Band-to-Band Tunneling Transistors: Scalability and Circuit Performance

Zachery Jacobson

Electrical Engineering and Computer Sciences
University of California at Berkeley

Technical Report No. UCB/EECS-2013-34
http://www.eecs.berkeley.edu/Pubs/TechRpts/2013/EECS-2013-34.html

May 1, 2013

Band-to-Band Tunneling Transistors: Scalability and Circuit Performance

By

Zachery A Jacobson

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Engineering – Electrical Engineering and Computer Sciences

and the Designated Emphases

in

Nanoscale Science and Engineering

and

Energy Science and Technology

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Tsu-Jae King Liu, Chair

Professor Sayeef Salahuddin

Professor Paul Wright

Spring 2012

Band-to-Band Tunneling Transistors: Scalability and Circuit Performance

By

Zachery A. Jacobson

Abstract

Band-to-Band Tunneling Transistors: Scalability and Circuit Performance

by

Zachery A Jacobson

Doctor of Philosophy in Engineering – Electrical Engineering and Computer Sciences

University of California, Berkeley

Professor Tsu-Jae King Liu, Chair

Continuing scaling of transistors as density approaches the terascale regime (10^{12} devices/cm^2) requires evaluating new devices that can perform on several metrics beyond density scaling, such as cost savings, performance improvements, and energy efficiency. A comprehensive review and evaluation of potential new devices is performed. Metrics such as processing cost, plan-view area scaling, and stage delay are benchmarked. One of the most promising devices, tunneling field effect transistors, is also the most confounding, as simulation and experimental results are orders of magnitude apart.

To better understand and evaluate tunnel field effect transistors (TFETs), a new TCAD analysis tool with dynamic nonlocal tunneling path determination is calibrated to experimental data. From this calibrated model, an optimal source design for TFETs is found where a moderate doping concentration (~10^{19} cm^{-3}) is found to be preferable to the higher doping concentrations more commonly used. Following this optimization, a study is performed to find the minimum device size, or the ultimate scalability, of TFETs. Using a raised source design allows TFETs to have a minimum device pitch (including contacts) of 29 nm.

A higher level of analysis is performed at the circuit level, where a Verilog-A based lookup table approach is used to evaluate the circuit performance of TFETs. Inverters, ring oscillators, SRAM, and Register Files are benchmarked and compared to UTB and FinFET technologies. TFETs are found to have advantages over standard CMOS for stage delays slower than 100ps in logic and for the 0.25V – 0.4V range in memory cells.

Table of Contents

Acknowledgements

The end of my graduate studies is a good time to stop and give thanks to all of those who helped me get to this milestone. As I finish my terminal degree, the end of over twenty years of schooling leaves me with some trepidation, but also excitement at the possibilities that await me. Without those who have guided me along the way, the arc of my life, and correspondingly this PhD dissertation, would have been far different.

First and foremost, I would like to thank my parents for all they have done for me. Their unending expression of love and pride in my choices allowed me to take risks in my educational path while knowing that whatever happened, I had their support.

Next, I can not adequately express in words my thanks to Prof. Tsu-Jae King Liu, my advisor. Prof. King is the ideal advisor. To me, a professor is a mentor, a teacher, and a scholar all in one. Prof. King exemplifies the best in each of these roles. As a mentor, she is patient and empathetic, while also pushing when she knows you can achieve more. As a teacher, Prof. King is professional and well versed in both industry trends and scholarly developments. As a scholar, she is knowledgeable and intellectually curious about research, keeping students on their toes until all of the data is fully understood. Her professionalism and dedication to her students is beyond reproach.

My qualification exam committee and dissertation committee both deserve great thanks. My qualification exam committee, Prof. Elad Alon, Prof. Paul Wright, Prof. Sayeef Salahuddin, and Prof. Tsu-Jae King Liu, were incredibly flexible in scheduling my exam and gave me valuable feedback to help me complete my research with the best of practices. As a perennial procrastinator, my dissertation committee, Prof. Paul Wright, Prof. Sayeef Salahuddin, and Prof. Tsu-Jae King Liu, are owed a special thanks for their patience and continued guidance throughout the process.

During my PhD, I worked with several industrial colleagues. At Intel, Dr. Kelin Kuhn, Dr. Rafael Rios, and Dr. Uygar Avci were most helpful in our weekly teleconferences. Over the course of almost two years, our interactions were always pleasant and motivating to complete our work. Dr. Seonghoon Jin of Synopsys is owed a special piece of gratitude for working with me at the start of my work with Prof. King. The access that he and his colleagues at Synopsys provided to the newest TCAD simulation tools started this project and led to all of the knowledge later gained.

In the Device Group at UC Berkeley, Prof. Jeff Bokor, Prof. Nathan Cheung, and Prof. Vivek Subramanian both taught courses and provided mentorship that I appreciated and am thankful for.

Many of those in the Electrical Engineering and Computer Sciences department deserve acknowledgement for the work they do. In particular, Ruth Gjerde, Pat Hernan, Dana Jantz, and Sam Rifkin were fantastic at supporting graduate students.

Although my dissertation research did not involve the UC Berkeley Microfabrication Lab, my previous interactions with lab staff were a strong influence and taught me much about experimental work. In particular, the equipment maintenance and process development staff, including Jay Morford, Joe Donnelly, Brian McNeil, Danny Pestal, Alan Briggs, and the late Jimmy Chang, made my research significantly more expedient. I give a special thanks and apology to Ben Lake of the machine shop, who was often in a tough position, tasked to get jobs completed ahead of schedule and under cost.

On a day-to-day basis, the graduate students I worked with provided some of the strongest influence on my work. In my early studies in the Javey group, Prof. Roie Yerushalmi, Dr. Johnny Ho, Dr. Zhiyong "Joseph" Fan, Dr. Lexi Ford, Dr. Kanghoon Jeon, and Prof. Paul Leu provided a consistent outlet for any and all frustrations related to research and adjusting to graduate student life. In the King Group, I made fantastic friends that will last a lifetime. Byron Ho, my desk neighbor, inspires me with his drive and dedication. Dr. Reinaldo Vega has an instinctual drive to push those around him to the boundaries of research questions. My TFET colleagues, Sung Hwan Kim and Peter Matheu, both were strong support for understanding the mechanisms behind these new devices. All of the King Group members influenced my research in one way or another, and I thank them. I especially thank Dr. Joanna Lai, my EEGSA "Big Sister", who was an empathetic mentor for issues that came up during graduate school.

Outside of King Group, I met many colleagues that made an impact. Dr. Anupama Bowonder and Dr. Pratik Patel both helped with my initial understanding of TFETs. Dr. Li-Wen Hung was extraordinarily knowledgeable about lab equipment and effective research methods. Gireeja Ranade, my EEGSA Co-President, demonstrated true leadership and capacity to give back to the graduate student community. My preliminary exam study group, including Dr. Reinaldo Vega, Dr. Tim Bakhishev, and Dr. David Carlton, helped get me through a very challenging study process. My interview study group of Dr. Volker Sorger and Amit Lakhani also eased the stress of a very challenging process with their tips and generous offerings of study tools.

Finally, I thank everyone who I interacted with at Berkeley. I learned so much from the people that made this work possible.

My work was supported for three years by the DoD Air Force Office of Scientific Research under the National Defense Science and Engineering Graduate Fellowship. I thank them for their support, as well as the support of Intel for my final year of studies.

Chapter 1

Introduction

1.1 The Never-Ending End of Transistor Scaling

Technology has transformed the world over the past century. Although the transistor was only invented in 1947 (65 years before the year of this dissertation's publication), transistors are now part of almost every person's daily life. As of November 2011, 5.9 billion people (87% of the Earth's population) now have a cell phone, a device that would never be possible without transistors [1]. The largest company in the world is now a technology company (Apple Inc.) whose products are all enabled by transistors [2]. Any changes, for better or worse, to the underlying pace of transistor technology improvement will have implications throughout many industries and the lives of people through the world.

The rapid growth of technology has been enabled by the continued miniaturization of transistors. The original transistor was invented at Bell Labs in the late 1940s (others had come close with similar devices in the prior two decades) [3-4]. The first transistor was a point-contact transistor that was centimeters in size, versus the nanometers of today's devices. Jack Kilby was one of the creators of the first Integrated Circuit in 1958 at Texas Instruments [5-6]. Integrated Circuits allow transistors to be built on a common substrate rather than as individual discrete components that needed to be wired together by hand, vastly reducing manufacturing complexity. Over time, engineers developed methods to fabricate smaller and smaller transistors. By doing so, not only did the devices perform better with high speeds, but the cost per device also decreased. This concept allows circuits to perform faster and with more functions in the same or smaller amount of area as the previous technology generation. This is known as scaling and is fundamentally responsible for the fast pace of technology improvement in the last half-century.

In 1965, Gordon Moore (later a founder of Intel Corporation) made a famous chart showing that the number of transistor components per integrated circuit increases exponentially over time [7]. Even in this early work, there were questions of how long this scaling trend could continue.

Fundamentally, Moore's "Law" is driven by cost and functionality. Of the years, many have questioned if scaling will be able to continue [8-11]. Scaling allows for the manufacturing of more devices for the same price. Increased costs for fabrication equipment, particularly lithography tools, at smaller lengths is one potential scaling roadblock [12]. Another is power density. If circuits need elaborate cooling systems to operate, these circuits will no longer have the functionality required for mobile systems. Power and power density are key impediments to scaling that will be explored further in this dissertation.

1.2 New Scaling Rules for CMOS Power

To understand how power and scaling are related, we need to understand how scaling affects circuits. Table 1.1 shows typical guidelines for MOSFET scaling [13].

Parameter	Constant Electric Field Scaling	General Scaling	Fixed Voltage Scaling
W, L, t_{OX}	1/S	1/S	1/S
V_{DD}, V_T	**1/S**	**1/U**	**1**
N_{SUB}	**S**	**S^2/U**	**S^2**
Area / Device	$1/S^2$	$1/S^2$	$1/S^2$
C_{OX}	S	S	S
C_{GATE}	1/S	1/S	1/S
k_n, k_p	S	S	S
I_{DSAT}	**1/S**	**1/U**	**1**
Current Density	**S**	**S^2/U**	**S^2**
R_{ON}	1	1	1
Intrinsic delay	1/S	1/S	1/S
Power	**$1/S^2$**	**$1/U^2$**	**1**
Power density	**1**	**S^2/U^2**	**S^2**

Table 1.1: MOSFET scaling guidelines (adapted from [13]). S is the scaling factor by which one dimension of a transistor is scaled (generally ~1.4 per process generation). U is the scaling factor by which voltages scale and has decreased over time. Bold lines show parameters that differ between the scaling scenarios.

Constant electric field scaling is an ideal model where S=U (*i.e.* voltage scale at the same rate as physical size scaling). Note that S and U are always greater than or equal to 1. Under this scenario, each technology generation has smaller area per device, higher performance (intrinsic delay), and reduced power consumption. However, S and U generally do not match exactly. This scenario is shown in the general scaling scenario. Assuming S>U>1, area and delay still scale as before, but power consumption now scales at a slower rate. If voltages can not scale at all, the fixed voltage scaling scenario occurs (U=1). In this case, power does not scale at all, but because each device continues to get smaller, power density rises with S^2. This was an issue that Moore presciently recognized in his 1965 paper [7]. Equation 1.1 and 1.2 explain the cause of this phenomenon by demonstrating the energy consumed in a circuit [14].

$$E_{TOTAL} = E_{SWITCHING} + E_{LEAKAGE} \quad (1.1)$$

$$E_{TOTAL} = \propto L_D f\ C\ V_{DD}^2 + L_D f\ I_{OFF} V_{DD} t_{DELAY} \quad (1.2)$$

In these equations, we separate the energy used in switching a transistor (*i.e.* from 0 to 1 or 1 to 0) and the energy used when a transistor is not switching, an energy we call leakage. $\propto$ refers to the probability of a switching event occuring. The more often a transistor switches, the more often switching energy is consumed. L_D refers to the number of logic stages a voltage signal must move through per complex logic function. f refers to fan-out, a measure of the average number of logic gates driven by each transistor. C refers to the capacitance per stage. V_{DD} is the supply voltage of the circuit. I_{OFF} is the current flowing through a transistor in the off-state. t_{DELAY} is the delay time for a transistor to switch.

Energy is equal to power multiplied by time. Under the constant electric field

scenario, voltage and gate capacitance both change by 1/S. This changes the switching energy by $1/S^3$. Since intrinsic delay (time) also changes by 1/S, this leads to a change in power consumption of $1/S^2$. However, under the fixed voltage scenario, only gate capacitance changes (by 1/S, as before). In this scenario, power is constant, and power density increases by S^2.

Clearly, the largest factor affecting power consumption is the voltage. For future scaling to continue without an increase in power density, voltages must reduce at least at the same rate as physical length scaling.

1.3 Dissertation Objectives

This dissertation seeks to solve the problem of continuing to scale physical transistor lengths while also scaling power density. Although traditional planar MOSFETs have been used for decades (since the migration from BJTs to MOSFETs for digital logic applications), new transistor structures will be necessary for scaling to continue. Of the many options available, Tunnel Field Effect Transistors (TFETs) are chosen as a potential candidate to replace or supplement traditional MOSFETs in integrated circuits. This hypothesis is tested by using simulation with TCAD models calibrated to fabricated devices. These models are then used to predict both the limits of physical device length of TFETs and the performance of these devices in circuits.

Chapter 2 is a comprehensive literature review of various charge-based devices that contend to replace or supplement traditional Complementary MOSFETs. The various histories are compiled and an understanding of the device principles of each class of structure is developed. Current work in the literature is compiled and used to benchmark 6 metrics. These metrics are on-state current, off-state current, energy, area, manufacturing complexity, and manufacturing cost. Several structures are shown to have some metrics that perform better than MOSFETs. TFETs are chosen for further study due to the large differences between simulated and fabricated devices.

In Chapter 3, simulation models are developed and calibrated to understand the physical processes behind TFETs. A collaboration with an industry vendor of transistor simulators (Synopsys Inc.) results in a calibrated model for band-to-band tunneling that includes dynamic tunneling path generation and support for heterostructures, a key feature in the fabricated device we chose to use for our calibration [15]. Simulations show that previous work on the optimal source doping concentration did not account for line tunneling [16] orthogonal to the gate, and a new lower optimal source doping concentration is found.

In Chapter 4, the scalability of Germanium Source TFETs is evaluated. A raised Germanium source allows for the tunneling area to be decoupled from the plan-view length of a device, allowing for scaling down to 29 nm of device length (isolation to isolation). Relevant parameters for a scaled device are optimized and presented.

In Chapter 5, circuit modeling is performed to understand how TFETs perform in

circuit applications. The previous simulation characteristics are input into a SPICE simulator, allowing the simulation of circuits containing hundreds of devices. The optimized TFET design is found to have performance better than that of a projected FinFET in 2018 at memory cell voltages below 0.45 V and logic speeds below ~1 GHz.

In Chapter 6, conclusions are made about the potential of TFETs in future technology. The contributions of this work are explained, and ideas for future work based on this dissertation are listed.

1.4 References

1. "Key Global Telecom Indicators for the World Telecommunication Service Sector," *International Telecommunication Union*, November 2011. [Online]. Available: http://www.itu.int/ITU-D/ict/statistics/at_glance/KeyTelecom.html .
2. J. B. Stewart, "Confronting a Law of Limits," *The New York Times*, 25 Feb. 2012: B1.
3. J. Vardalas, "Twists and Turns in the Development of the Transistor," *IEEE-USA's Today's Engineer*, May 2003. [Online] Available: http://www.todaysengineer.org/2003/May/history.asp .
4. "History of The Transistor (the 'Crystal Triode')," *The Porticus Center*. [Online]. Available: http://www.porticus.org/bell/belllabs_transistor.html .
5. "Integrated Circuit." [Online]. Available: http://en.wikipedia.org/wiki/Integrated_circuit .
6. "The Chip that Jack Built," *Texas Instruments*. [Online] Available: http://www.ti.com/corp/docs/kilbyctr/jackbuilt.shtml .
7. G. E. Moore, "Cramming more components onto integrated circuits," *Electronics*, vol. 38, pp. 114-117, 1965.
8. L. B. Kish, "End of Moore's law: thermal (noise) death of integration in micro and nano electronics," *Physical Letters A*, vol. 305, no. 3-4, pp. 144-149, 2002.
9. S. Tally, "One and done: Single-atom transistor is end of Moore's Law; may be beginning of quantum computing," *Purdue University News Service*, 19 Feb. 2012. [Online]. Available: http://www.purdue.edu/newsroom/research/2012/120219KlimeckAtom.html .
10. S. Hansell, "Counting Down to the End of Moore's Law," *The New York Times*, 22 May 2009. [Online]. Available: http://bits.blogs.nytimes.com/2009/05/22/counting-down-to-the-end-of-moores-law/ .
11. M. Kaku, *Physics of the Future: How Science Will Shape Human Destiny and Our Daily Lives by the Year 2100*, New York: Doubleday, 2011.
12. J. M. Rabaey, A. Chandrakasan, and B. Nikolić, *Digital Integrated Circuits: A Design Perpsective,* Upper Saddle River: Prentice Education, Inc., pp. 122-128, 2003.
13. B. J. Lin, "Lithography till the end of Moore's Law," in *Proceedings of the 2012 ACM International Symposium on Physical Design,* pp. 1-2, 2012.

14. H. Kam, T.-J. King Liu, E. Alon, and M. Horowitz, "Circuit-level requirements for MOSFET-replacement devices," in *Proceedings of the IEEE International Electron Devices Meeting*, pp. 427-428, 2008.
15. S. H. Kim, H. Kam, C. Hu, and T.-J. K. Liu, "Germanium-source tunnel field effect transistors with record high I_{ON}/I_{OFF}," in *Symposium on VLSI Technology Conference Digest*, pp. 178-179, 2009.
16. W. G. Vandenberghe, A. S. Verhulst, G. Groeseneken, B. Sorée, and W. Magnus, "Analytical model for point and line tunneling in a tunnel field-effect transistor," in *Proceedings of SISPAD*, pp. 137-140, 2008.

Chapter 2

Ultimate Device Scalability: Future Device Structures for the Terascale Regime

2.1 Introduction

Scaling transistor density into the terascale (10^{12} devices/cm^2) regime will require choosing device architectures and individual device structures that support increases in functionality while also continuing to reduce cost per function. In this chapter, mainstream and research structures are examined for feasibility for integration and implementation into the terascale regime.

2.2 Motivation

For more than forty years, logic device density has experienced exponential growth (a phenomenon known as Moore's Law [1]). This growth has enabled the information technology revolution of the last half-century. To continue this pace of innovation, transistor logic density must continue density scaling and include cost savings, performance improvements, and functionality additions.

Traditional Complementary Metal Oxide Semiconductor (CMOS) scaling is approaching fundamental limits. For the last several process nodes, V_T scaling and V_{SUPPLY} scaling have not kept pace with physical device pitch scaling due to the thermodynamic limit of 60 mV/dec of subthreshold slope in standard CMOS. As active power usage scales with $C \times V_{SUPPLY}^2$, the loss of voltage scaling prevents energy scaling from continuing. Planar CMOS gate scaling has also slowed, leaving much of the scaling to overall device pitch changes apart from the gate length [2]. Frequency scaling has also slowed, since power usage increases linearly with increased frequency.

As traditional CMOS scaling limits are reached, there are many technologies that are being considered to supplant or integrate with CMOS to continue functional scaling. This chapter's goal is to review future device technologies that could enable further scaling of devices beyond the limits of traditional bulk CMOS, while continuing gains in transistor performance. When examining these technologies, first the current research progress of each technology is assessed. Second, metrics to assess the potential of each technology and the technological challenges associated with each technology are developed. Third, these metrics are used to evaluate the technologies objectively. Finally, the technologies are benchmarked, both keeping in mind their existing progress and projecting their future long-range potential 5-10+ years out.

2.3 Scope

The literature review used to prepare this chapter confined the scope of replacement devices to those that could be either direct replacements or complements to existing CMOS logic. In some cases, circuit designs may need to change to

accommodate different operating modes (*e.g.* the unidirectional current of Tunnel FETs). However, this chapter does not look at designs that would require substantial architecture changes, such as reversible computing.

The use of other materials in existing structures was studied only for devices in which the basic operation of the device was vastly different than standard Silicon CMOS (*e.g.* HEMTs and GaN channel devices were included, but not III-V-channel MOS or Germanium-channel MOS devices).

