UC Santa Barbara
UC Santa Barbara Electronic Theses and Dissertations

Title
Nonradiative Recombination in Semiconductor Alloys

Permalink
https://escholarship.org/uc/item/1c126345

Author
SHEN, XUAN X

Publication Date
2018

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

University of California
Santa Barbara

Nonradiative Recombination in Semiconductor Alloys

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

Doctor of Philosophy
in
Physics

by

Jimmy-Xuan Shen

Committee in charge:

Professor Chris G. Van de Walle , Co-chair
Professor S. James Allen , Co-chair
Professor Leon Balents

June 2018

The Dissertation of Jimmy-Xuan Shen is approved.

S. James Allen , Co-chair

Leon Balents

Chris G. Van de Walle , Co-chair

June 2018

Nonradiative Recombination

in Semiconductor Alloys

Copyright © 2018

by

Jimmy-Xuan Shen

Acknowledgements

First and foremost I want to thank my advisor Chris Van de Walle. I am grateful for his time, ideas and tireless editing that have made my Ph.D. possible. In addition to the science we have been able to do together, I am sure his unique perspectives from experience in both academic and industrial research will have an even greater impact on me for the rest of my life.

The members of the Van de Walle group have contributed greatly to my personal and professional life at UCSB. The group has been a source of friendships and advice, I am grateful to have been part of such a diverse group both in terms of personal and academic background.

I would also like to thank my family: First, my soon-to-be wife Doreen for her love and understanding while enduring living thousands of miles away for the duration of our Ph.D.s, I am grateful that we will both be working in the same place very soon. Also, my parents for their support and advice and allowing me to have a life where I can pursue my interests and passions. Their choice of moving across the Pacific twice in pursuit of a better life has allowed me to meet so many interesting and hard-working people throughout my life.

Lastly, I would like to thank the funding agencies that have made my Ph.D. possible. The work on nitride alloys was supported by the National Science Foundation (NSF), while the work on nonradiative recombination was supported by the U.S. Department of Energy (DOE).

Curriculum Vitæ
Jimmy-Xuan Shen

Education

2018	Ph.D. in Physics (Expected), University of California, Santa Barbara.
2012	M.A. in Applied and Engineering Physics, Cornell University
2010	B.A. in Physics, University of Toronto

Publications

Band bowing and the direct-to-indirect crossover in random BAlN alloys, **J.-X. Shen**, D. Wickramaratne, C. G. Van de Walle, Phys. Rev. Mater. 1, 065001 (2017).

Calcium as a nonradiative recombination center in InGaN, **J.-X. Shen**, D. Wickramaratne, C. E. Dreyer, A. Alkauskas, E. Young, J. S. Speck, and C. G. Van de Walle, Appl. Phys. Express 10, 021001 (2017).

The effects of La 5d and 4f states on the electronic and optical properties of $LaAlO_3$, **J.-X. Shen**, Andre Schleife, Anderson Janotti, and Chris G. Van de Walle, Phys. Rev. B 94, 205203 (2016).

Iron as a source of efficient Shockley-Read-Hall recombination in GaN, D. Wickramaratne, **J.-X. Shen**, C. E. Dreyer, M. Engel, M. Marsman, G. Kresse, S. Marcinkevicius, A. Alkauskas, and C. G. Van de Walle, App. Phys. Lett. 109, 162107 (2016).

Comment on Comparative study of ab initio nonradiative recombination rate calculations under different formalisms, D. Wickramaratne, **J.-X. Shen**, A. Alkauskas, and Chris G. Van de Walle, Phys. Rev. B 97, 077301 (2017).

Three-dimensional spin texture in hybrid perovskites and its impact on optical transitions, Xie Zhang, **J.-X. Shen** (equal contribution), and Chris G. Van de Walle, J. Phys. Chem. Lett. (2018)(in press)

Stability of exfoliated $Bi_2Sr_2Dy_xCa_{1-x}Cu_2O_{8+\delta}$ *studied by Raman microscopy*, L. J. Sandilands, **J.-X. Shen**, G. M. Chugunov, S. Y. F. Zhao, S. Ono, Y. Ando, and K. S. Burch, Phys. Rev. B 82, 064503 (2010).

Manuscripts submitted or in preparation

Origins of the unexpectedly strong Auger recombination in halide perovskites, **J.-X. Shen**, Xie Zhang, and Chris G. Van de Walle.
(under review)

Impact of spin-orbit coupling and phonons on the Auger recombination rate in InAs from first principles, **J.-X. Shen**, D. Steiauf, A. McAllister, Guangsha Shi, Emmanouil Kioupakis, Anderson Janotti, and Chris G. Van de Walle.
(under review)

Influence of structural disorder in free-carrier absorption of WO_3, W. Wang, H. Peelaers, **J.-X. Shen**, and Chris G. Van de Walle.
(under review)

Radiative recombination in hybrid perovskites, Xie Zhang, **J.-X. Shen** (equal contribution), and Chris G. Van de Walle.
(under review)

Electrical and optical properties of iron in nitride semiconductors, D. Wickramaratne, C. E. Dreyer, **J.-X. Shen**, Georg Kresse, A. Alkauskas, and C. G. Van de Walle.

Thermodynamics of of boron incorporation in gallium nitride, **J.-X. Shen**, D. Wickramaratne, Mark Turiansky, and Chris G. Van de Walle.

Defect identification based on first-principles calculations for deep level transient spectroscopy, D. Wickramaratne, C. E. Dreyer, B. Monserrat, **J.-X. Shen**, A. Alkauskas, and C. G. Van de Walle.

First-principles calculations of photoionization cross sections of defects in semiconductors, D. Wickramaratne, C. E. Dreyer, **J.-X. Shen**, John L. Lyons, A. Alkauskas, and C. G. Van de Walle.

Abstract

Nonradiative Recombination
in Semiconductor Alloys

by

Jimmy-Xuan Shen

The nonradiative recombination of electrons and holes in semiconductors is inherently detrimental to the performance of optoelectronic technologies. Two types of recombination mechanism cause the loss of carriers at different carrier density regimes: Shockley-Read-Hall recombination dominates at low carrier-densities and Auger recombination dominates at high carrier densities. Shockley-Read-Hall recombination can be considered as the independent capture of electrons and holes by a crystal defect or impurity via interactions with lattice vibrations. Auger recombination is a three-carrier process that involves an electron and a hole recombining across the band gap with the excess energy of that recombination going to a third carrier (either an electron or a hole). In this thesis, we discuss the simulation of these two distinct types of nonradiative recombination mechanisms using first-principles calculations by presenting case studies of the nonradiative recombination in several different material systems.

Recently, unexpectedly large concentrations of calcium have been found in InGaN quantum wells, likely due to unintentional contamination during the polishing process or from the In source. We assess the role of Ca impurities in pure GaN and InGaN alloys and identify it as a deep donor. Using our methodology for simulating the Shockley-Read-Hall recombination we will demonstrate that the Ca impurity readily assists in nonradiative recombination and is a detrimental recombination center in lower band gap InGaN alloys.

For Auger recombination, we look at two material systems (InAs and $CH_3NH_3PbI_3$) where the spin-orbit interactions play a large role in the electronic structure. Both InAs and $CH_3NH_3PbI_3$ exhibit a resonance between the band gap and the spin-orbit splitting, and we examine how this splitting affects the Auger recombination in each case. In the case of InAs, we also examine the impact of the indirect, phonon-assisted, Auger process on the recombination rate. For $CH_3NH_3PbI_3$, the Rashba-type linear-k splitting at the band edges has been flagged as a key feature in the band structure. We demonstrate how this splitting influences the Auger process, and propose how Auger recombination can be suppressed in this material.

List of Acronyms

Abbreviations

1BZ first Brillouin zone

CBM conduction-band minimum

CCD configuration coordinate diagram

DFPT density functional perturbation theory

DFT density functional theory

eeh electron-electron-hole

epME electron-phonon matrix element

GGA generalized gradient approximation

HF Hartree-Fock

hhe hole-hole-electron

HSE Heyd, Scuseria and Ernzerhof

IQE internal quantum efficiency

LDA local density approximation

LED light-emitting diode

$MAPbI_3$ methyl-ammonium lead iodide

MBE molecular-beam-epitaxy

NCPP norm-conserving pseudopotential

PAW projector augmented wave

PBE Perdew, Burke and Ernzerhof

PP pseudopotential

SOC spin-orbit coupling

SRH Shockley-Read-Hall

VASP Vienna Ab initio Simulation Package

VBM valence-band maximum

Contents

Curriculum Vitae v

Abstract vii

List of Acronyms ix

1 Introduction 1

1.1 Semiconductors and Recombination Processes . . . 1
1.2 Light-Emitting Diodes . . . 2
1.3 ABC Model . . . 3
1.4 Simulation of Nonradiative Mechanisms . . . 7
1.5 Permissions and Attributions . . . 9

2 Theoretical Background 10

2.1 The Many-Body Problem . . . 10
2.1.1 The Many-Body Schrödinger Equation . . . 10
2.1.2 The Born-Oppenheimer Approximation . . . 11
2.2 Density Functional Theory . . . 12
2.2.1 Hohenberg-Kohn Theorems . . . 13
2.2.2 Kohn-Sham Equations . . . 16
2.2.3 The Exchange-Correlation Functional . . . 19
2.2.4 Hybrid Functionals . . . 20
2.3 The Wave Function in DFT . . . 23
2.3.1 Blochs Theorem and Periodicity . . . 24
2.3.2 The Plane-Wave Pseudopotential Approach . . . 25
2.3.3 Projector Augmented Wave Potentials . . . 27
2.4 Shockley-Read-Hall Recombination . . . 30
2.4.1 Transition Levels and Configuration-Coordinate Diagrams . . . 31
2.4.2 Modeling the Capture Process . . . 34
Evaluating the Electron-Phonon Matrix Element with PAWs . . . 38
2.5 Auger Recombination . . . 41

Considerations for Spin-Orbit Coupling . . . 47

3 Electronic structure of BAlN alloys **51**
3.1 Introduction . . . 52
3.2 Methodology . . . 54
3.2.1 First-Principles Calculations . . . 54
3.2.2 Determination of Alloy Band Edges . . . 55
3.3 Results and Discussions . . . 59
3.3.1 Parent Compounds . . . 59
3.3.2 Alloy Structure . . . 61
3.3.3 Alloy Band Structure . . . 63
3.3.4 Implications for Applications of BAlN Alloys . . . 66
3.4 Conclusions . . . 66

4 SRH recombination at calcium defects in InGaN **68**
4.1 Introduction . . . 69
4.2 Computational Approach . . . 70
4.3 Results and Discussions . . . 71
4.3.1 Thermodynamics of Ca Defects in GaN . . . 71
4.3.2 SRH Recombination at Ca Defects in GaN . . . 73
4.3.3 SRH Recombination in InGaN . . . 74
4.4 Conclusions . . . 77

5 Auger recombination in InAs **78**
5.1 Introduction . . . 79
5.2 Computational Approach . . . 80
5.3 Results and Discussions . . . 83
5.3.1 Auger Coefficients in pure InAs . . . 83
5.3.2 Auger Coefficients in InAs alloys . . . 84
5.3.3 Effects of Spin-Orbit Coupling . . . 86
5.3.4 Effects of Phonons . . . 87
5.4 Conclusions . . . 89

6 Auger Recombination in Halide Perovskites **90**
6.1 Introduction . . . 92
6.2 Methodology . . . 95
6.3 Results and Discussions . . . 96
6.3.1 Coincidental Resonance . . . 98
6.3.2 Effects of Lattice Distortions on Auger . . . 100
6.3.3 Implications for Device Efficiency . . . 104
6.4 Conclusions . . . 104

7 Looking Forward **107**
7.1 Auger in IR detectors 107
7.2 Band-Gap Independence of Indirect Auger 108
7.3 Improving the SRH Calculations 108
7.4 Improving the Auger Calculations 109

Bibliography **110**

Chapter 1

Introduction

1.1 Semiconductors and Recombination Processes

Semiconductors are distinguished from insulators by the ability to be doped. Doping occurs when defects or impurities are present; the dopants can be donors, which provide an electron to the conduction band, or acceptors, which remove an electron from the valence band (leaving behind a hole). The electrons and holes in these semiconductors are considered *free* carriers. Much of the physics of a semiconductor is determined by the various processes that govern how these free carriers interact with each other and with the surrounding atomic environment [1]. The description of these processes is, fundamentally, a quantum mechanics problem. The existence of holes in semiconductors was first observed in the form of positive Hall coefficients which confirmed Pauli's theoretical predictions that there is a correspondence between excess electrons and electron "deficiencies" [2]. However, as late as 1931, Pauli had the opinion that "one shouldn't work on semiconductors, that is a filthy mess; who knows whether any semiconductors exist." [3]

The study of the quantum mechanical description of the electronic structure of semi-

conductors was initiated in 1931 by Allen Wilson [4]. Systematic investigations of the recombination of electrons and holes in semiconductors began in the 1950's in the works by Shockley, Read and Hall [5, 6] on defect recombination. This type of work was very relevant at the time as semiconductor transistors based on germanium had been invented in 1947, and the defect-assisted recombination of holes and electrons dramatically affected the carrier lifetimes. This initial work on defect recombination was followed shortly by the description of radiative recombination, where the energy of an electron-hole annihilation is given to a photon, in 1954 [7], and Auger recombination, where the energy of an electron-hole annihilation is used to excite another carrier, in 1959 [8].

While the theoretical description of recombination processes in semiconductors was in principle established in the 1950s, reliable and predictive calculations have become possible only in recent years. A practical methodology for calculating the electronic structure of solids was developed in the 1960s, in the form of density functional theory [9]. Many more decades would pass before computers (based on the transistors that inspired the initial studies by Shockley, Read and Hall!) became powerful enough to allow us to make *ab initio* predictions of the recombination rates in semiconductors.

1.2 Light-Emitting Diodes

Recombination of electrons and holes is the process by which these charge carriers annihilate each other. Since holes describe a missing electron, the recombination of carriers can be considered as an electron moving from the conduction band to the valence band. The states that participate in the recombination process are typically near the valence-band maximum (VBM) and conduction-band minimum (CBM), so the energy released by a recombination event is on the order of the band gap of the system. Different recombination events are classified according to the mode by which the energy is released:

in a *radiative* process the energy is emitted as a photon, while in a *nonradiative* process the energy is ultimately released in the form of lattice vibrations (i.e., phonons).

To anchor our discussions, we will examine the consequences of these recombination mechanism in the context of light-emitting diodes (LEDs). In addition to being the foundation of two separate multi-billion dollar industries (display technologies, and solid-state lighting), LEDs are also a very useful system for understanding the fundamental physical properties of materials due to the simplicity of their basic design. The fundamental components of an LED are shown in Fig. 1.1. With the advent of solid-state lighting, gallium nitride (GaN) has become one of the most widely used materials for LEDs. GaN can be intentionally doped with Si impurities that acts as donors to create n-type material (n-GaN), or with Mg impurities that act as acceptors to create p-type material (p-GaN). The electrons and holes that are generated in the n-GaN and p-GaN regions are then injected into an active region with a lower band gap (typically InGaN). In order to enhance the capture of electrons in the active region and reduce carrier leakage, a electron-blocking layer with a higher band gap (typically AlGaN) is used between the active region and the p-GaN layer.

1.3 ABC Model

The basic design in Fig. 1.1 allows us to populate the active region with both electrons and holes, and the different recombination mechanisms will allow them to annihilate. Unlike incandescent light sources, which rely on black body radiation for the generation of photons and in which much of the energy is lost as heat, there is in principle no intrinsic limitation on the efficiency of LEDs. If each electron that is injected into the active region results in the emission of a photon at the energy of the band gap, then we would consider the LED to have 100 % quantum efficiency. However during the

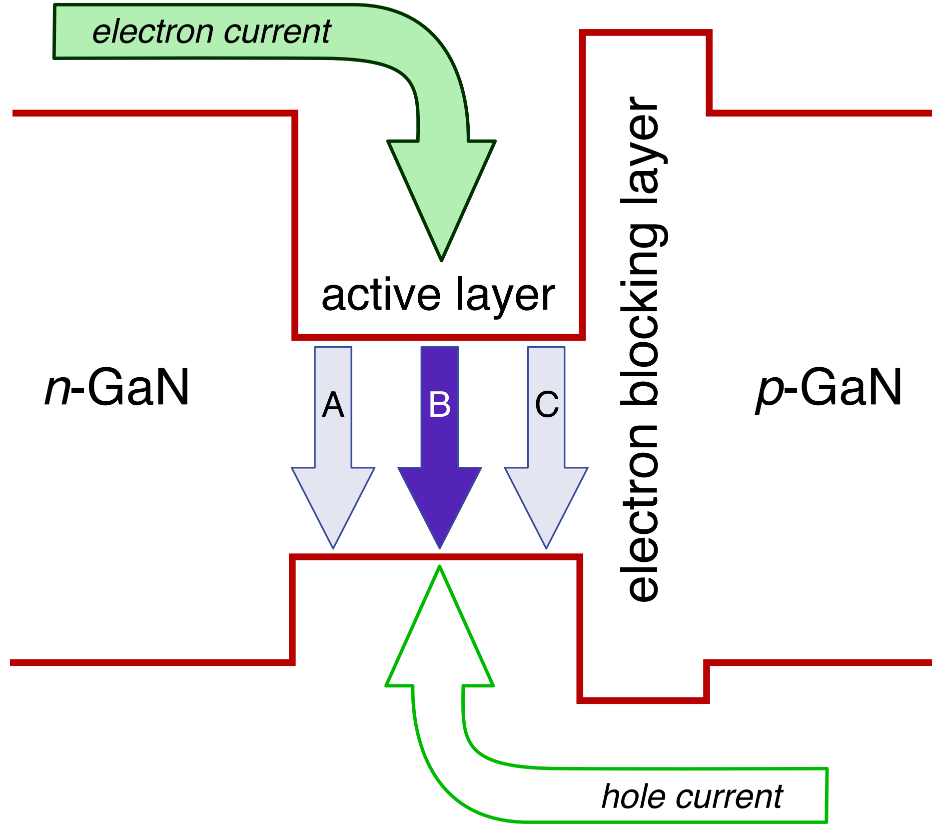


Figure 1.1: Schematic of a basic LED device structure. The energy band diagram of the VBM and CBM are shown in red. The flows of injected electrons and holes are shown by green arrows. Photons are generated inside the active layer (usually comprised of multiple quantum wells). The three possible options for electron-hole recombination are labeled A (defect-related recombination), B (radiative recombination), and C (Auger recombination).

conversion process between electronic energy to optical energy, there are a number of possible avenues of energy loss.

To separate the different contributions to efficiency loss in an LED, we split the quantum efficiency (η_Q) of an LED into two contributing factors, the internal quantum efficiency (IQE) and the extraction efficiency, η_{ext}:

$$\eta_Q = \eta_{IQE}\eta_{\text{ext}} \,. \tag{1.1}$$

The extraction efficiency is dependent on the device geometry while the IQE is an intrinsic property of the material in the active region. The IQE is the quantity that places fundamental limits on the efficiency of LEDs operating at a given band gap.

The different contributions to the IQE can be understood using the well-known ABC model for carrier recombination described below. The injected densities of electrons (n) and holes (p) can be assumed to be equal under typical LED operating conditions. The three different recombination mechanisms [defect-assisted Shockley-Read-Hall (SRH) recombination, radiative recombination, and Auger recombination] have rates proportional to different power laws of the carrier density, as shown in Fig. 1.2

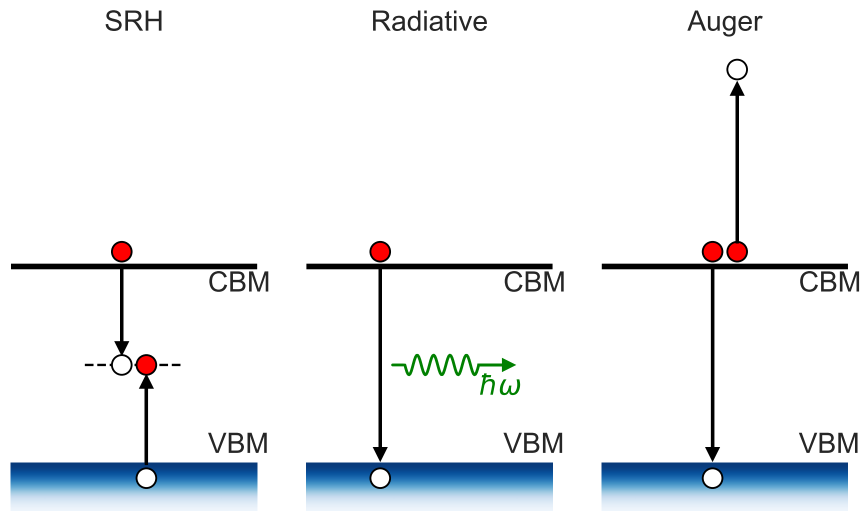


Figure 1.2: Schematic of the different forms of recombination occurring in semiconductors. The red circles represent electrons and white circles represent holes. In the Shockley-Read-Hall (SRH) process, electrons/holes from the VBM/CBM recombine at a defect level. In the radiative recombination process, the charge carriers recombine across the band gap and emit a photon. In the Auger recombination process, the excess energy from the band-gap recombination is transferred to a third charge carrier.

In the case of SRH recombination, the electrons and holes are independently captured

by a defect level. As such, the charge state of the defect is unchanged after a complete SRH event, and the defect can be continuously recycled. Since the capture of an electron and the capture of a hole are independent events, SRH recombination is a first-order process that depends linearly on n. For radiative recombination, an electron and a hole must annihilate each other to produce a photon, which is a second order (n^2) process. The Auger recombination process also involves the annihilation of an electron-hole pair; however, since the excess energy is given to another carrier, Auger recombination is a third-order (n^3) process.

According to the ABC model, the IQE is given by the ratio of the radiative recombination rate divided by the sum of all three rates:

$$\eta_{IQE} = \frac{R_{rad}}{R_{SRH} + R_{rad} + R_{Auger}} = \frac{Bn^2}{An + Bn^2 + Cn^3} \,. \tag{1.2}$$

The rate coefficients of the SRH, radiative, and Auger recombination processes are given by A, B, and C, respectively. A fundamental assumption of the ABC model is that these rate coefficients are independent of the carrier density.

In this thesis, we will focus on the simulation of nonradiative recombination mechanisms, namely SRH and Auger recombination, which are represented by the A and C coefficients in the ABC model. A schematic of the IQE curve as a function of carrier density is shown in Fig. 1.3. The IQE is peaked at

$$n_0 = \sqrt{\frac{A}{C}} \,, \tag{1.3}$$

which corresponds to a maximum efficiency of

$$\eta_0^{IQE} = \frac{B}{B + 2\sqrt{AC}} \,. \tag{1.4}$$

In the best GaN devices, the peak efficiency can be higher than 90% [10]. SRH recombination primarily affects the efficiency at low carrier concentrations, while Auger recombination limits the efficiency at high carrier densities.

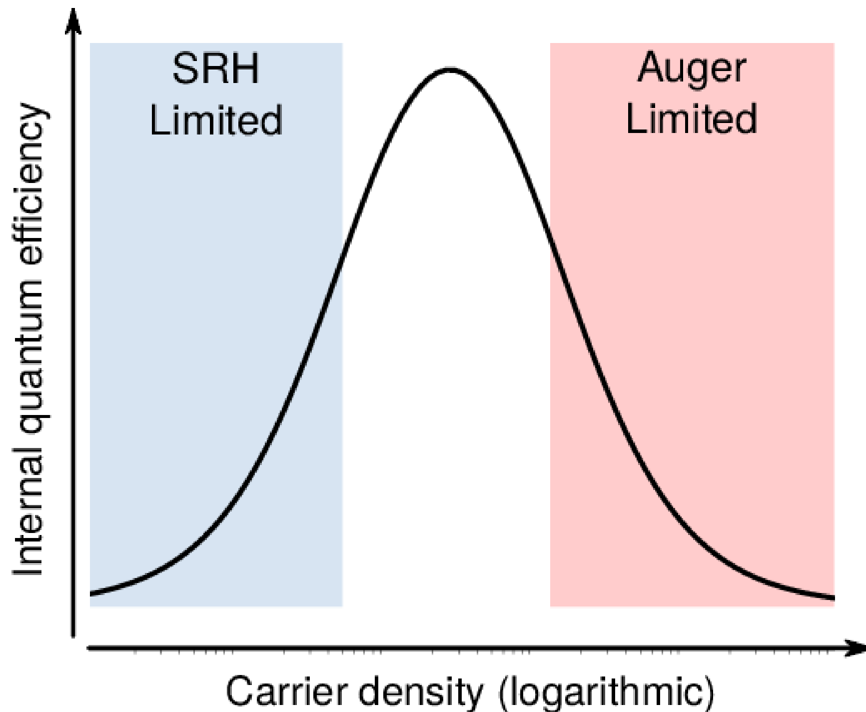


Figure 1.3: Schematic of the IQE as a function of the carrier density.

1.4 Simulation of Nonradiative Mechanisms

The focus of this thesis is on the calculation of the nonradiative recombination rates, represented by the A and C terms in Eq. (1.2). As we can see from Eq.(1.3), the A and C constants alone are enough to determine the carrier concentration at which the IQE is peaked. The recombination processes can be considered as a time-dependent perturbation of the electronic structure of the material of interest. The perturbing interaction will change based on the recombination mechanism we are interested in. For

SRH recombination, the capture process of electrons and holes is facilitated by lattice vibrations associated with a defect/impurity in the lattice, so the interaction is mediated by electron-phonon interactions. For Auger recombination, each recombination event sees two carriers recombine across the gap and the excess energy is transferred to a third carrier [See Fig. 1.2]. While the interactions for the two different recombinations mechanisms are dramatically different, the recombination rate in either case is given by *Fermi's golden rule*:

$$R_{i\to f} = \frac{2\pi}{\hbar} \left| \left\langle i \,\middle|\, \hat{H}_{int} \,\middle|\, f \right\rangle \right|^2 \delta(E_f - E_i)\,, \tag{1.5}$$

where i and f represent the initial and final states of the recombination and E_i and E_f represent their energies. In cases where multiple final states are possible, a total rate is obtained by summing (or integrating) over the final states; the energy conserving delta function then yields a density of states.

The basic methodology for how to calculate these nonradiative recombination rates was pioneered in the Van de Walle group by Prof. Kioupakis [11] and Prof. Alkauskas [12]. The work covered in this thesis will extend those methodologies to handle materials systems that were previously inaccessible. In Chapter 2, we will describe these methodologies and the extensions in detail. The buildup of the necessary components from first-principles calculations that are needed in our recombination-rate calculations is described in Sec. 2.1–2.3. The methodology for calculating SRH recombination using computationally efficient projector-augmented-wave potentials is detailed in Sec. 2.4. The extension of the methodology for Auger recombination to handle systems with spin-orbit coupling is discussed in Sec. 2.5.

The results presented in this thesis are based on four projects, which will be described in detail in subsequent chapters. Since both the SRH and Auger recombination rates are sensitive to the value of the band gap, we will first examine how the band gap can change

in a materials as we slowly change the chemical composition, using the wide-band-gap insulator BAlN as the model system (Chapter 3). The SRH methodology was used to investigate the effects of Ca in InGaN (Chapter 4). The first-principles methodology for spin-orbit coupling was used to study the Auger process in InAs, a low-band-gap semiconductor with large spin-orbit splitting (Chapter 5). Finally, we discuss the exotic band-structure features of the prototypical hybrid perovskite, $CH_3NH_3PbI_3$, and how these features influence Auger recombination (Chapter 6).

1.5 Permissions and Attributions

1. The content of Chapter 3 previously appeared in:

   Band bowing and the direct-to-indirect crossover in random BAlN alloys, **J.-X. Shen**, D. Wickramaratne, C. G. Van de Walle, Phys. Rev. Mater., 1, 065001 (2017).

2. The content of Chapter 4 previously appeared in:

   Calcium as a nonradiative recombination center in InGaN, **J.-X. Shen**, D. Wickramaratne, C. E. Dreyer, A. Alkauskas, E. Young, J. S. Speck, and C. G. Van de Walle, Appl. Phys. Express 10, 021001 (2017).

3. The content of Chapter 6 has been submitted to *Advanced Energy Materials*.

Chapter 2

Theoretical Background

2.1 The Many-Body Problem

2.1.1 The Many-Body Schrödinger Equation

The goal of first-principles calculations is to use only the most basic information about a material and predict its physical properties. On a fundamental level, a material can be viewed as composed only of positively charged nuclei and negatively charged electrons. Both the nuclei and the electrons are governed by the laws of quantum mechanics from which all of the properties of the system are derived or computed. Consider a system of N nuclei with index I at the coordinates $\mathbf{R}_I$, with masses M_I, and atomic numbers Z_I, along with N_e electrons, indexed by i, and described by the coordinate $\mathbf{r}_i$. Ignoring spin, the full Hamiltonian of the system is given by:

$$\hat{H} = \sum_{I=1}^{N} \frac{|\mathbf{P}_I|^2}{2M_I} + \sum_{i=1}^{N_e} \frac{|\mathbf{p}_i|^2}{2m_e} + \sum_{i>j} \frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|} + \sum_{I>J} \frac{Z_I Z_J e^2}{|\mathbf{R}_I - \mathbf{R}_J|} - \sum_{i,I} \frac{Z_I e^2}{|\mathbf{r}_i - \mathbf{R}_i|} \tag{2.1}$$

where m_e is the mass of the electron and e is the charge. The first two terms represent the kinetic energy of the nuclei and the electrons. The final three terms, in order, represent

the *electron-electron*, *nuclear-nuclear*, and *electron-nuclear* interactions respectively.

In principle, all of the properties of a system of interacting electrons in a solid can be obtained by solving the Schrödinger equation with the above Hamiltonian,

$$\hat{H}\Psi = E\Psi \,, \tag{2.2}$$

where E is the total energy of the system, and Ψ is the many-body wave function. Neglecting spin, Ψ is complex function of the spatial coordinate of the electrons ($\mathbf{r}_i$) and nuclei ($\mathbf{R}_I$) in the system, $\Psi(\mathbf{r}_i, ..., \mathbf{r}_{N_e}, \mathbf{R}_I, ..., \mathbf{R}_N)$. However, as Dirac pointed out [13], this "Theory of Everything" has a catch, as the "equations are too complicated to be soluble". This complexity lies within the many-body nature of the wave function itself. As an illustration, consider a simple iron atom with $N_e = 26$ electrons. The many-body wave function depends on the three spatial coordinates of the nucleus, and the 26 times 3 spatial coordinates for the electrons. Choosing a coarse grid of 10 points per coordinate axis yields 10^{81} numbers to be stored, an impossibly large quantity of data for a single atom. While the many-body nature of this wave function make the problem difficult to approach directly, a quantitative description of interacting electrons is not impossible, as there are approximations that work surprisingly well for a variety of materials.

2.1.2 The Born-Oppenheimer Approximation

From a dynamics point of view, there is a strong separation of time scales between the electronic and nuclear motions since the electrons are lighter than the nuclei by three orders of magnitude. As such, we can invoke the Born-Oppenheimer approximation, where we assume that the many-body wave function Ψ, i.e., the solution of the full

Hamiltonian [Eq. (2.1)], is separable, i.e.,

$$\Psi(\vec{\mathbf{r}}, \vec{\mathbf{R}}) \equiv \psi(\vec{\mathbf{r}})\chi(\vec{\mathbf{R}}) \tag{2.3}$$

where ψ is the wave function of the electrons, and χ is the wave function of the nuclei. We have used the notation $\vec{\mathbf{R}} \equiv \mathbf{R}_1, ..., \mathbf{R}_N$ and $\vec{\mathbf{r}} \equiv \mathbf{r}_1, ..., \mathbf{r}_{N_e}$ to make the equations more compact. The Born-Oppenheimer approximation allows us to treat the electrons and nuclei independently. In many cases, we are interested in the properties of solid-state systems, where the atoms are in their lowest energy configuration. As such, we can treat the atomic positions as parameters and only consider the Hamiltonian for the electrons in the system:

$$\hat{H} = \sum_{i=1}^{N_e} \frac{|\mathbf{p}_i|^2}{2m_e} + \sum_{i>j} \frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|} - \sum_{i,I} \frac{Z_I e^2}{|\mathbf{r}_i - \mathbf{R}_i|} \tag{2.4}$$

Note that nuclear coordinates still enter the electronic Hamiltonian equation as an external potential experienced by the electrons. While we have just eliminated the nuclear degree of freedom, solving the electron many-body problem is still intractable. Clearly, further approximations are required.

2.2 Density Functional Theory

Almost immediately after Dirac proclaimed in 1929 that practical approximations were needed to simulate quantum systems [13], physicists and chemists began devising the theoretical framework needed to calculate realistic quantum systems. The idea that ground-state energy can be expressed in terms of electron density alone was implemented in Thomas-Fermi theory in 1928 [14], and the idea to solve the Schrödinger equation with a determinant of single-particle wave functions was implemented in what we now call Hartree-Fock (HF) theory in 1935 [15, 16]. By the 1960's, quantum chemists

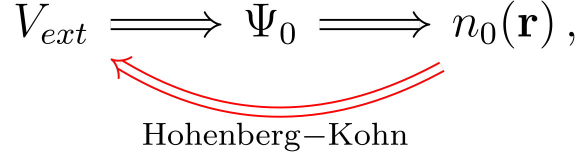
routinely used HF theory to solve the many-body quantum system. Density functional theory (DFT) first appeared in 1964–65 [17, 18], and has since been one of most widely used theoretical results in physics. Broadly speaking, DFT allows us to replace the the many-body Schrödinger equation [Eq. (2.2)] with a system of auxiliary single-particle Schrödinger equations. This allows us to dramatically reduce the size of the wave function in our iron atom example, since each single particle wave function only requires the storage of 10^3 numbers for a total of 26×10^3 for the whole system. This approximation makes solving the Schrödinger equation [Eq. (2.4)] much more tractable.

