UC Riverside
UC Riverside Electronic Theses and Dissertations

Title
Modeling, Design, and Analysis of III-V Nanowire Transistors and Tunneling Transistors

Permalink
https://escholarship.org/uc/item/6z37b72x

Author
Khayer, Mohammad Abul

Publication Date
2011

Peer reviewed|Thesis/dissertation

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

UNIVERSITY OF CALIFORNIA
RIVERSIDE

# Modeling, Design, and Analysis of III-V Nanowire Transistors and Tunneling Transistors

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

Doctor of Philosophy

in

Electrical Engineering

by

Mohammad Abul Khayer

March 2011

Dissertation Committee:
Dr. Roger K. Lake, Chairperson
Dr. Alexander A. Balandin
Dr. Javier E. Garay

Copyright by
Mohammad Abul Khayer
2011

The Dissertation of Mohammad Abul Khayer is approved:

---

---

---

**Committee Chairperson**

University of California, Riverside

# Acknowledgments

At first, I would like to express my heartiest gratitude and deep appreciation to my PhD advisor Professor Roger Lake, Department of Electrical Engineering, University of California at Riverside, for his continuous support, interactive guidance and suggestions during the past four years as his student. Without his strenuous engagement in this thesis, this work would not be possible to materialize. My experience of working and interacting with Prof. Lake on developing modeling and simulation tools, analytical models, deriving theory, understanding and solving the crucial problems for nanoscale devices is invaluable. I am grateful to my Doctoral committee members for their support and suggestions. I appreciate discussions with my former lab mates Dr. Khairul Alam and Dr. Nicolas Bruque.

To my family: I thank to my mom (Amma), brothers (Elias, Delwar, Shahid), sisters (Saleha and Rehana) for their support throughout graduate school. And lastly, to my dearest and loving wife Sohana, who has dealt with all the ups and downs by supporting and encouraging me without any condition, while pursuing her PhD in Mechanical Engineering. I would be delighted to convey my special thanks to my cute and beautiful son Aayan, whom while taking in my lap I was being able to work in my laptop.

The text of this dissertation, in part or in full, is a reprint of the material as it appears in the following journals and/or proceedings:

- IEEE Transactions on Electron Devices [1,2]. Reprinted with permission from [1,2]. © [2008] IEEE and © [2009] IEEE.
- IEEE Int. Electron Devices Meeting Tech. Dig. [3]. Reprinted with permission from [3]. © [2008] IEEE.
- IEEE Electron Devices Letters [4]. Reprinted with permission from [4]. © [2009] IEEE.
- Journal of Applied Physics [5,6]. Reprinted with permission from [5,6]. Copyright [2010], American Institute of Physics.
- Applied Physics Letters [7]. Reprinted with permission from [7]. Copyright [2009], American Institute of Physics.

The co-author R. K. Lake, listed in the above publications directed and supervised the research which forms the basis for this dissertation. This work is supported by the Microelectronics Advanced Research Corporation Focus Center on Nano Materials (FENA).

To my beloved family

ABSTRACT OF THE DISSERTATION

Modeling, Design, and Analysis of III-V Nanowire Transistors and Tunneling Transistors

by

Mohammad Abul Khayer

Doctor of Philosophy, Graduate Program in Electrical Engineering
University of California, Riverside, March 2011
Dr. Roger Lake, Chairperson

The aggressive downsizing of the transistor will continue for at least another decade. The critical feature size (physical gate length) of transistors will drop to 5 nm in 2020 (referred to as the 11 nm technology node). In the sub-10 nm range, a variety of low-dimensional materials and structures are being considered to increase device and circuit performance. Examples are semiconductor nanowires (NWs), carbon nanotubes (CNTs), and single-atomic layers of carbon called graphene.

In order to investigate the performance, understand the physics, propose device design, and guide experiments of nanometer scale complementary metal-oxide semiconductor (CMOS) devices with one-dimensional (1-D) novel channel materials, such as III-V NWs, a generalized quantum mechanical modeling and simulation approach is undertaken in this dissertation. We have developed models and simulation tools, derived theory to understand and investigate III-V NW field-effect transistors (FETs) for next generation high-speed, low-power logic applications. These alternative materials and geometries are being investigated for two different types of transistors, (a) standard FETs, and (b) band-to-band tunneling FETs (TFETs).

In the first part of the dissertation, we have investigated the key device metrics such as the quantum capacitance, the drive current, the charge, the power-delay product, the energy-delay product, and the switching frequency of NW FETs based on InSb, InAs, InP, and GaN materials. We have identified two operational regimes for these nanoscale devices, namely, the quantum capacitance limit (QCL) and the classical capacitance limit (CCL). It is shown that $n$-type NW FETs upto $\sim$50 nm in core diameter operate in the QCL, and the corresponding $p$-type devices operate in the CCL. Drive currents at a fixed gate overdrive for the $n$- and $p$-type devices are found to be well-matched. Significant performance improvement in terms of device metrics are predicted for devices operating in the QCL.

In the second part of the dissertation, we have investigated III-V NW and CNT TFETs. A generalized approach to quickly determine the drive current as a function of materials, diameter, and electric field is developed. It is found that a CNT with the same bandgap as a NW can provide 10$\times$ drive current at the same electric field. We have developed a general non-equilibrium Green's function (NEGF) based approach within recursive Green's function (RGF) algorithm to investigate the effects of 'band-tails' on the subthreshold characteristics of TFETs. Band-tails can result from heavy doping, impurities, and phonons. We show that band-tails resulting from necessary heavy doping of the source are not a show-stopper for TFETs.

# Contents

**Approval** iii

**Acknowledgments** v

**Dedication** vi

**Abstract** vii

**1 Introduction** 1

1.1 Silicon CMOS scaling: From long channel to ultra-short channel devices 2

1.2 Beyond silicon . . . . . . . . . . . . . . . . . . . . . . . . . . . 4

1.3 III-V CMOS: Experimental exploration . . . . . . . . . . . . . . . . . 6

1.4 Nanowire field-effect transistors . . . . . . . . . . . . . . . . . . . . 6

1.5 Simulation issues and challenges . . . . . . . . . . . . . . . . . . . . 8

1.6 Outline of the thesis . . . . . . . . . . . . . . . . . . . . . . . . . 10

**2 Electronic bandstructures** 13

2.1 Zinc blende materials systems . . . . . . . . . . . . . . . . . . . . 14

2.2 Electronic structures calculation . . . . . . . . . . . . . . . . . . . . 15

2.2.1 General description of the $\mathbf{k} \cdot \mathbf{p}$ method . . . . . . . . . . . . . 17

2.2.2 Spin-orbit coupling . . . . . . . . . . . . . . . . . . . . . . . 19

2.2.3 Multiband $\mathbf{k} \cdot \mathbf{p}$ method . . . 20
2.2.4 2-band effective mass model . . . 31
2.2.5 Evanescent dispersions . . . 32

**3 Charge and transport calculation** **34**
3.1 Recursive Green's function (RGF) algorithm for charge calculation . . 34
3.1.1 RGF algorithm for calculating $G^R$ . . . 36
3.1.2 RGF algorithm for calculating $G^<$ . . . 38
3.1.3 Calculation of surface self energies . . . 40
3.2 Transport calculation . . . 42
3.2.1 Ballistic transport . . . 42
3.2.2 Diffusive transport . . . 44

**4 Nanowire MOSFETs: High-speed, low-power applications** **46**
4.1 Introduction . . . 47
4.2 Model . . . 48
4.3 Energy-wave vector characteristics . . . 48
4.3.1 Effects of the discretization length . . . 50
4.4 Bandgap and effective mass . . . 51
4.5 Quantum capacitance . . . 53
4.6 Current-voltage characteristics . . . 54
4.7 Summary . . . 60

**5 Nanowire MOSFETs: Device metrics for $n$- and $p$-type devices** **61**
5.1 Introduction . . . 62
5.2 Model . . . 64
5.3 Quantum vs. classical capacitance regime . . . 64

5.4 Numerical calculation . . . . . . . . . . . . . . . . . . . . . . . . . . 64
5.5 Analytical calculation . . . . . . . . . . . . . . . . . . . . . . . . . . 68
5.6 Results . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 74
5.6.1 Quantum and classical capacitance . . . . . . . . . . . . . . . . . 74
5.6.2 Device metrics . . . . . . . . . . . . . . . . . . . . . . . . . . . 84
5.6.3 Diameter dependent performance: InSb NW FETs . . . . . . 88
5.6.4 Diameter dependent performance: InAs NW FETs . . . . . . 96
5.7 Summary . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 108

**6 Nanowire MOSFETs: High-speed, high-power applications 110**
6.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 110
6.2 Model . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 111
6.3 Results . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 112
6.3.1 Energy dispersions . . . . . . . . . . . . . . . . . . . . . . . . . 112
6.3.2 Electron density, quantum capacitance and current . . . . . . 114
6.4 Summary . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 116

**7 Nanowire (NW) band-to-band tunneling FETs (TFETs) 118**
7.1 Introduction . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 119
7.2 TFET definition and working principle . . . . . . . . . . . . . . . . . 121
7.3 Why TFET? . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 121
7.4 Drive currents: Effects of materials, diameters, and electric fields . . . 123
7.4.1 Model . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 123
7.4.2 Analytical calculations . . . . . . . . . . . . . . . . . . . . . . . 123
7.4.3 Cold-carrier injection . . . . . . . . . . . . . . . . . . . . . . . . 126
7.4.4 InSb and InAs NW TFETs . . . . . . . . . . . . . . . . . . . . 127
7.4.5 InP and GaN NW TFETs . . . . . . . . . . . . . . . . . . . . . 132

7.5 Subthreshold characteristics: Effects of ‘band-tails’ . . . . . . . . . . . 142
7.5.1 Model . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 143
7.5.2 What is ‘band-tail’? . . . . . . . . . . . . . . . . . . . . . . . . 144
7.5.3 Numerical calculation . . . . . . . . . . . . . . . . . . . . . . . 146
7.5.4 Band-tails: Density of states . . . . . . . . . . . . . . . . . . . 148
7.5.5 Band-tails: Effects of heavy doping in the source . . . . . . . . 150
7.5.6 Band-tails: Effects of heavy doping in the source and the channel 153
7.6 Summary . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 160

**8 SUMMARY AND FUTURE DIRECTION 165**

8.1 Summary . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 165
8.2 Future work . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 167

# List of Tables


2.1 *Material Constants* . . . . . . . . . . 16

5.1 *Figures of merit of the single-moded n-type InSb and InAs NW FETs from analytical calculations.* $L_G$*=10 nm,* $\alpha_G$*=0.8,* $m^*$*=0.034*$m_0$ *and 0.042*$m_0$*, respectively for InSb and InAs NW FETs [1],* $E_{FS}-\varepsilon_1$*=0.12 eV (Table 5.4).* $\tilde{\alpha_G}$*=0.73 and 0.71, respectively for InSb and InAs NW FETs. All values are calculated at maximum gate overdrive. Reproduced with permission from [2]. © [2009] IEEE.* . . . . . . . . . . 73

5.2 *Figures of merit of the proposed InSb and InAs NW FETs with a 10 nm long gate. The switching delay time and the intrinsic cut-off frequency are defined as* $\tau_D = (C_{GS}V_{DD})/I_D$ *and* $f_T = g_m/(2\pi C_{GS})$ *with* $g_m$ *being the transconductance are also listed. Reproduced with permission from [1]. © [2008] IEEE.* . . . . . . . . . . 84

5.3 *Figures of merit of the InSb and InAs NW FETs with NW diameters of 4 nm from numerical calculations. Gate length of 10 nm is considered. A source Fermi energy of* $E_{FS}$*=0.2eV is considered to calculate these values. All values are calculated at maximum gate overdrive. Reproduced with permission from [2]. © [2009] IEEE.* . . . . . . . . . . 85

5.4 *Charge under the gate, $|n-p|$, gate source capacitance, $C_{GS}$ (aF/nm), $E_{FS}-\varepsilon_1$ at maximum gate overdrive for InSb and InAs NW FETs with NW diameters of 4nm and with a gate length of 10 nm. Threshold current, $I_{th}$ for each device is equal to 0.63 $\mu A$. Transconductance for the devices are also shown. All values are calculated at the maximum gate overdrive. Reproduced with permission from [2]. © [2009] IEEE.* 86

5.5 *Figures of merit of 10 nm NW diameter InSb and InAs NW FETs from numerical calculations. Gate length of 10 nm is considered. A source Fermi energy of $E_{FS}$=0.2eV is taken to calculate these values. Threshold current, $I_{th}$ for each device is equal to 0.64 $\mu A$. All values are calculated at maximum gate overdrive. Reproduced with permission from [2]. © [2009] IEEE.* . . . . . . . . . . . . . . . . . . . . . . . . 87

5.6 *Figures of merit of the proposed InSb NW FETs with a 10 nm long gate. The intrinsic cut-off frequency is defined as $f_T = g_m/(2\pi C_{GS})$ with $g_m$ being the transconductance. $C_{GS}$ is the series combination of the geometric ($C_G$) and the quantum ($C_Q$) capacitances. $I_{ON}$ from Figs. 5.10 are also listed. $P \cdot \tau$ is calculated as $\int Q dV_{GS}$, where Q is the total channel charge extracted from the simulations. Accordingly the gate delay time $\tau = \int Q dV_{GS}/(V_{DD} I_{ON})$. Reproduced with permission from [3]. © [2008] IEEE.* . . . . . . . . . . . . . . . . . . . . . . . . 95

5.7 *Figures of merit of the proposed InAs NW FETs with a 10 nm long gate. The intrinsic cut-off frequency is defined as $f_T = g_m/(2\pi C_{GS})$ with $g_m$ being the transconductance. $C_{GS}$ is the series combination of the geometric ($C_G$) and the quantum ($C_Q$) capacitances. $I_{ON}$ from Figs. 5.16 are also listed. $P \cdot \tau$ is calculated as $\int QdV_{GS}$, where $Q$ is the total channel charge extracted from the simulations. Accordingly the gate delay time $\tau = \int QdV_{GS}/(V_{DD}I_{ON})$. Reproduced with permission from [5]. Copyright [2010], American Institute of Physics.* . . . . . . 107

# List of Figures

2.1 *Zinc blende crystal structure.* . . . . . . . . . . . . . . . . . . . . . . . 14

2.2 *Discretization scheme for NW.* . . . . . . . . . . . . . . . . . . . . . . . 31

3.1 *Partitioning the structure into left contact, the device, and the right contact. The block indices* $\{-\infty, ...., -1, 0\}$ *lie on the left contact. The block indices* $\{1, 2, ....., N-1, N\}$ *belong to the device, and* $\{N+1, N+2, ....., \infty\}$ *lie on the right contact.* . . . . . . . . . . . . . . . . . . . . 36

4.1 *(a) Schematic diagram of the simulated gate-all-around NW transistor, (b) The electrostatics of the seminumerical ballistic FET model used in the calculation. The diagrams are not to scale. Reproduced with permission from [1].* © *2008 IEEE.* . . . . . . . . . . . . . . . . . . . . 49

4.2 *E-k dispersion relationship calculated using the discretized 8-band* $\boldsymbol{k} \cdot \boldsymbol{p}$ *model for InSb NWs with 2 nm (a), and 4 nm (b) diameter and for InAs NWs with 2 nm (c), and 4 nm (d) diameter. Rectangular cross sections are considered. Reproduced with permission from [1].* © *[2008] IEEE.* 50

4.3 *Effects of the discretization length on the E-k dispersion of InAs for (a) the conduction band and (b) the heavy hole band.* . . . . . . . . . 51

4.4 *Lowest conduction band electron effective masses vs. electron energy calculated from the E-k dispersion for InSb NWs with 2 nm (a), and 4 nm (b) diameter and for InAs NWs with 2 nm (c), and 4 nm (d) diameter. Reproduced with permission from [1]. © [2008] IEEE.* . . . 52

4.5 *(a) Bandgap as a function of wire diameter for the simulated InSb and InAs NWs, and (b) The wire diameter dependence of the lowest conduction band electron effective mass at the zone center. Reproduced with permission from [1]. © [2008] IEEE.* . . . . . . . . . . . . . . . . 53

4.6 *Quantum capacitance as a function of the Fermi level for the simulated InSb NW FETs with two NW diameter, 2 nm (a), and 4 nm (b) and the same for the InAs NW FETs with two NW diameter, 2 nm (c), and 4 nm (d). Corresponding gate capacitances are also shown. All four devices are operating in quantum capacitance limit. Reproduced with permission from [1]. © [2008] IEEE.* . . . . . . . . . . . . . . . . 55

4.7 $I_{DS}$ *vs.* $V_{DS}$ *characteristics of the simulated InSb NW FETs with two NW diameter, 2 nm (a), 4 nm (b), and the same for the simulated InAs NW FETs with two NW diameter, 2 nm (c), 4 nm (d). The highest bias voltages have been taken such that flat band transport can be considered. Reproduced with permission from [1]. © [2008] IEEE.* 58

4.8 *log(*$I_{DS}$*) vs.* $V_{GS}$ *transfer characteristics of the simulated InSb NW FETs with two NW diameter, 2 nm (a), 4 nm (b), and the same for the simulated InAs NW FETs with two NW diameter, 2 nm (c), 4 nm (d). The highest bias voltages have been taken such that flat band transport can be considered. Reproduced with permission from [1]. © [2008] IEEE.* . . . . . . . . . . . . . . . . . . . . . . . . . . . 59

4.9 *Number of modes as a function of mode energy for the simulated InSb and InAs NWs. All four devices are single moded within the energy range considered. Reproduced with permission from [1]. © [2008] IEEE.* 59

5.1 *Small signal equivalent circuit corresponding to Eq. (5.2). Reproduced with permission from [2]. © [2009] IEEE.* . . . . . . . . . . . . . . . . 68

5.2 *E-k dispersion relationship calculated using the discretized 8-band* $\boldsymbol{k} \cdot \boldsymbol{p}$ *model for (a) InSb NWs, and (b) InAs NWs with 4 nm NW diameter.* $E = 0$ *corresponds to the lowest conduction band energy of the bulk InSb or InAs materials. Rectangular cross sections are considered. Reproduced with permission from [2]. © [2009] IEEE.* . . . . . . . . . . 75

5.3 *Calculated transmission as a function of energy for the simulated 4 nm diameter (a) InSb, and (b) InAs NW FETs.* $E = 0$ *corresponds to the lowest conduction band energy of the bulk InSb or InAs materials. Reproduced with permission from [2]. © [2009] IEEE.* . . . . . . . . . . 76

5.4 *Quantum capacitance under the gate as a function of* $|E_{FS} - \varepsilon_1|$ *where* $E_{FS}$ *is the source Fermi level and* $\varepsilon_1$ *is the energy of the fundamental mode under the gate for (a) InSb NW FETs, and (b) InAs NW FETs. Figures are presented for both n-type and p-type devices. Corresponding gate capacitances are also shown. Reproduced with permission from [2]. © [2009] IEEE.* . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 78

5.5 *log($I_{DS}$) vs. $V_{GS}$ transfer characteristics of the simulated (a) InSb NW FETs, and (b) InAs NW FETs with 4 nm NW diameter for both n- and p-type devices. The drain bias voltages are fixed at half of the energy bandgap for each NW as shown. For the p-type devices, the polarity of the voltage is reversed. Reproduced with permission from [2]. © [2009] IEEE.* . . . . . 80

5.6 *Carrier density, $Q_{top}$ vs. $V_{GS}$ plots for the simulated (a) InSb NW FETs, and (b) InAs NW FETs with 4 nm NW diameter. For the p-type devices, the polarity of the voltage is reversed. Reproduced with permission from [2]. © [2009] IEEE.* . . . . . 81

5.7 *Band-edge energies under the gate as a function of the gate bias for simulated (a) InSb NW FETs, and (b) InAs NW FETs. The source Fermi energy $E_{FS}$ is at 0.2 eV for all devices and is shown by the horizontal dashed line. The drain biases are fixed at half of the energy bandgap for each NW as shown. For the p-type devices, the polarity of the voltage is reversed. Reproduced with permission from [2]. © [2009] IEEE.* . . . . . 82

5.8 *Number of spin-degenerate conduction modes as a function of the mode energy for the simulated InSb NW FETs with 10 nm (a), and 60 nm (b) NW diameters. All the devices contain multiple modes within the energy range considered. Reproduced with permission from [3]. © [2008] IEEE.* . . . . . 91

5.9 *Quantum capacitance as a function of the source Fermi level for the simulated InSb NW FETs with 10 nm (a), and 60 nm (b) NW diameters. The corresponding geometrical capacitance,* $C_G$ *is also shown.* $C_G$ *is calculated from* $C_G = (2\pi\epsilon_r\epsilon_0)/(ln((2d_{ox} + d_{NW})/d_{NW}))$ *assuming a coaxial gate geometry. All the devices are operating in the quantum capacitance limit within source Fermi level (*$E_F - E_c = E_{FS}$*) range of 0.1-0.2 eV. Reproduced with permission from [3]. © [2008] IEEE.* . . . 92

5.10 *ON-current as a function of NW diameter with source Fermi level (*$E_{FS} = E_F - E_c$*) of 0.1 eV (a), and 0.2 eV (b). The drain bias voltage is fixed at 0.5 V as shown. The gate voltage is fixed at a gate overdrive of 0.2V, i.e.,* $V_{GS} - V_T = 0.2$ *where* $V_T$ *is the threshold voltage for each device. Reproduced with permission from [3]. © [2008] IEEE.* 93

5.11 *Simulated* $P \cdot \tau$ *(power-delay product) as a function of NW diameter with source Fermi level (*$E_{FS} = E_F - E_c$*) of 0.1 eV (a), and 0.2 eV (b).* $P \cdot \tau$ *is calculated as* $\int QdV_{GS}$*, where Q is the total channel charge extracted from the simulations. Accordingly the gate delay time* $\tau = \int QdV_{GS}/(V_{DD}I_{ON})$*. Device operating regime, quantum capacitance limit (QCL), is indicated in the plots. These curves show a significant performance improvement in terms of the power-delay product for the devices scaled towards the quantum capacitance limit with* $E_{FS}$*=0.1 eV. Inset shows the gate delay time as a function of NW diameter extracted from the simulations. A gate length of 10 nm was considered to calculate these values. Reproduced with permission from [3]. © [2008] IEEE.* . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 93

5.12 *Simulated $P \cdot \tau$ (power-delay product) as a function of NW diameter with source Fermi level ($E_F - E_c$) of 0.3 eV. $P \cdot \tau$ is calculated as $\int QdV_{GS}$, where Q is the total channel charge extracted from the simulations. Accordingly the gate delay time $\tau = \int QdV_{GS}/(V_{DD}I_{ON})$. Device operating regimes, quantum capacitance limit (QCL), and classical capacitance limit (CCL) are indicated in the plots. These curves show a significant performance improvement in terms of the power-delay product for the devices scaled towards the quantum capacitance limit. Inset shows the gate delay time as a function of NW diameter extracted from the simulations. A gate length of 10 nm was considered to calculate these values. Reproduced with permission from [3]. © [2008] IEEE.* . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 94

5.13 *Simulated $E \cdot \tau$ (energy-delay product) as a function of NW diameter with source Fermi level ($E_{FS} = E_F - E_c$) of 0.1 eV (a), and 0.2 eV (b). $E \cdot \tau$ is the product of the $P \cdot \tau$ and the gate delay $\tau$. These curves show a significant performance improvement in terms of the energy-delay product for the devices scaled towards the quantum capacitance limit with $E_{FS}$ =0.1 eV. Reproduced with permission from [3]. © [2008] IEEE.* . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 95

5.14 *Number of spin-degenerate conduction modes as a function of the mode energy for the simulated InAs NW FETs with 10 nm (a), and 60 nm (b) NW diameters. All the devices contain multiple modes within the energy range considered. Reproduced with permission from [5]. Copyright [2010], American Institute of Physics.* . . . . . . . . . . . . . . . . . 100

5.15 *Quantum capacitance under the gate as a function of* $E_{FS} - \varepsilon_1$ *where* $E_{FS}$ *is the source Fermi level and* $\varepsilon_1$ *is the energy of the fundamental mode under the gate for the simulated InAs NW FETs with 10 nm (a), and 60 nm (b) NW diameters. The corresponding geometrical capacitance,* $C_G$ *is also shown.* $C_G$ *is calculated from* $C_G = (2\pi\epsilon_r\epsilon_0)/(ln((2d_{ox} + d_{NW})/d_{NW}))$ *assuming a coaxial gate geometry. All the devices are operating in the quantum capacitance limit within source Fermi level range of 0.1 - 0.2 eV. Reproduced with permission from [5]. Copyright [2010], American Institute of Physics.* . . . . . . 101

5.16 *ON-current as a function of NW diameter with source Fermi level of 0.1 eV (a), and 0.2 eV (b). The drain bias voltage is fixed at 0.5 V as shown. The gate voltage is fixed at a gate overdrive of 0.2V, i.e.,* $V_{GS} - V_T = 0.2$ *where* $V_T$ *is the threshold voltage for each device. Reproduced with permission from [5]. Copyright [2010], American Institute of Physics.* . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 103

5.17 *Simulated power-delay product* $(P \cdot \tau)$ *and the delay time* $(\tau)$ *as a function of NW diameter with source Fermi level of 0.1 eV (a), and 0.2 eV (b). Device operating regime, quantum capacitance limit (QCL), is indicated in the plots. These curves show a significant performance improvement in terms of the power-delay product for the devices scaled towards the quantum capacitance limit with* $E_{FS}$*=0.1 eV. A gate length of 10 nm was considered to calculate these values. Reproduced with permission from [5]. Copyright [2010], American Institute of Physics.* . . . 104

5.18 *Simulated power-delay product ($P \cdot \tau$) as a function of NW diameter with source Fermi level of 0.3 eV. Device operating regimes, quantum capacitance limit (QCL), and classical capacitance limit (CCL) are indicated in the plots. These curves show a significant performance improvement in terms of the power-delay product for the devices scaled towards the quantum capacitance limit. Inset shows the gate delay time as a function of NW diameter extracted from the simulations. A gate length of 10 nm was considered to calculate these values. Reproduced with permission from [5]. Copyright [2010], American Institute of Physics.* . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 105

5.19 *Simulated energy-delay product ($E \cdot \tau$) as a function of NW diameter with source Fermi level of 0.1 eV (a), and 0.2 eV (b). $E \cdot \tau$ is the product of the $P{\cdot}\tau$ and the gate delay $\tau$. These curves show a significant performance improvement in terms of the energy-delay product for the devices scaled towards the quantum capacitance limit with $E_{FS}$ =0.1 eV. Reproduced with permission from [5]. Copyright [2010], American Institute of Physics.* . . . . . . . . . . . . . . . . . . . . . . . . . . . . 106

6.1 *(a) E-k relations calculated using the 3-D discretized 8-band* $\boldsymbol{k}{\cdot}\boldsymbol{p}$ *model, (b) Number of conduction band modes as a function of the mode energy for a 2-nm diameter GaN NW FET. Figs. (c-d) are similar to figs. (a-b) except with a diameter of 4-nm. Reproduced with permission from [6]. Copyright [2010], American Institute of Physics.* . . . . . . 113

6.2 *(a) Electron density, (b) the quantum capacitance (corresponding gate capacitances are also shown) as a function of the source Fermi energy ($E_{FS} = E_F - E_C$, with $E_C$ being the lowest conduction band-edge) for 2- and 4-nm diameter GaN NW FETs. Fig. (b) shows that both the NW FETs are operating in the classical capacitance limit (CCL) with $E_{FS}$=0.2 eV as $C_G \ll C_Q$. (c) Band-edge under the gate as a function of the gate voltages for 2- and 4-nm diameter GaN NW FETs. (d) log($I_{DS}$) vs. $V_{GS}$ transfer characteristics of 2- and 4-nm diameter GaN NW FETs. The drain and gate voltages are fixed to $E_G/(2q)$ and $E_G/q$, respectively for each NW FET, to maximize the drive current. The OFF-current is significantly reduced due to the large bandgap for GaN NWs, which blocks the leakage current. Reproduced with permission from [6]. Copyright [2010], American Institute of Physics.* . . . . . . 117

7.1 *Energy band profile of TFETs at (a) off state and (b) on state. Figure (a) shows how energy gap chops off the Fermi tail at the off-state and thereby minimize high energy thermal distribution of carriers. Figure (b) shows energy pass window through which carriers can tunnel through. Sufficiently large window can be created with heavily doped source and drain.* . . . . . . . . . . . . . . . . . . . . . . . . . . 122

7.2 *Illustration of electron tunneling from the valence band to the conduction band. In the bandgap, the wavevector $\kappa$ becomes imaginary with a magnitude shown by the curve. $E_V$ and $E_C$ are the valence and conduction band edges, respectively. Reproduced with permission from [4]. © [2009] IEEE.* . . . . . . . . . . . . . . . . . . . . . . . . . . 124

7.3 *Energy band profile of TFETs at on state showing how the Fermi-tail is essentially suppressed by the bandgap.* . . . . . . . . . . . . . . . . . 126

7.4 *Smallest imaginary wavevectors in the bandgap of 10 nm diameter InSb and InAs NWs and a (23,0) CNT. The valence band edges of all structures are set to* $E = 0$. *The curves shown couple the highest valence band mode and the lowest conduction band mode. Other such curves with larger imaginary wavevectors exist which couple lower valence modes to higher conduction modes, but they are irrelevant for tunneling since they are exponentially suppressed. The CNT curve is 2-fold degenerate. Reproduced with permission from [4].* © *[2009] IEEE.* 129

7.5 *(a) Calculated analytical and numerical tunneling currents as a function of the electric field for 2nm InSb and InAs NW FETs and for a (25,0) CNT FET. The current axis for the CNT is on the left, and the current axis for the InSb and InAs NWs is on the right (note the different scales). The inset shows the band profile and window of integration,* $\Delta E$. *(b) Wavevector integral,* $\mathcal{E}_I \equiv \frac{2}{q}\int_0^{E_G} dE'\kappa(E')$, *for InSb, InAs NWs, and for CNTs as a function of the NW/CNT diameter. The* $\mathcal{E}_I$ *axis for the NWs is on the left, and the* $\mathcal{E}_I$ *axis for the CNTs is on the right. Reproduced with permission from [4].* © *[2009] IEEE.* 131

7.6 *Enhancement mode FET biasing. (a) No bias applied. (b) On state; $V_{GS} = V_{DS} = V_{DD}$. (c) Off state; $V_{GS} = 0$ and $V_{DS} = V_{DD}$. (d) Source electric field, $\mathcal{E}_S$, needed to achieve an ON-current of 1$\mu$A and 10$\mu$A, repectively, for NW FETs and CNT FETs as a function of diameter, and (e) Drain electric field, $\mathcal{E}_D$, required to obtain an ON-OFF current ratio of $10^5$ for various diameter NW/CNT TFETs. For (d) and (e), the fields for the NWs are on the left axis, and the fields for the CNTs are on the right axis. Reproduced with permission from [4]. © [2009] IEEE.* . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 132

7.7 *Smallest imaginary wavevectors in the bandgap of 2 nm and 10 nm diameters InP NWs. The valence band edges of all structures are set to $E = 0$. The curves shown couple the highest valence band mode and the lowest conduction band mode. Other such curves with larger imaginary wavevectors exist which couple lower valence modes to higher conduction modes, but they are irrelevant for tunneling since they are exponentially suppressed. Reproduced with permission from [7]. Copyright [2009], American Institute of Physics.* . . . . . . . . . . . . . . . . 133

7.8 *Number of conduction band modes (right axis) and quantum capacitance (left axis) as a function of source Fermi energy for 2 nm and 10 nm diameter InP NWs. With a source Fermi energy of 0.1 eV, all the devices have $C_Q < C_G$, and thus operate in the quantum capacitance limit. Reproduced with permission from [7]. Copyright [2009], American Institute of Physics.* . . . . . . . . . . . . . . . . . . . . . . . 134

7.9 *(a) Calculated analytical and numerical tunneling currents as a function of the drain electric field for InP NW FETs. Inset shows the band profile and window of integration, $\Delta E$. (b) Wavevector integral, $\mathcal{E}_I \equiv \frac{2}{q}\int_0^{E_G} dE'\kappa(E')$ for the devices as a function of NW diameter. Reproduced with permission from [7]. Copyright [2009], American Institute of Physics.* . . . . . . . . . . 135

7.10 *Standard enhancement mode FET biasing. (a) No bias applied. (b) On state; $V_{GS} = V_{DS} = V_{DD}$. (c) Off state; $V_{GS} = 0$ and $V_{DS} = V_{DD}$. (d) Source electric field needed to achieve an ON-current of 1$\mu$A, and (e) Drain electric field (left vertical axis) and the drain underlap (right vertical axis) needed to acquire ON-OFF current ratio of $10^6$ for various diameter InP NW TFETs. Reproduced with permission from [7]. Copyright [2009], American Institute of Physics.* . . . . . . 137

7.11 *Bandgap as a function of NW diameter for the simulated GaN NWs. Bulk GaN gandgap is also shown in the plot for comparison. Reproduced with permission from [6]. Copyright [2010], American Institute of Physics.* . . . . . . . . . . 138

7.12 *(a) Smallest imaginary wave vectors in the bandgap of 2- and 8-nm diameter GaN NWs. The valence band-edges of all structures are set to E=0. The curves shown couple the highest valence band mode and the lowest conduction band mode. Other such curves with larger imaginary wave vectors exist which couple lower valence band modes to higher conduction band modes, but they are irrelevant for tunneling since they are exponentially suppressed. For smaller NW diameter, the imaginary band contains more area, and vice versa. Imaginary bands are considered as a leakage/tunneling path for electrons. The tunneling current will be exponentially damped with a damping coefficient proportional to the area contained by the imaginary band. (b) Wave vector integral,* $\varepsilon_I = (2/q)\int dE^{'}\kappa(E^{'})$*, for GaN NWs as a function of NW diameter.The smaller the NW diameter, the larger the values of the wave vector integral (damping coefficient), and the smaller the tunneling currents will be. Reproduced with permission from [6]. Copyright [2010], American Institute of Physics.* . . . . . . . . . . . . . . . . . . . 139

7.13 *Enhancement-mode FET biasing. (a) No bias applied, (b) ON state; $V_{GS} = V_{DS} = V_{DD}$. (c) OFF state; $V_{GS} = 0$ and $V_{DS} = V_{DD}$. $E_G$ is the bandgap of the NWs, and $\alpha$ is the gate control parameter. (d) Source electric field, $\varepsilon_S$ (left vertical axis) and source-channel underlap, $x_{uS}$ (right vertical axis) needed to achieve ON-current of 1 $\mu$A for GaN NW TFETs as a function of NW diameter. (e) Drain electric field, $\varepsilon_D$ (left vertical axis) and the channel-drain underlap, $x_{uS}$ (right vertical axis) required to obtain an ON-OFF current ratio of $10^6$ for GaN NW TFETs as a function of NW diameter. With larger NW diameter, tunneling is higher and consequently less field is required to produce or suppress the desired drive current or undesired leakage current, respectively. Reproduced with permission from [6]. Copyright [2010], American Institute of Physics.* . . . . . . . . . . . . . . . . . . . 141

7.14 *(a) The band profile (not to scale) for a $p^+$-i-$n^+$ TFET in the off state. 'Band-tail' resulting from heavy doping promotes carrier tunneling from the source to the gate region. Two tunneling processes in the subthreshold conduction mode are indicated: (1) mid-gap trap assisted tunneling, and (2) tunneling by carrier hopping. (b) Density of states as a function of energy of 2nm wire diameter InSb NW showing the band-edges and the band-tails. Away from the 1-D band-edge singularity, the band-tails decay in the bandgap as $e^{-|E-E_{C,V}|/\alpha_A}$, with $E_C$ and $E_V$ being the conduction band-edge and valence band-edge, respectively. A range of values for $\alpha_A$ are investigated. In this case, $\alpha_A = kT$, with $k$ being the Boltzmann constant and $T$ is the room temperature (300K).* 145

7.15 *Density of states as a function of energy of 2nm wire diameter InSb NW showing the band-edges and the band-tails with $\eta_0$=20 meV. Away from the 1-D band-edge singularity, the band-tails decay in the bandgap as $e^{-|E-E_{C,V}|/\alpha_A}$, with $E_C$ and $E_V$ being the conduction band-edge and valence band-edge, respectively. Plots for a range of values for $\alpha_A$ are investigated.* . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 149

7.16 *Density of states as a function of energy of 2nm wire diameter InSb NW showing the band-edges and the band-tails with $\eta_0$=50 meV. Away from the 1-D band-edge singularity, the band-tails decay in the bandgap as $e^{-|E-E_{C,V}|/\alpha_A}$, with $E_C$ and $E_V$ being the conduction band-edge and valence band-edge, respectively. Plots for a range of values for $\alpha_A$ are investigated.* . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 149

7.17 *Current density as a function of energy of InSb NW TFETs with $\Sigma^R$ used in the source region for two gate lengths (a) 10 nm and (b) 20 nm. For all plots, $E_{VS}$=0 eV is the valence band-edge in the source which is taken as reference. $\eta_0$=20meV and $\alpha_A$=0.54kT. Source Fermi level of 0.1 eV is used. Three gate biases are considered, i.e., $V_{GS}$=0 V (dotted lines), $V_{GS}$=0.125 V (dash dotted lines), and $V_{GS}$=0.25 V (solid lines). On top of each plot, level for conduction band-edge under the gate, $E_{CG}$ for three bias conditions are indicated by the dashed vertical lines.* . . . 151

7.18 *Current density as a function of energy of InSb NW TFETs with* $\Sigma^R$ *used in the source region for two gate lengths (a) 10 nm and (b) 20 nm. For all plots,* $E_{VS}$*=0 eV is the valence band-edge in the source which is taken as reference.* $\eta_0$*=50meV and* $\alpha_A$*=kT. Source Fermi level of 0.1 eV is used. Three gate biases are considered, i.e.,* $V_{GS}$*=0 V (dotted lines),* $V_{GS}$*=0.125 V (dash dotted lines), and* $V_{GS}$*=0.25 V (solid lines). On top of each plot, level for conduction band-edge under the gate,* $E_{CG}$ *for three bias conditions are indicated by the dashed vertical lines.* . . . 152

7.19 $I_D$-$V_G$ *characteristics on a log scale of InSb NW TFETs for a range of values for* $\alpha_A$ *with* $\Sigma^R$ *used in the source region for two gate lengths (a) 10 nm and (b) 20 nm.* $\eta_0$*=20meV. For comparison,* $I_D$-$V_G$ *curves with* $\Sigma^R$*=0 throughout the device is also presented. For all plots, a source Fermi level of 0.1 eV is used.* . . . . . . . . . . . . . . . . . . . . . . . 154

7.20 $I_D$-$V_G$ *characteristics on a log scale of InSb NW TFETs for a range of values for* $\alpha_A$ *with* $\Sigma^R$ *used in the source region for two gate lengths (a) 10 nm and (b) 20 nm.* $\eta_0$*=50meV. For comparison,* $I_D$-$V_G$ *curves with* $\Sigma^R$*=0 throughout the device is also presented. For all plots, a source Fermi level of 0.1 eV is used.* . . . . . . . . . . . . . . . . . . . . . . . 155

7.21 *Subthreshold swing, S as a function of* $\alpha_A$ *for InSb NW TFETs with* $\Sigma^R$ *used only in the source region for two gate lengths, 10 nm and 20 nm.* $\eta_0$*=20meV. A source Fermi level of 0.1 eV is used.* . . . . . . . . 156

7.22 *Subthreshold swing, S as a function of* $\alpha_A$ *for InSb NW TFETs with* $\Sigma^R$ *used only in the source region for two gate lengths, 10 nm and 20 nm.* $\eta_0$*=50meV. A source Fermi level of 0.1 eV is used.* . . . . . . . . 157

7.23 *Current density as a function of energy of InSb NW TFETs with* $\Sigma^R$ *used in the source, source-to-channel, and the gate region for two gate lengths (a) 10 nm and (b) 20 nm. For all plots,* $E_{VS}$*=0 eV is the valence band-edge in the source which is taken as reference.* $\eta_0$*=20meV and* $\alpha_A$*=0.54kT. Source Fermi level of 0.1 eV is used. Three gate biases are considered, i.e.,* $V_{GS}$*=0 V (dotted lines),* $V_{GS}$*=0.125 V (dash dotted lines), and* $V_{GS}$*=0.25 V (solid lines). On top of each plot, level for conduction band-edge under the gate,* $E_{CG}$ *for three bias conditions are indicated by the dashed vertical lines.* . . . . . . . . . . . . . . . . . . . 158

7.24 *Current density as a function of energy of InSb NW TFETs with* $\Sigma^R$ *used in the source, source-to-channel, and the gate region for two gate lengths (a) 10 nm and (b) 20 nm. For all plots,* $E_{VS}$*=0 eV is the valence band-edge in the source which is taken as reference.* $\eta_0$*=50meV and* $\alpha_A$*=kT. Source Fermi level of 0.1 eV is used. Three gate biases are considered, i.e.,* $V_{GS}$*=0 V (dotted lines),* $V_{GS}$*=0.125 V (dash dotted lines), and* $V_{GS}$*=0.25 V (solid lines). On top of each plot, level for conduction band-edge under the gate,* $E_{CG}$ *for three bias conditions are indicated by the dashed vertical lines.* . . . . . . . . . . . . . . . . . . . 159

7.25 $I_D$*-*$V_G$ *characteristics on a log scale of InSb NW TFETs for a range of values for* $\alpha_A$ *with* $\Sigma^R$ *used in the source, source-to-channel, and the gate region for two gate lengths (a) 10 nm and (b) 20 nm.* $\eta_0$*=20meV. For comparison,* $I_D$*-*$V_G$ *curves with* $\Sigma^R$*=0 throughout the device is also presented. For all plots, a source Fermi level of 0.1 eV is used.* . . . . 161

7.26 $I_D$-$V_G$ *characteristics on a log scale of InSb NW TFETs for a range of values for* $\alpha_A$ *with* $\Sigma^R$ *used in the source, source-to-channel, and the gate region for two gate lengths (a) 10 nm and (b) 20 nm.* $\eta_0$*=50meV. For comparison,* $I_D$-$V_G$ *curves with* $\Sigma^R$*=0 throughout the device is also presented. For all plots, a source Fermi level of 0.1 eV is used.* . . . . . 162

7.27 *Subthreshold swing, S as a function of* $\alpha_A$ *for InSb NW TFETs with* $\Sigma^R$ *used only in the source, source-to-channel, and the gate region for two gate lengths, 10 nm and 20 nm.* $\eta_0$*=20meV. A source Fermi level of 0.1 eV is used.* . . . . . 163

7.28 *Subthreshold swing, S as a function of* $\alpha_A$ *for InSb NW TFETs with* $\Sigma^R$ *used only in the source, source-to-channel, and the gate region for two gate lengths, 10 nm and 20 nm.* $\eta_0$*=50meV. A source Fermi level of 0.1 eV is used.* . . . . . 164

# Chapter 1

# Introduction

In order to investigate the performance, understand the physics, propose device design, and guide experiments of nanometer scale complementary metal-oxide semiconductor (CMOS) devices with novel channel materials, such as III-V and carbon, a generalized modeling and simulation approach is undertaken in this work. In Sec. 1.1, a brief overview of the journey of conventional silicon (Si) metal-oxide semiconductor field-effect transistors (MOSFETs) over a period of past four decades, is presented. In Sec. 1.2, beyond silicon materials and device structures are reviewed. In Sec. 1.3, the experimental demonstration of III-V CMOS devices is outlined. In Sec. 1.4, brief description of one dimensional (1-D) nanostructures such as nanowires (NWs) based on III-V materials, and their FET application is presented. In Sec. 1.5, the issues and challenges associated with the simulation of nanoscale III-V MOSFETs are highlighted. Finally, in Sec. 1.6, the outline of this thesis is presented.

## 1.1 Silicon CMOS scaling: From long channel to ultra-short channel devices

Planar MOSFETs have experienced a steady, exponential downsizing of their critical dimensions since their first demonstration in 1960 [8]. The exponential down scaling of feature sizes, and hence the exponential increase of the transistor count in an integrated circuit (IC), was first discovered by Gordon Moore in 1965 [9]. His findings, which later became known as Moore's Law, states that the number of transistors per technology node doubles every twenty four months. Moore's law has been serving as the guiding principle for the semiconductor chip industries for over 30 years [10]. Over this period of time, the printed gate lengths of the MOSFETs have been scaled down from 100 $\mu$m to 20 nm [10], the later refers to the gate lengths of the 22 nm technology node devices in commercial IC chips which will be available by the end of 2011. Due to the steady improvement of their performance through scaling, MOSFETs have become the leading IC technology for high performance and low power logic applications. Over this long period of time, the technology has faced numerous challenges, which were solved by vibrant research, ingenuous design, and excellent engineering.

According to ITRS [10], by 2015, the printed gate length of MOSFETs will be less than 10 nm. Sustaining Moore's law will become challenging for these future technology nodes, where the key technical issue for scaling devices is the off-state leakage current [11]. Demonstration of planar MOSFETs with gate lengths as short as 5 nm is found in the literature [12], however, due to high off-state leakage current, they are not an obvious option for the future ICs. The origin of the scaling limit for single gate, bulk CMOS is identified to be the inherent poor electrostatics of the planar geometry, and the poor transport properties of carriers in the channel.

Consequently, an intense and vibrant research efforts both in academia and industry have been directed toward exploring novel channel materials and device architectures for future technology nodes.

Scaling is further pushed using silicon-on-insulator (SOI) technology. Ultra-thin body SOI MOSFETs show the maturity of this technology [13]. The electrostatics of the nanoscale CMOS devices drastically improves when additional gates are incorporated. Examples are dual-gate [14], tri-gate [15], and FinFET [16]. Due to the improved short-channel effects of these non-planar devices, it is speculated that one of them will be the basic device architecture for future technology nodes.

Sustaining Moore's law depends on the gate insulator scaling, which has been steadily reduced down to a thickness of 1.2 nm until 65 nm technology node. Any further reduction in the dielectric thickness severely degrades oxide reliability and results in exponential increase in gate leakage current. Intel's innovation of high-$\kappa$ (which replaced $SiO_2$) and metal gate (which replaced poly-Si gate) (HKMG) [17] circumvented these issues and pushed the scaling of the technology nodes down to 32 nm. The replacement of silicon channel by novel materials with improved transport properties is being viewed as a strong and promising option to continue downscaling until middle of this decade. Research activities in this area has substantially been increased in recent years - both in academia and industry; with many new and novel materials systems and devices now being proposed. We will review the recent experimental advancements in the area of novel channel materials such as III-V CMOS devices in the next section.

## 1.2 Beyond silicon

Can we sustain progress in information technology hardware by developing some new switching devices that can be made smaller, faster, and cheaper than silicon transistors? Is the decade-long era of exponentially compounding improvements in the cost and performance of computing devices about to end? These were some of the questions raised by a panel session, “Looking Beyond Moore’s Law, A Technical Perspective” [18]. Looking into the past 20 years, the scaling of planar bulk silicon MOSFETs has been very successful that offered roughly 17% device performance enhancement every year [10]. In the coming 15-20 years, the ITRS projects the future device performance to closely follow this historical trend [19]. However, theoretical calculations suggest that by simply maintaining the same planar bulk geometry and/or Si channel, the MOSFET may not be able to keep up with the required performance beyond the year 2015 [19]. The saturation of drive current and the increase in off-state leakage current upon dimension scaling are the limiting factors for future technology nodes. By using new materials in the channel which have higher electron mobility and therefore high injection velocity, the drive current can be increased, and thus the continued scaling can be sustained. Therefore, looking into the future scaling of MOSFETs, it becomes important that higher mobility materials such as III-V together with innovative device structures such as 1-D NWs which may perform better than even very highly strained Si.

Although there is no hard limit to the scaling of FETs down to sub-10nm gate lengths (11 nm technology node), it is generally agreed that performance-dimension scaling will require significant innovations to enhance the transport characteristics of channel materials [20]. New CMOS device structures and materials will be introduced in the future nodes as they are proven economically necessary [21]. Integration of III-

V materials onto Si with novel device structures such as NWs and materials such as InSb and InAs, can take the technology node to sub-10 nm era with the continued increment in performance.

There has been extensive research on many other alternative devices, such as spintronic devices [22], single electron transistors [23], resonant tunneling diodes [24–43], band-to-band tunneling transistors [4,7,44–54], nanoparticle and molecular based devices [55–58], and NW FETs [1, 2, 59–70]. But it is not yet clear which of these new and novel materials and devices will lead the future technology. However, it is certain that new materials and novel architectures will have to be incorporated onto Si technology to sustain Moore's law for technologies beyond 2020.

Due to their small electron effective mass, III-V materials such as InAs, InSb, and other ternary compound materials such as InGaAs are being investigated as high mobility channel materials for high performance nMOS. For beyond 10 nm technology node various other new materials and novel architectures are being considered. Examples are semiconductor NWs (based on both III-V and Si), carbon nanotubes and single layer of carbon called graphene [10]. NWs are 1-D structures that can be grown with uniform diameter in the nanoscale, without the use of nanometer patterning or lithography technology. In this thesis, i) III-V materials such as InSb and InAs are being investigated as novel channel materials for NW-based FETs for high-speed, low-power logic applications, and ii) III-V materials such as InP and GaN are being investigated as channel materials in NW FETs for high-speed, high-power microwave applications.

## 1.3 III-V CMOS: Experimental exploration

Robust and highly manufacturable new process technologies, such as chemical vapor deposition (CVD), metalorganic CVD, atomic layer deposition (ALD), advanced lithography tools (EBL, FIB), hetero-epitaxy and metal gates, have opened up the opportunity to integrate III-V compound semiconductors onto Si technology. In Refs. [71–76] both $n$- and $p$-type III-V MOSFETs have been realized experimentally as potential candidates for future CMOS devices.

III-V compound semiconductor materials have become the most attractive candidates for the next generation FETs also because of the unique possibilities they offer for the rational control of fundamental properties such as dimension, composition, and doping during growth [77]. They can be considered promising candidates for future high-speed, low-power electronic devices because of their narrow bandgap ($E_g = 0.35$ eV for InAs, for example) implying a small electron effective mass, which yields a very high mobility. For undoped InAs, the mobility can be as high as $\mu_n = 33000 cm^2/Vs$, compared with $\mu_n = 1500 cm^2/Vs$ for Si [66]. Furthermore, the large intravalley separation energy of InAs allows for a high saturated velocity, $v_{sat} = 4 \times 10^7 cm/s$.

## 1.4 Nanowire field-effect transistors

A semiconductor NW is a solid rod with a diameter less than 100-200 nm composed of one or several semiconductor materials. The emergence of bottom-up chemical methods to fabricate one-dimensional semiconductor structures has enabled the synthesis of nanoscale wires that can serve as both devices and interconnects in nanoelectronic circuits. NW FETs have been of particular interest recently, both as vehicles for the investigation of basic carrier-transport behavior and as potential future high-

performance electronic devices.

NWs can be synthesized using many techniques, such as molecular/chemical beam epitaxy, vapor phase epitaxy/chemical vapor deposition, and laser ablation. The particular method used influences the potential material compositions, doping and crystal quality, and growth rate.

Most efforts on NW devices have focused on making NW FETs. Researchers have typically either studied individual NWs [61] randomly placed on a substrate or have tried to fabricate transistors from arrays of vertical NWs including wrap-gated Si [62] and InAs NWs [64], even incorporating first attempts at bandgap engineering into the channel of the device. The carrier mobilities of such NW FETs have been shown to match or exceed those of planar silicon, and the devices can achieve gain. Furthermore, p- and n-type nanowires can be assembled into crisscross arrays where the junctions of the crossed wires serve as on-off switches.

A wide range of NW-based devices and systems, including transistors and circuits [67, 68, 78], light emitters [79–82], and sensors [83] have been explored. NW FETs based on III-V [63–66, 78, 84, 85] and II-VI [69, 86] compound semiconductor materials have demonstrated promising FET characteristics in various gate geometries, namely, top-gate [70,87], back-gate [78,85], wrap-around-gate [65,69], and core-shell [88] structures.

InSb and InAs NWs, in particular, are attractive candidates for NW FETs due to their high electron mobility at room temperature [89] and low contact resistance [90]. The low contact resistance is particularly important in nanoscale systems. Contacts often define performance at the nanoscale. Reports on fabricated both $n$- and $p$-type InSb quantum well FETs show that InSb-based quantum well FETs can achieve equivalent high performance with lower dynamic power dissipation [72–75]. Recent discovery of the observation of Quantum Spin Hall effect in III-V [91] and II-VI [92]

materials have also motivated the field of spintronics largly due to the fact that there is the possibility of low power logic devices design using the spin degree of freedom of the electron [93, 94]. Furthermore, III-V materials also have the potential to be used as channel material in NW band-to-band tunneling FETs (TFETs) for high-speed, low-power [4, 47, 48, 95] and high-speed, high-power [6, 7] applications.

