A Multiscale Study of P-I-N Devices for Computing and Sensing Applications

---

A Dissertation

Presented to

the faculty of the School of Engineering and Applied Science

University of Virginia

---

in partial fulfillment

of the requirements for the degree

Doctor of Philosophy

by

Sheikh Ziauddin Ahmed

December 2021

APPROVAL SHEET

This

Dissertation

is submitted in partial fulfillment of the requirements
for the degree of

Doctor of Philosophy

Author: Sheikh Ziauddin Ahmed

This Dissertation has been read and approved by the examing committee:

Advisor: Avik W. Ghosh

Advisor:

Committee Member: Joe C. Campbell

Committee Member: Benton H. Calhoun

Committee Member: Mona Zebarjadi

Committee Member: Kyusang Lee

Committee Member: Nikhil Shukla

Committee Member:

Accepted for the School of Engineering and Applied Science:

Jennifer L. West, School of Engineering and Applied Science

December 2021

To my family and friends

Abstract

The advent of 5G communication and Internet of Things applications is driving the demand for faster and more efficient computers and communication devices. The stalling of Moore's law due to CMOS scalability issues is also propelling research into novel device architectures and computing approaches. Semiconductor devices involved in computing and sensing technologies are commonly based on *p-i-n* junctions. *p-i-n* Tunnel Field Effect Transistors (TFETs) are one of the major candidates for sub-thermal operation of transistors, while *p-i-n* Avalanche Photodiodes (APDs) are widely deployed in applications like single photon detection, fiber optic communication and LIDAR systems due to their high gain-bandwidth product and low noise. An ideal III-V planar TFET operates on ballistic transport, but their performance is compromised by higher-order scattering processes. The desired transport mechanism in TFETs is band to band tunneling, with Auger generation (impact ionization) being undesirable. This research initially investigates the effects of the TFET non-idealities. The later chapters explore and establish the origins of observed low excess noise in *p-i-n* III-V digital alloy APDs. Design principles are then proposed based on these observations. In these APDs, impact ionization and unipolar transport are the key mechanisms for attaining gain, while tunneling promotes bipolarity and is detrimental to their performance. Since the materials and structures of the two devices are similar, it is possible to extend the simulation tools developed for TFETs to APDs.

To investigate the effect of TFET non-idealities, a chemistry-based analytical model is clearly needed that has a proper tunneling equation which accounts for multiple transverse

modes, and includes accurate material chemistry, electrostatics and temperature effects. Presently, there is no such good model that captures all these effects. A quasi-analytical model for planar III-V TFETs is developed to study the effect of these higher-order processes, It is then demonstrated using the model that the observed discrepancy between experimental and theoretical TFET devices is due to the presence of non-ideal processes. Furthermore, the model shows that the minimum subthreshold swing in planar TFETs is limited by the Auger generation process.

Currently, the underlying physics of low noise III-V APDs are not well understood. A solid understanding of these physics will enable us to design better performing APDs. Thus, the material bandstructures of the III-V digital alloys are studied to investigate the origin of the low excess noise, utilizing an Environment-Dependent Tight Binding Model coupled with a band unfolding technique. It is shown that a combination of "minigaps", increased effective mass, and large separation between light-hole and split-off bands lead to reduced excess noise. Thereafter, both quantum kinetic Non-Equilibrium Green's Function and Boltzmann transport formalisms are used to show that these properties prevent hole ionization in many of these digital alloys. The resulting unipolarity creates low excess noise in the electron injected APDs. Based on these simulations, this thesis proposes some empirical design criteria for attaining low excess noise using digital alloys. Also, this work attributes the low noise in quaternary random alloys to large effective masses and separation between light-hole and split-off bands using bandstructure studies. Furthermore, it is demonstrated that strain generates opposite movement of bands in the alloy binary constituents which lead to the formation of the minigaps. It is then possible to control the minigap sizes by modulation of the strain, which can possibly lead to lower excess noise in APDs.

Optoelectronic devices, like APDs, are increasingly being incorporated into photonic integrated circuits. To accurately model the digital alloy APDs in these circuits, a physics-based multiscale compact model for these APDs is developed. The model provides a framework to study the digital alloys starting from material properties to device transport and eventually

their performance in circuits. Lastly, in this research work, a simple one-dimensional Non-Equilibrium Green's Function model incorporating impact ionization is developed, forming a matrix-based theoretical quantum transport framework for studying APDs. This method automatically captures quantum effects like tunneling which are not accounted for by semi-classical tools and can incorporate the effects of complicated bandstructures that generate non-parabolic, energy and voltage-dependent effective mass tensors, as seen in digital alloy superlattices. This framework will allow the development of quantum transport based accurate APD models that are more advanced than state-of-the-art semi-classical approach based models.

Acknowledgments

I am highly grateful to my advisor, Avik Ghosh, for guiding me in the right direction towards achieving the qualities of being a good researcher and engineer. His infectious enthusiasm and unlimited passion have been major driving forces throughout my graduate studies. I thank him for his time, patience and advice beyond my research.

I am thankful to my committee members, Joe C. Campbell, Benton H. Calhoun, Mona Zebarjadi, Nikhil Shukla, and Kyusang Lee, for their constructive advice on my research work that helped me prepare my dissertation proposal and defense.

I would like to thank my fellow graduate students, past and present, from the VINO group — for their companionship, feedback, and friendship. I consider myself fortunate to have been acquainted with some brilliant colleagues: Yaohua Tan, Samiran Ganguly, Jiyuan Zheng. Countless hours of discussions and arguments with them heavily influenced my thesis work.

I am really grateful to Seth Bank (University of Texas-Austin), John P. David (University of Sheffield), Yuan Yuan, Daniel S. Truesdell, and Bingtian Guo for productive collaboration and fruitful discussions.

Finally but most importantly, I would like to thank my family and friends. I am forever indebted to my parents, Sheikh Shahabuddin Ahmed and Shamima Akhter, for their constant support, wise counsel, and sympathetic ears for not only my Ph.D. journey but throughout my life. I appreciate the love and support from my sister, Halima Ferdousy. I am thankful to my two nieces, Marzia Ferdousy Mahbub and Nurjahan Ferdousy Mahbub, for being a

constant source of inspiration and joy. Last but not the least, I want to thank all my friends, who are now scattered around the world but are always a phone call away, for their friendship, support, and belief in me.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Motivation

The volume of data generated in this world is growing at an exponential rate, driven by applications such as Internet of Things, social media, streaming and even genomics driven medicine [1]. Consequently, the demand for more efficient sensors and computers is booming. Computing in the post-Moore's law era is now increasingly reliant on novel architectures and systems rather than simply improving switching devices. With the scalability of complementary metal oxide semiconductor (CMOS) transistors limited by the Boltzmann limit, new device architectures like Tunnel field-effect transistors (TFETs) [2] and Negative Capacitance FET [3], and approaches like photonic-based computing [4] and quantum computing [5] are being considered for the next generation of computers. Also, photodetectors are increasingly being incorporated onto photonic integrated circuits for Internet of Things and 5G communication applications [6–8]. The demand for efficient optical detectors is constantly growing due to rapid developments in telecommunication, light imaging, detection and ranging (LIDAR) systems and other military and research fields [9–15]. Traditional *p-i-n* diodes form the foundation for many of the devices, like Avalanche Photodiodes (APDs) and TFETs, that are required for these approaches and architectures. The structures and band

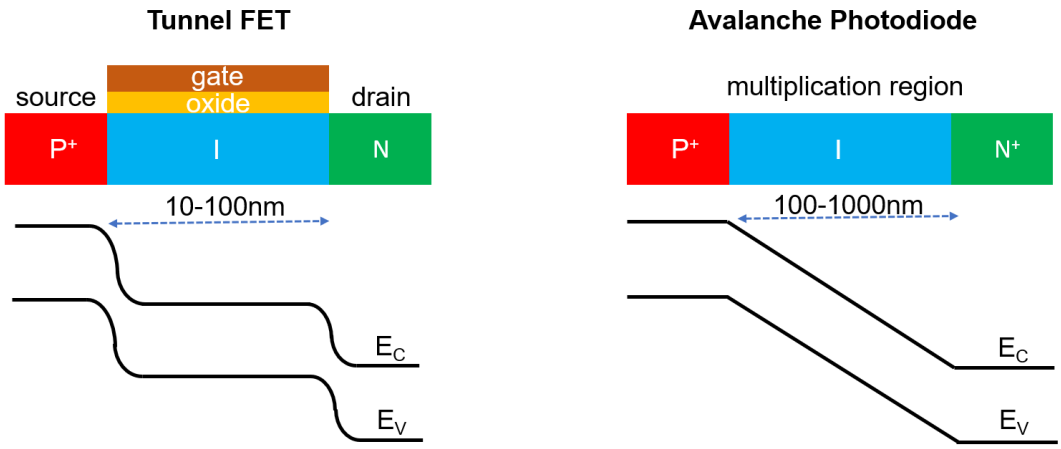

Figure 1.1: Structure schematics and band diagram of a Tunnel FET and a Avalanche Photodiode.

diagrams of these devices are shown in Fig. 1.1. Some key metrics and transport mechanisms of these devices are given in Table. 1.1.

TFETs are electronic switches that are capable of theoretically beating the Boltzmann limit of 60mV/dec that sets the limit of subthreshold swing (SS) in metal-oxide-semiconductor field-effect transistors (MOSFETs), as illustrated in Fig. 1.2. They operate by the abrupt opening of gate controlled transmission channels through band-to-band (Zener) tunneling across reverse bias $pn/p\text{-}i\text{-}n$ junctions [16, 17]. Ideally, the sharper turn on (low turn on voltage V_{DD}) characteristics of TFETs decrease the dynamic power dissipation, as the power dissipation $P_{diss} \propto V_{DD}^2$, which allows for greater transistor density in chips. Unfortunately,

Device	**TFET**	**APD**
Desired Mechanism	Tunneling	Impact Ionization (Aug. Gen.)
Undesired Mechanism	Impact Ionization (Aug. Gen.)	Tunneling
Operating Voltage	0.5-1V	10-200V
Figure of Merit	$SS < 60mV/decade$ & $I_{ON} = 1 - 10\mu A/\mu m$	$k << 1$ & GB Product 300-350GHz
Length	$10 - 100nm$	100-1000nm
Gain	1	$100 - 10^6$

Table 1.1: Key metrics of Tunnel FETs and Avalanche Photodiodes.

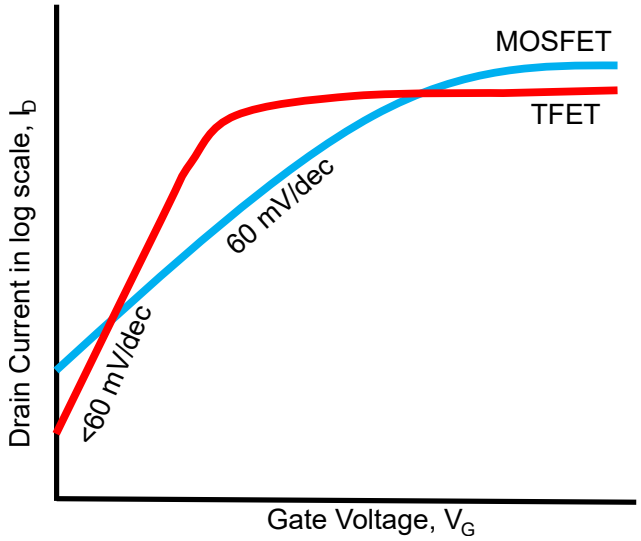

Figure 1.2: Comparison of TFET and MOSFET transfer characteristics.

none of the reported TFETs in the literature, to the best of our knowledge, approach the dual needs of high current for fast speed and a low SS over several decades from ON to OFF, as shown in Fig. 1.3, in order to get low voltage operation and thereby low dynamic power dissipation. Some experimental tunnel transistors have demonstrated below 60 mV/dec but

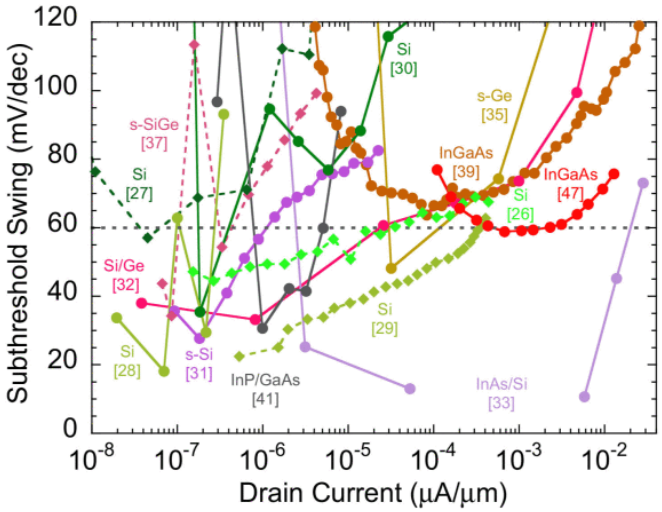

Figure 1.3: Reported Subthreshold swing vs. drain current [16] of experimental TFETs. A winning device needs to have subthreshold swing < 60 mV/decade over several decades and an on current $\sim$ 1-10 μA/μm. ©2014 IEEE

for extremely low currents [18, 19]. However, the low SS of TFETs cannot be sustained for at least four orders of magnitude of current required for driving logic circuits, nor is the low current adequate for speed and drivability. Previous attempts to increase the ON current with heterojunction Type-II staggered gap engineering while keeping the SS low have not been completely successful [16, 20]. More recently there have been reports of some III-V TFETs which have SS $<$ 60 mV/dec for two to three orders magnitude of current in the 1-100 $nA/\mu m$ range, which is significantly lower than the ON current of state-of-the-art transistors [16, 21–23]. It is therefore imperative to understand the limiting processes in TFETs, thereby identifying ways to minimize these processes in order for further improvement in device performance. For this study a sophisticated model for TFETs that accurately captures material chemistry, tunneling physics and non-ideal processes is essential.

APDs are often deployed as photodetectors these days instead of the conventional p-i-n photodiodes due to their higher sensitivity which is enabled by their intrinsic gain mechanism [24]. The materials and device structures of TFETs and APDs are similar with the key difference being the potential profile in the intrinsic region. For TFETs, the intrinsic region has a flat constant potential whereas it has a potential gradient in APDs. Due to the

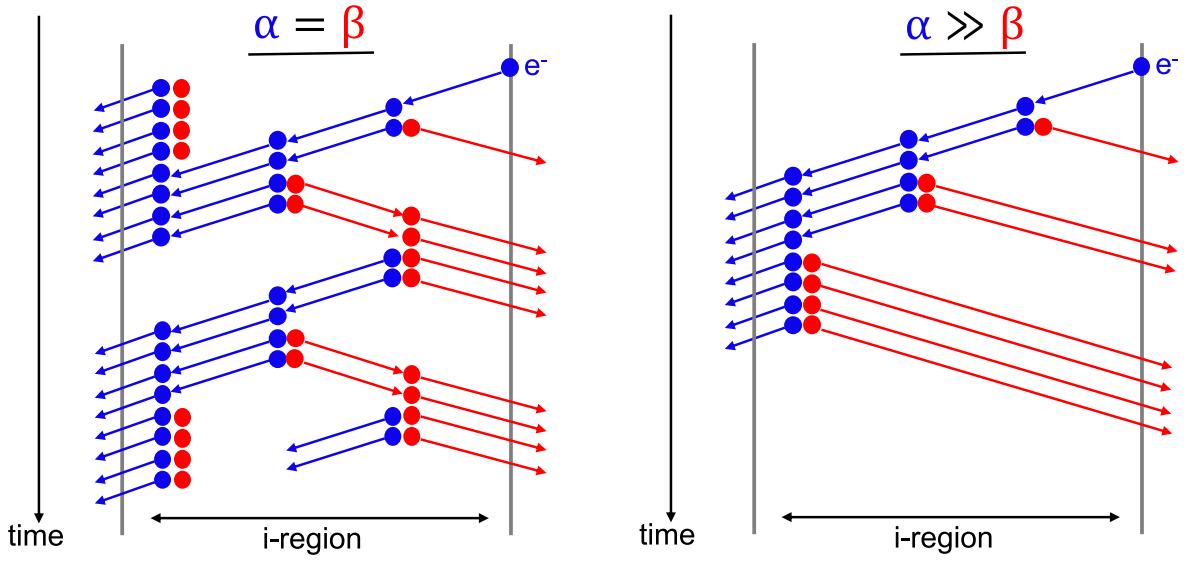

Figure 1.4: Illustration of impact ionization events in the case of $k = 1$ (left) and $k << 1$ (right), assuming electron injection.

stochastic nature of the impact ionization process of APDs, it adds an excess noise factor $F(M) = kM + (1-k)(2-1/M)$ to the shot noise current, $\langle i_{shot}^2 \rangle = 2qIM^2F(M)\Delta f$ [25–27]. Here, q is the electron charge, I is the total photo plus dark current, M the average multiplication gain and Δf is the bandwidth. This popular McIntyre's formula for $F(M)$ can be re-expressed in a more intuitive way as [28]

$$F(M) - 1 = \underbrace{\left(\frac{M-1}{M}\right)}_{unipolar} + \underbrace{k\left[\frac{(M-1)^2}{M}\right]}_{bipolar}. \tag{1.1}$$

In the first term, for a gain M there are $M-1$ excess particles and $(M-1)/M$ fraction of the total particles contribute to the excess noise. The later quadratic term arises from secondary ionization processes by impact ionization of the minority carriers. Fig. 1.4 illustrates the impact ionization process with equal electron and hole contribution (left), and for a hole-suppressed impact ionization (right). A low value of k, which is the ratio of hole ionization coefficient β to the electron ionization coefficient, α, is desirable for designing low-noise electron injected APDs. This ratio stipulates that for pure electron injection, a significantly lower hole ionization than the electron ionization rate leads to reduced shot noise. If impact ionization is caused by pure hole injection, k in the equation will be replaced

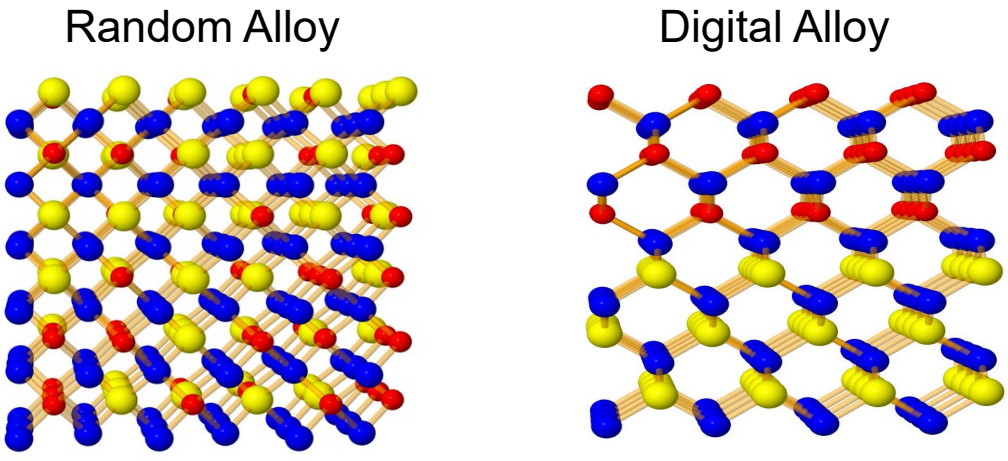

Figure 1.5: Lattice structures for InAlAs random and digital alloys [29]. ©2018 IEEE

by $1/k$. There are multiple techniques for attaining low k and hence low excess noise [24].

There have been recent reports of several III-V random and digital alloy APDs which exhibit extremely low noise currents and a high gain-bandwidth product in the short-infrared wavelength spectrum [30–35]. Digital alloys are essentially short-period superlattices with binary components stacked alternately in a periodic manner. Lattice structures of InAlAs random and digital alloys are shown in Fig. 1.5 [29]. The low noise in the digital alloys has been attributed to the presence of tiny gaps, called 'minigaps', inside the valence band of the material bandstructure that inhibit hole impact ionization [29, 36, 37]. However, other digital alloys with minigaps do not demonstrate low excess noise and this behavior is not well understood. Also, the reason behind the low noise in the quaternary random alloys has not yet been addressed. Thus, a comprehensive analysis of these materials, using sophisticated simulation tools, is necessary to understand their underlying physics.

1.2 Dissertation organization

In this work, I theoretically investigate the material and transport properties of modern III-V *p-i-n* TFETs and APDs to explain their experimental behavior. The structure of this dissertation is as follows: Chapter 1 discusses the motivation behind this research. Chapter 2 presents a quasi-analytical model for TFETs that includes the microscopic physics and chemistry of interfaces and non-idealities. The model is used to study the effect of the non-idealities on TFET performance. Furthermore, this chapter explores methods of mitigating the effect of Auger generation in these transistors. Chapter 3 investigates the origin of low excess noise in various III-V random and digital alloys using bandstructure calculations. The role of minigaps on III-V digital alloy transport is studied using a Non-Equilibrium Green's Function model I developed and a Boltzmann transport solver. Chapter 3 also proposes some empirical design principles of III-V digital alloys for use in low noise photodetectors. The formation of the minigaps is explained using a tight binding picture in chapter 4. Moreover,

this chapter investigates the role of strain in modulating the minigap sizes in the digital alloys. Chapter 5 introduces a physics-based SPICE compatible compact model for *p-i-n* APDs built from parameters extracted from the Environment-Dependent Tight Binding model calibrated to *ab-initio* Density Functional Theory and Monte Carlo methods. This model can accurately capture the physical characteristics of these APDs in integrated photonics circuit simulations. A quantum transport framework for treating impact ionization is developed in Chapter 6 using the Non-Equilibrium Green's Function formalism. Chapter 7 concludes with proposed future research topics. A summary of the main topics that will be covered in this dissertation is given in Fig. 1.6. These topics are divided into two broad categories- tool building and problem solving. The next chapters discusses these topics in more details.

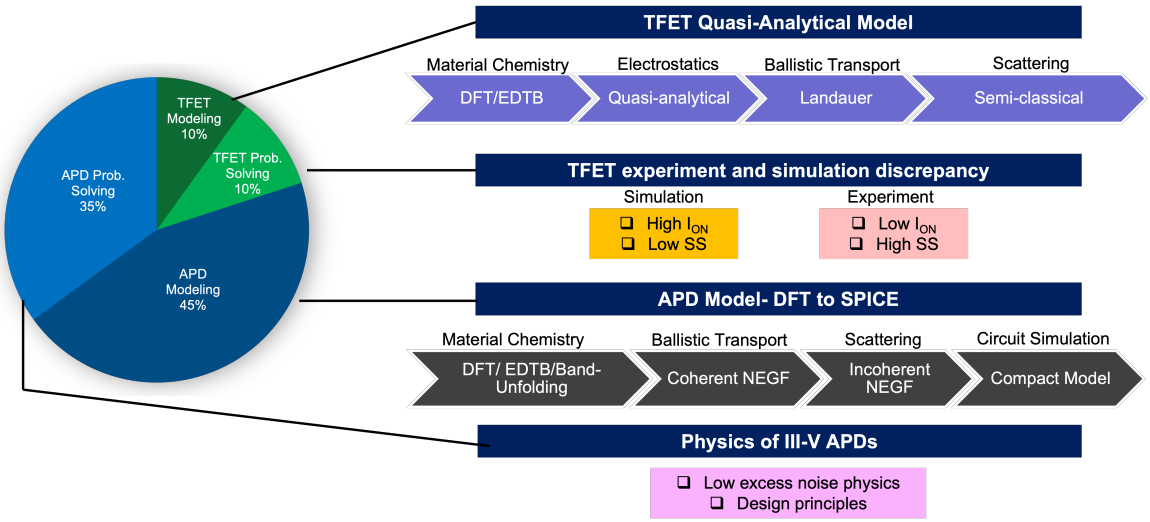

Figure 1.6: A summary of the key topics covered in this dissertation. These topics can be classified into two categories- tool building and problem solving.

Chapter 2

Modeling tunnel field effect transistors - from interface chemistry to non-idealities

When it comes to the down-scaling of semiconductor transistors, Moore's Law has had a spectacular run, one that has unfortunately reached its inevitable slow-down. For CMOS devices, power dissipation has become a major bottleneck for digital applications [38], constrained ultimately by the fundamental Boltzmann limit of $k_B T \ln 10/q \sim 60$ mV/decade that sets the steepness or subthreshold swing for the gate transfer characteristic [39] in conventional MOSFETs. To overcome the theoretical limit of SS for low power applications, novel transistor architectures such as TFETs [2], Mott Transition FETs [40], Graphene Klein Tunnel FETs [41] and Negative capacitance FETs [3], have been proposed and widely investigated in the past decade. Among those novel device architectures, one of the most widely studied are TFETs that operate on the abrupt opening of gate controlled transmission channels through band-to-band (Zener) tunneling across a reverse bias pn junction [16, 17]. Unfortunately, none of the reported TFETs in the literature, to the best of our knowledge, approach the dual needs of high current for fast speed and a low subthreshold swing over

several decades from ON to OFF in order to get low voltage operation and thereby low dynamic power dissipation. A number of TFET designs across a wide variety of device structures and materials have been investigated theoretically [16]. In this chapter, we develop a chemistry-based TFET model that captures material chemistry, accurate tunneling physics and effects of non-idealities. This model is then used to understand the simulation and experimental discrepancy.

The usual approach to studying the detailed physics of these devices is atomistic quantum transport based on Non-equilibrium Green's Functions (NEGF) [42–45]. Atomistic quantum transport models are capable of revealing physical insights but are challenging in many aspects. For instance, many quantum transport studies use empirical models such as $k \cdot p$ and tight-binding models that are numerically-efficient but are fitted to bulk properties and have questionable transferability at surfaces and interfaces. Additionally, the intrinsic numerical complexity of atomistic quantum transport modeling causes them to become excessively time-consuming for systems with realistic device sizes. As a result, most studies using quantum transport modeling are limited to ballistic transport of small-sized devices which predicted much lower off-current and SS compared to experimental results [16]. Quantum transport calculations with scatterings are even more computationally expensive [46], and are thus more size constrained. This high computational burden makes it difficult to consider processes such as trap and defect-assisted tunneling [47], electron-phonon scattering, [48] and electron-electron interactions such as Auger generation [49]. Such processes have zeroth order effects on device performance, for instance by introducing leakage currents that raise the current floor and limit the SS of heterostructure TFET devices. For a fast intuitive way to estimate the TFET device physics including crucial high-order processes, a proper physics based compact model is needed. Accurate quasi-analytical modeling is also needed to bridge numerical modeling and experimental data with circuit-level studies and simulations.

The central component of any TFET structure is band to band tunneling (BTBT) across a new channel that opens at the source end of a *p*-*i*-*n* junction. To model BTBT current,

many existing analytical TFET models [50–53] use Kane's approach [54], which uses WKB approximation to estimate the tunneling current through triangular barriers within a simple effective mass approximation. However, these models typically do not explicitly include the Fermi tails but incorporate them through fitting functions. Moreover, Kane's model requires extra model parameters to compensate for errors caused by the over-simplified bands and 1D electrostatics. There is thus a pressing need for a chemistry-based analytical model based on a proper tunneling equation that accounts for multiple transverse modes, avoids 'fudge factors' and is rooted in chemical modeling and realistic electrostatics, used thereafter to calculate temperature-dependent BTBT current across complicated junctions [55]. To understand the discrepancies between theoretical ballistic current and experimental current, higher order effects such as defects have been included (albeit sparingly) in previous TFET analytical models. These include Trap Assisted Tunneling (TAT) studies by Sajjad *et al.* [47], and Auger generation in perpendicular TFETs by Teherani *et al.* [56].

In this chapter, we present a physics-based analytical model for planar TFETs. This analytical model makes use of the potential obtained by solving the pseudo-2D Poisson's equation. A simplified two-band model is used to describe the electronic properties of the channel, source and drain materials. The material parameters are extracted from tight binding band structures that have been calibrated with first principles band structures and wave functions. Using the two-band model and the approximated potential, ballistic band to band tunneling is calculated using the modified Simmons equation. On top of the ballistic model, we introduce the impact of trap and phonon assisted tunneling, and Auger effect. Furthermore, we identify the source carrier concentration and source valence band transport effective mass as the two main determinants of Auger generation. The effects of the aforementioned factors on OFF current, ON current and subthreshold swing are also discussed. This chapter has been reproduced from Ref. [57] co-authored with Y. Tan, D. S. Truesdell, B. H. Calhoun and A. W. Ghosh, and Ref. [58] co-authored with D. S. Truesdell, Y. Tan, B. H. Calhoun and A. W. Ghosh.

2.1 A quasi-analytical Model

The geometry of a n-type double-gated TFET is shown in Fig. 2.1 (a). The source, channel and drain regions are p^+, i and n^+ doped, respectively. The doping concentrations are N_S for source, N_{ch} for channel, and N_D for drain. The channel region is rectangular with a width of t_{ch} and channel length of L_{ch}. The gate oxide has a thickness of t_{ox}. The dielectric constants for source, drain, channel and gate oxides are ε_S, ε_D, ε_{ch} and ε_{ox}. In this work, we consider both homojunction and heterojunction TFETs - the former targeting a pristine interface for low OFF current while the latter allowing a thin tunnel barrier across a staggered gap (Type II) junction for large ON current. The homojunction TFET has an $In_{0.53}Ga_{0.47}As$ channel. The heterojunction TFET has a GaSb source and an InAs channel/drain.

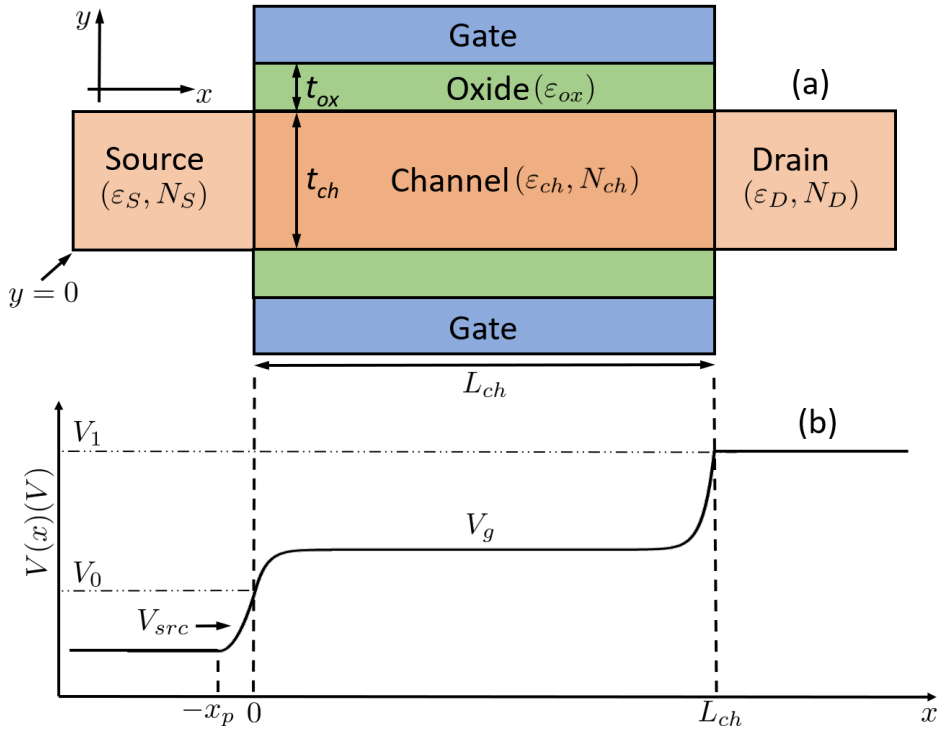

Figure 2.1: Planar TFET structure considered in this work(a) and potential in the device(b). In the analytical model, 2D Poisson's equation in the channel region is solved and approximated by analytical equations.

2.1.1 TFET Surface Potential

For our device structure the surface potential, shown in Fig. 2.1 (b), is obtained by solving Poisson's equation with appropriate boundary conditions. Since the tunneling current in a TFET is dominated by the source/channel junction, the potential at that junction is critical. While we know the dopings in the different regions, the potential in the channel and thus the voltage barrier across the source-channel region is complicated, and needs to be computed including the gate and drain capacitances. We will now simplify the solution to 2D Poisson's equation with suitable boundary conditions to extract the 1D channel potential at the gate-channel interfaces, culminating in Eq. 2.8.

The surface potential V_0 at the boundary of the heavily doped source region can be calculated by solving 1D Poisson's equation with a uniform doping concentration N_S and dielectric constant ϵ_S, assuming a homogeneous potential along the vertical y direction. The solution to this equation can be written as $V_{src}(x) = qN_S(x + x_p)^2/2\epsilon_S$ where x_p is the depletion width at the junction along the x-direction as shown in Fig. 2.1 (b). This gives us one boundary condition at the source/channel interface, the other at the drain/channel interface is obtained from the offset in the local quasi-Fermi levels between source and drain

$$V_0 = V_{src}(0) = \frac{qN_S}{2\epsilon_S} x_p^2 \tag{2.1}$$

$$V_1 = \frac{kT}{q} \ln\left(\frac{N_S N_D}{n_{iS} n_{iD}}\right) + V_{DS} + \Delta E, \tag{2.2}$$

where N_S, N_D represent the doping concentrations in the source and drain regions, and n_{iS} and n_{iD} are the intrinsic carrier concentrations of the source and drain materials, while ΔE is the band offset between source and drain materials. V_0 is the solution of 1D Poisson's equation at the source and will need to be estimated shortly. For V_1, we assume the potential is constant in the drain region. Compared with the rigorous solution in the drain, this approximation leads to a negligible difference in the current because the band-to-band tunneling occurs at the source-channel interface.

For the potential in the channel, a pseudo-2D Poisson's equation is solved in the rectangular channel region shown in Fig. 2.1 (a). It is assumed that mobile charge carriers do not affect the electrostatics of the device [59] compared to the fixed charges (dopants). The 2D Poisson's equation in the channel region can be written as

$$\frac{\partial^2 V(x,y)}{\partial x^2} + \frac{\partial^2 V(x,y)}{\partial y^2} = \frac{qN_{ch}}{\epsilon_{ch}} \tag{2.3}$$

where $V(x,y)$ is the electrostatic potential of the region, N_{ch} is the effective doping and ϵ_{ch} is the dielectric constant of the material. Considering a parabolic variation of the potential in the y-direction ($y = 0$ being the bottom channel-gate interface and $y = t_{ch}$ the top) the 2D potential can be approximated by the second order polynomial in y [60],

$$V(x,y) = a_0(x) + a_1(x)y + a_2(x)y^2 \tag{2.4}$$

We first use the continuity of potential and displacement field in the y direction to convert the 2D Poisson equation into an equivalent 1D equation for the channel potential $V_{ch}(x)$. The boundary conditions are set by the gate potential V_g and field at the lower $y = 0$ and upper $y = t_{ch}$ gate-channel interfaces

$$\begin{aligned} V(x,0) &= V_{ch}(x) \\ V(x,t_{ch}) &= V_{ch}(x) \\ E_y(x,0) &= -\frac{\eta}{t_{ch}}(V_G - V_{ch}(x)) \\ E_y(x,t_{ch}) &= -\frac{\eta}{t_{ch}}(V_{ch}(x) - V_G) \end{aligned} \tag{2.5}$$

The parameter $\eta = C_{ox}/C_{ch}$ represents the ratio between the gate capacitance C_{ox} and the channel capacitance $C_{ch} = \epsilon_{ch}/t_{ch}$. The gate potential V_g is referenced with respect to the flatband condition $V_G = V_{GS} - V_{fb}$, where $V_{fb} = \phi_m + \chi + E_g/2$, ϕ_m, χ and E_g representing the gate metal work function, electron affinity and the bandgap of the channel material,

respectively. By applying these four vertical boundary conditions at the gate/channel interfaces, we can find the coefficients in equation (2.4). Since the electric fields are largest at the channel-gate interfaces $y = 0,\ t_{ch}$, the tunneling electrons are preferably attracted to those interfaces. We thereafter focus on the channel potential $V_{ch}(x) = V(x, y = 0)$. Substituting in Eq. 2.3, we find that $V_{ch}(x)$ satisfies a 1D Poisson equation

$$V_{ch}^{''}(x) - k^2 V_{ch}(x) = -k^2 V_g \tag{2.6}$$

with

$$\begin{aligned} k &= \sqrt{2\eta/t_{ch}^2} \\ k^2 V_g &= k^2 V_G - \frac{qN_{ch}}{\epsilon_{ch}} \end{aligned} \tag{2.7}$$

Here, the characteristic length in the channel region is given by $1/k$ and V_g represents the solution of 1D approximation of Poisson's equation using the long-channel approximation. The solution V_{ch} can be written as [61]

$$V_{ch}(x) = be^{kx} + ce^{-kx} + V_g \tag{2.8}$$

with boundary conditions $V_{0,1}$ at the two ends (Eq. 2.2), with V_0 still unknown.

$$\begin{aligned} b &= \frac{1}{2\sinh(kL_{ch})}\left(-V_0 e^{-kL_{ch}} - V_g(1 - e^{-kL_{ch}}) + V_1\right) \\ c &= \frac{1}{2\sinh(kL_{ch})}\left(V_0 e^{kL_{ch}} + V_g d(1 - e^{kL_{ch}}) - V_1\right) \end{aligned} \tag{2.9}$$

The related unknown variable is x_p, the width of the depletion region width in the source. x_p can be obtained using the continuity of the displacement field at source/channel interface

$$\epsilon_S \frac{dV_{src}}{dx} = \epsilon_{ch} \frac{dV_{ch}}{dx}. \tag{2.10}$$

By substituting $V_{ch}(x)$ from Eq. 2.8 and the form of $V_{src}(x)$ discussed before Eq. 2.2, we get a nonlinear equation in V_0 or equivalently a quadratic equation in x_p with a positive solution

$$x_p = \frac{-1 + \sqrt{1 + 2P/qN_S\epsilon_S}}{Q} \tag{2.11}$$

where,

$$P = \epsilon_{ch}^2 k^2 \coth(kL_{ch}) \left[V_g \coth(kL_{ch}) - \frac{V_g - V_1}{\sinh(kL_{ch})} \right]$$

$$Q = \frac{\epsilon_{ch} k \coth(kL_{ch})}{\epsilon_S} \tag{2.12}$$

Our key equation is thus the channel potential (Eq. 2.8), with coefficients b and c from Eq. 2.9. The long channel potential V_g is related to the applied gate bias and the doping through Eq. 2.7 and the definition of V_G after Eq. 2.5. The channel potentials at the two ends V_0 and V_1 are obtained from Eqs. 2.2, 2.11 and 2.12.

The potential model presented in this section can be applied to both homojunction TFET and heterojunction TFETs. For a long channel, $kL_{ch} \gg 1$, we get $x_p \approx \sqrt{2\epsilon_S V_g/qN_S}$, as we expect for depletion widths across conventional PN junctions with V_g replacing the built-in potential. The channel potential $V_{ch}(x) \approx V_0 + (V_1 - V_0) \exp k(x - L_{ch})$ with $V_0 \approx V_g$, meaning that the potential stays pinned to V_g for much of the channel length and switches to V_1 only within a distance $\sim 1/k$ of the drain end. For short channels $kL_{ch} \ll 1$, $x_P \approx \sqrt{2\epsilon_S V_1/qN_S}$ and $V_{ch}(x) \approx V_0 + (V_1 - V_0)x/L_{ch}$. Also, note that the equations can be generalized to a gate all around cylindrical nanowire TFET by using polar coordinates. The decay length $\lambda = 1/k$ in Eq. 2.7 gets replaced by the logarithmic capacitance of the cylindrical wire of diameter d_{ch}, namely, $\lambda = \sqrt{[2\epsilon_{ch} d_{ch}^2 \ln(1 + 2t_{ox}/d_{ch}) + \epsilon_{ox} d_{ch}^2]/16\epsilon_{ox}}$.

Fig. 2.2 shows the band diagrams of a homojunction and a heterojunction TFET under different gate bias conditions. We assume both TFETs have 100 nm channel length. Material parameters are extracted from previous Non-Equilibrium Green's function calculations. The

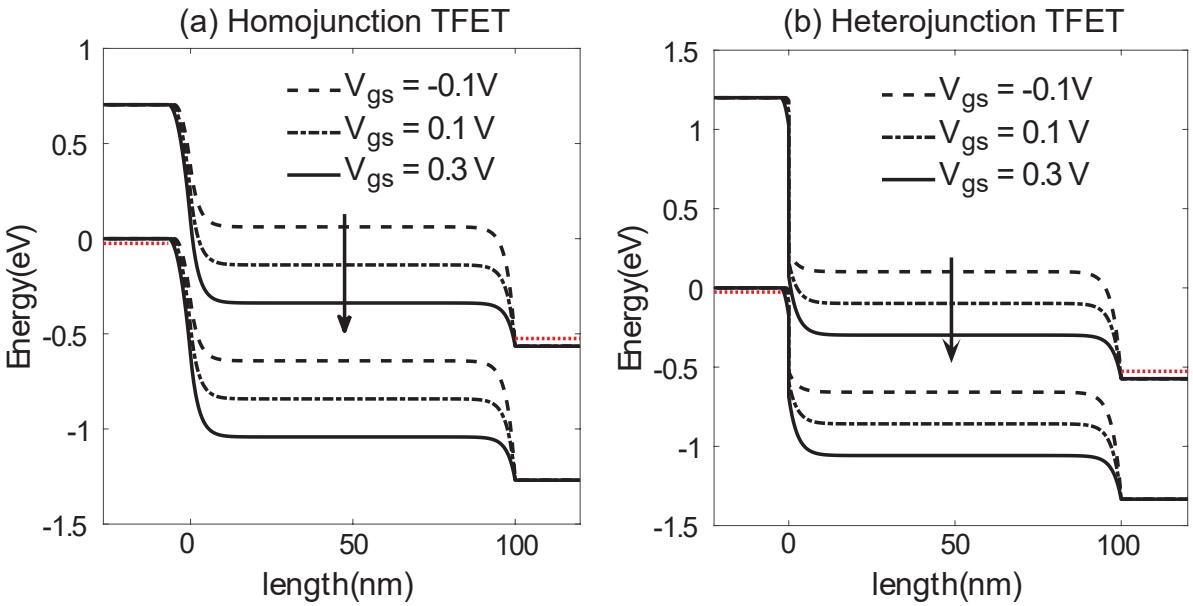

Figure 2.2: Band diagram of the ON and OFF states in (a) homojunction and (b) heterojunction TFETs. The source region has $x < 0$ nm, drain region has $x > 100$ nm, and the channel region has 0 nm $\leq x \leq$ 100 nm.

homojunction TFET has InGaAs as channel with a band gap of 0.74 eV [43], and the heterojunction TFET has GaSb as source with a 1.2 eV band gap [44] and InAs as channel with a band gap of 0.76 eV [45].

2.1.2 Two-band Model for the junction

To model band to band tunneling correctly, a single band effective mass model is clearly insufficient. We use a simplified two-band $k \cdot p$ model which can generate more accurate real and complex bands for the direct band gap III-V group materials considered in this work.

In this two-band model, the bands considered represent the conduction band and light hole band at the Γ point, critical for the band to band tunneling process. For III-V materials, the complex band connecting conduction band (Γ) and light hole band has the smallest imaginary wave vector, as shown in Fig. 2.3 (b). Carriers tunneling from the light hole band to the lowest conduction band across the junction clearly dominate the current. The heavy hole and split-off bands connected to higher conduction bands have complex bands with

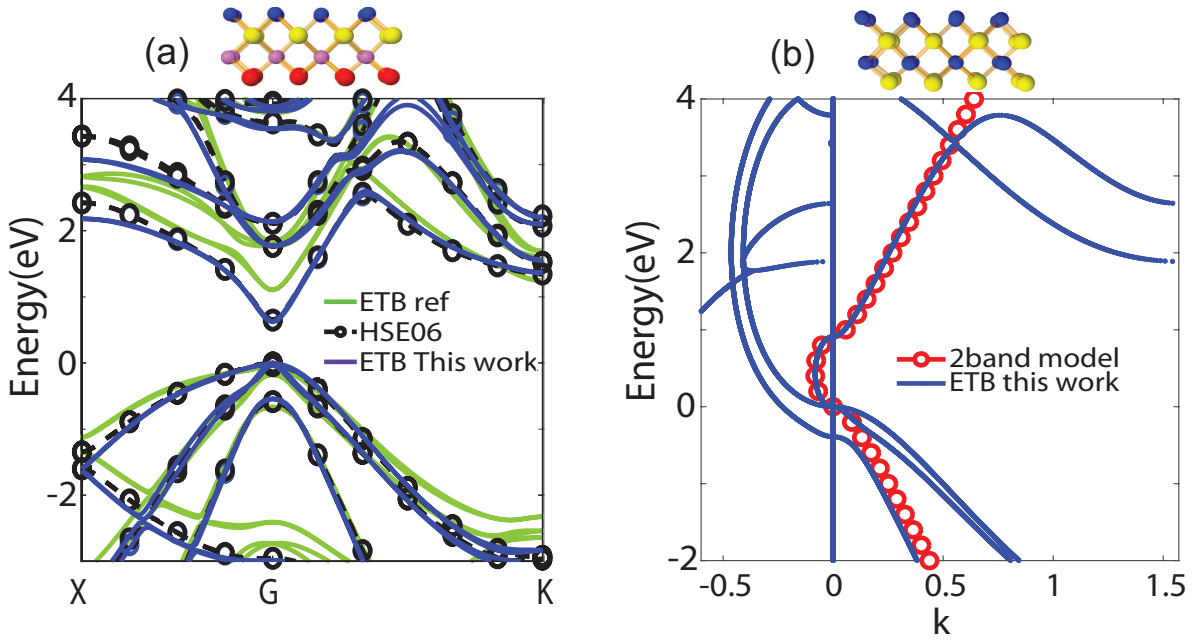

Figure 2.3: (a) Comparison of band structure of InAsGaSb superlattice. The tight binding calculation (blue lines) in this work agree well with the first principle band structure with hybrid functional calculation (HSE06) in black dotted lines. While the tight binding calculation using previous parameters(green line) show large discrepancies compared with first principle calculations. (b) Real and complex band structure of strained InAs two-band model vs tight binding. By adjusting the parameters of the two-band model, both the real bands and complex bands from the two-band model agree well with the tight binding model.

much larger imaginary wave vectors. These bands decay much faster in real space and can thus be ignored. The two-band model in this work can be written as

$$H(\mathbf{k}) = \begin{bmatrix} E_c(\mathbf{k}_\parallel) & Ak_x \\ Ak_x & E_v(\mathbf{k}_\parallel) \end{bmatrix}, \tag{2.13}$$

The equation generates two parabolic bands with a common tunneling effective mass m^* that is obtained by setting $A^2/(E_c - E_v) = \hbar^2/2m^*$. While this model ignores the separate masses for conduction and valence band within a single material, we assume band-to-band tunneling occurs from a single light hole valence band in the source to a single conduction band in the channel, which means a single mass separately set in each material suffices to capture the dominant TFET current. The E_c and E_v dispersions are set by the potential we

just worked out, shown in Fig. 2.1b.

$$\begin{aligned} E_c(\mathbf{k}_\parallel, x) &= E_c - qV(x) + \hbar^2\mathbf{k}_\parallel^2/2m^* \\ E_v(\mathbf{k}_\parallel, x) &= E_v - qV(x) - \hbar^2\mathbf{k}_\parallel^2/2m^* \\ V(x) &= \begin{cases} V_{src}(x), & x \leq x_0 \\ V_{ch}(x), & x_0 < x < L_{ch} \\ V_1, & x > x_0 \end{cases} \end{aligned} \tag{2.14}$$

The required material parameters for the two-band model are the electron tunneling effective mass m^* and band edges E_c and E_v. The bold $\mathbf{k}$'s correspond to vectors, and italic k's correspond to scalars.