Additionally, this chapter is restricted to devices relying on charge-based transport. Although spin-based transport devices are of increasing interest, they would require a radical architecture shift from the existing architecture used today for CMOS and are discussed elsewhere (for example, see ITRS Emerging research devices [3]).

Finally, some devices were not included in this review due to well-recognized scaling limitations. For example, JFETs were not included, since the primary motivation of this work is extreme scalability of devices. Similarly, although organic-based devices have excellent cost scaling per device, the potential for physical scaling and high performance operation is unlikely. Carbon-based nanoelectronic structures, such as nanotubes and graphene-based devices, were also not included due to current concerns about manufacturability at the terascale level of integration.

2.4 New Devices for Terascale Computing

2.4.1 Front-End

2.4.1.1 Non-Conventional Materials

HEMT

The High Electron Mobility Transistor, or HEMT, increases device mobility by separating charge carriers from the ionized dopant atoms, thus reducing ionized impurity scattering. This is accomplished by confining carriers in an undoped quantum well.

History

Early work on HEMT devices occurred at Fujitsu in 1979 under the direction of Dr. Takashi Mimura [4],[5]. While working on creating a GaAs n-MOSFET, Dr. Mimura realized that electron inversion or accumulation was difficult due to the presence of a high concentration of surface states at the gate dielectric interface. Simultaneously, Bell Labs had developed a modulation-doped heterojunction superlattice where potential wells of undoped GaAs captured electrons from donors in AlGaAs layers [6]. These electrons move with high mobility in the undoped GaAs wells due to a lack of ionized impurity scattering. By combining these two concepts, Dr. Mimura realized that with a stack comprised of a Schottky metal gate, doped n-AlGaAs region, undoped thin AlGaAs region, and GaAs, a structure similar to a MOS gate is formed and results in a device with reduced scattering and higher mobility. Additionally, by altering the

thickness of the doped AlGaAs layer, a depletion mode device is formed. (A thicker AlGaAs layer results in an electron accumulation layer at the dielectric interface.)

In roughly the same time frame as the Mimura work, Delagebeaudeuf and Linh at Thomson-CSF demonstrated a two-dimensional electron gas effect in a MESFET device, similar to a HEMT [7],[8]. This device was the first "inverted" HEMT, where the Schottky gate is deposited on an undoped GaAs channel layer grown over a doped AlGaAs layer.

Further work resulted in new designs, such as AlGaAs/InGaAs pseudomorphic HEMTs (pHEMTs, not to be confused with p-type HEMTs) [5]. Traditional HEMTs are constrained to using materials with matching lattice constants. pHEMTs use very thin layers of materials with mismatched lattice constants, improving performance.

Most of the initial uses of HEMTs were for military and aerospace applications, but demand for HEMTs increased in the 1990s when Direct Broadcast Satellite television receivers began using HEMT amplifiers. More recent uses of HEMTs include radar systems, radio astronomy, and cell phone communications.

Device Principles

HEMTs use the properties of a heterojunction to form a conductive channel with greater mobility than a traditional MOSFET. In a HEMT, a heterojunction with a wide bandgap semiconductor is fabricated on top of a narrow bandgap semiconductor, such as AlGaAs / GaAs [9]. The electrons from the n-doped wide bandgap region (AlGaAs in this example) diffuse into the GaAs, which has a lower conduction band than AlGaAs. The GaAs is undoped, so carriers experience reduced scattering, increasing mobility. This layer of carriers is called a two-dimensional electron gas, or 2DEG.

Due to difficulties in forming a gate dielectric on these materials, HEMTs use a Schottky gate contact over the wide bandgap semiconductor. This Schottky contact results in higher gate leakage for HEMTs than traditional MOSFETs.

Recent Work

Several challenges exist for HEMTs. First, although drive currents are high, operating voltages for most HEMTs are much higher than traditional CMOS, which poses an issue for low-power operation. Gate leakage is a key concern, as Schottky gates have very high gate leakage due to the lack of a dielectric barrier. Band-to-band tunneling due to the narrow bandgap is an issue, as is high source and drain resistance [10]. There is also the issue of integration of p-type devices. Finally, the use of III-V wafers also adds fabrication cost and manufacturing complexity.

Recent work has focused on the use of HEMTs at lower operating voltages. Dewey et al. shows drive currents that are able to match 40 nm MOSFETs at V_{DD}=1 V as well as 0.5 V [11].

Gate leakage can be improved with new gate dielectric materials. Work from Radosavljevic et al. has shown improvements in gate leakage by using TaSiOx rather than a Schottky gate [12]. Kim et al. showed that using a delta doping that is located further away from the gate and removing a portion of it during etch allows for a large reduction in gate leakage while reducing drive current by only a small amount [13].

To reduce band-to-band tunneling, the bandgap of the channel material can be modified. Kim and Del Alamo showed that using an InAs subchannel sandwiched between two InGaAs layers reduces band-to-band tunneling due to energy level quantization within the InAs forming a larger effective bandgap [14]. In addition, they also showed that the source resistance is improved with good Drain Induced Barrier Lowering (DIBL) and Subthreshold Slope (SS). High source and drain resistance can also be addressed through different annealing or diffusion techniques [10].

To form p-type devices on the same wafer, wafer bonding has been used by Chung et al. to attach GaN to Silicon wafers, where a p-type Si device can be used [15]. Since alignment issues are less of an issue with large power devices, this strategy may be less successful for logic applications where density is critical.

Scaling

Several groups have addressed scaling of HEMTs [16],[17],[18],[19]. Using electron beam lithography and multiple etch steps, Waldron et al. showed that it is possible to reduce a HEMT down to 30 nm gate-to-contact spacing, but it is difficult to make the gate length small without improvements to the etch processes [16].

Kharche et al. found that InAs is projected to scale well, as quantum well width scaling brings improvements in I_{ON}/I_{OFF} due to lower I_{OFF} [17]. The reduced well width brings the electron peak closer to the gate, allowing for better gate control.

Oh and Wong showed that if issues with gate leakage and process integration at small gate lengths can be solved (along with finding a symmetric p-type device), HEMT devices can have lower delay or lower Energy-Delay Product (EDP) [18]. However, others, including Skotnicki and Boeuf, have shown that when DIBL and SS are included into an effective current metric, strained Silicon performs better than III-V HEMTs [19].

Conclusions

HEMTs' main advantage when compared to CMOS is that the increased electron concentration allows for higher drive current in devices. However, to be competitive with CMOS, certain significant challenges need to be resolved, such as gate leakage, device pitch, and the lack of equivalent p-type devices. Progress has been made in scaling L_G down to 30 nm, but strained Silicon will continue to improve, possibly at a faster rate than HEMT technology can catch up. III-V MOSFETs that combine the high electron velocity of III-V materials with low gate leakage will also be a competing option [20]. Cost scaling is also expected to be a factor, since the substrates and/or specialized processing, such as MOCVD, would likely be a significant additional manufacturing cost. HEMTs are expected to dominate specialized applications where speed and frequency are more critical than power consumption and manufacturing costs, such as communications, military and aerospace.

Gallium Nitride

Gallium Nitride is a III-V material with many properties that make it appealing as a channel material. It has a high breakdown voltage, high electron mobility, and high saturation velocity. Perhaps more importantly, a two-dimensional electron gas is

induced by polarization at the AlGaN/GaN interface (creating a HEMT spontaneously), unlike an AlGaAs/GaAs HEMT that requires intentional doping to form charge.

History

Gallium Nitride (GaN) crystals were synthesized in 1932 by W. D. Johnson by passing ammonia over heated Gallium [21]. However, large crystals of GaN were not synthesized until 1969, when Maruska and Tietjen grew GaN on sapphire with hydride vapor phase epitaxy [22].

A variety of devices can be constructed using the properties of GaN. An early switching device made using GaN was a MESFET created by Khan at APA Optics in 1993 [23]. Soon after, Khan demonstrated a GaN/AlGaN-based HEMT [24]. GaN nanotubes (similar to carbon nanotubes) have also been formed, with some of the earliest examples including Goldberger at UC Berkeley in 2003 [25] and Hu at NIMS [26].

Device Principles

GaN devices can be considered as spontaneously formed HEMT devices. Gallium Nitride and Aluminum Gallium Nitride are polar materials due to the large size difference between Gallium and Nitrogen. When AlGaN is deposited on GaN, the tensile stress of AlGaN puts on GaN causes piezoelectric polarization to occur. This polarization leads to the formation of electrons and holes, whose charges normally cancel each other out. However, due to the heterojunction at the interface, the AlGaN/GaN interface collects the electrons as an electron gas that can be used as a conduction channel, very similar to a traditional HEMT [27], [28].

Recent Work

Challenges for GaN-based devices are similar to those of HEMTs. These include generating high drive current at low voltages, reducing gate leakage, and integrating p-type devices. Additional challenges include finding the best way to create enhancement mode devices, and reliability.

Significant work has been undertaken to tackle the challenges of GaN HEMT devices. Using N-face surfaces of GaN rather than Ga-face allowed Nidhi et al. to demonstrate depletion mode GaN devices producing about 1 mA/μm at V_{DS}=1 V [28]. Xin and Chang have shown that high-κ dielectrics such as ALD HfO_2 or Al_2O_3 can reduce the gate leakage [29],[30]. Chung et al. demonstrated a 2 layer transfer process to integrate p-type Silicon MOSFETs with n-type GaN [15].

To create enhancement-mode devices, researchers have used several methods to change the threshold voltage. Cai et al. used fluorination through the use of a CF_4 etch to passivate surface states, changing the threshold voltage by 5 V and allowing the creation of both enhancement and depletion mode devices, which was demonstrated by creating ring oscillators [31]. Ota et al. used piezoneutralization (a layer inserted beneath the gate to neutralize polarization charges underneath the gate) to adjust V_T [32]. Silicon Nitride was used by Derluyn et al. to passivate the surface charge in AlGaN as another method of adjusting V_T [33]. Kanamura et al. used the piezoelectric effect of i-AlN on n-GaN to create an enhancement mode device while increasing the

2DEG density [34]. Finally, Im et al. demonstrated that a superlattice of AlN/GaN changes the biaxial stress to make enhancement mode devices with better on-state resistance [35].

Joh and del Alamo studied reliability concerns for GaN devices, and at V_{DS}=5 V, hot carriers caused reductions in on-state current and changes in V_T, with greater voltages causing faster degradation [36]. In addition, lattice defects form due to excessive stress from the inverse piezoelectric effect.

Scaling

Short channel devices with gate lengths down to 20 nm (with 40 nm source/drain offsets) have been developed by Shinohara et al. with record high on-current of 2.7 mA/μm [37]. Both enhancement and depletion mode devices were fabricated with high uniformity. However, voltages are still high to achieve these results (3-5 V). Uren et al. found punchthrough effects occurring in devices with a 0.17 μm gate length due to leakage through the GaN buffer layer. Uren proposed that buffer layers should be insulating to prevent this and confine the channel potential [38]. Park and Rajan found that N-polar GaN HEMTs suppressed DIBL better than Ga-polar HEMTs due to the N-polar device's superior electrostatics from its inverted structure [39].

Conclusions

Gallium Nitride based HEMTs' major advantages are high electron mobility (although not as high as GaAs), higher critical breakdown voltage, and a higher thermal conductivity than GaAs [27]. Gallium Nitride based devices have been suggested to be useful in RF or high voltage applications. Some groups have also thought that these devices could also be useful as traditional MOSFET replacements [30].

There are several challenges that would need to be overcome to replace traditional MOSFETs. First, Gallium Nitride (GaN) devices are typically depletion-mode rather than enhancement-mode. Next, to avoid HEMT gate leakage, gate dielectrics need to be developed that are compatible with Gallium Nitride. Note that Gallium Nitride only holds an advantage for n-type devices so p-type devices, such as Silicon or Germanium, would need to be fabricated on the same wafer. Additionally, reliability issues due to hot carriers need to be better studied. Scaling also needs further study, as devices have mostly been long channel up to this point. The inability to create GaN ingots as cost effective substrates (or Silicon Carbide ingots coupled with GaN deposition) means that sophisticated (and potentially expensive) techniques would need to be developed to integrate GaN on more conventional substrates.

GaN based devices seem to perform best as power or RF solutions where voltages are too high for logic applications. Thus, GaN seems best suited for telecommunications and radar applications. For solutions like WiMax base stations and power electronics, GaN could also be useful. However, the limitations of GaN combined with integration challenges (particularly very thick buffer layers for growing defect-free GaN), GaN does not appear to be appealing for future development in logic.

Ferroelectric Gate Stacks

Ferroelectric gate stacks use a ferroelectric capacitor in series with a traditional gate oxide to create a region of negative gate capacitance, resulting in subthreshold swing below 60 mV per decade.

History

In 2008, Sayeef Salahuddin at Purdue theorized that using a ferroelectric capacitor in series with a normal capacitor should stabilize the ferroelectric material and allow an overall negative gate capacitance effect to occur [40]. By making C_{GATE} negative, Salahuddin predicted it would be possible to overcome the limitation that normally prevents operation with a subthreshold slope less than 60 mV per decade at room temperatures [41],[40]. A device was fabricated by Salvatore at EPFL using the ferroelectric dielectric P(VDF-TrFE) and was found to achieve 13 mV/dec behavior at low current in 2008 [42]. Rusu at EPFL demonstrated a device in 2010 with sub-60 mV/dec behavior over 2.5 decades of current [43].

Device Principles

In a standard dielectric, an energy versus charge curve shows a $Q^2/2C$ relationship. In a ferroelectric dielectric, the energy versus charge curve has two minima, resulting in a region of negative capacitance between the minima. By biasing the device in this region of negative capacitance, it is possible to create overall negative gate capacitance. The equation for Subthreshold Slope is given in Equation 2.1.

$$SS = \ln(10)\frac{kT}{q}\left(1 + \frac{C_{DEP}}{C_{GATE}}\right) \quad (2.1)$$

By making C_{GATE} negative, Salahuddin predicted it would be possible to overcome the limitation that normally prevents operation with a subthreshold slope less than 60 mV per decade at room temperatures [41],[40]. The device would be fabricated by placing the ferroelectric dielectric between the gate metal and conventional dielectric.

Recent Work

Research on ferroelectric devices is very recent, with significant challenges today in simply observing the negative capacitance effect and achieving sub-60 mV behavior. Only a select few groups so far (primarily [42], [43]) have been able to achieve sub-60 mV/dec subthreshold swing. However, Tanakamaru et al. looked at ferroelectrics for SRAM and was unable to achieve sub-60 mV/dec operation [44]. Khan et al. demonstrated a proof of concept device, which also does not show sub-60 mV/dec behavior, but was focused on demonstrating clear proof of the negative capacitance effect [45]. Complicating the research landscape, it is not possible to directly measure negative capacitance – only an enhancement in total capacitance. Krowne et al. incorrectly interpreted a lack of measurable negative capacitance as an indication that ferroelectrics do not cause negative capacitance, but instead cause highly nonlinear biasing behavior when in a series capacitor stack [46].

Scaling

As the mechanisms for generating and operating negative capacitance are better understood, fundamental limitations may be uncovered for these devices. Jin et al. simulated the potential for scalability of these devices and found that the subthreshold slope rises as the gate length is scaled down, with substantial increases below 50 nm negating their primary benefit [47].

Conclusions

The potential for sub 60 mV/dec operation, if expanded over multiple decades of current, is a large potential advantage for ferroelectric gate stack devices. However, the basic understanding of how to fabricate and design these devices still needs further study. Hysteresis is also a concern, as a large hysteresis would prevent voltage scaling, further negating the underlying value of pursuing sub 60 mV/dec operation. While scaling studies that have suggested the technology scales poorly, more research is necessary, and the costs for incorporating a ferroelectric layer are relatively minimal. Even if there is a gate length limitation, if hysteresis can be reduced, these devices could find a use in a large gate length, ultra low power application.

Electro Chemical Devices

Electro chemical devices (ECDs) use a chemical reaction to control the flow of current through a device.

History

Devices as small as those with a single molecule were theorized as early as 1974 [48]. Modern ECDs were experimentally demonstrated by Collier in 1999 based on the concept of Chemically Assembled Electronic Nanocomputers (CAENs) [49]. Using Ag_2S as a filament, Terebe was able to produce various logic functions, including AND, OR, and NOT [50].

Device Principles

Electro-chemical devices, like all switching devices, control the flow of current through a device. However, they use a chemical mechanism (for example, the reaction of Cu ions precipitating out of Cu_2S as a voltage is applied) to create a conducting bridge. Some are two terminal devices, and others are three terminal devices. They are often used as memory devices, which can either be write-once (irreversible), or write many times (reversible).

Recent Work

Challenges for ECDs mostly focus on switching speeds, reliability (expressed in cycles, which is often called endurance), and circuit fabrication issues.

Terebe et al. attained 1 MHz operation in 2005 using Ag_2S, although this is orders of magnitudes from what would be necessary to compete with scaled CMOS logic [50]. Thomson et al. in 2006 showed switching speeds as fast as 0.1 microseconds [51]. Sakamoto et al. in 2007 switched to Ta_2O_5 and was able to increase $V_{PROGRAM}$ to over 1 V while keeping switching times in the 10^{-5} to 10^{-4} seconds range [52].

Improvement of reliability, as expressed in mean cycles before failure, remains difficult. For three terminal devices, Sakamoto showed that a gate isolated from a

filament can be used to control the filament's conductivity, although endurance was lacking at only 50 cycles [53]. In 2007, Sakamoto et al. had been able to increase endurance to 10,000 cycles by using a Cu_2S electrolyte [52].

A wide variety of circuit concepts have been explored. Molecular FETs, using molecules such as Roaxane, have been shown for over a decade to be able to achieve basic logic functionality, such as AND and OR, but at very low currents (less than 1 nA) [49]. Using $CuSO_4$, basic look-up tables can be created to mimic FPGA technology [54]. (Unfortunately, endurance cycles of the FPGA in [53] are still low, at about 100 cycles, and retention and switching times need to be improved.) With a nanowire in a porous alumina membrane, Liang et al. was able to connect several nanowires in parallel to create similar FPGA-like devices [55]. A solid electrolyte in the back-end between a via and metal line have also been explored, but mostly as a nonvolatile memory on top of logic [56]. Endurance would again need to be improved beyond 100 cycles.

Scaling

Studies of scaling are limited, as fabricating larger single devices is still challenging. Most work has focused on the use of ECDs for memory applications. Kim and Nishi found that as cell size shrinks, the ratio of on-state to off-state resistance increases, implying that a larger area gives rise to more filament formation, negatively impacting performance [57].

Conclusions

ECDs hold an advantage in their size, which can be on the order of individual atoms. However, these devices are still at too early a stage of development for use as a CMOS replacement without major breakthroughs. Unfortunately, electrochemical devices currently demonstrate low current, low endurance cycles, or poor switching speeds, and in some cases, all three negative qualities [58]. There have only been limited studies on scaling, and costs may or may not be significant, depending on the materials and process technologies needed for fabrication. Electrochemical devices may be more suited for use as nonvolatile memory devices.

2.4.1.2 Carrier Transport Mechanisms

Impact Ionization

Impact Ionization transistors are gated p-n diodes that rely on avalanche breakdown to create carriers in the channel. This mechanism creates positive feedback, allowing for sub-60 mV/dec subthreshold slopes.

History

Impact Ionization FET devices, also known as IMOS, were simulated and fabricated in 2002 at Stanford University [59]. Kailash Gopalakrishnan, under the direction of Prof. James Plummer, was searching for a gain mechanism that was internal to the device with sufficient gate control. Gopalakrishnan simulated devices showing subthreshold slopes down to 5 mV/dec and fabricated devices with about a 10 mV/dec

swing.

Device Principles

Impact Ionization transistors use the avalanche mechanism of breakdown in reverse-biased diodes to achieve carrier transport. They are also gated p-i-n diodes, but they have a larger intrinsic region that is partially not gated. The device works by modulating the channel length with the gate. At high gate voltages, the gate inverts a portion of the channel, reducing the effective channel length, and increasing the electric field from the drain to the source. Eventually, the device breaks down with impact ionization, causing current to quickly flow to the drain.

Recent Work

Challenges for IMOS devices include high drain voltage requirements, reliability, and circuit issues.

To cause avalanche breakdown, a high V_{DS} is required, which increases power dissipation. Nematian, Fathipour, and Nayeri found that this drain voltage could be reduced if other materials with reduced bandgaps are used (such as SiGe) [60].

Reliability and variability can be an issue. Abelein et al. found that carriers with such high energy levels can cause large changes in V_T after multiple cycles [61].

