INTERBAND TUNNEL TRANSISTORS

A Dissertation

Submitted to the Graduate School
of the University of Notre Dame
in Partial Fulfillment of the Requirements
for the Degree of

Doctor of Philosophy

by

Qin Zhang

______________________________
Alan C. Seabaugh, Director

Graduate Program in Electrical Engineering

Notre Dame, Indiana

April 2009
INTERBAND TUNNEL TRANSISTORS

Abstract

by

Qin Zhang

Interband tunnel transistors have been attracting increasing attention because of their potential to achieve subthreshold swings below the 60 mV/decade thermionic limit, and realize high performance and low power dissipation simultaneously. This work explores the design and modeling of semiconducting and graphene nanoribbon-based tunnel transistors, to understand the performance measures and guide the experimental development. Experiments in the formation of Ge interband junctions are also described.

Analytic expressions for Zener tunneling in one-, two-, and three-dimensional semiconductors are derived to establish the guidelines for tunnel transistor design. An analytic expression is derived, showing that the subthreshold swing of interband tunnel transistors is a function of the gate-to-source voltage and can be less than the thermionic limit of 60 mV/decade in MOSFET. Based on this expression, a new fully-depleted interband tunnel transistor structure is proposed and designed. The low subthreshold swing is verified by Synopsys TCAD simulation. Germanium interband tunnel transistors are shown by simulation to exhibit improved on-state performance vs. Si, because of the smaller bandgap and effective mass. To realize the proposed Ge interband tunnel transistor, a rapid melt growth process was developed to form submicron $p^+n^+$ Ge tunnel
junctions. Transmission electron microscopy (TEM) reveals the regrown film and a contact microstructure consistent with the Al-Ge phase diagram. Negative differential resistances are observed which indicate the junction was abrupt heavily-doped.

A graphene nanoribbon (GNR) tunnel transistor is first proposed and modeled analytically by quasi-1D Poisson equation. An improved numerical model is developed that treats energy-dependent transmission coefficients, direct source-to-drain tunneling, and self-consistent channel electrostatics. Graphene nanoribbons have a width-tunable bandgap and ultra-thin body layer, which is especially favorable for tunnel transistor applications. It is shown by simulation that the GNR tunnel transistors at the long channel limit can operate at 0.1 V with an ultra-low subthreshold swing of 2.8 mV/decade, but the subthreshold swing and off-state current are degraded at short channel length due to direct source-to-drain tunneling. Smaller ribbon widths (down to a certain limit) can significantly improve the off-state behavior without considerably affecting the on-state current density and speed. For 20 nm channel length, GNR tunnel transistors with ribbon width of 2 and 3 nm can achieve high performance and low operating power simultaneously, meeting 2012 ITRS targets.
To my parents
5.3 Geometry dependence.................................................................................51
5.4 Quantum capacitance..............................................................................55
5.5 Temperature dependence.......................................................................58
5.6 Comparison with nano-CMOS technology............................................59

CHAPTER 6: CONCLUSIONS AND SUGGESTED FUTURE WORK.............61
6.1 Summary.................................................................................................61
6.2 Suggested future research......................................................................63

APPENDIX A.................................................................................................64
A.1 DERIVATION OF 1D ZENER TUNNELING CURRENT IN DIRECT
SEMICONDUCTORS (Eq. 2.3).......................................................................64
A.2 DERIVATION OF 2D ZENER TUNNELING CURRENT IN DIRECT
SEMICONDUCTORS (Eq. 2.11)....................................................................65

APPENDIX B: GERMANIUM RAPID-MELT-GROWTH TUNNEL JUNCTION
PROCESS TRAVELER..................................................................................66

APPENDIX C: MATLAB PROGRAM TO SELF-CONSISTENTLY CALCULATE
BAND DIAGRAMS AND TUNNELING CURRENT DENSITIES
FOR GNR TFETS......................................................................................75

APPENDIX D: LOW-SUBTHRESHOLD-SWING TUNNEL TRANSISTORS......79

APPENDIX E: FULLY-DEPLETED GE INTERBAND TUNNEL TRANSISTOR:
MODELING AND JUNCTION FORMATION.............................................84

APPENDIX F: GRAPHENE NANORIBBON TUNNEL TRANSISTORS.........91

REFERENCES.............................................................................................95
FIGURES

Figure 1.1: Scaling of transistor physical gate length and technology node by Moore’s law [1]-[2] and ITRS 2008 [3]. The solid data points are provided by R. Chau et al. [2]: the data above 65 nm node are from measurements and the data below 65 nm are projections. ..........................................................2

Figure 1.2: MOSFET active and passive power density vs. gate length, after W. Haensch et al. [4]. ..........................................................3

Figure 1.3: Schematic device structure (a) and band diagrams for an nTFET at (b) off-state, (c) on-state, and (d) saturation regime......................................………..4

Figure 2.1: Zener tunneling in a reverse-biased p⁺n+ junction for (a) 1-D semiconductors and (b) GNRs. The red dotted line in (b) shows the linear E-k relation at high k limit for GNRs. (c) The triangular potential barrier seen by tunneling electrons..........................................................9

Figure 2.2: Computed room temperature Zener tunneling current density vs. electric field and ribbon width in GNR p⁺n+ junctions at a reverse bias of 0.1 V. .............13

Figure 2.3: Zener tunneling current density per volt of reverse bias vs. electric field for Si and Ge tunnel diodes. Close agreement is obtained between Eq. (2.9) and measurements with a fitted effective mass of 0.16 m₀ for Si and 0.02 m₀ for Ge..........................................................15

Figure 2.4: Computed Zener tunneling current density vs. electric field at a reverse bias of 0.25 V for 2D InAs tunnel junctions. The bandgap of InAs is increased to 0.5eV for a 2.5 nm quantum well structure [46]...............................17

Figure 2.5: Computed Zener tunneling current density vs. electric field and semiconductor material at a reverse bias of 0.1 V. The solid line shows the current per unit width of 3D semiconductors, assuming a lateral junction depth is 2 nm. The dashed line shows the current per ribbon width of GNR tunnel junctions for widths of 3, 5, and 10 nm. The inset gives the material parameters used in Eq. (2.3), (2.5), and (2.9)...............................18
Figure 2.6: Tunneling probability vs. electric field comparing GNR Zener tunneling and 3D semiconductor Zener tunneling with the same bandgap, using Eq. (2.2) and (2.5). ………………………………………………………………………………………………………………19

Figure 3.1: Band diagram for a MIS structure at zero gate bias, where $X_m$ is the metal work function, $X_{ox}$ is the electron affinity of the insulator, and $X_e$ is the electron affinity of the semiconductor. ………………………………………………………………23

Figure 3.2: Cross section of complementary TFETs, (a) n-TFET and (b) p-TFET, where the ultrathin body is heavily-doped to form a $p^+ n^+$ tunnel junction and the gate is fully-depletes the channel. The computed energy-band diagrams for Ge TFETs along the center of the channel are shown for both off-state (solid line, $|V_{DS}| = 0.5 \, \text{V}, V_{GS} = 0$) and on-state (dashed line, $|V_{DS}| = 0.5 \, \text{V}, V_{GS} = 0.5 \, \text{V}$). For the n-TFET, the $n^+$ doping is $1.8 \times 10^{20} /\text{cm}^3$, and the $p^+$ doping is $3.2 \times 10^{20} /\text{cm}^3$; for the p-TFET, the $n^+$ doping is $3.4 \times 10^{20} /\text{cm}^3$, and the $p^+$ doping is $1.1 \times 10^{20} /\text{cm}^3$ …………………………………………………………………………………26-27

Figure 3.3: Simulated transfer characteristics of n-TFET at $V_{DS} = 0.5 \, \text{V}$ for Si and Ge using the geometry in Fig. 3.2 (a), with comparison to a 45 nm n-MOSFET projection (dashed line) [50]. The dopings for Ge n-TFET are the same as in Fig. 3.2(a), and the doping for the Si n-TFET is $10^{20} /\text{cm}^3$ for both the $n^+$ and $p^+$ sides. …………………………………………………………………………………………………29

Figure 3.4: Junction maximum electric field vs. position to the Ge/Al$_2$O$_3$ interface in the channel and as a function of abruptness in the on-state. The on-state current density degrades approximately one order of magnitude as the abruptness degrades from 0 to 4 nm/decade. …………………………………………………………………………30

Figure 3.5: (a) Simulated common-source load curves and (b) voltage transfer characteristics for a complementary pair of Ge tunnel transistors in a static CMOS TFET inverter using the geometry of Fig. 3.2, n-TFET, $n^+$ doping of $1.8 \times 10^{20} /\text{cm}^3$, and $p^+$ doping of $3.2 \times 10^{20} /\text{cm}^3$, while for the p-TFET, the $n^+$ doping is $3.4 \times 10^{20} /\text{cm}^3$ with $p^+$ doping of $1.1 \times 10^{20} /\text{cm}^3$ …………………………………………………………………………………………………31

Figure 3.6: (a) Simulated common-source load curves and (b) voltage transfer characteristics for a complementary pair of Ge tunnel transistors in a static CMOS TFET inverter using the geometry of Fig. 3.2 with identical n- and p-TFET junction dopings, $n^+ = p^+ = 1 \times 10^{20} /\text{cm}^3$ …………………………………………………………………………………………………31
Figure 4.1: Scale drawing showing a cross section of a $300 \times 300$ nm Ge tunnel diode process using SODs and RTP: (a) an $n^+$ phosphorus doped layer is formed on the $p^-$ substrate, (b) an Al film is patterned by e-beam lithography, and (c) lift off, (d) a SiN$_x$ cap is deposited and a $p^+n^+$ junction is formed by rapid thermal annealing, (e) a 100 nm via is open by e-beam lithography and etch, and finally (f) bondpads are patterned by lift-off.

Figure 4.2: SEM image of a $300 \times 300$ nm Al/Ge emitter with a SiN$_x$ cap and $100 \times 100$ nm via after 600 ºC, 1 s rapid thermal annealing.

Figure 4.3: Schematic e-beam mask layout for Ge tunnel diode fabrication and measurement with device sizes from $300 \times 300$ nm to $10 \times 10$ µm.

Figure 4.4: (a) High resolution TEM image of a Ge $p^+n^+$ junction after 600 ºC rapid thermal annealing. No dislocations are observed. (b) TEM of the $10 \times 10$ µm AlGe-p$^+n^+$ tunnel junction. The interface is flat in some areas and curved in other areas, like the one shown. These images were obtained by T. Kosel, Notre Dame.

Figure 4.5: (a) STEM image of the regrown Ge epilayer between the Ge substrate and the eutectic mixture after 600 ºC rapid thermal annealing. X-ray maps of (b)Al and (c) Ge showing the darker regions in (a) are the Al-doped Ge-rich phase, and the lighter regions are the Al-rich phases as expected from the phase diagram. Measurements courtesy of T. Kosel, Notre Dame.

Figure 4.6: Current-voltage characteristics of Ge tunnel junctions formed by 600 ºC rapid melt regrowth of 50 nm Al on $n^+$ Ge.

Figure 4.7: Top view optical micrographs of the Ge substrate after phosphorus SOD spinning, 800 ºC, 5 min rapid thermal annealing, and 5 min buffered HF etch to form the $n^+$ layer.

Figure 5.1: Schematic cross section of an interband $p$-channel GNR TFET, where the GNR is heavily-doped or electrostatically-doped to form a $p^+n^+$ tunnel junction and the gate is placed over the fully-depleted $p$-side.

Figure 5.2: Energy band diagrams for a GNR $p$-TFET at (a) off-state and (b) on-state with a supply voltage of 0.1 V. The calculated gate oxide thickness is 1 nm, the GNR width is 5 nm, and the gate length is 20 nm.

Figure 5.3: Calculated transfer characteristics of GNR $n$- (solid line) and $p$-TFETs (dashed line) at room temperature using the same geometry as Fig. 5.2.
Figure 5.4: Tight-binding energy dispersion relation of the first and second subband of a graphene nanoribbon with the width of $15a$; $a = 0.246$ nm is the lattice constant of graphene. The solid lines are the real $E-k$ relation and the dots are the imaginary $E-k$ relation in the bandgap. The dashed lines show Eq. (5.3) fit to the first subband in the bandgap. The tight-binding simulation is courtesy of Tian Fang, Notre Dame. 49

Figure 5.5: Calculated transfer characteristics of GNR $p$-TFETs with (a) $w = 5$ nm for different channel lengths, and (b) $L_{CH} = 20$ nm for different ribbon widths. 52

Figure 5.6: (a) Calculated intrinsic gate delay and (b) total gate delay vs. $I_{OFF}$ performance of GNR $p$-TFETs with $L_{CH} = 20$ nm for different ribbon widths at a supply voltage of 0.1 V. 53

Figure 5.7: Calculated intrinsic gate delay vs. $I_{OFF}$ performance of GNR $p$-TFETs with $L_{CH} = 20$ nm and (a) $w = 5$ nm, (b) $w = 2$ nm at a supply voltage of 0.1 V. The intrinsic gate delay $\tau_{INT}$ (solid line) is separated to $\tau_{S}$ (dashed line) and $\tau_{D}$ (dotted line), which corresponds to the charge injection from the source and drain respectively. 54

Figure 5.8: Calculated gate capacitance vs. $V_{GS}$ of GNR $p$-TFETs with $L_{CH} = 20$ nm and (a) $w = 5$ nm, (b) $w = 2$ nm at $V_{DS} = -0.1$ V. The gate capacitance $C_{G}$ (solid line) is separated to $C_{GS}$ (dashed line) and $C_{GD}$ (dotted line), which corresponds to the charge injection from the source and drain respectively. 55

Figure 5.9: Calculated gate capacitance vs. $V_{GS}$ of GNR $p$-TFETs with $L_{CH} = 20$ nm and (a) $w = 5$ nm, (b) $w = 2$ nm at $V_{DS} = -0.02$ V. The gate capacitance $C_{G}$ (red line) is separated to $C_{GS}$ (blue line) and $C_{GD}$ (black line), which corresponds to the charge injection from the source and drain respectively. 58

Figure 5.10: Calculated drain current density vs. $V_{DS}$ of GNR $p$-TFETs with $L_{CH} = 20$ nm and (a) $w = 5$ nm, (b) $w = 2$ nm. 58

Figure 5.11: On-state current density (solid line) and on-to-off ratio (dotted line) vs. temperature for 0.5 nm $EOT$ gate oxide, 20 nm gate length GNR TFETs with ribbon width of 2 and 5 nm. 59
Table 3.1: Speed and power estimates comparing $n$MOSFETs with TFETs. The MOSFET projections are based on the 2008 Updates ITRS Roadmap [3] targets. The TFETs use the geometry of Fig. 3.2 and are projected as a function of channel material using *Synopsis TCAD* to compute the leakage currents. For the tunnel currents, Eq. (3.10) is used with the electric field computed from *Synopsis*.

Table 5.1: Speed and power estimates comparing 2012 $n$MOSFETs targets from the 2008 Updates ITRS roadmap [3] with 2 nm and 3 nm GNR TFETs using the geometry in Fig. 5.1.
ACKNOWLEDGEMENTS

I would like to sincerely thank many people for their great help on my Ph.D. study and research. I want to thank my advisor, Dr. Alan C. Seabaugh, for giving me the opportunity to study at University of Notre Dame. Without his brilliant guidance and continuous support in the last six years, I could not finish this dissertation. I want to thank all my committee members, Dr. Thomas H. Kosel, Dr. Debdeep Jena and Dr. Bernstein, for their willingness to read through my dissertation and their thoughtful comments and suggestions. I want to thank Dr. Joerg Appenzeller, Purdue University, who gave me the first idea of this work; D.J. Frank, IBM, for discussions on MOSFET performance measures; and Brian Doyle, Intel, for providing the SIMS measurements. I also want to thank Tian Fang, Yeqing Lu, Dr. Patrick Fay, Dr. Debdeep Jena, and Dr. Siyuranga Koswatta (IBM) for their help in clarifying my thinking.

I especially want to thank all my group members, Sajid Kabeer, Qingmin Liu, Zoltan Racz, Surajit Sutar, Dane Wheeler, Bin Wu, Yan Yan, Jialin Zhao, and Wei Zhao, for their help either on the experiments or theoretical discussions, and also the great friendship, making my study at Notre Dame a pleasure.

Special thanks go to Keith Darr, Mike Thomas, and Mark Richmond, for their patient training in the lab. Also, special thanks to all the secretaries in the department of electrical engineering. I also want to thank many people outside of our group, Lili Ji,
Yong Tang, VJ, Kejia Wang, Ning Su, Ze Zhang, Xiu Xing, Chunxin Lian, Ying Cao for their great help with my work.

Finally, I would like to thank the financial support from Amberwave Systems and the Nano-electronics Research Initiative through the Midwest Institute for Nanoelectronics Discovery.
CHAPTER 1

INTRODUCTION

Complementary Metal Oxide Semiconductor (CMOS) technology has been scaling down in size for 40 years under the guidance of Moore’s law [1] which dictates that the density of transistors on an integrated circuit (IC) doubles every 18-24 months while gate length halves every 3-4 years. Figure 1.1 shows the scaling of transistor physical length and technology node vs. year in agreement with Moore’s law until the 65 nm node [2], but delayed beyond it [3]. The technology node, defined as the smallest half-pitch of contacted metal lines, is used by the semiconductor industry as an overall indicator for IC feature scaling [3]. Projected by 2008 International Technology Roadmap for Semiconductors (ITRS) [3], the 45 nm (with gate length of 24 nm) and 32 nm technology node (with gate length of 18 nm) will be delayed 3 to 4 years to achieve.

This is due to the power dissipation issue more and more difficult to solve for CMOS while keeping the performance gain of 30% per generation [4]. Because the subthreshold swing of MOSFET is limited to 60 mV/decade, it is difficult to scale the supply voltage $V_{DD}$ below 1 V while increasing the speed and maintaining low off-state power dissipation. According to the generalized field scaling rule [5], the power density increases as the square of the scaling factor if $V_{DD}$ is kept constant. Transistor scaling without reducing the supply voltage makes charge control ever more difficult. Shrinking
of the gate length brings about undesired short-channel effects (SCE). First, as gate length is reduced, the drain bias has greater influence on the channel potential, reducing the threshold voltage and increasing the off-state leakage; this phenomenon is called drain-induced barrier lowering (DIBL). Second, thinner gate oxides are needed to maintain gate control at shorter gate lengths. The gate oxide thickness has been scaled to 1.2 nm at the 65 nm node, getting close to the physical limit [3]. The ultra-thin gate oxide increases the gate leakage by band-to-band tunneling across $n^+$PolySi gate/SiO$_2$/pSi body.

At the 65 nm node (gate length of 40 nm), the power dissipated in leakage passive mechanisms has become comparable to the active power, Fig. 1.2 [4], and become the primary limit to further scaling.
To continue scaling the MOSFET beyond the 65 nm technology node, new materials and device geometries are being implemented. Low subthreshold swing field-effect transistors are offering another alternative. Generally there are two kinds of low subthreshold swing field-effect transistors that can overcome 60 mV/decade subthreshold swing limit at room temperature: one is using new gate stack design to provide an intrinsic gate voltage amplification, i.e., suspended-gate MOSFET [6]-[8] and ferroelectric MOSFET [9]-[10]; the other one is using new current mechanisms rather than MOSFET diffusion current in the subthreshold region, i.e., Impact-Ionization MOSFET [11]-[13] and interband tunnel transistor [14]-[26]. In this dissertation, interband tunnel transistors or tunnel field-effect transistors (TFETs) will be discussed.

Figure 1.2 MOSFET active and passive power density vs. gate length, after W. Haensch et al. [4].
Interband tunnel transistors use an electric field to gate the Zener tunneling current of a $p^+n^+$ junction, a mechanism which can be configured to lower subthreshold swing and power. A schematic device structure of an $n$-type TFET is shown in Fig. 1.3(a). A heavily-doped $p^+n^+$ junction is formed in the ultra-thin semiconductor body; a gate is aligned to the junction on the $n^+$ side and fully depleted the channel at zero gate bias. So at zero gate bias, the band gap in the channel blocks electrons from the source to tunnel, and the transistor is OFF, Fig. 1.3(b). When a positive gate voltage is applied, the conduction band in the channel is pushed below the valence band at the source, turning the interband tunneling and the transistor ON, Fig. 1.3(c). Because the source-channel junction is reverse-biased, the on-state current mechanism is Zener tunneling which will

![Figure 1.3](image-url)

**Figure 1.3** Schematic device structure (a) and band diagrams for an $n$TFET at (b) off-state, (c) on-state, and (d) saturation regime.
be discussed in details in chapter 2. Further increasing the drain voltage, the tunneling current can saturate since the gate screens the drain field, Fig. 1.3(d).

The first TFET was reported in 1965 [14], attracting intense attention only recently [16]-[26] after Appenzeller et al. [15] experimentally demonstrated a carbon nanotube TFET with subthreshold swing less than 60 mV/decade in 2004. This low subthreshold swing showed the potential of TFET to realize high performance and low power dissipation simultaneously. In 2007, Choi et al. [18] demonstrated a planar Si TFET with subthreshold swing of 52.8 mV/decade, but only 12.1 $\mu$A/$\mu$m on-state current density at a 1 V supply voltage, limited by the large bandgap of silicon. In 2008, Krishnamohan et al. [20] and Mayer et al. [21] reported two more TFETs with subthreshold swing less than 60 mV/decade and improved on-state current in SiGe, but still lower than the high-performance MOSFET. Recently, graphene nanoribbon (GNR) TFETs have been proposed and analyzed [25]-[26]. These transistors are based on the one-dimensional (1D) semiconducting properties of GNRs, with width-tunable bandgap [27]-[28] and monolayer-thin body [29].