2.2.1 Hohenberg-Kohn Theorems

The theoretical foundation of DFT was established in 1964 by Hohenberg and Kohn [17]. They considered the problem of electrons moving in an external potential V_{ext}, where the electrons interact through the Coulomb interaction. Within the Born-Oppenheimer approximation, the three terms in the Hamiltonian of Eq. (2.4) can be written as:

$$\hat{H} = T_e + V_{ext} + V_{ee}\,, \tag{2.5}$$

where T_e is the kinetic energy operator, and V_{ee} represents the total energy from electron-electron interactions. In principle, this Hamiltonian above can be solved to obtain the ground-state wave function (Ψ_0), and the density operator can be evaluated using Ψ_0 to obtain the electron density $n_0(\mathbf{r})$. The first Hohenberg-Kohn theorem completes the cycle by proving that the ground-state electron density uniquely determines the external potential that induced it [17]. This can be visualized as:



and thus there is an one-to-one correspondence between V_{ext} and n_0. This one-to-one correspondence establishes that n_0 can be used as the key quantity for computing all expectation values for the ground state of the system.

Assuming that the ground state is nondegenerate, the proof, by contradiction, starts by assuming there are two different Hamiltonians $\hat{H}$ and $\hat{H}'$, with external potentials V_{ext} and V'_{ext} and the corresponding unique ground-state wave functions Ψ_0 and Ψ'_0. Additionally, assume that (†) these two different wave functions give the same charge density $n_0(\mathbf{r})$. Since Ψ'_0 is the ground state of $\hat{H}'$, the energy expectation value of any other wave function (namely Ψ_0) will be greater:

$$\langle \Psi'_0 | \hat{H}' | \Psi'_0 \rangle < \langle \Psi_0 | \hat{H}' | \Psi_0 \rangle \ . \tag{2.6}$$

Since the two Hamiltonians only differ in the external potential, we can rewrite Eq. (2.6) as:

$$\langle \Psi'_0 | \hat{H}' | \Psi'_0 \rangle < \langle \Psi_0 | \hat{H} | \Psi_0 \rangle + \langle \Psi_0 | \left[V'_{ext}(\mathbf{r}) - V_{ext}(\mathbf{r}) \right] | \Psi_0 \rangle \ . \tag{2.7}$$

The integrals can be evaluated to give:

$$E' < E + \int [V'_{ext}(\mathbf{r})) - V_{ext}(\mathbf{r})] n_0(\mathbf{r}) \, d\mathbf{r} \, . \tag{2.8}$$

Similarly, if we look at the minimum energy of the unprimed state,

$$\langle \Psi_0 | \hat{H} | \Psi_0 \rangle < \langle \Psi'_0 | \hat{H} | \Psi'_0 \rangle \ . \tag{2.9}$$

we get:

$$E < E' + \int [V_{ext}(\mathbf{r})) - V'_{ext}(\mathbf{r})] n_0(\mathbf{r})\, d\mathbf{r}\,, \tag{2.10}$$

where n_0 is assumed to be the same in both equations. Adding the two equations above leads to:

$$E + E' < E' + E\,, \tag{2.11}$$

which is clearly contradictory. Thus, the assumption above (†) must be false, which means that n_0 uniquely determines the external potential, and therefore uniquely determines solutions of the entire system.

The first Hohenberg-Kohn theorem proves the existence of a one-to-one mapping between the external potential and the ground-state density, but it does not provide a practical way of obtaining the ground-state density. The second Hohenberg-Kohn theorem addresses this issue. The second Hohenberg-Kohn theorem states that there exists a functional $F_{\mathrm{HK}}[n_0]$, accounting for the kinetic and interaction energies in Eq. (2.5):

$$F_{\mathrm{HK}}[n_0] = T_e[n_0] + V_{ee}[n_0]\,, \tag{2.12}$$

which is a *universal* functional that is the same for all systems of interacting electrons. If this functional is known, the energy of of a system can be written as a functional of the density,

$$E[n_0] = \int V_{ext} n_0(\mathbf{r})\, d\mathbf{r} + F_{\mathrm{HK}}[n_0]\,, \tag{2.13}$$

The density that minimizes this functional corresponds to the exact ground-state charge density n_0 of the system.

This theorem can be proven by considering a trial electron density $\tilde{n}(\mathbf{r})$ such that

$$\int \tilde{n}(\mathbf{r})\, d\mathbf{r} = N\,. \tag{2.14}$$

From the statement of the first Hohenberg-Kohn theorem, the trial electron density $\tilde{n}$ determines an external potential $\tilde{V}_{ext}$ and its own wave functions $\tilde{\Psi}$. Following the variational approach we used above, we get that:

$$\left\langle \tilde{\Psi} \middle| \hat{H} \middle| \tilde{\Psi} \right\rangle = \int V_{ext}\tilde{n}(\mathbf{r})\, d\mathbf{r} + F_{\mathrm{HK}}[\tilde{n}] > E_0\,, \tag{2.15}$$

Thus, if we have the system-independent functional F_{HK} we can simply minimize $\left\langle \tilde{\Psi} \middle| \hat{H} \middle| \tilde{\Psi} \right\rangle$ to obtain the solutions to our Hamiltonian.

2.2.2 Kohn-Sham Equations

The Hohenberg-Kohn theorems prove the existence of a universal functional $F_{\mathrm{HK}}[n]$, but does not provide any way of determining its form. To this end, Kohn and Sham [18] outlined an approximation scheme in which the interacting electron system is mapped onto a system of non-interacting particles. Since the system is non-interacting, the density can be constructed as a sum of the individual particle densities,

$$n(\mathbf{r}) = \sum_{i=1}^{N_e} |\psi_i(\mathbf{r})|^2\ . \tag{2.16}$$

In the case of the iron atom, this is the crucial approximation that reduces the 10^{78} numbers needed to store the wave function to 26×10^3 numbers. Within this approximation, we can write down the energy functional,

$$E[n_0] = T[n] + V_{\mathrm{Hartree}}[n] + E_{\mathrm{XC}}[n] + \int V_{ext}n(\mathbf{r})\, d\mathbf{r}\,, \tag{2.17}$$

Here, the kinetic energy term is simply the sum of the kinetic energies of the non-interacting particles,

$$T[n] = \sum_{i=1}^{N_e} \int \psi_i^*(\mathbf{r}) \left(\frac{-\hbar^2}{2m_e} \nabla^2 \right) \psi_i(\mathbf{r})\, d\mathbf{r}\,. \tag{2.18}$$

The V_{Hartree} and E_{XC} terms together make up the functional $F_{\text{HK}}[n]$ for the electron-electron interactions. V_{Hartree} is the contribution to the total energy from Coulomb repulsion between electrons,

$$V_{\text{Hartree}}[n] = \frac{e^2}{2} \int \int \frac{n(\mathbf{r})n(\mathbf{r})}{|\mathbf{r} - \mathbf{r}'|}\, d\mathbf{r} d\mathbf{r}'\,. \tag{2.19}$$

The additional effects of the exchange interaction and correlation are added through the exchange-correlation functional $E_{\text{XC}}[n]$. While we know this functional exists, its true form is not known, and many approximations have been developed.

Within the Kohn-Sham framework, the many-body problem is reduced to solving an auxiliary one-electron Schrödinger equation, known as the Kohn-Sham (KS) equation,

$$\underbrace{\left[\frac{-\hbar^2}{2m_e} \nabla^2 + V_{ext}(\mathbf{r}) + e^2 \int \frac{n(\mathbf{r})}{|\mathbf{r} - \mathbf{r}'|}\, d\mathbf{r}' + \frac{\delta E_{\text{XC}}[n]}{\delta n(\mathbf{r})} \right]}_{\hat{H}_{\text{KS}}} \psi_i(\mathbf{r}) = \varepsilon_i \psi_i(\mathbf{r})\,, \tag{2.20}$$

where the single-particle states ψ_i are solutions of the single-particle Hamiltonian $\hat{H}_{\text{KS}}$. Clearly, solving the Kohn-Sham equations requires *a priori* knowledge of the electron density, but this quantity can only be obtained from solutions of the Kohn-Sham equations. Hence, solving for the ground-state charge density requires a self-consistent procedure. In practice, this is done by starting with a trial density n_{tr}; one common choice is to start with the charge density of the isolated atoms arranged into the periodic lattice of

the solid. With this charge density, the Hamiltonian in the Kohn-Sham equation can be constructed. Then, the Hamiltonian is solved to obtain the single-particle wave functions ψ_i, which can be used to update the charge density n using Eq. (2.16). This calculated density can be used as the new trial density, and the Kohn-Sham equations can be updated and solved again. The process above (illustrated in Fig. 2.1) is repeated until the total-energy difference between each step is small enough to consider the solution "converged".

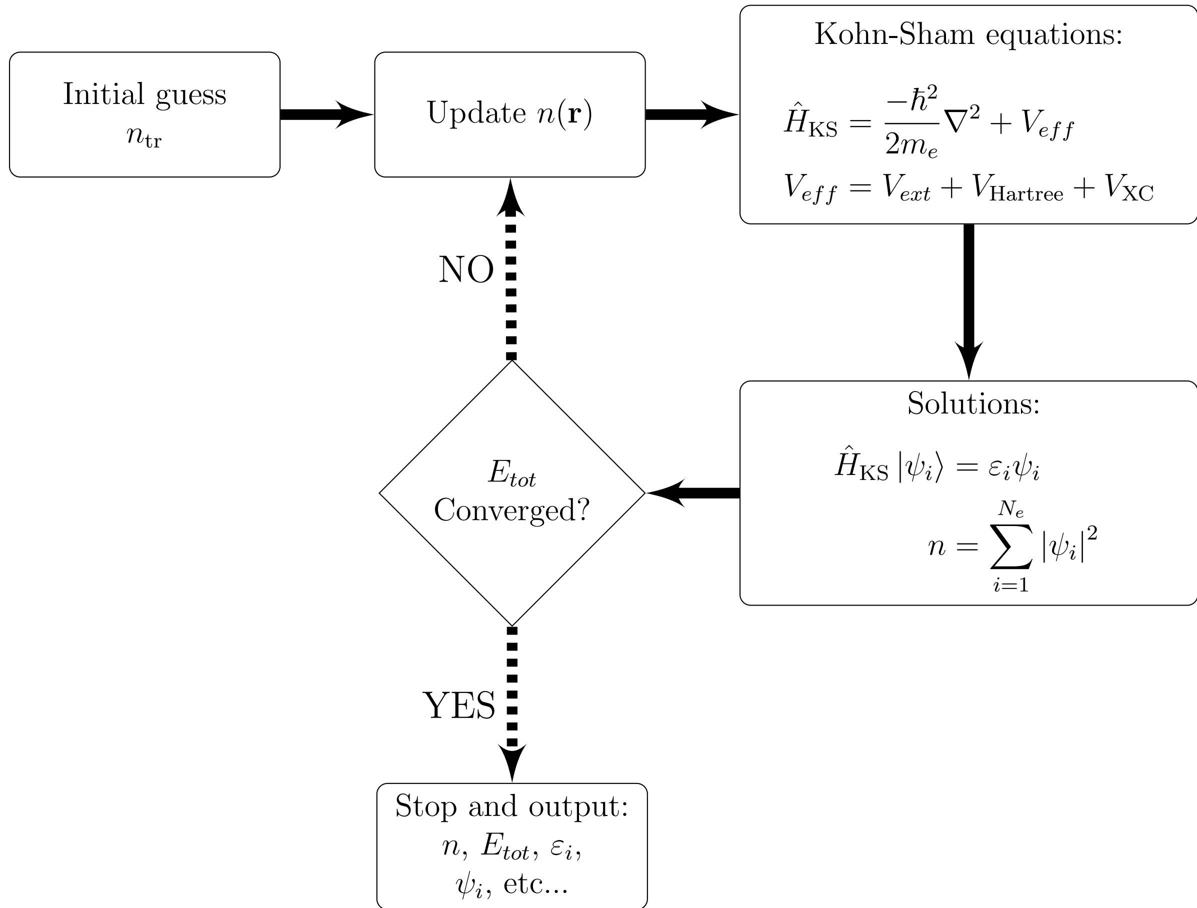


Figure 2.1: Schematic computational diagram of the Kohn-Sham procedure.

2.2.3 The Exchange-Correlation Functional

In principle, the Kohn-Sham scheme outlined above provides the exact solution to the ground-state energy of a many-body system. However, V_{XC} is not known and must be approximated. The earliest approximation of V_{XC} was also introduced by Kohn and Sham [18] and is commonly known as the local density approximation (LDA). LDA takes advantage of the fact that E_{XC} can be calculated explicitly for the homogeneous electron gas and simply replaces the exchange-correlation contributions at each point in space with those of a homogeneous electron gas with the same density

$$E_{\mathrm{XC}}^{\mathrm{LDA}}[n] = \int n(\mathbf{r})\mathcal{E}_{\mathrm{XC}}(n(\mathbf{r}))\, d\mathbf{r}\,. \tag{2.21}$$

Here, $\mathcal{E}_{\mathrm{XC}}(n(\mathbf{r}))$ is the exchange-correlation energy per electron of a homogeneous electron gas with density $n(\mathbf{r})$. The exchange contribution can be expressed analytically,

$$E_{\mathrm{X}}^{\mathrm{LDA}}[n] = -\frac{3}{4}\left(\frac{3}{\pi}\right)^{1/3}\int n(\mathbf{r})^{4/3}\, d\mathbf{r}\,, \tag{2.22}$$

and the correlation energy can be obtained from quantum Monte Carlo calculations [19]. While the approximation taken in the LDA seems dramatic, LDA often works well for solid-state systems with delocalized electrons. However, LDA consistently underestimates the the bond lengths and the band gaps.

Another common approximation scheme is the generalized gradient approximation (GGA), or the semi-local approximation. Within GGA, the exchange-correlation energy is determined by $n(\mathbf{r})$ and $\nabla n(\mathbf{r})$,

$$E_{\mathrm{XC}}^{\mathrm{GGA}}[n] = \int f\left[n(\mathbf{r}),\, \nabla n(\mathbf{r})\right]\, d\mathbf{r}\,. \tag{2.23}$$

While GGA generally gives more accurate results than LDA, the bond lengths tend to be overestimated, and the band gap is still systematically underestimated. The local (LDA) and semi-local (GGA) approximations have their limitations, but they are still widely used to date. The total energies obtained from LDA and GGA are often accurate enough for predicting materials properties that depend only on the total energies. Moreover, the quantities of interest are always energy differences. where the errors in the approximation of $E_{\mathrm{XC}}[n]$ can often cancel out. LDA and GGA have been consistently successful in predicting the structural properties and phase stability of many materials.

While the total energies obtained are surprisingly accurate, both LDA and GGA suffer from the well-known "band-gap error", where the standard local and semi-local DFT underestimate the band gap of semiconductors and insulators. The extent of this error can be several electron-volts; as a result, small-band-gap materials are often mistakenly predicted to be metallic using LDA and GGA. If we want to make predictions about the electronic and optical properties of semiconductors, this systematic underestimation must be addressed.

2.2.4 Hybrid Functionals

The fundamental limitation of the local (LDA) and semi-local (GGA) functionals is the so-called *self-interaction error*. The self-interaction error can be illustrated by considering a simple single-electron system like the hydrogen atom. One contribution to the energy of the Kohn-Sham system is the Hartree energy given in Eq. (2.19). It is obvious in the case of the hydrogen atom that the Hartree term accounts for unphysical interactions of the electron with itself. For the hypothetical *exact* functional, contributions from the exchange-correlation functional should exactly cancel this self-interaction error (as is the case in Hartree-Fock). Unfortunately, the semi-local functionals do not approximate

the exact functional well enough for this cancellation to occur [20].

It turns out that the missing energy is precisely the discontinuity of the function at the highest occupied electron number [21], as the band gap (E_g) is given by:

$$E_g = \varepsilon_{\text{CBM}} - \varepsilon_{\text{VBM}} + \left.\frac{\delta E_{\text{XC}}}{\delta n}\right|_{N+\delta} - \left.\frac{\delta E_{\text{XC}}}{\delta n}\right|_{N-\delta}, \tag{2.24}$$

where the eigenenergies (ε) of the conduction-band minimum (CBM) and valence-band maximum (VBM) are obtained for DFT. The discontinuity comes from evaluating the functional derivative $\delta E_{\text{XC}}/\delta n$ at $N+\delta$ and $N-\delta$ in the limit as $\delta \to 0$. It is well known that semi-local functionals fail to capture the derivative discontinuity [20]. However, this discontinuity is present in the exchange functional from HF theory [21]. As such, one obvious approach to correct for the band-gap underestimation of semi-local functionals is to incorporate some degree of HF exchange into the exchange-correlation functional to reproduce the derivative discontinuity; an illustration of this idea is shown in Fig. 2.2. This idea was first proposed by Becke in 1993 [22], but usage was limited at the time since evaluation of the HF exchange is orders of magnitude more expensive computationally. In recent years, due to the exponential increase in parallel computing power and advances in algorithmic design, this approach has become increasingly popular.

In a hybrid functional, the exchange-correlation potential ($V_{\text{XC}}^{\text{hybrid}}$) is calculated by mixing the exchange energy from a standard DFT calculation ($E_{\text{X}}^{\text{DFT}}$) with a fraction of the HF exchange (E_{X}^{HF}):

$$E_{\text{XC}}^{\text{hybrid}} = E_{\text{XC}}^{\text{DFT}} + \alpha(E_{\text{X}}^{\text{HF}} - E_{\text{X}}^{\text{DFT}})\,. \tag{2.25}$$

The *mixing parameter* α is the fraction of HF exchange used in the calculation of the exchange-correlation potential. One of most successful hybrid functionals used for the

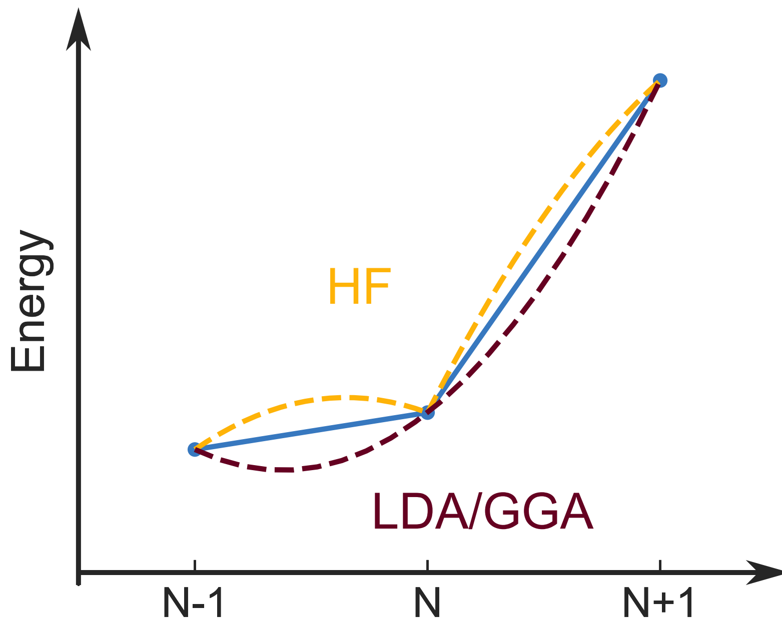


Figure 2.2: Illustration of the derivative discontinuity problem. The semi-local functionals give a continuous derivative at N, while HF tends to overestimate the discontinuity. A compromise can be had by mixing these two methods to approximate the behavior of the exact functional.

calculation of electronic properties of bulk and defect properties is the screened hybrid functional proposed by Heyd, Scuseria and Ernzerhof, commonly called the HSE functional [23, 24]. The HSE functional uses a range-separated approach, where the long-range and short-range exchange potentials are treated differently. HSE uses the GGA functional of Perdew, Burke and Ernzerhof (PBE) [25] as the starting point. The exchange potential falls off as $1/r$ and can be split into:

$$\frac{1}{r} = \underbrace{\frac{\mathrm{erf}(\omega r)}{r}}_{\text{long range}} + \underbrace{\frac{\mathrm{erfc}(\omega r)}{r}}_{\text{short range}}, \tag{2.26}$$

where the term containing the error function describes the long-range exchange potential,

and the term containing the complimentary error function describes the short-range exchange potential. The adjustable parameter ω defines the range separation in Eq. (2.26). The long-range exchange potential and the correlation potential are calculated with the PBE functional, while the HF exchange is only mixed with the short-range exchange potential. Range separating the exchange interaction gives a dramatic computational advantage since only the short-range HF exchange integrals need to be evaluated.

Thus, the exchange-correlation energy is given by:

$$E_{\mathrm{XC}}^{\mathrm{HSE}} = \alpha E_{\mathrm{X}}^{\mathrm{HF,SR}}(\omega) + (1-\alpha) E_{\mathrm{X}}^{\mathrm{PBE,SR}}(\omega) + E_{\mathrm{X}}^{\mathrm{PBE,LR}}(\omega) + E_{\mathrm{C}}^{\mathrm{PBE}} \, . \tag{2.27}$$

The superscripts SR and LR represent short-range and long-range, respectively. In principle, the two parameters in Eq. (2.27), ω and α, can both be tuned to achieve better agreement with experiments. The typical implementation of HSE, known as HSE06 [24], uses $\alpha = 0.25$ and $\omega = 0.2$ Å^{-1} and has been used to obtain quantitatively accurate physical parameters (such as band gaps and formation enthalpies) of semiconductors [26].

2.3 The Wave Function in DFT

The eigenenergies ε_i and the wave functions ψ_i of the single-particle states obtained in the Kohn-Sham DFT scheme are technically auxiliary variables [only the eigenenergy of the VBM (ε_{VBM}) in an finite system has physical meaning)]. Nonetheless, the Kohn-Sham non-interacting system tends to capture enough of the physics of the real many-particle system that the Kohn-Sham states are often used to help understand how the actual wave functions of the solid-state system behave. In this section, we will examine

how the wave function in the solid state is treated in DFT implementations.

2.3.1 Blochs Theorem and Periodicity

The problem of solving a quantum system with a small number of atoms is well-defined in the Kohn-Sham scheme: we simply have to solve for the eigenenergies and wave functions of a single non-interacting Hamiltonian. However, explicit solutions of this single-particle problem are still difficult to obtain for systems with large numbers of atoms. Fortunately, in the case of crystalline solids, we can take advantage of periodicity since the external potential is given by a periodic lattice of positively charged nuclei. The periodicity in the external potential means that solutions to the Kohn-Sham equations need to be expressed as Bloch waves [27], which are the product of a plane wave and a lattice-periodic function u:

$$\psi_{j,\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}} u_{j,\mathbf{k}}(\mathbf{r})\,. \tag{2.28}$$

Note that due to the periodicity of the external potential, the crystallographic wave vector $\mathbf{k}$ is a valid quantum number that describes the state along with the band index j. A key property of Bloch wave functions is the periodicity of the wave function for any reciprocal lattice vector $\mathbf{G}$,

$$\psi_{j,\mathbf{k}+\mathbf{G}} = \psi_{j,\mathbf{k}}\,. \tag{2.29}$$

The equivalence of $\mathbf{k}$ and $\mathbf{k}+\mathbf{G}$ means that while there are an infinite range of wave vectors, we only need to consider the wave vectors (commonly called k points) in the first Brillouin zone (1BZ).

Since the resulting wave functions $\psi_{j,\mathbf{k}}$ depend explicitly on the wave vector $\mathbf{k}$, the electron density $n(\mathbf{r})$ can be written as an integral over the 1BZ,

$$n(\mathbf{r}) = \sum_{j}^{n} \int_{1\mathrm{BZ}} \left|\psi_{j,\mathbf{k}}(\mathbf{r})\right|^2 \, d\mathbf{k}\,. \tag{2.30}$$

In principle, we are still stuck because we have just traded the countably large number (all the electrons in a crystal $\approx 10^{23}$) of wave functions to solve for a uncountably large continuum of solutions for each $\mathbf{k}$. However, while the non-interacting single particle systems give independent solutions, all of the eigen-energy and Bloch wave functions vary slowly with $\mathbf{k}$. This means that all of the integrals above can be approximated with a summation over a finite, in fact modest, set of k points:

$$n(\mathbf{r}) = \sum_{j}^{n} \sum_{k}^{N_{\mathbf{k}}} |\psi_{j,\mathbf{k}}(\mathbf{r})|^2 \, d\mathbf{k}\,. \tag{2.31}$$

The number of k points needed can be further reduced by exploiting the symmetries in the problem; integrations (or summations) need to be carried out only over the *irreducible Brillouin zone*. A common scheme for selecting optimal k points is the Monkhorst-Pack method [28]. In fractional coordinates, the Monkhorst-Pack method uses a rectangular grid with $M_x \times M_y \times M_z$ spaced evenly in the 1BZ.

2.3.2 The Plane-Wave Pseudopotential Approach

In any practical implementation of DFT, solutions to the Kohn-Sham equation [Eq. (2.20)] must be expanded in a set of linearly independent basis functions. Since the solutions for solid-state systems are Bloch periodic, the natural basis for expanding the lattice-periodic part of a wave function [$u_i(\mathbf{r})$] is a plane-wave basis set. The periodicity in reciprocal space [Eq. (2.29)] means that only plane waves with wave vectors $\mathbf{k} + \mathbf{G}$ (where $\mathbf{G}$ is a reciprocal lattice vector). The lattice-periodic part of the wave function can then be expressed as a Fourier expansion:

$$u_i(\mathbf{r}) = \sum_{\mathbf{G}} c_i(\mathbf{G}) e^{i\mathbf{G}\cdot\mathbf{r}}\,. \tag{2.32}$$

In practical DFT implementations the infinite series in Eq. (2.32) is truncated by limiting the energy associated with the wave vector with some cut-off energy

$$\frac{\hbar^2}{2m_e}\left|\mathbf{k}+\mathbf{G}\right|^2 < E_{cut}\,. \tag{2.33}$$

A major drawback of the plane-wave method is that a high cut-off energy, and thus a large number of plane-wave coefficients, is needed to capture the rapid fluctuations of the wave function near the atomic core. This issue can be circumvented by using using the pseudopotential (PP) framework, where only a few of the electrons in the outer shell of the atom are treated as Kohn-Sham states. The core electrons are essentially eliminated, so that the Schrödinger equation will have a modified effective potential whose solutions are identical to the all-electron wave functions outside of the core. Since the core electrons are no longer considered, the valence electrons can be described by pseudo-wave functions that are smoother near the atomic core. This approach is much more computationally efficient since the number of plane waves needed to describe the smoothed functions is greatly reduced.

There exist a large variety of schemes for constructing PPs. A common method utilizes the concept of norm-conserving pseudopotentials (NCPPs) [29]. The NCPPs are constructed by solving the Kohn-Sham equations for an isolated atom. Since the isolated atom is a spherically symmetric system, the Kohn-Sham equations can be solved efficiently on a radial grid with all of the electrons treated explicitly. Then, the NCPP for angular momentum l is constructed on a Cartesian grid so that:

1. The resulting valence orbitals are identical with the corresponding all-electron orbitals for r larger than some l-dependent cut-off radius r_l^{cut}.

2. The norm of the atomic orbitals inside the sphere of radius r_l^{cut} is conserved between

the rapidly fluctuating all-electron wave functions and the smoother pseudized wave functions.

Note that condition 1 automatically ensures that the eigenvalues of a pseudized orbital are the same as for the all-electron orbital since the eigenenergies directly determine the asymptotic decay of the orbitals. The PPs are obtained on a dense radial grid and then translated to a Cartesian grid to be used for the plane-wave calculations. For the construction of the PP inside of the core region different procedures have been suggested, among which the Troullier-Martins [30] approach is widely used.

NCPPs have been used countless times to compute the properties of solid state materials. Since the potential is smooth, it can be stored on a sparse Cartesian grid along with the Kohn-Sham wave functions, which makes the problem much more computationally tractable. This gives us a convenient and easily accessible basis of orthonormalized wave functions, which can be used to evaluate the matrix elements of various interactions.

2.3.3 Projector Augmented Wave Potentials

A major drawback of the traditional pseudopotential approach is that the smoothing of the wave functions near the nucleus is not physically meaningful, and important information about the atomic core is lost. To overcome this, the projector augmented wave (PAW) method [31] was developed. The PAW method allows the wave function to retain its complex structure in the atomic core but simultaneously requires fewer plane-wave coefficients for convergence, compared to NCPPs. In the PAW method, space is divided into *core* and *non-core* regions by an augmentation sphere Ω_a around each atom indexed by a. The pseudo-wave function $|\tilde{\psi}\rangle$ is defined to be identical to the true Kohn-Sham wave function $|\psi\rangle$ (often called the *all-electron* wave function) everywhere outside the

augmentation spheres. We seek a linear transformation $\mathcal{T}$ between $|\tilde{\psi}\rangle$ and $|\psi\rangle$,

$$\mathcal{T}\,|\tilde{\psi}\rangle = |\psi\rangle \ , \tag{2.34}$$

We want to find solutions $|\tilde{\psi}_i\rangle$ to the Schrödinger equation for the augmented KS Hamiltonian:

$$\mathcal{T}^\dagger \hat{H}_{\mathrm{KS}} \mathcal{T}\,|\tilde{\psi}_i\rangle = \varepsilon_i \mathcal{T}^\dagger \mathcal{T}\,|\tilde{\psi}_i\rangle \ . \tag{2.35}$$

Note that the linear transformation $\mathcal{T}$ is not unitary (i.e. $\mathcal{T}^\dagger\mathcal{T}$ is not the identity). This makes PAWs cumbersome to work with since the augmented wave functions $|\tilde{\psi}_i\rangle$ are not orthogonal.

Since the wave functions are already smooth in the space between the atoms (outside the Ω_a's), $\mathcal{T}$ should only modify the wave function in the core region, and we can define

$$\mathcal{T} = \sum_a \mathcal{T}_a \, , \tag{2.36}$$

where each $\mathcal{T}_a$ is only nonzero inside Ω_a. Inside the Ω_a's, we can expand the true wave function $|\psi\rangle$ in terms of the atomic-like *partial wave* orbitals, ϕ_a^i, with the requirement that:

$$|\phi_i^a\rangle = (1 + \mathcal{T}_a)\,|\tilde{\phi}_i^a\rangle \tag{2.37}$$

Once $\mathcal{T}_a$ is defined for each atom a, $\mathcal{T}$ is completely defined. The partial waves define a complete basis, so we can expand the augmented wave function at each atom,

$$|\tilde{\psi}_n\rangle = \sum_i P_{ni}^a\,|\tilde{\phi}_i^a\rangle \ (\quad \text{inside } \Omega_a\,) \tag{2.38}$$

with expansion coefficients P_{ni}^a. Using Eq. (2.34), we find that the true wave function

$|\psi\rangle$ expanded in the $|\phi_i\rangle$,

$$|\psi_n\rangle = \mathcal{T}\,|\tilde{\psi}_n\rangle = \sum_i P^a_{ni}\,|\phi^a_i\rangle \qquad \text{(inside } \Omega_a \text{)}, \tag{2.39}$$

has the exact same expansion coefficients as $\tilde{\psi}$ in Eq. (2.38).

Since we require the $\mathcal{T}$ to be linear the coefficients P^a_{ni} must be linear functionals of $|\tilde{\psi}^a_n\rangle$, meaning that

$$P^a_{ni} = \langle \tilde{p}^a_i|\tilde{\psi}_n\rangle \ , \tag{2.40}$$

where $|\tilde{p}_i\rangle$ are a fixed set of functions called the *smooth projector functions*. Since there is no overlap between the augmentation spheres, inside each sphere Ω_a Eq. (2.38) becomes

$$|\tilde{\psi}^a_n\rangle = \sum_i |\tilde{\phi}^a_i\rangle\,\langle \tilde{p}^a_i|\tilde{\psi}_n\rangle \ . \tag{2.41}$$

Thus the smooth projector functions must satisfy

$$\sum_i |\tilde{\phi}^a_i\rangle\,\langle \tilde{p}^a_i| = 1 \tag{2.42}$$

inside each augmentation sphere. From Eqs. (2.39)–(2.42), the $\mathcal{T}_a$ operators can be written as

$$\hat{\mathcal{T}} = 1 + \sum_a \sum_i (|\phi^a_i\rangle - |\tilde{\phi}^a_i\rangle)\,\langle \tilde{p}^a_i| \tag{2.43}$$

We can now evaluate matrix elements for a local operator $\hat{O}$ taken between the states $|\psi_n\rangle$ and $|\psi_m\rangle$, with the general expression:

$$\langle \hat{O}\rangle = \langle \psi_n|\hat{O}|\psi_m\rangle = \langle \tilde{\psi}_n|\hat{O}|\tilde{\psi}_m\rangle + \sum_a \sum_{i,j} \left(\langle \tilde{\phi}^a_i|\hat{O}|\phi^a_j\rangle - \langle \phi^a_i|\hat{O}|\tilde{\phi}^a_j\rangle\right) D^a_{ij} \tag{2.44}$$

where

$$D_{ij}^{a} = \langle \tilde{\psi}_n | \tilde{p}_i^a \rangle \langle \tilde{p}_j^a | \tilde{\psi}_m \rangle \ , \tag{2.45}$$

The first term in 2.44 is just the matrix element evaluated using the pseudo-wave functions on a cartesian grid, and the second term contains the corrections to the partial wave orbitals due to the PAWs. The PAW formalism provides a mathematically rigorous way to evaluate important quantities such as wave function overlaps and matrix elements of the operators. However, since the $\tilde{\psi}$'s and ϕ's are stored on different grids, the evaluation of matrix elements is often complicated. The impact of including the corrections to the partial wave orbitals due to the PAWs is usually small, since the weight of the wave functions close to the core is typically low ($< 15\%$); it is therefore often a good approximation to neglect these corrections.