IMOS has issues with increased C_{GD} due to high Miller capacitance (the drain couples to the entire intrinsic region of the device), as shown by Tura and Woo [62]. Also, the devices do not fully saturate with high drain voltages. As such, they will not exhibit full rail-to-rail swing, another difficulty for using these devices in logic applications.

Scaling

Some issues with scaling IMOS include the need of an ungated intrinsic region. This will hinder future scaling. Savio et al. showed that Silicon IMOS will not scale well below 50 nm in Silicon [63]. Additionally, as shown by Shen et al., the need of time and space for the carriers to build enough energy for carrier multiplicative effects to occur, limits both the fundamental scaling length of these devices, and the switching time [64].

Conclusions

Impact Ionization FETs have advantages of low subthreshold slope with relatively high current compared to other sub-60 mV/dec devices, like TFETs. However, there are many challenges, such as reliability, avalanche onset delay, and high drain voltages. Scalability is a significant concern, although cost would be similar to traditional MOSFETs. These devices could be useful for very specific circuit applications where a variable V_T can be used, such as a low power write once memory element.

Tunnel FET

Tunnel FETs use quantum-mechanical tunneling of electrons from the source to the channel as the primary carrier transport mechanism, allowing for sub-60 mV/dec subthreshold slopes.

History

The origins of three-terminal devices come from the band-to-band tunneling

component in a Trench Transistor Cell [65]. A three-terminal tunnel device using this effect was proposed by Sanjay Banerjee at Texas Instruments in 1987 [66]. This device required gate overlap of the source, a situation later known as line tunneling. Later in 1992, Toshio Baba at NEC proposed a surface tunnel transistor using GaAs and AlGaAs that utilized point tunneling [67]. In 1995, William Reddick at Cambridge proposed a Silicon device using point tunneling [68]. All of these devices showed low currents and no subthreshold slope under 60 mV/dec. In 2004, Jorge Appenzeller at IBM showed experimental characteristics less than 60 mV/dec with carbon nanotube based devices [69].

Device Principles

Tunnel Field Effect Transistors (TFETs) use the tunneling of electrons as the carrier transport method for device operation. They are generally designed as gated p-i-n diodes, where the gate is used to modulate an effective tunneling barrier height [70]. Ideally, these devices would have a very low off-state current (proportional to reverse-biased diode leakage), a very low subthreshold slope, and acceptable on-current.

TFETs can be generally classified as point and/or line tunneling devices [71]. In a point tunneling device, the source does not appreciably deplete, but the gate causes the channel region to invert, resulting in tunneling from the source to the channel. In a line tunneling device, the source is inverted (generally by engineering an overlapped gate with an optimized source doping profile), resulting in tunneling into the inversion layer, similar to Gate-Induced Drain Leakage (GIDL).

Recent Work

TFETs' major challenges are to achieve significantly better subthreshold slope than 60 mV/dec and provide drive current comparable to MOSFET devices. Miller capacitance is also a challenge due to the p-n diode nature of TFETs, similar to IMOS. Ambipolar operation (tunneling occurring at the drain when the gate is reverse biased) and circuit design challenges (due to asymmetric device operation) will also require further understanding.

Few devices are able to achieve subthreshold slope of less than 60 mV/dec. Appenzeller et al. showed this with carbon nanotubes in 2004 [69]. Lu et al. also demonstrated this effect using DNA functionalization on carbon nanotubes [72]. Choi et al. demonstrated this with a purely Silicon device in 2007 [73]. This was followed by a sub-60 mV/dec device by Mayer in 2008 [74]. Jeon et al. used a silicided source to achieve sub-60 mV/dec switching in 2010 [75]. Leonelli et al. also demonstrated sub-60 mV/dec behavior with FinFET devices [76]. Kim et al. used a Germanium source to achieve sub-60 mV/dec operation in 2009 [77].

Even the best experimental devices (such as the device from Kim et al.) show I_{ON} in the μA/μm range [77], but do not meet the mA/μm requirements for future CMOS devices. Some strategies, such as that proposed by Kim, use a recessed Germanium source [77] and have the potential for increased drive current. Another approach, from Mookerjea et al., is to use one material (creating a homojunction rather than heterojunction) with a lower bandgap, allowing for a higher tunneling rate and hence,

higher tunneling current [78].

In addition, drain capacitance may increase due to enhanced Miller capacitances. Since the entire channel is only electrically coupled to the drain (rather than the source and drain, as in a typical MOSFET), the drain experiences increased C_{GD} and the source experiences decreased C_{GS} [79],[80]. Increased overshoot is also possible in these devices.

A homojunction TFET (with a single source, channel, and drain material) has ambipolar characteristics. To remove the ambipolar effect, the source and drain must be asymmetric, either by use of a heterostructure or offset drain [81],[82].

As an asymmetric device, TFETs can only conduct tunneling current in one direction, making circuit design more difficult. Groups have examined new SRAM and logic layouts and found that additional transistors (for example, a 7T SRAM cell) may be necessary to have sufficient noise margins for operation [83].

Scaling

Although simulations of advanced device structures show increased on-state current, many simulated cases require extremely abrupt junctions or doping profiles that have not been achieved in TFETs to date [84]. Some of these device structures require multiple junctions underneath the gate, which would reduce future scalability. In general, tunneling current is proportional to the barrier height (determined by bandgap/heterojunctions) and tunneling width (dependent on electrostatics and doping concentrations/gradients). TFETs with gate overlap have improved tunneling area (as well as electric field in the tunneling region), but also have reduced scalability. One solution is decoupling the overlap area from the gate length by using a raised source where tunneling is contained completely within the source [85].

Conclusions

Experimental TFET results show very low subthreshold slope at very low currents, in sharp contrast to the simulation results which show low subthreshold slope but with on-currents in the 0.1 mA/μm to over 1 mA/μm range. Unfortunately, in practice, devices have not been able to simultaneously show both subthreshold slopes below 60 mV/dec and high on-state currents. Devices that do achieve reasonably high on-currents do not see subthreshold slopes below 60 mV/dec, making off-state currents very high and eliminating the advantage over conventional MOSFETs. It is important to note that TFETs are not symmetric, so additional challenging lithographic steps are necessary, complicating fabrication. Scaling seems robust for TFETs, with few cost increases compared to MOSFETs. However, unless TFETs improve to better match their simulation results, they have limited application for logic devices.

2.4.1.3 Structures

Ultra Thin Body

Ultra Thin Body (UTB) devices isolate individual transistors from each other by

placing an oxide layer beneath the devices. By making the device depth very thin, the gate can control the entire body of the device, allowing it to be fully depleted by the gate.

History

Early work on MOSFETs constructed on isolated Silicon occurred in 1978 by K. Izumi at NTT in Japan [86]. Izumi implanted oxygen below devices. He then used annealing to form a Silicon oxide layer beneath the devices. However, this work did not publish individual device characteristics, although it did show ring oscillator results. Izumi also remarked that reduced leakage was measured and the devices could be useful for high performance logic.

In 1979, activity began increasing in SOI devices. K.F. Lee at Stanford University reported thin film devices grown on laser-annealed polysilicon [87]. These devices were targeted at low-cost and large-area applications where Silicon substrates were not possible. Later that year, A.F. Tasch at Texas Instruments demonstrated devices with much smaller gate lengths, down to 5 microns [88]. In 1982, S.D.S. Malhi realized that by using an ultra thin body thickness, the channel could be nearly intrinsic, resulting in enhanced mobility for carriers [89]. In 1983, Lim and Fossum studied how back-gate bias could be used to optimize performance in thin SOI devices [90]. Also in 1983, J.-P. Colinge created large single crystal SOI films using laser recrystallization [91].

Device Principles

Traditional MOSFETs experience short channel effects as the channel length scales. At sufficiently short channel lengths, the gate is no longer able to control the channel effectively due to the drain's electric field. Using higher doping in the channel can somewhat counter this effect, but leads to decreased mobility, hindering device performance.

UTB devices are one potential solution to improve gate control [92]. The devices are formed on Extremely Thin Silicon On Insulator substrates. By using these very thin bodies, the channel is very thin, and leakage paths between the source and drain are reduced. The channel is fully depleted, leading to improved subthreshold swing. With improved gate control, channel doping can be reduced, improving mobility.

Recent Work

UTB devices exhibit great promise for dimensional scaling, but they have limitations due to lack of strain (a key driver of on-current improvement in modern CMOS), increased source drain resistance, and threshold voltage control.

Ang et al. demonstrated enhanced mobility in NMOS by using SiC for S/D regions to strain the channel region. It was observed that a recessed S/D allows for better drive current. However, the mobility is dependent on orientation and device width [93]. Another method to add stress and enhance drive current was exhibited in PMOS by Chui et al. using Ge condensation by annealing. This method eliminates the need for a Si recessed etch before the SiGe epitaxy [94]. Lastly, a quasi-SOI was made by Tian et al., for both NMOS and PMOS which ideally minimizes SCEs, eliminates potential coupling, shows high mobility, and a decreasing $C_{PARASITIC}$. However, the

trade off is poorer quality of the epitaxially grown silicon in the channel, and small window for design of the source drain extensions [95].

Scaling

Scaling, especially the body thickness, is a concern for ultra thin body devices. As channel length scales, body thickness must decrease. Severe mobility degradation is exhibited below a 3.5nm thickness [96]. Also, quantization effects in the inversion layer can lead to an increase of threshold voltage and thus a decrease in on-state current. The effect of just a single unintentional impurity can degrade the drive current and increase threshold voltage, as shown by Vasileska et al. [97]. Self-heating is an additional concern. In 1989, McDaid et al. showed that negative differential resistance in SOI output characteristics was due to the reduced thermal conductance of buried oxide [98]. More recently, Fiegna et al. showed that thermal resistance increases as the gate length and body thickness are scaled [99]. Reducing the back oxide thickness reduces thermal resistance, potentially offering some room for improvement.

Threshold voltage control has been recently studied in Ultra Thin Body devices. Ren et al. demonstrated excellent variability and mismatch control with gate lengths down to 30 nm [100]. Liu et al. and Andrieu et al. demonstrated Ultra Thin Body and BOX (UTBB), which allows back bias to change the threshold voltage with good variability control[101], [102].

Conclusions

Ultra Thin Body devices offer better gate control and improved mobility over bulk devices. However, there are limitations to the improvement due to increased source and drain resistance and ability to strain the channel. Scaling may be a concern in the future, as well as additional costs due to the use of ETSOI substrates. For manufacturing processes currently using SOI technology, UTB represents the next step for scaling, albeit one with limited future scaling potential.

Multi Gate

Multiple gate FETs improve electrostatic confinement by wrapping the gate around the channel. This improves short channel control, reducing leakage and improving scalability.

History

Multiple gate FETs were initially discussed in the late 1980s as Surrounding Gate Transistors (SGTs). In 1987, K. Hieda at Toshiba realized a "triple-gate" structure by creating a trench isolated transistor with a gate that surrounded the channel [103]. The purpose of this device was to reduce the bird's beak effect common in LOCOS oxidation methods and was designed for use as the switching transistors in memory cells. In 1988, H. Takato at Toshiba demonstrated that this device would be useful for logic due to the excellent gate control [104]. Digh Hisamoto at Hitachi later produced several works where the Silicon fin had a reduced width, known as a Depleted Lean Transitor (DELTA), which further reduced short channel effects and increased gate control [105],[106],[107].

Device Principles

Multiple Gate FETs utilize multiple gates wrapped around the channel for better channel control in short channel devices. This approach reduces short channel effects, improving subthreshold characteristics and DIBL. These devices have been incorporated in the ITRS roadmap, and as of the 2009 ITRS revision, are projected to be utilized starting somewhere between 2015 and 2020. Multiple gates are also generally three-dimensional, allowing for increased channel area in a given plan-view area.

Recent Work

The many types of multiple gate FETs have individual advantages and challenges [108]. In general, increasing the electrostatic confinement improves short channel control and scalability. However, fabrication complexity often results as confinement is increased from fins to gate all around devices like nanowires.

Traditional FinFETs for integration and node insertion have been realized by Chang et al. for the 32 nm node, with 25 nm gate length devices showing 1296 μA/μm for n-type and 925 μA/μm for p-type [109]. Another type of FinFET, by Zhang, Fossum, and Mathew, would use the Ultra Thin Body region between fins to also conduct current. It has worse off-state characteristics, but might be useful for I/O circuits where maximum on-state current is required [110]. Independently gated double gates, even when arranged in a fin, likely will not scale well due to contact spacing issues, as shown by Mathew, et al. [111]. Wu et al. showed that stacked fin technologies would be difficult to fabricate when combined with raised source drain and stress technologies, as well as having issues with contact area scalability [112].

Gate-All-Around nanowire-like devices show comparable performance to p-type partially depleted SOI, but still lag in drive current compared to n-type devices, as demonstrated by Bangsaruntip, et al. [113]. Yeo et al. and Fang et al. demonstrated Twin Silicon Nanowire FETs (TSNFETs) that show excellent subthreshold slope and DIBL due to their high quality gate oxides and excellent gate control [114],[115]. 3D stacked nanowires by Dupre et al. with FinFET like characteristics show excellent possibilities for scaling, although independent gate control of several layers is unlikely [116]. Orientation effects shown by Singh et al. demonstrate that for p-type nanowires, the <010> direction is superior to the <110> direction by 1.84x the I_{ON} mean [117]. Mobility is degraded in Silicon nanowires due to phonon-scattering as nanowires get thinner, especially smaller than 6 nm [118]. Mobility has been shown to be linear with wire radius for InAs nanowires in the 7-18 nm range [119].

A unique device called the VeSFET, or Vertical Slit FET, uses regular arrays of pillars to form a very compact structure that can implement a logic function, such as AND or OR, in a single device [120]. Srivastava, Saubagya, and Singh simulate this device, which is very interesting for using space so efficiently, but difficult to fabricate.

Conclusions

Multiple Gate FETs hold promise for continued scaling beyond planar devices. However, multiple gate devices have increased process complexity when compared to planar, and the fabrication and process costs associated with Multiple Gate FETs need

to be considered. Nanowire-based and gate-all-around devices offer maximum gate control for scaling, but their structures are more complex to manufacture.

Metal Source / Drain

Metal Source / Drain technology uses Schottky barriers instead of doped sources and drains to reduce parasitic resistance.

History

The use of Schottky contacts for the source and drain was proposed and demonstrated by M.P. Lepselter and S.M. Sze at Bell Labs in 1968 [121]. They used Platinum Silicide source and drain regions and an n-type body region to demonstrate the first Schottky based S/D devices.

Device Principles

Metal Source Drain devices, also known as Schottky Barrier Field Effect Transistors (SB-FETs), traditionally use Schottky barriers rather than diode junctions as the source and drain. Using metal rather than a doped semiconductor reduces parasitic resistances, but requires band-edge metal work functions to be able to match on currents of traditional MOSFETs [122].

Recent Work

Metal Source / Drain devices have Schottky barriers at the source and drain. To achieve comparable drive current to standard MOSFETs, the barrier height must be reduced (the barrier workfunction must approach band-edge), with the amount of reduction dependent on the drive current requirements. Methods for reducing the barrier height include device structure optimization, Fermi level depinning, implants, and dopant segregation.

Connelly et al. has shown that a Schottky Barrier Height (SBH) less than 0.1 eV is needed to compete with traditional MOSFETs [123]. Underlap is preferred for these devices rather than conventional overlap of the source and drain, first due to parasitic capacitance, and second because the abrupt profile of a metal-semiconductor junction allows for a slight underlap while still maintaining gate control.

SB-FETs utilizing a double-gated structure can meet ITRS benchmarks with poorer SBH than single gate structures, suggesting GAA structures may be more attractive with metal S/D devices (neglecting the associated volume efficiency issues with GAA) [124].

Chen et al. created a planar structure where silicide was grown on top of the source and drain, and a small finger of silicide spread toward the gate to improve source drain extension resistance. This method also allows the use of strain with embedded SiGe [125].

Fermi level depinning can reduce Schottky barrier height. A thin layer of nitride was used by Connelly et al. to depin the Fermi level and reduce SBH to 0.2 eV, at the cost of increased R_{EXT} [126],[127]. Another method by Tao et al. uses a Selenium monolayer to reduce the SBH down to less than 0.08 eV[128].

Vega and King-Liu also show that a fluorine implant can reduce SBH close to 0

eV, but in FETs, the fluorine implant resulted in higher resistance, resulting in an ultimately lower drive current [129],[130].

Dopant-Segregated Schottky (DSS) MOSFETs use dopants at the Metal-Semiconductor interface to reduce the SBH. Experimentally, this was demonstrated by Kinoshita et al. in 2004 and was shown by Qiu et al. to be achieved for both implantation to and implantation through silicide [131],[132]. An excellent table showing different SBH for different anneal conditions is shown by Qiu et al. [132]. DSS FinFET and nanowires have also been demonstrated by Kaneko et al. and Chin et al. [133],[134].

Scaling

Vega and King-Liu examined the scaling potential of DSS devices for double gate structures. For High Performance applications, at a 10 nm gate length, a conventional double gate Raised Source Drain (RSD) structure will achieve higher performance than a DSS structure unless the epitaxial layer is doped less than 10^{20} cm^{-3} [135]. For low standby power applications, optimal parameters for various SBHs are shown [135]. For low operating power, Vega showed the advantages of dual high K / low K spacer technology for DSS structures, which allows fringing fields to be enhanced such that the source and drain can be underlapped, reducing parasitic capacitance and increasing effective gate length [136].

Conclusions

Metal Source / Drain devices show some advantages if SBH can be reduced to within 0.1 eV of band-edge. However, simulation studies of High Performance devices show that conventional Raised Source / Drain structures are better performing than DSS structures (even without accounting for the potential loss of strain effects with Metal Source Drains). Although some results suggest Metal Source / Drain devices can outperform conventional MOSFETs (especially if the SBH is 0 or negative [123]), lack of strain and increased parasitic capacitance effects are likely to result in lower performance overall. At very small gate lengths, while conventional raised source drain double gate devices scale better, fabrication costs may make Metal Source / Drain an attractive option.

Novel Vertical Devices

The category of vertical FETs is used to describe unique process integration schemes that allow the channel to be fabricated such that current flows in the vertical direction, either into or out of the wafer, as opposed to the normal lateral flow of current in traditional MOSFETs. This category was also used to apply to unique integration schemes for fabricating multiple layers of devices stacked in the vertical direction.

Recent Work

Sacchetto et al. used the Bosch etch process to create a vertical nanowire structure [137]. This process uses a repetition of dry etch then passivation to create a scalloped effect, which allows for the creation of multiple layers of crystalline devices

without the need for epitaxy. This process could also be useful for III-V devices, only requiring an additional oxidation step to thin and separate the nanowires.

Fukuzumi et al. created a "macaroni" FET by first creating layers of gate material and oxide spacer, then etching a vertical hole, and finally filling the vertical hole with gate oxide and channel material [138]. Unfortunately, this process is difficult to accomplish with crystalline semiconductors using epitaxial growth. This device also seems best suited for memory devices, since all gates on a level are shorted together, resulting in a bit-line/word-line arrangement.

Thelander et al. shows how Vapor-Liquid-Solid/Vapor-Solid-Solid growth allows for vertical FET creation on a substrate. Specific processes such as polishing and manipulating conformal versus nonconformal deposition are needed to contact, gate, and isolate individual devices in a wrap-gate arrangement [139].

Conclusions

These vertical devices all offer tradeoffs in process flow complexity for the ability to make very specialized vertical structures. Vertical structures offer some packing density benefits for certain circuit configurations (for example, memory), but at the cost of increased fabrication complexity.

Junctionless Accumulation Mode

A Junctionless Accumulation Mode (JAM) device is a fully depleted device, with either multigate or gate-all-around structures, where the source-drain regions have the same doping type as the channel region. Both traditional top-down and bottom-up nanowire devices have been studied for their advantages in fabrication complexity [140], [141].

History

Top-down fabricated multigate devices were originally studied by Lee et al. in the Colinge group as "junctionless transistors" [140], [142]. The devices have high n-type doping (in the 10^{19} cm^{-3} range), with uniform doping throughout the channel, source, and drain [143]. Devices operate with a very small cross-section (on the order of 5 nm x 5 nm) to permit fully-depleted operation at a desired gate voltage. In addition to being relatively easy to fabricate, these devices also have a lower electric field than a conventional CMOS device [144]. Iqbal et al. published a reference paper showing optimized geometrical and doping parameters for guidance in designing these devices [145].

