, 2(10):103803, 2018.

[90] D. N. Krasikov, A. V. Scherbinin, A. A. Knizhnik, A. N. Vasiliev, B. V. Potapkin, and T. J. Sommerer. Theoretical analysis of non-radiative multiphonon recombination activity of intrinsic defects in CdTe. *Journal of Applied Physics*, 119(8):085706, 02 2016.

[91] Dmitry Krasikov. Selenium lowers bulk recombination. *Nature Energy*, 4(6):442–443, 2019.

[92] Dmitry Krasikov, Da Guo, Samuel Demtsu, and Igor Sankin. Comparative study of as and cu doping stability in cdsete absorbers. *Solar Energy Materials and Solar Cells*, 224:111012, 2021.

[93] Dmitry Krasikov and Igor Sankin. Defect interactions and the role of complexes in the cdte solar cell absorber. *Journal of Materials Chemistry A*, 5(7):3503–3513, 2017.

[94] DN Krasikov, AV Scherbinin, AA Knizhnik, AN Vasiliev, BV Potapkin, and TJ Sommerer. Theoretical analysis of non-radiative multiphonon recombination activity of intrinsic defects in cdte. *Journal of Applied Physics*, 119(8), 2016.

[95] H Krebs, K-H Müller, and G Zürn. Über kristallisierte metallpolyphosphide. i. darstellung und struktur des cdp4. *Zeitschrift für anorganische und allgemeine Chemie*, 285(1-2):15–28, 1956.

[96] Georg Kresse and Jürgen Furthmüller. Efficient iterative schemes for ab initio total-energy calculations using a plane-wave basis set. *Physical review B*, 54(16):11169, 1996.

[97] Georg Kresse and Jürgen Hafner. Ab initio molecular dynamics for liquid metals. *Physical review B*, 47(1):558, 1993.

[98] Georg Kresse and Daniel Joubert. From ultrasoft pseudopotentials to the projector augmented-wave method. *Physical review b*, 59(3):1758, 1999.

[99] Darius Kuciauskas, Marco Nardone, Alexandra Bothwell, David Albin, Carey Reich, Chungho Lee, and Eric Colegrove. Why increased cdsete charge carrier lifetimes and radiative efficiencies did not result in voltage boost for cdte solar cells. *Advanced Energy Materials*, 13(35):2301784, 2023.

[100] Shou-Yi Kuo, Ming-Yang Hsieh, Fang-I Lai, Yu-Kuang Liao, Ming-Hsuan Kao, and Hao-Chung Kuo. Modeling and optimization of sub-wavelength grating nanostructures on cu (in, ga) se2 solar cell. *Japanese Journal of Applied Physics*, 51(10S):10NC14, 2012.

[101] Hans Landolt and Richard Börnstein. *Numerical data and functional relationships in science and technology: Crystal and solid state physics. Group 3*, volume 7. Springer-Verlag, 1966.

[102] Deng-Bing Li, Sandip S Bista, Rasha A Awni, Sabin Neupane, Abasi Abudulimu, Xiaoming Wang, Kamala K Subedi, Manoj K Jamarkattel, Adam B Phillips, Michael J Heben, et al. 20%-efficient polycrystalline cd (se, te) thin-film solar cells with compositional gradient near the front junction. *Nature Communications*, 13(1):7849, 2022.

[103] Deng-Bing Li, Sandip S Bista, Zhaoning Song, Rasha A Awni, Kamala K Subedi, Niraj Shrestha, Puja Pradhan, Lei Chen, Ebin Bastola, Corey R Grice, et al. Maximize cdte solar cell performance through copper activation engineering. *Nano Energy*, 73:104835, 2020.

[104] Deng-Bing Li, Canglang Yao, SN Vijayaraghavan, Rasha A Awni, Kamala K Subedi, Randy J Ellingson, Lin Li, Yanfa Yan, and Feng Yan. Low-temperature and effective ex situ group v doping for efficient polycrystalline cdsete solar cells. *Nature Energy*, 6(7):715–722, 2021.

[105] Dongxiang Liao and Angus Rockett. Cd doping at the $CuInSe_2$/CdS heterojunction. *Journal of Applied Physics*, 93(11):9380–9382, 2003.

[106] Nidia Gabaldon Limas and Thomas A Manz. Introducing ddec6 atomic population analysis: part 2. computed results for a wide range of periodic and nonperiodic materials. *RSC advances*, 6(51):45727–45747, 2016.

[107] Zhun Liu, Guangren Na, Fuyu Tian, Liping Yu, Jingbo Li, and Lijun Zhang. Computational functionality-driven design of semiconductors for optoelectronic applications. *InfoMat*, 2(5):879–904, 2020.

[108] Joseph J Loferski. Theoretical considerations governing the choice of the optimum semiconductor for photovoltaic solar energy conversion. *Journal of Applied Physics*, 27(7):777–784, 1956.

[109] David C Look. Electrical and optical properties of p-type zno. *Semiconductor Science and Technology*, 20(4):S55, 2005.

[110] Olle Lundberg, Marika Edoff, and Lars Stolt. The effect of Ga-grading in CIGS thin film solar cells. *Thin Solid Films*, 480:520–525, 2005.

[111] Jie Ma, Darius Kuciauskas, David Albin, Raghu Bhattacharya, Matthew Reese, Teresa Barnes, Jian V Li, Timothy Gessert, and Su-Huai Wei. Dependence of the minority-carrier lifetime on the stoichiometry of cdte using time-resolved photoluminescence and first-principles calculations. *Physical review letters*, 111(6):067402, 2013.

[112] Jie Ma, Jihui Yang, Su-Huai Wei, and Juarez LF Da Silva. Correlation between the electronic structures and diffusion paths of interstitial defects in semiconductors: The case of cdte. *Physical Review B*, 90(15):155208, 2014.

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

[114] R Mallick, X Li, C Reich, X Shan, W Zhang, T Nagle, L Bok, E Bicakci, N Rosenblatt, D Modi, et al. Arsenic-doped cdsete solar cells achieve world record 22.3% efficiency. *IEEE Journal of Photovoltaics*, 2023.

[115] Mohammad M Mandurah, Krishna C Saraswat, C Robert Helms, and Theodore I Kamins. Dopant segregation in polycrystalline silicon. *Journal of applied physics*, 51(11):5755–5763, 1980.

[116] Arun Mannodi-Kanakkithodi, Anand Chandrasekaran, Chiho Kim, Tran Doan Huan, Ghanshyam Pilania, Venkatesh Botu, and Rampi Ramprasad. Scoping the polymer genome: A roadmap for rational polymer dielectrics design and beyond. *Materials Today*, 21(7):785–796, 2018.

[117] Arun Mannodi-Kanakkithodi, Ji-Sang Park, Nari Jeon, Duyen H. Cao, David J. Gosztola, Alex B. F. Martinson, and Maria K. Y. Chan. Comprehensive computational study of partial lead substitution in methylammonium lead bromide. *Chemistry of Materials*, 31(10):3599–3612, 2019.

[118] Arun Mannodi-Kanakkithodi, Ji-Sang Park, Alex B. F. Martinson, and Maria K. Y. Chan. Defect energetics in pseudo-cubic mixed halide lead perovskites from first-principles. *The Journal of Physical Chemistry C*, 124(31):16729–16738, 2020.

[119] Arun Mannodi-Kanakkithodi, Michael Y. Toriyama, Fatih G. Sen, Michael J. Davis, Robert F. Klie, and Maria K.Y. Chan. Machine-learned impurity level prediction for semiconductors: the example of cd-based chalcogenides. *npj Computational Materials*, 6(1), 2020.

[120] Arun Mannodi-Kanakkithodi, Xiaofeng Xiang, Laura Jacoby, Robert Biegaj, Scott T Dunham, Daniel R Gamelin, and Maria KY Chan. Universal machine learning framework for defect predictions in zinc blende semiconductors. *Patterns*, 3(3):100450, 2022.