Thoughout this thesis we will use PAW and NCPP wave functions to solve the Kohn-Sham equations in different situations. The approximations in NCPP works well for s/p-bonded bulk systems. However in special cases where localized orbitals, especially those with high l quantum numbers are involved, a full PAW treatment is required.

2.4 Shockley-Read-Hall Recombination

The focus of this thesis is on calculating recombination rates directly from the information available in first-principles calculations. The recombination rate due to an interaction Hamiltonian $\hat{H}$ is given by *Fermi's golden rule*:

$$R_{i \to f} = \frac{2\pi}{\hbar} \left| \left\langle \Psi_i \,\middle|\, \hat{H}_{int} \,\middle|\, \Psi_f \right\rangle \right|^2 \delta(E_f - E_i) \, , \tag{2.46}$$

where Ψ_i and Ψ_f are the initial and final states of the electron with energy E_i and E_f. The delta function, δ, ensures the conservation of energy. The precise expression for a specific

recombination process depends on the Hamiltonian $\hat{H}_{int}$ of the interaction. In principle, the wave functions must be the many-body wave functions. However, we can simplify the situation dramatically in the case of the interactions we are interested in (SRH and Auger recombination), since each interaction event only changes the occupation of a few states.

Defect levels in the middle of the band gap can interact with carriers at the band edges via electron-phonon interactions. These interactions allow for the capture of electrons (see Fig. 2.6) and holes. The theory of defect-assisted carrier loss was formulated independently by Shockley and Read [5] and Hall [6] in 1952. The rate coefficient of the capture process is different for electrons (C_n) and holes (C_p). At a given impurity density N and for $n = p$, the SRH recombination rate is characterized by the coefficient A:

$$A = N \frac{C_n C_p}{C_n + C_p} \, . \tag{2.47}$$

From Eq. (2.47), it is clear that the SRH process is rate-limited by the smaller of the two carrier capture coefficients, either C_n or C_p.

2.4.1 Transition Levels and Configuration-Coordinate Diagrams

As we have discussed above, SRH recombination requires a defect level in the band gap of the semiconductor to facilitate the capture of electrons and holes. In this section, we will cover what the defect level in Fig. 1.2 means, and how we obtain this quantity from first-principles calculations.

In the dilute limit, the formation energy (E^f) of a given defect is related to the concentration of that defect via the Boltzmann relation [32]. The environment surrounding a defect will affect its electronic properties, and this is reflected in a dependence on the Fermi level (E_{F}), which is the chemical potential of the reservoir of electrons in the

crystal. The Fermi level of semiconductors typically resides within the band gap of the material. The formation energy of a defect (denoted by X) in the charge state q is given by:

$$E^f[X^q] = E_{\rm tot}[X^q] - E_{\rm tot}[{\rm bulk}] + \sum_i n_i \mu_i + qE_{\rm F} + \Delta^q\,, \tag{2.48}$$

where $E_{\rm tot}[X^q]$ is the total energy of the supercell containing the defect with charge q (attained by addition/removal of electrons), and $E_{\rm tot}[{\rm bulk}]$ is the total energy of the pristine supercell. The chemical potentials μ_i represent the energy cost of adding ($n_i = -1$) or removing ($n_i = +1$) an atom in the process of creating a defect. Finally, the finite-size charge correction Δ^q can be determined using the Freysoldt scheme [32].

At a given Fermi level, the most stable state of a defect is determined by the charge state with the lowest formation energy . Thus the formation energy of a defect is given by the lower hull of the different linear functions expressed by Eq. (2.48) for the different charge states. The charge-state transition levels correspond to the kinks in the formation-energy plot where the slope changes. For a hypothetical deep acceptor defect that has $q = +1$ and $q = 0$ as the stable states in the band gap, the formation energy is simply a piecewise linear function with a single kink where the formation energies of these two charge states are equal. The Fermi level at which this kink occurs is the defect transition level in the SRH recombination process (see Fig. 2.3).

The two different charge states at a given transition level will relax to two distinct atomic configurations (labeled A^- and A^0 in the case of the deep acceptor). We define a configuration coordinate Q that captures the change in atomic configuration between the two different defect structures.

We construct a so-called configuration coordinate diagram (CCD) by plotting the energy as a function of the configuration coordinate. The energy in each charge state

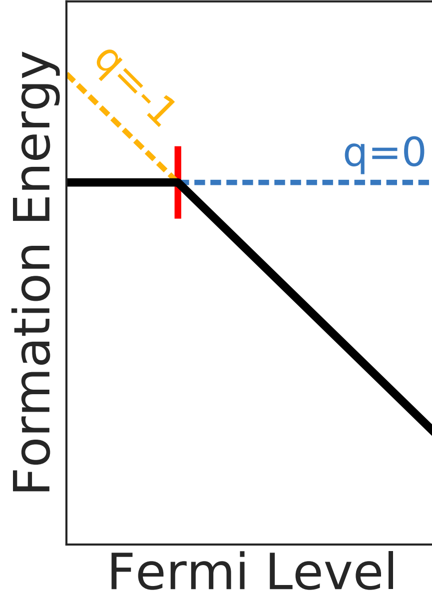


Figure 2.3: Schematic of the formation energy plot for a deep acceptor.

is assumed to be parabolic around the relaxed structure, so the two charge states are represented by two parabolas in the CCD. The configuration-coordinate offset between the minima of the two parabolas is given by the mass-weighted difference in their atomic configurations:

$$\Delta Q^2 = \sum_{\alpha} m_\alpha \left| \mathbf{R}_{\alpha;\mathrm{A}^-} - \mathbf{R}_{\alpha;\mathrm{A}^0} \right|^2 , \tag{2.49}$$

where α is the atomic index, m is mass and $\mathbf{R}$ the atomic position of a given atom. The variation of this this scalar quantity represents a continuous interpolation between the ground-state structures of the two charge states, as shown in Fig. 2.4.

The schematic of the CCD for a deep acceptor type defect is shown in Fig. 2.5. The curvature of parabolas can be obtained from additional calculations of interpolated structures between the two distinct ground-state structures. The energy offset ΔE is the energy difference between the defect level and the band edge of the captured carrier.

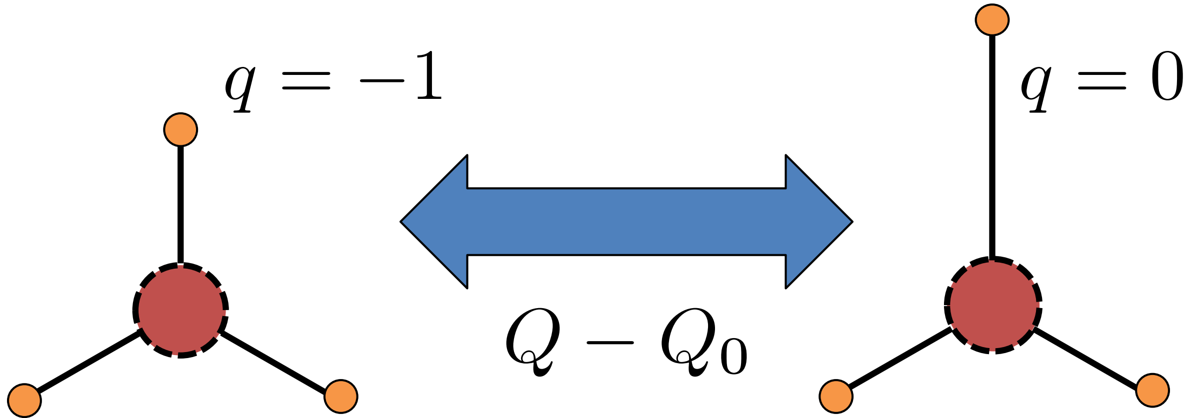


Figure 2.4: Illustration of the parameterization of the local distortions of a charged defect as the charge state is changed.

Conversely, this ΔE can be thought of as the ionization energy of the defect. In the case of a deep acceptor capturing a hole, the ΔE represents the energy difference between the defect level and the VBM.

While we are focusing on the nonradiative capture of electrons and holes, it is worthwhile to note that these CCDs are useful tools in studying the optical properties of defects. The absorption and emission energies are represented in the CCD as vertical energy transitions between the defect configurations, and can be directly compared with peaks in absorption and luminescence spectra. The energy difference ΔE corresponds to the optical transition energy when no phonons are involved and is often called the zero-phonon line in optical spectroscopy.

2.4.2 Modeling the Capture Process

A fully quantum mechanical treatment of carrier capture from first principles has been developed in the Van de Walle group [12]. To illustrate the methodology, the process of the electron capture by a deep donor is shown in Fig. 2.6. The process is illustrated in two different representations: (a) using a band diagram and (b) using a CCD. In the CCD representation, the capture of the electron can be viewed in terms of two distinct steps.

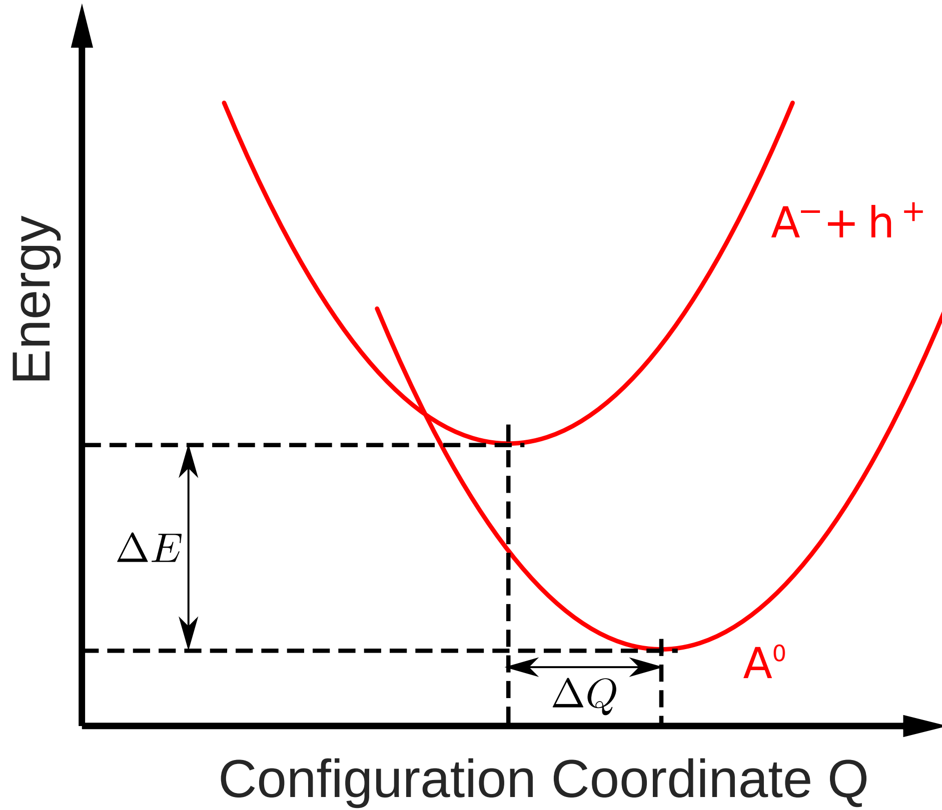


Figure 2.5: Schematic of a CCD for a deep acceptor defect. The negative charge state, with a free hole at the VBM ss the excited state and and the neutral defect is considered the ground state.

In step (1), a hole is captured by the given defect in the negative charge state (A^-) in a capture event mediated by electron-phonon interactions. This leaves the system in a high-energy vibronic state whose energy is subsequently lowered, in step (2), by emission of phonons into the crystal lattice. The vibronic relaxation in step (2) occurs on a time scale of a few picoseconds [33], while the electron capture event in step (1) is the rate-limiting step, with a typical time scale on the order of nanoseconds [34].

The carrier capture rate can be obtained from Fermi's golden rule [Eq. (2.46)] with

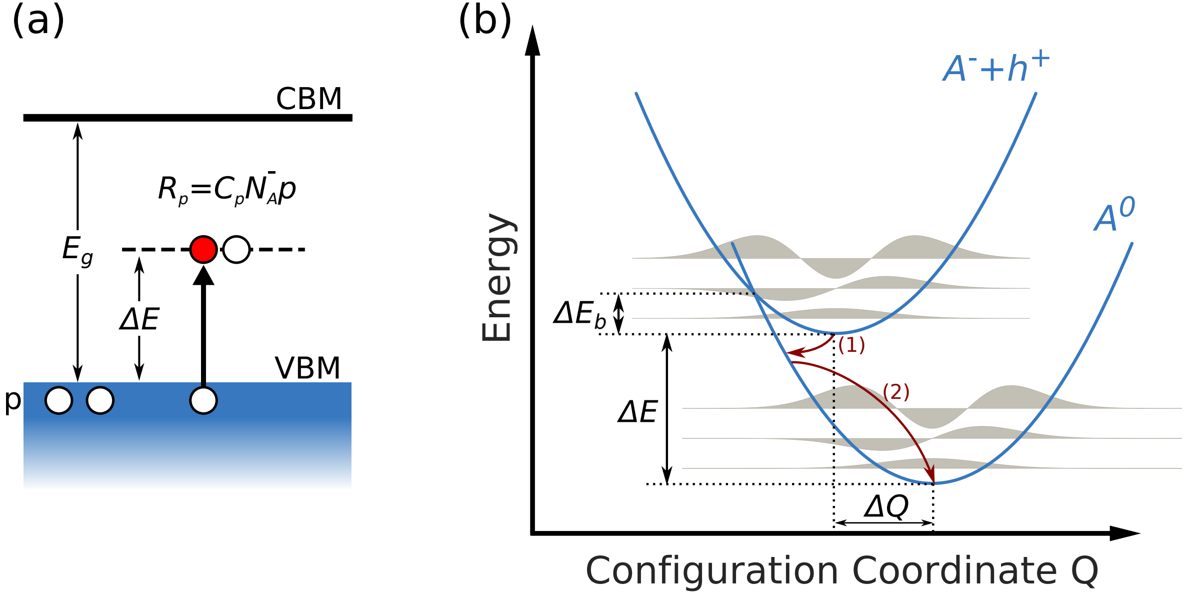


Figure 2.6: Schematic illustration of the capture process of a hole at a deep acceptor. The capture process is shown in two different representations: (a) band diagram and (b) configuration coordinate diagram. For illustration purposes, we assume the defect is unoccupied in the neutral charge state and singly occupied in the −1 charge state. The ionization energy of the acceptor is ΔE, and the difference between the atomic structures of the two charge states is given by the difference in the generalized configuration coordinate ΔQ. In (b), process (1) represents the capture of a hole via electron-phonon interactions, and process (2) is the relaxation of the vibronic state via the emission of phonons.

the electron-phonon interaction as the perturbation, giving the expression

$$R_{i\to f} = \frac{2\pi}{\hbar} g \sum_m w_m \sum_n |\Delta H^{e-ph}_{im;fn}| \delta(E_{im} - E_{fn})\,. \tag{2.50}$$

In Eq. (2.50), w_m is the thermal occupation of the vibrational state m of the initial charge state, and E_{im} and E_{fn} are total energies of the initial and the final state; g is the degeneracy factor of the final state, which reflects the fact that there might exist a few equivalent energy-degenerate (or nearly degenerate) atomic configurations of the final state. $\Delta H^{e-ph}_{im;fn}$ is the electron-phonon coupling matrix element.

In principle $\Delta H^{e-ph}_{im;fn}$ depends on both the electronic degrees of freedom $\{x\}$ and the

ionic degrees of freedom $\{Q\}$ [35]. The $\{x\}$ and $\{Q\}$ degrees of freedom are coupled to each other; however, certain approximations allow us to separate them in our calculations. The first of these approximations is the *linear coupling approximation* [36], where $\Delta H^{e-ph}_{im;fn}$ is expanded as a Taylor series of the atomic configuration around the minimum energy configuration $\{Q_0\}$. To first order, the matrix element $\Delta H^{e-ph}_{im;fn}$ is given by:

$$\Delta H^{e-ph}_{im;fn} = \sum_k \underbrace{\left\langle \Psi_i \middle| \frac{\partial \hat{H}}{\partial Q_k} \middle| \Psi_f \right\rangle}_{W^k_{if}} \langle \chi_{im} | Q_k - Q_{0;k} | \chi_{fn} \rangle \tag{2.51}$$

The sum k runs over all atomic degrees of freedom, which can be very large for a supercell defect calculation, and $Q_{0;k}$ represents the relaxed atomic configuration. The evaluation of the *electron-phonon* coupling matrix element W^k_{if} will be discussed in Sec. 2.3.3.

The key approximation in Fig. 2.6 is called the *single phonon approximation*. A phonon can be described by a generalized displacement vector connecting two states i and f,

$$\Delta \mathbf{Q} = \sum_\alpha \sqrt{m_\alpha} (\mathbf{R}_{\alpha;i} - \mathbf{R}_{\alpha;f}) \, , \tag{2.52}$$

where α is the atomic index, m is the mass and $\mathbf{R}$ is the atomic position. The special phonon in the single phonon approximation connects the relaxed atomic configurations of the two defect charge states before and after carrier capture. Alkauskas *et al.* [12] showed that this approximation captures enough of the relevant physics to give quantitative predictions of the capture coefficients. We have confirmed the validity of this approximation [37] by comparing our calculated capture rates with an alternative method which more fully includes the effects of all phonons [38, 39].

The ionic displacement matrix element [second integral in Eq. (2.51)] is analogous to the evaluation of the position operator between two off-set harmonic potentials with

different frequencies. This is a frequently encountered problem in quantum chemistry, which was solved a long time ago within that community; detailed expressions can be found in Refs. [40] and [41]. Only the evaluation of the matrix element W_{if}^{k} remains.

Evaluating the Electron-Phonon Matrix Element with PAWs

Methods for evaluating electron-phonon interactions in solids from *ab initio* calculations are well established [42]. We use the frozen-phonon method, whereby all the atoms in a supercell are displaced according to the eigenmode of the special phonon [42]. The coupling between of the atomic displacements and the electronic properties of the system is represented by the electron-phonon matrix element (epME). The epME of the special phonon (parameterized by Q) is given by:

$$W_{if} = \left\langle \Psi_i \middle| \frac{\partial \hat{H}}{\partial Q} \middle| \Psi_f \right\rangle . \tag{2.53}$$

In the equation above, the $\Psi_{\{i,f\}}$ are in principle many-body wave functions, and $\hat{H}$ is the many-body Hamiltonian. However, since we are simulating the capture of a single charge carrier, the initial and final states differ in the occupation of a single electron, and we can replace the many-body Hamiltonian and wave functions with their single-particle counterparts $\hat{H}_{\mathrm{KS}}$ (which will be denoted $\hat{h}$) and $\psi_{\{i,f\}}$ [12],

$$W_{if} = \left\langle \psi_i \middle| \frac{\partial \hat{h}}{\partial Q} \middle| \psi_f \right\rangle . \tag{2.54}$$

The typical outputs of first-principles calculations are eigenvalues and eigenstates of the Kohn-Sham Hamiltonian, and the $\frac{\partial \hat{h}}{\partial Q}$ in Eq. (2.54) is not readily accessible. However, the expression can be modified by evaluating the initial wave function $\psi_i(\delta Q)$ at some finite (but small) atomic displacement δQ along the eigenmode of the special phonon. The wave function at δQ can be expressed as the wave function of the relaxed structure

ψ_i plus the correction from first order perturbation theory:

$$|\psi_i(\delta Q)\rangle = |\psi_i\rangle + \sum_{m \neq i} \frac{|\psi_m\rangle \left\langle \psi_m \,\middle|\, \frac{\partial \hat{h}}{\partial Q} \delta Q \,\middle|\, \psi_i \right\rangle}{\varepsilon_i - \varepsilon_m} \,, \tag{2.55}$$

where m is dummy band index and the ε's are the corresponding eigenvalues. If we take the inner product of the expression above with the final state and move the displacement parameter δQ outside the integral, we get:

$$\langle \psi_f \,|\, \psi_i(\delta Q)\rangle = \frac{\left\langle \psi_f \,\middle|\, \frac{\partial \hat{h}}{\partial Q} \,\middle|\, \psi_i \right\rangle}{\varepsilon_i - \varepsilon_f} \delta Q \,. \tag{2.56}$$

which can be rearranged to:

$$\left\langle \psi_f \,\middle|\, \frac{\partial \hat{h}}{\partial Q} \,\middle|\, \psi_i \right\rangle = (\varepsilon_i - \varepsilon_f) \frac{\langle \psi_f \,|\, \psi_i(\delta Q)\rangle}{\delta Q} \,. \tag{2.57}$$

The equation above allows us to obtain the epME from two first-principles DFT calculations at slightly different atomic configurations. In practice, we calculate the wavefunction overlap $\langle \psi_f \,|\, \psi_i(\delta Q)\rangle$ for a few values of δQ and obtain the slope.

We use the PAW-based Vienna Ab initio Simulation Package (VASP) to perform the structural relaxation and defect-formation energy-calculations. This choice was made because PAWs, especially in the VASP implementation, allow us to perform the more accurate HSE hybrid functional calculations at a reasonable computational cost. However, this increase in accuracy and speed come at the cost of a more complicated data structure used to store the wave functions, making the evaluation of the overlap in Eq. (2.57) more complicated. To circumvent this problem, previous efforts in the Van de Walle group used a NCPP-based code called Car-Parrinello Molecular Dynamics (CPMD) [43] to obtain the pseudo-wave functions once the atomic structure has been relaxed in VASP [12, 44, 45],

and used these pseudo-wave functions in place of the *all-electron* wave functions (ψ) in order to evaluate the epMEs. Aside from the the fact that this entails some inconsistency, this process also suffers from the limitation that the CPMD code is orders of magnitude slower than VASP, so even doing static calculations with the VASP-relaxed structure is time consuming. Additionally, because CPMD uses NCPP, the features of the wave function near the core are significantly modified. While this has minimal effects for defects with mostly s and p orbital character, it becomes a problematic approximation if the defect states involves d-orbitals that are closer to the atomic core. Hence, performing these calculations consistently with the VASP code using PAWs is desirable.

Because the typical output of DFT codes are the pseudo-wave functions ($\tilde{\psi}$) and they are conveniently stored on a Cartesian grid, we would like to express the wave function overlap in terms of the $\tilde{\psi}$'s. The wave function overlap in Eq. (2.57) can be considered a special case of Eq. (2.44) with $\hat{O} = \hat{I}$, the identity:

$$\langle \psi_n | \hat{I} | \psi_m \rangle = \langle \tilde{\psi}_n | \tilde{\psi}_m \rangle + \sum_a \sum_{i,j} \langle \tilde{\psi}_n | \tilde{p}_i^a \rangle \left(\langle \phi_i^a | \tilde{\phi}_j^a \rangle - \langle \tilde{\phi}_i^a | \phi_j^a \rangle \right) \langle \tilde{p}_j^a | \tilde{\psi}_m \rangle \ . \tag{2.58}$$

Hence, we can define an overlap operator $\hat{S}$,

$$\hat{S} = \hat{I} + \sum_a \sum_{i,j} | \tilde{p}_i^a \rangle \left(\langle \phi_i^a | \tilde{\phi}_j^a \rangle - \langle \tilde{\phi}_i^a | \phi_j^a \rangle \right) \langle \tilde{p}_j^a | \ , \tag{2.59}$$

which only depends on the atomic (ϕ) and pseudoatomic ($\tilde{\phi}$) orbitals. The atomic orbital overlaps $\phi_{\{i,j\}}$ are naturally expressed on a radial grid, so evaluating their overlap is straightforward. Now we can express the epME in terms of the $\tilde{\psi}$'s:

$$\left\langle \psi_f \left| \frac{\partial \hat{h}}{\partial Q} \right| \psi_i \right\rangle = (\varepsilon_i - \varepsilon_f) \frac{\left\langle \tilde{\psi}_f \left| \hat{S} \right| \tilde{\psi}_i(\delta Q) \right\rangle}{\delta Q} \ . \tag{2.60}$$

The above process of evaluating the epME using PAWs was implemented in VASP with the help of Manuel Engel and Prof. Georg Kresse from the University of Vienna. With this improved methodology, we have studied the SRH recombination for the Ca defect in GaN, the details of which will discussed in the Chapter 4.

2.5 Auger Recombination

Similarly to the SRH recombination rates, Auger recombination rates can also be calculated with Fermi's golden rule. The method for calculating the Auger coefficient from first principles was first developed by Kioupakis *et al.* in the Van de Walle group [11]. We will give an overview of the methodology and then discuss how it can be extended to systems with strong spin-orbit splitting.

Auger recombination is nominally a three-carrier interaction; if the additional Auger carrier is an electron the process is termed the electron-electron-hole (eeh) process, and if the carrier is a hole the process is termed the hole-hole-electron (hhe) process. While the interaction only involves three charge carriers, the matrix element that enters Fermi's golden rule [Eq. (2.46)]must involve the final state of the electron. The initial (I) and final (F) states of a recombination event are given by two different anti-symmetrized many-body wave functions Ψ_I and Ψ_F. Since the net result of an Auger recombination event is two electrons moving to different states, Ψ_I and Ψ_F only differ in the occupations of two orbitals, and the rate is given by:

$$R_{I\to F} = \frac{2\pi}{\hbar} \left| \left\langle \Psi_I \,\middle|\, \hat{H}_{int} \,\middle|\, \Psi_F \right\rangle \right|^2 \delta(E_F - E_I)\,, \tag{2.61}$$

Auger recombination is facilitated by the two-body screened-Coulomb interaction $\hat{H}_{int} = \hat{W}(\mathbf{r}_1, \mathbf{r}_2)$, which only depends on two positional arguments $\mathbf{r}_1$ and $\mathbf{r}_2$ (we refer to

Ref. [11] for details on how the screened-Coulomb matrix element is calculated). From Eq. (4.22) of Grosso *et al.* [46], we find for the evaluation of a two-body operator between many-body states where two orbitals differ ($m \to \mu$ and $n \to \nu$):

$$\left\langle \Psi_{m,n} \middle| \hat{W} \middle| \Psi_{\mu,\mu} \right\rangle = \underbrace{\left\langle \psi_m \psi_n \middle| \hat{W} \middle| \psi_\mu \psi_\nu \right\rangle}_{M^d_{mn\mu\nu}} - \underbrace{\left\langle \psi_m \psi_n \middle| \hat{W} \middle| \psi_\nu \psi_\mu \right\rangle}_{M^x_{mn\mu\nu}} . \tag{2.62}$$

Just as in the case of SRH recombination, we have reduced the many-body matrix elements to matrix elements that only depend on single-particle states. For simplicity, we will call the two integrals on the right side of the expression the *direct* ($M^d_{mn\mu\nu}$) and *exchange* ($M^x_{mn\mu\nu}$) terms.

We can represent all the states in the band structure by a combined index $\mathbf{i} = (\mathbf{k}, n)$, where $\mathbf{k}$ is the momentum quantum number and n is the band index. For each state $\mathbf{i}$ there are two possible spin quantum numbers $\sigma_\mathbf{i} = \pm\frac{1}{2}$. The Coulomb interaction does not flip spins, so the total Auger recombination rate, obtained by summing over all initial and final states, is given by:

$$\begin{aligned} R^{\text{direct}} = & \frac{2\pi}{\hbar} \sum_{\mathbf{1234}} \sum_{\sigma_1\sigma_2\sigma_3\sigma_4} f_\mathbf{1} f_\mathbf{2} (1 - f_\mathbf{3})(1 - f_\mathbf{4}) \\ & \left| M^d_{\mathbf{1234}} \delta_{\sigma_1\sigma_3} \delta_{\sigma_2\sigma_4} - M^x_{\mathbf{1234}} \delta_{\sigma_1\sigma_4} \delta_{\sigma_2\sigma_3} \right|^2 \\ & \delta(\varepsilon_\mathbf{1} + \varepsilon_\mathbf{2} - \varepsilon_\mathbf{3} - \varepsilon_\mathbf{4}) , \end{aligned} \tag{2.63}$$

where $f_\mathbf{i}$ denotes the Fermi occupation number of the electronic state indexed by $\mathbf{i}$ with energy $\varepsilon_\mathbf{i}$. We have used the superscript "direct" to distinguish this rate from the "indirect" phonon-assisted Auger process that will be discussed later. After the implicit spin summation over the $2^4 = 16$ combined spin configurations, we recover the expression for

the Auger rate from Ref. [11]:

$$
\begin{aligned}
R^{\text{direct}} =& 2\frac{2\pi}{\hbar}\sum_{\mathbf{1234}} f_{\mathbf{1}}f_{\mathbf{2}}(1-f_{\mathbf{3}})(1-f_{\mathbf{4}}) \\
& \left(\left|M^{d}_{\mathbf{1234}}-M^{x}_{\mathbf{1234}}\right|^2+\left|M^{d}_{\mathbf{1234}}\right|^2+\left|M^{x}_{\mathbf{1234}}\right|^2\right) \\
& \delta(\varepsilon_{\mathbf{1}}+\varepsilon_{\mathbf{2}}-\varepsilon_{\mathbf{3}}-\varepsilon_{\mathbf{4}})\,.
\end{aligned} \tag{2.64}
$$

The screened Coulomb interaction $\hat{W}$ conserves momentum between the initial and final states. This fact, combined with the Dirac delta function in Eq. (2.64), enforces both energy and momentum conservation of each Auger recombination event, as shown in Fig. 2.7 (a).

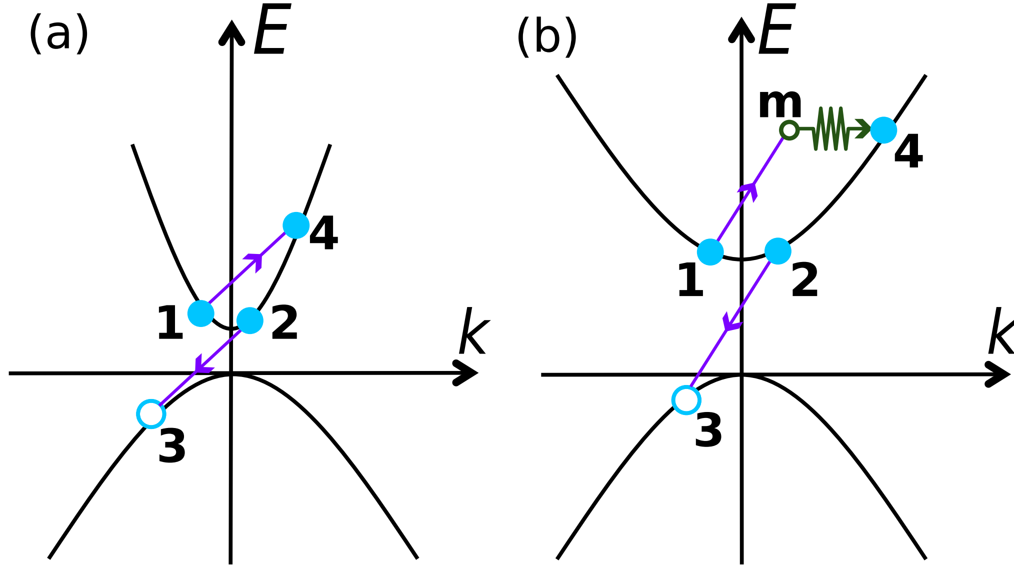


Figure 2.7: Schematic of the (a) direct and (b) phonon-assisted eeh Auger recombination processes.

The strict requirements for energy and momentum conservation turn out to be difficult to satisfy in many wide-band-gap semiconductors (for instance, gallium nitride) resulting in low direct Auger recombination coefficients [47]. In such cases, a phonon-assisted process, which uses the extra momentum of a phonon to facilitate the Auger process, can

become dominant [48]. The rate of the phonon-assisted process is given by a second-order Fermi's golden rule:

$$R_{I\to F} = \frac{2\pi}{\hbar}\left|\sum_M \frac{\mathcal{H}_{IM}\mathcal{H}_{MF}}{E_M - E_I}\right|^2 \delta(E_F - E_I)\,, \tag{2.65}$$

where M represents an intermediate state. The two matrix elements will consist of one screened-Coulomb matrix element plus one epME. An example of such a second-order interaction is shown in Fig. 2.7(b), where $\hat{H}_{\mathrm{IM}} = \left\langle \Psi_I \middle| \hat{W} \middle| \Psi_M \right\rangle$ represents the screened-Coulomb interaction and $\hat{H}_{\mathrm{MF}} = \left\langle \Psi_M \middle| \frac{\partial \hat{H}}{\partial Q} \middle| \Psi_F \right\rangle$ represents the electron-phonon interaction. The electron-phonon scattering process only involves a single electron so the many-body wave functions Ψ_M and Ψ_F can only differ in the occupation of a single orbital. Hence, the scattering rate to the intermediate state Ψ_M state is completely determined by the scattering via single-particle states indexed by $\mathbf{m}$.

By summing over all the initial and final states, and following the same derivation as for the direct Auger process above, we obtain an expression for the phonon-assisted Auger rate (also outlined in Ref. [11]):

$$\begin{aligned} R^{\mathrm{indirect}} =& 2\frac{2\pi}{\hbar}\sum_{\mathbf{1234},\nu\mathbf{q}} f_{\mathbf{1}} f_{\mathbf{2}}(1-f_{\mathbf{3}})(1-f_{\mathbf{4}})\left(n_{\nu\mathbf{q}} + \frac{1}{2} \pm \frac{1}{2}\right) \\ & \times |\tilde{M}_{\mathbf{1234};\nu\mathbf{q}}|^2 \delta(\epsilon_{\mathbf{1}} + \epsilon_{\mathbf{2}} - \epsilon_{\mathbf{3}} - \epsilon_{\mathbf{4}} \mp \hbar\omega_{\nu\mathbf{q}})\,. \end{aligned} \tag{2.66}$$

Here, ν and $\mathbf{q}$ are the band index and wave number of phonons, the $n_{\nu\mathbf{q}}$ are the phonon-occupation numbers from Bose-Einstein statistics:

$$n_{\nu\mathbf{q}} = \frac{1}{e^{\hbar\omega_{\nu\mathbf{q}}/k_{\mathrm{B}}T} - 1}\,, \tag{2.67}$$

and the phonon energy ($\hbar\omega_{\nu\mathbf{q}}$) is accounted for in the energy-conservation delta function.

