Nanowire Tunnel Field Effect Transistors: Prospects and Pitfalls

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

Doctor of Philosophy

in

Electrical Engineering

by

Somaia Sarwat Sylvia

March 2014

Dissertation Committee:
Dr. Roger K. Lake, Chairperson
Dr. Alexander A. Balandin
Dr. Jing Shi
The Dissertation of Somaia Sarwat Sylvia is approved:

__________________________________________

__________________________________________

__________________________________________

Committee Chairperson

University of California, Riverside
Acknowledgments

First of all, I would like to express my sincere gratitude to my supervisor Dr. Roger Lake for his constant guidance throughout my PhD years. I thank him for all that he taught me about quantum mechanics and NEGF. It is from him that I came to learn about how to be persistent in doing research, write a paper or present the results and above all, how to be a good researcher. Without his mentoring and support, this work would not have been possible. I also thank Dr. Khairul Alam and Dr. M. Abul Khayer for stretching out their helping hands whenever I needed.

Thanks to my group mates from LATTE for making a pleasant and friendly group to work with. Special thanks to Dr. Sonia Ahsan for all the good times together. I want to thank all the past and present members of UCR-Bangladesh group for bringing splendid enjoyment and pleasure to our life abroad.

I cannot thank enough my parents for what they have done to make my life wonderful. They made the toughest sacrifices so that I can fulfill my dreams. It is of no doubt that having them as parents is the best thing that happened to me. I want to express my profound thankfulness to my dearest husband for always being there by my side. With him, I had an amazing time throughout the graduate student years. I want to thank my elder sister (Soaiba Sarwat Synthia) and all the other members of my entire family for always wishing the very best for me. With deepest respect I remember those members of my family who are no longer here to rejoice the happy
moments of my reaching the milestone.

And finally, I want to thank almighty Allah for blessing me with all that I got in my life.

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

- IEEE Transactions on Electron Devices [1, 2]. Reprinted with permission from [1, 2]. © [2012] IEEE.

The co-author Roger K. Lake, listed in the above publications directed and supervised the research which forms the basis for this dissertation. The remaining co-authors listed provided technical expertise and support as collaborators. This work is supported in part by the Center on Functional Engineered Nano Architectonics (FENA) and the Center on Function Accelerated nanoMaterial Engineering (FAME), one of the six centers of STARnet, a Semiconductor Research Corporation program sponsored by MARCO and DARPA. I would like to acknowledge the computational resources provided by Information Technology at Purdue University, West Lafayette, IN.
To my family
The tunnel field effect transistor (TFET) has the potential to operate at lower voltages and lower power than the field effect transistor (FET). The TFET can circumvent the fundamental thermal limit of the inverse subthreshold slope ($S$) by exploiting interband tunneling of non-equilibrium "cold" carriers. The conduction mechanism in the TFET is governed by band-to-band tunneling which limits the drive current. TFETs built with III-V materials like InAs and InSb can produce enough tunneling current because of their small direct bandgap. Our simulation results show that although they require highly degenerate source doping to support the high electric fields in the tunnel region, the devices achieve minimum inverse subthreshold slopes of 30 mV/dec. In subthreshold, these devices experience both regimes of voltage-controlled tunneling and cold-carrier injection. Numerical results based on a discretized 8-band k.p model are compared to analytical WKB theory.

For both regular FETs and TFETs, direct channel tunneling dominates the leakage current when the physical gate length is reduced to 5 nm. Therefore, a survey of materials is performed to determine their ability to suppress the direct tunnel current
through a 5 nm barrier. The tunneling effective mass gives the best indication of
the relative size of the tunnel currents. Si gives the lowest overall tunnel current for
both the conduction and valence band and, therefore, it is the optimum choice for
suppressing tunnel current at the 5 nm scale.

Our numerical simulation shows that the finite number, random placement, and
discrete nature of the dopants in the source of an InAs nanowire (NW) TFET affect
both the mean value and the variance of the drive current and the inverse subthreshold
slope. The discrete doping model gives an average drive current and an inverse
subthreshold slope that are less than those predicted from the homogeneous doping
model. The doping density required to achieve a target drive current is higher in
the discrete doping model compared to the homogeneous doping model. The relative
variation in the ON current decreases as the average doping density and/or NW
diameter increases. For the largest 8 nm NW studied, the coefficient of variation
in the ON current is \(\sim 15\%\) at a doping density of 1.5 \(\times 10^{20}\) cm\(^{-3}\). Results from
full self-consistent non-equilibrium Green’s function calculations and semi-classical
calculations are compared.
Contents

Approval iii

Acknowledgments v

Dedication vi

Abstract vii

1 Introduction 1
   1.1 Motivation .............................................................. 1
   1.2 Background .......................................................... 3
   1.3 Definition of the Problems ......................................... 5
   1.4 Layout ................................................................. 6

2 Doping, Tunnel Barriers, and Cold Carriers in InAs and InSb Nanowire
   Tunnel Transistors ......................................................... 8
   2.1 Introduction ............................................................ 9
   2.2 WKB Theory of the Inverse Subthreshold Slope .................. 11
   2.3 Device Structure and Computational Approach .................... 15
   2.4 Results ............................................................... 17
   2.5 Conclusion ......................................................... 22
3 Material Selection for Minimizing Direct Tunneling in Nanowire Transistors

3.1 Introduction ................................................. 25

3.2 Computational Approach ................................. 26

3.3 Results ...................................................... 32

3.4 Conclusion .................................................. 40

4 Effect of Random, Discrete Source Dopant Distributions on Nanowire Tunnel FETs

4.1 Introduction .................................................. 43

4.2 Computational Approach ................................. 44

4.3 Results ...................................................... 50

4.4 Conclusion .................................................. 60

5 Conclusion .................................................... 62
### List of Tables

2.1 Length (in nm) of the drain underlaps for different diameters . . . . . 15

2.2 Doping required to achieve 1 $\mu$A of drive current and corresponding performance metrics for InAs and InSb NW-TFETs. . . . . . . . . 20

4.1 Comparison of the ON currents and the subthreshold slopes predicted by the continuum and the discrete doping models . . . . . . . . . . . . 51
List of Figures

2.1 (a) Schematic diagram for a p-i-n TFET. (b) Energy band diagrams in three modes of operation. In the off state (dotted black lines), the channel conduction band, $E_{c,g}$ is above the source valance band, $E_{v,s}$ (bands uncrossed) creating a large tunneling barrier for valance band electrons in the source. In subthreshold (dashed green line), when the bands are crossing, a channel opens for the electrons and $S$ can be below the thermal limit. In the on state (solid red line), $E_{v,s} - E_{c,g} > 0$ and the valance band states are aligned with the empty conduction band states enabling significant Zener tunneling of electrons through the barrier. ................................................................. 9

2.2 Transfer characteristics for (a) 4 (b) 6 (c) 8 and (d) 10 nm InAs NW-TFETs. Arrows indicate the direction of increasing source doping and numbers indicate the inverse subthreshold slope (mV/dec). The densities considered are: $1.5 \times 10^{19}$ cm$^{-3}$ (black line with diamonds), $5 \times 10^{19}$ cm$^{-3}$ (red line with squares), $1.5 \times 10^{20}$ cm$^{-3}$ (blue line with circles), $5 \times 10^{20}$ cm$^{-3}$ (green line with diagonal crosses). Bandgap, $E_g$ for each diameter is also shown. Inset of (a) compares output characteristics in the on state with (dashed line) and without (solid line) underlap for $N_A = 1.5 \times 10^{20}$ cm$^{-3}$. ...................................................... 18
2.3 Maximum on-currents at $V_{OD} = 0.2$ V as a function of doping for different diameter InAs and InSb NW-TFETs.

2.4 Band diagram for 4 nm InAs with (a,c) $N_A = 1.5 \times 10^{20}$ cm$^{-3}$ and (b,d) $N_A = 5 \times 10^{20}$ cm$^{-3}$. (a,b) are at the on state, $V_{OD} = 0.2$ V. (c,d) are at the off state, $V_{OD} = -0.4$ V. In all figures, the red (solid) line is the transmission coefficient. In (a,b), the green (dashed) line is the energy resolved current density. Blue (dashed) lines indicate the Fermi level positions. The transmission coefficients in (a) and (b) are plotted on the same scale so that they can be compared. The current densities in (a) and (b) are also plotted on the same scale. The transmission coefficients in (c) and (d) are plotted on the same scale.

2.5 Contributing terms to $S$ for InAs NW-TFET with $N_A = 1.5 \times 10^{20}$ cm$^{-3}$. $D_{II}$ calculated from Eq. (2.10) and $D_I$ calculated from Eq. (2.9) using numerical values for $\lambda$ and $\Delta$. Values are normalized to $1/V_T$ where $V_T$ is the thermal voltage, $k_B T/q$, at $T = 300$ K. The right axis of each plot labels the $S$ values calculated numerically (solid line with red diamonds) and analytically (solid line with orange squares) using the WKB approximation (Eq. (2.8)).

2.6 Transfer characteristics for (a) 4 (b) 6 (c) 8 and (d) 10 nm InSb NW-TFETs. Arrows indicate the direction of increasing source doping and numbers indicate the inverse subthreshold slope. The doping densities are mostly same as in Fig. 2.2: $5 \times 10^{18}$ cm$^{-3}$(orange line with triangles), $1.5 \times 10^{19}$ cm$^{-3}$(black line with diamonds), $5 \times 10^{19}$ cm$^{-3}$(red line with squares), $1.5 \times 10^{20}$ cm$^{-3}$(blue line with circles), $5 \times 10^{20}$ cm$^{-3}$(green line with diagonal crosses). The bandgap $E_g$ for each diameter is also shown.
3.1 Potential energy profile for the NMOS. ................................................. 27
3.2 Bandstructure for a 6 nm (a) InAs (b) Si and (c) Ge nanowire. Positive and negative values of wavevectors indicate real and imaginary dispersions, respectively. The red solid line indicates the bands that were used for the analytical calculations. ............................... 29
3.3 Intraband tunneling currents calculated numerically as function of diameter for (a) electrons and (b) holes. ................................. 32
3.4 Tunneling effective mass for (a) electrons and (b) holes calculated from two-band model. ................................................................. 33
3.5 Bandgap of different materials as a function of diameter. .............. 34
3.6 Comparisons of the full-band, numerically calculated \(E - \kappa\) relations with those from Eq. (3.5) for (a) the conduction band and (b) the valance band for 6 nm Si and 10 nm InAs nanowires. The dashed lines are from the full-band, numerical calculations and the solid lines are from Eq. (3.5). |0.4| eV corresponds to the top of the barrier and 0 eV corresponds to the source conduction or valence band edge. The solid lines are calculated from Eq. (3.5) by fixing \(\kappa\) to the numerical value at the source Fermi energy (|0.1| eV). ........................................ 35
3.7 (a) Electron and (b) hole current density versus energy where 0 eV indicates the band edge and 0.4 eV is the top of the barrier for electrons and -0.4 eV is the top of the barrier for holes. All the NWs have the same diameter of 3 nm. .................................................. 37
3.8 Electron current calculated analytically and numerically. Transmission calculated from Eq. (3.1) is plotted with the dashed line and transmission calculated from Eq. (3.2) is plotted with the dotted line. Solid lines are from NEGF calculations. ........................................ 38
3.9 Hole current calculated analytically and numerically. Transmission calculated from Eq. (3.1) is plotted with the dashed line and transmission calculated from Eq. (3.2) is plotted with the dotted line. Solid lines are from NEGF calculations.

4.1 Device schematic of the p-i-n TFET.

4.2 Histogram showing the number of dopant atoms in the discrete doping region for a 4 nm NW, \( N_A = 1.5 \times 10^{20} \text{ cm}^{-3} \). The expected number of atoms in the 20 nm source is 48. The dashed line shows a fitted normal distribution for 30 randomly selected samples used for the device simulations in this work.

4.3 Semi-classical calculation of the drain current vs. gate voltage for a 4 nm TFET. The spectrum of grey solid lines show the results for 30 devices and the blue line with triangles is their average. The red line with circles is the result from the continuum doping model. The gate voltage range is the same for all devices. The 3 numbered plots correspond to the 3 band diagrams shown in Fig. 4.6b,c,d.

4.4 Band diagrams at the center of the nanowire in the ON state for a 4 nm TFET with continuous doping (thick red line) and the thirty samples with random doping (gray lines). \( E_c, E_v, E_f \) denote conduction, valence and Fermi energy, respectively. The dashed line at 25 nm indicates the source-channel interface. The inset shows correlation coefficient \( r \) between \( I_{ON} \) and \( N_{dop} \) as function of distance from source-channel interface. The maximum value of \( r \) is at 6 nm.
4.5 Band diagrams for a 4 nm TFET that illustrate the reduction in degeneracy resulting from the discrete doping model. For the solid blue lines, dopant atoms are placed every 10 nm at the center of the nanowire, so that the macroscopic doping density is $N_A = 6.25 \times 10^{18} \text{ cm}^{-3}$. For the dotted red lines, the source is homogeneously doped with the same $N_A$. The 3 figures are plotted at different positions inside the nanowire: (a) center (b) 1 nm away from the center (c) 2 nm away from the center.