[121] Thomas A Manz and Nidia Gabaldon Limas. Introducing ddec6 atomic population analysis: part 1. charge partitioning theory and methodology. *RSC advances*, 6(53):47771–47801, 2016.

[122] Thomas A Manz and David S Sholl. Improved atoms-in-molecule charge partitioning functional for simultaneously reproducing the electrostatic potential and chemical states in periodic and nonperiodic materials. *Journal of chemical theory and computation*, 8(8):2844–2867, 2012.

[123] Alan H Marshak and David Assaf III. A generalized Einstein relation for semiconductors. *Solid-State Electronics*, 16(6):675–679, 1973.

[124] Brian McCandless, Wayne Buchanan, Gowri Sriramagiri, Christopher Thompson, Joel Duenow, David Albin, Søren A Jensen, John Moseley, M Al-Jassim, and Wyatt K Metzger. Enhanced p-type doping in polycrystalline cdte films: deposition and activation. *IEEE Journal of Photovoltaics*, 9(3):912–917, 2019.

[125] Deborah L. McGott, Christopher P. Muzzillo, Craig L. Perkins, Joseph J. Berry, Kai Zhu, Joel N. Duenow, Eric Colegrove, Colin A. Wolden, and Matthew O. Reese. 3d/2d passivation as a secret to success for polycrystalline thin-film solar cells. *Joule*, 5(5):1057–1073, 2021.

[126] Helmut Mehrer. *Diffusion in solids: fundamentals, methods, materials, diffusion-controlled processes*, volume 155. Springer Science & Business Media, 2007.

[127] E Menéndez-Proupin, M Casanova-Páez, AL Montero-Alejo, Mauricio A Flores, and W Orellana. Symmetry and thermodynamics of tellurium vacancies in cadmium telluride. *Physica B: Condensed Matter*, 568:81–87, 2019.

[128] Eduardo Menéndez-Proupin, A Amézaga, and N Cruz Hernández. Electronic structure of cdte using gga+ usic. *Physica B: Condensed Matter*, 452:119–123, 2014.

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

[130] W. K. Metzger, S. Grover, D. Lu, E. Colegrove, J. Moseley, C. L. Perkins, X. Li, R. Mallick, W. Zhang, R. Malik, J. Kephart, C.-S. Jiang, D. Kuciauskas, D. S. Albin, M. M. Al-Jassim, G. Xiong, and M. Gloeckler. Exceeding 20% efficiency with in situ group v doping in polycrystalline cdte solar cells. *Nature Energy*, 4:837–845, 2019.

[131] Samuel A. Miller, Maxwell Dylla, Shashwat Anand, Kiarash Gordiz, G. Jeffrey Snyder, and Eric S. Toberer. Empirical modeling of dopability in diamond-like semiconductors. *npj Computational Materials*, 4(1):71, 2018.

[132] Rekar O Mohammed and Gavin C Cawley. Over-fitting in model selection with gaussian process regression. In *International Conference on Machine Learning and Data Mining in Pattern Recognition*, pages 192–205. Springer, 2017.

[133] Daniel Mutter and Scott T Dunham. Calculation of defect concentrations and phase stability in Cu_2ZnSnS_4 and $Cu_2ZnSnSe_4$ from stoichiometry. *IEEE Journal of Photovoltaics*, 5(4):1188–1196, 2015.

[134] Akira Nagaoka, Darius Kuciauskas, and Michael A Scarpulla. Doping properties of cadmium-rich arsenic-doped cdte single crystals: Evidence of metastable ax behavior. *Applied Physics Letters*, 111(23), 2017.

[135] Akira Nagaoka, Kensuke Nishioka, Kenji Yoshino, Ryoji Katsube, Yoshitaro Nose, Taizo Masuda, and Michael A Scarpulla. Comparison of sb, as, and p doping in cd-rich cdte single crystals: Doping properties, persistent photoconductivity, and long-term stability. *Applied Physics Letters*, 116(13), 2020.

[136] Akira Nagaoka, Kensuke Nishioka, Kenji Yoshino, Darius Kuciauskas, and Michael A Scarpulla. Arsenic doped cd-rich cdte: equilibrium doping limit and long lifetime for high open-circuit voltage solar cells greater than 900 mv. *Applied Physics Express*, 12(8):081002, 2019.

[137] Pabitra K. Nayak, Suhas Mahesh, Henry J. Snaith, and David Cahen. Photovoltaic solar cell technologies: analysing the state of the art. *Nature Reviews Materials*, 4(4):269–285, 2019.

[138] Ch Neumann, A Nöthe, and NO Lipari. Two-photon magnetoabsorption of znte, cdte, and gaas. *Physical Review B*, 37(2):922, 1988.

[139] Andrew Y. Ng. Feature selection, l1 vs. l2 regularization, and rotational invariance. In *Proceedings of the Twenty-First International Conference on Machine Learning*, ICML '04, page 78, New York, NY, USA, 2004. Association for Computing Machinery.

[140] Shigeru Niki, Miguel Contreras, Ingrid Repins, Michael Powalla, Katsumi Kushiya, Shogo Ishizuka, and Koji Matsubara. CIGS absorbers and processes. *Progress in Photovoltaics: Research and Applications*, 18(6):453–466, 2010.

[141] Cun-Zheng Ning, Letian Dou, and Peidong Yang. Bandgap engineering in semiconductor alloy nanomaterials with widely tunable compositions. *Nature Reviews Materials*, 2(12), 2017.

[142] Takahito Nishimura, Hiroki Sugiura, Kazuyoshi Nakada, and Akira Yamada. Accurate control and characterization of cu depletion layer for highly efficient Cu(In, Ga)Se_2 solar cells. *Progress in Photovoltaics: Research and Applications*, 27(2):171–178, 2019.

[143] PISTORIUS CWFT; CLARK JB; COETZER J; KRUGER GJ; KUNZE OA. High-pressure phase relations and crystal structure determination for zinc phosphide, zn3p2, and cadmium phosphide, cd3p2. *HIGH TEMPER.-HIGH PRESS.; G.B.; DA. 1977; VOL. 9; NO 4; PP. 471-482; BIBL. 23 REF.*, 1977.

[144] Fumiyasu Oba and Yu Kumagai. Design and exploration of semiconductors from first principles: A review of recent advances. *Applied Physics Express*, 11(6):060101, may 2018.

[145] Fumiyasu Oba, Atsushi Togo, Isao Tanaka, Joachim Paier, and Georg Kresse. Defect energetics in zno: A hybrid hartree-fock density functional study. *Physical Review B*, 77(24):245202, 2008.

[146] S Oehling, HJ Lugauer, M Schmitt, H Heinke, U Zehnder, A Waag, CR Becker, and G Landwehr. p-type doping of cdte with a nitrogen plasma source. *Journal of applied physics*, 79(5):2343–2346, 1996.

[147] AA Ojo and IM Dharmadasa. The effect of fluorine doping on the characteristic behaviour of cdte. *Journal of Electronic Materials*, 45:5728–5738, 2016.

[148] Kirill Okhotnikov, Thibault Charpentier, and Sylvian Cadars. Supercell program: a combinatorial structure-generation approach for the local-level modeling of atomic substitutions and partial occupancies in crystals. *Journal of cheminformatics*, 8:1–15, 2016.

[149] Walter Orellana, Eduardo Menéndez-Proupin, and Mauricio A Flores. Self-compensation in chlorine-doped cdte. *Scientific reports*, 9(1):9194, 2019.

[150] Runhai Ouyang, Stefano Curtarolo, Emre Ahmetcik, Matthias Scheffler, and Luca M. Ghiringhelli. Sisso: A compressed-sensing method for identifying the best low-dimensional descriptor in an immensity of offered candidates. *Phys. Rev. Materials*, 2:083802, Aug 2018.