The upper (lower) sign represents the phonon-emission (-absorption) process. The generalized matrix element that appears in Eq. (2.66) ($\tilde{M}$, with the subscript omitted) is given by:

$$|\tilde{M}|^2 = \left|\tilde{M}^1 + \tilde{M}^2 + \tilde{M}^3 + \tilde{M}^4 - \tilde{M}^5 - \tilde{M}^6 - \tilde{M}^7 - \tilde{M}^8\right|^2 + \left|\tilde{M}^1 + \tilde{M}^2 + \tilde{M}^3 + \tilde{M}^4\right|^2 + \left|\tilde{M}^5 + \tilde{M}^6 + \tilde{M}^7 + \tilde{M}^8\right|^2$$

where each term ($\tilde{M}_1$ to $\tilde{M}_8$) represents an phonon-assisted screened-Coulomb interaction listed below and depicted in Fig. 2.8:

$$\tilde{M}^1_{\mathbf{1234};\nu\mathbf{q}} = \sum_{\mathbf{m}} \frac{g_{\mathbf{1m};\nu} M^d_{\mathbf{m234}}}{\epsilon_{\mathbf{m}} - \epsilon_{\mathbf{1}} \pm \hbar\omega_{\nu\mathbf{q}} + i\eta} \tag{2.68}$$

$$\tilde{M}^2_{\mathbf{1234};\nu\mathbf{q}} = \sum_{\mathbf{m}} \frac{g_{\mathbf{2m};\nu} M^d_{\mathbf{1m34}}}{\epsilon_{\mathbf{m}} - \epsilon_{\mathbf{2}} \pm \hbar\omega_{\nu\mathbf{q}} + i\eta} \tag{2.69}$$

$$\tilde{M}^3_{\mathbf{1234};\nu\mathbf{q}} = \sum_{\mathbf{m}} \frac{M^d_{\mathbf{12m4}} g_{\mathbf{m3};\nu}}{\epsilon_{\mathbf{m}} - \epsilon_{\mathbf{3}} \mp \hbar\omega_{\nu\mathbf{q}} + i\eta} \tag{2.70}$$

$$\tilde{M}^4_{\mathbf{1234};\nu\mathbf{q}} = \sum_{\mathbf{m}} \frac{M^d_{\mathbf{123m}} g_{\mathbf{m4};\nu}}{\epsilon_{\mathbf{m}} - \epsilon_{\mathbf{4}} \mp \hbar\omega_{\nu\mathbf{q}} + i\eta} \tag{2.71}$$

$$\tilde{M}^5_{\mathbf{1234};\nu\mathbf{q}} = \sum_{\mathbf{m}} \frac{g_{\mathbf{1m};\nu} M^x_{\mathbf{m234}}}{\epsilon_{\mathbf{m}} - \epsilon_{\mathbf{1}} \pm \hbar\omega_{\nu\mathbf{q}} + i\eta} \tag{2.72}$$

$$\tilde{M}^6_{\mathbf{1234};\nu\mathbf{q}} = \sum_{\mathbf{m}} \frac{g_{\mathbf{2m};\nu} M^x_{\mathbf{1m34}}}{\epsilon_{\mathbf{m}} - \epsilon_{\mathbf{2}} \pm \hbar\omega_{\nu\mathbf{q}} + i\eta} \tag{2.73}$$

$$\tilde{M}^7_{\mathbf{1234};\nu\mathbf{q}} = \sum_{\mathbf{m}} \frac{M^x_{\mathbf{12m4}} g_{\mathbf{m3};\nu}}{\epsilon_{\mathbf{m}} - \epsilon_{\mathbf{3}} \mp \hbar\omega_{\nu\mathbf{q}} + i\eta} \tag{2.74}$$

$$\tilde{M}^8_{\mathbf{1234};\nu\mathbf{q}} = \sum_{\mathbf{m}} \frac{M^x_{\mathbf{123m}} g_{\mathbf{m4};\nu}}{\epsilon_{\mathbf{m}} - \epsilon_{\mathbf{4}} \mp \hbar\omega_{\nu\mathbf{q}} + i\eta}. \tag{2.75}$$

where $\mathbf{m}$ is the single-particle state index of an intermediate state and η is the inverse lifetime of that intermediate state. The direct (M^d) and indirect (M^x) screened Coulomb matrix elements are defined in Eq. (2.62). The electron-phonon matrix elements (g) between two states, for example $\mathbf{1}$ and $\mathbf{m}$, can be indexed by ν alone since the momentum

difference is precisely the momentum $\mathbf{q}$ of the phonon. To connect with our notation from Sec. 2.4 the epMEs are given by:

$$g_{\mathbf{ab};\nu} = \left\langle \psi_{\mathbf{a}} \left| \frac{\partial \hat{h}}{\partial Q_{\mathbf{q},\nu}} \right| \psi_{\mathbf{b}} \right\rangle . \tag{2.76}$$

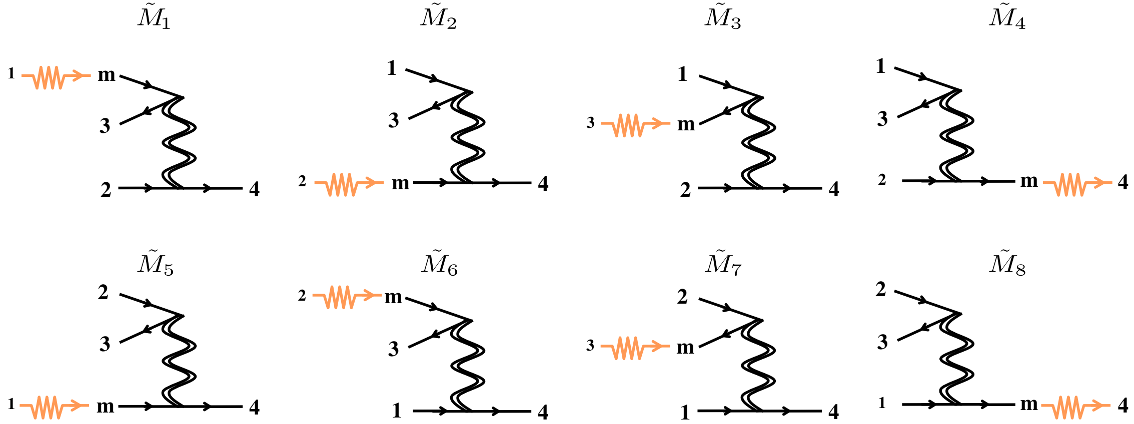


Figure 2.8: Schematic diagrams corresponding to each of the microscopic indirect Auger recombination processes ($\tilde{M}^1$–$\tilde{M}^8$). The expressions are given in Eqs. (2.68)–(2.75).

Unlike the case for the SRH recombination, where we are interested in a particular mode in the a supercell containing a defect, Auger recombination requires the calculation of the epME on a dense grid of phonons. For this purpose, density functional perturbation theory (DFPT) [49] is much better suited compared to the frozen-phonon approach we used for SRH recombination.

Under typical LED operating conditions, the density of electrons (n) and holes (p) can be assumed to be equal. For a particular density we can calculate the Fermi occupation factors $f_{\mathbf{i}}$. Using Eq. (2.64) and Eq. (2.66), we can determine the Auger recombination rates at a given carrier concentration. The Auger coefficient is given by each rate devided

by the cube of the carrier concentration:

$$C_{\{n,p\}}^{\{\text{direct},indirect\}} = \frac{R_{\{n,p\}}^{\{\text{direct},indirect\}}}{n^3}\,, \tag{2.77}$$

where n/p denote the charge of the Auger carrier so it corresponds to the eeh/hhe process.

Considerations for Spin-Orbit Coupling

In materials with heavier atoms, such as indium or lead, relativistic effects can have a dramatic impact on the band structure and thus the Auger coefficients. The two main types of relativistic effects that need to be accounted for in solid-state systems are scalar-relativistic effects and spin-orbit coupling (SOC). The Kohn-Sham equation contains neither of these effects, but their inclusion into the DFT framework is well established [50, 51].

The relativistic effects can be incorporated into DFT by way of an effective perturbing Hamiltonian [50]:

$$\Delta\hat{H}_{rel} = \Delta\hat{H}_{\text{SR}}[n] + \Delta\hat{H}_{\text{SOC}}\,. \tag{2.78}$$

The scalar relativistic term is included by default in most modern DFT implementations. The SOC term is given by:

$$\Delta\hat{H}_{\text{SOC}} = \sum_{l\neq 0} |l\rangle\, V_l(r)\hat{\mathbf{L}}\cdot\hat{\mathbf{S}}\,\langle l|\,, \tag{2.79}$$

where $\hat{\mathbf{L}}$ is the orbital angular momentum operator and $\hat{\mathbf{S}}$ is the spin operator. A rigorous determination of the $V_l(r)$ term involves treatment of the Dirac equations and will not be discussed here. We simply note that the perturbing Hamiltonian can be obtained for NCPPs [51] and rewritten in the basis of solutions to the Kohn-Sham equation [Eq. (2.20)] with two insertions of the identity. The new Hamiltonian can be written and diagonalized

at each k point in reciprocal space:

$$\hat{H}_{\mathrm{KS}}(\mathbf{k}) + \Delta\hat{H}_{\mathrm{SOC}}(\mathbf{k}) \tag{2.80}$$

$$= \epsilon_{\mathbf{k},i,\sigma_i}\delta_{i,j}\delta_{\sigma_i,\sigma_j} + \sum_{i,j} |\psi_{i,\sigma_i}(\mathbf{k})\rangle \langle\psi_{i,\sigma_i}(\mathbf{k})|\Delta\hat{H}_{\mathrm{SOC}}|\psi_{j,\sigma_j}(\mathbf{k})\rangle \langle\psi_{j,\sigma_j}(\mathbf{k})| \ , \tag{2.81}$$

where $\hat{H}_{\mathrm{KS}}(\mathbf{k})$ is the Hamiltonian without SOC and $\psi_{i,\sigma_i}(\mathbf{k})$ are the wave function solutions of $\hat{H}_{\mathrm{KS}}(\mathbf{k})$ indexed by the momentum ($\mathbf{k}$), band (i), and spin (σ_i) quantum numbers. The eigenvalues for the system with SOC are obtained by diagonalizing this new Hamiltonian.

In the non-SOC case, the two spin states $\psi_{i,\uparrow}$ and $\psi_{i,\uparrow}$ are degenerate, so they are pure product states of the wave function $\psi_i(\mathbf{k})$ and the spin:

$$|\psi_i(\mathbf{k})\rangle \otimes |{\uparrow}\rangle \ . \tag{2.82}$$

With the inclusion of SOC, the perturbing Hamiltonian couples to the spin and σ_i is no longer a valid quantum number. Hence, the spin-degenerate wave functions must be replaced with a spinor wave function where the two spin components are independent:

$$|\psi_i^{\uparrow}(\mathbf{k})\rangle \otimes |{\uparrow}\rangle + |\psi_i^{\downarrow}(\mathbf{k})\rangle \otimes |{\downarrow}\rangle \ , \tag{2.83}$$

where we have condensed the crystal momentum and band number into a combined index $\mathbf{i} = (\mathbf{k}, n)$.

Assuming we have N spin-degenerate bands in our non-SOC system, the transformation from the size-$2N$ spin-degenerate basis of $\hat{H}_{\mathrm{KS}}$ (indexed by j and σ_j) to the size-$2N$

spin basis of $\hat{H}_{\text{KS}} + \Delta\hat{H}_{\text{SOC}}$ is given by:

$$U_{i,j\sigma_j} = \Big(|\psi_i^{\uparrow}\rangle \otimes |\uparrow\rangle + |\psi_i^{\downarrow}\rangle \otimes |\downarrow\rangle\Big) \langle\psi_j| \otimes \langle\sigma_j| \ . \tag{2.84}$$

Without the degenerate spin, we will not be able to do the implicit spin summation that converted Eq. (2.63) to Eq. (2.64). As such, the direct Auger rate for systems with SOC is given by:

$$\begin{aligned} R^{\text{direct}} =& \frac{2\pi}{\hbar} \sum_{\mathbf{1234}} f_{\mathbf{1}} f_{\mathbf{2}} (1 - f_{\mathbf{3}})(1 - f_{\mathbf{4}}) \\ & \left|\mathcal{M}^{d}_{\mathbf{1234}} - \mathcal{M}^{x}_{\mathbf{1234}}\right|^2 \delta(\varepsilon_{\mathbf{1}} + \varepsilon_{\mathbf{2}} - \varepsilon_{\mathbf{3}} - \varepsilon_{\mathbf{4}}) \, . \end{aligned} \tag{2.85}$$

Here the Arabic-numeral is the index of the new spinor basis, and the matrix elements ($\mathcal{M}$) are transformed from the spin-degenerate basis by $U_{i,j\sigma_j}$. The direct and exchange terms in the new basis are given by:

$$\begin{aligned} \mathcal{M}^{d}_{\mathbf{1234}} \equiv & \left\langle \mathbf{12}|\hat{W}|\mathbf{34} \right\rangle \\ =& \langle\psi^{\uparrow}_{\mathbf{1}}\psi^{\uparrow}_{\mathbf{2}}|\hat{W}|\psi^{\uparrow}_{\mathbf{3}}\psi^{\uparrow}_{\mathbf{4}}\rangle + \langle\psi^{\uparrow}_{\mathbf{1}}\psi^{\downarrow}_{\mathbf{2}}|\hat{W}|\psi^{\uparrow}_{\mathbf{3}}\psi^{\downarrow}_{\mathbf{4}}\rangle \\ & \langle\psi^{\downarrow}_{\mathbf{1}}\psi^{\uparrow}_{\mathbf{2}}|\hat{W}|\psi^{\downarrow}_{\mathbf{3}}\psi^{\uparrow}_{\mathbf{4}}\rangle + \langle\psi^{\downarrow}_{\mathbf{1}}\psi^{\downarrow}_{\mathbf{2}}|\hat{W}|\psi^{\downarrow}_{\mathbf{3}}\psi^{\downarrow}_{\mathbf{4}}\rangle \; and \end{aligned} \tag{2.86}$$

$$\begin{aligned} \mathcal{M}^{x}_{\mathbf{1234}} \equiv & \left\langle \mathbf{12}|\hat{W}|\mathbf{43} \right\rangle \\ =& \langle\psi^{\uparrow}_{\mathbf{1}}\psi^{\uparrow}_{\mathbf{2}}|\hat{W}|\psi^{\uparrow}_{\mathbf{4}}\psi^{\uparrow}_{\mathbf{3}}\rangle + \langle\psi^{\uparrow}_{\mathbf{1}}\psi^{\downarrow}_{\mathbf{2}}|\hat{W}|\psi^{\uparrow}_{\mathbf{4}}\psi^{\downarrow}_{\mathbf{3}}\rangle \\ & \langle\psi^{\downarrow}_{\mathbf{1}}\psi^{\uparrow}_{\mathbf{2}}|\hat{W}|\psi^{\downarrow}_{\mathbf{4}}\psi^{\uparrow}_{\mathbf{3}}\rangle + \langle\psi^{\downarrow}_{\mathbf{1}}\psi^{\downarrow}_{\mathbf{2}}|\hat{W}|\psi^{\downarrow}_{\mathbf{4}}\psi^{\downarrow}_{\mathbf{3}}\rangle \, . \end{aligned} \tag{2.87}$$

The same basis transformation can be applied to the epME (g) in the expression for

the phonon-assisted Auger rate to give:

$$
\begin{aligned}
R^{\text{indirect}} =& \frac{2\pi}{\hbar} \sum_{\mathbf{1234},\nu\mathbf{q}} f_{\mathbf{1}} f_{\mathbf{2}} (1 - f_{\mathbf{3}})(1 - f_{\mathbf{4}}) \left(n_{\nu\mathbf{q}} + \frac{1}{2} \pm \frac{1}{2} \right) \\
& \times |\tilde{\mathcal{M}}_{\mathbf{1234};\nu\mathbf{q}}|^2 \delta(\epsilon_{\mathbf{1}} + \epsilon_{\mathbf{2}} - \epsilon_{\mathbf{3}} - \epsilon_{\mathbf{4}} \mp \hbar\omega_{\nu\mathbf{q}}) \,,
\end{aligned} \tag{2.88}
$$

where the new generalized matrix element is given by:

$$
|\tilde{\mathcal{M}}|^2 = \left| \tilde{\mathcal{M}}^1 + \tilde{\mathcal{M}}^2 + \tilde{\mathcal{M}}^3 + \tilde{\mathcal{M}}^4 - \tilde{\mathcal{M}}^5 - \tilde{\mathcal{M}}^6 - \tilde{\mathcal{M}}^7 - \tilde{\mathcal{M}}^8 \right|^2 . \tag{2.89}
$$

Each of the terms ($\tilde{\mathcal{M}}^1$ to $\tilde{\mathcal{M}}^8$) are defined similarly to Eq. (2.68)–(2.75) in the new spinor basis. In practice, both the spinor wave functions and the transformation matrices are stored during the calculation. The spinor wave functions are used directly in the calculation of $\tilde{\mathcal{M}}^d$ and $\tilde{\mathcal{M}}^x$, while the transformation matrices are used to transform the epME obtained from the DFPT calculation on the non-SOC system.

Chapter 3

Electronic structure of BAlN alloys

The band gap is one of the most important features of the electronic structure of a material. As we will see later, the SRH and Auger recombination mechanisms both show strong dependence on the value of the band gap. In this chapter we will present a case study of how the band gap of a material can change based on the chemical composition. In an alloy system, translational symmetry is broken so Eq. (2.29) is no longer satisfied for G, the reciprocal lattice vectors of the unit cell. Alloys can be simulated in supercells that explicitly include the chemical and positional disorder. While the wave function no longer has the periodicity of the unit cell, some features of the bulk wave function still remain in the alloy wave function. We will focus on the wave function characteristics of the VBM and CBM states, and use the information to track the eigenvalues of the band-edge states as the chemical composition changes.

The material system we will focus on is boron aluminum nitride (BAlN), an alloy of boron nitride and aluminum nitride. Boron-containing nitride alloys such as BAlN are being explored as novel members of the nitride family of materials for electronic and optoelectronic applications. Using hybrid density functional calculations we determine structural properties, band gaps and band-gap bowing of random wurtzite BAlN alloys.

The fundamental band gap of BN is indirect while AlN is a direct-band-gap semiconductor. This leads to a crossover in the band gap from direct to indirect at 28% boron. We find that the direct band gap experiences very large bowing, leading to a fundamental gap that changes very little up to 17% boron incorporation.

3.1 Introduction

Ternary alloys based on wurtzite AlN, GaN and InN have enabled the development of light-emitting diodes [52], laser diodes [53] and high-power electronics [54]. Large-band-gap nitride alloys are required to push optoelectronic devices into the ultraviolet (UV) spectrum, and large polarization discontinuities at the interface of III-nitride heterostructures would enable devices that host high-density two-dimensional electron gases (2DEGs). Use of boron nitride as a member of the III-nitride family may help in achieving these goals. Wurtzite BN is predicted to have the largest spontaneous polarization [55] among the III-nitrides [56]. Alloying boron into AlN could expand the range of band gaps and polarization charges that can be accessed by alloys of the III-nitrides, making BAlN alloys a promising material for an array of technological applications.

The ground state of BN is hexagonal [57], but the wurtzite phase [see Fig. 3.1(a)] is of high interest in light of experimental efforts to alloy boron at low concentrations into wurtzite AlN using metal-organic vapor phase epitaxy (MOVPE) [58] or molecular-beam-epitaxy (MBE) [59, 60]. While the large lattice mismatch between BN and AlN leads to a large miscibility gap [61], BAlN alloys are still expected to be stable in the wurtzite phase [58] for low boron concentrations. Given the potential technological impact of BAlN alloys, it is essential to know their structural and electronic properties and how they compare to those of the parent compounds, AlN and BN.

Since BN is not stable in the wurtzite phase, no experimental information is available to enable predictions of the properties of wurtzite BAlN alloys, highlighting the

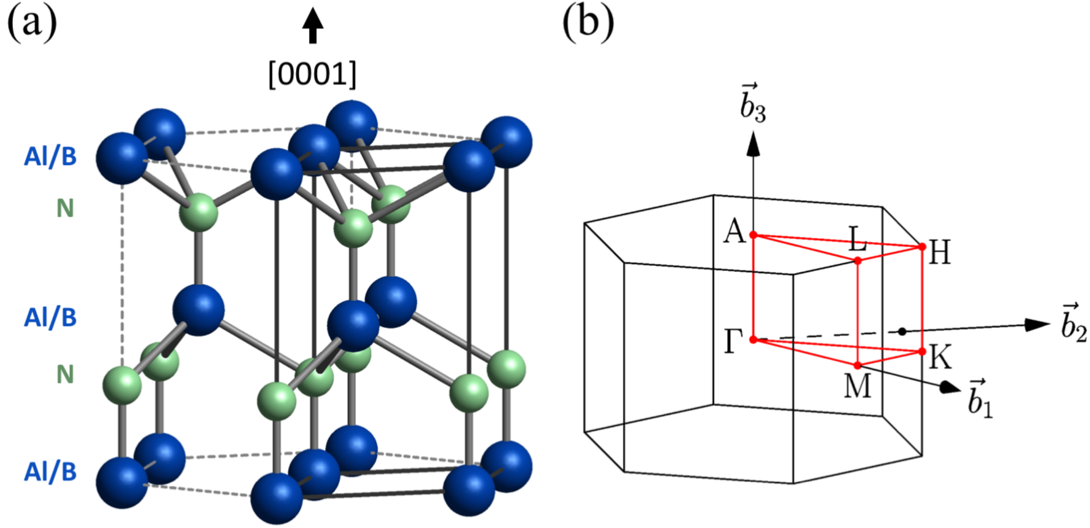


Figure 3.1: (a) The crystal structure and (b) Brillouin zone of wurtzite AlN and BN. The cation sites (blue) can be occupied by either Al or B, while the anion sites (green) are occupied by N. The solid black lines in (a) indicate the primitive cell and the red lines in (b) indicate the standard high symmetry path.

importance of predictive first-principles calculations. The electronic structure of zinc-blende [62, 63, 64] and wurtzite [65] BAlN alloys has been examined in a number of density functional calculations. These studies used the local density approximation (LDA), which is known to significantly underestimate the band gaps of semiconductors. Kumar *et al.* [64] attempted to correct for the band gap by using the modified Becke-Johnson functional for their calculations of zinc-blende BAlN alloys. However, the band gaps of the parent compounds were still severely underestimated, affecting a quantitative description of the alloy electronic structure. Zhang *et al.* [65] used a scissor shift to correct the LDA band gaps in their calculations of wurtzite BAlN alloys. However, their procedure for identifying the character of conduction-band extrema is unclear. In addition, some of their alloy band structures included low-dispersion bands that could be indicative of localized states, which can lead to a spurious reduction in the calculated band gap. Our present work overcomes these problems by using a hybrid functional to consistently calculate structural as well as electronic properties of wurtzite $B_xAl_{1-x}N$ alloys. Hybrid

density functional theory (DFT) has previously been successfully applied to the parent compounds, wurtzite AlN [66] and hexagonal BN [55].

Since AlN is a direct-band-gap semiconductor and BN is indirect, a crossover from direct to indirect band gap will occur. If information about the wurtzite phase of BN is available, one can attempt linear interpolation of the band gaps to identify the boron concentration at which the direct-to-indirect crossover occurs. However, we will see that the direct band gap of BAlN is strongly nonlinear as a function of boron content, i.e., it exhibits a large bowing, while the bowing of the indirect band gap is much weaker. Explicit alloy calculations are therefore essential. We identify the direct and indirect band gaps and predict the direct-to-indirect crossover to occur at 28% boron incorporation. While maintaining a direct gap up to high boron concentration is favorable for optoelectronic applications, a key conclusion from our work is that the value of this gap barely increases above the AlN value.

3.2 Methodology

3.2.1 First-Principles Calculations

Our DFT calculations use the hybrid functional of Heyd, Scuseria and Ernzerhof (HSE) [24] as implemented in the VASP code [67, 68]. The mixing parameter α is set to 0.33, which leads to an accurate description of the band gaps and structural properties of the parent compounds, AlN and BN, in their respective ground-state phases. Our calculations use projector augmented wave (PAW) potentials [31] and a plane-wave cutoff energy of 420 eV. For the calculations of the primitive unit cells of the parent compounds we use a $8 \times 8 \times 6$ Γ-centered k point grid. To account for the Van der Waals interactions in our calculations of hexagonal BN (*h*-BN) we use the semi-empirical

Grimme-D3 method [69]. For the alloy supercell calculations, a $2 \times 2 \times 2$ Monkhorst-Pack grid is used.

The calculations of $B_xAl_{1-x}N$ alloys are performed using a $3 \times 3 \times 2$ wurtzite supercell that contains 72 atoms. For each boron concentration, we generate ten BAlN alloy structures with a random distribution of B atoms. The lattice mismatch between the two parent compounds is large (19%). The lattice parameters are varied linearly as a function of boron content in accordance with Vegard's law. For example, the in-plane lattice parameter, $\mathrm{a}_{\mathrm{B}_x\mathrm{Al}_{(1-x)}\mathrm{N}}$ of the BAlN alloy is:

$$a_{\mathrm{B}_x\mathrm{Al}_{1-x}\mathrm{N}} = x a_{\mathrm{BN}} + (1-x) a_{\mathrm{AlN}} \tag{3.1}$$

where x is the boron content, a_{BN} is the in-plane lattice parameter of wurtzite BN (*wz*-BN) and a_{AlN} is the in-plane lattice parameter of *wz*-AlN. The atomic positions within the supercell were allowed to relax using HSE until all of the forces are below 20 meV/Å. We verified the accuracy of Vegard's law for a subset of alloy structures by allowing both the lattice parameters and atomic positions to relax. In all cases, the variation in the lattice parameters was close to linear, i.e., it followed Vegard's law.

3.2.2 Determination of Alloy Band Edges

The conduction-band minima of wurtzite AlN and BN lie at different high-symmetry points in the Brillouin zone (BZ) [Fig. 3.1(b)]; the CBM is at Γ in AlN and at K in BN. At a critical boron concentration, a direct-to-indirect crossover will occur. The alloy calculations are performed in supercells, which have a smaller Brillouin zone than the parent compounds. Zone folding will occur, which complicates the identification of direct and indirect band gaps in the alloy band structures. Calculations of direct-to-indirect crossovers in alloy band gaps can be based on identifying band degeneracies [70, 64].

However, examining the degeneracies is challenging for alloys of wz-AlN because some conduction-band extrema at other high-symmetry points are similar in energy to the value at K, and many supercell configurations will fold these points together. In addition, the large size mismatch between Al and B leads to large lattice relaxations that lower the symmetry and split degeneracies, making it even more challenging to identify conduction band states that correspond to AlN or BN via band degeneracies. To overcome these challenges, we utilize a different approach, which is in the spirit of unfolding the band structure of an alloy supercell by projecting the wave functions onto the supercell of a pure material [71, 72].

Our projection scheme is based on $3 \times 3 \times 2$ BAlN alloy supercells and pristine AlN and BN supercells strained to the same size. Since we are interested in BAlN alloys at low B content, we project onto wave functions in an AlN supercell. The K point, which occurs at the corner of the Brillouin zone (1/3,1/3,0) for the 4-atom primitive cell of AlN, is folded to Γ in the $3 \times 3 \times 2$ supercell, as illustrated in Fig. 3.2.

The band structure of the primitive cell of AlN is shown in Fig. 3.3(a). The conduction-band states at Γ and K (Γ_{c} and $\mathrm{K_c}$) as well as the valence-band state at Γ (Γ_{v}) are all folded to Γ (Γ^{AlN}) in the AlN supercell. The supercell states that correspond to states of interest in the primitive cell can be identified by their eigenenergies, atomic orbital contributions, and degeneracies. For the BAlN supercells, we project the states at Γ (Γ^{SC}) onto the states of the AlN supercell that we identify with Γ_{v}, Γ_{c} and $\mathrm{K_c}$. Each band in the alloy supercell gives us a projection weight defined as:

$$\left|\langle \Gamma^{\mathrm{AlN}}, n | \Gamma^{\mathrm{SC}}, m \rangle\right| \ , \tag{3.2}$$

where $|\Gamma^{\mathrm{SC}}, m\rangle$ is a state at Γ in the BAlN alloy supercell with band index m, and $|\Gamma^{\mathrm{AlN}}, n\rangle$

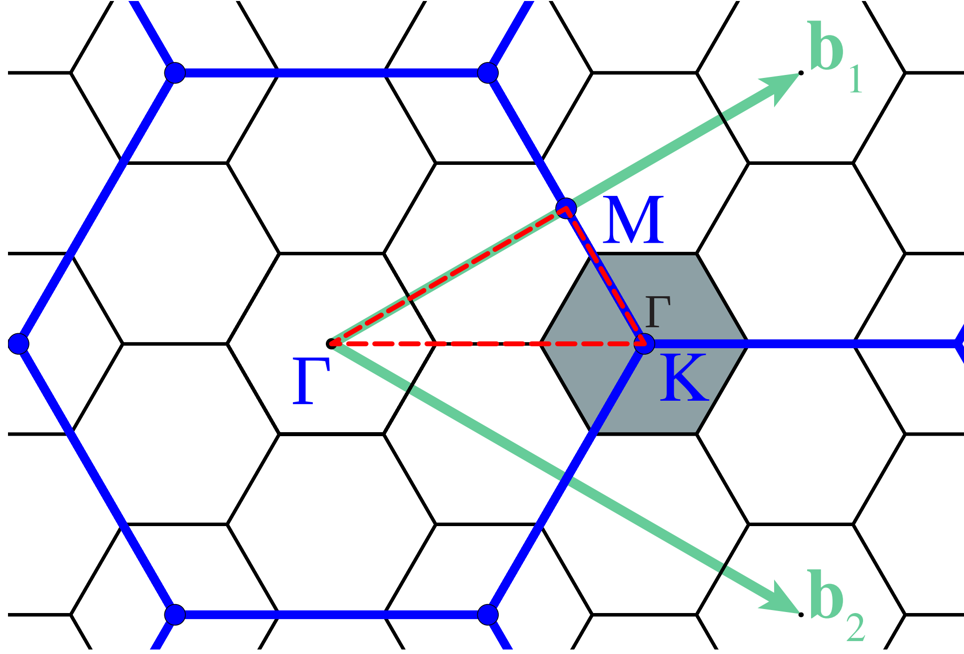


Figure 3.2: In-plane reciprocal lattice for the wurtzite structure. The Brillouin zone corresponding to the primitive unit cell is indicated in blue, and the reciprocal lattice vectors $\mathbf{b_1}$ and $\mathbf{b_2}$ are shown. The high-symmetry path is shown in red. The Brillouin zone of the $3 \times 3 \times 2$ supercell is shown in black. Note that the K point corresponding to the primitive unit cell is folded onto the Γ point of the supercell.

is one of the states at Γ identified above for the (strained) pristine AlN supercell.

We compute these projection weights for each band index m in the BAlN alloy supercell. This allows us to identify the states at Γ^{SC} in the BAlN supercell that have the largest projection weight associated with the Γ_{v}, Γ_{c} and K_{c} states. Once the states of interest have been identified, the direct band gap is taken to be the energy difference between the eigenvalues of the states that have strongest projection weight onto Γ_{c} and Γ_{v}, and the indirect band gap is given by the energy difference between the eigenvalues of the BAlN states that have the strongest projection weight onto K_{c} and Γ_{v}.

To take different atomic arrangements in the alloy into account, ten randomly constructed atomic configurations are generated at each boron content. We do not use Special Quasi-random Structures (SQS) [73] here. While the SQS approach is often pro-

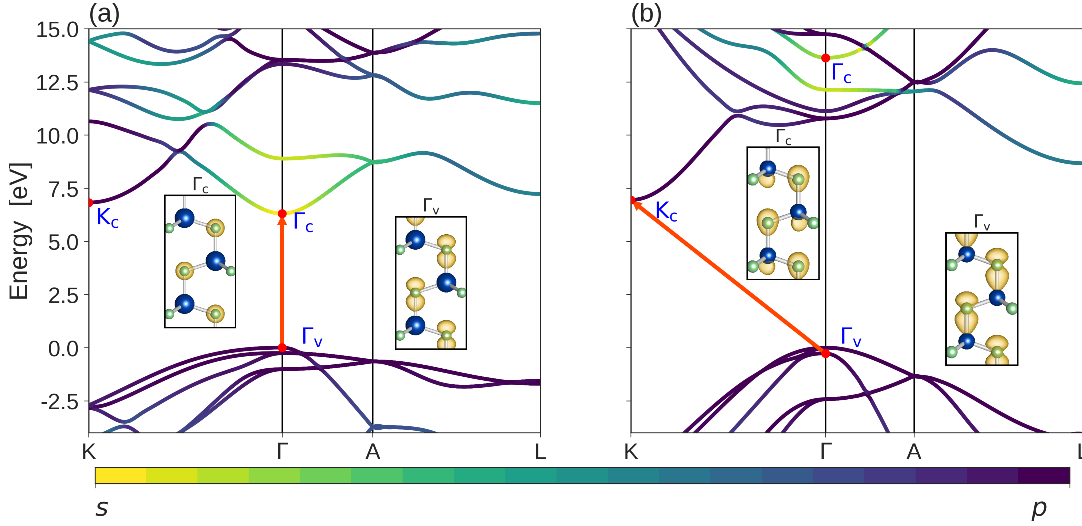


Figure 3.3: Band structures of wurtzite bulk (a) AlN and (b) BN calculated with the HSE functional. The color of each band indicates the angular momentum character of the states, according to the color bar below the plot. The VBM at Γ was used as the zero-energy reference in each plot. The isosurfaces correspond to the Γ_v and Γ_c wave functions are shown for each material.

ductively used to model alloy properties, the shape of the SQS supercell would preclude performing the projections onto states of the parent compounds that allow identifying the states corresponding to the CBM at Γ and K. Full atomic relaxation is allowed for each of our ten random configurations. *Average* direct and indirect band gaps at a given boron content are then obtained by averaging the band gaps calculated for these ten atomic configurations of our alloy supercells. We feel that this approach to determining the band edges approximates the results that would be obtained from experimental measurements that probe the electronic structure of the alloy.