The key trade-off in these devices is between the mobility gain due to reduced electric field and the mobility loss due to increased doping. Rios et al. showed experimentally that for a low doped case, the low field effects win and the mobility is ~30% higher. However, for a high doped case, the mobility is reduced due to impurity scattering. Rios also points out that these devices have a mixed threshold behavior where a low value governs the subthreshold turn-on and a higher one determines the extrapolated threshold of the accumulation regime [146]. In addition, when measuring temperature dependence, dV_T/dT is very poor compared to traditional inversion mode

devices, which would reduce voltage scaling, a critical criteria in future devices [147]. Geometric variation will also be problematic, as junctionless nanowires have been shown to have significantly higher threshold voltage variation than a comparably sized inversion mode device as the width scales [148].

One additional use of these structures is for operation as impact ionization devices, as shown by Lee et al. in the Colinge group. A lower V_D (just above the bandgap) can be used because the drain bias is fully dropped at the drain rather than throughout the channel [149]. Sub-60 mV/dec subthreshold swing operation can be achieved at drain biases as low as 1.75 V, although the I_{ON} is lower than that of traditional impact ionization devices of the same size. The region where impact ionization occurs is much larger in the junctionless device, which might make it possible to achieve sub-60 mV/dec behavior at less than 1 V with a germanium device [149].

Similarly to the top-down junctionless devices, Vapor-Liquid-Solid (VLS) grown bottom-up nanowires have been fabricated with single doping concentrations throughout the source, channel, and drain, in this case due to difficulty in controlling the doping profile during growth and fabrication. Unfortunately, even with surface passivation, work by the Lieber group has shown bottom-up p and n-type Silicon and Germanium nanowires continue to display poor off-state characteristics [150]. In addition, there are assumptions made in these Si and Ge bottom-up research devices that should be considered, including assuming no quantization effects when calculating transconductance [141],[151] and using an idealized theoretical capacitance [152],[153].

In all of these devices, the source and drain would ideally be more highly doped (for improved access resistance) and the channel more lightly doped (for improved mobility). This means that, in the limit, a junctionless devices has an undoped channel and a heavily doped source and drain, which is a traditional multiple gate device. The one benefit that these junctionless devices (with similar doping in S/D and channel) hold is in ease in manufacturing. If costs can be reduced sufficiently such that the reduction in current (leading to a larger device width and thus larger area per device) is offset with significantly reduced costs in processing (fewer mask steps, no epitaxy, etc.), then these devices could help continue Moore's law, at least on a cost basis.

Relays

Mechanical relays use physical movement from an OFF to an ON position to regulate current.

History

Although mechanical relays pre-date solid-state devices by many decades, Micro-Electro-Mechanical Systems (MEMS) technology that could compete with CMOS in scalability was first demonstrated in 1978 by Kurt E. Petersen [154]. Petersen demonstrated three devices: an optical display, a 4-terminal micromechanical switch, and a measurement method of Young's modulus.

Device Principles

Relays use mechanical movement to physically short or open an electrical connection between two contacts. Relays can be placed either in the front-end or back-end of a traditional CMOS process [155]. Ideal MEMS relays show no off-state current, sharp subthreshold slope, and low gate leakage. Resistance matters less in these devices than in CMOS because the relatively slow mechanical beam movement delay is the limiting factor, not the faster RC delay time constant.

Relays have a pull-in voltage (V_{PI}) determined by the actuation force (usually electrostatic) overcoming mechanical force and a pull-out voltage (V_{PO}) determined by mechanical force overcoming adhesion forces. To actuate relays, several different methods can be used, including thermal, magnetic, piezo, and electrostatic [156] [157],[158],[159],[160]. While MEMS devices can be engineered for $V_{PI} = V_{PO}$ practical implementations frequently display hysteresis (where $V_{PI} > V_{PO}$).

Recent Work

Relays have the advantages of steep subthreshold slope and negligible I_{OFF}. In addition, 4-terminal relays allow pass-gate logic, which potentially reduces the number of devices needed per function [161]. However, relay operating voltages are high and need to be reduced to achieve the benefits of low active power. In addition relays are currently very large (for example, 7.5 μm x 7.5 μm [162]) and need to be reduced to sizes comparable with CMOS. Reliability needs to be demonstrated for use in high activity factor logic. Variation and hysteresis also need to be reduced to values comparable to CMOS.

In 2010, Kam et al. showed that relays have the potential to be 10x more energy efficient than CMOS (albeit at lower frequencies) [163]. Relay circuits were demonstrated by Spencer et al. including a full adder in 12 NEMS relays with a single mechanical delay [164]. Hossein et al. has designed a 16-bit relay multiplier, which promises to achieve lower energy per operation than CMOS, as well as experimentally demonstrated a 7:3 compressor composed of 98 relays [188].

High pull-in voltages remain an issue with relays. Lee et al. used an insulating liquid (such as oil) to reduce V_{PI} with the liquid's higher dielectric constant, but relay reliability was still worse than that achieved with an ALD process [158]. Carbon nanotubes can also be used to reduce voltage, but can be difficult to fabricate, as discussed by Dadgour et al. [165],[166]. A suspended gate MOS fabricated by Abele et al. combines relay with MOS for enhanced efficiency, but current is low, and both mechanical and RC delay are issues with this device [167].

Shen et al. used simulation tools to show that scaled relays with feature sizes down to 10 nm will have pull-in voltages of less than 0.25 V [168]. Pott et al. explained that relays will ultimately be limited by contact asperities (surface roughness), but scaling to the 65 nm node would result in an actuation area 67% larger than a comparable MOSFET [156].

Reliability up to 65 B cycles has been demonstrated with appropriate contacts. Joshi et al. demonstrate back-end relays to 10^{11} cycles (In comparison, for a device operating a 100 MHz with an activity factor of 1%, 10^{15} cycles would be needed for 10

years of operation) [155].

Variation in contact resistance has been shown to have little impact on energy-performance characteristics, even when comparing best and worst in class contact materials [169]. Dadgour et al. showed that a 10% variation in beam length and width (for carbon nanotubes) has a dramatic effect on pull-in voltage distribution [165].

Scaling

Spencer et al. demonstrated 4-terminal relay circuits such as adders and then produced a theoretical layout for a 90 nm technology node relay [164]. At the 90 nm node, a 32-bit adder was projected to require 7000 μm^2 of area versus 2000 μm^2 for a traditional CMOS Sklansky adder. However, for equivalent delay, parallelism is required that would make the area penalty approximately 100x. Spencer notes that a more optimized device layout could reduce this penalty.

Chen et al. used cantilever relays to achieve similar simulated throughputs with only a 6 - 25x area overhead [169]. Lee et al. calculated the scaling limits of these cantilever beams and found poly-Si would be difficult to scale to beam lengths below 80 nm [170]. TiNi is more elastic and able to scale to ~30 nm. Lee notes that vertical structures may be advantageous for area efficiency.

Several papers, including by Akarvardar et al., on scaling and materials for higher mechanical switching speeds have shown that materials already in use in the semiconductor industry, Silicon and Germanium, offer higher performance (quality factor, beam velocity, switching current, etc.) than other materials such as Gold, Copper, and other metals [171].

Conclusions

Relays offer extremely high I_{ON}/I_{OFF} ratios with excellent subthreshold slopes. However, scalability and reliability need to be proven before these devices can find acceptance. Fabrication costs may be improved, as many steps (such as ion implantation and epitaxy) would not be necessary, although release etch processing may add cost and complexity. With current state-of-the-art relay technology, these devices could still find usage for non-volatile memory, or FPGA applications.

2.4.2 Non-Traditional Devices

There are other non-traditional techniques that can be explored to improve transistor density. For example, recrystallization can be used to grow crystalline or polycrystalline semiconductor material for devices above traditional MOSFETs. Alternatively, wafer bonding allows stacking of devices for increased areal density.

2.4.2.1 Thin Film Devices

One example of a non-traditional structure using recrystallization is to fabricate

transistors in the back-end on top of the normal device layers. The value of a these Thin-Film Transistors (TFTs) is not only to place additional devices in the back-end, but also to lower parasitics with layouts that decrease RC delay in interconnects.

Varadarajan et al. [172] used metal induced crystallization to form crystalline semiconductors termed WireFETs in the back-end. Unfortunately, the metal used (Aluminum on Silicon) doped the semiconductor, making it impossible to turn off the device. Other materials, or structures similar to a Junctionless Accumulation Mode (JAM) device, could make this technique more competitive. However, care would need to be taken to make sure that the active region is thin enough to be fully depleted, while the contacts are wide enough to prevent parasitic resistance issues, similar to problems with JAM devices.

Another method to induce crystallization is the use of poly-Germanium seeds. Subramanian and Saraswat used this method in 1997 to create laterally crystallized TFTs [173]. In 1999, Subramanian et al. demonstrated TFTs scaled down to 100nm, which were single grain and showed very low leakage (below 1 pA/μm) [174]. Mobility and on-state current of these devices remains lower than traditional MOSFET, but the ease of fabrication could allow for inexpensive additional layers of devices.

Metal-Induced Crystallization through a Cap layer (MICC) is another method of forming polycrystalline Silicon. Oh et al. demonstrated this technique using Nickel mediated crystallization, although currents were below 100 μA/μm at high voltages [175].

Work has also been done in the memory space due to the challenges of scaling memory. As an example, Jung et al. have shown that laser-induced epitaxy can be used to form high density 3D crystalline Silicon SRAM [176]. Further evaluation of techniques used for novel SRAM and other memory devices may prove useful for logic applications.

2.4.2.2 Stacking for Density

Wafer bonding is the process of bonding two substrates together to improve areal efficiency. Wafer bonding (without through silicon vias connecting individual devices) also has the benefit of allowing different materials of crystalline substrates to be used. For example, wafer bonding is a potential solution for the problem of lack of a good p-type device, such as with most HEMTs. With excellent alignment, wafer bonding strategies can also be used to enhance performance in carefully designed circuits, such as SRAM.

There has been significant work with wafer bonding in the memory space. Devices on with different orientations can be bonded together [177]. In addition to the density benefits, Batude et al. has also shown that these substrates can have positive interactions, for example by using the bottom device's gate to shift the V_T of the top device [177]. These structures can provide both SRAM stabilization and area efficiency.

Nho et al. designed a 3D SRAM architecture that reduces bitline capacitance, improving performance, although $V_{CC,MIN}$ was not evaluated [178]. Hsu and Wu showed a similar 3D design that reduces both latency and energy consumption [179].

For the highest performance, pre-fabrication wafer bonding may allow different devices to be tuned for maximum performance, which was shown by Yokoyama for InGaAs III-V on Silicon [180].

2.5 Evaluation Metrics

To benchmark these devices for the terascale regime, a set of metrics was developed to comprehensively examine their potential. It is difficult to develop benchmark metrics for devices that operate on different principles and in different regimes. Furthermore, many of the potential devices discussed in this work are at the early stages of their development. There are often large conflicts between simulation results and early experimental results as models are refined and new physical effects are discovered [181]. Frequently, early simulation models can neglect effects that limit long-term attainable performance [182].

Six metrics were used to compare these devices. These are I_{ON}, I_{OFF}, switching energy, fabrication complexity, fabrication cost, and density scaling. Four of the metrics are device level; the remaining two are circuit level. For device types with significant differences between simulation and experimental results, both sets of results were sometimes used. Since best in class performance can change over time, the published work with the most well-behaved totality of current devices is used. In some cases, the resulting scores were modified using guidance from industrial and academic experts, so all numbers besides I_{ON} and I_{OFF} have a partially qualitative nature.

I_{OFF} and I_{ON} are used as initial metrics for evaluation. These values are easily found using published data and form a concrete basis for the other metrics. Note that some works normalize current differently, especially for non-planar devices. I_{OFF} and I_{ON} for all of the devices shown have been recalculated using gate perimeter normalization. For stacked devices such as nanowires, a single gate perimeter is used to allow for the advantageous drawn pitch of these devices. To determine the on current value, I_{DSAT} is used, with V_{GS}-V_T = V_{DS} = V_{DD}. Supply voltage is determined by either the published work's typical voltage or the typical voltage value used for that class of devices. The off current value is determined by the device's minimum quoted I_{OFF} if possible.

Energy in Joules per switching event is used as the next set of metrics of evaluation. This was determined by inputting each device's current characteristics into a Verilog-A lookup table, which was then used in a set of SPICE simulations. For capacitance, experimental data was used when possible and calculated when not possible. Some device types are n-type only. In these cases, a standard P-type Germanium MOSFET was used. The energy dissipated was measured and averaged

during a high to low switching event and a low to high switching event.

Fabrication complexity was the next metric, as manufacturability is also a concern for future devices. A detailed comparison of major processing steps was completed for each device. Steps needed by all devices, such as isolation, were not included.

Fabrication cost is a related metric. Moore's law can also be a considered an economic motivation, as devices have not only increased performance at smaller sizes, but also cost less per device to manufacture. Each fabrication step was evaluated and assigned a value, depending on projected manufacturing cost. These values were totaled and used as the normalized fabrication cost per wafer. For devices that require a p-type Germanium MOSFET for p-type operation, these costs were also included.

Finally, the last metric is area scaling. To determine area scaling, design rules scaled to 2018 ITRS projections were assumed (design rules were taken from [183]). These design rules were then used for each device to determine both the size of a standard 6T SRAM cell (Figure 2.3) as well as a 2-input NAND gate. These areas were added to form the area metric, as both memory and logic elements will be necessary in a circuit design.

2.6 Benchmarking Results

Benchmarking figures and tables (Figures 2.1-2.3, Table 2.1) illustrate the various tradeoffs associated with each technology using the metrics defined in section 2.5.

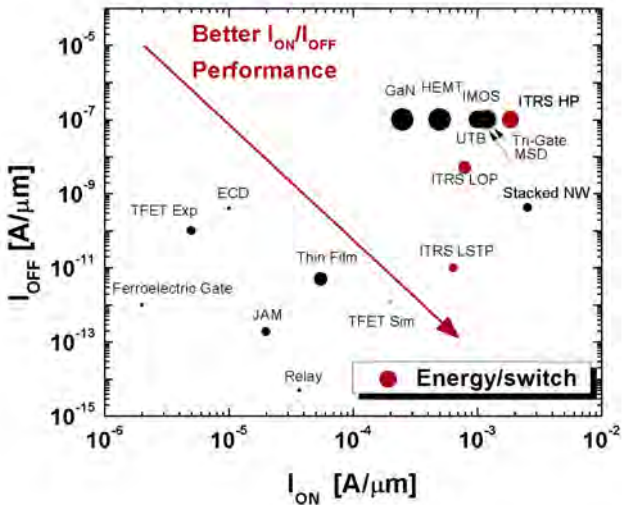


Figure 2.1: Benchmarking normalized current and energy. Devices in the bottom right have excellent I_{ON}/I_{OFF} ratios. Devices with small bubbles have lower energy per switching event. Red dots indicate ITRS targets. V_{SUPPLY} is 0.6 V when possible.

	Polarity	I_{OFF} [nA/μm]	I_{ON} [mA/μm]	Reference
ITRS 2018 HP	N/P	100	1.85	[3]
ITRS 2018 LOP	N/P	5	0.792	[3]
ITRS 2018 LSTP	N/P	0.01	0.643	[3]
HEMT	N/Ge-P*	100	0.5	[184]
GaN	N/Ge-P*	100	0.25	[37]
Fe-Gate	N/P	0.001	0.002	[42]
ECD	N/P	0.4	0.01	[185]
IMOS	N/P	100	1	[63]
TFET Simulation	N/P**	0.001	0.2	[85]
TFET Experiment	N/P**	0.1	0.005	[186]
UTB	N/P	100	1	[187]
Tri-Gate	N/P	100	1.2	[109]
Stacked NW	N/P	0.42	2.546	[115]
DSS	N/P	100	1.2	[135]
JAM	N/P	0.0002	0.02	[142]
Relay	N/P	0.00001	0.037	[161]
Thin Film	N/P	0.01	0.055	[174]

* A p-type Germanium FET was used in energy simulations.
** P-type TFETs have been fabricated, but require a different structure than n-type TFETs.
Table 2.1: Normalized I_{OFF} and I_{ON} with references for each class of device.

Looking first at the I_{ON}/I_{OFF} current ratios in Figure 2.1, high I_{ON} is necessary not only for intrinsic delay, but also to drive metal lines in integrated circuits. Low I_{OFF} and large I_{ON}/I_{OFF} ratios are needed for low power operation. HEMT and GaN devices both have some difficulty matching modern Si drive currents at low voltages. Ferroelectric gate devices also have low current, although they can have very low I_{OFF}. ECD devices with the highest I_{ON}/I_{OFF} ratios generally have significantly lower drive current. IMOS has been fabricated with higher drive current, close to modern devices. Simulated TFETs come close to ITRS Low Standby Power specifications, but still need more drive current. Most UTB devices have relatively close I_{ON} to the planar specifications in ITRS. Compared to the ITRS High Performance specifications, Tri-gate and Stacked nanowires come close to meeting the I_{ON}/I_{OFF} target. DSS MOSFETs also come close with optimistic Schottky Barrier Heights. JAM devices have had limited drive current demonstrated thus far. Relays have very low off current, but currently on a per μm scale have low drive current. Finally, thin film devices also have poor drive current.

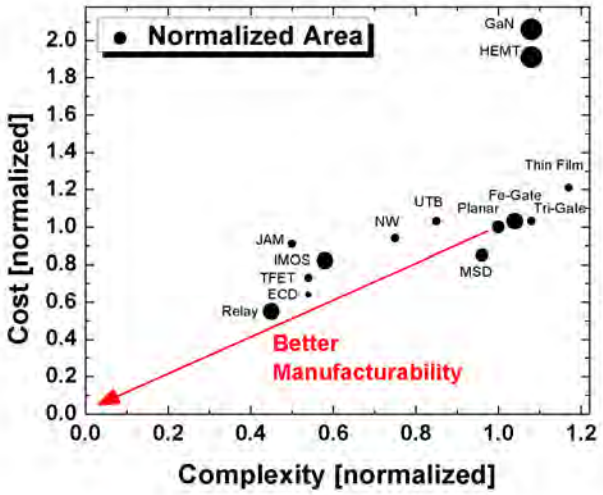
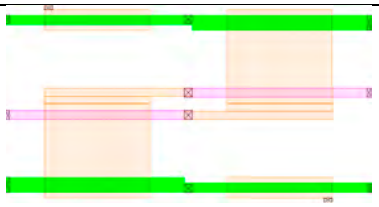
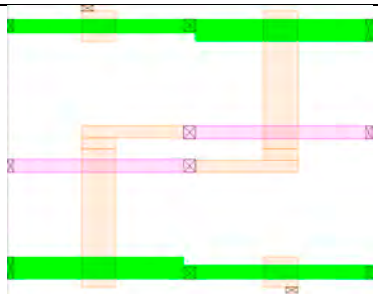


Figure 2.2: Benchmarking normalized fabrication cost, complexity, and cell area. Devices that require less manufacturing complexity and cost are located in the bottom left corner, with larger bubbles indicating greater normalized area.

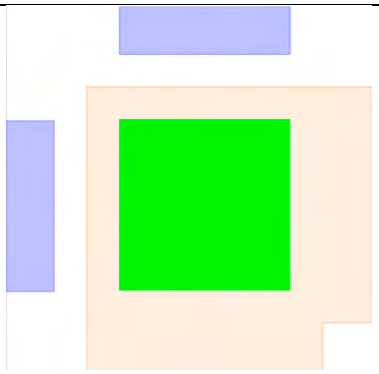
Considering the complexity metrics in Figure 2.2, ideally the complexity and cost of devices would be equivalent or lower than current planar technology. As illustrated in Figure 2.2, many of the technologies that are less optimized such as IMOS, TFET, ECD, and JAM, show reduced complexity and costs compared to planar, largely because they do not incorporate technology enhancements (such as strain). Relays and nanowires could likely be less expensive, as they may not require expensive substrates. Modifications to standard planar technology, such as MSD, Fe-Gate, and Thin Film, bring costs and complexity similar to planar devices. Tri-Gate and Ultra Thin Body are slightly more costly than planar. In the case of Tri-gate this is due to more complex fabrication for the non-planar process, while in the case of Ultra Thin Body there is less complexity but higher cost for the FDSOI wafers. HEMTs and GaN devices, which use expensive III-V wafers or III-V epitaxy, are significantly more expensive than planar devices.



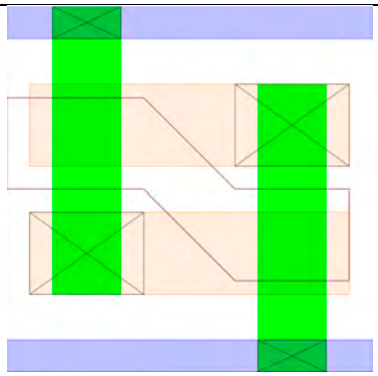
HEMT SRAM layout. The large gate lengths and source/drain to gate distances make scaling difficult.