4.6 Band diagrams and energy resolved current density in the ON state at the center of the nanowire for the (a) continuous doping case and three numbered transfer curves as marked in Fig. 4.3: (b) case I, (c) case II and (d) case III. Darker colors represent lower values of current. $E_c, E_v, E_f$ denote the conduction, valence and Fermi energies, respectively. In (b), (c) and (d), band lines from continuous doping model are shown in dashed white lines for comparison.

4.7 Percentage $CV$ for ON current ($CV_{I_{ON}}$) as function of doping for different NW sizes as indicated on the figure. The isolated red round data point corresponds to a set of samples with a fixed (expected) number of dopants in a 4 nm NW. The empty squares with dotted line are the variations obtained from NEGF self-consistent calculations labeled NSC. All other results shown are from the SC model.

4.8 (a) Percentage $CV$ for $N_{dop}$ and (b) Screening length inside the source ($\lambda_{source}$) vs. source doping. The screening length $\lambda$ is determined from the electrostatic potential resulting from the homogeneous doping model. It is the distance from the source-channel interface to the point where the potential is $kT$ above the potential of the source.
4.9 (a) Local density of states (LDOS) calculated along the cross-section of an ideal NW. (b) Band edges at the center of a 4 nm NW TFET with doping of $1.5 \times 10^{20}\text{cm}^{-3}$. Blue solid lines are obtained from semi-classical model and green dashed lines are obtained from NEGF self-consistent model. The dotted red line is the Fermi level position.
Chapter 1

Introduction

1.1 Motivation

In 1965, Gordon Moore noted that the number of transistors per unit area had doubled approximately every two years. His observation, later established as the famous “Moore’s Law”, has served as the guiding principle for the semiconductor chip industries for over 40 years. In accordance with Moore’s Law, metal oxide semiconductor (MOS) field effect transistors (FETs) have experienced steady, exponential downsizing of their critical dimensions, thereby increasing the number of transistors in a chip and enhancing the computing speed. As the size was reduced, the operating voltage was also reduced to maintain constant electric fields. However the voltage did not scale at the same rate as the transistor size. For conventional MOSFETS, there is a fundamental limit on the operating voltage set by the minimum inverse subthreshold slope ($S$) of 60 mV/dec at room temperature. A value of $S = 60$ mV/dec means that 60 mV of gate voltage is required to reduce the current by an order magnitude. If an ON/OFF current ratio of $10^4$ is desired, and if we assume that two-thirds of the maximum gate voltage is required to obtain a high ON current [3], the supply volt-
age cannot be reduced below 720 mV. Reducing the voltage range below this would either mean degraded ON current or increased leakage current. Thus, the voltage scaling cannot keep pace with the aggressive transistor scaling. As a result, IC power consumption is rising fast, and it has become a limiting factor to continued scaling.

In conventional MOSFETs, the minimum limit of 60 mV/dec results from the fact that the switching is controlled by carriers from a thermally broadened Fermi function. One approach to circumvent this fundamental limit is to use logic devices that operate out of thermal equilibrium with the environment [4–6]. The TFET is the physical implementation of this concept exploiting “cold” carriers to reduce switching energy and power [7–10].

Two different mechanisms act to circumvent the fundamental thermal limitation on $S$ in a TFET [2, 11]. One is the tunneling probability across the barrier for which the $S$ was shown to vary as the square of gate bias [11, 12]. However, the quadratic dependence indicates that $S$ will increase rapidly with gate bias. The second mechanism depends on the injection of non-equilibrium carriers around the source Fermi level where the high energy tail of the carrier distribution is cut-off by a band edge. In this case, $S$ varies approximately linearly with gate bias indicating that the low value can be sustained for a wider range of voltage [11]. The band pass filter like behavior in the source side of a Tunnel FET (TFET) was phrased as an “effective cooling of the Fermi function” [10,13,14], and the phenomenon came to be known as “cold carrier injection.”

The inverse subthreshold slope of a TFET can be reduced below the fundamental thermal limit of 60 mV/dec since it is not limited by the thermal voltage, $V_t$. The first experimental observation of an $S$ below 60 mV/dec was reported for a CNT-TFET by Appenzeller et al. [9] in 2004. Since then, many other groups [15–18] demonstrated that steep subthreshold switches can be built from materials like Si
and Ge which lie at the center of semiconductor technology. But even before the first experimental realization of a TFET, several other advantages were achieved from a tunneling device. In the TFET, the tunneling barrier width and junction fields vary as a function of gate bias [12,19]. The first measurements on a vertical Si Esaki tunneling transistor [20] confirmed fast switching due to an exponential increase in current with gate voltage. The source and drain contacts in TFETs are of opposite polarity and kept in reverse bias resulting in very small leakage current [19–21] and reduced short channel effects [22]. Since the tunnel barriers can be very small (less than 100 Å [7]), the gate lengths can be much smaller than those of regular MOSFETs. The smaller device dimension and the smaller capacitance will result in higher switching speed [23].

So far, Tunnel FETs have shown promising characteristics for use as future alternate logic devices and they have been listed as one of the emerging research devices by the International Technology Roadmap for Semiconductors (ITRS) [24]. However, there are practical design issues that need consideration before making chips out of TFETs. Therefore, in accordance with the growing interest for using the TFET as the next generation switch, this dissertation is focused on the quantum mechanical simulation of TFETs at the nanometer scale to address some of the critical design issues for their practical implementation.

1.2 Background

In order to maximize the performance of a TFET, the channel potential must be tightly controlled by the gate potential [8]. This means that the conduction (or valence) band in the channel can be pulled below (or lifted above) the valence (or conduction) band in the source with a very small change in gate voltage. This will result in a very sharp increase in current forcing $S$ to be close to zero. It is known that
a one-to-one gate control of the channel potential can be achieved if the TFET is operating in the quantum capacitance limit meaning that $C_{ox} >> C_q$, where $C_{ox}$ and $C_q$ are the oxide capacitance and quantum capacitance, respectively. $C_q$ is proportional to the density of states in the channel [14] which is very small for a one dimensional material system. This implies that 1-D nanowires (NWs) will more readily achieve the quantum capacitance limit. The geometry of a NW allows for a gate-all-around (GAA) geometry which maximizes the electrostatic gate control of the channel. All these factors make NWs the ideal choice for implementing a TFET.

The selection of a suitable material is another area of research for TFETs. The drive current results from band-to-band tunneling through the tunnel barrier. Si and Ge based TFETs have demonstrated good subthreshold characteristics, but they have low drive currents because of their indirect bandgaps [15,25–27]. Large band-to-band tunneling is possible with a small effective mass $m^*$ and a small direct bandgap [8]. However, a small bandgap can also deteriorate the performance by enhancing off state leakage through thermal emission and direct tunneling through the channel. Finding the optimum material combinations, geometry, and doping densities that maximize drive current while minimizing leakage current is an active area of research.

Group III-V materials have gained much interest due to their small effective mass ($m^*$) and direct bandgap. With an n-type $In_{0.53}Ga_{0.47}As$ channel, the first III-V based TFET [28] beat all the previous records of ON currents while still having a good ON/OFF current ratio ($>10^4$). Since then, many other III-V compounds (InAs, GaSb, AlGaSb, InP) and different combinations of them in heterostructure TFETs have been studied [29–32]. Both lateral and vertical tunneling have been considered and have shown to perform well at just 0.5 V. Graphene based TFETs [33–35] with novel architectures [36–38] have also shown promise. Most recently, monolayer transition metal dichalcogenide based heterojunction TFETs have been
1.3 Definition of the Problems

It is well known that group III-V compounds can result in high ON currents due to their small, direct bandgap. Khayer et. al. [40] showed that in order to achieve a target 1 $\mu$A ON current, the source electric field will have to vary from $\sim 0.05$ Vnm$^{-1}$ to $\sim 0.37$ Vnm$^{-1}$ for an InSb NW as the diameter is reduced from 10 nm to 2 nm. For the same diameter range, the electric field requirement is $\sim 0.1$ Vnm$^{-1}$ to $\sim 0.45$ Vnm$^{-1}$ for an InAs NW. In order to achieve these high electric fields, heavily degenerate source doping is required which may hinder steep subthreshold switching [11]. Therefore, whether the InAs and InSb TFETs can supply enough ON current while maintaining a high ON/OFF ratio and low $S$ values remains unknown.

The issue of current leakage in the OFF state is also important. With the recent technological advancements, we are likely to reach the 10 nm process generation within a few years [41]. In fact, the ITRS projects the transistor length to be only 5.9 nm by the year 2026. But the questions are, “How far can we scale down?” and “What happens when the channel length is reduced down to 5 nm?” As the transistor channel lengths are scaled down below 10 nm, direct source-to-drain tunneling becomes the dominant leakage mechanism limiting scaling [42]. The question then arises, “How well can different material systems block the leakage?” This question applies to both FETs and TFETs.

The small volume of NW TFETs exacerbates the problem of statistical variability. In NW systems, there will be a countable number of dopant atoms near the source-channel interface. In TFETs, the current is exponentially dependent on the potential profile of the tunnel barrier which is determined by the source doping. Due to the
random nature of the dopants, the tunnel barrier and hence the current will vary from device to device. It is not known how large the current variations are or how the variations depend on NW diameter or average source doping.

The objective of this thesis is to look into the aforementioned questions using theoretical simulations. For calculation of the Hamiltonians, a full-band discretized k.p model [43–45] and tight binding model [46,47] have been used. A three dimensional (3D) finite difference solution of Poisson’s equation is used for the electrostatics. Transport is calculated coherently using the non-equilibrium Green’s function formalism (NEGF) implemented in a recursive algorithm [48]. Random doping effects are simulated with a self-consistent Schrödinger-Poisson solver. Other than these, several other analytical formulae have been used.

1.4 Layout

The rest of this thesis is organized as follows. In Chapter 2, simulation results are presented for InAs and InSb NW TFETs with varying source doping and diameter. At the required heavy doping necessary to produce sufficient drive current, the source is highly degenerate, and the source Fermi level can be several hundred meV below the source valence band edge. At these heavy dopings, the inverse subthreshold slopes can still be below 60 mV/dec. The subthreshold region is analyzed numerically and compared against the standard analytical theory.

In Chapter 3, a survey of materials in a nanowire geometry is performed to determine their ability to suppress the direct tunnel current through a 5 nm barrier. The materials investigated are InAs, InSb, InP, GaAs, GaN, Si, Ge, and carbon nanotubes. A semianalytic approach to calculating tunnel current is demonstrated, which requires considerably less computation than a full-band numerical calculation.
In Chapter 4, the effect of discrete, random dopants in the source of a NW TFET is analyzed while keeping the channel undoped. We perform a statistical study based on numerical simulations for TFETs treating the source dopants near the source-channel interface discretely. Since sufficient drive current is an important issue for TFETs, the effects of the random nature of the source dopant distribution is investigated in terms of the drive current for a variety of NW sizes and doping densities. The subthreshold behavior is discussed for a few exemplary cases.

Finally, conclusions of this dissertation are given in Chapter 5.
Chapter 2

Doping, Tunnel Barriers, and Cold Carriers in InAs and InSb Nanowire Tunnel Transistors

InAs and InSb nanowire tunnel field effect transistors require highly degenerate source doping to support the high electric fields in the tunnel region. For a target on-current of 1 $\mu$A, the source Fermi energy lies in the range of 0.1 - 0.22 eV below the valence band edge depending on the material and diameter. Despite the large degeneracy, the devices achieve minimum inverse subthreshold slopes of $\sim$ 30 mV/dec. In sub-threshold, these devices experience both regimes of ‘voltage-controlled tunneling’ and ‘cold-carrier injection’. The reduction of the inverse subthreshold slope from each of these two processes is quantified. Numerical results based on a discretized 8-band k-p model are compared to analytical WKB theory. The standard WKB theory gives good qualitative agreement with the full-band, numerical simulations.
2.1 Introduction

Tunnel field effect transistors (TFETs) offer a way to circumvent the fundamental limit on drain bias for conventional field effect transistors (FETs) set by the inverse subthreshold slope ($S$) of 60 mV/dec at room temperature [49]. The basic working principle for a TFET is illustrated in Fig. 2.1. Two mechanisms act to circumvent the 60 mV/dec limit [11,49,50]. The first is voltage control of the source-channel tunnel barrier. The gate bias controls the voltage of the source-channel junction biased in Zener breakdown. Increasing the electric field in the junction has an exponential effect on the tunnel current. We refer to this process as ‘voltage-controlled tunneling.’