## 1.5 Simulation issues and challenges

Physics based simulation for nanoscale devices offer important insight into their operation and help their design optimization. Numerical simulation also helps to guide experiments and can become a tool for explaining experimental results. Device modeling at nanoscale consists of solving Schrödinger equation and the Poisson equation self-consistently. Schrödinger equation quantum mechanically calculates the charge densities and their transmission probabilities, and the Poisson equation solves for the potential and ensures that the charge profile is consistent with the potential profile.

The basis for calculating transport in semiconductor heterostructures is the underlying electronic bandstructures. Electronic bandstructure calculation can be performed in various ways [96]. The *ab initio* methods, such as Hartree-Fock or Density Functional Theory (DFT), calculate the electronic structure from first principles, i.e., without the need for empirical fitting parameters. These methods use a variational approach to calculate the ground state energy of a many-body system where the system is defined at atomic level. In contrast to *ab initio* method, the empirical methods, such as tight-binding [97] and the $\mathbf{k} \cdot \mathbf{p}$ method [98] involve empirical parameters to fit experimental data. Of them, the $\mathbf{k} \cdot \mathbf{p}$ method is widely used for direct bandgap semiconductors due to its simplicity and capability to capture essential physics of the materials in the band extrema. $\mathbf{k} \cdot \mathbf{p}$ method is based upon perturbation the-

ory [99, 100]. In this method, the energy is calculated near a band maximum or minimum by considering the wavevector as a perturbation. In this work, an 8-band $\mathbf{k} \cdot \mathbf{p}$ method is used to calculate the electronic bandstructure of the semiconductor NWs.

Non-equilibrium Green's function (NEGF) formalism of transport calculation has been used over a decade with great success in quantum mechanical device simulation for both silicon and compound semiconductor devices. Recursive Green's function (RGF) algorithm has proven to be the most efficient method of solving the NEGF equation for layered device structures. Atomistic tight-binding modeling of nanostructures using $s$, $p$ and $d$ atomic orbital-like basis has been successful for the predictive simulation of 1-D structures such as resonant tunneling diodes [101], and zero dimensional system such as quantum dots [102]. However, atomistic tight-binding modeling of nanostructures using $s$, $p$ and $d$ type atomic orbital basis can be computationally very expensive for large structures [103]. Modeling a realistically sized device using such computationally demanding approach can take months to generate transport data. Using perturbation approach such as $\mathbf{k} \cdot \mathbf{p}$ method, the computational burden can be reduced without compromising accuracy, at least for direct gap semiconductors. For instance, consider an InSb NW with a core/shell structures. The number of orbitals per layer ($\mathrm{N}_{xy}$) within $sp^3$ basis become respectively 1568, 3528, 14792 for 2, 4, 10 nm core diameters. RGF algorithm requires inverting a full $\mathrm{N}_{xy} \times \mathrm{N}_{xy}$ matrix for each layer. Inverting, factorizing, and multiplying such matrices are computationally very demanding [103]. If instead, a discretized $\mathbf{k} \cdot \mathbf{p}$ method is used to calculate the electronic structures, then with a proper discretization length (say 1 nm), $\mathrm{N}_{xy}$ can be reduced by 16 and computer memory required to store the data can be saved by 256. However, a compromise has to be done in using the discretization length between accuracy and computational burden.

## 1.6 Outline of the thesis

This thesis is divided into the following chapters:

- Chapter 2 (page 13): This chapter presents the material systems that we have investigated. We have considered technologically important materials systems for high-speed, low-power and high-speed, high-power applications. Examples are InSb, InAs, InP, and GaN binary materials. It outlines the methods that we use to calculate the electronic structures of the NWs, namely 8-band $\mathbf{k} \cdot \mathbf{p}$ method and 2-band effective mass method. We use a discretization technique to solve the $\mathbf{k} \cdot \mathbf{p}$ equations in three dimensions. The 8-band $\mathbf{k} \cdot \mathbf{p}$ method becomes computationally very demanding for larger diameter NWs. We use a simple 2-band effective mass model to calculate the electronic bandstructures for larger diameter NWs for $n$-type NW FETs. We verify the validity of this model for the conduction bands by comparing the zone center effective mass with those calculated with the $\mathbf{k} \cdot \mathbf{p}$ method. It is found that the bandgap and the zone center effective mass for the NWs reduce with an increase of the NW diameters. The approach and the results based on in this chapter are published in [1–3, 5, 6, 104].

- Chapter 3 (page 34): This chapter details the method for determining retarded Green's function $G^R$ and matrix for electron correlation operator $G^<$ that are used to calculate the charge and transport characteristics within non-equilibrium Green's function (NEGF) formalism. Recursive Green's function (RGF) algorithm is used to solve the NEGF equations. RGF algorithm is, arguably, the most efficient technique known to solve the NEGF equations. The full details and derivations for NEGF equations can be found in [105–107]. Fi-

nally the self-consistency between charge calculation and potential calculation is described. The results of this chapter is published in [2, 4, 108].

• Chapter 4 (page 46): This chapter presents simulated results of NW FETs based on InSb and InAs materials. These NW FETs offer high-speed and low-power operation. We have calculated the energy dispersion using 8-band $\mathbf{k} \cdot \mathbf{p}$ method. Parameters such as bandgaps and lowest conduction band zone center effective masses that have been extracted from the calculated $E$-$k$ relations are presented. These results are important for experimental works based on these materials. The results of this chapter are published in [1].

• Chapter 5 (page 61): In this chapter, device performance metrics for InSb, InAs NW FETs are presented. Both $n$- and $p$-type devices are considered. Two operational regimes for these nanoscale devices are identified - the quantum capacitance limit (QCL) and the classical capacitance limit (CCL). Device performance in these two opposite limits are described. It is shown that devices operating in the QCL offer significantly improved performance in terms of energy, power and delay. Optimum design solutions are also presented which will guide experimental works based on these materials. The results of this chapter are published in [1–3, 5, 6, 104].

• Chapter 6 (page 110): GaN-based NW FETs are being investigated for high-speed and high-power applications. The simulated results on GaN NW FETs are presented in this chapter. It is found that deeply scaled GaN NW FETs operate in the CCL, in stead of operating in the QCL. This is a result of the high electron effective mass and consequent high density of states. The results of this chapter are published in [6].

- Chapter 7 (page 118): This chapter presents the calculations and results of InSb, InAs, InP, and GaN NW band-to-band tunneling FETs (TFETs). In the first part, a generalized approach that we have developed to calculate the tunneling current for TFETs as a function of material, diameter, and electric field is presented. The approach is based on calculating the imaginary wave vectors in the bandgap. The approach allows one to perform a quick, preliminary evaluation of a large design-space to narrow down the possibilities for more in-depth modeling and analysis or designing split lots for empirical testing. In the second part, a generalized method to calculate the effects of 'band-tails' on the subthreshold characteristics of the NW TFETs are presented. The 'band-tails' can result from heavy doping, impurities, and/or phonons. The results of this chapter is published in [4, 6, 7, 108].

- Chapter 8 (page 165): In this chapter, summary of the thesis is presented and future works directly related to this dissertation is suggested.

# Chapter 2

# Electronic bandstructures

Full device modeling starts with calculating the electronic structures which basically defines the material under investigation in nanoscale. Therefore, electronic structures calculation is crucial for any nanoscale device modeling. In this chapter, we present the materials systems that we have investigated in Sec. 2.1. In Sec. 2.2, we present the details of the method that we use to calculate the electronic structures of III-V nanowires (NWs), namely, an 8-band $\mathbf{k} \cdot \mathbf{p}$ model. We start with a general description of the $\mathbf{k} \cdot \mathbf{p}$ in Sec. 2.2.1, then we introduce the spin-orbit coupling in for III-V materials in Sec. 2.2.2, and discuss the details of the multi-band $\mathbf{k} \cdot \mathbf{p}$ method and the discretization techniques in Sec. 2.2.3. Finally, we also present a 2-band effective mass model that we use to calculate the electronic structures of larger diameter NWs in Sec. 2.2.4, and the model for calculating the imaginary wave vector in the bandgap of the materials is presented in Sec. 2.2.5.

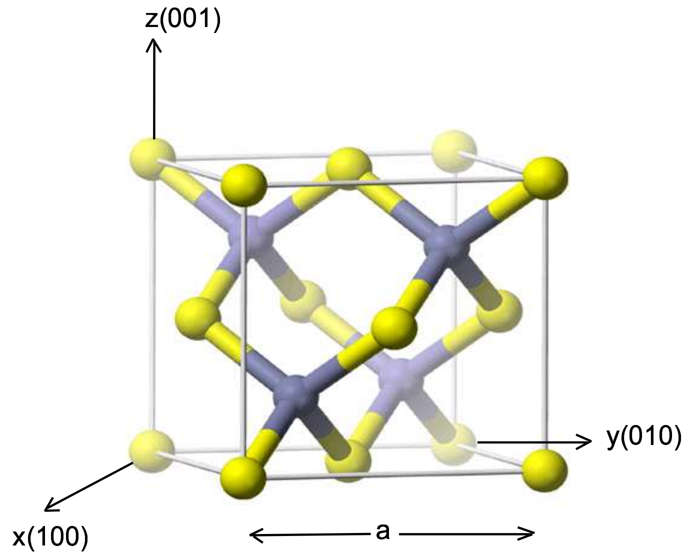



Figure 2.1: *Zinc blende crystal structure.*

## 2.1 Zinc blende materials systems

III-V materials such as InAs and InSb exhibit zinc blende crystal structure [109]. The zinc blende crystal consists of two interpenetrating face-centered-cubic (FCC) lattices, one having a group III element atom (i.e., In) and the other a group V element atom (i.e., As). A zinc blende crystal is characterized by a single lattice constant, $a$. Fig. 2.1 shows a typical zinc blende crystal structure. It is based on an FCC lattice of anions and cations. The cations occupy one of the tetrahedral holes. Zinc blende is best thought of as an FCC array of anions-cations occupying one half of the tetrahedral holes. Each ion is 4-coordinate and has local tetrahedral geometry.

The choice of InAs and InSb for high-speed and low-power NW FETs is motivated by their physical properties: first of all, they are direct gap materials, the bandgap is small, only 0.35 eV for InAs and 0.23 eV for InSb, and problems with electrical contacting of wires should be minimal. Secondly, the low electron effective mass, $m^* =$

$0.023m_0$ for InAs and $m^* = 0.013m_0$ for InSb, provides strong quantum confinement effects and a large energy level separation in the wires. The mobility is also high due to the low effective mass, which is of interest for high-speed device application.

Another feature which is distinctive for these semiconductor materials (InAs) [110] is that the Fermi level is known to pin in the conduction band at the surface, at least for bulk material. Because of this, there is an accumulation of carriers at the surface. This behavior is in contrast to GaAs, which has a surface depletion that limits the minimum feasible diameter of a wire unless the surface is passivated. In principle, any metal should thus result in a good, Ohmic contact to InAs. However, a disadvantage with the strong pinning in the conduction band might be that it prevents realization of p-type InAs NWs. Most recently, experimental techniques have been developed to unpinning the Fermi level for the design of III-V FET source/drain contacts [111]. We have investigated InSb/InP and InAs/InP core/shell NWs as the channel for the NW FETs and NW TFETs. The core region contains InSb or InAs materials and the shell is InP, which acts as the dielectric for the NW FETs.

III-V materials such as InP and GaN have direct wide bandgaps, large breakdown fields, and high saturation velocities and are considered for applications in high-speed, high-power nanoelectronics. These materials also exhibit zinc blende crystal structure. We have investigated InP/AlP and GaN/AlN core-shell NWs as channel for high-speed and high-power NW FETs and NW TFETs. The material parameters for these materials systems are presented in Table 2.1 [109].

## 2.2 Electronic structures calculation

Electronic bandstructure calculation can be performed in various ways [96]. The *ab initio* methods, such as Hartree-Fock or Density Functional Theory (DFT), calcu-

Table 2.1: *Material Constants*

| Symbols | InAs | InSb | InP | GaN |
|---|---|---|---|---|
| $E_g$(300K) eV | 0.417 | 0.235 | 1.4236 | 3.3 |
| $\Delta_0$ (eV) | 0.39 | 0.81 | 0.108 | 0.017 |
| $m_e$(emu) | 0.026 | 0.0135 | 0.0795 | 0.15 |
| $\gamma_1$ | 20.0 | 34.8 | 5.08 | 2.67 |
| $\gamma_2$ | 8.5 | 15.5 | 1.6 | 0.75 |
| $\gamma_3$ | 9.2 | 16.5 | 2.1 | 1.1 |
| $E_p$ (eV) | 21.1 | 23.3 | 20.7 | 25.0 |
| $a$(300$^0$K) $\AA$ | 6.0583 | 6.4794 | 5.8697 | 4.50 |

late the electronic structure from first principles, i.e., without the need for empirical fitting parameters. These methods use a variational approach to calculate the ground state energy of a many-body system where the system is defined at atomic level.

In contrast to *ab initio* method, the empirical methods, such as tight-binding [97] and the $\mathbf{k} \cdot \mathbf{p}$ method [98] involve empirical parameters to fit experimental data. Of them, the $\mathbf{k} \cdot \mathbf{p}$ method is widely used for direct bandgap semiconductors due to its simplicity and capability to capture essential physics of the materials in the band extrema. $\mathbf{k} \cdot \mathbf{p}$ method is based upon perturbation theory [99,100]. In this method, the energy is calculated near a band maximum or minimum by considering the wavevector as a perturbation. It uses a minimal set of parameters that are determined empirically from experiments. By means of a perturbative approach, it provides a continuation in the wave vector $k$ of the energy bands in the vicinity of some special point in the Brillouin zone (BZ). In our work, we use a three dimensionally (3-D) discretized 8-band $\mathbf{k} \cdot \mathbf{p}$ method to calculate the electronic bandstructures of III-V NWs.

The electronic wave functions that satisfy the Schrödinger equation with a periodic lattice potential in a bulk crystal are given by Bloch's theorem:

$$\psi_{nk}(r) = e^{ikr} u_{nk}(r). \tag{2.1}$$

The cell-periodic Bloch function $u_{nk}(r)$ depend on the band index $n$ and the envelope function wave vector $k$. The wave function $\psi(r)$ form a complete set of states as do the wave functions based on Bloch functions at any other wave vector, including the wave vectors at special points in the BZ [98]. In treating the electronic properties of direct gap semiconductors, it is usual to consider the zone-center Γ-point Bloch functions $u_{n0}(r)$ for the wave function expansions, and drop the reference to the $k$=0 index for these functions.

The general form of the wave functions may be considered to be linear combination of a finite number of band wave functions of the form

$$\psi(r) = f_n(z)e^{ik_x x}e^{ik_y y}u_n(r) \equiv F_n(r)u_n(r). \tag{2.2}$$

The envelope functions $F_n(r)$ are typically considered to be slowly varying, whereas the cell periodic and more oscillatory Bloch functions satisfy Schrödinger's equation with band-edge energies.

### 2.2.1 General description of the $\mathbf{k} \cdot \mathbf{p}$ method

The $\mathbf{k} \cdot \mathbf{p}$ method can be derived from one-electron Schrödinger equation,

$$H\psi_n(\mathbf{r}) = \left(\frac{p^2}{2m} + V(\mathbf{r})\right)\psi_n(\mathbf{r}) = E_n\psi_n(\mathbf{r}), \tag{2.3}$$

where $H$, $\psi_n(\mathbf{r})$, and $E_n$ denote, respectively, the one-electron Hamiltonian, and the wave function and energy of an electron in an eigen state labeled by $n$. Using the Bloch theorem, the solutions of Eq. 2.3 are expressed, in reduced zone scheme, as given in Eq. 2.2. When $\psi_{n\mathbf{k}}$ is substituted into Eq. 2.3, we obtain an equation in $u_{n\mathbf{k}}$

of the form,

$$\left(\frac{p^2}{2m} + \frac{\hbar\mathbf{k}\cdot\mathbf{p}}{m} + \frac{\hbar^2k^2}{2m} + V\right)u_{n\mathbf{k}} = E_{n\mathbf{k}}u_{n\mathbf{k}}. \tag{2.4}$$

At $\mathbf{k}_0 = (0,0,0)$, Eq. 2.4 reduces to

$$\left(\frac{p^2}{2m} + V\right)u_{n\mathbf{0}} = E_{n\mathbf{0}}u_{n\mathbf{0}} \quad (n = 1,2,3,...). \tag{2.5}$$

Similar equation can be obtained for $\mathbf{k}$ equal to any point $\mathbf{k}_0$. The solutions of Eq. 2.5 form a complete and orthonormal set of basis functions. Once $E_{n\mathbf{0}}$ and $u_{n\mathbf{0}}$ are known, the terms $\hbar\mathbf{k}\cdot\mathbf{p}/m$ and $\hbar^2k^2/(2m)$ can be treated as perturbation using either degenerate or non-degenerate perturbation theory [96]. Since the perturbation terms are proportional to $k$, the method works best for small values of $k$ (in the zone center).

Once the $E$-$k$ is known for a particular material system, effective of a non-degenerate band can be easily determined using $\mathbf{k}\cdot\mathbf{p}$ method. Let us assume that the band structure has an extremum at the energy $E_{n\mathbf{0}}$ and the band is non-degenerate at this energy. Using standard non-degenerate perturbation theory, the eigen functions $u_{n\mathbf{k}}$ and eigen values $E_{n\mathbf{k}}$ at a neighboring point $\mathbf{k}$ can be expanded to second order in $k$ in terms of the unperturbed wave functions $u_{n\mathbf{0}}$ and energies $E_{n\mathbf{0}}$ by treating the terms involving $k$ in Eq. 2.4 as perturbations.

$$u_{n\mathbf{k}} = u_{n\mathbf{0}} + \frac{\hbar}{m}\sum_{n'\neq n}\frac{\langle u_{n\mathbf{0}}|\mathbf{k}\cdot\mathbf{p}|u_{n'\mathbf{0}}\rangle}{E_{n\mathbf{0}} - E_{n'\mathbf{0}}}u_{n'\mathbf{0}}, \tag{2.6}$$

and

$$E_{n\mathbf{k}} = E_{n\mathbf{0}} + \frac{\hbar^2k^2}{2m} + \frac{\hbar^2}{m^2}\sum_{n'\neq n}\frac{|\langle u_{n\mathbf{0}}|\mathbf{k}\cdot\mathbf{p}|u_{n'\mathbf{0}}\rangle|^2}{E_{n\mathbf{0}} - E_{n'\mathbf{0}}}. \tag{2.7}$$

For small values of $k$, the energy $E_{n\mathbf{k}}$ becomes,

$$E_{n\mathbf{k}} = E_{n\mathbf{0}} + \frac{\hbar^2 k^2}{2m^*}, \tag{2.8}$$

where $m^*$ is defined as the 'effective mass' of the band. Comparing Eqs. 2.7 and 2.8, we obtain an expression for the effective mass,

$$\frac{1}{m^*} = \frac{1}{m} + \frac{2}{m^2 k^2} \sum_{n' \neq n} \frac{|\langle u_{n\mathbf{0}} | \mathbf{k} \cdot \mathbf{p} | u_{n'\mathbf{0}} \rangle|^2}{E_{n\mathbf{0}} - E_{n'\mathbf{0}}}. \tag{2.9}$$

Obvious is the fact from Eq. 2.9 is that the an electron in a solid has a different mass from that of an electron in free space because of coupling between electronic states in different bands via the $\mathbf{k} \cdot \mathbf{p}$ term.

### 2.2.2 Spin-orbit coupling

Before going into details of calculating the electronic dispersions using $\mathbf{k} \cdot \mathbf{p}$ method, it is useful to describe the spin-orbit coupling and the spin-orbit Hamiltonian. It is well-known that the electron spin can be coupled to the orbital angular momentum via the spin-orbit interaction. The spin-orbit coupling is a relativistic effect which scales with the atomic number of the atom. Thus for semiconductors containing heavier atoms such as As and Sb, one expects the spin-orbit coupling to be significant and has to include it in the unperturbed Hamiltonian, in particular, for states near $k$=0. The Hamiltonian for spin-orbit coupling is given by [96]

$$H_{so} = \frac{\hbar}{4c^2 m^2} (\nabla V \times \mathbf{p}) \cdot \sigma, \tag{2.10}$$

where the components of $\sigma$ are the Pauli spin matrices:

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}; \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}; \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{2.11}$$

Eq. 2.10 can also be written in the following form as an operator on the cell-periodic function [112]

$$H_{so} = \frac{\hbar}{4c^2m^2}(\nabla V \times \mathbf{p}) \cdot \sigma + \frac{\hbar}{4c^2m^2}(\nabla V \times \mathbf{k}) \cdot \sigma, \tag{2.12}$$

The first term is independent of $k$ and is analogous to the atomic spin-orbit splitting term. The second term is proportional to $\mathbf{k}$ and is the additional spin-orbit energy coming from the crystal momentum. The spin-orbit splitting in semiconductors occurs in the valence band. The splitting is more for crystals whose constituent atoms have higher atomic number. In fact, the spin-orbit splitting energy $\Delta$ of semiconductors increases as the fourth power of the atomic number of the constituent atoms. That is because the atomic number is equal to the number of protons, which determines the electric field seen by the valence electrons.

### 2.2.3 Multiband $\mathbf{k} \cdot \mathbf{p}$ method

The multiband effective mass equation [113] is widely used to describe the bandstructure of the low-dimensional structures. It can be expressed as,

$$i\hbar\frac{\partial}{\partial t}\Psi_\nu(\mathbf{r}, t) = \sum_{\nu'} H_{\nu\nu'}(-i\nabla)\Psi_{\nu'}(\mathbf{r}, t) + U(\mathbf{r}, t)\Psi_\nu(\mathbf{r}, t), \tag{2.13}$$

where $\nu$ represents band index and $H_{\nu,\nu'}(\mathbf{k})$ is defined as

$$H_{\nu,\nu'}(\mathbf{k}) = \begin{cases} \dfrac{\hbar^2 k^2}{2m_0} + E_{\nu 0}, & \nu' = \nu \\ \dfrac{\hbar \mathbf{P}_{\nu\nu'} \cdot \mathbf{k}}{m_0}, & \nu' \neq \nu. \end{cases} \tag{2.14}$$

$P_{\nu\nu'}$ is called the momentum matrix element between bands $\nu$ and $\nu'$ and is defined as

$$P_{\nu\nu'} = -i\hbar \langle \bar{u}_{\nu,0} | \nabla \bar{u}_{\nu',0} \rangle. \tag{2.15}$$

Here $u_{\nu,0}$ represents the zone center Bloch functions for the $\nu$th band. The relation between the actual wavefunction and the multiband envelope functions follows readily from,

$$\Psi_0(\mathbf{r}, t) = \sum_{\nu} u_{\nu,0}(\mathbf{r}, t) \Psi_\nu(\mathbf{r}, t). \tag{2.16}$$

Eq. (2.13) can be solved for the perfect spatially uniform semiconductor using a variety of well known techniques, but in a quantum well (QWell), and quantum wire (QWire) the crystal composition and/or strain varies from region to region and approximations are needed in order to solve Eq. (2.13). Many such approximate methods are now well known and are extensively used [114]. In our work, we apply an 8-band $\mathbf{k} \cdot \mathbf{p}$ method [115]. The choice of how many bands will be needed depends on the details of the problem to be solved. For our work, we included eight basis functions in the set, namely, the spin-up and spin-down $s$ and $p$ atomic orbital-like states. These are arranged in the following order: $|S \uparrow\rangle$, $|X \uparrow\rangle$, $|Y \uparrow\rangle$, $|Z \uparrow\rangle$, $|S \downarrow\rangle$, $|X \downarrow\rangle$, $|Y \downarrow\rangle$ and $|Z \downarrow\rangle$. As a result, the multiband effective mass equation is transformed into eight coupled differential equations. 2.2.

In the basis of the eight zone-centre Bloch functions described above, the matrix

$H$ takes the form:

$$H = \begin{bmatrix} G & \Gamma \\ -\Gamma^* & G^* \end{bmatrix}, \tag{2.17}$$

where $G$ and $\Gamma$ are both $4 \times 4$ matrices.

The matrices $G(\mathbf{k})$ and $\Gamma$ are defined as follows [115]:

$$G(\mathbf{k}) = G_1(\mathbf{k}) + G_2(\mathbf{k}) + G_{so}, \tag{2.18}$$

where

$$G_1 = \begin{bmatrix} E_c & iPk_x & iPk_y & iPk_z \\ -iPk_x & E_{v'} & 0 & 0 \\ -iPk_y & 0 & E_{v'} & 0 \\ -iPk_z & 0 & 0 & E_{v'} \end{bmatrix}, \tag{2.19}$$

$$G_2 = \begin{bmatrix} A'k^2 & Bk_yk_z & Bk_xk_z & Bk_xk_y \\ Bk_yk_z & L'k_x^2+ M(k_y^2+k_z^2) & N'k_xk_y & N'k_xk_z \\ Bk_zk_x & N'k_xk_y & L'k_y^2+ M(k_x^2+k_z^2) & N'k_yk_z \\ Bk_xk_y & N'k_xk_z & N'k_yk_z & L'k_z^2+ M(k_x^2+k_y^2) \end{bmatrix}, \tag{2.20}$$

and

$$G_{so} = -\frac{\Delta}{3} \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & i & 0 \\ 0 & -i & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}. \tag{2.21}$$

The matrix $\Gamma$ is:

$$\Gamma = -\frac{\Delta}{3} \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & i \\ 0 & 1 & -i & 0 \end{bmatrix}. \tag{2.22}$$

All the parameters appearing in Eqs. (2.19)-(2.22) are real. The parameters $A'$, $B$,

$L^{'}$, $M$ and $N^{'}$ are known as Kane parameters and are defined in [115]. The parameter $B$, which is equal to zero for crystals with inversion symmetry, is set to zero in our calculation as well. The parameter $P$, which is known as the optical matrix element, is proportional to the momentum matrix element between the conduction band and the valence band:

$$P = -i\frac{\hbar}{m_0}\langle S|p_i|i\rangle, \tag{2.23}$$

where $i = X, Y$ or $Z$. The parameter $P$ plays an important role in calculating the optical transition strength. The quantity $\Delta$ is the spin orbit splitting parameter, while $E_c$ and $E_{v'}$ are the band edge energies in the absence of the spin-orbit coupling. Spurious solutions of the $\mathbf{k} \cdot \mathbf{p}$ equations are eliminated by methods suggested in Refs. [116, 117]

For a bulk crystal for $k = 0$, solving the dispersion relation to second order in $\mathbf{k}$ allows one to obtain:

$$\begin{aligned}
\frac{m_0}{m_{hh}(100)} &= \gamma_1 - 2\gamma_2, \qquad \frac{m_0}{m_{lh}(100)} = \gamma_1 + 2\gamma_2 \\
\frac{m_0}{m_{hh}(111)} &= \gamma_1 - 2\gamma_3, \qquad \frac{m_0}{m_{lh}(111)} = \gamma_1 + 2\gamma_3 \\
\frac{m_0}{m_e} &= \frac{2m_0}{\hbar^2}\left(A^{'} + \frac{P^2(E_g + \frac{2}{3}\Delta)}{E_g(E_g + \Delta)}\right) \\
\frac{m_0}{m_{so}} &= \gamma_1 - \frac{2m_0 P^2 \Delta}{3\hbar^2 E_g(E_g + \Delta)},
\end{aligned} \tag{2.24}$$

where $m_{hh(lh)}(ijk)$ is the heavy (light) hole band edge effective mass in the $(ijk)$ crystallographic direction, $m_e$ is the conduction band effective mass and $m_{so}$ is the split-off band effective mass. The $\gamma_i$ are known as the Luttinger parameters [115]. In

terms of the $\gamma_i$, the constants in the matrix are given by:

$$\begin{aligned} L^{'} &= -\frac{\hbar^2}{2m_0}(1+\gamma_1+4\gamma_2)+\frac{P^2}{E_g} \\ M &= -\frac{\hbar^2}{2m_0}(1+\gamma_1-2\gamma_2) \\ N^{'} &= -\frac{3\hbar^2}{m_0}\gamma_3+\frac{P^2}{E_g}, \end{aligned} \tag{2.25}$$

where $E_g = E_c - E_v$ is the band gap with $E_v = E_{v'} + \frac{\Delta}{3}$. The mathematical relations described in Eqs. (2.25)-(2.26) define all the $8 \times 8$ Hamiltonian parameters in terms of the experimental bandgap, effective masses and the spin-orbit splitting.

For bulk structures, each component of $\mathbf{k}$ in Eq. (2.17) is a number, but for low dimensional structures, $k_n$ along each confined direction $n$ is replaced by the differential operator $-i\partial/\partial x_n$. We write the discrete first and second order derivatives of the wavefunction $\psi(x,y) \to \psi_{i,j}$ as follows, using a finite difference method [118]:

$$\begin{aligned} \frac{\partial\psi}{\partial x} &= \frac{\psi_{i+1,j}-\psi_{i-1,j}}{2\Delta x} \\ \frac{\partial^2\psi}{\partial x^2} &= \frac{1}{\Delta x^2}(\psi_{i-1,j}-2\psi_{i,j}+\psi_{i+1,j}) \\ \frac{\partial^2\psi}{\partial x\partial y} &= \frac{1}{4\Delta x^2}(\psi_{i-1,j-1}-\psi_{i+1,j-1}-\psi_{i-1,j+1}+\psi_{i+1,j+1}) \end{aligned} \tag{2.26}$$

In the $\mathbf{k}\cdot\mathbf{p}$ Hamiltonian, the product of any material parameter and the derivative should be written such that the Hamiltonian remains Hermitian. We use a technique as discussed in [119]. Using this method, the product of any material parameter $Q$

and the first order derivative is treated as follows:

$$\begin{aligned}
Q\frac{\partial}{\partial x}\psi_{i,j,k} &= \frac{1}{2}\left[Q_{i,j}\frac{\partial}{\partial x}\psi_{i,j,k} + \frac{\partial}{\partial x}\left(Q_{i,j}\psi_{i,j,k}\right)\right] \\
&= \frac{1}{2}\left[Q_{i,j}\frac{\psi_{i+1,j,k}-\psi_{i-1,j,k}}{2\Delta x} + \frac{Q_{i+1,j}\psi_{i+1,j,k}-Q_{i-1,j}\psi_{i-1,j,k}}{2\Delta x}\right] \\
&= \left[\frac{-Q_{i,j}-Q_{i-1,j}}{4\Delta x}\right]\psi_{i-1,j,k} + \left[\frac{Q_{i,j}+Q_{i+1,j}}{4\Delta x}\right]\psi_{i+1,j,k} \qquad (2.27)
\end{aligned}$$

Here, $i$, $j$, and $k$ are the indices in the $x$, $y$, and $z$ direction respectively. The material parameter $Q$ is a function of distance in the plane ($x$,$y$) and the wavefunction is a function of $x$, $y$, and $z$. Furthermore, the product $Q\psi$ is assumed to be an aggregate function and the derivative for that is computed as shown. It is clear that the well-known ‘product rule‘ of differentiation is not followed in Eq. (2.27). Similarly, we can calculate the other two first order derivative terms.

$$\begin{aligned}
Q\frac{\partial}{\partial y}\psi_{i,j,k} &= \left[\frac{-Q_{i,j}-Q_{i,j-1}}{4\Delta y}\right]\psi_{i,j-1,k} + \left[\frac{Q_{i,j}+Q_{i,j+1}}{4\Delta y}\right]\psi_{i,j+1,k} \\
Q\frac{\partial}{\partial z}\psi_{i,j,k} &= \left[\frac{-Q_{i,j}}{2\Delta z}\right]\psi_{i,j,k-1} + \left[\frac{Q_{i,j}}{2\Delta z}\right]\psi_{i,j,k+1} \qquad (2.28)
\end{aligned}$$

For the product of the first order derivative terms, we proceed as follows:

$$Q\frac{\partial}{\partial x}\frac{\partial}{\partial y}\psi_{i,j,k} = \frac{1}{2}\left[\frac{\partial}{\partial x}\left(Q_{i,j}\frac{\partial}{\partial y}\psi_{i,j,k}\right) + \frac{\partial}{\partial y}\left(Q_{i,j}\frac{\partial}{\partial x}\psi_{i,j,k}\right)\right] \qquad (2.29)$$

The first term of Eq. (2.29) is evaluated as:

$$\begin{aligned}
\frac{1}{2}\frac{\partial}{\partial x}\left(Q_{i,j}\frac{\partial}{\partial y}\psi_{i,j,k}\right) &= \frac{1}{2}\cdot\frac{1}{2\Delta x}\left(Q_{i+1,j}\left[\frac{\partial\psi}{\partial y}\right]_{i+1,j,k} - Q_{i-1,j}\left[\frac{\partial\psi}{\partial y}\right]_{i-1,j,k}\right) \\
&= \frac{1}{4\Delta x}\left[Q_{i+1,j}\frac{\psi_{i+1,j+1,k}-\psi_{i+1,j-1,k}}{2\Delta y} - Q_{i-1,j}\frac{\psi_{i-1,j+1,k}-\psi_{i-1,j-1,k}}{2\Delta y}\right] \\
&= \left[\frac{Q_{i+1,j}}{8\Delta x\Delta y}\right]\psi_{i+1,j+1,k} + \left[\frac{-Q_{i+1,j}}{8\Delta x\Delta y}\right]\psi_{i+1,j-1,k} \\
&+ \left[\frac{-Q_{i-1,j}}{8\Delta x\Delta y}\right]\psi_{i-1,j+1,k} + \left[\frac{Q_{i-1,j}}{8\Delta x\Delta y}\right]\psi_{i-1,j-1,k}
\end{aligned} \tag{2.30}$$

The second term of Eq. (2.29) is evaluated as:

$$\begin{aligned}
\frac{1}{2}\frac{\partial}{\partial y}\left(Q_{i,j}\frac{\partial}{\partial x}\psi_{i,j,k}\right) &= \frac{1}{2}\cdot\frac{1}{2\Delta y}\left(Q_{i,j+1}\left[\frac{\partial\psi}{\partial x}\right]_{i,j+1,k} - Q_{i,j-1}\left[\frac{\partial\psi}{\partial x}\right]_{i,j-1,k}\right) \\
&= \frac{1}{4\Delta y}\left[Q_{i,j+1}\frac{\psi_{i+1,j+1,k}-\psi_{i-1,j+1,k}}{2\Delta x} - Q_{i,j-1}\frac{\psi_{i+1,j-1,k}-\psi_{i-1,j-1,k}}{2\Delta x}\right] \\
&= \left[\frac{Q_{i,j+1}}{8\Delta x\Delta y}\right]\psi_{i+1,j+1,k} + \left[\frac{-Q_{i,j+1}}{8\Delta x\Delta y}\right]\psi_{i-1,j+1,k} \\
&+ \left[\frac{-Q_{i,j-1}}{8\Delta x\Delta y}\right]\psi_{i+1,j-1,k} + \left[\frac{Q_{i,j-1}}{8\Delta x\Delta y}\right]\psi_{i-1,j-1,k}
\end{aligned} \tag{2.31}$$

Putting back Eq. (2.30) and Eq. (2.31) into Eq. (2.29) results:

$$\begin{aligned}
Q\frac{\partial}{\partial x}\frac{\partial}{\partial y}\psi_{i,j,k} &= \left[\frac{Q_{i+1,j}+Q_{i,j+1}}{8\Delta x\Delta y}\right]\psi_{i+1,j+1,k} + \left[\frac{-Q_{i+1,j}-Q_{i,j-1}}{8\Delta x\Delta y}\right]\psi_{i+1,j-1,k} \\
&+ \left[\frac{-Q_{i-1,j}-Q_{i,j+1}}{8\Delta x\Delta y}\right]\psi_{i-1,j+1,k} + \left[\frac{Q_{i-1,j}+Q_{i,j-1}}{8\Delta x\Delta y}\right]\psi_{i-1,j-1,k}
\end{aligned} \tag{2.32}$$

We can follow the same prescription to determine other product terms.

$$
\begin{aligned}
Q\frac{\partial}{\partial y}\frac{\partial}{\partial z}\psi_{i,j,k} &= \left[\frac{Q_{i,j}+Q_{i,j+1}}{8\Delta y\Delta z}\right]\psi_{i,j+1,k+1}+\left[\frac{-Q_{i,j}-Q_{i,j+1}}{8\Delta y\Delta z}\right]\psi_{i,j+1,k-1} \\
&+ \left[\frac{-Q_{i,j}-Q_{i,j-1}}{8\Delta y\Delta z}\right]\psi_{i,j-1,k+1}+\left[\frac{Q_{i,j}+Q_{i,j-1}}{8\Delta y\Delta z}\right]\psi_{i,j-1,k-1} \\
Q\frac{\partial}{\partial z}\frac{\partial}{\partial x}\psi_{i,j,k} &= \left[\frac{Q_{i,j}+Q_{i+1,j}}{8\Delta z\Delta x}\right]\psi_{i+1,j,k+1}+\left[\frac{-Q_{i,j}-Q_{i-1,j}}{8\Delta z\Delta x}\right]\psi_{i-1,j,k+1} \\
&+ \left[\frac{-Q_{i,j}-Q_{i+1,j}}{8\Delta z\Delta x}\right]\psi_{i+1,j,k-1}+\left[\frac{Q_{i,j}+Q_{i-1,j}}{8\Delta z\Delta x}\right]\psi_{i-1,j,k-1} \qquad (2.33)
\end{aligned}
$$

We calculate the second order derivative terms as follows:

$$
\begin{aligned}
Q\frac{\partial^2}{\partial x^2}\psi_{i,j,k} &= \frac{\partial}{\partial x}\left(Q_{i,j}\frac{\partial}{\partial x}\psi_{i,j,k}\right) \\
&= \frac{1}{\Delta x}\left(\left[Q_{i+\frac{1}{2},j}\frac{\partial\psi}{\partial x}\right]_{i+\frac{1}{2},j,k}-\left[Q_{i-\frac{1}{2},j}\frac{\partial\psi}{\partial x}\right]_{i-\frac{1}{2},j,k}\right) \\
&= \frac{1}{\Delta x}\left[\frac{Q_{i+1,j}+Q_{i,j}}{2}\cdot\frac{\psi_{i+1,j,k}-\psi_{i,j,k}}{\Delta x}-\frac{Q_{i,j}+Q_{i-1,j}}{2}\cdot\frac{\psi_{i,j,k}-\psi_{i-1,j,k}}{\Delta x}\right] \\
&= \left[\frac{Q_{i-1,j}+Q_{i,j}}{2(\Delta x)^2}\right]\psi_{i-1,j,k}+\left[\frac{-Q_{i-1,j}-2Q_{i,j}-Q_{i+1,j}}{2(\Delta x)^2}\right]\psi_{i,j,k} \\
&+ \left[\frac{Q_{i,j}+Q_{i+1,j}}{2(\Delta x)^2}\right]\psi_{i+1,j,k} \qquad (2.34)
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
Q\frac{\partial^2}{\partial y^2}\psi_{i,j,k} &= \left[\frac{Q_{i,j-1}+Q_{i,j}}{2(\Delta y)^2}\right]\psi_{i,j-1,k}+\left[\frac{-Q_{i,j-1}-2Q_{i,j}-Q_{i,j+1}}{2(\Delta y)^2}\right]\psi_{i,j,k} \\
&+ \left[\frac{Q_{i,j}+Q_{i,j+1}}{2(\Delta y)^2}\right]\psi_{i,j+1,k} \\
Q\frac{\partial^2}{\partial z^2}\psi_{i,j,k} &= \left[\frac{Q_{i,j}}{(\Delta z)^2}\right]\psi_{i,j,k+1}+\left[\frac{-Q_{i,j}}{(\Delta z)^2}\right]\psi_{i,j,k}+\left[\frac{Q_{i,j}}{(\Delta z)^2}\right]\psi_{i,j,k-1} \qquad (2.35)
\end{aligned}
$$

Correspondingly, the Hamiltonian in Eq. (2.18) can be divided into eleven Hermitian

matrices as follows:

$$
\begin{aligned}
G\left(\frac{\partial}{\partial x},\frac{\partial}{\partial y},\frac{\partial}{\partial z}\right) &= G_0 + G_x\frac{\partial}{\partial x} + G_y\frac{\partial}{\partial y} + G_z\frac{\partial}{\partial z} - G_{xx}\frac{\partial^2}{\partial x^2} - G_{yy}\frac{\partial^2}{\partial y^2} - G_{zz}\frac{\partial^2}{\partial z^2} \\
&- G_{xy}\frac{\partial^2}{\partial x\partial y} - G_{yz}\frac{\partial^2}{\partial y\partial z} - G_{zx}\frac{\partial^2}{\partial z\partial x} + G_{so},
\end{aligned}
\tag{2.36}
$$

where

$$
G_0 = \begin{bmatrix} E_c & 0 & 0 & 0 \\ 0 & E_{v'} & 0 & 0 \\ 0 & 0 & E_{v'} & 0 \\ 0 & 0 & 0 & E_{v'} \end{bmatrix}, \quad G_x = \begin{bmatrix} 0 & P & 0 & 0 \\ -P & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},
\tag{2.37}
$$

$$
G_y = \begin{bmatrix} 0 & 0 & P & 0 \\ 0 & 0 & 0 & 0 \\ -P & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad G_z = \begin{bmatrix} 0 & 0 & 0 & P \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -P & 0 & 0 & 0 \end{bmatrix},
\tag{2.38}
$$

$$
G_{xx} = \begin{bmatrix} A' & 0 & 0 & 0 \\ 0 & L' & 0 & 0 \\ 0 & 0 & M & 0 \\ 0 & 0 & 0 & M \end{bmatrix}, \quad G_{yy} = \begin{bmatrix} A' & 0 & 0 & 0 \\ 0 & M & 0 & 0 \\ 0 & 0 & L' & 0 \\ 0 & 0 & 0 & M \end{bmatrix},
\tag{2.39}
$$

$$G_{zz} = \begin{bmatrix} A' & 0 & 0 & 0 \\ 0 & M & 0 & 0 \\ 0 & 0 & M & 0 \\ 0 & 0 & 0 & L' \end{bmatrix}, \quad G_{xy} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & N' & 0 \\ 0 & N' & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \tag{2.40}$$

$$G_{yz} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & N' \\ 0 & 0 & N' & 0 \end{bmatrix}, \quad G_{zx} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & N' \\ 0 & 0 & 0 & 0 \\ 0 & N' & 0 & 0 \end{bmatrix},$$

$G_{so}$ and $\Gamma$ are defined as in Eqs. 2.21 and 2.22

To make the discretized $\mathbf{k}\cdot\mathbf{p}$ Hamiltonian look like a nearest neighbor tight-binding Hamiltonian for a layered structure, the NW should be discretized as shown in Fig 2.2. It should be discretized in planes of sites. The indices of the first plane should run from 1 to N, the indices of the second plane should run from N+1 to 2N, etc. The matrix $[D_1]$ in Fig. 2.2 is the block of the Hamiltonian matrix of the isolated 1st plane of sites. The matrix $[t_{1,2}]$ is the block of the Hamiltonian matrix that couples plane 1 to plane 2. If the NW consisted only of the 4 planes shown in Fig. 2.2, then the total Hamiltonian matrix would have the form

$$H = \begin{bmatrix} [D_1] & [t_{1,2}] & 0 & 0 \\ [t_{1,2}]^\dagger & [D_2] & [t_{2,3}] & 0 \\ 0 & [t_{2,3}]^\dagger & [D_3] & [t_{3,4}] \\ 0 & 0 & [t_{3,4}]^\dagger & [D_4] \end{bmatrix} \tag{2.41}$$

Considering the components of the **k** vector in Eq. (2.17) to be numbers and di-

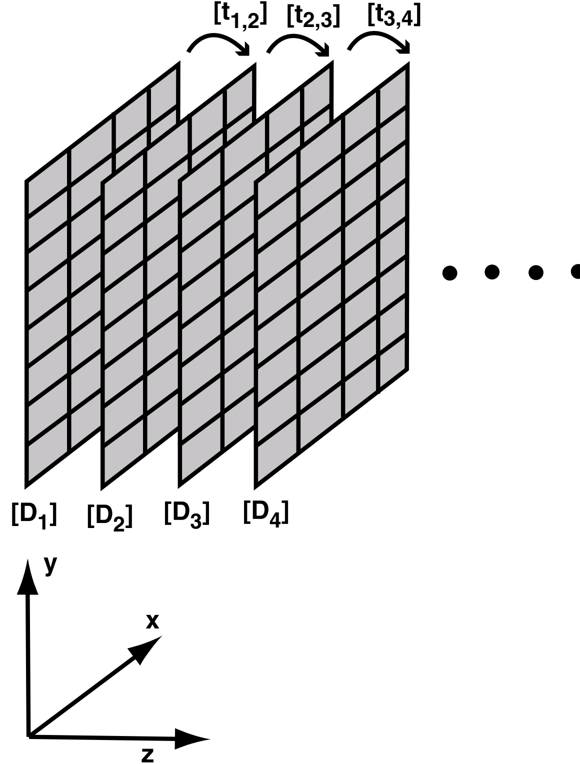



Figure 2.2: *Discretization scheme for NW.*

agonalizing the matrix for a bulk crystal for $k = 0$, one can obtain the dispersion relationships, namely, the **k** dependent eigenvalues $E_n(\mathbf{k})$.

### 2.2.4 2-band effective mass model

For larger diameter NWs, we use a 2-band effective mass model to calculate the mode energies. 2-band effective mass model is found to reasonably accurately model the conduction bands (CBs) of the III-V NWs. This is because the CBs of III-V materials resemble $s$-type atomic orbital-like states, which are parabolic in the zone center. $E$-$k$ dispersions for larger diameter (10 nm - 60 nm) $n$-type NW FETs are calculated by

a 2-band effective mass method [120],

$$E(1+\alpha E) = \frac{\hbar^2 k^2}{2m^*}, \tag{2.42}$$

where $\alpha = \frac{1}{E_G}(1 - \frac{m^*}{m_0})^2$. $E_G = 0.35$ eV is the bulk bandgap, $m^* = 0.026m_0$ is the bulk electron effective mass, and $m_0$ is the bare electron mass. The accuracy of such a method are justified [5] by comparing the zone center electron effective mass with those calculated with the 8-band $\mathbf{k}\cdot\mathbf{p}$ method.

### 2.2.5 Evanescent dispersions

The evanescent dispersions are numerically calculated within the 8-band $\mathbf{k}\cdot\mathbf{p}$ model using the generalized eigenvalue problem as described in [121]. The evanescent dispersion gives the imaginary wave vectors in the bandgap of a material, which is used to calculate carrier tunneling probability through the bandgap - a topic that will be discussed in more details in the results section. To calculate the imaginary wave vectors throughout the bandgap, we numerically solve [122]

$$[\mathbf{A} + z\mathbf{B}] \begin{bmatrix} C_i \\ C_{i-1} \end{bmatrix} = 0, \tag{2.43}$$

where

$$\mathbf{A} = \begin{bmatrix} (D_0 - E\mathbf{I}) & t^\dagger \\ -\mathbf{I} & \mathbf{0} \end{bmatrix}, \tag{2.44}$$

and

$$\mathbf{B} = \begin{bmatrix} t & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}., \tag{2.45}$$

Here **I** and **0** are, respective, unitary and zero matrices of the same size of the Hamiltonians $[D_0]$ and $[t]$.

# Chapter 3

# Charge and transport calculation

With the aggressive scaling down of device dimensions, charge and transport calculation require an atomistic quantum model with an open boundary condition. Non-equilibrium Green's function (NEGF) method has been used for a over a decade with great success in quantum device simulation for both silicon and compound semiconductor devices. Efficient algorithms such as recursive Green's function (RGF) algorithm are required to solve the NEGF equations efficiently for layered device structures. In the following sections, we derive and discuss the RGF algorithm in orthogonal basis for charge calculation. We also derive and discuss the models for transport calculations for such layered structured devices.

## 3.1 Recursive Green's function (RGF) algorithm for charge calculation

If we assume that there is no scattering in the device then all observables of interest, such as electron density, can, in principle, be obtained from the retarded Green's function of the open device. In the derivation of RGF we divide the 2-terminal device

domain into $N$ blocks in the longitudinal direction as shown in Fig. 3.1.

The blocks $\{-\infty, ...., -1, 0\}$ lie on the left contact. The blocks $\{1, 2, ....., N-1, N\}$ belong to the device, and $\{N+1, N+2, ....., \infty\}$ lie on the right contact. Each block can be just a single atomic layer or may be a unit cell. The block size is chosen so that the matrix elements are non-zero only between nearest neighbor blocks. For a self-consistent calculation between electrostatics and charge density, the electron concentration at the $i^{th}$ block can be calculated from the diagonal of the density matrix in the chosen basis representation

$$\begin{aligned} \varrho_{i,i} &= \int \frac{dE}{2\pi}\{A^L_{i,i} f^L + A^R_{i,i} f^R\} \\ &= \int \frac{dE}{2\pi}\{\left[G^R_{i,1}\Gamma_{1,1}G^A_{1,i}\right] f^L + \left[G^R_{i,N}\Gamma_{N,N}G^A_{N,i}\right] f^R\}. \end{aligned} \tag{3.1}$$

Here $A^L$ and $A^R$ are the left and right spectral functions,

$$\begin{aligned} A^L &= G^R\Gamma^L G^A \\ A^R &= G^R\Gamma^R G^A \end{aligned} \tag{3.2}$$

$f^L$ and $f^R$ are the left and right contact Fermi functions, and the block matrix notation is used. To calculate the charge density, we use the RGF algorithm as described in Lake et al. [123].

To derive the algorithm for the first and the last column blocks of retarded Green's function we define two Green's functions: (i) the right-connected Green's function that takes care of everything to the right with coupling to the left set to zero and (ii) the left-connected Green's function that takes care of everything to the left with coupling to the right set to zero.

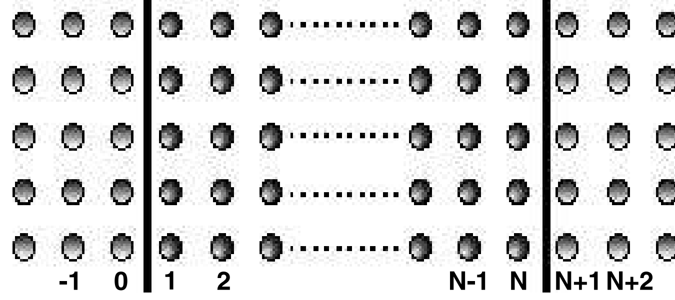



Figure 3.1: *Partitioning the structure into left contact, the device, and the right contact. The block indices* $\{-\infty, ...., -1, 0\}$ *lie on the left contact. The block indices* $\{1, 2, ....., N-1, N\}$ *belong to the device, and* $\{N+1, N+2, ....., \infty\}$ *lie on the right contact.*

### 3.1.1 RGF algorithm for calculating $G^R$

To derive the recursive algorithm for calculating the exact Green's function $G^R$, we define two retarded Green's functions: (i) the right-connected retarded Green's function $g^{\triangleright R}$ that takes into account everything on the right exactly with the coupling elements immediately to the left set to zero, (ii) the left connected retarded Green's function $g^{\triangleleft R}$ that takes into account everything exactly on the left with the coupling elements immediately on the right set to zero. The symbols $\triangleright$ and $\triangleleft$ stand for right-connected and left-connected, respectively. We write the Dyson's equation at the $i$-th block for the right connected Green's function,

$$g_{i,i}^{\triangleright R} = g_{i,i}^{0R} + g_{i,i}^{0R} t_{i,i+1} g_{i+1,i}^{\triangleright R}, \tag{3.3}$$

where $t$ is the coupling matrix from layer to layer, and $g_{i,i}^{0R} = \left[EI - D_{i,i} - U_{i,i} - \Sigma_{i,i}^{R}\right]^{-1}$ is the Green's function of the isolated block with $D$ being the layer Hamiltonian and $U$ being the potential energy matrix. $\Sigma^R$ is the self-energy. It can be energy dependent or independent. Using Dyson's equation, $g_{i+1,i}^{\triangleright R}$ can be written in terms of the

right-connected Green's function of the $(i+1)$-th block,

$$g^{\triangleright R}_{i+1,i} = g^{\triangleright R}_{i+1,i+1} t_{i+1,i} g^{\triangleright R}_{i,i}. \tag{3.4}$$

Substituting Eq. 3.4 in Eq. 3.3, we get the expression for the right-connected Green's function

$$g^{\triangleright R}_{i,i} = \left[EI - D_{i,i} - U_{i,i} - \Sigma^R_{i,i} - t_{i,i+1} g^{\triangleright R}_{i+1,i+1} t_{i+1,i}\right]^{-1}. \tag{3.5}$$

The calculation starts at the $N$-th block with $g^{\triangleright R}_{N+1,N+1}$ set to the right surface Green's function, $g^s_{N+1,N+1}$. Then we march from right to left across the device and generate the right-connected Green's function at each block. Following the similar foot steps, we can derive the expression for the left-connected Green's function

$$g^{\triangleleft R}_{i,i} = \left[EI - D_{i,i} - U_{i,i} - \Sigma^R_{i,i} - t_{i,i-1} g^{\triangleleft R}_{i-1,i-1} t_{i-1,i}\right]^{-1}. \tag{3.6}$$

Here the calculation starts at the 1st block by setting $g^{\triangleleft R}_{0,0}$ to the left surface Green's function. Then we walk from left to right across the device to generate the left-connected Green's function at each block.