This dissertation explores the design and modeling of semiconducting and GNR-based TFETs, to understand the performance measures and guide the experimental development. Experiments in the formation of Ge tunnel junctions are also described. In chapter 2, the analytic expressions for Zener tunneling in 1D, 2D, and 3D semiconductors are derived, which establish the guidelines for TFET design (chapter 3-5). The graphene nanoribbon is also discussed and compared with conventional 3D semiconductors. Chapter 3 outlines how the subthreshold swing limit arises in MOSFETs and then shows analytically that certain gate-controlled TFETs have intrinsically smaller subthreshold
swing [18]. These findings were published in a paper attached as Appendix D. Fully-depleted Si and Ge TFETs were proposed and simulated by Synopsys TCAD [30] to verify that low subthreshold swing is possible, and improved performance relative to Si can be achieved in Ge (chapter 3). To realize the proposed Ge interband tunnel transistor requires the development of an abrupt, heavily-doped lateral tunnel junction, which is the subject of chapter 4. A rapid melt growth process for forming submicron $p^n+n^+$ Ge tunnel junctions is developed and described. Ge $p^n+n^+$ diodes with sizes ranging from 30 $\times$ 30 $\mu$m to 300 $\times$ 300 nm were fabricated. Transmission electron microscopy (TEM) revealed the regrown film and a contact microstructure consistent with the Al-Ge phase diagram. The negative differential resistances (NDR) in the current-voltage measurements are observed which indicated the junction was heavily-doped. However, the peak-to-valley current ratio (PVR) of these devices was low and not improved by reducing the area down to 0.1 $\mu$m$^2$, which was likely a result of point defects or junction doping nonuniformity, as TEM shows no dislocations at the regrown junction. The findings on Ge interband tunnel transistor modeling and junction formation were published in a paper included as Appendix E.

Graphene nanoribbons (GNRs) have a width-tunable bandgap [27]-[28] and ultra-thin body layer [29], which is especially favorable for TFET applications, and more amenable to planar processing and large-scale integration than carbon nanotube (CNT) TFETs. In chapter 5, the GNR tunnel transistor is analyzed. Ribbon widths between 2 and 5 nm are considered, corresponding to energy bandgaps in the range of 0.69 to 0.28 eV. It is shown by simulation that the GNR tunnel transistors at long channel limit can operate at 0.1 V with an ultra low subthreshold swing of 2.8 mV/decade, but the subthreshold swing and
off-state current are limited at short channel length by direct source-to-drain tunneling. Smaller ribbon widths (down to a certain limit) can significantly improve the off-state behavior without considerably affecting the on-state current density and speed. For 20 nm channel length (36 nm technology node), GNR tunnel transistors with ribbon width of 2 and 3 nm can achieve high performance and low operating power simultaneously, meeting 2012 ITRS targets [3]. A portion of the results in this chapter are published in the first GNR TFET paper included as Appendix F. Chapter 6 gives a summary and suggestions for future research on interband tunnel transistors.

Other conference presentations and publications (the last paper was submitted) relevant to this work are listed below:


2. Q. Zhang, S. Sutar, T. Kosel, and A. Seabaugh, “Rapid melt growth of Ge tunnel junctions for interband tunnel transistors,” 2007 ISDRS.

3. Q. Zhang and A. Seabaugh, “Can the interband tunnel FET outperform Si CMOS?” 2008 DRC.


CHAPTER 2

ZENER TUNNELING THEORY

Zener tunneling, electrons passing “from one ‘energy band’ into another”, was first introduced by Clarence Zener in 1934 [31] to explain dielectric breakdown, a precipitous rise in current as the field strength is increased. Specifically, in a heavily-doped semiconductor $p^+n^+$ junction, electrons in the $p^+$ valence band can tunnel into the $n^+$ conduction band under a reverse bias, an effect now called Zener tunneling and the primary transport mechanism in tunnel transistors. Zener tunneling current is determined by integrating the product of charge flux and the tunneling probability from the energy states on the $p^+$ side to those on the $n^+$ side, where the tunneling probability is calculated by applying the WKB (Wentzel-Kramers-Brillouin) approximation to the triangular potential at the $p^+n^+$ junction.

This chapter derives analytic expressions for Zener tunneling in one-dimensional (1D), 2D and 3D semiconductors, which establish the principles for tunnel transistor design. The graphene nanoribbon, as a special 1D semiconductor is also discussed and compared with conventional 3D semiconductors.
2.1 Zener tunneling current in 1D semiconductors

2.11 General Zener tunneling current-voltage (I-V) relation in 1D semiconductors

In a reverse-biased heavily-doped $p^+n^+$ junction shown in Fig. 2.1 (a), the electrons at the valence band of the $p^+$ side can tunnel through the forbidden band into the conduction band of the $n^+$ side. This tunneling process is similar to that of a particle penetrating a triangular potential barrier with a height higher than its energy by the semiconductor bandgap $E_G$ and a slope given by the electron charge times the electric field $q\xi$, Fig. 2.1(c).

![Diagram of Zener tunneling in a reverse-biased $p^+n^+$ junction](image)

Figure 2.1 Zener tunneling in a reverse-biased $p^+n^+$ junction for (a) 1-D semiconductors and (b) GNRs. The red dotted line in (b) shows the linear $E-k$ relation at high $k$ limit for GNRs. (c) The triangular potential barrier seen by tunneling electrons.
The wave vector \( k_x(x) \) is given by

\[
k_x(x) = \sqrt{\frac{2m^*}{\hbar^2}(E - U(x))}
\]  

(2.1)

where \( m^* \) is the effective mass, \( \hbar \) is the reduced Planck constant, \( U(x) \) is the potential barrier energy and \( E \) is the particle’s energy: \( U(x) = E + q\zeta x \ (0 < x < d), \ d = E_G / q\zeta \). The wave vector \( k_x(x) \) is an imaginary number since \( U(x) > E \). The tunneling probability is calculated using the WKB approximation [32], \( T_{WKB} \approx \exp\left(-2\int_0^d|k_x(x)|dx\right) \).

\[
T_{WKB}^{1D} = \exp\left(-\frac{4\sqrt{2m^*E_G^2/2}}{3q\hbar\zeta}\right)
\]  

(2.2)

For a 1D direct bandgap semiconductor \( p^+ n^+ \) junction, \( m^* \) is replaced by the reduced effective mass \( m^*_e = (1/m_e^* + 1/m_h^*)^{-1} \), where \( m_e^* \) is the electron effective mass and \( m_h^* \) is the hole effective mass, and \( \zeta \) is the maximum electric field at the junction [33].

Zener tunneling current is obtained by integrating the product of charge flux and the tunneling probability from energy states on the \( p^+ \) side to the \( n^+ \) side,

\[
I^{1D} = \int q v_g(k_x) \rho(k_x)(f_V - f_C)T_{WKB}^{1D}dk_x
\]

\[
= \frac{q^2}{\pi\hbar} T_{WKB}^{1D} V_T \ln\left(\frac{1}{2} \left(1 + \cosh\frac{V_R}{V_T}\right)\right)
\]  

(2.3)

where \( v_g(k_x) = 1/\hbar(dE/dk_x) \) is the group velocity, and \( \rho(k_x) = 1/\pi \) is the 1D density of states. The Fermi-Dirac distributions, \( f_V \) and \( f_C \), at the valence band of the \( p^+ \) side and conduction band of the \( n^+ \) side are respectively, \( f_V(E) = \frac{1}{1 + \exp((E - qV_R)/qV_T)} \) and
\[ f_c(E) = \frac{1}{1 + \exp(E/qV_T)}, \] where \( V_R \) is the reverse bias and \( V_T \) is the thermal voltage, \( kT/q \) (\( k \) is Boltzmann’s constant). The derivation details are included in Appendix A.1.

If \( V_R \gg V_T \), the tunneling current can be expressed as
\[ I^{1D} = \frac{q^2}{\pi \hbar} T^{1D}_{WKB} (V_R - V_T \ln 4). \]

Further, at temperature \( T = 0 \) K, the current is proportional to the reverse bias
\[ I^{1D} = \frac{q^2}{\pi \hbar} T^{1D}_{WKB} V_R, \] similar to the Landauer expression [34].

2.12 Zener tunneling current in graphene nanoribbons (GNRs)

The graphene nanoribbon is a special 1D semiconductor material, where the \( E-k \) relation is not parabolic but given by \([35]\) \( E = \pm \hbar v_F \sqrt{k_x^2 + k_n^2} \) where the subscript \( n \) refers to the \( n \)th subband, \( v_F \), the Fermi velocity is \( 10^8 \) cm/s \([36]\), \( k_n \), the \( n \)th transverse wave vector \( k_n = n\pi/3w \) \([27]\) is quantized by the ribbon width \( w \), so that the bandgap \( E_G = 2\hbar v_F k_1 = 2\pi \hbar v_F / 3w = 1.375 / w \) (nm) eV \([27]-[28]\). Figure 2.1 (b) shows the first subband diagram of a GNR \( p^+ n^+ \) junction, in which the tunneling current has the same form as in a conventional 1D semiconductor, Eq. (2.3), but with different expression for the tunneling probability \( T^{1D}_{WKB} = T^{GNR}_{WKB} \) due to the special \( E-k \) relation of GNRs.

In a heavily-doped GNR \( p^+ n^+ \) junction, if only the first subband for tunneling is considered,
\[ k_x = \frac{1}{2\hbar v_F} \left( E - U(x) \right)^2 - k_1^2 = \frac{1}{\hbar v_F} \sqrt{\left( q\xi x / 2 \right)^2 - (E_G / 2)^2} \] (2.4)

therefore,
Figure 2.2 shows the room temperature tunneling current per μm of ribbon width as a function of the electric field for different ribbon widths at a reverse bias of 0.1 V, using Eq. (2.3) and (2.5). With a supply voltage of 0.1 V, a tenth of the MOSFET supply voltage, graphene nanoribbon tunnel junctions can achieve a current density close to 1000 μA/μm at an electric field of 4 MV/cm, which is comparable to a high-performance MOSFET.

The expression gives the ribbon width \( w_m \) at which the current density is maximized at a fixed electric field. As the electric field \( \zeta \) increases from 0.5 MV/cm to 5 MV/cm, \( w_m \) decreases from 9 nm to 3 nm [37]. At low electric field, the transmission coefficient dominates in the current density calculation, so narrower GNRs with larger bandgaps which leads to lower tunneling probability have lower current density, see Fig. 2.2. However, the current relationship reverses at high electric fields where \( 1/w \) dependence of the charge flux term becomes dominant. This analysis does not work for GNR TFET design, because the electric field at the source-to-channel junction is not a free variable, but a function of the ribbon width.
2.2 Zener tunneling current in 3D and 2D semiconductors

For three-dimensional (3D) direct semiconductors, the total energy can be divided into two parts \( E = E_x + E_\perp \), where \( E_x = \hbar^2 k_x^2 / 2m_r \) corresponds to the energy in the tunneling direction and \( E_\perp = \frac{\hbar^2 k_\perp^2}{2m_r} \) is the transverse energy which is conserved during the tunneling process. Therefore,

\[
k_x(x) = \sqrt{\frac{2m^*}{\hbar^2}(E_x - U(x))} = \sqrt{\frac{2m^*}{\hbar^2}(-q\zeta x - E_\perp)} \quad (2.6)
\]

\[
T^{3D}_{WKB} = \exp\left(-\frac{4\sqrt{2m^*(E_x + E_\perp)^{3/2}}}{3q\hbar\zeta}\right) \approx T^{1D}_{WKB} \exp\left(-\frac{E_\perp}{E}\right) \quad (2.7)
\]
where $\bar{E} = \frac{q \hbar \xi}{2\sqrt{2m_r E_G}}$ is a factor of which determines the impact of the transverse-energy-state carriers on the tunneling magnitude. The larger the factor $\bar{E}$ is, the less degradation of the tunneling probability by carriers with the transverse energy. The effect on tunneling by the transverse states is minimized by high electric field, small effective mass, and narrow bandgap.

Similar to 1D Zener tunneling, the 3D Zener tunneling current density is also calculated by integrating the product of charge flux and the tunneling probability from the energy level at the $p^+$ side to the $n^+$ side [38]

$$J^{3D} = \int \int q v_g(k_x) \rho_x(k_x) \rho_{\perp}(k_{\perp}) dk_x 2\pi k_{\perp} dk_{\perp} (f_f - f_c) T^{3D}_{WKB},$$

(2.8)

where $v_g(k_x) = 1/\hbar dE_x/dk_x$ is the group velocity, $\rho_x(k_x) = 1/\pi$ and $\rho_{\perp}(k_{\perp}) = 1/4\pi^2$ is the density of states in the tunneling direction and the transverse direction respectively. To simplify the calculation, $f_f - f_c \approx 1$ is assumed and then

$$J^{3D} \approx \int \int \frac{q m_r^*}{2\pi^2 \hbar} T^{3D}_{WKB} \exp(-\frac{E}{E}) dE_x dE_{\perp}.$$ Integrating $E_x$ from 0 to $qV_R, E_{\perp}$ from 0 to $(qV_R-E_x)$, and assuming $E \ll qV_R$,

$$J^{3D} = \frac{\sqrt{2m_r^* q^3 \xi V_R}}{8\pi^2 \hbar^2 E_G^{1/2}} \exp(-\frac{4\sqrt{2m_r^* E_G^{3/2}}}{3q \xi \hbar})$$

(2.9)

Equation (2.9) is lower than Sze and Ng’s equation [39] by a factor of 2 in the prefactor. This is likely an error in Sze and Ng’s book, as there is no reference to an original derivation.

Although Eq. (2.9) is derived for direct semiconductors, it also fits well to indirect semiconductors [40]. Using Eq. (2.9), the dependence of tunnel current density on reverse
bias is computed versus junction internal field for Si and Ge, Fig. 2.3. For the case of Si, measurements were taken from 8 different $p^+ n^+$ tunnel junctions [41]-[42], spanning over 8 orders of magnitude in current density. Equation (2.9) is in excellent agreement with the measurements using only the tunneling reduced effective mass as a fitting parameter and the maximum electric fields calculated from the measured doping profile in [41]-[42]. The calculations in Fig. 2.3 are also in reasonable agreement with the recent Si TFET demonstration [19] for a 10 nm channel thickness and a 4 MV/cm maximum electric field. For the Ge case, the maximum electric field $\zeta$ in Eq. (2.9) is calculated assuming abrupt doping, because of missing information on the doping profile, so the calculated current density is slightly higher than measurements [43]-[45]. If a 4 nm/decade dopant slope is assumed, the Ge data point provided by this work (Chapter 4) will shift to the calculated curve, as the electric field decreases from 2.8 to 1.6 MV/cm.

![Figure 2.3](image_url)

Figure 2.3 Zener tunneling current density per volt of reverse bias vs. electric field for Si and Ge tunnel diodes. Close agreement is obtained between Eq. (2.9) and measurements with a fitted effective mass of 0.16 $m_0$ for Si and 0.02 $m_0$ for Ge.
For application in ultra-thin body TFETs, it is interesting to understand 2D Zener tunneling. In a 2D system, the transverse energy can be further divided into two parts: 

\( E_\perp = E_y + E_{zn} \), where the \( y \) direction is free and the \( z \) direction is quantized. Substituting \( E_\perp \) for \( E_y + E_{zn} \) in Eq. (2.7), gives the tunneling probability as

\[
T_{\text{WKB}}^{2D} = T_{\text{WKB}}^{1D} \exp\left(-\frac{E_{zn}}{E}\right) \exp\left(-\frac{E_y}{E}\right) \tag{2.10}
\]

and the tunneling current density is

\[
J_{\text{2D}} = \int qv_y (k_x) \rho_z (k_x) \rho_y (k_y) dk_x dk_y (f_y - f_e) T_{\text{WKB}}^{2D}
\]

\[
= \frac{q \sqrt{m^*}}{2 \sqrt{2 \pi^2} h^2} \exp\left(-\frac{E_{zn}}{E}\right) \int_0^{qV_R} \int_0^{qV_R - E_x} \frac{1}{\sqrt{E_y}} \exp\left(-\frac{E_y}{E}\right) dE_y dE_x \tag{2.11}
\]

where the integral of the error function \( \text{Erf} \) can be solved numerically. The derivation details are included in Appendix A.2.

Figure 2.4 shows the computed Zener tunneling current density vs. electric field at a reverse bias of 0.25 V for 2D InAs tunnel junctions, using Eq. (2.11). The bandgap of InAs is 0.5 eV in the calculation, for a 2.5 nm quantum well structure [46].
2.3 Comparison of Zener tunneling current in 3D semiconductors and GNRs

Figure 2.5 compares the tunneling current per unit width as a function of internal field for Si, Ge, InAs, InSb and GNR lateral tunnel junctions, using the material parameters in the inset table. For 3D semiconductors, a 2 nm junction depth is assumed and quantization is neglected. The solid line, Eq. (2.9) shows that narrower band-gap materials with smaller effective mass can achieve tunneling current density approaching that of a high performance MOSFET. For the same electric field, Ge Zener tunneling current is more than two orders of magnitude greater than Si, while InAs and InSb is about three orders of magnitude higher. The dashed lines show the current per ribbon width of GNR tunnel junctions for widths of 3, 5 and 10 nm, respectively, using Eq. (2.3)
and (2.5), which is higher than the widely-explored low-mass semiconductors, even higher than InSb which has a smaller bandgap.

![Figure 2.5 Computed Zener tunneling current density vs. electric field and semiconductor material at a reverse bias of 0.1 V. The solid line shows the current per unit width of 3D semiconductors, assuming a lateral junction depth is 2 nm. The dashed line shows the current per ribbon width of GNR tunnel junctions for widths of 3, 5, and 10 nm. The inset gives the material parameters used in Eq. (2.3), (2.5), and (2.9).](image)

Figure 2.5 shows the computed Zener tunneling current density versus electric field and semiconductor material at a reverse bias of 0.1 V. The solid line represents the current per unit width of 3D semiconductors, assuming a lateral junction depth of 2 nm. The dashed line illustrates the current per ribbon width of GNR tunnel junctions for widths of 3, 5, and 10 nm. The inset contains the material parameters used in Eqs. (2.3), (2.5), and (2.9).

Figure 2.6 compares the tunneling transmission probability of GNR tunnel junctions, with that of semiconductor tunnel junctions having the same bandgap, using Eqs. (2.2) and (2.5); only minor differences are found, indicating that the higher tunneling current density of GNR tunnel junctions shown in Fig. 2.5 benefits from the charge flux in the 1D tunneling transport.
In this chapter, the analytic expressions for Zener tunneling in 1D, 2D, and 3D semiconductors were derived, which established the principles for tunnel transistor design. The Zener tunneling expression for bulk semiconductors (3D) in Sze and Ng’s book [39] were corrected and used for Si and Ge TFET simulations in chapter 3. The formula for Zener tunneling in GNRs (1D) was published [37], and used for GNR TFET design in chapter 5.

Figure 2.6 Tunneling probability vs. electric field comparing GNR Zener tunneling and 3D semiconductor Zener tunneling with the same bandgap, using Eq. (2.2) and (2.5).
CHAPTER 3

LOW SUBTHRESHOLD SWING TUNNEL TRANSISTORS

In a field-effect transistor, the minimum voltage swing needed to turn a transistor from on to off is an important figure of merit which ultimately determines how low in power a device technology can be. This characteristic is usually quantified by measuring how many millivolts it takes to change the drain current by one order of magnitude. The measure of this characteristic is called the subthreshold swing and is given in units of mV/decade.

In this chapter, the fundamental limit for the subthreshold swing in the MOSFET is discussed. In seeking lower subthreshold swings, the interband tunnel transistor is analyzed and shown to provide a lower subthreshold swing. Fully-depleted Si and Ge tunnel transistors are modeled to verify that low subthreshold swing can be achieved, and improved on-state current density relative to Si is seen in Ge tunnel transistors, which is comparable to 45 nm node high-performance MOSFETs at a supply voltage of 0.5 V.
3.1 Subthreshold device physics

3.11 MOSFET subthreshold device physics

In a field-effect transistor, the subthreshold swing, $S$, is defined as the voltage required to increase or reduce the drain current by one decade in the subthreshold region [47],

$$S = \left( \frac{d \log I_D}{d V_{GS}} \right)^{-1}$$

(3.1)

where $I_D$ is the subthreshold drain current and $V_{GS}$ is the gate-source voltage. In the MOSFET, when the gate voltage is below the threshold voltage, the semiconductor surface is in weak inversion and the corresponding diffusion-dominated drain current is called the subthreshold current. For an $n$-channel MOSFET, the subthreshold current, $I_D$, is analyzed by Sze and Ng [48],

$$I_D = -qD_n \frac{dn}{dx} = qAD_n \frac{n(0) - n(L)}{L},$$

(3.2)

where $D_n$ is the electron diffusivity, $n(x)$ is the spatially-dependent electron density, $L$ is the gate length, and $A$ is the cross section of the device in the direction of the current flow, given by

$$A = Wd_{2DEG}.$$  

(3.3)

Here, $W$ is the width of the device and $d_{2DEG}$ is the effective channel thickness that corresponds to the distance over which the potential decreases by $kT/q$ from the semiconductor surface. The effective channel thickness can be written as

$$d_{2DEG} = \frac{kT}{q\varepsilon_s} = \frac{kT}{q} \left( \frac{2qN_A \psi_s}{\varepsilon_s} \right)^{1/2},$$

(3.4)
where $\xi_s$ is the electric field at the surface, $N_A$ is the acceptor density in the transistor body, $\Psi_s$ is the surface potential, and $\varepsilon_s$ is the semiconductor dielectric constant. The electron densities in the channel at the source, $n(0)$, and the drain $n(L)$ can be written as

$$n(0) = n_{p0} e^{q\Psi_s/kT}$$
and
$$n(L) = n_{p0} e^{q(\Psi_s - V_{DS})/kT},$$

(3.5)

where $n_{p0}$ is the electron density in the body, and $V_{DS}$ is the drain-source voltage. Substituting Eqs. (3.3), (3.4), and (3.5) into (3.2) gives

$$I_D = qW \frac{kT}{q} (2qN_A\Psi_s/\varepsilon_s)^{-1/2} D_n \frac{(n_{p0} e^{q\Psi_s/kT} - n_{p0} e^{q(\Psi_s - V_{DS})/kT})}{L}$$

$$= q \frac{W}{L} D_n \frac{kT}{q} (2qN_A/\varepsilon_s)^{-1/2} n_{p0} (1 - e^{q\Psi_{ox}/kT}) \Psi_s^{-1/2} \exp\left(\frac{q\Psi_s}{kT}\right)$$

(3.6)

where $C$ is a coefficient (with unit of $AV^{1/2}$) which is independent of the gate-source voltage. The surface potential is related to the gate-source voltage by

$$V_{GS} = \Psi_{ox} + \Psi_s + V_{FB},$$

(3.7)

where $\Psi_{ox}$ is the voltage across the oxide and $V_{FB}$ is the flat band voltage. Figure 3.1 shows the band diagram for a metal-insulator-semiconductor (MIS) structure at zero gate bias, so the flat band voltage is just $-(\Psi_{ox} + \Psi_s)$. 

22
Figure 3.1 Band diagram for a MIS structure at zero gate bias, where $X_m$ is the metal work function, $X_{ox}$ is the electron affinity of the insulator, and $X_s$ is the electron affinity of the semiconductor.