By identifying the direct and indirect band gaps of BAlN as a function of boron content we can evaluate the bowing parameters for each transition. For example, the

bowing parameter b^{dir} for the direct band gap of BAlN is defined as:

$$E_g^{\mathrm{dir}}(\mathrm{B}_x\mathrm{Al}_{1-x}\mathrm{N}) = xE_g^{\mathrm{dir}}(\mathrm{BN}) + (1-x)E_g^{\mathrm{dir}}(\mathrm{AlN}) - b^{\mathrm{dir}}x(1-x) \quad (3.3)$$

where $E_g^{\mathrm{dir}}(\mathrm{BN})$ is the direct band gap of wz-BN and $E_g^{\mathrm{dir}}(\mathrm{AlN})$ is the direct band gap of AlN (both at Γ). A similar bowing parameter can be defined for the indirect band gap using the constituent indirect band gaps (corresponding to the CBM at K) of wurtzite AlN and BN.

3.3 Results and Discussions

3.3.1 Parent Compounds

We first examine the structural and electronic properties of AlN and BN. AlN is stable in the wurtzite structure (Fig. 3.1). The ground state of BN is the hexagonal structure, in which B and N atoms are arranged in an in-plane, sp^2 bonded honeycomb lattice. The BN layers are weakly bonded to each other by Van der Waals forces in an AB stacking configuration. BN can also be stabilized in the wurtzite structure, which is of primary interest for the present study.

The structural and electronic properties of AlN and BN in the wurtzite and hexagonal phases are summarized in Table 3.1. Our choice of $\alpha = 0.33$ for the HSE mixing parameter yields lattice parameters for wz-AlN and h-BN that are generally within 1% of the experimental parameters for each material, with only a slightly larger deviation (1.5%) for the c lattice parameter of h-BN. We also obtain a direct band gap of 6.18 eV for wz-AlN, which is within the range of experimentally measured gaps (6.12 eV − 6.19 eV) [74, 75]. For wz-BN, no experimental values are available, but $\alpha = 0.33$ produces an

Table 3.1: Lattice parameters a and c and band gaps E_g of bulk AlN and BN in their wurtzite (wz) and hexagonal (h) phases. Values calculated with the HSE functional are compared with results from experimental studies. The nature of the transition (VBM $\rightarrow$ CBM) corresponding to the direct and indirect gaps E_g is shown in parentheses. The calculated values for the transitions of interest for the wurtzite materials are shown in the last two columns.

Material	Method	a [Å]	c [Å]	E_g [eV]	$\Gamma_v \rightarrow \Gamma_c$ [eV]	$\Gamma_v \rightarrow K_c$ [eV]
wz-AlN	HSE	3.08	4.93	6.18 ($\Gamma \rightarrow \Gamma$)	6.18	6.82
	Exp.	3.11[a]	4.98[a]	6.12–6.19[b]	-	-
h-AlN	HSE	3.26	4.10	5.42 ($\Gamma \rightarrow \Gamma$)	-	-
	Exp.	-	-	-	-	-
wz-BN	HSE	2.52	4.17	6.84 ($\Gamma \rightarrow$ K)	13.90	7.21
	Exp.	2.55[c]	4.21–4.22[c]	-	-	-
h-BN	HSE	2.49	6.56	5.98 ($\sim$ H $\rightarrow\sim$ M)	-	-
	Exp.	2.50–2.51[d]	6.66-6.68[d]	5.75–6.08[e]	-	-

[a] Ref. [77]
[b] Ref. [74, 75]
[c] Ref. [78, 79]
[d] Ref. [80, 81]
[e] Ref. [82, 76]

indirect band gap of 5.98 eV for h-BN, very close to the indirect band gap of 6.08 eV [76] measured in bulk h-BN.

The HSE-calculated band structures of wurtzite AlN and BN are shown in Fig. 3.3. The orbital contribution to the states in the band structure is illustrated by a distinct color: yellow for s-states and purple for p-states. The fundamental band gap of AlN is direct at Γ. The CBM of AlN at Γ is primarily composed of s states while the VBM at Γ is primarily composed of p states. The wurtzite crystal field splits the VBM of AlN into a doubly degenerate Γ_6 and a singly degenerate Γ_1 state. The topmost Γ_1 valence-band state in AlN (labeled Γ_v) has strong p_z character, as evidenced by the isosurface shown

in the inset of Fig. 3.3(a).

Wurtzite BN, unlike AlN, is an indirect band gap semiconductor with the VBM at Γ and CBM at K. For the purposes of expressing the direct and indirect band gaps of BAlN alloys with respect to those of the parent compounds (and extracting bowing parameters) we need to focus on the band gaps of *wz*-AlN and *wz*-BN determined relative to the band edges that have the same character. Since we are interested in BAlN alloys with low B content, we focus on identifying valence- and conduction-band states in BN that have the same character as in AlN. In *wz*-BN, the Γ_1 valence-band state with p_z character occurs at 0.27 eV below the VBM; this state is labeled Γ_v in Fig. 3.3(b). For the conduction band, the singly degenerate Γ_c conduction-band state in *wz*-AlN has its equivalent in a state in *wz*-BN [also labeled Γ_c in Fig. 3.3(b)] that is composed primarily of *s* states at 13.90 eV above Γ_v.

3.3.2 Alloy Structure

In a BAlN alloy, each B and Al cation is tetrahedrally coordinated by four N atoms. In our *wz*-AlN calculations the Al-N bond length is 1.90 Å for the axial bond along the *c*-axis and 1.89 Å for the other bonds (which we refer to as planar bonds). In *wz*-BN the axial bond length is 1.58 Å and the planar bond length 1.56 Å. We have analyzed the distribution of nearest-neighbor Al-N and B-N bond lengths as a function of boron concentration. For each concentration and configuration we average over the planar and axial Al-N and B-N bond lengths. The results are illustrated in Fig. 3.4 for the lowest and highest boron concentrations investigated in our study. While Fig. 3.4 shows the bond length averaged over ten atomic configurations, the bond-length distribution for each individual configuration closely resembles the average. The nearest-neighbor bond lengths exhibit a bimodal distribution, peaked near the bond lengths of the parent

compounds. For the lowest boron concentration that we investigate ($x = 0.03$) the distribution in the B-N bond lengths is consistent with the bond lengths one would expect from ternary alloys in the dilute limit [83]. As the boron content increases up $x = 0.17$, the distributions of the Al-N and B-N bond lengths are both broadened, as illustrated in Fig. 3.4.

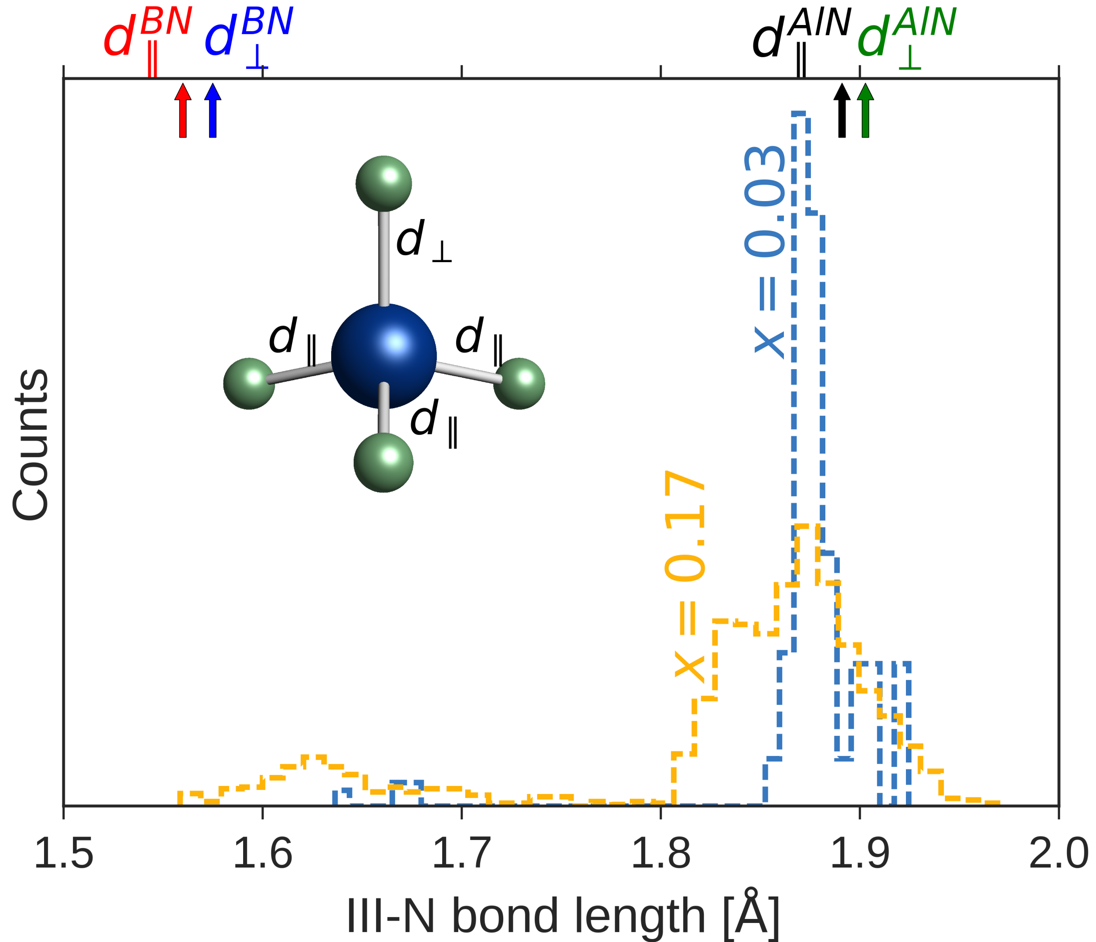


Figure 3.4: Distribution of nearest neighbor cation-nitrogen bonds for boron concentrations of x=0.03 and x=0.17. The in-plane ($d_{\parallel}$) and out-of-plane ($d_{\perp}$) bond lengths for AlN and BN are indicated on the top axis.

3.3.3 Alloy Band Structure

We now use the projection scheme outlined in Sec. 3.2.2 to identify the direct and indirect band gaps of $B_xAl_{1-x}N$ alloys. We limit our calculations to a boron content of x=0.17; for higher boron concentrations the band-edge states for the alloy exhibit larger contributions from BN, and our scheme of identifying states based on projecting on AlN states becomes less reliable. We checked for the presence of localized states within the band gap by plotting the wave functions of states in the vicinity of the band edges. Our alloy band structures calculated consistently within HSE do not show any evidence of localized states.

Since AlN has a direct gap and BN an indirect gap, we expect to see a direct-to-indirect crossover in the band gap of wurtzite BAlN alloys at a critical boron concentration. Linear interpolation between the gaps of the parent compounds would place this critical boron concentration at x=0.08 (see Fig. 3.5). However, alloy band gaps exhibit bowing, which can be particularly large if the lattice mismatch is large, and these non-linearities will affect the crossover. One expects different bowing parameters [Eq. (3.3)] for the direct and indirect band gaps of BAlN. A least-squares fit of the calculated BAlN band gaps to a second order polynomial leads to a bowing parameter of 8.55 eV for the direct band gap and a bowing parameter of 1.49 eV for the indirect gap. Using these bowing parameters to describe the alloy band gaps leads to the direct-to-indirect crossover occurring at 28% boron, illustrated by the vertical line in Fig. 3.5.

The description of the band gap with Eq. (3.3) using a single bowing parameter is only an approximation, particularly in the case of large lattice mismatch and strong bowing. A better description might be obtained by using a higher-order polynomial; alternatively, a composition-dependent bowing parameter can be defined by determining a value for b based on Eq. (3.3) at each composition x. It turns out that, over the range of boron

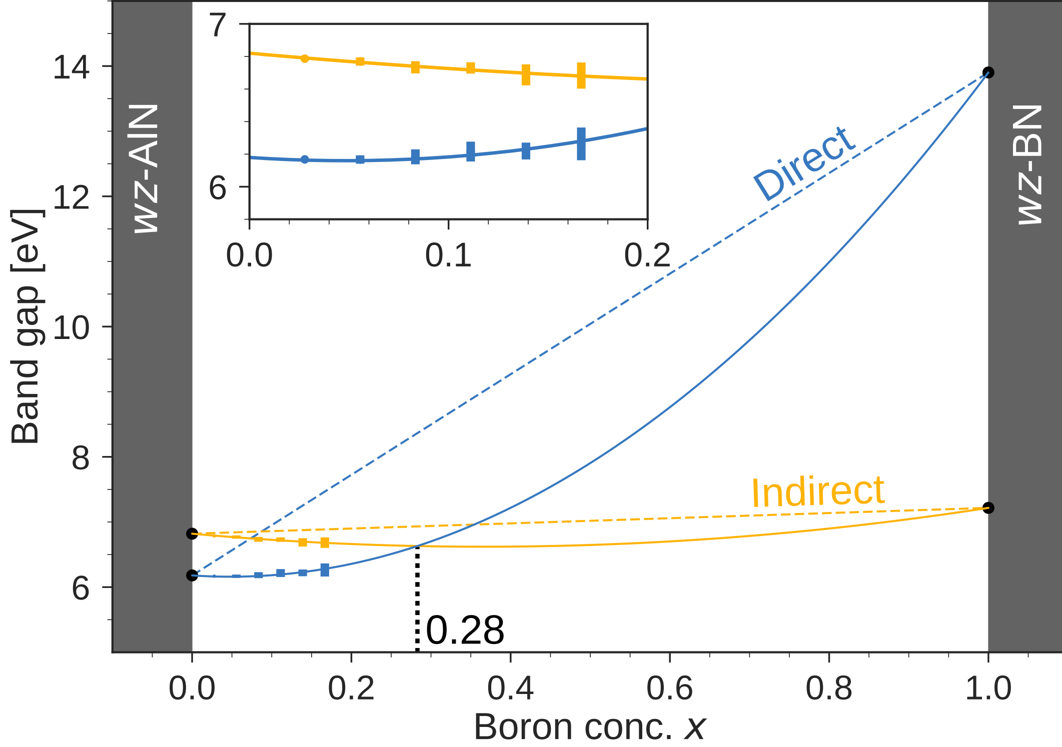


Figure 3.5: Direct $\Gamma_v \rightarrow \Gamma_c$ and indirect $\Gamma_v \rightarrow K_c$ band gaps of wurtzite $B_xAl_{1-x}N$ alloys as a function of boron concentration x. Linear interpolations of the direct and indirect band gaps are indicated by dashed lines. The spread of calculated values at each concentration is illustrated by the vertical bars. The solid curves indicate a quadratic fit to the calculated data [Eq. (3.3)]. The crossover between direct and indirect band gap at 28% is indicated by the vertical dotted line. The inset zooms in on the concentration range between 0 and 0.2.

concentrations considered in our study, the bowing parameter is not particularly sensitive to the composition. For the direct gap, the bowing parameter b^{dir} varies from 8.38 eV at x=0.03 to 8.67 eV at x=0.17, small variations compared to the value of $b^{\text{dir}} = 8.55$ eV obtained by a fit over the entire range. Similarly, for the indirect gap, the bowing parameter varies from 1.65 eV at x=0.03 to 1.46 eV at x=0.17 (compared to 1.49 eV obtained from the full-range fit). We conclude the direct and indirect band gaps of the BAlN band gaps are well described by a single bowing parameter.

The spread (standard deviation) in the direct and indirect band gaps for each set

of atomic configurations at a given boron content is illustrated in the inset of Fig. 3.5, and tabulated in Table 3.2. For boron concentrations up to 17% this spread is quite small. At the lowest simulated boron concentration (x =0.03) only one Al atom in the supercell is replaced by B, which means that only a single atomic configuration needs to be calculated. At the next boron concentration (x =0.06), the spread in band-gap values is still very low. The band-gap value in the structure where the boron atoms are furthest apart compared to the structure where the boron atoms are closest together (i.e., on nearest-neighbor cation sites) is 41 meV higher for the direct band gap and 62 meV higher for the indirect band gap. At x=0.17 the spread in the direct band gap is 100 meV and the spread in the indirect band gap is 80 meV.

The success of our projection scheme is based on our ability to project states of the alloy onto states of AlN and thus distinguish states that have Γ_c and K_c character. As an example, for the K_C state, the (normalized) magnitude of the projection [Eq. (3.2)] goes from 0.98 at x=0.03 to 0.53 at x =0.17. This is of course accompanied by an increased energy spread of the projections, but this distribution is quite peaked and for concentrations up to x=0.17 this provides an unambiguous identification of the band edges. For concentrations beyond x=0.17 the magnitude of the projections onto the AlN states decreases since the conduction-band states attain more contributions from BN. Hence, our ability to clearly distinguish between Γ_c and K_c states decreases. For this reason, we limit our calculations to a concentration of 17% boron.

Table 3.2: Standard deviation of the computed direct and indirect band gaps in units of meV.

x	0.03	0.06	0.08	0.11	0.14	0.17
direct	0	26	46	61	52	100
indirect	0	25	37	35	65	80

3.3.4 Implications for Applications of BAlN Alloys

Finally, we comment on the implications of our results for applications of BAlN alloys. Our calculations indicate it is possible to maintain a direct band gap of the BAlN alloy up to x=0.28, a much larger concentration than would be estimated based on simple linear interpolation.

However, it is doubtful that this actually offers any benefit in the case of optoelectronic devices, where an important goal would be to increase the band gap above the AlN value. We find that, due to large band-gap bowing, the band gap is actually slightly *reduced* at low B concentrations (by 0.02 eV at x=0.05). After that it increases only by a nominal amount before reaching the direct-to-indirect crossover at x=0.28, where an increase of the direct gap by 0.45 eV is calculated. Since BAlN alloys probably exhibit poor miscibility due to the large lattice mismatch between AlN and BN [61], it is doubtful that at the B concentrations for which homogeneous alloys can be achieved any increase in band gap will be observed.

Finally, we note that the band edges of BAlN alloys may be sensitive to the effects of short-range order [84]. These effects could be explored in future work if experimental evidence for short-range order emerges.

3.4 Conclusions

We have investigated the structural and electronic properties of wurtzite BAlN alloys using first-principles calculations based on hybrid density functional theory and a projection scheme to identify band edges. We find a large bowing of the direct band gap (bowing parameter b^{dir}=8.55 eV) for the direct band gap and a weaker bowing (b^{ind}=1.49 eV) for the indirect band gap. Our results indicate that BAlN alloys will have a direct band gap up to a boron concentration of x=0.28, a much larger concentration than would

be estimated based on linear interpolation. However, the increase in the direct gap over the AlN value is very modest.

Chapter 4

SRH recombination at calcium defects in InGaN

In Chapter 3, we looked at how the band gap of a material can change as a function of the alloy concentration. This change in the band gap can have dramatic effects on the Shockley-Read-Hall (SRH) and Auger recombination coefficients. In this chapter we will look at how the changing band gap of an alloy (InGaN in this case) influences the SRH coefficients calculated using the first-principle methodology outlined in Sec. 2.4.

Calcium (Ca) can be unintentionally incorporated during growth of semiconductor devices. Using hybrid functional first-principles calculations we assess the role of Ca impurities in GaN. Ca substituted on the cation site acts as a deep acceptor with a level $\sim$1 eV above the GaN valence-band maximum. We found that for Ca concentrations of 10^{17} cm^{-3}, the SRH recombination A coefficient in InGaN exceeds 10^{6} s^{-1} for band gaps less than 2.5 eV. A coefficients of this magnitude can lead to significant reductions in the efficiency of light-emitting diodes (LEDs).

4.1 Introduction

The group-III nitrides are key materials for LEDs [85]. Although the internal quantum efficiency (IQE) of III-nitride LEDs in the blue region of the spectrum exceeds 90%, the efficiency in the green and yellow region of the spectrum is much lower [86]. Within the ABC model the IQE is defined in Eq. (1.2), which we restate here:

$$IQE = \frac{Bn^2}{An + Bn^2 + Cn^3}\,, \tag{4.1}$$

where A is the SRH recombination coefficient, B is the radiative coefficient and C is the Auger coefficient. Deep-level defects are common in III-nitride materials, and they can facilitate nonradiative SRH recombination, which may affect the peak efficiency of nitride LEDs.

The defects responsible for SRH could be native point defects. For instance, gallium vacancies and their complexes have been suggested as nonradiative recombination centers in nitride LEDs [87], and recent computational studies have identified a microscopic mechanism by which such complexes act as SRH centers in InGaN [44]. However, impurities should also be taken into consideration as potential sources of nonradiative recombination. Building on the recent observation that the presence of calcium is correlated with efficiency loss [88], we have conducted a comprehensive investigation of the microscopic mechanisms by which Ca impurities cause nonradiative recombination.

Alkaline earth impurities such as calcium have previously been observed to be unintentionally incorporated during growth of GaInNAs using MBE [89, 90, 91]. In these studies, calcium was identified through secondary-ion-mass-spectroscopy (SIMS) with concentrations ranging from 10^{15} to 10^{17} cm^{-3}. Wafer polishing steps were identified as the source of the unintentional Ca incorporation. It was suggested that Ca rides on the

growth surface until growth conditions that support Ca incorporation occur. Calcium contamination can potentially also occur through the In source. Indium is commercially extracted as a byproduct of zinc [92], a process that involves leaching of zinc using sulphuric acid followed by a neutralization step using calcium carbonate. In recent work by Young *et al.* [88] calcium was found with concentrations as high as 10^{18} cm^{-3} in InGaN layers grown by ammonia MBE. The origin of Ca was determined to be contamination on the GaN templates prior to MBE growth. Systematic studies showed that Ca exhibits strong surface segregation in the MBE environment.

Here we report a first-principles study of the structural and electronic properties of Ca in GaN, and identify a mechanism by which Ca leads to SRH recombination in InGaN.

4.2 Computational Approach

Our calculations are based on density functional theory (DFT) [17, 18] with the hybrid functional of Heyd, Scuseria and Ernzerhof (HSE) [93], as implemented in the VASP code [67, 68]. The fraction of screened Fock exchange α was set to 31% which results in a GaN band gap of 3.48 eV, in agreement with the experimental value. A 400 eV energy cutoff was used and spin polarization was included. Ga d states were treated as part of the core. Impurity calculations were performed in a 96-atom GaN supercell, with a (2×2×2) Monkhorst-Pack t grid. For a generic impurity, the likelihood of incorporation is determined by its formation energy given by Eq. (2.48); for the example of substitutional Ca on the Ga site in GaN this is defined as:

$$E^{\rm form}({\rm Ca}_{\rm Ga}^{q}) = E_{\rm tot}({\rm Ca}_{\rm Ga}^{q}) - E_{\rm tot}({\rm GaN}) + \mu_{\rm Ga} - \mu_{\rm Ca} + q(E_{\rm F}) + \Delta^{q}\,. \qquad (4.2)$$

$E^{\text{form}}(\text{Ca}_{\text{Ga}}^{q})$ is the total energy of the 96-atom GaN supercell with Ca on the Ga site in charge state q, $E_{\text{tot}}(\text{GaN})$ is the total energy of the pristine GaN supercell, μ_{Ga} and μ_{Ca} are the chemical potentials of Ga and Ca, which depend on the growth conditions. We will show results for Ga-rich conditions (which are basically a "worst case" for incorporation of Ca on the Ga site) and with μ_{Ca} chosen to correspond to equilibrium with Ca_3N_2 (ΔH_f=−3.95 eV in HSE). The Fermi level (E_{F}) is referenced to the valence-band maximum (VBM) of GaN, and Δ^q is a finite-size correction obtained using the Freysoldt scheme [94].

Calculations of nonradiative capture coefficients utilize our previously developed methodology to describe nonradiative capture of carriers by multiphonon emission [12] (see Sec. 2.4.2). The electron-phonon matrix elements for electron and hole capture are obtained using the projector-augmented wave all-electron wave functions [31].

4.3 Results and Discussions

4.3.1 Thermodynamics of Ca Defects in GaN

Three configurations of the Ca impurity in GaN were considered: Ca on the Ga site (Ca_{Ga}), Ca on the N site (Ca_{N}), and Ca in an interstitial configuration (Ca_{i}). The formation energies for Ca_{Ga} and Ca_{i} are shown in Fig. 4.1. Substitutional Ca on the N site has a high formation energy, due to the large mismatch in the ionic radii between Ca and N; Ca_{N} is thus unlikely to incorporate in appreciable concentrations. When calcium is incorporated as an interstitial it is stable in the q=2+ charge state across the entire GaN band gap and thus acts as a shallow double donor. Since having a thermodynamic transition level within the gap is a prerequisite for nonradiative capture, Ca_{i} does not contribute to SRH recombination.

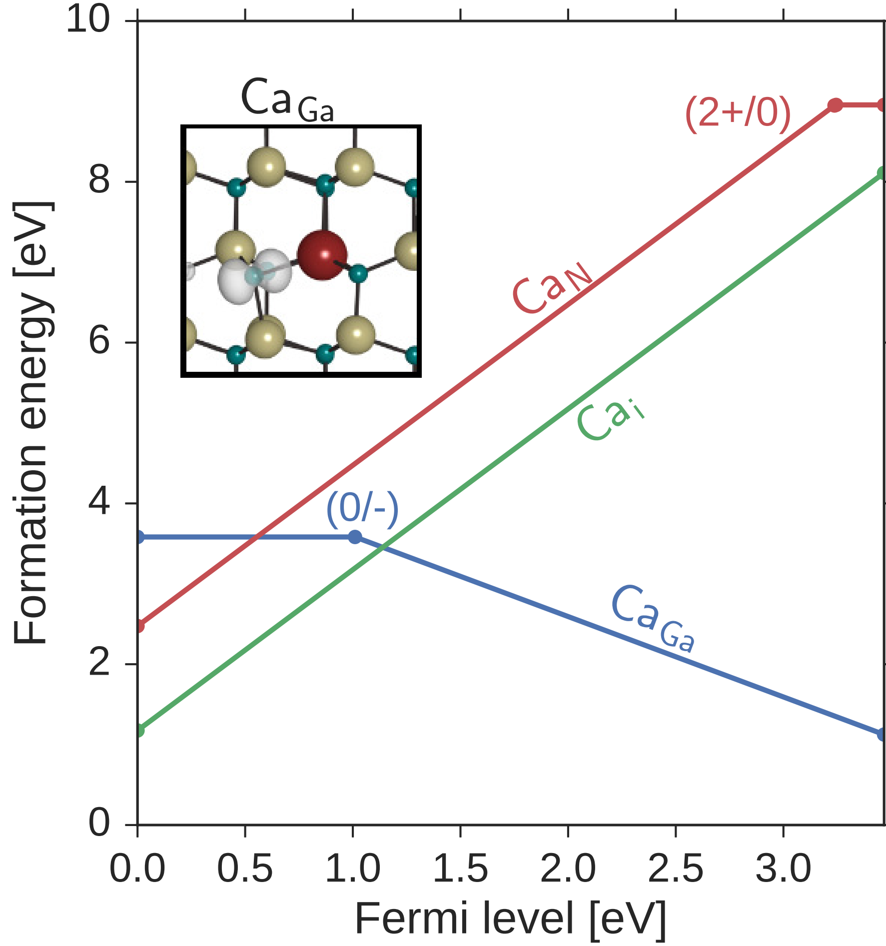


Figure 4.1: Formation energies vs. Fermi level for Ca_{Ga}, Ca_N and Ca_i in various charge states under Ga-rich conditions. The atomic geometry and impurity wave function of neutral Ca_{Ga} are illustrated in the inset.

Consistent with prior calculations [95] we find that substitutional Ca on the Ga site acts as a deep acceptor, with a $(0/-)$ acceptor level at 1.01 eV above the VBM. The larger ionic radius of Ca compared to Ga leads to an outward relaxation of the nearest-neighbor nitrogen atoms when Ca is substituted on the Ga-site. In the negative charge state the four nearest neighbor nitrogen atoms are uniformly displaced outwards by 13% of the Ga-N bond length. In the neutral charge states we find an asymmetric distortion of the Ca-N bonds: one of the in-plane Ca-N bonds increases by 15% of the Ga-N bond length, while the other three are increased by 12–13% of the Ga-N bond length.

4.3.2 SRH Recombination at Ca Defects in GaN

We now focus on the role of Ca_{Ga} as a recombination center. A configuration coordinate diagram (CCD) illustrating electron and hole capture into the $(0/-)$ level of Ca_{Ga} is shown in Figure 4.2. The transition energy ΔE (which is also called the "zero-phonon line" in the context of optical transitions) is given by the position of the $(0/-)$ transition level (from Fig. 4.1) relative to the relevant band edge: the conduction-band minimum (CBM) in the case of electron capture, leading to a transition energy ΔE^n, and the valence-band maximum (VBM) in the case of hole capture, leading to a transition energy ΔE^p.

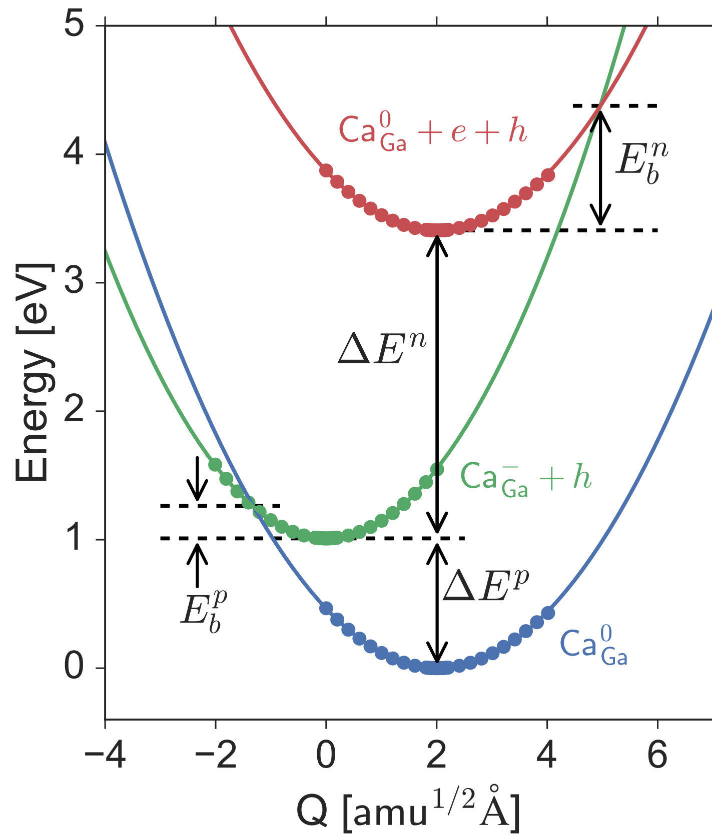


Figure 4.2: Configuration coordinate diagram illustrating electron and hole capture processes. ΔE is the transition energy and E_b is the classical barrier for the nonradiative process; the superscripts indicate whether the quantity applies to electron (n) or hole (p) capture. Solid circles denote calculated values, solid lines parabolic fits.

A complete recombination cycle requires capture of an electron by the neutral impurity, followed by capture of a hole by the negatively charged impurity. The overall recombination rate is thus governed by the *slower* of the two processes. Nonradiative capture rates decrease roughly exponentially with the energy of the transition [36]. In GaN, the (0/−) level of Ca_{Ga} is much closer to the VBM, and therefore the nonradiative capture coefficient for holes, C_p, is expected to be orders of magnitude larger than that for electrons, C_n. Slow electron capture will thus limit the SRH recombination rate for Ca_{Ga} in GaN.

4.3.3 SRH Recombination in InGaN

The situation dramatically changes in InGaN, which constitutes the active layer in visible LEDs. As the In content in InGaN increases the band gap decreases; we now examine how the position of the Ca_{Ga} (0/−) level shifts with respect to the band edges as a function of In content. We continue to use the notation "Ca_{Ga}" as a generic designation for incorporation of Ca on a group-III site in the InGaN alloy. We follow the procedure outlined in Ref. [96], where it was shown that the primary effect on the defect level is due to the lattice expansion that occurs upon increasing the In content of the alloy. We thus calculate the thermodynamic transition level for Ca_{Ga} using lattice parameters corresponding to InGaN alloys with In contents of 5%, 10% and 25%. To apply this information to actual InGaN alloys, we align the VBM in the expanded cell with that of unstrained GaN using the absolute valence-band deformation potential of GaN [97]. The band edges in the InGaN alloy, again aligned with respect to unstrained GaN, are taken from Ref. [98].

The results are shown in Fig. 4.3. We observe that the energy difference between the (0/−) level and the CBM is indeed significantly reduced in InGaN alloys.