We create the exact diagonal block $G^R_{1,1}$ from [106]

$$G^R_{1,1} = \left[EI - D_{1,1} - U_{1,1} - \Sigma^R_{1,1} - t_{1,0} g^{\triangleleft R}_{0,0} t_{0,1} - t_{1,2} g^{\triangleright R}_{2,2} t_{2,1}\right]^{-1}, \tag{3.7}$$

where $\Sigma^{\triangleleft RB}_{1,1} = t_{1,0} g^{\triangleleft R}_{0,0} t_{0,1}$ being the self-energy that includes the open boundary on the left of the device. $g^s_{0,0}$ ($g^{\triangleleft R}_{0,0} = g^s_{0,0}$) is the left surface Green's function. With $G^R_{1,1}$, we march back across the device creating $G^R_{i,i}$ using

$$G^R_{i,i} = g^{\triangleright R}_{i,i} + g^{\triangleright R}_{i,i} t_{i,i-1} G^R_{i-1,i-1} t_{i-1,i} g^{\triangleright R}_{i,i}, \tag{3.8}$$

which is derived by combining the Dyson's equation for $G^R_{i,i}$

$$G^R_{i,i} = g^{\triangleright R}_{i,i} + g^{\triangleright R}_{i,i}(-t_{i,i-1})G^R_{i-1,i}, \tag{3.9}$$

with the alternate form of the Dyson's equation for $G^R_{i-1,i}$

$$G^R_{i-1,i} = G^R_{i-1,i-1}(-t_{i-1,i})g^{\triangleright R}_{i,i}. \tag{3.10}$$

### 3.1.2 RGF algorithm for calculating $G^<$

Once $G^R$ is calculated recursively, we proceed to calculate $G^<$ using recursive Green's function algorithm. To derive the recursive algorithm for calculating the exact Green's function $G^<$, we define two more Green's functions: (i) the right-connected Green's function $g^{\triangleright <}$ representing the open boundary self-energy which takes into account the effect of the semi-infinite right contacts on the device, (ii) the left-connected Green's function $g^{\triangleleft <}$ representing the open boundary self-energy which takes into account the effect of the semi-infinite left contacts on the device

First, we note that for elastic incoherent scattering, $G^<$ can be factored into components injected from the left and the right contacts. This can be seen from Eqs. (73) - (76) of Ref. [106] or Eq. (13) of [124]. The empirical scattering model has a local, diagonal form for the self-energies [125],

$$\Sigma^R_{i,i} = -i\frac{\Gamma_{i,i}}{2}, \tag{3.11}$$

$$\Sigma^<_{i,i} = \frac{\Gamma_{i,i}}{A_{i,i}}G^<_{i,i}, \tag{3.12}$$

where $A_{i,i} = i(G^R_{i,i} - G^{R\dagger}_{i,i})$. The form for $\Sigma^<$ conserves current. The general form for

$G^<$ is

$$G^< = G^R \Sigma^< G^{R\dagger} + G^R \Sigma^{<B} G^{R\dagger}. \tag{3.13}$$

The first term on the right is from incoherent scattering, and the second term on the right is a source term due to injection from the contacts [106]. The starting point for the recursive Green's function algorithm for $G^<$ is the equation for the right-connected Green's function $g^{\triangleright <}$

$$g^{\triangleright <}_{i,i} = g^{\triangleright R}_{i,i} \left( \sigma^<_{i,i} + t_{i,i+1} g^{\triangleright <}_{i+1,i+1} t_{i+1,i} \right) g^{\triangleright R\dagger}_{i,i}, \tag{3.14}$$

where $\sigma^<_{i,i} = \frac{\Gamma_{i,i}}{A_{i,i}} G^<_{i,i}$. The calculation starts at the $N$-th block with $g^{\triangleright <}_{N+1,N+1} = i f^D a^{\triangleright}_{N+1,N+1}$ where $a^{\triangleright}_{N+1,N+1} = i(g^s_{N+1,N+1} - g^{s\dagger}_{N+1,N+1})$ and $f^D$ is the drain Fermi factor. Then we walk down the device from right to left and generate the right-connected Green's function $g^{\triangleright <}$ at each block. We follow the similar steps to generate the left-connected Green's function $g^{\triangleleft <}$

$$g^{\triangleleft <}_{i,i} = g^{\triangleleft R}_{i,i} \left( \sigma^<_{i,i} + t_{i,i-1} g^{\triangleleft <}_{i-1,i-1} t_{i-1,i} \right) g^{\triangleleft R\dagger}_{i,i}, \tag{3.15}$$

where the calculation starts at the 1st block by setting $g^{\triangleleft <}_{0,0} = i f^S a^{\triangleleft}_{0,0}$ where $a^{\triangleleft}_{0,0} = i(g^s_{0,0} - g^{s\dagger}_{0,0})$ and $f^S$ is the source Fermi factor. Then we march across the device from left to right to generate the left-connected Green's function $g^{\triangleleft <}$ at each block.

With the right-connected and left-connected Green's functions $g^{\triangleright <}$, $g^{\triangleleft <}$, the full Green's function $G^<_{1,1}$ can be calculated from [107]

$$G^<_{1,1} = G^R_{1,1} \left( \sigma^<_{1,1} + t_{1,0} g^{\triangleleft <}_{0,0} t_{0,1} + t_{1,2} g^{\triangleright <}_{2,2} t_{2,1} \right) G^{R\dagger}_{1,1}, \tag{3.16}$$

where $\Sigma^{\triangleleft <B}_{1,1} = t_{1,0} g^{\triangleleft <}_{0,0} t_{0,1}$ being the open boundary self-energy on the left of the

device. With $G^<_{1,1}$, we walk across the device creating $G^<_{i,i}$ using

$$G^<_{i,i} = g^{\triangleright<}_{i,i} + g^{\triangleright R}_{i,i} t_{i,i-1} G^R_{i-1,i-1} t_{i-1,i} g^{\triangleright<}_{i,i} + g^{\triangleright R}_{i,i} t_{i,i-1} G^<_{i-1,i-1} t_{i-1,i} g^{\triangleright R\dagger}_{i,i} + g^{\triangleright<}_{i,i} t_{i,i-1} G^{R\dagger}_{i-1,i-1} t_{i-1,i} g^{\triangleright R\dagger}_{i,i}. \tag{3.17}$$

The Eqs. 3.17 and 3.12 must be iterated until convergence is achieved, at which point the current is conserved.

### 3.1.3 Calculation of surface self energies

The self-energy calculation begins with the calculation of the surface Green function. A simple method for calculating the surface Green function is the linear iterative approach.

$$\begin{aligned} g^s_{0,0} &= \left[(EI + i\eta) - D_{0,0} - D_{0,-1} g^s_{0,0} D_{-1,0}\right]^{-1} \\ g^s_{N+1,N+1} &= \left[(EI + i\eta) - D_{N+1,N+1} - D_{N+1,N+2} g^s_{N+1,N+1} D_{N+2,N+1}\right]^{-1} \end{aligned}$$

The linear approach iteratively solves the above equations from an initial guess $g^s_{a,a} = [(EI + i\eta) - D_{a,a}]^{-1}$. The process is updated by taking the average of the old and new value of $g^s_{a,a}$. Here, $D_{a,a}$ is the unit cell Hamiltonian of the leads that includes the flat-band potential also, $D_{a,b}$ is the unit cell to unit cell coupling matrix in the leads, $\eta$ is an imaginary energy that is non-zero only in the leads, and a and b label the unit cell in the leads. Although this method is easy to implement, it is computationally expensive. Our calculations use the decimation technique [126, 127] to calculate the

surface Green function

$$\begin{aligned}
D^{m}_{a,b} &= D^{m-1}_{a,b}\left[H_{a,a}\right]^{-1}D^{m-1}_{a,b} \\
D^{m}_{b,a} &= D^{m-1}_{b,a}\left[H_{a,a}\right]^{-1}D^{m-1}_{b,a} \\
D^{m}_{a,a} &= D^{m-1}_{a,a}+D^{m-1}_{a,b}\left[H_{a,a}\right]^{-1}D^{m-1}_{b,a}+D^{m-1}_{b,a}\left[H_{a,a}\right]^{-1}D^{m-1}_{a,b} \\
Ds^{m}_{a,a} &= Ds^{m-1}_{a,a}+D^{m-1}_{a,b}\left[H_{a,a}\right]^{-1}D^{m-1}_{b,a}.
\end{aligned} \tag{3.18}$$

Here $H_{a,a}$ is defined as

$$H_{a,a}=(EI+i\eta)-D^{m-1}_{a,a}, \tag{3.19}$$

$m$ is the iteration number, and $Ds_{a,a}$ is the unit cell Hamiltonian that couples the leads to the device. Equations (3.18) define an iterative Hamiltonian for a chain of unit cells with lattice constant $2^m$a, where a is the zero-order lattice constant. The iterative sequence begins with an initial guess $D^0_{a,a}=Ds^0_{a,a}=D_{a,a}$, $D^0_{a,b}=D_{a,b}$, and $D^0_{b,a}=D_{b,a}$. Convergence is obtained when the maximum value of any element of the $D^m_{a,b}$ matrix is less than $10^{-10}$. A typical linear iterative approach takes about 120 iterations to converge while the decimation method converges in about 12 iterations. After convergence of Eqs. (3.18) at say iteration $n$, the surface Green function is

$$g^s_{a,a}=\left[EI-Ds^n_{a,a}\right]^{-1}. \tag{3.20}$$

Once the surface Green functions $g^s_{0,0}$ and $g^s_{N+1,N+1}$ are converged, the non-zero blocks of the self-energy matrices are calculated using

$$\begin{aligned}
\Sigma_{1,1} &= t_{1,0}g^s_{0,0}t_{0,1} \\
\Sigma_{N,N} &= t_{N,N+1}g^s_{N+1,N+1}t_{N+1,N}.
\end{aligned} \tag{3.21}$$

Since the coupling matrices cross the device/leads boundary and the imaginary potential $i\eta$ is non-zero only in the contacts, the energy dependency of the coupling matrices can be appropriately treated via the potential energy matrix.

## 3.2 Transport calculation

Once the retarded Green's function $G^R$ and the matrix of the electron correlation operator $G^<$ are determined, the next step is to self-consistently calculate charge and the potential. The potential is calculated using 3-D Poisson equations,

$$\nabla \cdot (\epsilon \nabla V) = -\rho \tag{3.22}$$

where $\epsilon = \epsilon_r \epsilon_0$ with $\epsilon_r$ being the relative permittivity and $\epsilon_0$ the free space permittivity. When the self-consistency between the charge and the potential is reached, we calculate the both the ballistic and the diffusive current. The methods are described in the next sections.

### 3.2.1 Ballistic transport

With continued shrinking of the channel length toward sub-10 nm dimension and introduction of high mobility channel materials [17, 71, 74], nanoscale channel lengths open up the possibility of near-ballistic MOSFET operation. Therefore, it is important to understand ballistic transport in both conventional and unconventional MOSFETs. The operation of MOSFETs in the ballistic regime has recently been explored by analytical models [128, 129] as well as by detailed numerical simulation [128, 130–132].

We have investigated ballistic transport characteristics of NW MOSFETs based on

III-V materials. Once the layer Hamiltonian and the layer-to-layer coupling matrices are calculated for the NW using the 3D discretized $\mathbf{k} \cdot \mathbf{p}$ formalism as described in Chapter 2, the NEGF calculation begins. The "device" Hamiltonian matrix $H_D$ is constructed from basis orbitals ($s$, $p_i$, with $i$=$x$, $y$, or $z$, including spin). The number of these orbitals determines the size of the matrix $[EI - H_D - \Sigma^\ell - \Sigma^r]$ which is the starting point of the NEGF calculations. At each energy $E$, the transmission is calculated from

$$T(E) \ = \ \mathrm{tr}\left\{\Gamma^B_{1,1} G^R_{1,N} \Gamma^B_{N,N} G^R_{1,N}\right\} \tag{3.23}$$

or

$$T(E) \ = \ \mathrm{tr}\left\{\Gamma^B_{1,1}\left[A_{1,1} \ - \ G^R_{1,1}\Gamma^B_{1,1}G^A_{1,1}\right]\right\} \tag{3.24}$$

where $\Gamma^B_{1,1} = i\left(\Sigma^\ell - \Sigma^{\ell\dagger}\right)$, $\Gamma^B_{N,N} = i\left(\Sigma^r - \Sigma^{r\dagger}\right)$, and $G^A = \left[G^R\right]^\dagger$. Expression (3.23) is the more commonly known expression for transmission and corresponds to what has become known as the Fisher-Lee [133] form of the transmission coefficient although the expression was written down 10 years earlier by Caroli et al. [134]. Eq. (3.24) is more numerically efficient since it only requires the calculation of the upper corner block of $G^R$.

To calculate current, a self-consistent approach is adopted. We calculate the surface Green function using Eq. (3.18) with $E$ replaced by $E - U_S$ for the source lead and $E - U_D$ for the drain lead where $U_S$ and $U_D$ are the source and drain potentials, respectively. The non-zero blocks of the self-energy matrices are calculated using the equations discussed above. The Green function and transmission are calculated using Eq. (3.23), respectively with $H_D$ replaced by $H_D + U$.

An adaptive energy grid approach is used in the transmission calculation to resolve resonance peaks. The energy window, $\mu_D - 40KT$ to $\mu_S + 40KT$, is linearly spaced where $KT$ is the thermal potential. Once a small energy window within a peak is

detected, the energy corresponding to the peak value of transmission is determined using an adaptive percentage deviation cutoff. Dense energy grids are taken around the peaks for energy integration of the current density

$$J(E) = T(E)\left[f_S(E-\mu_S) - f_D(E-\mu_D)\right]. \tag{3.25}$$

Here $f_S$ and $f_D$ are the source and drain Fermi functions, respectively. The current is calculated using the following equation

$$I = \frac{e}{\hbar}\int \frac{dE}{2\pi} J(E). \tag{3.26}$$

The approach captures the Stark effect, but not nonequilibrium self consistency.

### 3.2.2 Diffusive transport

Ballistic transport, where carrier scattering mechanisms are ignored, is an ideal case of transport in nanoscale devices. At room temperature, there is unavoidable carrier scattering due to phonons. Diffusive transport therefore needs to be considered in realistically large sized devices. The interactions of the free carriers with their environment (impurities, open surfaces, or lattice vibrations) strongly affect the performance of NW MOSFETs. For example, electron-phonon scattering is expected to drastically deteriorate the on-current of devices [103]. Modeling such many-body, inelastic interaction phenomena at a quantum mechanical level is computationally very challenging. The NEGF formalism [106, 125, 135–137] has established itself as a powerful approach to account for electron-phonon scattering in device simulations.

For a two-terminal device (source and drain), current conservation implies that the current that enters the simulation domain at layer $i$ is the same as the current

that leaves the structure at layer $i+1$. In block matrix notation for $t$ and $G^<$, the current expression is given as [106]

$$I_i = \frac{q}{\hbar}\int \frac{dE}{2\pi}\mathrm{Re}\left\{\mathrm{tr}\left[\mathrm{t_{i,i+1}G^<_{i+1,i}(E)}\right]\right\}, \tag{3.27}$$

where $q$ is the magnitude of the electron charge, $\hbar$ is the reduced Planck's constant, and tr{...} indicates a trace over all orbital states. To calculate the layer-by-layer current for the device, the off-diagonal blocks of $G^<$ are created using Langreth's rule [138] as follows,

$$G^<_{i+1,i} = g^{\rhd R}_{i+1,i+1}t_{i+1,i}G^<_{i,i} + g^{\rhd <}_{i+1,i+1}t_{i+1,i}G^{R\dagger}_{i,i}. \tag{3.28}$$

# Chapter 4

# Nanowire MOSFETs: High-speed, low-power applications

We have investigated NW MOSFETs based on InSb and InAs NWs as channel for high-speed and low-power applications. We calculate the energy dispersions for the NWs using 8-band $\mathbf{k}\cdot\mathbf{p}$ method that has been described in detail in Chapter 2. In this chapter, we discuss the calculated and simulation results of InSb and InAs NW FETs. We present the parameters such as bandgaps and lowest conduction band zone center electron effective masses that have been extracted from the calculated $E$-$k$ relations. These results are important for experimental works based on these materials.

We start this chapter with a brief introduction in Sec. 4.1. The model for the NW FETs is discussed in Sec. 4.2. The energy-wave vector relations for the NWs is presented in Sec. 4.3. We have also investigated the effects of the discretization length on the $E$-$k$ dispersions and the results are presented in Sec. 4.3.1. Sec. 4.4 presents the results on bandgap and effective masses. The calculated quantum capacitance is presented in Sec. 4.5. Finally, Sec. 4.6 presents the calculated current-voltage characteristics for InSb and InAs NW FETs following the summary in Sec. 4.7.

## 4.1 Introduction

Because of their high electron mobility at room temperature [139], and low contact resistance [140], InSb and InAs hold the promise to be the candidates for nanowire field effect transistors (NW FETs). Semiconductor nanowires (NWs) and carbon nanotubes (CNTs) are being considered as additions to the traditional Si technology to maintain the aggressive down scaling of the devices. Reports by Intel and Qinetiq on fabricated InSb quantum well field effect transistors (FETs) show that InSb-based quantum well FETs can achieve equivalent high performance with lower dynamic power dissipation and, therefore, should be promising device technology to complement scaled Si-based devices for very low power, ultra-high speed logic applications [72–74]. Although there are series of reports on the experimental realizations of InAs and InSb NWFETs [72–74, 141–145], there are few attempts to theoretically model them.

These materials have very small effective masses ($0.013m_0$ for bulk InSb, and $0.026m_0$ for bulk InAs, for example) which give high electron mobility (∼30,000 $cm^2V^{-1}s^{-1}$ for InSb and ∼20,000 $cm^2V^{-1}s^{-1}$ for InAs @ $n_s = 1 \times 10^{12}/cm^2$) at room temperature. However, in the ballistic limit, current drive is not determined by mobility, but by the density of states. The low effective mass which gives the high mobility becomes a liability in the ballistic limit, since it also gives a low density of states. Therefore, the current drive of InAs and InSb based NW FETs has been questioned. Furthermore, both of these materials have a small bandgaps which limit the voltage range of operation. In this work, we investigate the bandgap and current drive of these NW FETs using full band calculations of the electronic states and a ballistic FET model for the transport.

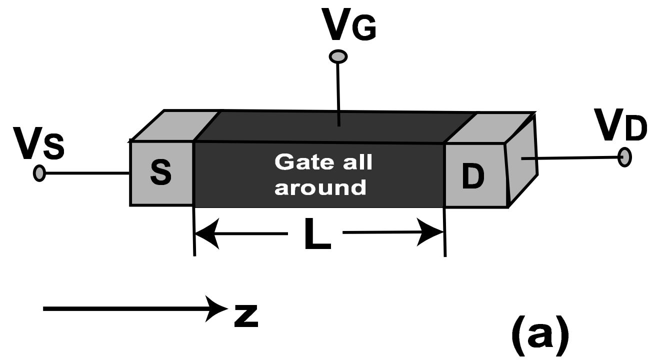
## 4.2 Model

A simple model for the ballistic gate-all-around core/shell NW MOSFET is shown in Fig. 4.1(a). Square [100] NWs are considered. The core region (also called the wire region) consists of the NW materials of interest. For our case, we have considered materials such as InSb, InAs, InP, and GaN. The shell region (also called the barrier region) consists of a high band-gap material (with respect to the core material) which helps to create one dimensional confinement in the core region. For our case, we consider InP as shell material for InSb and InAs, AlP as shell material for InP, and AlN as shell material for GaN. The parameters for such materials are presented in Chapter 2.

Fig. 4.1(b) shows the top-of-the barrier ballistic model that have been used to calculate the transport for the NW MOSFETs. The model consists of three capacitors, $C_G$, $C_S$, and $C_D$, which describe the electrostatic coupling between the top of the barrier and the gate, the source, and the drain, respectively.

## 4.3 Energy-wave vector characteristics

Fig. 4.2 shows the $E$-$k$ dispersion relations calculated using the 3D discretized 8-band $\mathbf{k} \cdot \mathbf{p}$ model for the two different cross-sections and materials. In each case, $E = 0$ corresponds to the lowest conduction band energy of the corresponding *bulk* InSb or InAs materials. We are pushing the validity of the 8-band $\mathbf{k} \cdot \mathbf{p}$ model for the 2 nm InSb NW. In InSb, the L-valley lies approximately 0.78 eV above the Γ-valley, and the 8-band $\mathbf{k} \cdot \mathbf{p}$ model does not include the effect of the L-valleys. Therefore, for InSb, the dispersions predicted by the 8-band $\mathbf{k} \cdot \mathbf{p}$ model cannot be trusted for energies above 0.8eV. The L-valleys of InAs, however, are approximately 1.15 eV



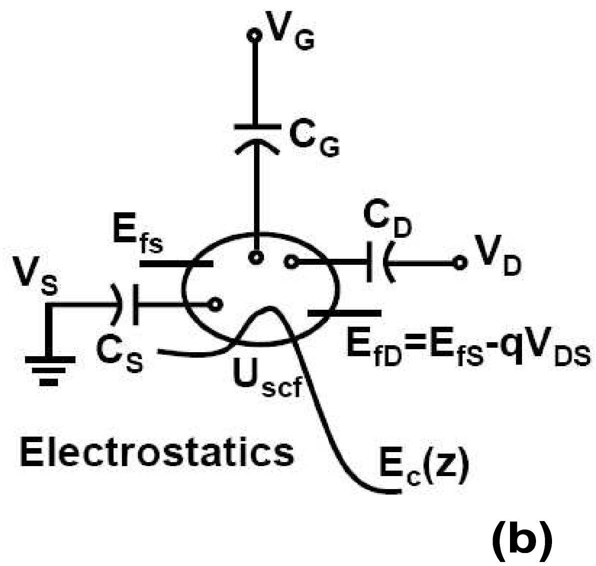




Figure 4.1: *(a) Schematic diagram of the simulated gate-all-around NW transistor, (b) The electrostatics of the seminumerical ballistic FET model used in the calculation. The diagrams are not to scale. Reproduced with permission from [1]. © 2008 IEEE.*

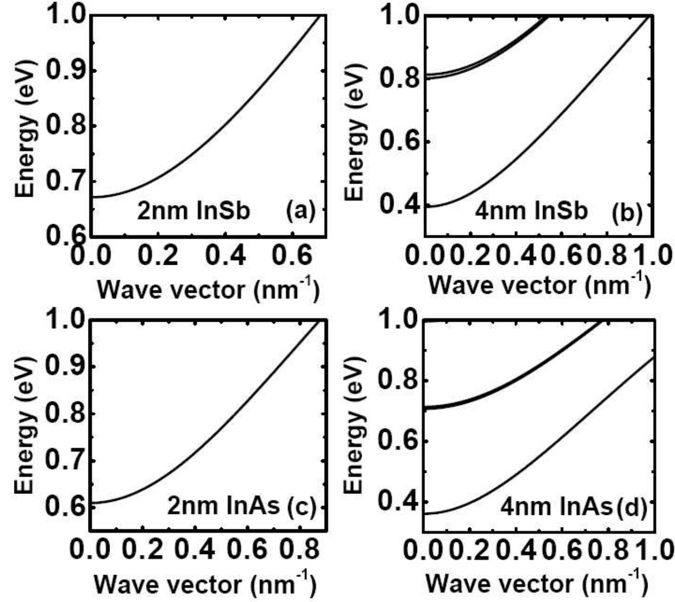



Figure 4.2: *E-k dispersion relationship calculated using the discretized 8-band* $\boldsymbol{k} \cdot \boldsymbol{p}$ *model for InSb NWs with 2 nm (a), and 4 nm (b) diameter and for InAs NWs with 2 nm (c), and 4 nm (d) diameter. Rectangular cross sections are considered. Reproduced with permission from [1]. © [2008] IEEE.*

above the Γ valley, so that our model is valid for InAs within the entire energy range under consideration. It is apparent from the $E - k$ plots that the dispersion quickly deviates from parabolic becoming significantly more linear above the energy of the mode minimum.

This increasing linearity is quantified by considering the effective mass of the fundamental mode as presented in the next section.

### 4.3.1 Effects of the discretization length

A correct calculation of electronic bandstructure is crucial for nanoscale device modeling as energy dispersion represents the material in quantum regime. In nanoscale, the energy dispersion of materials depends only on the material parameters and the geometry of the structure. As explained in Chapter 2, we use a 3-D discretized 8-band

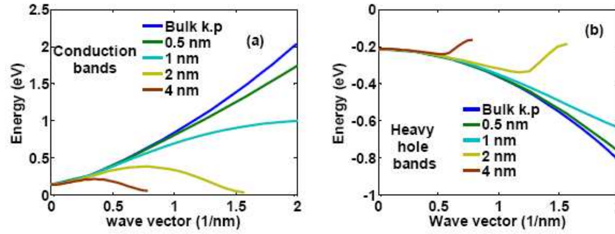



Figure 4.3: *Effects of the discretization length on the E-k dispersion of InAs for (a) the conduction band and (b) the heavy hole band.*

$\mathbf{k}\cdot\mathbf{p}$ method to calculate the electronic bandstructure of the NWs. The discretization scheme is presented in detail in Chapter 2, Sec. 2.2.3.

We use a central difference technique while discretizing the partial derivative as discussed in Sec. 2.2.3. We investigate the effects of the discretization length by calculating the bulk bandstructure with our 3-D discretized codes. Fig. 4.3 presents the calculation for InAs. The bulk dispersion is compared to the dispersions calculated from the discretized Hamiltonian for different discretization lengths as shown in Fig. 4.3. It has been found that a discretization length of 1 nm gives sufficient bandwidth.

After understanding the limitations imposed by the discretization length, we calculate the bandstructure for various III-V core-shell NWs using a 1 nm grid spacing for different diameter of the NW.

## 4.4 Bandgap and effective mass

The bandgap and effective masses for the lowest conduction band electrons are extracted from Fig. 4.2. The effective masses are calculated from the $E$-$k$ relation using $(1/m^*) = (1/\hbar^2)(\partial^2 E/\partial k^2)$. At the zone center, the effective mass is usually low for these materials. However, as we move away from the zone center, it increases rapidly as the bands become more linear as shown in Fig. 4.4.

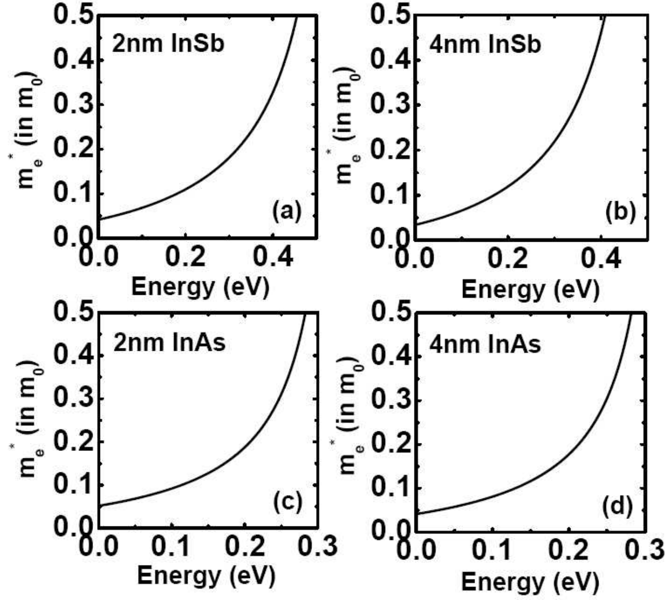



Figure 4.4: *Lowest conduction band electron effective masses vs. electron energy calculated from the E-k dispersion for InSb NWs with 2 nm (a), and 4 nm (b) diameter and for InAs NWs with 2 nm (c), and 4 nm (d) diameter. Reproduced with permission from [1]. © [2008] IEEE.*

Fig. 4.5 shows the NW diameter dependence of (a) the bandgap and (b) the lowest mode electron effective mass at the zone center. The bandgap for the InSb NW is 1.17 eV with 2 nm wire diameter and it falls to 0.37 eV with the 12 nm wire diameter. For InAs, the bandgap is 1.23 eV for 2 nm wire diameter and it falls to 0.54 eV for 12 nm wire diameter. The electron effective mass at the zone center for the InSb NW with core thickness of 2 nm is $0.042m_0$ and that with core thickness of 12 nm is $0.023m_0$. For InAs NW, the mass is $0.052m_0$ with 2 nm core thickness and $0.028m_0$ with core thickness of 12 nm.

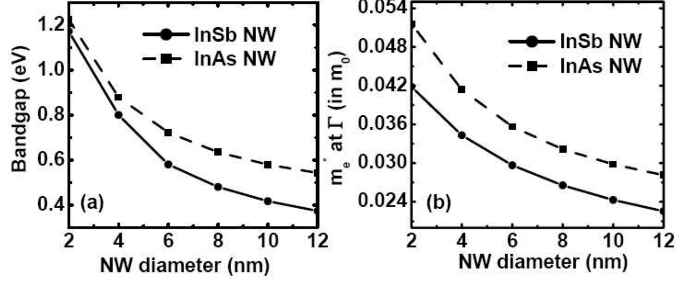



Figure 4.5: *(a) Bandgap as a function of wire diameter for the simulated InSb and InAs NWs, and (b) The wire diameter dependence of the lowest conduction band electron effective mass at the zone center. Reproduced with permission from [1]. © [2008] IEEE.*

## 4.5 Quantum capacitance

Quantum capacitance is defined as the thermally broadened density-of-states at the Fermi energy. For NWs, it is a result of the 1-D density-of-states. Mathematically, the quantum capacitance per unit length is defined as

$$C_Q = \frac{q^2 \partial n}{\partial E_F} \tag{4.1}$$

Ignoring thermal broadening, Eq. (4.1) becomes

$$C_Q = \frac{2}{\pi}\left(\frac{dE}{dk}\right)^{-1} = \frac{4q^2}{hv_F} \tag{4.2}$$

where $v_F = \frac{1}{\hbar}\frac{dE}{dk}$ evaluated at $E_F$ and spin degeneracy is assumed. After some calculations, it is shown that

$$C_Q = \frac{2q^2 L_G}{hv_{FS}} = \frac{q^2 L_G}{h}\sqrt{\frac{2m^*}{(E_{FS} - \varepsilon_1)}}. \tag{4.3}$$

Devices are said to operate in the quantum capacitance limit (QCL) when their quantum capacitance is much lower than their geometric gate capacitance. In the

QCL, nanoscale devices are found to offer improved device performance compared to the devices operating in the classical capacitance limit [1–3, 5, 146, 147]. It is to be noted from Eq. 4.3 that quantum capacitance can be reduced by reducing the effective mass. Materials such as InSb and InAs have ultra-small bulk electron effective mass, and therefore can be designed in nanoscale to operate in the QCL [1–3, 5].

Fig. 4.6 shows the quantum capacitance as a function of the Fermi level for the 4 NWs. The quantum capacitance is calculated from the expression, $C_Q = q^2(\partial n/\partial E_f)$, where $q$ is the electron charge and $n$ is the electron density. The geometrical gate capacitance, $C_G$, for the corresponding NW FET is also given in each figure. $C_G$ is calculated from $C_G = (2\pi\epsilon_r\epsilon_0)/(ln((2T_{ox} + T_{NW})/T_{NW}))$ assuming a coaxial gate geometry [148], where $T_{NW}$ is the diameter of the NW core and $T_{ox}$ is the thickness of the shell. It is clear from Fig. 6 that all four devices are operating in the quantum capacitance limit.

## 4.6 Current-voltage characteristics

Fig. 4.7 shows the $I_{DS}$ vs. $V_{DS}$ characteristics and Fig. 4.8 shows the $\log(I_{DS})$ vs. $V_{GS}$ transfer characteristics of the $n$-type InSb and InAs NW FETs. For each NW FET, the maximum $V_{DS}$ is chosen to be $E_G/(2q)$ and the maximum $V_{GS}$ is that which produces the flat band condition between the source and gate region. Because of the small quantum capacitance, the movement of the band-edge under the gate with gate bias is essentially linear under the entire gate bias swing.

The $I$-$V$ characteristics of $n$-type InSb and InAs NW FETs are evaluated by using a seminumerical ballistic FET model proposed by A. Rahman *et al.* [149] for ballistic planar MOSFETs and extended by J. Wang *et al.* [150] for ballistic high electron mobility transistors. The model is illustrated in Fig. 4.1(b). It consists of three

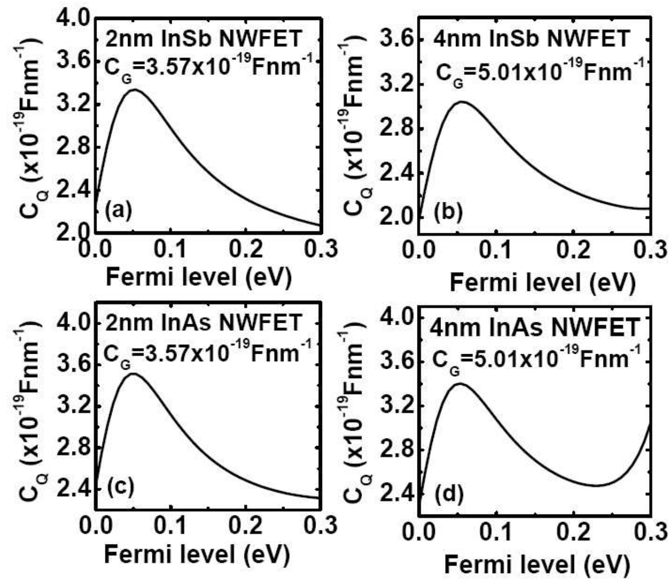



Figure 4.6: *Quantum capacitance as a function of the Fermi level for the simulated InSb NW FETs with two NW diameter, 2 nm (a), and 4 nm (b) and the same for the InAs NW FETs with two NW diameter, 2 nm (c), and 4 nm (d). Corresponding gate capacitances are also shown. All four devices are operating in quantum capacitance limit. Reproduced with permission from [1].* © *[2008] IEEE.*

capacitors, $C_G$, $C_S$, and $C_D$, which describe the electrostatic coupling between the top of the barrier and the gate, the source, and the drain, respectively. The potential at the top of the barrier is calculated as [149]

$$U_{scf} = \left[ C_G V_G + C_D V_D + C_S V_S + Q_{top} \right] / C_\Sigma \tag{4.4}$$

where $V_G$, $V_D$, and $V_S$ are the applied biases at the gate, drain, and source terminals, respectively, $C_\Sigma = C_G + C_D + C_S$, and $Q_{top}$ is the mobile charge at the top of the barrier. $Q_{top}$ is governed by the $U_{scf}$, the source and the drain Fermi levels, $E_{fS}$, $E_{fD}$ and the $E$-$k$ dispersion for the channel material. We calculate the gate control parameter, $\alpha_g = C_G/C_\Sigma$ and the drain control parameter, $\alpha_d = C_D/C_\Sigma$ numerically from our full-band simulations. After self-consistency between $U_{scf}$ and $Q_{top}$ is achieved, the drain current is calculated from [150]

$$I = M \frac{q k_B T}{2\pi\hbar} \left[ \Im_0(\eta_s) - \Im_0(\eta_s - U_D) \right], \tag{4.5}$$

where $M$ stands for spin degeneracy, $k_B$ is the Boltzmann constant, $T$ is the temperature. The function $\Im_i(x)$ is the Fermi-Dirac integral, $U_D = V_{DS}/(k_B T/q)$ and $\eta_s = (E_{fS} - \varepsilon(0) + qU_{scf})/k_B T$, where $E_{fS}$ is the source Fermi level and $\varepsilon(0)$ is the lowest conduction subband level at the top of the barrier.

In all cases, we choose a work function of the gate metal such that the difference, at zero bias, between the band-edge under the gate and the band-edge at the source is one half of the bandgap, $E_G/2$, which is a function of NW diameter and material. We will use the term 'band-edge' to mean the minimum energy of the fundamental conduction-band mode. A maximum value of $V_{DS} = E_G/(2q)$ was chosen to block the gate-drain interband tunneling leakage mechanism which is not included in the

simple transport model that we use. In this work, we have limited our investigation to NWs with a minimum bandgap of 0.8 eV, which under our worst-case biasing scheme, results in a minimum initial source-gate barrier of 0.4 eV.

To determine the maximum current, we set $V_{DS}$ to $E_G/(2q)$, and we apply sufficient gate voltage to drive the source-gate barrier to zero, i.e., we drive the band-edge under the gate down to the level of the band-edge in the source. Under these conditions, for a single moded, spin-degenerate wire, the second term in Eq. (4.5) is negligible, and the current is given by

$$I = \frac{2qk_BT}{h} \ln\left(1 + e^{E_F/k_BT}\right) \tag{4.6}$$

where the energy zero is the band-edge. For $e^{E_F/k_BT} >> 1$, Eq. (4.6) reduces to

$$I \approx \frac{2q}{h} E_F \tag{4.7}$$

which shows that the current is only a function of $E_F$ and independent of the NW material. The reason is that the group velocity exactly cancels the density of states making the current per unit energy independent of the dispersion.

There are only two ways to increase the current, increase the number of contributing modes and increase $E_F$. However, $E_F$ is the Fermi level of the source. To make full use of this $E_F$, one must be able to drive the band-edge of the gate down to that of the source. In the classical capacitance limit, this is not possible. When the gate region accumulates (or strongly inverts) the band-edge becomes relatively fixed when the density of states is high. In the quantum capacitance limit, however, even under accumulation or strong inversion, the band edge under the gate is still controllable with the gate voltage. To be in the quantum capacitance limit requires a low density

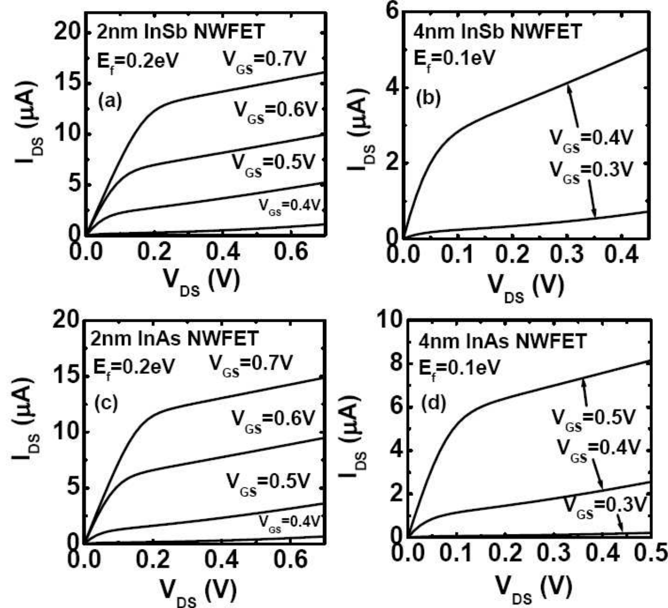



Figure 4.7: *$I_{DS}$ vs. $V_{DS}$ characteristics of the simulated InSb NW FETs with two NW diameter, 2 nm (a), 4 nm (b), and the same for the simulated InAs NW FETs with two NW diameter, 2 nm (c), 4 nm (d). The highest bias voltages have been taken such that flat band transport can be considered. Reproduced with permission from [1].* © *[2008] IEEE.*

of states which results from a low effective mass.

The low effective masses of InAs and InSb NWs have both a negative and positive effect on the maximum current drive. Low effective mass negatively affects the current by reducing the number of modes which lie below or within several $k_BT$ of $E_F$ that contribute to the current. In our case, the number of modes is 1 (with 2 spins), as shown in Fig. 4.9. Low effective mass positively affects the current by reducing the quantum capacitance well below the geometric gate capacitance so the NW FET operates in the quantum capacitance limit. In this limit, the gate voltage can drive the band-edge under the gate down to the level of the band-edge of the source to access the full $E_F$ of the source and maximize the current flowing in a single mode.

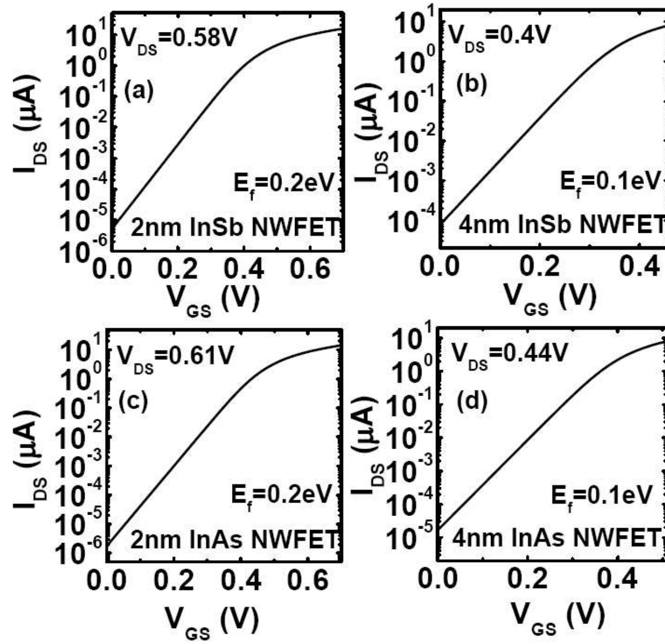



Figure 4.8: *log($I_{DS}$) vs. $V_{GS}$ transfer characteristics of the simulated InSb NW FETs with two NW diameter, 2 nm (a), 4 nm (b), and the same for the simulated InAs NW FETs with two NW diameter, 2 nm (c), 4 nm (d). The highest bias voltages have been taken such that flat band transport can be considered. Reproduced with permission from [1]. © [2008] IEEE.*

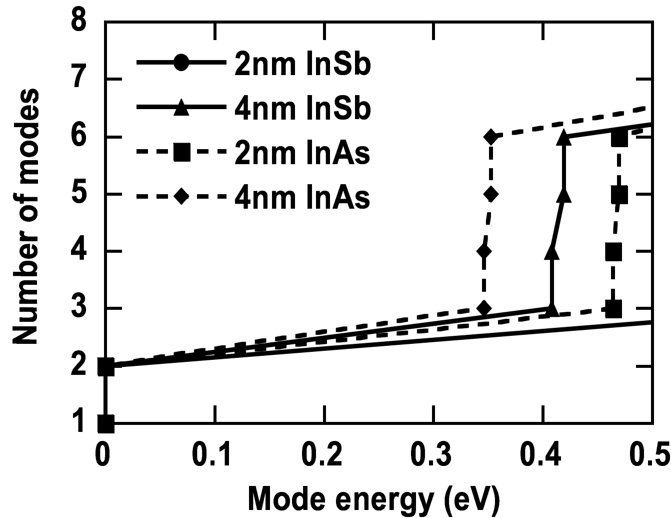



Figure 4.9: *Number of modes as a function of mode energy for the simulated InSb and InAs NWs. All four devices are single moded within the energy range considered. Reproduced with permission from [1]. © [2008] IEEE.*

## 4.7 Summary

The electronic structure of highly scaled InSb and InAs core-shell NW FETs has been modeled. The effects of the discretization length on the electronic structures are investigated. It is found that a 1 nm grid spacing gives a reasonable bandwidth. The important material parameters for the NWs, such as bandgap, zone center electron effective masses, etc, are extracted from the $E$-$k$ plots.

Structures with band gaps of 0.8 eV or more limit the NW core cross-sections to 4nm or less. In this case, all devices are single moded and operate in the quantum capacitance limit. Furthermore, when the NW FET source and gate region are degenerate such that $E_F$ is several $k_BT$ above the band edge, the NW FETs have the following properties. The transconductance is independent of the drain current and proportional to the conductance quantum with values in the range of 55-60 $\mu$S. The intrinsic cut-off frequency is weakly dependent on the drain current scaling as $\sqrt{I_D}$ with values in the 4 - 7 THz range good for RF applications. The gate delay time depends strongly on the drive current scaling as $I_D^{-3/2}$ with values ranging from 25 fs to 132 fs which is competitive for digital applications.

# Chapter 5

# Nanowire MOSFETs: Device metrics for $n$- and $p$-type devices

For NW MOSFETs, we have derived and investigated the device metrics such as drive current, $I_D$, on-off current ratio, switching frequency, $f_T$, delay time, $\tau_D$, power-delay product, $P \cdot \tau_D$, and energy-delay product, $E \cdot \tau_D$. These metrics are sufficient to analyze and benchmark a device for high-speed and low-power logic applications. These matrics capture the four fundamental device parameters for logic applications, namely, speed, switching energy, scalability and off-state leakage.

We have identified two opposite operation regime for both $n$- and $p$-type NW MOSFETs, namely, quantum capacitance limit (QCL) and the classical capacitance limit (CCL). We have investigated detail device operation in these two operational regimes and found that devices operating in the QCL offer significantly improved performance compared to their counterparts which are operating in the CCL. We have calculated the device metrics for both $n$- and $p$-type NW MOSFETs in these two regimes and compare their performance based on their operational regimes. In this chapter, we will present numerical and analytical calculations on the device metrics,

introduce the concept of quantum capacitance (QC) in nanoscale NW MOSFETs, and compare and contrast device operations in QCL and CCL.

We start this chapter with a brief introduction in Sec. 5.1. In Sec. 5.2, we present the device model that has been considered to derive and investigate the device metrics. Sec. 5.3 defines the two operational regimes - QCL and CCL - for the NW MOSFETs. Sec. 5.4 presents the detailed numerical calculations to evaluate the device metrics for NW MOSFETs. In Sec. 5.5, we present the analytical expressions for the device metrics that we have derived and finally Sec. 5.6 presents the calculated results on the device metrics for both $n$- and $p$-type NW MOSFETs and compare the values from numerical and analytical calculations following the summary in Sec. 5.7.

## 5.1 Introduction

Semiconductor nanowires (NWs) are being considered as additions to the traditional Si technology to maintain the aggressive scaling of devices. InSb and InAs materials, in particular, have been the focus of research because of their (i) high electron mobility at room temperature [139], and (ii) low contact resistance [140]. Recent reports by Intel on fabricated both $n$- and $p$-type InSb quantum well field effect transistors (FETs) show that InSb-based quantum well FETs can achieve equivalent high performance with lower dynamic power dissipation [72–75]. There are series of reports on the experimental realizations of InSb and InAs NW field effect transistors (NW FETs) [72–75, 141–145, 151].

Quantum capacitance (QC) is an important concept in nanoscale devices, first introduced by Luryi [152]. QC in NW FET devices accounts for the fact that the gate field penetrates through the wire since it is not completely screened on the wire surface as is the case in an ideal macroscopic conductor. Devices are said to operate

in the quantum capacitance limit (QCL) when their QC is less than their classical capacitance (CC). In highly scaled nanotransistors based on NWs or nanotubes exhibiting 1-D transport, the QCL limit can be reached [3, 147, 149]. Significant performance improvement in terms of the power-delay product has been predicted in devices operating in the QCL [1, 3, 147].

In this work, we compare and contrast the operation of devices with identical geometries which operate in the two different capacitance limits, the QCL and the classical capacitance limit (CCL). $n$-type InSb and InAs NW FETs up to 50 nm diameters operate in the quantum capacitance limit (QCL) [3], and the corresponding $p$-type NW FETs operate in the classical capacitance limit (CCL). The $n$- and $p$-type devices have identical geometries and differ only in their carrier type. Thus, we can compare the performance of two NW FETs which differ only in their QC.

Comparing $n$-type NW FETs operating in the QCL and identical geometry $p$-type NW FETs operating in the CCL, we find that the drive currents at a fixed gate overdrive are well matched. This is surprising since the density-of-states of the $p$-type FET is much larger than the density-of-states of the $n$-type FET. However, this effect has been observed both theoretically and experimentally by others [151,153]. Despite the matched current drive, the $p$-type devices operating in the CCL have twice the delay times, twice the power-delay products, and 4-5 times the energy-delay products of the $n$-type devices operating in the QCL. These effects have nothing to do with mobility since all transport is assumed ballistic. The effects arise solely from the density-of-states and electrostatics. We will describe, analyze, and explain these effects in the discussion below.

## 5.2 Model

The model for these devices is described in detail in Chapter 4, section 4.1. To be specific, the model consists of three capacitors, $C_G$, $C_S$, and $C_D$, which describe the electrostatic coupling between the top of the barrier and the gate, the source, and the drain, respectively.

## 5.3 Quantum vs. classical capacitance regime

Quantum capacitance (QC) is an important concept in nanoscale devices, first introduced by Luryi [152]. QC in NW FET devices accounts for the fact that the gate field penetrates through the wire since it is not completely screened on the wire surface as is the case in an ideal macroscopic conductor. Devices are said to operate in the quantum capacitance limit (QCL) when their QC is less than their classical capacitance (CC). In highly scaled nanotransistors based on NWs or nanotubes exhibiting 1-D transport, the QCL limit can be reached [3, 147, 149]. Significant performance improvement in terms of the power-delay product has been predicted in devices operating in the QCL [1, 3, 147]. In the following sections, we will derive and discuss performance metrics of devices operating in the QCL and show how they outperform devices operating the CCL.

## 5.4 Numerical calculation

An 8-band $\mathbf{k}\cdot\mathbf{p}$ method as described by Gershoni et al. [115] is used to calculate the electronic bandstructures of the NWs. The details of the $\mathbf{k}\cdot\mathbf{p}$ method are described in Chapter 2. Eight basis functions are included namely, the spin-up and spin-down $s$ and $p$ atomic orbital-like states. The multiband effective mass equation is transformed

into eight coupled differential equations for the envelope function $F_n$

$$\sum_{n'=1}^{8} H_{nn'}(\mathbf{r},\nabla)F_{n'}(\mathbf{r}) = EF_n(\mathbf{r}). \tag{5.1}$$

$H$ is the multiband Hamiltonian matrix and $E$ is the eigenenergy. We use a finite difference method [154] with a constant grid spacing of 1 nm to discretize the Hamiltonian. Spurious solutions of the $\mathbf{k}\cdot\mathbf{p}$ equations are eliminated by methods described in Refs. [155, 156]. The material parameters for the binary semiconductors are taken from Ref. [109].

The $I$-$V$ characteristics of both $n$- and $p$-type InSb and InAs NW FETs are evaluated by using a semiclassical ballistic FET model described by A. Rahman *et al.* [149] for ballistic planar MOSFETs and extended by J. Wang *et al.* [150] for ballistic high electron mobility transistors. The model is illustrated in Fig. 4.1(b). The model is described in Sec. 5.2. The potential at the top of the barrier is calculated as [149]

$$U_{scf} = \left[C_G V_G + C_D V_D + C_S V_S + Q_{top}\right]/C_\Sigma, \tag{5.2}$$

where $V_G$, $V_D$, and $V_S$ are the applied biases at the gate, drain, and source terminals, respectively, $C_\Sigma = C_G + C_D + C_S$, and $Q_{top}$ is the mobile charge at the top of the barrier. $Q_{top}$ is governed by $U_{scf}$, the source and the drain Fermi levels, $E_{FS}$, $E_{FD}$, and the $E$-$k$ dispersion for the channel material. In terms of energy, Eq. (5.2) becomes

$$E_{CG} = -q\left[\alpha_G V_{GS} + \alpha_D V_{DS} + Q_{top}/C_\Sigma\right], \tag{5.3}$$

where $E_{CG}$ is the conduction band-edge under the gate, $q$ is the magnitude of the electron charge, $\alpha_G = \frac{C_G}{C_\Sigma}$, $\alpha_D = \frac{C_D}{C_\Sigma}$, and the source is at ground. $\alpha_G$ and $\alpha_D$ are referred to as the gate and drain control parameters, respectively. They are calculated nu-

merically from $\alpha_G = \frac{C_G}{C_\Sigma} = \left|\frac{\Delta E_{CG}}{\Delta q V_{GS}}\right|_{\Delta V_{DS}=0, \Delta N=0}$ and $\alpha_D = \frac{C_D}{C_\Sigma} = \left|\frac{\Delta E_{CG}}{\Delta q V_{DS}}\right|_{\Delta V_{GS}=0, \Delta N=0}$, with $\Delta N$ being the induced charge in the channel, using a three dimensional Laplace solver [157].