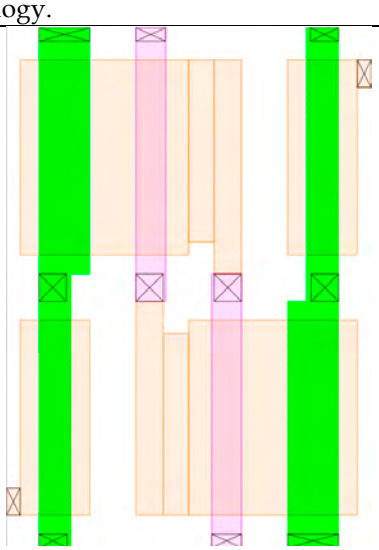
GaN SRAM layout. Similarly to the HEMT, the source/drain to gate distances make scaling this SRAM cell challenging.



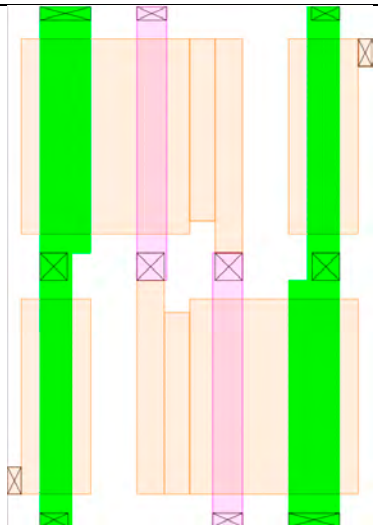
Electrochemical devices can incorporate a memory element in one device. Note that these memory elements in present research may be much slower than current SRAM technology.



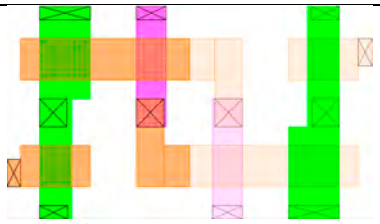
Thin film devices can incorporate several layers of devices in a smaller layout area.



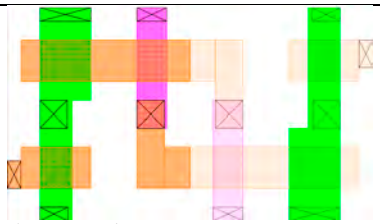
Impact Ionization FET SRAM layout. The 50 nm gate length limitation and gate/drain underlap impact the scalability of this device.



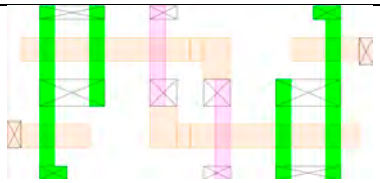
Ferroelectric-Gate FET SRAM layout. The gate length limitation to keep a low subthreshold swing impact the SRAM cell's ultimate scalability.



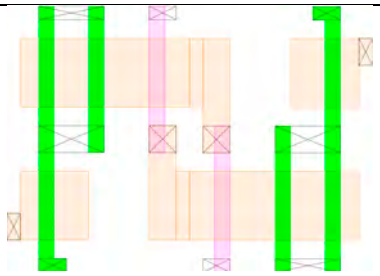
Traditional Planar MOSFET SRAM layout. Gate length scaling is limited to 14 nm, limiting how small the SRAM cell area can be made.



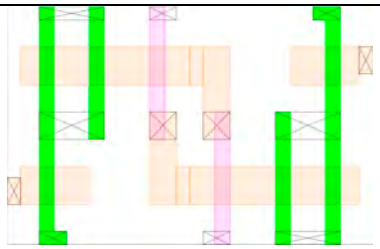
Ultra Thin Body MOSFET SRAM layout. Similarly to the Planar device, UTB gate lengths are limited to greater than 10.6 nm.



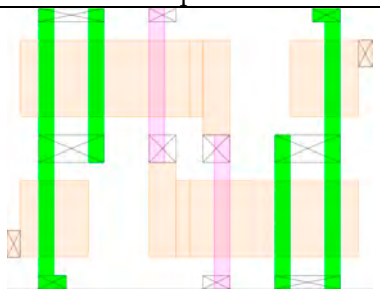
Tri-gate MOSFET has superior scalability in SRAM, although fin quantization and the associated fin pitch spacing should be considered.



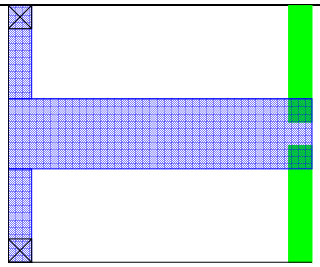
Stacked Nanowires have high drive current, but experience width quantization.



Metal Source Drain devices require double gate to scale well, leading to a similar fin quantization scaling issue.

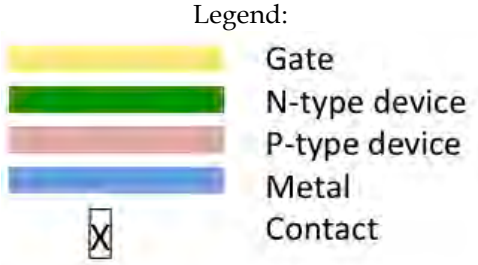


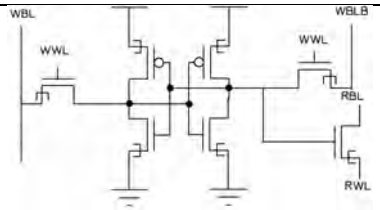
Junctionless Accumulation Mode devices may also experience quantization issues.



Relays can implement a memory element in one device.

Tunnel FETs may require an additional transistor to facilitate SRAM operation, as in the 7T design.





Tunnel FET 7T SRAM cell schematic separates read and write functions.

Figure 2.3: SRAM cell layouts for the 14 different device technologies benchmarked. Cell layouts are not comparable to scale (*i.e.* individual width to length ratios are accurate, but each device is scaled to fit in the same space). Design rules were taken from [183], scaled from the 22 nm node.

Cell area is also illustrated in Figure 2.3. Several devices are projected to scale more poorly due to larger spacers, such as in IMOS, HEMT, and GaN devices. A few devices simply are not projected to be able to continue scaling, such as IMOS, Ferroelectric gate, UTB, and planar CMOS. Relays show respectable scaling down to the 65 nm node, but contact pitch between devices may still be larger, since the minimum feature size and material properties limit how small serpentine springs or cantilever beams can be produced. Devices using multiple gates, such as Tri-gate, must account for fin pitch when designing cell layouts.

2.7 Conclusions

There are many options to continue CMOS scaling. Each has its own tradeoffs. Many are still in the research stage, but hold promise if solutions can be found to some of their drawbacks. In all of these devices, it is important to keep in mind that research should not aim for a fixed target; the point where a technology will replace standard CMOS will be a future technology node, and must be a viable solution for at least one technology node past that to justify the switching costs. Further fundamental (mostly in the case of materials) and engineering research must continue to sustain the inexorable scaling of switching devices.

2.8 References

1. Moore, G. E. "Cramming More Components Onto Integrated Circuits." *Electronics*, 114-117 (1965).
2. Shahidi, G. G. "Device scaling for 15 nm node and beyond." *Device Research Conference, 2009. DRC 2009*, 247 -250 (2009).
3. International Technology Roadmap for Semiconductors (ITRS) 2011. http://public.itrs.net/.
4. Mimura, T. "The early history of the high electron mobility transistor (HEMT)." *Microwave Theory and Techniques, IEEE Transactions on*, **50** 780 -782 (2002).
5. "History of HEMTs," Available: http://www.iue.tuwien.ac.at/phd/vitanov/node11.html.
6. Dingle, R., Störmer, H. L., Gossard, A. C., & Wiegmann, W. "Electron mobilities in modulation-doped semiconductor heterojunction superlattices." *Applied Physics Letters*, **33** http://link.aip.org/link/?APL/33/665/1 665-667 (1978).
7. Delagebeaudeuf, D., Delescluse, P., Etienne, P., Laviron, M., Chaplart, J., & Linh, N. T. "Two-dimensional electron gas m.e.s.f.e.t. structure." *Electronics Letters*, **16** 667 - 668 (1980).

8. Delagebeaudeuf, D. & Linh, N. T. "Charge control of the heterojunction two-dimensional electron gas for MESFET application." *Electron Devices, IEEE Transactions on*, **28** 790 - 795 (1981).
9. "High Electron Mobility Transistors (HEMTs)," Available: http://www.mwe.ee.ethz.ch/en/about-mwe-group/research/vision-and-aim/high-electron-mobility-transistors-hemt.html.
10. Ok, I., Veksler, D., Hung, P. Y., Oh, J., Moore, R. L., McDonough, C., Geer, R. E., Gaspe, C. K., Santos, M. B., Wong, G., Kirsch, P., Tseng, H. H., Bersuker, G., Hobbs, C., & Jammy, R. "Reducing Rext in laser annealed enhancement-mode $In_{0.53}Ga_{0.47}As$ surface channel n-MOSFET." *VLSI Technology Systems and Applications (VLSI-TSA), 2010 International Symposium on*, 38 -39 (2010).
11. Dewey, G., Kotlyar, R., Pillarisetty, R., Radosavljevic, M., Rakshit, T., Then, H., & Chau, R. "Logic performance evaluation and transport physics of Schottky-gate III-V compound semiconductor quantum well field effect transistors for power supply voltages (VCC) ranging from 0.5v to 1.0v." *Electron Devices Meeting (IEDM), 2009 IEEE International*, 1 -4 (2009).
12. Radosavljevic, M., Chu-Kung, B., Corcoran, S., Dewey, G., Hudait, M. K., Fastenau, J. M., Kavalieros, J., Liu, W. K., Lubyshev, D., Metz, M., Millard, K., Mukherjee, N., Rachmady, W., Shah, U., & Chau, R. "Advanced high-K gate dielectric for high-performance short-channel In0.7Ga0.3As quantum well field effect transistors on silicon substrate for low power logic applications." *Electron Devices Meeting (IEDM), 2009 IEEE International*, 1 -4 (2009).
13. Kim, T.-W., Kim, D.-H., & del Alamo, J. A. "30 nm In0.7Ga0.3As Inverted-Type HEMTs with reduced gate leakage current for logic applications." *Electron Devices Meeting (IEDM), 2009 IEEE International*, 1 -4 (2009).
14. Kim, D.-H. & del Alamo, J. A. "Scalability of Sub-100 nm InAs HEMTs on InP Substrate for Future Logic Applications." *Electron Devices, IEEE Transactions on*, **57** 1504 -1511 (2010).
15. Chung, J. W., Lee, J.-k., Piner, E. L., & Palacios, T. "Seamless On-Wafer Integration of Si(100) MOSFETs and GaN HEMTs." *Electron Device Letters, IEEE*, **30** 1015 -1017 (2009).
16. Waldron, N., Kim, D.-H., & del Alamo, J. A. "A Self-Aligned InGaAs HEMT Architecture for Logic Applications." *Electron Devices, IEEE Transactions on*, **57** 297 -304 (2010).
17. Kharche, N., Klimeck, G., Kim, D.-H., del Alamo, J. A., & Luisier, M. "Performance analysis of ultra-scaled InAs HEMTs." *Electron Devices Meeting (IEDM), 2009 IEEE International*, 1 -4 (2009).
18. Oh, S. & Wong, H.-S. P. "Effect of Parasitic Resistance and Capacitance on Performance of InGaAs HEMT Digital Logic Circuits." *Electron Devices, IEEE Transactions on*, **56** 1161 -1164 (2009).
19. Skotnicki, T. & Boeuf, F. "How can high mobility channel materials boost or degrade performance in advanced CMOS." *VLSI Technology (VLSIT), 2010 Symposium on*, 153 -154 (2010).
20. del Alamo, J. A. "Nanometre-scale electronics with III-V compound

semiconductors." *Nature*, **479** http://dx.doi.org/10.1038/nature10677 317--323 (2011).
21. Shur, M. S. "GaN-based devices." *Electron Devices, 2005 Spanish Conference on*, 15 - 18 (2005).
22. Maruska, H. P. & Tietjen, J. J. "The Preparation and Properties of Vapor-Deposited Single Crystalline GaN." *Applied Physics Letters*, **15** 327-329 (1969).
23. Asif Khan, M., Kuznia, J. N., Bhattarai, A. R., & Olson, D. T. "Metal semiconductor field effect transistor based on single crystal GaN." *Applied Physics Letters*, **62** 1786 -1787 (1993).
24. Asif Khan, M., Bhattarai, A., Kuznia, J. N., & Olson, D. T. "High electron mobility transistor based on a GaN /$Al_xGa_{1-x}N$ heterojunction." *Applied Physics Letters*, **63** 1214 - 1215 (1993).
25. Goldberger, J., He, R., Zhang, Y., Lee, S., Yan, H., Choi, H.-J., & Yang, P. "Single-crystal gallium nitride nanotubes." *Nature*, **422** http://dx.doi.org/10.1038/nature01551 599--602 (2003).
26. Hu, J., Bando, Y., Golberg, D., & Liu, Q. "Gallium Nitride Nanotubes by the Conversion of Gallium Oxide Nanotubes." *Angewandte Chemie International Edition*, **42** http://dx.doi.org/10.1002/anie.200351001 3493--3497 (2003).
27. Eastman, L. F. & Mishra, U. K. "The toughest transistor yet [GaN transistors]." *Spectrum, IEEE*, **39** 28 -33 (2002).
28. Nidhi, Dasgupta, S., Brown, D. F., Keller, S., Speck, J. S., & Mishra, U. K. "N-polar GaN-based highly scaled self-aligned MIS-HEMTs with state-of-the-art $f_T.L_G$ product of 16.8 GHz-μm." *Electron Devices Meeting (IEDM), 2009 IEEE International*, 1 - 3 (2009).
29. Xin, X., Shi, J., Liu, L., Edwards, J., Swaminathan, K., Pabisz, M., Murphy, M., Eastman, L. F., & Pophristic, M. "Demonstration of Low-Leakage-Current Low-On-Resistance 600-V 5.5-A GaN/AlGaN HEMT." *Electron Device Letters, IEEE*, **30** 1027 - 1029 (2009).
30. Chang, Y. C., Chang, W. H., Chiu, H. C., Chang, Y. H., Tung, L. T., Lee, C. H., Hong, M., Kwo, J., Hong, J. M., & Tsai, C. C. "Inversion-channel GaN MOSFET using atomic-layer-deposited Al_2O_3 as gate dielectric." *VLSI Technology, Systems, and Applications, 2009. VLSI-TSA '09. International Symposium on*, 131 -132 (2009).
31. Cai, Y., Cheng, Z., Tang, W. C. W., Chen, K. J., & Lau, K. M. "Monolithic integration of enhancement-and depletion-mode AlGaN/GaN HEMTs for GaN digital integrated circuits." *Electron Devices Meeting, 2005. IEDM Technical Digest. IEEE International*, 4 774 (2005).
32. Ota, K., Endo, K., Okamoto, Y., Ando, Y., Miyamoto, H., & Shimawaki, H. "A normally-off GaN FET with high threshold voltage uniformity using a novel piezo neutralization technique." *Electron Devices Meeting (IEDM), 2009 IEEE International*, 1 - 4 (2009).
33. Derluyn, J., Van Hove, M., Visalli, D., Lorenz, A., Marcon, D., Srivastava, P., Geens, K., Sijmus, B., Viaene, J., Kang, X., Das, J., Medjdoub, F., Cheng, K., Degroote, S., Leys, M., Borghs, G., & Germain, M. "Low leakage high breakdown e-mode GaN DHFET on Si by selective removal of in-situ grown Si_3N_4." *Electron Devices Meeting (IEDM), 2009 IEEE International*, 1 -4 (2009).

34. Kanamura, M., Ohki, T., Kikkawa, T., Imanishi, K., Imada, T., Yamada, A., & Hara, N. "Enhancement-Mode GaN MIS-HEMTs With n-GaN/i-AlN/n-GaN Triple Cap Layer and High- Gate Dielectrics." *Electron Device Letters, IEEE*, **31** 189 -191 (2010).
35. Im, K.-S., Ha, J.-B., Kim, K.-W., Lee, J.-S., Kim, D.-S., Hahm, S.-H., & Lee, J.-H. "Normally Off GaN MOSFET Based on AlGaN/GaN Heterostructure With Extremely High 2DEG Density Grown on Silicon Substrate." *Electron Device Letters, IEEE*, **31** 192 - 194 (2010).
36. Joh, J. & del Alamo, J. A. "Mechanisms for Electrical Degradation of GaN High-Electron Mobility Transistors." *Electron Devices Meeting, 2006. IEDM '06. International*, 1 -4 (2006).
37. Shinohara, K., Regan, D., Corrion, A., Brown, D., Burnham, S., Willadsen, P. J., Alvarado-Rodriguez, I., Cunningham, M., Butler, C., Schmitz, A., Kim, S., Holden, B., Chang, D., Lee, V., Ohoka, A., Asbeck, P. M., & Micovic, M. "Deeply-scaled self-aligned-gate GaN DH-HEMTs with ultrahigh cutoff frequency." *Electron Devices Meeting (IEDM), 2011 IEEE International*, 19.1.1 -19.1.4 (2011).
38. Uren, M. J., Nash, K. J., Balmer, R. S., Martin, T., Morvan, E., Caillas, N., Delage, S. L., Ducatteau, D., Grimbert, B., & De Jaeger, J. C. "Punch-through in short-channel AlGaN/GaN HFETs." *Electron Devices, IEEE Transactions on*, **53** 395 - 398 (2006).
39. Park, P. S. & Rajan, S. "Simulation of Short-Channel Effects in N- and Ga-Polar AlGaN/GaN HEMTs." *Electron Devices, IEEE Transactions on*, **58** 704 -708 (2011).
40. Salahuddin, S. & Datta, S. "Use of Negative Capacitance to Provide Voltage Amplification for Low Power Nanoscale Devices." *Nano Letters*, **8** http://pubs.acs.org/doi/abs/10.1021/nl071804g 405-410 (2008).
41. Salahuddin, S. & Datta, S. "Can the subthreshold swing in a classical FET be lowered below 60 mV/decade?." *Electron Devices Meeting, 2008. IEDM 2008. IEEE International*, 1 -4 (2008).
42. Salvatore, G. A., Bouvet, D., & Ionescu, A. M. "Demonstration of subthrehold swing smaller than 60mV/decade in Fe-FET with P(VDF-TrFE)/SiO2 gate stack." *Electron Devices Meeting, 2008. IEDM 2008. IEEE International*, 1 -4 (2008).
43. Rusu, A., Salvatore, G. A., Jimenez, D., & Ionescu, A. M. "Metal-Ferroelectric-Metal-Oxide-semiconductor field effect transistor with sub-60mV/decade subthreshold swing and internal voltage amplification." *Electron Devices Meeting (IEDM), 2010 IEEE International*, 16.3.1 -16.3.4 (2010).
44. Tanakamaru, S., Hatanaka, T., Yajima, R., Takahashi, M., Sakai, S., & Takeuchi, K. "A 0.5V operation, 32% lower active power, 42% lower leakage current, ferroelectric 6T-SRAM with V_{TH} self-adjusting function for 60% larger Static Noise Margin." *Electron Devices Meeting (IEDM), 2009 IEEE International*, 1 -4 (2009).
45. Khan, A. I., Bhowmik, D., Yu, P., Kim, S. J., Pan, X., Ramesh, R., & Salahuddin, S. "Experimental Evidence of Ferroelectric Negative Capacitance in Nanoscale Heterostructures." *Applied Physics* Letters, **99** 113501 (2011).
46. Krowne, C. M., Kirchoefer, S. W., Chang, W., Pond, J. M., & Alldredge, L. M. B. "Examination of the Possibility of Negative Capacitance Using Ferroelectric Materials in Solid State Electronic Devices." *Nano Letters*, **11** http://pubs.acs.org/doi/abs/10.1021/nl1037215 988-992 (2011).