For thin oxides, $\psi_{ox}$ is much less than $\psi_s$, and

$$\frac{d \psi_s}{d V_{GS}} \approx 1$$  \hspace{1cm} (3.8)

Finally, Eqs. (3.6) and (3.8) are used to find the subthreshold swing $S$, which by definition is

$$S \equiv \ln 10 (d \ln I_D / dV_{GS})^{-1} > \ln 10 \left( \frac{q}{kT} \frac{d \psi_s}{dV_{GS}} \right)^{-1} \approx kT \ln 10 \frac{q}{dV_{GS}}$$  \hspace{1cm} (3.9)

According to this well-known analysis [48], the MOSFET has a subthreshold swing, given in mV/decade of current, which is limited by the thermal voltage to 60 mV/decade at room temperature.

This analysis has been extended to double gate (DG) MOSFETs by Chen [49], where an analytical subthreshold swing model for symmetric DG MOSFETs is derived and
verified using the 2D device simulator Medici. A worsening of the subthreshold swing is predicted with scaling below 30 nm, which suggests that low power circuit design will become increasingly more challenging as gate length decreases. The need for transistors which can scale below 30 nm in gate length and operate with smaller subthreshold swings is evident.

3.12 Tunnel transistor subthreshold device physics

The device geometry and operating principle of interband tunnel transistor has been introduced in chapter 1. In an interband tunnel transistor, the drain current flows across a degenerately-doped $p^+n^+$ tunnel junction in the Zener tunneling direction, a transport mechanism which has been discussed in chapter 2 (see Eq. (2.9) for 3D Zener tunneling),

$$I = a V_{\text{eff}} \xi \exp\left(-\frac{b}{\xi}\right)$$  \hspace{1cm} (3.10)

where $a$ and $b$ are coefficients determined by the materials properties of the junction and the cross-sectional area of the device. Specifically, $a = A q^3 \sqrt{2m^* E_G}/8\pi^2 h^2$ and $b = 4\sqrt{2m^* E_G^{3/2}}/3qh$ . $V_R$ in Eq. (2.9) is replaced by $V_{\text{eff}}$ here, which is the reverse bias of the source-channel junction for the transistor.

The derivative of the tunneling current expression of Eq. (3.10) with respect to the gate-source voltage can be used to determine an expression for the subthreshold swing of a field-effect tunneling transistor.

$$S = \ln 10 \left[ \frac{1}{V_{\text{eff}} V_{\text{GS}}} + \frac{\xi + b}{\xi^2 V_{\text{GS}}} \right]^{-1}$$  \hspace{1cm} (3.11)
There are two terms in the denominator of Eq. (3.11) which are maximized to achieve a low subthreshold swing and these terms are not limited by $kT/q$. According to the first term, the transistor should be engineered so that the gate-source voltage directly controls the tunnel junction bias, which suggests transistor geometry with thin and/or high-k gate dielectric (dielectric constant $\varepsilon > 4$) and an ultrathin body to assure the gate-field directly modulates the channel. For an equivalent oxide thickness approaching 1 nanometer, $dV_{\text{eff}}/dV_{GS} \approx 1$, and the first term in the denominator of Eq. (3.11) is approximately inversely related to $V_{GS}$. As a consequence, the lowest subthreshold swing will occur at small gate-source voltage. A second way to lower the subthreshold voltage swing is described by the second term in the denominator of Eq. (3.11) which says that the derivative of the junction electric field with respect to the gate-source voltage should be maximized. This is a primary basis for gate control in Bhuwalka’s proposed interband tunnel transistor [16]. In Bhuwalka’s transistor, increasing the gate-source voltage reduces the tunneling width and increases the junction electric field. Accordingly, the subthreshold swing in a tunnel transistor increases with gate-source voltage. This characteristic has been predicted in simulations by Wang [17] and Bhuwalka [16]. In practice, the tunnel junction bias and the junction electric field are coupled and cannot be engineered independently.

There have been four demonstrations of a sub-60-mV-per-decade tunnel transistor at room temperature, reported by Appenzeller et al. [15] in carbon nanotubes, Choi et al. [19] in Si, and Krishnamohan et al. [20] and Mayer et al. [21] in SiGe. The subthreshold swings were measured as a function of gate-source voltage, which is consistent with our prediction.
3.2 Interband tunnel transistor modeling

3.21 Device structure and operation principle

Using the design guidance of Eq. (3.11), a new transistor geometry is proposed here toward achieving low subthreshold swings. Scaled cross sections of the complementary tunneling field-effect transistor (TFET) pair are shown in Fig. 3.1(a), for an $n$-TFET, and 3.2(b), for a $p$-TFET. The transistors utilize a Si/Ge-on-insulator (SOI or GOI) structure, in which a lateral abrupt $p^+ n^+$ tunnel junction is formed in a 2 nm semiconductor body. A 20 nm long gate is oriented adjacent to the junction to fully-deplete the underlying semiconductor. In the simulations, a 1 nm thick Al$_2$O$_3$ dielectric is used as the gate oxide ($\varepsilon_R = 9$, $E_G = 8.7$ eV). Gold with a work function of 4.8 eV is used for the $n$-TFET gate metal, and Al with a work function of 4.25 eV is used for the $p$-TFET gate. The channel is fully depleted at zero gate bias.
Figure 3.2 Cross section of complementary TFETs, (a) n-TFET and (b) p-TFET, where the ultrathin body is heavily-doped to form a $p^+ n^+$ tunnel junction and the gate is fully-depletes the channel. The computed energy-band diagrams for Ge TFETs along the center of the channel are shown for both off-state (solid line, $|V_{DS}| = 0.5 \text{ V, } V_{GS} = 0$) and on-state (dashed line, $|V_{DS}| = 0.5 \text{ V, } V_{GS} = 0.5 \text{ V}$). For the n-TFET, the $n^+$ doping is $1.8 \times 10^{20} / \text{cm}^3$, and the $p^+$ doping is $3.2 \times 10^{20} / \text{cm}^3$; for the p-TFET, the $n^+$ doping is $3.4 \times 10^{20} / \text{cm}^3$, and the $p^+$ doping is $1.1 \times 10^{20} / \text{cm}^3$.

The device operation can be understood from the simulated band diagrams shown in Figs. 3.2(a) and 3.2(b). At zero gate bias, the $n^+$ (n-TFET) or $p^+$ (p-TFET) body is fully depleted and the transistor is normally off (solid lines) with no direct interband tunneling path. With a positive gate bias for n-TFET and a negative gate bias for p-TFET, interband tunneling is turned on (dashed lines) in the Zener tunneling direction. In this geometry, the gate screens the drain field and the current injection is set by the gate-source bias.

3.22 $I$-$V$ characteristics of Si and Ge tunnel transistors

*Synopsys CAD 2005 Tool* [30] is used to compute the 2D electrostatics for the transistors of Figs. 3.2(a) and 3.2(b). The channel is divided into 16 layers (with unit...
thickness of 2 nm/16 = 1.25 Å). For each layer, the band diagram is calculated, from which the reverse voltage and the maximum electric field at the junctions are read and entered into the tunneling current relation of Eq. (3.10) to determine the current density in mA/µm$^2$ in each incremental layer; the current density is then summed over all 16 layers. The tunneling path may not be restricted to each divided layer, but since the electric field does not change much across the junction (see Fig. 3.4), this integral method is a good approximation. Direct tunneling from source to drain is not included, which is only important for narrow bandgap semiconductors (see Fig. 5.5, where GNR with ribbon width of 2 nm has a bandgap of 0.69 eV, similar to Ge).

Transfer characteristics of Si and Ge n-TFET at $V_{DS} = 0.5$ V are shown in Fig. 3.3, and compared with Frank’s Si CMOS performance projection for 45 nm technology node [50]. Consistent with 1D analysis in section 3.12, the subthreshold swing in the TFET is a function of gate bias and lower than 60 mV/decade within a small range of gate-to-source voltage. The average subthreshold swing is more meaningful in comparisons of TFETs with MOSFETs. In MOSFETs, the threshold current $I_{th} = \frac{W}{L} I_{CC}$, where $I_{CC}$ is a constant number if not considering SCE, and typically is 100 nA for Si MOSFETs [51]. So the threshold current density $I_{th}/W$ is 100 nA divided by the gate length, which is 5 µA/µm in Fig.3.3. And using this constant current method [51], the threshold voltage of the Ge TFET is 0.05V. A swing of 50 mV changes the current by more than 3 orders of magnitude, giving an effective subthreshold swing less than 17 mV/decade.

Although Si TFET shows more than two orders of magnitude lower current density than the high performance MOSFET, the Ge TFET can provide comparable on-state current of nearly 450 µA/µm at $V_{DS} = V_{GS} = 0.5$ V. The off-state current density,
determined from Synopsis, is 3.6 nA/μm, even lower than the low operating power technology requirement by ITRS, which is 18.3 nA/μm [3].

![Simulated transfer characteristics of n-TFET at $V_{DS} = 0.5$ V for Si and Ge using the geometry in Fig. 3.2 (a), with comparison to a 45 nm n-MOSFET projection (dashed line) [50]. The dopings for Ge n-TFET are the same as in Fig. 3.2(a), and the doping for the Si n-TFET is $10^{20}$/cm$^3$ for both the $n^+$ and $p^+$ sides.](image)

Figure 3.3 Simulated transfer characteristics of $n$-TFET at $V_{DS} = 0.5$ V for Si and Ge using the geometry in Fig. 3.2 (a), with comparison to a 45 nm $n$-MOSFET projection (dashed line) [50]. The dopings for Ge $n$-TFET are the same as in Fig. 3.2(a), and the doping for the Si $n$-TFET is $10^{20}$/cm$^3$ for both the $n^+$ and $p^+$ sides.

Figure 3.4 illustrates the influence of the junction abruptness on the electric field in TFETs. In the on-state, as the abruptness degrades from 0 to 4 nm/decade, the maximum electric field at the junction decreases from 3.9 to 2.2 MV/cm and the on-state current density degrades almost one order of magnitude. Quantization of the ultra-thin Ge body should also influence the current density: the tunneling probability will decrease due to the increase in bandgap, while the tunneling density of states will increase from 3D to 2D. The quantization can cause current to increase or decrease depending on the energetic position of the quantized state.
3.23 Voltage-transfer characteristic of static Ge tunnel transistor inverters

Figure 3.5 (a) shows the simulated load curves of the static Ge tunnel transistor inverter at a supply voltage of 0.5 V. For channel currents under 170 μA/μm, in both the n- and p-TFET, the current saturates with high output resistance. The voltage transfer characteristic is extracted from the load curves and is shown in Fig. 3.5(b). The noise margin is around 100 mV, better than CMOS noise margin requirements of approximately 10% of the supply voltage [52]. The output characteristics and noise margin can be improved by lowering both the $n^+$ and $p^+$ doping to $10^{20}$ /cm$^3$ and on-state current density, as shown in Fig. 3.6. The improved gate control with lower tunnel-junction doping results both in better turn-on, saturation characteristics, and higher noise margin, but at the expense of lowering the on-state current density.
Figure 3.5 (a) Simulated common-source load curves and (b) voltage transfer characteristics for a complementary pair of Ge tunnel transistors in a static CMOS TFET inverter using the geometry of Fig. 3.2, \( n \)-TFET, \( n^+ \) doping of \( 1.8 \times 10^{20} / \text{cm}^3 \), and \( p^+ \) doping of \( 3.2 \times 10^{20} / \text{cm}^3 \), while for the \( p \)-TFET, \( n^+ \) doping is \( 3.4 \times 10^{20} / \text{cm}^3 \) with \( p^+ \) doping of \( 1.1 \times 10^{20} / \text{cm}^3 \).

Figure 3.6 (a) Simulated common-source load curves and (b) voltage transfer characteristics for a complementary pair of Ge tunnel transistors in a static CMOS TFET inverter using the geometry of Fig. 3.2 with identical \( n \)- and \( p \)-TFET junction dopings, \( n^+ = p^+ = 1 \times 10^{20} / \text{cm}^3 \).
3.3 Comparison with conventional nano-CMOS technology

Table 3.1 compares \( n \)-TFET performance with scaled \( n \)MOSFET targets (highlighted in gray) from the 2008 International Technology Roadmap for Semiconductors (ITRS) [3]. The on-state MOSFET currents are computed from Frank’s model [50] at a supply voltage of 0.5 V, and the off-state MOSFET currents are computed from Hanson’s model [26] with \( V_{GS} = 0 \) and \( V_{DS} = 0.5 \). From Hanson’s model, the off-state current at 32 nm node (2013) is 2.02 \( \mu A/\mu m \), \(~3 \times \) of ITRS requirement which is 0.64 \( \mu A/\mu m As \). As the development of high-k/metal gate stack technology improves, the gate leakage could be much lower than the off-state leakage at this stage; Chui has claimed a gate leakage less than \( 10^{-3} \) A/cm\(^2\) for HfO\(_2\)/Ge structure at 1 V bias, where the equivalent thickness of the oxide is less than 1 nm [54]. This gate leakage level is equal to \( 2 \times 10^{-7} \) \( \mu A/\mu m \) at 20 nm gate length. So the gate leakage is neglected here for the total leakage calculation. For the TFET, the transistor properties are evaluated for Si and Ge with a maximum internal junction field of 4 MV/cm. The off-state current is simulated by Synopsys’ generation-recombination model which is in agreement with measurements, a comparison with Choi et al. Si TFET [19]. Quantization effects are not included in these calculations; quantization would raise the effective band-gap and lower the off-currents relative to these predictions. The speed is estimated by \( C_{OX}V_{DD}/I_{ON} \) without quantum capacitance considerations, and in the last six rows, power and energy consumptions (dynamic and leakage) are calculated for an \( n \) stage inverter chain (\( n = 50 \)) with an activity factor \( \alpha \) (\( \alpha = 2\% \)), after Hanson et al. [53]. The activity factor of 2\% accounts for an average switching of 1 in 50 transistors per cycle.
TABLE 3.1 SPEED AND POWER ESTIMATES COMPARING NMOSFETs WITH TFETS. THE MOSFET PROJECTIONS ARE BASED ON THE 2008 UPDATES ITRS ROADMAP [3] TARGETS. THE TFETS USE THE GEOMETRY OF FIG. 3.2 AND ARE PROJECTED AS A FUNCTION OF CHANNEL MATERIAL USING SYNOPSIS TCAD TO COMPUTE THE LEAKAGE CURRENTS. FOR THE TUNNEL CURRENTS, EQ. (3.10) IS USED WITH THE ELECTRIC FIELD COMPUTED FROM SYNOPSIS.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>MOSFET 2013 (32 nm node)</th>
<th>Si</th>
<th>Ge</th>
<th>unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gate length L_G</td>
<td>18</td>
<td>20</td>
<td>20</td>
<td>nm</td>
</tr>
<tr>
<td>Gate width W ~ 10L_G</td>
<td>180</td>
<td>200</td>
<td>200</td>
<td>nm</td>
</tr>
<tr>
<td>Equivalent oxide thickness EOT</td>
<td>0.65</td>
<td>1</td>
<td>1</td>
<td>nm</td>
</tr>
<tr>
<td>Supply voltage V_DD</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>V</td>
</tr>
<tr>
<td>On current I_ON</td>
<td>701</td>
<td>1.2</td>
<td>440</td>
<td>µA/µm</td>
</tr>
<tr>
<td>Off current I_OFF</td>
<td>2.02</td>
<td>2.7E-06</td>
<td>0.0036</td>
<td>µA/µm</td>
</tr>
<tr>
<td>Oxide capacitance density C_OX</td>
<td>53.1</td>
<td>34.5</td>
<td>34.5</td>
<td>fF/µm²</td>
</tr>
<tr>
<td>Gate capacitance C_G ~ C_OX L_G</td>
<td>0.96</td>
<td>0.69</td>
<td>0.69</td>
<td>fF/µm</td>
</tr>
<tr>
<td>Intrinsic speed τ ~ C_G V_DD I_ON</td>
<td>0.68</td>
<td>288</td>
<td>0.78</td>
<td>ps</td>
</tr>
<tr>
<td>Leakage P_leak ~ n I_leak V_DD</td>
<td>50.50</td>
<td>6.8E-05</td>
<td>0.09</td>
<td>µW/µm</td>
</tr>
<tr>
<td>Dynamic P_dyn ~ 1/2 n I_ON V_DD α</td>
<td>175</td>
<td>0.300</td>
<td>110</td>
<td>µW/µm</td>
</tr>
<tr>
<td>Total P ~ P_leak + P_dyn</td>
<td>226</td>
<td>0.300</td>
<td>110</td>
<td>µW/µm</td>
</tr>
<tr>
<td>Leakage E_leak ~ (n I_leak) V_DD (n τ)</td>
<td>1722</td>
<td>1</td>
<td>4</td>
<td>aJ/µm</td>
</tr>
<tr>
<td>Dynamic E_dyn ~ 1/2 (n C_G) V_DD² α</td>
<td>120</td>
<td>86</td>
<td>86</td>
<td>aJ/µm</td>
</tr>
<tr>
<td>Total E ~ E_leak + E_dyn</td>
<td>1842</td>
<td>87</td>
<td>90</td>
<td>aJ/µm</td>
</tr>
</tbody>
</table>

ITRS 2008 Updates [3]  
logic depth n = 50, activity factor α = 2%

Since the MOSFET and the TFET are compared at the same supply voltage and similar capacitance, the speeds are determined by the on-state current. Silicon is not attractive for the TFET channel because of its low on-state tunnel current, only 1.2 µA/µm. With this low current, the speed is more than 400× lower than the MOSFET. In contrast, the Ge tunnel transistor shows an on-state current density as high as 440 µA/µm at 0.5 V supply voltage, and comparable speed to the 2013 nMOSFET with 18 nm gate length. The off-state current density for the Ge tunnel transistor is 0.0036 µA/µm, much lower than the MOSFET. The low off-state current of the Ge TFET dissipates 2× less power.
with energy dissipation more than $20\times$ lower than the 2013 MOSFET for a 50-stage inverter chain with an activity factor of 2%.

In summary, this chapter discussed the physics of both subthreshold MOSFETs and TFETs, and showed that TFETs could achieve lower subthreshold swings than 60 mV/decade at room temperature, the theoretical limit in MOSFETs. These findings were published in [18], Appendix D. Fully-depleted Si and Ge TFETs were proposed and simulated in two dimensions by *Synopsys TCAD* [30] to verify that the low subthreshold swings were possible, and improved performance relative to Si could be achieved in Ge. The Ge TFETs could have comparable performance to 45 nm node high-performance MOSFETs at a supply voltage of 0.5 V, but lower power and energy dissipation due to the low subthreshold swing. These simulation results were published in a paper in *Solid-State Electronics* (vol. 53), Appendix E.
CHAPTER 4

GERMANIUM TUNNEL JUNCTION FABRICATION

To realize the proposed Ge interband tunnel transistor it requires the development of an abrupt, heavily-doped lateral tunnel junction. Toward this end, a rapid melt growth process for forming submicron $p^+n^+$ Ge tunnel junctions is developed in which evaporated Al contacts on $n^+$ Ge are liquefied in a rapid thermal processor to dissolve back and regrow $p^+$ Ge and form the tunnel junction. Prior work [55] has shown high peak current density exceeding 1 mA/µm$^2$ for micron-scale diodes, but with low PVR, less than 1.5 at room temperature.

In this chapter, Ge $p^+n^+$ junctions with sizes ranging from 30 × 30 µm to 300 × 300 nm are fabricated. TEM reveals the regrown film and a contact microstructure consistent with the Al-Ge phase diagram. The low PVR of devices produced by this growth method is likely a result of point defects or junction doping nonuniformity, as TEM shows no dislocations at the regrown junction. The PVR of these junctions does not improve as the device area is reduced from 100 to 0.1 µm$^2$, a size smaller than the formation scale for grains in the Al-Ge system.
4.1 Demonstration of first submicron Ge tunnel diodes

4.11 Process flow

Figure 4.1 shows the process flow of fabricating $300 \times 300$ nm Ge tunnel diodes. The process started with a $p^-$ Ge substrate and used a phosphorus $1 \times 10^{21}$ /cm$^3$ spin-on diffusant (SOD) followed by rapid thermal annealing to form the $n^+$ side of the junction [55]-[56], Fig. 4.1(a). Secondary Ion Mass Spectroscopy (SIMS) measurements showed that a density of $6 \times 10^{19}$ /cm$^3$ was achieved for the $n^+$ doping. A 50 nm thick Al film was then patterned by electron beam (e-beam) lithography and lift-off. The Al serves both as a contact and the acceptor dopant source, Fig. 4.1(b) and (c). A 50 nm plasma Si$_3$N$_4$ cap was applied to act as a microcrucible after Liu et al. [57] and hold the Al-Ge melt during the annealing, Fig. 4.1(d). The rapid thermal annealing of Al in contact with Ge above the eutectic temperature [58] caused Ge to be dissolved into the liquefied Al. On cooling the Ge regrew epitaxially until the eutectic temperature was reached, forming an abrupt, heavily-doped $p^+$ layer. As the temperature further decreased, Al-rich and Ge-rich phases were nucleated, leaving a eutectic mixture above the Ge $p^+n^+$ junction. To contact the submicron devices, $100 \times 100$ nm vias were written by e-beam lithography and reactive ion etched (RIE) to open contact vias to the AlGe, Fig. 4.1(e). Figure 4.2 shows the scanning electron microscopy (SEM) image of a $300 \times 300$ nm device with a $100 \times 100$ nm via. Finally, Ti/Au bondpads were patterned and deposited by lift-off, Fig. 4.1(f). A detailed process traveler is attached as Appendix B.
Figure. 4.1 Scale drawing showing a cross section of a 300 × 300 nm Ge tunnel diode process using SODs and RTP: (a) an $n^+$ phosphorus doped layer is formed on the $p^-$ substrate, (b) an Al film is patterned by e-beam lithography, and (c) lifted off, (d) a SiN$_x$ cap is deposited and a $p^+ n^+$ junction is formed by rapid thermal annealing, (e) a 100 nm via is open by e-beam lithography and etch, and finally (f) bondpads are patterned by lift-off.
Figure 4.2 SEM image of a $300 \times 300$ nm Al/Ge emitter with a SiN$_x$ cap and $100 \times 100$ nm via after 600 °C, 1 s rapid thermal annealing.