The charge under the gate, $Q_{top}$, for electrons, is calculated from

$$Q_{top} = -(q/2) \int_{\varepsilon_1(0)}^{\infty} dE N_{1D}(E - \varepsilon_1(0)) \left[ f(E - \eta) + f(E - \eta + U_D) \right], \tag{5.4}$$

where $U_D = qV_{DS}$, $\eta = E_{FS} - \varepsilon_1(0) + qU_{scf}$, $E_{FS}$ is the source Fermi level, and $\varepsilon_1(0)$ is the 1st subband level at the top of the barrier at zero gate and drain bias. The factor of '1/2' comes from the fact that under the gate in the ballistic limit, the left contact only fills the right moving states under the gate, and the right contact only fills the left moving states under the gate. The density-of-states corresponding to the left or right moving states is one half the full density of states. A similar expression can also be used to calculate $Q_{top}$ for holes. We use the Green's function method to calculate $N_{1D}$,

$$N_{1D}(E) = \frac{1}{a} \mathrm{tr} \left[ \frac{-1}{\pi} \mathrm{Im}\{G(E)\} \right], \tag{5.5}$$

where $a$ is the discretization length and tr{...} is the trace over all states within a single discretized layer along the transport direction (including spin).

The exact Green's function $G$ of the device is calculated as

$$G(E) = \left[ EI - H_D - \Sigma_L - \Sigma_R \right]^{-1}, \tag{5.6}$$

where $H_D$ is the device Hamiltonian, and $\Sigma_L = t_{1,0} g_{0,0}^L t_{0,1}$, and $\Sigma_R = t_{0,1} g_{N,N}^R t_{1,0}$ are the self-energies to the left and the right contact, respectively. $t_{1,0}$ and $t_{0,1}$ are the coupling matrices between the device and the left and the right contact, respectively. The $g$'s are the surface Green's functions calculated using a decimation technique

[126, 127, 158].

To evaluate the number of modes contributing to the drive current for the devices, we calculate transmission at each energy point $E$ [133],

$$T(E) \;=\; \mathrm{tr}\left\{\Gamma_L G \Gamma_R G^{\dagger}\right\}, \tag{5.7}$$

where $\Gamma_L = i(\Sigma_L - \Sigma_L^{\dagger})$ and $\Gamma_R = i(\Sigma_R - \Sigma_R^{\dagger})$.

After self-consistency between $U_{scf}$ and $Q_{top}$ is achieved, the drain current for the devices is calculated as [123]

$$I = \frac{q}{h}\int_0^{\infty} dE T(E)\left[f(E-\eta) - f(E-\eta+U_D)\right]. \tag{5.8}$$

To compare the performance of different NW FETs, we set the maximum gate bias voltages such that the gate overdrive ($V_{OD}$) for each device is fixed at 0.2 V, i.e., $(V_{GS} - V_T)_{\max}$=0.2 V, where $V_T$ is the threshold voltage for each device. $V_T$ is determined as follows. For each single-moded $n$-type device, $V_T$ is taken as the $V_{GS}$ when the conduction band-edge under the gate ($E_{CG}$) reaches the energy $E_{FS} + kT$. At this $V_{GS}$, the single-moded $n$-type devices have the same threshold current, $I_{th}$. For the $p$-type devices, $V_T$ is the gate voltage that produces this $I_{th}$. Thus, by definition, the n-type and p-type devices have the same current when $\mathrm{V}_{GS}$=$\mathrm{V}_T$.

The source Fermi level with respect to the source conduction band edge must be large enough such that the maximum current is not limited by source injection saturation at the maximum gate overdrive of 0.2 V. Actual values for the n-type and p-type source Fermi levels will be discussed in Sec. 5.6.2.

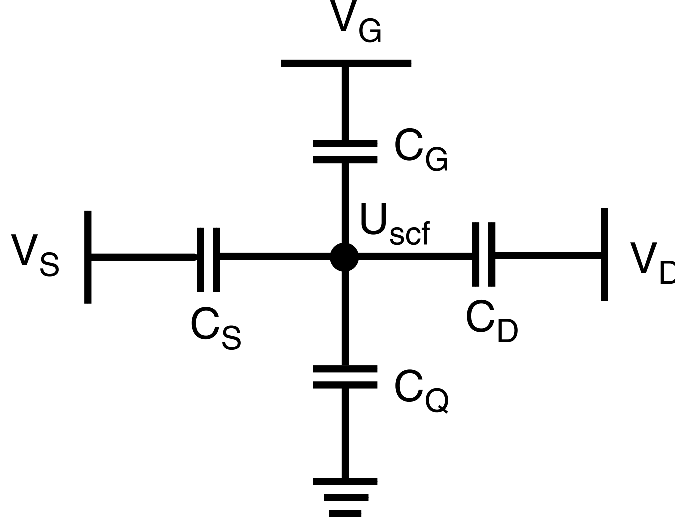



Figure 5.1: *Small signal equivalent circuit corresponding to Eq. (5.2). Reproduced with permission from [2].* © *[2009] IEEE.*

## 5.5 Analytical calculation

Several analytical expressions are now derived to provide insight into the effect of QC and the performance of single-moded devices deep in the QCL. The expressions related to the QC show how the QC affects the movement of the band-edge under the gate with gate voltage. To understand the performance metrics of NW FETs, analytical expressions are also derived for the current, the charge, the power-delay product, the energy-delay product, the gate delay time, and the cut-off frequency for a single-moded device operating in the QCL, ignoring thermal broadening, and assuming parabolic dispersion. The expressions for the the power-delay product, the energy-delay product, the gate delay time, and the cut-off frequency are fundamental limits for these devices.

In equilibrium, the QC per unit length is defined as $C_Q = \frac{q^2 \partial n}{\partial E_F}$, and under the

gate with large $V_{DS} >> k_B T$ applied

$$C_Q = \frac{q^2 \partial n}{\partial(E_{FS} - \varepsilon_1)}, \tag{5.9}$$

where $\varepsilon_1$ is the energy level of the fundamental mode including the effect of the self-consistent potential, i.e., $\varepsilon_1 = \varepsilon_1(V_{GS} = 0) - qU_{scf}$. The gate capacitance to ground, $C_{GS}$, is the series combination of the geometric ($C_G$) and quantum ($C_Q$) gate capacitance [159]. Thus, when $C_Q << C_G$, $C_{GS} \approx C_Q$ and the device is said to be in the QCL. Physically, the QC is the energy broadened density-of-states evaluated at the Fermi level.

The QC affects how the band-edge under the gate responds to the gate bias. Taking the derivative of Eq. (5.3), we obtain

$$\left| \frac{\partial E_{CG}}{\partial q V_{GS}} \right| = \frac{\alpha_G}{1 + \frac{C_Q}{C_\Sigma}}, \tag{5.10}$$

where we used the fact that the QC defined in Eq. (5.9) is also equal to $C_Q = q\frac{\partial Q_{top}}{\partial E_{CG}}$. Eq. (5.10) shows that in the QCL ($C_Q << C_G$), the gate control of the band-edge under the gate is unaffected by charge in the channel, i.e., there is no screening. $E_{CG}$ moves linearly with gate voltage in proportion to the gate control parameter $\alpha_G$. However, for $C_Q >> C_G$, the gate voltage has little effect on the band-edge under the gate, i.e., screening is significant.

Eq. (5.10) shows that Eq. (5.2) is consistent with the equivalent circuit model of the QC in series with the geometric gate capacitance shown in Fig. 5.1. In the circuit of Fig. 5.1, the change in $U_{scf}$ resulting from a change in gate voltage is $\Delta U_{scf} = \frac{\alpha_G}{1 + C_Q/C_\Sigma} \Delta V_G$ in agreement with Eq. (5.10).

We now derive expressions valid for a single-moded device operating in the QCL,

ignoring thermal broadening, and assuming parabolic dispersion. Under the gate, in the ballistic limit, ignoring thermal broadening, the QC per unit length is given by $C_Q = q^2\partial n/\partial(E_{FS} - \varepsilon_1) = q^2 N_{1D}^+(E_{FS})$ where $N_{1D}^+$ is the one dimensional (1D) density-of-states with positive velocity which is just the standard 1D density-of-states divided by 2. $N_{1D}^+$ can be written in several ways, two of which will be useful later,

$$N_{1D}^+(E_{FS}) = \frac{2}{hv_{FS}} = \frac{1}{h}\sqrt{\frac{2m^*}{(E_{FS} - \varepsilon_1)}}. \tag{5.11}$$

In Eq. (5.11), $v_{FS}$ is the velocity of an electron under the gate injected at the source Fermi level, $E_{FS}$. Therefore, the QC under the gate is,

$$C_Q = \frac{2q^2 L_G}{hv_{FS}} = \frac{q^2 L_G}{h}\sqrt{\frac{2m^*}{(E_{FS} - \varepsilon_1)}}. \tag{5.12}$$

The maximum drain current occurs when $V_{DS}$ is biased to its maximum value of $V_{DD}$, and the energy of the fundamental mode under the gate is well below the Fermi level of the source. Under these conditions, for a single-moded, spin-degenerate wire, the second term in Eq. (5.8) is negligible, and Eq. (5.8) becomes,

$$I = \frac{2qk_BT}{h}\ln\left(1 + e^{(E_{FS}-\varepsilon_1)/k_BT}\right). \tag{5.13}$$

When the energy of the fundamental mode, $\varepsilon_1$, is pulled several $k_BT$ below the Fermi level of the source, Eq. (5.13) reduces to

$$I_D \approx \frac{2q}{h}(E_{FS} - \varepsilon_1). \tag{5.14}$$

Thus, the current is proportional to the energy difference between the source Fermi level and the fundamental mode, and it is independent of any material parameters

such as the effective mass or density-of-states. To calculate other quantities such as the power-delay and energy-delay products, quantities such as the current will be needed as a function of voltage rather than energy. The energy scale is converted to a gate voltage using Eq. (5.10) to write

$$E_{FS} - \varepsilon_1 \; = \; q(V_{GS} - V_T)\,\frac{\alpha_G}{1 + \langle C_Q/C_\Sigma \rangle} \; = \; q\tilde{\alpha}_G(V_{GS} - V_T), \tag{5.15}$$

where we define the reduced gate control parameter $\tilde{\alpha}_G \doteq \frac{\alpha_G}{1+\langle C_Q/C_\Sigma \rangle}$ where $\langle C_Q/C_\Sigma \rangle$ is an average value. We will see that for the structures that we consider, the value of $\langle C_Q/C_\Sigma \rangle$ is approximately 0.1. Thus, the current as a function of voltage is

$$I_D \approx \frac{2q^2}{h}\tilde{\alpha}_G(V_{GS} - V_T), \tag{5.16}$$

with the understanding that $V_{GS} \geq V_T$ and $V_{DS} >> k_B T$ so that the device is in saturation, and there is no injection into the channel from the drain. The transconductance is defined as $g_m = dI_D/dV_{GS}$ which from Eq. (5.16) is

$$g_m = \frac{\tilde{\alpha}_G 2q^2}{h}, \tag{5.17}$$

i.e., $g_m$ is the quantum of conductance times the reduced gate control parameter.

The charge under the gate, $Q_{top}$ is

$$\begin{aligned} Q_{top} = \int_{\varepsilon_1}^{E_{FS}} dE\, N_{1D}^{+}(E) \; &= \; \frac{2q\sqrt{2m^*}L_G}{h}\sqrt{E_{FS} - \varepsilon_1} \\ &= \; \frac{2q\sqrt{2m^*}L_G}{h}\sqrt{q\tilde{\alpha}_G(V_{GS} - V_T)}. \end{aligned} \tag{5.18}$$

The power-delay product corresponds to the energy required to charge the gate. If one assumes that all of the charge on the gate is imaged by the charge in the channel,

one obtains a lower limit for this quantity given by [147]

$$P \cdot \tau_D = \int_0^{V_T+V_{OD}} dV_{GS} Q_{top}. \tag{5.19}$$

Substituting Eq. (5.18) into (5.19) and integrating gives

$$P \cdot \tau_D = \frac{4L_G\sqrt{2m^*}}{3h\tilde{\alpha}_G} \left(q\tilde{\alpha}_G V_{OD}\right)^{3/2}. \tag{5.20}$$

The gate delay time, $\tau_D$ is defined as [147]

$$\tau_D = \frac{P \cdot \tau_D}{V_{DD} I_D}. \tag{5.21}$$

Substituting Eqs. (5.20) and (5.16) into (5.21), $\tau_D$ is evaluated as,

$$\tau_D = \frac{2L_G\sqrt{2m^*}}{3q\tilde{\alpha}_G V_{DD}} \sqrt{q\tilde{\alpha}_G V_{OD}}. \tag{5.22}$$

Since we are using a lower limit for $P \cdot \tau_D$, Eq. (5.22) provides a lower limit for $\tau_D$.

The energy-delay is the product of the power-delay and the delay time. Multiplying Eqs. (5.20) and (5.22), $E \cdot \tau_D$ is

$$E \cdot \tau_D = \frac{16qm^* L_G^2}{9hV_{DD}} V_{OD}^2. \tag{5.23}$$

Again, this expression should be viewed as a lower limit for the energy-delay product.

The intrinsic cut-off frequency is calculated as $f_T = g_m/(2\pi C_{GS})$, where $C_{GS}$ is the series combination of the geometric ($C_G$) and quantum ($C_Q$) gate capacitance [159]. Assuming that $C_Q << C_G$ so that $C_{GS} \approx C_Q$, one obtains the upper limit of the

intrinsic cut-off frequency,

$$f_T \approx \tilde{\alpha}_G \frac{v_{FS}}{2\pi L_G} = \frac{\tilde{\alpha}_G}{2\pi\tau_G} = \frac{\tilde{\alpha}_G}{2\pi L_G}\sqrt{\frac{2(E_{FS}-\varepsilon_1)}{m^*}} = \frac{\tilde{\alpha}_G}{2\pi L_G}\sqrt{\frac{2q\tilde{\alpha}_G V_{OD}}{m^*}}. \tag{5.24}$$

In Eq. (5.24), $\tau_G = L_G/v_{FS}$ is the gate transit time. The cut-off frequency has the form of one over the transit time divided by $2\pi$.

Table 5.1 presents values of $I_D$, $P \cdot \tau_D$, $\tau_D$, $E \cdot \tau_D$, and $f_T$ for single-moded $n$-type NW FETs evaluated using the above expressions with material and device parameters corresponding to the n-type NW FETs numerically modeled in Sec. 5.2. The key concept to understanding the high intrinsic performance of single-moded NW FETs is the 'decoupling of the charge and the current.' This is apparent from Eqs. (5.16) and (5.18). The current in a a single-moded NW FET is material independent. It does not depend on the density-of-states or the charge in the channel. The charge in the channel depends on the density-of-states through the effective mass. For a given current, the charge could be anything depending on the value of the effective mass. To optimize performance, one wants maximum current with minimum charge. This is obtained by minimizing the effective mass. Minimizing the effective mass serves two purposes. (i) It minimizes the charge in the channel according to Eq. (5.18), and (ii) it allows one to achieve a higher Fermi level in the source for a given doping. This maximizes the current by maximizing the source Fermi level, $E_{FS}$ according to Eq. (5.14).

Table 5.1: *Figures of merit of the single-moded n-type InSb and InAs NW FETs from analytical calculations.* $L_G$*=10 nm,* $\alpha_G$*=0.8,* $m^*$*=0.034*$m_0$ *and 0.042*$m_0$*, respectively for InSb and InAs NW FETs [1],* $E_{FS}-\varepsilon_1$*=0.12 eV (Table 5.4).* $\tilde{\alpha_G}$*=0.73 and 0.71, respectively for InSb and InAs NW FETs. All values are calculated at maximum gate*

*overdrive. Reproduced with permission from [2]. © [2009] IEEE.*

| NW | $I_D$ ($\mu$A) | $f_T$ (THz) | $P \cdot \tau_D$ ($10^{-20}$J) | $\tau_D$ (fs) | $E \cdot \tau_D$ ($10^{-35}$Js) |
|---|---|---|---|---|---|
| InSb | 9.2 | 12 | 1.8 | 4.9 | 8.9 |
| InAs | 9.2 | 11 | 2.1 | 5.1 | 11 |

## 5.6 Results

We present the results of numerical and analytical calculations on the quantum and classical capacitance and the device metrics in this section.

### 5.6.1 Quantum and classical capacitance

Fig. 5.2 shows the $E$-$k$ dispersion relations calculated using the 3-D discretized 8-band $\mathbf{k} \cdot \mathbf{p}$ model for the two different materials. In each case, $E = 0$ corresponds to the lowest conduction band energy of the corresponding *bulk* InSb or InAs materials. It is apparent from the $E - k$ plots that the dispersion for the electrons quickly deviates from parabolic becoming significantly more linear above the energy of the mode minimum. Considerable band mixing in the excited hole subbands takes place for both the NWs.

Fig. 5.3 shows the transmission plots as a function of energy for the electrons and for the holes calculated from Eq. (5.7). There are two details worth commenting on for clarification. First, the initial turn-on of the transmission is 2 since the 8-band k·p model explicitly includes spin in the basis. The trace in Eq. (5.7) traces over all of the basis orbitals which include spin. Second, the non-monotonic behavior of

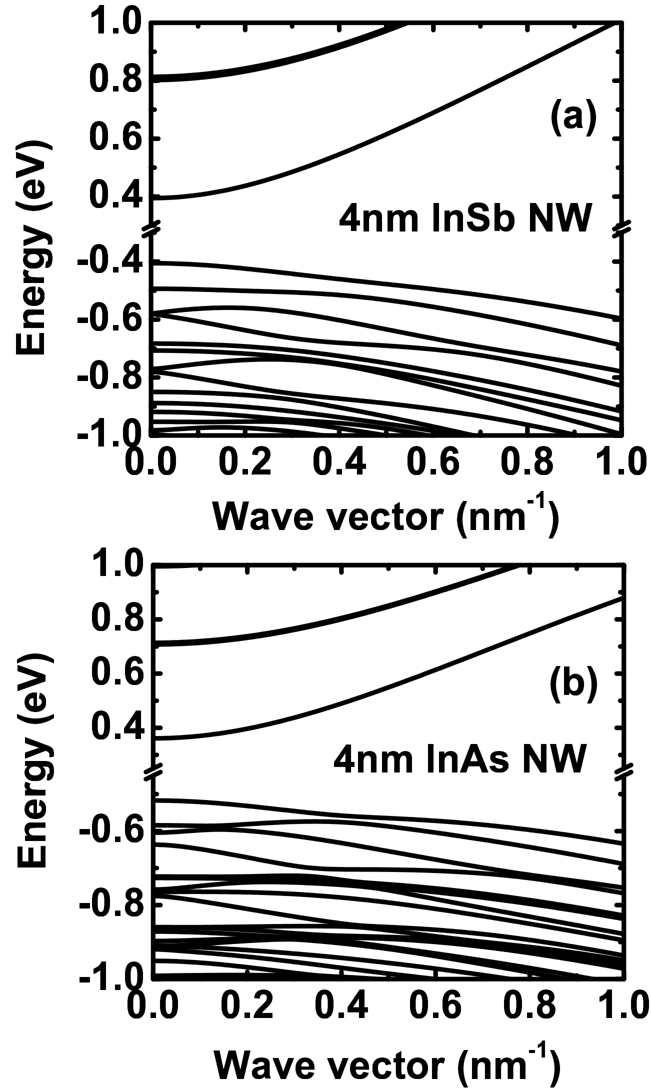



Figure 5.2: *E-k dispersion relationship calculated using the discretized 8-band* $\boldsymbol{k} \cdot \boldsymbol{p}$ *model for (a) InSb NWs, and (b) InAs NWs with 4 nm NW diameter.* $E = 0$ *corresponds to the lowest conduction band energy of the bulk InSb or InAs materials. Rectangular cross sections are considered. Reproduced with permission from [2].* © *[2009] IEEE.*

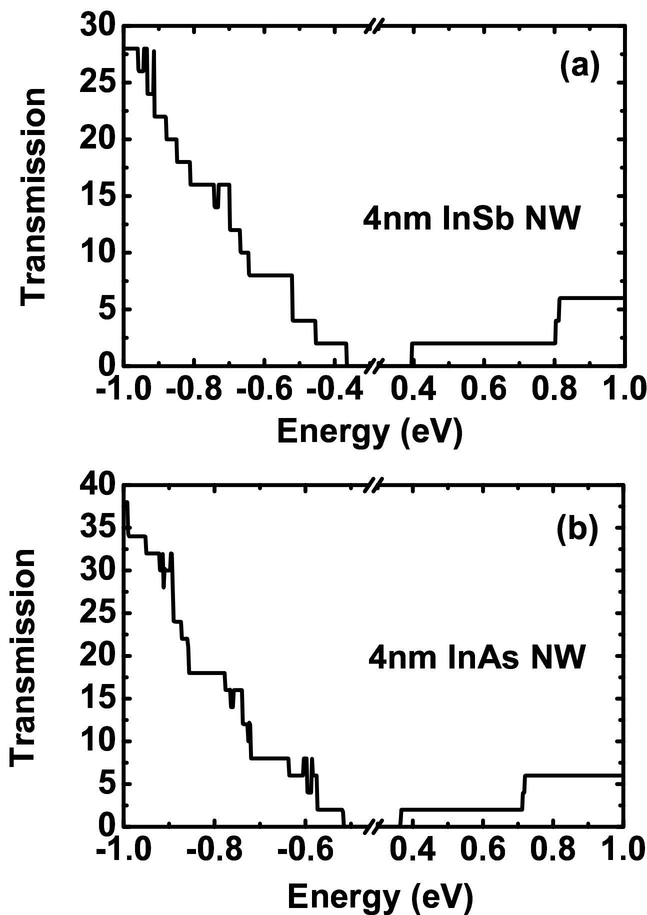



Figure 5.3: *Calculated transmission as a function of energy for the simulated 4 nm diameter (a) InSb, and (b) InAs NW FETs.* $E = 0$ *corresponds to the lowest conduction band energy of the bulk InSb or InAs materials. Reproduced with permission from [2]. © [2009] IEEE.*

the transmission for the p-type NWs is the result of the finite bandwidth and non-monotonic nature of the energy versus wavevector relations of the hole modes as seen in Fig. 5.2. Some modes have multiple regions of positive velocity, and as the energy moves down below a local extremum, a mode can turn off. The same effect has been observed in Si NWs [160].

Fig. 5.3 gives insight into the number of modes contributing to the drive current. For the n-type InSb and InAs NWs, the second set of modes occur at 0.35 eV and

0.4 eV above the fundamental mode, respectively. In all cases for the $n$-type devices, with a maximum gate overdrive of 0.2 V, the NW FETs are single-moded (2 spins) as shown in Fig. 5.3. For the $p$-type InSb and InAs NWs, the second set of modes occur at 57 meV and 89 meV below the fundamental mode, respectively. Thus, the p-type devices are not single-moded, and there is a contribution from the higher modes to the current, charge and, in particular, to the quantum capacitance that we discuss next.

Fig. 5.4 shows the quantum capacitance under the gate as a function of the source Fermi level for both $n$- and $p$-type devices. The quantum capacitance is calculated from Eq. (5.9) with the charge, $-qn = Q_{top}$ calculated from Eq. (5.4). The geometrical gate capacitance, $C_G$, for the corresponding NW FET is also shown in each figure. $C_G$ is calculated from $C_G = (2\pi\epsilon_r\epsilon_0)/(ln((2T_{ox} + T_{NW})/T_{NW}))$ assuming a coaxial gate geometry [148], where $T_{NW}$ is the diameter of the NW core and $T_{ox}$ is the thickness of the shell. The effect of discrete, well-spaced modes on the quantum capacitance is clearly depicted in Fig. 5.4. For all $n$-type devices, $C_Q << C_G$ for the entire energy range considered and the devices are operating in the QCL. For all $p$-type devices, however, $C_Q$ is comparable to or larger than $C_G$ and the devices are operating in the CCL.

Fig. 5.5 shows the $\log(I_{DS})$ vs. $V_{GS}$ transfer characteristics of both the $n$- and $p$-type InSb and InAs NW FETs calculated from Eq. (5.8). For the $p$-type devices, the polarity of $V_{GS} - V_T$ has been reversed. For all devices, $V_{DS}$ is chosen to be large enough so that there is no back injection from the drain. The currents are balanced for the $n$- and $p$-type devices. Maximum currents (at $V_{OD} = 0.2$ V) for each device are shown in Table 5.3. The $n$- and $p$-type device currents are well-matched. This is surprising since the $n$- and $p$-type devices have very different densities-of-states and they fall into different capacitance regimes. Naively, one would expect that, in the

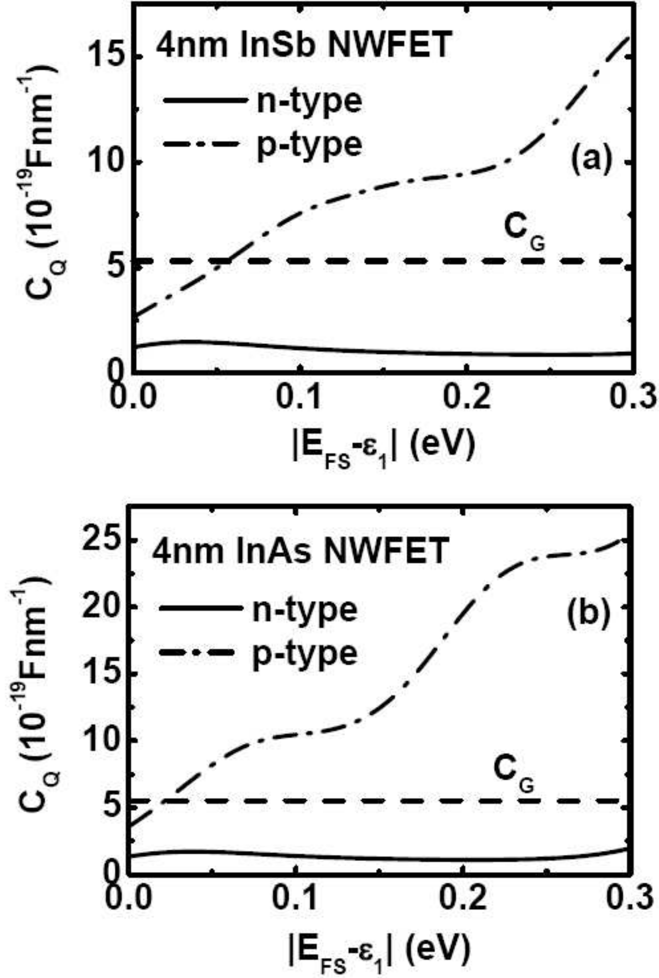



Figure 5.4: *Quantum capacitance under the gate as a function of* $|E_{FS} - \varepsilon_1|$ *where* $E_{FS}$ *is the source Fermi level and* $\varepsilon_1$ *is the energy of the fundamental mode under the gate for (a) InSb NW FETs, and (b) InAs NW FETs. Figures are presented for both n-type and p-type devices. Corresponding gate capacitances are also shown. Reproduced with permission from [2].* © *[2009] IEEE.*

ballistic limit, the device with the larger density-of-states and more closely spaced modes would carry more current.

Devices with higher density-of-states do contain more charge in the channel. Fig. 5.6 shows the carrier density, $Q_{top}$ vs. $V_{GS}$ for the InSb and InAs NW FETs. The carrier density is higher for the $p$-type devices. The $p$-type devices not only have more occupied modes, but the dispersion of each mode is considerably flatter than those in the conduction band (see Fig. 5.2). This results in a larger density-of-states associated with each valence band mode. The net result is more charge in the channel for the $p$-type devices for a given gate overdrive and current.

Fig. 5.7 shows how the band-edges under the gate change with applied gate bias. These curves are explained by Eq. (5.10). Initially, the channel in both the $n$- and $p$-type devices is empty, there is a significant source-channel barrier to electron or hole flow, and, thus, the quantum capacitance, $|\partial Q_{top}/\partial(E_{FS} - \varepsilon_1)|$, is exponentially reduced. Therefore, initially, the band-edges for both $n$- and $p$-type devices move identically with gate voltage with a slope whose magnitude is given by the gate control parameter $\alpha_G$.

When the first set of excited modes under the gate are shifted by the gate bias such that they start to become populated by the source Fermi level at $V_{GS} = V_T$, the quantum capacitance turns on and is as shown in Fig. (5.4). The single-moded, $n$-type devices remain in the QCL for the entire range of gate bias. At threshold, for the $n$-type devices, the magnitude of the slope $\left|\frac{\Delta E_{CG}}{\Delta q V_{GS}}\right|$ reduces slightly from $\alpha_G = 0.8$ to $\tilde{\alpha}_G \approx 0.7$. For the analytical results shown in Table 5.1, values of $\tilde{\alpha}_G$ were obtained by reading off the slopes of the curves in Figs. 5.7a,b in the regions $0 < V_{GS} - V_T < 0.2$.

When the first set of excited modes of the $p$-type devices are shifted by the gate bias such that they start to become populated by the source Fermi level, the quantum capacitance becomes larger than the classical capacitance and the $p$-type devices

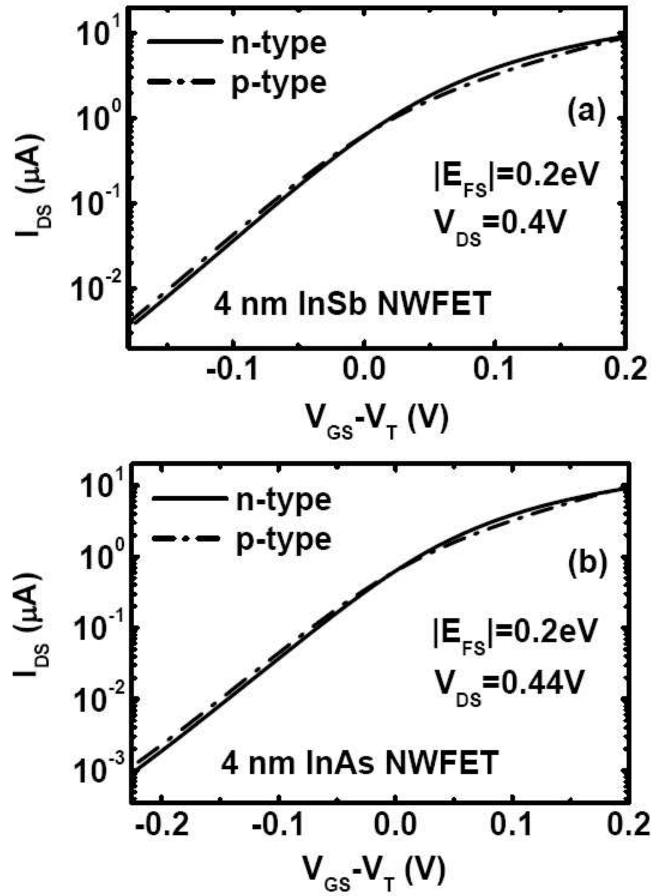



Figure 5.5: *$log(I_{DS})$ vs. $V_{GS}$ transfer characteristics of the simulated (a) InSb NW FETs, and (b) InAs NW FETs with 4 nm NW diameter for both n- and p-type devices. The drain bias voltages are fixed at half of the energy bandgap for each NW as shown. For the p-type devices, the polarity of the voltage is reversed. Reproduced with permission from [2]. © [2009] IEEE.*

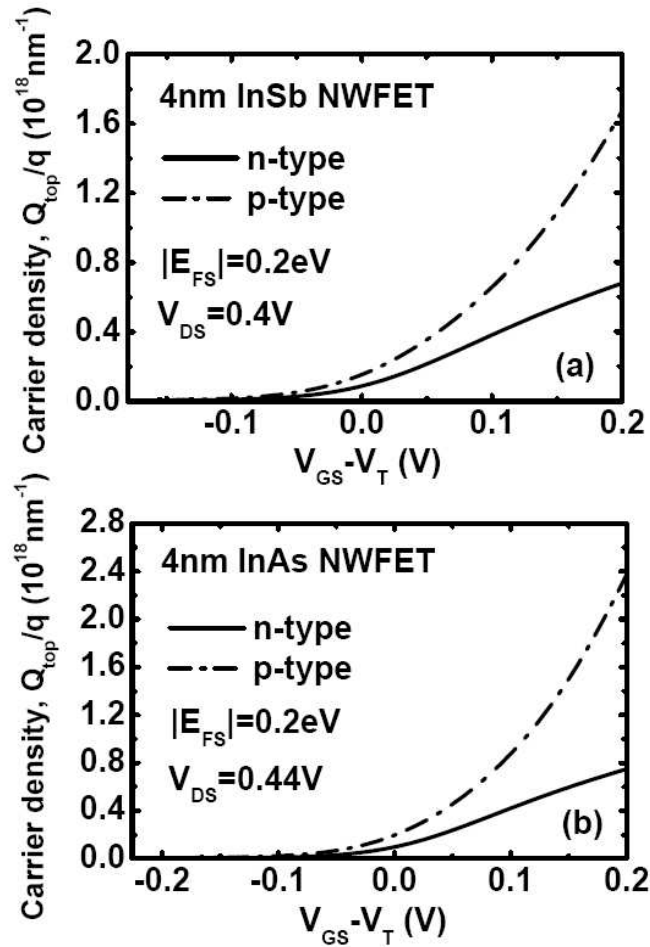



Figure 5.6: *Carrier density, $Q_{top}$ vs. $V_{GS}$ plots for the simulated (a) InSb NW FETs, and (b) InAs NW FETs with 4 nm NW diameter. For the p-type devices, the polarity of the voltage is reversed. Reproduced with permission from [2].* © *[2009] IEEE.*

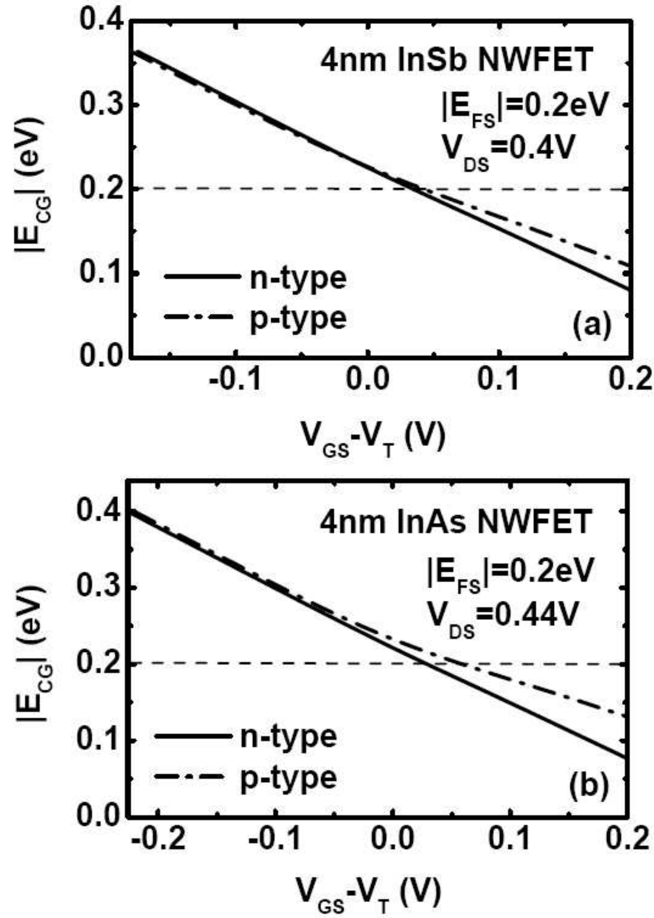



Figure 5.7: *Band-edge energies under the gate as a function of the gate bias for simulated (a) InSb NW FETs, and (b) InAs NW FETs. The source Fermi energy $E_{FS}$ is at 0.2 eV for all devices and is shown by the horizontal dashed line. The drain biases are fixed at half of the energy bandgap for each NW as shown. For the p-type devices, the polarity of the voltage is reversed. Reproduced with permission from [2].* © *[2009] IEEE.*

move into the CCL with $C_Q > C_\Sigma$ as shown in Fig. 5.4. When this happens, the denominator of Eq. (5.3) increases, and the magnitude of the slope $|\partial E_{CG}/\partial qV_{GS}|$ decreases as shown in Fig. 5.7. In the ideal classical limit with an infinite density-of-states, $C_Q$ becomes infinite, the slope $|\partial E_{CG}/\partial qV_{GS}|$ goes to zero, and the band-edge under the gate becomes fixed, independent of the gate voltage. This decrease in $|\partial E_{CG}/\partial qV_{GS}|$ resulting from charging of the channel is the expected behavior of devices operating in the CCL.

Consider an $n$-type device. The positive gate bias lowers the band-edge in the channel which results in charging of the channel. The charge results in a self-consistent potential, $U_{scf}$, which works against the gate bias to raise the band-edge. The larger the density-of-states, the larger the negative feedback will be. This negative feedback is absent, or much reduced, in the QCL. Therefore, the gate bias moves the band-edge less in the $p$-type devices than in the $n$-type devices, and the band-edge in the channel presents a larger barrier to the source in the $p$-type devices than in the $n$-type devices for the same gate overdrive. Therefore, even though the density-of-states and carrier density is higher in the $p$-type channel than in the $n$-type channel, the barrier to source injection is also higher. For a NW in the ballistic limit, ignoring thermal broadening, the current resulting from a single mode $m$ is $\frac{2q}{h}(E_{FS} - \varepsilon_m)$ where $\varepsilon_m$ is the mode energy in the channel. For a $p$-type channel, a larger $U_{scf}$ results in smaller energy differences $|(E_{FS} - \varepsilon_m)|$. Values for the energy differences $|(E_{FS} - \varepsilon_1)|$, electron densities, and hole densities at maximum gate overdrive are shown in Table 5.4. Thus, although there are more modes carrying current, each mode is carrying less current than the single mode in the $n$-type device. As a result, the currents of the $n$-type and $p$-type devices tend to be similar for the same gate overdrive as shown in Table **??**.

The negative feedback due to charging affects how far the bands move under the gate, which, in turn, determines how large the source Fermi level must be to avoid

source injection saturation. At the maximum gate overdrive of 0.2V, the conduction band edge under the gate is pushed 0.12 eV below the Fermi level of the source for both the InAs and InSb NW FETs. For the p-type devices with greater charging, the band edges move less. At the maximum gate overdrive, the InSb and InAs bandedges under the gate are pushed 91 meV and 69 meV above the Fermi level of the source, respectively. Therefore to ensure that the current is not limited by source injection saturation, the source Fermi level should be at least $\sim 0.15$ eV above the conduction band edge of the source for the 4nm n-type NW FETs and at least $\sim 0.1$ eV below the valence band edge of the source for the 4nm p-type NW FETs. The 10nm n-type FETs are still single moded, so the requirement of a source Fermi level $\geq \sim 0.15$ eV remains the same. The valence bands of the 10nm p-type NW FETs, with a higher density of modes, move less so that a source Fermi level $\geq\sim 50$ meV is sufficient to prevent source injection saturation.

### 5.6.2 Device metrics

Table 5.2 presents the calculated results of on-current, on-off current ratios, transconductance, intrinsic cut-off frequency, and delay time for both InSb and InAs NW MOSFETs with diameters of 2 nm and 4 nm for different source Fermi levels. The intrinsic cut-off frequencies in Table. 5.2 ranges from 4-7 THz for all devices.

Table 5.2: *Figures of merit of the proposed InSb and InAs NW FETs with a 10 nm long gate. The switching delay time and the intrinsic cut-off frequency are defined as $\tau_D = (C_{GS}V_{DD})/I_D$ and $f_T = g_m/(2\pi C_{GS})$ with $g_m$ being the transconductance are also listed. Reproduced with permission from [1]. © [2008] IEEE.*

| NW | D (nm) | $E_f$ (eV) | $I_{on}$ ($\mu$A) | $I_{on}/I_{off}$ ($\times 10^6$) | $g_m$ ($\mu$S) | $f_T$ (THz) | $\tau_D$ (fs) |
|---|---|---|---|---|---|---|---|
| InSb | | 0.1 | 8 | 71.7 | 56.0 | 5.5 | 118 |
| | 2 | 0.2 | 15 | 2.95 | 57.8 | 6.5 | 54.3 |
| | | 0.3 | 23 | 0.09 | 57.2 | 6.9 | 33.0 |
| | | 0.1 | 8.2 | 0.1 | 60.4 | 5.4 | 87.4 |
| | 4 | 0.2 | 15 | $4.2\times 10^{-3}$ | 59.7 | 6.1 | 41.2 |
| | | 0.3 | 23 | $1.4\times 10^{-4}$ | 59.2 | 6.4 | 25.6 |
| InAs | | 0.1 | 7.7 | 210 | 57.5 | 5.5 | 132 |
| | 2 | 0.2 | 15 | 8.83 | 59.4 | 6.4 | 60.0 |
| | | 0.3 | 23 | 0.28 | 58.9 | 6.7 | 37.2 |
| | | 0.1 | 8 | 0.47 | 58.5 | 4.9 | 105 |
| | 4 | 0.2 | 15 | 0.02 | 60.3 | 5.7 | 49.1 |
| | | 0.3 | 24 | $6.5\times 10^{-4}$ | 71.7 | 6.0 | 34.7 |

Table 5.3 summarizes the on-current, the power-delay product, $P \cdot \tau_D$, the delay time, $\tau_D$, the energy-delay product $E \cdot \tau_D$, and the intrinsic cut-off frequency, $f_T$, for both the InSb and InAs NW FETs with NW diameters of 4 nm and with a gate lengths of 10 nm. The numerical values of $P \cdot \tau_D$, $\tau_D$, $E \cdot \tau_D$, and $f_T$ for the single-moded $n$-type InSb and InAs NW FETs shown in Table 5.3 are predicted fairly well by the values from the analytical models shown in Table 5.1.

Table 5.3: *Figures of merit of the InSb and InAs NW FETs with NW diameters of 4 nm from numerical calculations. Gate length of 10 nm is considered. A source Fermi energy of $E_{FS}$=0.2eV is considered to calculate these values. All values are calculated*

*at maximum gate overdrive. Reproduced with permission from [2].* © *[2009] IEEE.*

| NW | $I_D$ ($\mu$A) | | $f_T$ (THz) | | $P \cdot \tau_D$ ($10^{-20}$J) | | $\tau_D$ (fs) | | $E \cdot \tau_D$ ($10^{-35}$Js) | |
|---|---|---|---|---|---|---|---|---|---|---|
| | $n$ | $p$ | $n$ | $p$ | $n$ | $p$ | $n$ | $p$ | $n$ | $p$ |
| InSb | 9.2 | 8.9 | 10 | 4.1 | 2.0 | 4.0 | 5.6 | 11 | 11 | 44 |
| InAs | 9.2 | 9.7 | 9.1 | 4.5 | 2.3 | 5.4 | 5.6 | 13 | 13 | 70 |

The $p$-type devices have twice the delay times, half the intrinsic cut-off frequencies, twice the power-delay products, and 4-5 times the energy-delay products of the $n$-type devices. Due to the higher density-of-states, the charge in the channel, for a fixed gate overdrive, is higher for the $p$-type devices compared to their $n$-type counterparts, (see Table 5.4), however, the currents are the same. This results in a higher delay time, power-delay and energy-delay products for the $p$-type devices compared to the $n$-type devices. These ratios and trends are identical to those of the 10 nm NW FETs as shown in Table 5.5.

Table 5.4: *Charge under the gate,* $|n - p|$*, gate source capacitance,* $C_{GS}$ *(aF/nm),* $E_{FS} - \varepsilon_1$ *at maximum gate overdrive for InSb and InAs NW FETs with NW diameters of 4nm and with a gate length of 10 nm. Threshold current,* $I_{th}$ *for each device is equal to 0.63* $\mu$*A. Transconductance for the devices are also shown. All values are calculated at the maximum gate overdrive. Reproduced with permission from [2].* © *[2009] IEEE.*

| NW | $g_m$ ($\mu$S) | | $\|n-p\|$ ($10^{-1}$nm$^{-1}$) | | $C_{GS}$ (aF·nm$^{-1}$) | | $\|E_{FS}-\varepsilon_1\|$ (eV) | |
|---|---|---|---|---|---|---|---|---|
| | $n$ | $p$ | $n$ | $p$ | $n$ | $p$ | $n$ | $p$ |
| InSb | 57 | 76 | 1.1 | 2.7 | 0.09 | 0.30 | 0.120 | 0.091 |
| InAs | 57 | 90 | 1.2 | 3.8 | 0.10 | 0.32 | 0.120 | 0.069 |

Table 5.5: *Figures of merit of 10 nm NW diameter InSb and InAs NW FETs from numerical calculations. Gate length of 10 nm is considered. A source Fermi energy of $E_{FS}$=0.2eV is taken to calculate these values. Threshold current, $I_{th}$ for each device is equal to 0.64 $\mu$A. All values are calculated at maximum gate overdrive. Reproduced with permission from [2].* *© [2009] IEEE.*

| NW | $I_D$ ($\mu$A) | | $\|n-p\|$ ($10^{-1}$nm$^{-1}$) | | $P\cdot\tau_D$ ($10^{-20}$J) | | $\tau_D$ (fs) | | $E\cdot\tau_D$ ($10^{-33}$Js) | |
|---|---|---|---|---|---|---|---|---|---|---|
| | $n$ | $p$ | $n$ | $p$ | $n$ | $p$ | $n$ | $p$ | $n$MOS | $p$MOS |
| InSb | 10 | 9.2 | 1.6 | 7.6 | 7.3 | 14 | 18 | 37 | 1.3 | 4.9 |
| InAs | 10 | 11 | 1.9 | 9.9 | 8.7 | 17 | 21 | 41 | 1.8 | 7.2 |

Finally, to place the performance metrics into context, we compare to the numbers predicted or projected for other materials and dimensionalities. For $n$-type InSb and InAs NW FETs with 4 nm NW diameter, the values of $P\cdot\tau_D$ are half the value of $5\times10^{-20}$ J, and the values of $P\cdot\tau_D$ with 10 nm NW diameter are close to the value of $5\times10^{-20}$ J predicted in Ref. [147] for a 3 nm diameter Si NW FET with a 10 nm gate length. The corresponding gate delay times for the $n$-type devices with 4 nm NW diameter are close to the value of 8 fs and the gate delay times with 10 nm NW

diameter are twice the value of 8 fs predicted in Ref. [147] for a 3 nm diameter Si NW FET with 10 nm long gate. The energy-delay product for $n$-type devices with 4 nm NW diameter is found to be 10 - 100 times lower than the projected experimental curve for a III-V planar $n$-channel MOSFET with a 10 nm channel width [74]. For $n$-type devices with 10 nm NW diameter, however, the energy-delay product falls on the projected experimental curve for a III-V planar $n$-channel MOSFET with a 10 nm channel width [74]. For $p$-type devices with 4 nm NW diameter, $E \cdot \tau_D$ is found to be 10 - 100 times lower than the projected experimental curve and with 10 nm NW diameter, $E \cdot \tau_D$ falls on the projected experimental curve for $p$-channel CNTFET and Si NW FET with 10 nm channel width [74].

### 5.6.3 Diameter dependent performance: InSb NW FETs

The diameter dependent performance of InSb NW FETs will be presented in this section. The diameter dependent performance metrics are investigated using an analytical two-band model (as described in Chapter 2) and a semiclassical ballistic transport model (as described in Chapter 4, section 4.1). The analysis of the diameter dependence of the current, the gate delay time, the power-delay product, and the energy-delay product of InSb NW FETs are presented and performance improvement in the QCL is identified. Device metrics are compared in devices operating in the QCL and CCL.

InSb is being considered as an attractive channel material for the next generation field effect transistors (FETs). Intel and Qinetiq report that InSb-based FETs can achieve equivalent high performance with lower dynamic power dissipation to complement scaled Si-based devices for very low power, ultra-high speed logic applications [1,2]. Although there are series of reports in the literature on the experimental real-

izations of NW FETs based on InSb material, there are few attempts to theoretically model them.

We have performed a theoretical study of the performance metrics of InSb nanowire (NW) FETs, and we present the first analysis of the diameter dependence of the current, gate delay, power-delay product, and energy-delay product of InSb NW FETs which operate in the quantum capacitance limit (QCL). Because of their small density of states, relatively large diameter ($\leq$ 60 nm) InSb NW FETs operate in the QCL. Both the energy-delay and power-delay products are reduced as the diameter is reduced, and optimum designs are obtained for diameters in the range of 10 - 40 nm.

We find, in agreement with J. Knoch et al. [3], a $\sim$ 40% lower $P \cdot \tau$ (power-delay product) for NW FETs operating in the QCL compared with devices operating in the classical capacitance limit (CCL). However, the ultra-small diameters (3 nm) necessary for Si NW FETs to operate in the QCL are not required for InSb. We also find reduced energy-delay products for NW FETs operating in the QCL. For the NW FETs with source Fermi levels such that $(E_F - E_c) = 0.2$ eV, the energy-delay product falls on the projected experimental curve for III-V planar HEMTs shown in Fig. 7 of Ref. [1]. Smaller source Fermi energies reduce the energy-delay product. The relationships between drain current, gate delay, power-delay product and energy-delay product are not obvious. The change in quantum capacitance with doping or diameter can override the change in current or transconductance giving counterintuitive trends which we elucidate below.

Each NW FET is composed of an InSb core and an InP shell which serves as the gate dielectric. The thickness of InP is assumed to be 6 nm (equivalent to 2 nm for $SiO_2$). The gate length is fixed at 10 nm. Unless otherwise stated, the NWs simulated in this work are all [001] oriented.

To calculate the mode energies and $E$-$k$ dispersions for various diameter (10 nm - 60 nm) NWs, we use an analytical 2-band dispersion relation [3] as described in Chapter 2. A semiclassical ballistic model, as described in chapter 4, is used to obtain the charge density, the self-consistent potential and the current. The drain bias voltage is fixed at 0.5 V for all devices. To compare the effect of different cross-sections and doping, we fix the gate overdrive to 0.2 V for all devices. The threshold voltages are determined from the linear $I_D$-$V_{GS}$ curves. For all diameters, two different Fermi levels in the source are modeled, $E_F - E_c = 0.1$ eV and 0.2 eV.

We consider NWs with diameters in the range of 10 – 60 nm. In each case, the core material contains InSb surrounded by a shell material of InP which is treated as the gate insulator. We first show the effect of confinement on the electronic properties of mode spacing, and quantum capacitance, and then we present calculations of the current, power-delay product, delay times, energy-delay product, and cutoff frequencies.

Fig. 5.8 shows the number of spin-degenerate modes as a function of the mode energy for 10 nm (a), and 60 nm (b) diameter NWs. We consider NW FETs with Fermi levels of up to 0.2 eV in the source. In all cases, the NW FETs contain multiple modes as shown in Fig. 5.8. We'll see the effects of these multiple modes on the quantum capacitance that we discuss next.

The quantum capacitance is $C_Q = \frac{e^2 \partial n}{\partial E_F}$ which is $e^2$ times the thermally broadened density of states at the Fermi level. Fig. 5.9 presents the quantum capacitance vs. source Fermi energy for the corresponding two NWs. The relationship between the number of populated modes and the quantum capacitance is obvious from Figs. 5.8 and 5.9. Figs. 5.8a and 5.9a show the effect of discrete, well-spaced modes on $C_Q$. The dashed lines in Figs. 5.9a - 5.9b show the geometrical classical gate capacitance, $C_G$. For $E_F - E_c = 0.1$ eV, all of the NWs with diameters $\leq 60$ nm have $C_Q < C_G$

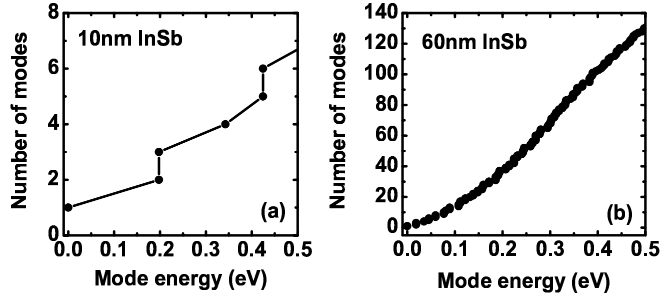



Figure 5.8: *Number of spin-degenerate conduction modes as a function of the mode energy for the simulated InSb NW FETs with 10 nm (a), and 60 nm (b) NW diameters. All the devices contain multiple modes within the energy range considered. Reproduced with permission from [3].* © *[2008] IEEE.*

and thus operate in the quantum capacitance limit.