47. Jin, R., Song, Y., Ji, M., Xu, H., Kang, J., Han, R., & Liu, X. "Characteristics of sub-100nm ferroelectric field effect transistor with high-k buffer layer." *Solid-State and Integrated-Circuit Technology, 2008. ICSICT 2008. 9th International Conference on*, 888 -891 (2008).
48. Aviram, A. & Ratner, M. A. "Molecular rectifiers." *Chemical Physics Letters*, **29** http://www.sciencedirect.com/science/article/pii/0009261474850311 277 - 283 (1974).
49. Collier, C. P., Wong, E. W., Belohradský, M., Raymo, F. M., Stoddart, J. F., Kuekes, P. J., Williams, R. S., & Heath, J. R. "Electronically Configurable Molecular-Based Logic Gates." *Science*, **285** http://www.sciencemag.org/content/285/5426/391.abstract 391-394 (1999).
50. Terabe, K., Hasegawa, T., Nakayama, T., & Aono, M. "Quantized conductance atomic switch." *Nature*, **433** http://dx.doi.org/10.1038/nature03190 47--50 (2005).
51. Thomson, A. F., Melville, D. O. S., & Blaikie, R. J. "Nanometre-scale electrochemical switches fabricated using a plasma-based sulphidation technique." *Nanoscience and Nanotechnology, 2006. ICONN '06. International Conference on*, (2006).
52. Sakamoto, T., Banno, N., Iguchi, N., Kawaura, H., Sunamura, H., Fujieda, S., Terabe, K., Hasegawa, T., & Aono, M. "A Ta_2O_5 solid-electrolyte switch with improved reliability." *VLSI Technology, 2007 IEEE Symposium on*, 38 -39 (2007).
53. Sakamoto, T., Banno, N., Iguchi, N., Kawaura, H., Kaeriyama, S., Mizuno, M., Terabe, K., Hasegawa, T., & Aono, M. "Three terminal solid-electrolyte nanometer switch." *Electron Devices Meeting, 2005. IEDM Technical Digest. IEEE International*, 475 - 478 (2005).
54. Sakamoto, T., Kaeriyama, S., Sunamura, H., Mizuno, M., Kawaura, H., Hasegawa, T., Terabe, K., Nakayama, T., & Aono, M. "A nonvolatile programmable solid electrolyte nanometer switch." *Solid-State Circuits Conference, 2004. Digest of Technical Papers. ISSCC. 2004 IEEE International*, 290 - 529 Vol.1 (2004).
55. Liang, C., Terabe, K., Hasegawa, T., Negishi, R., Tamura, T., & Aono, M. "Ionic--Electronic Conductor Nanostructures: Template-Confined Growth and Nonlinear Electrical Transport." *Small*, **1** http://dx.doi.org/10.1002/smll.200500155 971--975 (2005).
56. Kozicki, M. N., Gopalan, C., Balakrishnan, M., & Mitkova, M. "A Low-Power Nonvolatile Switching Element Based on Copper-Tungsten Oxide Solid Electrolyte." *Nanotechnology, IEEE Transactions on*, **5** 535 -544 (2006).
57. Kim, S.-W. & Nishi, Y. "Copper Sulfide-based Resistance Change Memory." *Non-Volatile Memory Technology Symposium, 2007. NVMTS '07*, 76 -78 (2007).
58. Hasegawa, T., Terabe, K., Sakamoto, T., & Aono, M. "Nanoionics Switching Devices: Atomic Switches." *MRS Bulletin*, **34** 929-934 (2009).
59. Gopalakrishnan, K., Griffin, P. B., & Plummer, J. D. "I-MOS: a novel semiconductor device with a subthreshold slope lower than kT/q." *Electron Devices Meeting, 2002. IEDM '02. Digest. International*, 289 - 292 (2002).
60. Nematian, H., Fathipour, M., & Nayeri, M. "A Novel impact Ionization MOS (I-MOS) structure using a silicon-germanium/silicon heterostructure channel." *Microelectronics, 2008. ICM 2008. International Conference on*, 228 -231 (2008).
61. Abelein, U., Born, M., Bhuwalka, K. K., Schindler, M., Schlosser, M., Sulima, T., &

Eisele, I. "Improved Reliability by Reduction of Hot-Electron Damage in the Vertical Impact-Ionization MOSFET (I-MOS)." *Electron Device Letters, IEEE*, **28** 65 -67 (2007).
62. Tura, A. & Woo, J. "Performance Comparison of Silicon Steep Subthreshold FETs." *Electron Devices, IEEE Transactions on*, **57** 1362 -1368 (2010).
63. Savio, A., Monfray, S., Charbuillet, C., & Skotnicki, T. "On the Limitations of Silicon for I-MOS Integration." *Electron Devices, IEEE Transactions on*, **56** 1110 -1117 (2009).
64. Shen, C., Lin, J.-Q., Toh, E.-H., Chang, K.-F., Bai, P., Heng, C.-H., Samudra, G. S., & Yeo, Y.-C. "On the Performance Limit of Impact-Ionization Transistors." *Electron Devices Meeting, 2007. IEDM 2007. IEEE International*, 117 -120 (2007).
65. Banerjee, S., Coleman, J., Richardson, B., & Shah, A. "A Band-to-Band Tunneling Effect in the Trench Transistor Cell." *VLSI Technology, 1987. Digest of Technical Papers. Symposium on*, 97 -98 (1987).
66. Banerjee, S., Richardson, W., Coleman, J., & Chatterjee, A. "A new three-terminal tunnel device." *Electron Device Letters, IEEE*, **8** 347 - 349 (1987).
67. Baba, T. "Proposal for Surface Tunnel Transistors." *Japanese Journal of Applied Physics*, **31** http://jjap.jsap.jp/link?JJAP/31/L455/ L455-L457 (1992).
68. Reddick, W. M. & Amaratunga, G. A. J. "Silicon surface tunnel transistor." *Applied Physics Letters*, **67** http://link.aip.org/link/?APL/67/494/1 494-496 (1995).
69. Appenzeller, J., Lin, Y.-M., Knoch, J., & Avouris, P. "Band-to-Band Tunneling in Carbon Nanotube Field-Effect Transistors." *Phys. Rev. Lett.*, **93** 196805 (2004).
70. Nayfeh, O. M., Chleirigh, C. N., Hennessy, J., Gomez, L., Hoyt, J. L., & Antoniadis, D. A. "Design of Tunneling Field-Effect Transistors Using Strained-Silicon/Strained-Germanium Type-II Staggered Heterojunctions." *Electron Device Letters, IEEE*, **29** 1074 -1077 (2008).
71. Vandenberghe, W., Verhulst, A. S., Groeseneken, G., Soree, B., & Magnus, W. "Analytical model for point and line tunneling in a tunnel field-effect transistor." *Simulation of Semiconductor Processes and Devices, 2008. SISPAD 2008. International Conference on*, 137 -140 (2008).
72. Lu, Y., Bangsaruntip, S., Wang, X., Zhang, L., Nishi, Y., & Dai, H. "DNA Functionalization of Carbon Nanotubes for Ultrathin Atomic Layer Deposition of High κ Dielectrics for Nanotube Transistors with 60 mV/Decade Switching." *Journal of the American Chemical Society*, **128** http://pubs.acs.org/doi/abs/10.1021/ja058836v 3518-3519 (2006).
73. Choi, W. Y., Park, B.-G., Lee, J. D., & Liu, T.-J. K. "Tunneling Field-Effect Transistors (TFETs) With Subthreshold Swing (SS) Less Than 60 mV/dec." *Electron Device Letters, IEEE*, **28** 743 -745 (2007).
74. Mayer, F., Le Royer, C., Damlencourt, J.-F., Romanjek, K., Andrieu, F., Tabone, C., Previtali, B., & Deleonibus, S. "Impact of SOI, $Si_{1-x}Ge_xOI$ and GeOI substrates on CMOS compatible Tunnel FET performance." *Electron Devices Meeting, 2008. IEDM 2008. IEEE International*, 1 -5 (2008).
75. Jeon, K., Loh, W.-Y., Patel, P., Kang, C. Y., Oh, J., Bowonder, A., Park, C., Park, C. S., Smith, C., Majhi, P., Tseng, H.-H., Jammy, R., Liu, T.-J. K., & Hu, C. "Si tunnel transistors with a novel silicided source and 46mV/dec swing." *VLSI Technology*

(VLSIT), 2010 Symposium on, 121 -122 (2010).
76. Leonelli, D., Vandooren, A., Rooyackers, R., Verhulst, A. S., Gendt, S. D., Heyns, M. M., & Groeseneken, G. "Performance Enhancement in Multi Gate Tunneling Field Effect Transistors by Scaling the Fin-Width." *Japanese Journal of Applied Physics*, **49** http://jjap.jsap.jp/link?JJAP/49/04DC10/ 04DC10 (2010).
77. Kim, S. H., Kam, H., Hu, C., & Liu, T.-J. K. "Germanium-source tunnel field effect transistors with record high ION/IOFF." *VLSI Technology, 2009 Symposium on*, 178 -179 (2009).
78. Mookerjea, S., Mohata, D., Mayer, T., Narayanan, V., & Datta, S. "Temperature-Dependent I-V Characteristics of a Vertical $In_{0.53}Ga_{0.47}As$ Tunnel FET." *Electron Device Letters, IEEE*, **31** 564 -566 (2010).
79. Mookerjea, S., Krishnan, R., Datta, S., & Narayanan, V. "On Enhanced Miller Capacitance Effect in Interband Tunnel Transistors." *Electron Device Letters, IEEE*, **30** 1102 -1104 (2009).
80. Mookerjea, S., Krishnan, R., Datta, S., & Narayanan, V. "Effective Capacitance and Drive Current for Tunnel FET (TFET) CV/I Estimation." *Electron Devices, IEEE Transactions on*, **56** 2092 -2098 (2009).
81. Wan, J., Le Royer, C., Zaslavsky, A., & Cristoloveanu, S. "SOI TFETs: Suppression of ambipolar leakage and low-frequency noise behavior." *Solid-State Device Research Conference (ESSDERC), 2010 Proceedings of the European*, 341 -344 (2010).
82. Krishnamohan, T., Kim, D., Raghunathan, S., & Saraswat, K. "Double-Gate Strained-Ge Heterostructure Tunneling FET (TFET) With record high drive currents and <60mV/dec subthreshold slope." *Electron Devices Meeting, 2008. IEDM 2008. IEEE International*, 1 -3 (2008).
83. Kim, D., Lee, Y., Cai, J., Lauer, I., Chang, L., Koester, S. J., Sylvester, D., & Blaauw, D. "Low power circuit design based on heterojunction tunneling transistors (HETTs)." *Proceedings of the 14th ACM/IEEE international symposium on Low power electronics and design*, http://doi.acm.org/10.1145/1594233.1594287 219--224 (2009).
84. Nagavarapu, V., Jhaveri, R., & Woo, J. C. S. "The Tunnel Source (PNPN) n-MOSFET: A Novel High Performance Transistor." *Electron Devices, IEEE Transactions on*, **55** 1013 -1019 (2008).
85. Kim, S. H., Agarwal, S., Jacobson, Z. A., Matheu, P., Hu, C., & Liu, T.-J. K. "Tunnel Field Effect Transistor With Raised Germanium Source." *Electron Device Letters, IEEE*, **31** 1107 -1109 (2010).
86. Izumi, K., Doken, M., & Ariyoshi, H. "C.M.O.S. devices fabricated on buried SiO2 layers formed by oxygen implantation into silicon." *Electronics Letters*, **14** 593 -594 (1978).
87. Lee, K. F., Gibbons, J. F., Saraswat, K. C., & Kamins, T. I. "Thin film MOSFET's fabricated in laser-annealed polycrystalline silicon." *Applied Physics Letters*, **35** http://link.aip.org/link/?APL/35/173/1 173-175 (1979).
88. Tasch, A. F., Holloway, T. C., Lee, K. F., & Gibbons, J. F. "Silicon-on-insulator m.o.s.f.e.t.s fabricated on laser-annealed polysilicon on SiO2." *Electronics Letters*, **15** 435 -437 (1979).
89. Malhi, S. D. S., Lam, H. W., Pinizzotto, R. F., Hamdi, A. H., & McDaniel, F. D.

"Novel SOI CMOS design using ultra thin near intrinsic substrate." *Electron Devices Meeting, 1982 International*, **28** 107 - 110 (1982).
90. Lim, H.-K. & Fossum, J. G. "Threshold voltage of thin-film Silicon-on-insulator (SOI) MOSFET's." *Electron Devices, IEEE Transactions on*, **30** 1244 - 1251 (1983).
91. Colinge, J.-P., Demoulin, E., Bensahel, D., Auvert, G., & Morel, H. "Transistors made in single-crystal SOI films." *Electron Device Letters, IEEE*, **4** 75 - 77 (1983).
92. An, X., Huang, R., Zhao, B., Zhang, X., & Wang, Y. "Design guideline of an ultra-thin body SOI MOSFET for low-power and high-performance applications." *Semiconductor Science and Technology*, **19** http://stacks.iop.org/0268-1242/19/i=3/a=009 347 (2004).
93. Ang, K.-W., Chui, K.-J., Bliznetsov, V., Wang, Y., Wong, L. Y., Tung, C.-H., Balasubramanian, N., Li, M.-F., Samudra, G., & Yeo, Y.-C. "Thin body silicon-on-insulator N-MOSFET with silicon-carbon source/drain regions for performance enhancement." *Electron Devices Meeting, 2005. IEDM Technical Digest. IEEE International*, 497 -500 (2005).
94. Chui, K.-J., Ang, K.-W., Madan, A., Wang, H., Tung, C.-H., Wong, L.-Y., Wang, Y., Choy, S.-F., Balasubramanian, N., Li, M. F., Samudra, G., & Yeo, Y.-C. "Source/drain germanium condensation for p-channel strained ultra-thin body transistors." *Electron Devices Meeting, 2005. IEDM Technical Digest. IEEE International*, 493 -496 (2005).
95. Tian, Y., Huang, R., Zhang, X., & Wang, Y. "A novel nanoscaled device concept: quasi-SOI MOSFET to eliminate the potential weaknesses of UTB SOI MOSFET." *Electron Devices, IEEE Transactions on*, **52** 561 - 568 (2005).
96. Gomez, L., Aberg, I., & Hoyt, J. L. "Electron Transport in Strained-Silicon Directly on Insulator Ultrathin-Body n-MOSFETs With Body Thickness Ranging From 2 to 25 nm." *Electron Device Letters, IEEE*, **28** 285 -287 (2007).
97. Vasileska, D. & Ahmed, S. S. "Narrow-width SOI devices: the role of quantum-mechanical size quantization effect and unintentional doping on the device operation." *Electron Devices, IEEE Transactions on*, **52** 227 - 236 (2005).
98. McDaid, L. J., Hall, S., Mellor, P. H., Eccleston, W., & Alderman, J. C. "Physical origin of negative differential resistance in SOI transistors." *Electronics Letters*, **25** 827 - 828 (1989).
99. Fiegna, C., Yang, Y., Sangiorgi, E., & O'Neill, A. G. "Analysis of Self-Heating Effects in Ultrathin-Body SOI MOSFETs by Device Simulation." *Electron Devices, IEEE Transactions on*, **55** 233 -244 (2008).
100. Ren, Z., Mehta, S., Cai, J., Wu, S., Zhu, Y., Kanarsky, T., Kanakasabapathy, S., Edge, L. F., Zhang, R., Lindo, P., Koshy, J., Tabakman, K., Kulkarni, P., Sardesai, V., Cheng, K., Khakifirooz, A., Doris, B., Bu, H., & Park, D.-G. "Assessment of fully-depleted planar CMOS for low power complex circuit operation." *Electron Devices Meeting (IEDM), 2011 IEEE International*, 15.5.1 -15.5.4 (2011).
101. Liu, Q., Yagishita, A., Loubet, N., Khakifirooz, A., Kulkarni, P., Yamamoto, T., Cheng, K., Fujiwara, M., Cai, J., Dorman, D., Mehta, S., Khare, P., Yako, K., Zhu, Y., Mignot, S., Kanakasabapathy, S., Monfray, S., Boeuf, F., Koburger, C., Sunamura, H., Ponoth, S., Reznicek, A., Haran, B., Upham, A., Johnson, R., Edge, L. F., Kuss, J., Levin, T., Berliner, N., Leobandung, E., Skotnicki, T., Hane, M., Bu, H., Ishimaru, K.,

Kleemeier, W., Takayanagi, M., Doris, B., & Sampson, R. "Ultra-thin-body and BOX (UTBB) fully depleted (FD) device integration for 22nm node and beyond." *VLSI Technology (VLSIT), 2010 Symposium on*, 61 -62 (2010).
102. Andrieu, F., Weber, O., Mazurier, J., Thomas, O., Noel, J.-P., Fenouillet-Bé andranger, C., Mazellier, J.-P., Perreau, P., Poiroux, T., Morand, Y., Morel, T., Allegret, S., Loup, V., Barnola, S., Martin, F., Damlencourt, J.-F., Servin, I., Cassé and, M., Garros, X., Rozeau, O., Jaud, M.-A., Cibrario, G., Cluzel, J., Toffoli, A., Allain, F., Kies, R., Lafond, D., Delaye, V., Tabone, C., Tosti, L., Bré andvard, L., Gaud, P., Paruchuri, V., Bourdelle, K. K., Schwarzenbach, W., Bonnin, O., Nguyen, B.-Y., Doris, B., Bœ anduf, F., Skotnicki, T., & Faynot, O. "Low leakage and low variability Ultra-Thin Body and Buried Oxide (UT2B) SOI technology for 20nm low power CMOS and beyond." *VLSI Technology (VLSIT), 2010 Symposium on*, 57 -58 (2010).
103. Hieda, K., Horiguchi, F., Watanabe, H., Sunouchi, K., Inoue, I., & Hamamoto, T. "New effects of trench isolated transistor using side-wall gates." *Electron Devices Meeting, 1987 International*, **33** 736 - 739 (1987).
104. Takato, H., Sunouchi, K., Okabe, N., Nitayama, A., Hieda, K., Horiguchi, F., & Masuoka, F. "High performance CMOS surrounding gate transistor (SGT) for ultra high density LSIs." *Electron Devices Meeting, 1988. IEDM '88. Technical Digest., International*, 222 -225 (1988).
105. Hisamoto, D., Kaga, T., Kawamoto, Y., & Takeda, E. "A fully depleted lean-channel transistor (DELTA)-a novel vertical ultra thin SOI MOSFET." *Electron Devices Meeting, 1989. IEDM '89. Technical Digest., International*, 833 -836 (1989).
106. Hisamoto, D., Kimura, S., Kaga, T., Nakagome, Y., Isoda, M., Nishida, T., & Takeda, E. "A new stacked cell structure for giga-bit DRAMs using vertical ultra-thin SOI (DELTA) MOSFETs." *Electron Devices Meeting, 1991. IEDM '91. Technical Digest., International*, 959 -961 (1991).
107. Hisamoto, D., Kaga, T., & Takeda, E. "Impact of the vertical SOI `DELTA' structure on planar device technology." *Electron Devices, IEEE Transactions on*, **38** 1419 - 1424 (1991).
108. Singh, N., Buddharaju, K. D., Manhas, S. K., Agarwal, A., Rustagi, S. C., Lo, G. Q., Balasubramanian, N., & Kwong, D.-L. "Si, SiGe Nanowire Devices by Top-Down Technology and Their Applications." *Electron Devices, IEEE Transactions on*, **55** 3107 - 3118 (2008).
109. Chang, C.-Y., Lee, T.-L., Wann, C., Lai, L.-S., Chen, H.-M., Yeh, C.-C., Chang, C.-S., Ho, C.-C., Sheu, J.-C., Kwok, T.-M., Yuan, F., Yu, S.-M., Hu, C.-F., Shen, J.-J., Liu, Y.-H., Chen, C.-P., Chen, S.-C., Chen, L.-S., Chen, L., Chiu, Y.-H., Fu, C.-Y., Huang, M.-J., Huang, Y.-L., Hung, S.-T., Liaw, J.-J., Lin, H.-C., Lin, H.-H., Lin, L.-T. S., Lin, S.-S., Mii, Y.-J., Ou-Yang, E., Shieh, M.-F., Su, C.-C., Tai, S.-P., Tao, H.-J., Tsai, M.-H., Tseng, K.-T., Wang, K.-W., Wang, S.-B., Xu, J. J., Yang, F.-K., Yang, S.-T., & Yeh, C.-N. "A 25-nm gate-length FinFET transistor module for 32nm node." *Electron Devices Meeting (IEDM), 2009 IEEE International*, 1 -4 (2009).
110. Zhang, W., Fossum, J. G., & Mathew, L. "The ITFET: A Novel FinFET-Based Hybrid Device." *Electron Devices, IEEE Transactions on*, **53** 2335 -2343 (2006).
111. Mathew, L., Du, Y., Thean, A. V.-Y., Sadd, M., Vandooren, A., Parker, C.,