The second process results from the injection of a non-equilibrium carrier distribution from the source into the channel. For an n-type channel, the valence band edge in the

![Figure 2.1: (a) Schematic diagram for a p-i-n TFET. (b) Energy band diagrams in three modes of operation. In the off state (dotted black lines), the channel conduction band, $E_{c,g}$, is above the source valance band, $E_{v,s}$ (bands uncrossed) creating a large tunneling barrier for valance band electrons in the source. In subthreshold (dashed green line), when the bands are crossing, a channel opens for the electrons and $S$ can be below the thermal limit. In the on state (solid red line), $E_{v,s} - E_{c,g} > 0$ and the valance band states are aligned with the empty conduction band states enabling significant Zener tunneling of electrons through the barrier.](image-url)
source cuts off the high energy thermal tail of the injected electron distribution. The low pass filter like behavior of the source injection provides an “effective cooling of the Fermi function” [10,13,14] and this is referred to as ‘cold-carrier injection’. The use of logic devices that operate out of thermal equilibrium with the environment is one of the suggested directions for reducing the minimum switching energy per bit [4–6]. The TFET is the physical implementation of this concept exploiting nonequilibrium ‘cold’ carriers to reduce switching energy and power.

To realize the device concept, a tight gate control over the channel potential is required [8,14,49]. Nanowires (NWs) offer the tightest gate control of the channel not only because of the proximity of the gate to the channel but also because nanowires allow access to the quantum capacitance limit [14]. InAs and InSb n-channel NW-TFETs are in the quantum capacitance limit for diameters ranging up to ~ 50 nm [44,51–54]. Their small bandgaps promote band to band tunneling for high drive current, and their bandgaps are widely adjustable as a function of NW diameter increasing to 0.8 - 0.9 eV for InSb and InAs NWs, respectively, as the diameters are reduced down to ~ 4 nm [44,55,56]. Furthermore, InAs NWs in particular, have been the focus of a long-term, sustained research effort [57]. Thus, the InAs or InSb materials appear to be the most promising candidates for implementation of a NW-TFET in a single material NW system. There are also many alternative source/channel materials and heterojunctions [33,58–62] and novel device structures [36,37] being investigated to boost the band to band (BTB) tunneling current in TFETs.

A semi-analytical approach to determine the source-channel electric field required to achieve a given drive current in a homojunction NW-TFET was described in Ref. [40]. For a 1 μA drive current in InAs and InSb NWs, the fields ranged from ~ 0.05 to 0.48 V/nm as diameters decreased from 10 nm to 2 nm. In the channel region,
the dielectric thickness and NW diameter determine the screening length that is on
the order of a few nanometers [40]. However, in the source, the high field must be
supported by high doping.

In this work, we vary the source doping and model its effect on the drive currents
and subthreshold currents in InAs and InSb NW-TFETs. At the required heavy
doping necessary to produce sufficient drive current, the source is highly degenerate,
and the source Fermi level can be several hundred meV below the source valence band
edge. At these heavy dopings, the inverse subthreshold slopes can still be below 60
mV/dec. The subthreshold region is analyzed in terms of the 2 processes described
above, gate-controlled tunneling and cold-carrier injection. Numerical results are
compared against the standard analytical theory.

In the next section, the standard theory of the inverse subthreshold slope based
on the WKB approximation is described and its predicted estimates are given. The
predictions are later used in Sec. 2.4 to compare with and analyze the numerical
results. Sec. 2.3 describes the numerical approach, and conclusions are given in Sec.
2.5.

2.2 WKB Theory of the Inverse Subthreshold Slope

The standard theory of the TFET inverse subthreshold slope starts from a coherent
tunneling expression with a WKB approximation for the transmission coefficient [11,
14, 49, 50]. We will consider an n-type NW-TFET in which the electrons are injected
from the valence band of the source into the conduction band of the channel. Then,
the coherent expression for the current is

\[ I_D = \frac{2q}{\hbar} \int_{E_{c,g}}^{E_{v,s}} dET(E) [f(E - \mu_s) - f(E - \mu_d)] \]  \hspace{1cm} (2.1)
where $E_{c,g}$ is the conduction band edge in the channel, $E_{v,s}$ is the valence band edge in the source, $f$ is the Fermi function, and $\mu_s$ and $\mu_d$ are the source and drain Fermi energies, respectively. The WKB expression for the transmission coefficient is

$$T = \exp \{ -\mathcal{E}_I / \mathcal{E} \} \quad (2.2)$$

where $\mathcal{E}$ is the electric field in the tunnel region and $\mathcal{E}_I$ is given by the integral of the smallest imaginary wavevector through the bandgap [40,63],

$$\mathcal{E}_I \equiv \frac{2}{q} \int_0^{E_G} dE' \kappa(E'). \quad (2.3)$$

$\mathcal{E}_I$ is often computed for a model triangular barrier using the effective mass as a fitting parameter giving a closed form expression of $\mathcal{E}_I = 4\sqrt{2m^*E_G^{3/2}}/(3qh)$ [49]. With the integrand of Eq. (2.3), $\kappa(E)$, obtained from a full-band model, we calculate $\mathcal{E}_I$ directly from Eq. (2.3) for different diameter nanowires as described in [40] where the values for $\mathcal{E}_I$ are also plotted.

In this WKB approximation, $T$ is independent of energy and can be pulled outside the integral in Eq. (2.1). Then, the integral of the Fermi functions can be explicitly evaluated. Since $\mu_d = \mu_s - qV_{DD}$, where $V_{DD}$ is the supply voltage, we set $f(E - \mu_d) = 0$ in the range of integration, and the current expression becomes,

$$I_D = \frac{2q}{h} e^{-\mathcal{E}_I / k_B T} \ln \left[ \frac{1 + e^{-(E_{c,g} - \mu_s)/k_B T}}{1 + e^{-(E_{v,s} - \mu_s)/k_B T}} \right]. \quad (2.4)$$

This equation is evaluated in 1D, where the effect of quantization is included in $\mathcal{E}_I$.

The inverse subthreshold slope is defined by

$$S \equiv \left[ \frac{d \log_{10}(I_D)}{dV_{GS}} \right]^{-1} = \ln(10) \times \left[ \frac{1}{I_D} \frac{dI_D}{dV_{GS}} \right]^{-1} \quad (2.5)$$
where $V_{GS}$ is the gate bias.

To evaluate the derivative of Eq. (2.4) with respect to $V_{GS}$, $\mathcal{E}$ and $E_{c,g}$ are written as linear functions of the gate voltage. The electric field in the junction is written as

$$\mathcal{E} = (E_G + \Delta)/(q\lambda) \quad (2.6)$$

where $E_G$ is the bandgap, $\lambda$ is the screening length, and $\Delta = E_{v,s} - E_{c,g} > 0$ is the tunneling window equal to the crossing of the bands. In the quantum capacitance limit, the change in $E_{c,g}$ with gate voltage is given by the gate control parameter $\alpha_g [52, 64, 65]$, i.e.

$$dE_{c,g}/dV_{GS} = -q\alpha_g \quad (2.7)$$

where, for a gate all-around NW FET or TFET, $0 < \alpha_g < 1.0$. For the rest of the discussion, we set $\alpha_g = 1$. To include $\alpha_g$, simply take the derived expressions for $S$ and divide by $\alpha_g$. Explicitly evaluating $S$ from (2.4) using (2.6) and (2.7) gives,

$$S = \ln(10)/[D_I + D_{II}] \quad (2.8)$$

where

$$D_I = q^2\lambda\mathcal{E}/(E_G + \Delta)^2, \quad (2.9)$$

and

$$D_{II} = \left\{ \frac{k_B T}{q} \left( 1 + e^{(E_{c,g} - \mu_s)/k_B T} \right) \ln \left[ \frac{1 + e^{-(E_{c,g} - \mu_s)/k_B T}}{1 + e^{-(E_{v,s} - \mu_s)/k_B T}} \right] \right\}^{-1}. \quad (2.10)$$

The two terms $D_I$ and $D_{II}$ represent the two different processes, ‘gate-controlled tunneling’ and ‘cold-carrier injection,’ respectively.

The second term, $D_{II}$, has two important limits. In the first limit, when $(E_{c,g} - \mu_s) >> k_B T$, $(E_{v,s} - \mu_s) >> k_B T$, and $(E_{v,s} - E_{c,g}) = \Delta >> k_B T$, $D_{II} \to q/k_B T$. In
this limit, Eq. (2.8) becomes

\[
S = \ln(10)/ \left[ \frac{\frac{q^2 \lambda \mathcal{E}_I}{(E_G + \Delta)^2} + \frac{q}{k_B T}}{q} \right].
\]  

(2.11)

If \( D_I \) is negligible, \( S \) approaches the usual thermal limit of an FET, \( S \to \ln(10) \times k_B T/q = 60 \text{ mV/dec} \) at \( T = 300 \text{ K} \). Changing only the third condition of the above three to \( (E_{v,s} - E_{c,g}) = \Delta << k_B T \), i.e. the bands are just about to uncross, then the second limit of \( D_{II} \) becomes \( D_{II} \to q/\Delta \). In this limit, Eq. (2.8) becomes

\[
S = \ln(10)/ \left[ \frac{\frac{q^2 \lambda \mathcal{E}_I}{(E_G + \Delta)^2} + \frac{q}{\Delta}}{q} \right].
\]  

(2.12)

Thus, as the bands uncross, \( \Delta \to 0, D_{II} \to \infty, \) and \( S \to 0 \). This is the limit when ‘cold carrier injection’ controls \( S \).

To determine which term dominates in Eq. (2.8), \( D_I \) or \( D_{II} \), the value of \( D_I \) is calculated for an 8 nm diameter InAs NW where \( \mathcal{E}_I = 0.45 \text{ V/nm} \) [40], \( E_G = 0.63 \text{ eV} \) [44], and from the numerical analysis, \( \lambda \approx 7 \text{ nm} \). Setting \( \Delta = 0 \), the maximum value for \( D_I \) is \( 8.0 = 0.2 \frac{q}{k_B T} (1/V) \). If \( D_{II} \) were negligible, ‘gate-controlled tunneling’ alone would give an inverse subthreshold slope of \( 5 \ln(10) k_B T/q = 300 \text{ mV/dec} \) at room temperature. This indicates that, of the two terms \( D_I \) and \( D_{II} \), the latter one will be responsible for reducing \( S \) significantly below 60 mV/dec, and this occurs when the second limit for \( D_{II} \) applies, \( \Delta << k_B T \), i.e. the bands are just about to uncross resulting in ‘cold carrier injection.’ If \( \Delta \) remains >> \( k_B T \) and the bands do not uncross, Eq. (2.11) applies, and reducing \( S \) below the ideal limit of \( \ln(10) \times k_B T/q \) depends on the size of \( D_I \), which is around \( 0.2q/k_B T \). In this highly degenerate limit,
the minimum room-temperature value for $S$ based on this standard WKB analysis is

$$S_{\text{min}} = \ln(10) \times \frac{k_B T}{q} \times \frac{1}{[1 + 0.2]} = 50 \text{ mV/dec.} \quad (2.13)$$

In this limit, only a small reduction ($\sim 0.8$) of $S$ below the ideal thermal limit can be expected.

These two different regimes of ‘gate-controlled tunneling’ and ‘cold-carrier injection’ will be analyzed for the InAs and InSb NW TFETs in Sec. 2.4 comparing numerical results with the analytical theory.

### 2.3 Device Structure and Computational Approach

The NW TFETs are [001] oriented InAs/InP and InSb/InP core-shell structures with square cross-sections and core diameters of 4, 6, 8 and 10 nm. All the NWs have a 2 nm shell surrounding the core. Each NW is doped p-i-n with the source p-type and the drain n-type. The length of the intrinsic channel and gate is 20 nm. A drain underlap region prevents back injection of holes from the drain in the off-state of the device. The length of this region is determined for an approximate on-off current ratio, $I_{\text{ON}}/I_{\text{OFF}} = 1 \times 10^6$, following the method given in Ref. [40]. The lengths of the underlaps are listed for different diameters in Table 2.1.

With a fixed drain doping of $1.5 \times 10^{19} \text{ cm}^{-3}$, the source doping, $N_A$, is varied from $5 \times 10^{18} \text{ cm}^{-3}$ to $5 \times 10^{20} \text{ cm}^{-3}$. The lower drain doping is chosen to reduce the
ambipolar behavior by providing a larger tunnel barrier for holes. The drain bias $V_{DD}$ is kept fixed at 0.6 V. A maximum gate overdrive $V_{OD} = V_{GS} - V_t = 0.2$ V is used for all devices where $V_t$ is the threshold voltage. The gate voltage is swept over 0.6 V such that $-0.4 \leq (V_{GS} - V_t) \leq 0.2$ V. The threshold voltage, $V_t$, is determined for each device by extrapolating the $I_D - V_{GS}$ curve down to $I_D = 0$.