As diameter increases, larger number of conduction modes become populated within a particular Fermi energy (Fig. 5.8), and the drain current increases. Fig. 5.10 shows the ON-current as a function of the NW diameter for two source Fermi energies, 0.1 eV (a), and 0.2 eV (b). The drain bias voltage is fixed at 0.5 V for each device, and the gate bias voltage is fixed at a gate overdrive of 0.2 V, $V_{GS} - V_T$=0.2 V, where $V_T$ is the threshold voltage and has been calculated from the linear $I_{DS}$-$V_{GS}$ characteristics for each device. The ON-current for all devices varies from 7 - 165 $\mu$A within a source Fermi energy variation of 0.1 - 0.2 eV (Table I).

The power-delay product increases and the gate delay decreases with an increase of the NW diameter as shown in Fig. 5.11. The power-delay product, $P \cdot \tau$, is calculated from $\int QdV_{GS}$, where $Q$ is the magnitude of the charge in the channel. The gate delay is obtained from $\tau = \int QdV_{GS}/(V_{DD}I_{ON})$. The gate delay time $\tau$ for all devices varies from 4 - 16 fs (Table I) within a source Fermi level range of 0.1-0.2 eV and decreases as the NW diameter increases (Figs. 5.11). This is due to the higher ON-current for the larger diameter NWs (Figs. 5.10). $P \cdot \tau$ varies from $2{\times}10^{-20}$ J to $68{\times}10^{-20}$ J for all devices (Table I) with a source Fermi level range of 0.1 - 0.2 eV.

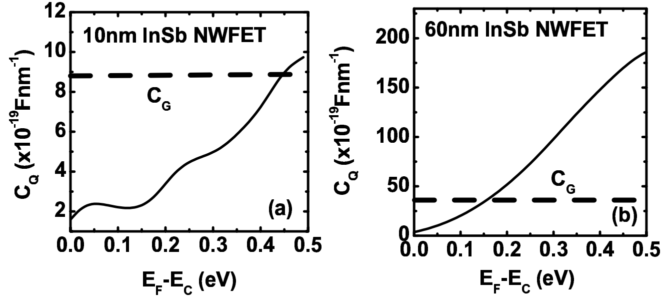



Figure 5.9: *Quantum capacitance as a function of the source Fermi level for the simulated InSb NW FETs with 10 nm (a), and 60 nm (b) NW diameters. The corresponding geometrical capacitance, $C_G$ is also shown. $C_G$ is calculated from $C_G = (2\pi\epsilon_r\epsilon_0)/(ln((2d_{ox} + d_{NW})/d_{NW}))$ assuming a coaxial gate geometry. All the devices are operating in the quantum capacitance limit within source Fermi level ($E_F - E_c = E_{FS}$) range of 0.1-0.2 eV. Reproduced with permission from [3]. © [2008] IEEE.*

For $E_F - E_c = 0.1$ eV, the values of $P \cdot \tau$ for diameters 10 - 50 nm all match closely the value of $5 \times 10^{-20}$ J predicted in Ref. [3] for a 3 nm diameter Si NW FET with a 10 nm gate.

Fig. 5.12 shows $P \cdot \tau$ as a function of the NW diameter with a source Fermi energy such that $E_F - E_c$ =0.3 eV. Inset shows the corresponding gate delay time $\tau$. The two regions QCL and CCL are indicated in the figure. We obtain a $\sim$ 40% lower $P \cdot \tau$ for NW FETs operating in the QCL compared with devices operating in the CCL, which is in agreement with a 3 nm diameter Si NW FET [3]. However, the ultra-small diameters (3 nm) necessary for Si NW FETs to operate in the QCL are not required for InSb. Although $C_Q$ is a function of gate voltage (see Fig. 5.9), one can understand the diameter dependence of $P \cdot \tau$ by approximating it as $C_Q V_{DD}^2/2$. The increase in $P \cdot \tau$ with diameter and Fermi energy results from the increase in $C_Q$.

Fig. 5.13 presents the energy-delay product as a function of NW diameter. As diameter and Fermi energy increase, the energy-delay product also increases. The energy-delay is the product of the power-delay and the delay time shown in Fig. 5.11.

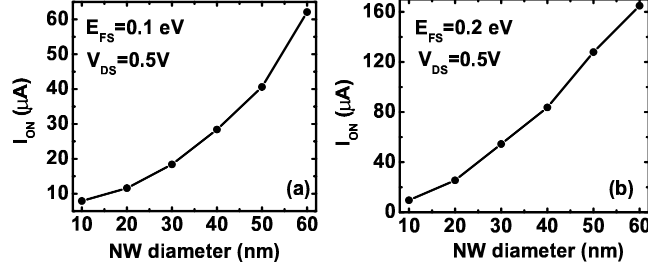



Figure 5.10: *ON-current as a function of NW diameter with source Fermi level ($E_{FS} = E_F - E_c$) of 0.1 eV (a), and 0.2 eV (b). The drain bias voltage is fixed at 0.5 V as shown. The gate voltage is fixed at a gate overdrive of 0.2V, i.e., $V_{GS} - V_T = 0.2$ where $V_T$ is the threshold voltage for each device. Reproduced with permission from [3].* © *[2008] IEEE.*

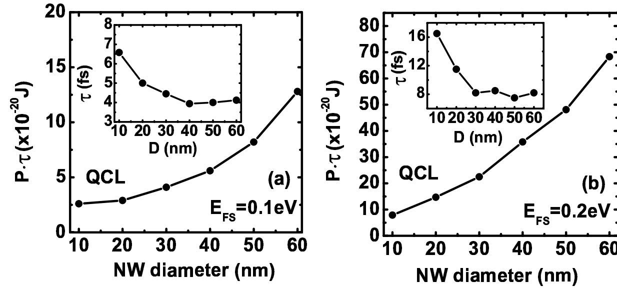



Figure 5.11: *Simulated $P \cdot \tau$ (power-delay product) as a function of NW diameter with source Fermi level ($E_{FS} = E_F - E_c$) of 0.1 eV (a), and 0.2 eV (b). $P \cdot \tau$ is calculated as $\int QdV_{GS}$, where Q is the total channel charge extracted from the simulations. Accordingly the gate delay time $\tau = \int QdV_{GS}/(V_{DD}I_{ON})$. Device operating regime, quantum capacitance limit (QCL), is indicated in the plots. These curves show a significant performance improvement in terms of the power-delay product for the devices scaled towards the quantum capacitance limit with $E_{FS}$=0.1 eV. Inset shows the gate delay time as a function of NW diameter extracted from the simulations. A gate length of 10 nm was considered to calculate these values. Reproduced with permission from [3].* © *[2008] IEEE.*

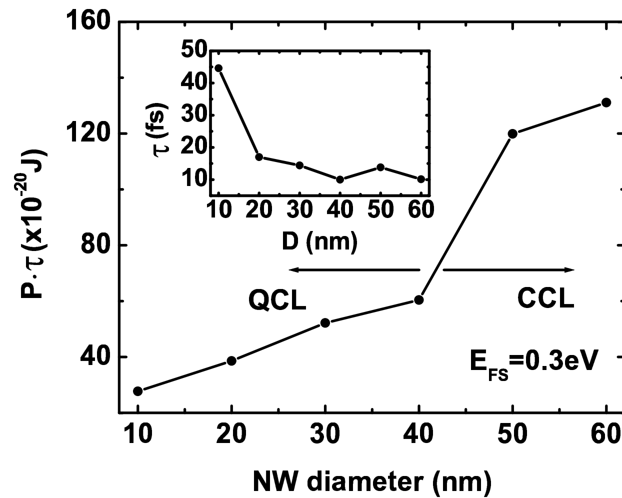



Figure 5.12: *Simulated $P \cdot \tau$ (power-delay product) as a function of NW diameter with source Fermi level ($E_F - E_c$) of 0.3 eV. $P \cdot \tau$ is calculated as $\int QdV_{GS}$, where $Q$ is the total channel charge extracted from the simulations. Accordingly the gate delay time $\tau = \int QdV_{GS}/(V_{DD}I_{ON})$. Device operating regimes, quantum capacitance limit (QCL), and classical capacitance limit (CCL) are indicated in the plots. These curves show a significant performance improvement in terms of the power-delay product for the devices scaled towards the quantum capacitance limit. Inset shows the gate delay time as a function of NW diameter extracted from the simulations. A gate length of 10 nm was considered to calculate these values. Reproduced with permission from [3].* ⓒ *[2008] IEEE.*

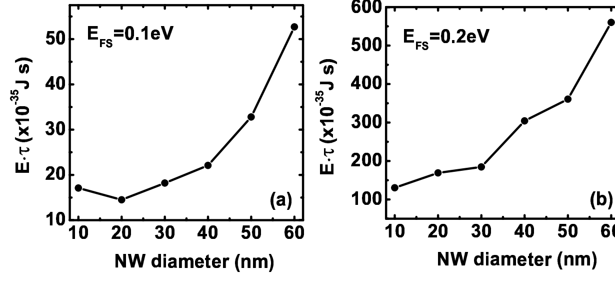



Figure 5.13: *Simulated $E \cdot \tau$ (energy-delay product) as a function of NW diameter with source Fermi level ($E_{FS} = E_F - E_c$) of 0.1 eV (a), and 0.2 eV (b). $E \cdot \tau$ is the product of the $P \cdot \tau$ and the gate delay $\tau$. These curves show a significant performance improvement in terms of the energy-delay product for the devices scaled towards the quantum capacitance limit with $E_{FS}$ =0.1 eV. Reproduced with permission from [3].* © *[2008] IEEE.*

It is interesting that the $1.4 \times 10^{-33}$ Js energy-delay product of the 10 nm NW FET with $E_F - E_c = 0.2$ eV falls on the projected experimental curve for a III-V planar HEMT with a 10 nm channel width shown in Fig. 7 of [1].

Both the energy-delay and power-delay products are optimal for diameters in the range of 10 - 40 nm with source Fermi levels of 0.1 eV. Increasing the source Fermi level increases $C_Q$ which causes a rapid degradation of the energy-delay product. Approximating $\tau$ as $\frac{C_Q V_{DD}}{I_D}$, the energy-delay product is proportional to $\frac{C_Q^2}{I_D}$. In this case, the increase in $C_Q$ overrides the increase in $I_D$.

Table I summarizes the ON-current, transconductance $g_m$, intrinsic cut-off frequency $f_T$, power-delay product, and gate delay time for the NW FETs with two source Fermi levels. For a particular source Fermi energy, the transconductance increases with an increase of the NW diameter due to the higher number of populated modes. The intrinsic cut-off frequency varies from 3 - 13 THz for all devices within a source Fermi level range of 0.1 - 0.2 eV, which is good for RF applications.

Table 5.6: *Figures of merit of the proposed InSb NW FETs with a 10 nm long gate.*

*The intrinsic cut-off frequency is defined as $f_T = g_m/(2\pi C_{GS})$ with $g_m$ being the transconductance. $C_{GS}$ is the series combination of the geometric ($C_G$) and the quantum ($C_Q$) capacitances. $I_{ON}$ from Figs. 5.10 are also listed. $P \cdot \tau$ is calculated as $\int Q dV_{GS}$, where $Q$ is the total channel charge extracted from the simulations. Accordingly the gate delay time $\tau = \int Q dV_{GS}/(V_{DD} I_{ON})$. Reproduced with permission from [3].* *© [2008] IEEE.*

| D (nm) | $E_{FS}$ (eV) | $I_{ON}$ ($\mu$A) | $g_m$ ($\mu$S) | $f_T$ (THz) | $P \cdot \tau$ ($10^{-20}$J) | $\tau$ (fs) |
|---|---|---|---|---|---|---|
| 10 | 0.1 | 7.9 | 57.6 | 5.2 | 2.6 | 7.0 |
| | 0.2 | 9.6 | 58.5 | 3.8 | 7.9 | 16 |
| 20 | 0.1 | 11.6 | 127.4 | 7.0 | 2.9 | 5.0 |
| | 0.2 | 25.5 | 205.3 | 6.5 | 14.7 | 11 |
| 30 | 0.1 | 18.4 | 221.6 | 7.2 | 4.1 | 4.0 |
| | 0.2 | 54.5 | 473.1 | 13 | 22.5 | 8.0 |
| 40 | 0.1 | 28.4 | 356.6 | 7.8 | 5.6 | 4.0 |
| | 0.2 | 83.7 | 634.1 | 8.6 | 35.8 | 8.0 |
| 50 | 0.1 | 40.6 | 519.3 | 7.9 | 8.2 | 4.0 |
| | 0.2 | 128 | 981.0 | 11 | 48.1 | 7.0 |
| 60 | 0.1 | 62.1 | 776.3 | 9.2 | 12.8 | 4.0 |
| | 0.2 | 165 | 1300 | 5.4 | 68.3 | 8.0 |

### 5.6.4 Diameter dependent performance: InAs NW FETs

The diameter dependent performance of InAs NW FETs will be presented in this section. The diameter dependent performance metrics are investigated using an analytical two-band model (as described in Chapter 2) and a semiclassical ballistic transport model (as described in Chapter 4, section 4.1). The analysis of the diameter dependence of the current, the gate delay time, the power-delay product, and the

energy-delay product of InAs NW FETs are presented and performance improvement in the QCL is identified. Device metrics are compared in devices operating in the QCL and CCL.

Semiconductor nanowires (NWs) are being considered as premier choice for next generation nanoelectronics. Due to its high electron mobility at room temperature [139] and low contact resistance [90], InAs is being investigated as high mobility channel material for high performance $n$MOS devices [1, 74, 161]. There are series of reports on the experimental realizations of InAs NW field effect transistors (FETs) [74, 141–145, 151, 162–164].

This material has very small effective mass of $0.026m_0$ for bulk InAs, which gives high electron mobility (∼20,000 cm$^2$V$^{-1}$s$^{-1}$ @ $n_s = 1 \times 10^{12}/cm^2$) at room temperature. The advantage that InAs material offers from its smaller effective mass is the high injection velocity under ballistic condition. However, smaller electron effective mass comes with a lower density-of-states which tends to reduce the drive current. Therefore, the current drive of InAs based NW FETs has been questioned.

We have performed a theoretical study of the performance metrics of InAs NW FETs, and we present the first analysis of the diameter dependence of the current, the gate delay time, the power-delay product, and the energy-delay product of InAs NW FETs which operate in the quantum capacitance limit (QCL). Devices are said to operate in QCL when their quantum capacitance (QC) is less than their classical capacitance (CC). In highly scaled nanotransistors based on NWs or nanotubes exhibiting 1-D transport, the QCL limit can be reached [1, 147, 149]. Significant performance improvement in terms of the power-delay product has been predicted in devices operating in the QCL [1, 147].

We find, in agreement with J. Knoch et al. [147], a $\sim 40\%$ lower $P \cdot \tau$ (power-delay product) for NW FETs operating in the QCL compared with devices operating

in the classical capacitance limit (CCL). However, the ultra-small diameters (3 nm) necessary for Si NW FETs to operate in the QCL are not required for InAs. We also find reduced energy-delay products for NW FETs operating in the QCL. For the NW FETs with source Fermi levels such that $E_{FS} = 0.2$ eV, the energy-delay product falls on the projected experimental curve for III-V planar HEMTs shown in Fig. 7 of Ref. [74]. Smaller source Fermi energies reduce the energy-delay product. The relationships between drain current, gate delay, power-delay product and energy-delay product are not obvious. The change in quantum capacitance with doping or diameter can override the change in current or transconductance giving counterintuitive trends which we elucidate below.

We consider NWs with diameters in the range of 10 – 60 nm. In each case, the core material contains InAs surrounded by a shell material of InP which is treated as the gate insulator. We first show the effect of confinement on the electronic properties of mode spacing, and quantum capacitance, and then we present calculations of the current, power-delay product, delay times, energy-delay product, and cutoff frequencies.

To calculate the mode energies and $E$-$k$ dispersions for various diameter (10 nm - 60 nm) NWs, we use an analytical 2-band dispersion relation [3] as described in Chapter 2. A semiclassical ballistic model, as described in chapter 4, is used to obtain the charge density, the self-consistent potential and the current. The drain bias voltage is fixed at 0.5 V for all devices. To compare the effect of different cross-sections and doping, we fix the gate overdrive to 0.2 V for all devices. The threshold voltages are determined from the linear $I_D$-$V_{GS}$ curves. For all diameters, two different Fermi levels in the source are modeled, $E_F - E_c = 0.1$ eV and 0.2 eV.

Fig. 5.14 shows the number of spin-degenerate modes as a function of the mode energy for 10 nm (a), and 60 nm (b) diameter NWs. We consider NW FETs with

Fermi levels of up to 0.2 eV in the source. Within this source Fermi energy, NW with diameter 10 nm is single-moded (with spin). We compare the lowest conduction band electron effective mass at zone center calculated from the analytical 2-band model with that calculated from an 8-band $\mathbf{k} \cdot \mathbf{p}$ model [1] for single-moded 10 nm diameter NW. The zone center electron effective mass calculated from the analytical model is $0.032m_0$ which is in good agreement with the value of $0.03m_0$ reported for the same NW diameter [1]. All other NW FETs contain multiple modes as shown in Fig. 5.14. We'll see the effects of these multiple modes on the quantum capacitance that we discuss next.

Fig. 5.15 presents the quantum capacitance under the gate as a function of the source Fermi energy for the corresponding two NWs. The quantum capacitance is calculated from Eq. (5.9) with the charge, $-qn = Q_{top}$ calculated from Eq. (5.4). $C_G$ is calculated from $C_G = (2\pi\epsilon_r\epsilon_0)/(ln((2T_{ox} + T_{NW})/T_{NW}))$ assuming a coaxial gate geometry [148], where $T_{NW}$ is the diameter of the NW core and $T_{ox}$ is the thickness of the shell. The dashed lines in Figs. 5.15a - 5.15b show the geometrical gate capacitance, $C_G$. The relationship between the number of populated modes and the quantum capacitance is obvious from Figs. 5.14 and 5.15. Figs. 5.14a and 5.15a show the effect of discrete, well-spaced modes on $C_Q$. For a source Fermi energy such that $E_{FS} = E_F - E_C = 0.1$ eV ($E_C$ is the lowest conduction band-edge), all of the NWs with diameters $\leq 60$ nm have $C_Q < C_G$ and thus operate in the quantum capacitance limit.

As diameter increases, larger number of conduction modes become populated within a particular Fermi energy (Fig. 5.14), and the drain current increases. Fig. 5.16 shows the ON-current as a function of the NW diameter for two source Fermi energies, 0.1 eV (a), and 0.2 eV (b). The drain bias voltage is fixed at 0.5 V for each device, and the gate bias voltage is fixed at a gate overdrive of 0.2 V, $V_{GS} - V_T$=0.2

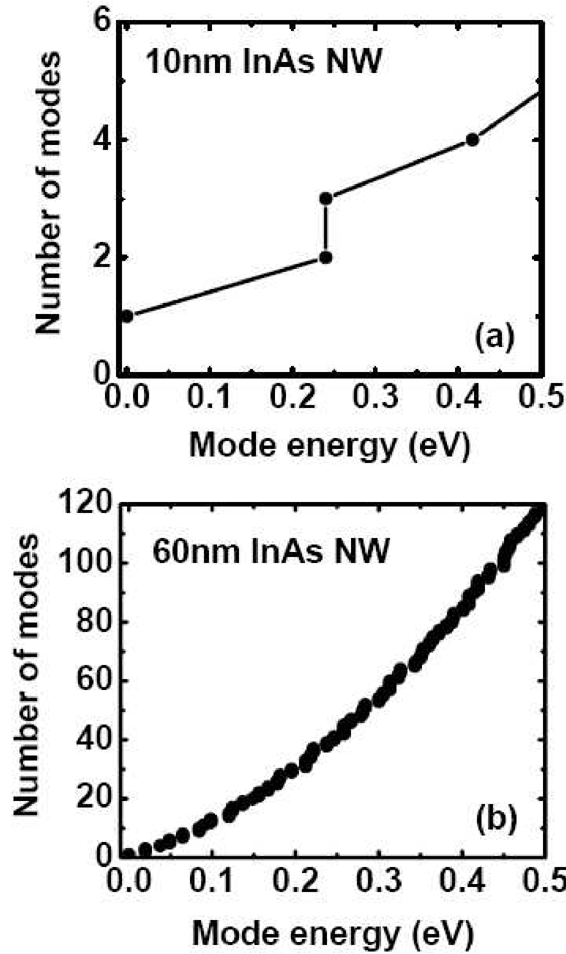



Figure 5.14: *Number of spin-degenerate conduction modes as a function of the mode energy for the simulated InAs NW FETs with 10 nm (a), and 60 nm (b) NW diameters. All the devices contain multiple modes within the energy range considered. Reproduced with permission from [5]. Copyright [2010], American Institute of Physics.*

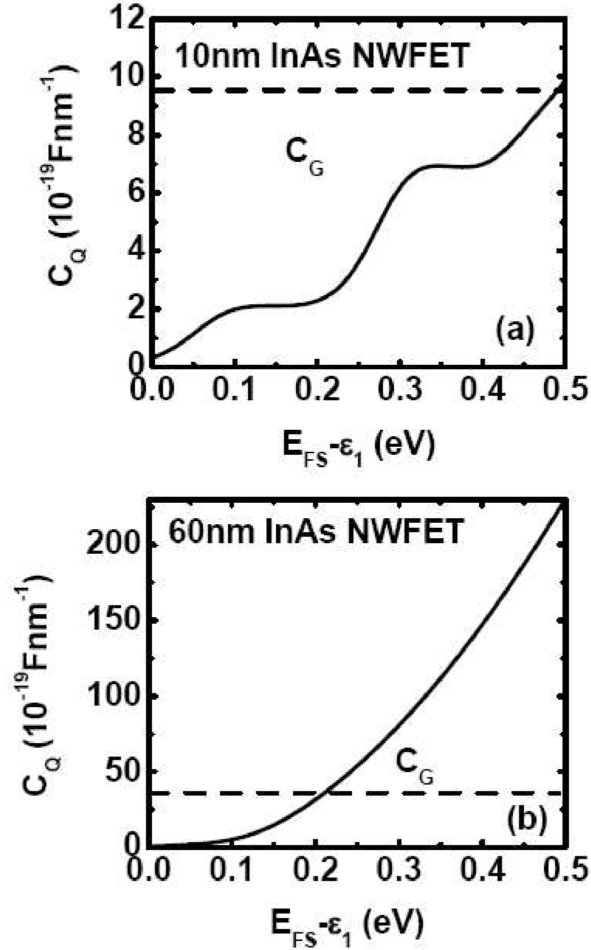



Figure 5.15: *Quantum capacitance under the gate as a function of $E_{FS} - \varepsilon_1$ where $E_{FS}$ is the source Fermi level and $\varepsilon_1$ is the energy of the fundamental mode under the gate for the simulated InAs NW FETs with 10 nm (a), and 60 nm (b) NW diameters. The corresponding geometrical capacitance, $C_G$ is also shown. $C_G$ is calculated from $C_G = (2\pi\epsilon_r\epsilon_0)/(ln((2d_{ox} + d_{NW})/d_{NW}))$ assuming a coaxial gate geometry. All the devices are operating in the quantum capacitance limit within source Fermi level range of 0.1 - 0.2 eV. Reproduced with permission from [5]. Copyright [2010], American Institute of Physics.*

V, where $V_T$ is the threshold voltage and has been calculated from the linear $I_{DS}$-$V_{GS}$ characteristics for each device. The ON-current for all devices varies from 7 - 195 $\mu$A within a source Fermi energy variation of 0.1 - 0.2 eV (Table 5.7).

The power-delay product increases and the gate delay decreases with an increase of the NW diameter as shown in Fig. 5.17. The power-delay product, $P \cdot \tau$, is calculated from $\int Q dV_{GS}$, where $Q$ is the magnitude of the charge in the channel. The gate delay is obtained from $\tau = \int Q dV_{GS}/(V_{DD} I_{ON})$. The gate delay time $\tau$ for all devices varies from 4 - 16 fs (Table 5.7) within a source Fermi level range of 0.1-0.2 eV and decreases as the NW diameter increases (Figs. 5.17). This is due to the higher ON-current for the larger diameter NWs (Figs. 5.16). $P \cdot \tau$ varies from $2{\times}10^{-20}$ J to $63{\times}10^{-20}$ J for all devices (Table 5.7) with a source Fermi level range of 0.1 - 0.2 eV. For $E_{FS} = 0.1$ eV, the values of $P \cdot \tau$ for diameters 10 - 40 nm all match closely the value of $5 \times 10^{-20}$ J predicted in Ref. [147] for a 3 nm diameter Si NW FET with a 10 nm gate.

Fig. 5.18 shows $P \cdot \tau$ as a function of the NW diameter with a source Fermi energy such that $E_{FS}$ =0.3 eV. Inset shows the corresponding gate delay time $\tau$. The two regions QCL and CCL are indicated in the figure. We obtain a $\sim$ 40% lower $P \cdot \tau$ for NW FETs operating in the QCL compared with devices operating in the CCL, which is in agreement with a 3 nm diameter Si NW FET [147]. However, the ultra-small diameters (3 nm) necessary for Si NW FETs to operate in the QCL are not required for InAs. Although $C_Q$ is a function of gate voltage (see Fig. 5.15), one can understand the diameter dependence of $P \cdot \tau$ by approximating it as $C_Q V_{DD}^2/2$. The increase in $P \cdot \tau$ with diameter and Fermi energy results from the increase in $C_Q$.

Fig. 5.19 presents the energy-delay product as a function of NW diameter. As diameter and Fermi energy increase, the energy-delay product also increases. The energy-delay is the product of the power-delay and the delay time shown in Fig. 5.17.

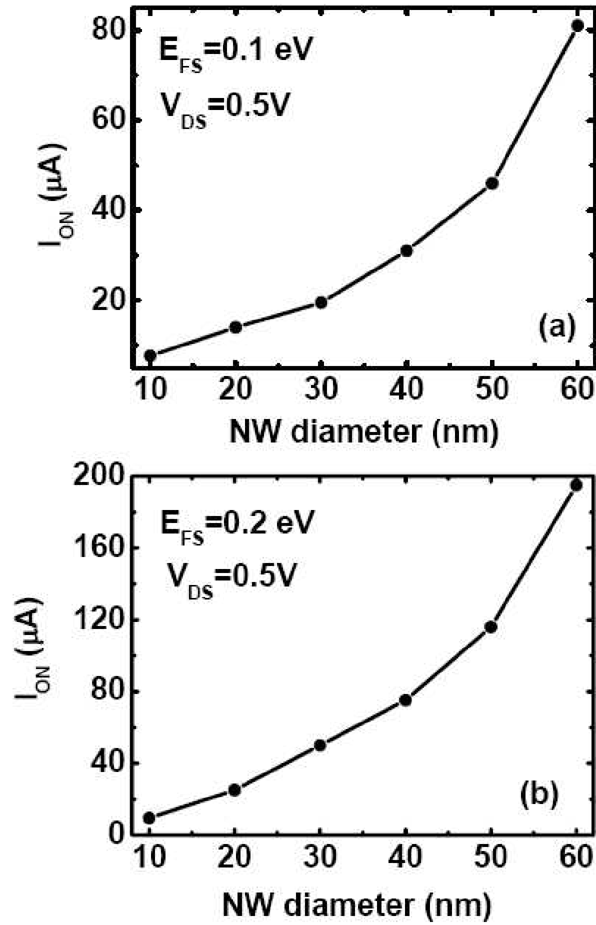



Figure 5.16: *ON-current as a function of NW diameter with source Fermi level of 0.1 eV (a), and 0.2 eV (b). The drain bias voltage is fixed at 0.5 V as shown. The gate voltage is fixed at a gate overdrive of 0.2V, i.e.,* $V_{GS} - V_T = 0.2$ *where* $V_T$ *is the threshold voltage for each device. Reproduced with permission from [5]. Copyright [2010], American Institute of Physics.*

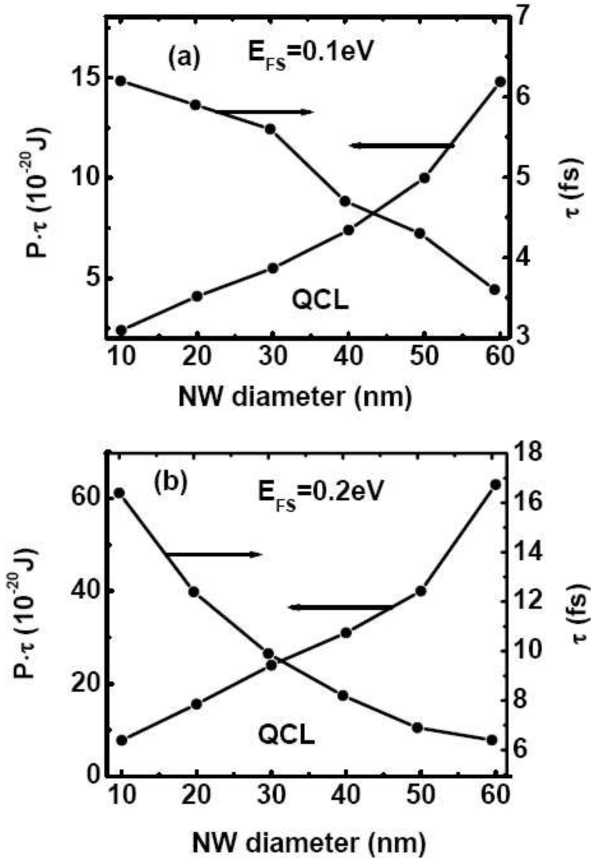



Figure 5.17: *Simulated power-delay product ($P \cdot \tau$) and the delay time ($\tau$) as a function of NW diameter with source Fermi level of 0.1 eV (a), and 0.2 eV (b). Device operating regime, quantum capacitance limit (QCL), is indicated in the plots. These curves show a significant performance improvement in terms of the power-delay product for the devices scaled towards the quantum capacitance limit with $E_{FS}$=0.1 eV. A gate length of 10 nm was considered to calculate these values. Reproduced with permission from [5]. Copyright [2010], American Institute of Physics.*

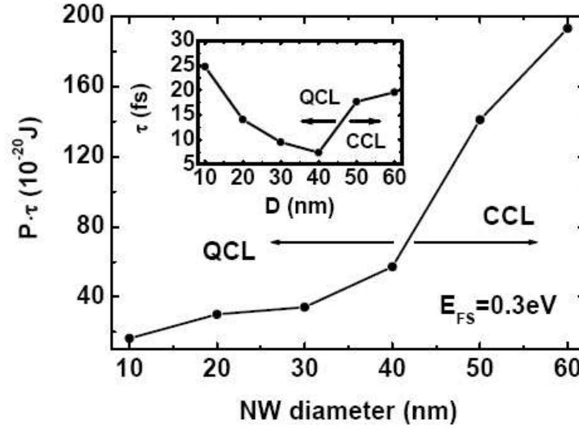



Figure 5.18: *Simulated power-delay product ($P \cdot \tau$) as a function of NW diameter with source Fermi level of 0.3 eV. Device operating regimes, quantum capacitance limit (QCL), and classical capacitance limit (CCL) are indicated in the plots. These curves show a significant performance improvement in terms of the power-delay product for the devices scaled towards the quantum capacitance limit. Inset shows the gate delay time as a function of NW diameter extracted from the simulations. A gate length of 10 nm was considered to calculate these values. Reproduced with permission from [5]. Copyright [2010], American Institute of Physics.*

The energy-delay product ranges from $1.5\times10^{-34}$ Js to $4.0\times10^{-33}$ Js within a source Fermi level range of 0.1 eV to 0.2 eV (Table 5.7). It is interesting that the $1.3\times10^{-33}$ Js energy-delay product of the 10 nm NW FET with $E_{FS} = 0.2$ eV falls on the projected experimental curve for a III-V planar HEMT with a 10 nm channel width shown in Fig. 7 of [74].

Both the energy-delay and power-delay products are optimal for diameters in the range of 10 - 40 nm with source Fermi levels of 0.1 eV. Increasing the source Fermi level increases $C_Q$ which causes a rapid degradation of the energy-delay product. Approximating $\tau$ as $\frac{C_Q V_{DD}}{I_D}$, the energy-delay product is proportional to $\frac{C_Q^2}{I_D}$. In this case, the increase in $C_Q$ overrides the increase in $I_D$.

Table 5.7 summarizes the doping densities, the ON-current, transconductance $g_m$, intrinsic cut-off frequency $f_T$, power-delay product, gate delay time, and energy-delay

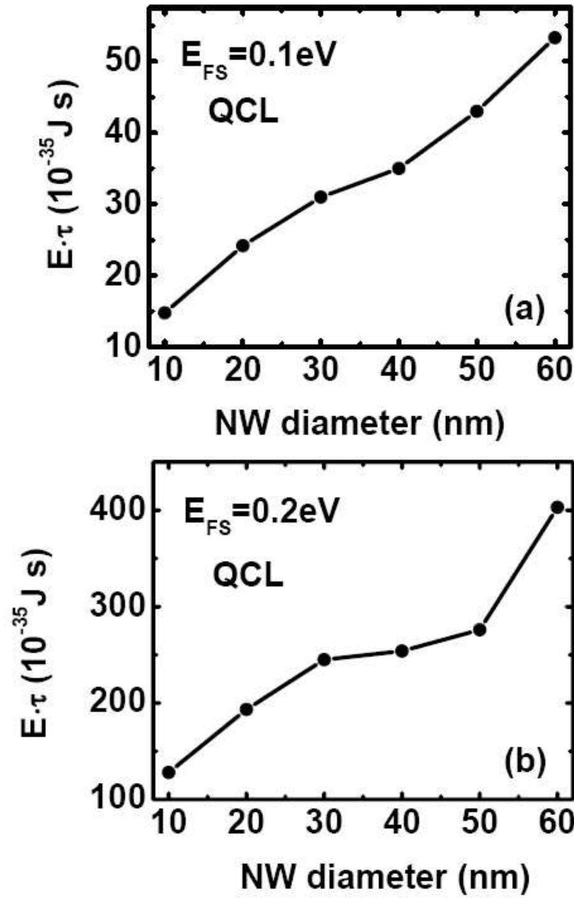



Figure 5.19: *Simulated energy-delay product ($E\cdot\tau$) as a function of NW diameter with source Fermi level of 0.1 eV (a), and 0.2 eV (b). $E\cdot\tau$ is the product of the $P\cdot\tau$ and the gate delay $\tau$. These curves show a significant performance improvement in terms of the energy-delay product for the devices scaled towards the quantum capacitance limit with $E_{FS}$ =0.1 eV. Reproduced with permission from [5]. Copyright [2010], American Institute of Physics.*

product for the NW FETs with two source Fermi levels. For a particular source Fermi energy, the transconductance increases with an increase of the NW diameter due to the higher number of populated modes. A similar behavior in transconductance and NW diameter has been observed in experimentally realized InAs NW FETs [165]. The intrinsic cut-off frequency varies from 9 - 52 THz for all devices within a source Fermi level range of 0.1 - 0.2 eV, which is good for RF applications. For single-moded NW FETs, our numerically calculated value of $f_T$ =9 THz is in good agreement with the reported value of 6.9 THz calculated using full quantum non-equilibrium Green function simulation [166]. The doping densities ($N_D$) as a function of NW diameter for two different source Fermi energies are presented in Table 5.7. The doping density increases as the NW diameter decreases. This trend is experimentally and theoretically confirmed by others [165].

The numerical values of $I_D$, $g_m$, $f_T$, $P \cdot \tau$, $\tau$, and $E \cdot \tau$ for the single moded 10 nm diameter NW FET with $E_{FS}$ =0.1 eV shown in Table 5.7 are predicted fairly well by the values from the analytical calculations (discussed in chapter 4).

Table 5.7: *Figures of merit of the proposed InAs NW FETs with a 10 nm long gate. The intrinsic cut-off frequency is defined as* $f_T = g_m/(2\pi C_{GS})$ *with* $g_m$ *being the transconductance.* $C_{GS}$ *is the series combination of the geometric* ($C_G$) *and the quantum* ($C_Q$) *capacitances.* $I_{ON}$ *from Figs. 5.16 are also listed.* $P \cdot \tau$ *is calculated as* $\int Q dV_{GS}$, *where* $Q$ *is the total channel charge extracted from the simulations. Accordingly the gate delay time* $\tau = \int Q dV_{GS}/(V_{DD} I_{ON})$. *Reproduced with permission from [5]. Copyright [2010], American Institute of Physics.*

| D (nm) | $E_{FS}$ (eV) | $N_D$ ($10^{17}$cm$^{-3}$) | $I_{ON}$ ($\mu$A) | $g_m$ ($\mu$S) | $f_T$ (THz) | $P \cdot \tau$ ($10^{-20}$J) | $\tau$ (fs) | $E \cdot \tau$ ($10^{-35}$Js) |
|---|---|---|---|---|---|---|---|---|
| 10 | 0.1 | 6.1 | 7.70 | 55.0 | 10 | 2.4 | 6.2 | 14.8 |
| | 0.2 | 16 | 9.50 | 55.0 | 9.4 | 7.8 | 16 | 128 |
| 20 | 0.1 | 0.84 | 14.0 | 99.8 | 18 | 4.1 | 5.9 | 24.2 |
| | 0.2 | 4.7 | 25.1 | 231 | 19 | 16 | 12 | 193 |
| 30 | 0.1 | 0.46 | 19.5 | 220 | 30 | 5.5 | 5.6 | 31 |
| | 0.2 | 3.2 | 50.0 | 419 | 22 | 25 | 9.9 | 245 |
| 40 | 0.1 | 0.35 | 31.0 | 364 | 38 | 7.4 | 4.7 | 35 |
| | 0.2 | 2.5 | 75.2 | 640 | 24 | 31 | 8.2 | 254 |
| 50 | 0.1 | 0.32 | 46 | 555 | 52 | 10 | 4.3 | 43 |
| | 0.2 | 2.3 | 116 | 1000 | 34 | 40 | 6.9 | 276 |
| 60 | 0.1 | 0.30 | 81 | 919 | 46 | 15 | 3.6 | 53.3 |
| | 0.2 | 2.1 | 195 | 1600 | 38 | 63 | 6.4 | 403 |

## 5.7 Summary

Deeply scaled $n$-type NW FETs based on InSb and InAs materials operate in the QCL, and the corresponding $p$-type NW FETs based on the same materials operate in the CCL. Drive currents at a fixed gate overdrive for the $n$- and $p$-type devices are well matched. However, the $p$-type devices with larger density-of-states have more charge in the channel. This results in the $p$-type devices having twice the power-delay products, twice the delay times, half the cut-off frequencies, and 4-5 times the energy-delay products of the single-moded $n$-type devices. Analysis of a single-moded NW FET in the QC regime, shows that the current and charge decouple. The intrinsic device metrics in this regime are optimized by using channel materials with a minimum effective mass.

Diameter dependent analysis of the device metrics for both InSb and InAs NW

FETs have been performed. It is found that NW FETs upto a NW diameter of $\sim$50 nm operate in the QCL. Both the energy-delay and power-delay products are minimized by choosing NW FETs with diameters of 10 - 40 nm with a source Fermi level in the range of $\sim 0.1$ eV. These NW FETs provide both ultra low-power switching and high speed.

# Chapter 6

# Nanowire MOSFETs: High-speed, high-power applications

We have investigated NW MOSFETs based on GaN NWs as channel for high-speed, high-power and microwave applications. We calculate the energy dispersions for the NWs using 8-band $\mathbf{k} \cdot \mathbf{p}$ method that has been described in detail in Chapter 2. In this chapter, we discuss the calculated and simulation results of GaN NW FETs. We start with a brief introduction of GaN-based NW FETs in Sec. 6.1. The model for the NW FETs is discussed in Sec. 6.2. The energy-wave vector relations for the NWs is presented in Sec. 6.3.1. The simulated results of carrier density, quantum capacitance and current are presented in Sec. 6.3.2 following summary in Sec. 6.4.

## 6.1 Introduction

One-dimensional (1-D) nanostructures such as nanowires (NWs) and nanotubes are of considerable interest due to their potential for next generation nanoelectronics and nanophotonics. Gallium nitride (GaN) NWs with direct wide bandgaps, large

breakdown fields, and high saturation velocities have attracted much attention for applications in high-speed, high-power nanoelectronics [167–179], nanophotonics [180, 181], HEMTs [182], NEMS [183] and memories [184]. There are are series of reports in literature on the experimental studies of GaN NWs [167–184] and few attempts to theoretically model them.

A 1-D semiconductor is expected to offer the right mix of properties for high performance FETs [1,2,5,185,186]. Due to the 1-D density of states, it is shown that NWs based on III-V materials such as InSb, InAs, InP [1–3,7], Si [147] and carbon [52] operate in the quantum capacitance limit (QCL) - a regime where tight control of the channel potential is attained with gate bias that is not accessible in conventional 2- or 3-D FETs [147]. In this work, however, we show that deeply scaled GaN NW FETs operate in the classical capacitance limit (CCL), in stead of operating in the QCL. Even for the ultra-small diameter (∼2 nm) GaN NW FETs, the geometric gate capacitance, $C_G$ is found to be smaller than the quantum capacitance, $C_Q$, and thus, the NW FETs operate in the classical capacitance limit (CCL). This is a result of the high electron effective mass and consequent high density of states. We will discuss how the CCL operation of these devices affect the device performance.

The purpose of this work is to perform theoretical study of the transport properties of deeply scaled GaN NW FETs and investigate their performance limit.

## 6.2 Model

The model for these devices is described in detail in Chapter 4, section 4.1. To be specific, the model consists of three capacitors, $C_G$, $C_S$, and $C_D$, which describe the electrostatic coupling between the top of the barrier and the gate, the source, and the drain, respectively.

## 6.3 Results

We present results for $n$-type GaN NW FETs on the calculations of energy dispersions, electron density, quantum capacitance, and current in the following sections. We consider NWs with diameters in the range of $2 - 8$ nm. Experimental realizations of such dimensions are reported [168, 173, 187]. In each case, the core material contains GaN surrounded by a shell material of AlN which is treated as the gate insulator. The thickness of the AlN is 4nm (equivalent to 1.6nm for $SiO_2$). The gate length is fixed at 10 nm. We first show the effect of confinement on the electronic properties of mode spacing, and quantum capacitance, and then we present calculations of the transport characteristics.

### 6.3.1 Energy dispersions

Fig. 6.1(a) shows the dispersion relationship for 2 nm core diameter GaN NW calculated within the 3-D discretized 8-band $\mathbf{k} \cdot \mathbf{p}$ method. $E$=0 corresponds to the lowest conduction band energy of the corresponding *bulk* GaN material. It is apparent from the $E$-$k$ plot that the dispersion for the electrons quickly deviates from parabolic, becoming linear above the energy of the mode minimum. Fig. 6.1(b) shows the number of spin-degenerate modes as a function of the mode energy extracted from Fig. 6.1(a). Fig. 6.1(c-d) presents the same with a GaN core diameter of 4 nm. We consider NW FETs with Fermi levels of up to 0.2 eV in the source. In all cases, the NW FETs contain multiple modes as shown in Fig. 6.1(b) and Fig. 6.1(d). We'll see the effects of these multiple modes on the quantum capacitance that we discuss next.

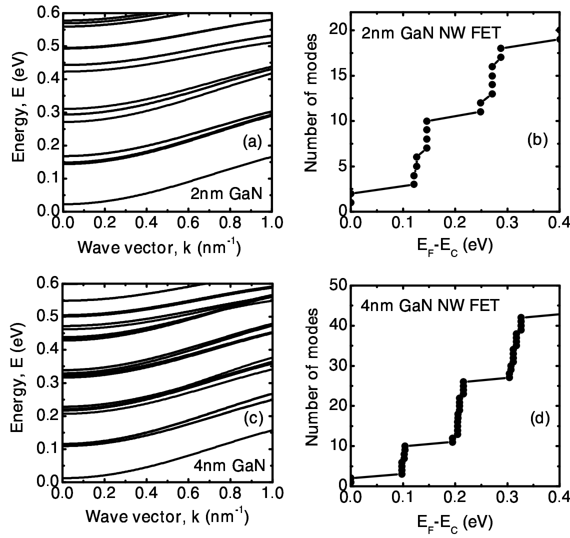



Figure 6.1: *(a) E-k relations calculated using the 3-D discretized 8-band* $\boldsymbol{k} \cdot \boldsymbol{p}$ *model, (b) Number of conduction band modes as a function of the mode energy for a 2-nm diameter GaN NW FET. Figs. (c-d) are similar to figs. (a-b) except with a diameter of 4-nm. Reproduced with permission from [6]. Copyright [2010], American Institute of Physics.*

### 6.3.2 Electron density, quantum capacitance and current

Fig. 6.2(a) presents the electron density under the gate as a function of the source Fermi energy for the two NWs. The carrier density is higher for the larger diameter GaN NW (4 nm). This is because larger diameter NW contains higher number of modes (Fig. 6.1(b-d)) that can be populated upon applying gate bias. Fig. 6.2(b) presents the quantum capacitance under the gate as a function of the source Fermi energy for the corresponding two NWs. The quantum capacitance is calculated from Eq. (5.9) with the charge, $-qn = Q_{top}$ calculated from Eq. (5.4). The corresponding geometrical gate capacitance, $C_G$ is also shown in the plot. The relationship between the number of populated modes and the quantum capacitance is obvious from Figs. 6.1(b-d) and 6.2(b). For a source Fermi energy such that $E_{FS} = E_F - E_C = 0.2$ eV ($E_C$ is the lowest conduction band-edge), all of the NWs have $C_G < C_Q$ and thus operate in the classical capacitance limit, in stead of operating in the quantum capacitance limit.

Fig. 6.2(c) shows how the band-edges under the gate change with applied gate bias. These curves are explained by Eq. (5.10). Initially, the channel is empty, there is a significant source-channel barrier to the electron flow, and thus, the quantum capacitance, $|\partial Q_{top}/\partial(E_{FS} - \varepsilon_1)|$ is exponentially reduced. Therefore, initially, the band-edges move identically with gate voltage with a slope whose magnitude is given by the gate control parameter $\alpha_G$. When the first set of excited modes under the gate are shifted by the gate bias such that they start to become populated by the source Fermi level at $V_{GS} = V_T$, the quantum capacitance turns on and is as shown in Fig. 6.2(b). The quantum capacitance becomes larger than the classical capacitance and the devices move into the CCL with $C_Q > C_\Sigma$ as shown in Fig. 6.2(b). This is a result of the high electron effective mass ($0.19\text{m}_0$ and $0.16\text{m}_0$ for the lowest conduction

band, respectively, for the 2- and 4-nm diameter GaN NWs), and consequent high density of states. When this happens, the denominator of Eq. (5.3) increases, and the magnitude of the slope $|\partial E_{CG}/\partial qV_{GS}|$ decreases as shown in Fig. 6.2(c). In the ideal classical limit with an infinite density-of-states, $C_Q$ becomes infinite, the slope $|\partial E_{CG}/\partial qV_{GS}|$ goes to zero, and the band-edge under the gate becomes fixed, independent of the gate voltage. This decrease in $|\partial E_{CG}/\partial qV_{GS}|$ resulting from charging of the channel is the expected behavior of devices operating in the CCL as shown in Fig. 6.2(c). Therefore, the band-edges in the channel presents a larger barrier to the source. This, in turn, affects the amount of current that flows from the source to the channel, as shown in Fig. 6.2(d).

Fig. 6.2(d) presents the log($I_{DS}$) vs. $V_{GS}$ transfer characteristics of the NW FETs calculated from Eq. 5.8. For all devices, $V_{DS}$ is chosen to be large enough ($E_G/(2q)$V) such that the gate drain leakage and back injection from drain can be minimized. It is clear from Fig. 6.2(d) that current gets saturated once $V_{GS}$=$V_T$ is applied. This is because the devices are operating in the CCL, where the source to channel barrier can not be reduced with applied gate bias as shown in Fig. 6.2(c). As explained earlier, in the classical capacitance limit, the band-edge under the gate becomes fixed independent of the gate voltage once $E_{CG} = E_{FS}$, as shown in Fig. 6.2(c). Thus, within the voltage range considered of $V_{GSmax}$=3.4V, the drive current is limited by the source to gate barrier, not source saturation. This can also be explained from Eq. 5.14. For a NW in the ballistic limit, ignoring thermal broadening, the current resulting from a single mode $m$ is $(2q/h)(E_{FS} - \varepsilon_m)$, where $\varepsilon_m$ is the mode energy in the channel. The energy difference $E_{FS} - \varepsilon_m$ is smaller (Fig. 6.2(c)) because of CCL operation of these devices. Thus, although there are modes (Fig. 6.1(b,d)) carrying current within $E_{FS}$=0.2eV, each mode is carrying less current. As a result, even though there are higher carrier densities under the gate (Fig. 6.2(a)) for higher gate

bias, current is not increased due to the source to the channel barrier that is posed due to the classical operation.

For the GaN NW FETs with a source Fermi level of 0.2 eV, the current density varies from 5 - 8.5 A/mm and the transconductance ranges from 11-12 $\mu$S. The off-current due to the thermionic emission over the gate barrier is $\sim 10^{-12}$A/mm. The power-delay products vary from $2.6\times10^{2}$ - $3.6\times10^{2}$ eV, and the intrinsic cut-off frequencies are in the range of 0.4 - 0.5 THz. The output power ranges from 0.06 - 0.07 mW. These NW FETs, thus, provide both ultra-high speed and high power.

## 6.4 Summary

The electronic bandstructures of deeply scaled GaN NWs have been modeled. The transport characteristics of GaN NW FETs have been investigated. It is shown that deeply scaled GaN NW FETs operate in the CCL in stead of operating in the QCL, which is a result of the high electron effective mass and high density of states. These devices offer both ultra-high speed and high power which has the potential for microwave application.

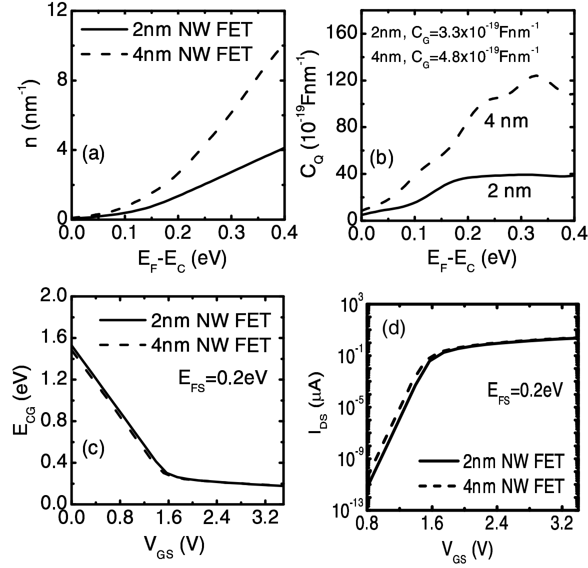



Figure 6.2: *(a) Electron density, (b) the quantum capacitance (corresponding gate capacitances are also shown) as a function of the source Fermi energy ($E_{FS} = E_F - E_C$, with $E_C$ being the lowest conduction band-edge) for 2- and 4-nm diameter GaN NW FETs. Fig. (b) shows that both the NW FETs are operating in the classical capacitance limit (CCL) with $E_{FS}$=0.2 eV as $C_G \ll C_Q$. (c) Band-edge under the gate as a function of the gate voltages for 2- and 4-nm diameter GaN NW FETs. (d) log($I_{DS}$) vs. $V_{GS}$ transfer characteristics of 2- and 4-nm diameter GaN NW FETs. The drain and gate voltages are fixed to $E_G/(2q)$ and $E_G/q$, respectively for each NW FET, to maximize the drive current. The OFF-current is significantly reduced due to the large bandgap for GaN NWs, which blocks the leakage current. Reproduced with permission from [6]. Copyright [2010], American Institute of Physics.*

# Chapter 7

# Nanowire (NW) band-to-band tunneling FETs (TFETs)

Band-to-band tunneling FETs (TFETs) are considered to be the potential successor for post-CMOS logic applications because of their ability to operate in the sub-60mV/dec limit. TFETs, therefore, offer ultra-low power operation at high speed. Nanowire (NW) and carbon nanotube (CNT) based TFETs are of special interest to the research community because of their ability to operate in the quantum capacitance limit (QCL). Because the tightest control of the gate to the channel potential can be achieved in the QCL, devices operating in this limit are good candidates for TFET applications.

Because tunneling of carriers take place through the bandgap (barrier), tunneling current is exponentially suppressed. Therefore, achieving sufficient drive current out of these devices is a fundamental issue. However, some authors have demonstrated barrier-free tunneling for carbon-based TFETs [188]. The same idea can be applied for NW-based TFETs using materials such as InSb and InAs.