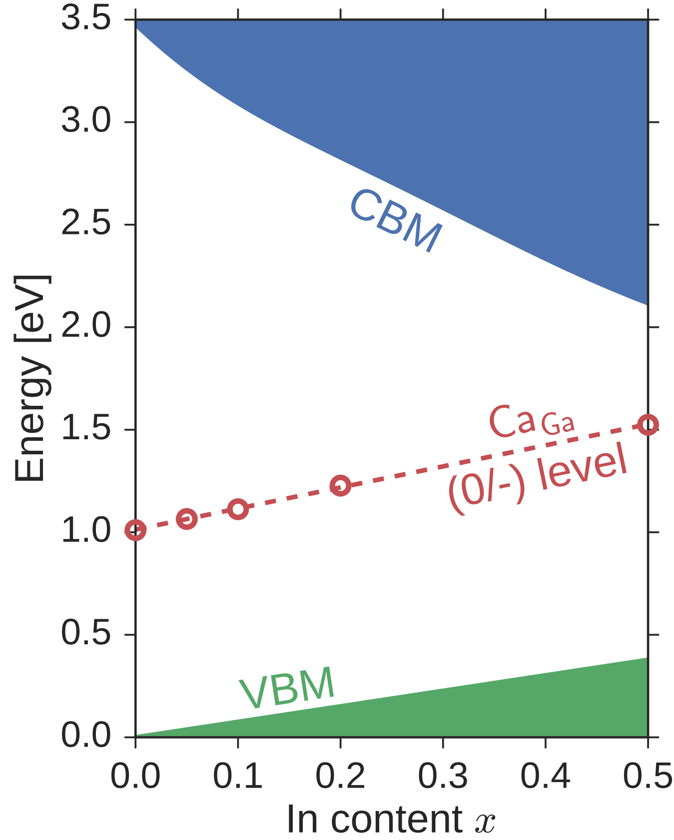


Figure 4.3: Calculated position of the $(0/-)$ transition level of Ca_{Ga} within the band gap of InGaN as a function of In content.

Using these results, we now determine the nonradiative electron and hole capture coefficients for Ca_{Ga} as a function of In content. The capture coefficients, C_n and C_p are shown as a function of the band gap of the InGaN alloy in Fig. 4.4, assuming a typical internal operating temperature of 390 K. Because the $(0/-)$ transition level increases only slightly with respect to the VBM of InGaN [Fig. 4.3] the hole capture coefficient, C_p, does not change dramatically over the plotted range. For electron capture, the transition energy decreases rapidly with indium content, resulting in a dramatic rise in C_n.

At a given impurity density N, the SRH recombination rate is characterized by the coefficient A where

$$A = N \frac{C_n C_p}{C_n + C_p}. \tag{4.3}$$

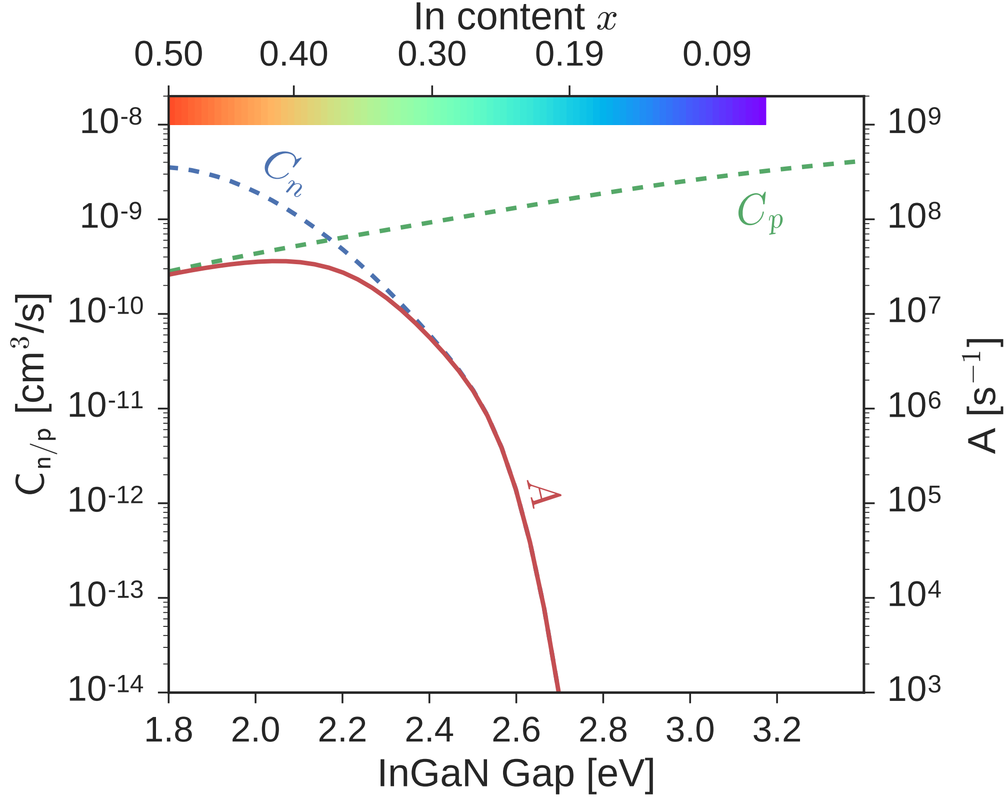


Figure 4.4: Electron (C_n) and hole (C_p) capture coefficients for the (0/−) level of Ca_{Ga} in InGaN calculated at 390 K. The SRH coefficient A is calculated for a Ca density of 10^{17} cm^{-3}.

The A coefficient, calculated for a Ca impurity concentration N=10^{17} cm^3, is also shown in Fig. 4.4. The figure shows that at low In content (large band gap) the rate-limiting step is electron capture, while at higher In content (low band gap), hole capture limits the rate. This crossover is directly related to the position of the Ca_{Ga} (0/−) level within the band gap. As the In content increases, the energy difference between the CBM and the (0/−) level decreases faster than the slight increase in the energy difference between the VBM and the (0/-) level [Fig. 4.3].

We can now assess the impact of unintentional Ca incorporation on the IQE of an LED, again assuming N=10^{17} cm^3. To do so, we compare the rate of nonradiative SRH recombination, An, with the rate of radiative band-to-band transitions, Bn^2; or,

alternatively, the magnitude of A with the magnitude of Bn. We assume an operating carrier density of $n = 10^{18}$ cm^{-3}. The recombination coefficient in an InGaN quantum well is given by $B = 4{\times}10^{-11}$ cm^{3}s^{-1} [99], and hence $Bn = 4{\times}10^7$ cm^{-3}s^{-1}. Figure 4.4 shows that in the blue region of the spectrum the A coefficient is sufficiently low to not cause loss of carriers. However, as the In content increases and the band gap decreases towards the green region of the spectrum, the A coefficient increases significantly: for a band gap of 2.3 eV the SRH coefficient is approximately 10^7 s^{-1}; comparing this value with Bn, we conclude that 20% of carriers would be lost due to SRH recombination. Control of Ca during nitride growth is thus very important. The study by Young *et al.* [88] showed that alternating high-temperature/low-temperature MBE layers can effectively incorporate Ca in the low-temperature layers and thus reduce the Ca concentration in the active layer to as low as 10^{14} cm^{-3}.

4.4 Conclusions

In conclusion, we have used first-principles calculations to determine the thermodynamic transition levels and nonradiative recombination properties due to Ca in GaN. We find that $\mathrm{Ca_{Ga}}$ behaves as a deep acceptor with a level $\sim$1 eV above the GaN VBM. For a Ca impurity concentration of 10^{17} cm^{-3}, we find SRH coefficients exceeding 10^6 s^{-1} for InGaN band gaps below 2.5 eV. SRH coefficients of this magnitude can severely limit the IQE of nitride LEDs [100].

Chapter 5

Auger recombination in InAs

As we see in Fig. 4.4, the Shockley-Read-Hall (SRH) process shows a strong dependence on the band gap of the material. The Auger recombination coefficient is also known to decrease rapidly as the band gap increases, at least for materials with band gaps below about 1 eV [47]. InAs, with a band gap of 0.35 eV, is observed to have one of the highest Auger coefficients. Still, a detailed analysis of the exact source of the high Auger recombination has been lacking, for example on the role (if any) of phonon-assisted Auger recombination in this material. In addition to improving the understanding of the Auger process in InAs, computing the Auger recombination coefficient in this well-studied material will serve as a benchmark for the extension of the first-principles Auger methodology to account for spin-orbit coupling (SOC), outlined in Sec. 2.5. This type of validation will lend confidence to our investigations of less well studied materials in Chapter 6.

We compute the direct and phonon-assisted Auger coefficients in indium arsenide InAs and related alloys from first principles. The direct process involves only Coulomb interaction, while the indirect process is mediated by phonons. We have adapted the methodology to explicitly include spin-orbit coupling. We show how the spin-orbit split-off states allow for efficient direct Auger recombination, greatly enhancing the excitation

rate of Auger holes due to a near-resonance between the spin-orbit splitting and the band gap of the material. We find that the direct *hhe* Auger recombination coefficient decays exponentially with increasing band gap, while the eeh decays faster than exponentially. Additionally, we modeled the phonon-assisted process in InAs and show that the direct Auger process is dominant.

5.1 Introduction

Auger recombination is an important nonradiative carrier recombination mechanism in semiconductors. It reduces the efficiency of light-emitting devices at high power and has been shown to play a significant role in the efficiency-droop and green-gap problems of nitride light-emitting diodes [101, 48, 99]. Auger processes also contribute to efficiency loss in photovoltaics [102] and limit the direct current performance of bipolar transistors [103]. Auger recombination is particularly critical in narrow-gap semiconductors such as InAs-based heterostructures. These materials find applications in mid-wavelength infrared (3–5 μm) diode lasers [104] as well as in interband cascade lasers (ICLs) with InAs/GaInSb quantum wells (QWs) [105]. The reduction in quantum efficiency due to Auger recombination tends to govern the lasing threshold, especially at high carrier concentrations and high temperatures [106]. Understanding the microscopic details of the Auger recombination mechanism in these materials will aid in identifying engineering solutions to mitigate its impact.

In the Auger process (Fig. 5.1), the energy resulting from an electron-hole pair recombination is not emitted as a photon but transferred to a third carrier, either an Auger electron (in the e–e–h process) or an Auger hole (h–h–e). In the direct e–e–h process illustrated in Fig. 5.1(b), an initial electron-hole pair (states 1 and 3) recombines across the band gap and transfers its energy to a third electron (state 2) that gets excited to

a high-energy state in the conduction band (state 4). Such direct Auger recombination is mediated purely by Coulomb interactions. However, additional second-order interactions involving the emission or absorption of a phonon [Fig. 5.1(c)] can also contribute significantly to the total Auger rate [48, 107].

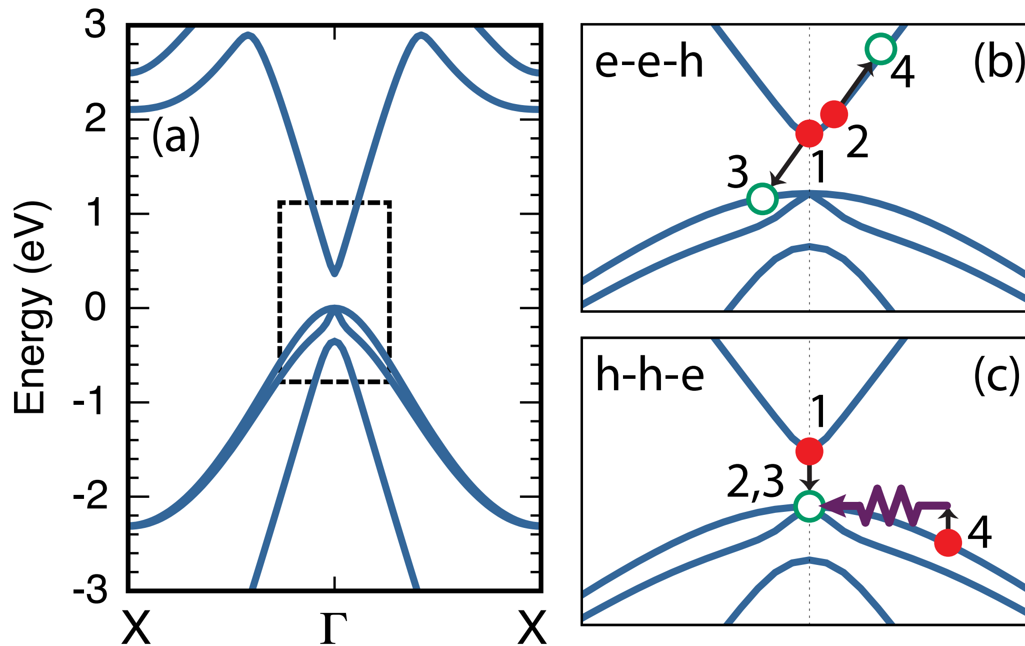


Figure 5.1: (a) Calculated band structure of InAs. The indicated region near the center of the Brillouin zone is enlarged in panels (b) and (c) to illustrate Auger processes: (b) shows the direct e–e–h process, (c) the phonon-assisted h–h–e process.

5.2 Computational Approach

First-principles calculations based on density functional theory (DFT) are well suited for the predictive modeling of optoelectronic materials [108]. Auger carriers are excited to states with energy approximately equal to the gap away from the band edges. DFT calculations can reliably predict the energies and wave functions throughout the Brillouin zone (BZ) and at higher energies. Based on such calculations, it was revealed that in

wide-band-gap semiconductors such as GaN Auger recombination is dominated by the indirect, phonon-assisted process [48, 11]. It is still an open question to what extent phonon-assisted processes contribute to Auger recombination in smaller-gap materials.

Calculations of the Auger rate in narrow-gap materials require an accurate treatment of spin-orbit (SO) interactions, which modify the band structure near the band extrema. In InAs, the magnitude of the SO splitting (0.39 eV [109]) is comparable to the band gap (0.35 at 300 K [110] and 0.43 eV at 0 K [109]) and enables additional direct Auger recombination processes that would otherwise be forbidden by energy and momentum conservation. In addition, the lack of inversion symmetry in InAs breaks the spin degeneracy of the bands. Hence, careful consideration of the SO interaction is needed.

We have performed a fully first-principles prediction of SO-coupling effects in direct and phonon-assisted Auger recombination, which required significant modifications to previously developed methodology [11]. When applied to InAs, the results show that SO interactions significantly enhance direct Auger recombination. The resulting rate for the direct Auger process in InAs is a factor of 50 larger than for the phonon-assisted process. We also find that, as the band gap of the material increases, the direct Auger process becomes comparatively less relevant. The Auger recombination coefficients decay approximately exponentially as a function of band gap, and this strong dependence means that the direct Auger process can be significantly reduced by alloying.

Our computational approach is based on time-dependent perturbation theory using Fermi's golden rule. The direct Auger recombination rate, $R^{\mathrm{direct}}_{\mathrm{Auger}}$, is given by

$$\begin{aligned} R^{\mathrm{direct}}_{\mathrm{Auger}} =& \frac{2\pi}{\hbar} \sum_{\mathbf{1234}} f_{\mathbf{1}} f_{\mathbf{2}} (1-f_{\mathbf{3}})(1-f_{\mathbf{4}}) \left| M_{\mathbf{1234}} \right|^2 \\ & \times \delta(\varepsilon_{\mathbf{1}} + \varepsilon_{\mathbf{2}} - \varepsilon_{\mathbf{3}} - \varepsilon_{\mathbf{4}}) \ , \end{aligned} \tag{5.1}$$

where the bold indices [$\mathbf{1} \equiv (n_1 \mathbf{k}_1)$, etc.] are composite band (n_1) and wave vector ($\mathbf{k}_1$)

indices, $\varepsilon_{\mathbf{i}}$ are the band energies, and $f_{\mathbf{i}}$ are the Fermi occupation numbers of the states. In the presence of SO coupling, a generalization of the formalism in Ref. [11] is necessary. The anti-symmetrized screened Coulomb (W) matrix elements are written in terms of the single-particle spinor states (such as $|\mathbf{1}\rangle = |\psi_{\mathbf{1}}^{\uparrow}\rangle \otimes |\uparrow\rangle + |\psi_{\mathbf{1}}^{\downarrow}\rangle \otimes |\downarrow\rangle$):

$$M_{\mathbf{1234}} \equiv \langle \mathbf{12}|W|\mathbf{34}\rangle - \langle \mathbf{12}|W|\mathbf{43}\rangle \,, \tag{5.2}$$

where the first term is a *direct* term and the second an *exchange* term, accounting for the anti-symmetry under fermion exchange. Details of the formalism for including SO interactions are presented in Sec. 2.3.3.

A similar generalization can be made for the phonon-assisted Auger rate:

$$\begin{aligned} R_{\text{Auger}}^{\text{phonon}} = \frac{2\pi}{\hbar} \sum_{\mathbf{1234}} f_{\mathbf{1}} f_{\mathbf{2}} (1 - f_{\mathbf{3}})(1 - f_{\mathbf{4}}) \left| \tilde{M}_{\mathbf{1234},\nu\mathbf{q}} \right|^2 \\ \times (n_{\nu\mathbf{q}} + \frac{1}{2} \pm \frac{1}{2}) \times \delta(\varepsilon_{\mathbf{1}} + \varepsilon_{\mathbf{2}} - \varepsilon_{\mathbf{3}} - \varepsilon_{\mathbf{4}} \pm \hbar\omega_{\nu\mathbf{q}}) \;, \end{aligned} \tag{5.3}$$

where $n_{\nu\mathbf{q}}$ are the Bose-Einstein phonon occupation numbers and $\omega_{\nu\mathbf{q}}$ are the vibrational frequencies of the phonon mode ν and phonon wave vector $\mathbf{q}$. The phonon-assisted matrix elements ($\tilde{M}_{\mathbf{1234},\nu \boldsymbol{q}}$) contain eight terms [11], which must again be generalized to account for SO effects (see Sec. 2.3.3).

Our first-principles calculations are based on DFT [17, 18]. The electronic structure is computed with projector augmented wave (PAW) potentials as implemented in the VASP [67, 68]. The Auger rate is highly sensitive to the eigenvalues near the band extrema, requiring an accurate description of the band structure. We use the effective-mass-fitted modified Becke-Johnson functional within the local density approximation (mBJLDA), as described in Ref. [111]. The phonon spectra and electron-phonon coupling matrix elements are obtained with density functional perturbation theory [49]. Since

electron-phonon interactions are not implemented in VASP at this time, we have used the Quantum ESPRESSO code [112]. Perturbations to the wave functions due to SO interactions are implemented using the vector part of the *ab initio* pseudopotentials [51].

The reciprocal space is sampled with a $60 \times 60 \times 60$ Γ-centered **k**-point mesh for the electron and phonon states. The screened Coulomb interaction is evaluated using a static model dielectric function [11] and the experimental dielectric constant of 12.3 for InAs [113]. We account for the screening by free carriers with the Debye model. For the evaluation of the phonon-assisted Auger rate we assume that the initial states of the three carriers are all at Γ—the same approximation as employed in Refs. [107, 99] to render the problem computationally tractable.

5.3 Results and Discussions

5.3.1 Auger Coefficients in pure InAs

We assume equal densities of free electrons and holes ($n = p$), which is typical in optoelectronic devices with high carrier injection. The Auger coefficients can be defined by $C_{\text{Auger}} = R_{\text{Auger}}/n^3$. The resulting direct and phonon-assisted coefficients are tabulated in Table 5.1. At the experimental band gap of 0.35 eV we computed a direct Auger coefficient (e–e–h + h–h–e) of 8.60×10^{-28} cm^{6}s^{-1}. The computed phonon-assisted coefficient, 1.7×10^{-29} cm^{6}s^{-1}, is more than an order of magnitude smaller. A survey of experimental literature [114, 115, 116, 117, 118] reveals significant variations among the reported Auger coefficients, ranging from $1 - 3 \times 10^{-28}$ cm^{6}s^{-1} in InAs based quantum wells [119] to bulk values of 6.0×10^{-27} in Ref. [116] and 6.5×10^{-26} cm^{6}s^{-1} in Ref. [114]. The model theory of Takeshima [120] predicted an Auger coefficient of 1.8×10^{-27} cm^{6}s^{-1} (at $T = 300$ K and $n = 10^{18}$ cm^{-3}).

Table 5.1: Computed Auger coefficients C (in units of $10^{-30}\text{cm}^6\text{s}^{-1}$) for InAs at 300 K with a carrier concentration of $10^{18}\,\text{cm}^{-3}$. The coefficients calculated without SO interactions are shown in parentheses.

	Direct	Direct (no SO)	Phonon	Phonon (no SO)
e–e–h	496	387	9	10
h–h–e	364	84	6	9
Total	860	471	15	19

While the experimental measurements of the Auger recombination rates are conducted at room temperature, the effective temperature of the carriers can be much higher than the lattice temperature. To evaluate the Auger coefficients at higher carrier temperatures, we performed a series of calculations for the direct Auger process between 150 K and 450 K, and fit the results to an Arrhenius law [121] [$C(T) = C_\infty \exp(-E_A/k_B T)$], as shown in Fig. 5.2. The resulting activation energies are $E_A^{\text{e-e-h}} = 0.021\,\text{eV}$ and $E_A^{\text{h-h-e}} = 0.011\,\text{eV}$.

5.3.2 Auger Coefficients in InAs alloys

Our results can be extrapolated to InAs-based alloys, for instance with GaAs (band gap 1.43 eV) or InSb (band gap 0.17 eV), both of which have band structures similar to InAs [111]. Explicit alloy calculations require large supercells and are computationally prohibitive with the **k**-point densities needed for the Auger calculations. We therefore approximate the alloy electronic structure by modifying the band gap of InAs while keeping all of the Coulomb and electron-phonon matrix elements fixed—commonly referred to as a *scissors shift*. In doing so we are neglecting additional scattering mechanisms due to alloy disorder and focusing purely on the band-gap dependence. Figure 5.3 shows the

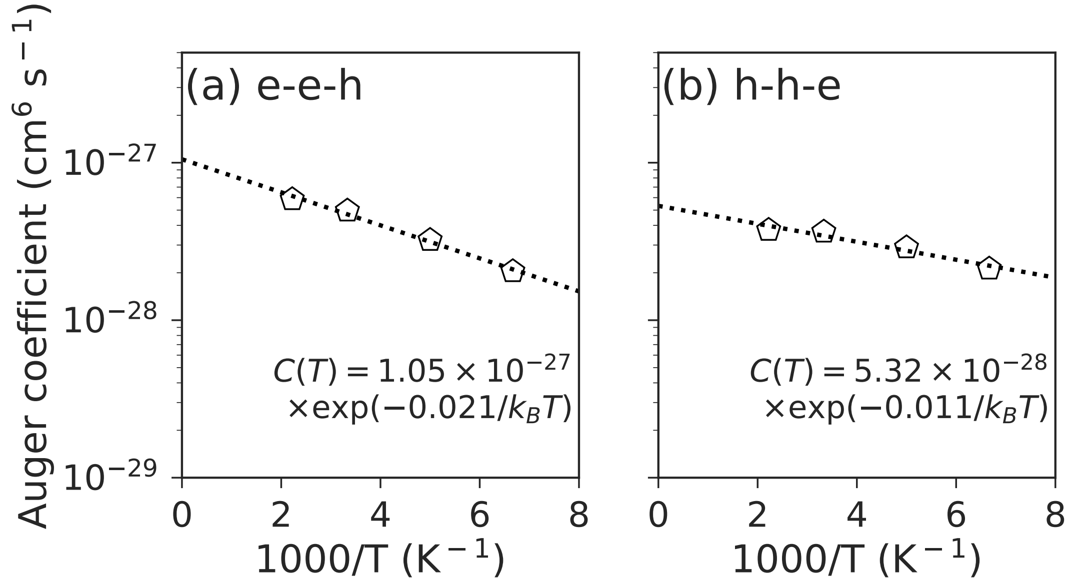


Figure 5.2: Computed direct Auger coefficients at the experimental band gap (E_g =0.35 eV) with a carrier concentration of 10^{18} cm^{-3}, for temperatures between 150 K and 450 K, fitted to an Arrhenius relationship.

calculated coefficients as a function of alloy band gap.

Figure 5.3 shows that direct Auger is stronger than indirect Auger across the entire range of band-gap values for both e–e–h and h–h–e. The decrease of the direct e–e–h Auger coefficient as a function of increasing band gap can be attributed to the lack of higher-energy states (roughly one band gap higher in energy than the conduction-band minimum) near the center of the BZ. This renders energy and momentum conservation difficult to satisfy when three of the states must come from the same highly dispersive conduction band [states 1, 2 and 4 in Fig. 5.1(b)]. The decrease in the direct h–h–e coefficient is less dramatic, as energy and momentum conservation is easier to satisfy by scattering among the multiple valence bands. The overall decay then follows an exponential trend, reflecting the tail of the Fermi-Dirac distributions. The same trends can be observed in previous studies of GaN [11] and GaAs [107].

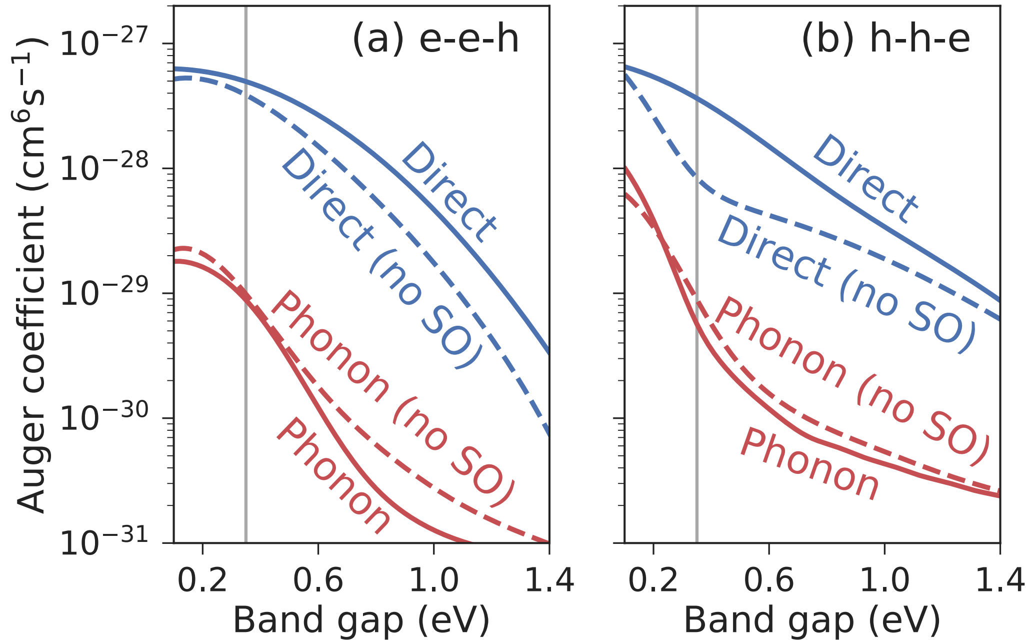


Figure 5.3: Auger coefficient vs. scissors-shift-adjusted band gap for direct (solid blue) and phonon-assisted (solid red) Auger recombination due to (a) e–e–h and (b) h–h–e processes. The calculated coefficients without SO interactions for the direct (dashed blue) and phonon-assisted (dashed red) process are included for comparison. The gray vertical line indicates the experimental band gap of InAs. The calculations are carried out at 300 K with a carrier concentration of 10^{18} cm^{-3}.

5.3.3 Effects of Spin-Orbit Coupling

Inclusion of SO effects modifies the band structure near the valence-band maximum and significantly enhances the h–h–e direct Auger coefficient. This is most evident for band-gap values close to the SO-splitting energy of InAs (see Fig. 5.3), where a near-resonance occurs. At the band gap of InAs, the h–h–e Auger coefficient is 3.64×10^{-28} cm^{6}s^{-1} when SO interaction is included (see Table 5.1), more than a factor of four greater than the coefficient computed without SO interactions (8.4×10^{-29} cm^{6}s^{-1}). The inclusion of SO interaction has a less dramatic effect on the e–e–h process (4.96×10^{-28} cm^{6}s^{-1} with SO interaction and 3.87×10^{-28} cm^{6}s^{-1} without). The near-coincidence of the band gap with the SO splitting makes InAs close to a *worst case scenario* in terms

of impact of SO coupling on Auger recombination.

5.3.4 Effects of Phonons

In contrast to the direct process, in the indirect process a phonon supplies the additional momentum (the phonons do not contribute much to energy conservation), in principle allowing the Auger electron/hole to be scattered anywhere in the BZ. At the carrier concentrations and temperatures considered here, the initial electron and hole states represent a very small portion of the BZ—0.03% for electrons and 2% for holes. Any small amount of momentum difference in the initial states can easily be compensated by the additional phonon, which justifies our approximation that the initial states are all at Γ. The small momentum differences will modify the matrix elements; however, any corrections to the phonon-assisted rate would not change our conclusion that the direct process dominates across the entire range of band gaps considered in Fig. 5.3.

Figure 5.4 shows contributions to the total Auger rate from phonons of different absolute wave numbers. At the 0.35 eV band gap of InAs [Fig. 5.4(a)], the vast majority of phonon-assisted Auger contributions comes from the center of the BZ, near Γ. In this regime, the energy and momentum conservation for the phonon-assisted Auger processes is very similar to the direct process, and the phonon-assisted Auger process behaves as a higher-order correction to the direct Auger recombination, with the phonon-assisted rate more than an order of magnitude smaller than the direct rate (Table 5.1).

In contrast, at larger band gaps [Fig. 5.4(b)], states from other regions of the BZ can be involved via a phonon. These states account for a much greater portion of the total phonon-assisted Auger rate, with barely any contributions coming from the zone center (since there are no states in the vicinity of Γ that would allow energy conservation to be satisfied with a low-wave-number phonon). For even higher band gaps, the majority of

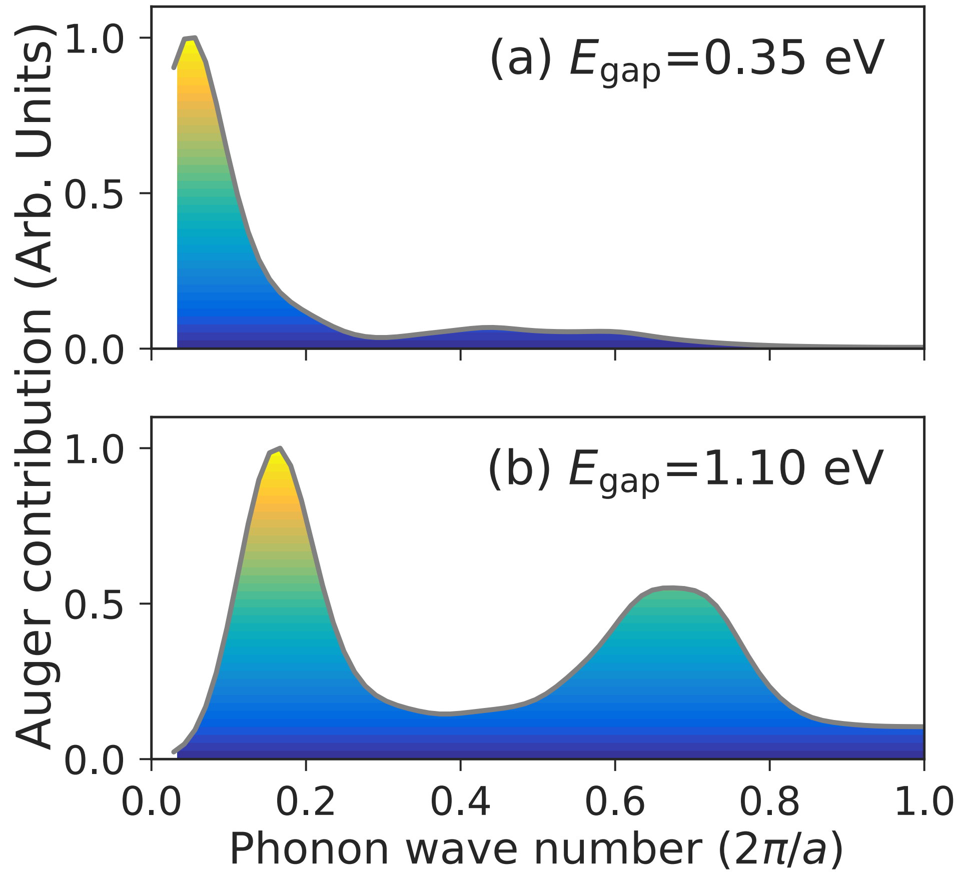


Figure 5.4: Normalized contributions to the phonon-assisted Auger rate plotted as a function of phonon wave number q for (a) a band gap of 0.35 eV, and (b) a band gap of 1.1 eV.

states involved in the recombination process are no longer near the zone center (where the direct process occurs), and indirect Auger becomes stronger relative to the direct process.

In wide-band-gap materials (e.g., GaN [11] with a band gap of 3.5 eV) the phonon-assisted process becomes the dominant process. This is because the direct Auger coefficient falls off exponentially at large band gaps, while the phonon-assisted coefficient decays much slower, roughly as an inverse power law above 0.5 eV. An extrapolation of our calculations shows the cross-over between the direct and phonon-assisted rates occurs at 1.8 eV. We acknowledge that this is well beyond the limit where our approximate

treatment of alloys is expected to be accurate. However, we believe the trends to be correct, and the observation that direct Auger recombination dominates in direct-gap materials at lower band gaps while phonon-assisted Auger becomes dominant at higher band gaps should be general [47].

Since the Auger coefficients generally decrease with increasing band gap, suppression of Auger recombination can in principle be achieved with band-gap engineering. This can be accomplished by alloying or straining the material, but is not always feasible if specific band-gap values are required. However, the strong dependence of the Auger coefficient on SO coupling opens an additional dimension in the design space, allowing for suppression of Auger recombination by eliminating the resonance between the SO splitting and the band gap [122].