The electrostatic potential within the device is calculated from a three-dimensional, finite difference solution of Poisson’s equation. Dirichlet boundary conditions are set at the metal-gate / InP interface. Von Neumann boundary conditions are used at all other exterior boundaries. The electron and hole densities are calculated semiclassically using the density-of-states that is numerically calculated from a discretized 8-band k-p model [44,45]. The source and drain Fermi levels are set at 0 and $-qV_{DD}$, respectively, and the Fermi level in the channel is set to an average of the two. The semiclassical approach is justified by the fact that, in the quantum capacitance limit, there is negligible screening in the channel, so that the potential in the channel is essentially independent of the charge in the channel. The converged potential is added to the diagonal elements of the Hamiltonian calculated using the discretized three dimensional 8-band k-p model as described in Ref. [44, 45]. The full-band k-p model takes care of the non-parabolicity in both the propagating and evanescent modes. Transmission and current are calculated using the non-equilibrium Green’s function (NEGF) formalism implemented in a recursive Green function algorithm [48]. The drain (right contact) self-energy, $\Sigma^R$, is set at the middle of the underlap region to avoid spurious reflection resulting from the finite width of the fundamental mode in the drain. Any electron that reaches this point is collected by the drain contact.
2.4 Results

To determine the effect of source doping on the drive current, the inverse subthreshold slopes, and the $I_{ON}/I_{OFF}$ current ratios, transfer characteristics for the NW-TFETs are computed numerically. The general principles governing the effect of doping on the drive currents and the $I_{ON}/I_{OFF}$ current ratios will be illustrated using the InAs transfer characteristics. The quantitative differences with the InSb transfer characteristics will be discussed afterwards.

The drain currents of the InAs NW-TFET are plotted versus $V_{OD}$ in Fig. 2.2. $V_{OD} = 0$ V corresponds to the threshold voltage, $V_t$. At threshold, the conduction band of the channel ($E_{c,g}$) is well below the Fermi level of the source ($\mu_s$). $\mu_s - E_{c,g}$ is approximately $5k_BT$ (at $T = 300$ K) for the majority of the InAs NWs. For the InSb NWs, there is more variation with $\mu_s - E_{c,g}$ ranging from 3-6 $k_BT$.

The effect of source doping on the on-current of the InAs NW-TFETs is given by the red lines with circles in Fig. 2.3. These values are taken from Fig. 2.2 at $V_{OD} = 0.2$ V. For all diameters, there is initially a strong dependence of the on-current on the source doping. As the doping increases beyond a certain diameter-dependent value, the doping dependence weakens. This can be explained with the help of Fig. 2.4.

In Fig. 2.4(a), a source doping of $1.5 \times 10^{20}$ cm$^{-3}$ gives a source Fermi level that lies 180 meV below the source valence band edge. The Fermi level is aligned with the thinnest part of the tunnel barrier and the peak in the transmission $T(E)$ (solid red curve). As a result, the energy distribution of the current (dashed green curve) peaks at the same energy as the transmission. When the doping is increased to $5 \times 10^{20}$ cm$^{-3}$ as shown in Fig. 2.4(b), the Fermi energy lies 360 meV below the source valence band edge, and it is now below the thinnest part of the tunnel barrier and the peak in the transmission curve. Electrons at an energy near the peak in the
Figure 2.2: Transfer characteristics for (a) 4 (b) 6 (c) 8 and (d) 10 nm InAs NW-TFETs. Arrows indicate the direction of increasing source doping and numbers indicate the inverse subthreshold slope (mV/dec). The densities considered are: 1.5 $\times$ 10$^{19}$ cm$^{-3}$ (black line with diamonds), 5 $\times$ 10$^{19}$ cm$^{-3}$ (red line with squares), 1.5 $\times$ 10$^{20}$ cm$^{-3}$ (blue line with circles), 5 $\times$ 10$^{20}$ cm$^{-3}$ (green line with diagonal crosses). Bandgap, $E_g$ for each diameter is also shown. Inset of (a) compares output characteristics in the on state with (dashed line) and without (solid line) underlap for $N_A = 1.5 \times 10^{20}$ cm$^{-3}$.

Transmission curve, being energetically far from the Fermi energy, contribute little to the current density (dashed green line). Thus, increasing the source doping beyond a certain point pushes the source Fermi energy below the transmission peak resulting in diminishing returns in terms of increased current drive.

Interpolating from the plots in Fig. 2.3, the dopings required to achieve target drive currents of 1 $\mu$A are shown in Table 2.2 for InAs. The corresponding degeneracy ($E_{v,s} - \mu_s$), minimum value of $S$, and $I_{ON}/I_{OFF}$ ratio are also shown. The drive currents per unit width for each diameter are also shown for benchmarking with other competing devices (see Fig. 3 in Ref. [49]).

While increasing doping increases the on-current by increasing the electric field, it can also increase the off-current by increasing the Fermi level degeneracy as illustrated
Figure 2.3: Maximum on-currents at $V_{OD} = 0.2$ V as a function of doping for different diameter InAs and InSb NW-TFETs.

In Figs. 2.4(c) and (d). The transmission curves (red solid lines) in both plots are plotted on the same scale. The transmission in the lower doping (c) is almost zero compared to the heavier doping (d). For the heavier doping, the bands never uncross. The channel conduction band in the off-state still resides below the source valence band ($E_{v,s} - E_{c,g} = \Delta = 0.13$ eV) allowing a window for the thermally energetic electrons to tunnel and contribute to the off-state leakage into channel. Since the bands never uncross, the values for the inverse subthreshold slope shown on the top curve of Fig. 2.2(a) remain close to and slightly below the ideal thermal limit as predicted by Eq. (2.11).

For the lower doping of $N_A = 1.5 \times 10^{20} \text{ cm}^{-3}$, the bands uncross at $V_{OD} \approx -0.35$ V in all four diameters of the InAs NWs. These devices exhibit the effects of both ‘voltage-controlled tunneling’ ($D_I$) and ‘cold-carrier injection’ ($D_{II}$). To understand
Figure 2.4: Band diagram for 4 nm InAs with (a,c) $N_A = 1.5 \times 10^{20}$ cm$^{-3}$ and (b,d) $N_A = 5 \times 10^{20}$ cm$^{-3}$. (a,b) are at the on state, $V_{\text{OD}} = 0.2$ V. (c,d) are at the off state, $V_{\text{OD}} = -0.4$ V. In all figures, the red (solid) line is the transmission coefficient. In (a,b), the green (dashed) line is the energy resolved current density. Blue (dashed) lines indicate the Fermi level positions. The transmission coefficients in (a) and (b) are plotted on the same scale so that they can be compared. The current densities in (a) and (b) are also plotted on the same scale. The transmission coefficients in (c) and (d) are plotted on the same scale.

Table 2.2: Doping required to achieve 1 $\mu$A of drive current and corresponding performance metrics for InAs and InSb NW-TFETs.

<table>
<thead>
<tr>
<th>Dia. (nm)</th>
<th>$I_{\text{ON}}$ ($\mu$A/$\mu$m)</th>
<th>$N_A$ (cm$^{-3}$) ($\times 10^{19}$)</th>
<th>Degeneracy (eV)</th>
<th>Min S (mV/dec)</th>
<th>$I_{\text{ON}}/I_{\text{OFF}}$ ($\times 10^8$)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>InAs</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>250</td>
<td>15</td>
<td>0.182</td>
<td>25</td>
<td>55</td>
</tr>
<tr>
<td>6</td>
<td>167</td>
<td>7.4</td>
<td>0.135</td>
<td>25</td>
<td>14</td>
</tr>
<tr>
<td>8</td>
<td>125</td>
<td>4.0</td>
<td>0.097</td>
<td>29</td>
<td>3.8</td>
</tr>
<tr>
<td>10</td>
<td>100</td>
<td>3.0</td>
<td>0.089</td>
<td>29</td>
<td>9.0</td>
</tr>
<tr>
<td><strong>InSb</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>250</td>
<td>10</td>
<td>0.215</td>
<td>26</td>
<td>6.0</td>
</tr>
<tr>
<td>6</td>
<td>167</td>
<td>4.0</td>
<td>0.151</td>
<td>27.5</td>
<td>47</td>
</tr>
<tr>
<td>8</td>
<td>125</td>
<td>2.0</td>
<td>0.111</td>
<td>27</td>
<td>14.5</td>
</tr>
<tr>
<td>10</td>
<td>100</td>
<td>1.14</td>
<td>0.085</td>
<td>28</td>
<td>26</td>
</tr>
</tbody>
</table>
the interplay of the two effects, the values of $D_I$ and $D_{II}$ are plotted in the subthreshold regime in Fig. 2.5. The values are plotted in units of $1/V_T$ where $V_T = k_BT/q$ is the thermal voltage at $T = 300$ K. $D_I$ is calculated from Eq. (2.9), and $D_{II}$ is calculated from Eq. (2.10) using numerical values for $E_{v,s}$, $E_{c,g}$, $E_G$, $\lambda$, $E_I$, and $\Delta$. The values of $S$ calculated analytically from Eq. (2.8) and numerically from the transfer curves of Fig. 2.2 are also plotted. As the gate voltage is reduced, the values for $D_{II}$ initially reach the first limit of $q/k_BT$ discussed in Sec. 2.2. As the bands start to uncross at $V_{OD} = -0.35$ V, $D_{II}$ approaches the second limit of $q/\Delta$ and becomes significantly larger than $q/k_BT$. This is the limit when ‘cold-carrier injection’ significantly reduces $S$ below the thermal limit. The value of $D_I$ resulting from ‘voltage-controlled tunneling’ is always small compared to the value of $D_{II}$, and it does not change much with gate bias. ‘Voltage controlled tunneling’ plays a small role in the initial part of the subthreshold region where $D_{II} \sim q/k_BT$. It reduces $S$ slightly below the thermal limit to $\sim 50$ mV/dec as predicted by Eq. (2.13). The large reduction in the inverse subthreshold slope results from ‘cold carrier injection’ when the bands uncross. As shown in Fig. 2.5, the WKB theory qualitatively reproduces the numerical values of $S$.

The transfer characteristics for the InSb NW-TFETs at different diameters and doping densities are plotted in Fig. 2.6. The density of states and bandgaps of the InSb NWs are less than those for the InAs NWs with the same diameter [44,52]. The smaller bandgap of InSb gives an increased on-current at a given doping density, and thus, the doping required to achieve 1 $\mu$A of drive current is less as shown in Table 2.2.

While increasing the drive current, the narrower bandgap of InSb also increases the off-current and $S$. However, for the diameters in the range of 4 through 10 nm and dopings corresponding to drive currents of 1 $\mu$A, the $I_{ON}/I_{OFF}$ ratios are greater
Figure 2.5: Contributing terms to S for InAs NW-TFET with $N_A = 1.5 \times 10^{20}$ cm$^{-3}$. $D_{II}$ calculated from Eq. (2.10) and $D_I$ calculated from Eq. (2.9) using numerical values for $\lambda$ and $\Delta$. Values are normalized to $1/V_T$ where $V_T$ is the thermal voltage, $k_BT/q$, at $T = 300$ K. The right axis of each plot labels the S values calculated numerically (solid line with red diamonds) and analytically (solid line with orange squares) using the WKB approximation (Eq. (2.8)).

than $1 \times 10^8$. Overall, for all diameters at their optimum doping, S is lower than the ideal thermally limited value of 60 mV/dec.

### 2.5 Conclusion

InAs and InSb NW-TFETs require highly degenerate source doping to support the high electric fields in the tunnel region. For a target on-current of 1 $\mu$A, the source Fermi energy lies in the range of 0.1 - 0.22 eV below the valence band edge depending on the material and diameter. In subthreshold, these devices experience both regimes of ‘voltage-controlled tunneling’ and ‘cold-carrier injection.’ While ‘voltage-controlled tunneling’ provides a small decrease in $S$, the large decrease below the thermal limit
Figure 2.6: Transfer characteristics for (a) 4 (b) 6 (c) 8 and (d) 10 nm InSb NW-TFETs. Arrows indicate the direction of increasing source doping and numbers indicate the inverse subthreshold slope. The doping densities are mostly same as in Fig. 2.2: $5 \times 10^{18}$ cm$^{-3}$ (orange line with triangles), $1.5 \times 10^{19}$ cm$^{-3}$ (black line with diamonds), $5 \times 10^{19}$ cm$^{-3}$ (red line with squares), $1.5 \times 10^{20}$ cm$^{-3}$ (blue line with circles), $5 \times 10^{20}$ cm$^{-3}$ (green line with diagonal crosses). The bandgap $E_g$ for each diameter is also shown.

is the result of ‘cold-carrier injection’ which occurs as the bands uncross. Since the bands do uncross in all of the devices achieving the target on-current of 1 $\mu A$ listed in Table 2.2, they all show a significant reduction in the subthreshold slopes below the ideal thermal limit even though the source doping is highly degenerate. The standard WKB theory gives good qualitative agreement with the full-band, numerical simulations.
Chapter 3

Material Selection for Minimizing Direct Tunneling in Nanowire Transistors

When the physical gate length is reduced to 5 nm, direct channel tunneling dominates the leakage current for both field effect transistors (FETs) and tunnel field effect transistors (TFETs). Therefore, a survey of materials in a nanowire (NW) geometry is performed to determine their ability to suppress the direct tunnel current through a 5 nm barrier. The materials investigated are InAs, InSb, InP, GaAs, GaN, Si, Ge and carbon nanotubes (CNTs). The tunneling effective mass gives the best indication of the relative size of the tunnel currents when comparing two different materials of any type. The indirect gap materials, Si and Ge, give the largest tunneling masses in the conduction band, and they give the smallest conduction band tunnel currents within the range of diameters considered. Si gives the lowest overall tunnel current for both the conduction and valence band and, therefore, it is the optimum choice for suppressing tunnel current at the 5 nm scale. A semi-analytic approach to calculating
tunnel current is demonstrated which requires considerably less computation than a full-band numerical calculation.