Stephens, T., Mora, R., Rai, R., Zavala, M., Sing, D., Kalpat, S., Hughes, J., Shimer, R., Jallepalli, S., Workman, G., Zhang, W., Fossum, J. G., White, B. E., Nguyen, B.-Y., & Mogab, J. "CMOS Vertical Multiple Independent Gate Field Effect Transistor (MIGFET)." *SOI Conference, 2004. Proceedings. 2004 IEEE International*, 187 - 189 (2004).
112. Wu, X., Chan, P. C. H., Zhang, S., Feng, C., & Chan, M. "A three-dimensional stacked fin-CMOS technology for high-density ULSI circuits." *Electron Devices, IEEE Transactions on*, **52** 1998 - 2003 (2005).
113. Bangsaruntip, S., Cohen, G. M., Majumdar, A., Zhang, Y., Engelmann, S. U., Fuller, N., Gignac, L. M., Mittal, S., Newbury, J. S., Guillorn, M., Barwicz, T., Sekaric, L., Frank, M. M., & Sleight, J. W. "High performance and highly uniform gate-all-around silicon nanowire MOSFETs with wire size dependent scaling." *Electron Devices Meeting (IEDM), 2009 IEEE International*, 1 -4 (2009).
114. Yeo, K. H., Suk, S. D., Li, M., Yeoh, Y.-y., Cho, K. H., Hong, K.-H., Yun, S., Lee, M. S., Cho, N., Lee, K., Hwang, D., Park, B., Kim, D.-W., Park, D., & Ryu, B.-I. "Gate-All-Around (GAA) Twin Silicon Nanowire MOSFET (TSNWFET) with 15 nm Length Gate and 4 nm Radius Nanowires." *Electron Devices Meeting, 2006. IEDM '06. International*, 1 -4 (2006).
115. Fang, W. W., Singh, N., Bera, L. K., Nguyen, H. S., Rustagi, S. C., Lo, G. Q., Balasubramanian, N., & Kwong, D.-L. "Vertically Stacked SiGe Nanowire Array Channel CMOS Transistors." *Electron Device Letters, IEEE*, **28** 211 -213 (2007).
116. Dupre, C., Hubert, A., Becu, S., Jublot, M., Maffini-Alvaro, V., Vizioz, C., Aussenac, F., Arvet, C., Barnola, S., Hartmann, J.-M., Garnier, G., Allain, F., Colonna, J.-P., Rivoire, M., Baud, L., Pauliac, S., Loup, V., Chevolleau, T., Rivallin, P., Guillaumot, B., Ghibaudo, G., Faynot, O., Ernst, T., & Deleonibus, S. "15nm-diameter 3D stacked nanowires with independent gates operation: ΦFET." *Electron Devices Meeting, 2008. IEDM 2008. IEEE International*, 1 -4 (2008).
117. Singh, N., Lim, F. Y., Fang, W. W., Rustagi, S. C., Bera, L. K., Agarwal, A., Tung, C. H., Hoe, K. M., Omampuliyur, S. R., Tripathi, D., Adeyeye, A. O., Lo, G. Q., Balasubramanian, N., & Kwong, D. L. "Ultra-Narrow Silicon Nanowire Gate-All-Around CMOS Devices: Impact of Diameter, Channel-Orientation and Low Temperature on Device Performance." *Electron Devices Meeting, 2006. IEDM '06. International*, 1 -4 (2006).
118. Kotlyar, R., Obradovic, B., Matagne, P., Stettler, M., & Giles, M. D. "Assessment of room-temperature phonon-limited mobility in gated silicon nanowires." *Applied Physics Letters*, **84** http://link.aip.org/link/?APL/84/5270/1 5270-5272 (2004).
119. Ford, A. C., Ho, J. C., Chueh, Y.-L., Tseng, Y.-C., Fan, Z., Guo, J., Bokor, J., & Javey, A. "Diameter-Dependent Electron Mobility of InAs Nanowires." *Nano Letters*, **9** http://pubs.acs.org/doi/abs/10.1021/nl803154m 360-365 (2009).
120. Srivastava, V. M., Saubagya, N., & Singh, G. "Circuit Design with Independent Double Gate Transistors." *Advances in Computer Engineering (ACE), 2010 International Conference on*, 289 -291 (2010).
121. Lepselter, M. P. & Sze, S. M. "SB-IGFET: An insulated-gate field-effect transistor using Schottky barrier contacts for source and drain." *Proceedings of the IEEE*, **56** 1400 - 1402 (1968).

122. Larson, J. M. & Snyder, J. P. "Overview and status of metal S/D Schottky-barrier MOSFET technology." *Electron Devices, IEEE Transactions on*, **53** 1048 - 1058 (2006).
123. Connelly, D., Faulkner, C., & Grupp, D. E. "Optimizing Schottky S/D offset for 25-nm dual-gate CMOS performance." *Electron Device Letters, IEEE*, **24** 411 - 413 (2003).
124. Connelly, D., Clifton, P., Faulkner, C., & Grupp, D. E. "Ultra-thin-body fully depleted SOI metal source/drain n-MOSFETs and ITRS low-standby-power targets through 2018." *Electron Devices Meeting, 2005. IEDM Technical Digest. IEEE International*, 972 -975 (2005).
125. Chen, H.-W., Ko, C.-H., Wang, T.-J., Ge, C.-H., Wu, K., & Lee, W.-C. "Enhanced Performance of Strained CMOSFETs Using Metallized Source/Drain Extension (M-SDE)." *VLSI Technology, 2007 IEEE Symposium on*, 118 -119 (2007).
126. Connelly, D., Faulkner, C., Clifton, P. A., & Grupp, D. E. "Fermi-level depinning for low-barrier Schottky source/drain transistors." *Applied Physics Letters*, **88** 012105 - 012105-3 (2006).
127. Connelly, D., Clifton, P., Faulkner, C., Owens, J., & Wetzel, J. "Self-Aligned Low-Schottky Barrier Deposited Metal S/D MOSFETs with Si3N4 M/Si Passivation." *Device Research Conference, 2008*, 83 -84 (2008).
128. Tao, M., Agarwal, S., Udeshi, D., Basit, N., Maldonado, E., & Kirk, W. P. "Low Schottky barriers on n-type silicon (001)." *Applied Physics Letters*, **83** 2593 -2595 (2003).
129. Vega, R. A. & Liu, T.-J. K. "DSS MOSFET With Tunable SDE Regions by Fluorine Pre-Silicidation Ion Implant." *Electron Device Letters, IEEE*, **31** 785 -787 (2010).
130. Vega, R. A. & Liu, T.-J. K. "Dopant-Segregated Schottky Junction Tuning With Fluorine Pre-Silicidation Ion Implant." *Electron Devices, IEEE Transactions on*, **57** 1084 - 1092 (2010).
131. Kinoshita, A., Tsuchiya, Y., Yagishita, A., Uchida, K., & Koga, J. "Solution for high-performance Schottky-source/drain MOSFETs: Schottky barrier height engineering with dopant segregation technique." *VLSI Technology, 2004. Digest of Technical Papers. 2004 Symposium on*, 168 - 169 (2004).
132. Qiu, Z., Zhang, Z., Ostling, M., & Zhang, S.-L. "A Comparative Study of Two Different Schemes to Dopant Segregation at NiSi/Si and PtSi/Si Interfaces for Schottky Barrier Height Lowering." *Electron Devices, IEEE Transactions on*, **55** 396 -403 (2008).
133. Kaneko, A., Yagishita, A., Yahashi, K., Kubota, T., Omura, M., Matsuo, K., Mizushima, I., Okano, K., Kawasaki, H., Izumida, T., Kanemura, T., Aoki, N., Kinoshita, A., Koga, J., Inaba, S., Ishimaru, K., Toyoshima, Y., Ishiuchi, H., Suguro, K., Eguchi, K., & Tsunashima, Y. "High-Performance FinFET with Dopant-Segregated Schottky Source/Drain." *Electron Devices Meeting, 2006. IEDM '06. International*, 1 -4 (2006).
134. Chin, Y. K., Pey, K.-L., Singh, N., Lo, G.-Q., Tan, K. H., Ong, C.-Y., & Tan, L. H. "Dopant-Segregated Schottky Silicon-Nanowire MOSFETs With Gate-All-Around Channels." *Electron Device Letters, IEEE*, **30** 843 -845 (2009).
135. Vega, R. A. & Liu, T.-J. K. "A Comparative Study of Dopant-Segregated Schottky and Raised Source/Drain Double-Gate MOSFETs." *Electron Devices, IEEE Transactions on*, **55** 2665 -2677 (2008).
136. Vega, R. A., Liu, K., & Liu, T.-J. K. "Dopant-Segregated Schottky Source/Drain Double-Gate MOSFET Design in the Direct Source-to-Drain Tunneling Regime."

Electron Devices, IEEE Transactions on, **56** 2016 -2026 (2009).
137. Sacchetto, D., Ben Jamaa, M. H., De Micheli, G., & Leblebici, Y. "Fabrication and Characterization of Vertically Stacked Gate-All-Around Si Nanowire FET Arrays." *Proceedings of the 39th European Solid-State Device Research Conference (ESSDERC)*, (2009).
138. Fukuzumi, Y., Matsuoka, Y., Kito, M., Kido, M., Sato, M., Tanaka, H., Nagata, Y., Iwata, Y., Aochi, H., & Nitayama, A. "Optimal Integration and Characteristics of Vertical Array Devices for Ultra-High Density, Bit-Cost Scalable Flash Memory." *Electron Devices Meeting, 2007. IEDM 2007. IEEE International*, 449 -452 (2007).
139. Thelander, C., Rehnstedt, C., Froberg, L. E., Lind, E., Martensson, T., Caroff, P., Lowgren, T., Ohlsson, B. J., Samuelson, L., & Wernersson, L.-E. "Development of a Vertical Wrap-Gated InAs FET." *Electron Devices, IEEE Transactions on*, **55** 3030 -3036 (2008).
140. Lee, C.-W., Afzalian, A., Akhavan, N. D., Yan, R., Ferain, I., & Colinge, J.-P. "Junctionless multigate field-effect transistor." *Applied Physics Letters*, **94** http://link.aip.org/link/?APL/94/053511/1 053511 (2009).
141. Cui, Y., Zhong, Z., Wang, D., Wang, W. U., & Lieber, C. M. "High Performance Silicon Nanowire Field Effect Transistors." *Nano Letters*, **3** http://pubs.acs.org/doi/abs/10.1021/nl025875l 149-152 (2003).
142. Colinge, J.-P., Lee, C.-W., Afzalian, A., Akhavan, N. D., Yan, R., Ferain, I., Razavi, P., O'Neill, B., Blake, A., White, M., Kelleher, A.-M., McCarthy, B., & Murphy, R. "Nanowire transistors without junctions." *Nat Nano*, **5** http://dx.doi.org/10.1038/nnano.2010.15 225--229 (2010).
143. Colinge, J. P., Lee, C. W., Afzalian, A., Dehdashti, N., Yan, R., Ferain, I., Razavi, P., O'Neill, B., Blake, A., White, M., Kelleher, A. M., McCarthy, B., & Murphy, R. "SOI gated resistor: CMOS without junctions." *SOI Conference, 2009 IEEE International*, 1 -2 (2009).
144. Colinge, J.-P., Lee, C.-W., Ferain, I., Akhavan, N. D., Yan, R., Razavi, P., Yu, R., Nazarov, A. N., & Doria, R. T. "Reduced electric field in junctionless transistors." *Applied Physics Letters*, **96** http://link.aip.org/link/?APL/96/073510/1 073510 (2010).
145. Iqbal, M. M.-H., Hong, Y., Garg, P., Udrea, F., Migliorato, P., & Fonash, S. J. "The Nanoscale Silicon Accumulation-Mode MOSFET:A Comprehensive Numerical Study." *Electron Devices, IEEE Transactions on*, **55** 2946 -2959 (2008).
146. Rios, R., Cappellani, A., Armstrong, M., Budrevich, A., Gomez, H., Pai, R., Rahhal-orabi, N., & Kuhn, K. "Comparison of Junctionless and Conventional Trigate Transistors With Down to 26 nm." *Electron Device Letters, IEEE*, **32** 1170 -1172 (2011).
147. Lee, C.-W., Borne, A., Ferain, I., Afzalian, A., Yan, R., Dehdashti Akhavan, N., Razavi, P., & Colinge, J.-P. "High-Temperature Performance of Silicon Junctionless MOSFETs." *Electron Devices, IEEE Transactions on*, **57** 620 -625 (2010).
148. Choi, S.-J., Moon, D.-I., Kim, S., Duarte, J. P., & Choi, Y.-K. "Sensitivity of Threshold Voltage to Nanowire Width Variation in Junctionless Transistors." *Electron Device Letters, IEEE*, **32** 125 -127 (2011).
149. Lee, C.-W., Nazarov, A. N., Ferain, I., Akhavan, N. D., Yan, R., Razavi, P., Yu, R., Doria, R. T., & Colinge, J.-P. "Low subthreshold slope in junctionless multigate

transistors." *Applied Physics Letters*, **96** http://link.aip.org/link/?APL/96/102106/1 102106 (2010).
150. Hu, Y., Xiang, J., Liang, G., Yan, H., & Lieber, C. M. "Sub-100 Nanometer Channel Length Ge/Si Nanowire Transistors with Potential for 2 THz Switching Speed." *Nano Letters*, **8** http://pubs.acs.org/doi/abs/10.1021/nl073407b 925-930 (2008).
151. Zheng, G., Lu, W., Jin, S., & Lieber, C. M. "Synthesis and Fabrication of High-Performance n-Type Silicon Nanowire Transistors." *Advanced Materials*, **16** http://dx.doi.org/10.1002/adma.200400472 1890--1893 (2004).
152. Xiang, J., Lu, W., Hu, Y., Wu, Y., Yan, H., & Lieber, C. M. "Ge/Si nanowire heterostructures as high-performance field-effect transistors." *Nature*, **441** http://dx.doi.org/10.1038/nature04796 489--493 (2006).
153. Liang, G., Xiang, J., Kharche, N., Klimeck, G., Lieber, C. M., & Lundstrom, M. "Performance Analysis of a Ge/Si Core/Shell Nanowire Field-Effect Transistor." *Nano Letters*, **7** http://pubs.acs.org/doi/abs/10.1021/nl062596f 642-646 (2007).
154. Petersen, K. E. "Dynamic micromechanics on silicon: Techniques and devices." *Electron Devices, IEEE Transactions on*, **25** 1241 - 1250 (1978).
155. Joshi, V., Khieu, C., Smith, C. G., Schepens, C., Csaszar, F., Lacey, D., Nagata, T., Renault, M., Van Kampen, R., Knipe, R., & Yost, D. "A CMOS compatible back end MEMS switch for logic functions." *Interconnect Technology Conference (IITC), 2010 International*, 1 -3 (2010).
156. Pott, V., Kam, H., Nathanael, R., Jeon, J., Alon, E., & Liu, T.-J. K. "Mechanical Computing Redux: Relays for Integrated Circuit Applications." *Proceedings of the IEEE*, **98** 2076 -2094 (2010).
157. Kam, H., Lee, D. T., Howe, R. T., & King, T.-J. "A new nano-electro-mechanical field effect transistor (NEMFET) design for low-power electronics." *Electron Devices Meeting, 2005. IEDM Technical Digest. IEEE International*, 463 -466 (2005).
158. Lee, J.-O., Kim, M.-W., Ko, S.-D., Kang, H.-O., Bae, W.-H., Kang, M.-H., Kim, K.-N., Yoo, D.-E., & Yoon, J.-B. "3-terminal nanoelectromechanical switching device in insulating liquid media for low voltage operation and reliability improvement." *Electron Devices Meeting (IEDM), 2009 IEEE International*, 1 -4 (2009).
159. Kam, H., Pott, V., Nathanael, R., Jeon, J., Alon, E., & Liu, T.-J. K. "Design and reliability of a micro-relay technology for zero-standby-power digital logic applications." *Electron Devices Meeting (IEDM), 2009 IEEE International*, 1 -4 (2009).
160. Chen, F., Spencer, M., Nathanael, R., Wang, C., Fariborzi, H., Gupta, A., Kam, H., Pott, V., Jeon, J., Liu, T.-J. K., Markovic, D., Stojanovic, V., & Alon, E. "Demonstration of integrated micro-electro-mechanical switch circuits for VLSI applications." *Solid-State Circuits Conference Digest of Technical Papers (ISSCC), 2010 IEEE International*, 150 -151 (2010).
161. Nathanael, R., Pott, V., Kam, H., Jeon, J., & Liu, T.-J. K. "4-terminal relay technology for complementary logic." *Electron Devices Meeting (IEDM), 2009 IEEE International*, 1 -4 (2009).
162. Chen, I-R., Hutin, L., Park, C., Lee, R., Nathanael, R., Yaung, J., Jeon, J., & King Liu, T.-J.. "Scaled Micro-Relay Structure with Low Strain Gradient for Reduced

Operating Voltage." *Electrochemical Society (ECS) Meeting, 221st*, p. 867 (2012).
163. Kam, H., Liu, T.-J. K., Stojanović and, V., Marković and, D., & Alon, E. "Design, Optimization, and Scaling of MEM Relays for Ultra-Low-Power Digital Logic." *Electron Devices, IEEE Transactions on*, **58** 236 -250 (2011).
164. Spencer, M., Chen, F., Wang, C. C., Nathanael, R., Fariborzi, H., Gupta, A., Kam, H., Pott, V., Jeon, J., Liu, T.-J. K., Marković and, D., Alon, E., & Stojanović and, V. "Demonstration of Integrated Micro-Electro-Mechanical Relay Circuits for VLSI Applications." *Solid-State Circuits, IEEE Journal of*, **46** 308 -320 (2011).
165. Dadgour, H., Cassell, A. M., & Banerjee, K. "Scaling and variability analysis of CNT-based NEMS devices and circuits with implications for process design." *Electron Devices Meeting, 2008. IEDM 2008. IEEE International*, 1 -4 (2008).
166. Dadgour, H. F. & Banerjee, K. "Hybrid NEMS-CMOS integrated circuits: A novel strategy for energy-efficient designs." *Computers Digital Techniques, IET*, **3** 593 -608 (2009).
167. Abele, N., Fritschi, R., Boucart, K., Casset, F., Ancey, P., & Ionescu, A. M. "Suspended-gate MOSFET: bringing new MEMS functionality into solid-state MOS transistor." *Electron Devices Meeting, 2005. IEDM Technical Digest. IEEE International*, 479 - 481 (2005).
168. Shen, X., Chong, S., Lee, D., Parsa, R., Howe, R. T., & Wong, H.-S. P. "2D analytical model for the study of NEM relay device scaling." *Simulation of Semiconductor Processes and Devices (SISPAD), 2011 International Conference on*, 243 -246 (2011).
169. Chen, F., Kam, H., Markovic, D., Liu, T.-J. K., Stojanovic, V., & Alon, E. "Integrated circuit design with NEM relays." *Computer-Aided Design, 2008. ICCAD 2008. IEEE/ACM International Conference on*, 750 -757 (2008).
170. Lee, D., Osabe, T., & Liu, T.-J. K. "Scaling Limitations for Flexural Beams Used in Electromechanical Devices." *Electron Devices, IEEE Transactions on*, **56** 688 -691 (2009).
171. Akarvardar, K., Elata, D., Parsa, R., Wan, G. C., Yoo, K., Provine, J., Peumans, P., Howe, R. T., & Wong, H.-S. P. "Design Considerations for Complementary Nanoelectromechanical Logic Gates." *Electron Devices Meeting, 2007. IEDM 2007. IEEE International*, 299 -302 (2007).
172. Varadarajan, V., Yasuda, Y., Balasubramanian, S., & Liu, T.-J. K. "WireFET Technology for 3-D Integrated Circuits." *Electron Devices Meeting, 2006. IEDM '06. International*, 1 -4 (2006).
173. Subramanian, V. & Saraswat, K. C. "Laterally crystallized polysilicon TFTs using patterned light absorption masks." *Device Research Conference Digest, 1997. 5th*, 54 -55 (1997).
174. Subramanian, V., Toita, M., Ibrahim, N. R., Souri, S. J., & Saraswat, K. C. "Low-leakage germanium-seeded laterally-crystallized single-grain 100-nm TFTs for vertical integration applications." *Electron Device Letters, IEEE*, **20** 341 -343 (1999).
175. Oh, J. H., Kang, D. H., Park, M. K., & Jang, J. "Low Off-State Drain Current Poly-Si TFT with Ni-Mediated Crystallization of Ultrathin a-Si." *Electrochemical and Solid-State Letters*, **12** http://link.aip.org/link/?ESL/12/J29/1 J29-J32 (2009).
176. Jung, S.-M., Jang, J., Cho, W., Moon, J., Kwak, K., Choi, B., Hwang, B., Lim, H., Jeong, J., Kim, J., & Kim, K. "The revolutionary and truly 3-dimensional 25F^2 SRAM