With the two-band model, for a given energy E, the k_x can be calculated analytically as

$$k_x\left(\mathbf{k}_\parallel, x\right) = \pm\frac{\sqrt{(E-\bar{E})^2 - \Delta^2}}{A} \tag{2.15}$$

with $\bar{E} = \left(E_c(\mathbf{k}_\parallel, x) + E_v(\mathbf{k}_\parallel, x)\right)/2$ and $\Delta = \left(E_c(\mathbf{k}_\parallel, x) - E_v(\mathbf{k}_\parallel, x)\right)/2$. This expression works for any given energy E, and can generate both real and complex k. For a given E, all possible k_xs are calculated within the first brillouin zone of $k_\parallel$ and their transmissions will eventually be summed up to get the total transmission over perpendicular states.

2.1.3 Accurate parametrization and band-unfolding at the junction

In this work, we extract the material parameters from tight binding calculations. It should be emphasized that the accuracy of tight binding has a significant impact on the results, since the tunneling current depends exponentially on the tunneling effective mass, which is ultimately a hybrid between the bulk light hole and conduction band effective masses on its sides. We employ a tight binding model that has been carefully calibrated not only with

band structure but also wavefunctions based on experiments as well as high accuracy first principles calculations [62, 63]. We fit our tight binding parameters with Density Functional Theory (DFT) within the HSE06 hybrid functional [64] approximation that is known to generate accurate bandstructure of semiconductors matching experiments. Previously, we demonstrated that one way to make tight-binding transferrable was to employ non-orthogonal basis sets to calibrate bond overlaps in Extended Hückel Theory [65, 66]. Here we employ an alternate way to endow orthogonal tight binding with transferrability between bulk geometries and systems with strain and interfaces, by additionally matching the radial wavefunctions with DFT.

Fig. 2.3 (a) shows excellent agreement between our tight binding calculations and hybrid functional(HSE06) results for systems with interfaces, in this case, an ultra small InAs/GaSb superlattice. In comparison, tight binding calculations using previous parameters [67] show an obvious discrepancy, because they are extracted by fitting to bulk InAs and GaSb bands without considering physical insights from wave functions, and without calibration to interfaces. The comparison suggests that the tight binding we developed has much better transferability for III-V superlattices and alloys. Fig. 2.3 (b) shows how well our simplified two-band model matches the conduction band, light hole band and the complex bands from tight binding calculations. For comparison studies with past NEGF calculations, we retain parameters m^* and E_g from previous work to keep the benchmarking standards the same.

For the alloy $In_{0.5}Ga_{0.5}As$ in the Homo-junction TFET, we studied two different cases - namely a random alloy and a digital alloy. The tight binding band structures and two-band model band structures for the two geometries are shown in Fig. 2.4. It can be seen from (a) and (b) that the tight binding band structures for a random alloy and a digital alloy share little resemblance with each other, because of their vastly different unit cell sizes. Here, we only consider a specific instance of a random alloy whereas, for a practical device, the TB band structure must be obtained by doing a Monte Carlo averaging to account for the distribution of the defects. In order to make a meaningful comparison between random and

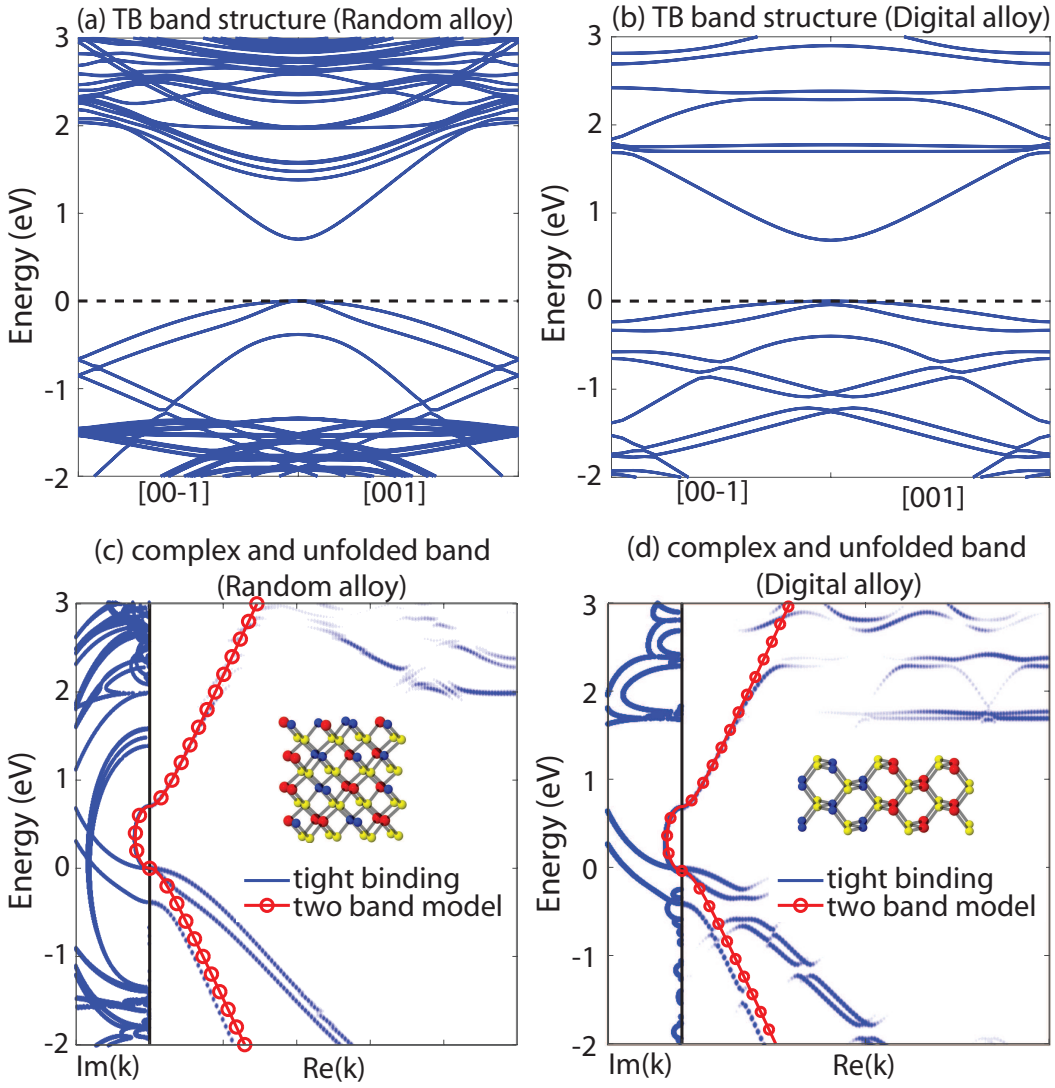

Figure 2.4: Real and complex band structure of strained $In_{0.5}Ga_{0.5}As$ alloys with two-band model and tight binding model. (a) original tight binding band structure and (c) unfolded real and complex band structure of a Random alloy. (b) original tight binding band structure and (d) unfolded real and complex band structure of a Digital alloy. By adjusting the parameters of the two-band model, both the real bands and complex bands from the two-band model agree well with the tight binding model.

digital alloy in (c) and (d), we used the technique of band unfolding [68–70] to simplify the real bands. In this technique, the wavefunctions of the supercell in the alloy Brillouin zone are fourier decomposed and the Fourier coefficients are used to generate the wavefunction probabilities at the corresponding k and energy points. A description of band-unfolding process is provided in Chapter 3. We see how some of the high energy conduction bands and

low energy valence bands are unfolded back to the Brillouin zone of the primitive zincblende unit cell. It can be seen that the unfolded direct conduction band and valence bands in digital and random alloy have similar band profiles. Notably, the digital alloy creates broken bands with minigaps due to coherent destructive interference, suggesting a strong interaction among the bands. To visualize the tunneling properties of InGaAs, the complex bands in a random and digital alloy are also shown in (c) and (d). More complex branches can be seen in the random alloy than the digital alloy, corresponding to the fact that there are more bands for the former. However, the two band model shows a much simpler band profile with only the most important complex band captured. By adjusting the parameters of the two-band model, we can match the direct conduction band, dominant light hole band and the complex branch connecting the two band-edges for both digital and random alloys. The agreement implies the two band model is a good approximation to model band to band tunneling even for III-V alloy materials.

2.1.4 BTBT Current Model

With the k vectors set by the two-band model, the BTBT current is ready to be calculated. The Simmons model has been very successful in the chemistry literature in quantitatively describing the tunneling current through a thin film. This model approximates the WKB tunneling current by Taylor expanding the barrier profile in the exponent around a rectangular shape and then summing over a continuum of transverse modes [55].

We modified the Simmons equation to account for a non-rectangular shape in the current integration over perpendicular k, and also retained a finite temperature dependence that sets the subthreshold swing for switching. The modified Simmons equation arises from Landauer equation [71]

$$I = \frac{q}{h} \int_{E_{cd}}^{E_{vs}} T\left(E\right) \left[f_S(E) - f_D(E)\right] dE \tag{2.16}$$

Here, $T(E)$ represents the sum of the transmission probability over all the transverse states at

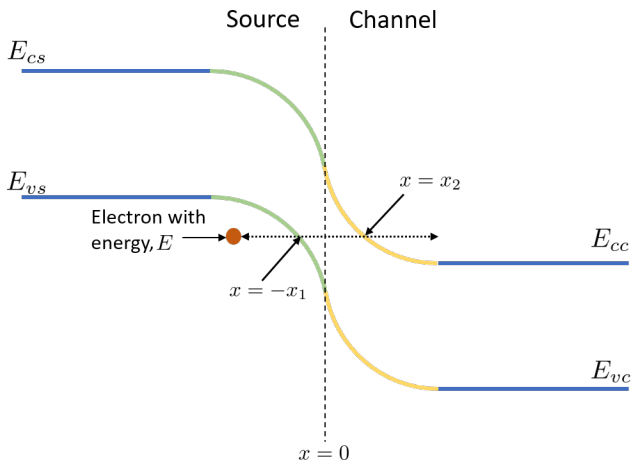

Figure 2.5: Band diagram of the tunneling junction. The distance between $-x_1$ and x_2 is the tunneling width for the electron shown.

a particular energy E and $T\left(k_{\|}, E\right)$ gives the transmission probability for a single transverse state. Here, the $k_{\|}$ is calculated by equation (2.15) in section 2.1.2, and the band to band , tunneling probability $T\left(k_{\|}, E\right)$ is estimated by WKB approximation

$$T\left(k_{||}, E\right) = (1 - R)\left|\exp\left(-2i\int k_x\left(x\right)dx\right)\right| \tag{2.17}$$

where,

$$\int k_x\left(x\right)dx = \int_{-x_1}^{0} k_x\left(x\right)dx + \int_{0}^{x_2} k_x\left(x\right)dx$$

In this equation, the $T\left(k_{||}, E\right) = 1$ for any real k_x, and $T\left(k_{||}, E\right) < 1$ for complex κ_x corresponding to the process of tunneling through a barrier. The boundary conditions for determining x_1 and x_2 are $E - E_v(-x_1) = 0$ and $E - E_c(x_2) = 0$, respectively. The band diagram of the tunneling region is shown in Fig. 2.5. To take account of added reflections at the interface not accounted for in the WKB approximation (arising from the kinetic energy pre-factor in the semiclassical approximation), we introduce a phenomenological correction factor $1 - R < 1$ with R the reflection coefficient at the interface. The reflection R is manually adjusted to mimic the effect of interface reflection and achieve better agreement with the reference I-V. The integral is separated into the source and channel regions in order to obtain

an analytic expression.

To obtain a closed-form solution of the source region integral, the source potential is at first approximated using linearization applied at the point $x = -x_1/2$

$$V_{src,li}(x) = \frac{qN_S}{\epsilon_S}\left[\frac{1}{2}\left(x_p - \frac{x_1}{2}\right)^2 + \left(x_p - \frac{x_1}{2}\right)\left(x + \frac{x_1}{2}\right)\right] \tag{2.18}$$

$$V_{src,li}(x) = V_{src}\left(-\frac{x_1}{2}\right) + V'_{src}\left(-\frac{x_1}{2}\right)\left(x + \frac{x_1}{2}\right) \tag{2.19}$$

Using equation (2.19) the result of the indefinite source integral can be written as

$$\begin{aligned}\int k_x(x)\,dx = \frac{-\epsilon_S}{2AqN_S\left(x_p - \frac{x_1}{2}\right)}\Big[&\tilde{E}\sqrt{\tilde{E}^2 - \Delta^2} \\ &-\Delta^2 \log\left(\tilde{E} + \sqrt{\tilde{E}^2 - \Delta^2}\right)\Big]\end{aligned} \tag{2.20}$$

where,

$$\tilde{E} = E - \frac{E_c + E_v}{2} + V_{src,li}(x)$$

The channel region integral is approximated by assuming that the be^{kx} term in the channel potential is negligible near the source/channel junction where the potential profile is starting to saturate (Fig. 2.2). Thus, the solution of the indefinite integral can be written as

$$\begin{aligned}\int k_x(x)\,dx = -\frac{1}{kA}\Big[&-\sqrt{\tilde{E}^2 - \Delta^2} + \sqrt{C_1^2 - \Delta^2}kx \\ &-C_1 \log\left(\tilde{E} + \sqrt{\tilde{E}^2 - \Delta^2}\right) \\ &+\sqrt{C_1^2 - \Delta^2}\log\left(C_1^2 - \Delta^2 + C_1 ce^{-kx}\right. \\ &\left.+\sqrt{C_1^2 - \Delta^2}\sqrt{\tilde{E}^2 - \Delta^2}\right)\Big]\end{aligned} \tag{2.21}$$

where,

$$C_1 = E - \frac{E_c + E_v}{2} + V_d \tag{2.22}$$

and

$$\tilde{E} = E - \frac{E_c + E_v}{2} + V_{ch}(x)$$

The total transmission $T(E)$ is calculated by integrating the tunneling probability over all k-states parallel to the interface in the first Brillouin zone over a 2D circular phase space

$$T(E) = 2\pi \int_0^{k_{||max}} k_{||} T\left(k_{||}, E\right) dk_{||} \tag{2.23}$$

A fitting equation is used to calculate the value of $k_{||max}$ since for large values of $k_{||max}$ the two-band model deviates from the actual bandstructure. Here, $k_{||max} = |c_1 E + c_2|$. The constants $c_1 = 5.6 \times 10^8\ m^{-1}eV^{-1}$, $c_2 = 5.1 \times 10^8\ m^{-1}$ and energy E has units of eV. The integral in equation (2.23) does not have an analytical solution. Thus, an approximate solution for $T(E)$ must be obtained by

$$T(E) = 2\pi \sum_{n=0}^{3} \int_{k_{||n}}^{k_{||n+1}} f\left(k_{||}\right) dk_{||} \tag{2.24}$$

where $k_{||n} = nk_{||max}/4$ and $f\left(k_{||}\right) = k_{||} T\left(k_{||}, E\right)$. The integral in equation (2.24) can be computed using Simpson's 3/8 rule.

$$\int_{k_{||n}}^{k_{||n+1}} f\left(k_{||}\right) dk_{||} = \frac{k_{||n+1} - k_{||n}}{8} \left[3f\left(\frac{2k_{||n} + k_{||n+1}}{3}\right) + 3f\left(\frac{k_{||n} + 2k_{||n+1}}{3}\right) + f\left(k_{||n}\right) + f\left(k_{||n+1}\right)\right]$$

2.1.5 Trap Assisted Tunneling

To account for critical high order effects near the source-channel junction, a trap assisted tunneling (TAT) process is included in our model. Previous phonon scattering (without traps)

models have shown higher on off currents but not a significant change in the subthrehsold swing [72]. Due to the existence of defects near a material interface, intermediate energy levels known as trap states form a quasi-continuous density of states in the band gap. Electrons can jump from the source valence band into the channel conduction band through these trap states by exchanging energy with optical phonons. This undesired flow of electrons creates a leakage current which has many adverse effects on TFET performance, namely higher off-current and higher SS. The trap current per unit width, I_{TAT}, is calculated using a Fowler-Nordheim type tunneling through a tilted barrier around the trap. The Shockley-Reed-Hall generation rate is given by [47]

$$G = \int \frac{\sigma_n \sigma_p v_{th} \left(n_i^2 - np\right)}{\sigma_n \frac{n+n_1}{1+\Gamma_p} + \sigma_p \frac{p+p_1}{1+\Gamma_n}} D_{it} dE \tag{2.25}$$

where Γ describes the electric field-enhancement of the trap assisted tunneling and thermionic emission processes, n_i is the intrinsic charge concentration, n, p are the electron and hole densities in the conduction and valence band, D_{it} is the interface trap density, assumed primarily mid-gap ($E_T \approx E_i$) and σs are the capture cross-sections. The quantity Γ, typically much larger than unity in a strong interfacial field, can be estimated by looking at the fractional change in emission coefficient $e_n = e_{n0} \exp\left(E_C - E\right)/k_B T$ in the presence of a Boltzmann (Frenkel-Poole) jump over a tilted barrier and a WKB tunneling (Fowler-Nordheim) through the tilted barrier.

$$\begin{aligned} \Gamma &= \int dE T(E) \frac{d(e_n/e_{n0})}{dE} \\ &= \frac{\Delta E_n}{k_B T} \int_0^1 exp\left[\frac{\Delta E_n}{k_B T} u - K_n u^{3/2}\right] du \\ K_n &= \frac{4}{3} \frac{\sqrt{2m^* \Delta E_n^3}}{q\hbar F} \end{aligned} \tag{2.26}$$

where $\Delta E_n = E_C - E_T$, $u = (E_C - E)/\Delta E_n$. Assuming the trap density of states is localized at mid-gap, i.e., a delta function of weight $D_{it}(E_i)$, we then get the trap assisted current per

unit width, under drain bias, to be [73]

$$I_{TAT} = \frac{q}{2} v_{rcmb} n_i \Gamma d \left[1 - e^{-qV_{DS}/k_B T}\right], \tag{2.27}$$

where d is the width of the trap active region along the transport direction, the recombination velocity $v_{rcmb} = \sigma v_{th} N_t$ and $\int D_{it} dE = \int N_t \delta(E - E_i) dE = N_t$. Here, N_t represents the trap density per unit area at the midgap energy, while D_{it} converts this into a density of states with a delta function profile at the trap energy.

This TAT model can be incorporated in our analytical ballistic TFET model since it requires only the electric field at the junction. The highest electric field F can be evaluated from the potential model described in section 2.1.1. In this work, we assume $\sigma = 5 \times 10^{-17}$ m^2, $d = 1$ nm and $\Delta E_n = E_g/4$, because under the action of an electric field the trap level slips past the source conduction band edge and has a smaller barrier to the bandedge of the intrinsic region [73].

2.1.6 Auger Current Model

In our BTBT current model, we now add current due to Auger generation. Auger involves charge scattering through Coulomb interaction. In fact, three particles exchange energy and momentum, as a consequence, one of the particles can transit from the valence band to conduction band, as illustrated by Fig. 2.6.

To increase the ON current of TFETs, researchers have proposed near broken gap (almost type III) heterojunctions, where the conduction band of the channel lies just above the valence band of the source. The resulting narrow triangular barriers increase the ON current upon the onset of band-to-band tunneling. Besides the creation of interfacial traps due to unsatisfied bond valency at the hetero-interface, these triangular barriers tend to increase the wavefunction overlap due to the steepness of the potential (i.e., the built-in electric field), which tends to shoot up the Auger generation (impact ionization) process depicted in Fig. 2.6

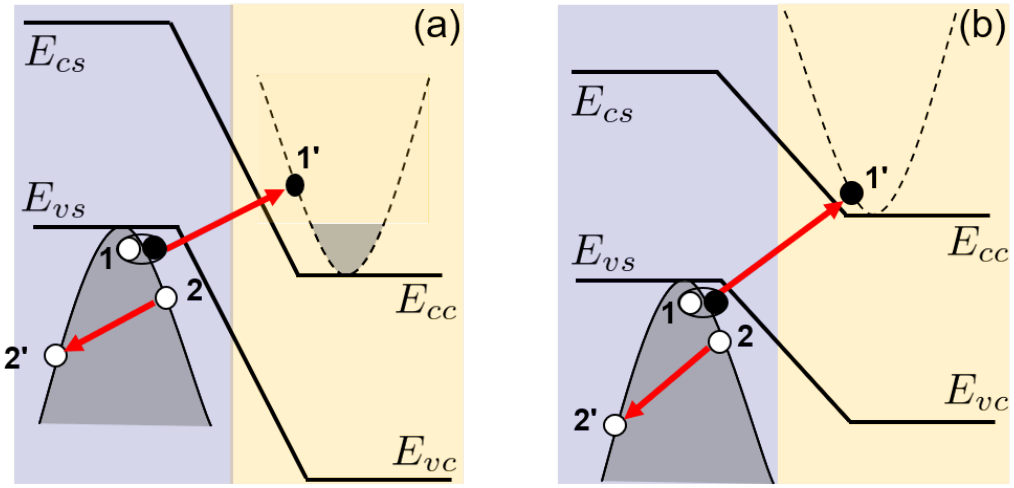

Figure 2.6: Auger generation process of near the source/channel junction in (a) on and (b) off states in a TFET.

(b) and increases the off current floor in the TFET. Through Auger generation process, extra holes in the source region and extra electrons in the channel region are generated. While the impact on the ON current (Fig. 2.6 (a)) is minor compared to the BTBT term, any change in an otherwise low off current hurts the ON-OFF ratio and the subthreshold swing [56]. The rate of Auger generation can be estimated by employing Fermi's Golden Rule

$$G = \frac{1}{A}\frac{2\pi}{\hbar}\sum_{1,1',2,2'} P\left(1,1',2,2'\right)|M|^2\delta\left(E_1 - E_{1'} + E_2 - E_{2'}\right). \tag{2.28}$$

As depicted in Fig. 2.6, 1 and 2 correspond to initial states of holes, while 1' and 2' correspond to final states in Auger generation process. The P is the occupancy of the initial and final states, given by

$$\begin{aligned} P_{HCHH}(1,1',2,2') &= \bar{f}_v(E_1)\bar{f}_v(E_2)f_c(E_{1'})f_v(E_{2'}) \\ &- f_v(E_1)f_v(E_2)\bar{f}_c(E_{1'})\bar{f}_v(E_{2'}) \\ &\approx \bar{f}_v(E_1)\bar{f}_v(E_2)f_c(E_{1'}) - \bar{f}_v(E_{2'}) \end{aligned} \tag{2.29}$$

where $\bar{f} = 1 - f$ is the hole occupancy. The subscripts refer to the bands involved, heavy-hole

and conduction band. Within an envelope function approximation for the matrix element M, the Auger generation rate can be written as [56]

$$\begin{aligned} G &= \frac{1}{A}\frac{4\pi}{\hbar}\sum_{1,1',2,2'} -\frac{p}{N_v}\exp\left(-\frac{E_{2'}-E_v}{kT}\right) \\ &\left(\frac{q^2}{2\epsilon A}\frac{\delta_{k_{\perp 1}-k_{\perp 1'}+k_{\perp 2}-k_{\perp 2'}}}{|k_{\perp 1}-k_{\perp 1'}|}\right)^2 \\ &(c_u K)^2\left|\langle\psi_{1'}|\psi_1\rangle\right|^2\delta\left(E_1-E_{1'}+E_2-E_{2'}\right) \end{aligned} \tag{2.30}$$

In this equation, $|\psi_1\rangle$ and $|\psi_{1'}\rangle$ correspond to the envelope functions of the initial valence state 1 and final conduction $1'$ as depicted in Fig. 2.6. The $c_u K$ involve the Bloch parts of the wavefunctions and are evaluated using a 8 band $k\cdot p$ model [56]. For III-V semiconductors, c_u is approximately $\sqrt{2\times 10^{-17}}\ cm$ and $K=|k_1-k_{1'}|$.

The summation $\sum_{1,1',2,2'}$ has to be treated differently according to the device structure. For bulk materials this summation sums over a twelve dimensional k-space since it involves four particles. For quantum well TFETs studied earlier [56], the states 1,1′, 2 and 2′ are quasi continuous in 2-dimensions, whereupon the $\sum_{1,1',2,2'}$ becomes

$$\int d^2k_{\perp 1}d^2k_{\perp 2}d^2k_{\perp 1'}d^2k_{\perp 2'}\sum_{k_{x_1},k_{x_{1'}},k_{x_2},k_{x_{2'}}} \tag{2.31}$$

Here k_{x_1},$k_{x_{1'}}$,k_{x_2},$k_{x_{2'}}$ are discrete states due to quantum confinement in the wells in the transport direction. For planar TFETs considered in this work, the states 1,1′, 2 and 2′ are quasi continuous in all 3-dimensions. The $\sum_{1,1',2,2'}$ becomes

$$\int d^2k_{\perp 1}d^2k_{\perp 2}d^2k_{\perp 1'}d^2k_{\perp 2'}\int dk_{x_1}dk_{x_2}dk_{x_{1'}}dk_{x_{2'}} \tag{2.32}$$

Since the device is considered as an infinite plane in y and z directions, $k_\perp$s are summed separately. While the x direction is transport direction, the states k_{x_1},$k_{x_{1'}}$,k_{x_2},$k_{x_{2'}}$ are quasi-continuous states which have exponential tails near the source-channel junction. Therefore,

the wave function overlaps $\langle\psi_{1'}|\psi_1\rangle$ depend on position x.

To estimate the wave function overlap, we make use of the band model in 2.1.2. The wave vectors are calculated using equation (2.15) and the right decaying wave function of 1 is $\psi_1 = \exp\left(-\kappa_x(x - x_1)\right)$ and left decaying wave function of 1' $\psi_{1'} = \exp\left(\kappa_x(x - x_1')\right)$. Here we only consider the Auger generation at the junction since the carriers (electrons) generated in the high-field region of a pn junction are swept out by the electric field to the channel. Thereafter the current-density from Auger processes can be estimated as

$$J_{aug} = qG \tag{2.33}$$

The Auger current can be obtained by enforcing momentum and energy conservation involving two holes (Fig. 2.6)

$$\begin{aligned} k_1 + k_2 &= k_1' + k_2' \\ E_1 + E_2 &= E_1' + E_2' \end{aligned} \tag{2.34}$$

We can eliminate k_2 and write E_2' in terms of k_1 and k_1' as independent variables (k_2' being set by E_2'). We can then minimize E_2' with respect to k_1, k_1', which gives us $k_1 = k_2$ and $J_{Aug} \propto \exp\left(-E_{2,min}'/kT\right) = \exp\left(-\frac{2\mu^{-1}+1}{\mu^{-1}+1}\frac{\Delta E}{kT}\right)$ which defines the Auger generation limited sub-threshold, where μ is the mass ratio $\mu = m_c^*/m_v^*$. Here ΔE is the energy separating the lowest conduction band in the channel and the highest valence band in the source. The Auger generation limited sub-threshold ranges from 30 meV/decade when $m_c^* \ll m_v^*$ and 60 meV/decade when $m_c^* \gg m_v^*$.

The Auger exponent is fairly easy to understand. If we make the effective mass m_c^* very small, then we have a highly localized conduction band and $k_1' = 0$, in which case we get $k_1 = k_2 = k_2'/2$ (co-moving electron and hole, or equivalently oppositely moving electrons in the conduction band or holes in the valence band) and $E_{2,min}' = 2\Delta E$. The double jump across the band-gap creates a very low subthreshold swing. On the other hand if $m_v^* \to 0$

then $k_1 = k_2 = 0$ and thus $k_2' = -k_1'$ and $E'_{2,min} = \Delta E$. Since the momentum is equally split, we are limited by the usual Boltzmann limit.

2.1.7 Phonon Assisted Tunneling

Phonon assisted tunneling (PAT) is fairly common in TFETs, including through defect states and traps that create levels in the gap. Electrons can enter these states and then jump up or down through multiphonon emission or absorption. These effects were already included phenomenologically in the Shockley-Reed-Hall coefficients earlier, but even in their absence, electron phonon coupling softens the band-edges through phonon sidebands to create a gradual turn-on characteristic. The celebrated Huang-Rhys model [71] gives us the transformation of a sharp energy level into a comb of sidebands, skewed more favorably towards emission rather than absorption through a Boltzmann ratio

$$\begin{aligned} \delta(E-\epsilon_0) &\rightarrow \sum_{l=-\infty}^{\infty} A_l \delta(E-\epsilon_0+\Delta-l\hbar\omega), \quad \Delta = M^2/\hbar\omega \\ A_l &= e^{l\hbar\omega/2k_BT - g(2N_\omega+1)} I_l(2g\sqrt{N_\omega(N_\omega+1)}) \end{aligned} \tag{2.35}$$

for a fixed energy level ϵ_0 and phonon frequency ω with electron-phonon coupling M, equilibrium phonon occupancy $N_\omega = 1/[e^{\hbar\omega/k_BT} - 1]$ [71]. $g = (M/\hbar\omega)^2$ is the renormalized coupling, and I_l is the modified Bessel function that kills multiphonon processes for large $|l| \gg 1$. Δ represents the polaronic shift needed to preserve the center-of-mass for the density of states in the presence of higher emission over absorption probabilities, while A_ls are the weight factors ensuring the integral of the density of states stays unchanged with phonon participation. Integrating over the conduction band-edge density of states and the phonon density of states D_{ph}, we get the transformation

$$\Theta(E-E_c) \rightarrow \sum_{l=-\infty}^{\infty} \int_0^{\infty} d\omega D_{ph}(\omega) A_l \Theta(E-E_c+\Delta-l\hbar\omega) \tag{2.36}$$

The end result is a smearing of the step function bandedge that can be phenomenologically described as $1/2 - \tan^{-1}[(E - E_C)/\Gamma]/\pi$, the integral of a Lorenzian. This smearing softens the corresponding step function like turn-on of the current in the gate transfer characteristic and increases the subthreshold swing near the band-edge where it was supposed to be the smallest. The resulting current is now given by

$$\begin{aligned} I &= \frac{q}{h}\int_{-\infty}^{E_{vs}} T(E)[f_S(E) - f_D(E)]dE \\ &\times \int_{E_{FD}}^{\infty} \frac{1}{\pi}\frac{(\Gamma/2)}{(E-E')^2 + (\Gamma/2)^2}dE' \end{aligned} \tag{2.37}$$

which involves the convolution with a Lorenzian. As a result, the subthreshold swing near the bandedge $S = \left(\partial \log_{10} I/\partial V_G\right)^{-1} \approx (\Delta\phi + \Gamma)\ln 10/q$, with $\Delta\phi = E_{vs} - E_{cc}$ being the band offset in the ON state arising from the Fermi function difference, while the broadening Γ arises from the convolutional smearing. In the absence of phonons, the subthreshold swing vanishes at the band-edge ($\Delta\phi = 0$), but the phonon induced broadening Γ softens that transition. Understandably, we expect a similar smearing in the presence of disorder, which creates a similar Urbach tail that smudges out the band edges.

2.2 Model Calibration and Results

In our study, we simulated an InGaAs homojunction TFET and a GaSb/InAs heterojunction TFET. For calibration with NEGF simulations, shown in Fig. 2.7, device dimensions used in the original papers were used. The homojunction and heterojunction TFETs in our simulations, as shown in Fig. 2.9 and 2.10, share the same device geometry, with the channel length $L_{ch} = 100nm$, the channel thickness $t_{ch} = 5nm$, oxide thickness $t_{ox} = 2nm$, and gate oxide dielectric constant of $\epsilon_r = 11.9$. For the homojunction TFET, doping concentrations of $2 \times 10^{19}\ cm^{-3}$, $10^{14}\ cm^{-3}$ and $10^{18}\ cm^{-3}$ are used for the source, channel and drain, respectively. For the heterojunction TFET, doping concentrations of $5 \times 10^{19}\ cm^{-3}$, 10^{14}

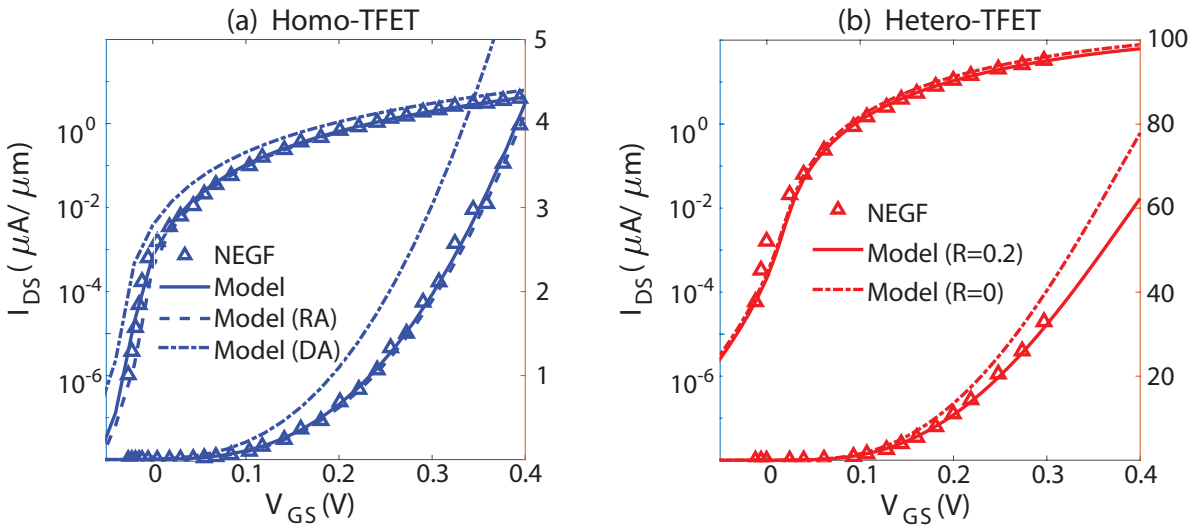

Figure 2.7: $I_D - V_{GS}$ characteristics of (a) Homojunction and (b) Heterojunction TFETs by analytical model (this work) and NEGF [43, 45]. In (a), the results presented by solid lines are obtained by analytical models with material parameters extracted from the NEGF simulations for benchmark purpose. The dashed lines are by analytical models with material parameters extracted from tight binding calculations by this work, from Fig. 2.4. In (b), the solid line uses $R = 0$, corresponding to zero reflection at the InAs/GaSb interface; the dashed line uses $R = 0.2$ to achieve better agreement with NEGF. Here DA represents digital alloy and RA represents random alloy.

cm^{-3} and 5×10^{17} cm^{-3} are used for the source, channel and drain, respectively.

In order to benchmark the accuracy of our analytical model, we compared the results of our model to previous tight binding based NEGF simulations for ballistic TFETs. The comparison is shown in Fig. 2.7. Our analytical model of a 100nm homojunction TFET was compared to NEGF simulations of Avci *et al.* [43], while our 30nm heterojunction TFET data was compared to simulations carried out by Long *et al.* [45]. For benchmarking purposes, we use parameters such as m^* and E_g from past NEGF work, with the understanding that at hetero-interfaces these parameters will change and will need to be corrected later from our two-band model fitted to DFT. The only free parameter we can adjust to calibrate our results for homo and heterojunctions are their gate work functions. The gate work function we used is 4.11 eV for the homojunction TFET and 4.642 eV for the heterojunction TFET. Compared with the NEGF calculations, we see that our analytical model is in excellent agreement for the homo-junction TFET. We get a slightly higher current for the hetero-junction TFET if we

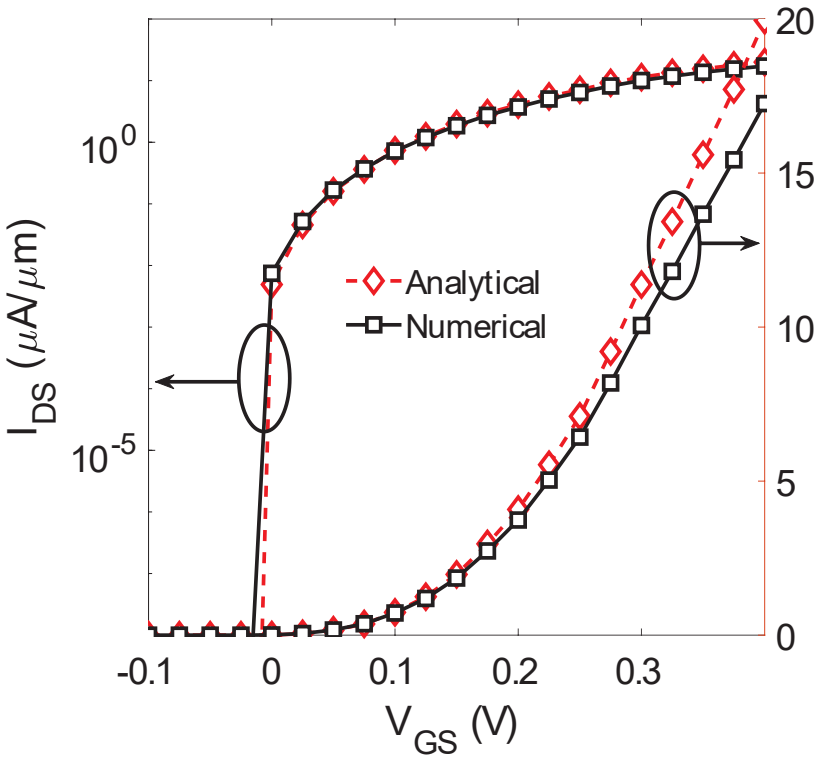

Figure 2.8: Comparison of analytical vs. numerical calculation of the $I_D - V_{GS}$ characteristics of a homojunction TFET at $V_{DS} = 0.3V$. The analytical result deviates from the numerical simulation by about 20%. This is a reasonable deviation due to the approximations made in the analytical result.

assume no reflection at the InAs/GaSb interface ($R = 0$). In reality, we expect some reflection at the material interface. To achieve a better agreement with NEGF for heterojunction TFET, we use a reflection parameter $R = 0.2$. The integrations in the model can be performed both numerically and analytically. A comparison between the two methods is shown in Fig. 2.8. Our homo-junction ballistic TFET analytical model yields a steep subthreshold slope of 8.4mV/dec, which is slightly higher than the NEGF data which has a SS of 6.5 mV/dec (some extraction error could exist); for heterojunction TFET, our analytical model leads to a SS of 14.9 mV/dec which is in good agreement with the NEGF SS of 14.4 mV/dec.

Since the tunneling current depends exponentially on the material parameters, it is important to get these parameters accurately - which becomes questionable at interfaces for conventional tight-binding models that are typically fitted to bulk bandstructures and lack explicit atom-like localized non-orthogonal orbital basis sets. For comparison, we show in

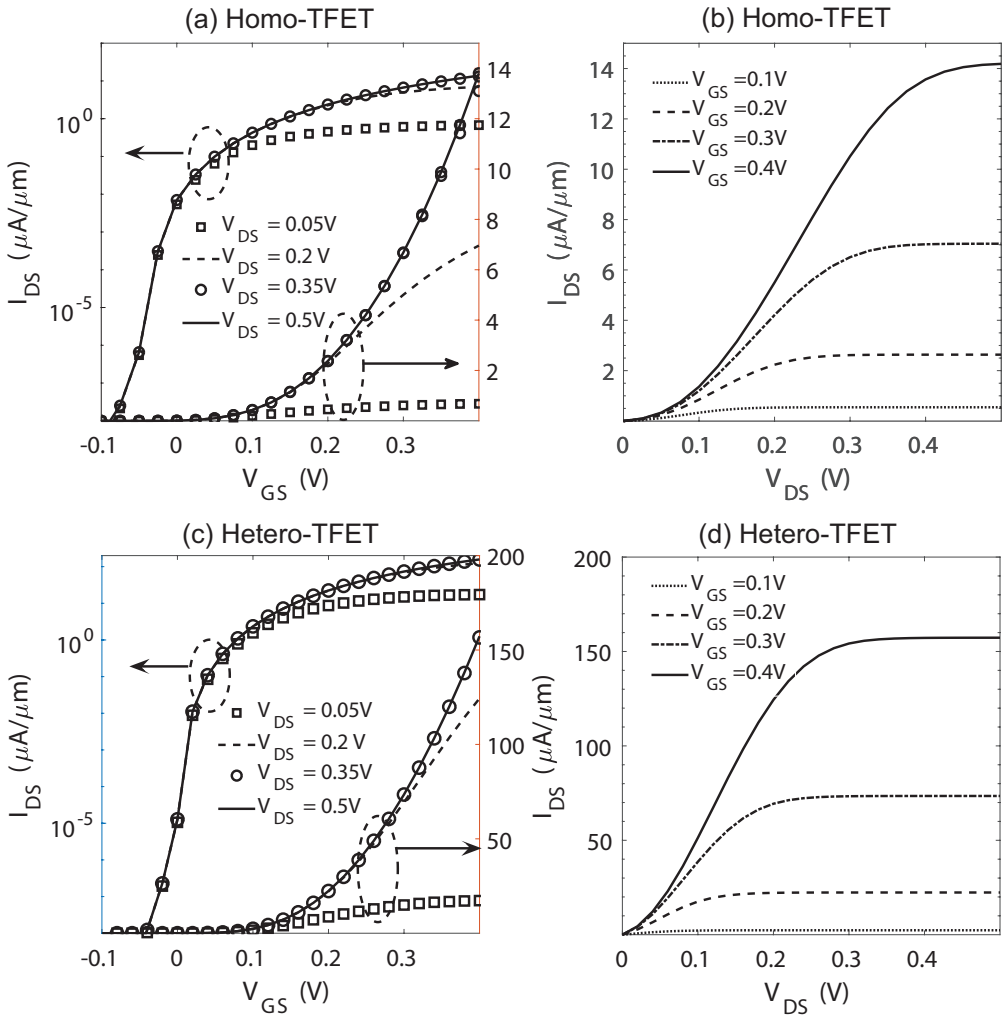

Figure 2.9: (a) Transfer characteristics ($I_D - V_{GS}$) and (b) Output characteristics ($I_D - V_{DS}$) of Homojunction TFET; (c) Transfer characteristics ($I_D - V_{GS}$) and (d) Output characteristics ($I_D - V_{DS}$) of Heterojunction TFET.

Fig. 2.7 (a) the $I_D - V_{GS}$ of the homojunction TFET using material parameters extracted from our own tight binding calculations fitted to DFT (Fig. 2.3). Based on parameters of the random InGaAs alloy extracted from our tight binding calculation, we find that its $I_D - V_{GS}$ is higher than previous ballistic NEGF calculations with tight binding parameters fitted

Material	InGaAs (ref [43])	InGaAs (random)	InGaAs (digital)	GaSb (ref [45])	InAs (ref [45])
$E_g(eV)$	0.740	0.704	0.730	1.20	0.76
$m^*(m_0)$	0.041	0.041	0.043	0.073	0.052

Table 2.1: Material parameters used in the TFET simulations.

to experimental data. This is mainly due to the smaller band gap in the random alloy, as shown in Table 2.1. However, the digital alloy agrees with previous NEGF calculations. The difference in material parameters is summarized in table 2.1.

Fig. 2.9 shows the $I_D - V_{GS}$ and $I_D - V_{DS}$ of a ballistic homojunction TFET and heterojunction TFET. Here we used the 100nm channel for both homo and heterojunction TFETs. The heterojunction TFET shows a larger on-current for the same bias due to a narrower tunneling barrier in the heterojunction TFET. For the homojunction TFET, we used the band gap and effective masses of random alloy InGaAs as shown in table 2.1. The $I_D - V_{GS}$ shows monotonous increasing behavior for both cases. The $I_D - V_{GS}$ shows a current saturation behavior for small V_{DS}, suggesting the integrated transmission in the TFET is saturated as well. The SS of the homojunction TFET with 100nm channel length is 9.3 mV/dec, and the heterojunction TFET with 100nm channel length has a smaller SS of 6.8 mV/dec. While these numbers are impressive, we will now see how non-idealities tend to affect these metrics.