3.1 Introduction

Many different materials are being considered for field effect transistor (FET) applications. Examples include Si, Ge, III-Vs, carbon nanotubes (CNTs), and graphene [57, 60, 66–69]. For both standard FETs [42] and tunnel FETs (TFETs) [54, 70], direct tunneling through the channel limits the off-current as the channel lengths drop below 10 nm. Therefore, it is timely to assess and compare the ability of different channel materials to block the direct tunnel current at the sub-10 nm scale.

At this length scale, extremely tight gate control over the channel potential is required. Nanowires offer the tightest gate control of the channel not only because of the proximity of the gate to the channel, but also because nanowires allow access to the quantum capacitance limit [14]. Recently, promising results have been experimentally demonstrated for scalable nanowire FETs in developing complex nanoprocessors [71]. Therefore, the nanowire geometry is considered for a variety of materials and diameters for both n-type (NMOS) and p-type (PMOS) channels. The materials investigated are InAs, InSb, InP, GaAs, GaN, Si, Ge and CNTs. The primary purpose of this work is to survey a broad range of materials and diameters and compare their ability to block the tunnel current both in the conduction and valence bands. A secondary objective is to demonstrate a semi-analytical approach to calculating the tunnel current which requires considerably less computation than a full-band numerical calculation.

Previous studies have focused on issues of electrostatics and gate control of the channel which are clearly critical issues. These studies also focused on planar struc-
tures, and they used either effective mass models [42,72,73] or full-band models [74]. The most recent study compared 5 nm gate, n-type FETs of different geometries and materials including Si, carbon nanotubes (CNTs), graphene nanoribbons, and In$_{0.75}$Ga$_{0.25}$As, using atomistic, full-band models [75]. It was found that all materials could provide good transistor performance provided that two conditions were met. The first condition is that their bandgaps must be sufficiently large to block the interband tunneling current. The second condition is that, for materials with smaller effective masses, the gate must modulate the potential of the source and drain extensions increasing the effective gate length. As acknowledged in [75], the second condition is identical to using source and drain underlaps to effectively increase the channel region. It has been shown that long source and drain underlaps can block both the direct tunneling current and the interband tunneling current of a CNTFET, so that even with a 2 nm metal gate, good on-off current ratios can be achieved [76]. However, this is not the device geometry of interest now. In this paper, we answer the question, ‘How well do different materials block the direct tunneling current in both n-type and p-type nanowire FETs or CNTFETs with actual 5 nm channels?’

3.2 Computational Approach

To focus on the single issue of channel tunneling and to make a clearly-defined, even comparison across a wide range of materials, a uniform, square, 5 nm long tunnel-barrier is considered. The actual potential barrier in appropriately designed structures can closely approximate this shape [75]. As will be shown in Sec. 3.3, the currents for different materials differ by up to 6 orders of magnitude. While a potential discontinuity reduces transmission, the reduction is of order unity. Therefore, the abruptness of the barrier does not affect the material comparisons or final conclusions.
The III-V materials are simulated for 3, 6 and 10 nm diameters, and their Hamiltonian matrix elements are calculated using the discretized three dimensional 8-band k-p model as described in Ref. [44,45]. All the compound semiconductors are considered zinc blende and their material parameters are taken from Ref. [77]. The (100) Si and Ge nanowires are simulated atomistically using a $sp^3s^*d^5$ tight-binding model as implemented in the newly developed code NEMO5 [78]. Due to the heavier computational burden involved with such simulations, the H-passivated Si and Ge nanowires are simulated only for 3, 4, 5 and 6 nm diameters. Hamiltonians for (7,0), (10,0) and (25,0) zigzag CNTs are calculated using an empirical $\pi$-bond model [46].

The tunnel barrier and source Fermi level for an n-type channel are shown in Fig. 3.1. The barrier and Fermi level for p-type channels are the same with the sign of all energies reversed. The total barrier height is 0.4 eV, and the height above the source Fermi level is 0.3 eV. The Fermi level in the source side is set to 0.1 eV above the conduction band edge for NMOS and 0.1 eV below the valance band edge for PMOS. In the off-state of a digital FET, the drain bias is the supply voltage, $V_{DD}$, so injection from the drain can be ignored. Therefore, the drain Fermi function is set to 0 for the n-FETs and 1 for the p-FETs.

Intraband tunnel current through the potential barrier is calculated coherently using the non-equilibrium Green’s function (NEGF) formalism implemented in a re-
cursive Green function algorithm [48]. The temperature is \( T = 300 \text{ K} \). Full-band, self-consistent calculations including phonon scattering have been performed for Si NW FETs with channel lengths of 15 nm. In the off-state, the current calculated coherently exceeded the current calculated with phonon scattering by a ratio of 1.35 for a 3 nm diameter wire, and the trend indicated that the ratio decreased towards unity as the diameter increased [79]. For 5 nm channels, the coherent tunneling process will be exponentially enhanced and thus be the dominant tunneling mechanism pushing the ratio very close to unity justifying a coherent tunneling calculation.

For wider diameter nanowires, full-band calculations of the current can be quite numerically demanding. Therefore, we assess two semi-analytical approaches to calculating the intraband tunnel current in nanowires. The first method uses the standard analytical expression of transmission through a finite potential barrier [80],

\[
T_{SB}(E) = \frac{4k^2\kappa^2}{4k^2\kappa^2 + (k^2 + \kappa^2)^2 \sinh^2 \kappa d},
\]

where \( d \) is the barrier width, and \( k \) and \( \kappa \) are the magnitudes of the real and imaginary wavevectors, outside and inside of the barrier, respectively, at energy \( E \). The wavevectors are calculated numerically from the full-band models.

The imaginary wavevectors \( \kappa \) in the bandgap are the decay constants that have an exponential effect on the tunnel current, and, thus, they govern its order-of-magnitude [81]. The imaginary wavevectors are calculated following Ref. [82] for the Si and Ge nanowires and Ref. [83] for the rest of the materials. For direct gap materials, like the III-Vs and CNTs, one single imaginary band connecting the bottom of conduction band and top of valence band can be found as shown by the red solid line in Fig. 3.2(a) for InAs.

Si and Ge have an indirect gap bulk bandstructure that results in a considerably
Figure 3.2: Bandstructure for a 6 nm (a) InAs (b) Si and (c) Ge nanowire. Positive and negative values of wavevectors indicate real and imaginary dispersions, respectively. The red solid line indicates the bands that were used for the analytical calculations.
more complex bandstructure in the bandgap. In Fig. 3.2(b), the bandstructure for the 6 nm Si nanowire is plotted. Although the bandgap of the Si in nanowire form becomes direct \(^1\) there is no single imaginary band that connects the band edges. Therefore, the minimal values for \(\kappa\), as shown by the red solid line in Fig. 3.2(b), which give the least action for tunneling, are used for evaluating Eq. (3.1). Since the tunneling barrier height is 0.4 eV, only the upper imaginary branch contributes to the electron tunneling, and only the lower imaginary branch contributes to the hole tunneling. Thus, in contrast to \(\text{interband}\) tunneling (tunneling from the conduction to the valence band or vice versa), phonon assistance is not required.

The bandstructure for the 6 nm Ge nanowire is shown in Fig 3.2(c). The conduction band edge is at \(X\) in the one-dimensional Brillouin zone. This connects to the minimal imaginary band at the top of the bandgap. The values for \(k\) and \(\kappa\) used to evaluate Eq. (3.1) are the ones indicated in red (solid lines). As with Si, the minimal values for \(\kappa\) are used. However, in contrast to Si, the imaginary band connecting to the conduction band minimum crosses the imaginary band connecting the direct bandgap at \(\Gamma\) a few hundred meV below the bandedge. To move from one branch to another requires phonon assistance. Nevertheless, the coherent tunneling expression (3.1) for electron tunneling will be shown to give good agreement with the numerical calculations. For hole tunneling, the real \(k\) is taken from the light hole valence band since it connects to the direct-gap imaginary band at \(\Gamma\). Having chosen the real and complex bands for tunneling, \(k\) and \(\kappa\) are calculated at each energy and used in Eq. (3.1) for transmission.

The second semi-analytic method uses the WKB expression for tunneling in which

\(^1\)Even though Si is indirect gap in its bulk bandstructure, it is often observed as direct gap in NWs as found in [25, 84–86].
the energy dependent transmission coefficient $T_{WB}$ is

$$T_{WB}(E) = \exp\{-2\kappa d\}. \tag{3.2}$$

In this approach, the transmission is calculated only from the complex bandstructure in the bandgap. Even though this method has been successfully used previously to calculate the \textit{interband} tunneling current [25, 40, 87], we find that Eq. (3.1) provides better agreement with the numerical calculation of the \textit{intraband} tunneling current. The same imaginary bands as described earlier and shown in Fig. 3.2 are used to evaluate Eq. (3.2).

The tunneling current, $I_D$, is calculated from

$$I_D = M \frac{S q}{h} \int_{E_c}^{E_c + 0.4} dE T(E) f(E - E_{FS}) \tag{3.3}$$

for n-type channels or from

$$I_D = M \frac{S q}{h} \int_{E_v}^{E_v - 0.4} dE T(E) [1 - f(E - E_{FS})] \tag{3.4}$$

for p-type channels where $S$ is 1 or 2 depending on whether or not spin is explicitly included in the Hamiltonian. $M$ is the number of propagating modes, and it is automatically included in the Hamiltonian in the numerical (NEGF) calculations. For the analytical calculations, its value is set to 4 for the Si and the Ge conduction bands and to 1 for all other cases. $E_c$ is the conduction band edge, $E_v$ is the valence band edge. The transmission $T(E)$ is calculated either from Eqs. (3.1) or (3.2) or numerically from NEGF.

To focus solely on the tunnel current, all currents are calculated by integrating the transmission coefficient over the energy range from the band edge of the source.
Figure 3.3: Intraband tunneling currents calculated numerically as function of diameter for (a) electrons and (b) holes.

to the top of the barrier as given in Eqs. 3.3 and 3.4. This is sufficient to capture 84% of the total current for GaN and 99.8% for the other materials except for Si and Ge. These two materials block the tunneling current so effectively that a significant fraction (as high as 72%) of the total current can flow above the barrier.

3.3 Results

The NEGF calculated currents versus diameter for the different materials are shown in Fig. 3.3(a) and (b) for electrons and holes, respectively. As might be expected, the off current is slightly lower for PMOS than for NMOS for most of the materials since holes tend to have heavier effective masses. For the Ge nanowires, the currents are slightly larger for the PMOS devices because Ge has a small light-hole effective
mass. Si has the lowest tunnel-currents for both NMOS and PMOS except for the smallest diameter NMOS where Ge has slightly lower current. Electron and hole currents are same for the CNTs because their Hamiltonians are calculated from a $\pi$-bond model that produces symmetric bandstructure with respect to the middle of the bandgap. Also, it is worth mentioning that, for the largest diameter CNT, the bandgap ($E_g = 0.397$ eV) is smaller than the $0.4$ eV barrier height assumed here. This would mean a large current due to band-to-band tunneling through the valance band in the channel. For the range of diameters considered, all of the other materials have bandgaps that are greater than the barrier height.

Insight can be obtained from the tunneling effective masses, $m^*$, and the bandgaps
plotted in Figs. 3.4 and 3.5. The tunneling masses are calculated from

\[ m^* = \frac{\hbar^2 \kappa^2 E_g}{2(E_c - E)(E - E_v)} \]  

(3.5)

following Kane’s two-band model [88], where \( E_g \) and \( \hbar \) are the bandgap and the reduced Planck’s constant, respectively. Energy, \( E \), is taken to be at the source Fermi energy. Note that as \( E \) approaches a band-edge, such as, for example, \( E_c, E_g \) in the numerator cancels \( E - E_v \) in the denominator, and Eq. 3.5 reduces to the usual expression for parabolic dispersion, \( E_c - E = \frac{\hbar^2 \kappa^2}{2m^*} \), as it should. As shown in Fig. 3.6, the single value of \( m^* \) combined with Eq. (3.5) provides a good fit to the \( E - \kappa \) relations within the energy ranges of interest.