A three-layer e-beam mask set was designed and is shown in Fig. 4.3, where the emitter layer was blue, the via layer was white and the bondpad layer was yellow. The current-voltage ($I$-$V$) characteristics were measured by putting one probe on the center bondpad and the other probe on the side bondpad, so that the small-area diode were measured in series with a large-area diode back-to-back and only the characteristics of the small-area diode were revealed [55]-[56].
4.12 Physical analysis of the regrown Ge tunnel junction

Figure 4.4(a) shows a TEM image of an AlGe-$p^+ n^+$ tunnel junction after 600 °C, 1 s rapid thermal annealing with a cooling rate of 30 °C/s. Close examination of the $p^+ n^+$ junction in this location and elsewhere shows no evidence of dislocations. However, Fig. 4.4(b) reveals the overall nonuniformity of the junction, a cross section through a 10 × 10 μm$^2$ device.

Scanning transmission electron microscopy (STEM) image and energy-dispersive x-ray spectroscopy (EDXS) are shown in Fig. 4.5 to confirm the presence of Al and locate the junction. EDXS analysis of the eutectic mixture above the regrown Ge layer shows that the darker regions are the Al-doped Ge-rich phase, and the lighter regions are the Al-rich phases as expected from the phase diagram [58]. Electron diffraction patterns show that the Ge-rich regions of the eutectic mixture are within a few degrees of the same crystallographic orientation as the Ge substrate, indicating that these regions are epitaxial extensions of the regrown layer.
Figure 4.4 (a) High resolution TEM image of a Ge $p^+n^+$ junction after 600 °C rapid thermal annealing. No dislocations are observed. (b) TEM of the 10 × 10 µm AlGe-$p^+n^+$ tunnel junction. The interface is flat in some areas and curved in other areas, like the one shown. These images were obtained by T. Kosel, Notre Dame.

Figure 4.5 (a) STEM image of the regrown Ge epilayer between the Ge substrate and the eutectic mixture after 600 °C rapid thermal annealing. X-ray maps of (b)Al and (c) Ge showing the darker regions in (a) are the Al-doped Ge-rich phase, and the lighter regions are the Al-rich phases as expected from the phase diagram. Measurements courtesy of T. Kosel, Notre Dame.
4.13 Ge tunnel diode $I$-$V$

Figure 4.6 shows the current-voltage ($I$-$V$) characteristics of Ge tunnel junctions with sizes ranging from $10 \times 10$ $\mu m^2$ to $300 \times 300$ $nm^2$. Negative differential resistances (NDR) are observed in all sizes, a signature of abrupt heavily-doped junctions. Series resistance shifts are responsible for the voltages differences between small and large area devices. For the $300 \times 300$ $nm^2$ diode, a peak current density of $0.15$ mA/$\mu m^2$ is achieved, corresponding to an effective doping of $\sim 3.5 \times 10^{19}$ /cm$^3$ [59], and $1$ mA/$\mu m^2$ Zener tunneling current is achieved at the reverse bias of $0.13$ V, where the maximum electric field of $2.9$ MV/cm is calculated assuming an ideal abrupt junction. Compared with the calculated current density of Ge tunnel diodes shown in Fig. 2.3, the measured current density is lower, which suggests that doping profiles are graded. If a 4 nm/decade abruptness is assumed, the electric field will decrease to $1.6$ MV/cm, in agreement with the current density calculation, discussed with reference to Fig. 2.3. The PVR is low and does not increase significantly as junction area is reduced. The low PVR could result from defect-assisted tunneling through point defect energy level in the bandgap or could be an indication that the junction doping is nonuniform across the growth interface (see the next section). Doping nonuniformity results in a peak current and peak voltage which depends on position. Currents in a nonuniform junction are the summation of $I$-$V$ behavior over each incremental area; these may sum to lower the overall PVR.
4.2 Nonuniformity of Ge tunnel junctions

The nonuniformity of the Ge junctions revealed in Fig. 4.4 (b) is likely influenced by nonuniformities observed in the P diffusion from SOD. Figure 4.7 shows that the phosphorus doped Ge $n^+$ layer was not smooth, which might result from the reaction of Ge and $P_2O_5$ at high temperature, creating GeO$_x$. The desorption of GeO$_x$ could produce surface roughness [55]. The nonuniformity of the Ge junctions could also arise from the complexity of Al-Ge phase diagram [58].

This nonuniformity of the Ge junctions is one of the possible reasons to explain the low PVR in the $I$-$V$ measurements, Fig. 4.6, and make a low yield. An alternative way to do Ge $n^+$ doping, i.e. implantation, and different $p^+$ metal dopants could be included in the future work.
Figure 4.7 Top view optical micrographs of the Ge substrate after phosphorus SOD spinning, 800 °C, 5 min rapid thermal annealing, and 5 min buffered HF etch to form the $n^+$ layer.

In this chapter, a rapid melt growth process was used to fabricate Ge $p^+n^+$ tunnel diodes with sizes ranging from $30 \times 30 \, \mu m$ to $300 \times 300 \, nm$. The regrown film and contact microstructure revealed by TEM were consistent with the Al-Ge phase diagram. The NDR in the current-voltage measurements was observed, indicating the junctions were heavily doped. But the PVR of these devices was low and not improved by reducing the area down to $0.1 \, \mu m^2$, which was likely a result of point defects or junction doping nonuniformity as TEM suggest no dislocations at the regrown junction. The findings on Ge $p^+n^+$ junction formation were published in *Solid-State Electronics* (vol. 53) and included as Appendix E.
CHAPTER 5

GRAPHENE NANORIBBON TUNNEL TRANSISTORS

To enhance the on-state tunneling current in TFETs, narrower bandgap materials with smaller effective masses are being considered [16] [20]-[21]; Knoch et al. show by simulation that 1D TFETs are preferred to bulk FETs because of the superior gate control and a reduction of transverse energy component in the 1D tunneling transport [22]-[23]. Graphene nanoribbons (GNRs) have a width-tunable bandgap [27]-[28] [35]-[36] [60]-[62] and monolayer-thin body, which is especially favorable for TFET applications, and more amenable to planar processing and large-scale integration than CNTs. In this chapter, a GNR TFET is proposed and modeled. Ribbon widths between 2 and 5 nm are considered, corresponding to energy bandgaps in the range of 0.69 to 0.28 eV. First, the GNR TFET is analytically modeled by a quasi-1D Poisson equation. Then a compact model is developed that treats energy dependent transmission coefficients, direct source-to-drain tunneling, and self-consistent channel electrostatics. It is shown that GNR TFETs in the long channel limit can operate at 0.1 V with an ultra-low subthreshold swing (2.8 mV/decade). Subthreshold swing and off-state current are degraded as the channel length scales down. Smaller ribbon widths (down to a certain limit) can significantly improve the off-state behavior while keeping high performance. Simulations
show that for 20 nm channel length, GNR TFETs with ribbon width of 2 and 3 nm can achieve high performance and low operating power simultaneously, compared to 2012 ITRS targets [3].

5.1 Device structure and electrostatic model for GNR TFETs

![Device structure diagram](image)

Figure 5.1 Schematic cross section of an interband $p$-channel GNR TFET, where the GNR is heavily-doped or electrostatically-doped to form a $p^+n^+$ tunnel junction and the gate is placed over the fully-depleted $p$-side.

Similar to a CNT TFET [22]-[23], the schematic of a $p$-channel GNR TFET is shown in Fig. 5.1. The source is heavily doped to be $n^+$ and the drain is $p^+$. The channel is also designed to be $p^+$, but is fully depleted at zero gate bias. Heavy doping concentrations are generated either by dopants or engineered electrostatically [23]. Due to the strong C-C bonds in GNRs (and CNTs), chemical doping is challenging, but rapid progress is being made in that direction [62]. On the other hand, electrostatic doping has successfully been demonstrated for achieving tunnel junctions in CNT FETs [15] and the same can be achieved for GNRs. The electrostatics of this quasi-1D geometry is solved by a surface potential method, where the GNR surface potential, $\Phi_F(x)$, is given by the 1D Poisson equation [63],

\[
\frac{d^2\Phi_F(x)}{dx^2} - \frac{\Phi_F(x) - \Phi_G - \Phi_{BL}}{\lambda^2} = -\frac{q(\pm N)}{\varepsilon_{GNR}}
\]  

(5.1)
Here, \( q\Phi_G \) is the gate potential energy, \( q\Phi_{bi} = E_G/\sqrt{2} - qV_T \ln(N/n) \) is the built-in potential energy, \( N \) is the doping concentration; the “+” sign is for donors and the “-” sign is for acceptors. The parameter \( \lambda = \sqrt{(\varepsilon_{GRN}/\varepsilon_{OX})t_{GRN}} t_{OX} \) is the relevant length scale for potential variations; \( \varepsilon_{GRN} \) and \( \varepsilon_{OX} \) are the GNR and oxide dielectric constants respectively and assumed the same values, and \( t_{GRN} = 0.35 \) nm [29] is the thickness of the GNR. Lastly, \( t_{OX} \) is the gate oxide thickness. Equation (5.1) is applied in all three regions: \( n^+ \) source, \( p^+ \) channel, and \( p^+ \) drain. The boundary conditions are: 1) zero electric field at \( x = \pm\infty \), 2) continuous electric field and potential at the source/channel and drain/channel junctions, 3) \( E_F - E_C = qV_T \) at the \( n^+ \) source, and \( E_V - E_F = qV_T \) at the \( p^+ \) drain, and 4) the valence band along the channel is aligned with the Fermi level at the source, at zero gate bias. This guarantees low off-state current at zero gate bias and a small turn-on voltage.

For this analysis, the supply voltage is limited to half the GNR’s band gap where ambipolar effects can be ignored. The computed band diagrams for a GNR pTFET with a supply voltage of 0.1 V are shown in Fig. 5.2(a) for the off-state at \( V_{GS} = 0 \) and in (b) for the on-state at \( V_{GS} = -0.1 \) V, for a ribbon width \( w_{GRN} = 5 \) nm, gate length \( LCH = 20 \) nm, and \( t_{OX} = 1 \) nm.

In the on-state, the 1D Zener tunneling current is given by Eqs (2.3) and (2.5), substituting \( V_R \) by \( \Delta \Phi \),

\[
I_D = \frac{q^2}{\pi \hbar} T_{WK}^{GRN} V_T \ln \left( \frac{1}{2} \left( 1 + \cosh \frac{\Delta \Phi}{V_T} \right) \right)
\]  

(5.2)

Using Eq. (5.2), taking \( \Delta \Phi \) and \( \zeta \) from the band diagram calculations in Fig. 5.2 (b), the drain current density at room temperature is calculated as a function of the gate-to-source
Figure 5.2 Energy band diagrams for a GNR $p$-TFET at (a) off-state and (b) on-state with a supply voltage of 0.1 V. The calculated gate oxide thickness is 1 nm, the GNR width is 5 nm, and the gate length is 20 nm.

Voltage and shown in Fig. 5.3, dashed line. Here, the influence of mobile charge in the channel on the electrostatics is neglected, where the mobile charge could be injected from the source by tunneling or from the drain by thermionic emission. The off-state leakage is assumed to arise from thermal emission over the barrier, $q\Phi_B$, between source and the drain, see Fig. 5.2. This assumption is only true at long channel limit or for large bandgap ribbons. The resulting leakage current for the 1D case is given by $I_{OFF} = \frac{q^2 V_T}{\hbar \pi} \exp(-\Phi_B / q V_T)$ [64], which is 0.13 pA, somewhat lower than 1 pA observed in the long channel GNR FET off-state [61]; the lower current might be anticipated from the higher barrier, $q\Phi_B$, of the TFET vs. the FET.
The complementary $n$TFET with symmetric geometry is also shown in Fig. 5.3, solid line. From this calculation, a 0.1 V swing can turn a 5 nm wide GNR TFET from on at 800 nA/µm to off at 26 pA/µm, with an on-to-off ratio of more than 7 orders of magnitude. The on-state current per unit ribbon width is given by Eq. (5.2) divided by the ribbon width. In Fig. 5.3, the threshold voltage of GNR TFET is 0.001 V using a current per unit width of $100 \text{nA} / L_{CH}$ [51], which is used for high-performance MOSFETs to define the threshold voltage.

![Figure 5.3 Calculated transfer characteristics of GNR $n$- (solid line) and $p$-TFETs (dashed line) at room temperature using the same geometry as Fig. 5.2.](image)

**Figure 5.3** Calculated transfer characteristics of GNR $n$- (solid line) and $p$-TFETs (dashed line) at room temperature using the same geometry as Fig. 5.2.

### 5.2 Self-consistent model for GNR TFETs

For high current densities, the carrier injection into the channel will generate significant corrections to the band profile along the channel and cannot be neglected. A self-consistent iterative procedure is necessary to get the corrected current density. To
achieve self-consistency, first, the electrostatic band diagram is calculated from section 5.1. A high-k gate oxide is used with dielectric constant $\varepsilon_{OX} = 16\varepsilon_0$ and a thickness $t_{OX} = 2$ nm for better gate control. Second, the energy-dependent tunneling probability from the source $T_s(E)$ is calculated from the band diagram using the WKB approximation. Figure 5.4 shows that the first subband imaginary energy dispersion relation of a GNR by tight-binding simulation (dots) is in agreement with an analytical expression (dashed line)

$$E^{im} = \pm \sqrt{(\hbar v_F k_x)^2 + (E_G/2)^2},$$

where $E^{im}$ is the energy in the bandgap, and $k_x$ is the imaginary wavevector.

Figure 5.4 Tight-binding energy dispersion relation of the first and second subband of a graphene nanoribbon with the width of $15a$; $a = 0.246$ nm is the lattice constant of graphene. The solid lines are the real $E$-$k$ relation and the dots are the imaginary $E$-$k$ relation in the bandgap. The dashed lines show Eq. (5.3) fit to the first subband in the bandgap. The tight-binding simulation is courtesy of Tian Fang, Notre Dame.

The potential barrier seen by the carrier for tunneling with energy $E$ is

$$(E_C - E, E - E_V)_< = \frac{E_G}{2} - |E^{im}| = \frac{E_G}{2} - \sqrt{\left(\hbar v_F k_x\right)^2 + \left(E_G/2\right)^2}$$

$(E_C - E, E - E_V)_<$ is the smaller one between $E_C - E$ and $E - E_V$. So
\[ k_x = \sqrt{\left[ E_C(x) - E - E_G / 2 \right]^2 - \left( E_G / 2 \right)^2} , \]  

(5.5)

and

\[ T_S(E) = \exp(-2 \left| k_x \right| dx) = \exp\left( -\frac{2}{\hbar v_F} \int_{x_i}^{x_f} \sqrt{\left( E_G / 2 \right)^2 - \left( E_C(x) - E - E_G / 2 \right)^2} dx \right) \]  

(5.6)

where \( x_i \) is the initial position and \( x_f \) is the final position of tunneling. Equation (5.6) is used to calculate the tunneling through all the paths, including the direct source to drain tunneling. The charge injection from both the source and the drain are considered and calculated separately by [65],

\[ Q_S = q \left[ \rho_{GNR}(E)(1 - f_S(E))T_S(E)dE \right] \]  

(5.7)

and

\[ Q_D = q \left[ \rho_{GNR}(E)\left( f_D(E) - 1\right)T_S(E) + 2\left( 1 - f_D(E)\right)\right]dE \]  

(5.8)

where \( \rho_{GNR}(E) = \frac{1}{\pi \hbar v_F} \frac{E}{\sqrt{E^2 - (E_G / 2)^2}} \) \((E > E_G/2)\) is the density of states for the first subband of the GNR. \( f_S \) and \( f_D \) denote the Fermi-Dirac distributions at the source and the drain respectively. Ballistic transport is assumed as in CNT MOSFETs [66]-[67], so that the charge is uniformly distributed in the channel and has a unit of c/m. Third, the total injected charge \( Q = Q_S + Q_D \) is added to the channel and the electrostatics are recomputed by Eq. (5.1) to determine the band diagram. Finally, the second and the third steps are repeated until convergence is achieved, \( \Phi_{n+1}^F - \Phi_{n}^F < 0.0001 \) eV, where \( \Phi_{n}^F \) is the channel surface potential calculated at the \( n \)th iteration. In this way, the potential profile \( \Phi_F(x) \) and the injected charge \( Q \) achieve the self-consistency. The drain current is calculated by [23] \( I_D = \frac{q}{\pi \hbar} \int \left( f_D(E) - f_S(E)\right)T_S(E)dE \) . A MATLAB program for this
iteration is attached as Appendix C. Computation time for the self-consistent calculations of each data point in Figs 5.5(a) and (b) is less than 1 minute on an Acer laptop with a 2.2 GHz AMD processor.

5.3 Geometry dependence

To design a GNR TFET, there are three important geometry variables to be considered: the ribbon width, the gate length, and the effective oxide thickness (EOT). Here, the effective oxide thickness is fixed to be 0.5 nm with high-k gate oxide $\varepsilon_{OX} = 16 \varepsilon_0$ and $t_{OX} = 2$ nm. Performance is scaled with varied ribbon widths and gate lengths.

The computed transfer characteristics of GNR $p$-TFETs are shown in Fig. 5.5 with (a) ribbon width $w = 5$ nm for different channel lengths, and (b) channel length $L_{CH} = 20$ nm for different ribbon widths. In Fig. 5.5 (a), the subthreshold swing $S$ and the off-state current $I_{OFF}$ are degraded at short channel length due to direct source to drain tunneling, but can be significantly improved in the long channel limit by suppression of direct tunneling. The subthreshold swing at $L_{CH} = 100$ nm is shown to be 2.8 mV/decade without scattering considerations. For short channel lengths, the off-state behavior can be improved by using narrower ribbon widths, Fig. 5.5(b). Down to a certain limit, the on-state current density is not significantly affected, which has favorable implications on device immunity to width variations, and device optimization.
The intrinsic gate delay, the ratio of the channel charge induced by the gate to the current to turn the transistor on, is an important measure in a high-performance MOSFET. A 17% reduction in gate delay per year has been a key goal for device scaling [67]. Figure 5.6 shows (a) the intrinsic gate delay and (b) total gate delay vs. $I_{OFF}$ performance of GNR $p$-TFETs with $L_{CH} = 20$ nm for different ribbon widths at a supply voltage $V_{DD} = 0.1$ V. The intrinsic gate delay is given by $\tau_{INT} = (Q_{ON} - Q_{OFF})/I_{ON}$, where $Q_{ON}$ is the on-state charge in the channel, $Q_{OFF}$ is the off-state charge, $V_{ON} - V_{OFF}$ is equal to the supply voltage 0.1 V, and $I_{ON}$ is the on-state current. The total gate delay is approximated by the sum of intrinsic gate delay and parasitic delay due to outer-fringe capacitance $\tau_{PAR} = C_{PAR}V_{DD}/I_{ON}$. The outer-fringe capacitance $C_{PAR}$ is given by [68]

$$C_{PAR}/w = \frac{2e_{SiC}}{\pi} \ln(1 + \frac{t_{GATE}}{t_{OX}})$$  \hspace{2cm} (5.9)
$C_{PAR}/w = 100$ aF/µm per side, assuming the ratio of gate thickness and oxide thickness $t_{GATE}/t_{OX} = 40$, and the dielectric constant of the spacers $\varepsilon_{sp} = 5\varepsilon_0$. It is shown that smaller ribbon widths can lower $I_{OFF}$ by sacrificing the intrinsic speed, but the total gate delay is not considerably affected. At $L_{CH} = 20$ nm, GNR TFETs with ribbon width of 2 nm and 3 nm can meet both the intrinsic speed requirement of 2012 high performance logic technology ($\tau_{INT} < 0.51$ ps) and the off-state leakage requirement of 2012 low operating power technology ($I_{OFF} < 18.3$ nA/µm) [3]. Interestingly, the intrinsic gate delay increases with $I_{OFF}$ for larger bandgap ribbon ($w = 2$ nm), decreases with $I_{OFF}$ for smaller bandgap ribbon ($w = 5$ nm) and almost keeps constant for intermediate bandgap ribbons ($w = 3$ nm, 4 nm), which is different from MOSFET characteristics where better performance can always be achieved by sacrificing the off-state leakage [69].

Figure 5.6 (a) Calculated intrinsic gate delay and (b) total gate delay vs. $I_{OFF}$ performance of GNR $p$-TFETs with $L_{CH} = 20$ nm for different ribbon widths at a supply voltage of 0.1 V.
This phenomenon can be explained by Figs. 5.7 and 5.8. Figure 5.7 calculates intrinsic gate delay vs. $I_{OFF}$ performance of GNR $p$-TFETs with $L_{CH} = 20$ nm and (a) $w = 5$ nm, (b) $w = 2$ nm at $V_{DD} = 0.1$ V. The intrinsic gate delay $\tau_{INT}$ (solid line) is separated to $\tau_S$ (dashed line) and $\tau_D$ (dotted line), which corresponds to the charge injection from the source and drain respectively. As $I_{OFF}$ increases, thermal carrier back-injection from the drain is increased, so $\tau_D = (Q_{D,ON} - Q_{D,OFF})/I_{ON}$ increases; this degrades performance; On the other hand, with increasing $I_{OFF}$ tunneling injection from the source is increased, so $\tau_S = (Q_{S,ON} - Q_{S,OFF})/I_{ON}$ decreases, which improves the performance. For $w = 5$ nm, there is enough source injection, the second effect dominant and the intrinsic gate delay improves; for $w = 2$ nm, the source injection is low, so the thermal back-injection from the drain is dominant and the intrinsic gate delay degrades. For intermediate bandgap ribbons ($w = 3$ and 4 nm) the two effects cancel each other, giving rise to a flat $\tau_{INT}$ vs. $I_{OFF}$ curve.

Figure 5.7 Calculated intrinsic gate delay vs. $I_{OFF}$ performance of GNR $p$-TFETs with $L_{CH} = 20$ nm and (a) $w = 5$ nm, (b) $w = 2$ nm at a supply voltage of 0.1 V. The intrinsic gate delay $\tau_{INT}$ (solid line) is separated to $\tau_S$ (dashed line) and $\tau_D$ (dotted line), which corresponds to the charge injection from the source and drain respectively.
Figure 5.8 Calculated gate capacitance vs. $V_{GS}$ of GNR $p$-TFETs with $L_{CH} = 20$ nm and (a) $w = 5$ nm, (b) $w = 2$ nm at $V_{DS} = -0.1$ V. The gate capacitance $C_G$ (solid line) is separated to $C_{GS}$ (dashed line) and $C_{GD}$ (dotted line), which corresponds to the charge injection from the source and drain respectively.