[151] Vasilios Palekis, Wei Wang, Sheikh Tawsif Elahi, Md Zahangir Alom, and Chris Ferekides. Thin film solar cells with n-type cdte absorber and p-type znte window layers. In *2021 IEEE 48th Photovoltaic Specialists Conference (PVSC)*, pages 1293–1297. IEEE, 2021.

[152] Jie Pan, Wyatt K Metzger, and Stephan Lany. Spin-orbit coupling effects on predicting defect properties with hybrid functionals: A case study in cdte. *Physical Review B*, 98(5):054108, 2018.

[153] Ji Sang Park, Sunghyun Kim, Zijuan Xie, and Aron Walsh. Point defect engineering in thin-film solar cells. *Nat. Rev. Mater.*, 3(7):194–210, June 2018.

[154] P D BIRKMIRERW SHAFARMANWN Paulson, RW Birkmire, and WN Shafarman. Optical characterization of $CuIn_{1-x}Ga_xSe_2$ alloy thin films by spectroscopic ellipsometry. *Journal of Applied Physics*, 94(2):879–888, 2003.

[155] 1J Perrenoud, L Kranz, C Gretener, F Pianezzi, S Nishiwaki, S Buecheler, and AN Tiwari. A comprehensive picture of cu doping in cdte solar cells. *Journal of Applied Physics*, 114(17), 2013.

[156] A Pietraszko and K Łukaszewicz. A refinement of the crystal structure of α"-cd3as2. *Acta Crystallographica Section B: Structural Crystallography and Crystal Chemistry*, 25(5):988–990, 1969.

[157] Johan Pohl and Karsten Albe. Intrinsic point defects in $CuInSe_2$ and $CuGaSe_2$ as seen via screened-exchange hybrid density functional theory. *Physical Review B*, 87(24):245203, 2013.

[158] Maciej P. Polak, Ryan Jacobs, Arun Mannodi-Kanakkithodi, Maria K. Y. Chan, and Dane Morgan. Machine learning for impurity charge-state transition levels in semiconductors from elemental properties using multi-fidelity datasets. *The Journal of Chemical Physics*, 156(11):114110, 03 2022.

[159] Jonathan D Poplawsky, Wei Guo, Naba Paudel, Amy Ng, Karren More, Donovan Leonard, and Yanfa Yan. Structural and compositional dependence of the cdte x se1- x alloy layer photoactivity in cdte-based solar cells. *Nature communications*, 7(1):12537, 2016.

[160] R Prasad, AK Das, and Udai P Singh. Bilayer cigs-based solar cell device for enhanced performance: a numerical approach. *Applied Physics A*, 127(2):140, 2021.

[161] Jorge López Puga, Martin Krzywinski, and Naomi Altman. Points of significance: Bayesian statistics. *Nature Methods*, 12(5):377–378, 2015.

[162] Hans J. Queisser and Eugene E. Haller. Defects in semiconductors: Some fatal, some vital. *Science*, 281(5379):945–950, 1998.

[163] Prajit Ramachandran, Barret Zoph, and Quoc V. Le. Searching for activation functions, 2018.

[164] CR Rao and Y Wu. Linear model selection by cross-validation. *Journal of Statistical Planning and Inference*, 128(1):231–240, 2005.

[165] Carey Reich. *Investigations to Improve CdTe-Based Solar Cell Open Circuit Voltage and Efficiency Using a Passivation and Selectivity Theoretical Framework*. PhD thesis, Colorado State University, 2022.

[166] Patrick Reinhard, Adrian Chirilă, Patrick Blösch, Fabian Pianezzi, Shiro Nishiwaki, Stephan Buechelers, and Ayodhya N Tiwari. Review of progress toward 20% efficiency flexible CIGS solar cells and manufacturing issues of solar modules. In *2012 IEEE 38th Photovoltaic Specialists Conference (PVSC) PART 2*, pages 1–9. IEEE, 2012.

[167] U Reislöhner, J Grillenberger, and W Witthuhn. Band-gap level of the cadmium vacancy in cdte. *Journal of crystal growth*, 184:1160–1164, 1998.

[168] Adachi Sadao. Properties of group-iv, iii-v and ii-vi semiconductors. *Hoboken (USA) Wiley&Sons*, 2005.

[169] M Said and MA Kanehisa. Excited states of acceptors in cdte and znte. *Journal of crystal growth*, 101(1-4):488–492, 1990.

[170] K Sakurai, R Scheer, S Nakamura, Y Kimura, T Baba, CA Kaufmann, A Neisser, S Ishizuka, A Yamada, K Matsubara, et al. Structural changes of CIGS during deposition investigated by spectroscopic light scattering: A study on Ga concentration and Se pressure. *Solar energy materials and solar cells*, 90(18-19):3377–3384, 2006.

[171] Pedro MP Salomé, Viktor Fjällström, Piotr Szaniawski, Joaquim P Leitao, Adam Hultqvist, Paulo A Fernandes, Jennifer P Teixeira, Bruno P Falcao, Uwe Zimmermann, Antonio F da Cunha, et al. A comparison between thin film solar cells made from co-evaporated $CuIn_{1-x}Ga_xSe_2$ using a one-stage process versus a three-stage process. *Progress in Photovoltaics: Research and Applications*, 23(4):470–478, 2015.

[172] M. D. Sampson, J. S. Park, R. D. Schaller, M. K. Y. Chan, and A. B. F. Martinson. Transition metal-substituted lead halide perovskite absorbers. *J. Mater. Chem. A*, 5:3578–3588, 2017.

[173] Michael A Scarpulla, Brian McCandless, Adam B Phillips, Yanfa Yan, Michael J Heben, Colin Wolden, Gang Xiong, Wyatt K Metzger, Dan Mao, Dmitry Krasikov, et al. Cdte-based thin film photovoltaics: Recent advances, current challenges and future prospects. *Solar Energy Materials and Solar Cells*, 255:112289, 2023.

[174] Michael A. Scarpulla, Brian McCandless, Adam B. Phillips, Yanfa Yan, Michael J. Heben, Colin Wolden, Gang Xiong, Wyatt K. Metzger, Dan Mao, Dmitry Krasikov, Igor Sankin, Sachit Grover, Amit Munshi, Walajabad Sampath, James R. Sites, Alexandra Bothwell, David Albin, Matthew O. Reese, Alessandro Romeo, Marco Nardone, Robert Klie, J. Michael Walls, Thomas Fiducia, Ali Abbas, and Sarah M. Hayes. Cdte-based thin film photovoltaics: Recent advances, current challenges and future prospects. *Solar Energy Materials and Solar Cells*, 255:112289, 2023.

[175] M Schenk and C Silber. Lattice parameter and microhardness investigations on cdte1-xsex. *Journal of Materials Science: Materials in Electronics*, 9(4):295–300, 1998.

[176] Karlheinz Seeger. *Semiconductor physics*. Springer Science & Business Media, 2013.

[177] MATTHIAS SEEGER. Gaussian processes for machine learning. *International Journal of Neural Systems*, 11(2):69–106, 2004.

[178] Selva Chandrasekaran Selvaraj, Sameer Gupta, Damien Caliste, and Pascal Pochet. Passivation mechanism in cdte solar cells: The hybrid role of se. *Applied Physics Letters*, 119(6), 2021.

[179] Ebru Şenadim, Hamide Kavak, and Ramazan Esen. The effect of annealing on structural and optical properties of ZnO thin films grown by pulsed filtered cathodic vacuum arc deposition. *Journal of Physics: Condensed Matter*, 18(27):6391, 2006.

[180] Abhishek Sharan, Marco Nardone, Dmitry Krasikov, Nirpendra Singh, and Stephan Lany. Atomically thin interlayer phase from first principles enables defect-free incommensurate sno2/cdte interface. *Applied Physics Reviews*, 9(4), 2022.

[181] Vinit Sharma, Pankaj Kumar, Pratibha Dev, and Ghanshyam Pilania. Machine learning substitutional defect formation energies in abo3 perovskites. *Journal of Applied Physics*, 128(3):034902, 2020.