Comparing the tunneling masses in Fig. 3.4 with the tunneling currents in Fig. 3.3, it is clear that the order of the magnitudes of the tunneling currents correlate closely with the order of the magnitudes of the inverse of the tunneling masses in Fig. 3.4. In other words, the larger the tunneling mass, the smaller the tunneling current. The CNTs provide the one discrepancy in which the tunneling mass of the smallest CNT is heavier than the tunneling hole mass of GaAs and InP; however, the tunnel current is slightly larger.
Comparing the bandgaps of the direct gap semiconductors shown in Fig. 3.5, the order of the bandgaps correlate closely with the order of the tunneling masses (there is a small reversal for GaAs and InP). Thus, the bandgaps give a good indication of the ordering of the tunnel currents, i.e. the larger the bandgap, the smaller the intraband tunnel current. However, this trend cannot be used to compare the direct bandgap semiconductors with CNTs. For example, even though the bandgaps of InP and GaAs are greater than the smallest CNTs, the tunneling masses are less. Most notably, the indirect gap materials, Si and Ge, stand out amongst themselves. When comparing them with the other direct-gap materials, only the tunneling masses matter. Once the indirect gap materials are included, the bandgaps can no longer be used to predict which material gives the smallest intraband tunnel current. Finally, for all but the smallest diameters, Si has the heaviest tunneling mass and the lowest tunnel current both for electrons and holes.

The energy distribution of the electron and hole currents are shown in Fig. 3.7. The energy distribution of the smallest (GaN) and largest (InSb) direct gap currents
are shown along with the indirect gap (Si and Ge) currents. In the tunneling region, $|E| < 0.4$ eV, the conduction-band tunnel currents of Si and Ge are an order of magnitude smaller than the closest direct gap material, GaN. Immediately above the barrier, there is a resonance effect present in Si and Ge that is significantly stronger than that observed in any of the other direct gap materials. This could be related to the fact that the degeneracy of the conduction band edge in both Si and Ge (100) nanowires is 4, i.e. there are 4 propagating modes at the conduction band edges [84,89]. In contrast, the conduction band of the direct-gap materials consists of one propagating mode. At higher energies, the Si and Ge current distributions converge to those of GaN. For those three materials, many modes contribute to the current at the energies above the barrier. This is in contrast to InSb which is still single moded at those energies, and this accounts for the difference in the magnitude of the curves above the barrier. The valence band shows similar trends. The tunnel current of Si is smallest everywhere while Ge and GaN are comparable.

A comparison of the tunneling currents calculated from the semi-analytical transmission coefficients, Eqs. (3.1) and (3.2), and the NEGF calculation are shown in Figs. 3.8 and 3.9. The current calculated from the exact analytical expression, Eq. (3.1), gives good agreement with the full numerical calculation with the largest error being a factor of 4.8 for the Si valence band. Also, the error is usually large for the narrower gap materials, and this is not unexpected. For bandgaps less than 0.8 eV, electrons from the source are tunneling through the barrier below midgap. Thus, the intraband tunneling becomes a virtual interband tunneling process which is outside of the validity of Eq. (3.1). In our case, the 6 and 10 nm InAs and InSb nanowires have bandgaps less than 0.8 eV, and these are the two other cases which also do not give good agreement between the numerical calculations and Eq. (3.1).

The agreement with the WKB expression, Eq. (3.2), is generally worse for all
Figure 3.7: (a) Electron and (b) hole current density versus energy where 0 eV indicates the band edge and 0.4 eV is the top of the barrier for electrons and -0.4 eV is the top of the barrier for holes. All the NWs have the same diameter of 3 nm.
Figure 3.8: Electron current calculated analytically and numerically. Transmission calculated from Eq. (3.1) is plotted with the dashed line and transmission calculated from Eq. (3.2) is plotted with the dotted line. Solid lines are from NEGF calculations.

of the semiconductors except InSb. It does, however, capture the correct order-of-magnitude, and this is to be expected, since the order-of-magnitude is determined by the imaginary wavevector in the bandgap and the exponential decay factor [90].

The good agreement between the full numerical calculation and the semi-analytic calculation of the transmission coefficient given by Eq. (3.1) indicates that Eq. (3.1) can be used to calculate the tunnel current in larger diameter structures which are difficult to simulate fully numerically. Since we have demonstrated that Eq. (3.1) is valid for a single square barrier, it can also be used to obtain the transmission through a spatially varying barrier. Spatially varying barriers can be approximated as a series of thin square barriers, and the transmission for each thin square barrier can be cascaded to obtain the total transmission [80].

While this discussion has focused on *intraband* tunneling through the channel of a
Figure 3.9: Hole current calculated analytically and numerically. Transmission calculated from Eq. (3.1) is plotted with the dashed line and transmission calculated from Eq. (3.2) is plotted with the dotted line. Solid lines are from NEGF calculations.
standard FET in the off-state, the results are also relevent to the off-state leakage current in sub-10 nm channel TFETs. In standard FETs, the source-to-drain tunneling is an *intraband* tunneling process. For a TFET in the off-state with a p-type source, an electron in the source valence band undergoes *interband* tunneling through the source-to-channel, high-field region into an evanescent conduction band state a few hundred meV below the conduction band edge of the channel. At this energy, it can tunnel through a low-field, 10 nm channel into the drain. The process of tunneling through the low-field channel region is determined by the evanescent wavevector \( \kappa \) in the channel, and this \( \kappa \) is identical to the \( \kappa \) governing the source-to-drain tunneling process in a standard FET. The injection mechanisms are different, but the tunneling transport through the channel region is the same governed by the same evanescent wavevector \( \kappa \). Thus, the current and effective mass comparisons shown in Figs. 3.3 and 3.4 also indicate the effectiveness of different materials in blocking the off-state leakage current in sub-10 nm channel TFETs.

### 3.4 Conclusion

For a 5 nm physical gate length, direct channel tunneling dominates the leakage current for both FETs and TFETs. Therefore, we have performed a survey of materials to determine their ability to suppress the direct tunnel current through a 5 nm barrier. The tunneling effective mass gives the best indication of the relative size of the tunnel currents when comparing two different materials of any type. For direct gap III-V materials, the bandgaps are correlated with the tunneling masses, so the relative size of the bandgaps can also be used to determine the relative size of the tunnel currents. However, this only holds within the direct-gap, III-V subset of materials. The indirect gap materials, Si and Ge, give the largest tunneling masses in the con-
duction band, and they give the smallest conduction band tunnel currents within the range of diameters considered. Si gives the lowest overall tunnel current for both the conduction and valence band and, therefore, is the optimum choice for suppressing tunnel current at the 5 nm scale. The semi-analytic expression for the tunnel current given by Eq. (3.1) gives good agreement with the numerical result. Therefore, this expression can be used to calculate the tunnel current in wider diameter NWs which, using a full band model, are intractable or very difficult to simulate fully numerically. The transmission coefficients can also be cascaded to simulate spatially varying barriers.
Chapter 4

Effect of Random, Discrete Source Dopant Distributions on Nanowire Tunnel FETs

The finite number, random placement, and discrete nature of the dopants in the source of an InAs nanowire tunnel field effect transistor (TFET) affect both the mean value and the variance of the drive current and the inverse subthreshold slope. The discrete doping model gives an average drive current and an average inverse subthreshold slope that are less than those predicted from the homogeneous doping model. Both result from a reduction of the effective degeneracy in the source. The doping density required to achieve a target drive current is higher in the discrete doping model compared to the homogeneous doping model, but the difference reduces as the diameter increases. The relative variation in the ON current decreases as the average doping density and/or NW diameter increases. For the largest 8 nm diameter NW studied, the coefficient of variation in the ON current is \(~15\%\) at a doping density of \(1.5 \times 10^{20}\) cm\(^{-3}\). Results from full self-consistent non-equilibrium Green’s function calculations
and semi-classical calculations are compared.

4.1 Introduction

Tunnel Field Effect Transistors (TFETs) offer a way to circumvent the fundamental limit on drain bias set by the inverse subthreshold slope ($S$) of 60 mV/dec at room temperature [49, 50]. A tight gate control over the channel potential is required in order to realize the device concept of a TFET [8, 14, 49]. Because of their 1-D geometry, nanowires (NWs) can offer the tightest gate control of the channel by allowing access to the quantum capacitance limit [14]. InAs and InSb n-channel NW TFETs are in the quantum capacitance limit for diameters ranging up to $\sim$50 nm [44, 51–54]. While their small bandgaps promote band to band tunneling for high drive current, the bandgaps are widely adjustable as a function of NW diameter increasing to $0.8 - 0.9$ eV for InSb and InAs NWs, respectively, as the diameters are reduced down to $\sim$4 nm [44, 55, 56]. Moreover, InAs NWs have been the focus of a sustained research investigation [57].

High fields are required to produce sufficient interband tunnel drive current [40]. In the source, the high field must be supported by high doping both in heterostructure, broken gap TFETs [59] and in homojunction TFETs [2] as well as in planar TFETs [91]. The small size of the NW geometry will result in a countable number of dopant atoms within a screening length of the source-channel interface. Depending on the actual number and position of the dopants in the source, the local electric field defining the source-to-channel tunnel barrier will vary from device to device. Since the drive current of a TFET depends exponentially on the electric field in the tunnel region, this will result in a random variation of the drive current between nominally identical devices.
Effects of random dopant fluctuations in MOSFET channels have been extensively studied both theoretically [92–96] and experimentally [97,98]. For Si NW transistors, studies of random discrete dopant distributions in the source/drain regions using quantum transport simulations reveal the non-negligible effects of the variations [99–103]. Sensitivity to random dopants and line edge roughness for Tunnel FETs have been investigated using TCAD simulations in [104] [105].

In this work, we analyze the effect of discrete, random dopants in the source while keeping the channel of the NW TFET undoped. We perform a statistical study based on numerical simulations for TFETs treating the source dopants near the source-channel interface discretely. Since sufficient drive current is an important issue for TFETs, the effects of the random nature of the source dopant distribution is investigated in terms of the drive current for a variety of NW sizes and doping densities. The subthreshold behavior will be discussed for a few exemplary cases. The paper is organized as follows. In the next section the numerical approach is described. Results and discussions are presented in Sec. 4.3. Finally, Sec. 4.4 concludes the work.

4.2 Computational Approach

The NW TFETs are [001] oriented InAs/InP core-shell structures with square cross sections and core diameters, $d_{NW}$, of 4, 6 and 8 nm. All of the NWs have a 2 nm shell surrounding the core. A schematic diagram for the device is given in Fig. 4.1. Each NW is doped p-i-n with the source p-type and the drain n-type. The length of the simulated intrinsic channel and gate is 10 nm. The source, as seen in Fig. 4.1 is divided into two regions. A discrete doping region near the source-channel interface and a continuous doping region far away from the interface. The continuous doping
region provides a flat potential energy for insertion of the source self energy in the transport calculations. The discrete doping region, $L_s^{dis}$ is 20 nm for the 4 nm NWs and 10 nm for the 6 and 8 nm NWs. The smaller length for the thicker NWs is chosen to save computational time and is justified by the reduced screening lengths discussed in Sec. 4.3. The source doping, $N_A$, is varied from $4 \times 10^{19}$ cm$^{-3}$ to $5 \times 10^{20}$ cm$^{-3}$. The value of $N_A$ for each diameter is chosen to be around the values suggested in [2].

Randomness in both the placement and the number of dopant atoms is considered. The microscopic treatment of random dopant distribution is taken into account in the following manner. First, a uniform 1 nm square mesh grid is created inside the discrete doping region. If the total number of nodes in the discrete doping area is $N$, then at each node, the probability, $P$, of having a dopant atom is $\frac{N_A \cdot d_{NW}^2 \cdot L_s^{dis}}{N}$. At each node, $P$ is compared to a random number, $p$ ($0 \leq p \leq 1$) generated by the uniformly distributed pseudorandom number generator. If $P$ is greater than $p$ at a given node, that node is assigned one dopant atom. The doping concentration for each such node is $\frac{1}{v}$, where $v$ is the volume associated with the node. For 1000 samples, this procedure results in the dopant number distribution as shown by the histogram in Fig. 4.2. For a total of $N$ nodes in the discrete doping area, the probability of getting $n$ dopant atoms will be: $P^n(1 - P)^{(N-n)} \frac{N!}{(N-n)n!}$ resulting in the analytical probability distribution function (PDF) of Fig. 4.2 (blue solid line) which resembles a normal distribution (red dotted line). The above procedure was repeated using a grid spacing of 0.1 nm, and the results did not change.
Figure 4.2: Histogram showing the number of dopant atoms in the discrete doping region for a 4 nm NW, $N_A = 1.5 \times 10^{20}$ cm$^{-3}$. The expected number of atoms in the 20 nm source is 48. The dashed line shows a fitted normal distribution for 30 randomly selected samples used for the device simulations in this work.