Moreover, to realize band-to-band tunneling for carriers, TFETs often have to

contain very heavily doped source and drain regions. We have developed generalized models to accurately determine tunneling current as a function of material, diameter and electric field. We have also developed general methods to model the effects of heavy doping on the subthreshold characteristics of NW TFETs. We begin this chapter with brief introduction in Sec. 7.1. The definition, working principle and motivation behind TFETs are described in Secs. 7.2 - 7.3. In Sec. 7.4, we present the results and the detail of the model that we have developed to determine the drive current as a function of material, diameter and electric field. Finally, Sec. 7.5 presents the results and the detail of the model that we have developed to investigate the effects of heavy doping on the subthreshold characteristics of NW TFETs, following the summary in Sec. 7.6.

## 7.1 Introduction

Short channel effects (SCEs) are the severe roadblocks for aggressive downsizing of silicon MOSFETs. Increased SCEs and high leakage power significantly degrade device on-state performance and on-off current ratio. On-state and off-state current go hand in hand due to the fundamental limit of 60 mV/dec in the subthreshold slope. One of the challenging shortcomings of nanoscale MOSFETs is, therefore, its high leakage power consumption.

For the standard MOSFETs, the gate voltage changes the electrostatic potential in the gate region. To the injected electrons, this can look like a small potential hill or a large potential hill depending on the gate voltage. Because of thermal energy, electrons have a thermal energy distribution which decays exponentially above the top of the barrier according to $e^{-E/kT}$ where $E$ is energy, $k$ is Boltzmann's constant, and $T$ is the temperature. Therefore, there are always some high-energy electrons

which can go over the top of the potential hill in the gate region and reach the drain. This thermal distribution sets a limit as to how much voltage is required to reduce the current by a factor of 10 (referred to as the inverse subthreshold slope (SS)). The ideal limit at $T$=300K (80$^o$F) is 60 mV/dec which means that 60 mV of gate voltage is required to reduce the current by a factor of 10, and 360 mV is required to reduce the current by a factor of $10^6$. This is a fundamental limit to the voltage required to switch off a FET. With aggressive technology scaling, the SS of a device essentially imposes a severe constraint on the minimal operational voltage needed to achieve a decent on-off current ratio at a certain on-current. A device offering an SS lower than the conventional limit of 60 mV/dec guarantees a significantly reduced off current while maintaining a reasonable large on-current in the on-state. This, in turn, enables the designers to meet the desired frequency specifications in high-performance integrated circuits (ICs) while significantly suppressing standby leakage power.

To go to lower voltages (and thus, lower power), one must inject a non-equilibrium distribution of electrons from the source into the gate region, a distribution that does not have the exponential high-energy tail. Such a scenario is referred to as 'cold carrier injection,' and it is achieved in TFETs [189, 190] by injecting electrons from the source to the gate region by quantum mechanical tunneling through the bandgap, a process first discovered by Leo Esaki for which he won the Nobel prize. Therefore, recently a great deal of research works have been done on TFETs [4, 7, 45–49, 51, 52, 188, 191–193]. TFETs based on both direct-gap (III-V) and indirect-gap (Si) semiconductor nanowires (NWs) and carbon nanotubes (CNTs) are being investigated theoretically [4, 7, 47–49, 51] and experimentally [45, 46, 52].

## 7.2 TFET definition and working principle

TFETs rely on tunneling of carriers through the bandgap. Fig. 7.1 shows the energy band profile of an $n$-type TFET structure both at on- and off-states. The device is composed of a $p^+$ source, an intrinsic (or very lightly doped) channel, and an $n^+$ drain as shown in Fig. 7.1(a) and (b). The channel of this device can be a semiconductor NW or CNT. In realizing an $n$-type TFET, the $p^+$ source doping is kept larger than the $n^+$ drain doping to favor the tunneling through the $p^+$-$i$ junction while significantly reducing the tunneling probability at the $n^+$-$i$ junction. The opposite is true for realizing a $p$-type TFET.

In equilibrium, the Fermi level is aligned with the valence band-edge in the source, $E_{VS}$ and the mid-gap in the channel. Due to a positive gate bias, the conduction band-edge in the channel, $E_{CG}$ under the gate moves down. With sufficiently large gate bias applied, $E_{VS}$ and $E_{CG}$ un-cross and eventually an energy pass window is created through which tunneling of carriers take place. A sufficiently high longitudinal electric field within the bandwidth can cause a significant amount of tunneling current to flow across the reverse biased $p^+$-$i$ junction.

## 7.3 Why TFET?

As mentioned earlier, TFETs are considered to be potential candidates for next generation ultra-low power logic applications. The most beneficial advantage that TFETs offer for low-power logic applications is that TFETs can be designed with appropriate materials to operate in the sub-60 mV/dec range. When an FET device overcomes the fundamental limit in the subthreshold slope of 60 mV/dec, which is a Boltzmann constraint, it's off-current can be reduced significantly without compromising a required

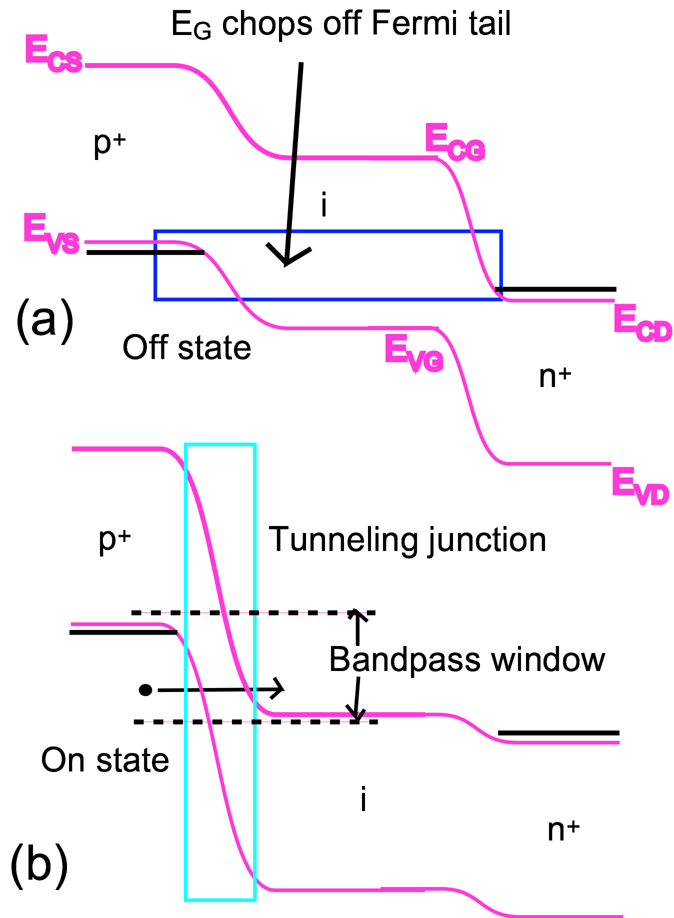



Figure 7.1: *Energy band profile of TFETs at (a) off state and (b) on state. Figure (a) shows how energy gap chops off the Fermi tail at the off-state and thereby minimize high energy thermal distribution of carriers. Figure (b) shows energy pass window through which carriers can tunnel through. Sufficiently large window can be created with heavily doped source and drain.*

on-current. Therefore, TFETs are the ideal devices for scaled down technology.

## 7.4 Drive currents: Effects of materials, diameters, and electric fields

In order to determine and comapre the tunneling current of TFETs, we have developed a general method which is based on calculating the imaginary wave vector in the bandgap. The method that we have developed allows one to analytically calculate the interband tunnel current for a constant electric field. One can survey material choices to determine maximum possible currents and required electric fields for specific on-off current ratios. In the following sections, we will present and discuss the model and the results. The model is applied to analyze and investigate InSb, InAs, InP, and GaN NW TFETs and CNT TFETs.

### 7.4.1 Model

The schematic structure of the model for evaluating tunneling current is presented in Fig. 7.2. The model illustrates tunneling of electron from the valence band to the conduction band. In the bandgap, the wave vector $\kappa$ becomes imaginary with a magnitude shown by the curve.

### 7.4.2 Analytical calculations

The band-to-band tunneling current for a direct gap semiconductor is given by the standard coherent tunneling expression

$$I_T = \frac{Mq}{h} \sum_{\text{spin}} \int dE T(E) \left[ f(E - \mu_L) - f(E - \mu_R) \right]. \tag{7.1}$$

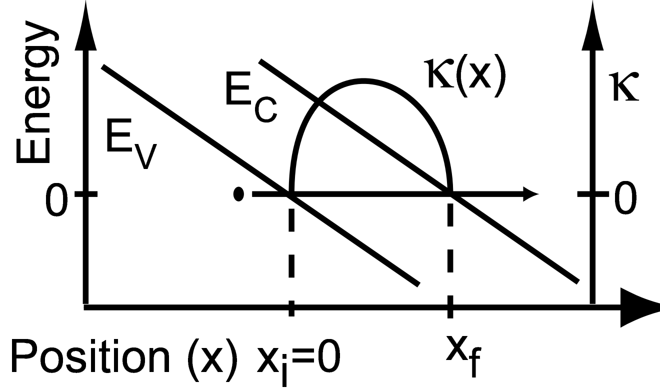



Figure 7.2: *Illustration of electron tunneling from the valence band to the conduction band. In the bandgap, the wavevector $\kappa$ becomes imaginary with a magnitude shown by the curve. $E_V$ and $E_C$ are the valence and conduction band edges, respectively. Reproduced with permission from [4]. © [2009] IEEE.*

where $f$ is the Fermi function, $\mu_L$ is the Fermi level on the left, $\mu_R$ is the Fermi level on the right, $q$ is the magnitue of the electron charge, $h$ is Planck's constant, $T(E)$ is the transmission coefficient, and $M$ is the orbital degeneracy of the fundamental mode. $M = 2$ for CNTs and $M = 1$ for nanowires. For the specific case of band-to-band tunneling, the transmission $T(E)$ is limited by the tunneling. If one can obtain a good expression for the tunneling transmission, one will have a good expression for the total transmission.

The interband tunneling process is illustrated in Fig. 7.2. For this process, the transmission can be approximated as $T = \exp\left\{-2\int_{x_i}^{x_f} dx'\kappa(x')\right\}$ where $\kappa$ is the magnitude of the imaginary wavevector within the bandgap [194]. Using the coordinates as shown in Fig. 7.2, the energy dependence of the valence band is $E_v(x) = -q\mathcal{E}x$ where $\mathcal{E}$ is the electric field. Changing the variable of integration from position to energy with respect to the valence band edge gives

$$T = \exp\left\{\frac{-2}{q\mathcal{E}}\int_0^{E_G} dE'\kappa(E')\right\} \equiv \exp\left\{-\mathcal{E}_I/\mathcal{E}\right\} \tag{7.2}$$

where $E = 0$ is the valence band edge, $E = E_G$ is the conduction band edge, and the integral $\mathcal{E}_I \equiv \frac{2}{q} \int_0^{E_G} dE' \kappa(E')$ is an intrinsic property of the nanowire. Once it is computed, one immediately has the tunneling coefficient for any electric field using Eq. (7.2). The transmission, $T$, is independent of energy as can be seen from Fig. 7.2, since injection from any other energy on the vertical axis results in an identical tunneling barrier.

To obtain the current, we could use Eq. (7.2) in Eq. (7.1), but in the spirit of this zero-order approximation, it is sufficient to define an energy window $\Delta E$ through which the current is tunneling. For drive current, depending on the source doping, this could be from either the valence band or the Fermi level of the source to the conduction band of the gate. For leakage current, this could be from the valence band of the gate to the Fermi level of the drain. Then, Eq. (7.1) reduces to

$$I_T = \frac{2Mq}{h} T \Delta E. \tag{7.3}$$

To verify the accuracy of Eq. (7.3), we compare its predictions to those of numerical simulations using Eq. (7.1) in which the transmission coefficient is calculated using the non-equilibrium Green function method as described in Refs. [135, 195].

The accuracy of both the analytical expression and the numerical expression depend on the accuracy of the bandstructure model. For the InAs and InSb nanowires, we use a discretized 8-band $\mathbf{k} \cdot \mathbf{p}$ model as described in Ref. [1]. The [100] nanowires are discretized on a uniform 1 nm grid. For the CNTs, we use a $\pi$-bond model with a nearest-neighbor Hamiltonian matrix element $V_{pp\pi} = -2.77$ eV [195]. The imaginary wavevector throughout the bandgap is calculated from the generalized eigenvalue problem as described in [121].

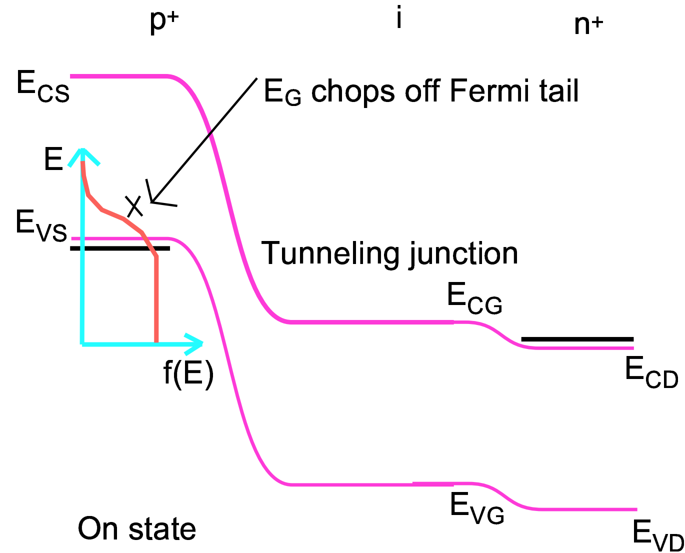



Figure 7.3: *Energy band profile of TFETs at on state showing how the Fermi-tail is essentially suppressed by the bandgap.*

### 7.4.3 Cold-carrier injection

The use of non-equilibrium carriers, in particular, cold carrier injection, may be a key enabler to extending the ITRS roadmap. In TFETs, cold-carriers can be exploited for tunneling which gives the advantage in terms of subthreshold slope lower than the fundamental limit of 60 mV/dec. The carriers are cold in the sense that the Fermi-tail is suppressed by the bandgap, as shown in Fig. 7.3. The high energy tail of the carriers, which is a direct consequence of the thermal distribution of carriers at room temperature, can be suppressed by employing the tunneling mechanism. Essentially then the carrier distribution will look like as the one at $T$=0K. The prototypical device which exploits cold carrier injection is the band-to-band tunnel FET (TFET).

### 7.4.4 InSb and InAs NW TFETs

There are two aspects of performance on which we will focus, the drive current, $I_{\text{on}}$ and the leakage current $I_{\text{off}}$. The drive current is limited by the band-to-band tunneling from the source to the channel. The leakage current considered is limited by band-to-band tunneling in the drain. This can also be the limiting mechanism for conventional nanowire or nanotube FETs [195]. Since adequate drive current has been an ongoing issue for TFETs, the small bandgaps of InAs, InSb, and CNTs should give a maximum drive current for this type of device. Furthermore, the small electron effective mass allows InAs and InSb NWs to reach the quantum capacitance limit at relatively large diameters [3]. Therefore, we will compare the tunneling currents in InAs and InSb nanowires and CNTs as a function of diameter and electric field.

To do this, we demonstrate a zero-order analytical approach for determining the tunnel current as a function of material, diameter, and electric field. The approach allows one to perform a quick, preliminary evaluation of a large design-space to narrow down the possibilities for more in-depth modeling and analysis [48] or designing split lots for empirical testing.

Fig. 7.4 shows the imaginary wavevector in the bandgap for 10 nm InAs and InSb nanowires and a (23,0) CNT. For each nanowire or nanotube, the tunneling current will be exponentially damped with a damping coefficient proportional to the area contained by the $E - \kappa$ curve. Thus, it is clear that, even though a 10 nm diameter InSb nanowire and a (23,0) CNT have similar bandgaps, tunneling will be considerably less for the InSb nanowire. To calculate the analytical expression for the transmission coefficient, we integrate these curves once for each material and diameter and then insert the value into Eq. (7.2) along with the relevant electric field.

The accuracy of the approach is assessed by comparing the current calculated

analytically using Eq. (7.3) to the current calculated numerically using Eq. (7.1) with $T(E)$ calculated exactly using NEGF. Linear potential drops, self-consistent potentials, and Fermi level degeneracy are considered. First, tunneling out of a full valence band into an empty conduction band with a linear potential drop is considered. The band diagram is shown in the inset of Fig. 7.5. The energy window lies between the valence band on the left and the conduction band on the right corresponding to $\Delta E = 0.5$ eV. In Eq. (7.1), $f(E - \mu_L) = 1$ and $f(E - \mu_R) = 0$. The current comparison, as a function of electric field, is shown in Fig. 7.5. The current axis for the CNT is on the left, and the current axis for the InSb and InAs NWs is on the right (note the different scales). The maximum ratio of the numerical to the analytical value is 1.57 occuring at $\mathcal{E} = 0.35$ V/nm for both InSb and InAs. At all other fields the ratio lies between 1 and 1.23 for the NWs. For the CNTs, for all fields, the ratio lies between 0.94 and 1.09. In a second test, we consider a self consistent potential for a 2 nm diameter InAs NW FET provided from Ref. [48] with an energy window between the valence band of the source and conduction band of the gate of 0.556 eV resulting from a potential drop between the source and the gate of 1.74 V. The Fermi level degeneracy in the source is $E_{VS} - E_{FS} = 0.0112$ eV from [48] resulting in a reduction of $\Delta E$ by 0.0112 eV from 0.556 eV to 0.545 eV. To obtain an electric field for Eq. (7.2), we approximate the potential drop as linear between points in the source and the gate where the potential is changed by $\pm k_B T$ from flat-band. The resulting electric field is 0.117 V/nm. The numerically calculated current obtained from Eq. (7.1) using the exact $f(E - \mu_L)$ with $\mu_L = E_{FS}$, $f(E - \mu_R) = 0$, and $T(E)$ calculated using NEGF, is 0.446 nA. The analytically calculated current from Eq. (7.3) is 0.469 nA. For all verification comparisons, the maximum ratio that we have observed between the numerical calculation using (7.1) and the analytical calculation using (7.3) is 1.6. This verifies that the analytical expression (7.3) captures well the

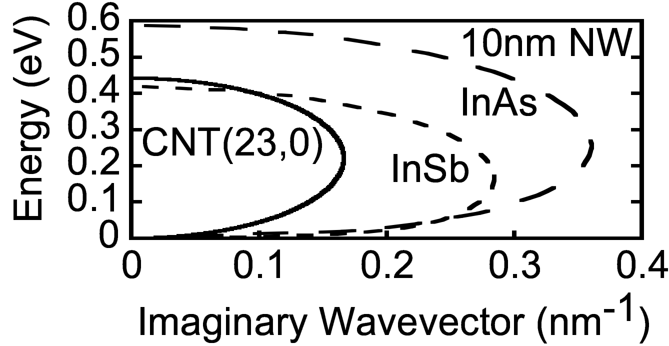



Figure 7.4: *Smallest imaginary wavevectors in the bandgap of 10 nm diameter InSb and InAs NWs and a (23,0) CNT. The valence band edges of all structures are set to $E = 0$. The curves shown couple the highest valence band mode and the lowest conduction band mode. Other such curves with larger imaginary wavevectors exist which couple lower valence modes to higher conduction modes, but they are irrelevant for tunneling since they are exponentially suppressed. The CNT curve is 2-fold degenerate. Reproduced with permission from [4]. © [2009] IEEE.*

dominant exponential dependence of current on the electric field, the band gap, and the dispersion, and that it is of sufficient accuracy for understanding current limits and for narrowing down the design space for more in-depth analysis.

Fig. 7.5(b) shows the value of the integral $\mathcal{E}_I \equiv \frac{2}{q}\int_0^{E_G} dE\kappa(E)$ as a function of nanowire material and diameter. Note that the axis for the CNT diameter is on top and the value for the CNT integral is at the right. With these values, one can immediately determine and compare the electric fields required to produce a desired drive current or prevent an undesired leakage current for different materials and diameter nanowires and CNTs.

We now demonstrate an example analysis using the integral values shown in Fig. 7.5(b) to understand the effect of material and diameter choice on the drive current, the leakage current, and the required electric fields. For this example, we choose a model device biasing scheme illustrated in Fig. 7.6. The device is p-i-n. In equilibrium, the Fermi level is aligned with the valence band edge in the source, the

conduction band edge in the drain, and midgap in the channel. Under bias, the drain voltage is $V_{DD}$. In the 'on' state, the gate voltage is also $V_{DD}$ which is chosen to give a tunneling window, $\Delta E$, of 0.2 eV between the *source* and the channel. To achieve this, the gate voltage must be sufficient to change the potential in the channel by $E_G/2+0.2$. The change in channel potential with respect to gate voltage is governed by the gate control parameter, $\alpha$, which lies in the range of $0.8-0.9$ [1]. Thus, $V_{DD}=(E_G/2+0.2)/\alpha$. Therefore, the tunneling window $\Delta E$ in the *drain* in the off state is somewhat larger than 0.2 eV, but for this analysis, we will assume that it is also 0.2 eV. Given this model biasing scheme, we now set targets for the 'on' current and the 'off' current and determine what electric fields are required to achieve the targets.

To set an achievable target for the 'on' current, one must first consider the implications of Eq. (7.3). Since $T\leq 1$, the maximum current achievable is $I_{\max}=2Mq\Delta E/h=15.5M$ $\mu$A for $\Delta E=0.2$ eV. To achieve a current of 10 $\mu$A in a nanowire with $M=1$ requires $T=0.65$. This requires an electric field of 0.76 V/nm in the 10 nm InAs NW and 0.43 V/nm in the 10 nm InSb NW. This means that 1 V must be dropped across 1.3 nm in the InAs NW and across 2.3 nm in the InSb NW. The feasibility of such fields can be understood by considering the screening length in a 1D wrap-around gate geometry [196], $\lambda=d\sqrt{\varepsilon/(8\varepsilon_{\rm ox})\ln(1+2d_{\rm ox}/d)+1/16}$ where $d$ is the NW diameter, $\varepsilon$ is the NW or CNT dielectric constant, $\varepsilon_{\rm ox}$ is the insulator dielectric constant, and $d_{\rm ox}$ is the insulator thickness. Assuming a 5nm $HfO_2$ insulator, $\lambda=3.4$ nm for both the 10nm InAs and InSb NWs. A drive current of 10 $\mu$A is not possible in the 10 nm InAs or InSb NWs under the biasing conditions shown. A drive current of 5 $\mu$A would require fields of 0.294 V/nm and 0.168 V/nm for 10nm InAs and InSb NWs, respectively, corresponding to 1V drops over 3.4nm and 6nm. Thus, 5 $\mu$A should be achievable for a 10nm InSb NW and near the limit of what

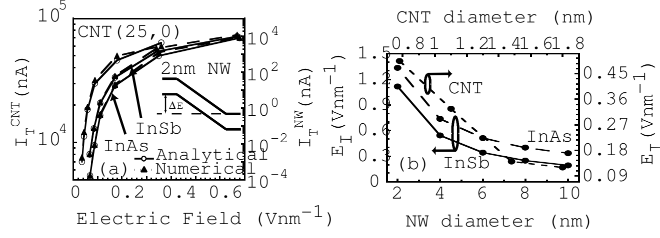



Figure 7.5: *(a) Calculated analytical and numerical tunneling currents as a function of the electric field for 2nm InSb and InAs NW FETs and for a (25,0) CNT FET. The current axis for the CNT is on the left, and the current axis for the InSb and InAs NWs is on the right (note the different scales). The inset shows the band profile and window of integration, $\Delta E$. (b) Wavevector integral, $\mathcal{E}_I \equiv \frac{2}{q}\int_0^{E_G} dE'\kappa(E')$, for InSb, InAs NWs, and for CNTs as a function of the NW/CNT diameter. The $\mathcal{E}_I$ axis for the NWs is on the left, and the $\mathcal{E}_I$ axis for the CNTs is on the right. Reproduced with permission from [4]. © [2009] IEEE.*

can be achieved for a 10nm InAs NW (for the bias condition of Fig. 7.6b). For demonstration, a current target of 1 $\mu$A is chosen. Fig. 7.6d shows the electric field required to achieve an 'on' current of 1 $\mu$A as a function of NW diameter. All of these fields are less than 1V/$\lambda$ so they should be achievable in all of the NWs.

Now considering the leakage process shown in Fig. 7.6(c), we ask, 'What drain fields are required to adequately shut the devices off with an on-off current ratio of $10^5$?' Taking the voltage drop from channel to drain to be approximately 1V, the inverse of the drain field gives the required drain underlap length $x_u$ shown in Fig. 7.6a. As the NW diameters go from 10nm to 2nm, the values of $x_u$ go from 33nm to 11nm for InAs and from 50nm to 12.5nm for InSb. Therefore, these NWs are capable of providing 1 $\mu$A of drive current with a $10^5$ on-off current ratio with drain underlaps less than 50nm for InSb and less than 33nm for InAs.

Because of the double degeneracy of the band edges and smaller integral, $\mathcal{E}_I$, an achievable on-current target for CNTs is 10 $\mu$A. The required fields for the on and off currents are shown in Figs. 7.6d and 7.6e, respectively. For the on current, the

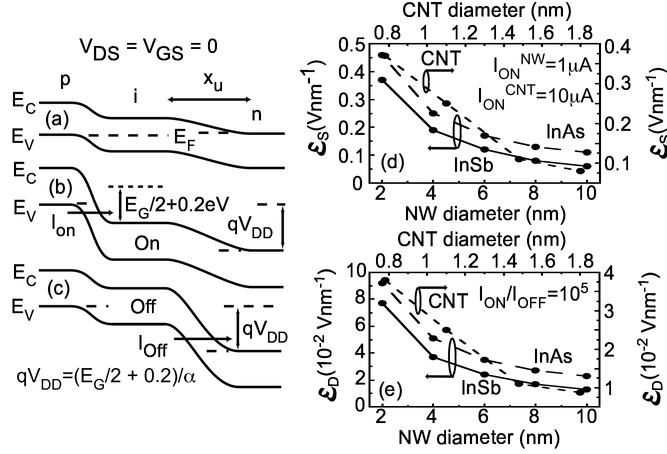



Figure 7.6: *Enhancement mode FET biasing. (a) No bias applied. (b) On state;* $V_{GS} = V_{DS} = V_{DD}$*. (c) Off state;* $V_{GS} = 0$ *and* $V_{DS} = V_{DD}$*. (d) Source electric field,* $\mathcal{E}_S$*, needed to achieve an ON-current of 1μA and 10μA, repectively, for NW FETs and CNT FETs as a function of diameter, and (e) Drain electric field,* $\mathcal{E}_D$*, required to obtain an ON-OFF current ratio of* $10^5$ *for various diameter NW/CNT TFETs. For (d) and (e), the fields for the NWs are on the left axis, and the fields for the CNTs are on the right axis. Reproduced with permission from [4]. © [2009] IEEE.*

source fields range from 0.37 V/nm for the 0.8 nm diameter CNT to 0.08 V/nm for the 1.8 nm CNT. All fields are less than 1V/$\lambda$. For the off current, the fields range from $8.6 \times 10^{-3}$ V/nm to $3.8 \times 10^{-2}$ V/nm corresponding to drain underlaps ranging from 26nm for the 0.8nm diameter CNT to 117 nm for the 1.8 nm CNT.

### 7.4.5 InP and GaN NW TFETs

InP is a direct bandgap material. It shows the potential to be used in high-speed, high-power electronic devices because of its excellent transport properties [197]. In this section, we present the theoretical investigation on the drive currents and leakage currents of InP NW TFETs based on calculating the imaginary wavevector in the bandgap as a function of NW diameter. Since adequate drive current has been an ongoing issue for TFETs, we investigate the tunneling currents in InP NW TFETs as a function of NW diameter and electric field. We show the feasibility of using such

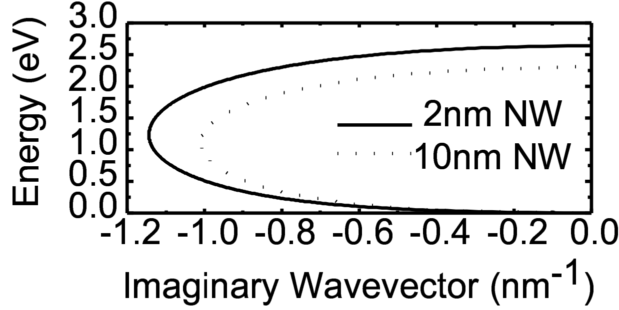



Figure 7.7: *Smallest imaginary wavevectors in the bandgap of 2 nm and 10 nm diameters InP NWs. The valence band edges of all structures are set to $E = 0$. The curves shown couple the highest valence band mode and the lowest conduction band mode. Other such curves with larger imaginary wavevectors exist which couple lower valence modes to higher conduction modes, but they are irrelevant for tunneling since they are exponentially suppressed. Reproduced with permission from [7]. Copyright [2009], American Institute of Physics.*

NWs in a TFET channel in terms of the drive current, leakage current and electric field.

The model that we use to investigate InP and GaN NW TFETs is presented in Sec. 7.4.1. For the NWs, we use a discretized 8-band $\mathbf{k} \cdot \mathbf{p}$ model as described in Ref. [1]. The [100] nanowires are discretized on a uniform 1 nm grid. The imaginary wavevector throughout the bandgap is calculated from the generalized eigenvalue problem as described in [121].

Fig. 7.7 shows the imaginary wavevector in the bandgap for 2 nm and 10 nm InP NWs. For each NW, the tunneling current will be exponentially damped with a damping coefficient proportional to the area contained by the $E - \kappa$ curve. Plotting the $E - \kappa$ curves allows one to quickly compare the tunneling properties of different NW diameters. To calculate the analytical expression for the transmission coefficient, we simply integrate these curves once for various diameter NWs and then insert the value into Eq. (7.2) along with the relevant electric field.

The diameter at which the QCL can be reached in a NW depends on the NW

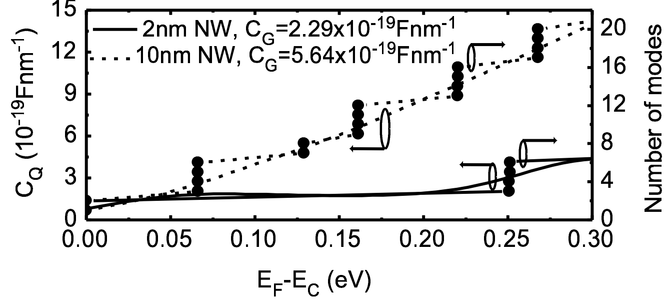



Figure 7.8: *Number of conduction band modes (right axis) and quantum capacitance (left axis) as a function of source Fermi energy for 2 nm and 10 nm diameter InP NWs. With a source Fermi energy of 0.1 eV, all the devices have $C_Q < C_G$, and thus operate in the quantum capacitance limit. Reproduced with permission from [7]. Copyright [2009], American Institute of Physics.*

material. Fig. 7.8 presents the number of modes (right vertical axis) and the quantum capacitance (left vertical axis) for 2 nm and 10 nm diameter InP NWs as a function of the source Fermi energy. With a source Fermi energy of 0.1eV, 2 nm diameter InP NW is single-moded (with spin). However, 10 nm diameter InP NW contains several spin degenerate modes. The second set of subbands for the 10 nm NW occur at 65 meV above the fundamental conduction band mode. It is clear that for source Fermi energy of 0.1eV, all NWs have the quantum capacitance $C_Q <$ the geometric gate capacitance $C_G$, and thus operate in the QCL. InP NWs, thus, reach the QCL at considerably larger diameters than Si NWs [147]. This is a result of the small electron effective mass and corresponding low density of states of InP.

We compare the analytically calculated current to the numerically calculated current to assess the accuracy of the approach presented in this letter. The analytically calculated current is obtained from Eq. (7.3), and the numerically calculated current is obtained from Eq. (7.1) with $f(E - \mu_L) = 1$, $f(E - \mu_R) = 0$, and $T(E)$ calculated using NEGF. A linear potential drop is considered with a $\Delta E = 0.5$ eV in Eq. (7.3). The current comparison, as a function of electric field, is shown in Fig. 7.9(a), and

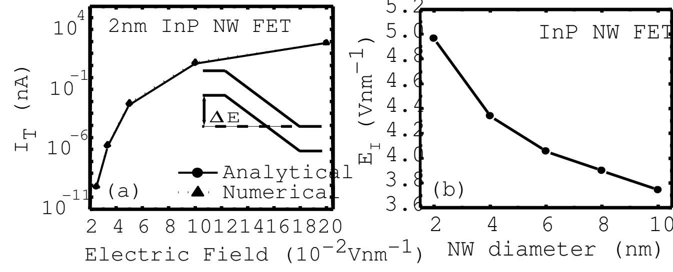



Figure 7.9: *(a) Calculated analytical and numerical tunneling currents as a function of the drain electric field for InP NW FETs. Inset shows the band profile and window of integration, $\Delta E$. (b) Wavevector integral, $\mathcal{E}_I \equiv \frac{2}{q}\int_0^{E_G} dE'\kappa(E')$ for the devices as a function of NW diameter. Reproduced with permission from [7]. Copyright [2009], American Institute of Physics.*

the potential profile is shown in the inset. The currents compare well over many orders of magnitude ranging from below fA to $\mu$A with electric fields ranging from $2 \times 10^5 - 2 \times 10^6$ V/cm.

Fig. 7.9(b) shows the value of the integral $\mathcal{E}_I \equiv \frac{2}{q}\int_0^{E_G} dE\kappa(E)$ for the InP NWs as a function of NW diameter. With these values, one can determine and compare the electric fields required to produce a desired drive current or prevent an undesired leakage current for different diameter NWs.

We now use the integral values shown in Fig. 7.9(b) to understand the source and drain electric fields that are required in NW TFETs. To do that, first we describe the model device biasing scheme that is illustrated in Fig. 7.10. The device is p-i-n. In equilibrium, the Fermi level is aligned with the valence band edge in the source, the conduction band edge in the drain, and midgap in the channel. Under bias, the drain voltage is $V_{DD}$. In the 'on' state, the gate voltage is also $V_{DD}$ which is chosen to give a tunneling window, $\Delta E$, of 0.25 eV between the *source* and the channel. To achieve this, the gate voltage must be sufficient to change the potential in the channel by $E_G/2 + 0.25$. The change in channel potential with respect to gate voltage is governed by the gate control parameter, $\alpha$, which lies in the range of $0.8 - 0.9$.

Thus, $V_{DD} = (E_G/2 + 0.25)/\alpha$. Therefore, the tunneling window $\Delta E$ in the *drain* in the off state is somewhat larger than 0.25 eV, but for this analysis, we will assume that it is also 0.25 eV. Given this model biasing scheme, we now set targets for the 'on' current and the 'off' current and determine what electric fields are required to achieve the targets.

The drive current in TFETs is limited by the band-to-band tunneling from the source to the channel. To set an achievable target for the 'on' current, one must first consider the implications of Eq. (7.3). Since $T \leq 1$, the maximum current achievable is $I_{\max} = 2Mq\Delta E/h = 19.3M$ $\mu$A for $\Delta E = 0.25$ eV. To achieve a current of 5 $\mu$A in a NW with $M = 1$ requires $T = 0.26$. This requires an electric field of 2.7 V/nm in the 10 nm InP NW. This means that 1 V must be dropped across 0.37 nm in the InP NW. The feasibility of such fields can be understood by considering the screening length in a 1D wrap-around gate geometry [196], $\lambda = d\sqrt{\varepsilon/(8\varepsilon_{\mathrm{ox}})\ln(1 + 2d_{\mathrm{ox}}/d) + 1/16}$ where $d$ is the NW diameter, $\varepsilon$ is the NW or CNT dielectric constant, $\varepsilon_{\mathrm{ox}}$ is the insulator dielectric constant, and $d_{\mathrm{ox}}$ is the insulator thickness. Assuming a 2nm $HfO_2$ insulator, $\lambda = 0.89$ nm for the 10nm InP NW. A drive current of 5 $\mu$A is not possible in the 10 nm InP NW under the biasing conditions shown. A drive current of 1 $\mu$A would require field of 1.2 V/nm for 10nm InP NW, corresponding to 1V drops over 0.83 nm. Thus, 1 $\mu$A should be achievable for a 10nm InP NW. For demonstration, a current target of 1 $\mu$A is chosen. Fig. 7.10d shows the electric field required to achieve an 'on' current of 1 $\mu$A as a function of NW diameter. All of these fields are less than 1V/$\lambda$ so they should be achievable in all of the NWs.

The leakage current occurs due to the band-to-band tunneling process from the channel to the drain. We consider an off-current limited by band-to-band tunneling in the drain. Considering the leakage process shown in Fig. 7.10c, we ask, 'What drain fields are required to adequately shut the devices off with an on-off current ratio

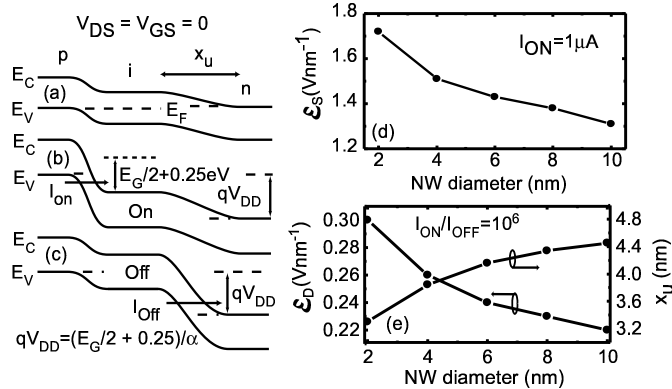



Figure 7.10: *Standard enhancement mode FET biasing. (a) No bias applied. (b) On state;* $V_{GS} = V_{DS} = V_{DD}$*. (c) Off state;* $V_{GS} = 0$ *and* $V_{DS} = V_{DD}$*. (d) Source electric field needed to achieve an ON-current of 1μA, and (e) Drain electric field (left vertical axis) and the drain underlap (right vertical axis) needed to acquire ON-OFF current ratio of* $10^6$ *for various diameter InP NW TFETs. Reproduced with permission from [7]. Copyright [2009], American Institute of Physics.*

of $10^6$?' Taking the voltage drop from channel to drain to be approximately 1V, the inverse of the drain field gives the required drain underlap length $x_u$ shown in Fig. 7.10a. As the NW diameters go from 10nm to 2nm, the values of $x_u$ go from 5nm to 3nm for InP NWs. Therefore, these NWs are capable of providing 1 $\mu$A of drive current with a $10^6$ on-off current ratio with drain underlaps less than 5nm.

Fig. 7.11 shows the bandgap as a function of the NW diameter for GaN NWs. The bandgap for the NWs varies from 4.25 - 4.0 eV (bulk GaN bandgap is 3.2eV at room temperature [109]) for NW diameters ranging from 2 - 8 nm. The confinement has little effect on the bandgap.

Fig. 7.12(a) shows the imaginary wave vector in the bandgap for 2- and 8-nm diameter GaN NWs. For each NW, the tunneling current will be exponentially damped with a damping coefficient proportional to the area contained by the $E$-$\kappa$ curve. To calculate the analytical expression for the transmission coefficient, we integrate these curves for each NW and then insert the value into Eq. 7.2 along with the relevant

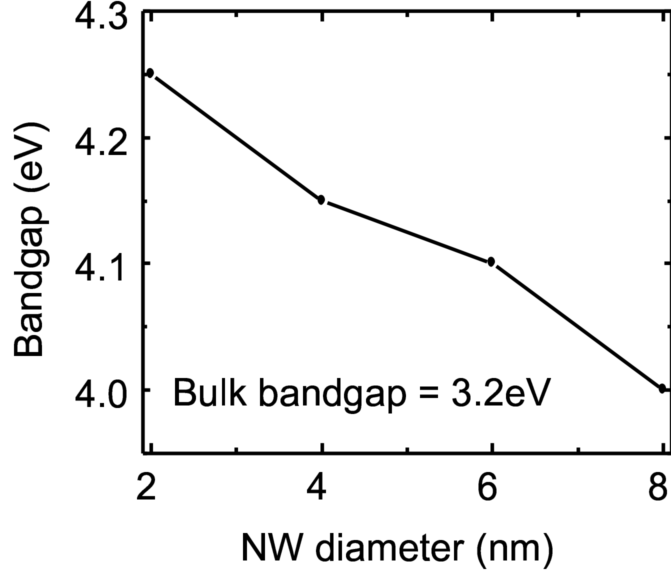



Figure 7.11: *Bandgap as a function of NW diameter for the simulated GaN NWs. Bulk GaN gandgap is also shown in the plot for comparison. Reproduced with permission from [6]. Copyright [2010], American Institute of Physics.*

electric field. Fig. 7.12(b) shows the value of the integral $\mathcal{E}_I \equiv \frac{2}{q}\int_0^{E_G} dE\kappa(E)$ as a function of NW diameter. With these values, one can determine and compare the electric fields required to produce a desired drive current or prevent an undesired leakage current for different NW diameters.

We now demonstrate an example analysis using the integral values shown in Fig. 7.12(b) to understand the effect of the NW diameter on the drive current, the leakage current, and the required electric fields. For this example we present a model biasing scheme as shown in Fig. 7.13. The device is *p*-*i*-*n*. In equilibrium, the Fermi level is aligned with the valence band edge in the source, the conduction band edge in the drain, and the midgap in the channel. Under bias, the drain voltage is $V_{DS}(=E_G/(2\text{q}))$. In the ON state, the gate voltage, $V_{GS}$ is chosen such that a tunneling window ΔE of 0.25 eV is achieved between the *source* and the channel.

Given this biasing scheme, we set the targer ON-current as 0.1$\mu$A for all devices

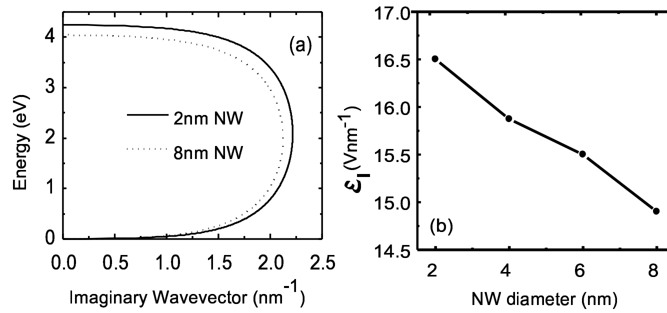



Figure 7.12: *(a) Smallest imaginary wave vectors in the bandgap of 2- and 8-nm diameter GaN NWs. The valence band-edges of all structures are set to E=0. The curves shown couple the highest valence band mode and the lowest conduction band mode. Other such curves with larger imaginary wave vectors exist which couple lower valence band modes to higher conduction band modes, but they are irrelevant for tunneling since they are exponentially suppressed. For smaller NW diameter, the imaginary band contains more area, and vice versa. Imaginary bands are considered as a leakage/tunneling path for electrons. The tunneling current will be exponentially damped with a damping coefficient proportional to the area contained by the imaginary band. (b) Wave vector integral,* $\varepsilon_I = (2/q)\int dE'\kappa(E')$, *for GaN NWs as a function of NW diameter.The smaller the NW diameter, the larger the values of the wave vector integral (damping coefficient), and the smaller the tunneling currents will be. Reproduced with permission from [6]. Copyright [2010], American Institute of Physics.*

and calculate the required source electric field to produce the target ON-current. The results are presented in Fig. 7.13(d). For NW diameter of 2 nm, the required source electric field to generate the target ON-current is 3.2 V/nm, and for NW diameter of 8 nm, the source field requirement is 2.8 V/nm. Considering the applied potential between the source and the channel as $E_G/2$, the source-channel underlap varies from 0.62 - 0.71 nm as NW diameter ranges from 2 - 8 nm. Then we calculate the drain electric field to adequately shut the device off with an ON-OFF current ratio of $10^6$. The results are shown in Fig. 7.13(e). For NW diameter of 2nm, the drain electric field needed to achieve the target off-current is 0.98 V/nm, and the field requirement for NW diameter of 8nm is 0.89 V/nm. Considering the applied potential between the the channel and the drain as $E_G/2$, the channel-drain underlap varies from 2.0 - 2.3 nm as NW diameter ranges from 2 - 8 nm. The fields calculated are all within the breakdown field of zinc-blende GaN. The feasibility of such fields is also evaluated by considering the screening length in a one-dimensional nanostructure [196].

In conclusion, we have presented a theoretical study on the drive currents and leakage currents of InP and GaN NW TFETs based on calculating the imaginary wavevector in the bandgap as a function of NW diameter. The integral of the $\kappa - E$ curve, $\mathcal{E}_I$, allows one to analytically calculate the interband tunnel current for any electric field. The analytical expression is verified with numerical calculations. One can then survey this material to determine maximum possible currents and required electric fields with specific on-off current ratios for applications in high-speed and high-power electronic devices.

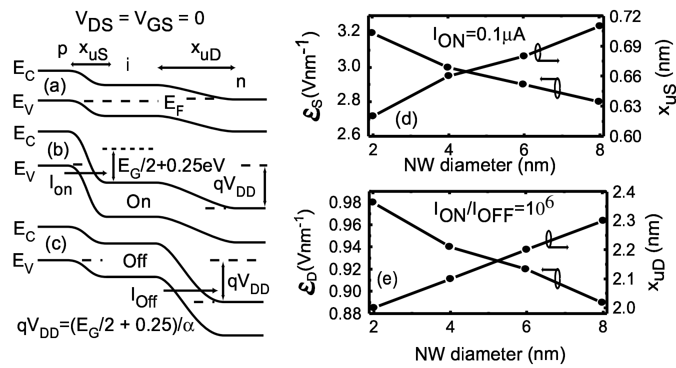



Figure 7.13: *Enhancement-mode FET biasing. (a) No bias applied, (b) ON state; $V_{GS} = V_{DS} = V_{DD}$. (c) OFF state; $V_{GS} = 0$ and $V_{DS} = V_{DD}$. $E_G$ is the bandgap of the NWs, and $\alpha$ is the gate control parameter. (d) Source electric field, $\varepsilon_S$ (left vertical axis) and source-channel underlap, $x_{uS}$ (right vertical axis) needed to achieve ON-current of 1 $\mu A$ for GaN NW TFETs as a function of NW diameter. (e) Drain electric field, $\varepsilon_D$ (left vertical axis) and the channel-drain underlap, $x_{uS}$ (right vertical axis) required to obtain an ON-OFF current ratio of $10^6$ for GaN NW TFETs as a function of NW diameter. With larger NW diameter, tunneling is higher and consequently less field is required to produce or suppress the desired drive current or undesired leakage current, respectively. Reproduced with permission from [6]. Copyright [2010], American Institute of Physics.*

## 7.5 Subthreshold characteristics: Effects of 'band-tails'

There are a number of questions concerning the design and ultimate performance of TFET devices. The drive current is limited by interband tunneling. An analysis on the drive currents as a function of materials, diameters and electric fields are presented in the previous sections. The off-current and inverse subthreshold slope are limited by several factors such as band tails resulting from heavy doping, impurities, and phonons.

To support the high electric fields required to achieve large interband tunnel current [4], heavy doping is required in the source and / or drain. When the doping is so high, the impurity bands merge with the conduction and valence bands giving rise to a large band-tail in the density of states which decays exponentially [198] into the bandgap. These band-tails can result from heavy doping, impurities, and phonons. The band-tails act as local scattering centers which can enhance carrier movement in the off-state, and thereby affect the subthreshold characteristics of NW TFETs.

The density of states in the presence of impurity states are calculated [199, 200] and their effects on the $I$-$V$ characteristics of resonant tunneling diodes are investigated [25, 201] analytically using parabolic band approximations. Recently the subthreshold slope of a NW TFET in the presence of an impurity state is being investigated analytically using parabolic approximation in the ballistic approach [50]. However, in order to understand the effects of inelastic scattering among electrons and impurities in the tunneling barrier [202], a full band quantum mechanical dissipative transport should be considered in the presence of incoherent scattering. To investigate the effects of realistic band-tails due to impurity states on the off-current and the subthreshold slope of NW TFETs, we have developed a generalized full band

quantum mechanical numerical model within the non-equilibrium Green's function (NEGF) formalism. The approach used a recursive Green's function (RGF) algorithm for both the retarded Green's function, $G^R$ [106] and the matrix of correlation operator, $G^<$ [107] in the presence of incoherent scattering within an 8-band $\mathbf{k} \cdot \mathbf{p}$ model [1, 2]. The method is applied to InSb NW TFETs. The method is, in detail, described in Chap. 3, Sec. 3.2.2.

In the following sections, we will present the generalized model that we have developed to investigate the effects of such band-tails on the subthreshold characteristics of NW TFETs. The model was applied to investigate the off-state and the subthreshold characteristics of NW TFETs based on InSb NWs. The simulated results will be presented in the following sections.

### 7.5.1 Model

We consider a square [100] InSb/InP core-shell NW with a core cross-section of 2 nm. The bandgap for this NW is 1.19 eV [1]. Fig. 7.14(a) shows schematic band diagram for a $p^+$-$i$-$n^+$ TFET in the subthreshold mode of conduction. $E_{CS(VS)}$ is the conduction(valence) band-edge at the source and $E_{CG(VG)}$ is the conduction(valence) band-edge at the gate region. $E_{VS}$=0 is taken as reference. The source and the drain regions are, respectively, 10 nm and 5 nm long. The source-to-gate and the gate-to-drain underlap regions are, respectively, 5 nm and 10 nm. Two gate lengths, 10 nm and 20 nm, are considered. In Fig. 7.14(a), two tunneling processes in the subthreshold conduction mode are indicated: (1) promoted by a mid-gap impurity state in the source, an electron can directly tunnel from the source to the gate region and subsequently collected by the drain, and (2) an electron can hop by the impurity states through the bandgap in the gate region and tunnel to the drain. These

tunneling mechanisms affect the subthreshold conduction of TFET devices which we investigate in these sections.

The impurity bands, when merged with the conduction and valence bands, give rise to large band-tails in the density of states which decays exponentially into the bandgap as $\exp\{-|E-E_{C,V}|/\alpha\}$, with $E_{C,V}$ being the conduction(valence) band-edge and $\alpha$ is the decay factor. This is illustrated in Fig. 7.14(b) for a 2nm core diameter InSb NW with $\alpha$=4kT.

The off current and the inverse subthreshold slope (S) are limited by such density-of-states tails (hereafter referred to 'band-tails') in the bandgap which result in mid-gap traps allowing trap-assisted band-to-band tunneling [50]. A larger value of $\alpha$ means a larger band-tail and a wider trap distribution and vice versa. It is important to investigate the effects of such impurity states in the bandgap to determine the super *cutoff* nature of TFETs in order to ensure a steep S ($\ll$60 mV/decade) at a reasonably higher on current.

### 7.5.2 What is 'band-tail'?

As mentioned earlier, band-tails can result from heavy doping, impurities and phonons. The presence of the band-tails in the bandgap of nanoscale materials behave like impurity states which can act as scatterers to enhance or diminish electron movement. Depending on the doping condition, there may be a single impurity in the bandgap which act as a local scattering center, or there may be series of impurities which act as scatterer chain. In either case, an electron finding itself in the scattering centers can tunnel through the bandgap by a mechanism well-known as 'hopping'. These unwanted tunneling of carriers in the off-state can severely degrade device subthreshold characteristics. In the off-state, TFET performance can be affected by these impurity

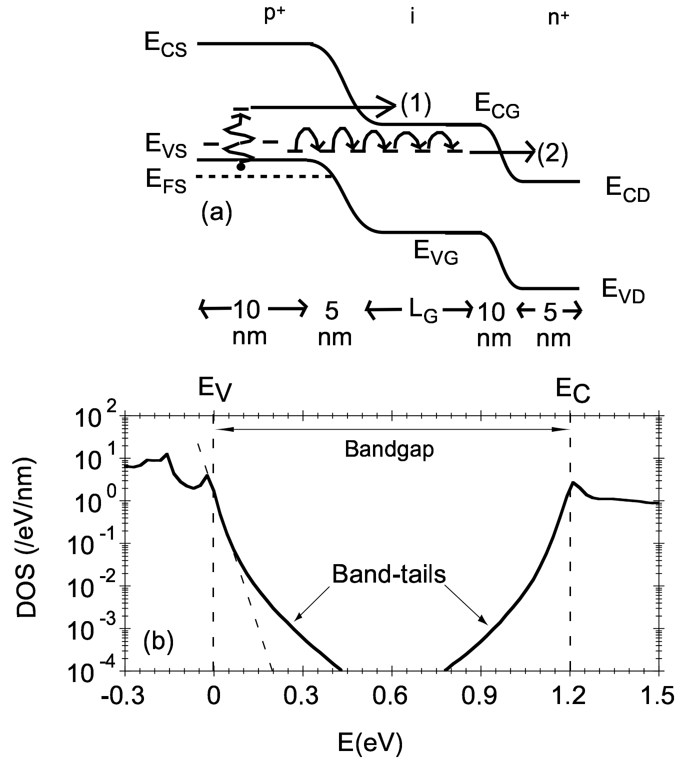



Figure 7.14: *(a) The band profile (not to scale) for a $p^+$-i-$n^+$ TFET in the off state. 'Band-tail' resulting from heavy doping promotes carrier tunneling from the source to the gate region. Two tunneling processes in the subthreshold conduction mode are indicated: (1) mid-gap trap assisted tunneling, and (2) tunneling by carrier hopping. (b) Density of states as a function of energy of 2nm wire diameter InSb NW showing the band-edges and the band-tails. Away from the 1-D band-edge singularity, the band-tails decay in the bandgap as $e^{-|E-E_{C,V}|/\alpha_A}$, with $E_C$ and $E_V$ being the conduction band-edge and valence band-edge, respectively. A range of values for $\alpha_A$ are investigated. In this case, $\alpha_A = kT$, with $k$ being the Boltzmann constant and $T$ is the room temperature (300K).*

bands which result from the band-tail like density of states.