Figure 5.8 shows the gate capacitance vs. $V_{GS}$ of GNR $p$-TFETs with $L_{CH} = 20$ nm and (a) $w = 5$ nm, (b) $w = 2$ nm at $V_{DS} = -0.1$ V. The gate capacitance $C_G = dQ/dV_{GS}$ (solid line) is separated to $C_{GS} = dQ_S/dV_{GS}$ (dashed line) and $C_{GD} = dQ_D/dV_{GS}$ (dotted line), which corresponds to the charge injection from the source and drain respectively. It is seen that for $w = 5$ nm, there is large source injection, so the intrinsic gate delay is improved for larger gate-to-source bias, which corresponds to larger $I_{OFF}$. But for $w = 2$ nm, the drain back-injection is always dominant, so intrinsic delay degrades for larger $|V_{GS}|$ and $I_{OFF}$.

5.4 Quantum capacitance

Traditionally, the gate capacitance $C_G$ is assumed equal to the gate oxide capacitance $C_{OX}$ at the on-state, which means the change of the semiconductor surface potential with the gate voltage is zero [68]. This is not a good approximation as the oxide thickness
scales down or for low-dimensional systems [22]-[23], [70]-[73]. Quantum capacitance \( C_Q \) must be introduced first to account for the 2D electron gas in a quantum well or inversion layer [71], or in a 1D system [22]-[23], [72]. The total gate capacitance in these cases is modeled as an oxide capacitance in series with the quantum capacitance.

In the off-state, since there is no charge, the surface potential of GNR in the channel, \( \Phi_F \), is equal to \( V_{GS} \), which is zero. As \( |V_{GS}| \) increases, charges \( Q \) are induced in the channel and induce a surface potential of the opposite sign, so the surface potential \( \Phi_F \) is not equal to but less than \( V_{GS} \) in the on-state:

\[
\Phi_F = V_{GS} - Q / C_{OX}
\]  

(5.10)

The change in channel charge with changing gate potential is given by

\[
\frac{dQ}{dV_{GS}} = \frac{dQ}{d\Phi_F} \frac{d\Phi_F}{dV_{GS}}
\]  

(5.11)

Substituting Eq. (5.10) into Eq. (5.11) gives

\[
\frac{dQ}{dV_{GS}} = \frac{C_{OX}(dQ / d\Phi_F)}{C_{OX} + dQ / d\Phi_F}
\]  

(5.12)

where \( dQ / d\Phi_F \) has the same unit as a capacitor, and is defined as quantum capacitance \( C_Q \), which is proportional to the density of states (DOS) [23], [73].

\[
\frac{dQ}{dV_{GS}} = \frac{C_{OX}C_Q}{C_{OX} + C_Q}
\]  

(5.13)

For conventional MOSFETs, the 3D DOS increases with the gate potential energy so \( C_Q >> C_{OX} \) in the on-state and \( C_G \approx C_{OX} \). As the oxide thickness shrinks, \( C_{OX} \) is increased
to be comparable to $C_Q$, so the effect of quantum capacitance needs to be considered. For 2D system, DOS is a constant and for a 1D system DOS decreases with the potential energy, $C_Q$ usually cannot be neglected [22]-[23], [72]-[73]. Furthermore, $C_Q \ll C_{OX}$ ($C_Q \approx C_G$) can be achieved in 1D system with ultra-thin gate oxide [22]-[23], [73], which means the change of the semiconductor surface potential with the gate voltage $d\Phi_F/dV_{GS}$ = $C_G/C_Q$ is close to “1”. This ideal case is called “quantum capacitance limit” [73], where the surface potential of the channel is only determined by the gate and will not be influenced by the drain.

In section 5.2-5.5, the effective oxide thickness $EOT$ is fixed at 0.5 nm, so the gate oxide capacitance density $C_{OX}/w = \varepsilon_{SiO_2}/EOT$ gives $C_{OX} = 350$ aF/µm and 140 aF/µm for $w = 5$ and 2 nm respectively. The red lines in Fig. 5.9 show the gate capacitance $C_G$ vs. $V_{GS}$ of GNR $p$-TFETs with $L_{CH} = 20$ nm and (a) $w = 5$ nm, (b) $w = 2$ nm at $V_{DS} = -0.02$ V. It is seen that at $V_{GS} = -0.04$ V, for $w = 5$ nm, $C_G = 98$ aF/µm < $C_{OX}$, which means the quantum capacitance cannot be neglected: $C_Q = 140$ aF/µm, so $C_G/C_Q = 0.7$, close to the “quantum limit”. Similarly, for the 2 nm wide GNR TFET, $C_G = 64$ aF/µm, $C_Q = 120$ aF/µm, $C_G/C_Q = 0.53$, not as good gate control as the 5 nm one. These capacitance calculations can explain the “slow turn-on” in Fig. 10, which is also observed in CNT TFET simulations [23]. For a small $V_{DS}$ and $V_{GS}$, the injection, mainly from drain, increases the quantum capacitance, so the movement of the surface potential in the channel with changing $V_{GS}$ slows down. Especially for the large bandgap ribbon ($w = 2$ nm), the source injection (blue line) is suppressed by the drain injection (black line) at low $V_{DS}$.

57
Figure 5.9 Calculated gate capacitance vs. $V_{GS}$ of GNR $p$-TFETs with $LCH = 20$ nm and (a) $w = 5$ nm, (b) $w = 2$ nm at $V_{DS} = -0.02$ V. The gate capacitance $C_G$ (red line) is separated to $C_{GS}$ (blue line) and $C_{GD}$ (black line), which corresponds to the charge injection from the source and drain respectively.

Figure 5.10 Calculated drain current density vs. $V_{DS}$ of GNR $p$-TFETs with $LCH = 20$ nm and (a) $w = 5$ nm, (b) $w = 2$ nm.

5.5 Temperature dependence

The previous calculations are all based on room temperature operation. Figure 5.11 shows the temperature characteristics. As discussed before, both the on-state and off-state
currents are due to tunneling, not thermal emission, so they both decrease with temperature due to the expansion of the Fermi function. As the temperature increases from room temperature to 100 °C, the on-to-off ratio does not degrade.

![Graph showing on-state current density and on-to-off ratio vs. temperature](image)

**Figure 5.11** On-state current density (solid line) and on-to-off ratio (dotted line) vs. temperature for 0.5 nm \( EOT \) gate oxide, 20 nm gate length GNR TFETs with ribbon width of 2 and 5 nm.

### 5.6 Comparison with conventional nano-CMOS technology

The intrinsic performance of the proposed GNR TFET (\( w = 2 \) nm and \( 3 \) nm, \( EOT = 0.5 \) nm, \( L_{CH} = 20 \) nm) is summarized and compared with ITRS projected 2012 high-performance and low-power \( n \)MOSFETs [3], Table 5.1. It is shown that the GNR TFETs can switch from on to off with only 0.1 V gate swing. The 2 nm wide GNR can provide 2.8× higher intrinsic speed than the high-performance MOSFET, 42× lower dynamic power and 100× lower off-state power dissipation than the low-power MOSFET. The 3 nm wide GNR can provide 5.5× higher intrinsic speed than the high-performance...
MOSFET, 41\times lower dynamic power and 5.2\times lower off-state power dissipation than the low-power MOSFET.

**TABLE 5.1 SPEED AND POWER ESTIMATES COMPARING 2012 NMOSFETS TARGETS FROM THE 2008 UPDATES ITRS ROADMAP [3] WITH 2 NM AND 3 NM GNR TFETS USING THE GEOMETRY IN FIG. 5.1**

<table>
<thead>
<tr>
<th></th>
<th>2012 nMOSFET</th>
<th>GNR TFET</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gate length L_G</td>
<td>20</td>
<td>22</td>
</tr>
<tr>
<td>Equivalent oxide thickness EOT</td>
<td>0.75</td>
<td>0.90</td>
</tr>
<tr>
<td>Supply voltage V_DD</td>
<td>1</td>
<td>0.7</td>
</tr>
<tr>
<td>Drive current I_D</td>
<td>1639</td>
<td>682</td>
</tr>
<tr>
<td>Off-state leakage current I_OFF</td>
<td>700</td>
<td>18.30</td>
</tr>
<tr>
<td>Intrinsic speed ~ I_D / C_G V_DD</td>
<td>1961</td>
<td>1205</td>
</tr>
<tr>
<td>Off-leakage power ~ I_OFF V_DD</td>
<td>0.7</td>
<td>0.013</td>
</tr>
<tr>
<td>Dynamic power ~ 1/2 I_D V_DD</td>
<td>820</td>
<td>238.7</td>
</tr>
</tbody>
</table>

In summary, the first GNR TFET was proposed and published in [25], Appendix F. A compact model was developed that treated energy dependent transmission coefficients, direct source-to-drain tunneling, and self-consistent channel electrostatics. The geometry dependent performance scaling with ribbon widths from 2 to 5 nm and gate lengths from 20 to 100 nm was discussed. For 20 nm channel length, GNR tunnel transistors with ribbon width of 2 and 3 nm could achieve high performance and low operating power simultaneously, meeting 2012 ITRS targets [3]. The simulations results from the compact model were submitted to *IEEE Electron Device Letters*. 
CHAPTER 6

CONCLUSIONS AND SUGGESTED FUTURE WORK

6.1 Summary

It has been a goal of this research to understand the physics of interband tunnel transistors and benchmark it for high-performance and low-power logic transistor applications. Analytic expressions for Zener tunneling in 1D, 2D, and 3D semiconductors were derived to establish the guidelines for tunnel transistor design. An analytic expression was derived, which showed that the subthreshold swing of interband tunnel transistors was a function of the gate-to-source voltage and can beat the MOSFET limit of 60 mV/decade about certain gate bias. Based on this expression, a new fully-depleted interband tunnel transistor structure was proposed and designed. The low subthreshold swing was verified by Synopsys TCAD simulation. Ge interband tunnel transistors were shown by simulation to exhibit improved on-state performance vs. Si, because of the smaller bandgap and effective mass. Compared with a 2013 nMOSFET at 0.5 V supply voltage, Ge interband tunnel transistors can save 2× in power and 20× in energy (for a 50-stage inverter chain with an activity factor of 2%) while not sacrificing speed. To realize the proposed Ge interband tunnel transistor, a rapid melt growth process was developed to form submicron $p^+n^+$ Ge tunnel junctions. Ge $p^+n^+$ diodes with sizes ranging from $30 \times 30 \ \mu$m to $300 \times 300 \ \text{nm}$ were fabricated. Transmission electron
microscopy revealed the regrown film and a contact microstructure consistent with the Al-Ge phase diagram. Negative differential resistances were observed which indicated the junction was abrupt heavily-doped. The peak-to-valley current ratio of these devices was low and not improved by reducing the area down to 0.1 μm², which is likely a result of point defects or junction doping nonuniformity, as TEM showed no dislocations at the regrown junction.

A GNR TFET was first proposed and then modeled analytically by quasi-1D Poisson equation. An improved compact model was developed that treated energy-dependent transmission coefficients, direct source-to-drain tunneling, and self-consistent channel electrostatics. Graphene nanoribbons have a width-tunable bandgap and ultra-thin body layer, which is especially favorable for TFET applications, and more amenable to planar processing and large-scale integration than CNTs. Ribbon widths between 2 and 5 nm were considered, corresponding to energy bandgaps in the range of 0.69 to 0.28 eV. It was shown by simulation that the GNR tunnel transistors at long channel limit could operate at 0.1 V with an ultra-low subthreshold swing of 2.8 mV/decade, but the subthreshold swing and off-state current were degraded at short channel length due to direct source to drain tunneling. Smaller ribbon widths (down to a certain limit) could significantly improve the off-state behavior without considerably affecting the on-state current density and speed. For 20 nm channel length, GNR tunnel transistors with ribbon width of 2 and 3 nm can achieve high performance and low operating power simultaneously, meeting 2012 ITRS targets. The 2 nm wide GNR can provide 2.8× higher intrinsic speed than the high-performance MOSFET, 42× lower dynamic power and 100× lower off-state power dissipation than the low-power MOSFET. The 3 nm wide
GNR can provide 5.5× higher intrinsic speed than the high-performance MOSFET, 41× lower dynamic power and 5.2× lower off-state power dissipation than the low-power MOSFET. This work contributes to the reduction in power dissipation and energy consumption in semiconductor devices beyond the 32 nm technology node.

6.2 Suggested future research

Low subthreshold swing tunnel transistors have been built in carbon nanotubes, Si and SiGe, but the on-state current density is still much lower than high-performance MOSFETs and the supply voltage thus far is not reduced. To enhance the on-state current density, narrow bandgap materials with smaller effective mass or graphene nanoribbons should be used for interband tunnel transistor applications. But the narrow bandgap may also increase the off-state leakage. Heterojunctions, i.e. InAs/InGaAs, 2 nm GNR/ 5 nm GNR can be considered to keep high electric fields at the tunnel junction while presenting a large bandgap for source to drain direct tunneling or thermal emission.

For the simulation, self-consistent models for tunneling devices are already included in Synopsys TCAD, which should be used for semiconductor tunnel transistor modeling in the future. The GNR TFET simulations can be further improved by considering scattering in the channel to make it more realistic. Finally, the first experimental demonstration of high-performance, low-power GNR TFETs will be more than exciting!
APPENDIX A

A.1 DERIVATION OF 1D ZENER TUNNELING CURRENT IN DIRECT SEMICONDUCTORS (Eq. 2.3)

\[ I^{1D} = \int qv_x(k_x)\rho(k_x)(f_f - f_e)T_{WKB}^{1D}dk_x \]

\[ = \int \frac{q}{\hbar} \frac{1}{dk_x} \frac{1}{\pi} \left( \frac{1}{1 + \exp((E - qV_R) / qV_T)} - \frac{1}{1 + \exp(E / qV_T)} \right)T_{WKB}^{1D}dk_x \]

\[ = \frac{q}{\pi\hbar} T_{WKB}^{1D} \left[ dE \left( \frac{1}{1 + \exp((E - qV_R) / qV_T)} - \frac{1}{1 + \exp(E / qV_T)} \right) \right]^{qV_R \mid_{E=0}}_{qV_R} \]

\[ = \frac{q^2}{\pi\hbar} T_{WKB}^{1D} V_T \left( -\ln 2 + \ln(1 + \exp(-V_R / V_T)) + \ln(1 + \exp(V_R / V_T)) - \ln 2 \right) \]

\[ = \frac{q^2}{\pi\hbar} T_{WKB}^{1D} V_T \ln \left( \frac{1 + \exp(-V_R / V_T)}{4} \frac{1 + \exp(V_R / V_T)}{4} \right) \]

\[ = \frac{q^2}{\pi\hbar} T_{WKB}^{1D} V_T \ln \left( \frac{2 + \exp(-V_R / V_T) + \exp(V_R / V_T)}{4} \right) \]

\[ = \frac{q^2}{\pi\hbar} T_{WKB}^{1D} V_T \ln \left( \frac{1}{2} \left( 1 + \cosh \frac{V_R}{V_T} \right) \right) \]
A.2 DERIVATION OF 2D ZENER TUNNELING CURRENT IN DIRECT SEMICONDUCTORS (Eq. 2.11)

\[ E_y = \frac{\hbar^2 k_y^2}{2m^*_y} \]

\[ \therefore \frac{dE_y}{dk_y} = \frac{\hbar^2}{2m^*_y} 2k_y = \frac{\hbar^2}{m^*_y} 2 \frac{m^*_y E_y}{\hbar} = \hbar \frac{2E_y}{m^*_y} \]

\[ \therefore dk_y = dE_y \frac{1}{\hbar} \sqrt{\frac{m^*_y}{2E_y}} \]

Assuming \( f_y - f_C \approx 1 \),

\[ J^{2D} = \iiint qv_y (k_x) \rho_x (k_x) \rho_y (k_y) dk_x dk_y (f_y - f_C) T_{WKB}^{2D} \approx \iiint q \left( \frac{1}{h} \right) \frac{1}{\pi} \frac{1}{2\pi} dk_x \left( dE_y \frac{1}{h} \sqrt{\frac{m^*_y}{2E_y}} \right) T_{WKB}^{1D} \exp(-\frac{E_{zn}}{E}) \exp(-\frac{E_y}{E}) \]

\[ = \frac{q \sqrt{m^*_y T_{WKB}^{1D}}}{2\sqrt{2\pi^2 h^2}} \exp(-\frac{E_{zn}}{E}) \int_1^1 \exp(-\frac{E_y}{E}) dE_y dE_y \]

\[ = \frac{q \sqrt{m^*_y T_{WKB}^{1D}}}{2\sqrt{2\pi^2 h^2}} \exp(-\frac{E_{zn}}{E}) \int_0^{qV_R-E_y} \int_0^{qV_R-E_y} \frac{1}{\sqrt{2\pi\hbar^2}} \exp\left(-\frac{E_y}{E}\right) dE_y dE_y \]

\[ = \frac{q \sqrt{m^*_y T_{WKB}^{1D}}}{2\sqrt{2\pi^2 h^2}} \exp\left(-\frac{E_{zn}}{E}\right) \sqrt{E} \int_0^{qV_R-E_y} \exp\left(\sqrt{E} (qV_R - E_y) / E\right) dE_y \]

where \( \int_0^{qV_R-E_y} \exp\left(-\frac{E_y}{E}\right) dE_y = \text{Erf}\left(\sqrt{E} (qV_R - E_y) / E\right) \) is given by Mathematica.
APPENDIX B

GERMANIUM RAPID-MELT-GROWTH TUNNEL JUNCTION PROCESS TRAVELER

Rapid Thermal Diffusion of Phosphorus Day 1 Date / /

- Remove SOD from refrigerator
- Allow SOD to warm up to room temperature. Normally several hours

Measure 4-Probe Resistivity

- Measure the resistivity of the device wafer _______ Ω cm ( ___type )

Clean-up

- Degrease the Ge wafer in hot acetone and methanol for 5 minutes each
- Blow dry
- Etch off surface oxide in buffered 10:1 HF 25 s
- Rinse in DI water and blow dry
- Bake the wafer at 100 °C for 2 min to further dry the surface
- Inspect

Spin-on Phosphosilicate Film

Cover spinner ring with aluminum foil

- Drip coat source wafers with Emulsitone Phosphosilicafilm 5 x 10²⁰ spin-on diffusant
- Spin program: 1000 rpm 10 s, 3000 rpm 30 s. (thickness ~ 550 nm)
- Bake the wafer at 150 °C for 20 min.

Rapid Thermal Diffusion

- Verify the RTA program by running on a test wafer, program 800C300S.RPD.
  Rapid thermal anneal: 800 °C 300 s, ramp rate 30°C/s, forming gas ambient (N₂:H₂ 95:5), expect doping concentration 2 x 10²⁰ cm⁻³.

---

1 Veeco Model FPP-5000
2 Phosphosilicafilm SOD expires 12 months after opening
3 RTP-600s Rapid Thermal Processing System by Modular Process Technology Corp.
### Time (s) Temp (°C) gas1: N₂ gas3 N₂:H₂

<table>
<thead>
<tr>
<th>State</th>
<th>Time (s)</th>
<th>Temp (°C)</th>
<th>gas1: N₂</th>
<th>gas3 N₂:H₂</th>
<th>Factors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Idle</td>
<td>90</td>
<td>25</td>
<td>0</td>
<td>10</td>
<td>1 2</td>
</tr>
<tr>
<td>Ramp</td>
<td>27</td>
<td>800</td>
<td>0</td>
<td>10</td>
<td>2 1</td>
</tr>
<tr>
<td>Hold</td>
<td>300</td>
<td>800</td>
<td>0</td>
<td>10</td>
<td>3 1</td>
</tr>
<tr>
<td>Idle</td>
<td>180</td>
<td>0</td>
<td>10</td>
<td>0</td>
<td>4 1</td>
</tr>
</tbody>
</table>

### Factors
- Rapid thermal anneal the device wafer, record actual annealing temperatures and times.

### Peak Temp. Cooling start Cooling

<table>
<thead>
<tr>
<th>Peak Temp.</th>
<th>Cooling start</th>
<th>Time (s) at</th>
<th>Cooling</th>
</tr>
</thead>
<tbody>
<tr>
<td>T_p  (°C)</td>
<td>time (s) at T_p</td>
<td>T_p - 50 °C</td>
<td>T_p - 100 °C</td>
</tr>
</tbody>
</table>

- Etch the Ge sample in buffered HF for 5 min to remove SOD.
- Rinse in DI.⁴ 2 min.
- Observe the surface.

### Alignment Mark Formation

<table>
<thead>
<tr>
<th>Day 2</th>
<th>Date / /</th>
</tr>
</thead>
</table>

### Clean-up
- Degrease the Ge wafer in hot acetone and methanol for 5 minutes each
- Blow dry

### PMMA and Copolymer MMA(8.5)MAA⁵ Coating

- Set hotplates to 150 °C and 180 °C
- Load Ge wafer on spinner, vacuum on. Use glass pipette and
- Dispense one-quarter pipette of copolymer MMA(8.5)MAA onto wafer
- Ramp 2000 rpm/s, spin 1000 rpm, 90 s
  (Expect ~2500 Å thickness – Coating should become uniform after 65 s)
- Hot plate bake 150 °C, 75 s
- Load sample on spinner, vacuum on. Use new glass pipette and
- Dispense one-quarter pipette of PMMA 9506 onto sample
- Ramp 2000 rpm/s, spin 4000 rpm, 45 s (expect ~1250 Å thick layer)
- Hot plate bake 180 °C, 75 s

### Electron-Beam Lithography Elionix ELS-7700, 75 keV

Mask: GeTD2007_Layer 0 AlignMark
- Record CON filenames: qzhang1/GeTD2007/align.con
- Record the die position
  - Dx, Dy: __________

### Develop
- Mix developer⁷ if necessary (50 ml MIBK: 150 ml IPA: 18 drops MEK)
- Develop wafers in 1MIBK: 3IPA: 0.015MEK solution for 30 s

---

⁴ deionized water, resistivity 18 MΩ cm
⁵ polymethyl methacrylate C₅H₉O₂, methyl mannose C₅H₁₀O₆, and methylarsonic acid CH₅AsO₃
⁶ 950,000 molecular weight
⁷ methyl isobutylketone C₅H₁₂O, methylethylketone C₄H₈O, isopropyl alcohol C₃H₇O
- Bottle rinse in IPA 30 s
- Blow dry in N₂
- Inspect under microscope to make sure alignment marks are visible
- Develop as needed, record total develop time
  - ______ s
- UV Ozone ⁸ clean 120 s  Record UV/ozone O₂ inlet pressure ______ psi
  99.999% O₂ purity, 10 psi inlet pressure, removes photoresists ~0.25 Å/s
- Load without delay

**Evaporate Alignment Marks (FC1800)⁹ # 2**

- Pump chamber to < 2 x 10⁻⁶ Torr  Record base pressure ___ x 10⁻⁷ Torr
- FC1800 system set points – check and record actual setting

<table>
<thead>
<tr>
<th>Evaporation parameters</th>
<th>Beam parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metal</td>
<td>Ti</td>
</tr>
<tr>
<td>Tooling*</td>
<td>180</td>
</tr>
<tr>
<td>Density (g/cm³)</td>
<td>4.5</td>
</tr>
<tr>
<td>Z-ratio</td>
<td>0.628</td>
</tr>
<tr>
<td>Rate (Å/s)</td>
<td>5</td>
</tr>
<tr>
<td>Thickness (Å)</td>
<td>200</td>
</tr>
</tbody>
</table>

* This tooling is for the flat bar.