After carrying out the above procedure for 1000 sample devices, 30 test devices were randomly selected to reduce the workload for further simulation and analysis. A normal PDF for the chosen samples is plotted in Fig. 4.2. While the normal distribution for the 1000 samples is centered around the expected mean number of atoms (48 in this case), the average of the 30 test samples is slightly higher (50 atoms). For these 30 sample devices, the electrostatic potential and current are calculated. This is repeated for each diameter and each target doping. A total of 330 devices are modeled as described below.

The threshold voltage, $V_t$, is determined for each diameter and target doping using the continuous doping model by extrapolating the $I_D - V_{GS}$ curve down to $I_D = 0$, where $I_D$ is the drain current and $V_{GS}$ is the gate bias. A maximum gate overdrive $V_{OD} = V_{GS} - V_t = 0.2$ V is used. The gate voltage is swept over 0.6 V such that $-0.4$ V $\leq (V_{GS} - V_t) \leq 0.2$ V. All 30 instances of the same device with random, discrete doping are compared within the same voltage range.

The Hamiltonian is constructed from the three-dimensionally discretized, 8-band,
The uniform grid spacing is 1 nm. The effect of the spatially varying electrostatic potential is included by shifting the diagonal elements of the Hamiltonian at each site.

The electrostatic potential within the device is calculated from a three-dimensional (3D) finite-difference solution of Poisson’s equation discretized on the same 1 nm grid as the \( k \cdot p \) Hamiltonian. Dirichlet boundary conditions are set at the metal-gate/InP interface. Von Neumann boundary conditions are used at all other exterior boundaries. The right boundary of the simulation domain is set in the channel 10 nm to the right of the source-gate interface. At this point, the von Neumann boundary conditions force the electric field along the \( z \)-direction in Fig. 4.1 to zero. This is adequate, since the source electric fields do not penetrate 10 nm into the channel, and our focus is on the effect of the discrete source dopants on the potential near the source-gate tunnel barrier.

In the channel, the charge density \( \rho \) is set to 0. Since InAs NWs operate in the quantum capacitance limit for a wide range of diameters [44,51–54], the channel potential is controlled by the gate. This assumption was verified by comparing the resulting channel potential to the potential resulting from fully self-consistent non-equilibrium calculations of \( \rho \) in the channel. The maximum change in the channel potential was 3.6 \% for a 4 nm NW.

Since the source doping is high, and the current is limited by interband tunneling, the electron and hole distributions in the source are assumed to be well-described by the equilibrium source Fermi distribution. This approximation has been used successfully in other tunneling devices [106–108]. The source is heavily degenerate \( p \)-type. Therefore, it is sufficient to take into account only the hole density and it is
calculated from

$$p_i = \int_{-10k_B T}^{(E_{v,i} + E_g/2)} dE N_i(E) [1 - f(E)]$$  \hspace{1cm} (4.1)$$

where $p_i$ is the hole density at site $i$, $E$ is the energy, $E_g$ is the band gap and $E_{v,i}$ is the valence band edge at site $i$. The source Fermi energy is the reference energy, so that, in the source, $E_F \equiv 0$. $N_i(E)$ is the density of states at site $i$, and it is calculated by two different methods.

The first method we denote as the NEGF self-consistent (NSC) method. In this method, the density of states in (4.1) is calculated from

$$N_i(E) = -\frac{1}{\pi \nu} \sum_\alpha \text{Im} G_{i,\alpha,i,\alpha}^R (E)$$  \hspace{1cm} (4.2)$$

where $G^R$ is the retarded Green function, $\alpha$ is the index of the four orbitals and the two spins at site $i$, and $\nu$ is the volume associated with site $i$. The NEGF calculation of $G^R$ and the solution of Poisson’s equation are iterated until convergence. To deal with the effects of energy quantization in the attractive potential wells for holes in the source, an imaginary potential $\Gamma = 20$ meV is added to the diagonal elements of the Hamiltonian. This procedure is similar to broadening the energy levels in an emitter quantum well [106]. An initial energy grid spacing of $\Gamma/10$ is used. A 5-point Gauss Lobatto integration method is used which refines the grid further [109]. The Newton-Raphson iteration scheme is used where the Jacobian matrix elements are calculated as $[\nabla \epsilon \nabla]_{i,j} + \frac{\delta \rho_i}{\delta V_j}$, where $V_j$ is the electrostatic potential at site $i$ and $\epsilon$ is the dielectric constant. A 3-point method is used for the second derivative. $\frac{\delta \rho_i}{\delta V_i}$ is approximated as $e \int_{-10k_B T}^{(E_{v,i} + E_g/2)} dE N_i(E) \frac{\partial f(E)}{\partial E}$ [76, 110]. The potential is updated using Anderson mixing of the potentials from the 10 previous iterations [111]. A mixing
A factor of 0.01 is used if the convergence error is larger than 100 mV, otherwise a factor of 0.1 is used. Convergence occurs when the maximum potential difference at any site between 2 iterations is less than 0.1 mV.

The second method we denote as the semi-classical (SC) method. In this method, the one-dimensional density of states of the ideal nanowire with no potential variation is calculated from

$$N(E) = \frac{-1}{\Omega \pi} \sum_i \sum_{\alpha} \text{Im} G_{i,\alpha;i,\alpha}^R(E)$$

where $G^R$ is again the retarded Green function calculated with the 8-band $k \cdot p$ model, $\alpha$ is the orbital and spin index, the sum over $i$ runs over all sites of the InAs wire, and $\Omega$ is the volume of the wire. This pre-calculated density-of-states is stored in an array. Local electrostatic potential variations $V_i$ are included by shifting the array at each site so that the density of states used in (4.1) is given by $N_i(E) = N(E - qV_i)$. At each iteration, the voltage $V_i$ is calculated as $[V] = [\nabla \epsilon \nabla]^{-1}[\rho]$. The new estimate for $[V]$ is obtained from the Anderson mixing scheme described above. The same convergence criterion is used for both the NSC and the SC calculations.

The semi-classical calculation is performed for all 330 devices. The NEGF self-consistent calculation is performed for 90 devices corresponding to 3 nominal dopant concentrations of the 4 nm NWs. The two methods result in quantitative differences, but the trends remain the same.

Once the potential is converged, the transmission and current are calculated. For the current calculation, the source (left contact) self energy, $\Sigma^L$ is placed in the continuous doping region where the potential energy is flat. It is calculated with the decimation method [112] using an imaginary convergence factor of 1 $\mu eV$ on the diagonal of the Hamiltonian. As discussed earlier and to reduce computation time for the 330 full-band NEGF simulations, the drain (right contact) self energy, $\Sigma^R$, is set.
at the middle of the channel region 10 nm from the source-gate interface. Thus, any electron that reaches the middle of the channel is collected by the drain. The current is determined by the potential profile of the source-to-channel, interband, tunnel barrier. The drain Fermi function is set equal to 0, i.e. there is no back-injection from the drain. The transmission and current are calculated with a recursive Green’s function algorithm as described in [76].

### 4.3 Results

The effect of discrete, random source dopant distributions on the average drive currents and their variation is analyzed for different doping concentrations and diameters. The average drive currents calculated for different doping concentrations and NW diameters are given in Table 4.1(a). As shown in Table 4.1(a), the average ON current shown in the fourth column resulting from the discrete doping model, $I_{ON}^{dis}$, is less than the ON current shown in the third column resulting from the continuum doping model, $I_{ON}^{cont}$. The percentage change in current ($= \frac{I_{ON}^{dis} - I_{ON}^{cont}}{I_{ON}^{cont}} \times 100\%$) can vary from -72% for a 4 nm diameter TFET to -3.7% for an 8 nm diameter. The three devices simulated with the NSC model show the same trends. The average current from the ensemble of devices with discrete, random doping is less than the current from the same device with homogeneous doping.

To analyze the effect of discrete, random source dopants on the subthreshold characteristics, transfer characteristics over the entire voltage range are computed numerically for 60 test devices with 4 and 6 nm diameters and $N_A = 1.5 \times 10^{20}$ cm$^{-3}$. The resulting minimum inverse subthreshold slopes $S$ are given in Table 4.1(b). For most instances in the ensemble, the minimum $S$ is less than that from the continuum doping model.
Table 4.1: Comparison of the ON currents and the subthreshold slopes predicted by the continuum and the discrete doping models

(a) ON currents. Results labeled NSC are from the NEGF self-consistent calculations. Otherwise, results are from the semi-classical calculations. The one entry labeled RP results from ensembles in which the position is random but the total number is fixed for all instances in the ensemble.

<table>
<thead>
<tr>
<th>Dia. (nm)</th>
<th>Doping (10^{19} \text{cm}^{-3})</th>
<th>Homogeneous</th>
<th>Discrete</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>5</td>
<td>0.18</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>1.4</td>
<td>0.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>0.68 (RP)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.25 (NSC)</td>
<td>0.50 (NSC)</td>
</tr>
<tr>
<td>30</td>
<td></td>
<td>3</td>
<td>1.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.85 (NSC)</td>
<td>1.46 (NSC)</td>
</tr>
<tr>
<td>50</td>
<td></td>
<td>4.0</td>
<td>3.23</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3.88 (NSC)</td>
<td>2.87 (NSC)</td>
</tr>
<tr>
<td>6</td>
<td>7.4</td>
<td>2.15</td>
<td>1.58</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>3.05</td>
<td>2.22</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>4.37</td>
<td>3.34</td>
</tr>
<tr>
<td>8</td>
<td>4</td>
<td>2.1</td>
<td>1.6</td>
</tr>
<tr>
<td></td>
<td>5.5</td>
<td>3.1</td>
<td>2.3</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>6.15</td>
<td>5.92</td>
</tr>
</tbody>
</table>

(b) Inverse subthreshold slopes; $N_A = 1.5 \times 10^{20}$ \text{cm}^{-3}$
Figure 4.3: Semi-classical calculation of the drain current vs. gate voltage for a 4 nm TFET. The spectrum of grey solid lines show the results for 30 devices and the blue line with triangles is their average. The red line with circles is the result from the continuum doping model. The gate voltage range is the same for all devices. The 3 numbered plots correspond to the 3 band diagrams shown in Fig. 4.6b,c,d.

The physics governing the effect of the discrete, random doping on the current will be illustrated using a 4 nm diameter TFET simulated with the semi-classical model followed by comparisons with the NEGF self-consistent model. An example set of current voltage characteristics is shown in Fig. 4.3. Almost all of the curves from the discrete doping model lie below the curve from the continuum doping model. The average ON current is 0.8 $\mu$A in contrast to 1.4 $\mu$A as predicted by the continuum doping model. The mean and the variance of the current distributions depend on the randomness in both (a) the number of dopants and (b) the position of the dopants. First, we will discuss the mean value of the current and then the variations from the mean.

The discrete doping model gives an average ON current and an average inverse subthreshold slope that are less than those from the homogeneous doping model. Both the reduction in $I_{ON}$ and the reduction in $S$ result from a reduction of the effective degeneracy in the source. The reduction in degeneracy is illustrated by the plot of
Figure 4.4: Band diagrams at the center of the nanowire in the ON state for a 4 nm TFET with continuous doping (thick red line) and the thirty samples with random doping (gray lines). $E_c$, $E_v$, $E_f$ denote conduction, valence and Fermi energy, respectively. The dashed line at 25 nm indicates the source-channel interface. The inset shows correlation coefficient $r$ between $I_{ON}$ and $N_{dop}$ as function of distance from source-channel interface. The maximum value of $r$ is at 6 nm.

The 30 band diagrams from discrete doping configurations (in gray) and the band diagram from the homogeneous doping distribution (in red) shown in Fig. 4.4. These bands are calculated at the center of the NWs. Most of the gray bands lie closer than the red band to the Fermi level in the source. The discrete doping tends to result in devices that are less degenerate in the source. This is true even though the average number of dopants (50) is slightly above the target (48). The lower degeneracy then results in lower drive currents and lower inverse subthreshold slopes for a fixed voltage range as one would expect [2].

Why does the discrete doping tend to give a less degenerate source? The randomness of the dopant positions gives rise to a non-uniform potential profile in the source. The effect is illustrated with Fig. 4.5 in which dopants have been deterministically placed in a line down the center of the source. At each dopant site, there exists a localized potential well at a lower potential (for holes) than the potential from the
Figure 4.5: Band diagrams for a 4 nm TFET that illustrate the reduction in degeneracy resulting from the discrete doping model. For the solid blue lines, dopant atoms are placed every 10 nm at the center of the nanowire, so that the macroscopic doping density is $N_A = 6.25 \times 10^{18} \text{ cm}^{-3}$. For the dotted red lines, the source is homogeneously doped with the same $N_A$. The 3 figures are plotted at different positions inside the nanowire: (a) center (b) 1 nm away from the center (c) 2 nm away from the center.

homogeneous doping model. Elsewhere, the doping is lower (zero actually) as compared to the continuous doping. Therefore, away from the center where there is no dopant, the potential from the discrete doping model results in a bandedge that is less degenerate. Thus, except for the small regions around the dopants, the potential resulting from the discrete doping model is less degenerate than that from the homogeneous doping model. Since the number of dopant sites is much less than the total number of sites present in the source, the overall degeneracy is reduced. This reduction in degeneracy explains the reduction of the average ON current and the reduction of the average inverse subthreshold slope resulting from the discrete doping model.