[182] Anna Shepidchenko, Biplab Sanyal, Mattias Klintenberg, and Susanne Mirbt. Small hole polaron in cdte: Cd-vacancy revisited. *Scientific reports*, 5(1):14509, 2015.

[183] S Shokhovets, G Gobsch, and O Ambacher. Conduction band parameters of zno. *Superlattices and Microstructures*, 39(1-4):299–305, 2006.

[184] Vallisree Sivathanu, Thangavel R, and Trupti Ranjan Lenka. Modeling and performance optimization of two-terminal cu2znsns4–silicon tandem solar cells. *International Journal of Energy Research*, n/a(n/a):1–11, 2021.

[185] David E Sommer and Scott T Dunham. Atomistic models of Cu diffusion in $CuInSe_2$ under variations in composition. *Journal of Applied Physics*, 123(11):115116, 2018.

[186] David E Sommer and Scott T Dunham. Understanding copper diffusion in $CuInSe_2$ with first-principles based atomistic and continuum models. *Journal of Applied Physics*, 130(23):235701, 2021.

[187] David E Sommer, Daniel Mutter, and Scott T Dunham. Defects in Na-, K-, and Cd-Doped $CuInSe_2$: Canonical thermodynamics based on ab initio calculations. *IEEE Journal of Photovoltaics*, 7(4):1143–1152, 2017.

[188] Conrad Spindler, Finn Babbe, Max Hilaire Wolter, Florian Ehré, Korra Santhosh, Pit Hilgert, Florian Werner, and Susanne Siebentritt. Electronic defects in Cu(In, Ga)Se_2: Towards a comprehensive model. *Physical Review Materials*, 3(9):090302, 2019.

[189] Adam Stokes, Mowafak Al-Jassim, David R Diercks, Brian Egaas, and Brian Gorman. 3-d point defect density distributions in thin film cu (in, ga) se2 measured by atom probe tomography. *Acta Materialia*, 102:32–37, 2016.

[190] Adam Stokes, Mowafak Al-Jassim, Andrew Norman, David Diercks, and Brian Gorman. Nanoscale insight into the p-n junction of alkali-incorporated Cu(In, Ga)Se_2 solar cells. *Progress in Photovoltaics: Research and Applications*, 25(9):764–772, 2017.

[191] AM Stoneham. Non-radiative transitions in semiconductors. *Rep. Prog. Phys.*, 44(12):1251, 1981.

[192] Synopsys Inc. Sentaurus device user guide. Version K-2023.12, Synopsys, Inc., Mountain View, 2023.

[193] Synopsys Inc. Sentaurus process user guide. Version V-2023.12, Synopsys, Inc., Mountain View, 2023.

[194] Yijun Tong, Xiaofeng Xiang, and Scott T Dunham. Overcoming the limitations of first-principles defect calculations in inorganic lead/tin halide perovskites. *(unpublished)*, 2024.

[195] Marko Topič, Franc Smole, and Jože Furlan. Band-gap engineering in CdS/Cu(In, Ga)Se_2 solar cells. *Journal of applied physics*, 79(11):8537–8540, 1996.

[196] Torres-Jaramillo, Santiago, Bernal-Correa, Roberto, and Morales-Acevedo, Arturo. Improved design of ingap/gaas//si tandem solar cells. *EPJ Photovolt.*, 12:1, 2021.

[197] RE Treharne, A Seymour-Pierce, K Durose, K Hutchings, S Roncallo, and D Lane. Optical design and fabrication of fully sputtered CdTe/CdS solar cells. In *Journal of Physics: Conference Series*, volume 286, page 012038. IOP Publishing, 2011.

[198] Mark E Turiansky, Audrius Alkauskas, Manuel Engel, Georg Kresse, Darshana Wickramaratne, Jimmy-Xuan Shen, Cyrus E Dreyer, and Chris G Van de Walle. Nonrad: Computing nonradiative capture coefficients from first principles. *Computer Physics Communications*, 267:108056, 2021.

[199] Joel B. Varley, Amit Samanta, and Vincenzo Lordi. Descriptor-based approach for the prediction of cation vacancy formation energies and transition levels. *The Journal of Physical Chemistry Letters*, 8(20):5059–5063, 2017.

[200] G. Varoquaux, L. Buitinck, G. Louppe, O. Grisel, F. Pedregosa, and A. Mueller. Scikit-learn: Machine learning without learning the machinery. *GetMobile: Mobile Comp. and Comm.*, 19(1):29–33, June 2015.

[201] Rama K. Vasudevan, Kamal Choudhary, Apurva Mehta, Ryan Smith, Gilad Kusne, Francesca Tavazza, Lukas Vlcek, Maxim Ziatdinov, Sergei V. Kalinin, and Jason Hattrick-Simpers. Materials science in the artificial intelligence age: high-throughput library generation, machine learning, and a pathway from correlations to the underpinning physics. *MRS Communications*, 9(3):821–838, 2019.

[202] Sahitya V Vegesna, Vinayak J Bhat, Danilo Bürger, Jan Dellith, Ilona Skorupa, Oliver G Schmidt, and Heidemarie Schmidt. Increased static dielectric constant in znmno and zncoo thin films with bound magnetic polarons. *Scientific Reports*, 10(1):6698, 2020.

[203] Zhongyu Wan, Quan-De Wang, Dongchang Liu, and Jinhu Liang. Data-driven machine learning model for the prediction of oxygen vacancy formation energy of metal oxide materials. *Phys. Chem. Chem. Phys.*, 23:15675–15684, 2021.

[204] Guohong Wang, Xiaoyan Yi, Teng Zhan, and Yang Huang. The algainp/algaas material system and red/yellow led. In *Light-Emitting Diodes*, pages 171–202. Springer, 2019.

[205] Jing Wang, Hangyan Chen, Su-Huai Wei, and Wan-Jian Yin. Materials design of solar cell absorbers beyond perovskites and conventional semiconductors via combining tetrahedral and octahedral coordination. *Advanced Materials*, 31(17):1806593, 2019.

[206] Shenghao Wang, Xia Hao, Muhammad Monirul Islam, Takuya Kato, Hiroki Sugimoto, Katsuhiro Akimoto, and Takeaki Sakurai. Influence of potassium treatment on electronic properties of $Cu(In_{1-x}Ga_x)(Se_{1-y}S_y)_2$ solar cells studied by steady state photo-capacitance and admittance spectroscopy. *Japanese Journal of Applied Physics*, 57(8S3):08RC13, 2018.

[207] Logan Ward, Alexander Dunn, Alireza Faghaninia, Nils E.R. Zimmermann, Saurabh Bajaj, Qi Wang, Joseph Montoya, Jiming Chen, Kyle Bystrom, Maxwell Dylla, Kyle Chard, Mark Asta, Kristin A. Persson, G. Jeffrey Snyder, Ian Foster, and Anubhav Jain. Matminer: An open source toolkit for materials data mining. *Computational Materials Science*, 152:60–69, 2018.

[208] Su-Huai Wei and SB Zhang. Chemical trends of defect formation and doping limit in ii-vi semiconductors: The case of cdte. *Physical Review B*, 66(15):155211, 2002.

[209] Su-Huai Wei and Alex Zunger. Band offsets and optical bowings of chalcopyrites and Zn-based II-VI alloys. *Journal of Applied Physics*, 78(6):3846–3856, 1995.

[210] Thomas P Weiss, Benjamin Bissig, Thomas Feurer, Romain Carron, Stephan Buecheler, and Ayodhya N Tiwari. Bulk and surface recombination properties in thin film semiconductors with different surface treatments from time-resolved photoluminescence measurements. *Scientific reports*, 9(1):1–13, 2019.