- Evaporation notes:
- Vent and unload wafers

**Lift-off**

- Transfer wafers into acetone
- Soak in hot acetone 5 min and start timer  Record actual soak time _____ min
- Acetone spray to complete lift-off
- Soak in methanol 2 min
- Blow dry with N₂

**Al/Ge Emitter Formation**

**Day 3**  **Date** / /

**Clean-up**

- Degrease the Ge wafer in hot acetone and methanol for 5 minutes each
- Rinse in D.I. water and blow dry
- Etch off surface oxide in buffered 10:1 HF 25 s
- Rinse in DI water and blow dry
- Bake the wafer at 100 °C for 2 min to further dry the surface

---

⁸ Jelight Company Model 144AX
⁹ Airco Temescal electron-beam evaporator
**PMMA and Copolymer MMA(8.5)MAA Coating**


- Set hotplates to 150 °C and 180 °C
- Load Ge wafer on spinner, vacuum on. Use glass pipette and
- Dispense one-quarter pipette of copolymer MMA(8.5)MAA onto wafer
- Ramp 2000 rpm/s, spin 1000 rpm, 90 s
  (Expect ~2500 Å thickness – Coating should become uniform after 65 s)
- Hot plate bake 150 °C, 75 s
- Load sample on spinner, vacuum on. Use new glass pipette and
- Dispense one-quarter pipette of PMMA 950 onto sample
- Ramp 2000 rpm/s, spin 4000 rpm, 45 s (expect ~1250 Å thick layer)
- Hot plate bake 180 °C, 75 s

**Electron-Beam Lithography Elionix ELS-7700, 75 keV**

Mask: GeTD2007_ Layer 1 Emitter

- Compute e-beam write time using Excel workbook below

<table>
<thead>
<tr>
<th>Settings</th>
<th>Calculations</th>
<th>Actual</th>
</tr>
</thead>
<tbody>
<tr>
<td># of Die*</td>
<td>Pattern Area (μm²)</td>
<td>Beam I (pA)</td>
</tr>
<tr>
<td>1</td>
<td>80000</td>
<td>250</td>
</tr>
<tr>
<td>1</td>
<td>120000</td>
<td>250</td>
</tr>
</tbody>
</table>

* Die is what is layed out under Mentor Graphics

The manufacturer recommended dose range is from 50–500 μC/cm² (Lili Ji recommends 600-800 μC/cm²)

Comments:
- Record the position of alignment marks
  - mark A______       B______
- Record actual write time in table above
- Record CON filenames: qzhang1/GeTD2007/emitter.con

**Develop**

- Mix developer if necessary (50 ml MIBK: 150 ml IPA: 18 drops MEK)
- Develop wafers in 1MIBK: 3IPA: 0.015MEK solution for 30 s
- Bottle rinse in IPA 30 s
- Blow dry in N₂
- Inspect under microscope to make sure alignment marks are visible
- Develop as needed, record total develop time
  - _______ s
- UV Ozone clean 120 s  
  Record UV/ozone O₂ inlet pressure ______ psi  
  99.999% O₂ purity, 10 psi inlet pressure, removes photoresists ~0.25 Å/s
- Load without delay

**Evaporate TD emitter (FC1800 # 2 )**

- Pump chamber to < 2 x 10⁻⁶ Torr  
  Record base pressure ___ x 10⁻⁷ Torr

69
- FC1800 system set points – check and record actual setting

<table>
<thead>
<tr>
<th>Evaporation parameters</th>
<th>Beam parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metal</td>
<td>Al</td>
</tr>
<tr>
<td>Tooling*</td>
<td>180</td>
</tr>
<tr>
<td>Density (g/cm³)</td>
<td>2.7</td>
</tr>
<tr>
<td>Z-ratio</td>
<td>1.08</td>
</tr>
<tr>
<td>Rate (Å/s)</td>
<td>7</td>
</tr>
<tr>
<td>Thickness (Å)</td>
<td>500</td>
</tr>
</tbody>
</table>

* This tooling is for the flat bar.

- Evaporation notes:
- Vent and unload wafer

**Lift-off**
- Transfer wafers into acetone
- Soak in hot acetone 5 min and start timer
  - Record actual soak time _____ min
- Acetone spray to complete lift-off
- Soak in methanol 2 min
- Blow dry with N₂

**Alpha-Step**
**Metal Height**
Step profile metal height and surface roughness in 3 locations; expect a thickness of 50 nm.

<table>
<thead>
<tr>
<th>Step height/roughness (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>left</td>
</tr>
<tr>
<td>center</td>
</tr>
<tr>
<td>right</td>
</tr>
</tbody>
</table>

**Al/Ge Emitter Formation – Continue**
**Day 4**
**Date / /**

**SEM Measurement**

**Clean-up**
- Degrease the Ge wafer in hot acetone and methanol for 5 minutes each
- Rinse in D.I. water and blow dry
- UV Ozone clean 120 s
  - Record UV/ozone O₂ inlet pressure _____ psi
  - 99.999% O₂ purity, 10 psi inlet pressure, removes photoresists ~0.25 Å/s

**Plasma-Enhanced Chemical Vapor Deposition**

---

10 KLA Tencor Model 500
11 Uniaxis 790
- Prepare, RCA clean, Si pilot wafer, load with samples
- Load samples
- Cap the Al/Ge with 50 nm SiNₓ (Recipe SiNxb)
  - SiH₄ 40 SCCM, NH₃ 4 SCCM
  - Heat exchanger 60 °C, substrate 250 °C
  - Power: 200 W
  - Pressure: 500 mTorr
  - Expected rate 20 nm/min
  - Time: 2.5 min (expecting 50 nm)

**Ellipsometry**
- Measure Si₃N₄ thickness using GEMP (Gaertner Ellipsometer Measurement Program) on Si pilot with incident angle 70°

<table>
<thead>
<tr>
<th>Wafer #</th>
<th>Location Si thickness (nm)</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Rapid Thermal Anneal**
- Run program 600C60.RPD
- Verify the RTA program by running on a test wafer
  600 °C, 1 s, ramp rate 100 °C/s, forming gas ambient (N₂:H₂ 95:5)

<table>
<thead>
<tr>
<th>Factors</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4.2</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Factors</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>4.2</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
</tr>
</tbody>
</table>

- Rapid thermal anneal each sample, record actual annealing temperatures and times

<table>
<thead>
<tr>
<th>Peak Temp.</th>
<th>Cooling start time (s) at T_p</th>
<th>Cooling Rate (°C/s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T_p (°C)</td>
<td>T_p - 50 °C</td>
<td>T_p - 100 °C</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>

inspect the surface

**Measurements**
- I-V measurement for big patterns
Via Etch
Day 5
Date / /

Clean-up
- Degrease the Ge wafer in hot acetone and methanol for 5 minutes each
- Blow dry

PMMA Coating
- Set hotplate to 180 °C
- Load Ge wafer on spinner, vacuum on.
- Using a new glass pipette, dispense one-quarter pipette of old PMMA 950 onto sample
- Ramp 2000 rpm/sec, spin 4000 rpm, 45 s (expect ~1250 Å thick layer)
- Bake 180 °C, 75 s
- Ramp 2000 rpm/sec, spin 4000 rpm, 45 s (expect ~1250 Å thick layer)
- Bake 180 °C, 75 s
- Ramp 2000 rpm/sec, spin 4000 rpm, 45 s (expect ~1250 Å thick layer)
- Bake 180 °C, 75 s
- Ramp 2000 rpm/sec, spin 4000 rpm, 45 s (expect ~1250 Å thick layer)
- Bake 180 °C, 75 s
- Total thickness ~ 5000 Å is expected

Electron-Beam Lithography Elionix ELS-7700, 75 keV
- Mask: GeTD2007_Layer 2 Via
  - Compute e-beam write time using Excel workbook below

<table>
<thead>
<tr>
<th># of Die</th>
<th>Pattern Area (μm²)</th>
<th>Beam I (pA)</th>
<th>Dose μC/cm²</th>
<th>Chip Size (μm)</th>
<th>Dots 60000</th>
<th>Dot space μm/dot</th>
<th>Time μs/dot</th>
<th>Write min.</th>
<th>Write totalmin</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>135000</td>
<td>250</td>
<td>400</td>
<td>300</td>
<td></td>
<td>5</td>
<td>0.4</td>
<td>36.0</td>
<td></td>
</tr>
</tbody>
</table>

* Die is what is layed out under Mentor Graphics

Comments:
- Record the position of alignment marks
  - mark A______ B______
- Record actual write time in table above
- Record CON filenames: gzhang1/GeTD2007/via.con

Develo
- Mix developer if necessary (50 ml MIBK: 150 ml IPA: 18 drops MEK)
- Develop wafers in 1MIBK: 3IPA: 0.015MEK solution for 30 s
- Bottle rinse in IPA 30 s
- Blow dry in N₂
- Inspect under microscope to make sure alignment marks are visible
• Develop as needed, record total develop time
  • _____ s
  • UV Ozone clean 120 s  Record UV/ozone O2 inlet pressure _____ psi

**SiNx RIE etch**

• Condition chamber (SF₆ 20 sccm @ 20 mTorr, 50 W for 2 min)
• Use SiNx on Si to determine the etch time
• Etch (SF₆ 20 sccm @ 20 mTorr, 50 W for _____ sec, DC _____ V (90-80)); overetch factor _____ (1.3)
• SiNx etch rate 1300 Å/min
• Etch stops when metal appears
• Remove PMMA by soaking in acetone/methanol

**SEM Imaging**

**Bond Pad Formation**

**Clean-up**

• Degrease the Ge wafer in hot acetone and methanol for 5 minutes each
• Blow dry

**PMMA and Copolymer MMA(8.5)MAA Coating**


• Set hotplates to 150 °C and 180 °C
• Load Ge wafer on spinner, vacuum on. Use glass pipette and
• Dispense one-quarter pipette of copolymer MMA(8.5)MAA onto wafer
• Ramp 2000 rpm/s, spin 1000 rpm, 90 s
  (Expect ~2500 Å thickness – Coating should become uniform after 65 s)
• Hot plate bake 150 °C, 75 s
• Load sample on spinner, vacuum on. Use new glass pipette and
• Dispense one-quarter pipette of PMMA 950 onto sample
• Ramp 2000 rpm/s, spin 4000 rpm, 45 s (expect ~1250 Å thick layer)
• Hot plate bake 180 °C, 75 s

**Electron-Beam Lithography Elionix ELS-7700, 75 keV**

Mask: GeTD2007_ Layer 3 Bondpad

• Compute e-beam write time using Excel workbook below

<table>
<thead>
<tr>
<th>Settings</th>
<th>Calculations</th>
<th>Actual</th>
</tr>
</thead>
<tbody>
<tr>
<td># of Die*</td>
<td>Pattern Area (µm²)</td>
<td>Beam I (pA)</td>
</tr>
<tr>
<td>1</td>
<td>360000</td>
<td>500</td>
</tr>
</tbody>
</table>

* Die is what is layed out under Mentor Graphics

Comments:

• Record the position of alignment marks
• mark A______       B______
• Record actual write time in table above
• Record CON filenames: qzhang1/GeTD2007/bondpad.con

Develop
• Mix developer if necessary (50 ml MIBK: 150 ml IPA: 18 drops MEK)
• Develop wafers in 1MIBK: 3IPA: 0.015MEK solution for 30 s
• Bottle rinse in IPA 30 s
• Blow dry in N2
• Inspect under microscope to make sure alignment marks are visible
• Develop as needed, record total develop time
• _______ s
• UV Ozone clean 120 s                    Record UV/ozone O2 inlet pressure ______ psi
  99.999% O2 purity, 10 psi inlet pressure, removes photoresists ~0.25 Å/s
• Load without delay

Evaporate Bondpad Metal (FC1800 # 2 )
• Pump chamber to < 2 x 10^-6 Torr                    Record base pressure ___ x 10^-7 Torr
• FC1800 system set points – check and record actual setting

<table>
<thead>
<tr>
<th>Evaporation parameters</th>
<th>Beam parameters</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metal</td>
<td>Ti</td>
</tr>
<tr>
<td>Tooling*</td>
<td>180</td>
</tr>
<tr>
<td>Density (g/cm³)</td>
<td>4.5</td>
</tr>
<tr>
<td>Z-ratio</td>
<td>0.628</td>
</tr>
<tr>
<td>Rate (Å/s)</td>
<td>5</td>
</tr>
<tr>
<td>Thickness (Å)</td>
<td>200</td>
</tr>
</tbody>
</table>

* This tooling is for the flat bar.

• Evaporation notes:
• Vent and unload wafer

Lift-off
• Transfer wafers into acetone
• Soak in hot acetone 5 min and start timer         Record actual soak time _____ min
• Acetone spray to complete lift-off
• Soak in methanol 2 min
• Blow dry with N2

Measurements
• I-V measurements for devices with size from 300 nm to 10 μm
APPENDIX C

MATLAB PROGRAM TO SELF-CONSISTENTLY CALCULATE BAND DIAGRAMS AND TUNNELING CURRENT DENSITIES FOR GNR TFETS

%Physical and Mathematica Constants
q=1.6e-19;hbar=1.05e-34;e0=8.84e-21;pi=3.14159;
%Graphene Constants
vF=1e6;tGNR=0.35;eGNR=16;
%Geometry Parameters
tox=2;w=3;L=20;
lamda=0.5916*sqrt(tox);Eg=1.375/w;
%Temperature
T=300;
Vt=0.0258*T/300;
%Electrical Parameters
fbi=-0.0258; Vds=-0.1; Vgs=0.0;
%Electrostatic Calculations
for j=1:1:36
    Qcs(1)=0;Qcd(1)=0;Qc(1)=0;
    Vgs(j)=-0.2+(j-1)*0.01;
    fsc(j)=+(Eg-Vgs(j))/2;
    fcd(j)=(2*Eg-Vgs(j)-Vds)/2;
    x1=-5:0.1:0;
    f1=fbi+fsc(j)*exp(x1/lamda);
    x2=0:0.1:L;
    f2=fbi+Eg-Vgs(j)-Qc(1)*lamda^2/eGNR/e0-fsc(j)*exp(-x2/lamda)+(Eg-Vds-fcd(j)-
        fbi)*exp((x2-L)/lamda);
    x3=L:0.1:(L+5);
    f3=Eg-Vds-fbi-(-Eg+Vds+fcd(j)+fbi)*exp((L-x3)/lamda);
    Vcs(1)=f2(L/0.1/2)-Eg;
%Self-Consistent Calculations
n=1;
while n<=200

%%%%%%Charge%%%%%%

Id2=0;Is2=0;

for E=(Vcs(n)-Eg/100):(-Eg/100):(Vcs(n)-Eg+Eg/100)

Ex=-E+Vcs(n)+Eg/2;
Id2=Id2+Eg/100*Ex/sqrt(Ex^2-(Eg/2)^2)*(1-1/(1+exp((E+Vds)/Vt)));
Ecy=E+2*Eg;
Ey=-Vcs(n)+Eg*3/2+E;
Is2=Is2+Eg/100*Ey/sqrt(Ey^2-(Eg/2)^2)*(1/(1+exp(Ecy/Vt)));
end

if Vcs(n)<fbi+0.00001 Is1=0;Id1=0;
else
    Is1=0;Id1=0;
    for i=1:1:99
        E(i)=fbi+(Vcs(n)-fbi)*i/100;
        if E(i)<(fbi+fsc(j))
            xi(i)=lamda*log((E(i)-fbi)/fsc(j));
        else
            C(i)=E(i)+Vgs(j)-fbi+Qc(n)*lamda^2/eGNR/e0;
            D=(Eg-Vds-fcd(j)-fbi)*exp(-L/lamda);
            y=[fsc(j),C(i),-D];
            r=roots(y);
            if r(1)<0
                xi(i)=log(r(2))*(-lamda);
            else
                xi(i)=log(r(1))*(-lamda);
            end
        end
    end
    if E(i)>(fcd(j)-Eg) && E(i)>Vcs(n)
        xf(i)=L-log((E(i)+Vds+fbi)/(-Eg+Vds+fcd(j)+fbi))*lamda;
    else
        A(i)=E(i)+Vgs(j)-fbi+Qc(n)*lamda^2/eGNR/e0;
        B=(Eg-Vds-fcd(j)-fbi)*exp(-L/lamda);
        y=[fsc(j),A(i),-B];
        r=roots(y);
        if r(1)<0
            xf(i)=log(r(2))*(-lamda);
        else
            xf(i)=log(r(1))*(-lamda);
        end
    end
end

S(i)=0;
for x=(xi(i)+(xf(i)-xi(i))/100):(xf(i)-xi(i))/100:(xf(i)-(xf(i)-xi(i))/100)
    if x<0
        Ec=fbi+fsc(j)*exp(x/lamda);
    else if x<L
        Ec=fbi+Eg*Vgs(j)-Qc(n)*lamda^2/eGNR/e0*exp(-x/lamda)+(Eg-Vds-fcd(j)-fbi)*exp((x-L)/lamda);
    else
        Ec=Eg-Vds-fbi+(-Eg+Vds+fcd(j)+fbi)*exp((-L-x)/lamda);
    end
end
S(i) = S(i) + 2/hbar/vF*q/1e9*sqrt((Eg/2)^2-(Ec-E(i)-Eg/2)^2)*(xf(i)-xi(i))/100;
end
T(i) = exp(-S(i));
Ex(i) = Vcs(n) - E(i) + Eg/2;
Is1 = Is1 + T(i)*(Vcs(n) - fbi)/100*Ex(i)/sqrt(Ex(i)^2-(Eg/2)^2)*(1-1/(1+exp(E(i)/Vt)));
Id1 = Id1 + T(i)*(Vcs(n) - fbi)/100*Ex(i)/sqrt(Ex(i)^2-(Eg/2)^2)*(1/(1+exp((E(i) + Vds)/Vt))-1);
end
end
Qcs(n+1) = q^2/hbar/vF/pi*(Is1-Is2)/w/tGNR/1e9;
Qcd(n+1) = q^2/hbar/vF/pi*(Id1+Id2*2)/w/tGNR/1e9;
Qc(n+1) = Qcs(n+1) + Qcd(n+1);

%%%%%%BandProf%%%%%%%%%
fsc(j) = (+Eg-Vgs(j)-Qc(n+1)*lamda^2/eGNR/e0)/2;
fcd(j) = (2*Eg-Vgs(j)-Qc(n+1)*lamda^2/eGNR/e0-Vds)/2;
x1 = 5:0.1:0;
f1 = fbi+fsc(j)*exp(x1/lamda);
x2 = 0:0.1:L;
f2 = fbi+Eg-Vgs(j)-Qc(n+1)*lamda^2/eGNR/e0-fsc(j)*exp(-x2/lamda)+(Eg-Vds-fcd(j)-fbi)*exp((x2-L)/lamda);
x3 = L:0.1:(L+5);
f3 = Eg-Vds-fbi+(-Eg+Vds+fcd(j)+fbi)*exp((L-x3)/lamda);
Vcs(n+1) = f2(L/0.1/2)-Eg;
if abs(Vcs(n+1)-Vcs(n)) <= 0.0001
  Q(j) = Qc(n+1); Qs(j) = Qcs(n+1); Qd(j) = Qcd(n+1);
n = 999;
end
n = n+1;
end

%Results: Band Diagram and Tunneling Current Density
if n == 1000
  v1 = f1-Eg;
  v2 = f2-Eg;
  v3 = f3-Eg;
  plot(x1,f1);
  hold on;
  plot(x2,f2);
  plot(x3,f3);
  plot(x1,v1);
  plot(x2,v2);
  plot(x3,v3);
  disp('finished');
  fsc(j) = -fbi+f2(1);
fcd(j)=f3(1);
It(1)=0;
for i=1:1:99
 E(i)=fbi+(-Vds-fbi-0.001)*i/100;
 if E(i)<(fbi+fsc(j))
 xi(i)=lamda*log((E(i)-fbi)/fsc(j));
 else C(i)=E(i)+Vgs(j)-Eg+Q(j)*lamda^2/eGNR/e0;
 D=(Eg-Vds-fcd(j)-fbi)*exp(-L/lamda);
 y=[fsc(j),C(i),-D];
 r=roots(y);
 if r(1)<0
 xi(i)=log(r(2))*(-lamda);
 else xi(i)=log(r(1))*(-lamda);
 end
end
if E(i)<(fcd(j)-Eg)
 A(i)=-fbi+E(i)+Vgs(j)+Q(j)*lamda^2/eGNR/e0;
 B=(Eg-Vds-fcd(j)-fbi)*exp(-L/lamda);
 y=[fsc(j),A(i),-B];
 r=roots(y);
 if r(1)<0
 xf(i)=log(r(2))*(-lamda);
 else xf(i)=log(r(1))*(-lamda);
 end
else xf(i)=L-lamda*log((E(i)+Vds+fbi)/(-Eg+Vds+fcd(j)+fbi));
end
S(i)=0;
for x=(xi(i)+(xf(i)-xi(i))/100):(xf(i)-xi(i))/100:(xf(i)-(xf(i)-xi(i))/100)
 if x<0
 Ec=fbi+fsc(j)*exp(x/lamda);
 else if x<L
 Ec=fbi+Eg-Vgs(j)-Q(j)*lamda^2/eGNR/e0-fsc(j)*exp(-x/lamda)+(Eg-Vds-fcd(j)-
 fbi)*exp((x-L)/lamda);
 else Ec=Eg-Vds-fbi+(-Eg+Vds+fcd(j)+fbi)*exp((-L-x)/lamda);
 end
 end
S(i)=S(i)+2/hbar/vF*q/1e9*sqrt((Eg/2)^2-(Ec-E(i)-Eg/2)^2)*(xf(i)-xi(i))/100;
 end
T(i)=exp(-S(i));
 It(i+1)=It(i)+q^2/pi/hbar*T(i)*(1/(1+exp(E(i)/Vt))-1/(1+exp((E(i)+Vds)/Vt)))*(-Vds-
 fbi-fbi)/100/w*10^9;
end
Current(j)=It(100);
else disp('not converged');
 end
end
APPENDIX D

Low-Subthreshold-Swing Tunnel Transistors
Qin Zhang, Wei Zhao, Student Member, IEEE, and Alan Seabaugh, Fellow, IEEE

Abstract—A formula is derived, which shows that the subthreshold swing of field-effect interband tunnel transistors is not limited to 60 mV/dec as in the MOSFET. This formula is consistent with two recent reports of interband tunnel transistors, which show lower than 60 mV/dec subthreshold swings and provide two simple design principles for configuring these transistors. One of these principles suggests placing the gate adjacent to the tunnel junction. Modelling of this configuration verifies that sub-60 mV/dec swing is possible.