5.4 Conclusions

In conclusion, we have developed methodology to include spin-orbit coupling in first-principles simulations [11] of direct and indirect Auger recombination. Our calculations for InAs show that inclusion of SO interactions enhances the h–h–e recombination coefficient by more than a factor of four. The direct Auger process is dominant in InAs, accounting for more than 98% of the total Auger rate. The direct process decreases exponentially with increasing band gap, making the phonon-assisted process relatively more important in larger-gap materials.

Chapter 6

Auger Recombination in Halide Perovskites

While the method of scissors-shifting the band gap to approximate alloying is relatively simple, it was able to capture the strong dependence of the Auger coefficient on the band gap as we have shown in Chapter 5. We can put this into a broader context by considering the Auger coefficients of different materials. The Auger recombination coefficient of various semiconductors have been tabulated in Ref. [47] and plotted vs. the band gap in Fig. 6.1. For the traditional semiconductors tabulated, the trend indicates that indirect-band-gap materials have a low Auger coefficient, around 10^{-30} cm^{6}s^{-1}, and direct-band-gap materials exhibits a exponential dependence of the Auger coefficient on the band gap for band gaps below 1 eV. In this chapter, we will investigate Auger recombination in halide perovskites, a class of materials that has recently received a great deal of attention. For the prototypical halide perovskite (methyl-ammonium lead iodide ($MAPbI_3$)), with the chemical formula $CH_3NH_3PbI_3$, we find an Auger coefficient of 7.3×10^{-29} cm^{6}s^{-1} which is much larger than expected for a material with a 1.6 eV gap.

The emergence of halide perovskites for photovoltaic applications has triggered great

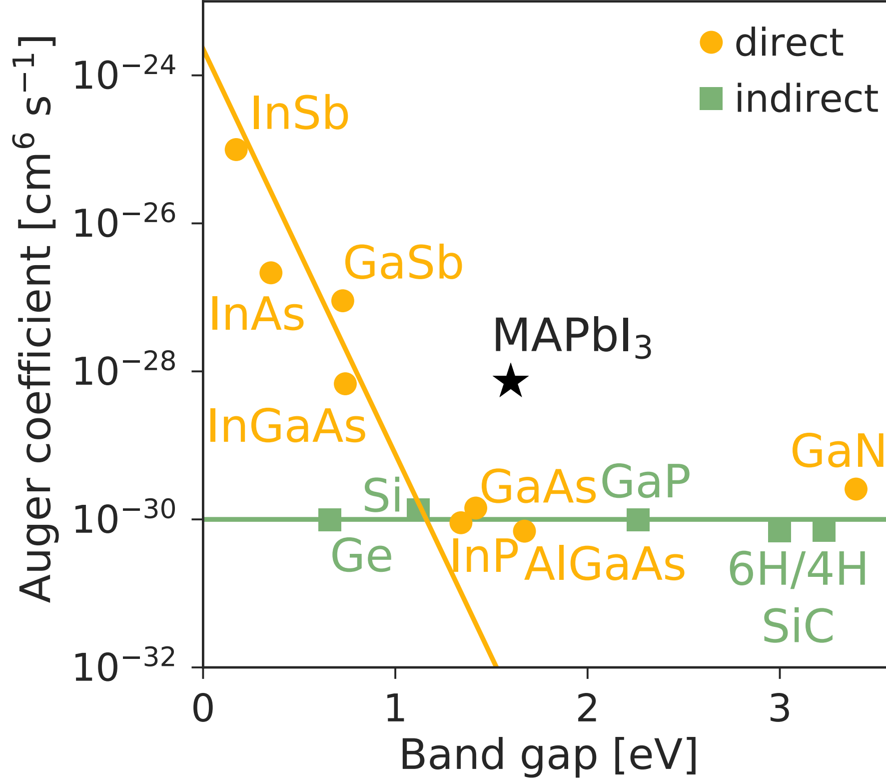


Figure 6.1: Auger coefficient vs. band gap for various direct (yellow circle) and indirect (green square) band gap materials. The lines are drawn to guide the eye. The computed Auger coefficient of $MAPbI_3$ is shown with a star.

interest in these materials for solid-state light-emission. Higher-order electron-hole recombination processes can critically affect the efficiency of such devices. In this chapter, we compute the Auger recombination coefficient of $MAPbI_3$ using first-principles calculations. We demonstrate that Auger recombination is responsible for the exceptionally high third-order recombination coefficient observed in experiment. We attribute the large Auger coefficient to a coincidental resonance between the band gap and interband transitions to a complex of higher-lying conduction bands. Additionally, we find that the distortions of PbI_6 octahedra contribute significantly to the high Auger coefficient, offering potential avenues for materials design.

6.1 Introduction

Halide perovskites comprise a group of materials with the chemical formula ABX_3, in which the A-site cation is either a metallic element or an organic molecule, the B-site is occupied by a metallic element and X is a halogen anion such as I^-, Br^-, and Cl^-. Methylammonium lead iodine ($MAPbI_3$, $MA=CH_3NH_3$) [123] is the archetypal member of this family, where the A-site molecule (MA^+) sits in a lattice of PbI_6 octahedra (see Fig. 6.2). These materials have attracted widespread attention as candidates for affordable and efficient photovoltaics; within a few years reported power conversion efficiencies have risen from 3% to over 20% [124, 125, 126], and a worldwide effort to improve the conversion efficiency of these materials is still underway.

Given the high solar conversion efficiency, detailed balance in the Shockley-Queisser formulation suggests that radiative recombination will be efficient [127]. Halide perovskites should therefore also make good light emitters, and light-emitting diodes (LEDs) have indeed been fabricated [128] and have been shown to cover the entire visible spectrum [129]. However, the operation of LEDs requires much higher carrier densities (typically around 10^{18} cm^{-3}) [130] compared to photovoltaics, where the carrier density is fundamentally limited by the availability of solar radiation. At these higher carrier densities, nonradiative recombination mechanisms might play a much more important role than in solar cells. The prospect of making efficient LEDs based on halide perovskites thus hinges on the behavior of carrier recombination at higher carrier densities.

Higher-order recombination processes play an important role in determining the quantum efficiency of an LED. Under typical LED operating conditions, the non-equilibrium densities of electrons (n) and holes (p) can be assumed to be equal. The internal quantum

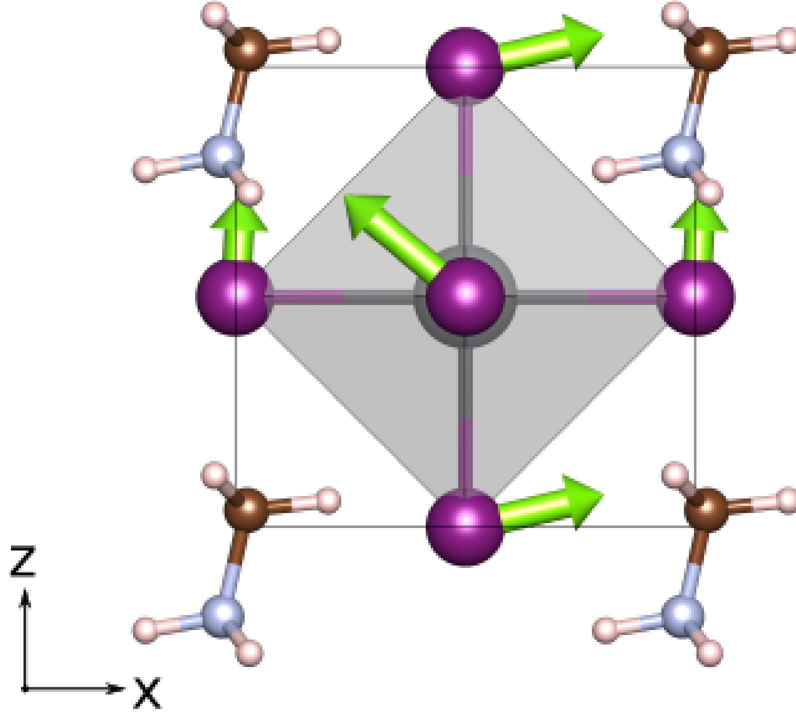


Figure 6.2: 12-atom unit cell of $MAPbI_3$ where the I atoms are at the ideal face-center positions. The arrows indicate displacement of the I atoms upon relaxation within a cubic cell shape. For clarity the length of the displacement vector is enhanced by a factor of five.

efficiency (IQE) is then described by:

$$\mathrm{IQE} = \frac{k_2 n^2}{k_1 n + k_2 n^2 + k_3 n^3}\,, \tag{6.1}$$

where k_1, k_2, and k_3 represent the mono-, bi- and trimolecular recombination coefficients respectively. For $MAPbI_3$, peak IQEs of greater than 70% have been achieved [128, 131]. The IQE peaks around $n = 10^{18}$ cm^{-3} and falls off rapidly as carrier density increases; the fall-off is due to the trimolecular recombination dominating at higher carrier densities. Transient spectroscopy [132, 133, 134] and photoluminescence [135, 131] measure-

ments have obtained trimolecular recombination coefficients ranging from 5.4×10^{-29} to 1.6×10^{-28} cm^{6}s^{-1}.

In this chapter we set out to determine from first principles whether Auger recombination can explain the observed efficiency loss at high carrier densities. Auger is a trimolecular process in which the energy released by electron-hole recombination promotes a third charge carrier to an excited state. This process is nonradiative, as the excess energy of the excited carrier is lost to phonons. If that third charge carrier is an electron, the Auger process is called an *electron-electron-hole* (eeh) process; in case it is a hole, it is called a *hole-hole-electron* (hhe) process.

In common semiconductor materials, the Auger coefficient C spans seven orders of magnitude (from 10^{-30} to 10^{-23} cm^{6}s^{-1}), decreasing exponentially with the band gap [47]. The measured k_3 coefficients in $MAPbI_3$ are two orders of magnitude higher than the Auger coefficients C in other materials with similar band gaps [47], prompting the question whether Auger is actually the responsible mechanism. Our first-principles calculations allow us to examine whether the Auger process is really so much stronger than in traditional semiconductors, and what the underlying mechanisms are.

We use first-principles methods (Sec. 6.2) to compute the Auger recombination coefficient in $MAPbI_3$, and investigate its relationship with key features in the band structure. We find that the unexpectedly high Auger coefficient in $MAPbI_3$ originates from unique features in its band structure induced by large spin-orbit coupling (SOC) [136] (Sec. 6.3.1). We also find that k-space splitting of the band edges significantly enhances Auger recombination, and that this splitting is correlated with structural distortions (Sec. 6.3.2). It turns out distortions can be engineered by strain or by alloying, resulting in a suppression of Auger recombination. In this spirit, we search for other perovskite iodides that minimize the degree of internal distortion, thus identifying promising candidates for higher-efficiency light emitters. Finally, we discuss how the Auger effect of

$MAPbI_3$ compares with other common semiconductors (Sec. 6.3.3), and demonstrate that the experimentally observed third-order recombination coefficient is indeed due to Auger recombination.

6.2 Methodology

Fully *ab initio* calculations of Auger recombination rates have recently been developed and applied to III-V semiconductors [11]. Our calculations of eigenvalues and wave functions are based on density functional theory (DFT) [17, 18] and norm-conserving pseudopotentials as implemented in the Quantum ESPRESSO package [112]. A plane-wave energy cutoff of 100 Ry along with a $6 \times 6 \times 6$ $\mathbf{k}$-point grid were used in the calculations of structure and charge density. We use the local density approximation (LDA) for the exchange-correlation functional. To account for the LDA underestimation of the band gap, the conduction-band energies are shifted relative to the valence bands (commonly referred to as a *scissors shift*). The overall shape of the scissors-shifted band structure agrees well with the quasiparticle band structure [137], at a fraction of the computational cost.

The large SOC and lack of inversion symmetry lead to non-collinear spins. The spin-orbit interactions are included via a perturbing Hamiltonian (ΔH_{SOC}) constructed using the vector part of the *ab initio* pseudopotentials [51]. The corrections to the spin-degenerate eigenvalues are obtained by diagonalizing ΔH_{SOC} in the spin-degenerate basis (see Sec. 2.5 for details). To ensure that the spinor wave functions can be represented as a linear combination of the spin-degenerate wave functions we included 40 unoccupied conduction bands in our band structure.

Each state in the band structure can be labeled by a general state index $\mathbf{I} \equiv [\mathbf{k}, m]$, where $\mathbf{k}$ is the crystal momentum in reciprocal space and m is the band index. A

$20 \times 20 \times 20$ k-space grid was used for the calculations of the Auger rates. As discussed in Sec. 2.5, the Auger rate is given by Fermi's golden rule [11]:

$$R_{\text{Auger}} = \frac{2\pi}{\hbar} \sum_{\mathbf{1234}} f_{\mathbf{1}} f_{\mathbf{2}} (1 - f_{\mathbf{3}})(1 - f_{\mathbf{4}}) \tag{6.2}$$
$$|M_{\mathbf{1}234}|^2 \, \delta(\varepsilon_{\mathbf{1}} + \varepsilon_{\mathbf{2}} - \varepsilon_{\mathbf{3}} - \varepsilon_{\mathbf{4}}),$$

where $\mathbf{I} = \mathbf{1}, \mathbf{2}, \mathbf{3}, \mathbf{4}$ represent the four single-particle states involved in the Auger process, $f_{\mathbf{I}}$ are Fermi occupation factors, and the δ function enforces energy conservation. The matrix element $M_{\mathbf{1}234}$ is the anti-symmetrized screened-Coulomb matrix element:

$$\langle \psi_{\mathbf{1}} \psi_{\mathbf{2}} | W | \psi_{\mathbf{3}} \psi_{\mathbf{4}} \rangle - \langle \psi_{\mathbf{1}} \psi_{\mathbf{2}} | W | \psi_{\mathbf{4}} \psi_{\mathbf{3}} \rangle, \tag{6.3}$$

computed from the single-particle wave functions $\psi_{\mathbf{I}}$. If we assume equal densities of electrons and holes, we can define an Auger coefficient as $C = R_{\text{Auger}}/n^3$. The matrix elements are nonzero only if momentum is conserved (i.e., $k_{\mathbf{1}} + k_{\mathbf{2}} = k_{\mathbf{3}} + k_{\mathbf{4}}$).

In our calculations of the Auger coefficients for the eeh (C_n) and hhe (C_p) processes, we treat the scissors-shifted band gap as an adjustable parameter to model the effect of alloying $MAPbI_3$ with other materials and to better understand the interplay between the Auger coefficient and features in the band structure. Results for $MAPbI_3$ itself are obtained at the experimental band-gap value of 1.60 eV [138].

6.3 Results and Discussions

The low-temperature ground-state phase of $MAPbI_3$ is orthorhombic ($Pnma$), where the PbI_6 octahedra are significantly tilted and deformed. At slightly above room temperature (327 K), $MAPbI_3$ undergoes a transition to the cubic phase ($Pm\bar{3}m$) [139]. In our

simulations we focus on this cubic phase with a 12-atom unit cell. While the overall cell shape is cubic, atomic relaxation leads to significant deviations from the ideal perovskite positions (see Fig. 6.2). Using the Pb atom as the reference point at the center of the cubic unit cell, we find that the three iodine atoms are displaced 0.28, 0.39 and 0.43 Å from their face-centered positions, and such internal distortions lead to an effective Rashba splitting [140], which significantly modifies the band structure near the band edges and breaks the spin degeneracy. This splitting can be observed in the band structure of the relaxed unit cell shown in Fig. 6.3.

The states near the conduction-band minimum (CBM) have predominantly Pb-p character, and the valence-band maximum (VBM) states predominantly I-p character. When SOC is included, the six spin-degenerate bands near the CBM at R split into two bands at the band edge and four bands at 1.29 and 1.46 eV above the CBM, and the band gap changes from a direct gap at R to a slightly indirect transition between states slightly off R. These key features of our band structure agree well with the quasiparticle band structure of Brivio *et al.* [137]

We choose 10^{18} cm^{-3} as a typical carrier density for an LED. At this density, the initial states are restricted to the portion of the Brillouin zone near the band extrema around R. Since Auger recombination must conserve energy and momentum, the process will be strong if final states are available with energies approximately one band gap away from the band edges. Such states (labeled as **4** in Fig. 6.3) are clearly available in the $MAPbI_3$ band structure, for both the eeh and hhe processes.

At the experimental band gap of 1.60 eV, we obtain Auger coefficients of $C_n = 2.7\times10^{-29}$ cm^{6}s^{-1} and $C_p = 4.6\times10^{-29}$ cm^{6}s^{-1}. These values are two orders of magnitude larger than Auger coefficients in semiconductors with similar band gaps [47]. We now explore the physical mechanisms responsible for this enhancement.

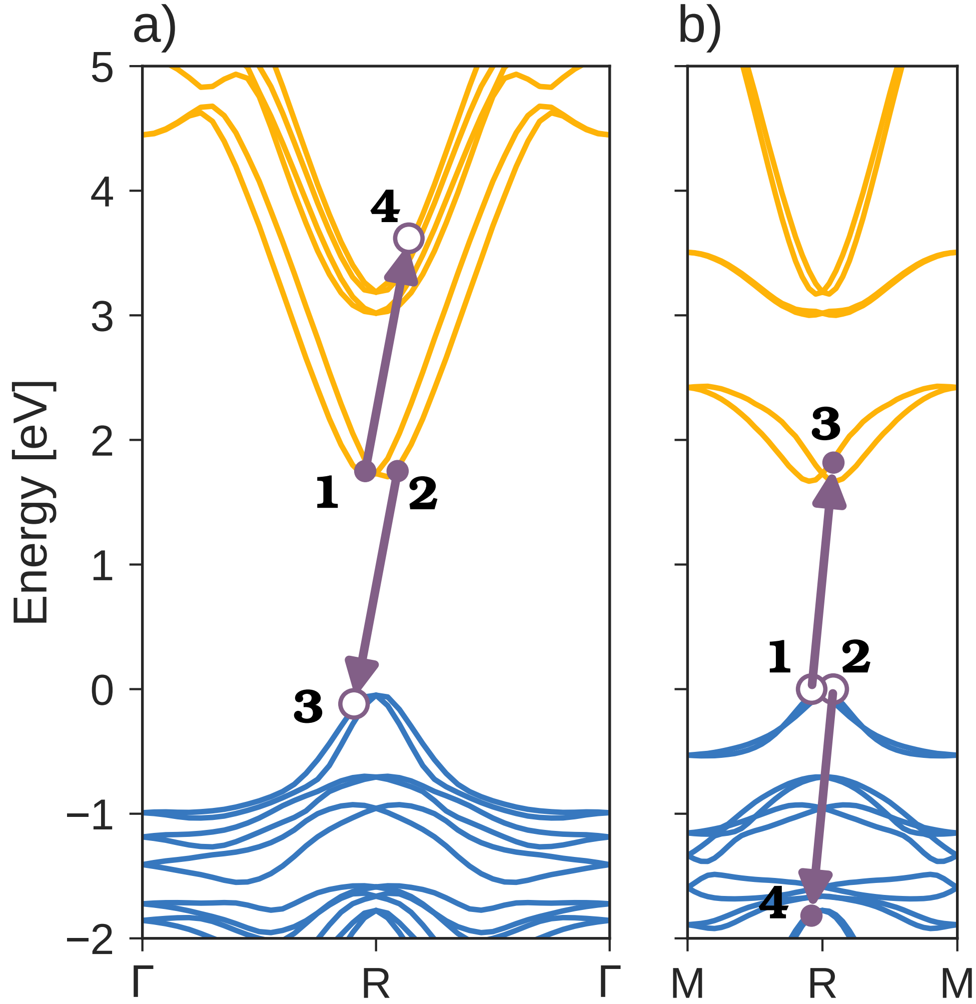


Figure 6.3: Band structure of $MAPbI_3$ including SOC, with the gap adjusted to the experimental value, plotted along the a) R → Γ and b) R → M directions. Auger recombination events for the eeh a) and hhe b) processes are indicated.

6.3.1 Coincidental Resonance

For the hhe process, an array of valence bands is available that can serve as final states for Auger recombination, explaining the high C_p value (4.6×10^{-29} cm^{6}s^{-1}). For the eeh process, Fig. 6.4 b) shows our calculated Auger coefficient as a function of band gap. This plot is obtained by computing C_n as a function of the scissors-adjusted gap, while keeping the matrix elements in Eq. (6.2) fixed. This provides an approximation to the Auger coefficients for materials that could be produced by alloying, but it also elucidates

the mechanisms responsible for the high Auger rate. If Auger transitions are restricted to the first two conduction bands (CB1 and CB2) [see Fig. 6.4 a)], the computed Auger coefficient falls off rapidly as a function of band gap—a behavior similar to intraband eeh processes in conventional direct-gap semiconductors [107]. If transitions to the higher-lying four bands (CB3–CB6) are included the Auger coefficient is dramatically increased, and shows a distinct maximum when the band gap is equal to the energy difference between CB1-CB2 and the higher-lying conduction bands (CB3-CB6) [Fig. 6.4 b)]; this happens to occur very close to the band gap of $MAPbI_3$, where C_n is enhanced by three orders of magnitude relative to the case where transitions to the higher-lying conduction bands would not occur. We conclude that most of the high eeh Auger recombination rate observed in $MAPbI_3$ can be attributed to the coincidental resonance of the band gap with interband transitions to a complex of higher-lying conduction bands.

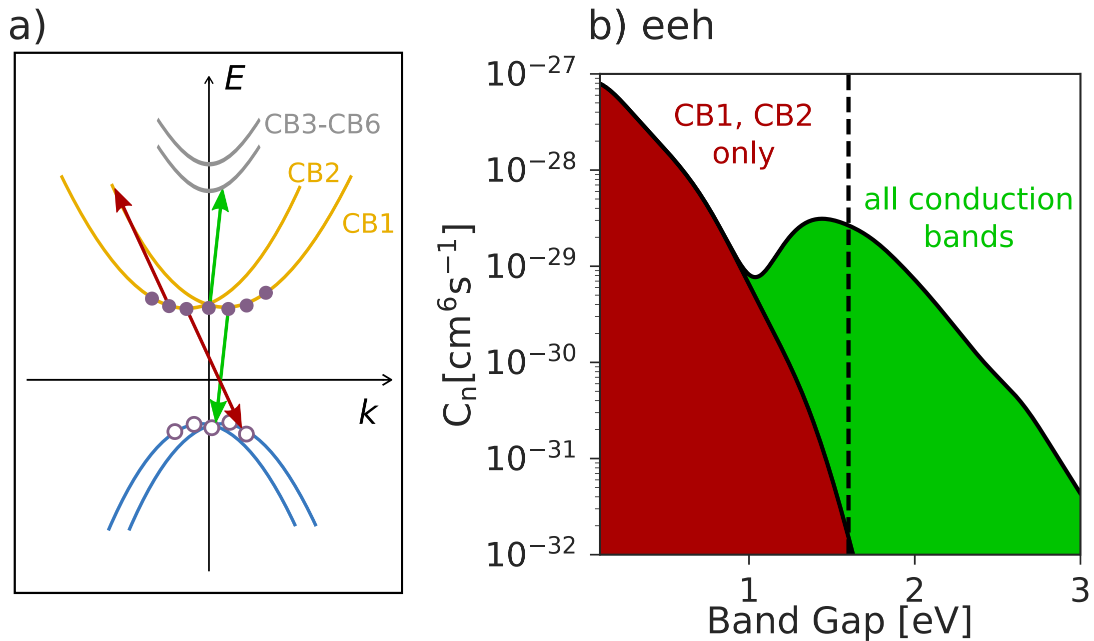


Figure 6.4: a) Schematic of Auger processes involving only the CB1-CB2 bands versus processes with transitions to the CB3-CB6 bands. b) Computed Auger coefficients as a function of scissors-adjusted band gap at a carrier density $n = 10^{18}$ cm^{-3}, for processes restricted to CB1-CB2 versus processes with transitions to all conduction bands bands. The dashed line represents the experimental band gap of 1.60 eV.

6.3.2 Effects of Lattice Distortions on Auger

Spin-orbit coupling not only leads to the splitting of the conduction bands that results in the enhancement of the eeh Auger rate, but (in combination with a breaking of inversion symmetry) it also causes a splitting of states near the band extrema [141]. We will see this splitting also has a distinct effect on the Auger coefficients.

The CBM and VBM of $MAPbI_3$ are each comprised of two spinor states. When the PbI_6 octahedra are undistorted (by fixing the Pb atom at the center of the unit cell, and the iodine atoms at the ideal face-center positions), the bands at the CBM and VBM are degenerate. When the atoms are allowed to relax, the bands split into two bands. Most noticeably, the internal distortions lead to a linear-k splitting of the band edges, which directly influences the distribution of states involved in the Auger process.

A schematic of how the CBM changes due to structural distortion is shown in Fig. 6.5. In the case of undistorted (*ideal*) Pb and I positions, the VBM and CBM are both at the high-symmetry R point. Structural relaxations lead to a linear-k splitting, which in turn results in a ring of states around R. The band edge of the relaxed structure has a higher dimensionality, and thus many more states available for the Auger process. Therefore, at the same carrier density, each of the states at the band edge for the relaxed structure will have a lower quasi-Fermi occupation factor than in the ideal structure. This is evident in the fact that the quasi-Fermi level is lower for electrons (and higher for holes) in the relaxed atomic configuration than in the ideal structure.

To examine the distributions of band-edge states that participate in Auger recombination, we can compute the extent of the region in the Brillouin zone occupied by electrons or holes. Table 6.1 shows results for % BZ corresponding to a volume in which 99% of the electrons(holes) reside, for the case of a carrier density of 10^{18} cm^{-3}. The table shows that in the ideal structure the charge carriers are much more concentrated

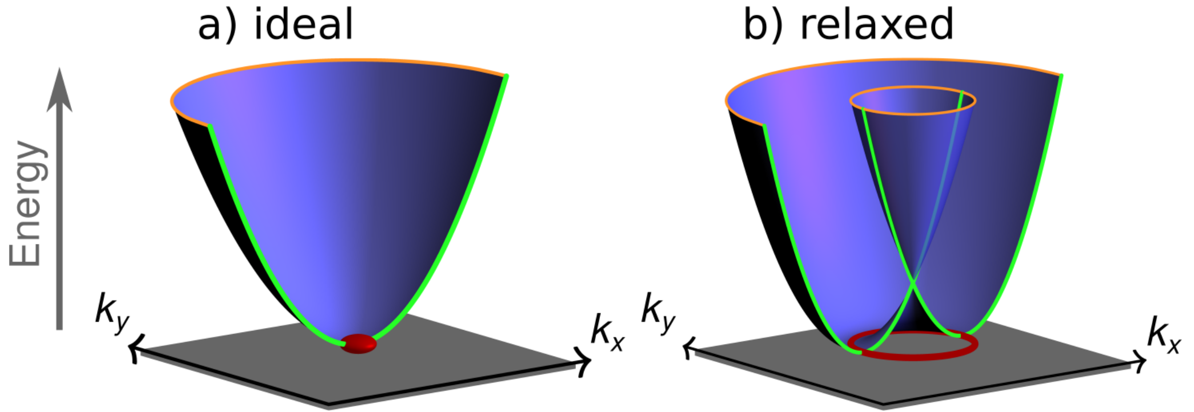


Figure 6.5: Schematic of the CBM energy dispersion of $MAPbI_3$ near the R point for a) the ideal and b) the relaxed structure. The band-edge states are shown in red.

near the R point, making it significantly harder to simultaneously satisfy energy and momentum conservation.

Table 6.1: Quasi-Fermi level μ_F (referenced to the CBM/VBM) and extent of Brillouin-zone filling (% BZ) for electrons/holes, for ideal and relaxed atomic configurations of $MAPbI_3$. The sign of μ_F is positive in case of degenerate doping for electrons (Fermi level above the CBM), and negative in case of degenerate doping for holes (Fermi level below the VBM). The computed Auger coefficients ($C_{n/p}$) are also listed. All values are obtained for a carrier density of 10^{18} cm^{-3}.

Carrier	Structure	μ_F [eV]	% BZ	$C_{n/p}$ [cm^6s^{-1}]
electrons	ideal	28.9	0.2	2.2×10^{-30}
	relaxed	-15.7	0.6	2.7×10^{-29}
holes	ideal	-26.2	0.3	4.6×10^{-30}
	relaxed	30.3	0.8	6.5×10^{-29}

The Auger coefficients versus band gap for the ideal and relaxed structures are shown in Fig. 6.6. The Auger coefficient is clearly significantly lower for the ideal structure. The Auger coefficients at the experimental band gap of 1.60 eV are listed in Table 6.1. The concentration of charge carriers near the R point causes the eeh Auger coefficient for the

ideal structure to be 92% lower than in the relaxed structure; the hhe Auger coefficient experiences a similar drop, by 86%. These Auger coefficients for the ideal structure are close to the value of $\approx 10^{-30}$ cm^6s^{-1} that seems to be a lower limit for Auger coefficients, as observed across a wide range of materials spanning a large range of band gaps [47].

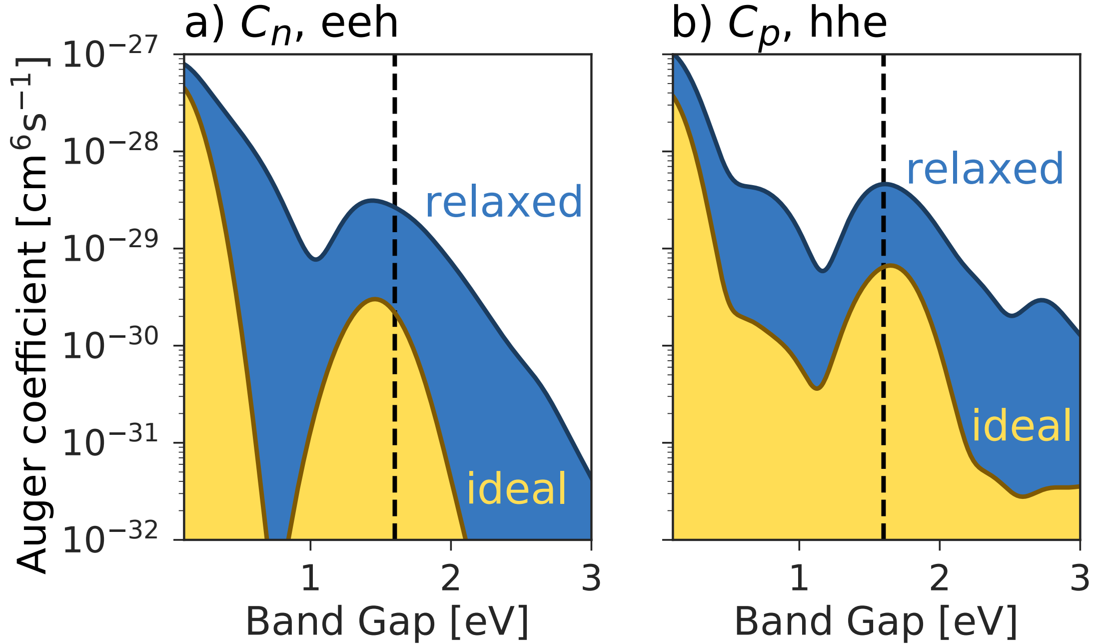


Figure 6.6: a) eeh and b) hhe Auger coefficients vs. scissors-shifted band gap for the ideal and relaxed structures.

The connection between distortions and higher Auger coefficient is evident. Schemes for controlling the octahedral distortion have been proposed, such as using anisotropic strain [142] or alloying with other halide perovskites [143]. Atomic substitutions can of course have other effects on the band structure; still, looking for materials with lower internal distortions can be a useful criterion in the search for materials with lower Auger coefficients.

The degree of distortion of the perovskite octahedra can be characterized by the

variance of the angles subtended by the bonds at the Pb atom [144]:

$$\sigma^2_{\theta(\mathrm{oct})} = \frac{1}{11}\sum_{i=1}^{12}(\theta_i - 90^\circ)^2\,. \tag{6.4}$$

We have computed the bond-angle variance $\sigma^2_{\theta(\mathrm{oct})}$ for a number of inorganic and hybrid iodide perovskites. The results, shown in Fig. 6.7, show a clear trend with the size of the cations: smaller distortions occur for larger A-site cations as well as for smaller B-site cations. A similar general trend is observed as a function of tolerance factor: the tolerance factor increases for larger A-site cations as well as for smaller B-site cations, and materials with larger tolerance factors tend to have smaller distortions [145]. We feel that the bond-angle variance provides a more direct measure of the distortions.