Such a band-tail can be modeled by incorporating energy dependent self-energy in the NEGF equations. A representative density of states plot including such band-tails is calculated for a 2nm InSb NW, and presented in Fig. 7.14(b). The density of states decays exponentially in the bandgap by a decay factor, as mentioned in the previous section, as $\alpha$. The decay factor $\alpha$ can be comparable to the thermal energy $k$T ($k$ is the Boltzmann constant and $T$ is the room temperature, 300K). Thus, the decay of carrier density below the band-edge can be comparable to the decay of carrier density above the band-edge due to the thermal tail of the Fermi factor. In a standard FET, it is this thermal tail which sets the ideal lower limit of 60 mV/decade on the inverse subthreshold slope. In a TFET, however, it is not clear whether the band-tails would act in the same way. In this investigation, we report on the effects of band-tails on the off current and the inverse subthreshold slope of NW TFETs. We show that even with $\alpha$ comparable with or greater than $k$T, there is still a significant reduction of the inverse subthreshold slope in TFETs from its ideal thermally-limited value in FETs.

### 7.5.3 Numerical calculation

In the mathematical model, the exponentially decaying band-tails result from the energy-dependent self-energy, $\Sigma^R_{i,i} = -i\frac{\Gamma_{i,i}}{2}$, $\Gamma$ being the matrix for level broadening at $i$-th layer of the device. $\Gamma = \hbar/\tau_{in}$ where $\tau_{in}$ is the inelastic scattering time, $\Gamma = \Gamma_0 exp\{-(E_C - E)/\alpha\}$ for $E < E_C$, $\Gamma = \Gamma_0 exp\{-(E - E_V)/\alpha\}$ for $E > E_V$, and $\Gamma = \Gamma_0$ otherwise. $\Gamma_0 = 2\eta_0$. $\eta_0$ defines the broadening. The choice of $\eta_0$ dictates the feasible inelastic scattering time. We will present the results for various values of $\eta_0$. For instance, $\eta_0$=50 meV gives an inelastic scattering time $\tau_{in}$=6.6 fs, while $\eta_0$=20 meV gives an inelastic scattering time $\tau_{in}$=16.5 fs.

The inelastic scattering time due to ionized impurity (II) scattering is calculated using [203]

$$\frac{1}{\tau_{in}} = \frac{\pi N_D}{\hbar}\left(\frac{q^2 L_D^2}{\epsilon}\right)^2 g_C(E) \tag{7.4}$$

where delta-doping is considered. Here $L_D$ is the Debye length which is given as $L_D = \sqrt{\frac{\epsilon kT}{q^2 N_D}}$ [203]. $q$ is the magnitude of the electron charge, $\hbar$ is the reduced Planck's constant. $\epsilon = \epsilon_0 \epsilon_r$ with $\epsilon_0$ being the permittivity of free space and $\epsilon_r$ is the relative permittivity of materials. $N_D$ is the doping density of the NWs at a certain Fermi level. $g_C(E)$ is the 1-D density of states. Using the parameters such as $N_D$=5×10$^{19}$cm$^{-3}$ and $g_C(E)$=3.47 eV$^{-1}$nm$^{-1}$ for a 2 nm core diameter InSb NW for a Fermi level of 0.1 eV [1], inelastic scattering time due to II scattering, $\tau_{in}$ becomes 17.8 fs.

The band-to-band tunneling current in the presence of incoherent scattering is calculated as (in block matrix notation for $t$ and $G^<$) [106]

$$I_i = \frac{q}{\hbar}\int \frac{dE}{2\pi}\mathrm{Re}\left\{\mathrm{tr}\left[\mathrm{t_{i,i+1}G^<_{i+1,i}(E)}\right]\right\}, \tag{7.5}$$

where tr{...} indicates a trace over all orbital states. Linear potential drop is considered for calculating tunneling current. The accuracy of such an approach is verified [4]. Here, $t$ is the layer-to-layer coupling matrix and $G^<$ is the matrix for electron correlation operator. $G^<$ is calculated using RGF algorithm in the presence of incoherent scattering [107]

$$G^<_{i,i} = g^{\triangleright<}_{i,i} + g^{\triangleright R}_{i,i} t_{i,i-1} G^R_{i-1,i-1} t_{i-1,i} g^{\triangleright<}_{i,i} + g^{\triangleright R}_{i,i} t_{i,i-1} G^<_{i-1,i-1} t_{i-1,i} g^{\triangleright R\dagger}_{i,i} + g^{\triangleright<}_{i,i} t_{i,i-1} G^{R\dagger}_{i-1,i-1} t_{i-1,i} g^{\triangleright R\dagger}_{i,i}. \tag{7.6}$$

To calculate the layer-by-layer current for the device, the off-diagonal blocks of $G^<$

are created using Langreth's rule [138] as follows,

$$G^{<}_{i+1,i} = g^{\triangleright R}_{i+1,i+1} t_{i+1,i} G^{<}_{i,i} + g^{\triangleright <}_{i+1,i+1} t_{i+1,i} G^{R\dagger}_{i,i}. \quad (7.7)$$

### 7.5.4 Band-tails: Density of states

In the following analysis, we will consider two values of $\eta_0$, specifically 20 meV and 50 meV, which gives an inelastic scattering time, $\tau_{in}$ of 16.5 fs and 6.6 fs, respectively. The formaer value of the $\eta_0$ gives an inelastic scattering time comparable to the one with II scattering.

The density of states as a function of energy for a 2 nm diameter InSb NW at equilibrium showing the band-edges and band-tails is presented in Fig. 7.15 with $\eta_0$=20 meV, and in Fig. 7.16 with $\eta_0$=50 meV. The density of states is calculated using the relation $N^{1-D}_{i,i}(E) = (1/a)\mathrm{tr}[(-1/\pi)\mathrm{Im}\{\mathrm{G}^{\mathrm{R}}_{\mathrm{i,i}}(\mathrm{E})\}]$ with $a$ being the discretization length, and tr{...} is the trace over all states within a single discretized layer along the transport direction (including spin). Away from the band-edge singularity (shown as the dashed vertical lines in Fig. 7.15 and 7.16), the band-tails decay in the bandgap as $e^{-|E-E_{C,V}|/\alpha_A}$, with $E_C$ and $E_V$ being the conduction band-edge and valence band-edge, respectively. $\alpha_A$ represents the initial decay from the band-edge singularity, and is extracted from the initial decay of the density-of-states plots as shown in Fig. 7.15 and Fig. 7.16. A range of values for $\alpha_A$ (0.5$k$T to $k$T) are investigated.

With $\eta_0$=20 meV, the maximum decay factor $\alpha_A$ that we can achieve is 0.54$k$T, and with $\eta_0$=50 meV, the maximum decay factor $\alpha_A$ that can be achieved is 1$k$T. Moreover, $\eta_0$=20 meV gives an inelastic scattering time of 16.5 fs which is approximately equal to the inelastic scattering time due to II scattering of 17.8 fs (calculated using Eq. 7.4), whereas $\eta_0$=50 meV gives an inelastic scattering time of 6.6 fs which is

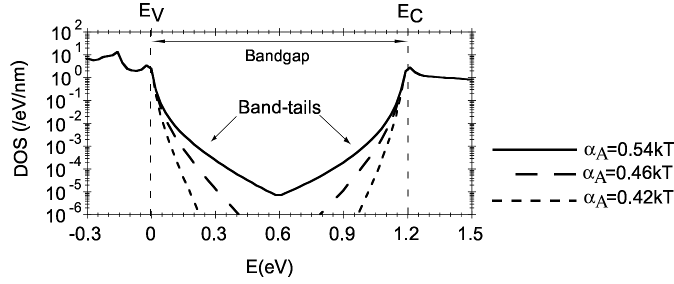



Figure 7.15: *Density of states as a function of energy of 2nm wire diameter InSb NW showing the band-edges and the band-tails with* $\eta_0$*=20 meV. Away from the 1-D band-edge singularity, the band-tails decay in the bandgap as* $e^{-|E-E_{C,V}|/\alpha_A}$*, with* $E_C$ *and* $E_V$ *being the conduction band-edge and valence band-edge, respectively. Plots for a range of values for* $\alpha_A$ *are investigated.*

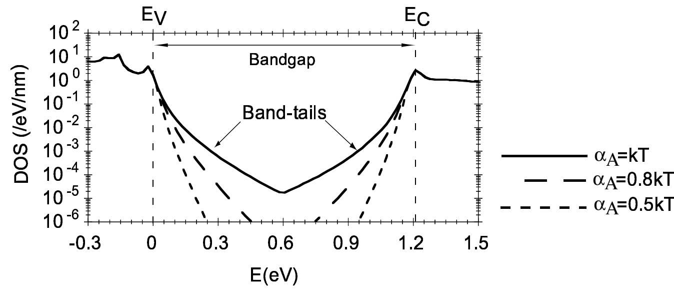



Figure 7.16: *Density of states as a function of energy of 2nm wire diameter InSb NW showing the band-edges and the band-tails with* $\eta_0$*=50 meV. Away from the 1-D band-edge singularity, the band-tails decay in the bandgap as* $e^{-|E-E_{C,V}|/\alpha_A}$*, with* $E_C$ *and* $E_V$ *being the conduction band-edge and valence band-edge, respectively. Plots for a range of values for* $\alpha_A$ *are investigated.*

less than that due to the II impurity scattering for the NWs based on InSb materials. We will show below the effects of broadening $\eta_0$ (and therefore that of $\tau_{in}$) on the subthreshold characteristics of InSb NW TFETs.

### 7.5.5 Band-tails: Effects of heavy doping in the source

Fig. 7.17 presents current density (in a log scale) as a function of energy for the NW TFETs considering $\Sigma^R$ used (impurity states and hence band-tails incorporated) in the source and the source-to-channel underlap region (Fig. 7.17a-b) with $\eta_0$=20meV, while Fig. 7.18(a-b) shows the same for $\eta_0$=50meV. Unless specified otherwise, a source Fermi level of 0.1 eV is used which corresponds to a doping density of $N_A$=5×10$^{19}$/cm$^3$. Such doping densities are experimentally realized for InSb NWs [204]. Two gate lengths, 10 nm and 20 nm, are considered. Plots are shown for three gate biases, i.e., $V_{GS}$=0 V (dotted lines), $V_{GS}$=0.125 V (dash dotted lines), and $V_{GS}$=0.25 V (solid lines). On top of each plot, $E_{CG}$ for three bias conditions are indicated as shown by the dashed vertical lines in Fig. 7.17 and Fig. 7.18. Current is dominated by tunneling through the bandgap for all cases. For a gate length of 20nm (Fig. 7.17(b) and Fig. 7.18(b)), current is significantly enhanced by the band-tails in the gate region.

Fig. 7.19 shows $I_D$-$V_G$ characteristics on a log scale for NW TFETs with a range of values for $\alpha_A$ with $\Sigma^R$ used in the source and the source-to-channel region with $\eta_0$=20meV and Fig. 7.20 shows the same for $\eta_0$=50meV. For comparison, $I_D$-$V_G$ curves with $\Sigma^R$=0 (coherent transport) throughout the device is also shown. $E_{VS}$ and $E_{CG}$ band un-crossing biases are indicated by the vertical ellipse. For a wider band-tail (higher values of $\alpha_A$, in other words, wider impurity distribution), the number of available states within the energy bandgap increases, which in turn, promotes

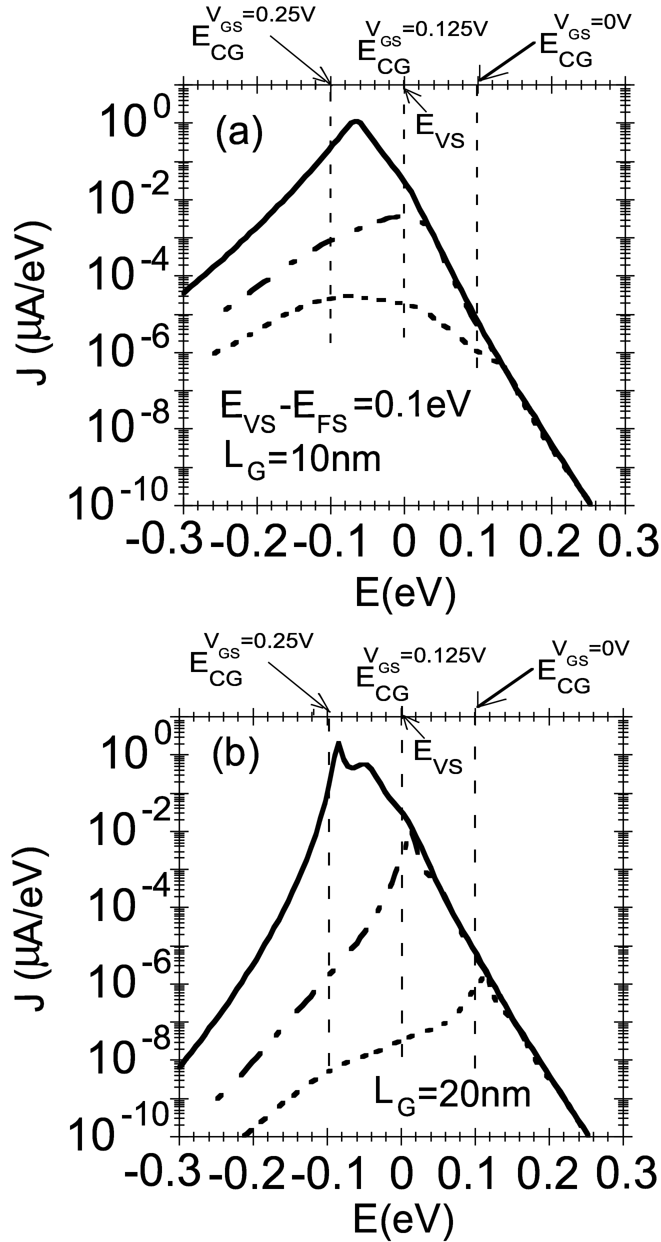



Figure 7.17: *Current density as a function of energy of InSb NW TFETs with $\Sigma^R$ used in the source region for two gate lengths (a) 10 nm and (b) 20 nm. For all plots, $E_{VS}$=0 eV is the valence band-edge in the source which is taken as reference. $\eta_0$=20meV and $\alpha_A$=0.54kT. Source Fermi level of 0.1 eV is used. Three gate biases are considered, i.e., $V_{GS}$=0 V (dotted lines), $V_{GS}$=0.125 V (dash dotted lines), and $V_{GS}$=0.25 V (solid lines). On top of each plot, level for conduction band-edge under the gate, $E_{CG}$ for three bias conditions are indicated by the dashed vertical lines.*

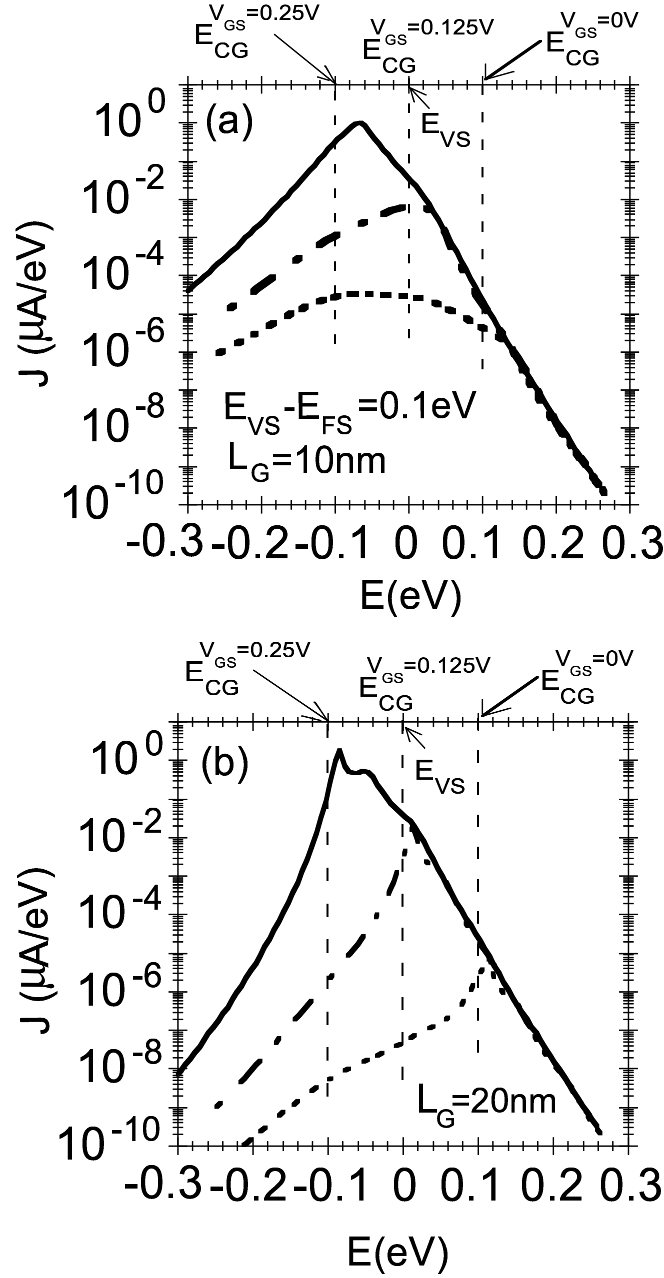



Figure 7.18: *Current density as a function of energy of InSb NW TFETs with $\Sigma^R$ used in the source region for two gate lengths (a) 10 nm and (b) 20 nm. For all plots, $E_{VS}$=0 eV is the valence band-edge in the source which is taken as reference. $\eta_0$=50meV and $\alpha_A$=kT. Source Fermi level of 0.1 eV is used. Three gate biases are considered, i.e., $V_{GS}$=0 V (dotted lines), $V_{GS}$=0.125 V (dash dotted lines), and $V_{GS}$=0.25 V (solid lines). On top of each plot, level for conduction band-edge under the gate, $E_{CG}$ for three bias conditions are indicated by the dashed vertical lines.*

larger tunneling current in the subthreshold mode of conduction. Note that for all devices, the off currents for the coherent transport is lower than those with incoherent scattering.

The subthreshold swing, S as a function of $\alpha_A$ for the NW TFETs with $\Sigma^R$ used in the source and source-to-channel underlap region is presented in Fig. 7.21 with $\eta_0$=20meV and that for $\eta_0$=50meV in Fig. 7.22. For all cases, as $\alpha_A$ increases, $S$ increases. It is shown that even with $\alpha_A$ comparable to $k$T (Fig. 7.22), it is possible to achieve values of S lower than the fundamental limit of 60 mV/dec, both for 10 nm and 20 nm gate length InSb NW TFETs.

### 7.5.6 Band-tails: Effects of heavy doping in the source and the channel

Fig. 7.23 presents current density (in a log scale) as a function of energy for the NW TFETs considering $\Sigma^R$ used in the source, source-to-channel underlap, and the gate region with $\eta_0$=20meV, and Fig. 7.24 shows the same for $\eta_0$=50meV. Unless specified otherwise, a source Plots are shown for three gate biases, i.e., $V_{GS}$=0 V (dotted lines), $V_{GS}$=0.125 V (dash dotted lines), and $V_{GS}$=0.25 V (solid lines). On top of each plot, $E_{CG}$ for three bias conditions are indicated as shown by the dashed vertical lines in Fig. 7.23 and Fig. 7.24. Current is dominated by tunneling through the bandgap for all cases. For a gate length of 20nm (Fig. 7.23(b) and Fig. 7.24(b)), current is significantly enhanced by the band-tails in the gate region.

Fig. 7.25 shows $I_D$-$V_G$ characteristics on a log scale for NW TFETs with a range of values for $\alpha_A$ with $\Sigma^R$ used in the source, source-to-channel, and the gate region with $\eta_0$=20meV and Fig. 7.26 shows the same with $\eta_0$=50meV. For comparison, $I_D$-$V_G$ curves with $\Sigma^R$=0 (coherent transport) throughout the device is also shown.

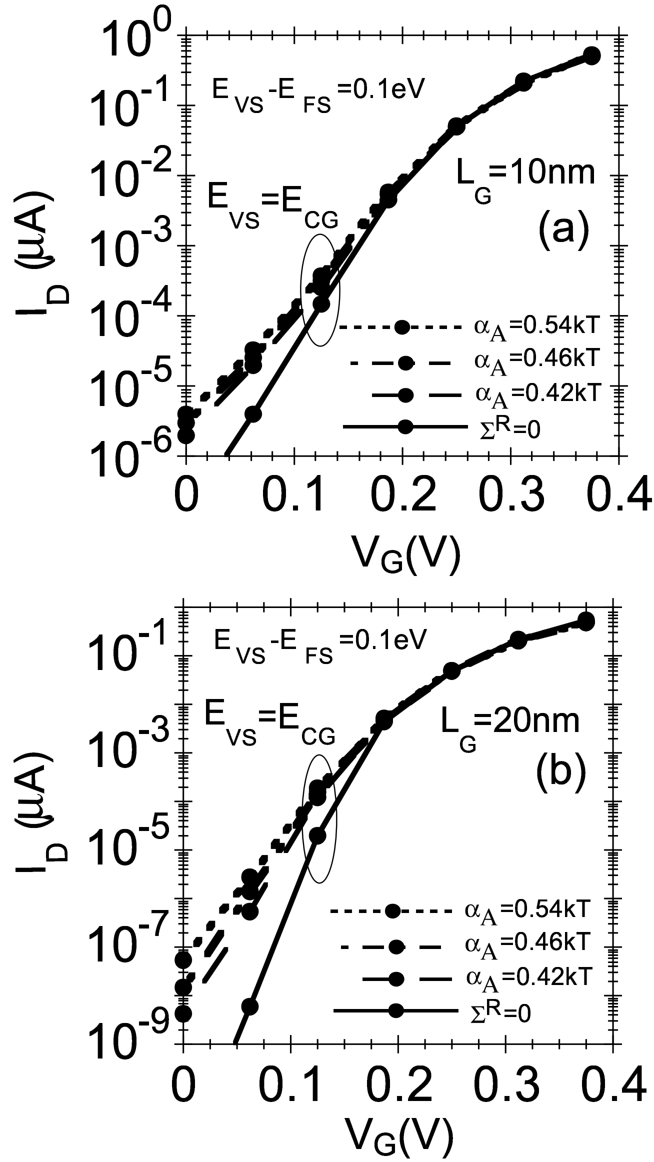



Figure 7.19: *$I_D$-$V_G$ characteristics on a log scale of InSb NW TFETs for a range of values for $\alpha_A$ with $\Sigma^R$ used in the source region for two gate lengths (a) 10 nm and (b) 20 nm. $\eta_0$=20meV. For comparison, $I_D$-$V_G$ curves with $\Sigma^R$=0 throughout the device is also presented. For all plots, a source Fermi level of 0.1 eV is used.*

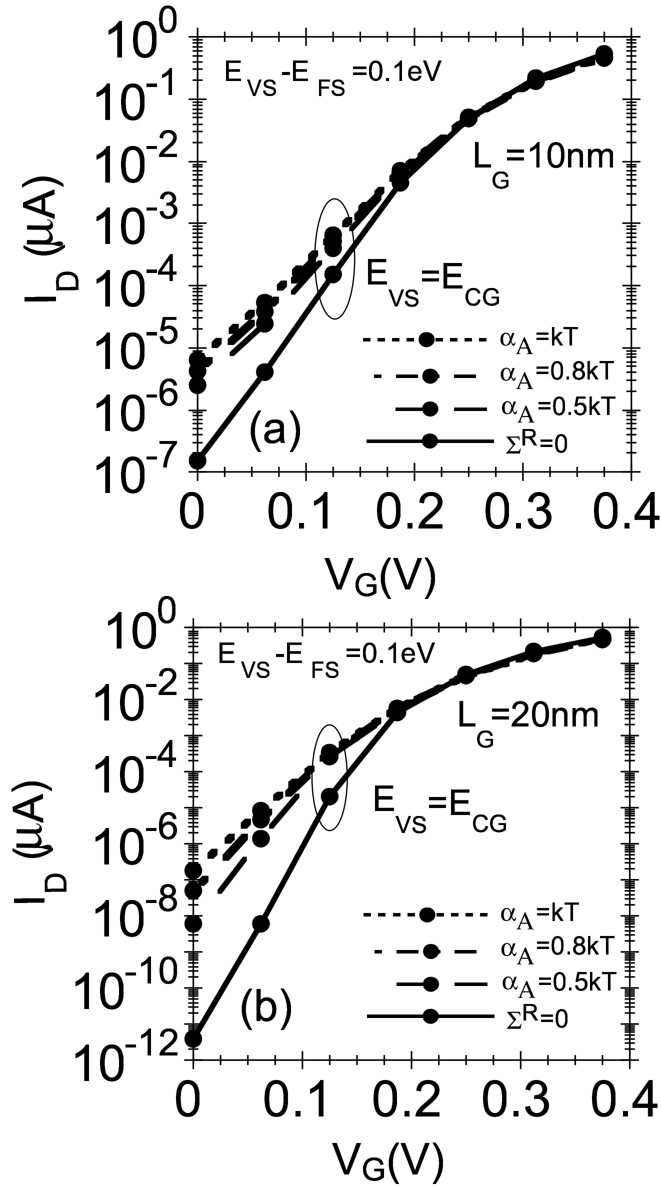



Figure 7.20: *$I_D$-$V_G$ characteristics on a log scale of InSb NW TFETs for a range of values for $\alpha_A$ with $\Sigma^R$ used in the source region for two gate lengths (a) 10 nm and (b) 20 nm. $\eta_0$=50meV. For comparison, $I_D$-$V_G$ curves with $\Sigma^R$=0 throughout the device is also presented. For all plots, a source Fermi level of 0.1 eV is used.*

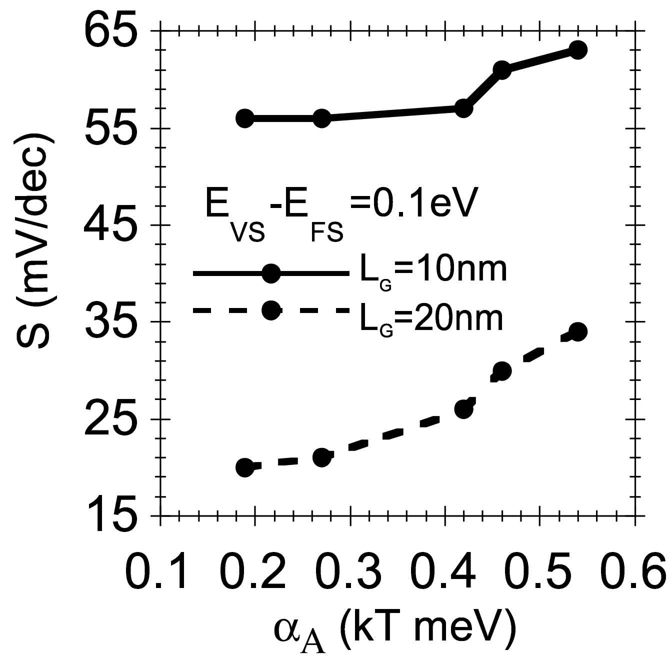



Figure 7.21: *Subthreshold swing, S as a function of $\alpha_A$ for InSb NW TFETs with $\Sigma^R$ used only in the source region for two gate lengths, 10 nm and 20 nm. $\eta_0$=20meV. A source Fermi level of 0.1 eV is used.*

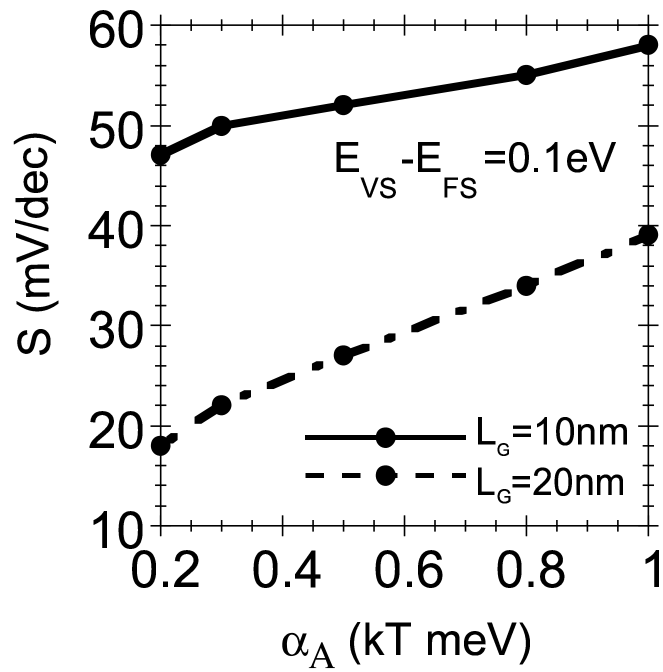



Figure 7.22: *Subthreshold swing, S as a function of $\alpha_A$ for InSb NW TFETs with $\Sigma^R$ used only in the source region for two gate lengths, 10 nm and 20 nm. $\eta_0$=50meV. A source Fermi level of 0.1 eV is used.*

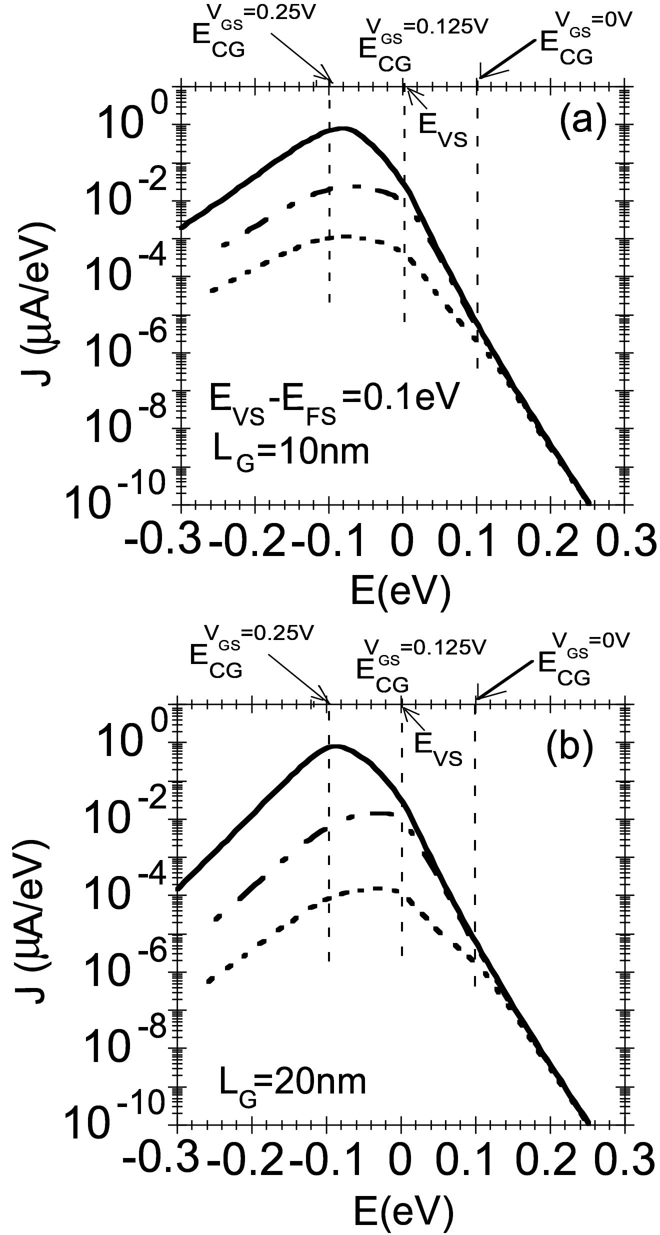



Figure 7.23: *Current density as a function of energy of InSb NW TFETs with $\Sigma^R$ used in the source, source-to-channel, and the gate region for two gate lengths (a) 10 nm and (b) 20 nm. For all plots, $E_{VS}$=0 eV is the valence band-edge in the source which is taken as reference. $\eta_0$=20meV and $\alpha_A$=0.54kT. Source Fermi level of 0.1 eV is used. Three gate biases are considered, i.e., $V_{GS}$=0 V (dotted lines), $V_{GS}$=0.125 V (dash dotted lines), and $V_{GS}$=0.25 V (solid lines). On top of each plot, level for conduction band-edge under the gate, $E_{CG}$ for three bias conditions are indicated by the dashed vertical lines.*

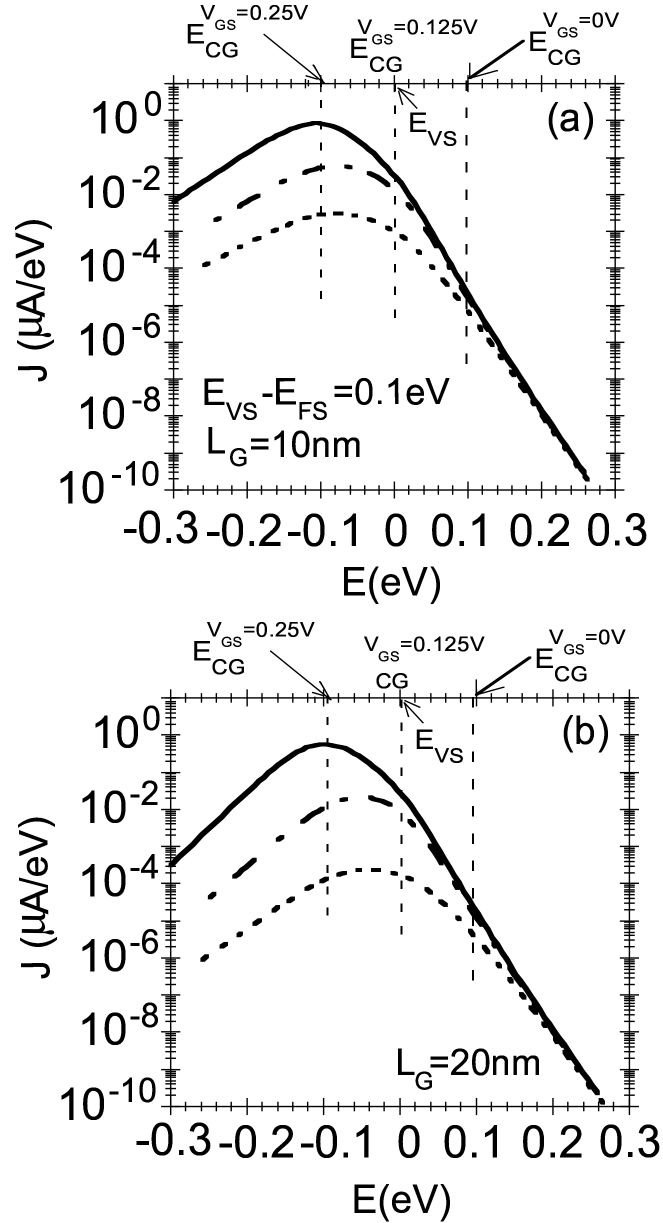



Figure 7.24: *Current density as a function of energy of InSb NW TFETs with $\Sigma^R$ used in the source, source-to-channel, and the gate region for two gate lengths (a) 10 nm and (b) 20 nm. For all plots, $E_{VS}$=0 eV is the valence band-edge in the source which is taken as reference. $\eta_0$=50meV and $\alpha_A$=kT. Source Fermi level of 0.1 eV is used. Three gate biases are considered, i.e., $V_{GS}$=0 V (dotted lines), $V_{GS}$=0.125 V (dash dotted lines), and $V_{GS}$=0.25 V (solid lines). On top of each plot, level for conduction band-edge under the gate, $E_{CG}$ for three bias conditions are indicated by the dashed vertical lines.*

$E_{VS}$ and $E_{CG}$ band un-crossing biases are indicated by the vertical ellipse. For a wider band-tail (higher values of $\alpha_A$, in other words, wider impurity distribution), the number of available states within the energy bandgap increases, which in turn, promotes larger tunneling current in the subthreshold mode of conduction. Note that for all devices with both the cases considered above, the off currents for the coherent transport is lower than those with incoherent scattering.

The subthreshold swing, S as a function of $\alpha_A$ for the NW TFETs with $\Sigma^R$ used in the source, source-to-channel underlap, and the gate region is presented in Fig. 7.27 with $\eta_0$=20meV and that with $\eta_0$=50meV in Fig. 7.28. As $\alpha_A$ increases, $S$ increases. For band-tails incorporated in the source, source-to-channel, and the gate region (Fig. 7.27 and Fig. 7.28), $S$ can be greater than 60 mV/dec (the fundamental limit for the conventional MOSFET) with a gate length of 10 nm.

## 7.6 Summary

In the first part of this chapter, we have presented a method for quickly comparing the tunneling properties of different materials for use as NW or CNT TFETs based on calculating the imaginary wavevector in the bandgap as a function of NW or CNT diameter. The integral of the $\kappa - E$ curve, $\mathcal{E}_I$, allows one to analytically calculate the interband tunnel current for a constant electric field. The analytical expression is verified with numerical calculations. One can then survey material choices to determine maximum possible currents and required electric fields for specific on-off current ratios. The analysis was applied to InSb, InAs, InP, GaN NWs and CNTs.

In the second part of this chapter, we have presented a generalized NEGF-RGF based algorithm for investigating the effects of band-tails in the source, and the channel on the subthreshold characteristics of NW TFETs. Band-tails can result

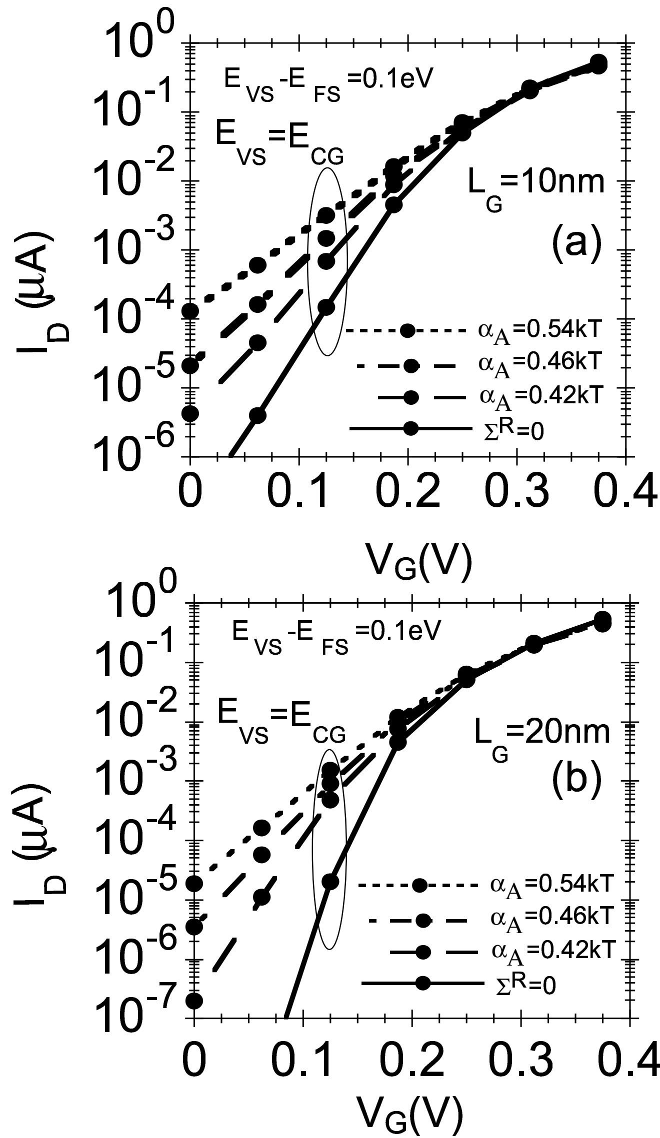



Figure 7.25: *$I_D$-$V_G$ characteristics on a log scale of InSb NW TFETs for a range of values for $\alpha_A$ with $\Sigma^R$ used in the source, source-to-channel, and the gate region for two gate lengths (a) 10 nm and (b) 20 nm. $\eta_0$=20meV. For comparison, $I_D$-$V_G$ curves with $\Sigma^R$=0 throughout the device is also presented. For all plots, a source Fermi level of 0.1 eV is used.*

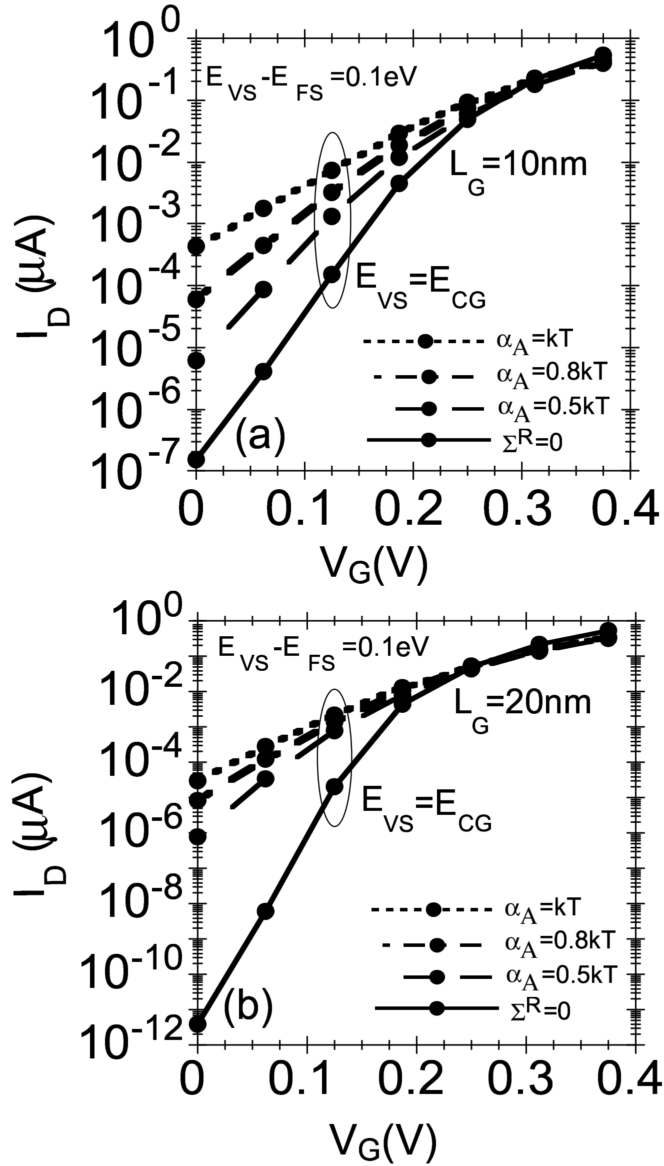



Figure 7.26: *$I_D$-$V_G$ characteristics on a log scale of InSb NW TFETs for a range of values for $\alpha_A$ with $\Sigma^R$ used in the source, source-to-channel, and the gate region for two gate lengths (a) 10 nm and (b) 20 nm. $\eta_0$=50meV. For comparison, $I_D$-$V_G$ curves with $\Sigma^R$=0 throughout the device is also presented. For all plots, a source Fermi level of 0.1 eV is used.*

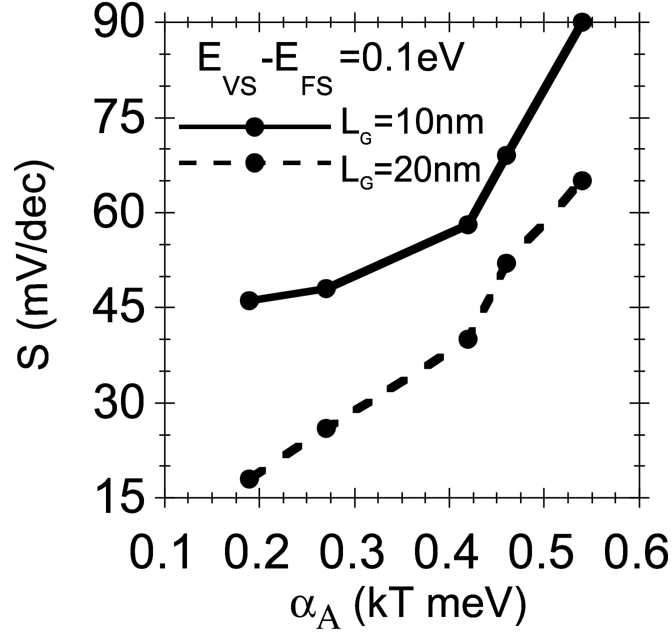



Figure 7.27: *Subthreshold swing, S as a function of $\alpha_A$ for InSb NW TFETs with $\Sigma^R$ used only in the source, source-to-channel, and the gate region for two gate lengths, 10 nm and 20 nm. $\eta_0$=20meV. A source Fermi level of 0.1 eV is used.*

from heavy doping, impurities and phonons. We have presented the analysis of the results based on numerical calculations for InSb NW TFETs. The analysis shows that band-tails resulting from the necessary heavy doping of the source of the TFETs are not a show-stopper.

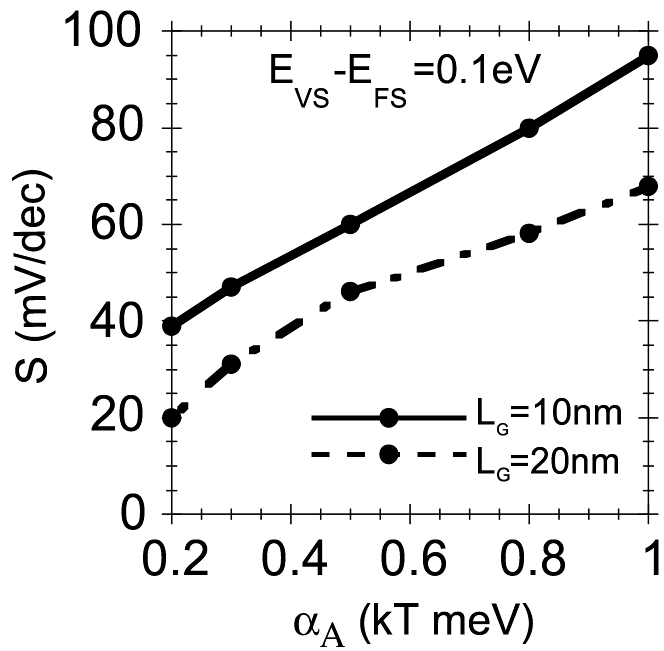



Figure 7.28: *Subthreshold swing, S as a function of $\alpha_A$ for InSb NW TFETs with $\Sigma^R$ used only in the source, source-to-channel, and the gate region for two gate lengths, 10 nm and 20 nm. $\eta_0$=50meV. A source Fermi level of 0.1 eV is used.*

# Chapter 8

# SUMMARY AND FUTURE DIRECTION

## 8.1 Summary

Lower dimensionality, such as in a semiconductor NW, creates quantum confinement that is beneficial to FET performance. At the nanometer scale, electrons in a semiconductor NW behave like waves in a microwave guide. At small enough diameter, there is, effectively, a single mode that is occupied with electrons contributing to the current. In this regime, the current can decouple from the charge density resulting in significant increases in speed and reduction in power. Two regimes of operation for conventional NW FETs, namely quantum and classical operation regime, have been extensively investigated in this dissertation. To perform this investigation, we have developed the following models/methods: (i) Full 3-D discretized 8-band $\mathbf{k} \cdot \mathbf{p}$ model, 2-band analytical model to calculate the electronic structures of materials, (ii) Numerical techniques based on non-equilibrium Green function (NEGF)-recursive Green function (RGF) method to calculate charge and transport involving both coherent and

incoherent scattering.

There are two primary questions regarding the performance of TFETs that have been addressed. Since the current is limited by inter-band tunneling, a process that reduces current exponentially with height or thickness of the tunnel barrier, it is not clear whether adequate current could be achieved in such devices. To answer this question, we have developed a method to quickly determine the maximum possible drive current as a function of semiconductor NW or CNT diameter. The analysis showed that 1 $\mu$A of current could be expected for the best semiconductor NWs, and 10 $\mu$A of current could be expected from a CNT with a similar bandgap and electric field. The method allows one to quickly compare the effect of different materials and geometries on the drive current and leakage current to determine if such materials and geometries are suitable candidates for high-performance TFETs

The second question is, 'How effectively cold can the injected electron distribution be?' The colder the distribution, the more the inverse subthreshold slope can be reduced beyond the 60 mV / decade thermal limit allowing reduced operating voltage and power. This question addresses the TFET's *raison d'être*. One of the fundamental physical mechanisms limiting the 'coldness' of the injected distribution is inelastic scattering of carriers due to the band-tail resulting from heavy doping and/or phonon. To address this issue, we have developed a generalized models based on NEGF-RGF method to investigate the effects of band-tails on the subthreshold characteristics of NW TFETs. The models can calculate the incoherent transport properties in the present of elastic scattering. The method can also be extended to calculate the incoherent transport in the presence of inelastic scattering.

In summary, the major and critical findings of this dissertation may be summarized as follows:

- $n$-type devices with diameter $\leq$50nm operate in the quantum capacitance limit and the corresponding $p$-type devices operate in the classical capacitance limit.
- Currents are well-matched for $n$- and $p$-type devices.
- Charge and current are decoupled. Current can be maximized by maximizing doping in the source, by increasing number of conduction modes, and by decreasing carrier masses.
- Both the energy-delay product and power-delay product are minimum at diameter range of 10-40 nm.
- A carbon nanotube with the same bandgap as an InSb nanowire has 10$\times$ more tunnel current at the same electric field.
- Heavy doping and band-tails are not a show-stopper for tunneling transistors.

## 8.2 Future work

A list of possible future works, directly related to this research, is presented below:

- **Ternary and quaternary III-V materials for next generation electronic and optoelectronic devices:** A great deal of research work has recently been done on III-V materials. The advantage that III-V materials offer is the materials engineering for proper material choice for device applications. Ternary materials such as InGaAs is considered as a potential successor for next generation low power logic applications. Recently, there have been a surge of experimental works that have been published in IEDM, TED and EDL on ternary materials such as InGaAs showing their potential for low power logic. By appropriate alloying, there are other materials combination that can be designed

and investigated for device applications. The $\mathbf{k} \cdot \mathbf{p}$ model (detailed description is presented in Chapter 2) that we have developed is a key tool for modeling electronic structures of such direct gap materials. With some modifications, the $\mathbf{k} \cdot \mathbf{p}$ model can be used to calculate the electronic structures of such ternary and quaternary materials. Quaternary materials are, especially, considered for nanoscale optoelectronics applications.

- **Strained III-V materials for high-speed, low-power nanoscale devices:** In alloying various III-V materials to find an appropriate material choice for a particular device application, usually two or three materials are sandwiched together. Generally, then, there appears to be lattice mismatch between different materials. Depending on the mismatch, compressive or tensile strain can be developed in the interface of such materials. Strain is predicted to affect electronic properties of materials by changing the effective mass of carriers. For instance, strained nanowires offer higher electron mobility compared to their unstrained counterpart. Such bandstructure effects due to strain can be modeled using the $\mathbf{k} \cdot \mathbf{p}$ model that we have developed by modifying the Hamiltonian. There will be an additional strain Hamiltonian that should be calculated using the material parameters due to strain, and add it to the unstrained Hamiltonian, and therefore, calculate the electronic structures. It will interesting to compare the strained data to the unstrained one to see the effects on the device characteristics.

- **Application to quantum well (QW) FETs:** We have investigated one dimensional nanowires. It will be a matter of interest to investigate the two dimensional (QW) planar structures based on these materials. Planar structures are compatible with the state-of-the-art technology. Generally, there will be

bandstructure effects due to the degree of quantization, and accordingly, the electronic properties of materials will be different. The $\mathbf{k} \cdot \mathbf{p}$ model that we have developed can easily be modified to calculate the electronic structures of 2-D structures. Then the transport models can be utilized to calculate the device transport characteristics. It will be interesting to compare the device characteristics for QW FETs and NW FETs, and investigate the dimensional effects on the device metrics.