Fig. 2.10 shows the $I_D - V_{GS}$ with various non-idealities such as trap assisted tunneling and Auger effect in homojunction and heterojunction TFETs. We see that the impact of the

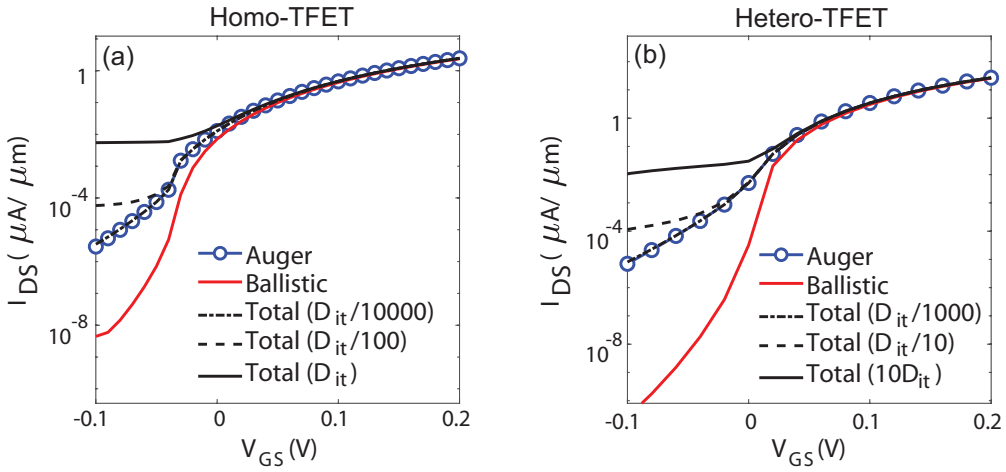

Figure 2.10: Comparison of non-ideal current with Ballistic, Auger and TAT current for Homojunction TFET. The Ballistic current is a steep switch, while the Auger effect increase the off-current and sub-threshold swing. The trap assisted tunneling current dominate the off-current when the density of trap is larger than $\frac{D_{it}}{10000} = 5 \times 10^{12} m^{-2} eV^{-1}$.

trap-assisted tunneling process (TAT) is a significant increase of the off-current, in agreement with past simulations [47].

Beyond trap assisted tunneling that is interface-specific, there is intrinsic leakage at high fields due to Auger generation that will also impact the off-current. In fact, Auger generation dominates the behavior of the off-current region when the traps are sufficiently low. Our analysis shows that when the trap concentration at the junction of a homo-TFET is lower than a critical interface trap density of $5 \times 10^{12} m^{-2} eV^{-1}$, the Auger current begins to dominate the off-current region of the TFET. In contrast, in hetero junction TFETs, the critical interface trap density is about one order of magnitude higher due to the intrinsically higher on-current. This Auger limited off-current defines a different limit of sub-threshold swing. Since the Auger generation current satisfies $J_{Aug} \propto \exp\left(-\frac{2\mu^{-1}+1}{\mu^{-1}+1}\frac{\Delta E}{kT}\right)$, the Auger limited subthreshold swing is $\mathrm{SS}_{Aug} \approx \frac{\mu^{-1}+1}{2\mu^{-1}+1} 60\mathrm{mV/dec}$.

The Auger current depends on the wave function overlap of the electron and hole states

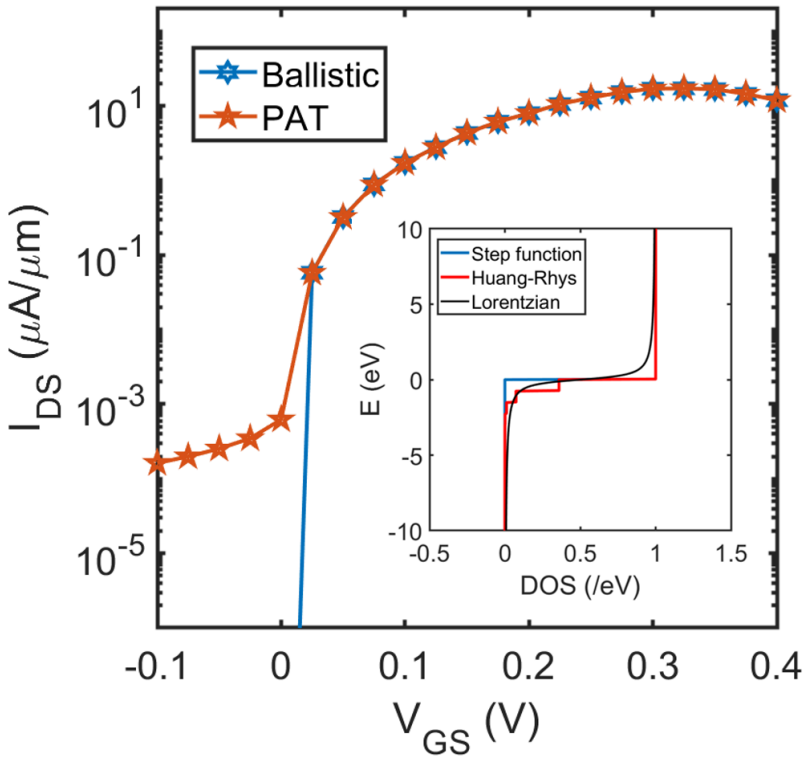

Figure 2.11: Transfer characteristics of a heterojunction Tunnel FET with phonon assisted tunneling in comparison to ballistic simulations.

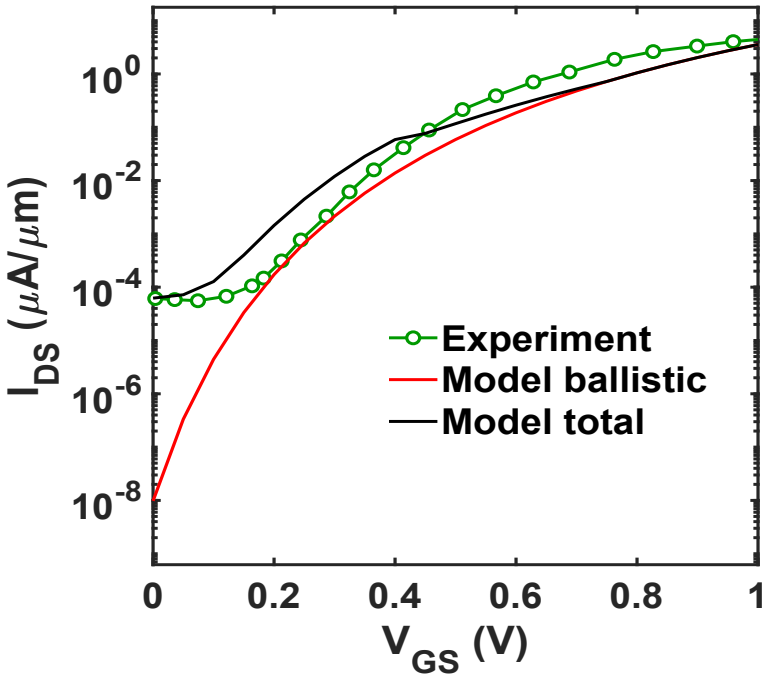

Figure 2.12: Model calibration with experimental homojunction InGaAs TFET at $V_{DS} = 0.3V$ [75].

across the junction. In a TFET, this wave function overlap depends on the junction width and barrier height. In a heterojunction TFET, the junction width is smaller compared with homojunction TFET. However, barrier height also affects the wave function overlaps since the states decay faster in a higher barrier. An ideal junction that minimizes the Auger current is still an abrupt junction. In Fig. 2.11 we have shown how PAT affects the transfer characteristics of a heterojunction TFET in comparison to ballistic results. The increase of OFF current and SS due to PAT depends on the broadening factor of the Lorentzian distribution, Γ. Here, we take $\Gamma = 0.6\ meV$ [74]. However, the broadening factor needs to be determined more accurately using experimental measurements to better model the effects of PAT. Temperature dependence of the transfer characteristics is the key indicator of the effect of PAT on TFETs.

The calibration of the model with a fabricated InGaAs homojunction TFET is shown in Fig 2.12 [75]. The calibrated model marginally overestimates the current in the subthreshold region and underestimates the ON current at some voltages. The effect of non-idealities on

the characteristics are shown in the figure. For fitting the experimental results, a trap density of $10^{11}\ cm^{-2}eV^{-1}$ is used. In order to fit the data more accurately (in the linear scale), we multiply a fitting factor of 0.5 to the Auger generation current to account for other factors, like Coulomb screening, that has not been considered in this model.

2.3 Optimizing Auger Generation

In addition to minimizing traps, the reduction of other leakage mechanisms like Auger generation is critical to realizing steep transfer characteristics. The Auger process strongly depends on the carrier concentration and material effective mass. The source doping, N_S, determines the number of electrons available to participate in the Auger generation process, while the valence band (VB) effective mass controls the wavefunction confinement and hence the wavefunction overlap between the valence and conduction band states that enter the scattering matrix element M [56, 57]. We study the effect of changing these parameters on I_{OFF}, I_{ON} and on/off ratio in the absence of the TAT component. For these simulations, we consider a channel length $L_{ch} = 100nm$, channel thickness $t_{ch} = 5nm$ and oxide thickness $t_{ox} = 2nm$. In VLSI circuits, the ON and OFF voltages are generally defined relative to the ground (0V) and power supply voltage (V_{DD}), respectively. Following this convention, in our simulations, we define I_{OFF} to be the current at $V_{GS} = 0V$ and $V_{DS} = V_{DD}$ and I_{ON} to be the current at $V_{GS} = V_{DS} = V_{DD}$.

The relationship between I_{OFF} and N_S is shown in Fig. 2.13(a). The off current increases by nearly an order of magnitude when source doping is increased by a similar number. The increased carrier concentration increases the probability of collision, which leads to a higher Auger generation rate. The OFF current decreases by a factor of 4.5 if N_S decreases from $1 \times 10^{20}\ cm^{-3}$ to $8 \times 10^{18}\ cm^{-3}$. The VB effective mass dependence of the off current is shown in Fig. 2.13(b), assuming the material bandgap remains unchanged. For our simulations, we consider the hole transport effective mass. Heavy-hole bands in III-V materials are

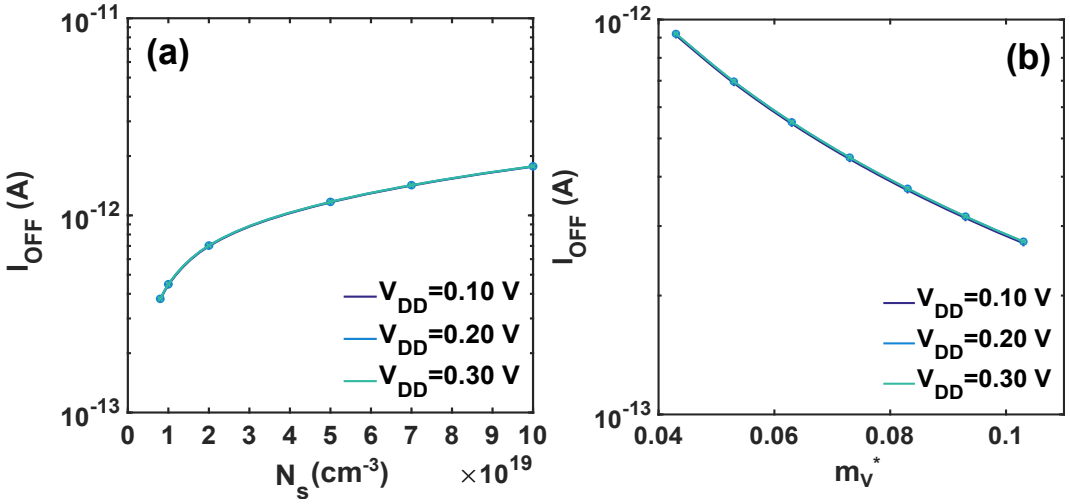

Figure 2.13: OFF current (including Auger Generation) vs. (a) source carrier concentration (b) source valence band effective mass for simulated double gate TFET, in the absence of traps (N_t=0), with parameter values from our previous paper [57].

known to be strongly anisotropic and in ultra-thin bodies, such as ours, the 2D hole subband structure strongly depends on the confinement plane [76, 77]. Long *et al.* have stated that for a 2nm thick p-GaSb/n-InAs TFET, the hole transport effective mass is 0.073 m_0 given $(1\bar{1}0)$ confinement and [110] transport [45]. Due to this confinement, the heavy hole effective mass decreases and becomes comparable to the light-hole effective mass. It is observed that

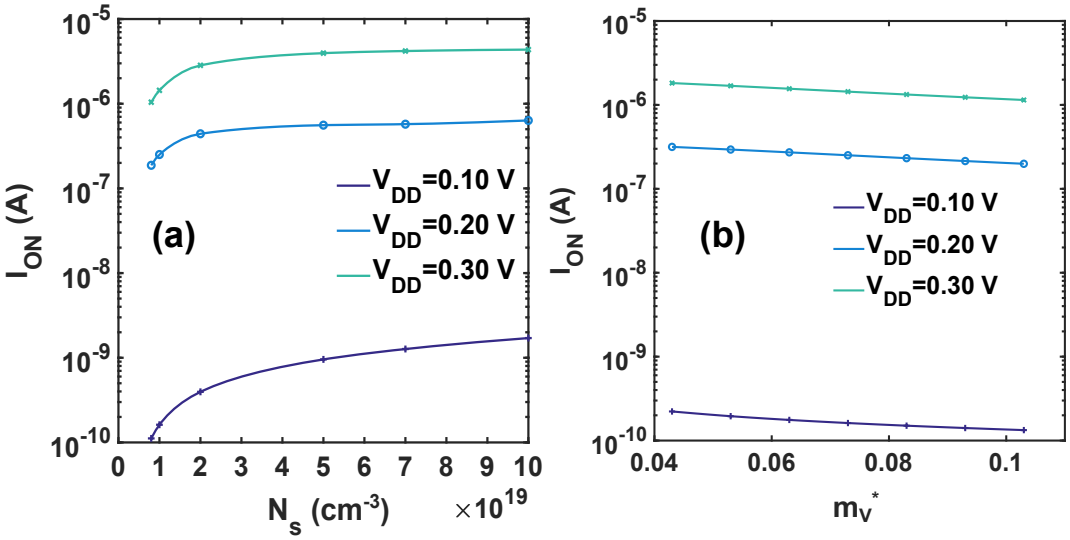

Figure 2.14: ON current vs. (a) source carrier concentration (b) source valence band effective mass. The ON current only contains the contribution from Auger generation and ballistic processes, but not trap assisted tunneling.

increasing m_v^* leads to a reduction in the OFF current. This happens because as the mass increases the wavefunction of the valence electrons gets more confined in real space. This leads to an overall reduction in wavefunction overlap between the valence band and conduction band states, thus reducing I_{OFF}. The valence band transport effective mass approaches the bulk heave hole effective mass if the channel thickness is increased. We expect that in this regime, the Auger generation will exhibit similar dependence on the effective mass. However, increasing the channel thickness to bulk values will adversely affect the electrostatic control of the gate. This will negatively impact the band-to-band tunneling and result in degraded performance of the TFET.

Fig. 2.14(a) and 2.14(b) exhibit the ON current dependence on the source doping and VB effective mass. In Fig. 2.14(a), it is observed that I_{ON} increases with doping concentration at low supply voltage, V_{DD}. For high supply voltages, the ON current becomes nearly constant after passing a certain critical doping. The tunneling probability depends on the tunneling width of the source/channel junction. A smaller tunneling width allows the electron to bridge the gap between the source valence band and channel conduction band (CB) easily. Increasing the source doping concentration is used as a method of decreasing the tunneling width by bringing the valence band and conduction band close together. As N_S is increased from $8 \times 10^{18}\ cm^{-3}$ to $5 \times 10^{19}\ cm^{-3}$ the tunneling width decreases and so the ON current will increase. However, as the concentration is increased to $1 \times 10^{20}\ cm^{-3}$ there is not much movement of the bands. The tunneling width remains the same. Thus, the ON current is seen to be constant at higher doping. Fig. 2.14(b) shows the dependence of I_{ON} on the VB effective mass. The I_{ON} predictably decreases with increase in m_v^*. The tunneling probability is proportional to the exponential of the negative square root of the effective mass. A lower effective mass increases the tunneling probability and hence the ON current. A higher V_{DD} pushes down the conduction band further towards the valence band. Thus, the tunneling width is reduced and hence the effect of effective mass on the ON current somewhat reduces.

At low supply voltages, the on/off ratio follows a positive trend with increasing N_S and

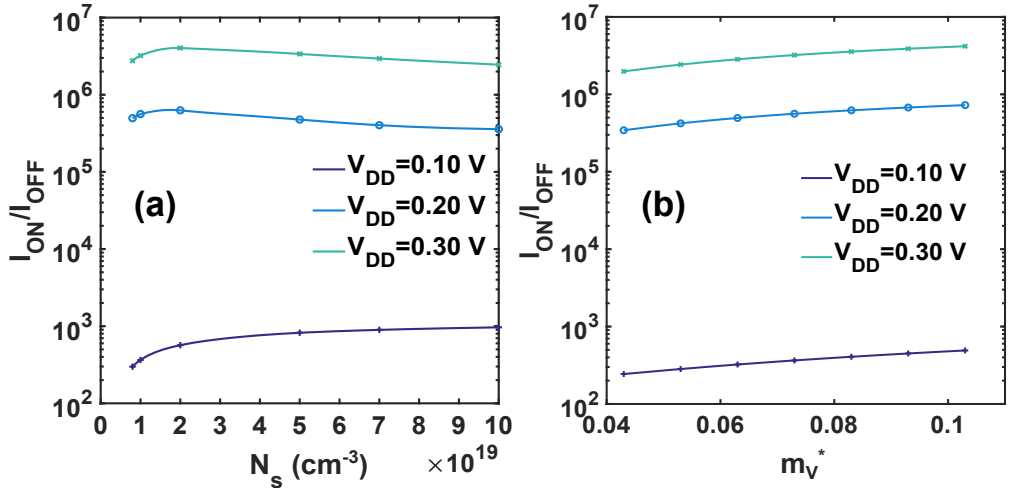

Figure 2.15: Ratio of ON current to OFF current vs. (a) source carrier concentration (b) source valence band effective mass. The trap-assisted tunneling component is excluded from both the currents in order to study the effect of Auger generation on these.

m_v^* as seen in Fig. 2.15(a) and 2.15(b). At low supply voltages, the VB and CB are far apart. Thus, increasing doping decreases the tunneling width and a reduced effective mass enhances the tunneling probability. At higher voltages, the I_{ON}/I_{OFF} starts decreasing after reaching a critical doping concentration which is observed in Fig. 2.15(a). This happens because the OFF current increases with higher doping due to increased Auger generation rate, but the ON current remains unchanged as the tunneling width remains fixed. The on/off ratio

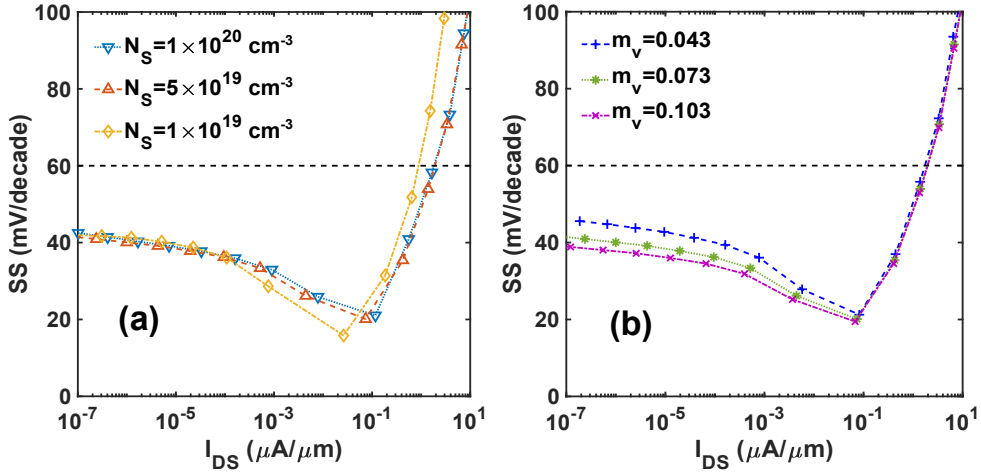

Figure 2.16: SS vs. Auger(Auger+Ballistic) current for (a) different dopings (b) different effective mass.

increases with increasing effective mass for higher supply voltages too as shown in Fig. 2.15(b). Another interesting observation from our simulation is that a SS<60 mV/decade for four orders of magnitude of current is attained at higher supply voltages of 0.2 and 0.3V even in the presence of Auger generation.

The subthreshold swing vs. I_{DS} for different carrier concentrations is shown in Fig. 2.16(a). We observe that lowering the concentration reduces the minimum SS. Reducing N_S by an order of magnitude from $1 \times 10^{20}\ cm^{-3}$ to $1 \times 10^{19}\ cm^{-3}$ decreases the minimum SS from 21 mV/dec to 15.82 mV/dec. The effect of VB effective mass on SS is shown in Fig. 2.16(b). Increasing m_v^* shifts the curve downwards at low currents as seen in the plot. At I_{DS}=10^{-5} $\mu A/\mu m$ increasing m_v^* from 0.043 to 0.103 decreases the SS from 42 mV/dec to 36 mV/dec. It has been previously shown that $\mathrm{SS}_{Aug} \approx \frac{\mu^{-1}+1}{2\mu^{-1}+1}60\mathrm{mV/dec}$ [56, 57]. Here, $\mu = m_c^*/m_v^*$. This relationship explains why increasing VB effective mass can decrease the SS. A higher m_v^* reduces the mass ratio and thus leads to a decrease in the subthreshold swing.

Achieving subthreshold swing below 60 mV/dec will be immensely impactful for any device considered as a replacement for standard CMOS transistors. Our study acts as a reference for improving the performance of TFETs and achieve SS<60 mV/dec. The most effective way of reducing SS and improving the on/off ratio is by reducing the trap density. Often increasing carrier concentration is considered as a method of increasing ON current. However, as seen from our results, this can be detrimental to TFET performance after a critical doping at high voltages, which must therefore limit the source doping. For further reduction of tunneling width, other mechanisms like electrostatic doping can be used to enhance the ON current by placing an additional gate over the source region near the channel. In addition to doping, the effective mass is also important for improving TFET performance. From our simulations, we see that at high supply voltages, a slightly higher VB effective mass can significantly reduce OFF current and improve on/off ratio without affecting the ON current. III-V tertiary and quaternary alloys can be studied for this purpose since changing their composition results in different properties.

2.4 Summary

In this chapter, we have presented an analytical model which captures the essential device physics of a TFET. The two-band $k.p$ model uses material parameters obtained from tight-binding complex band calculations fitted to DFT at interfaces and accurately represent the material bandstructure. The correct potential model and calibration of the model with NEGF simulations for both homo and heterojunction TFETs allow us to precisely calculate the drain current at finite temperature using a modified Simmons equation. The inclusion of trap-assisted tunneling and Auger generation processes into the model can explain the considerable observed discrepancy between ballistic vs. experimental TFETs. These tools can be used to understand the effects of higher order processes in TFETs and explore ways to mitigate their deleterious effects in order to improve the performance of practical TFETs.

We conclude from our simulations that the most significant leakage mechanism affecting TFET performance is the TAT process. Improvement in device characteristics can be obtained by reducing trap density, but even that does not take it to the ballistic regime, as the leakage current is dominated by the Auger generation process at low trap densities. The device performance can be enhanced further by lowering the doping concentration and increasing source valence band effective mass in order to reduce Auger generation. To keep the ON current sufficiently high new device architectures like vertical Tunnel FETs or the use of electrostatic doping can be explored to compensate for the lower doping.

Key Contributions:

- Developed accurate chemistry-based TFET model.
- Demonstrated that TAT and Auger generation processes lead to higher off current and subthreshold swing in TFETs.
- Identified that Auger generation can be modulated using source doping and valence band effective mass.

Chapter 3

Designing Low Noise III-V Avalanche Photodiodes

The demand for efficient optical detectors is constantly growing due to rapid developments in telecommunication, LIDAR systems, and other military and research fields [9–15]. Photodetectors are increasingly being incorporated in photonic integrated circuits for Internet of Things and 5G communications [6–8]. These applications require higher sensitivity in comparison to traditional *p-i-n* photodiodes [24]. Avalanche photodiodes (APDs) are often deployed instead due to their higher sensitivity, enabled by their intrinsic gain mechanism. However, the stochastic nature of the impact ionization process of APDs adds an excess noise factor $F(M)$ [25–27]. A low value of k, which is the ratio of hole ionization coefficient β to the electron ionization coefficient, α, is desirable for designing low-noise electron injected APDs. This ratio stipulates that for pure electron injection, a significantly lower hole ionization than the electron ionization rate leads to reduced shot noise. If impact ionization is caused by pure hole injection, k will be replaced by $1/k$.

Recently, several III-V digital alloys, i.e., short-period superlattices with binary components stacked alternately in a periodic manner, were found to exhibit extremely low noise currents and a high gain-bandwidth product in the short-infrared wavelength spectrum [30–32].

Characterization of InAlAs, AlInAsSb and AlAsSb digital alloy APDs have shown very small values of k [30–32], whereas other digital alloys, like InGaAs and AlGaAs, demonstrate much higher k value [78, 79]. Based on previous full-band Monte Carlo simulations [29, 36, 37], the low k has been attributed to the presence of superlattice minigaps inside the valence band of the material bandstructure, along with an enhanced effective mass arising from the lower band-width available to the holes. Such valence band minigaps often co-exist with similar (but not symmetrical) minigaps in the conduction band. However, electrons in the conduction band typically have very low effective mass, which allows quantum tunneling and enhanced phonon scattering to circumvent minigaps in the conduction band. Furthermore, certain digital alloys showing minigaps do not exhibit low noise, and the reason behind that has not yet been addressed. Our postulate is that a combination of valence band minigaps, a large separation between light-hole and split-off bands, and corresponding enhanced hole effective mass tend to limit hole ionization coefficient. A comprehensive analysis is clearly necessary to understand the carrier impact ionization in these materials. In this chapter, the physics of low excess noise in III-V APDs is investigated. Based on our investigation, some empirical design criteria are proposed that can be used to design new materials for APDs with low noise.

In this chapter, we employ a fully atomistic, Environment-Dependent Tight Bindng (EDTB) model, [63] calibrated to Density Functional Theory (DFT) bandstructure as well as wavefunctions, to compute the bandstructures of several III-V digital and random alloys. Using a full three-dimensional quantum kinetic Non-Equilibrium Green's Method (NEGF) formalism with the EDTB Hamiltonian as input, we compute the ballistic transmission across the digital alloys that accounts for intraband quantum tunneling across minigaps and light-hole/split-off bands offset. Additionally, a full-band Boltzmann transport solver is employed to determine the energy resolved carrier density distribution under the influence of an electric field in order to study the effect of optical phonon scattering in these short-period superlattices. The calculations are performed using computational resources at the University

of Virginia and XSEDE [80]. Using these transport formalisms, we elucidate the impact of minigap sizes, light-hole/split-off band offset and effective masses on carrier transport in the valence band.

Our simulations demonstrate that the squashing of subbands into tighter bandwidths, such as arising from minigap formation, or the engineering of large light-hole/split-off band offset lead to the suppression in transport of one carrier type, by resisting quantum tunneling or phonon-assisted thermal jumps. For InAlAs, the improved performance is primarily due to the minigaps generated by the digital alloy periodicity and the corresponding enhanced effective mass. For AlInAsSb and AlAsSb, the gain is a combination of minigaps, large effective mass and LH/SO offset. The LH/SO offsets in these two alloys results arise from the strong spin-orbit coupling due to the Sb atoms, which we believe also leads to the low excess noise recently observed in Sb-containing quaternary random alloy APDs [33,34]. A quantitative comparison of the various alloy gains measured is presented in the last two columns of Table IV.

The unique superlattice structure of the digital alloys opens the possibility for designing new low-noise alloy combinations for the detection of other frequency ranges. Ideally, it is easier and cheaper to at first computationally study the suitability of the alloys for achieving low noise before actually fabricating these. For this purpose, we need a set of design criteria for judging the alloy performance using theoretically calculated parameters. Based on our simulations, we propose five simple inequalities that can be used to judge the suitability of digital alloys for use in low-noise APDs. We judge the aptness of five existing digital alloys-InAlAs, InGaAs, AlGaAs, AlInAsSb and AlAsSb. We observe that the inequalities provide a good benchmark for gauging the applicability of digital alloys for use in low-noise APDs. Additionally, we also study the effect of Sb atoms on the LH/SO offset of the quaternary random alloys that demonstrate low excess noise. We computed the bandstructure of these random alloys to determine the limiting factor of the hole impact ionization in these materials. A significant portion of this chapter is reproduced from Ref. [81] co-authored with Y. Tan, J.

Zheng, J. C. Campbell and A. W. Ghosh.

3.1 Simulation Method

3.1.1 Environment Dependent Tight Binding and Band Unfolding for atomistic description

In order to understand the influence of minigap filtering in digital alloy structures, an accurate band structure over the entire Brillouin zone is required. The periodic structure of the InAlAs digital alloy is shown in Fig. 3.1(a), and Fig. 3.1(b) shows the typical structure of a *p-i-n* APD. We have developed an EDTB Model to accurately calculate the band structure of alloys [62,63]. Traditional tight binding models are calibrated directly to bulk bandstructures near their high symmetry points and not to the underlying chemical orbital basis sets [62]. These models are not easily transferable to significantly strained surfaces and interfaces

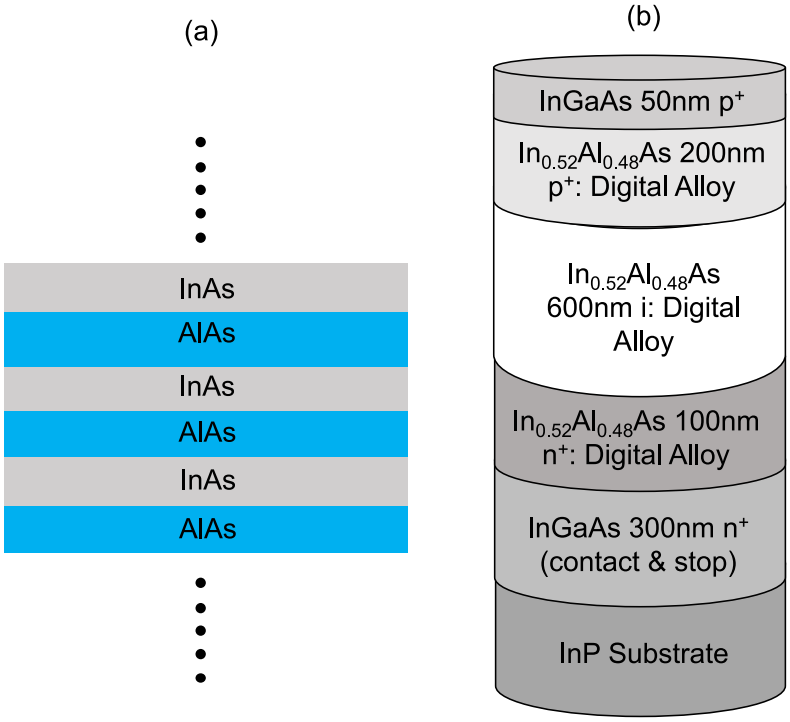

Figure 3.1: (a) Digital alloy structure (b) typical structure of an APD.

where the environment has a significant impact on their material chemistry. In other words, the tight binding parameters work directly with the eigenvalues (E-k) and not with the full eigenvectors. While the crystallographic point group symmetry is enforced by the angular transformations of the orbitals, the radial components of the Bloch wavefunctions, which determine bonding and tunneling properties, are left uncalibrated. Previously, in order to incorporate accuracy of radial components, an Extended Hückel theory [65, 66] was used that incorporated explicit Wannier basis sets created from non-orthogonal atomic orbitals that were fitted to DFT for the bulk Hamiltonian. The fitted basis sets were transferrable to other environments by simply recomputing the orbital matrix elements that the bonding terms were assumed to be proportional to. As an alternative, the EDTB model employs conventional orthogonal Wannier like basis sets. The tight binding parameters of this model are generated by fitting to both Hybrid functional (HSE06) [64] band structures and orbital resolved wave functions. Our tight binding model can incorporate strain and interface induced changes in the environment by tracking changes in the neighboring atomic coordinates, bond lengths and bond angles. The onsite elements of each atom have contributions from all its neighboring atoms. The fitting targets include unstrained and strained bulk III-V materials as well as select alloys. We have previously shown in the past that our tight binding model has the capability of matching the hybrid functional band structures for bulk, strained layers and superlattices [57, 63].

The band structures of the alloys contain a massive number of spaghetti-like bands due to the large supercell of the system that translates to a small Brillouin zone with closely separated minibands and minigaps. In order to transform the complicated band structure into something tractable, we employ the technique of band unfolding [68–70]. This method involves projecting the eigenvalues back to the extended Brillouin zone of the primitive unit cell of either component, with weights set by decomposing individual eigenfunctions into multiple Bloch wavefunctions with different wave vectors in the Brillouin zone of the original primitive unit cell. The supercell eigenvector $|\vec{K}m\rangle$ is expressible in terms of the linear

combination of primitive eigenvectors $|\vec{k_i}n\rangle$. The eigenstate E_p of an atom with wave vector k can be expressed as a linear combination of atomic-orbital wavefunctions. The supercell electron wavefunction $|\psi^{SC}_{m\vec{K}}\rangle$ can be written as a linear combination of electron wavefunctions in the primitive cell as [30]

$$|\psi^{SC}_{m\vec{K}}\rangle = \sum_n a\left(\vec{k_i}, n; \vec{K}, m\right) |\psi^{PC}_{n\vec{k_i}}\rangle \tag{3.1}$$

$$\vec{k_i} \in \{\vec{\tilde{k}_i}\}$$

where, $|\psi^{PC}_{n\vec{k_i}}\rangle$ is the electron wavefunction for the wave vector $\vec{k_i}$ in the nth band of the primitive cell. Here, $\vec{K}$ and $\vec{k}$ denote the reciprocal vector in supercell and primitive cell, respectively. The folding vector $\vec{G}_{\vec{k}\to\vec{K}}$ contains the projection relationship and is expressed as

$$\vec{K} = \vec{k} - \vec{G}_{\vec{k}\to\vec{K}} \ . \tag{3.2}$$

The projection of the supercell wavefunction $|\psi^{SC}_{m\vec{K}}\rangle$ into the primitive cell wavefunction $|\psi^{PC}_{n\vec{k_i}}\rangle$ is given as

$$P_{m\vec{K}} = \sum_n |\ \langle\psi^{SC}_{m\vec{K}}|\psi^{PC}_{n\vec{k_i}}\rangle\ |^2 \ . \tag{3.3}$$

Plotting these projection coefficients gives a cleaner picture of the band evolution from the individual primitive components to the superlattice bands.

3.1.2 Non-Equilibrium Green's Function Method for coherent transmission

Under the influence of a large electric field, it is possible for carriers to move across minigaps by means of quantum tunneling. Such transport involves a sum of complex transmissions limited

by wavefunction symmetry between several minibands. We make use of the Non-Equilibrium Green's Function formalism to compute the ballistic transmission and study the influence of minigaps on quantum tunneling in digital alloys. The digital alloys we are interested in studying are translationally invariant in the plane perpendicular to the growth direction and have finite non-periodic hopping in the transport (growth) direction. Thus, we need a device Hamiltonian H whose basis is Fourier transformed into k-space in the perpendicular $x-y$ plane but is in real space in the z growth direction, i.e., $H\left(r_z, k_x, k_y\right)$. Conventionally, this can be done with a DFT Hamiltonian in real space, $H\left(r_z, r_x, r_y\right)$, which is Fourier transformed along the transverse axes to get $H\left(r_z, k_x, k_y\right)$. However, DFT Hamiltonians are complex and sometimes do not match with bulk material bandstructure. Thus, it is simpler to utilize a tight binding Hamiltonian whose $E-\vec{k}$s are calibrated to bulk bandstructure, and inverse transform along the growth direction.

The matrix elements of the 3D EDTB Hamiltonian are given in the basis of symmetrically orthogonalized atomic orbitals $|nb\mathbf{R}\rangle$. Here $\mathbf{R}$ denotes the position of the atom, n is the orbital type (s, p, d or s^*) and b denotes the type of atom (cation or anion). The Hamiltonian can also be represented in $k-$space basis $|nb\mathbf{k}\rangle$ by Fourier transforming the elements of the real-space Hamiltonian. The 3D Hamiltonian is then converted into a quasi-1D Hamiltonian [82]. The Hamiltonian elements can be represented in the basis $\left|nbj\mathbf{k}_{||}\right\rangle$ with "parallel" momentum $\mathbf{k}_{||}=(k_x, k_y)$ and "perpendicular" position $x_j = a_L/4$ as parameters. For a zinc-blende crystal, the distance between nearest-neighbor planes is one-fourth the lattice constant a_L. The 3D Hamiltonian is converted to the quasi-1D one by means of a partial Fourier transform [82, 83]:

$$\left|nbj\mathbf{k}_{||}\right\rangle = L_{BZ}^{-1/2} \int dk_z e^{-ik_z j a_L/4} \left|nb\mathbf{k}\right\rangle \ . \tag{3.4}$$

Here $L_{BZ} = 8\pi/a_L$ is the length of the one-dimensional (1D) Brillouin zone over which the k_z integral is taken. The quasi-1D Hamiltonian is position dependent in the growth direction. Thus, we are able to utilize the accurate bandstructure capability of the EDTB.

In presence of contacts, the time-independent open boundary Schrödinger equation reads

$$(EI - H - \Sigma_1 - \Sigma_2)\Psi = S_1 + S_2 \tag{3.5}$$

where, E represents energy, I denotes identity matrix and $\Sigma_{1,2}$ are the self-energies for the left and right contacts, respectively, describing electron outflow, while $S_{1,2}$ are the inflow wavefunctions. The solution to this equation is $\Psi = G(S_1 + S_2)$, where the Green's function [84]

$$G(E) = [EI - H - \Sigma_1 - \Sigma_2]^{-1} \ . \tag{3.6}$$

Here H includes the applied potential, added to the onsite 1D elements. Assuming the contacts are held in local equilibria with bias-separated quasi-Fermi levels $E_{F1,2}$, we can write the bilinear thermal average $\langle S_i S_i^\dagger \rangle = \Gamma_i f(E - E_{Fi})$ where f is the Fermi-Dirac distribution and $\Gamma_{1,2} = i(\Sigma_{1,2} - \Sigma_{1,2}^\dagger)$ denoting the broadening matrices of the two contacts. The equal time current $I = q(d/dt + d/dt')Tr\langle \Psi^\dagger(t)\Psi(t')\rangle|_{t=t'}$ then takes the Landauer form $I = (q/h)\int dE T(f_1 - f_2)$, where the coherent transmission between the two contacts is set by the Fisher-Lee formula

$$T(E) = Tr\left[\Gamma_1 G \Gamma_2 G^\dagger\right] \tag{3.7}$$

where Tr represents the trace operator. The energy resolved net current density from the layer m to layer $m+1$ is expressed as [82]:

$$\begin{aligned} J_{m,m+1}(E) \ = \ & -\frac{iq}{h}\int \frac{\mathbf{k}_{||}}{(2\pi)^2} Tr[G^{n,p}_{m+1,m}H_{m,m+1} \\ & -G^{n,p}_{m,m+1}H_{m+1,m}] \end{aligned} \tag{3.8}$$

where, $G^n = \langle \psi^\dagger \psi \rangle$ and $G^p = \langle \psi \psi^\dagger \rangle$ represent electron (n) and hole density (p) correspondingly and $H_{m,m+1}$ is the tight binding hopping element between layers m and $m+1$ along the transport/growth direction.

3.1.3 Boltzmann Transport Model for incoherent scattering

The NEGF approach is particularly suited to ballistic transport where coherent quantum effects dominate. Incoherent scattering requires a self-consistent Born approximation which is computationally quite involved. We need a practical treatment of scattering. Under an external electric field, the carrier distributions in digital alloys no longer follow a local Fermi distribution, but re-distribute over real-space and momentum space. To understand the carrier distribution under electric field in digital alloys, we employed the multi-band Boltzmann equation.

$$\begin{aligned} \vec{v}\cdot\nabla_{\mathbf{r}} f_n + \vec{F}\cdot\nabla_{\mathbf{k}} f_n &= \sum_{m,\vec{p}'} S\left(\vec{p}',\vec{p}\right) f_m\left(\vec{p}'\right)\left[1-f_n\left(\vec{p}\right)\right] \\ &\quad - \sum_{m,\vec{p}'} S\left(\vec{p},\vec{p}'\right) f_n\left(\vec{p}\right)\left[1-f_m\left(\vec{p}'\right)\right] \end{aligned} \tag{3.9}$$

Here, $f = f(\mathbf{r},\mathbf{k})$ is the carrier distribution, n and m are band indices, $\vec{p}$ and $\vec{p}'$ are the momenta of the carriers, and $S\left(\vec{p}',\vec{p}\right)$ is the scattering rate. The left hand side of this equation alone describes the ballistic trajectory in the phase space of carriers under an electric field. The right hand side of the equation corresponds to the scattering processes, including intra-band and inter-band scattering.

In a homogeneous system where the electric field is a constant, the distribution function is independent of position, $\nabla_{\mathbf{r}} f = 0$ and the equation is reduced to

$$\begin{aligned} \vec{F}\cdot\nabla_{\mathbf{k}} f_n &= \sum_{m,\vec{p}'} S\left(\vec{p}',\vec{p}\right) f_m\left(\vec{p}'\right)\left[1-f_n\left(\vec{p}\right)\right] \\ &\quad - \sum_{m,\vec{p}'} S\left(\vec{p},\vec{p}'\right) f_n\left(\vec{p}\right)\left[1-f_m\left(\vec{p}'\right)\right] \ . \end{aligned} \tag{3.10}$$

For APDs, it is critical to consider optical phonon scattering, which is the dominant process besides tunneling that allows carriers to overcome the minigap arising in the band structures of digital alloys. The optical phonon has a non-trivial energy of $\hbar\omega_{opt}$ that can be absorbed

or emitted by carriers. The scattering rate $S(\vec{p}', \vec{p})$ has the form set by Fermi's Golden Rule

$$S(\vec{p}', \vec{p}) = \frac{2\pi}{\hbar} |H_{\vec{p},\vec{p}'}|^2 \delta_{\vec{p}',\vec{p}\pm\vec{\beta}} \delta\left(E(\vec{p}') - E(\vec{p}) \pm \hbar\omega_{opt}\right) \ . \tag{3.11}$$

The $E(\vec{p})$ and $E(\vec{p}')$ are band structures of digital alloy calculated by the tight binding model. $H_{p,p'}$ can be calculated by evaluating electron-phonon coupling matrix elements explicitly. In this work, we extract a constant effective constant scattering strength $H_{\vec{p},\vec{p}'}$ from experimental mobility μ. The scattering lifetime τ, which is $1/S(\vec{p}', \vec{p})$, can be extracted from the mobility using $\mu = q\tau/m^*$. Due to the lack of experimental mobilities of the digital alloys, we considered the average of the binary constituent room temperature mobilities for extracting the lifetime. A simple average is done since the binary constituents in periods of most of the digital alloys considered here are equally divided. In using room temperature values, the underlying assumption is that the dominant scattering mechanism here is phonon scattering due to the large phonon population. Ionized impurity scattering is considered to be much lower due to digital alloys having clean interfaces [31]. It is then possible to extract $H_{\vec{p},\vec{p}'}$ from the scattering lifetime. To get the equilibrium solution, we solve Eq. 3.10 self-consistently, starting from an initial distribution $f = \delta_{\vec{k},0}$.

A detailed model of carrier transport in APDs also requires a NEGF treatment of impact ionization self-energies and a Blanter-Buttiker approach to extract shot noise, but we leave that to future work. Our focus here is on conductive near-ballistic transport, and the role of quantum tunneling and perturbative phonon scattering in circumventing this.