[211] Thomas Paul Weiss, Romain Carron, Max H Wolter, Johannes Löckinger, Enrico Avancini, Susanne Siebentritt, Stephan Buecheler, and Ayodhya N Tiwari. Time-resolved photoluminescence on double graded Cu(In, Ga)Se_2–impact of front surface recombination and its temperature dependence. *Science and technology of advanced materials*, 20(1):313–323, 2019.

[212] Wolfgang SM Werner, Kathrin Glantschnig, and Claudia Ambrosch-Draxl. Optical constants and inelastic electron-scattering data for 17 elemental metals. *Journal of Physical and Chemical Reference Data*, 38(4):1013–1092, 2009.

[213] Christopher KI Williams and Carl Edward Rasmussen. *Gaussian processes for machine learning*, volume 2. MIT press Cambridge, MA, 2006.

[214] Matthew D Witman, Anuj Goyal, Tadashi Ogitsu, Anthony H McDaniel, and Stephan Lany. Defect graph neural networks for materials discovery in high-temperature clean-energy applications. *Nature Computational Science*, pages 1–12, 2023.

[215] Yelong Wu, Guangde Chen, Youzhang Zhu, Wan-Jian Yin, Yanfa Yan, Mowafak Al-Jassim, and Stephen J Pennycook. Lda+ u/gga+ u calculations of structural and electronic properties of cdte: Dependence on the effective u parameter. *Computational Materials Science*, 98:18–23, 2015.

[216] Xiaofeng Xiang, Aaron Gehrke, and Scott Dunham. Understanding the dopability of as in selenium-alloyed cadmium telluride solar cells. In *2023 IEEE 50th Photovoltaic Specialists Conference (PVSC)*, pages 1–3. IEEE, 2023.

[217] Xiaofeng Xiang, Dylan Soh, and Scott Dunham. Exploration of deep learning models for accelerated defect property predictions and device design of cubic semiconductor crystals. *The Journal of Physical Chemistry C*, 2024.

[218] Xiaofeng Xiang, David E Sommer, Aaron Gehrke, and Scott T Dunham. Coupled process and device modeling of Cu(In, Ga)Se_2 solar cells. In *2021 IEEE 48th Photovoltaic Specialists Conference (PVSC)*, pages 1707–1711. IEEE, 2021.

[219] Xiaofeng Xiang, David E Sommer, Aaron Gehrke, and Scott T Dunham. Coupled process/device modeling and point defect engineering of cu (in, ga) se 2 solar cells. *IEEE Journal of Photovoltaics*, 2024.

[220] Xiaofeng Xiang, Yijun Tong, Aaron Gehrke, and Scott Dunham. Point defects in cdte and cdtese alloy: a first principles investigation with dft+u. *arXiv*, 04 2024.

[221] Tian Xie and Jeffrey C Grossman. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. *Physical review letters*, 120(14):145301, 2018.

[222] Qing-Song Xu and Yi-Zeng Liang. Monte carlo cross validation. *Chemometrics and Intelligent Laboratory Systems*, 56(1):1–11, 2001.

[223] Ji-Hui Yang, Ji-Sang Park, Joongoo Kang, Wyatt Metzger, Teresa Barnes, and Su-Huai Wei. Tuning the fermi level beyond the equilibrium doping limit through quenching: The case of cdte. *Physical Review B*, 90(24):245202, 2014.

[224] Ji-Hui Yang, Lin Shi, Lin-Wang Wang, and Su-Huai Wei. Non-radiative carrier recombination enhanced by two-level process: a first-principles study. *Scientific reports*, 6(1):1–10, 2016.

[225] Ji-Hui Yang, Wan-Jian Yin, Ji-Sang Park, James Burst, Wyatt K Metzger, Tim Gessert, Teresa Barnes, and Su-Huai Wei. Enhanced p-type dopability of p and as in cdte using non-equilibrium thermal processing. *Journal of Applied Physics*, 118(2):025102, 2015.

[226] Ji-Hui Yang, Wan-Jian Yin, Ji-Sang Park, Jie Ma, and Su-Huai Wei. Review on first-principles study of defect properties of cdte as a solar cell absorber. *Semiconductor Science and Technology*, 31(8):083002, 2016.

[227] Ji-Hui Yang, Wan-Jian Yin, Ji-Sang Park, Wyatt Metzger, and Su-Huai Wei. First-principles study of roles of cu and cl in polycrystalline cdte. *Journal of Applied Physics*, 119(4), 2016.

[228] Shih-Chi Yang, Mario Ochoa, Ramis Hertwig, Abdessalem Aribia, Ayodhya N Tiwari, and Romain Carron. Influence of Ga back grading on voltage loss in low-temperature co-evaporated Cu(In, Ga)Se_2 thin film solar cells. *Progress in Photovoltaics: Research and Applications*, 29(6):630–637, 2021.

[229] Wan-Jian Yin, Tingting Shi, and Yanfa Yan. Unusual defect physics in $CH_3NH_3PbI_3$ perovskite solar cell absorber. *Applied Physics Letters*, 104(6):063903, 2014.

[230] Lixin Zhang, Juarez LF Da Silva, Jingbo Li, Yanfa Yan, TA Gessert, and Su-Huai Wei. Effect of copassivation of cl and cu on cdte grain boundaries. *Physical review letters*, 101(15):155501, 2008.

[231] Jian H Zhao, TE Schlesinger, and AG Milnes. Determination of carrier capture cross sections of traps by deep level transient spectroscopy of semiconductors. *Journal of applied physics*, 62(7):2865–2870, 1987.

[232] Yuan Zhao, Mathieu Boccard, Shi Liu, Jacob Becker, Xin-Hao Zhao, Calli M Campbell, Ernesto Suarez, Maxwell B Lassise, Zachary Holman, and Yong-Hang Zhang. Monocrystalline cdte solar cells with open-circuit voltage over 1 v and efficiency of 17%. *Nature Energy*, 1(6):1–7, 2016.