- **Numerical algorithm development for simulating realistically large devices:** In the 8-band $\mathbf{k} \cdot \mathbf{p}$ method, 8 atomic orbital-like states are considered as basis for calculating the electronic structures of III-V materials. In three dimensional grid, each grid point (atom) consists of 8 orbital. With the increase of the device dimension, depending on the grid size, the Hamiltonian can be as large as to be difficult to handle with the computational resources available. NEGF equations for charge and transport calculations require to inverse, factorize, and multiply such large matrices. It can take months to simulate such realistically large devices using the computer cluster available. The challenges with numerical calculations are briefly presented in Chapter 1, Section 1.5, page 8. A project can be taken focusing on developing numerical strategies and algorithm, and implementing them by efficient coding, and thereby calculating the device transport. Such an investigation will present the device transport data for realistically sized devices which can be used to theoretically support experimental works based on these materials.

- **Phonon scattering in NW devices:** In the second part of this dissertation, we have considered the effects of heavy doping on the subthreshold characteristics of NW TFETs based on III-V materials. We have developed a general

NEGF-RGF based method to include the effects of band-tails due to heavy doping in the source and/or channel, and thereby, investigate the effects of inelastic scattering within the coherent transport regime on the device subthreshold characteristics. The model can be extended and modified to take into account the effects of incoherent scattering and therefore the effects of phonon on the subthreshold behavior of NW TFETs. One of the physical mechanisms limiting the 'coldness' of the injected carrier distribution in TFET devices is phonon scattering - a process by which electrons transfer thermal energy to and from the nuclei of the crystal lattice. The NEGF-RGF model that we have developed can be modified to calculate the interband tunneling including phonon absorption and emission both in the tunnel barrier region and in the source and gate region. The theoretical approach involves the use of NEGF formalism, and the phonon scattering is treated in the self-consistent Born approximation. Such an investigation will determine the fundamental limit of the inverse subthreshold slope obtainable from interband tunneling cold carrier injection. Such a limit will be an important guide to researchers, designers, and the semiconductor industry for determining which future technologies to pursue.

- **Designing materials systems for efficient tunneling devices:** III-V based NWs have been investigated for TFET devices which offer the potential for next generation ultra-low power logic applications. One of the fundamental limit for TFET devices is the issue with obtaining sufficient drive current. Because drive current is achieved by tunneling of carriers through the barrier region, tunneling current will naturally be exponentially suppressed with the height and/or width of the tunneling barrier. In order to quickly determine how much drive current can be achieved from a particular device based on the materi-

als structures, we have developed a general method based on calculating the imaginary wave vector in the bandgap. The model suggests that it is possible to engineer the materials based on III-V group to achieve a 'barrier-free' tunneling. For instance, Sb-riched InSb behaves as a semi-metal. Therefore, it is possible to design semiconductor-metal-semiconductor structure (in stead of semiconductor-semiconductor structure) to achieve a nearly barrier-free tunneling which will give a sufficiently large drive current. The idea of barrier-free tunneling based on III-V materials is being motivated by a work recently published in APL based on all-carbon material.

# Bibliography


[1] M. A. Khayer and R. K. Lake, "Performance of n-type InSb and InAs nanowire field effect transistors," *IEEE Trans. Electron Devices*, vol. 55, no. 11, p. 2939, 2008.

[2] ——, "The quantum and classical capacitance limits of InSb and InAs nanowire FETs," *IEEE Trans. Electron Dev.*, vol. 56, no. 10, p. 2215, 2009.

[3] ——, "The quantum capacitance limit of high-speed, low-power InSb nanowire field effect transistors," *IEEE Int. Electron Devices Meeting Tech. Dig.*, p. 193, 2008.

[4] ——, "Drive currents and leakage currents in InSb and InAs nanowire and carbon nanotube band-to-band tunneling fets," *IEEE Electron Dev. Lett.*, vol. 30, no. 12, p. 1257, 2009.

[5] ——, "Diameter dependent performance of high-speed, low-power InAs nanowire field-effect trasistors," *J. Appl. Phys.*, vol. 107, p. 014502, 2010.

[6] ——, "Modeling and performance analysis of GaN nanowire field-effect trasistors and band-to-band tunnleing field-effect trasistors," *J. Appl. Phys.*, vol. 108, p. 104503, 2010.

[7] ——, "Performance analysis of InP nanowire band-to-band tunneling field-effect transistors," *Appl. Phys. Lett.*, vol. 95, p. 073504, 2009.

[8] D. Kahng and M. M. Atalla, "Silicon-silicon dioxide field induced surface devices," *Proc. IRE-AIEE Solid-State Devices Res. Conf. Pittsburgh, PA*, 1960.

[9] G. E. Moore, "Cramming more components onto integrated circuits," *Electronics*, p. 114, 1965.

[10] "International technology roadmap for semiconductors," *2008 Edition, www.itrs.net*, 2008.

[11] S. E. Thompson, R. S. Chau, T. Ghani, K. Mistry, S. Tyagi, and M. T. Bohr, "In search of "forever" continued transistor scaling one new material at a time," *IEEE Trans. on Semiconductor Manufacturing*, vol. 18, no. 1, p. 26, Feb. 2005.

[12] H. Wakabayashi, T. Ezaki, M. Hane, S. Yamagami, N. Ikarashi, K. Takeuchi, t. Yamamoto, T. Mogami, T. Ikezawa, T. Sakamoto, and H. Kawaura, "Transport properties of sub-10nm planar bulk CMOS devices," *IEEE Int. Electron Dev. Meeting Tech. Dig.*, p. 429, Dec. 2004.

[13] K. Uchida, J. Koga, and S. I. Takagi, "Experimental study on carrier transport mechanisms in double- and single-gate ultrathin-body MOSFETs - coulomb scattering, volume inversion, and $\delta$T-induced scattering," *IEEE Int. Electron Dev. Meeting Tech. Dig.*, Dec. 2003.

[14] R. S. Shenoy and K. C. Saraswat, "Optimization of extrinsic source/drain resistance in ultrathin body double-gate FETs," *IEEE Trans. Nanotechnol.*, vol. 2, no. 4, p. 265, 2003.

[15] B. S. Doyle, S. Datta, M. Doczy, S. Hareland, B. Jin, dJ. Kavalieros, T. Linton, A. Murthy, R. Rios, and R. Chau, "High performance fully-depleted tri-gate CMOS transistors," *IEEE Electron Dev. Lett.*, vol. 24, no. 4, p. 263, 2003.

[16] Y. K. Choi, T. J. King, and C. Hu, "Nanoscale CMOS spacer for FinFET for the terabit era," *IEEE Electron Dev. Lett.*, vol. 23, no. 1, p. 25, 2002.

[17] R. Chau, S. Datta, M. Doczy, B. Doyle, J. Kavalieros, and M. Metz, "High-$\kappa$/metal-gate stack and its MOSFET characteristics," *IEEE Electron Dev. Lett.*, vol. 25, no. 6, p. 408, 2004.

[18] T. N. Theis, "Beyond the silicon transistors: Personal observations," *Computing in Science and Engineering)*, p. 25, 2003.

[19] K. C. Saraswat, D. Kim, T. Krishnamohan, and A. Pethe, "Performance limitations of Si bulk CMOS and alternatives for future ULSI," *J. of the Indian Inst. of Sci.*, vol. 87, no. 3, p. 387, 2007.

[20] A. Khakifirooz and D. A. Antoniadis, "Transistor performance scaling: The role of virtual source velocity and its mobility dependence," *IEEE Int. Electron Dev. Meeting Tech. Dig.*, p. 667, 2006.

[21] P. G. Collins and P. Avouris, "Nanotubes for electronics," *Sci. Amer.*, p. 62, 2000.

[22] B. B. Aein, D. Datta, S. Salahuddin, and S. Datta, "Proposal for an all-spin logic device with built-in memory," *Nature Nanotech.*, vol. 5, p. 266, 2010.

[23] K. Yano, T. Ishii, T. Hashimotoi, E. Murai, and K. Seki, "Room-temperature single-electron memory," *IEEE Trans. Electron Dev.*, vol. 41, p. 1628, 1994.

[24] T. Ohshima and R. A. Klehi, "Operation of bistable phase-locked single-electron tunneling logic elements," *J. Appl. Phys.*, vol. 80, p. 912, 1996.

[25] R. T. Syme, “Modelling the effect of an impurity band on the I-V characteristics of resonant tunneling diodes,” *Semicon. Sci. Technol.*, vol. 7, p. 71, 1992.

[26] G. Neofotistos, R. Lake, and S. Datta, “Inelastic-scattering effects on single-barrier tunneling,” *Phys. Rev. B*, vol. 43, no. 3, p. 2442, 1990.

[27] L. Ding and P. Mazumder, “Noise tolerant quantum mos circuits using resonant tunneling diodes,” *IEEE Trans. Nanotech.*, vol. 3, no. 1, pp. 134 – 146, 2004.

[28] M. T. Bjork, B. J. Ohlsson, C. Thelander, A. I. Persson, K. Deppert, L. R. Wallenberg, and L. Samuelson, “Nanowire resonant tunneling diodes,” *Appl. Phys. Lett.*, vol. 81, no. 23, pp. 4458 – 4460, 2002.

[29] E. R. Brown, J. R. Söderstrom, C. D. Parker, L. J. Mahoney, K. M. Molvar, and T. C. McGill, “Oscillations up to 712 ghz in inas/alsb resonant-tunneling diodes,” *Appl. Phys. Lett.*, vol. 58, no. 20, pp. 2291–2293, 1991.

[30] E. R. Brown, C. D. Parker, and T. C. L. G. Sollner, “Effect of quasibound-state lifetime on the oscillation power of resonant tunneling diodes,” *Appl. Phys. Lett.*, vol. 54, no. 10, pp. 934–6, 1989.

[31] E. R. Brown, W. D. Goodhue, and T. C. L. G. Sollner, “Fundamental oscillations up to 200 ghz in resonant tunneling diodes and new estimates of their maximum oscillation frequency from stationary-state theory,,” *J. Appl. Phys.*, vol. 64, no. 3, pp. 1519–1528, 1988.

[32] P. Thompson, N. Jin, R. Yu, S.-Y. Chung, A. T. Rice, P. R. Berger, R. Lake, H. Chi, and D. S. Simons, “Mbe growth techniques to form si/sige resonant interband tunneling diodes with high peak current densities and high peak to valley current ratios,” in *Third International SiGe(C) Epitaxy and Heterostructure Conference Digest*, 2003.

[33] M. N. Feiginov, “Displacement currents and the real part of the high-frequency conductance of the resonant-tunneling diode,” *Appl. Phys. Lett.*, vol. 78, no. 21, pp. 3301–3, 2001.

[34] ——, “Effect of the coulomb interaction on the response time and impedance of the resonant-tunneling diodes,” *Appl. Phys. Lett.*, vol. 76, no. 20, pp. 2904 – 2906, 2000.

[35] Y. Hu and S. P. Stapleton, “Capacitance of a resonant tunneling diode,” *Jpn. J. Appl. Phys.*, vol. 31, no. 1, pp. 23 – 25, 1992.

[36] ——, “Quantum capacitance of resonant tunneling diode,” *Appl. Phys. Lett.*, vol. 58, no. 2, pp. 167 – 169, 1991.

[37] T. Wei, S. Stapleton, and O. Berolo, "Equivalent circuit and capacitance of double barrier resonant tunneling diode," *J. Appl. Phys.*, vol. 73, no. 2, pp. 829 – 834, 1993.

[38] N. Shimizu, T. Waho, and T. Ishibashi, "Capacitance anomoly in the negative differential resistance region of resonant tunneling diodes," *Jpn. J. Appl. Phys.*, vol. 36, no. 3B, pp. L330 – L333, 1997.

[39] T. Wei, S. Stapleton, and O. Berolo, "Scattering parameter measurements of resonant tunneling diodes up to 40 ghz," *IEEE Trans. Elect. Dev.*, vol. 42, no. 7, pp. 1378 – 1380, 1995.

[40] N. Jin, S.-Y. Chung, A. T. Rice, P. R. Berger, P. E. Thompson, C. Rivas, R. Lake, S. Sudirgo, J. J. Kempisty, B. Curanovic, S. L. Rommel, K. D. Hirschman, S. K. Kurinec, P. H. Chi, , and D. S. Simons, "Diffusion barrier cladding in si/sige resonant interband tunneling diodes and their patterned growth on pmos source/drain regions," *IEEE Trans. Elect. Dev.*, vol. 50, no. 9, pp. 1876 – 1884, 2003.

[41] K. Eberl, R. Duschl, O. G. Schmidt, U. Denker, and R. Haug, "Si-based resonant inter- and intraband tunneling diodes," *J. Crystal Growth*, vol. 227-228, pp. 770–776, 2001.

[42] S. L. Rommel, T. E. Dillon, P. R. Berger, R. Lake, P. E. Thompson, K. D. Hobart, A. C. Seabaugh, and D. S. Simons, "Si-based interband tunneling devices for high-speed logic and low power memory applications," in *1998 IEDM Technical Digest*. New York: IEEE, 1998, p. 1035.

[43] R. Lake, G. Klimeck, R. C. Bowen, C. Fernando, M. Leng, T. Moise, and Y. C. Kao, "Interface roughness, polar optical phonons, and the valley current of resonant tunneling diodes," *Superlatt. and Microstruct.*, vol. 20, no. 3, pp. 279–85, 1996.

[44] B. M. Borg, K. A. Dick, B. Ganjipour, M. E. Pistol, L. E. Wernersson, and C. Thelander, "InAs/GaSb heterostructure nanowires for tunnel field-effect transistors," *Nano Lett.*, vol. 10, p. 4080, 2010.

[45] M. T. Bjork, J. Knoch, H. Schmid, H. Riel, and W. Riess, "Silicon nanowire tunneling field-effect transistors," *Appl. Phys. Lett.*, vol. 92, p. 193504, 2008.

[46] K. R. Kim, D. H. Kim, K. W. Song, G. Baek, H. H. Kim, J. I. Huh, J. D. Lee, and B. G. Park, "Silicon-based field-induced band-to-band tunneling effect transistors," *IEEE Electron Dev. Lett.*, vol. 25, no. 6, p. 439, 2004.

[47] M. Luisier and G. Klimeck, “Simulation of nanowire tunneling transistors: From the wkb approximation to full-band phonon-assisted tunneling,” *J. Appl. Phys.*, vol. 107, p. 084507, 2010.

[48] ——, “Atomistic full-band design study of InAs band-to-band tunneling field-effect transistors,” *IEEE Electron Dev. Lett.*, vol. 30, no. 6, p. 602, 2009.

[49] S. O. Koswatta, M. Lundsrom, and D. E. Nikonov, “Influence of phonon scattering on the performance of $p$-$i$-$n$ band-to-band tunneling transistors,” *Appl. Phys. Lett.*, vol. 92, p. 043125, 2008.

[50] N. N. Mojumder and K. Roy, “Band-to-band tunneling ballistic nanowire FET: circuit-compatible device modeling and design of ultra-low-power digital circuits and memories,” *IEEE Trans. Electron Dev.*, vol. 56, no. 10, p. 2193, 2009.

[51] S. O. Koswatta, M. S. Lundstrom, and D. E. Nikonov, “Band-to-band tunneling in a carbon nanotube metal-oside-semiconductor field-effect transistor is dominated by phonon-assisted tunneling,” *Nano Lett.*, vol. 7, no. 5, p. 1160, 2007.

[52] 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, no. 19, p. 196805, 2004.

[53] J. Appenzeller, Y.-M. Lin, J. Knoch, Z. Chen, and P. Avouris, “Comparing carbon nanotube transistors - the ideal choice: a novel tunneling device design,” *IEEE Trans. Elect. Dev.*, vol. 52, no. 12, pp. 2568 – 2576, 2005.

[54] C. Rivas, R. Lake, G. Klimeck, W. R. Frensley, M. V. Fischetti, P. E. Thompson, S. L. Rommel, and P. R. Berger, “Full-band simulation of indirect phonon assisted tunneling in a silicon tunnel diode with delta-doped contacts,” *Appl. Phys. Lett.*, vol. 78, no. 6, pp. 814–6, 2001.

[55] E. Emberly and G. Kirczenov, “Principles for the design and operation of a molecular wire transistor,” *J. Appl. Phys.*, vol. 88, p. 5280, 2000.

[56] N. P. Guisinger, R. Basu, A. S. Baluch, and M. C. Hersam, “Molecular electronics on silicon: An ultrahigh vacuum scanning tunneling microscopy study,” *Annals of the New York Academy of Sciences*, vol. 1006, pp. 227 – 234, 2003. [Online]. Available: doi:10.1196/annals.1292.015

[57] Y. Xue and M. A. Ratner, “Microscopic theory of single-electron tunneling through molecular-assembled metallic nanoparticles,” *Phys. Rev. B*, vol. 68, no. 23, pp. 235 410/1–5, 2003.

[58] M. D. Ventra, S. T. Pantelides, and N. D. Lang, "The benzene molecule as a molecular resonant-tunneling transistor," *Appl. Phys. Lett.*, vol. 76, no. 23, pp. 3448 – 3450, 2000.

[59] A. Dehon, "Nanowire-based programmable architectures," *J. Emerg. Technol. Comput. Syst.*, vol. 1, no. 2, pp. 109–162, 2005.

[60] H. Yan, S. H. Park, G. Finkelstein, J. H. Reif, and T. H. LaBean, "Dna-templated self-assembly of protein arrays and highly conductive nanowires," *Science*, vol. 301, no. 5641, pp. 1882 – 1884, 2003.

[61] J. Xiang, W. Lu, Y. Hu, Y. Wu, H. Yan, and C. M. Lieber, "Ge/si nanowire heterostructures as high-performance field effect transistors," *Nature*, vol. 441, p. 489, 2006.

[62] V. Schmidt, H. Riel, S. Senz, S. Karg, and W. R. U. Gösele, "Realization of a silicon nanowire vertical surround-gate field-effect transistor," *Small*, vol. 2, no. 1, p. 85, 2006.

[63] S. A. Dayeh, C. Soci, P. K. L. Yu, E. T. Yu, and D. Wang, "Transport properties of inas nanowire field effect transistore: The effects of surface states," *J. Vac. Sci. Technol. B*, vol. 25, no. 4, p. 1432, 2007.

[64] S. A. Dayeh, D. P. R. Aplin, X. Zhou, P. K. L. Yu, E. T. Yu, and D. Wang, "High electron mobility inas nanowire field-effect transistors," *Small*, vol. 3, no. 2, p. 326, 2007.

[65] T. Bryllert, L. E. Wernersson, L. E. Fröberg, and L. Samuelson, "Vertical high-mobility wrap-gated inas nanowire transistor," *IEEE Electron Device Lett.*, vol. 27, no. 5, p. 323, 2006.

[66] E. Lind, A. I. Persson, and L. S. nd L. E. Wersersson, "Improved subthreshold slope in an inas nanowire heterostructure field-effect transistor," *Nano Lett.*, vol. 6, no. 9, p. 1842, 2006.

[67] Y. Huang, X. Duan, Y. Cui, L. J. Lauhon, K. H. Kim, and C. M. Lieber, "Logic gates and computation from assembled nanowire building blocks," *Science*, vol. 294, no. 5545, p. 1313, 2001.

[68] Z. Zhong, D. Wang, Y. Cui, M. W. Bockrath, and C. M. Lieber, "Nanowire crossbar arrays as address decoders for integrated nanosystems," *Science*, vol. 302, no. 5649, p. 1377, 2003.

[69] H. T. Ng, J. Han, T. Yamada, P. Nguyen, Y. P. Chen, and M. Meyyappan, "Single crystal nanowire vertical surround-gate field-effect transistor," *Nano lett*, vol. 4, no. 7, p. 1247, 2004.

[70] Q. Wang, A. Javey, R. Tu, H. Dai, H. H. Kim, P. Mclntyre, T. Krishnamohan, and K. Saraswat, "Germanium nanowire field-effect transistors with $SiO_2$ and high-k $HfO_2$ gate dielectrics," *Appl. Phys. Lett.*, vol. 83, no. 12, p. 2432, 2003.

[71] R. Chau, B. Doyle, S. Datta, J. Kavalieros, and K. Zhang, "Integrated nanoelectronics for the future," *Nature*, vol. 6, p. 810, 2007.

[72] S. Datta, T. Ashley, J. Brask, L. Buckle, M. Doczy, M. Emeny, D. Hayes, K. Hilton, R. Jefferies, T. Martin, T. J. Phillips, D. Wallis, P. Wilding, and R. Chau, "85nm gate length enhancement and depletion mode InSb quantum well transistors for ultra high speed and very low power digital logic applications," *IEEE Int. Electron Devices Meeting Tech. Dig.*, 2005.

[73] T. Ashley, A. R. Barnes, L. Buckle, S. Datta, A. B. Dean, M. T. Emeny, M. Fearn, D. G. Hayes, K. P. Hilton, R. Jefferies, T. Martin, K. J. Nash, T. J. Phillips, W. H. A. Ting, and R. Chau, "InSb quantum well transistors for high-speed, low power applications," *IEEE Int. Electron Devices Meeting Tech. Dig.*, p. 751, 1997.

[74] R. Chau, S. Datta, M. Doczy, B. Doyle, B. Jin, J. Kavalieros, A. Majumder, M. Metz, and M. Radosavljevic, "Benchmarking nanotechnology for high-performance and low-power logic transistor applications," *IEEE Trans. Nanotechnol.*, vol. 4, no. 2, p. 153, 2005.

[75] M. Radosavljevic, T. Ashley, A. Andreev, S. D. Coomber, G. Dewey, M. T. Emeny, M. Fearn, D. G. Hayes, K. P. Hilton, M. K. Hudait, R. Jefferies, T. Martin, R. Pillarisetty, W. Rachmady, T. Rakshit, S. J. Smith, M. J. Uren, D. J. Wallis, P. J. Wilding, and R. Chau, "High-performance 40nm gate length insb p-channel compressively strained quantum well field effect transistors for low-power ($v_{CC}$=0.5v) logic applications," *IEEE Int. Electron Devices Meeting Tech. Dig.*, p. 727, 2008.

[76] G. Dewey, M. K. Hudait, K. Lee, R. Pillarisetty, W. Rachmady, M. Radosavljevic, T. Rakshit, and R. Chau, "Carrier transport in high-mobility III-V quantum well transistors and performance impact for high-speed low-power logic applications," *IEEE Electron Dev. Lett.*, vol. 29, no. 10, p. 1094, 2008.

[77] Y. Xia, P. Yang, Y. Sun, Y. Wu, B. Mayers, B. Gates, Y. Yin, F. Kim, and H. Yan, "One dimensional nanostructures: synthesis, characterization, and applications," *Adv. Mater.*, vol. 15, no. 5, p. 353, 2003.

[78] X. Duan, Y. Huang, Y. Cui, J. Wang, and C. M. Lieber, "Indium phosphide nanowires as building blocks for nanoscale electronic and optoelectronic devices," *Nature*, vol. 409, p. 66, 2001.

[79] M. Gudiksen, L. J. Lauhon, J. Wang, D. C. Smith, and C. M. Lieber, "Growth of nanowire superlattice structures for nanoscale photonics and electronics," *Nature*, vol. 415, no. 6872, p. 617, 2002.

[80] J. Wang, M. S. Gudiksen, X. Duan, Y. Cui, and C. M. Lieber, "Highly polarized photoluminescence and photodetection from single indium phosphide nanowires," *Science*, vol. 293, no. 5534, p. 1455, 2001.

[81] M. H. Huang, S. Mao, H. Feick, H. Yan, Y. Wu, H. Kind, E. Weber, R. Russo, and P. Yang, "Room-temperature utraviolet nanowire nanolasers," *Science*, vol. 292, no. 5523, p. 1897, 2001.

[82] X. Duan, Y. Huang, R. Agarwal, and C. M. Lieber, "Single-nanowire electrically driven lasers," *Nature*, vol. 421, no. 5523, 2003.

[83] Y. Cui, Q. Wei, H. Park, and C. M. Lieber, "Nanowire nanosensors for highly sensitive and selective detection of biological and chemical species," *Science*, vol. 293, no. 5533, p. 1289, 2001.

[84] C. Thelander, M. T. Björk, M. W. Larsson, A. E. Hansen, L. R. Wallenberg, and L. Samuelson, "Electron transport in inas nanowires and heterostructure nanowire devices," *Solid State Communications*, vol. 131, p. 573, 2004.

[85] Y. Huang, X. Duan, Y. Cui, and C. M. Lieber, "Gallium nitride nanowire nanodevices," *Nano Lett.*, vol. 2, p. 101, 2002.

[86] J. Goldgerber, D. Sirbuly, M. Law, and P. Yang, "Zno nanowire transistors," *J. Phys. Chem. B*, vol. 109, no. 1, p. 9, 2005.

[87] D. Wang, Y. Chang, Q. Wang, J. Cao, D. Farmer, R. Gordon, H. Dai, and K. Saraswat, "Surface chemistry and electrical properties of germanium nanowires," *J. Am. Chem. Soc.*, vol. 126, no. 37, p. 11589, 2004.

[88] L. J. Lauhon, M. S. Gudiksen, D. Wang, and C. M. Lieber, "Eptaxial core-shell and core-multishell nanowire heterostructures," *Nature*, vol. 420, no. 7, p. 57, 2002.

[89] S. M. Sze, *Physics of Semiconductors*, 2nd ed. New York: John Wiley and Sons, 1981.

[90] J. M. Woodall, J. L. Freeouf, G. D. Pettit, T. Jackson, and P. Kirchner, "Ohmic contacts to $n$-GaAs using graded band gap layers of $Ga_{1-x}In_xAs$ grown by molecular beam epitaxy," *J. Vac. Sci. Technol*, vol. 19, p. 626, 1981.

[91] C. Liu, T. L. Hughes, K. W. X. L. Qi, and S. C. Zhang, "Quantum spin hall effect in inverted type *ii* semiconductors," *arXiv:0801.2831v1[cond-mat.mes-hall]*, 2008.

[92] M. König, S. Wiedmann, C. Brüne, A. Roth, and H. Buhmann, "Quantum spin hall insulator state in **H**g**T**e quantum wells," *Science*, vol. 318, p. 766, 2007.

[93] S. A. Wolf, D. D. Awschalom, R. A. Buhrman, J. M. Daughton, S. V. Mòlnar, M. L. Roukers, A. Y. Chtchelkanova, and D. M. Treger, "Spintronics: A spin-based electronics vision for the future," *Science*, vol. 294, p. 1488, 2001.

[94] S. Murakami, N. Nagaisa, and S. C. Zhang, "Dissipationless quantum spin current at room temperature," *Science*, vol. 301, p. 1348, 2003.

[95] B. M. Borg, K. A. Dick, B. Ganjipour, M.-E. Pistol, L.-E. Wernersson, and C. Thelander, "InAs/GaSb heterostructure nanowires for tunnel field-effect transistors," *Nano Lett.*, vol. 10, no. 10, p. 4080, 2010.

[96] P. Y. Yu and M. Cardona, *Fundamentals of semiconductors.* Berlin: Springer-Verlag, 1999.

[97] 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, p. 405, 1974.

[98] J. M. Luttinger and W. Kohn, "Motion of electrons and holes in perturbed periodic fields," *Phys. Rev.*, vol. 97, no. 4, p. 869, 1955.

[99] E. O. Kane, *J. Phys. Chem. Solids*, vol. 1, p. 82, 1956.

[100] ——, *J. Phys. Chem. Solids*, vol. 1, p. 249, 1957.

[101] G. Klimeck, "NEMO 1-D: The first NEGF-based TCAD tool," *Simulation of Semiconductor Processes and Devices - SISPAD 2004*, p. 9, 2004.

[102] G. Klimeck, F. Oyafuso, T. B. Boykin, R. C. Bowen, and P. v. Allmen, "Development of a nanoelectronic 3-D (NEMO 3-D) simulator for multimillion atom simulations adn its application to alloyed quantum dots," *Computer Modeling in Engineering and Sciences (CMES)*, vol. 3, no. 5, p. 601, 2002.

[103] M. Luisier and G. Klimeck, "Atomistic full-band simulation of silicon nanowire transistors: Effects of electron-phonon scattering," *Phys. Rev. B*, vol. 80, p. 155430, 2009.

[104] M. A. Khayer and R. K. Lake, "Modeling and performance analysis of high-speed, low-power InAs nanowire field-effect transistors," *Physica Status Solidi C.*, vol. DOI 10.1002/pssc.200983879, p. 1, 2010.

[105] R. K. Lake and R. R. Pandey, Non-Equilibrium Green Functions in Electronic Device Modeling at http://arxiv.org/abs/cond-mat/0607219.

[106] R. Lake, G. Klimeck, R. C. Bowen, and D. Jovanovic, "Single and multiband modeling of quantum electron transport through layered semiconductor devices," *J. Appl. Phys.*, vol. 81, no. 12, pp. 7845–7869, 1997.

[107] R. Lake, D. Jovanovic, and C. Rivas, "Nonequilibrium green functions in semiconductor device modeling," in *Progress in Nonequilibrium Green Functions*, M. Bonitz and D. Semkat, Eds. New Jersey: World Scientific, 2003, pp. 143–158.

[108] M. A. Khayer and R. K. Lake, "Effects of heavy doping on the subthreshold characteristics of nanowire band-to-band tunneling transistors," *Unpublished.*

[109] I. Vurgaftman, J. R. Meyer, and L. R. Ram-Mohan, "Band parameters for iii-v compound semiconductors and their alloys," *J. Appl. Phys.*, vol. 89, no. 11, p. 5815, 2001.

[110] M. Noguchi, K. Hirakawa, and T. Ikoma, "Intrinsic electron accumulation layers on reconstructed clean inas (100) surfaces," *Phys. Rev. Lett.*, vol. 66, no. 17, p. 2243, 1991.

[111] J. Hu, X. Guan, D. Choi, J. S. Harris, K. Saraswat, and H. S. P. Wong, "Fermi level depinning for the design of iii-v fet source/drain contacts," *Int. Symp. on VLSI Technology, Systems, and Applications*, p. 123, 2009.

[112] D. Mamaluy, D. Vasileska, M. Sabathil, T. Zibold, and P. Vogl, "Contact block reduction method for ballistic transport and carrier densities of open nanostructures," *Phys. Rev. B*, vol. 71, p. 245321, 2005.

[113] J. M. Luttinger and W. Kohn, "Motion of electrons and holes in perturbed periodic fields," *Phys. Rev.*, vol. 97, no. 4, p. 869, 1955.

[114] P. C. Sercel and K. J. Vahalla, "Analytical formalism for determining quantum wire and quantum dot band structure in the multi-band envelope function approximation," *Phys. Rev. B*, vol. 42, p. 3690, 1990.

[115] D. Gershoni, C. H. Henry, and G. A. Baraff, "Calculating the optical properties of multidimensional heterostructures: application to the modeling of quaternary quantum well lasers," *IEEE J. Quantum Electron.*, vol. 29, no. 9, p. 2433, 1993.

[116] B. A. Foreman, "Elimination of spurious solutions from eight-band $\mathbf{k}\cdot\mathbf{p}$ theory," *Phys. Rev. B*, vol. 56, no. 20, p. 748, 1997.

[117] K. I. Kolokolov and C. Z. Ning, "$\mathbf{k}\cdot\mathbf{p}$ hamiltonian without spurious-state solutions," *Phys. Rev. B*, vol. 68, no. 16, p. 161308, 2003.

[118] D. Mamaluy, D. Vesileska, M. Sabathil, T. Zibold, and P. Vogi, "Contact block reduction method for ballistic transport and carrier densities of open nanostructures," *Phys. Rev. B*, vol. 71, no. 24, p. 245321, 2005.

[119] R. Eppenga, M. F. H. Schuurmans, and S. Colak, "New $\mathbf{k} \cdot \mathbf{p}$ theory for $\mathbf{G}$a$\mathbf{A}$s/$\mathbf{Ga_{1-x}Al_xAs}$-type quantum wells," *Phys. Rev. B*, vol. 36, no. 3, p. 1554, 1987.

[120] B. M. Askerov, *Electron Transport Phenomena in Semiconductors*, 2nd ed. Singapore: World Scientific Publishing, 1994.

[121] T. B. Boykin, "Generalized eigenproblem method for surface and interface states: the complex bands of gaas and alas," *Phys. Rev. B*, vol. 54, p. 8107, 1996.

[122] C. Rivas, R. Lake, W. R. Frensley, G. Klimeck, P. E. Thompson, , K. D. Hobart, S. L. Rommel, and P. R. Berger, "Full-band modeling of the excess current in a delta-doped silicon tunnel diode," *J. Appl. Phys.*, vol. 94, no. 8, pp. 5005–5013, 2003.

[123] R. Lake, G. Klimeck, R. C. Bowen, and D. Jovanovic, "Single and multiband modeling of quantum electron transport through layered semiconductor devices," *J. Appl. Phys.*, vol. 81, no. 12, p. 7845, 1997.

[124] A. Svizhenko, M. P. Anantram, T. R. Govindan, B. Biegel, and R. Venugopal, "Two-dimensional quantum mechanical modeling of nanotransistors," *J. Appl. Phys.*, vol. 91, no. 4, pp. 2343–2354, 2002.

[125] R. K. Lake and R. R. Pandey, *Non-Equilibrium Green Functions in Electronic Device Modeling.* Stevenson Ranch: American Scientific Publishers, 2006, vol. 3, pp. 409–443. [Online]. Available: http://arxiv.org/abs/cond-mat/0607219

[126] M. Galperin, S. Toledo, and A. Nitzan, "Numerical computation of tunneling fluxes," *J. Chem. Phys.*, vol. 117, pp. 10 817–10 826, 2002.

[127] M. P. L. Sancho, J. M. L. Sancho, and J. Rubio, "Highly convergent schemes for the calculation of bulk and surface green functions," *J. Phys. F*, vol. 15, pp. 851–858, 1985.

[128] K. Natori, "Ballistic metal-oxide-semiconductor field-effect transistors," *J. Appl. Phys.*, vol. 76, p. 4879, 1994.

[129] S. Datta, F. Assad, and M. S. Lundstrom, "The Si MOSFET from a transmission viewpoint," *Superlatt. Microstruct.*, vol. 23, p. 771, 1998.

[130] Z. B. Ren, R. Venugopal, S. Goasguen, S. Datta, and M. S. Lundstrom, “A two dimensional simulator for quantum transport in double-gate MOSFETs,” *IEEE Trans. Electron Dev.*, vol. 50, no. 9, p. 1914, 2003.

[131] R. Venugopal, M. Paulsson, S. Goasguen, S. Datta, and M. S. Lundstrom, “A simple quantum mechanical treatment of scattering in nanoscale transistors,” *J. Appl. Phys.*, vol. 93, no. 9, pp. 5613–5625, 2003.

[132] J. Knoch, B. Lengeler, and J. Appenzeller, “Quantum simulations of an ultrashort channel single-gated $n$-MOSFET on SOI,” *IEEE Trans. Electron Dev.*, vol. 49, no. 7, p. 1212, 2002.

[133] D. S. Fisher and P. A. Lee, “Relation between conductivity and transmission matrix,” *Phys. Rev. B*, vol. 23, no. 12, pp. 6851–4, 1981.

[134] C. Caroli, R. Combescot, P. Nozieres, and D. Saint-James, “Direct calculation of the tunneling current,” *J. Phys. C: Solid State Physics*, vol. 4, pp. 916–929, 1971.

[135] R. Lake, G. Klimeck, R. C. Bowen, and D. Jovanovic, “Single and multi-band modeling of quantum electron transport through layered semiconductor devices,” *J. Appl. Phys.*, vol. 81, no. 12, pp. 7845–7869, 1997.

[136] R. Lake, G. Klimeck, and R. C. Bowen *Modeling of Wide Cross-Sectional Area, 1-D, Steady-State High-bias Quantum Devices: Theory*, (preprint, April 1995) .

[137] R. Lake, in *Quantum Transport in Ultrasmall Devices*, edited by D. K. Ferry, H. L. Grubin, and C. Jacoboni (Plenum Press, New York, 1995), p. 521-524; R. Lake, G. Klimeck, and R. C. Bowen, *Modeling of Wide Cross-Sectional Area, 1-D, Steady-State, High-bias Quantum Devices: Theory*, (preprint) .

[138] D. C. Langreth, *Linear and Non-linear Electron Transport in Solids, NATO ASI Series, Vol. 17*, J. T. Devreese and E. van Doren, Eds. New York: Plenum, 1976.

[139] S. M. Sze, *Physics of Semiconductors*, 2nd ed. New York: John Wiley and Sons, 1981.

[140] J. M. Woodall, J. L. Freeouf, G. D. Pettit, T. Jackson, and P. Kirchner, “Ohmic contacts to $n$-GaAs using graded band gap layers of $Ga_{1-x}In_xAs$ grown by molecular beam epitaxy,” *J. Vac. Sci. Technol*, vol. 19, p. 626, 1981.

[141] E. Lind, A. I. Persson, L. Samuelson, and L. E. Wersersson, “Improved subthreshold slope in an InAs nanowire heterostructure field-effect transistor,” *Nano Lett.*, vol. 6, no. 9, p. 1842, 2006.

[142] S. A. Dayeh, D. P. R. Aplin, X. Zhou, P. K. L. Yu, E. T. Yu, and D. Wang, "High electron mobility InAs nanowire field-effect transistors," *Small*, vol. 3, no. 2, p. 326, 2007.

[143] T. Bryllert, L. E. Wernersson, L. E. Fröberg, and L. Samuelson, "Vertical high-mobility wrap-gated InAs nanowire transistor," *IEEE Electron Device Lett.*, vol. 27, no. 5, p. 323, 2006.

[144] C. Thelander, M. T. Björk, M. W. Larsson, A. E. Hansen, L. R. Wallenberg, and L. Samuelson, "Electron transport in InAs nanowires and heterostructure nanowire devices," *Solid State Communications*, vol. 131, p. 573, 2004.

[145] H. T. Ng, J. Han, T. Yamada, P. Nguyen, Y. P. Chen, and M. Meyyappan, "Single crystal nanowire vertical surround-gate field-effect transistor," *Nano lett*, vol. 4, no. 7, p. 1247, 2004.

[146] J. Appenzeller, J. Knoch, M. T. Björk, H. Riel, and W. Riess, "Toward nanowire electronics," *IEEE Trans. Elect. Dev.*, vol. 55, no. 11, pp. 2827 – 2845, 2008.

[147] J. Knoch, W. Riess, and J. Appenzeller, "Outperforming the conventional scaling rules in the quantum capacitance limit," *IEEE Electron Device Lett.*, vol. 29, no. 4, p. 372, 2008.

[148] S. Ramo, J. R. Whinnery, and T. V. Duzer, *Fields and Waves in Communication Electronics*, 3rd ed. New York: New York, NY:John Wiley and Sons Inc., 1994.

[149] A. Rahman, J. Guo, S. Datta, and M. S. Lundstrom, "Theory of ballistic nanotransistors," *IEEE Trans. Electron Devices*, vol. 50, no. 9, p. 1853, 2003.

[150] J. Wang and M. Lundstrom, "Ballistic transport in high electron mobility transistors," *IEEE Trans. Electron Devices*, vol. 50, no. 7, p. 1604, 2003.

[151] N. Li, E. S. Harmon, J. Hyland, D. B. Salzman, T. P. Ma, Y. Xuan, and P. D. Ye, "Properties of InAs metal-oxide-semiconductor structure with atomic-layer-deposited $Al_2O_3$ dielectric," *Appl. Phys. Lett.*, vol. 92, p. 143507, 2008.

[152] S. Luryi, "Quantum capacitance devices," *Appl. Phys. Lett.*, vol. 52, no. 16, p. 501, 1987.

[153] A. Rahman, G. Klimeck, and M. Lundstrom, "Novel channel materials for ballistic nanoscale mosfets - bandstructure effects," *IEEE Int. Electron Devices Meeting Tech. Dig.*, 2005.

[154] D. Mamaluy, D. Vesileska, M. Sabathil, T. Zibold, and P. Vogi, "Contact block reduction method for ballistic transport and carrier densities of open nanostructures," *Phys. Rev. B*, vol. 71, no. 24, p. 245321, 2005.

[155] B. A. Foreman, "Elimination of spurious solutions from eight-band $\mathbf{k}\cdot\mathbf{p}$ theory," *Phys. Rev. B*, vol. 56, no. 20, p. 748, 1997.

[156] K. I. Kolokolov and C. Z. Ning, "$\mathbf{k}\cdot\mathbf{p}$ hamiltonian without spurious-state solutions," *Phys. Rev. B*, vol. 68, no. 16, p. 161308, 2003.

[157] J. Adams, "Mudpack: multigrid fortran software for the efficient solution of linear elliptic partial differential equations," *Applied Math. and Comput.*, vol. 34, p. 113, 1989.

[158] N. A. Bruque, K. Alam, R. R. Pandey, R. K. Lake, J. P. Lewis, X. Wang, F. Liu, C. S. Ozkan, M. Ozkan, and K. L. Wang, "Self-assembled carbon nanotubes for electronic circuit and device applications," *J. Nanoelectronics Optoelectronics*, vol. 1, no. 1, pp. 74 – 81, 2006. [Online]. Available: http://dx.doi.org/10.1166/jno.2006.007

[159] P. J. Burke, "Carbon nanotube devices for ghz to thz applicatons," *Proc. SPIE*, vol. 5593, p. 52, 2004.

[160] Y. Zheng, C. Rivas, R. K. Lake, K. Alam, T. B. Boykin, and G. Klimeck, "Electronic properties of Si nanowires," *IEEE Trans. Electron Dev.*, vol. 52, no. 6, p. 1097, 2005.

[161] M. V. Fischetti, "Monte carlo simulation of transport in technologically significant semiconductors of the diamond and zinc-blende structures - part I: homogeneous transport," *IEEE Trans. Electron Devices*, vol. 38, no. 3, p. 634, 1991.

[162] Z. Zanolli, M.-E. Pistol, L. E. Fröberg, and L. Samuelson, "Quantum confinement effects in InAs-InP core-shell nanowires," *J. Phys.: Condens. Matter*, vol. 19, p. 295219, 2007.

[163] D. B. Suyatin, J. Sun, A. Fuhrer, D. Wallin, L. E. Fröberg, L. S. Karlsson, I. Maximov, L. R. Wallenberg, L. Samuelson, and H. Q. Xu, "Electrical properties of self-assembled branched InAs nanowire junctions," *Nano Lett.*, vol. 8, no. 5, p. 1100, 2008.

[164] A. C. Ford, J. C. Ho, Y.-L. Chueh, Y.-C. Tseng, Z. Fan, J. Guo, J. Bokor, and A. Javey, "Diameter-dependent electron mobility of InAs nanowires," *Nano Lett.*, vol. 9, no. 1, p. 360, 2009.

[165] S. A. dayeh, E. T. Yu, and D. Wang, "Transport coefficients of InAs nanowires as a function of diameter," *Small*, vol. 5, no. 1, p. 77, 2009.

[166] K. Alam, "Transport and performance of a gate all around InAs nanowire transistor," *Semiconductor Science Technol.*, vol. 24, p. 085003, 2009.

[167] J. Li, S. J. Cai, G. Z. Pan, Y. L. Chen, C. P. Wen, and K. L. Wang, "High breakdown voltage GaN HFET field plate," *Electronics Lett.*, vol. 37, no. 3, p. 196, 2001.

[168] W. Han, S. Fan, Q. Li, and Y. Hu, "Synthesis of gallium nitride nanorods through a carbon nanotube-confined reaction," *Science*, vol. 277, p. 1287, 1997.

[169] A. Motayed, M. Vaudin, A. V. Davydov, J. Melngailis, M. He, and S. N. Mohammad, "Diameter-dependent transport properties of gallium natride nanowire field effect transistors," *Appl. Phys. Lett.*, vol. 90, p. 043104, 2007.

[170] J. W. Lee, M. H. Ham, and J. M. Myoung, "Gate-controlled transport in gan nanowire devices with high-$k$ $Si_3N_4$ gate dielectrics," *Solid State Communications*, vol. 145, p. 327, 2008.

[171] C. Y. Chang, G. C. Chi, W. M. Wang, L. C. Chen, K. H. Chen, R. Ren, and S. J. Pearton, "Electrical transport properties of single GaN and InN nanowires," *J. Electronic Materials*, vol. 35, no. 4, p. 738, 2006.

[172] H. Y. Cha, H. Wu, M. Chandrashekhar, Y. C. Choi, S. Chae, G. Koley, and M. G. Spencer, "Fabrication and characterization of pre-aligned gallium nitride nanowire field-effect transistors," *Nanotechnology*, vol. 17, p. 1264, 2006.

[173] Y. Huang, X. Duan, Y. Cui, and C. M. Lieber, "Gallium nitride nanowire nanodevices," *Nano Lett.*, vol. 2, no. 2, p. 101, 2002.

[174] G. Chen, A. Kolmakov, Y. Zhang, M. Moskovits, R. Munden, M. A. Reed, G. Wang, D. Moses, and J. Zhang, "Current rectification in a single GaN nanowire with a well-defined *p-n* junction," *Appl. Phys. Lett.*, vol. 83, no. 8, p. 1578, 2003.

[175] A. Motayed, M. He, A. V. Davydov, J. Melngailis, and S. N. Mohammad, "Realization of reliable GaN nanowire transistors utilizing dielectrophoretic alignment technique," *J. Appl. Phys.*, vol. 100, p. 114310, 2006.

[176] H. Wu, H. Y. Cha, M. Chandrashekhar, M. G. Spencer, and G. Koley, "High-yield GaN nanowire synthesis and field-effect transistor fabrication," *J. Electronic Materials*, vol. 35, no. 4, p. 670, 2006.

[177] V. O. Turin and A. A. Balandin, "Performance degradation of GaN field-effect transistors due to thermal boundary resistance at GaN/substrate interface," *Electronics Lett.*, vol. 40, no. 1, 2004.

[178] C. Lian, H. G. Xing, Y. C. Chang, and N. Fichtenbaum, "Electrical transport properties of wafer-fused *p*-GaAs/*n*-GaN heterojunctions," *Appl. Phys. Lett.*, vol. 93, p. 112103, 2008.

[179] K. A. Abdullah, M. J. Abdullah, F. K. Yam, and Z. Hassan, "Electrical characteristics of GaN-based metal-oxide-semiconductor structures," *Microelectronic Eng.*, vol. 81, p. 201, 2005.

[180] S. Han, W. Jin, D. Zhang, T. Tang, C. Li, X. Liu, Z. Liu, B. Lei, and C. Zhou, "Photoconduction studies on gan nanowire transistors under uv and polarized uv illumination," *Chemical Physics Lett.*, vol. 389, p. 176, 2004.

[181] J. C. Johnson, H. J. Choi, K. P. Knutsen, R. D. Schaller, P. Yang, and R. J. Saykally, "Single gallium nitride nanowire lasers," *Nature Lett.*, vol. 1, p. 106, 2002.

[182] T. Zimmermann, D. Deen, Y. Cao, J. Simon, P. Fay, D. Jena, and H. G. Xing, "AlN/GaN insulated-gate HEMTs with 2.3 A/mm output current and 480 mS/mm transconductance," *IEEE Electron Dev. Lett.*, vol. 29, no. 7, p. 661, 2008.

[183] C. Y. Nam, P. Jaroenapibal, D. Tham, D. E. Luzzi, S. Evoy, and J. E. Fischer, "Diameter-dependent electromechanical properties of GaN nanowires," *Nano Lett.*, vol. 6, no. 2, p. 153, 2006.

[184] H. Y. Cha, H. Wu, S. Chae, and M. G. Spencer, "Gallium nitride nanowire nonvolatile memory device," *J. Appl. Phys.*, vol. 100, p. 024307, 2006.

[185] J. Appenzeller, J. Knoch, M. T. Björk, H. Riel, H. Schmid, and W. Riess, "Towards nanowire electronics," *IEEE Trans. Electron Dev.*, vol. 55, no. 11, p. 2827, 2008.

[186] C. Thelander, P. Agarwal, S. Brongersma, J. Eymery, L. F. Feiner, A. Forchel, M. Scheffler, W. Riess, B. J. O. ad U. Gösele, and L. Samuelson, "Nanowire-based one-dimensional electronics," *Mater. Today*, vol. 9, no. 10, p. 28, 2006.

[187] M. H. Ham, J. H. Choi, W. Hwang, C. Park, W. Y. Lee, and J. M. Myoung, "Contact characteristics in GaN nanowire devices," *Nanotechnology*, vol. 17, p. 2203, 2006.

[188] Y. Yoon and S. Salahuddin, "Barrier-free tunneling in a carbon heterojunction transistor," *Appl. Phys. Lett.*, vol. 97, no. 03, p. 033102, 2010.

[189] J. Appenzeller, Y. M. Lin, J. Knoch, Z. Chen, and P. Avouris, "Comparing carbon nanotube transistors - The ideal choice: A novel tunneling device design," *IEEE Trans. Electron Dev.*, vol. 52, no. 12, p. 2568, 2005.

[190] S. O. Koswatta, D. E. Nikonov, and M. S. Lundstrom, "Computational study of carbon nanotube *p*-*i*-*n* tunnel FETs," *IEEE IEDM Tech. Dig.*, p. 518, 2005.

[191] K. Ganapathi, Y. Yoon, and S. Salahuddin, "Analysis of InAs vertical and lateral band-to-band tunneling transistors: Leveraging vertical tunneling for improved performance," *Appl. Phys. Lett.*, vol. 97, no. 03, p. 033504, 2010.

[192] Y. Yoon and S. Salahuddin, "Inverse temperature dependence of subthreshold slope in graphene nanoribbon tunneling transistors," *Appl. Phys. Lett.*, vol. 96, no. 01, p. 013510, 2010.

[193] S. O. Koswatta, M. S. Lundstrom, and D. E. Nikonov, "Band-to-band tunneling in a carbon nanotube metal-oxide-semiconductor field-effect transistor is dominated by phonon-assisted tunneling," *Nano Letters*, vol. 7, no. 5, pp. 1160–1164, 2007. [Online]. Available: http://dx.doi.org/10.1021/nl062843f

[194] W. A. Harrison, "Tunneling from an independent-particle point of view," *Phys. Rev.*, vol. 123, no. 1, pp. 85 – 89, 1961.

[195] K. Alam and R. K. Lake, "Leakage and performance of zero-schottky-barrier carbon nanotube transistors," *J. Appl. Phys.*, vol. 98, no. 6, p. 064307, 2005.

[196] C. P. Auth and J. D. Plummer, "Scaling theory for cylindrical, fully-depleted, surrounding-gate mosfet's," *IEEE Elect. Dev. Lett.*, vol. 18, no. 2, pp. 74 – 76, 1997.

[197] A. Maharjan, K. Pemasiri, P. Kumar, A. Wade, L. M. Smith, H. E. Jackson, J. M. Yarrison-Rice, A. Kogan, S. Paiman, Q. Gao, H. H. Tan, and C. Jagadish, "Room temperature photocurrent spectroscopy of single zincblende and wurtzite InP nanowires," *Appl. Phys. Lett.*, vol. 94, pp. 193 115 – 193 118, 2009.

[198] T. Walter, R. Herberholz, C. Muller, and H. W. Schock, "Determination of defect distributions from admittance measurements and application to Cu(In,Ga)$Se_2$ based heterojuctions," *J. Appl. Phys.*, vol. 80, no. 8, p. 4411, 1996.

[199] E. O. Kane, "Thomas-Fermi approach to impure semiconductor band structures," *Phys. Rev.*, vol. 131, no. 1, p. 79, 1963.

[200] B. I. Halperin and M. Lax, "Impurity-band tails in the high-density limit: I. minimum counting methods," *Phys. Rev.*, vol. 148, no. 2, p. 722, 1966.

[201] R. A. Logan and A. G. Chynoweth, "Effect of degenerate semiconductor band structures on current-voltage characteristics of silicon tunnel diodes," *Phys. Rev.*, vol. 131, no. 1, p. 89, 1963.

[202] G. Neofotistos, R. Lake, and S. Datta, "Inelastic-scattering effects on single-barrier tunneling," *Phys. Rev. B*, vol. 43, no. 3, p. 2442, 1990.

[203] M. Lundstrom, *Fundamentals of carrier transport*. United Kingdom, 2nd edition: Cambridge university press, 2000.

[204] W. V. Schoenfeld, M. J. Antonell, and C. R. Abernathy, "Carbon doping of InSb using $CBr_4$ during growth by gas-source molecular beam epitaxy," *Appl. Phys. Lett.*, vol. 72, no. 10, p. 1235, 1998.