3.2 Results and Discussion

There are three common ways to achieve low noise and high gain-bandwidth product - selecting a semiconductor with favorable impact ionization coefficients, scaling the multiplication region to exploit the non-local aspect of impact ionization, and impact ionization engineering using appropriately designed heterojunctions [24]. Typically, the lower hole impact ionization

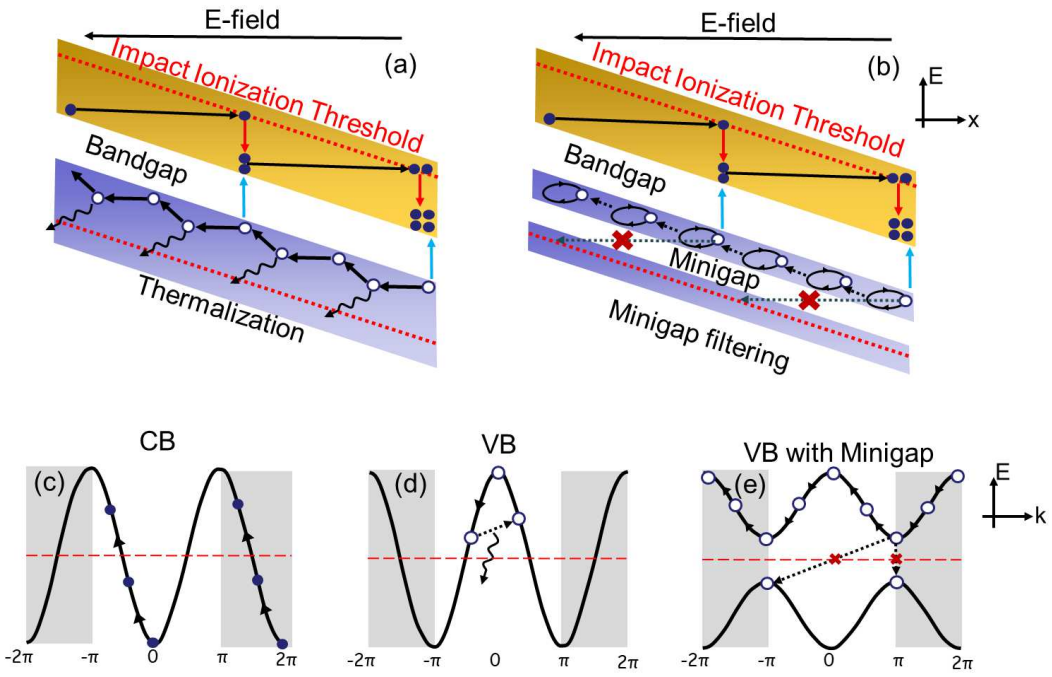

Figure 3.2: Impact ionization process in normal (random alloy) APD and superlattice APD. In both APDs, it is easier for electrons to gain energy and reach the impact ionization threshold (c). In normal APDs (a), holes find it harder to gain high energy compared to electrons because of thermalization. The hole energy is reduced by thermalization due to various scattering processes as shown in (d). In superlattice APD (b), the existence of minigaps makes it harder for holes to reach higher energies. The minigaps acts as barrier that prevent holes from moving to the lower valence bands. In the plots, the y-axis E is the total energy (kinetic+potential) meaning in between inelastic scattering events the particles travel horizontally.

coefficient in semiconductors is due to stronger scattering in the valence bands, as depicted in Fig. 3.2(a). Previously, the lowest noise with favorable impact ionization characteristics was realized with Si in the visible and near-infrared range, [85–88] and InAs [89–93] and HgCdTe [94,95] in the mid-infrared spectrum. In comparison, InGaAs/InAlAs [96,97] random alloy APDs exhibit significantly higher noise than Si, HgCdTe or InAs, which are the highest performance telecommunications APDs. In the recent past, digital alloy InAlAs APDs have demonstrated lower noise compared to their random alloy counterpart [30]. This seems a surprise, as the suppression of one carrier type (the opposite of ballistic flow expected in an ordered structure) is necessary for low excess noise. Initially, the low value of k in InAlAs was attributed to the presence of minigaps [29]. However, minigaps were also observed in InGaAs

digital alloy APDs which have higher excess noise [78, 98]. So, a clearer understanding of the minigap physics was needed, and hence a comprehensive study was required. Thus, we studied the physics of these alloys in details to get more insight into these materials.

Our recent results suggest that well defined minigaps introduced in the valence band of digital alloys suppress the density of high energy holes and thereby reduce the impact ionization greatly, as shown in Fig. 3.2(b). In a regular low-noise electron-injected APD, the electron ionization coefficient is much higher than the hole ionization coefficient. Thus, electrons can easily climb to higher kinetic energies in the conduction band, depicted in Fig. 3.2(c), and participate in the impact ionization process by gaining the impact ionization threshold energy. On the other hand, holes lose energy by various inelastic scattering processes (Fig. 3.2(d)), collectively known as thermalization. Thermalization prevents holes from reaching their secondary impact ionization threshold. In superlattice APDs, minigaps provide an additional filter mechanism that prevents holes from reaching the threshold energy required to initiate secondary impact ionization.

The effect of minigaps is shown in Fig. 3.2(e). However, not all digital alloy APDs exhibit low noise. The excess noise $F(M)$ vs. multiplication gain characteristics of experimental InGaAs, AlGaAs, InAlAs, AlInAsSb and AlAsSb digital alloy APDs are shown in Fig. 3.3 [30–32, 78, 79]. InGaAs APDs have the highest excess noise while AlAsSb has the lowest. The dotted lines represent the theoretical $F(M)$ vs. M calculated using the well known McIntyre's formula [25]. In order to understand the underlying physics in these digital alloys, an in-depth analysis of the material bandstructure and its effect on carrier transport is required.

We calculate the atomistic DFT-calibrated EDTB bandstructure of these materials and unfold their bands using the techniques described in section 3.1.1, to understand the underlying physics of their noise performance. In Fig. 3.4, we show the periods of the different digital alloys considered- (a) 6ML InGaAs, (b) 6ML AlGaAs, (c) 6ML InAlAs, (d) 10ML $Al_{0.7}In_{0.3}AsSb$ and (e) 5ML AlAsSb. Here, 6ML InGaAs includes 3ML InAs and 3ML GaAs,

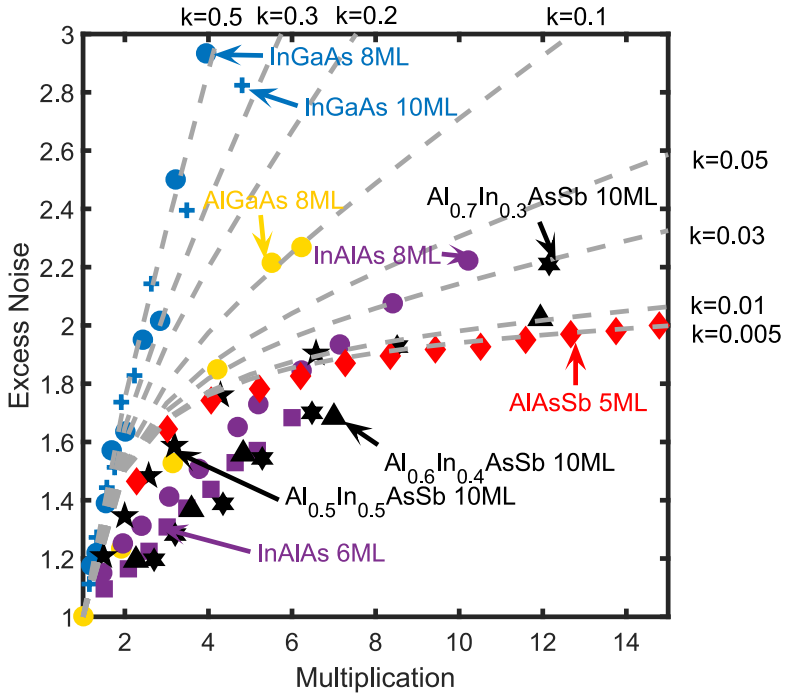

Figure 3.3: Experimentally measured Excess noise vs. Multiplication gain of InGaAs, AlGaAs, InAlAs, AlInAsSb and AlAsSb digital alloys are shown here [30–32, 78, 79]. The dotted lines for the corresponding k's are plotted using McIntyre's formula [25].

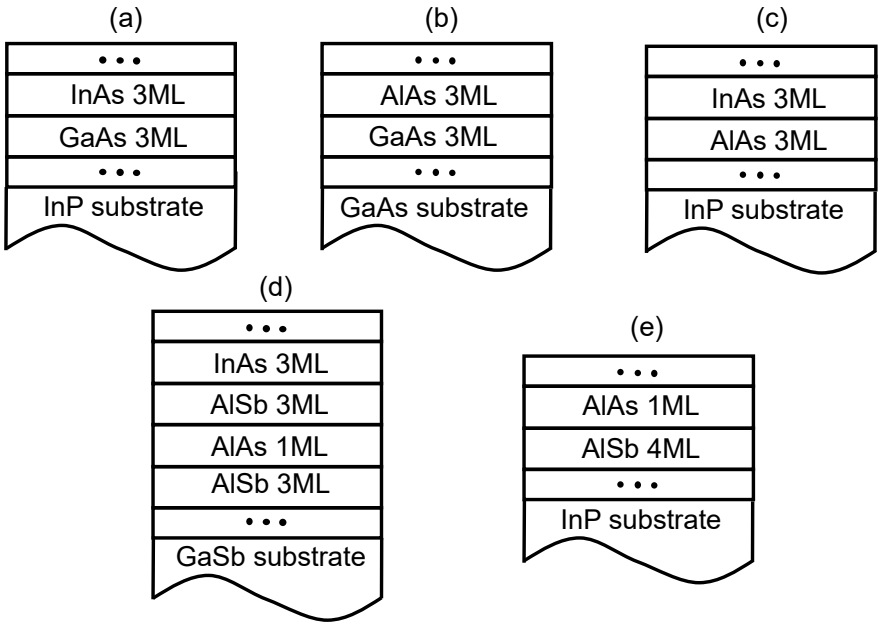

Figure 3.4: Lattice structures of (a) InGaAs, (b) AlGaAs, (c) InAlAs, (d) AlInAsSb and (e) AlAsSb digital alloys considered in this paper.

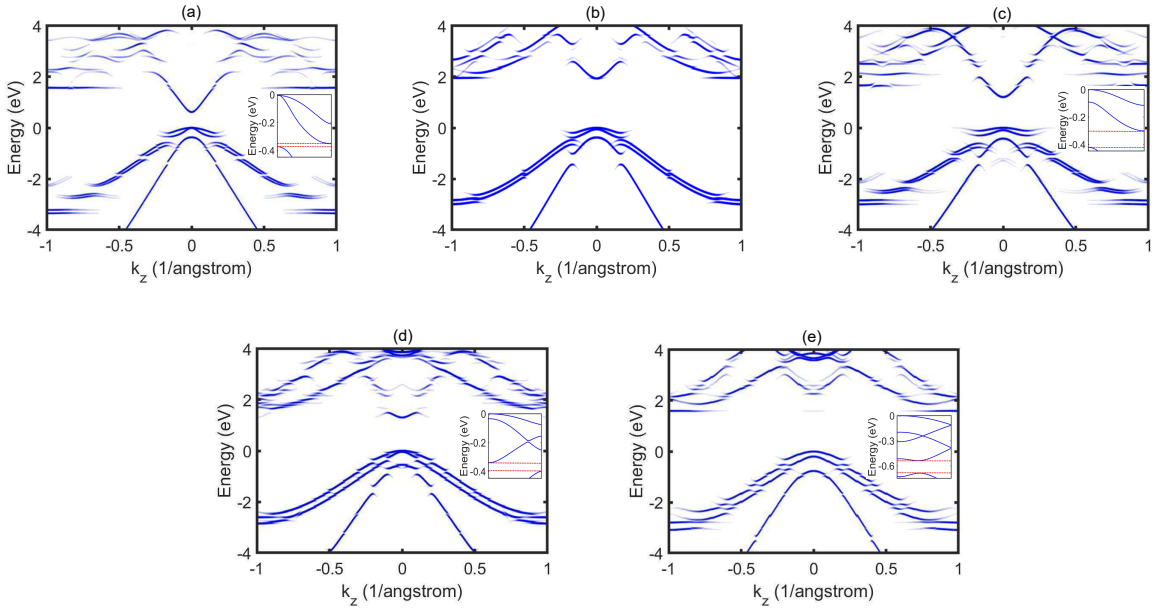

Figure 3.5: Unfolded bandstructure of (a) 6ML InGaAs (b) 6ML AlGaAs (c) 6ML InAlAs (d) 10ML AlInAsSb (e) 5ML AlAsSb. The minigaps of InGaAs, InAlAs, AlInAsSb and AlAsSb real bandstructures are shown in the insets.

Material	E_G (eV)	ΔE_b (eV)	ΔE_m (eV)	**HH** m^*	**LH** m^*	**SO** m^*	ΔE_{LS} (eV)
InGaAs	0.63	0.34	0.03	0.31	0.13	0.045	0.35
AlGaAs	1.94	1.03	0.34	0.45	0.31	0.12	0.33
InAlAs	1.23	0.30	0.12	0.5	0.4	0.1	0.31
AlInAsSb	1.19	0.33	0.06	0.42	0.38	0.08	0.48
AlAsSb	1.6	0.56	0.1	0.45	0.3	0.13	0.54

Table 3.1: Material parameters of the different digital alloys simulated in this work.

6ML AlGaAs has 3ML AlAs and 3ML GaAs, and 6 ML InAlAs has 3ML InAs and 3ML AlAs. 10ML $Al_{0.7}In_{0.3}AsSb$ consists of 3ML AlSb, 1ML AlAs, 3ML AlAs and 3ML InAs in its period. AlAsSb has 4ML AlSb and 1ML AlAs. The unfolded bandstructures of these alloys are shown in Fig. 3.5. We observe that minigaps exist in at least one of the valence bands (heavy-hole, light-hole or split-off) for all the material combinations. The InAlAs valence band structure is magnified in Fig. 3.6. The minigap between the LH and SO band is denoted in the figure. Additionally, the large separation between the LH and SO bands at the Γ point is highlighted.

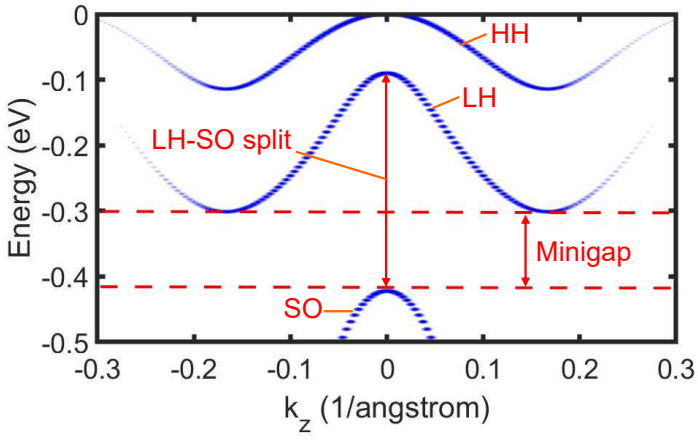

Figure 3.6: A magnified picture of the InAlAs valence band shows the minigap closest to the valence band edge. The split between the LH and SO at the Γ point is also highlighted.

The role of the minigaps on hole localization is not identical across different alloys. For instance, the presence of minigaps in material bandstructure is not sufficient to realize low noise in APDs. Taking a closer look at the bandstructures, we observe that the positions in energy of the minigaps with respect to the valence band edge differ from one material to another. Additionally, the minigap sizes of the different alloys vary in magnitude. A complementary effect of the minigap size is the flattening of the energy bands, *i.e.*, a large minigap size results in flatter bands around the gap. This, in turn, results in an increased effective mass which tends to inhibit carrier transport. Table 3.1 lists the energy location of the minigap with respect to the valence band edge ΔE_b, the minigap size ΔE_m, the light-hole (LH) and split-off (SO) band effective masses and the energy difference between the LH and SO bands ΔE_{LS} at the Γ point for the digital alloys studied.

We can see in the table that there are significant variations in minigap size and position between different materials. At first glance, there seems to be no direct correlation between these variations and the excess noise, prompting us to do added transport analyses. Under high electric field, a carrier must gain at least the threshold energy, E_{TH}, in order to impact ionize. Typically, E_{TH} is assumed to be approximately 1.5 times the material bandgap, E_G. Thus, in the presence of minigaps, electrons/holes must bypass these gaps by some

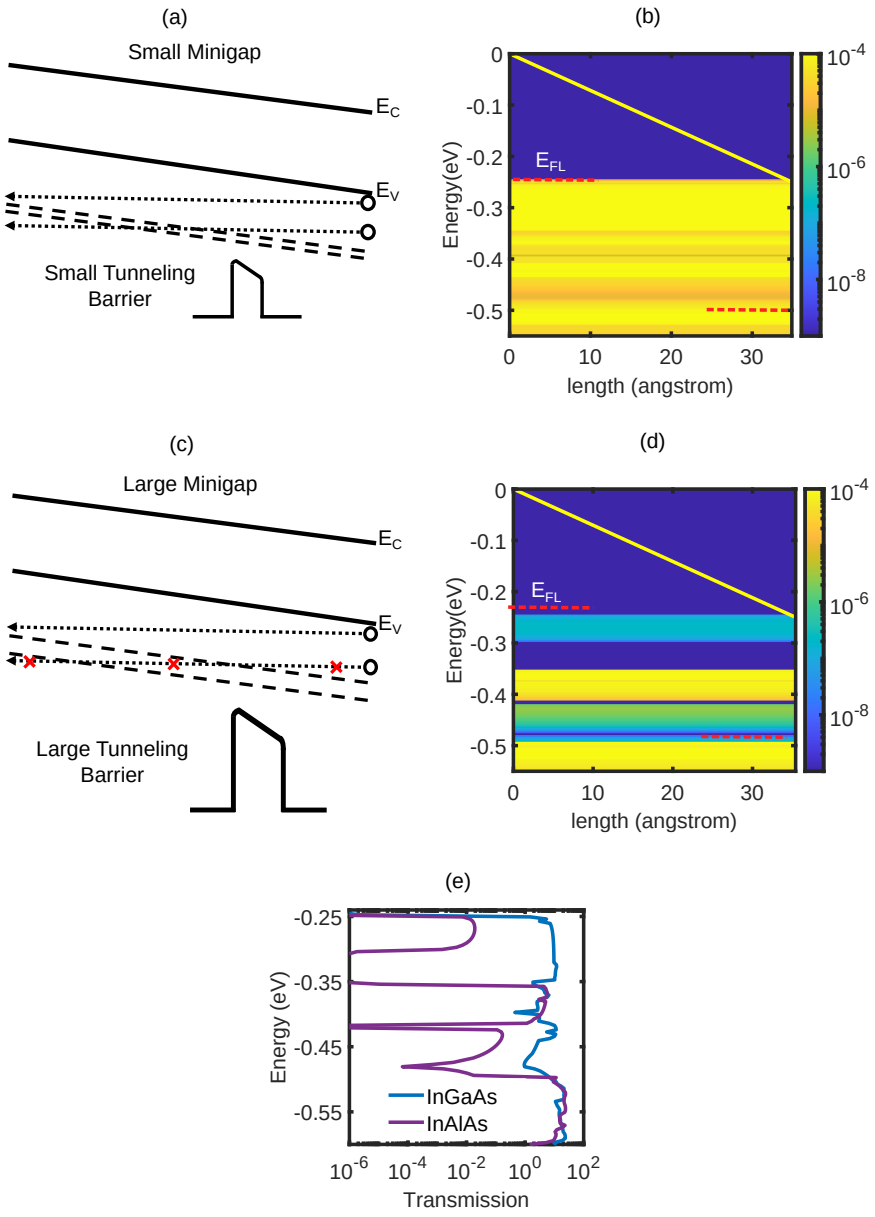

Figure 3.7: Small minigaps in the valence band, as shown in (a), create a small tunneling barrier which can be overcome by holes with low mass. The spectral current density for InGaAs, which has a small minigap and small LH effective mass, is shown in (b). The current spectrum for InGaAs in the Fermi window is continuous. The creation of a large tunneling barrier by a larger minigap is shown in (c). This barrier prevents hole transmission. InAlAs has a larger minigap and LH m^*. Regions of low current density is observed within the Fermi window in the InAlAs spectral current density in (d). The large minigap in InAlAs results in reduced transmission as shown in the $T(E)$ vs. (E) plot of (e). The simulations for (b), (d) and (e) were conducted under bias of $V = 0.25V$.

transport mechanism in order to gain energy equivalent to E_{TH}. The two such major transport mechanisms are quantum mechanical tunneling and optical phonon scattering. Our transport study must incorporate these two mechanisms to understand the effectiveness of minigaps on the APD excess noise.

We employ the NEGF formalism described in Section 3.1.2 to compute the ballistic transmission in the valence band as a function of energy, $T(E)$, dominated by tunneling processes. The effect of different minigap sizes is highlighted in Fig. 3.7. For our simulation, we set the quasi-Fermi level of the left contact at $-qV$ below the valence band edge and the quasi-Fermi level of the right contact at another $-qV$ below. This is done in order to only observe the intraband tunneling inside the valence band, which is responsible for overcoming minigaps under ballistic conditions. In Fig. 3.7 (a), We demonstrate that a small minigap in the valence band creates a small tunneling barrier for the holes. A hole with a small enough effective mass will be able to tunnel across this barrier and render it ineffective. That is the case for InGaAs, which has a LH effective mass of $0.13m_0$ and $\Delta E_m = 0.03eV$. The spectral current density for InGaAs under a bias $V = 0.25V$ is shown in Fig. 3.7 (b). We observe that the current spectrum in the valence band is continuous in the Fermi energy window and there is no drop in transmission due to the minigap. For a large minigap, the holes encounter a larger tunneling barrier, as shown in Fig. 3.7 (c), preventing them from gaining the threshold energy E_{TH} for secondary impact ionization. This case is operational in InAlAs digital alloys, as shown in the spectral density plot in Fig. 3.7 (d). InAlAs has a minigap size of $0.12eV$ and LH effective mass of $0.4m_0$. Within the Fermi window, we see that there are regions with extremely low current due to low tunneling probability across the minigap. This is further demonstrated by the $T(E)$ vs. E plot in Fig. 3.7 (e). Here, it is observed that there are regions of low transmission for InAlAs whereas the InGaAs transmission is continuous. This signifies that the minigaps in the InAlAs valence band are large enough to prevent holes from gaining kinetic energy, resulting in a low hole ionization coefficient.

In order to investigate the role of minigaps in the remaining digital alloys, we look at

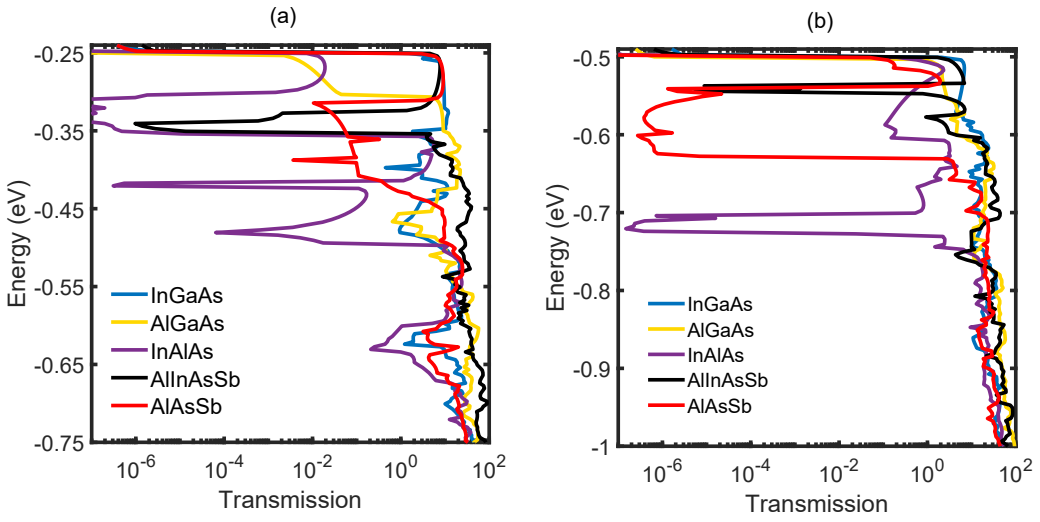

Figure 3.8: The Transmission $T(E)$ vs. Energy E for all the digital alloys at $V = 0.25V$ in (a) and $V = 0.5V$ in (b). A 21×21 grid for transverse wavevectors is used.

the transmission vs. energy plots for all the alloys. The $T(E)$ vs. E characteristics for the five digital alloys are shown in Fig. 3.8 for two bias conditions, (a) $V = 0.25V$ and (b) $V = 0.5V$. We use a 21×21 grid for the transverse wavevectors (k_x, k_y) within the first Brillouin zone. For this simulation, the structure length for InGaAs, AlGaAs, InAlAs and AlAsSb is considered to be two periods. For AlInAsSb, we consider one period length. This allows us to keep the structure lengths as close as possible. We consider lengths of $3.48nm$ InGaAs, $3.42nm$ AlGaAs, $3.54nm$ InAlAs, $3.06nm$ AlInAsSb and $3.08nm$ AlAsSb channels. The channel sizes chosen are small compared to actual device lengths in order to keep the computation tractable. For both the bias conditions in Fig. 3.8, we see there are energy ranges for InAlAs, AlInAsSb and AlAsSb in which the transmission probability drops drastically. This low tunneling probability can be attributed to two factors. The first factor is the presence of a sizeable minigap in all directions in the material bandstructure. The other contributing factor is the separation between the LH and SO bands. This factor is partly responsible for the low transmission regions in AlInAsSb and AlAsSb, whose minigap sizes (from Table 3.1) are smaller than InAlAs but also demonstrate lower excess noise. InGaAs and AlGaAs do not have any large drop in transmission for both biases. This characteristic implies that either the minigap size is too small to affect the carrier transport like in InGaAs

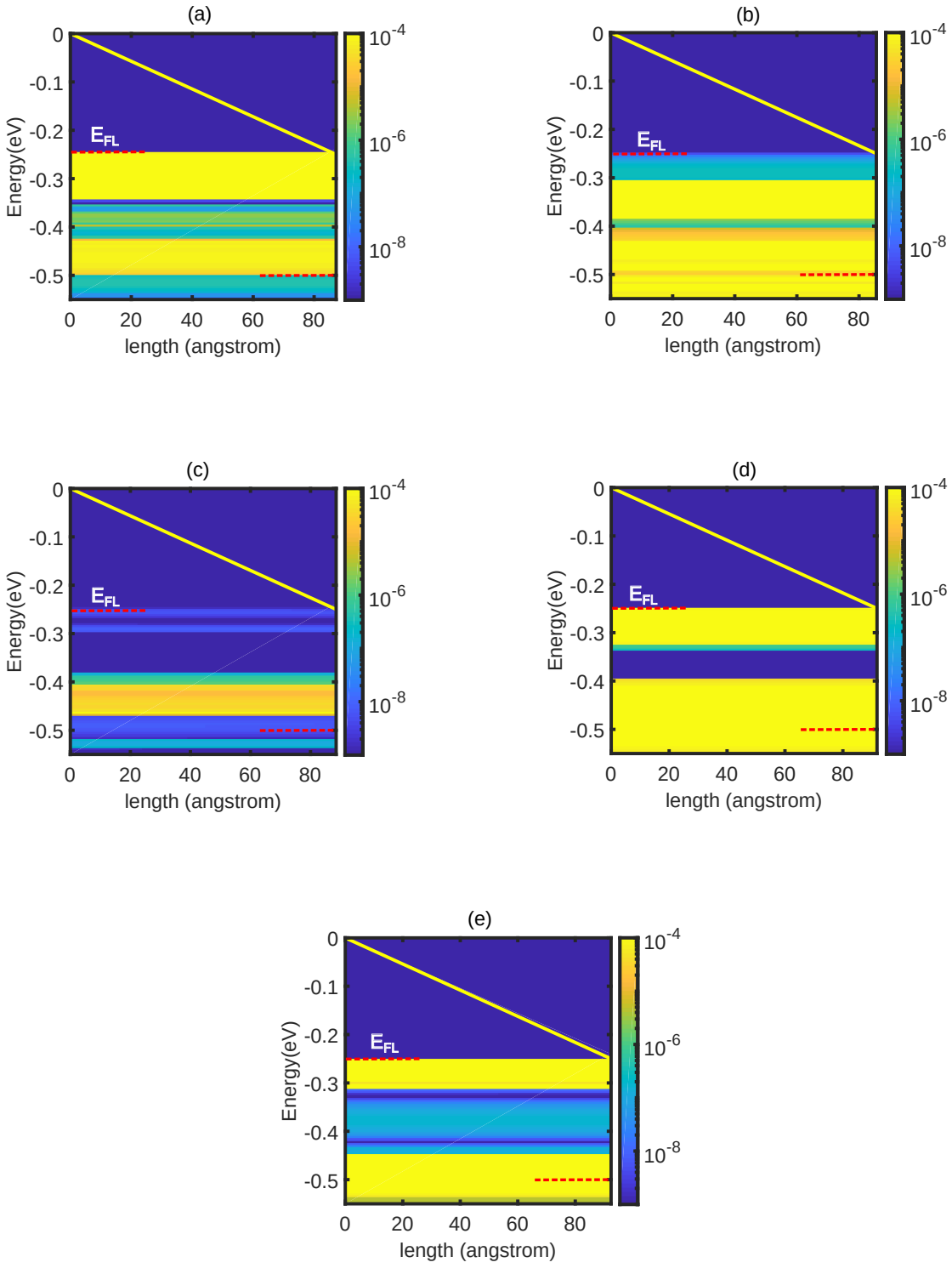

Figure 3.9: Energy resolved current spectral density in the valence band for (a) InGaAs, (b) AlGaAs, (c) InAlAs, (d) AlInAsSb and (e) AlAsSb. The bias for the simulation is set to $V = 0.25V$ and total period length is 30 monolayers.

or there is no minigap at all as in AlGaAs.

For further confirmation of these observations, we compute the spectral current density for the case of constant total period length of all the structures. The period size of each unit cell stays the same, but the number of unit cells is increased to make the total period length the same for all alloys. We consider the case with a total period of 30MLs and voltage bias of $0.25V$. The current spectral density plots for the five digital alloys using a 15×15 transverse wavevector grid are shown in Fig. 3.9. Smaller number of grid points are used here to save computation time. In the figure, a very small minigap is observed for InGaAs within the Fermi window and a continuous spectrum is seen for AlGaAs. Regions of low transmission/current are observed for InAlAs, AlInAsSb and AlAsSb. These observations are consistent with our previous calculations. We can thus infer that at least under fully coherent transport including tunneling, holes will not be able to gain sufficient kinetic energy to achieve impact ionization.

Besides tunneling processes, it is possible for carriers to jump across energy gaps through inelastic scattering. In APDs, the dominant scattering mechanism is intervalley optical phonon scattering. Using the BTE model described in Section 3.1.3, the effect of phonon scattering in digital alloys is studied. The carrier mobilities and optical phonon energies of the binary constituents of the alloys used in the BTE simulations are listed in Table 3.2. An effective scattering strength $H_{\vec{p},\vec{p}'}$ is obtained from the mobility values as described in Section 3.1.3. For our BTE simulations, we use the heavy-hole effective masses outlined in Table. 3.1. We compute the carrier density distribution in the valence band under a high electric field of $1MV/cm$, by solving the three-dimensional Boltzmann equation with the

Material	μ_h (cm^2/Vs)	E_{opt} (meV)
InAs	500	30
AlAs	200	50
GaAs	400	35
AlSb	400	42

Table 3.2: Electron/hole mobilities and optical phonon energies of binary compounds that form the digital alloys [99, 100].

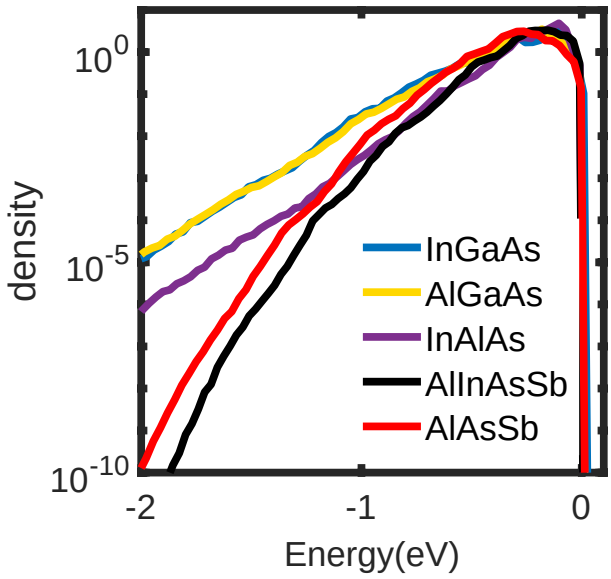

Figure 3.10: Carrier density distribution vs. Energy for the valence band in the presence of optical phonon scattering computed using BTE simulation. InAlAs, AlInAsSb and AlAsSb have lower occupation probability compared to InGaAs and AlGaAs. This prevents holes from gaining the ionization threshold energy.

entire set of tight binding energy bands within the Brillouin zone of the digital alloy. The optical phonon energy and mobilities of each alloy are taken to be the average of the binary constituent optical phonon energies and their mobilities. The energy resolved carrier density distribution for all the alloys is shown in Fig. 3.10. The valence band plot in Fig. 3.10 shows that the occupation probability for InAlAs, AlInAsSb and AlAsAsb is lower than the other two alloys at high energies. The optical phonon energies of these alloys are not sufficiently large to overcome their minigaps and thus prevent holes from ramping their kinetic energies up to E_{TH}.

The top few valence bands of InGaAs are shown on the left side of Fig. 3.11(a) and the valence band carrier density distribution is projected onto the bottom. The bands are inverted for a better view. For a clearer understanding, the InGaAs carrier density distribution contour is also shown on the right. The valence band carrier distributions for the other alloys are shown in Fig. 3.11(b) AlGaAs, (c) InAlAs, (d) AlInAsSb and (e) AlAsSb. By studying the contours of each material, we observe that the densities for InAlAs, AlInAsSb and AlAsSb

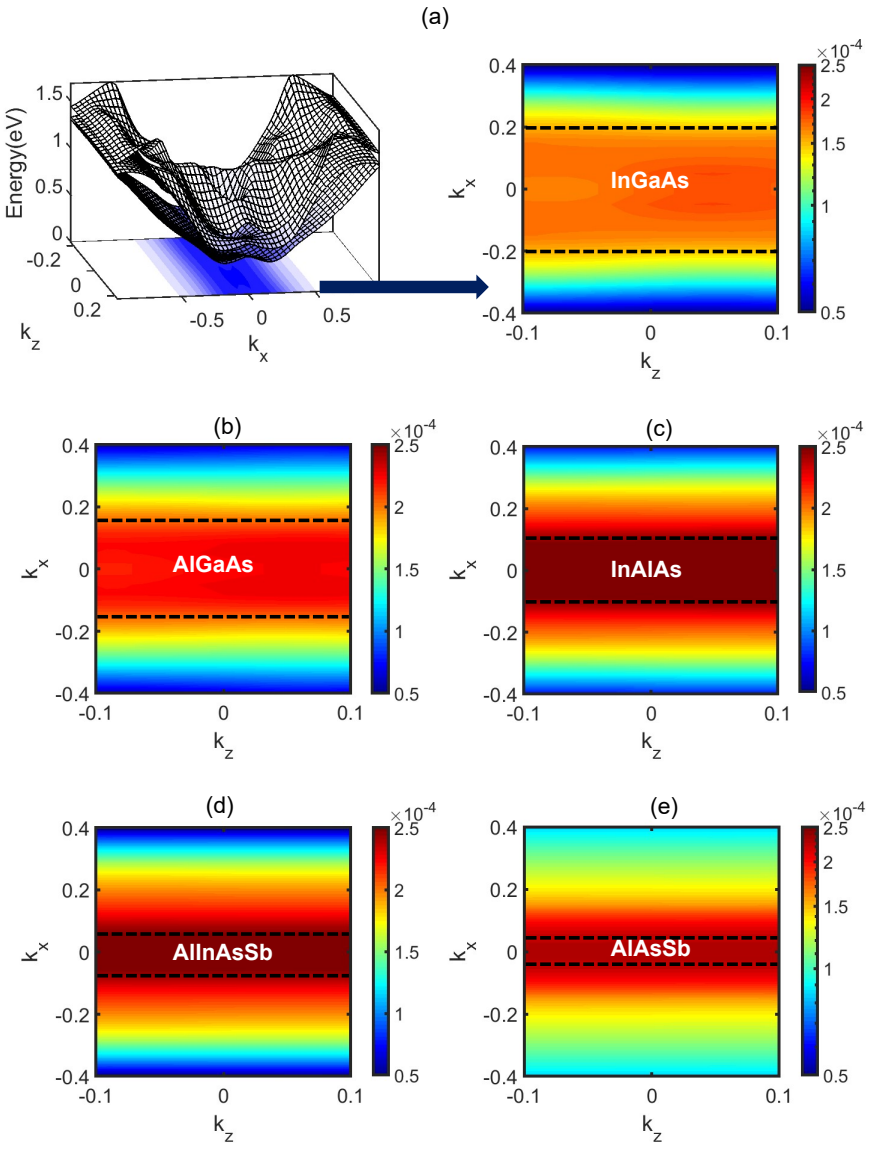

Figure 3.11: Carrier density distribution for (a) InGaAs, (b) AlGaAs, (c) InAlAs, (d) AlInAsSb and (e) AlAsSb.

are more localized compared to thåt of AlGaAs and InGaAs. This is once again consistent with the lower hole impact ionization of InAlAs, AlInAsSb and AlAsSb.

For InGaAs and AlGaAs, the bandwidths are large enough to allow both holes and electrons to reach E_{TH} easily. The resulting values of k for these materials are quite high. Correspondingly, these two alloys have higher excess noise. In contrast, in InAlAs, AlInAsSb and AlAsSb, it is easy for electrons to reach the threshold energy, but the holes are confined close to the valence band edge. This results in asymmetric ionization coefficients, which give

a low k, leading in turn to low excess noise.

Armed with these results, we attempt to paint a clearer picture of how the minigaps and band splitting can reduce the excess noise in APDs. Specifically, we propose a set of empirical inequalities that can be used to judge the excess noise performance of a digital alloy.

3.3 Empirical Inequalities

Based on our experimental results and theoretical calculations, five inequalities are proposed that use only material parameters like effective mass and minigap size obtained from our material bandstructures as inputs. In this work, the transport is in the [001] direction. Since the minigaps considered lie in the LH band, we use the unfolded LH effective mass value in the $\Gamma - [001]$ direction for the inequalities. The masses are obtained using the relationship $\hbar^2k^2/2m^* = E(1+\alpha E)$ where $\alpha = [(1-m^*/m_0)^2]/E_G$ [101]. In reality, the effective masses are complicated tensors that cannot be included in these empirical inequalities but are captured by the NEGF simulations described in Section. 3.1.2. A digital alloy material should favor low noise if it satisfies the majority of these inequalities. The four main inequalities are:

$$\text{Inequality (1)} \qquad \Delta E_b/E_{TH} << 1$$

$$\text{Inequality (2)} \qquad E_{opt}/\Delta E_m << 1$$

$$\text{Inequality (3)} \qquad exp\left(-\frac{4\sqrt{2m_l}\Delta E_m^{3/2}}{3q\hbar F}\right) << 1$$

$$\text{Inequality. (4)} \qquad exp\left(-\frac{4\sqrt{2m_l}\Delta E_{LS}^{3/2}}{3q\hbar F}\right) << 1$$

Here, ΔE_b represents the energy difference between the VB maximum and the first minigap edge in the VB, E_{opt} is the optical phonon energy and ΔE_m gives the size of the minigap.

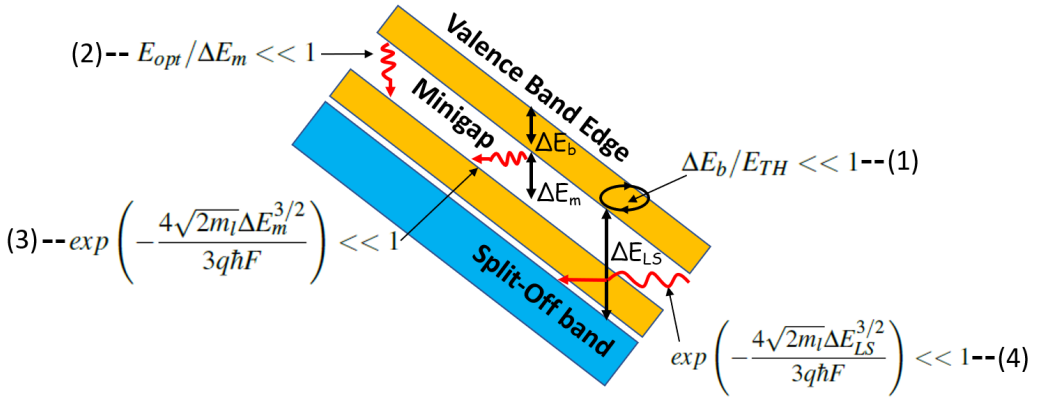

Figure 3.12: Criteria for designing low noise digital alloy APDs. Inequality (1) states that the bandwidth to the first minigap is lower than the ionization threshold energy. Inequality (2) asserts that the optical phonon energy has to be less than the minigap size. The tunneling probability for holes to jump across the minigap or from the light-hole band to the split-off band must be low. These are described by Inequality (3) and Inequality (4).

The longitudinal effective mass of the band in which the minigap exists is represented by m_l. ΔE_{LS} signifies the energy difference between the LH and SO bands at the Γ point. A pictorial view of the different energy differences and inequalities mentioned above is shown in Fig. 3.12.

The first inequality, Inequality (1), states that the energy bandwidth ΔE_b must be less than the ionization threshold energy E_{TH}. This means a carrier cannot gain sufficient kinetic energy to impact ionize before reaching the minigap. When a carrier reaches a minigap, it faces a barrier (Fig. 3.12), which it can overcome by phonon scattering or quantum tunneling. Inequality (2) sets the condition for phonon scattering across the minigap. If the E_{opt} of the material is less than ΔE_m, then the phonon scattering of the carriers across the minigap is inhibited because carriers cannot gain sufficient energy to jump across the gap. It is possible for the carrier to still overcome the minigap by tunneling, and the condition for that is given in Inequality (3), in terms of the tunneling probability across the minigap under the influence of an electric field. To compute the tunneling probability, we consider a triangular barrier in the minigap region and use the well-known Fowler-Nordheim equation. Together Inequalities (2) and (3) give the effectiveness of the minigap in limiting hole ionization in

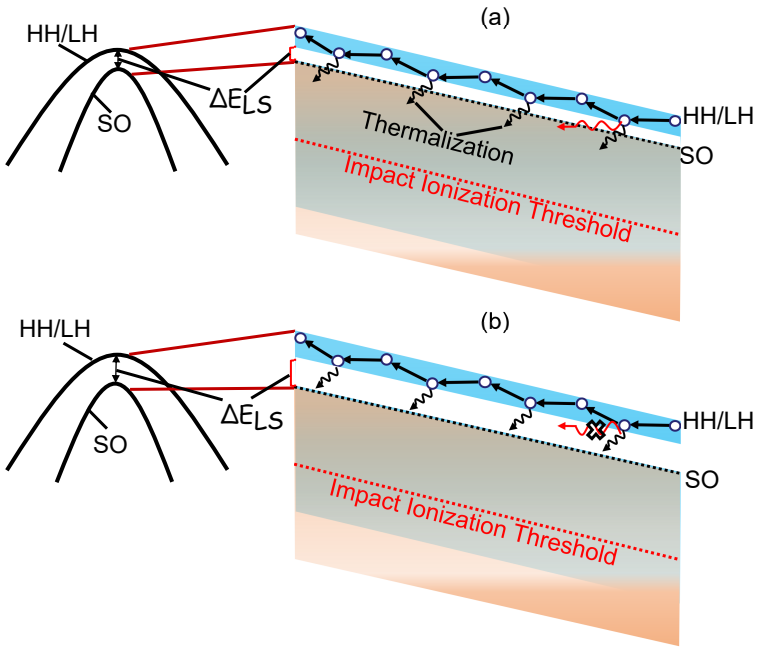

Figure 3.13: Effect of spin-orbit coupling on LH/SO separation. (a) Weak coupling results in small ΔE_{LS} and (b) strong coupling results in large ΔE_{LS}.

digital alloys.

Electron injected digital alloys can, in fact, achieve low noise even in the absence of minigaps, for instance, in a material with a large separation ΔE_{LS} between the LH and SO bands, like AlAsSb. Holes within HH/LH bands are limited near the valence band edge by thermalization (hole-phonon scattering) due to the heavy effective masses in these bands, preventing them from reaching the ionization threshold energy within the band. An alternate pathway to ionization involves the split-off band. Since the split-off band has a low effective mass, holes require much smaller momentum to reach higher energies in this band, so that holes entering this band from HH/LH can quickly gain their ionization threshold energy. The separation between HH/LH and SO bands is controlled by spin-orbit coupling, as shown in Fig. 3.13. Strong spin-orbit coupling due to the inclusion of heavy elements, like antimony or bismuth, can increase the separation ΔE_{LS}, as shown in Fig. 3.13(b). When ΔE_{LS} is large, it becomes very difficult for holes to reach the threshold energy. Inequality (4) is accordingly important for APDs in which electron impact ionization is the dominant process and is a measure of hole tunneling from the light-hole to the split-off band.

An inherent fifth inequality, satisfied by these five alloys, is

$$E_{SC} < E_{TH} \tag{3.12}$$

E_{SC} is the energy gained by a hole between successive phonon scattering events, expressed as $E_{SC} = \lambda_{mfp}/F$. The z-directed mean free path, $\lambda_{mfp} = v_{sat}\tau_{SC}/2$, where v_{sat} is the saturation velocity and τ_{SC} is the scattering lifetime. E_{SC} values of the five alloys at electric fields of $100kV/cm$ and $500kV/cm$ are given in Table. 3.3. We extract τ_{SC} for an alloy by assuming a virtual crystal approximation of the component binary alloy scattering times. τ_{SC} values for InAs, GaAs, AlAs and AlSb are $0.08ps$, 0.09, $0.08ps$ and $0.11ps$, respectively [102]. A similar average is done for the ternary alloy saturation velocities. Due to the unavailability of AlSb v_{sat}, InAs v_{sat} is used for AlInAsSb and AlAs v_{sat} for AlAsSb. InAs, GaAs and AlAs v_{sat} values used are $5 \times 10^4 m/s$, $9 \times 10^4 m/s$ and $8 \times 10^4 m/s$, respectively [103].

In order to validate these inequalities as design criteria, we apply them to the set of digital alloys mentioned in this paper. We consider a high electric field of $1MV/cm$ for Inequalities (3) and (4). The values of the left sides of the inequalities for the five alloys- InGaAs, AlGaAs, InAlAs, AlInAsSb and AlAsSb, are given in the first four columns of Table 3.4, while the measured k is provided as reference in columns 5 and 6. The table cells are colored green or red. Green cells aid in noise suppression (left sides of the inequalities are relatively small), and

Material	E_{SC} (eV) at $100kV/cm$	E_{SC} (eV) at $500kV/cm$	E_{TH} (eV)
InGaAs	0.029	0.149	0.95
AlGaAs	0.036	0.181	3.91
InAlAs	0.028	0.138	1.85
AlInAsSb	0.024	0.119	1.79
AlAsSb	0.038	0.19	2.4

Table 3.3: E_{SC} values at $F = 100kV/cm$ and $F = 500kV/cm$, and E_{TH} of the five alloys. For a material with equal conduction and valence band effective masses, considering parabolic bands, the threshold energy $E_{TH} = 1.5E_G$ [104]. The same assumption is made here for the fifth inequality as this standard practice in the APD literature.