Index Terms—Silicon-on-insulator (SOI), subthreshold swing, tunnel transistor.

I. INTRODUCTION

In a field-effect transistor (FET), the minimum voltage swing needed to turn a transistor "on" is an important figure of merit that ultimately sets the minimum power supply voltages and the minimum power dissipation of a technology. The subthreshold swing is defined as the gate voltage required to change the drain current by one order of magnitude, one decade. In the MOSFET, the subthreshold swing is limited to \( (kT/q) \ln 10 \) or 60 mV/dec at room temperature, and with scaling, the subthreshold swing increases \([1]\) \((k\) is Boltzmann’s constant, \(q\) is the electron charge, and \(T\) is the temperature). This increase in subthreshold swing as gate length is decreased marks a significant fundamental limitation of conventional FETs. In this letter, we show analytically that 1) certain interband tunnel transistors can have subthreshold swings below 60 mV/dec and 2) that the subthreshold swing is minimized by engineering the transistor geometry to enhance the gate control of the tunnel-junction bias-voltage and internal electric field.

In recent years, a growing number of transistors have been reported, which explore the use of the field effect to gate an interband tunneling current \([2]–[7]\). Several of these authors have characterized the subthreshold swing. In the carbon nanotube field-effect transistor of Appenzeller et al. \([5]\), a subthreshold swing of 40 mV/dec was measured in a double-gated transistor in which interband tunneling current flows through an n-p-n channel modulated by a top gate. In simulations, Bisrowka et al. \([4]\) have shown in a vertical gate p-i-n Si/SiGe interband tunneling transistor that subthreshold swings of 44 mV/dec can be achieved. In a lateral embodiment of the p-i-n Si interband tunnel transistor, Wang et al. \([3]\) have shown

Fig. 1. Silicon tunneling current density versus electric field for a reverse-biased tunnel junction bias of 1 V showing the excellent experimental agreement between (1), solid line, and theory predictions (after Tan et al. \([10]\) using the data of Stork and Isaac \([11]\) and Fair and Wistow \([12]\). The inset shows the p+-n tunnel-junction energy band diagram under bias \(V_{th}\).

by simulation that subthreshold swings of 15 mV/dec and even smaller are achievable and conclude that the subthreshold swing is not limited by \(kT/q\). In this letter, we outline the reasons for the low subthreshold swing in tunnel transistors and analyze a new transistor configuration for low subthreshold swing transistor action.

II. SUBTHRESHOLD SWING OF TUNNEL TRANSISTORS

The subthreshold swing \( S \) of a FET is defined by \( S = \ln 10 \, \frac{dV_{GS}}{d(\ln I_D)} \) \([8]\), where \( V_{GS} \) is the transistor gate-source voltage and \( I_D \) is the drain current. In an interband tunneling transistor, the drain current flows across a degenerately doped p+-n tunnel junction, a transport mechanism that is particularly well described, in the reverse, Zener tunneling direction, by Sze \([9]\)

\[
I = \alpha V_{th} \xi \exp \left( -\frac{b}{\xi} \right)
\]

(1)

where \( V_{th} \) is the tunnel-junction bias (see the inset of Fig. 1), \( \xi \) is the electric field, and \( \alpha \) and \( b \) are coefficients determined by the materials' properties of the junction and the cross-sectional area of the device \( A \). Specifically, \( \alpha = A\mu^2\sqrt{2m^*/E_g}/4\pi^2R^2 \) and \( b = 4\sqrt{m^*/E_g}/3\phi \), where \( m^* \) is the carrier effective mass, \( E_g \) is the energy band gap, and \( \phi \) is Planck’s constant divided by \( 2\pi \). This simple analytic expression for interband tunneling \((1)\) is in phenomenal agreement with the measurements of reverse biased p+-n tunnel junctions \([10]–[12]\)

Manuscript received January 4, 2006. This work was supported by the Office of Naval Research under Contract N00014-02-1-0924. The review of this letter was arranged by Editor E. Sangaengi.

The authors are with the Department of Electrical Engineering, University of Notre Dame, Notre Dame, IN 46556 USA.

Digital Object Identifier 10.1109/LED.2006.871855

0741-3106/06 $20.00 © 2006 IEEE
The derivative of the tunneling current expression of (1) with respect to the gate–source voltage can be used to determine an expression for the subthreshold swing of a field-effect tunneling transistor.

\[ S = \ln \left( \frac{1}{V_{GS}} \frac{\partial V_{t}}{\partial V_{GS}} \right) + \frac{\xi + b}{\xi^2} \frac{\partial V_{t}}{\partial V_{GS}} \]  

There are two terms in the denominator of (2), which can be maximized in order to achieve a low subthreshold swing, and these terms are not limited by \( kT/q \). According to the first term, the transistor should be engineered so that the gate–source voltage directly controls the tunnel-junction bias \( V_{t} \); this suggests a transistor geometry with a thin and/or high-\( \lambda \) gate dielectric and an ultrathin body to assure that the gate field directly modulates the channel. For an equivalent oxide thickness approaching 1 nm, \( \partial V_{t}/\partial V_{GS} \approx 1 \) and the first term in the denominator of (2) is approximately inversely related to \( V_{GS} \). Accordingly, the subthreshold swing in a tunnel transistor increases with gate–source voltage, a characteristic that has been observed in both measurements [5] and simulations [3], [4].

A second way to lower the subthreshold voltage swing is described by the second term in the denominator of (2), which says that the derivative of the junction electric field on the gate–source voltage should be maximized. This is a primary basis for gate control in the interband tunnel transistor of Bhuiyan et al. [4]. In such a device, the gate–source voltage changes the tunneling width and increases the junction electric field. In practice, the tunnel-junction bias and the junction electric field are coupled and cannot be engineered independently.

III. LOW SUBTHRESHOLD-SWING DEVICE DESIGN

Using the design guidance of (2), we explore a new transistor geometry toward achieving low subthreshold-swing operation. A schematic cross section of the transistor is shown in Fig. 2 and is comprised of a silicon-on-insulator (SOI) structure in which a lateral \( p^+/n^- \) tunnel junction is formed in an ultrathin semiconductor body; a gate is placed on the p-side. The gate is aligned to the junction on the p-side to facilitate control of the p-side electrostatic potential; the p region is fully depleted for zero gate bias.

This device geometry is amenable to an analytic solution for the potential profile using a two-dimensional (2-D) Poisson equation, a solution method described by Young [13]. This analysis ignores the effects of statistical dopant fluctuations, which will become increasingly important as the transistor dimensions are minimized. On the p-side, the potential \( \phi(x, y) \) is given by

\[ \frac{\partial^2 \phi(x, y)}{\partial x^2} + \frac{\partial^2 \phi(x, y)}{\partial y^2} = \frac{qN_A}{\varepsilon_{Si}} \]  

for \( 0 \leq x \leq L \) and \( 0 \leq y \leq L \) (3)

where \( N_A \) is the acceptor doping concentration on the p-side, \( \varepsilon_{Si} \) is the Si dielectric constant, \( L \) is the thickness of the Si layer, and \( L \) is the gate length. Using a parabolic approximation at low drain-to-source voltage \( V_{DS} \) gives \( \phi(x, y) = \phi_b(y) + c_1(y)x + c_2(y)x^2 \), where the coefficients \( c_1 \) and \( c_2 \) are functions of \( y \) only. The n-side is unmodulated by the gate and, therefore, can be treated as a one-dimensional (1-D) problem with the potential given by

\[ \frac{\partial^2 \phi(y)}{\partial y^2} = -\frac{qN_D}{\varepsilon_{Si}} \]  

for \( 0 \leq x \leq L \) and \( L \leq y \leq L + W \) (4)

where \( N_D \) is the donor doping concentration on the n-side, and \( W \) is the depletion length on the n-side. Four boundary conditions are applied: 1) constant potential on the surface between the semiconductor and the gate oxide; 2) zero electric field in the x-direction on the surface between the semiconductor and the buried oxide; 3) continuous potential and electric field at the p-n junction; and 4) 0.3-eV Schottky barrier at the drain contact. Equations (3) and (4) can then be solved to obtain the surface potential and energy band diagrams versus drain–source and gate–source bias.

The transistor operation can be understood from the computed energy band diagrams in Fig. 3 for a drain–source bias of
TABLE I  
CALCULATED SUBTHRESHOLD SWINGS FOR THE Si INTERBAND TUNNEL TRANSISTOR OF FIG. 2 VERSUS GATE VOLTAGE

<table>
<thead>
<tr>
<th>$V_{GS}$ (mV)</th>
<th>$\phi$ (mV/cm)</th>
<th>$V_{ON}$ (mV)</th>
<th>$S$ (mV/dec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>3.38</td>
<td>6</td>
<td>13.3</td>
</tr>
<tr>
<td>0</td>
<td>3.40</td>
<td>16</td>
<td>33.5</td>
</tr>
<tr>
<td>-10</td>
<td>3.42</td>
<td>26</td>
<td>56.4</td>
</tr>
<tr>
<td>-20</td>
<td>3.44</td>
<td>35</td>
<td>60.9</td>
</tr>
<tr>
<td>-30</td>
<td>3.47</td>
<td>45</td>
<td>81.4</td>
</tr>
</tbody>
</table>

Fig. 4. Simulated transfer characteristic of the proposed Si interband tunneling transistor using the Synopsis 2-D simulator.

$-0.2$ V: these band diagrams correspond to the surface of the SOI at $x = 0$. The transistor off condition is shown in Fig. 3(a); for a gate-source voltage of zero, the p-side is depleted and the interband tunneling probability is low. When a negative gate voltage is applied, the valence band on the p-side is raised above the Fermi level on the n-side turning the interband tunneling and the transistor on [Fig. 3(b)]. In this transistor geometry, the gate screens the drain field enabling current saturation and isolation.

Using these energy band simulations and this transistor geometry to compute the derivatives in (2), the bias dependence of the subthreshold swing can be estimated, given that $b$ is 33 MW/cm for Si, and $V_{GS}$ and $\phi$ are extracted from the simulated band diagrams [Table I]. The subthreshold swing is dependent on the applied gate-source bias, and it is apparent that this device can achieve smaller than 60 mV/dec over a small bias range near zero gate voltage. For the tunnel transistor, it is clear that the minimum subthreshold swing will not be the most important parameter. Rather, it will be necessary to define the minimum voltage swing needed to achieve a given ON-OFF current ratio.

The Synopsis device simulator [17] [former Integrated Systems Engineering (ISE)] is used to compute the 2-D energy band and channel electric-field profiles to verify agreement with the analytic model. The computed electric-field distribution is then used with (1) to determine the current distribution in the channel, allowing a comparison with our analytic calculations. The $I_D-V_{GS}$ transfer characteristic, for the device geometry previously given, is shown in Fig. 4. The 2-D simulations are in good agreement with the analytic model; low subthreshold swing $\sim 7$ mV/dec occurs near $V_{GS} = 0.07$ V, and the low subthreshold swings can be kept within a small gate-source voltage, 40 mV, in reasonable agreement with the results shown in Table I. From Fig. 4, a voltage swing of 140 mV provides an ON-OFF current ratio of $10^6$ compared to a MOSFET swing of 180 mV. The voltage swing of 140 mV in the tunnel transistor should be reduced by changing the channel materials to, e.g., Ge or InAs while increasing the current drive.

IV. DISCUSSION

Band-to-band Si tunnel transistors have been demonstrated with ON-state current density as high as $100 \mu$A/µm [6], which is below the MOSFET's 1000 µA/µm (90-nm process node) [14]. Simulations indicate that higher interband current density (850 µA/µm) can be obtained using Ge channels [4]. Tunnel transistors have much lower off leakage, <0.1 nA/µm [6], than the MOSFET, $\sim 150$ nA/µm [15], for the same technology node, 90 nm, due to the high barrier imposed by the p-n junction and resulting in much lower static power dissipation. The low subthreshold swing of tunnel transistors is shown in Table I, while low power applications; subthreshold logic circuits [16] are of particular relevance to this type of transistor.

V. CONCLUSION

An analytic expression was derived, which showed that the subthreshold swing of interband tunnel transistors was a function of the gate-source voltage and can exceed the MOSFET limit of 60 mV/dec about certain gate biases. Based on this expression, a new SOI tunnel transistor was proposed and simulated, outlining some of the attributes of the low subthreshold swing in interband tunneling transistors.

REFERENCES

APPENDIX E

Fully-depleted Ge interband tunnel transistor: Modeling and junction formation

Qin Zhang, Surajit Sutar, Thomas Kosel, Alan Seabaugh *

Department of Electrical Engineering, University of Notre Dame, Z15 Fitzpatrick Hall, Notre Dame, IN 46556-5637, USA

ARTICLE INFO

Article history:
Received 30 April 2008
Received in revised form 28 August 2008
Accepted 29 September 2008
Available online 13 November 2008

The review of this paper was arranged by Prof. S. Creely/Moncure

Keywords:
Intera band tunnel transistor
Low subthreshold swing
Ge tunnel junction
Rapid melt growth

ABSTRACT

Complementary fully-depleted Ge interband tunneling field-effect transistors (TFETs) and static inverters are modeled to quantify TFET performance relative to Si MOSFETs. SYMPOSYS TCAD is used to compute the two-dimensional electrostatics and determine the tunnel junction electric field. This electric field is used in an analytic expression to compute the tunnel current. When the speed and power performance of TFETs are compared with the nMOSFET at the same supply voltage, 0.3 V. For a gate length of 20 nm, Ge tunnel transistors can provide similar speed in comparison to 45-nm-node nMOSFETs (18 nm gate length), but saves more than 2X in power and lowers energy over by 7x. Toward demonstrating these transistors, a process for forming submicron p+n Ge tunnel junctions has been utilized in which Al-doped p+ Ge is regrown on n+ Ge, following melt-back of a patterned Al deposition. Transmission electron microscopy (TEM) reveals the regrown film and contact microstructure consistent with the Al-Ge phase diagram. The low peak-to-valley current ratio (PVR) of devices produced by this growth method is likely a result of point defects or junction doping non-uniformity as TEM suggest no dislocations are seen at the regrown junction. The PVR of these junctions does not improve as the device area is reduced from 100 to 0.1 μm², a size smaller than the formation scale for grains in the Al-Ge system.

© 2008 Elsevier Ltd. All rights reserved.

1. Introduction

Interband tunnel transistors can achieve a room temperature subthreshold swing below the metal–oxide–semiconductor field-effect transistor (MOSFET) limit of 60 mV/decade [1]. Lowering subthreshold swing is the most effective way to reduce power dissipation in devices, because it enables use of a lower supply voltage. Sub-60-mV/decade subthreshold swing was first realized in a carbon nanotube tunnel FET (TFET) [2] and recently a subthreshold swing of 52.8 mV/decade has been demonstrated in a planar Si TFET [3]. In addition to lowering subthreshold swing, TFETs offer lower off-state current than MOSFETs, as a higher off-state leakage current is present in semiconductor structures. However, the Si TFET [3] achieves an on-state current density of 12.1 μA/μm at 1 V supply voltage, over two orders of magnitude lower than a high-performance MOSFET.

The method for deriving the dependence of tunnel current density on semiconductor properties is given by Mull [5]. Using the triangular barrier model approximation of Sze and Ng [6], the following relation is obtained:

\[ J = \frac{q^2 V_T}{4\pi \varepsilon_0} \sqrt{\frac{m^*}{2eE_{G}}}, \]  

where in good agreement with experimental data. \( V_T \) is the maximum electric field at the junction, \( V_E \) is the reverse bias, \( m^* \) is the tunneling reduced effective mass, \( \varepsilon_0 \) is the dielectric constant, and the constants \( q \) and \( h \) are electron charge and Planck's constant, respectively. Using Eq. (1), the dependence of tunnel current density is computed versus wavelength. The model used here for Si and Ge. The case of Si measurements (open and closed circles) from eight different p-n junctions [8,9] spanning over eight orders of magnitude in current density, are shown in Fig. 1. Eq. (1) (dashed line) is in excellent agreement with the calculations using only the tunneling reduced effective mass as a fitting parameter. The calculations in Fig. 1 are also in reasonable agreement with the recent SitFET demonstration [3] if a 10 nm channel thickness and a 4 μm/cm maximum electric field are assumed. For the Ge case, the current density is slightly higher than measurements [10–12], which might be caused by the junction non-uniformity. Fig. 1b shows that Ge tunnel junctions can have about two orders of magnitude higher current density than Si at the same electric field.

In this paper, Section 2. Ge-based fully-depleted TFETs are designed and modeled using Synopsys. The simulation results of TFETs are compared with Si TFETs and high-performance Si MOSFETs in Section 3. To achieve high tunneling current density designed for Ge TFETs, submicron abrupt heavily-doped Ge tunnel junctions are explored using rapid melt regrowth (Section 4).

2. Ge interband tunnel transistor modeling

The two-dimensional (2D) electrostatics of Ge interband tunnel transistors and inverters are simulated with the SYMPOSYS TCAD 2005 tool, used in conjunction with Eq. (1), to determine the tunneling current density.
Fig. 1. (a) Tunnel current density per reverse bias vs. electric field for Si and Ge tunnel diodes. Close agreement is obtained between Eq. (1) and measurements with a fitted effective mass of 0.02m₀, for Si while for Ge, a theoretical mass, 0.02m₀,[13] is used. (b) Tunnel current density vs. electric field for a reverse junction bias of 0.5 V, showing Ge can provide two orders of magnitude higher current than Si at the same electric field.

transistor current-voltage relations. Scaled cross-sections of the complementary Ge tunnel-transistor pair are shown in Fig. 2a, for an n-TFTET and Fig. 2b, for a p-TFTET. The transistors utilize a germanium-on-insulator (GOI) structure, in which a laterally abrupt p'n' tunnel junction is formed in a 2 nm Ge body. A 20 nm long gate is aligned adjacent to the junction to fully-deplete the underlying semiconductor. In the simulations, a 1 nm thick Al₂O₃ dielectric is used as the gate oxide (ε₀ = 9, εₛ = 8.7 eV). Gold with a work function of 4.8 eV is used for the n-TFTET gate metal, and Al with a work function of 4.26 eV is used for the p-TFTET gate.

The device operation can be understood from the simulated band diagrams shown in Fig. 2a and b. At zero gate bias, the n⁺ (n-TFTET) or p− (p-TFTET) body is fully depleted and the transistor is normally off (solid lines) with no direct interband tunneling path. With a 0.5 V gate bias, interband tunneling is turned on (dashed lines) in the Zener tunneling direction. In this geometry, the gate screens the drain field and the current injection is set by the gate-source bias.

SYNOPSYS is used to compute the 2D electrostatics for the transistors of Fig. 2a and b. The channel is divided into 16 layers (with unit thickness of 2 nm/16 = 0.125 Å). For each layer, the band diagram is calculated, from which the reverse voltage and the maximum electric field at the junctions are read and entered into the tunneling current relation of Eq. (1) to determine the current density in mA/cm² in each incremental layer; the current is then summed over all 16 layers. The tunneling path may not be restricted to each divided layer, but since the electric field does not change much across the junction (see Fig. 4), this integral method is a good approximation. Shown in Fig. 3 is the dependence of the channel current on gate-to-source bias for drain-to-source biases of 0.5 and 0.1 V, respectively. For both the n- and p-TFTET, an on-state current density of nearly 450 µA/µm is achieved at Vₛ = -1.5 V, and the off-state current density, determined from SYNOPSYS, is 3.6 nA/µm. The threshold voltage is 0.05 V using a constant current method with a current of 100 nA at a typical number for MOSFETs [14]. A swing of 50 mV changes the current by more than 3 orders of magnitude, giving an effective subthreshold swing less than 17 mV/decade.

An n-TFTET is used to increase the influence of the junction abruptness. At on-state, as the abruptness decreases from 0 to 4 nm/decade, the maximum electric field at the junction decreases from -64 MV/cm to 2.2 MV/cm and the on-state current density decreases almost one order of magnitude. Quantization on the ultra thin Ge body should also influence the current density: the tunneling probability will decrease due to the increase in band gap, while the tunneling density of states will increase from 3D to 2D. The quantization can
3. Interband tunnel transistor performance estimates

Table 1 compares n-TFET performance with scaled nMOSFET targets (highlighted in italic) from the International Technology Roadmap for Semiconductors (ITRS) [17]. The on-state MOSFET currents are computed from Panek's model [17] at a supply voltage of 0.5 V, and the off-state MOSFET currents are computed from Hansen's model [18] with \( V_{dd} = 0 \) and \( V_{gs} = 0.5 \). From Hansen's model, the off-state current at the 65 nm node (2007) is 0.29 \( \mu A/\mu m \), with in the ITRS requirement which is 0.34 \( \mu A/\mu m \); but the 45 nm (2010) and 32 nm (2011) nodes are \( \approx \times 3 \) of the ITRS requirement. As the development of high-k metal gate stack technology, the gate leakage could be much lower than the off-state leakage at this stage; Chui has claimed a gate leakage less than \( 10^{-7} A/cm^{2} \) for HfO\(_2\)/Ge structure at 1 V bias, where the equivalent thickness of the oxide is less than 1 nm [19]. This gate leakage level is equal to \( 2 \times 10^{-7} \mu A/\mu m \) at 20 nm gate length, so the gate leakage is neglected here for the total leakage calculation. For the TFET, the transistor properties are evaluated for Si and Ge with a maximum internal junction field of 4 MV/cm. The off-state current is simulated by SYNOPSYS' generation-recombination model which is in agreement with measurements for the case of Si [4]. Quantization effects are not included in these calculations; quantization would raise the effective band gap and lower the off-currents relative to these predictions. The speed is calculated by CVT, and in the last six rows, power and energy consumptions (dynamic and leakage) are calculated for an n-stage inverter chain (\( n = 50 \)) with an activity factor \( \alpha = 2 \), after Hansen et al. [18]. The activity factor of 2 accounts for an average switching of 1 in 50 MOSFETs per cycle.