The NEGF self-consistent calculations show the same trends, except that, as shown in Table 4.1(a), the resulting current is even less, both for homogeneous doping and for discrete doping. The local density of states (LDOS) in the NSC model is peaked in the center of the NW, whereas the LDOS of the SC model is uniform.
Thus, the NW center is less degenerate in the NSC model resulting in lower current than the SC model. This comparison is discussed further at the end of this section.

While a shift in the average values is interesting, it is not particularly important from a device engineering point of view, since the averages can be shifted to their target values by increasing the target doping. Furthermore, Table 4.1(a) indicates that the difference in the average current predicted by the homogeneous doping model and the discrete doping model decreases with increasing diameter and doping density. What is more problematic from an engineering point of view are large statistical variations from the mean resulting from the randomness of the dopants.

The fluctuation in $I_{ON}$ should depend on the fluctuation in the number of dopant atoms. In order to determine how strongly the two quantities are related, a correlation coefficient $r$ between the ON current $I_{ON}$ and the dopant number $N_{dop}$ is considered. The formula for $r$ is

$$r = \frac{\sum_{i=1}^{N_{tot}} (N_{dop,i} - \overline{N_{dop}})(I_{ON,i} - \overline{I_{ON}})}{(N_{tot} - 1)\sigma(N_{dop})\sigma(I_{ON})}$$

(4.4)

where $N_{tot} = 30$ is the total number of test devices, $\sigma(x)$ is the standard deviation of the quantity $x$, $\overline{x}$ is the mean of all observations and subscript $i$ denotes the $i^{th}$ device. $r$ is +1 in the case of a perfect positive linear relationship, -1 in the case of a perfect negative linear relationship and some value in between for other cases.

One expects that dopant fluctuations far from the interface will have little effect on the current [104,105]. Therefore, to determine a maximum value for $r$, $r$ is calculated as a function of distance from the source-channel interface. The total number of dopants $N_{dop}$ contained within a given distance from the source-channel interface is calculated for each of the 30 devices of an ensemble. From these 30 values, the average value, $\overline{N_{dop}}$, and the standard deviation, $\sigma(N_{dop})$ are determined. Using these values...
for each distance from the interface, \( r \) is calculated, and an example plot is shown in the inset of Fig. 4.4.

Both the value of \( r_{\text{max}} \) and its position have a physical interpretation. The position of \( r_{\text{max}} \) follows the screening length calculated from the homogeneous doping model. This indicates that fluctuations of the dopant number within a screening length of the interface have a maximal impact on the tunnel current. The values of \( r_{\text{max}} \) for all diameters and doping range between 0.53 and 0.83. The fact that the correlation between \( I_{\text{ON}} \) and \( N_{dop} \) is always less than 1, indicates that \( I_{\text{ON}} \) is affected not only by fluctuations in the number of dopants, but also by fluctuations in their position.

Insight into the effect of the random position of the dopants on the source-to-channel tunnel barrier can be obtained from the band diagrams plotted in Fig. 4.6 corresponding to the three numbered transfer characteristics in Fig. 4.3. The band to band tunneling transmission can be estimated as

\[
T = \exp \left\{ -2 \int_{0}^{w} dx' \kappa(x') \right\} [63],
\]

where \( w \) is the tunneling barrier width and \( \kappa \) is the magnitude of the imaginary wavevector within the bandgap. Hence, the transmission and the current depend exponentially on the barrier width determined by the local electric field. Smaller tunneling widths give larger transmission and current. The three bands from the 3 devices in Fig. 4.6b, c, d have different tunnel barrier widths. Fig. 4.6b is for the device with the lowest current in Fig. 4.3. The tunneling width is nearly 4 nm wider than that of the homogeneously doped model, and therefore, the current is reduced by more than an order of magnitude. For 4.6c, \( w \) is same in both the discrete and homogeneous doping models, and, hence, the currents are very close differing by only 0.07 \( \mu \)A. For the last plot in Fig. 4.6, the electric field near the tunnel region is 0.35 eV/nm whereas for the homogeneous doping case, it is 0.14 eV/nm. The higher electric field results in a thinner tunnel barrier and a 0.5 \( \mu \)A larger current.

In the literature, \( \sigma \) has been used to quantify and compare the dispersion in
Figure 4.6: Band diagrams and energy resolved current density in the ON state at the center of the nanowire for the (a) continuous doping case and three numbered transfer curves as marked in Fig. 4.3: (b) case I, (c) case II and (d) case III. Darker colors represent lower values of current. $E_c, E_v, E_f$ denote the conduction, valence and Fermi energies, respectively. In (b), (c) and (d), band lines from continuous doping model are shown in dashed white lines for comparison.
Figure 4.7: Percentage $CV$ for ON current ($CV_{I_{ON}}$) as function of doping for different NW sizes as indicated on the figure. The isolated red round data point corresponds to a set of samples with a fixed (expected) number of dopants in a 4 nm NW. The empty squares with dotted line are the variations obtained from NEGF self-consistent calculations labeled NSC. All other results shown are from the SC model.

threshold voltage [92,94]. However, since we want to compare the dispersion in ON currents as a function of NW diameter and doping, and the average ON currents are very different, as shown in column 4 of Table 4.1(a), an alternative measure of dispersion is needed. The statistical metric to use in such cases is the coefficient of variation, $CV$. By definition, $CV = \frac{\sigma}{\mu}$ where $\mu$ is the mean of the available data. It is a normalized measure of dispersion.

Fig. 4.7 illustrates how $CV$ for the ON current, $CV_{I_{ON}}$, decreases with both increasing diameter and doping. The decreasing values of $CV_{I_{ON}}$ with increasing NW size and doping can be explained in terms of the $CV$ for the number of dopants, $CV_{Ndop}$, and the screening lengths shown in Fig. 4.8. As shown in Fig. 4.8a, the variations in the number of dopant atoms decrease as the macroscopic doping density $N_A$ is increased. Furthermore, the screening length shown in Fig. 4.8b also decreases as $N_A$ is increased. Therefore, the variation in current also decreases as $N_A$ increases. The screening length in the source also decreases as the diameter increases causing
the geometry to become more two-dimensional, and the electric field at the tunnel junction to be higher for a given doping. The reduced screening length also serves to reduce the variation in the current.

For the largest NW with the highest doping, the variation remains around 15%. The solid isolated circle in Fig. 4.7 shows the CV for the 4 nm NW when randomness exists only in the spatial configuration of the dopant atoms; the total number, $N_{dop}$, is fixed to the expected number in all 30 samples. In this case, the variation is reduced, but it is still quite high.

The NSC calculated average ON currents for the 4 nm TFET for the 30 samples doped with $N_A = 1.5 \times 10^{20}$ cm$^{-3}$, $3 \times 10^{20}$ cm$^{-3}$, and $5 \times 10^{20}$ cm$^{-3}$ are listed in Table 4.1(a). The average currents are smaller than those from the SC approach. The local density of states used in the NSC calculation determined from Eq. (4.2) and plotted in Fig. 4.9a is peaked in the center of the NW. The SC LDOS is uniform across the cross section. The average values of the SC and NSC LDOS are the same.
Figure 4.9: (a) Local density of states (LDOS) calculated along the cross-section of an ideal NW. (b) Band edges at the center of a 4 nm NW TFET with doping of $1.5 \times 10^{20} \text{cm}^{-3}$. Blue solid lines are obtained from semi-classical model and green dashed lines are obtained from NEGF self-consistent model. The dotted red line is the Fermi level position.

as is clear from Eq. (4.3). Since the NSC LDOS is larger at the NW center, the degeneracy at the center is less than that from the SC model. This is seen in a comparison of the band diagrams taken at the center of the NWs in Fig. 4.9b. The potentials from the two methods differ by a rigid shift, and therefore, the same two doping configurations result in the same trends in the current and its variation for both models. The CV values for the 4 nm NW calculated from the NSC model and the SC model are compared in Fig. 4.7. The CV values from the NSC model follow those of the SC model, and they approach each other at the highest doping value. The CV values from the NSC model are slightly higher than those of the SC model. Thus, the SC model gives slightly more optimistic values than the NSC model.

4.4 Conclusion

The effects of microscopic dopant fluctuations in the source have been investigated for InAs TFETs in a NW geometry. The small device cross sectional area gives rise to a finite number of dopant atoms in the source. The effect of both the random
number and the random placement of source dopants has been analyzed. The discrete doping model gives an average ON current and an average inverse subthreshold slope that are considerably less than those predicted from the homogeneous doping model. Both the reduction in $I_{ON}$ and the reduction in $S$ result from a reduction of the effective degeneracy in the source. The doping density required to achieve a target drive current is higher in the discrete doping model compared to the homogeneous doping model. The ON current variation from sample-to-sample results from the variation in the local electric field defining the tunnel barrier. The effects of the discrete, random dopants are particularly significant within a screening length of the source-channel interface. The relative variation in the ON current decreases as the average doping density and/or NW diameter increases. For the largest 8 nm diameter NW considered, the coefficient of variation in the ON current is 30% at a doping of $5.5 \times 10^{19}$ cm$^{-3}$ and 15% at a doping of $1.5 \times 10^{20}$ cm$^{-3}$. Although demonstrated for InAs TFETs, the general trends in the statistical variation of drive current with doping and diameter should apply to other materials as well.
Chapter 5

Conclusion

Today, tunnel FET is the most promising steep-slope switch. They have the potential to be operated below 0.5 V and can offer significant power reduction in power dissipation. Their working mechanism depends on quantum tunneling, and it consumes low OFF current. Hence, they are suitable for both low operating power and low standby power applications. The greatest challenges with TFETs are getting enough ON current without sacrificing the OFF current while still maintaining the low inverse subthreshold swing \( S \).

In this dissertation, we have performed quantum mechanical simulations that address these issues. The key findings of this work are summarized below.

- InAs and InSb NW TFETs require highly degenerate source doping to support the high electric fields in the tunnel region for any reasonable ON current. For NW sizes varying between 4 - 10 nm and for a target ON current of 1 \( \mu \)A, the source Fermi energy lies in the range of 0.1 – 0.22 eV below the valence band edge depending on the material and diameter.

- In the subthreshold, these devices experience both regimes of “voltage-controlled
tunneling” and “cold-carrier injection.” While “voltage-controlled tunneling” provides a small decrease in $S$, the large decrease below the thermal limit is the result of “cold-carrier injection” which occurs as the bands uncross.

- We found that the bands do uncross in all of the devices achieving the target on-current of 1 $\mu$A and they all show a significant reduction in the subthreshold slopes below the ideal thermal limit even though the source doping is highly degenerate.

- The standard WKB theory gives good qualitative agreement with the full-band numerical simulations.

- For a 5 nm physical gate length, direct channel tunneling dominates the leakage current for both FETs and TFETs and channel material selection becomes a critical choice.

- We found that the tunneling effective mass gives the best indication of the relative size of the tunnel currents when comparing two different materials of any type.

- For direct-gap III-V materials, the bandgaps are correlated with the tunneling masses, so the relative size of the bandgaps can also be used to determine the relative size of the tunnel currents. However, this only holds within the direct-gap III-V subset of materials.

- The indirect-gap materials, Si and Ge, give the largest tunneling masses in the conduction band, and they give the smallest conduction band tunnel currents within the range of diameters considered. Si gives the lowest overall tunnel current for both the conduction and valence bands and, therefore, is the optimum choice for suppressing tunnel current at the 5 nm scale.
• The semianalytic expression for the tunnel current given by Eq. (3.1) gives good agreement with the numerical result. Therefore, this expression can be used to calculate the tunnel current in wider diameter NWs which, using a full-band model, are intractable or very difficult to simulate fully numerically.

• The small device cross sectional area in a NW TFET gives rise to a finite number of dopant atoms in the source. The effect of both the random number and the random placement of source dopants has been analyzed.

• The discrete doping model gives an average ON current and an average inverse subthreshold slope that are considerably less than those predicted from the homogeneous doping model. Both the reduction in $I_{ON}$ and the reduction in S result from a reduction of the effective degeneracy in the source. The doping density required to achieve a target drive current is higher in the discrete doping model compared to the homogeneous doping model but their difference reduces as the NW size increases.

• The ON current variation from sample-to-sample results from the variation in the local electric field defining the tunnel barrier.

• The effects of the discrete, random dopants are particularly significant within a screening length of the source-channel interface. The relative variation in the ON current decreases as the average doping density and/or NW diameter increases. For the largest 8 nm diameter NW considered, the coefficient of variation in the ON current is 30% at a doping of $5.5 \times 10^{19}$cm$^{-3}$ and 15% at a doping of $1.5 \times 10^{20}$cm$^{-3}$.

• Although demonstrated for InAs TFETs, the general trends in the statistical
variation of drive current with doping and diameter should apply to other materials as well.
Bibliography