Material	Inq. 1	Inq. 2	Inq. 3	Inq. 4	k (DA)	k (RA)
InGaAs	0.38	1.08	0.88	0.006	0.3 [78]	0.5 [78]
AlGaAs	1	∞	1	7.2×10^{-4}	0.1 [79]	0.2 [79]
InAlAs	0.16	0.33	0.17	5.6×10^{-4}	0.05 [30]	0.2 [30]
AlInAsSb	0.17	0.59	0.53	7.9×10^{-7}	0.01 [31]	0.018 [34]
AlAsSb	0.22	0.45	0.3	3.4×10^{-7}	0.005 [32]	0.05 [105]

Table 3.4: Suitability of digital alloys for attaining low noise is judged using the proposed inequalities. Here, the color green means beneficial for low noise and red indicates it is detrimental. The impact of the inequality in determining the experimentally determined ionization coefficient ratio k of the material is depicted by the color shades. A darker shade indicates that the inequality has a greater impact on the value of k. The experimental random alloy k's of the five alloys are given in column 6.

red is detrimental to reducing noise (left sides are larger and corresponding inequalities not satisfied). Additionally, the color intensities highlight the strength of that inequality (how far the left side is from equality with the right side). A lighter shade represents a smaller impact, while a darker shade means that condition has a greater effect on the impact ionization noise. For example, in the case of InGaAs, Inequality (1) is shaded light green which means it does not affect noise performance significantly. However, the remaining inequalities for InGaAs are shaded dark red, indicating their key role in the high noise and hence high k of InGaAs. The inequalities for AlGaAs, which has a slightly lower k, have a lighter shade of red. There are no minigaps for AlGaAs in the light-hole band. There is a minigap in the SO band of AlGaAs which is very deep in the valence band and there are other available states at that energy. Thus, holes can gain sufficient momentum to jump to other bands and bypass the minigap. So, we consider $\Delta E_m = 0$ for it. We accordingly expect that AlGaAs has a lower noise. However, since the LH effective mass for AlGaAs is greater than InGaAs, it has lower hole impact ionization and thus lower noise compared to InGaAs. The remaining alloys have significantly lower noise compared to these two.

The boxes for InAlAs, AlInAsSb and AlAsSb are all green. This means these three alloys are quite favorable for attaining low excess performance. InAlAs has a minigap size $\Delta E_m = 0.12eV$ which is larger than its optical phonon energy. It also has a large LH effective

mass which prevents quantum tunneling across the minigap, as well as the LH-SO separation ΔE_{LS} which is comparable to that of AlGaAs and InGaAs. AlInAsSb has a low value for Inequality (1), so that box is shaded dark green. However, for Inequalities (2) and (3), the values for AlInAsSb are higher than that of InAlAs and are thus shaded in a lighter color. AlInAsSb has a larger LH-SO separation than InAlAs and hence its Inequality (4) has a darker shade. In AlAsSb, the values for Inequalities (1)-(3) have medium shades as they lie between the maximum and minimum values in each of these columns for the corresponding inequalities. However, AlAsSb has a large $\Delta E_{LS} = 0.54eV$, so its Inequality (4) is shaded dark green. Based on the inequality values, it would seem InAlAs would have the lowest noise since it has the darkest shades. However, looking at the Inequality (4) values for these three materials, we can infer that the LH-SO separation plays a critical role in reducing noise. Here, AlAsSb has the lowest $k = 0.005$ and also the largest ΔE_{LS}. On the contrary, InAlAs has the highest $k = 0.1$ and the smallest ΔE_{LS}. Finally, inequality 5, discussed in the context of split-off states (Eq. 3.12), is trivially satisfied by all five studied alloys. While important, it is thus not tabulated here, as it does not alter the status quo.

In short, the values of the inequalities in Table 3.4 give a fairly good understanding of the excess noise performance of the set of digital alloys considered in this paper. They can potentially serve as empirical design criteria for judging new digital alloys in consideration as potential material candidates for digital alloy superlattice APDs. Based on the inequalities we can say that materials with large minigaps, heavy effective mass and large LH/SO offsets are suitable for low noise APDs.

3.4 Physics of Low Noise Quaternary Random Alloy APDs

Thus far, it was believed that digitization of the III-V alloys resulted in their low excess noise. However, there has been recent experimental observations of low noise AlGaAsSb

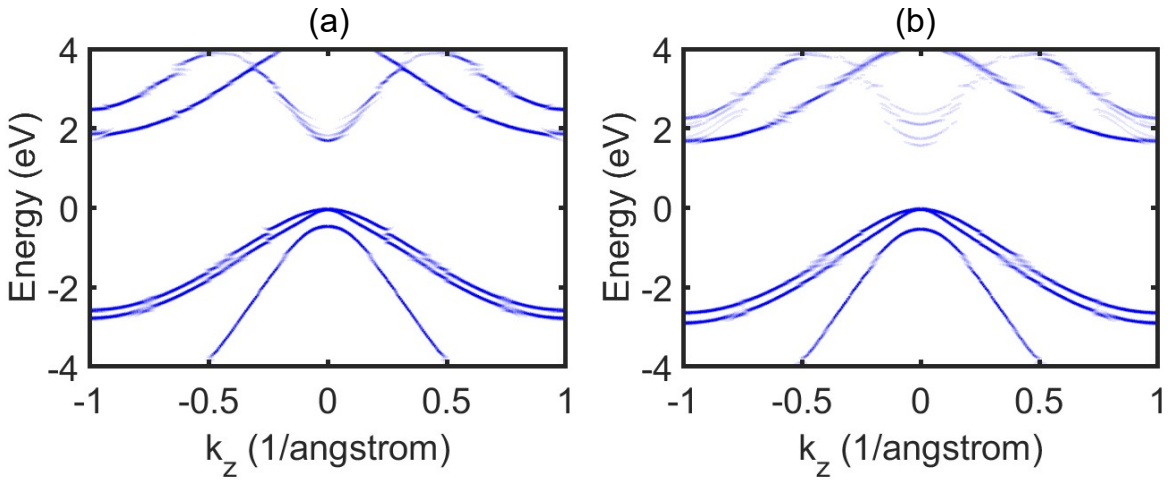

Figure 3.14: Unfolded bandstructure of (a) $Al_{0.85}Ga_{0.15}As_{0.56}Sb_{0.44}$ and (b)$Al_{0.79}In_{0.21}As_{0.74}Sb_{0.26}$ random alloys.

($k = 0.01 - 0.04$) and AlInAsSb ($k = 0.01$) random alloy (RA) APDs [33, 34]. Low noise AlGaAsSb digital alloy APD has also been reported [35]. The noise performance of these APDs is comparable to AlInAsSb and AlAsSb digital alloy (DA) APDs. Hole impact ionization in these alloys are suppressed, leading to the low k values. In this section, we look at the bandstructure of these random alloys to investigate the origin of their low k values.

The unfolded bandstructures of $Al_{0.85}Ga_{0.15}As_{0.56}Sb_{0.44}$ and $Al_{0.79}In_{0.21}As_{0.74}Sb_{0.26}$ random alloys are depicted in Fig. 3.14. The calculated bandgaps of these alloys are $1.59eV$ and

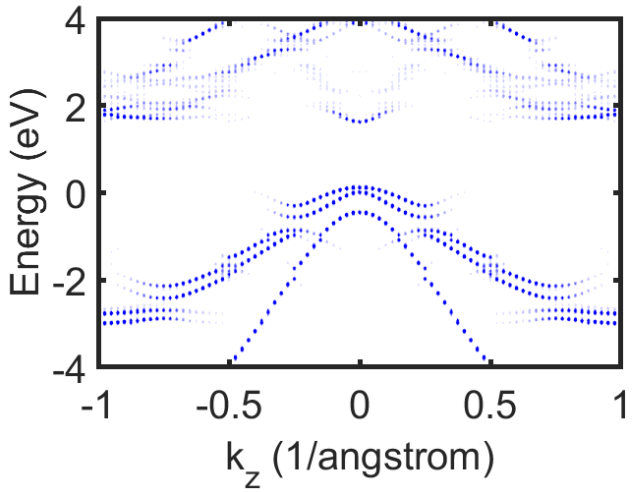

Figure 3.15: Unfolded bandstructure of 4ML $Al_{0.85}Ga_{0.15}As_{0.56}Sb_{0.44}$ digital alloy.

$1.73eV$, respectively. The HH effective masses in the $Al_{0.85}Ga_{0.15}As_{0.56}Sb_{0.44}$ RA is $0.4m_0$ and for $Al_{0.79}In_{0.21}As_{0.74}Sb_{0.26}$ is $0.42m_0$. There are no minigaps present in the valence band of these random alloys and thus cannot be the root for the low hole impact ionization. Therefore, we extracted the LH/SO offset for these two alloys. For $Al_{0.85}Ga_{0.15}As_{0.56}Sb_{0.44}$, we got $\Delta E_{LS} = 0.5eV$ and for $Al_{0.79}In_{0.21}As_{0.74}Sb_{0.26}$ the offset $\Delta E_{LS} = 0.44eV$. Both the values are larger than the ternary digital alloy LH/SO offsets given in Table 3.1. The AlInAsSb RA LH/SO offset is comparable to the AlInAsSb DA LH/SO offset. Thus, this large LH/SO offset and large HH effective mass prevent holes from gaining kinetic energy in the $Al_{0.79}In_{0.21}As_{0.74}Sb_{0.26}$ RA. We computed the bandstructure of a 4ML $Al_{0.85}Ga_{0.15}As_{0.56}Sb_{0.44}$ DA, shown in Fig. 3.15, to extract its LH/SO offset. For this alloy we get $\Delta E_{LS} = 0.44eV$. The $Al_{0.85}Ga_{0.15}As_{0.56}Sb_{0.44}$ RA offset is comparatively larger than it's DA counterpart which also exhibits low noise. Also, minigaps are present in the DA bandstructure. Thus, the low noise in the $Al_{0.85}Ga_{0.15}As_{0.56}Sb_{0.44}$ RA can be attributed to the LH/SO offset and large HH mass. A combination of minigaps, LH/SO offset and higher effective mass results in the low

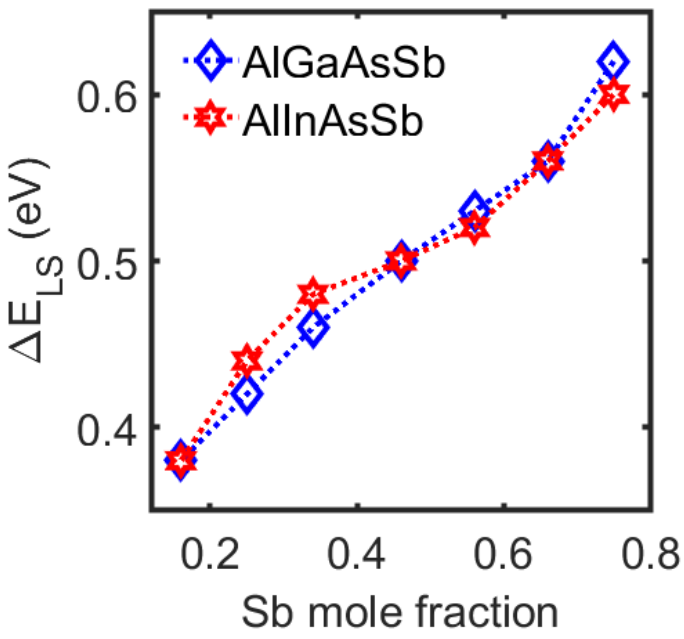

Figure 3.16: LH/SO offset as a function of Sb mole fraction for AlGaAsSb and AlInAsSb random alloys.

noise of the $Al_{0.85}Ga_{0.15}As_{0.56}Sb_{0.44}$ DA, which is consistent with our earlier observations.

We believe that the large LH/SO offset seen in the AlGaAsSb and AInAsSb RAs result from the inclusion of Sb atoms which have high spin-orbit coupling. We calculated the LH/SO offset for the two alloys for different compositions of antimonide. The LH/SO offset as a function of Sb mole fraction is shown in Fig. 3.16. We observe that ΔE_{LS} increases with the Sb mole fraction for both the alloys. For low Sb mole fractions, the ΔE_{LS} of AlGaAsSb and AlInAsSb approach the values of AlGaAs and InAlAs, respectively. This shows that the origin of the increased LH/SO offset is due to the presence of the heavier Sb atoms. It has been demonstrated that alloying GaAs with Bismuth leads to reduced hole impact ionization [106]. Based on the observations of this paper and our above results, we can postulate that introduction of heavy atoms with strong spin-orbit coupling in the alloys result in lower hole impact ionization. Consequently, it is then possible to get a lower excess noise performance in APDs using such alloys. A summary of the k values and key valence band features of the various III-V random and digital alloys studied in this dissertation is provided in Table 3.5.

Material	**Alloy Type**	k **value**	**VB features**
InGaAs	RA	0.5	low m_v^*
InGaAs	DA	0.3	tiny minigaps+low m_v^*
AlGaAs	RA	0.2	high m_v^*
AlGaAs	DA	0.1	high m_v^*
InAlAs	RA	0.2	high m_v^*
InAlAs	DA	0.05	minigaps+high m_v^*
AlAsSb	RA	0.05	high m_v^*+large LH/SO offset
AlAsSb	DA	0.005	minigaps+ high m_v^*+ large LH/SO offset
AlInAsSb	RA	0.018	high m_v^*+large LH/SO offset
AlInAsSb	DA	0.01	small minigaps+ high m_v^*+ large LH/SO offset
AlGaAsSb	RA	0.01-0.04	high m_v^*+large LH/SO offset
AlGaAsSb	DA	0.01	minigaps+ high m_v^*+ large LH/SO offset

Table 3.5: Summary of k values and key valence band features for various III-V random and digital alloys.

3.5 Summary

This chapter investigates the origin of low excess noise in III-V APDs using state-of-the-art bandstructure calculations. Furthermore, the digital alloy valence band carrier transport is studied using NEGF and BTE formalisms. Based on our simulation results, we explain how minigaps and LH/SO offset impede hole impact ionization in APDs and improve their excess noise performance. When these gaps/offsets are sufficiently large they cannot be bridged across by quantum tunneling or phonon scattering processes. Furthermore, we propose five inequalities as empirical design criteria for digital alloys with low noise performance capabilities. Material parameters calculated computationally are used as inputs for these. We validate these criteria by explaining the excess noise performance of several experimentally fabricated digital alloy APDs. The design criteria can be used to computationally design new digital alloy structures and benchmark them before actually fabricating these. Furthermore, we explain the origin of low excess noise in quaternary random alloy APDs using bandstructure calculations.

Key Contributions:

- Developed quasi-1D ballistic NEGF model for III-V digital alloys.
- Demonstrated that a combination of minigaps, large valence band effective mass and LH/SO offset lead to excess low noise in III-V APDs.
- Developed empirical design criteria for III-V APDs.

Chapter 4

Strain Modulated Valence Band Engineering in III-V Digital Alloys

The low k in the few digital alloys mentioned earlier can be ascribed to multiple factors - the generation of 'minigaps' in the material valence band, a corresponding enhanced valence band effective mass and finally, a large separation between the light-hole and split-off bands [29, 36, 37]. These properties prevent holes from gaining energy, keeping them localized near the valence band edge and resulting in a low hole ionization rate. However, in these electron injected APDs, electrons in the conduction band can easily move to higher energies, bypassing conduction band minigaps, in order to impact ionize due to their low effective mass. This leads to a notable increase in the electron ionization rate.

Minigaps are seen to arise naturally in the first-principles unfolded bandstructures calculated for the superlattice stack. However, their chemical origin is not well understood and requires an in-depth analysis. While the presence of minigaps is not a necessary condition for high photogain with low excess noise, it may well prove to be a sufficient condition in many cases. It is thus useful to identify ways to engineer such minigaps deterministically with various design knobs, such as alloying and strain.

In this chapter, we use a simple sp^3 tight binding model to illustrate how strain alters the

bonding chemistry in APD digital alloy materials and plays a crucial role in the formation of minigaps. We then employ the more elaborate EDTB model [62, 63] with band unfolding techniques [68–70] to investigate the role of strain in the formation and modulation of these digital alloy minigaps. The model captures the short range atomistic modifications at the hetero-interfaces, and long-range band modulation by the superlattice potential. Furthermore, we study the relationship between biaxial strain and minigap size and their overall impact on carrier transport. When it comes to simulating carrier transport, it is worth keeping in mind that there are two primary mechanisms by which carriers can bypass the minigaps, namely, quantum tunneling and optical phonon scattering. The impact of tunneling through a minigap is readily captured by computing the ballistic transmission in 3-D using NEGF formalism described in the previous chapter. We study the effect of phonon scattering in the digital alloys using a multi-band Boltzmann transport model that outputs the energy resolved carrier occupation probability, which we calculate to explore the effect of minigaps on phonon scattering. This study provides a convenient design principle towards efficient photodetectors, and for overall tunability of electron wavefunction in digital superlattices. This chapter is reproduced from Ref. [107] co-authored with Y. Tan, J. Zheng, J. C. Campbell and A. W. Ghosh.

4.1 Formulation of Theory

Previous studies of digital alloys like InAlAs, AlInAsSb and AlAsSb [29, 30, 36, 37] have demonstrated that valence band minigaps present in the material bandstructure play a part in reducing excess noise by limiting hole carrier transport. However, the role of minigap is firmly established for one material combination, InAlAs. For the other materials either a systematic experimental comparison between digital and random alloy superlattices does not exist, or when it does, the random shows low noise as well and is attributed to an energy separated split-off band [81]. Nonetheless, a deterministic creation of a strong minigap can

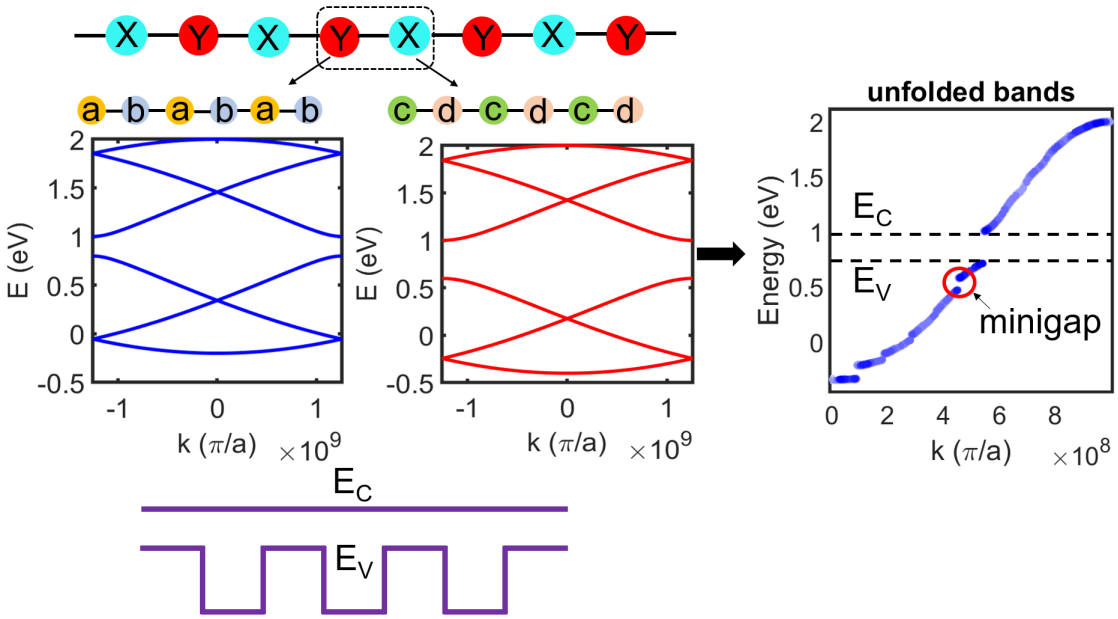

Figure 4.1: In a toy bandstructure model, we consider an unit cell (dotted box) consisting two arbitrary binary materials. By adjusting onsite energies and hopping parameters a material system with zero conduction band offset and large valence band offset is created. The large valence band offset results in minigaps in the valence band (within red circle) as shown in the unfolded band structure.

significantly aid APD gain by suppressing one carrier type. In this section, we explore the formation of these gaps using a one dimensional simple 'toy' model.

We consider an arbitrary alloy consisting of two materials X and Y stacked alternately like a digital alloy, as shown in Fig. 4.1. Each of these materials is essentially a dimer consisting of a set of two atoms. For X, the component atoms are a and b, and for Y they are c and d. The resulting Hamiltonian of the unit cell for this material then looks like:

$$H = \begin{pmatrix} \alpha_X & -\beta_X & & & & -\gamma_{YX1} \\ -\beta_X^\dagger & \alpha_X & -\beta_X & & & \\ & -\beta_X^\dagger & \alpha_X & -\gamma_{XY1} & & \\ & & -\gamma_{XY2} & \alpha_Y & -\beta_Y & \\ & & & -\beta_Y^\dagger & \alpha_Y & -\beta_Y \\ -\gamma_{YX2} & & & & -\beta_Y^\dagger & \alpha_Y \end{pmatrix} \tag{4.1}$$

where,

$$\alpha_X = \begin{pmatrix} E_X & -t_1 \\ -t_1 & E_X \end{pmatrix}, \beta_X = \begin{pmatrix} 0 & 0 \\ t_2 & 0 \end{pmatrix} \tag{4.2}$$

$$\alpha_Y = \begin{pmatrix} E_Y & -t_3 \\ -t_3 & E_Y \end{pmatrix}, \beta_Y = \begin{pmatrix} 0 & 0 \\ t_4 & 0 \end{pmatrix} \tag{4.3}$$

$$\gamma_{XY1} = \begin{pmatrix} 0 & 0 \\ t_5 & 0 \end{pmatrix}, \gamma_{XY2} = \begin{pmatrix} 0 & t_6 \\ 0 & 0 \end{pmatrix} \tag{4.4}$$

$$\gamma_{YX1} = \begin{pmatrix} 0 & t_7 \\ 0 & 0 \end{pmatrix}, \gamma_{YX2} = \begin{pmatrix} 0 & 0 \\ t_8 & 0 \end{pmatrix} \tag{4.5}$$

For each material, we consider the onsite energies, $E_{X,Y}$ to be constant, while the hopping parameters $t_{1,2,3,4}$ between the dimer elements vary. $t_{5,6,7,8}$ represent the coupling between material X and material Y. Here, we set $E_X = 0.9$, $E_Y = 0.8$, $t_1 = 0.6$, $t_2 = 0.5$, $t_3 = 0.7$, $t_4 = 0.5$, $t_5 = -0.4$, $t_6 = -0.6$, $t_7 = -0.4$ and $t_8 = -0.6$ in eV. These parameter values are chosen such that there is a large valence band offset between X and Y but the conduction band offset is zero, as depicted in Fig. 4.1. The resulting unfolded bandstructure is shown on the right side of the figure. We observe that a clear minigap forms in the valence band (highlighted with a red circle), while correspondingly large minigaps do not arise in the conduction band. This simple example illustrates that sizeable minigaps can be engineered selectively in one band by creating large onsite energy variations in the frontier atomic orbitals that generate that band. We will now explore how such large offsets can be deliberately engineered in the III-V digital alloys using strain.

4.2 Strain Engineered Minigaps in Digital Alloys

In bulk heterojunctions, band discontinuities form at the interface owing to the alignment of Fermi levels of the constituent components, resulting in band offsets. The band offset sizes can be manipulated if the position of band edges can be altered [108]. This is achieved by means of hydostratic pressure [109], applying biaxial strain [110–114] and alloying [115, 116]. In digital alloys, it is biaxial strain that results in the opening of the minigaps, as we will describe next.

It is well known that biaxial strain in semiconductors removes the degeneracy of the valence bands and results in the splitting of the heavy-hole (HH) and light-hole (LH) bands [116, 117].

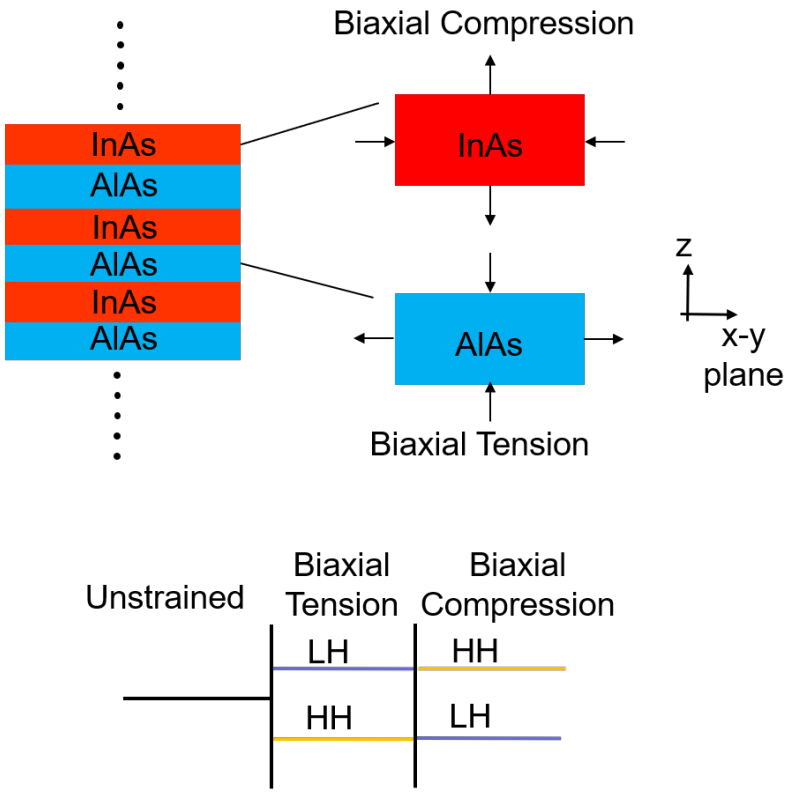

Figure 4.2: InAlAs digital alloy consists of InAs and AlAs grown on InP substrate. Thus, InAs experiences biaxial compression and AlAs experiences biaxial tension. Biaxial strain results in splitting of the HH and LH bands. Since InAs and AlAs experience opposite kinds of strain, their bands move in opposite direction. This results in opening of minigaps in InAlAs.

Let us consider the case of InAlAs digital alloy to understand how minigaps form. In Fig. 4.2, we see that InAlAs DA consists of InAs and AlAs layers stacked alternately. The alloy is grown on an InP substrate having a lattice constant 5.87$\mathring{A}$ [63]. Compared to InP, the lattice constant of InAs at 6.06$\mathring{A}$ is greater, while that of AlAs at 5.66$\mathring{A}$ is smaller. As a result, AlAs experiences biaxial tension in the $x-y$ plane, while InAs undergoes biaxial compression. In the (001) z -direction, InAs undergoes expansion and AlAs undergoes compression. As we will see shortly, biaxial tension results in LH bands moving up and HH moving down in energy, as depicted at the bottom of Fig. 4.2. The opposite happens for biaxial compression. As the bands in the alternately strained layers move in opposite directions, the band offset increases, resulting in the formation of the minigaps.

Fig. 4.3(a), shows the bandstructure of the strained InAs and AlAs (grown on InP substrate) computed with the $sp^3s^*d^5$ EDTB model. We observe a large valence band offset at the Γ point between the strained AlAs and InAs. The unfolded bandstructure of a 6-monolayer InAlAs showing the resulting valence band minigaps, computed with the EDTB model, is depicted in Fig. 4.3(b). The unit cell of the InAlAs DA considered consists of 3ML AlAs and 3ML InAs. In order to comprehend the movement of these bands, a closer look at the orbital chemistry is required.

In a bulk zinc blende semiconductor, each atom is tetrahedrally bonded to four neighboring

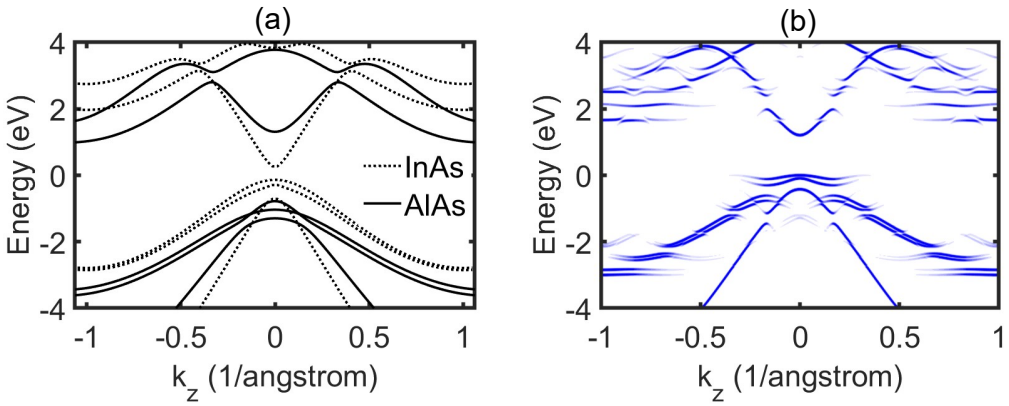

Figure 4.3: (a) Bandstructure of InAs and AlAs grown on InP substrate (b) unfolded bandstructure of 6ML InAlAs with InP as the substrate.

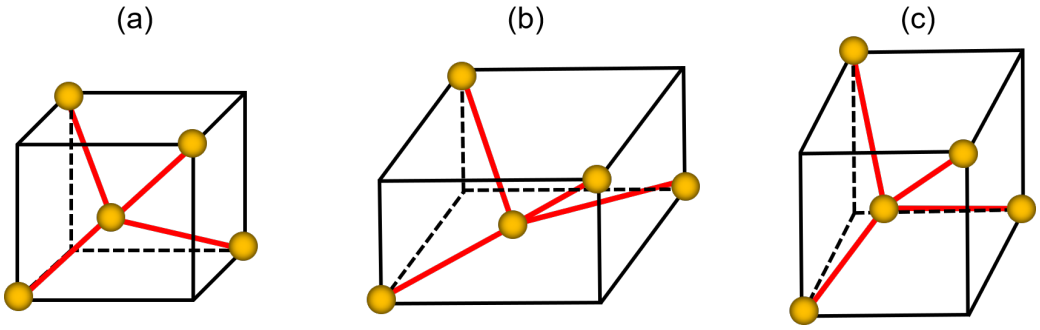

Figure 4.4: (a) Unstrained zinc-blende crystal, (b) under biaxial tension in the $x-y$ plane, and (c) under biaxial compression in the $x-y$ plane.

atoms. The bonds connecting these atoms point toward the $\langle 111 \rangle$ directions of the cube that bounds around the tetrahedron. Every bond consists of 25% contribution each from the s, p_x, p_y and p_z orbitals [117]. Fig. 4.4(a) shows the chemical bonds in the unit cell of an unstrained zinc-blende crystal. The bonds have cubic point group symmetry, so the valence bands are degenerate at the Γ point. However, under biaxial tension (uniaxial compression along z) all the bonds are equally rotated towards the $x-y$ plane (Fig. 4.4b), while under biaxial compression, they move away from the $x-y$ plane (Fig. 4.4c). Near the valence band edge, bonding states arising from the overlap of the directional p orbitals mainly contribute to the formation of the bands there. The spherical s orbitals contribute to the conduction band edge states. Considering only contributions from the p orbitals and projecting one of the tetrahedral bonds along a principal direction ($\langle 100 \rangle$, $\langle 110 \rangle$ or $\langle 111 \rangle$), the out-of-plane orbital forms the LH states, i.e., p_z orbital along (001) or z direction. Then, HH states are formed by the in-plane orbitals, for instance, p_x and p_y orbitals if we are looking from the $z-$direction. We can then explain the effect of strain on these p orbitals using a simple sp^3 tight binding model. The ignored virtual s* and d orbitals end up being important quantitatively, the former for indirect band-gap semiconductors like Si, the latter to nail down its transverse effective masses. However, they have less qualitative relevance to direct bandgap III-V materials. We use the full $sp^3s^*d^5$ set for our numerical evaluations, but a simplified sp^3 for the current qualitative arguments.

Chadi and Cohen [118] and Harrison [119] used sp^3 tight-binding model to investigate the electronic band structure of various diamond and zinc-blende crystals. In the model, the valence band orbitals form the conduction and valence bands. Each atom in the primitive cell contributes an s, p_x, p_y and p_z orbital. The resulting Hamiltonian is an 8×8 matrix without inclusion of spin-orbit coupling. At the Γ point, the sp^3 Hamiltonian can be simplified to:

$$H = \begin{pmatrix} E_{SC} & V_{SS} & 0 & 0 & 0 & 0 & 0 & 0 \\ V_{SS} & E_{SA} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & E_{PC} & V_{XX} & 0 & 0 & 0 & 0 \\ 0 & 0 & V_{XX} & E_{PA} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & E_{PC} & V_{YY} & 0 & 0 \\ 0 & 0 & 0 & 0 & V_{YY} & E_{PA} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & E_{PC} & V_{ZZ} \\ 0 & 0 & 0 & 0 & 0 & 0 & V_{ZZ} & E_{PA} \end{pmatrix} \tag{4.6}$$

This Hamiltonian can be simplified into four 2×2 matrices. Each of these matrices represents the interaction of an aforementioned orbital type between the two constituent atoms. The eigenstates at the valence band edge can be computed from the Hamiltonians of the p_x, p_y and p_z orbitals:

$$H_1 = \begin{pmatrix} E_{PC} & V_{XX} \\ V_{XX} & E_{PA} \end{pmatrix} H_2 = \begin{pmatrix} E_{PC} & V_{YY} \\ V_{YY} & E_{PA} \end{pmatrix}$$

$$H_3 = \begin{pmatrix} E_{PC} & V_{XX} \\ V_{XX} & E_{PA} \end{pmatrix} \tag{4.7}$$

Here, $E_{A,C}$ represent the on-site energy of the anion and cation, respectively, and V_{ii} is the interaction constant representing the orbital overlap. The valence band states at the Γ point can be computed by diagonalizing these matrices to get:

$$
\begin{aligned}
E_1 &= \frac{E_{PC}+E_{PA}}{2} - \sqrt{\left(\frac{E_{PC}-E_{PA}}{2}\right)^2 + V_{XX}^2} \\
E_2 &= \frac{E_{PC}+E_{PA}}{2} - \sqrt{\left(\frac{E_{PC}-E_{PA}}{2}\right)^2 + V_{YY}^2} \\
E_3 &= \frac{E_{PC}+E_{PA}}{2} - \sqrt{\left(\frac{E_{PC}-E_{PA}}{2}\right)^2 + V_{ZZ}^2}
\end{aligned}
\tag{4.8}
$$

For an unstrained system, $V_{XX} = V_{YY} = V_{ZZ}$, which results in degenerate bands. This is consistent with the observation that bulk semiconductors are symmetric along all the cubic axes. A pictorial view of the p_x, p_y and p_z orbital overlaps is shown in Fig. 4.5(a), (b) and (c). Each p orbital bond consists of head-on (σ) and side-on (π) couplings, as shown in Fig. 4.5(d). The interaction constant V_{ii} is written in terms of contributions from these bonds. In the figure, θ represents the azimuthal angle between the bond and relevant axis for the constant

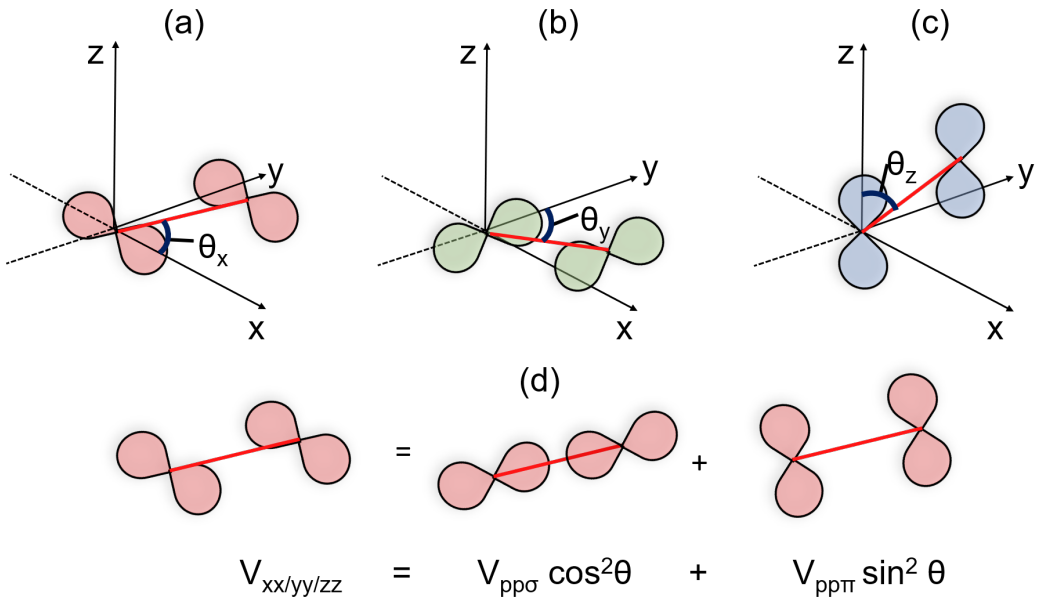

Figure 4.5: Orbital overlap and azimuthal θ angle for (a) p_x, (b) p_y and (c) p_z orbitals. In (d) the σ and π components of the bond are shown.

we are considering, i.e., x-axis for V_{XX}. These interaction constants can then be written in terms of the directional cosines (l, m, n) [120]:

$$\begin{aligned} V_{XX} &= l^2 V_{pp\sigma} + (1 - l^2) V_{pp\pi} \\ V_{YY} &= m^2 V_{pp\sigma} + (1 - m^2) V_{pp\pi} \\ V_{ZZ} &= n^2 V_{pp\sigma} + (1 - n^2) V_{pp\pi} \end{aligned} \tag{4.9}$$

where, $(l, m, n) = (1, 1, 1)/\sqrt{3}$ for an unstrained system.

The strain tensor of a system can be broken down into three components- a hydrostatic strain and two kinds of shear strain [108]. The hydrostatic strain results in the overall shifting of the energy bands as the crystal symmetry is not broken. However, biaxial shear strain results in the breaking of crystal symmetry, lifting band degeneracy at the Γ point and resulting in band warping as well. Under biaxial strain in the x-y plane, the traceless shear strain tensor can be written as

$$\frac{1}{3} \begin{pmatrix} e_{xx} - e_{zz} & 0 & 0 \\ 0 & e_{xx} - e_{zz} & 0 \\ 0 & 0 & -2\,(e_{xx} - e_{zz}) \end{pmatrix} \tag{4.10}$$

where, $e_{xx} = a_{||}/a_i - 1$ and $e_{zz} = -D_{001} e_{xx}$. Here, $a_{||}$ and a_i represent the substrate and epilayer lattice constants, respectively. Also, the Poisson's ratio $D = 2C_{12}/C_{11}$ where C_{11} and C_{12} are elastic constants [112]. Considering $\epsilon = e_{xx} - e_{zz}$ the directional cosines change to $(l, m, n) = (1 + \epsilon, 1 + \epsilon, 1 - 2\epsilon)/\sqrt{3}$. As a result, $V_{XX} = V_{YY}$, but these are not equal to V_{ZZ}. Using Eq. 4.9 it is then possible to show the effect of biaxial strain on the bandstructure.

Under biaxial tension, as the bond rotates towards the $x - y$ plane, the overlap between the p_x/p_y orbitals of the two atoms increases while the overlap of the p_z orbitals decreases. The azimuthal angles θ_x and θ_y decrease while θ_z increases. One can think of the p_x/p_y

orbitals of the two atoms becoming more head-on while p_z orbitals becoming more parallel. This increases the contribution of the σ components of the p_x/p_y orbitals and weakens for the p_z orbital. On the contrary, the contribution of the π bond of the p_z orbital overlap increases but diminishes for the p_x/p_y orbitals. As a result, V_{XX}, V_{YY} will increase while V_{ZZ} will decrease, as can be inferred by placing the values of the directional cosines in Eq. 4.9. Using Eq. 4.8, we can then see that the HH states go down in energy while the LH states go up under biaxial tension. The situation is reversed under biaxial compression. The bond rotates away from the $x-y$ plane, increasing θ_x/θ_y and reducing θ_z. This in turn leads to lower V_{XX}/V_{YY} and higher V_{ZZ}. As a result, HH bands rise in energy while LH states go down. This simplified picture explains the movement of the bands in the InAlAs digital alloy,

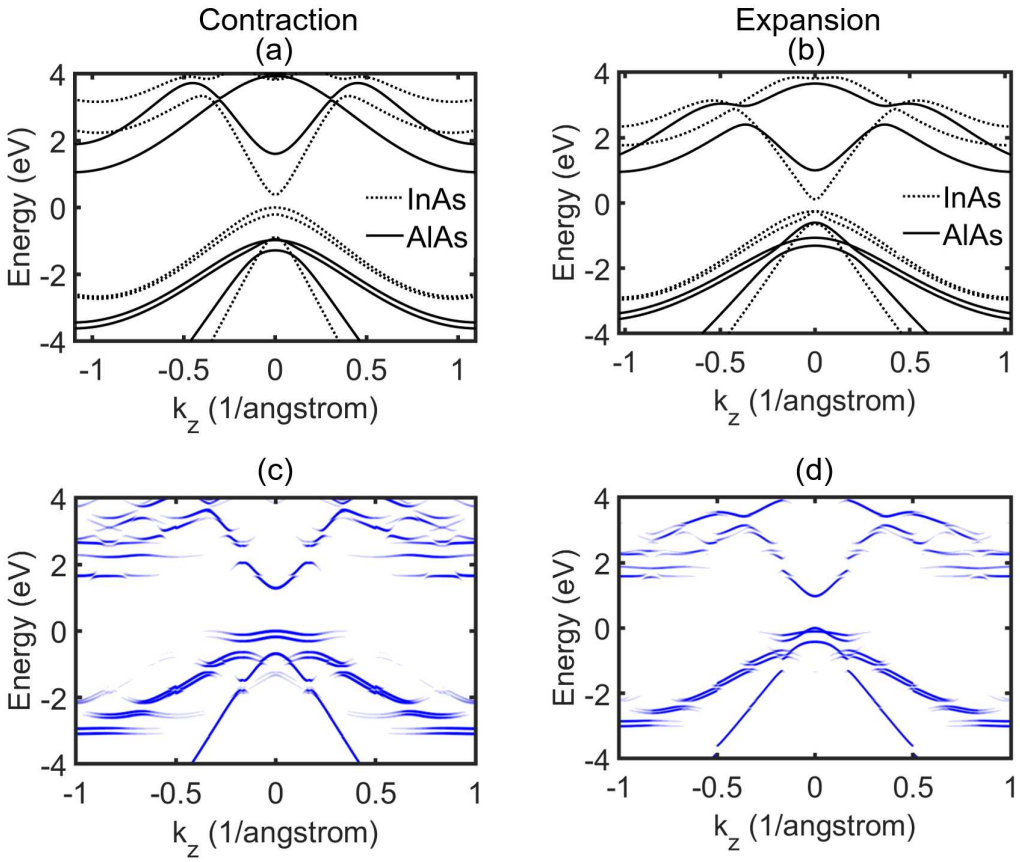

Figure 4.6: Bandstructure of strained InAs and AlAs for (a) "contraction"-where substrate lattice constant is 3% less than InP lattice constant and (b) "expansion"- substrate lattice constant is 3% more than InP lattice constant. The unfolded bandstructure of 6ML InAlAs under contraction and expansion is shown in (c) and (d), respectively.

and subsequently, the essential physics of the minigap formation in the sp^3 basis.

Having a detailed understanding of the underlying physics of the minigaps, it is then essential to see how we can control the minigap size in these digital alloys. Since the minigap formation is related to biaxial strain, we must study how the bandstructure of these alloys change with strain. We compute the bandstructure for two cases: "contraction"- the substrate lattice constant is smaller than the real substrate (InP for InAlAs) lattice constant, and "expansion"- in which the substrate lattice constant is greater. The bandstructure for strained InAs and AlAs under contraction and expansion is shown in Fig. 4.6(a) and (b). Under these conditions, the binary constituents experience unequal biaxial strains, and due to their different values of Poisson's ratio, D_{001}, they also undergo different amounts of strain in the z direction. Thus, their valence bands move by different amounts. InAs has a higher D_{001} than AlAs [112] and hence bands of InAs are more responsive to strain. We notice that the valence band offset under contraction is large compared to the expansion case. Consequently, the valence bands of InAlAs under contraction become flatter and the minigaps increase in size, as depicted in Fig. 4.6(c). However, we can see in Fig. 4.6(d) that under expansion the InAlAs top valence band effective mass decreases and the minigaps become smaller. The 2D energy contours of the top band of InAlAs in the $x-y$ plane for regular, contraction and expansion cases are depicted in Fig. 4.7. We observe that for the regular and contraction cases, the top bands are highly anisotropic. Under biaxial strain, in the in-plane (x and y) directions the

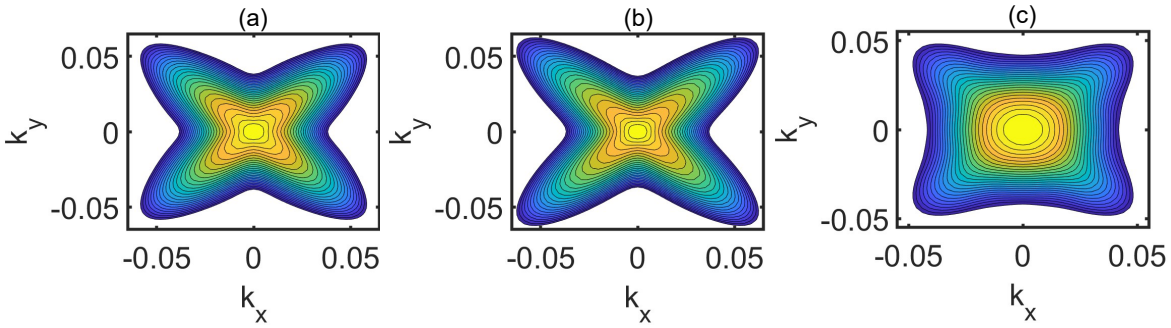

Figure 4.7: 2D energy contour in the $x-y$ plane of the top band of InAlAs for (a) regular (b) contraction and (c) expansion.The energy range for the contour is from $0.025eV$ to $0.5eV$ below the valence band edge.