Since the MOSFET and the TFET are compared at the same supply voltage and similar capacitance, the speeds are determined by the on-state current. Silicon is not attractive for the TFET channel because of its low on-state current, only 1.2 \( \mu A/\mu m \). With this low current the speed is more than \( 400 \times \) lower than the MOS FET. In contrast, the Ge tunnel transistor shows an on-state current density as high as 440 \( \mu A/\mu m \) at 0.5 V supply voltage, and comparable speed to the 2010 nMOSFET with 18 nm gate length. The off-state current density for the Ge tunnel transistor is 0.0036 \( \mu A/\mu m \), much lower than the MOSFET. The low off-state current of the Ge TFET dissipates \( 2 \times \) less power with energy dissipation more than \( 20 \times \) lower than the 2010 MOSFET for a 50-stage inverter chain with an activity factor of 2.

4. Ge tunnel junctions

4.1. Device fabrication

To realize the proposed Ge interband tunnel transistor require the development of an abrupt, heavily-doped lateral tunnel junction. Toward this end, a rapid melt growth process for forming Ge interband tunnel junctions has been developed in which evaporated Al contacts on n+ Ge are liquified in a rapid thermal process to dissolve back and regrow p' Ge and form the tunnel junction [20]. Prior work [20] has shown high current density exceeding 1 mA/\( \mu m^{2} \) for micron-scale diodes, but with low peak-to-valley ratios (PVR), under 1.5 at room temperature. In this work, Ge p'n junctions with sizes ranging from 30 \( \times \) 30 \( \mu m \) to 300 \( \times \) 300 \( \mu m \) were prepared to explore the tunnel junction dependence on junction size.

The process uses a phosphorus \( 1 \times 10^{19} \) cm\(^{-3} \) spin-on diffusion followed by rapid thermal annealing to form the n+ side of the junction [20]. Secondary ion mass spectroscopy (SIMS...
Fig. 5. (a) Simulated load curves for a complementary pair of Ge tunnel transistors in a static CMOS TET inverter using the geometry of Fig. 2a; n-TET, n' doping of 1.8 x 10^{18} cm^{-3}, and p' doping of 3.2 x 10^{20} cm^{-3}, while for the p-TET, the n' doping is 3.4 x 10^{18} cm^{-3} with p' doping of 1.1 x 10^{20} cm^{-3}; (a) common-source load characteristics and (b) voltage transfer characteristics.

Fig. 6. (a) Simulated load curves for a complementary pair of Ge tunnel transistors in a static CMOS TET inverter using the geometry of Fig. 2 with identical n- and p-TET junction dopings, n' = p' = 1 x 10^{18} cm^{-3}; (a) common-source load characteristics and (b) voltage transfer characteristics.

Table 1

<table>
<thead>
<tr>
<th>Parameter</th>
<th>MOSFET*</th>
<th>2007</th>
<th>2010</th>
<th>2013(^{a})</th>
<th>Tunnel transistor</th>
<th>Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gate length ( L_g )</td>
<td>25</td>
<td>18</td>
<td>13</td>
<td>20</td>
<td>20 mm</td>
<td></td>
</tr>
<tr>
<td>Gate width ( W ) = 10( L_g )</td>
<td>250</td>
<td>190</td>
<td>130</td>
<td>200</td>
<td>200 mm</td>
<td></td>
</tr>
<tr>
<td>Equivalent oxide thickness ( ED )</td>
<td>1.1</td>
<td>0.65</td>
<td>0.5</td>
<td>1</td>
<td>1 mm</td>
<td></td>
</tr>
<tr>
<td>Supply voltage ( V_{DS} )</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5</td>
<td>0.5 V</td>
<td></td>
</tr>
<tr>
<td>On current ( I_{ON} )</td>
<td>428</td>
<td>701</td>
<td>1053</td>
<td>1.2</td>
<td>440 (\mu A/\mu m^2)</td>
<td></td>
</tr>
<tr>
<td>Off current ( I_{OFF} )</td>
<td>2.09</td>
<td>2.62</td>
<td>1.88</td>
<td>2.7E-05</td>
<td>0.0036 (\mu A/\mu m^2)</td>
<td></td>
</tr>
<tr>
<td>Oxide capacitance density ( COX \approx \varepsilon / \varepsilon_0 )</td>
<td>31.4</td>
<td>531</td>
<td>65.1</td>
<td>34.5</td>
<td>34.5 (F/\mu m^2)</td>
<td></td>
</tr>
<tr>
<td>Gate capacitance ( C_{G} = C_{OX,GC} )</td>
<td>0.78</td>
<td>0.96</td>
<td>0.90</td>
<td>0.99</td>
<td>0.99 (F/\mu m)</td>
<td></td>
</tr>
<tr>
<td>Intrinsic speed ( S = C_{OX}/K_{ox} )</td>
<td>0.02</td>
<td>0.08</td>
<td>0.43</td>
<td>288</td>
<td>288 ps</td>
<td></td>
</tr>
<tr>
<td>Leakage ( P_{leak} \approx \mu I_{leak} )</td>
<td>3.25</td>
<td>50.50</td>
<td>47.00</td>
<td>6.8E-05</td>
<td>0.09 (\mu W/\mu m)</td>
<td></td>
</tr>
<tr>
<td>Dynamic ( P_{d} ) ( \approx L_{D}/2K_{ox}V_{DD}^2 )</td>
<td>107</td>
<td>173</td>
<td>263</td>
<td>0.300</td>
<td>110 (\mu W/\mu m)</td>
<td></td>
</tr>
<tr>
<td>Total ( P = P_{d} + P_{leak} )</td>
<td>114</td>
<td>226</td>
<td>310</td>
<td>0.300</td>
<td>110 (\mu W/\mu m)</td>
<td></td>
</tr>
<tr>
<td>Leakage ( E_{leak} \approx (I_{leak} / V_{DD}) \times (\text{It}) )</td>
<td>332</td>
<td>1722</td>
<td>1002</td>
<td>1</td>
<td>4 (\mu A/\mu m^2)</td>
<td></td>
</tr>
<tr>
<td>Dynamic ( E_{d} ) ( \approx L_{D} / 2(n_{G}C_{OX}) )</td>
<td>68</td>
<td>120</td>
<td>112</td>
<td>86</td>
<td>86 (\mu A/\mu m^2)</td>
<td></td>
</tr>
<tr>
<td>Total ( E = E_{d} + E_{leak} )</td>
<td>430</td>
<td>1642</td>
<td>1114</td>
<td>87</td>
<td>90 (\mu A/\mu m^2)</td>
<td></td>
</tr>
</tbody>
</table>

Logic depth \( n = 50 \), activity factor \( a = 2\%.\)

\(^{a}\) ITRS 2007 Edition.

\(^{b}\) UTB FD ultra thin body fully depleted.
measurements show that a density of $6 \times 10^{20}$ cm$^{-3}$ is achieved for the n$^+$ doping. A 50 nm thick Al film is then patterned by electron beam lithography and liftoff, serving as the acceptor dopant source. A 50 nm plasma Si$_3$N$_4$ cap is applied to act as a micromembrane before Liu et al. [21] and hold the Al–Ge melt during the annealing. The rapid thermal annealing of Al in contact with Ge above the eutectic temperature [22] causes Ge to be dissolved into the liquefied Al. On cooling the Ge regrows epitaxially until the eutectic temperature is reached, forming an abrupt, heavily-doped p$^+$ layer. As the temperature further decreases, Al-rich and Ge-rich phases are nucleated, leaving a eutectic mixture above the Ge p$^+$ junction. To measure the submicron devices, 100 x 100 nm vias were written by electron beam lithography and etch, and Ti/Au bondpads were patterned by liftoff.

Fig. 7 shows a transmission electron microscopy (TEM) image of an AlxGe$_{1-x}$ tunnel junction after 600°C, 1 s rapid thermal annealing with a cooling rate of 30°C/s. Energy-dispersive X-ray spectroscopy (EDXS) is used to confirm the presence of Al and locate the junction. Close examination of the p$^+$nn$^+$ junction in this location and elsewhere shows no evidence of dislocations. Fig. 8 shows the overall non-uniformity of the junction, a cross section through a 10 x 10 μm$^2$ device. EDXS analysis of the eutectic mixture above the regrown Ge layer shows that the darker regions are the Al-doped Ge-rich phase, and the lighter regions are the Al-rich phases as expected from the phase diagram [22]. Electron diffraction patterns show that the Ge-rich regions of the eutectic mixture are within a few degrees of the same crystallographic orientation as the Ge substrate, indicating that these regions are nucleated epitaxially off the regrown layer.

4.2. Current–voltage characteristics

Fig. 9 shows the current–voltage (I–V) characteristics of Ge tunnel junctions with sizes ranging from 10 x 10 μm$^2$ to 300 x 300 μm$^2$. Negative differential resistances are observed, a signature of abrupt heavily-doped junctions. Series resistance shifts are responsible for the voltages differences between small and large area devices. For the 300 x 300 μm$^2$ diode, a peak current density of 0.15 mA/μm$^2$ is achieved, corresponding to an effective doping of $\sim 2.5 \times 10^{19}$ cm$^{-3}$ [23]; and 1 mA/μm$^2$ Zener tunneling current is achieved at the reverse bias of 0.13 V, where the maximum electric field of 2.87 MV/cm is calculated assuming an ideal abrupt junction. Compared with the calculated current density of Ge tunnel diodes shown in Fig. 1, the measured current density is lower, which might be caused by the non-ideal junction abruptness. If a 4 nm/decade abruptness is assumed, the electric field will decrease to 1.6 MV/cm, in agreement with the current density calculation. The PVR is low and does not increase significantly as junction area is reduced. The low PVR appears to be a result of point defects in the junction doping non-uniformity. Doping non-uniformity results in a peak current and peak voltage which depends on position. Measurements on a non-uniform junction is the superposition of I–V behavior over each incremental area; these sum to lower the overall PVR.

5. Conclusions

Complementary fully-depleted Ge interband tunnel transistors are designed and simulated, showing low subthreshold swing, low off-state current and on-state current density as high as 440 μA/μm. Compared with a 2010 nMOSFET at 0.5 V supply voltage, Ge interband tunnel transistors can save 2× in power and 20× in energy for a 50-stage inverter chain with an activity factor of 2% while not sacrificing speed. Submicron Ge tunnel junctions were
fabricated by a rapid melt growth technique, and show a clean doping interface without observable dislocations. The low PVR is explained by point defects or junction non-uniformity.

Acknowledgements

The authors would like to thank A. Lochtefeld for useful discussions and the support of AmberWave Systems. The authors would also like to thank D.J. Frank, IBM, for discussions on MOSFET performance measures and Brian Doyle, Intel, for providing the SIMS measurements.

References

APPENDIX F

Graphene Nanoribbon Tunnel Transistors
Qin Zhang, Tian Fang, Huili Xing, Alan Seabaugh, and Debdeep Jena

Abstract—A graphene nanoribbon (GNR) tunnel field-effect transistor (TFET) is proposed and modeled analytically. Ribbon widths between 3 and 10 nm are considered, corresponding to energy bandgaps in the range of 0.46 to 0.14 eV. It is shown that a 5-nm ribbon width TFET can switch from on to off with only 0.1 V gate swing. The transistor achieves 800 pA/μm ON-state current and 26 pA/μm OFF-state current, with an effective subthreshold swing of 0.19 mV/dec. Compared to a projected 2009 nMOSFET, the GNR TFET can provide 5 times higher speed, 26 times lower dynamic power, and 2000 times lower OFF-state power dissipation. The high performance of GNR TFETs results from their narrow bandgaps and their 1-D nature.

Index Terms—Graphene, subthreshold swing, tunnel transistor, 1-D.

I. INTRODUCTION

The tunneling field-effect transistor (TFET) is attracting attention because of its low subthreshold swing and low OFF-state leakage [1]–[7]. Choi et al. [4] have experimentally demonstrated a Si n-channel TFET with a 32.8 mV/dec subthreshold swing. However, the ON-state current density in this device is two orders of magnitude lower than a high-performance MOSFET. To enhance the ON-state tunneling current, narrower bandgap materials with smaller effective masses are being considered [6], [7]. Knoch et al. [1], [2] show by simulation that 1-D TFETs are preferred to bulk TFETs because of the superior gate control and a reduction of transverse energy component in the 1-D tunneling transport. Graphene nanoribbons (GNRs) have a width-tunable narrow bandgap [8]–[13], which is particularly favorable for TFET applications, and more amenable to planar processing and large-scale integration than carbon nanotubes (CNTs). In this letter, the GNR TFET is analyzed. It is shown that TFETs based on GNRs are capable of higher drive currents than pFETs or group IV channel TFETs with lower subthreshold swing, higher speed, and orders of magnitude lower power dissipation.

II. DEVICE STRUCTURE AND OPERATION

Similar to a CNT TFET [1], [2], the schematic of a p-channel GNR TFET is shown in Fig. 1. The source is heavily doped to be n+, and the drain is p+. The channel is also designed to be p+, but is fully depleted at zero gate bias.

Fig. 1. Schematic cross section of an interband p-channel GNR TFET, where the GNR is heavily doped or electrostatically doped to form a p+n+ tunnel junction, and the gate is placed over the fully depleted p-side.

Heavy doping concentrations are generated either by dopants or electrostatically engineered [1]. Due to the strong C-C bonds in GNRs (and CNTs), chemical doping is challenging, but rapid progress is being made in that direction [13]. On the other hand, electrostatic doping has successfully been demonstrated for achieving tunnel junctions in CNT TFETs [3], and the same can be achieved for GNRs. The electrostatics of this quasi-1-D geometry is solved by a surface potential method, where the GNR surface potential, ΦS(x), is given by the 1-D Poisson equation [14]

\[ \frac{d^2 \Phi_S(x)}{dx^2} + \frac{\Phi_S(x) - \Phi_G - \Phi_{BI}}{\varepsilon_{GNR}} = \frac{q(\pm N)}{\varepsilon_{GNR}} \]

Here, \( \Phi_G \) is the gate potential, \( \Phi_{BI} = E_G/2 - qV_T \ln(N/n_i) \) is the built-in potential, \( E_G = 2\pi \hbar v_F / \hbar \omega_{GNR} \) [9], [10] is the GNR’s bandgap, \( \hbar = 10^8 \text{ cm/s} \) is the Fermi velocity of carriers in graphene [15], \( \omega_{GNR} \) is the ribbon width, \( V_T \) is the thermal voltage, \( n_i \) is the intrinsic carrier concentration, \( N \) is the doping concentration; the "+" sign is for donors, and the "-" sign is for acceptors. The parameter \( \lambda = \sqrt{\varepsilon_{GNR} / \varepsilon_{OX}} \omega_{GNR} / \omega_{OX} \) is the relevant length scale for potential variations; \( \varepsilon_{GNR} \) and \( \varepsilon_{OX} \) are the GNR and oxide dielectric constants, respectively, and \( \omega_{GNR} = 0.25 \text{ nm} \) [16] is the thickness of the GNR. Lastly, \( \omega_{OX} \) is the gate oxide thickness. Equation (1) is applied in all three regions: n+ source, p+ channel, and p+ drain. The boundary conditions are the following: 1) zero electric field at \( x = \pm \infty \); 2) continuous electric field and potential at the source/channel and drain/channel junction; 3) \( E_F - E_C = qV_T \) at the n+ source, and \( E_V - E_F = qV_T \) at the p+ drain; and 4) the valence band along the channel is aligned with the Fermi level at the source, at zero gate bias. This guarantees low OFF-state current at zero gate bias and a small turn-on voltage. For this analysis, the supply voltage is limited to half the GNR’s band gap where ambipolar effects can be ignored. The computed band diagrams for a GNR pTFET with a supply voltage of 0.1 V are shown in Fig. 2(a) for the OFF-state at \( V_{GS} = 0 \) and in Fig. 2(b) for the ON-state at \( V_{GS} = -0.1 \) V, for a ribbon width \( w_{GNR} = 5 \text{ nm} \), gate length \( L_{CH} = 20 \text{ nm} \), and \( \varepsilon_{GNR} = 1 \text{ nm} \).

Manuscript received June 9, 2008, revised September 1, 2008. Current version published November 21, 2008. The review of this letter was arranged by Editor J. Cai.

The authors are with the Department of Electrical Engineering, University of Notre Dame, Notre Dame, IN 46556 USA (e-mail: qzh@nd.edu, tfang@nd.edu, seabaugh@nd.edu, djena@nd.edu).

Digital Object Identifier 10.1109/LED.2008.2006530
At the ON-state, the 1-D Zener tunneling current is obtained by integrating the product of charge flux and the tunneling probability within the energy window $\Delta \Phi$.

$$I_D = \int f_q \rho_{GNR}(k) \times \frac{\hbar}{\pi} \left[ f(E - E_F) - f(E - E_F^+) \right] T_{WBK} dk$$

$$= \frac{\hbar}{\pi} \rho_{GNR}(k) \int f(E - E_F) - f(E - E_F^+) \times T_{WBK} dk$$

$$= \frac{\hbar}{\pi} T_{WBK} \times \left[ \ln \left( \exp \left( \frac{\Delta \Phi - (E_F - E_F^+)}{qV_T} \right) + 1 \right) - \ln \left( \exp \left( \frac{E_F - E_F^+}{qV_T} \right) + 1 \right) + \ln \left( \exp \left( \frac{V_{DS}/V_T}{1} \right) + 1 \right) - \ln \left( \exp \left( \frac{\Delta \Phi - qV_{DS}/V_T}{1} \right) + 1 \right) \right]$$

(2)

where $\rho_q = 1/\hbar \frac{dE}{dk}$, $\rho_{GNR}(k) = 1/\pi$ is the 1-D density of states, and $T_{WBK}$ is the tunneling probability, calculated by applying the WKB approximation to a triangular potential with a barrier height of $E_G$ [17].

$$T_{WBK} = \exp \left( -\frac{\pi E_G^2}{4 \hbar \nu \nu_q} \right)$$

(3)

is dependent on the bandgap $E_G$, and the electric field $\xi$ across the tunnel junction. Using (2) and (3), taking $\Delta \Phi$ and $\xi$ from the band diagram calculations in Fig. 2(b), the drain-current density at room temperature is calculated as a function of the gate-to-source voltage and shown in Fig. 3, dashed line. Here, the effect of carrier injection into the channel on the channel electrostatics is neglected. The OFF-state leakage is assumed to arise from thermal emission over the barrier $\varphi_B$ between source and the drain (see Fig. 2). The resulting leakage current for the ON-D case is given by $I_{OFF} = q^2\tau \varphi_B / h r \exp(-\varphi_B / qV_T)$ [18], which is 0.13 pA, somewhat lower than 1 pA observed in the GNR FET OFF-state [12]; this might be anticipated from the higher barrier $\varphi_B$ of TFET versus the FET.

The complementary nTFET with symmetric geometry is also shown in Fig. 3, solid line. From this calculation, a 0.1-V swing can turn the transistor from off at 800 $\mu$A/µm to

off at 26 pA/µm, with an on-to-off ratio of more than seven orders of magnitude. The subthreshold swing is 0.19 mV/dec using a current per unit width of 100 nA/µm [19] to define the threshold voltage, indicated in Fig. 3. Comparison of the 20-nm gate length GNR TFET to the 200-nm nMOSFET target from the 2007 Edition of the International Technologic Roadmap for Semiconductors (ITRS) Roadmap [20] is given in Table I. It is shown that the intrinsic switching speed of the GNR TFET is more than five times that of the nMOSFET target. The OFF-state channel leakage of the GNR TFET is more than four orders of magnitude lower, and the supply voltage is one order of magnitude lower, which makes the off-leakage power more than five orders of magnitude lower and the dynamic power also more than one order of magnitude lower than the high-performance nMOSFET. Certainly, these predictions represent upper bounds on the expected performance, as thermal gate oxide leakage, interface charge, and parasitic resistance will all act to degrade this projection.

III. ON-STATE CURRENT DENSITY OF TUNNEL TRANSISTORS

Recently, GNR FETs of sub-10-nm width have shown ON-state current density as high as 2000 $\mu$A/µm [12]. From the Zener tunneling current density expression by Sze [18], it is not surprising that the GNR TFET can have an ON-state current density as high as 800 $\mu$A/µm (Fig. 3), while the Si TFET shows 12.1 $\mu$A/µm for $V_{DS} = 1$ V, and only 0.4 $\mu$A/µm for the comparable $V_{DS} = 0.1$ V and $V_{DS} - V_{TH} = 0.1$ V [4].

$$J_D = \frac{\sqrt{2 m^* q^2 V}}{4 \pi^2 \hbar^2 E_g^2 / 2} \exp \left( \frac{4 \sqrt{m^* q^2 V}}{3 \varphi_B} \right)$$

(4)

where $V$ is the reverse bias. The ON-state current density of a bulk tunnel transistor is the product of $J_D$ and channel thickness, where $V$ is the supply voltage $V_D$. Fig. 4(a) show the calculated ON-state tunneling current density versus electric field and channel material at a supply voltage of 0.1 V. The solid line shows the current per unit gate width for bulk TFETs assuming the channel thickness is 5 nm. The dashed line show the current per ribbon width of GNR TFETs for the width of 3 nm.
### TABLE 1

<table>
<thead>
<tr>
<th>Speed and Power Estimates Comparing 2009 nMOSFETs Targets From the 2007 Edition</th>
<th>GNR TFETs Using the Geometry of Fig. 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Equivalent oxide thickness EOT</td>
<td>0.75</td>
</tr>
<tr>
<td>Supply voltage $V_{DD}$</td>
<td>1</td>
</tr>
<tr>
<td>Drive current $I_D$</td>
<td>1639</td>
</tr>
<tr>
<td>Off-state leakage current $I_{OSS}$</td>
<td>7.0</td>
</tr>
<tr>
<td>Intrinsic speed $= I_D / V_{OSS}$</td>
<td>1561</td>
</tr>
<tr>
<td>Off-leakage power $= I_{OSS} V_{DD}$</td>
<td>0.70</td>
</tr>
<tr>
<td>Dynamic power $= 1/2 I_D V_{DD}$</td>
<td>820</td>
</tr>
</tbody>
</table>

**REFERENCES**


REFERENCES