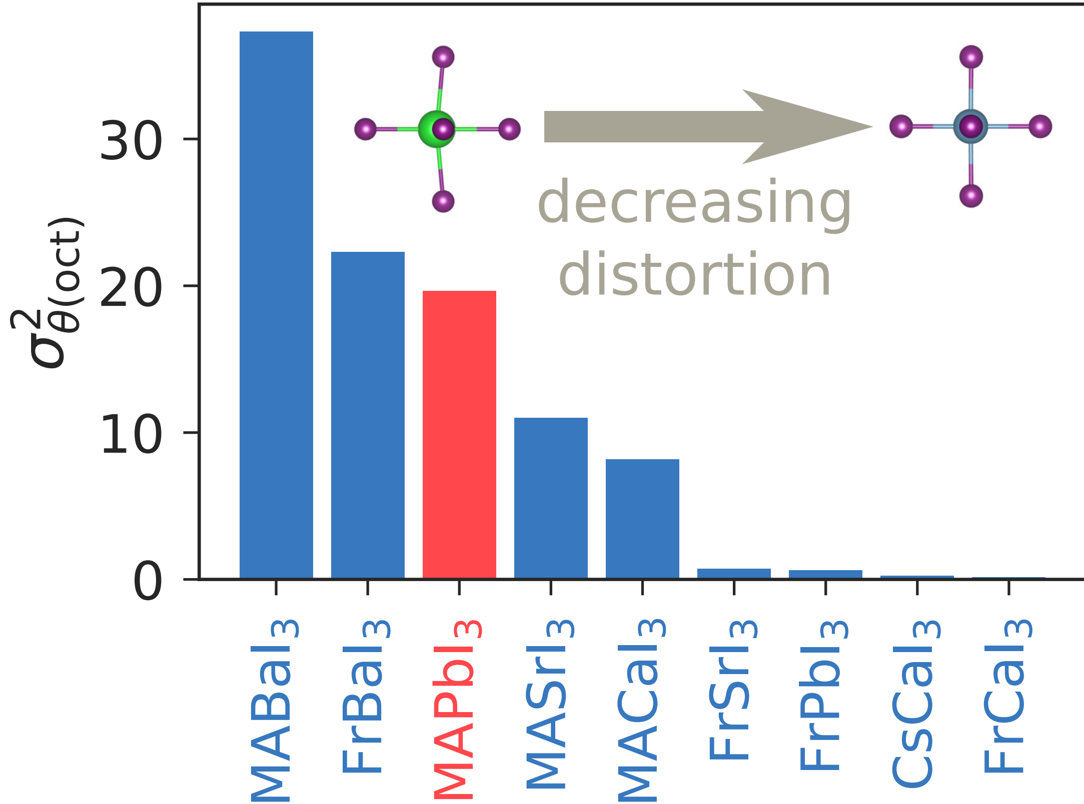


Figure 6.7: Bond-angle variance of several inorganic and hybrid halide perovskites. The calculations were done using LDA assuming the cubic phase.

6.3.3 Implications for Device Efficiency

We now discuss the impact of the Auger coefficient on the efficiency of LEDs. At the experimental band gap of 1.60 eV, the combined Auger coefficient (eeh+hhe) from our calculations is 7.3×10^{-29} cm^{6}s^{-1} for the relaxed structure and 8.6×10^{-30} cm^{6}s^{-1} for the ideal structure. In Fig. 6.8 we plot the IQE as a function of the carrier density n [Eq.(6.1)] using our calculated Auger coefficient as k_3. For the radiative recombination coefficient k_2 we use a calculated value of 8.4×10^{-11} cm^{3}s^{-1} [146], and for k_1 the value reported by Richter *et al.* [131] We compare our results with an IQE curve calculated using experimentally determined recombination coefficients $k_1 = 5 \times 10^6$ s^{-1}, $k_2 = 8.1 \times 10^{-11}$ cm^{3}s^{-1}, and $k_3 = 1.1 \times 10^{-28}$ cm^{6}s^{-1}, obtained in Ref. [131] from fitting transient absorption (TA) measurements. In Ref. [131] it was also demonstrated this curve matched results from photoluminescence quantum yield measurements. The calculated and experimental IQE agree very well with each other, thus confirming that the third-order recombination observed in $MAPbI_3$ is primarily due to Auger recombination.

Figure 6.8 indicates that Auger recombination causes the IQE of $MAPbI_3$ to drop off above a carrier density of 2.5×10^{17} cm^{-3}, and also reduces the peak IQE. However, as discussed in Sec. 6.3.2 and Fig. 6.6, Auger recombination can be dramatically reduced if lattice distortions are suppressed. This reduced Auger coefficient results in an IQE with a peak that increases from 69% to 86%, with the carrier density at peak efficiency shifting to 7.6×10^{17} cm^{-3} (Fig. 6.8).

6.4 Conclusions

We have computed the Auger recombination coefficient of the prototypical halide perovskite, $MAPbI_3$. Our calculations result in a total Auger coefficient (eeh+hhe) of 7.3×10^{-29} cm^{6}s^{-1}, in good agreement with the third-order recombination coefficients

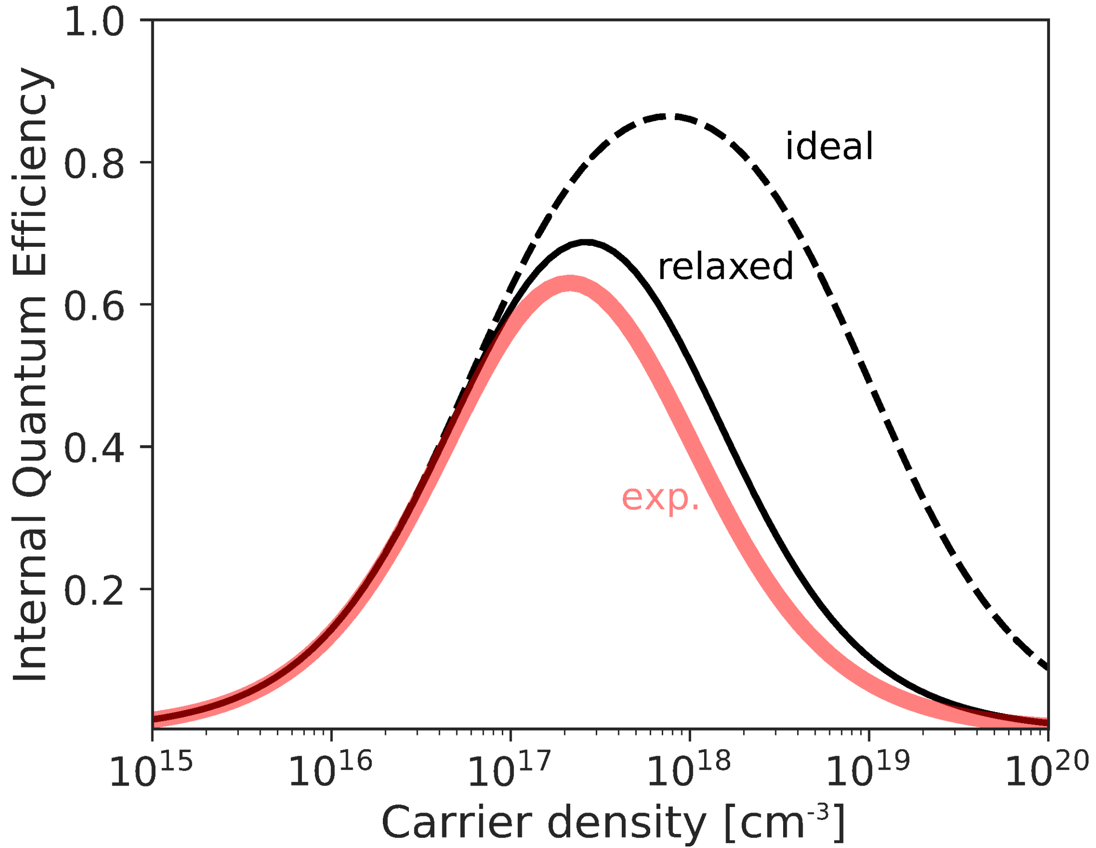


Figure 6.8: Internal quantum efficiency calculated as a function of carrier density [Eq. (6.1)] using k_3 values corresponding to our calculated Auger coefficients for the relaxed (solid curve) and ideal (dashed black curve) perovskite structures, and compared to experimental values.

reported by experiments [135, 132, 133], confirming that Auger is the dominant loss mechanism in $MAPbI_3$ at high carrier concentrations.

The high Auger recombination coefficient for the eeh process is mainly due to the coincidental resonance of the band gap with states that are roughly one band gap away from the band edges. We also found that if the distortions in the metal-halide lattice are removed, Auger recombination can be dramatically suppressed. To aid future attempts at engineering this feature, we computed the octahedral distortions in a number of halide perovskites. In order for halide perovskites to reach quantum efficiencies comparable to those of traditional III-V semiconductor devices, materials will have to be identified that

avoid the coincidental resonance that is present in $MAPbI_3$, and that minimize distortions of the metal-halide lattice.

Chapter 7

Looking Forward

Throughout this thesis, we have demonstrated methodology for studying the nonradiative recombination of carriers in different materials. We have shown that first-principles calculations can be a powerful tool in understanding how basic properties such as atomic distortions and band gap can influence the recombination processes in a given material. The work presented here can be extended to address additional problems in materials research and to gain better understandings of the underlying physics, and the methodologies can also be improved to allow for more efficient calculations.

7.1 Auger in IR detectors

Materials systems such as InSb, InAsSb, and InAs/GaSb quantum wells emitting in the infrared regime are of great interest in photonics. One potential application is in next-generation on-chip energy-efficient CO, CO_2, and CH_4 gas sensors. As we have shown in Chapter 5, Auger recombination in these lower-band-gap materials is dominated by the direct Auger process. Strategies to suppress the Auger recombination involve various kinds of band-structure engineering, aimed at making energy and momentum

conservation more difficult to satisfy. Our methodology allows testing many of these strategies, such as (i) reducing the band-edge density of states, (ii) eliminating the final states for the excited carriers. Such investigations can produce an assessment of the effectiveness of specific approaches in a well-controlled environment.

7.2 Band-Gap Independence of Indirect Auger

A striking observation of the available Auger data in Fig. 6.1 is that in the higher-band-gap materials, where phonon-assisted Auger is the dominant recombination mechanism, the Auger coefficient is essentially independent of band gap. Our calculations do show evidence that the phonon-assisted Auger coefficient is less sensitive to the value of the band gap at higher band gaps (Fig. 5.3). The physical origin of this band-gap-independence of the recombination coefficient across many materials is still not clear, and requires further investigation.

7.3 Improving the SRH Calculations

The SRH methodology relies on the *single-phonon approximation* (discussed in Sec .2.4.2). In a separate study [37], we have shown that this approximation yields quantitatively accurate results when compared to methods that consider all lattice degrees of freedom [38, 39]. As discussed in the original paper by Alkauskas *et al.* [12], the single-phonon approximation is most reliable when the local displacement between the different charge states of a defect is large relative to the natural vibrations of the crystal. At these large displacements, the potential-energy curve we have used to calculate the vibrational contributions to the electron-phonon matrix element in Eq. (2.51) could well have anharmonic contributions. Fortunately, in the nitride materials that we have investigated

thus far, there was little anharmonicity in the configuration coordinate diagrams for the special phonon. However, there is evidence that the anharmonicity is an issue for defects in methyl-ammonium lead iodide ($MAPbI_3$) [147]. Considering that the defect tolerance of $MAPbI_3$ (i.e., the fact that $MAPbI_3$ exhibits low carrier trapping despite the relatively crude synthesis process) remains one of the mysteries of this material [148], this is a promising direction for future projects.

7.4 Improving the Auger Calculations

A significant barrier to performing calculations of the Auger recombination coefficient on a larger scale is the fact that a dense k-point mesh is needed to capture the distribution of carriers at the band edges. In order to reduce the computational cost, some type of interpolation scheme is needed. This can be accomplished by projecting onto either maximally localized Wannier functions [149] or onto atomic orbital basis functions [150]. Both interpolation schemes are regularly used to evaluate the band structure and the electron-phonon matrix elements on a fine grid, but have yet to be applied to the four-body Coulomb scattering matrix elements needed in the Auger calculations. A promising direction of future research will be to implement this type of interpolation scheme for the Auger calculations. This could decrease the computational load by more than an order of magnitude and allow studying many more materials in the future.

Bibliography

[1] B. K. Ridley, *Quantum Processes in Semiconductors (Oxford Science Publications)*, Oxford University Press, 2000.

[2] W. Pauli, *Z. Angew. Phys.* 1925, **31**, 765.

[3] *Out of the Crystal Maze*, Oxford University Press, 1992.

[4] A. H. Wilson, *Proc. R. Soc. Lond. A* 1931, **133**, 458.

[5] W. Shockley, W. T. Read, *Phys. Rev.* 1952, **87**, 835.

[6] R. N. Hall, *Phys. Rev.* 1952, **87**, 387.

[7] W. van Roosbroeck, W. Shockley, *Phys. Rev.* 1954, **94**, 1558.

[8] A. R. Beattie, P. T. Landsberg, *Proc. R. Soc. A Math. Phys. Eng. Sci.* 1959, **249**, 16.

[9] W. Kohn, L. J. Sham, *Phys. Rev.* 1965, **140**, A1133.

[10] M. A. Reshchikov, J. D. McNamara, A. Behrends, M. S. Mohajerani, A. Bakin, A. Waag, *Phys. Status solidi C* 2013, **10**, 507.

[11] E. Kioupakis, D. Steiauf, P. Rinke, K. T. Delaney, C. G. Van de Walle, *Phys. Rev. B* 2015, **92**, 035207.

[12] A. Alkauskas, Q. Yan, C. G. Van de Walle, *Phys. Rev. B* 2014, **90**, 075202.

[13] P. A. M. Dirac, *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 123, The Royal Society, 1929 pages 714–733.

[14] E. Fermi, *Z. Angew. Phys.* 1928, **48**, 73.

[15] F. R. S. D. R. Hartree, W. Hartree, *Proc. R. Soc. Lond. A* 1935, **150**, 9.

[16] V. Fock, *Z. Angew. Phys.* 1930, **61**, 126.

[17] P. Hohenberg, W. Kohn, *Phys. Rev.* 1964, **136**, B864.

[18] W. Kohn, L. J. Sham, *Phys. Rev.* 1965, **140**, A1133.

[19] D. M. Ceperley, B. J. Alder, *Phys. Rev. Lett.* 1980, **45**, 566.

[20] J. P. Perdew, R. G. Parr, M. Levy, J. L. Balduz Jr, *Phys. Rev. Lett.* 1982, **49**, 1691.

[21] R. M. Dreizler, E. Engel, *Density Functional Theory: An Advanced Course*, Springer, 2011.

[22] A. D. Becke, *The Journal of chemical physics* 1993, **98**, 1372.

[23] J. Heyd, G. E. Scuseria, M. Ernzerhof, *J. Chem. Phys.* 2003, **118**, 8207.

[24] J. Heyd, G. E. Scuseria, M. Ernzerhof, *J. Chem. Phys.* 2006, **124**, 219906.

[25] J. P. Perdew, K. Burke, M. Ernzerhof, *Phys. Rev. Lett.* 1996, **77**, 3865.

[26] M. Marsman, J. Paier, A. Stroppa, G. Kresse, *J. Phys. Condens. Matter* 2008, **20**, 064201.

[27] F. Bloch, *Z. Angew. Phys.* 1929, **52**, 555.

[28] H. J. Monkhorst, J. D. Pack, *Phys. Rev. B* 1976, **13**, 5188.

[29] D. Hamann, M. Schlüter, C. Chiang, *Phys. Rev. Lett.* 1979, **43**, 1494.

[30] N. Troullier, J. L. Martins, *Phys. Rev. B* 1991, **43**, 1993.

[31] P. E. Blöchl, *Phys. Rev. B* 1994, **50**, 17953.

[32] C. Freysoldt, B. Grabowski, T. Hickel, J. Neugebauer, G. Kresse, A. Janotti, C. G. Van de Walle, *Rev. Mod. Phys.* 2014, **86**, 253.

[33] A. Matulionis, J. Liberis, I. Matulionienė, M. Ramonas, E. Šermukšnis, J. H. Leach, M. Wu, X. Ni, X. Li, H. Morkoç, *Appl. Phys. Lett.* 2009, **95**, 192102.

[34] C. H. Henry, D. V. Lang, *Phys. Rev. B* 1977, **15**, 989.

[35] R. Pässler, *Phys. Status Solidi B* 1975, **68**, 69.

[36] A. M. Stoneham, *Theory of Defects in Solids*, Oxford University Press, 2001.

[37] D. Wickramaratne, J.-X. Shen, A. Alkauskas, C. G. Van de Walle, *Phys. Rev. B* 2018, **97**, 077301.

[38] L. Shi, K. Xu, L. W. Wang, *Phys. Rev. B* 2015, **91**, 1.

[39] J. Li, H.-F. Zhu, Y.-Y. Zhang, Z.-K. Yuan, S. Chen, X.-G. Gong, *Phys. Rev. B* 2017, **96**, 104103.

[40] B. Zapol, *Chem. Phys. Lett.* 1982, **93**, 549.

[41] F. Ansbacher, *Zeitschrift für Naturforschung A* 1959, **14**, 889.

[42] F. Giustino, *Rev. Mod. Phys.* 2017, **89**, 015003.

[43] *CPMD, http://www.cpmd.org/,*
Copyright IBM Corp 1990-2008,
Copyright MPI für Festkörperforschung Stuttgart 1997-2001.

[44] C. E. Dreyer, A. Alkauskas, J. L. Lyons, J. S. Speck, C. G. Van de Walle, *Appl. Phys. Lett.* 2016, **108**, 141101.

[45] A. Alkauskas, C. E. Dreyer, J. L. Lyons, C. G. Van de Walle, *Phys. Rev. B* 2016, **93**, 201304.

[46] G. Grosso, G. P. Parravicini, *Solid State Physics*, Academic Press, Amsterdam, 2 edition edn., 2013.

[47] K. A. Bulashevich, S. Yu. Karpov, *Phys. Status Solidi C* 2008, **5**, 2066.

[48] E. Kioupakis, P. Rinke, K. T. Delaney, C. G. Van de Walle, *Appl. Phys. Lett.* 2011, **98**, 161107.

[49] S. Baroni, S. de Gironcoli, A. Dal Corso, P. Giannozzi, *Rev. Mod. Phys.* 2001, **73**, 515.

[50] W. P. Huhn, V. Blum, *Phys. Rev. Materials* 2017, **1**, 033803.

[51] M. S. Hybertsen, S. G. Louie, *Phys. Rev. B* 1986, **34**, 2920.

[52] S. Nakamura, M. R. Krames, *Proc. IEEE* 2013, **101**, 2211.

[53] A. Pourhashemi, R. M. Farrell, D. A. Cohen, J. S. Speck, S. P. DenBaars, S. Nakamura, *Appl. Phys. Lett.* 2015, **106**, 111105.

[54] U. Mishra, P. Parikh, Yi-Feng Wu, *Proc. IEEE* 2002, **90**, 1022.

[55] C. E. Dreyer, J. L. Lyons, A. Janotti, C. G. Van de Walle, *Appl. Phys Express* 2014, **7**, 031001.

[56] C. E. Dreyer, A. Janotti, C. G. Van de Walle, D. Vanderbilt, *Phys. Rev. X* 2016, **6**, 021038.

[57] R. Geick, C. H. Perry, G. Rupprecht, *Phys. Rev.* 1966, **146**, 543.

[58] X. Li, S. Sundaram, Y. El Gmili, F. Genty, S. Bouchoule, G. Patriache, P. Disseix, F. Réveret, J. Leymarie, J.-P. Salvestrini, R. D. Dupuis, P. L. Voss, A. Ougazzaden, *J. Cryst. Growth* 2015, **414**, 119.

[59] A. Nakajima, Y. Furukawa, H. Yokoya, H. Yonezu, *J. Cryst. Growth* 2005, **278**, 437.

[60] T. L. Williamson, N. R. Weisse-Bernstein, M. A. Hoffbauer, *Phys. Status Solidi C* 2014, **11**, 462.

[61] L. K. Teles, L. M. R. Scolfaro, J. R. Leite, J. Furthmüller, F. Bechstedt, *Appl. Phys. Lett.* 2002, **80**, 1177.

[62] J.-C. Zheng, H.-Q. Wang, C. H. A. Huan, A. T. S. Wee, *J. Phys. Condens. Matter* 2001, **13**, 5295.

[63] L. Teles, J. Furthmüller, L. Scolfaro, A. Tabata, J. Leite, F. Bechstedt, T. Frey, D. As, K. Lischka, *Phys. E Low-dimensional Syst. Nanostructures* 2002, **13**, 1086.

[64] S. Kumar, S. Joshi, B. Joshi, S. Auluck, *J. Phys. Chem. Solids* 2015, **86**, 101.

[65] M. Zhang, X. Li, *Phys. Status Solidi B* 2017, **254**, 1600749.

[66] J. L. Lyons, A. Janotti, C. G. Van de Walle, *Phys. Rev. B* 2014, **89**, 035204.

[67] G. Kresse, J. Furthmüller, *Phys. Rev. B* 1996, **54**, 11169.

[68] G. Kresse, J. Furthmüller, *Comput. Mater. Sci.* 1996, **6**, 15.

[69] S. Grimme, J. Antony, S. Ehrlich, H. Krieg, *J. Chem. Phys.* 2010, **132**, 154104.

[70] J. W. Nicklas, J. W. Wilkins, *Appl. Phys. Lett.* 2010, **97**, 091902.

[71] V. Popescu, A. Zunger, *Phys. Rev. B* 2012, **85**, 085201.

[72] Y. Zhang, A. Mascarenhas, L.-W. Wang, *Phys. Rev. B* 2008, **78**, 235202.

[73] S.-H. Wei, L. G. Ferreira, J. E. Bernard, A. Zunger, *Phys. Rev. B* 1990, **42**, 9622.

[74] J. Li, K. B. Nam, M. L. Nakarmi, J. Y. Lin, H. X. Jiang, P. Carrier, S.-H. Wei, *Appl. Phys. Lett.* 2003, **83**, 5163.

[75] D. Brunner, H. Angerer, E. Bustarret, F. Freudenberg, R. Hopler, R. Dimitrov, O. Ambacher, M. Stutzmann, *J. Appl. Phys.* 1997, **82**, 5090.

[76] G. Cassabois, P. Valvin, B. Gil, *Nat. Photonics* 2016, **10**, 262.

[77] S. Iwama, K. Hayakawa, T. Arizumi, *J. Cryst. Growth* 1982, **56**, 265.

[78] A. Nagakubo, H. Ogi, H. Sumiya, K. Kusakabe, M. Hirao, *Appl. Phys. Lett.* 2013, **102**, 241909.

[79] V. L. Solozhenko, D. Hausermann, M. Mezouar, M. Kunz, *Appl. Phys. Lett.* 1998, **72**, 1691.

[80] Y. Andreev, T. Lundström, *J. Alloys Compd.* 1994, **216**, L5.

[81] V. L. Solozhenko, *High Press. Res.* 1995, **13**, 199.

[82] D. A. Evans, A. G. McGlynn, B. M. Towlson, M. Gunn, D. Jones, T. E. Jenkins, R. Winter, N. R. J. Poolton, *J. Phys. Condens. Matter* 2008, **20**, 75233.

[83] C. K. Shih, W. E. Spicer, W. A. Harrison, A. Sher, *Phys. Rev. B* 1985, **31**, 1139.

[84] L. Bellaiche, A. Zunger, *Phys. Rev. B* 1998, **57**, 4425.

[85] S. Nakamura, M. R. Krames, *Proc. IEEE* 2013, **101**, 2211.

[86] K. A. Bulashevich, A. V. Kulik, S. Yu. Karpov, *Phys. Status Solidi A* 2015, **212**, 914.

[87] S. F. Chichibu, A. Uedono, T. Onuma, T. Sota, B. A. Haskell, S. P. DenBaars, J. S. Speck, S. Nakamura, *Appl. Phys. Lett.* 2005, **86**, 021914.

[88] E. C. Young, N. Grandjean, T. E. Mates, J. S. Speck, *Appl. Phys. Lett.* 2016, **109**, 212103.

[89] A. J. Ptak, D. J. Friedman, S. Kurtz, R. C. Reedy, M. Young, D. B. Jackrel, H. B. Yuen, S. R. Bank, M. A. Wistey, J. S. Harris, *J. Vac. Sci. Technol. B* 2006, **24**, 1540.

[90] J. B. Hurst, S. D. Lewis, M. M. Oye, A. L. Holmes, A. J. Ptak, R. C. Reedy, *J. Vac. Sci. Technol. B* 2007, **25**, 1058.

[91] M. M. Oye, S. R. Bank, A. J. Ptak, R. C. Reedy, M. S. Goorsky, A. L. Holmes, *J. Vac. Sci. Technol. B* 2008, **26**, 1058.

[92] A. Alfantazi, R. Moskalyk, *Miner. Eng.* 2003, **16**, 687.

[93] J. Heyd, G. E. Scuseria, M. Ernzerhof, *J. Chem. Phys.* 2006, **124**, 219906.

[94] C. Freysoldt, J. Neugebauer, C. Van de Walle, *Phys. Rev. Lett.* 2009, **102**, 016402.

[95] J. Neugebauer, C. G. Van de Walle, *J. Appl. Phys.* 1999, **85**, 3003.

[96] C. E. Dreyer, A. Janotti, C. G. Van de Walle, D. Vanderbilt, *Phys. Rev. X* 2016, **6**, 021038.

[97] C. G. Van de Walle, J. Neugebauer, *Appl. Phys. Lett.* 1997, **70**, 2577.

[98] P. G. Moses, C. G. Van de Walle, *Appl. Phys. Lett.* 2010, **96**, 021908.

[99] E. Kioupakis, Q. Yan, D. Steiauf, C. G. Van de Walle, *New J. Phys.* 2013, **15**, 125006.

[100] J. Danhof, U. T. Schwarz, T. Meyer, C. Vierheilig, M. Peter, *Phys. Status Solidi B* 2012, **249**, 600.

[101] Y. C. Shen, G. Mueller, S. Watanabe, N. Gardner, A. Munkholm, M. Krames, *Appl. Phys. Lett.* 2007, **91**, 141101.

[102] M. Green, *IEEE Trans. Electron Devices* 1984, **31**, 671.

[103] H.-W. Chiang, J. C. Rode, P. Choudhary, M. J. W. Rodwell, *J. Appl. Phys.* 2014, **116**, 164509.

[104] D. H. Chow, R. H. Miles, T. C. Hasenberg, A. R. Kost, Y.-H. Zhang, H. L. Dunlap, L. West, *Appl. Phys. Lett.* 1995, **67**, 3700.

[105] W. W. Bewley, C. L. Canedy, C. S. Kim, M. Kim, C. D. Merritt, J. Abell, I. Vurgaftman, J. R. Meyer, *Opt. Express* 2012, **20**, 3235.

[106] J. R. Meyer, C. L. Felix, W. W. Bewley, I. Vurgaftman, E. H. Aifer, L. J. Olafsen, J. R. Lindle, C. A. Hoffman, M.-J. Yang, B. R. Bennett, B. V. Shanabrook, H. Lee, C.-H. Lin, S. S. Pei, R. H. Miles, *Appl. Phys. Lett.* 1998, **73**, 2857.

[107] D. Steiauf, E. Kioupakis, C. G. Van de Walle, *ACS Photonics* 2014, **1**, 643.

[108] J. M. Rondinelli, E. Kioupakis, *Annu. Rev. Mater. Res.* 2015, **45**, 491.

[109] I. Vurgaftman, J. R. Meyer, L. R. Ram-Mohan, *J. Appl. Phys.* 2001, **89**, 5815.

[110] Z. M. Fang, K. Y. Ma, D. H. Jaw, R. M. Cohen, G. B. Stringfellow, *J. Appl. Phys.* 1990, **67**, 7034.

[111] Y.-S. Kim, M. Marsman, G. Kresse, F. Tran, P. Blaha, *Phys. Rev. B* 2010, **82**, 205212.

[112] P. Giannozzi, S. Baroni, N. Bonini, M. Calandra, R. Car, C. Cavazzoni, D. Ceresoli, G. L. Chiarotti, M. Cococcioni, I. Dabo, A. Dal Corso, S. de Gironcoli, S. Fabris, G. Fratesi, R. Gebauer, U. Gerstmann, C. Gougoussis, A. Kokalj, M. Lazzeri, L. Martin-Samos, N. Marzari, F. Mauri, R. Mazzarello, S. Paolini, A. Pasquarello, L. Paulatto, C. Sbraccia, S. Scandolo, G. Sclauzero, A. P. Seitsonen, A. Smogunov, P. Umari, R. M. Wentzcovitch, *J. Phys. Condens. Matter* 2009, **21**, 395502.

[113] S. Adachi, *J. Appl. Phys.* 1982, **53**, 8775.

[114] V. L. Dalal, W. A. Hicinbothem, H. Kressel, *Appl. Phys. Lett.* 1974, **24**, 184.

[115] K. L. Vodopyanov, H. Graener, C. C. Phillips, T. J. Tate, *Phys. Rev. B* 1992, **46**, 13194.

[116] J. R. Lindle, J. R. Meyer, C. A. Hoffman, F. J. Bartoli, G. W. Turner, H. K. Choi, *Appl. Phys. Lett.* 1995, **67**, 3153.

[117] J. E. L. Hollis, *Proc. Phys. Soc.* 1967, **91**, 151.

[118] S. Marchetti, M. Martinelli, R. Simili, *J. Phys.: Condens. Matter* 2002, **14**, 3653.

[119] W. W. Bewley, J. R. Lindle, C. S. Kim, M. Kim, C. L. Canedy, I. Vurgaftman, J. R. Meyer, *Appl. Phys. Lett.* 2008, **93**, 041118.

[120] M. Takeshima, *Phys. Rev. B* 1982, **25**, 5390.

[121] A. McAllister, D. Åberg, A. Schleife, E. Kioupakis, *Appl. Phys. Lett.* 2015, **106**, 141901.

[122] K. J. Cheetham, A. Krier, I. P. Marko, A. Aldukhayel, S. J. Sweeney, *Appl. Phys. Lett.* 2011, **99**, 141110.

[123] Z. Xiao, R. A. Kerner, L. Zhao, N. L. Tran, K. M. Lee, T.-W. Koh, G. D. Scholes, B. P. Rand, *Nat. Photonics* 2017, **11**, 108.

[124] A. Kojima, K. Teshima, Y. Shirai, T. Miyasaka, *J. Am. Chem. Soc.* 2009, **131**, 6050.

[125] H. Zhou, Q. Chen, G. Li, S. Luo, T.-b. Song, H.-S. Duan, Z. Hong, J. You, Y. Liu, Y. Yang, *Science* 2014, **345**, 542.

[126] N. J. Jeon, J. H. Noh, W. S. Yang, Y. C. Kim, S. Ryu, J. Seo, S. I. Seok, *Nature* 2015, **517**, 476.

[127] O. D. Miller, E. Yablonovitch, S. R. Kurtz, *IEEE J. Photovoltaics* 2012, **2**, 303.

[128] S. D. Stranks, H. J. Snaith, *Nat. Nanotechnol.* 2015, **10**, 391.

[129] S. Adjokatse, H.-H. Fang, M. A. Loi, *Mater. Today* 2017, **20**, 413.

[130] S. P. DenBaars, *Solid State Lumin.*, pages 263–291, Springer Netherlands, Dordrecht, 1993.

[131] J. M. Richter, M. Abdi-Jalebi, A. Sadhanala, M. Tabachnyk, J. P. Rivett, L. M. Pazos-Outón, K. C. Gödel, M. Price, F. Deschler, R. H. Friend, *Nat. Commun.* 2016, **7**, 13941.

[132] J. Fu, Q. Xu, G. Han, B. Wu, C. H. A. Huan, M. L. Leek, T. C. Sum, *Nat. Commun.* 2017, **8**, 1300.

[133] C. Wehrenfennig, G. E. Eperon, M. B. Johnston, H. J. Snaith, L. M. Herz, *Adv. Mater.* 2014, **26**, 1584.

[134] Y. Yang, M. Yang, Z. Li, R. Crisp, K. Zhu, M. C. Beard, *J. Phys. Chem. Lett.* 2015, **6**, 4688.

[135] R. L. Milot, G. E. Eperon, H. J. Snaith, M. B. Johnston, L. M. Herz, *Adv. Funct. Mater.* 2015, **25**, 6218.

[136] J. Even, L. Pedesseau, J.-M. Jancu, C. Katan, *J. Phys. Chem. Lett.* 2013, **4**, 2999.

[137] F. Brivio, K. T. Butler, A. Walsh, M. van Schilfgaarde, *Phys. Rev. B* 2014, **89**, 155204.

[138] A. M. A. Leguy, P. Azarhoosh, M. I. Alonso, M. Campoy-Quiles, O. J. Weber, J. Yao, D. Bryant, M. T. Weller, J. Nelson, A. Walsh, M. van Schilfgaarde, P. R. F. Barnes, *Nanoscale* 2016, **8**, 6317.

[139] M. T. Weller, O. J. Weber, P. F. Henry, A. M. Di Pumpo, T. C. Hansen, *Chem. Commun.* 2015, **51**, 4180.

[140] X. Zhang, Q. Liu, J.-W. Luo, A. J. Freeman, A. Zunger, *Nat. Phys.* 2014, **10**, 387.

[141] M. Kepenekian, J. Even, *J. Phys. Chem. Lett.* 2017, **8**, 3362.

[142] L. Leppert, S. E. Reyes-Lillo, J. B. Neaton, *J. Phys. Chem. Lett.* 2016, **7**, 3683.

[143] J.-H. Lee, N. C. Bristowe, J. H. Lee, S.-H. Lee, P. D. Bristowe, A. K. Cheetham, H. M. Jang, *Chem. Mater.* 2016, **28**, 4259.

[144] K. Robinson, G. V. Gibbs, P. H. Ribbe, *Science* 1971, **172**, 567.

[145] V. M. Goldschmidt, *Naturwissenschaften* 1926, **14**, 477.

[146] X. Zhang, J.-X. Shen, W. Wang, C. G. Van de Walle, 2018.

[147] L. Martiradonna, *Nat. Mater.* 2018, **17**, 377.

[148] D. Meggiolaro, S. G. Motti, E. Mosconi, A. J. Barker, J. Ball, C. A. R. Perini, F. Deschler, A. Petrozza, F. De Angelis, *Energy Environ. Sci.* 2018, **11**, 702.

[149] N. Marzari, A. A. Mostofi, J. R. Yates, I. Souza, D. Vanderbilt, *Rev. Mod. Phys.* 2012, **84**, 1419.

[150] L. A. Agapito, S. Ismail-Beigi, S. Curtarolo, M. Fornari, M. B. Nardelli, *Phys. Rev. B* 2016, **93**, 035104.