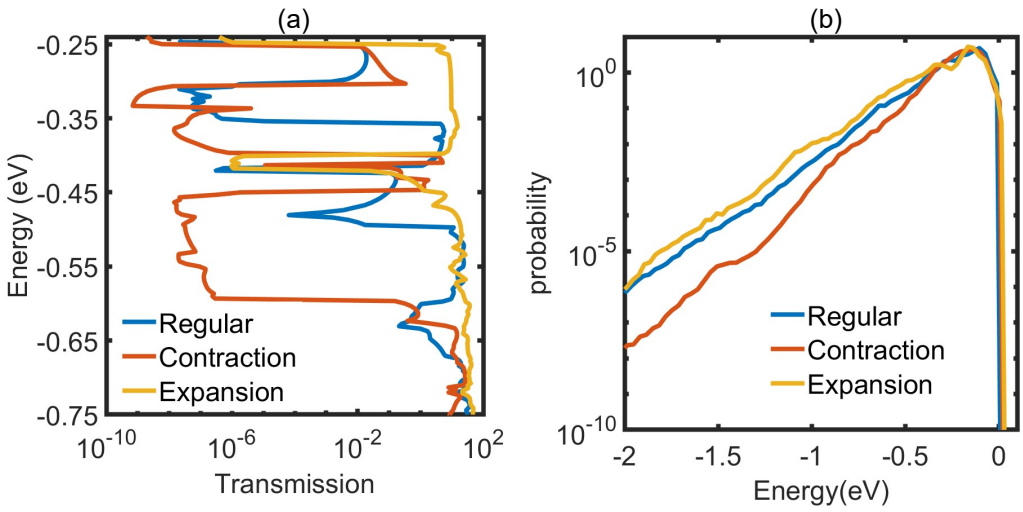

Figure 4.8: (a) Transmission vs. Energy plot and (b) Carrier Occupation Probability vs. Energy for 6ML InAlAs with regular, compressive and tensile strain.

bands move in the opposite direction to that of the out-of-plane (z) direction [117]. As a result, under contraction, the effective mass in the $x-y$ plane decreases. If Fig. 4.7(b) is compared to Fig. 4.7(a), we observe the contour lines become more elliptical, which indicates the lowering of the mass under contraction. The effective mass increases for expansion as the contour lines becomes flatter in Fig. 4.7(c). This observed anisotropic nature of the bands can be utilized to explore the use of digital alloys like InAlAs in other applications such as transistors.

One key aspect we need to study is the impact of the strain on the carrier transport of digital alloys. Since we have been primarily concerned with the effect of strain in the valence bands, we look at the effect on carrier transport in InAlAs valence band in Fig. 4.8. Fig. 4.8(a) depicts the ballistic transmssion vs. energy spectrum in the valence band under regular, expansion and contraction conditions in 6ML InAlAs. The transmission has been computed using the NEGF formalism. We observe that as we go from expansion to regular to contraction case, the transmission gaps increase in size due to the increasing size of the minigaps and enhanced effective mass. As a result, the probability to tunnel across the minigaps decreases and the holes will be more localized near the valence band. This will help in reducing the

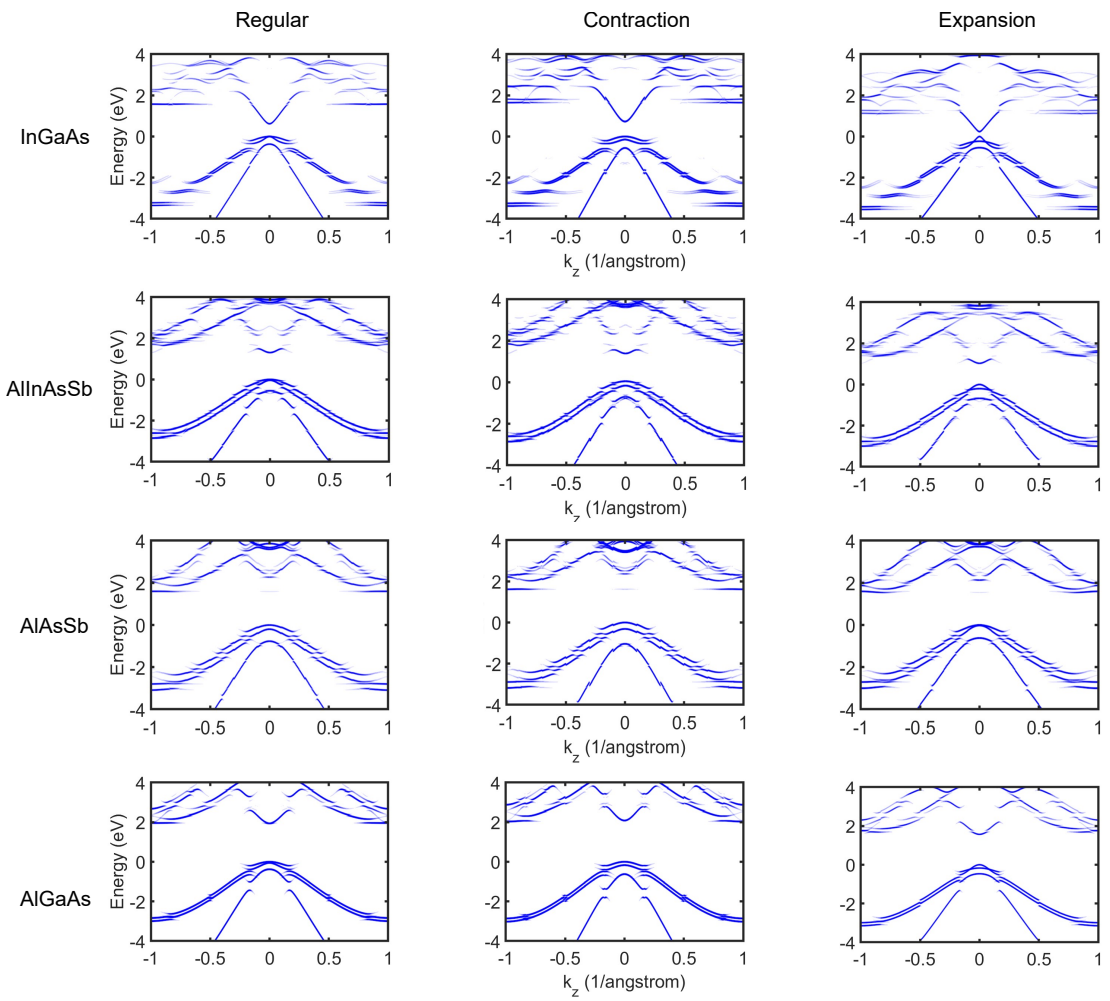

Figure 4.9: Bandstructure of InGaAs, AlInAsSb, AlAsSb and AlGaAs for regular, contraction and expansion cases.

excess noise in APDs. Another mechanism by which holes can bypass the minigaps is optical phonon scattering. We look at the effect of this scattering using a Boltzmann Transport solver. The carrier probability vs. energy with optical phonon scattering under an electric field of $1MV/cm$ is shown in Fig. 4.8(b). Under expansion condition, holes have a higher probability of occupying higher energy states compared to the regular and contraction cases. Under contraction, the probability is the lowest. Therefore, this is a further indication that contraction prevents holes from reaching higher energies. It is then possible to design better low noise electron injected digital alloy APDs with lower hole impact ionization by applying contraction to materials like InAlAs.

In addition to InAlAs, we also computed the bandstructures of 6ML InGaAs, 10ML AlInAsSb, 5ML AlAsSb and 6ML AlAsSb digital alloys under regular, contraction and expansion conditions along the 001 direction. These bandstructures are shown in Fig. 4.9. The primary binary constituents for these alloys are: InAs and GaAs for InGaAs, InAs and AlSb for AlInAsSb, AlAs and AlSb for AlAsSb, and AlAs And GaAs for AlGaAs. For the regular bandstructures, InGaAs and AlAsSb has InP substrate, AlInAsSb has GaSb substrate and AlGaAs has GaAs substrate. The lattice constants for all these materials are taken from the paper by Tan *et al.* [63]. For InGaAs, AlInAsSb and AlAsSb, one of the binary constituents has a lattice constant that is greater than the substrate lattice constant while the other constituent lattice constant is smaller. Thus, the binary components experience alternating types of strain. This is not the case for AlGaAs. For all the material combinations, we observe that the effective mass of the top valence band increases under contraction and reduces for expansion. This is mainly because under contraction HH states move up in energy, whereas under expansion they move down, leaving LH states as the top states in the valence band. For InGaAs, we see that the minigap increases in size with contraction from $0.03eV$ to $0.16eV$ which is similar to the behavior of InAlAs described earlier. In AlInAsSb, there is a separation between the HH and LH bands around the Γ point under contraction. A similar gap is seen for AlAsSb under regular condition. This gap size increases under contraction. The gaps vanish for both AlInAsSb and AlAsSb under expansion. The minigap sizes also increase under contraction by about $0.04eV$ for AlInAsSb and $0.02eV$ for AlAsSb. However, for AlGaAs we do not observe any minigaps in the light-hole band. This is primarily because the HH/LH bands of the binary constituents in AlGaAs move in the same direction under strain as both experience the same type of biaxial strain. Thus, by band engineering in the digital alloys using biaxial strain, their performance in APDs can be enhanced or possibly used for other applications.

4.3 Summary

In this study, we demonstrate that large band offsets result in the formation of minigaps in III-V digital alloys. This band offset results from biaxial strain. Using an orbital chemistry picture, we explained how these minigaps are created. Furthermore, we illustrated that we can engineer the bandstructure by tuning the biaxial strain in a wide range of digital alloys. As a general rule, we observe that decreasing the substrate lattice constant can enhance the performance of digital alloys in APDs.

Key Contributions:

- Demonstrated that strain results in the formation of minigaps.
- Illustrated that minigap size can be modulated by varying strain.

Chapter 5

A Multiscale Compact Model of *p-i-n* Avalanche Photodiodes

The rapid growth of Internet of Things (IoT) applications is resulting in the connection of more and more devices to the internet. An estimated 20.4 billion IoT devices connected through machine-to-machine technology by the end of the year 2020 [6]. Furthermore, the advent of 5G communication technology will enable faster communication between wireless devices along with a reduction of over 90% in energy consumption compared to 4G systems [7]. Thus, it is expected that 5G technology will enable an exponential growth in IoT devices and systems in the near future. This boom in telecom and data communication applications will drive the demand for more efficient and cheaper photonic integrated circuits (PICs) [8]. Only scalable, integrated photonic technologies can meet the huge demand coming from 5G and IoT technologies. Currently, there is a significant push to integrate III-V photonic devices onto the silicon platforms that form the backbone of modern electronic devices. The prospect of coupling optical transmitters and receivers with the state-of-the-art CMOS technology for compact IoT devices is highly enticing. In the communication area, APDs can achieve superior performance than conventional *p-i-n* photodiodes. The high internal gain of APDs, which arises from impact ionization, translates into greater receiver sensitivity

and a dynamic operating range with an accompanying increase in loss margins [121–124]. In addition to communications [10], APDs have been used in a wide range of applications including imaging, [11, 12] and single photon detection [9, 125].

This chapter proposes a simple circuit model of a *p-i-n* APD that is calibrated to both state-of-the art first-principles material studies, as well as experimental devices. The model incorporates accurate material parameters like material effective mass and bandgap as well as fitting parameters for calibration. Our model enables accurate simulation of the circuit behavior of state-of-the-art digital alloy APDs in PICs, all the way from first-principles studies of underlying materials to circuits. Circuit models for both *p-i-n* APDs and separate absorption, charge, and multiplication (SACM) APDs have been reported [126–128]. Previously reported *p-i-n* APD circuit models [126] use bulk material parameters that fail to capture the quantum effects, such as minigaps, seen in these short-period superlattices. Thus, it is necessary to develop a simple physics-based circuit model that can include the interesting physical properties observed in today's newer materials.

The various tools and models used for the simulation of APDs in this chapter are described in the following sections. We calculate the material properties of the digital alloys using the EDTB model. The transport properties of these digital alloy APDs are calculated using full-band Monte Carlo simulations, which show a good match with experimental results. Finally, we show simulations performed using this circuit model along with calibrations to experimental data. This chapter is reproduced from Ref. [28] ©2021 IEEE co-authored with S. Ganguly, Y. Yuan, J. Zheng, Y. Tan, J. C. Campbell and A. W. Ghosh.

5.1 Model

In this work, we consider short period III-V digital alloys for our simulation. In particular, we will study the characteristics of a digital alloy InAlAs *p-i-n* APD. Fig. 5.1(a) shows a schematic cross-section of the device [29]. For SPICE modeling, we consider a simplified

structure shown in Fig. 5.1(b). The typical electric field profile of this device is given in Fig. 5.1(c). We can see that the highest electric field is in the intrinsic region where the avalanche multiplication occurs. For the simulations in this work, the electric field is considered to be uniform in this region.

The properties of III-V digital alloy APDs differ largely from their bulk counterparts due to the band unfolding imposed by the overall superlattice periodicity. It is therefore imperative to develop a solid modeling framework that accounts for the band modification, device geometry and various scattering processes in our SPICE model. We use state-of-the-art band structure and transport models calibrated to first-principles based results and experiments, which makes our tools very reliable for simulating existing and emerging APD structures. As a first step, the detailed band structure of the material is calculated using the EDTB model, described earlier in detail in Chapter 3. The resulting band structure is fed into a full band Monte Carlo simulator. The Monte Carlo simulator and the APD SPICE models are described in the following subsections.

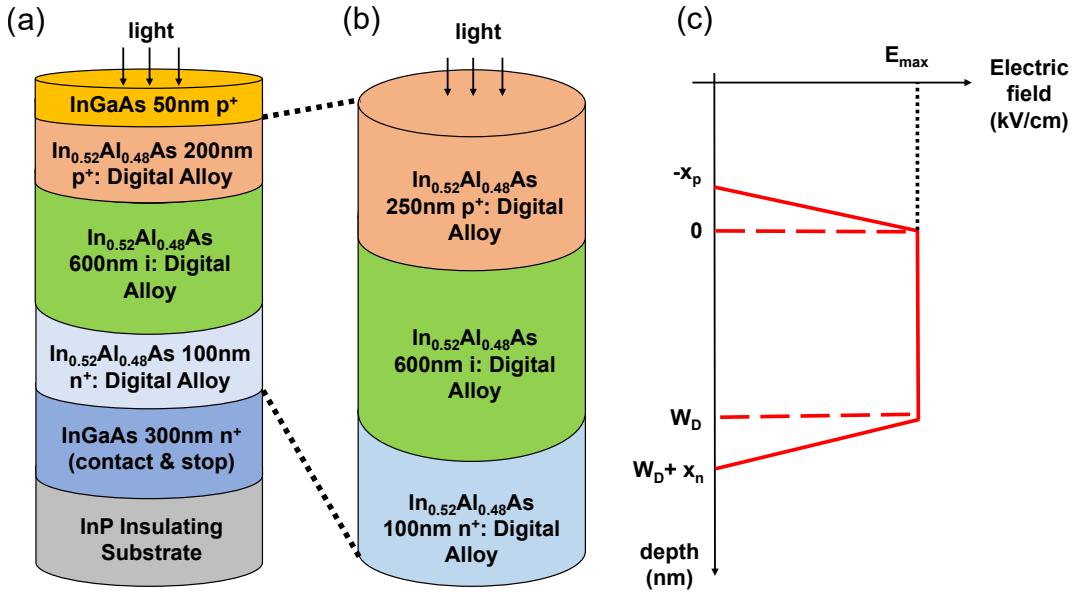

Figure 5.1: (a) Schematic diagram of experimental InAlAs digital alloy APD (b) Schematic diagram of simplified device considered for SPICE model (c) Electric field profile of the simulated *p-i-n* APD.

5.1.1 Monte Carlo Simulation

The Monte Carlo simulation tracks the transport behavior of injected electrons. The flowchart of the method used for extracting device parameters is shown in Fig. 5.2. Initially, the scattering rates are calculated using Fermi's Golden Rule, incorporating the full bandstructure obtained from the EDTB model. The Monte Carlo simulation results used in this chapter

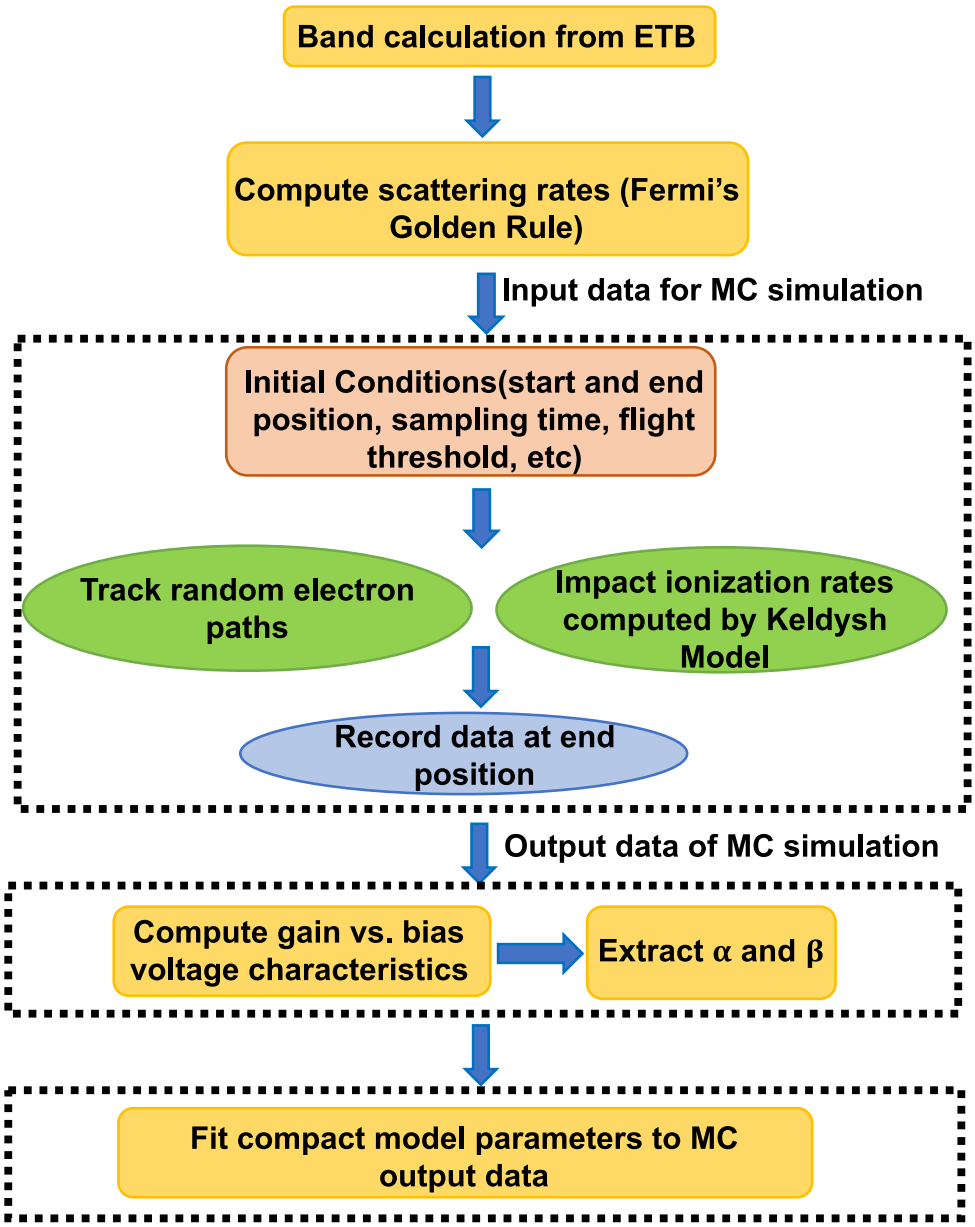

Figure 5.2: Flowchart of Monte Carlo simulation process.

consider deformational potential scattering. The deformational scattering rate $P^{def}_{\nu\nu',n}(\mathbf{k}, \Omega_{\mathbf{k}\pm\mathbf{q}})$ from a point $\mathbf{k}$ in band ν to a region $\Omega_{\mathbf{k}'}$ in band ν' centered around $\mathbf{k}$' is expressed as [36,37]

$$P^{def}_{\nu\nu',n}(\mathbf{k}, \Omega_{\mathbf{k}\pm\mathbf{q}}) = \frac{\pi}{\rho\omega_{nq}}|\Delta^n(\nu', \mathbf{k}, \mathbf{q}, \nu)|^2|I(\nu, \nu'; \mathbf{k}, \mathbf{k} \pm \mathbf{q})|^2 \\ D_{\nu'}(E', \Omega_{\mathbf{k}'}\left(N_{nq} + \frac{1}{2} \mp \frac{1}{2}\right) \tag{5.1}$$

where, ρ represents the lattice density, $\mathbf{q}$ is the phonon wave vector of mode n and the deformation potential is $\Delta^n(\nu', \mathbf{k}, \mathbf{q}, \nu)$.

The path of a single electron through the multiplication region is then tracked under the effects of electric field and random scattering events. The impact ionization rates used in the simulation are computed using the Keldysh model [129], described below. The impact ionization rate using this model is expressed by [36,37]

$$P_{\nu\nu'}(\mathbf{k}, \mathbf{k}') = S(E - E_{th})^{\gamma} \tag{5.2}$$

where, E_{th} is the threshold energy, S is the softness parameter and γ is an approaching index. We can approximate the threshold energy by $E_{TH} = E_G(2\mu + 1)/(\mu + 1)$, where E_G is the material bandgap and μ is the ratio of the electron effective mass to the hole effective mass. This expression assumes a two-band model, i.e., a conduction band and a valence band. For equal conduction and valence band mass, E_{TH} is equal to 1.5 times the bandgap E_G. The expression is widely used to approximate the threshold energy in APDs. These parameters were adjusted by fitting the gain curves and excess noise of the Monte Carlo simulation results with experimental results. The simulation is repeated for multiple electrons and carrier transport properties are computed by averaging over the many trajectories [130]. From the ensemble Monte Carlo simulation, we obtain the gain versus bias voltage characteristics, as shown in Fig. 5.3 for an InAlAs APD.

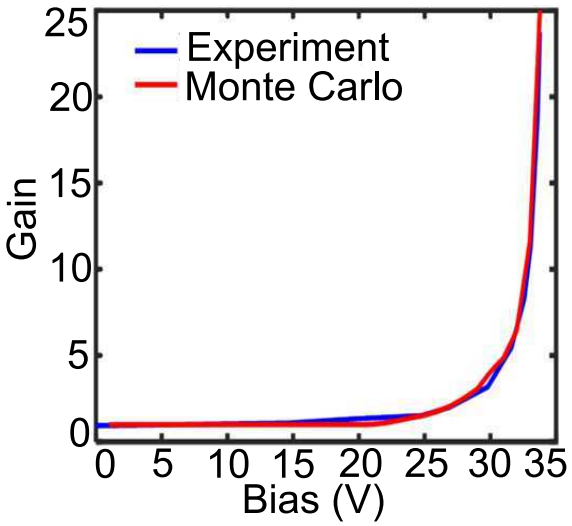

Figure 5.3: Gain vs. reverse bias voltage characteristics for the InAlAs APD.

The electron and hole ionization coefficients, α and β respectively, can be calculated using the following equations:

$$\alpha(\mathcal{E}) = \frac{1}{W_D}\left[\frac{M_n(V)-1}{M_n(V)-M_p(V)}\right] ln\left[\frac{M_n(V)}{M_p(V)}\right] \tag{5.3}$$

$$\beta(\mathcal{E}) = \frac{1}{W_D}\left[\frac{M_p(V)-1}{M_p(V)-M_n(V)}\right] ln\left[\frac{M_p(V)}{M_n(V)}\right] \tag{5.4}$$

where, $M_n(V)$ and $M_p(V)$ represent the gain of the electrons and holes, respectively, $\mathcal{E} = \mathcal{V}/\mathcal{W}_\mathcal{D}$ is the electric field and W_D is the width of the multiplication region. This ratio is an important metric for determining the excess noise factor $F(M)$ of APDs. By varying parameters like temperature and repeating the MC simulations, we can extract the relationship of the ionization coefficients with these parameters. Then the extracted ionization coefficients, their ratio and the gain can be re-expressed in terms of simpler empirical functions that are then used for the compact model:

$$k = c_1 e^{c_2 T} \tag{5.5}$$

$$\alpha(T,\mathcal{E}) = c_3 exp\left[-c_4 T - \left(\frac{c_5}{\mathcal{E}}\right)^n\right] \tag{5.6}$$

$$M = \frac{k-1}{k - exp\left[\alpha(1-k)W_D\right]} \tag{5.7}$$

where $M = M_n$ for pure electron injection and M_p for pure hole injection. In other words, the ionization coefficients are reduced to six material-dependent constants c_{1-5} and n that are extracted from numerical data on the gain curve, such as from experiments or the ensemble Monte Carlo approach described earlier. The fitted parameters are used to calculate the voltage dependent impact ionization gain in the circuit model.

5.1.2 Circuit Model

In this section, we describe the SPICE compatible physics-based circuit model. The circuit layout of the compact model is displayed in Fig. 5.4. The unit cell consists of a simple diode

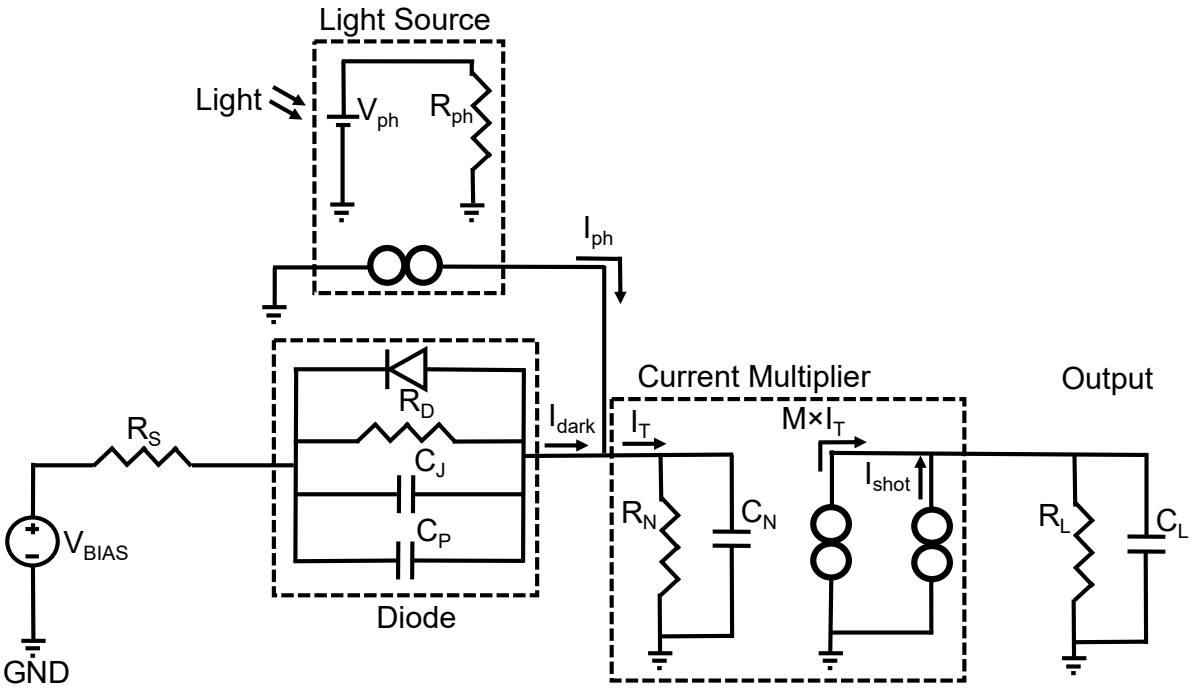

Figure 5.4: Schematic diagram of avalanche photodiode model and testbench used in the SPICE simulations.

coupled to a light source unit and a current multiplier unit to model the APD characteristics. For our model, we consider a dark current that consists of two components: the reverse bias saturation current and the tunneling current. The dominating current at low reverse bias is the current through the resistor R_D. The value of R_D can be calibrated from experimental $I-V$ characteristics or theoretically computed by diode reverse bias saturation current using material parameters. The tunneling current flowing through the diode symbol, I_{TUN}, considering a triangular barrier, is calculated using the Fowler-Nordheim equation.

$$I_{TUN} = \frac{\sqrt{2m^*}q^3\mathcal{E}V_{bias}A}{4\pi^2\hbar^2E_G^{1/2}} exp\left[-\frac{\Theta\sqrt{2m^*}E_G^{3/2}}{q\hbar\mathcal{E}}\right] \tag{5.8}$$

Parameter	**Value**
η	40%
c_1	0.012
R	0.01
c_2	0.0147 K^{-1}
m^*	$0.08m_0$
c_3	2.2×10^7 cm^{-1}
E_G	1.25 eV
c_4	0.004 K^{-1}
n	0.9
c_5	3.5×10^6 Vcm^{-1}
a_p	1.57×10^4 Vcm^{-1}
A	31.4 nm^2
λ	1.08 μm
R_D	1.5×10^{11} Ω
I_{L0}	5.3×10^{-13} A
ζ	0.3414 V^{-1}
k	0.01
Θ	0.8
T	300 K
W_P	250 nm
C_J	5.77 pF
R_L	0.1 Ω
C_L	1 pF

Table 5.1: Table of major parameter values used in SPICE model.

The parameters m^*, V_{bias}, A and E_G represent the tunneling effective mass, the bias voltage across the p-i-n structure, cross-sectional area and the bandgap, respectively. The fitting parameter Θ is used to account for the difference in barrier shape with that of an experimental device. For a triangular barrier we can consider $\Theta = 4/3$. Another current component, I_L, is used to model other various leakage currents through the diode like Shockley-Read-Hall generation and trap-assisted tunneling

$$I_L = I_{L0} e^{\zeta V_{bias}} \tag{5.9}$$

For calibration with InAlAs APD experimental data and simulations in this work, we consider I_{L0} and ζ as fitting parameters. The total dark current, I_{dark}, is then given by

$$I_{dark} = \frac{V_{bias}}{R_D} + I_{TUN} + I_L \tag{5.10}$$

The photocurrent I_{ph} is modeled as a voltage controlled current source. The photon source is modeled as a voltage source, V_{ph}, connected to a large resistor, R_{ph}. The source input power P_{IN}, which is proportional to V_{ph}, is determined by

$$P_{IN} = V_{ph} \times 1amp \tag{5.11}$$

$$I_{ph} = q\eta \frac{P_{IN}(1-R)}{hc/\lambda} \left[1 - e^{-a_p W_p}\right] \tag{5.12}$$

Eq.5.12 gives the photocurrent, I_{ph}. In the equation, R is the reflectivity of the absorption region, a_p is the absorption coefficient, W_P is the p-region width and λ gives the wavelength of the input light source. η represents the internal quantum efficiency. For a given material, the theoretical value of η can be numerically calculated using [131]

$$\eta(\lambda, T) = \frac{E_g \int_{E_G}^{\infty} D(E) f(E) dE}{\int_0^{\infty} E D(E) f(E) dE} \tag{5.13}$$

where, $D(E)$ and $f(E)$ are the photon density of states and occupancy (given by Bose-Einstein statistics) functions.

The reason a voltage source is preferred for the photon source is that internally SPICE models use all nodal equations as $I = GV$ type matrices when the simulations are set up, i.e., the voltages V are independent variables, and the currents I are dependent. Therefore, any current-controlled voltage source or current-controlled current source element needs to be converted to an equivalent voltage-controlled current source or voltage-controlled voltage source element using Thevenin equivalent. This matters only for large simulations or to guard against convergence issues, particularly if the resistance, R, is very small since this makes the conductance G matrix singular. In this case, since R is very large this is not a consideration.

The current multiplier models the avalanche multiplication of the APD. The multiplier takes the sum of the unmultiplied dark and photo current given by eq. 5.14 as input and multiplies that by the gain M, which is a function of the reverse bias voltage given by Eq. 5.7. The output current of the APD is finally obtained by eq. 5.15

$$I_T = I_{dark} + I_{ph} \tag{5.14}$$

$$I_{out} = M \times I_T \tag{5.15}$$

The noise sources of the APDs must also be included in the model to accurately model their behavior in circuit applications. In the model, noise currents are modeled as separate current sources. Here, we can consider only the dominant shot noise of the APDs. For a given bandwidth Δf, the noise variance of the shot noise is given by

$$\langle i_{shot}^2 \rangle = 2q(I_{dark} + I_{ph})\langle M^2 \rangle F(M)\Delta f \tag{5.16}$$

It is possible to extract the APD bandwidth, Δf, from MC simulation. The average impulse function for a large number of short input photon pulses, shaped like Gaussian pulses,

can be extracted from MC simulation [132]. The decaying tail of the impulse function can be fitted by the function $h(t) = 2\pi\eta Mq\Delta f exp(-2\pi\Delta f t)u(t)$, where $u(t)$ is a unit step function, having a value of one for all positive t and zero for the rest. Here, Δf is the frequency at which APD's gain drops to 0.707 of its maximum value, which is the 3 dB bandwidth. It is important to point out that the bandwidth of a circuit, where APD is a component, can be dominated by other circuit elements besides the APD. However, those extrinsic effects can be captured by our compact circuit model where the APD enters as a component block.

For transient calculations, capacitances are added in parallel to the diode. C_J represents the intrinsic capacitance of the device and C_P is the parasitic capacitance. A resistor R_S is added in series to the bias voltage source to model the contact resistances. A small resistor and capacitor, R_N and C_N, are used at the output of the diode to help with convergence in SPICE. R_L and C_L are the load resistor and capacitor used to extract the output. The values used in this chapter are listed in Table 5.1 and are obtained from a previous publication [133]. These can be computed from MC simulations.

In summary, given any digital alloy combination, our compact model can take material parameters like bandgap and effective mass as input and outputs relevant metrics of APD circuit simulation like dark current, gain, quantum efficiency and noise current. The dark current is obtained from eq, (11), gain from eq. (8), quantum efficiency from eq. (14) and APD excess noise current from eq. (17). These metrics are calculated based on calibrations to MC simulation or experimental results.

5.2 Results and Discussion

The temperature dependent reverse bias I-V characteristics are shown in Fig. 5.5(a). At low bias the current is dominated by the diffusion current density which is a constant in the device. The large current at high bias is attributed to the increasing tunneling current and increasing gain of the device. At high bias, the current reduces with increased temperature

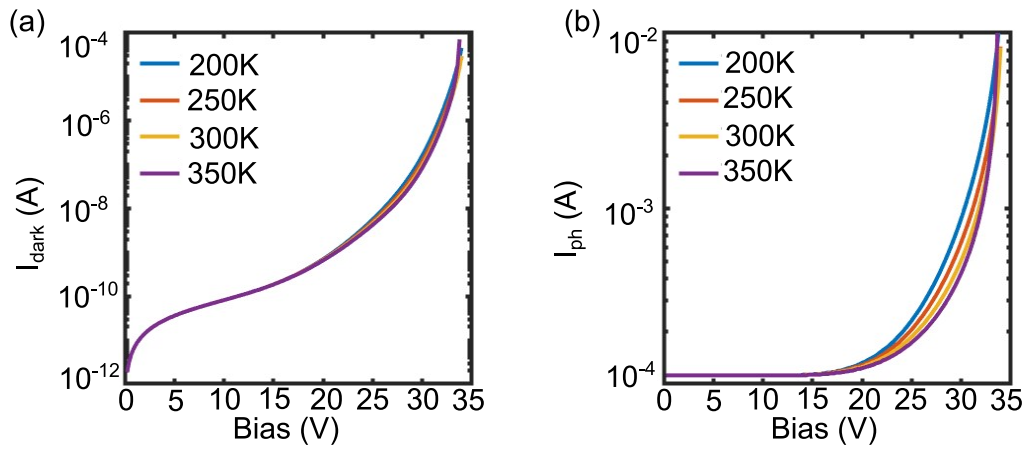

Figure 5.5: (a) Dark current vs. reverse bias voltage characteristics (b) Photo current vs. reverse bias voltage characteristics. Both plots are given as a function of temperature.

due to increased phonon scattering occurring in the device, captured through our scattering rates that enter the Monte Carlo simulations and are eventually lumped into the c_{1-5} and n parameters. Fig. 5.5(b) shows the photo-current vs. reverse bias voltage for different temperatures. It exhibits a similar trend at high bias like the dark current due to the increasing gain and phonon scattering. For this plot, we considered $P_{IN} = 1mW$.

In Fig. 5.6(a), the simulated reverse bias dark current characteristics of the InAlAs digital

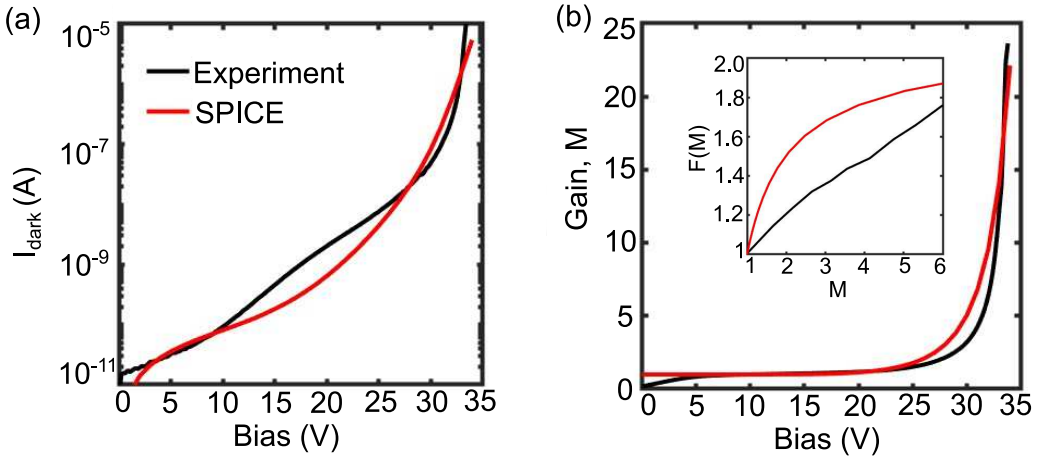

Figure 5.6: (a) Calibration of the simulated dark current characteristics using SPICE model of a 6-monolayer InAlAs digital alloy APD (b) Calibrated of the simulated gain vs. bias characteristics of this APD. Inset-Excess noise factor vs. gain of computed using McIntyre's Formula compared to the experimental results.

alloy APD using the calibrated compact model are plotted and compared to the experimental measurements [29]. The calibrated parameters are given in Table 5.1. The simulated values agree well at low and high bias. There is a small discrepancy in the middle region that is due to the use of the approximate equations used in calculating some of the parameters, which result in overestimation of the scattering processes involved. Additionally, some discrepancy is due to the simplification of the structure considered in the simulation. This is primarily to allow for faster convergence of the SPICE model. Fig. 5.6(b) shows the simulated gain versus bias characteristics compared to the experimental data. The values are in good agreement with each other. To calculate the excess noise factor versus gain characteristics, we use the standard McIntyre's formula [25]. The comparison between the calculated and experimental F versus M is shown in the inset of Fig. 5.6(b). We observe that McIntyre's formula overestimates the excess noise at low gain values compared to experiments. The mismatch can be attributed to simplifications, like ignoring the 'dead-space' effect. For APDs with large multiplication regions, the dead-space effect can be ignored and good agreement can be achieved using McIntyre's formula.

We plot the output I-V characteristics for different input powers in Fig. 5.7. A higher input power of the light source results in more electron-hole pair generation in the absorption

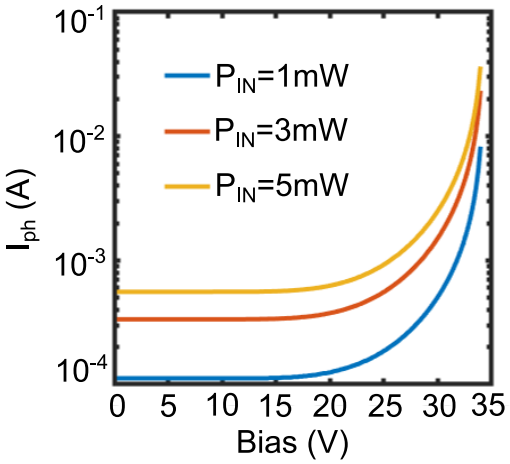

Figure 5.7: Photocurrent vs. reverse bias voltage characteristics as a function of input power.

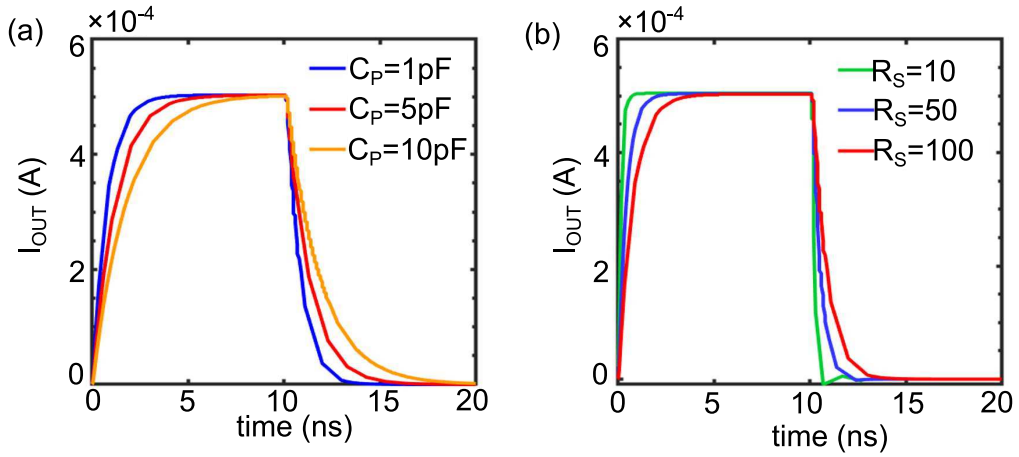

Figure 5.8: (a) Transient response of output current with varying parasitic capacitance, C_P (b) Transient response of output current as a function of contact resistance, R_S.

region. This increases the current at low bias. At a very high bias, the APD reaches the avalanche breakdown region where the output current is not affected much by the input power magnitude due to the high carrier generation by the impact ionization process.

The transient characteristics of the InAlAs digital alloy APD considered in this compact model are shown in Figs. 5.8 and 5.9. The bias voltage used is 30V. The transient response of the output current for different parasitic capacitance is simulated in Fig. 5.8(a). We can see that the parasitic capacitance present in the APD can significantly affect its responsiveness

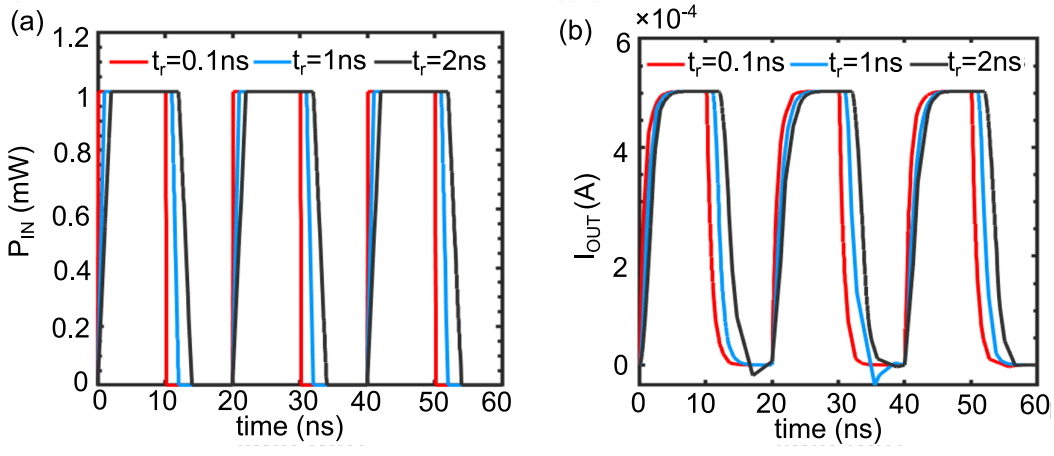

Figure 5.9: (a) Input power (light source) vs. time for different rise and fall times (b) output current vs. time for the different rise and fall times.

to the input light. A larger capacitance results in longer rise and fall times. In a receiver, this will translate to lower detection speed and will affect the overall speed of an integrated photonic device. The effect of different contact resistances on the output transient response is shown in Fig. 5.8(b). Increased contact resistance also leads to higher rise and fall times. However, the effect of R_S is not as significant as the parasitic capacitances.

Fig. 5.9 shows input and output characteristics of the APD for different rise and fall times. It is seen that a longer rise/fall time results in slower discharge at the end of each input cycle and can also result in overshoots while discharging. This shows that the switching pattern of the light source can also affect the refresh rate of the APD.