Appendix A

SDEVICE EXAMPLE (CIGS SOLAR CELLS)

```
##########################################################################################
Electrode{
   { Name="Base"      Voltage= 0.0}
   { Name="Emitter"       Voltage= 0.0}
}
##########################################################################################
File{
#-Predefined paranmeters
   Grid     = "@tdr@"
   Plot     = "@tdrdat@"
   Current = "@plot@"
   Output  = "@log@"
   Param    = "@parameter@"
   DevFields="@tdr@"
   PMIUserFields = "@tdr@"
   IlluminationSpectrum= "am15g_1.2um_equi-10nm.txt" #AM1.5 G Spectrum
}
##########################################################################################
Physics (MaterialInterface = "Molybdenum/CuInGaSe")
{
 Recombination( surfaceSRH )
}

Physics (MaterialInterface = "CuInGaSe/CdS")
{
 Recombination( surfaceSRH )
}

Physics (Region="CIGS_Region"){
  IncompleteIonization(Dopants = "VcuConcentration IncuCuthConcentration GacuCuthConcentration")
  Traps (
#Gacu
      (Donor fromValBand Sfactor="PMIUserField6" EnergyMid=0.73335 eXsection=@exsection@ hXsection=@hxsection@)

#CuIII
      (Acceptor fromValBand Sfactor="PMIUserField10" EnergyMid=0.3801 eXsection=@exsection@ hXsection=@hxsection@)
  )
}

Physics{
   HeteroInterface
   Fermi
   Mobility( ConstantMobility )
```

```
   EffectiveIntrinsicDensity (NoBandGapNarrowing)
   Recombination(Radiative ,SurfaceSRH)
   Optics(
        ComplexRefractiveIndex( WavelengthDep(real imag) )
        OpticalGeneration (
                        QuantumYield(Stepfunction(Bandgap))
                    ComputeFromSpectrum (
            )
                        )
        Excitation (
                Theta= 180
                Window ("L1") (
                        Origin = (0,4.5)
                        OriginAnchor = Center
                        Line ( Dx = @Width@ )
                )
        )
        OpticalSolver (
                TMM (
                        PropagationDirection = Perpendicular
                        NodesPerWavelength = 50
                        LayerStackExtraction (
                                WindowName = "L1"
                        )
                )
        ) * end of OpticalSolver
  )  * end of Optics
}

##############################################################################################################

Plot{
xMoleFraction
yMoleFraction
BandGap
eMobility
hMobility
LatticeTemperature
hEffectiveStateDensity
eEffectiveStateDensity
IntrinsicDensity
DonorPlusConcentration
AccepMinusConcentration

#—OpticalGeneration:

  AbsorbedPhotonDensity
  OpticalGeneration
  TotalConcentration
  DopingConcentration
  ComplexRefractiveIndex
  OpticalAbsorptionHeat
```

```
#—Density and Currents, etc:

   eDensity hDensity
   TotalCurrent eCurrent hCurrent
   eMobility hMobility
   eVelocity hVelocity
   eQuasiFermi hQuasiFermi
   eTrappedCharge hTrappedCharge
   eGapStatesRecombination hGapStatesRecombination
   DeepLevels

#—Fields and charges:

   ElectricField/Vector Potential SpaceCharge
   ElectricField

#—Doping Profiles:

   Doping DonorConcentration AcceptorConcentration

#—Driving forces:

   eGradQuasiFermi/Vector hGradQuasiFermi/Vector
   eEparallel hEparallel eENormal hENormal

#—Band structure/Composition:

   BandGap
   BandGapNarrowing
   ConductionBand ValenceBand

#—Recombination:

   TotalRecombination
   SRHRecombination
   AugerRecombination
   SurfaceRecombination

}

################################################################################################################

Math {
   Extrapolate
   Derivatives
   RelErrControl
   Notdamped=100
   Iterations=40
   SubMethod=ParDiSo
   Method=Blocked
   Stacksize= 200000000
   Traps(Damping=100)
   Number_of_Threads = maximum
```

```
}

####################################################################################################

Solve {
   Coupled(Iterations=100){ Poisson}
   Coupled{ Poisson Electron Hole}

   Quasistationary(
     InitialStep=1e-3 Increment=1.2
      Minstep=1e-13 MaxStep=0.01
      Goal{ Name="Base" Voltage= 1.0}
   ){ Coupled{ Poisson Electron Hole} }
}
```

Appendix B

ML-ASSISTED DOPANT SCREENING FOR DEVICE DESIGN (EXAMPLE OF CDTE SOLAR CELLS)

B.1 Device Simulation Details

Sentaurus TCAD is used for device simulation [192]. The model consists of three semiconductor layers, CdTe, CdS and ZnO. Critical parameters of these layers are listed in Table B.1. Mo is used as back contact. The back surface recombination velocity is set to 10^5 cm/s. The Transfer Matrix Method (TMM) is used in Sentaurus Device to calculate the propagation of plane waves through layered media under the AM 1.5 Global spectrum with an integrated power of 1000 W/m^2. There are two deep level defects considered in CdTe, V_{Cd} and Te_{Cd}. We use NONRAD [5, 198] to compute carrier capture for both two defects. The hole/electron carrier capture cross sections found for V_{Cd} (-1/0) are 8.66 x 10^{-16} and 1.34 x 10^{-16}, respectively. The hole/electron carrier capture cross sections for Te_{Cd} (+1/0) are 1.19 x 10^{-15} and 3.73 x 10^{-15}, respectively. These capture cross sections are then implemented into SRH recombination model in device simulation to simulate defect influence on carrier lifetime.

Table B.1: Critical parameters in CdTe Device Model

	CdTe	CdS	ZnO
Thickness (nm)	4000	50	50
Bandgap (eV)	1.50	2.4	3.6
Carrier Density (cm^{-3})	varied	1.1 x 10^{18}	10^{18}
Dielectric Constant	10.2 [138]	8.4 [113]	9.0 [202]
Effective mass e/h	0.095/0.84 [226]	0.25/0.70 [113]	0.24/0.59 [183]
Mobility e/h (cm^2/V s)	320/80 [76]	100/25 [113]	200/25 [109]
Lifetime e/h (ns)	Defect-assisted SRH Model	0.01/0.1	0.1/0.1

Table B.2: ML-assisted Dopant Screening for Device Design (Example of CdTe Solar Cells, Full Table)

Dopant	Condition	Dominant Defects	Deep Level?	Type	Fermi Level (eV)	Carrier Density @ 300K (cm^{-3})
Li	Cd-rich	Li_Cd, Li_int_A, Li_int_B, Li_int_neut	Yes	p-type	0.37	9.85e+12
Li	Te-rich	Li_Cd, Li_int_A	Yes	p-type	0.12	1.64e+17
Na	Cd-rich	Na_Cd, Na_int_A, Na_int_B	Yes	n-type	1.18	-5.20e+12
Na	Te-rich	Na_Cd, Na_int_A	Yes	p-type	0.28	3.86e+14
K	Cd-rich	K_int_A, K_Cd, K_int_B	Yes	n-type	1.34	-2.30e+15
K	Te-rich	K_Cd	Yes	p-type	0.19	1.47e+16
Rb	Cd-rich	Rb_int_A, Rb_Cd	Yes	n-type	1.27	-1.40e+14
Rb	Te-rich	Rb_Cd	Yes	p-type	0.26	8.11e+14
Cs	Cd-rich	Cs_Cd	Yes	n-type	1.27	-1.39e+14
Cs	Te-rich	Cs_Cd	Yes	p-type	0.38	6.65e+12
Be	Cd-rich	Be_Cd	No	n-type	1.20	-8.42e+12
Be	Te-rich	Be_Cd	No	p-type	0.62	6.17e+08
Mg	Cd-rich	Mg_Cd	No	n-type	1.26	-1.15e+14
Mg	Te-rich	Mg_Cd	No	p-type	0.56	7.92e+09
Ca	Cd-rich	Ca_Cd	No	n-type	1.27	-1.40e+14
Ca	Te-rich	Ca_Cd	No	p-type	0.50	6.75e+10
Sr	Cd-rich	Sr_Cd	No	n-type	1.27	-1.39e+14
Sr	Te-rich	Sr_Cd	No	p-type	0.42	1.58e+12
Ba	Cd-rich	Ba_Cd	No	n-type	1.27	-1.39e+14
Ba	Te-rich	Ba_Cd	No	p-type	0.36	1.67e+13
B	Cd-rich	B_Cd, B_Te	Yes	n-type	1.27	-1.39e+14
B	Te-rich	B_Cd	Yes	p-type	0.32	7.53e+13
Al	Cd-rich	Al_Cd	Yes	n-type	1.31	-5.94e+14
Al	Te-rich	Al_Cd	Yes	n-type	1.28	-2.20e+14
Ga	Cd-rich	Ga_Cd	Yes	n-type	1.25	-6.07e+13
Ga	Te-rich	Ga_Cd	Yes	n-type	1.19	-6.67e+12
In	Cd-rich	In_Cd	Yes	n-type	1.27	-1.34e+14
In	Te-rich	In_Cd	Yes	n-type	1.12	-3.99e+11
Tl	Cd-rich	Tl_Cd, Tl_int_A	Yes	n-type	1.25	-7.96e+13
Tl	Te-rich	Tl_Cd	Yes	n-type	0.92	-1.86e+08
C	Cd-rich	C_Te	Yes	n-type	1.27	-1.39e+14
C	Te-rich	C_Cd, C_Te	Yes	p-type	0.32	7.53e+13
Si	Cd-rich	Si_Te	Yes	n-type	1.26	-1.09e+14

Si	Te-rich	Si_Cd	Yes	p-type	0.65	1.96e+08
Ge	Cd-rich	Ge_Cd, Ge_Te	Yes	n-type	1.26	-8.49e+13
Ge	Te-rich	Ge_Cd	Yes	p-type	0.55	1.03e+10
Sn	Cd-rich	Sn_Cd	Yes	n-type	1.27	-1.38e+14
Sn	Te-rich	Sn_Cd	Yes	p-type	0.73	9.86e+06
Pb	Cd-rich	Pb_Cd	Yes	n-type	1.27	-1.38e+14
Pb	Te-rich	Pb_Cd	Yes	p-type	0.65	2.78e+08
N	Cd-rich	N_Te	No	p-type	0.25	1.16e+15
N	Te-rich	N_Te	No	p-type	0.32	7.53e+13
P	Cd-rich	P_Te	No	p-type	0.21	4.82e+15
P	Te-rich	P_Te	No	p-type	0.20	9.70e+15
As	Cd-rich	As_Te	No	p-type	0.23	2.22e+15
As	Te-rich	As_Te	No	p-type	0.24	1.61e+15
Sb	Cd-rich	Sb_Te	No	n-type	1.27	-1.35e+14
Sb	Te-rich	Sb_Cd	No	p-type	0.32	7.53e+13
Bi	Cd-rich	Bi_Te	Yes	p-type	0.39	6.27e+12
Bi	Te-rich	Bi_Cd	No	n-type	1.27	-1.36e+14
O	Cd-rich	O_Te	Yes	n-type	0.87	-3.38e+07
O	Te-rich	O_Te, O_int_A	Yes	p-type	0.32	7.54e+13
S	Cd-rich	S_Te	No	n-type	1.01	-5.67e+09
S	Te-rich	S_Te	No	p-type	0.44	7.12e+11
Se	Cd-rich	Se_Te	No	n-type	0.94	-4.16e+08
Se	Te-rich	Se_Te	No	p-type	0.50	6.44e+10
F	Cd-rich	F_int_A, F_Te	Yes	p-type	0.76	3.33e+06
F	Te-rich	F_int_A	No	p-type	0.12	1.98e+17
Cl	Cd-rich	Cl_Te	Yes	n-type	1.33	-1.42e+15
Cl	Te-rich	Cl_Te, Cl_int_A	Yes	p-type	0.54	1.65e+10
Br	Cd-rich	Br_Te	Yes	n-type	1.43	-7.17e+16
Br	Te-rich	Br_Te	Yes	n-type	1.20	-9.46e+12
I	Cd-rich	I_Te	Yes	n-type	1.46	-2.26e+17
I	Te-rich	I_Te	Yes	n-type	1.24	-4.66e+13
Sc	Cd-rich	Sc_Cd	Yes	n-type	1.29	-3.52e+14
Sc	Te-rich	Sc_Cd	Yes	n-type	1.25	-6.05e+13
Ti	Cd-rich	Ti_Cd	Yes	n-type	1.27	-1.38e+14
Ti	Te-rich	Ti_Cd	Yes	n-type	0.81	-3.36e+06
V	Cd-rich	V_Cd	Yes	n-type	1.27	-1.33e+14
V	Te-rich	V_Cd	Yes	n-type	0.89	-5.30e+07
Cr	Cd-rich	Cr_Cd	Yes	n-type	0.82	-4.42e+06
Cr	Te-rich	Cr_Cd	Yes	p-type	0.50	6.47e+10

Mn	Cd-rich	Mn_Cd	No	n-type	1.22	-1.89e+13
Mn	Te-rich	Mn_Cd	No	p-type	0.71	2.49e+07
Fe	Cd-rich	Fe_Cd	Yes	n-type	0.93	-3.43e+08
Fe	Te-rich	Fe_Cd	Yes	p-type	0.62	8.50e+08
Co	Cd-rich	Co_Cd	Yes	n-type	0.91	-1.21e+08
Co	Te-rich	Co_Cd	Yes	p-type	0.51	5.51e+10
Ni	Cd-rich	Ni_Cd, Ni_int_neut	Yes	n-type	1.27	-1.39e+14
Ni	Te-rich	Ni_Cd	Yes	p-type	0.45	4.62e+11
Cu	Cd-rich	Cu_Cd	No	n-type	1.26	-9.97e+13
Cu	Te-rich	Cu_Cd	No	p-type	0.14	8.37e+16
Zn	Cd-rich	Zn_Cd	No	n-type	1.22	-2.32e+13
Zn	Te-rich	Zn_Cd	No	p-type	0.61	9.83e+08
Y	Cd-rich	Y_Cd	No	n-type	1.38	-1.11e+16
Y	Te-rich	Y_Cd	No	n-type	1.37	-6.09e+15
Zr	Cd-rich	Zr_Cd	Yes	n-type	1.27	-1.39e+14
Zr	Te-rich	Zr_Cd	Yes	n-type	1.06	-4.19e+10
Nb	Cd-rich	Nb_Cd	Yes	n-type	1.27	-1.39e+14
Nb	Te-rich	Nb_Cd	Yes	p-type	0.37	1.14e+13
Mo	Cd-rich	Mo_Cd	Yes	n-type	1.27	-1.39e+14
Mo	Te-rich	Mo_Cd	Yes	p-type	0.32	7.53e+13
Tc	Cd-rich	Tc_Cd	Yes	n-type	1.27	-1.39e+14
Tc	Te-rich	Tc_Cd	Yes	p-type	0.32	7.53e+13
Ru	Cd-rich	Ru_Cd	Yes	n-type	1.27	-1.39e+14
Ru	Te-rich	Ru_Cd	Yes	p-type	0.32	7.53e+13
Rh	Cd-rich	Rh_Cd	Yes	n-type	1.27	-1.39e+14
Rh	Te-rich	Rh_Cd	Yes	p-type	0.48	1.45e+11
Pd	Cd-rich	Pd_int_neut, Pd_int_B	No	n-type	1.27	-1.39e+14
Pd	Te-rich	Pd_Cd	Yes	p-type	0.33	4.75e+13
Ag	Cd-rich	Ag_Cd, Ag_int_B	Yes	p-type	0.25	1.15e+15
Ag	Te-rich	Ag_Cd, Ag_int_B, Ag_int_neut	Yes	p-type	0.12	1.77e+17
La	Cd-rich	La_Cd	No	n-type	1.27	-1.75e+14
La	Te-rich	La_Cd	No	n-type	1.26	-9.90e+13
Hf	Cd-rich	Hf_Cd	Yes	n-type	1.27	-1.32e+14
Hf	Te-rich	Hf_Cd	Yes	n-type	1.18	-4.86e+12
Ta	Cd-rich	Ta_Cd	Yes	n-type	1.27	-1.39e+14
Ta	Te-rich	Ta_Cd	Yes	p-type	0.32	7.51e+13
W	Cd-rich	W_Cd	Yes	n-type	1.27	-1.39e+14
W	Te-rich	W_Cd	Yes	p-type	0.32	7.53e+13
Re	Cd-rich	Re_Cd	Yes	n-type	1.27	-1.39e+14

Re	Te-rich	Re_Cd	Yes	p-type	0.32	7.53e+13
Os	Cd-rich	Os_Cd	Yes	n-type	1.27	-1.39e+14
Os	Te-rich	Os_Cd	Yes	p-type	0.32	7.53e+13
Ir	Cd-rich	Ir_Cd, Ir_int_neut	Yes	n-type	1.27	-1.39e+14
Ir	Te-rich	Ir_Cd	Yes	p-type	0.32	7.53e+13
Pt	Cd-rich	Pt_int_neut, Pt_Te, Pt_Cd	Yes	n-type	1.27	-1.39e+14
Pt	Te-rich	Pt_Cd	Yes	p-type	0.32	7.45e+13
Au	Cd-rich	Au_Te, Au_Cd	Yes	n-type	1.27	-1.39e+14
Au	Te-rich	Au_Cd	Yes	p-type	0.26	9.86e+14
Hg	Cd-rich	Hg_Cd	No	n-type	1.27	-1.39e+14
Hg	Te-rich	Hg_Cd	No	p-type	0.50	8.32e+10