One of the crucial elements of APD performance is the excess noise that is generated due to the random nature of the impact ionization process. Any noise can significantly degrade the photon detection ability of these photodetectors. Thus, it is necessary to include the noise in the circuit model in order to accurately model PICs. This noise is reduced by allowing impact ionization to be initiated by carrier injection of the carrier with the highest ionization coefficient. This essentially translates to a lower value of the ionization coefficient ratio k.

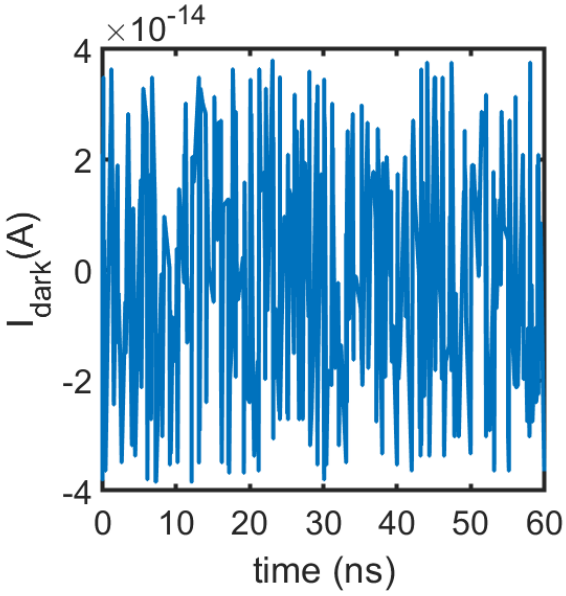

Figure 5.10: Simulated shot noise of the InAlAs APD.

Fig. 5.10 shows the dominant shot noise in the dark current simulated using the SPICE model. The noise current is added as a separate current source in the circuit. We consider $\Delta f = 1Hz$ since it is generally application dependent. Many methods are used to reduce the shot noise in the APDs. In the InAlAs digital alloy APD, the noise is minimized due to the presence of "minigaps" in the valence band which prevents holes from ionizing and thus suppress $F(M)$ [29] . Although in this treatment we only consider shot noise as the dominant source in digital alloy APDs, we can in principle include other types of noise relevant for different geometries as added current sources.

Despite the model being developed for *p-i-n* APDs, it can be modified to simulate other structures such as Separate Absorption Charge Multiplication (SACM) APDs as well. The potential profile of the SACM APDs can be obtained using added electrostatic solvers, which can then be used as input to the MC model. The fitting parameters of the compact model, described in sections 5.1.1 and 5.1.2, can be adjusted to match the results of the MC simulations. The compact model might require some additional resistance and capacitance elements to account for the behavior of the absorption and charge layers of the SACM structure. This will attach additional components to the unmultiplied dark current. The equation for gain given in Sec. 5.1.2 will still be valid, with slight alterations to the fitting parameters for calibration, as the physics of the multiplication region remains unchanged. Also, empirical equations might be required to capture the voltage dependencies of some of the parameters used in the unmultiplied photocurrent.

5.3 Summary

In this chapter, we report an elementary physics-based compact model of APDs. This model captures the essential device physics calculated by first-principles methods and simulates their effect on circuit behavior. The compact model ultimately treats the APD as a black box that can be included in a modular fashion into bigger photonic integrated circuits, with

its input being material and geometric parameters and output being relevant performance metrics such as excess noise, gain, dark current and quantum efficiency. In particular, we simulate the behavior of a new class of digital alloy APDs which exhibit low excess noise at very high bias, which makes them highly suitable for a wide range of applications.

Key Contributions:

- Developed physics-based compact model for III-V digital alloy *p-i-n* APDs.

Chapter 6

Matrix Based Quantum Mechanical Treatment of Impact Ionization

High-field phenomena, such as impact ionization in semiconductor devices are generally simulated using the Monte Carlo simulation method. This technique can accurately simulate the behavior of bulk semiconductors devices. The model treats carriers, like electrons, as particles and simulate their behavior using semi-classical transport equations. The semi-classical picture fails to take into account quantum effects, like tunneling, observed in nanoscale materials like the digital alloy superlattices. Furthermore, in the Monte Carlo model, the carrier ionization rate is calculated using the Keldysh equation that incorporates the ionization threshold energy as a parameter [36]. For a material bandstructure with two parabolic energy bands, a conduction band and a valence band, the threshold energy can be estimated using an analytical equation derived from energy and momentum conservation laws [104]. The equation uses the effective masses of the two bands and the bandgap of the material to estimate the threshold energy. This approach works well for bulk materials. However, the bandstructure of superlattices, like the DAs, includes many complicated non-parabolic energy bands. This gives rise to complex mass tensors that vary with both energy and electric field. In order to simulate these nanoscale materials/devices using Monte Carlo,

certain approximations have to be made and some adjustable parameters must be used to match experimental data. To encapsulate these exotic physics, a quantum transport approach, like NEGF, is required. This work develops a matrix-based NEGF framework for impact ionization that treats carriers as waves and consequently captures the resulting characteristics. To the best of our knowledge there is no such model for impact ionization.

Earlier works derive fit functions for the impact ionization rate based on generalized quantum kinetic equations that improve upon the Keldysh model [134, 135]. The fit formula can capture the effect of a more realistic bandstructure in Monte Carlo simulations. However, this approach cannot capture quantum effects, like tunneling, which requires a matrix based NEGF model. In NEGF, scattering processes like phonon scattering can be included by introducing an additional artificial contact to the channel within the self-consistent Born approximation [136]. The carriers can flow into the contact at a particular energy and is then reintroduced back into the channel at a different energy. The behavior of this contact is described by scattering self-energy terms. The technique for including phonon scattering in NEGF models is well developed [136]. However, it is necessary to develop a simple method by which the impact ionization physics can be included in NEGF models.

This chapter develops a simple one dimensional matrix theory based NEGF model for simulating impact ionization processes under steady-state conditions. We start by describing the general process of including scattering processes within the NEGF framework. Then, we describe the methodology for simulating impact ionization in a basic four-level system. Finally, we extend the methodology to a one-dimensional semiconductor with two parabolic bands. This framework can be extended to materials with a more complete bandstructure and lays the groundwork for eventually developing a quantum transport model for multi dimensional nanoscale devices/materials that incorporates impact ionization.

6.1 Scattering in NEGF

In matrix based NEGF formalism, a device/channel is represented by a Hamiltonian H that is coupled with real contacts whose electron outflow is described by their self-energies Σ. An electrostatic potential can be added to the device using the matrix U. A detailed description of NEGF method for coherent transport has already been provided in Section 3.1.2. In this section, we provide a brief description of incorporating non-coherent transport within the NEGF framework.

Non-coherent processes can be incorporated into a quantum transport model by including a virtual terminal 's' as shown in Fig. 6.1. The inflow and outflow to and from this terminal are described by the additional in and out scattering terms Σ_s^{in} and Σ_s^{out}. For regular contacts, the in and out scattering terms can be described by the broadening matrices, $\Sigma_{1,2}^{in} = \Gamma_{1,2} f_{1,2}$ and $\Sigma_{1,2}^{out} = \Gamma_{1,2}(1 - f_{1,2})$. However, since there is no Fermi function describing the scattering 'virtual' terminal, there is no simple connection between $\Sigma_s^{in,out}$ and Γ_s. Expressions for these terms can be derived using the "Büttiker probe" approach within the self-consistent Born

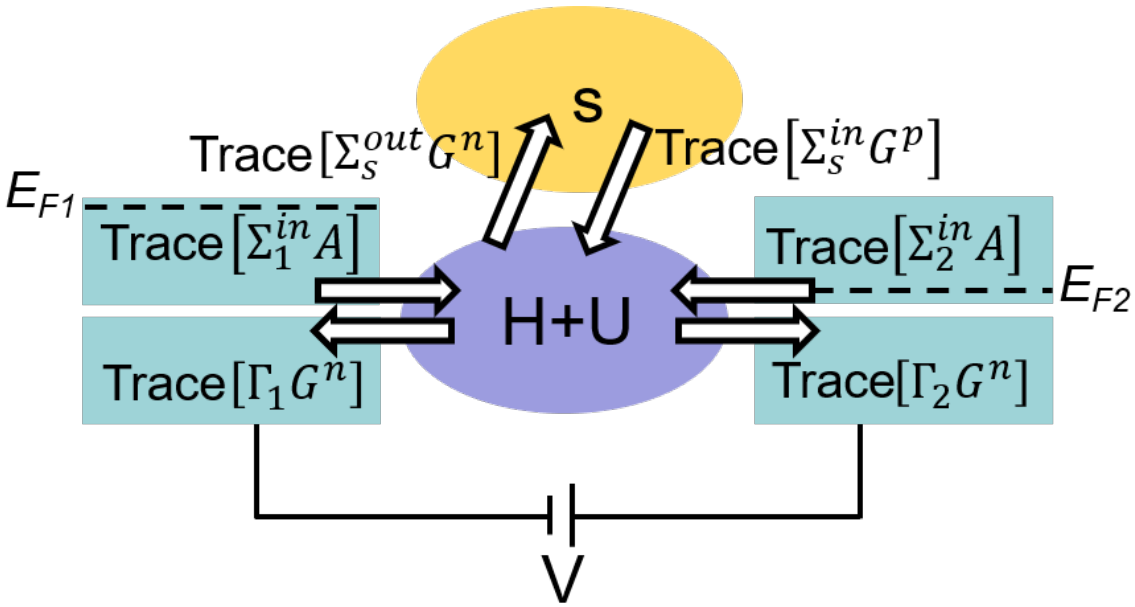

Figure 6.1: Inflow and outflow in non-coherent NEGF transport.

approximation. The Büttiker probe takes in electrons from the channel at a certain energy and reinserts them back into the channel at a different energy based on the scattering mechanism physics. The in and out scattering functions for the virtual terminal at a particular energy can be generally written as [71, 137]:

$$\begin{aligned} \Sigma_s^{in}(E) &= D \otimes G^n(E') \\ \Sigma_s^{out}(E) &= D \otimes G^p(E') \end{aligned} \tag{6.1}$$

where, $D = \langle U_i U_j^* \rangle$ represents the ensemble average of the correlation between the random interaction potentials at the points i and j. The $\otimes$ sign means an element by element multiplication. The choice of the D matrix decides whether the momentum is conserved or relaxed in the scattering process. E' denotes the energy state from which the carrier is flowing in for $\Sigma_s^{in}(E)$ or the energy state to which the carrier is flowing into for $\Sigma_s^{out}(E)$. The correlation functions G^n and G^p denote the electron and hole carrier concentrations here.

A key condition that needs to be satisfied by the virtual contact is that the net current flowing through it must be zero. The energy resolved scattering terminal current I_{SC} is expressed in terms of the Meir-Wingreen formula as [71]:

$$I_{SC}(E) = \frac{q}{h} Tr \left[\Sigma_s^{in}(E) G^p(E) - \Sigma_s^{out}(E) G^n(E) \right]. \tag{6.2}$$

The carrier redistribution in energy and momentum space due to scattering is expressed by this I_{SC} term. This current integrated over all energy must be zero, ensuring there is no leakage from this fictitious contact.

The scattering broadening matrix Γ_s can be expressed as the sum of the in and out scattering functions of the scattering contact, $\Gamma_s(E) = \Sigma_s^{in}(E) + \Sigma_s^{out}(E)$. We can then derive

an expression for the scattering self-energy Σ_s from Γ_s that involves an Hilbert transformation:

$$\Sigma_s(E) = -\frac{i}{2}\Gamma_s(E) + hilbert\left(\Gamma_s(E)\right). \tag{6.3}$$

The retarded Green's function is then recomputed involving the scattering self-energy

$$G(E) = \left[EI - H - \Sigma_1 - \Sigma_2 - \Sigma_s\right]^{-1}. \tag{6.4}$$

The applied potential U is added to the Hamiltonian H. The redistributed electron and hole concentrations should then be recomputed. The expressions for these quantities for dissipative quantum transport are given below.

$$\begin{aligned} G^n(E) &= G(\Sigma_1^{in} + \Sigma_2^{in} + \Sigma_s^{in})G^\dagger \\ G^p(E) &= G(\Sigma_1^{out} + \Sigma_2^{out} + \Sigma_s^{out})G^\dagger \end{aligned} \tag{6.5}$$

Finally, the new terminal currents for the regular contacts can be calculated using the Meir-Wingreen formula

$$I_{1,2} = \frac{q}{h}\int dE Tr\left[\Sigma_{1,2}^{in}A - \Gamma_{1,2}G^n\right] \tag{6.6}$$

where $A = G^n + G^p$ is the spectral function which is essentially the density of states and $I_1 = -I_2$ due to current conservation. These terminal currents include the effect of the scattering processes involved. To incorporate impact ionization into the NEGF framework, we need to derive the scattering self-energies for impact ionization.

6.2 Impact Ionization in a Four-level System

For electron impact ionization, the scattering self-energies allow a high kinetic energy electron to drop down and transfer its energy to an electron in the valence band, and multiply the electron current (and hole current) as shown in Fig. 6.2. The process involves four energy states, three in the conduction band and one in the valence band. In the literature, impact ionization scattering terms are related to the carrier concentrations as n^2p or p^2n. For the matrix based NEGF theory, these relationships will be described in terms of the electron concentration G^n and hole concentration G^p. In order to derive the impact ionization self-energies, we start with a simple four energy level system and study the impact ionization process here. Based on our understanding of this system, we can then extend the self-energy expressions to a 1D semiconductor with energy bands.

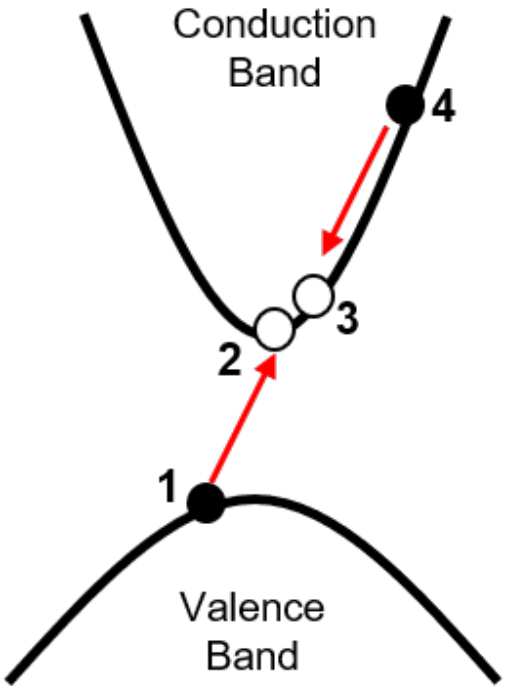

Figure 6.2: Schematic of electron impact ionization.

An illustration of the four-level system considered in this section is given in Fig. 6.3. We designed the system such that the initial states are connected to contact 1 and the final

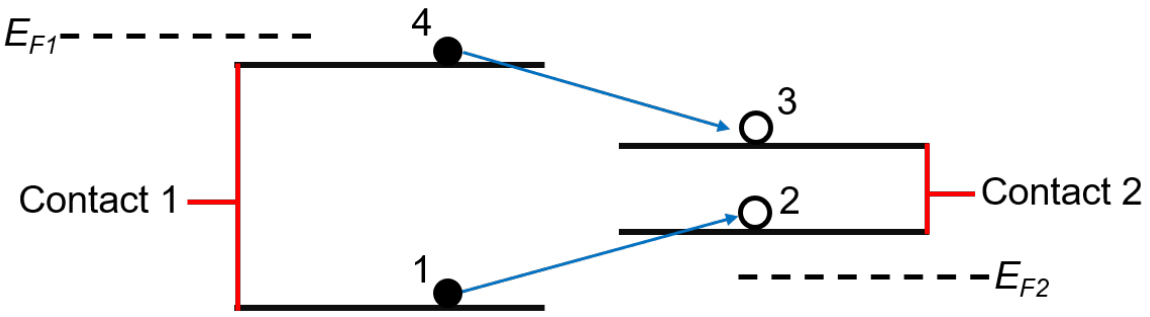

Figure 6.3: Schematic of a four level system under impact ionization.

states are connected to contact 2. The system consists of four energy levels whose onsite energies are denoted by $\epsilon_1 - \epsilon_4$. There is no coupling between the different energy levels. This ensures there is no current flowing through the system under ballistic conditions. The four level Hamiltonian can be written as:

$$H = \begin{pmatrix} \epsilon_1 & 0 & 0 & 0 \\ 0 & \epsilon_2 & 0 & 0 \\ 0 & 0 & \epsilon_3 & 0 \\ 0 & 0 & 0 & \epsilon_4 \end{pmatrix}. \tag{6.7}$$

Current flows through this system only when the conditions of electron impact ionization are satisfied. States 1 and 4 in this system represent the valence band filled state and high energy conduction band state, respectively. The low energy conduction band states are represented by states 2 and 3. The quasi-Fermi levels at the two contacts, E_{F1} and E_{F2}, need to be set such that electrons are injected into the device from the left contact and are extracted from the right contact. In the system, $E_{F1} = E_{F0} + V/2$ and $E_{F2} = E_{F0} - V/2$, where V is the applied voltage across the terminals and E_{F0} is the equilibrium Fermi level of the system. For impact ionization to happen, E_{F1} must be above the levels 1 and 4, and E_{F2} should be below 2 and 3. Electrons are then injected into the low energy state 1 and the high energy state 4. These electrons move to the empty states at the energy levels 2 and 3, respectively, due to the impact ionization process, and are swept away by the right contact.

The carriers must satisfy energy conservation, *i.e.*, $E_4 - E_3 = E_2 - E_1$. Here, the superscripts of E represent the corresponding energy level in the system. The values of $\epsilon_1 - \epsilon_4$ must be chosen to satisfy this energy conservation.

The next step involves defining the in/out scattering functions Σ_s^{in} and Σ_s^{out} for impact ionization in this system. For this purpose, we look at the system using the Büttiker probe approach. The Büttiker probe for impact ionization will take in electrons from the states 1 and 4, and insert those to the states 2 and 3 when the above mentioned conditions are satisfied. Thus, we need outscattering functions for the states 1 and 4, and inscattering functions for the states 2 and 3. These functions can be expressed as

$$\begin{aligned}
\Sigma_s^{out,1}(E) &= D \otimes \int dE^{'''} dE^{''} dE^{'} G_2^p(E^{'''}) G_3^p(E^{''}) G_4^n(E^{'}) \delta(E^{'''} - E - E^{'} + E^{''}) \quad (6.8)\\
\Sigma_s^{out,4}(E) &= D \otimes \int dE^{'''} dE^{''} dE^{'} G_2^p(E^{'''}) G_3^p(E^{''}) G_1^n(E^{'}) \delta(E^{'''} - E^{'} - E + E^{''})\\
\Sigma_s^{in,2}(E) &= D \otimes \int dE^{'''} dE^{''} dE^{'} G_1^n(E^{'''}) G_4^n(E^{''}) G_3^p(E^{'}) \delta(E - E^{'''} - E^{''} + E^{'})\\
\Sigma_s^{in,3}(E) &= D \otimes \int dE^{'''} dE^{''} dE^{'} G_1^n(E^{'''}) G_4^n(E^{''}) G_2^p(E^{'}) \delta(E^{''} - E - E^{'} + E^{'''})
\end{aligned}$$

where, D is treated as a multiplicative constant for now. The indices 1 to 4 represent the four states in the system. Energy conservation is satisfied using the delta function. The scattering terminal I_{SC} current can be written as

$$I_{SC} = \int dE Tr \left[\Sigma_s^{in,2}(E) G_2^p(E) + \Sigma_s^{in,3}(E) G_3^p(E) - \Sigma_s^{out,1}(E) G_1^n(E) - \Sigma_s^{out,4}(E) G_4^p(E)\right] \quad (6.9)$$

and it satisfies the condition $I_{SC} = 0$. Equations 6.8 and 6.9 are scalar equations. The Σ_s^{in} and Σ_s^{out} matrices is then expressed using the equations:

$$\Sigma_s^{in}(E) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & \Sigma_s^{in,2}(E) & 0 & 0 \\ 0 & 0 & \Sigma_s^{in,3}(E) & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{6.10}$$

$$\Sigma_s^{out}(E) = \begin{pmatrix} \Sigma_s^{out,1}(E) & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \Sigma_s^{out,4}(E) \end{pmatrix}. \tag{6.11}$$

Afterwards, the terminal currents of the four-level system is computed using the equations described in Section 6.1. The current vs. voltage characteristics of such a system is shown in Fig. 6.4. From the plot, we can see that under ballistic conditions, there is no current flowing through the system (Contact 1_{ns} and Contact 2_{ns} currents are zero). The terminal current including impact ionization shows a sharp jump at $V = 1V$. At this voltage E_{F1} is above state 4 and E_{F2} is below state 2 and energy conservation is also satisfied, resulting

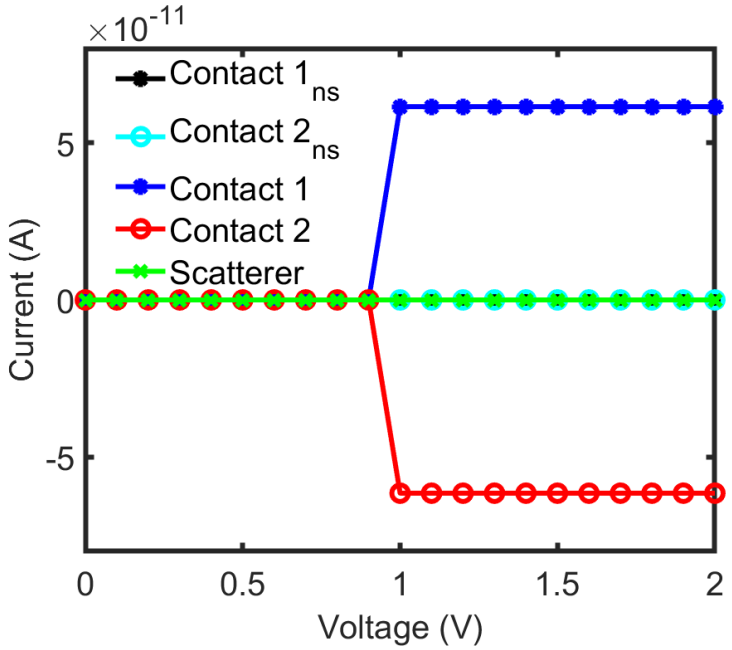

Figure 6.4: Current vs. Voltage characteristics of a four-level system.

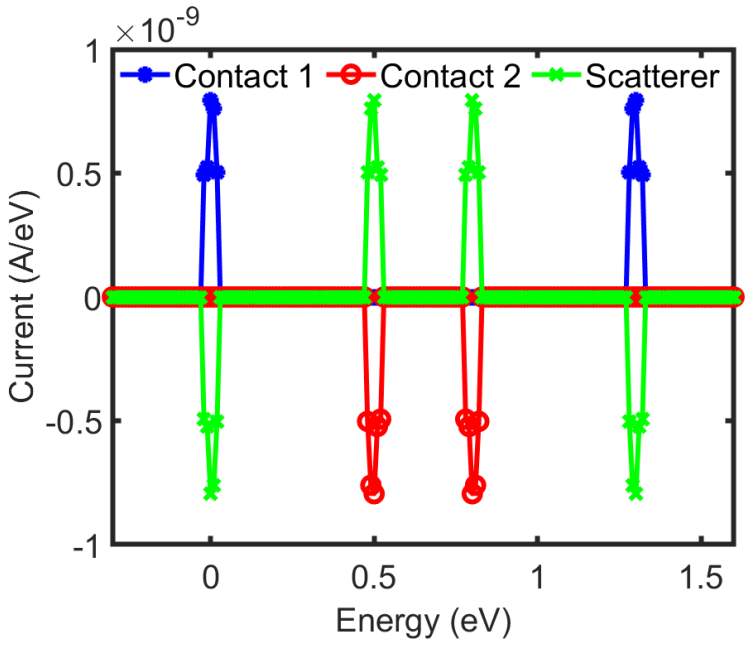

Figure 6.5: Energy resolved current for a four-level system at $V = 1.5V$.

in the jump. The $I_{SC} = 0$ condition is also satisfied in this plot. For our simulations, we set $\epsilon_1 = 0eV$, $\epsilon_2 = 0.5eV$, $\epsilon_3 = 0.8eV$, $\epsilon_4 = 1.3eV$, $E_{F0} = 0.86eV$, $D = 10$, and temperature $T = 3K$. The energy resolved currents are plotted in Fig. 6.5. We can see that the left contact (blue) injects electrons in the states 1 and 4. The virtual contact (green) takes these electrons and reinserts them into states 2 and 3, which are then carried away by the right contact (red).

Equipped with the insight from this NEGF model of a four-level system, we extend the model to a 1D semiconductor next. The following section describes that model and the results obtained using it.

6.3 Impact ionization in a 1D semiconductor

Real semiconductors have energy bands instead of discrete energy levels. Here, we extend the matrix based NEGF theory for impact ionization to such a material. We study a one dimensional dimer chain with parabolic energy bands. Since the threshold energy for parabolic

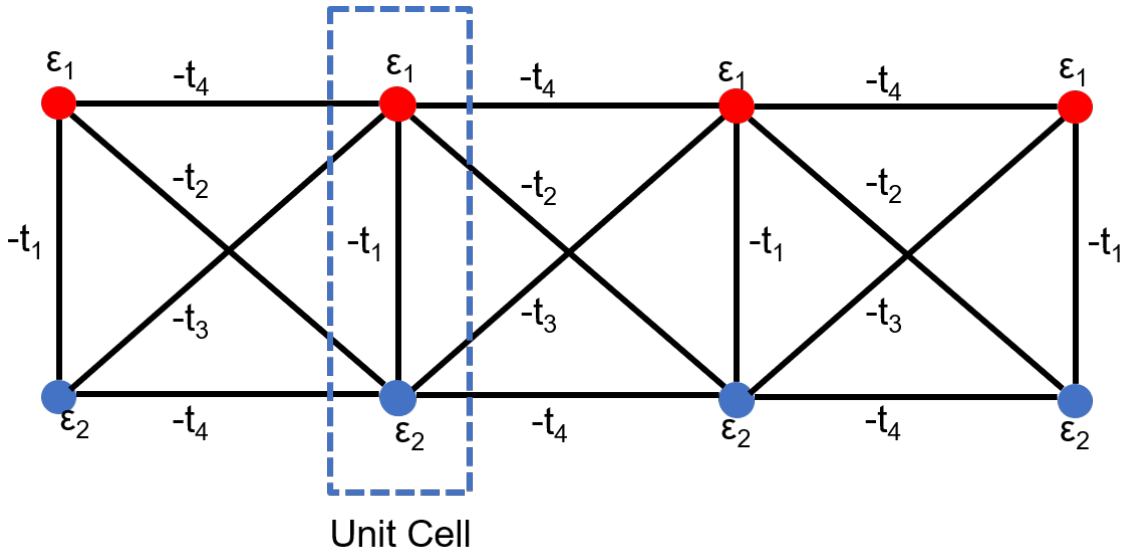

Figure 6.6: One dimensional cross linked dimer chain structure with parabolic bands.

bands can easily be calculated with an analytical expression, it is easy to test the validity of this model. The schematic of a one dimensional cross-linked dimer chain is shown in Fig. 6.6. The cross-links allow us to have different effective mass ratios between the conduction and valence bands. The Hamiltonian for this 1D chain is

$$H = \begin{pmatrix} \alpha & \beta & 0 & & & \\ \beta^{+} & \alpha & \beta & 0 & & \\ 0 & \beta^{+} & \alpha & \beta & & \\ & 0 & \cdots & \cdots & \cdots & \\ & & & & \cdots & \beta \\ & & & & \beta^{+} & \alpha \end{pmatrix} \tag{6.12}$$

where,

$$\alpha = \begin{pmatrix} \epsilon_1 & -t_1 \\ -t_1 & \epsilon_2 \end{pmatrix}, \ \beta = \begin{pmatrix} -t_4 & -t_2 \\ -t_3 & -t_4 \end{pmatrix}.$$

The bandstructure for the unit cell of the dimer chain is depicted in Fig. 6.7. We can create asymmetry between the conduction band and valence band effective masses by varying

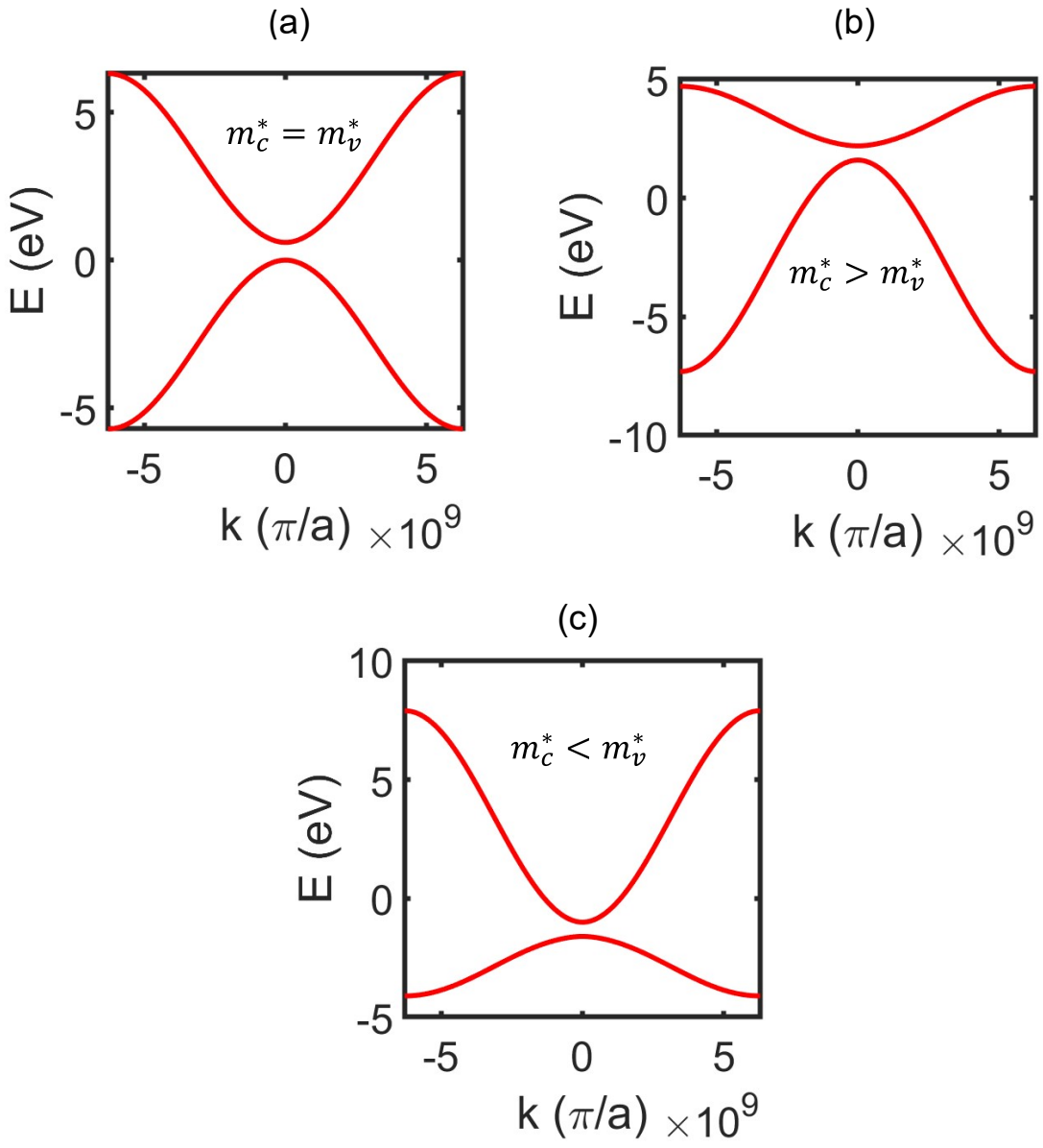

Figure 6.7: Bandstructure for the dimer chain for (a) $m_c^* = m_v^*$ mass (b) $m_c^* > m_v^*$ and $m_c^* < m_v^*$.

the different couplings in the Hamiltonian. For our simulations, we set $\epsilon_1 = 0.6$, $\epsilon_1 = 0$, $t_1 = 3$, $t_2 = -1$ and $t_2 = -2$. Then, we can vary the value of t_4 to create the mass symmetry. Fig. 6.7(a) shows the bandstructure for equal mass for which $t_4 = 0$. In Fig. 6.7(b) we set $t_4 = -0.8$ which results in $m_c^* > m_v^*$ and $t_4 = 0.8$ causes $m_c^* < m_v^*$ like in Fig. 6.7(c).

For electron impact ionization in a semiconductor, three of the energy states are in the conduction band - the high energy electron and the two empty states into which electrons

flow after ionization. The remaining low energy state is in the valence band. We need to distinguish between these states when we extend the model of the four-level system to that with conduction and valence bands. This is done by setting limits to the integrals in the equation of the virtual scattering terminal current. For a semiconductor with energy bands, we can write I_{SC} as

$$\begin{aligned} I_{SC} \quad = Tr\Bigg[\Bigg\{ & D \otimes \int_{E_C}^{\infty} dE_4 G^n(E_4) \int_{-\infty}^{E_V} dE_1 G^n(E_1) \int_{E_C}^{\infty} dE_3 G^p(E_3) \int_{E_C}^{\infty} dE_2 G^p(E_2) + \\ & D \otimes \int_{E_C}^{\infty} dE_2 G^p(E_2) \int_{E_C}^{\infty} dE_4 G^n(E_4) \int_{-\infty}^{E_V} dE_1 G^n(E_1) \int_{E_C}^{\infty} dE_3 G^p(E_3) - \\ & D \otimes \int_{-\infty}^{E_V} dE_1 G^n(E_1) \int_{E_C}^{\infty} dE_3 G^p(E_3) \int_{E_C}^{\infty} dE_2 G^p(E_2) \int_{E_C}^{\infty} dE_4 G^n(E_4) - \\ & D \otimes \int_{E_C}^{\infty} dE_3 G^p(E_3) \int_{E_C}^{\infty} dE_2 G^p(E_2) \int_{E_C}^{\infty} dE_4 G^n(E_4) \int_{-\infty}^{E_V} dE_1 G^n(E_1) \Bigg\} \\ & \times \delta(E_4 - E_3 - E_2 + E_1) \Bigg] \end{aligned} \tag{6.13}$$

where, E_1 and E_4 denote the energies of the initial valence and conduction band electrons, respectively. E_2 and E_3 represent the energies of the empty states to which the electrons flow into. The limits E_C and E_V represent the equilibrium conduction and valence band edges. From I_{SC}, the in and out scattering functions for impact ionization are then defined as

$$\begin{aligned} \Sigma_s^{in}(E_2) &= D \otimes \left(\int_{E_C}^{\infty} dE_4 G^n(E_4) \int_{-\infty}^{E_V} dE_1 G^n(E_1) \int_{E_C}^{\infty} dE_3 G^p(E_3) \right) \delta(E_4 - E_3 - E_2 + E_1) \\ \Sigma_s^{in}(E_3) &= D \otimes \left(\int_{E_C}^{\infty} dE_2 G^p(E_2) \int_{E_C}^{\infty} dE_4 G^n(E_4) \int_{-\infty}^{E_V} dE_1 G^n(E_1) \right) \delta(E_4 - E_3 - E_2 + E_1) \\ \Sigma_s^{out}(E_4) &= D \otimes \left(\int_{-\infty}^{E_V} dE_1 G^n(E_1) \int_{E_C}^{\infty} dE_3 G^p(E_3) \int_{E_C}^{\infty} dE_2 G^p(E_2) \right) \delta(E_4 - E_3 - E_2 + E_1) \\ \Sigma_s^{out}(E_1) &= D \otimes \left(\int_{E_C}^{\infty} dE_3 G^p(E_3) \int_{E_C}^{\infty} dE_2 G^p(E_2) \int_{E_C}^{\infty} dE_4 G^n(E_4) \right) \delta(E_4 - E_3 - E_2 + E_1). \end{aligned} \tag{6.14}$$

In equations 6.13 and 6.14 an element wise multiplication is carried out within the bracketed

G^n/G^p terms. To conserve momentum, the D matrix must be chosen appropriately [138]. When the random potential is well-correlated throughout the channel in real space, *i.e.*, having the same value at all points of the matrix D, the momentum is conserved (Fourier transform of D into momentum space is a delta function ensuring there is no momentum loss). The equation of D is given below and in this study we consider d_0 to be an adjustable parameter.

$$D = d_0 \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & \cdots & \cdots \\ 1 & 1 & 1 & 1 & 1 & 1 & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & 1 & 1 & 1 & 1 & 1 & 1 \\ \cdots & \cdots & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix} \tag{6.15}$$

To compute the terminal currents with impact ionization in a 1D semiconductor, we consider the setup as shown in Fig. 6.8. The quasi-Fermi level of the left contact E_{F1} is set above the conduction band edge on the left. This facilitates the injection of electrons into the channel, mimicking the process of electron injection due to photon absorption. The

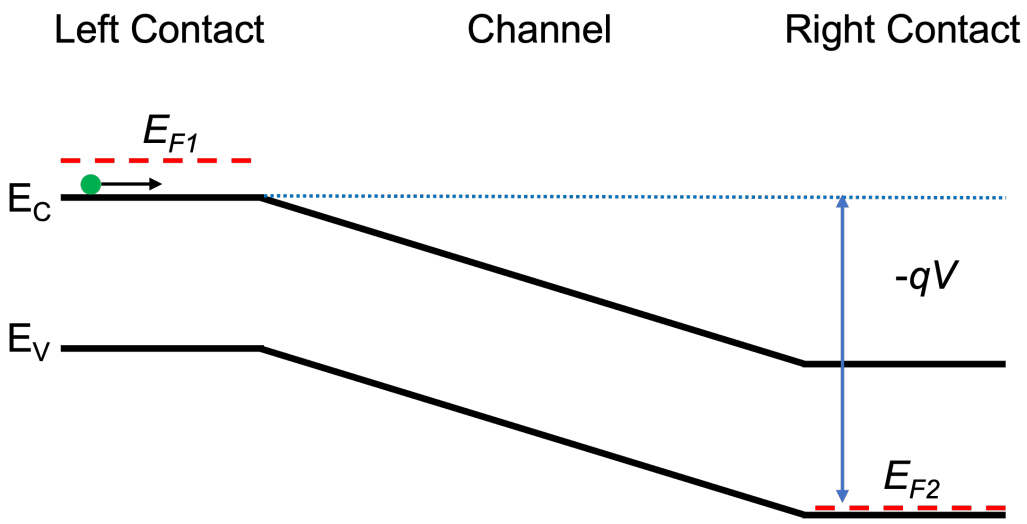

Figure 6.8: Potential diagram of a 1D semiconductor device for studying impact ionization.

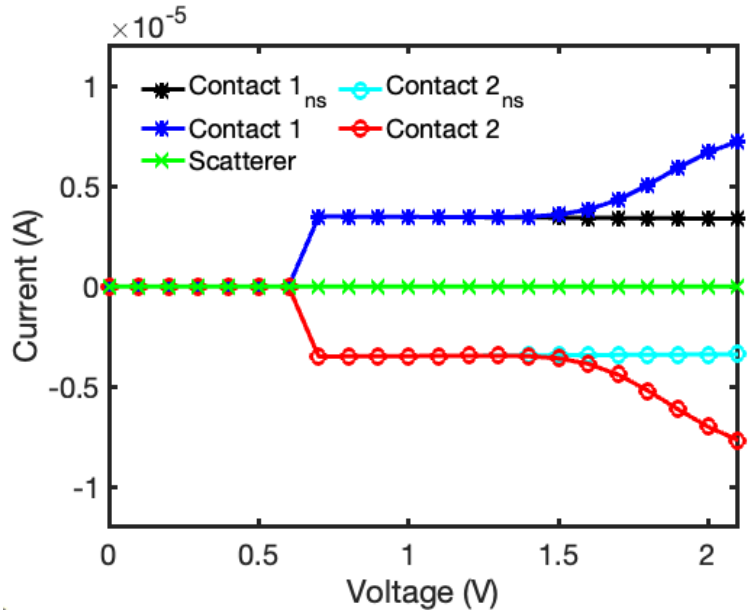

Figure 6.9: Current vs. voltage characteristics of a 1D semiconductor with impact ionization.

quasi-Fermi level of the right contact E_{F2} is fixed to the right valence band edge. The voltage V is the applied potential across the channel.

The simulated current vs. voltage characteristics of a 1D semiconducting dimer chain with a length of 80 dimers is shown in Fig. 6.9. The ballistic currents (without impact ionization), labeled Contacts 1_{ns} and 2_{ns}, are initially zero as the right-side conduction band is above the left-side one. The current increases as the right-sided conduction band goes the left-sided conduction band with bias. We set $E_{F1} = E_{C,LEFT} + 0.1eV$ in our simulation. Under impact ionization, the terminal currents (labeled Contact 1 and 2) increase after reaching the threshold voltage. This clearly indicates that electron impact ionization is happening in the system. In our simulation, we set $d_0 = 5$, the temperature $T = 3K$, and the the semiconductor considered has a bandgap $E_G = 0.6eV$. For the plot above, conduction and valence band effective masses are considered to be equal ($t_4 = 0$).

Fig. 6.10(a) depicts the impact ionization current (total terminal current-ballistic current) vs. voltage characteristics for different effective mass ratios μ, where $\mu = m_c^*/m_v^*$. We observe that the turn on voltage for the impact ionization increases with increasing μ. The impact ionization current increases with voltage because carriers with lower kinetic energy

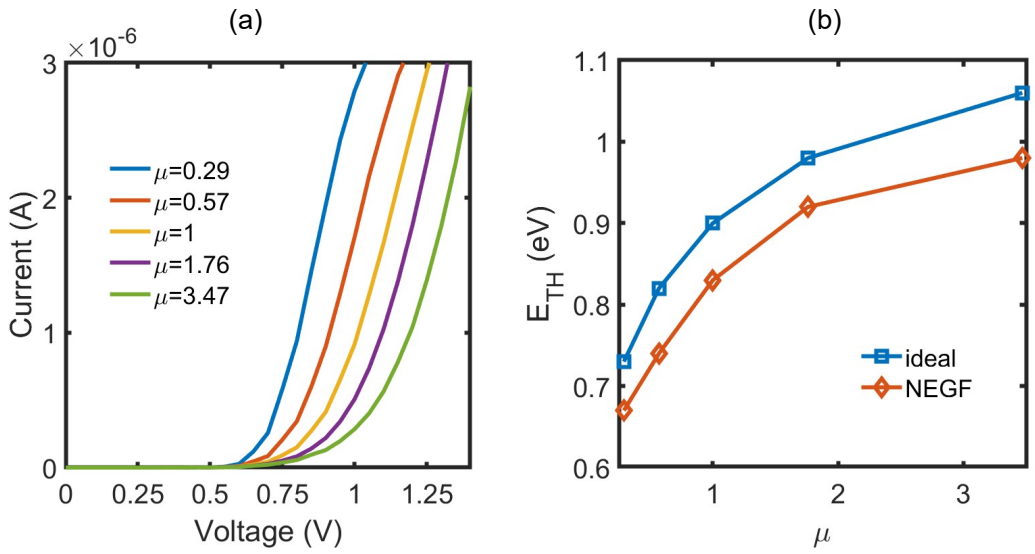

Figure 6.10: (a) Impact Ionization current vs. voltage characteristics for different mass ratios and (b) ideal and NEGF simulation threshold energy E_{TH} as a function of the mass ratio μ.

are able to impact ionize at higher voltages. The impact ionization behavior of the 1D semiconductor can be verified by studying the relationship between the threshold energy and the effective mass ratio of the bands. For a semiconductor with two parabolic bands, a conduction band and a valence band, the threshold energy can be approximated using the equation $E_{TH} = [(2\mu + 1)/(\mu + 1)]E_G$. The threshold energy as function of μ is plotted in Fig. 6.10(b) for a semiconductor with $E_G = 0.6eV$. E_{TH} approaches a value of $2E_G$ as $\mu \to \infty$ and E_{TH} goes towards E_G as $\mu \to 0$. From Fig. 6.10(a) we can extract an effective threshold voltage of the impact ionization current for the different mass ratios. This is done by drawing a tangent from the linearly rising part of each curve to the intersecting point on the voltage axis. The intersection point is defined as the threshold voltage. Then, the threshold energy for the NEGF simulations is defined as $E_{TH,NEGF} = eV_{TH,NEGF}$. The NEGF threshold energy exhibits the same trend as the ideal threshold energy, as seen in Fig. 6.10(b). The offset between the two threshold energies can be attributed to the extra kinetic energy of the injected electrons due to the quasi-Fermi level of the left contact E_{F1} being above E_C.

This model can also be used to simulate hole impact ionization by changing the integral

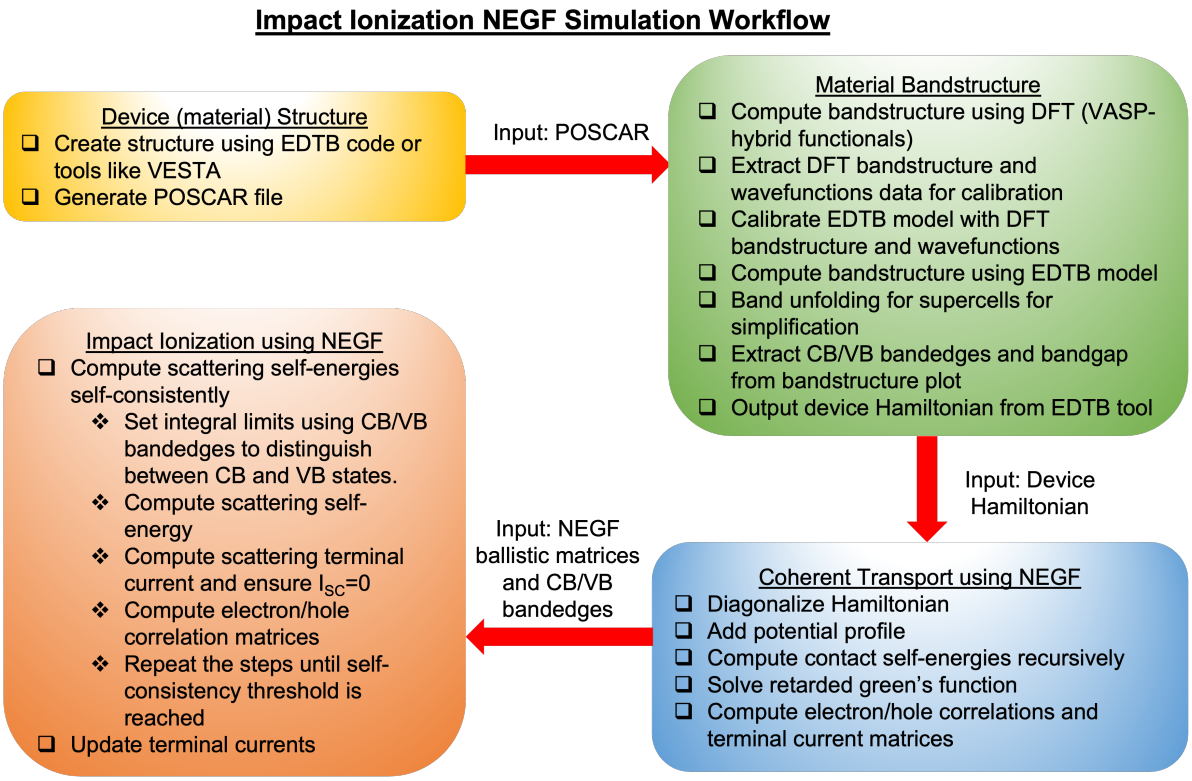

Figure 6.11: Workflow for full scale quantum transport impact ionization model.

limits in Eqs. 6.13 and 6.14. It is possible to extend this matrix based quantum mechanical treatment of impact ionization to devices with complicated material bandstructures and quantum effects like tunneling across minigaps. A tentative workflow for such a full-scale quantum transport simulator is provided in Fig. 6.11. This chart highlights the general methodology for developing such a tool based on the EDTB model in the future.

6.4 Summary

In this chapter, we develop a matrix based quantum transport model for impact ionization using the NEGF formalism. Initially, we create a model of a four-level system to comprehend how impact ionization works within this NEGF framework. The model is then extended to a one dimensional semiconductor with two parabolic bands. The model exhibits behavior that is expected of such a material. This framework lays the groundwork for developing a matrix

based quantum transport model that includes impact ionization for devices like APDs. The model would naturally allow us to account for multiple unfolded bands, their mass tensors, complex tunneling as well as non-parabolicities through an atomistic matrix Hamiltonian.

Key Contributions:

- Developed matrix-based quantum transport model for impact ionization.

Chapter 7

Conclusion and Future Works

The primary focus of this dissertation is understanding the device physics of modern *p-i-n* junction based devices to explain their experimental characteristics. This work explores the discrepancy between experimental and theoretical TFETs using a chemistry-based quasi-analytical model developed in this dissertation. It was identified that non-ideal processes, mainly trap-assisted tunneling and Auger generation, degrades the performance of experimental TFETs. Moreover, it is demonstrated that the Auger generation process limits the lower limit of subthreshold swing in a trap free III-V planar TFET. It is possible to attenuate the effect of Auger generation by tuning the source doping and effective masses. A key future research topic in this area is exploring device designs and material physics to mitigate the effects of Auger generation without affecting the on-state TFET performance. Expanding our study on Auger generation, also known as impact ionization, we then examined III-V digital alloy APDs.

This dissertation carries out a detailed theoretical investigation of the underlying physics of low noise III-V APDs using first-principles based tools. The materials properties of III-V digital and random alloys are studied using an Environment-Dependent Tight Binding model coupled with a band unfolding technique. The ballistic transport in the digital alloys is studied using a quasi-1D NEGF model developed as part of this thesis. The effect of

phonon scattering in these alloys is studied using a Boltzmann transport model. Based on my simulations, it is demonstrated that a combination of minigap, increased effective mass and sizeable offset between the light-hole and split-off bands results in hole localization near the valence band edge. This leads to a lower hole impact ionization coefficient that results in the low excess noise in these APDs. An interesting finding from this work is that introduction of heavy atoms like antimony, which has large spin-orbit coupling, results in the large light-hole/split-off offset. In the future, a detailed investigation on tuning this offset by varying the heavy atom composition or type in the III-V alloys and its consequent effects on APD noise can be carried out.

The origin of minigaps of the digital alloys is also investigated here and it is shown that alternating strain types result in the opening of these gaps. This work shows that the biaxial strain in the digital alloys can be used as a knob for controlling the minigap sizes. This knob can be used to enhance the low noise performance of digital alloys in APDs or allow the use of these alloys for other electronic or photonic applications. Furthermore, as part of this thesis, a multiscale compact model for *p-i-n* digital alloy APDs is developed that can be used to study digital alloys from the materials to the circuit level. The model can be used to study and simulate digital alloy APD based photonic integrated circuits.

Lastly, in this dissertation,a first of its kind matrix based quantum transport model for impact ionization using the NEGF formalism is developed. This framework can incorporate the effects of complicated bandstructures that generate non-parabolic, energy and voltage-dependent effective mass tensors, and quantum processes like tunneling. In the future, this model can be extended to include the full 3D EDTB Hamiltonian to accurately capture material properties. This can be done by integrating the impact ionization self-energies with the ballistic NEGF model developed in this dissertation. Furthermore, this model can be developed further to incorporate noise using the Blanter-Buttiker approach, which would allow us to study the noise performance of APDs using the quantum transport framework. This model lays the foundation for developing a powerful full-scale quantum transport model

for devices like APDs that incorporate impact ionization and the associated noise.

Appendices

Appendix A

Publications

A.1 Journal Articles

1. **S. Z. Ahmed**, Y. Tan, J. Zheng, J. C. Campbell & A. W. Ghosh, "Biaxial Strain Modulated Valence Band Engineering in III-V Digital Alloys," *arXiv preprint arXiv:2111.04247*, 2021.

2. **S. Z. Ahmed**, Y. Tan, J. Zheng, J. C. Campbell & A. W. Ghosh, "Atomistic transport modeling, design principles and empirical rules for Low Noise III-V Digital Alloy Avalanche Photodiodes," *arXiv preprint arXiv:2109.01995*, 2021.

3. **S. Z. Ahmed**, S. Ganguly, Y. Yuan, J. Zheng, Y. Tan, J. C. Campbell & A. W. Ghosh, "A Physics Based Multiscale Compact Model of p-i-n Avalanche Photodiodes," *Journal of Lightwave Technology*, 39.11, 3591-3598, 2021.

4. **S. Z. Ahmed**, D. S. Truesdell, Y. Tan, B. H. Calhoun & A. W. Ghosh, "A comprehensive analysis of Auger generation impacted planar Tunnel FETs." *Solid-State Electronics*, 107782, 2020.

5. D.S. Truesdell, **S. Z. Ahmed**, A. W. Ghosh & B. H Calhoun, "Minimum-Energy Digital Computing with Steep Subthreshold Swing Tunnel FETs." *IEEE Journal on Exploratory Solid-State Computational Devices and Circuits*, 6.2, 138-145, 2020.

6. J. Zheng, **S. Z. Ahmed**, Y. Yuan, A. Jones, Y. Tan, A. K. Rockwell, S. D. March, S. R. Bank, A. W. Ghosh & J. C. Campbell, "Full band Monte Carlo simulation of AlInAsSb digital alloys." InfoMat, 2: 1236– 1240, 2020.

7. J. Zheng, A. H. Jones, Y. Tan, A. K. Rockwell, S. D. March, **S. Z. Ahmed**, C. A. Dukes, A. W. Ghosh, S. R. Bank & J. C. Campbell, "Characterization of band offsets in $Al_xIn_{1-x}As_ySb_{1-y}$ alloys with varying Al composition." *Applied Physics Letters*, 115(12), 122105, 2019.

8. **S. Z. Ahmed**, Y. Tan, D. S. Truesdell, B. H. Calhoun & A. W. Ghosh, "Modeling tunnel field effect transistors– From interface chemistry to non-idealities to circuit level performance." *Journal of Applied Physics*, 124(15), 154503, 2018.

A.2 Conferences Proceedings

1. **S. Z. Ahmed**, J. Zheng, Y. Tan, J. C. Campbell & A. W. Ghosh, "A Comprehensive Modeling Approach of Electronic Properties in III-V Digital Alloys." *In 2021 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD)*, 2021.

2. **S. Z. Ahmed**, J. Zheng, Y. Tan, J. C. Campbell & A. W. Ghosh, "Development of Low Noise III-V Digital Alloys for Improved Photodetection." *In CLEO: Applications and Technology*, pp. JTu3A-18, 2021.

3. **S. Z. Ahmed**, J. Zheng, J. C. Campbell, Y. Tan & A. W. Ghosh, "Physics of Strain Engineered Minigaps in III-V Digital Alloys." *APS March Meeting 2021*, 2021.

4. **S. Z. Ahmed**, J. Zheng, Y. Tan, J. C. Campbell & A. W. Ghosh, "Design methodology of high-gain III-V digital alloy avalanche photodiodes." *In SPIE Physics and Simulation of Optoelectronic Devices XXIX*, Vol. 11680, p. 116800V, 2021.

5. S. Ganguly, F. F. Tonni, **S. Z. Ahmed**, P. Ghuman, S. Babu, N. K. Dhar, A .K. Sood & A. W. Ghosh, "Dissipative Quantum Transport Study of A Bi-Layer Graphene-CdTe-HgCdTe Heterostructure for MWIR Photodetector." *In 2019 EEE Research and Applications of Photonics in Defense Conference (RAPID)*, 2021.

6. **S. Z. Ahmed**, S. Ganguly, Y. Yuan, J. Zheng, J. C. Campbell & A. W. Ghosh, "A multiscale compact model of low noise pin avalanche photodiodes." *In SPIE Advanced Photon Counting Techniques XIV*, Vol. 11386, p. 1138607, 2020.

7. S. Ganguly, **S. Z. Ahmed**, A. W. Ghosh, A. K. Sood, J. Zeller, P. Ghuman, S. Babu & N. K. Dhar, "Density functional theory based bandstructure analysis of graphene-HgCdTe heterostructure mid-wave infrared detector for Earth science applications." *In SPIE Image Sensing Technologies: Materials, Devices, Systems, and Applications VII*, Vol. 11388, p. 1138803, 2020.

8. **S. Z. Ahmed**, J. Zheng, Y. Tan, J. C. Campbell & A.W. Ghosh, "Understanding the Role of Minigaps in APDs: Towards Designing a Better Photodetector." *In 2019 IEEE Photonics Conference*, 2019.

9. S. Ganguly, **S. Z. Ahmed**, A. W. Ghosh, P. Ghuman, S. Babu, N. K. Dhar & A .K. Sood, "First Principles Investigation into Graphene-PbSe MidWave IR (MWIR) Photodetector Physics." *In 2019 IEEE Photonics Conference*, 2019.

10. **S. Z. Ahmed**, Y. Tan, J. Zheng, J. C. Campbell & A.W. Ghosh, "APD performance enhancement: Minigap engineering in digital alloys." *In 2018 IEEE Photonics Conference*, 2018.

11. **S. Z. Ahmed**, Y. Tan, D. S. Truesdell, B. H. Calhoun & A. W. Ghosh, "Auger effect limited performance in tunnel field effect transistors." *In 2017 Fifth Berkeley Symposium on Energy Efficient Electronic Systems & Steep Transistors Workshop (E3S)*, 2017.

Bibliography

[1] S. Jordan, "The promise of silicon photonics," May 2019.

[2] A. C. Seabaugh and Q. Zhang, "Low-voltage tunnel transistors for beyond cmos logic," *Proceedings of the IEEE*, vol. 98, pp. 2095–2110, Dec 2010.

[3] A. I. Khan, C. W. Yeung, C. Hu, and S. Salahuddin, "Ferroelectric negative capacitance mosfet: Capacitance tuning & antiferroelectric operation," in *2011 International Electron Devices Meeting*, pp. 11.3.1–11.3.4, Dec 2011.

[4] C. Sun, M. T. Wade, Y. Lee, J. S. Orcutt, L. Alloatti, M. S. Georgas, A. S. Waterman, J. M. Shainline, R. R. Avizienis, S. Lin, *et al.*, "Single-chip microprocessor that communicates directly using light," *Nature*, vol. 528, no. 7583, pp. 534–538, 2015.

[5] R. Barends, J. Kelly, A. Megrant, A. Veitia, D. Sank, E. Jeffrey, T. C. White, J. Mutus, A. G. Fowler, B. Campbell, *et al.*, "Superconducting quantum circuits at the surface code threshold for fault tolerance," *Nature*, vol. 508, no. 7497, pp. 500–503, 2014.

[6] S. Li, L. Da Xu, and S. Zhao, "5g internet of things: A survey," *Journal of Industrial Information Integration*, vol. 10, pp. 1–9, 2018.

[7] M. Z. Chowdhury, M. K. Hasan, M. Shahjalal, E. B. Shin, and Y. M. Jang, "Opportunities of optical spectrum for future wireless communications," in *2019 International Conference on Artificial Intelligence in Information and Communication (ICAIIC)*, pp. 004–007, Feb 2019.

[8] A. Y. Liu and J. Bowers, "Photonic integration with epitaxial iii–v on silicon," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 24, pp. 1–12, Nov 2018.

[9] A. Tosi, N. Calandri, M. Sanzaro, and F. Acerbi, "Low-noise, low-jitter, high detection efficiency ingaas/inp single-photon avalanche diode," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 20, no. 6, pp. 192–197, 2014.

[10] J. C. Campbell, "Advances in photodetectors," in *Optical Fiber Telecommunications VA*, pp. 221–268, Elsevier, 2008.

[11] N. Bertone and W. Clark, "Avalanche photodiode arrays provide versatility in ultrasensitive applications," *Laser focus world*, vol. 43, no. 9, 2007.

[12] P. Mitra, J. D. Beck, M. R. Skokan, J. E. Robinson, J. Antoszewski, K. J. Winchester, A. J. Keating, T. Nguyen, K. K. M. B. D. Silva, C. A. Musca, J. M. Dell, and L. Faraone, "Adaptive focal plane array (AFPA) technologies for integrated infrared microsystems," in *Intelligent Integrated Microsystems* (R. A. Athale and J. C. Zolper, eds.), vol. 6232, pp. 70 – 80, International Society for Optics and Photonics, SPIE, 2006.

[13] G. M. Williams, "Optimization of eyesafe avalanche photodiode lidar for automobile safety and autonomous navigation systems," *Optical Engineering*, vol. 56, no. 3, pp. 1 – 9, 2017.

[14] M. Nada, F. Nakajima, T. Yoshimatsu, Y. Nakanishi, S. Tatsumi, Y. Yamada, K. Sano, and H. Matsuzaki, "High-speed iii-v based avalanche photodiodes for optical communications-the forefront and expanding applications," *Applied Physics Letters*, vol. 116, no. 14, p. 140502, 2020.

[15] K. Pasquinelli, R. Lussana, S. Tisa, F. Villa, and F. Zappa, "Single-photon detectors modeling and selection criteria for high-background lidar," *IEEE Sensors Journal*, vol. 20, no. 13, pp. 7021–7032, 2020.

[16] H. Lu and A. Seabaugh, "Tunnel field-effect transistors: State-of-the-art," *IEEE Journal of the Electron Devices Society*, vol. 2, pp. 44–49, July 2014.

[17] A. W. Ghosh, "Transmission engineering as a route to subthermal switching," *IEEE Journal of the Electron Devices Society*, vol. 3, pp. 135–143, May 2015.

[18] J. Appenzeller, Y.-M. Lin, J. Knoch, and P. Avouris, "Band-to-band tunneling in carbon nanotube field-effect transistors," *Phys. Rev. Lett.*, vol. 93, p. 196805, Nov 2004.

[19] D. Sarkar, X. Xie, W. Liu, W. Cao, J. Kang, Y. Gong, S. Kraemer, P. M. Ajayan, and K. Banerjee, "A subthermionic tunnel field-effect transistor with an atomically thin channel," *Nature*, vol. 526, no. 7571, p. 91, 2015.

[20] R. Pandey, S. Mookerjea, and S. Datta, "Opportunities and challenges of tunnel fets," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 63, pp. 2128–2138, Dec 2016.

[21] X. Zhao, A. Vardi, and J. A. del Alamo, "Sub-thermal subthreshold characteristics in top–down ingaas/inas heterojunction vertical nanowire tunnel fets," *IEEE Electron Device Letters*, vol. 38, pp. 855–858, July 2017.

[22] E. Memisevic, M. Hellenbrand, E. Lind, A. R. Persson, S. Sant, A. Schenk, J. Svensson, R. Wallenberg, and L.-E. Wernersson, "Individual defects in inas/ingaassb/gasb nanowire tunnel field-effect transistors operating below 60 mv/decade," *Nano Letters*, vol. 17, no. 7, pp. 4373–4380, 2017. PMID: 28613894.

[23] T. Vasen, P. Ramvall, A. Afzalian, G. Doornbos, M. Holland, C. Thelander, K. Dick, L.-E. Wernersson, and M. Passlack, "Vertical gate-all-around nanowire gasb-inas core-shell n-type tunnel fets," *Scientific reports*, vol. 9, no. 1, p. 202, 2019.

[24] J. C. Campbell, "Recent advances in avalanche photodiodes," *Journal of Lightwave Technology*, vol. 34, pp. 278–285, Jan 2016.

[25] R. McIntyre, "Multiplication noise in uniform avalanche diodes," *IEEE Transactions on Electron Devices*, no. 1, pp. 164–168, 1966.

[26] M. Teich, K. Matsuo, and B. Saleh, "Excess noise factors for conventional and superlattice avalanche photodiodes and photomultiplier tubes," *IEEE Journal of Quantum Electronics*, vol. 22, no. 8, pp. 1184–1193, 1986.

[27] N. Z. Hakim, B. E. A. Saleh, and M. C. Teich, "Generalized excess noise factor for avalanche photodiodes of arbitrary structure," *IEEE Transactions on Electron Devices*, vol. 37, no. 3, pp. 599–610, 1990.

[28] S. Z. Ahmed, S. Ganguly, Y. Yuan, J. Zheng, Y. Tan, J. C. Campbell, and A. W. Ghosh, "A physics based multiscale compact model of pin avalanche photodiodes," *Journal of Lightwave Technology*, vol. 39, no. 11, pp. 3591–3598, 2021.

[29] J. Zheng, Y. Yuan, Y. Tan, Y. Peng, A. K. Rockwell, S. R. Bank, A. W. Ghosh, and J. C. Campbell, "Digital alloy inalas avalanche photodiodes," *Journal of Lightwave Technology*, vol. 36, no. 17, pp. 3580–3585, 2018.

[30] J. Zheng, Y. Tan, Y. Yuan, A. Ghosh, and J. Campbell, "Strain effect on band structure of inalas digital alloy," *Journal of Applied Physics*, vol. 125, no. 8, p. 082514, 2019.

[31] S. R. Bank, J. C. Campbell, S. J. Maddox, A. K. Rockwell, M. E. Woodson, M. Ren, A. Jones, S. March, J. Zheng, and Y. Yuan, "Digital alloy growth of low-noise avalanche photodiodes," in *2018 IEEE RAPID*, pp. 1–3, IEEE, 2018.

[32] X. Yi, S. Xie, B. Liang, L. W. Lim, J. S. Cheong, M. C. Debnath, D. L. Huffaker, C. H. Tan, and J. P. David, "Extremely low excess noise and high sensitivity $\mathrm{alas_{0.56}sb_{0.44}}$ avalanche photodiodes," *Nature Photonics*, vol. 13, no. 10, pp. 683–686, 2019.

[33] J. Taylor-Mew, V. Shulyak, B. White, C. H. Tan, and J. S. Ng, "Low excess noise of $\mathrm{al_{0.85}\ ga_{0.15}as_{0.56}sb_{0.44}}$ avalanche photodiode from pure electron injection," *IEEE Photonics Technology Letters*, vol. 33, no. 20, pp. 1155–1158, 2021.

[34] S. H. Kodati, S. Lee, B. Guo, A. H. Jones, M. Schwartz, M. Winslow, N. A. Pfiester, C. H. Grein, T. J. Ronningen, J. C. Campbell, and S. Krishna, "Alinassb avalanche photodiodes on inp substrates," *Applied Physics Letters*, vol. 118, no. 9, p. 091101, 2021.

[35] S. Lee, S. H. Kodati, B. Guo, A. H. Jones, M. Schwartz, M. Winslow, C. H. Grein, T. J. Ronningen, J. C. Campbell, and S. Krishna, "Low noise $\mathrm{al_{0.85}\ ga_{0.15}as_{0.56}sb_{0.44}}$ avalanche photodiodes on inp substrates," *Applied Physics Letters*, vol. 118, no. 8, p. 081106, 2021.

[36] J. Zheng, S. Z. Ahmed, Y. Yuan, A. Jones, Y. Tan, A. K. Rockwell, S. D. March, S. R. Bank, A. W. Ghosh, and J. C. Campbell, "Full band monte carlo simulation of alinassb digital alloys," *InfoMat*, vol. 2, no. 6, pp. 1236–1240, 2020.

[37] J. Zheng, Y. Yuan, Y. Tan, Y. Peng, A. Rockwell, S. R. Bank, A. W. Ghosh, and J. C. Campbell, "Simulations for inalas digital alloy avalanche photodiodes," *Applied Physics Letters*, vol. 115, no. 17, p. 171106, 2019.

[38] K. Bernstein, R. K. Cavin, W. Porod, A. Seabaugh, and J. Welser, "Device and architecture outlook for beyond cmos switches," *Proceedings of the IEEE*, vol. 98, pp. 2169–2184, Dec 2010.

[39] W. Y. Choi, B. G. Park, J. D. Lee, and T. J. K. Liu, "Tunneling field-effect transistors (tfets) with subthreshold swing (ss) less than 60 mv/dec," *IEEE Electron Device Letters*, vol. 28, pp. 743–745, Aug 2007.

[40] N. Shukla, A. V. Thathachary, A. Agrawal, H. Paik, A. Aziz, D. G. Schlom, S. K. Gupta, R. Engel-Herbert, and S. Datta, "A steep-slope transistor based on abrupt electronic phase transition," *Nature communications*, vol. 6, p. 7812, 2015.

[41] R. N. Sajjad and A. W. Ghosh, "Manipulating chiral transmission by gate geometry: switching in graphene with transmission gaps," *ACS nano*, vol. 7, no. 11, pp. 9808–9813, 2013.

[42] M. Luisier and G. Klimeck, "Performance comparisons of tunneling field-effect transistors made of insb, carbon, and gasb-inas broken gap heterostructures," in *Electron Devices Meeting (IEDM), 2009 IEEE International*, pp. 1–4, IEEE, 2009.

[43] U. E. Avci, S. Hasan, D. E. Nikonov, R. Rios, K. Kuhn, and I. A. Young, "Understanding the feasibility of scaled iii-v tfet for logic by bridging atomistic simulations and experimental results," in *2012 Symposium on VLSI Technology (VLSIT)*, pp. 183–184, June 2012.

[44] J. Z. Huang, P. Long, M. Povolotskyi, G. Klimeck, and M. J. W. Rodwell, "P-type tunnel fets with triple heterojunctions," *IEEE Journal of the Electron Devices Society*, vol. 4, pp. 410–415, Nov 2016.

[45] P. Long, J. Z. Huang, M. Povolotskyi, G. Klimeck, and M. J. W. Rodwell, "High-current tunneling fets with ($1\bar{1}0$) orientation and a channel heterojunction," *IEEE Electron Device Letters*, vol. 37, pp. 345–348, March 2016.

[46] M. Luisier and G. Klimeck, "Atomistic full-band simulations of silicon nanowire transistors: Effects of electron-phonon scattering," *Physical Review B*, vol. 80, no. 15, p. 155430, 2009.

[47] R. N. Sajjad, W. Chern, J. L. Hoyt, and D. A. Antoniadis, "Trap assisted tunneling and its effect on subthreshold swing of tunnel fets," *IEEE Transactions on Electron Devices*, vol. 63, pp. 4380–4387, Nov 2016.

[48] H. Carrillo-Nunez, R. Rhyner, M. Luisier, and A. Schenk, "Effect of surface roughness and phonon scattering on extremely narrow inas-si nanowire tfets," in *Solid-State Device Research Conference (ESSDERC), 2016 46th European*, pp. 188–191, IEEE, 2016.

[49] J. T. Teherani, W. Chern, S. Agarwal, J. L. Hoyt, and D. A. Antoniadis, "A framework for generation and recombination in tunneling field-effect transistors," in *2015 Fourth Berkeley Symposium on Energy Efficient Electronic Systems (E3S)*, pp. 1–3, Oct 2015.

[50] L. Zhang and M. Chan, "Spice modeling of double-gate tunnel-fets including channel transports," *IEEE Transactions on Electron Devices*, vol. 61, pp. 300–307, Feb 2014.

[51] R. Vishnoi and M. J. Kumar, "Compact analytical drain current model of gate-all-around nanowire tunneling fet," *IEEE Transactions on Electron Devices*, vol. 61, pp. 2599–2603, July 2014.

[52] R. Vishnoi and M. J. Kumar, "An accurate compact analytical model for the drain current of a tfet from subthreshold to strong inversion," *IEEE Transactions on Electron Devices*, vol. 62, pp. 478–484, Feb 2015.

[53] H. Lu, D. Esseni, and A. Seabaugh, "Universal analytic model for tunnel fet circuit simulation," *Solid-State Electronics*, vol. 108, no. Supplement C, pp. 110 – 117, 2015. Selected papers from the 15th Ultimate Integration on Silicon (ULIS) conference.

[54] E. Kane, "Zener tunneling in semiconductors," *Journal of Physics and Chemistry of Solids*, vol. 12, no. 2, pp. 181 – 188, 1960.

[55] J. G. Simmons, "Electric tunnel effect between dissimilar electrodes separated by a thin insulating film," *Journal of Applied Physics*, vol. 34, no. 9, pp. 2581–2590, 1963.

[56] J. T. Teherani, S. Agarwal, W. Chern, P. M. Solomon, E. Yablonovitch, and D. A. Antoniadis, "Auger generation as an intrinsic limit to tunneling field-effect transistor performance," *Journal of Applied Physics*, vol. 120, no. 8, p. 084507, 2016.

[57] S. Z. Ahmed, Y. Tan, D. S. Truesdell, B. H. Calhoun, and A. W. Ghosh, "Modeling tunnel field effect transistors—from interface chemistry to non-idealities to circuit level performance," *Journal of Applied Physics*, vol. 124, no. 15, p. 154503, 2018.

[58] S. Z. Ahmed, D. S. Truesdell, Y. Tan, B. H. Calhoun, and A. W. Ghosh, "A comprehensive analysis of auger generation impacted planar tunnel fets," *Solid-State Electronics*, vol. 169, p. 107782, 2020.

[59] C. Shen, S. L. Ong, C. H. Heng, G. Samudra, and Y. C. Yeo, "A variational approach to the two-dimensional nonlinear poisson's equation for the modeling of tunneling transistors," *IEEE Electron Device Letters*, vol. 29, pp. 1252–1255, Nov 2008.

[60] K. K. Young, "Short-channel effect in fully depleted soi mosfets," *IEEE Transactions on Electron Devices*, vol. 36, pp. 399–402, Feb 1989.

[61] M. G. Bardon, H. P. Neves, R. Puers, and C. V. Hoof, "Pseudo-two-dimensional model for double-gate tunnel fets considering the junctions depletion regions," *IEEE Transactions on Electron Devices*, vol. 57, pp. 827–834, April 2010.

[62] Y. P. Tan, M. Povolotskyi, T. Kubis, T. B. Boykin, and G. Klimeck, "Tight-binding analysis of si and gaas ultrathin bodies with subatomic wave-function resolution," *Phys. Rev. B*, vol. 92, p. 085301, Aug 2015.

[63] Y. Tan, M. Povolotskyi, T. Kubis, T. B. Boykin, and G. Klimeck, "Transferable tight-binding model for strained group iv and iii-v materials and heterostructures," *Phys. Rev. B*, vol. 94, p. 045311, Jul 2016.

[64] J. Heyd, G. E. Scuseria, and M. Ernzerhof, "Hybrid functionals based on a screened coulomb potential," *The Journal of chemical physics*, vol. 118, no. 18, pp. 8207–8215, 2003.

[65] D. Kienle, J. I. Cerda, and A. W. Ghosh, "Extended hückel theory for band structure, chemistry, and transport. i. carbon nanotubes," *Journal of Applied Physics*, vol. 100, no. 4, p. 043714, 2006.

[66] D. Kienle, K. H. Bevan, G.-C. Liang, L. Siddiqui, J. I. Cerda, and A. W. Ghosh, "Extended hückel theory for band structure, chemistry, and transport. ii. silicon," *Journal of Applied Physics*, vol. 100, no. 4, p. 043715, 2006.

[67] J.-M. Jancu, R. Scholz, F. Beltram, and F. Bassani, "Empirical spds* tight-binding calculation for cubic semiconductors: General method and material parameters," *Physical Review B*, vol. 57, no. 11, p. 6493, 1998.

[68] Y. Tan, F. W. Chen, and A. W. Ghosh, "First principles study and empirical parametrization of twisted bilayer mos_2 based on band-unfolding," *Applied Physics Letters*, vol. 109, no. 10, p. 101601, 2016.

[69] T. B. Boykin, N. Kharche, and G. Klimeck, "Brillouin-zone unfolding of perfect supercells having nonequivalent primitive cells illustrated with a si/ge tight-binding parameterization," *Physical Review B*, vol. 76, no. 3, p. 035310, 2007.

[70] T. B. Boykin, N. Kharche, G. Klimeck, and M. Korkusinski, "Approximate bandstructures of semiconductor alloys from tight-binding supercell calculations," *Journal of Physics: Condensed Matter*, vol. 19, no. 3, p. 036203, 2007.

[71] A. Ghosh, *Nanoelectronics - a Molecular View*. World Scientific, 2016.

[72] S. O. Koswatta, M. S. Lundstrom, and D. E. Nikonov, "Influence of phonon scattering on the performance of p-i-n band-to-band tunneling transistors," *Applied Physics Letters*, vol. 92, no. 4, p. 043125, 2008.

[73] R. N. Sajjad and D. Antoniadis, "A compact model for tunnel fet for all operation regimes including trap assisted tunneling," in *2016 74th Annual Device Research Conference (DRC)*, pp. 1–2, June 2016.

[74] Q. Shao, C. Zhao, J. Zhang, L. Zhang, and Z. Yu, "Compact model of nanowire tunneling fets including phonon-assisted tunneling and quantum capacitance," *arXiv preprint arXiv:1412.1902*, 2014.

[75] G. Dewey, B. Chu-Kung, J. Boardman, J. M. Fastenau, J. Kavalieros, R. Kotlyar, W. K. Liu, D. Lubyshev, M. Metz, N. Mukherjee, P. Oakey, R. Pillarisetty, M. Radosavljevic, H. W. Then, and R. Chau, "Fabrication, characterization, and physics of iii -v heterojunction tunneling field effect transistors (h-tfet) for steep sub-threshold swing," in *2011 International Electron Devices Meeting*, pp. 33.6.1–33.6.4, Dec 2011.

[76] N. Neophytou, A. Paul, and G. Klimeck, "Bandstructure effects in silicon nanowire hole transport," *IEEE Transactions on Nanotechnology*, vol. 7, no. 6, pp. 710–719, 2008.

[77] G. Klimeck, R. C. Bowen, and T. B. Boykin, "Strong wavevector dependence of hole transport in heterostructures," *Superlattices and Microstructures*, vol. 29, no. 3, pp. 187–216, 2001.

[78] A. Rockwell, M. Ren, M. Woodson, A. Jones, S. March, Y. Tan, Y. Yuan, Y. Sun, R. Hool, S. Maddox, *et al.*, "Toward deterministic construction of low noise avalanche photodetector materials," *Applied Physics Letters*, vol. 113, no. 10, p. 102106, 2018.

[79] Y. Yuan, J. Zheng, Y. Tan, Y. Peng, A.-K. Rockwell, S. R. Bank, A. Ghosh, and J. C. Campbell, "Temperature dependence of the ionization coefficients of inalas and algaas digital alloys," *Photonics Research*, vol. 6, no. 8, pp. 794–799, 2018.

[80] J. Towns, T. Cockerill, M. Dahan, I. Foster, K. Gaither, A. Grimshaw, V. Hazlewood, S. Lathrop, D. Lifka, G. D. Peterson, R. Roskies, J. Scott, and N. Wilkins-Diehr, "Xsede: Accelerating scientific discovery," *Computing in Science & Engineering*, vol. 16, pp. 62–74, sep 2014.

[81] S. Z. Ahmed, Y. Tan, J. Zheng, J. C. Campbell, and A. W. Ghosh, "Atomistic transport modeling, design principles and empirical rules for low noise iii-v digital alloy avalanche photodiodes," *arXiv preprint arXiv:2109.01995*, 2021.

[82] J. A. Støvneng and P. Lipavský, "Multiband tight-binding approach to tunneling in semiconductor heterostructures: Application to Γx transfer in gaas," *Phys. Rev. B*, vol. 49, pp. 16494–16504, Jun 1994.

[83] B. Stickler, *Theory and modeling of spin-transport on the microscopic and the mesoscopic scale*. na, 2013.

[84] S. Datta, "Nanoscale device modeling: the green's function method," *Superlattices and microstructures*, vol. 28, no. 4, pp. 253–278, 2000.

[85] C. Lee, R. Logan, R. Batdorf, J. Kleimack, and W. Wiegmann, "Ionization rates of holes and electrons in silicon," *Physical review*, vol. 134, no. 3A, p. A761, 1964.

[86] J. Conradi, "The distribution of gains in uniformly multiplying avalanche photodiodes: Experimental," *IEEE Transactions on Electron Devices*, vol. 19, no. 6, pp. 713–718, 1972.

[87] W. Grant, "Electron and hole ionization rates in epitaxial silicon at high electric fields," *Solid-State Electronics*, vol. 16, no. 10, pp. 1189–1203, 1973.

[88] T. Kaneda, H. Matsumoto, and T. Yamaoka, "A model for reach-through avalanche photodiodes (rapd's)," *Journal of Applied Physics*, vol. 47, no. 7, pp. 3135–3139, 1976.

[89] A. Marshall, C. Tan, M. Steer, and J. David, "Electron dominated impact ionization and avalanche gain characteristics in inas photodiodes," *Applied Physics Letters*, vol. 93, no. 11, p. 111107, 2008.

[90] A. R. Marshall, P. J. Ker, A. Krysa, J. P. David, and C. H. Tan, "High speed inas electron avalanche photodiodes overcome the conventional gain-bandwidth product limit," *Optics express*, vol. 19, no. 23, pp. 23341–23349, 2011.

[91] W. Sun, S. J. Maddox, S. R. Bank, and J. C. Campbell, "Record high gain from inas avalanche photodiodes at room temperature," in *72nd Device Research Conference*, pp. 47–48, IEEE, 2014.

[92] W. Sun, Z. Lu, X. Zheng, J. C. Campbell, S. J. Maddox, H. P. Nair, and S. R. Bank, "High-gain inas avalanche photodiodes," *IEEE Journal of Quantum Electronics*, vol. 49, no. 2, pp. 154–161, 2012.

[93] P. J. Ker, A. R. Marshall, A. B. Krysa, J. P. David, and C. H. Tan, "Inas electron avalanche photodiodes with 580 ghz gain-bandwidth product," in *2012 17th Opto-Electronics and Communications Conference*, pp. 220–221, IEEE, 2012.

[94] J. D. Beck, C.-F. Wan, M. A. Kinch, and J. E. Robinson, "Mwir hgcdte avalanche photodiodes," in *Materials for Infrared Detectors*, vol. 4454, pp. 188–197, International Society for Optics and Photonics, 2001.

[95] J. D. Beck, C.-F. Wan, M. A. Kinch, J. E. Robinson, P. Mitra, R. E. Scritchfield, F. Ma, and J. C. Campbell, "The hgcdte electron avalanche photodiode," in *Infrared Detector Materials and Devices*, vol. 5564, pp. 44–53, International Society for Optics and Photonics, 2004.

[96] Yu Ling Goh, Jo Shien Ng, Chee Hing Tan, W. K. Ng, and J. P. R. David, "Excess noise measurement in $in_{0.53}ga_{0.47}as$," *IEEE Photonics Technology Letters*, vol. 17, no. 11, pp. 2412–2414, 2005.

[97] C. Lenox, P. Yuan, H. Nie, O. Baklenov, C. Hansing, J. Campbell, A. Holmes Jr, and B. Streetman, "Thin multiplication region inalas homojunction avalanche photodiodes," *Applied physics letters*, vol. 73, no. 6, pp. 783–784, 1998.

[98] S. Z. Ahmed, Y. Tan, J. Zheng, J. C. Campbell, and A. W. Ghosh, "Apd performance enhancement: Minigap engineering in digital alloys," in *2018 IEEE Photonics Conference (IPC)*, pp. 1–2, IEEE, 2018.

[99] J. Piprek, *Semiconductor optoelectronic devices: introduction to physics and simulation*. Elsevier, 2013.

[100] M. S. Shur, *Handbook series on semiconductor parameters*, vol. 1. World Scientific, 1996.

[101] W. Fawcett, A. Boardman, and S. Swain, "Monte carlo determination of electron transport properties in gallium arsenide," *Journal of Physics and Chemistry of Solids*, vol. 31, no. 9, pp. 1963–1990, 1970.

[102] Q. L. Yang, H. X. Deng, S. H. Wei, S. S. Li, and J. W. Luo, "Materials design principles towards high hole mobility learning from an abnormally low hole mobility of silicon," 2020.

[103] V. Palankovski, *Simulation of Heterojunction Bipolar Transistors*. 12 2000.

[104] B. K. Ridley, *Quantum processes in semiconductors*. Oxford university press, 2013.

[105] J. Xie, S. Xie, R. C. Tozer, and C. H. Tan, "Excess noise characteristics of thin alassb apds," *IEEE Transactions on Electron Devices*, vol. 59, no. 5, pp. 1475–1479, 2012.

[106] Y. Liu, X. Yi, N. J. Bailey, Z. Zhou, T. B. Rockett, L. W. Lim, C. H. Tan, R. D. Richards, and J. P. David, "Valence band engineering of gaasbi for low noise avalanche photodiodes," *Nature Communications*, vol. 12, no. 1, pp. 1–8, 2021.

[107] S. Z. Ahmed, Y. Tan, J. Zheng, J. C. Campbell, and A. W. Ghosh, "Biaxial strain modulated valence band engineering in iii-v digital alloys," *arXiv preprint arXiv:2111.04247*, 2021.

[108] Y. Peter and M. Cardona, *Fundamentals of semiconductors: physics and materials properties*. Springer Science & Business Media, 2010.

[109] S.-H. Wei and A. Zunger, "Predicted band-gap pressure coefficients of all diamond and zinc-blende semiconductors: Chemical trends," *Phys. Rev. B*, vol. 60, pp. 5404–5411, Aug 1999.

[110] F. H. Pollak and M. Cardona, "Piezo-electroreflectance in ge, gaas, and si," *Phys. Rev.*, vol. 172, pp. 816–837, Aug 1968.

[111] C. G. Van de Walle and R. M. Martin, "Theoretical calculations of heterojunction discontinuities in the si/ge system," *Phys. Rev. B*, vol. 34, pp. 5621–5634, Oct 1986.

[112] C. G. Van de Walle, "Band lineups and deformation potentials in the model-solid theory," *Phys. Rev. B*, vol. 39, pp. 1871–1883, Jan 1989.

[113] S.-H. Wei and A. Zunger, "Fingerprints of cupt ordering in iii-v semiconductor alloys: Valence-band splittings, band-gap reduction, and x-ray structure factors," *Phys. Rev. B*, vol. 57, pp. 8983–8988, Apr 1998.

[114] P. R. C. Kent, G. L. W. Hart, and A. Zunger, "Biaxial strain-modified valence and conduction band offsets of zinc-blende gan, gap, gaas, inn, inp, and inas, and optical bowing of strained epitaxial ingan alloys," *Applied Physics Letters*, vol. 81, no. 23, pp. 4377–4379, 2002.

[115] G. Bir and G. Pikus, *Symmetry and Strain-induced Effects in Semiconductors*. A Halsted Press book, Wiley, 1974.

[116] C. Mailhiot and D. L. Smith, "Long-wavelength infrared detectors based on strained inas-ga$_{1-x}$in$_x$sb type-ii superlattices," *Journal of Vacuum Science & Technology A*, vol. 7, no. 2, pp. 445–449, 1989.

[117] Y. Sun, S. E. Thompson, and T. Nishida, "Physics of strain effects in semiconductors and metal-oxide-semiconductor field-effect transistors," *Journal of Applied Physics*, vol. 101, no. 10, p. 104503, 2007.

[118] D. J. Chadi and M. L. Cohen, "Tight-binding calculations of the valence bands of diamond and zincblende crystals," *physica status solidi (b)*, vol. 68, no. 1, pp. 405–419, 1975.

[119] W. A. Harrison, *Electronic structure and the properties of solids: the physics of the chemical bond*. Courier Corporation, 2012.

[120] J. C. Slater and G. F. Koster, "Simplified lcao method for the periodic potential problem," *Phys. Rev.*, vol. 94, pp. 1498–1524, Jun 1954.

[121] S. D. Personick, "Receiver design for digital fiber optic communication systems, parts i and ii," *Bell system technical journal*, vol. 52, no. 6, pp. 843–886, 1973.

[122] R. Smith and S. Personick, "Receiver design for optical fiber communication systems," in *Semiconductor devices for optical communication*, pp. 89–160, Springer, 1980.

[123] S. Forrest, "Sensitivity of avalanche photodetector receivers for high-bit-rate long-wavelength optical communication systems," in *Semiconductors and Semimetals*, vol. 22, pp. 329–387, Elsevier, 1985.

[124] B. Kasper and J. Campbell, "Multigigabit-per-second avalanche photodiode lightwave receivers," *Journal of lightwave technology*, vol. 5, no. 10, pp. 1351–1364, 1987.

[125] X. Jiang, M. Itzler, K. O'Donnell, M. Entwistle, M. Owens, K. Slomkowski, and S. Rangwala, "Inp-based single-photon detectors and geiger-mode apd arrays for quantum communications applications," *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 21, no. 3, pp. 5–16, 2014.

[126] W. Chen and S. Liu, "Pin avalanche photodiodes model for circuit simulation," *IEEE Journal of Quantum Electronics*, vol. 32, no. 12, pp. 2105–2111, 1996.

[127] J.-J. Jou, C.-K. Liu, C.-M. Hsiao, H.-H. Lin, and H.-C. Lee, "Time-delay circuit model of high-speed pin photodiodes," *IEEE Photonics technology letters*, vol. 14, no. 4, pp. 525–527, 2002.

[128] A. Banoushi, M. Kardan, and M. A. Naeini, "A circuit model simulation for separate absorption, grading, charge, and multiplication avalanche photodiodes," *Solid-state electronics*, vol. 49, no. 6, pp. 871–877, 2005.

[129] G. Dunn, G. Rees, J. David, S. Plimmer, and D. Herbert, "Monte carlo simulation of impact ionization and current multiplication in short gaas diodes," *Semiconductor science and technology*, vol. 12, no. 1, p. 111, 1997.

[130] J. Zheng, L. Wang, X. Wu, Z. Hao, C. Sun, B. Xiong, Y. Luo, Y. Han, J. Wang, H. Li, *et al.*, "A pmt-like high gain avalanche photodiode based on gan/aln periodically stacked structure," *Applied Physics Letters*, vol. 109, no. 24, p. 241105, 2016.

[131] S. Ganguly, M.-H. Jang, Y. Tan, S.-S. Yoo, M. C. Gupta, and A. W. Ghosh, "A multi-scale materials-to-systems modeling of polycrystalline pbse photodetectors," *Journal of Applied Physics*, vol. 126, no. 14, p. 143103, 2019.

[132] A. S. Huntington, *InGaAs Avalanche Photodiodes for Ranging and Lidar*. Woodhead Publishing, 2020.

[133] Y. Yuan, J. Zheng, Y. Tan, Y. Peng, A.-K. Rockwell, S. R. Bank, A. Ghosh, and J. C. Campbell, "Temperature dependence of the ionization coefficients of inalas and algaas digital alloys," *Photonics Research*, vol. 6, no. 8, pp. 794–799, 2018.

[134] J. R. Madureira, D. Semkat, M. Bonitz, and R. Redmer, "Impact ionization rates of semiconductors in an electric field: The effect of collisional broadening," *Journal of Applied Physics*, vol. 90, no. 2, pp. 829–836, 2001.

[135] R. Redmer, J. R. Madureira, N. Fitzer, S. M. Goodnick, W. Schattke, and E. Schöll, "Field effect on the impact ionization rate in semiconductors," *Journal of Applied Physics*, vol. 87, no. 2, pp. 781–788, 2000.

[136] S. Datta, *Quantum transport: atom to transistor*. Cambridge university press, 2005.

[137] S. Datta, *Lessons from Nanoelectronics: A New Perspective on Transport—Part B: Quantum Transport*. World Scientific, 2018.

[138] R. Golizadeh-Mojarad and S. Datta, "Nonequilibrium green's function based models for dephasing in quantum transport," *Physical Review B*, vol. 75, no. 8, p. 081301, 2007.