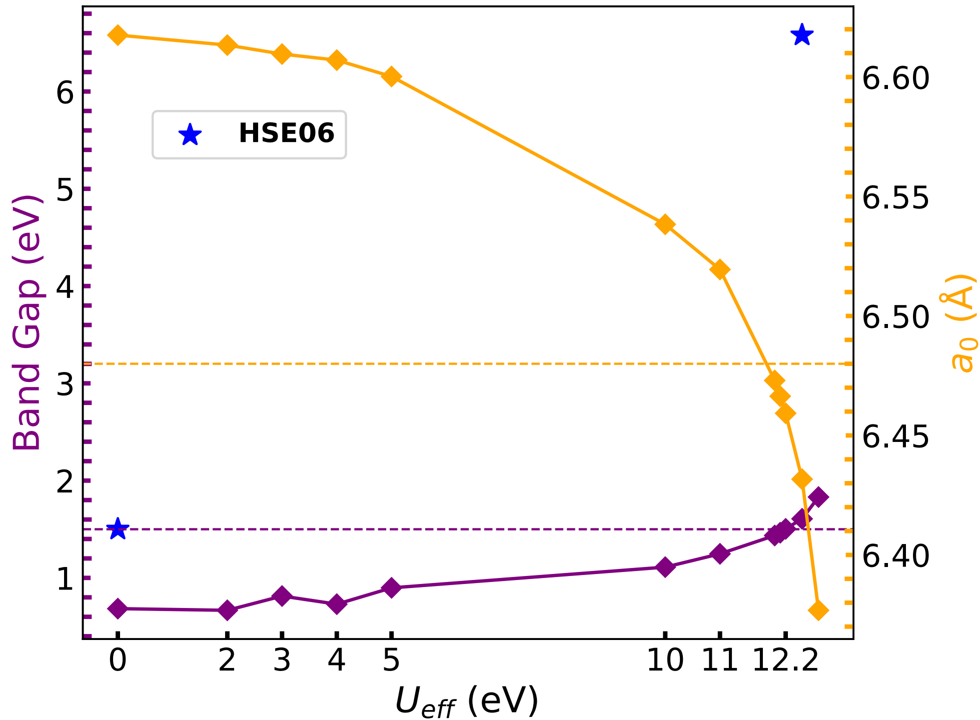


Figure C.1: Dependence of lattice parameter a_0 and bandgap on U of different methods: standard DFT (U=0 eV) and GGA+U. The stars represent the HSE06 values. The dashed lines represent the experimental values.

Appendix C

POINT DEFECTS ENGINEERING IN CDTE AND CDSETE ALLOY

C.1 Hubbard U Values Determination

The determination of the optimal U value for each material remains a matter of debate because a U value is often optimal for some calculated properties, but not all of them[215]. Here, we systematically vary the U value and compare the calculated results with experimental data, such as bandgap, lattice constant, and formation enthalpy. The comparison with other methods is shown

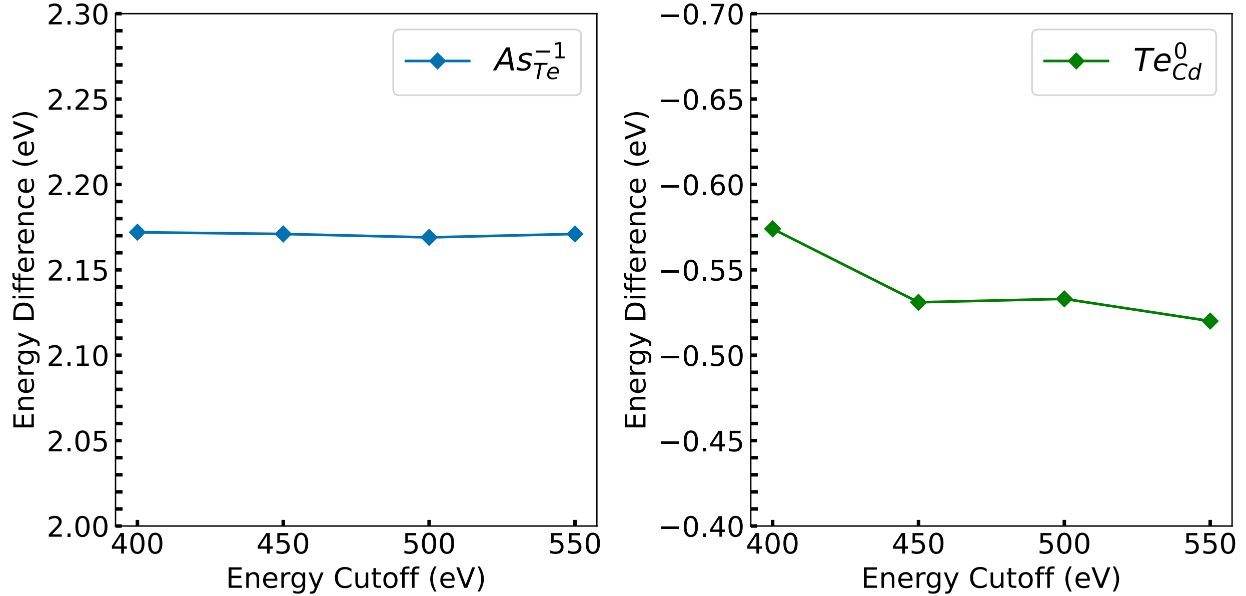


Figure C.2: Energy difference ($E_{bulk} - E_{defect}$) of defects (As_{Te}^{-1} and Te_{Cd}^{0}) without correction with energy cutoffs ranging from 400 eV to 550 eV.

in Fig C.1. The U value we chose (U=12.2 eV) show excellent agreement with the experimental bandgap and lattice constant. Besides, the experimental formation enthalpy of CdTe is -1.30 eV[18], which is close to -1.28 eV obtained in this work.

C.2 Energy Cutoff Validation

The energy cutoff for the plane-wave basis set in this work is set to 450 eV. To ensure that this energy cutoff is reasonable, we compared the energy difference between the bulk CdTe supercell and the CdTe supercell with defects (such as As_{Te}^{-1} and Te_{Cd}^{0}) without correction across a range of energy cutoffs from 400 eV to 550 eV. This comparison plot shows athat the error between large energy cutoff (550 eV) and 450 eV is within 11 meV, confirming (Fig C.2) that our chosen energy cutoff is appropriate.

Table C.1: Mean Values and Standard Deviation (σ) of CdSeTe Alloy Bulk Energy E_{bulk} and Lattice Constants a_0.

Alloy	E_{bulk} (eV)	σ_{bulk} (eV)	a_0 (Å)	σ_{lat} (Å)
$CdSe_{0.25}Te_{0.75}$	-44.25	0.048	6.34	0.003
$CdSe_{0.50}Te_{0.50}$	-47.84	0.039	6.21	0.013

Table C.2: CdSeTe Alloy Linear Regression Model: coefficients ΔE for each feature. All features are standardized before training. For nearest neighbor counting features, the count is specific to Se atoms. The training error is 0.1 meV while the testing error is 0.6 meV.

Feature	ΔE (meV)	Feature	ΔE (meV)
Se ratio	57	1NN Se	8
(Se ratio)^2	24	2NN Se	0.4
3NN Se	2	4NN Se	6

C.3 Se Distribution Dependence in CdSeTe Alloy

Utilizing a supercell program [148], we generated 30 configurations for each alloy composition, prioritizing those with the highest frequency of occurrence. In Table C.1, we list the average values and standard deviations of bulk energy and lattice constant of 25% and 50% CdSeTe alloy (30 configurations for each). In Table C.2, we use a linear regression model to explore the relationship between the Se local arrangement and bulk energy. The data consists of 66 points, with Se ratios ranging from 0% to 50%. The small standard deviation values and coefficients in Table C.2 imply that there is minimal Se-Se interaction and negligible dependence of lattice constants and bulk energy on Se arrangement.