technology with the smallest S^3 (stacked single-crystal Si) cell, 0.16um^2, and SSTFT (atacked single-crystal thin film transistor) for ultra high density SRAM." *VLSI Technology, 2004. Digest of Technical Papers. 2004 Symposium on*, 228 - 229 (2004).
177. Batude, P., Vinet, M., Pouydebasque, A., Le Royer, C., Previtali, B., Tabone, C., Hartmann, J.-M., Sanchez, L., Baud, L., Carron, V., Toffoli, A., Allain, F., Mazzocchi, V., Lafond, D., Thomas, O., Cueto, O., Bouzaida, N., Fleury, D., Amara, A., Deleonibus, S., & Faynot, O. "Advances in 3D CMOS sequential integration." *Electron Devices Meeting (IEDM), 2009 IEEE International*, 1 -4 (2009).
178. Nho, H. H., Horowitz, M., & Wong, S. S. "A high-speed, low-power 3D-SRAM architecture." *Custom Integrated Circuits Conference, 2008. CICC 2008. IEEE*, 201 -204 (2008).
179. Hsu, C.-L. & Wu, C.-F. "High-performance 3D-SRAM architecture design." *Circuits and Systems (APCCAS), 2010 IEEE Asia Pacific Conference on*, 907 -910 (2010).
180. Yokoyama, M., Yasuda, T., Takagi, H., Yamada, H., Fukuhara, N., Hata, M., Sugiyama, M., Nakano, Y., Takenaka, M., & Takagi, S. "Thin Body III--V-Semiconductor-on-Insulator Metal--Oxide--Semiconductor Field-Effect Transistors on Si Fabricated Using Direct Wafer Bonding." *Applied Physics Express*, **2** http://apex.jsap.jp/link?APEX/2/124501/ 124501 (2009).
181. Seabaugh, A. C. & Zhang, Q. "Low-Voltage Tunnel Transistors for Beyond CMOS Logic." *Proceedings of the IEEE*, **98** 2095 -2110 (2010).
182. Vandenberghe, W. G., Soree, B., Magnus, W., Groeseneken, G., & Fischetti, M. V. "Impact of field-induced quantum confinement in tunneling field-effect devices." *Applied Physics Letters*, **98** http://link.aip.org/link/?APL/98/143503/1 143503 (2011).
183. Haran, B. S., Kumar, A., Adam, L., Chang, J., Basker, V., Kanakasabapathy, S., Horak, D., Fan, S., Chen, J., Faltermeier, J., Seo, S., Burkhardt, M., Burns, S., Halle, S., Holmes, S., Johnson, R., McLellan, E., Levin, T. M., Zhu, Y., Kuss, J., Ebert, A., Cummings, J., Canaperi, D., Paparao, S., Arnold, J., Sparks, T., Koay, C. S., Kanarsky, T., Schmitz, S., Petrillo, K., Kim, R. H., Demarest, J., Edge, L. F., Jagannathan, H., Smalley, M., Berliner, N., Cheng, K., LaTulipe, D., Koburger, C., Mehta, S., Raymond, M., Colburn, M., Spooner, T., Paruchuri, V., Haensch, W., McHerron, D., & Doris, B. "22 nm technology compatible fully functional 0.1 μm2 6T-SRAM cell." *Electron Devices Meeting, 2008. IEDM 2008. IEEE International*, 1 -4 (2008).
184. Kim, D.-H. & del Alamo, J. A. "30 nm E-mode InAs PHEMTs for THz and future logic applications." *Electron Devices Meeting, 2008. IEDM 2008. IEEE International*, 1 -4 (2008).
185. Sakamoto, T., Iguchi, N., & Aono, M. "Nonvolatile triode switch using electrochemical reaction in copper sulfide." *Applied Physics Letters*, **96** http://link.aip.org/link/?APL/96/252104/1 252104 (2010).
186. Dewey, G., Chu-Kung, B., Boardman, J., Fastenau, J. M., Kavalieros, J., Kotlyar, R., Liu, W. K., Lubyshev, D., Metz, M., Mukherjee, N., Oakey, P., Pillarisetty, R., Radosavljevic, M., Then, H. W., & Chau, R. "Fabrication, characterization, and physics of III-V heterojunction tunneling Field Effect Transistors (H-TFET) for steep sub-threshold swing." *Electron Devices Meeting (IEDM), 2011 IEEE International*, 33.6.1 - 33.6.4 (2011).

187. Majumdar, A., Ren, Z., Koester, S. J., & Haensch, W. “Undoped-Body Extremely Thin SOI MOSFETs With Back Gates.” *Electron Devices, IEEE Transactions on*, **56** 2270 - 2276 (2009).

188. Fariborzi, H., Chen, F., Stojanovic, V., Nathanael, R., Jaeseok Jeon, Tsu-Jae King Liu. "Design and demonstration of micro-electro-mechanical relay multipliers," *Solid State Circuits Conference (A-SSCC), 2011 IEEE Asian* , pp.117-120, 14-16 Nov. 2011

Chapter 3

Germanium Source Tunnel Field Effect Transistors: Simulation, Calibration, and Design Optimization

3.1 Introduction

As discussed in Chapter 2, Tunnel Field-Effect Transistors (TFETs) are a potential CMOS replacement or complement in future technology nodes. TFETs have been investigated for ultra-low-power applications because they can in principle achieve a higher on/off current ratio (I_{ON}/I_{OFF}) than MOSFETs at low operating voltages [1]. They are able to achieve less than 60 mV/decade operation due to their different carrier transport; instead of the high energy tail of electrons passing over a barrier, electrons tunnel through a barrier. This effect has been demonstrated in numerous experimental devices.

N-channel TFETs with record high I_{ON}/I_{OFF} (>10^6) for low-voltage (0.5V) operation have been demonstrated by using polycrystalline Germanium (poly-Ge) in the source region to achieve a small effective tunneling band gap [2]. The Ge-source TFET design is attractive because it can be relatively easily integrated into a conventional CMOS process by adapting process techniques currently used in high-volume production for embedded $Si_{1-x}Ge_x$ source/drain stressors [3]. The heterostructure design (*i.e.* using a small band gap material only in the source region) provides for low off-state leakage current, in contrast to a pure Ge TFET design [4-9].

TFETs typically have been designed and simulated for lateral tunneling across the source-channel interface [5]. This has been referred to as "point tunneling" [10]. A heavily doped source with a steep doping profile is needed to maximize I_{ON}/I_{OFF} for this design [11-14], which inherently has a relatively small tunneling area and hence low I_{ON}. If the gate overlaps the source, and the source is not very heavily doped, tunneling can occur completely within the source. This has been referred to as "line tunneling" and it can provide for higher drive current due to increased tunneling area [10].

To adequately study TFETs, proper simulation software is necessary to understand the complex interplay of geometrical and electrical factors in influencing such performance metrics as subthreshold slope, on-state drive current, gate capacitance, and leakage current. Over time, simulation software has improved from local band-to-band tunneling models, which predict where tunneling occurs based on simple electric field magnitude, to complex dynamic nonlocal models, which take into account band structure and electric field direction to determine all locations where tunneling will occur. This work presents the first application of a dynamic nonlocal band-to-band tunneling model calibrated to a Germanium source Si TFET design.

Initially, the tunneling parameters must be calibrated to an experimental device, in this case a polycrystalline Germanium source n-TFET design [2]. An initial study of the design optimization for this device is first performed. Then, an advanced study, which takes into account Fermi statistics and quantization, appropriate for a highly doped source on an Ultra Thin Body Silicon-On-Insulator wafer, is performed, with a focus on source design optimization. Specifically, the impacts of the gate-to-source

overlap ($L_{OVERLAP}$), Ge-source thickness (T_{GE}), and source dopant concentration (N_{SOURCE}) are investigated.

3.2 Band-to-Band Tunneling

Tunneling is a quantum mechanical process where electrons move through potential energy barriers. Band-to-band tunneling is the effect when electrons travel from the valence band to the conduction band (or vice versa) through the forbidden energy band gap. Understanding the nature of this band-to-band tunneling is important for understanding the approximations made in various simulation models. This understanding is also useful when optimizing design parameters of TFETs for maximum performance.

3.2.1 Tunneling Theory

Quantum mechanical tunneling occurs due to a non-zero probability for transition through a barrier. To adequately model tunneling, this probability must be found and multiplied by the number of electrons in a given volume of space to find the net tunneling rate of electrons.

Kane is well known for having derived one of the first expressions for tunneling probability as it relates to band-to-band tunneling (known as Zener tunneling in the literature of this time period) [15]. It should be noted that Keldysh derived the same integral for the tunneling probability one year earlier [16], but it is understood that Kane's derivation is simpler while being more rigorous.

Kane used a time-independent Schrödinger equation in the presence of a uniform electric field. He represents the basis function using Bloch functions and used perturbation theory to evaluate the tunneling probability. Kane's derivation is quite complex, but results in an equation for tunneling per cm^3, as demonstrated in Equation 3.1.

$$G_{BTBT} = \frac{E^2 m^{*\left(\frac{1}{2}\right)}}{18\pi\hbar^2 E_G^{\frac{1}{2}}} \exp\left\{\frac{-\pi m^{*\left(\frac{1}{2}\right)} E_G^{\frac{3}{2}}}{2\hbar E}\right\} \tag{3.1}$$

In this equation, E represents the (uniform) electric field strength, m* represents the effective mass (reduced mass in Kane's nomenclature), and E_G is the bandgap energy. Commonly, this equation is reduced to the following to allow for a two variable calibration to experimental data, as shown in Equation 3.2.

$$G_{BTBT} = AE^2 \exp\left\{\frac{-B}{E}\right\} \tag{3.2}$$

Keldysh showed that this equation was only relevant for direct tunneling (without assistance from a phonon), and that a slightly modified version is necessary to

account for the phonon's effects in indirect tunneling [17]. Equation 3.3 shows this modified form.

$$G_{BTBT} = AE^{2.5} \exp\left\{\frac{-B}{E}\right\} \quad (3.3)$$

In both cases, the A parameter is the linear parameter, and the B parameter is the exponential parameter, allowing relatively straightforward calibration.

3.3 Simulation Methods

TCAD simulation tools allow for calibration of models to experimental data, optimization of device designs, and variability analysis of designs. The Synopsys line of TCAD device simulators models Band-to-Band tunneling current as Generation/Recombination rates. For each mesh point, the simulator calculates the probability of tunneling using a variation of Kane or Keldysh's equation as well as an analysis of the band structure at that mesh point. This analysis of the band structure differs for the different models and makes a large impact on both the quantity and location of predicted tunneling current.

3.3.1 MEDICI

MEDICI is a one and two-dimensional simulator created by Synopsys, Inc. [18]. It contains a band-to-band tunneling model based on Kane's equation, as shown in Equation 3.4 and 3.5.

$$G_{BTBT} = \frac{D_{TUNNEL} A_{BTBT} \mathrm{E}^{\mathrm{C.BTBT}}}{\mathrm{E}_{\mathrm{G}}^{0.5}} \exp\left\{\frac{-B_{BTBT} E_G^{1.5}}{E}\right\} \quad (3.4)$$

where:

$$D_{TUNNEL} = \frac{1}{1+\exp\left(\frac{E_{V,1}-E_{Fp,1}}{kT}\right)} - \frac{1}{1+\exp\left(\frac{E_{C,2}-E_{Fn,2}}{kT}\right)} \quad (3.5)$$

Several modifications have been performed on Kane's original formula. The bandgap of the semiconductor has been isolated from the A and B parameters, allowing for bandgap narrowing effects due to high doping to be considered without modifying the A and B parameters. In addition, the linear exponent can be changed to fit Keldysh's model.

To determine where tunneling occurs, a search along the direction opposite to the electric field is performed. If there is an electric potential increase of E_G/q or greater, MEDICI will calculate the tunneling rate for that data point.

To calculate the electric field (E), MEDICI offers several options. The local field at each mesh point is the default option. The average field along the tunneling path is another options. A third option uses the average field for the pre-exponential electric field $E^{\mathrm{C.BTBT}}$, but a path integral for the electric field in the exponential term E.

Finally, D_{TUNNEL} is used to take into account the probability of having a filled

state to tunnel from and an empty state to tunnel to. This pre-exponential factor is used to reduce the tunneling rate at zero bias.

In general, the MEDICI model does not handle heterostructures well. The bandgap used is the bandgap for each mesh point. Thus, an electron tunneling from Germanium to Silicon will result in different probabilities at the beginning and end of the tunneling path. In addition, although a basic modification is allowed to use Keldysh's model, a comprehensive indirect tunneling solution that incorporates phonon energies is not permitted.

3.3.2 Sentaurus Local Tunneling

Sentaurus Device can simulate one, two, and three-dimensional structures [19]. It contains several tunneling models, including both local and nonlocal models. The Schenk model is a basic phonon-assisted band-to-band tunneling model [20]. No nonlocal dependence is used, meaning that no search for a region where the conduction band equals the valence band is performed. Equation 3.6 shows the generation rate using the Schenk model.

$$G_{BTBT} = AE^{3.5}\frac{\left(np-n_{i,eff}^2\right)}{(n+n_{i,eff})(p+n_{i,eff})}\left[\frac{\left(E_C^{\mp}\right)^{-1.5}\exp\left(\frac{-E_C^{\mp}}{E}\right)}{\exp\left(\frac{\hbar\omega}{kT}\right)-1}+\frac{\left(E_C^{\pm}\right)^{-1.5}\exp\left(\frac{-E_C^{\pm}}{E}\right)}{1-\exp\left(\frac{\hbar\omega}{kT}\right)}\right] \quad (3.6)$$

The Hurkx model is similar to the MEDICI model, again without the basic nonlocal support present in the MEDICI model, as shown in Equation 3.7 and 3.8 [21].

$$G_{BTBT} = D\,A\left(\frac{\mathrm{E}}{1\frac{\mathrm{V}}{\mathrm{cm}}}\right)^{\mathrm{P}}\exp\left\{\frac{-B\,E_G(T)^{1.5}}{(E_G(300K)^{1.5}E}\right\} \quad (3.7)$$

where:

$$D = \frac{\left(np-n_{i,eff}^2\right)}{(n+n_{i,eff})(p+n_{i,eff})}(1-|\alpha|)+\alpha \quad (3.8)$$

The Hurkx model also does not take into account any nonlocal dependence. Its advantage over the basic MEDICI model is that it incorporates temperature dependence in calculating bandgap. Ideally, one could calibrate the model to a device tested at 300K and be able to simulate how the device would behave at other temperatures.

Finally, Sentaurus Device includes a very simple Kane formula for band-to-band tunneling. This formula does not incorporate the zero bias correction factor that is present in the other models. Equation 3.9 shows the simple model.

$$G_{BTBT} = A\,E^{\mathrm{P}}\exp\left\{\frac{-B}{E}\right\} \quad (3.9)$$

3.3.3 Sentaurus Nonlocal Tunneling

Tunneling probability (and thus, tunneling current as modeled as recombination

and generation rates) is dependent on the band profile along the entire tunneling path, not just the endpoints. Sentaurus Device's Nonlocal tunneling model allows a user to model this dependence by creating a special nonlocal mesh that is overlaid on top of the existing mesh with a specific length and angle. As such, it requires that a user predetermine the angle of tunneling. The generation rate is then modeled as shown in Equation 3.10, in this case for the generation of holes in the valence band.

$$R_V - G_V = \frac{A_{CV}}{2kq}\vartheta\left[\varepsilon - \mathrm{E_V(u)}, \frac{\mathrm{dE_V}}{\mathrm{du}}(\mathrm{u})\right]\vartheta\left[\varepsilon - \mathrm{E_C(l)}, \frac{\mathrm{dE_C}}{\mathrm{dl}}(\mathrm{l})\right]\Gamma_V(\mathrm{u,l},\varepsilon)[\mathrm{T_p(u)} + \mathrm{T_n(l)}]\left[\left(1+\exp\left(\frac{\varepsilon-\mathrm{E_{F,p}(u)}}{\mathrm{kT_P(u)}}\right)\right)^{-1} - \left(1+\exp\left(\frac{\varepsilon-\mathrm{E_{F,n}}(l)}{\mathrm{kT_n(l)}}\right)\right)^{-1}\right] \quad (3.10)$$

3.4 Sentaurus Dynamic Nonlocal Tunneling Model

A nonlocal band-to-band tunneling model applicable to arbitrary tunneling barriers involving nonuniform electric field and abrupt/graded heterojunctions recently has been developed [19]. Tunneling paths are dynamically determined according to the gradient of the band energy. The model accounts for the direct and phonon-assisted tunneling processes, which exactly reduces to the Kane's and Keldysh's model [15],[17] in the uniform electric field limit. Equation 3.11 shows the recombination rate for direct tunneling of holes, which is similar to the traditional Nonlocal Tunneling model.

$$R_V = |\nabla E_V(0)|\mathrm{C_d}\exp\left(-2\int_0^l \kappa dx\right)\left[\left(1+\exp\left(\frac{\varepsilon-\mathrm{E_{F,n}}(l)}{\mathrm{kT(l)}}\right)\right)^{-1} - \left(1+\exp\left(\frac{\varepsilon-\mathrm{E_{F,p}}(0)}{\mathrm{kT(0)}}\right)\right)^{-1}\right] \quad (3.11)$$

3.5 Initial Calibration to Experimental Data

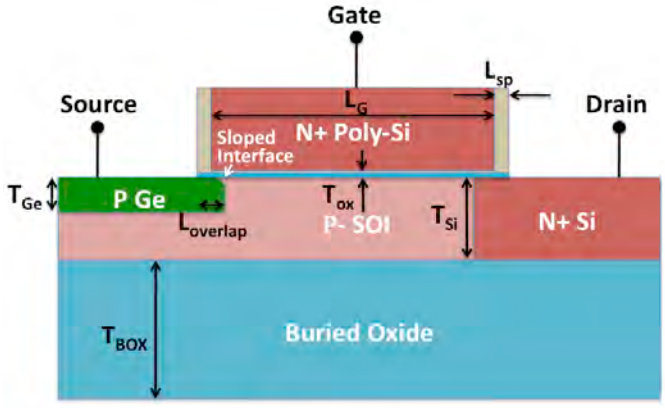


Figure 3.1: Schematic cross-sectional view of Germanium-source tunneling transistor.

A recently reported germanium-source TFET [2] was used to validate this model, schematically shown in Figure 3.1. A and B coefficients for Germanium of 1.46 x 10^{17} $cm^{-3}s^{-1}$ and 3.59 x 10^{6} V/cm respectively were used. To fit the experimental data accurately, several models and parameters were enabled and adjusted. N_{SOURCE} was selected to be in the low 10^{18} cm^{-3} range, consistent with sheet resistance and Hall measurements of a Ge film formed using the same deposition process conditions as for the TFET source region in [2]. A low-temperature deposition at 425 °C is used to selectively deposit the doped Ge source, resulting in an abrupt source doping profile [2]. Interface fixed charge (Q_F) exists at the Ge-SiO_2 gate-dielectric interface of this device due to the isotropic dry etch process used to undercut the gate stack prior to Ge deposition. To accurately reproduce the small kink in the measured I_D-V_{GS} curve, traps (with areal density N_T) were placed along the Ge-Si interface. Table 3.1 shows the relevant geometrical and electrical parameter values used to match the model to the experimental data. The measured and simulated I_D-V_{GS} characteristics for a long channel (5 μm) TFET are plotted in Figure 3.2.

Parameters			
Geometrical	**Value**	**Electrical**	**Value**
L_G	5 μm	ϕ_M	3.9 eV
$T_{OX\ OVER\ CHANNEL}$	3 nm	Q_F	6.5 x 10^{12} cm^{-2}
$T_{OX\ OVER\ SOURCE}$	2 nm	$N_{Channel}$	10^{15} cm^{-3}
T_{BOX}	200 nm	N_{Source}	3 x 10^{18} cm^{-3}
T_{Si}	70 nm	$N_{BulkTraps}$	0 cm^{-3}
T_{Ge}	21 nm	$N_{IntTraps}$	7.25 x 10^{12} cm^{-2}
Ge/Si Interface	Sloped	E_{Traps}	$E_{V,Ge}$ + 0.1 eV
L_{Sp}	8 nm		
$L_{Overlap}$	18 nm		

Table 3.1: TFET device parameters for initial model calibration to experimental data.

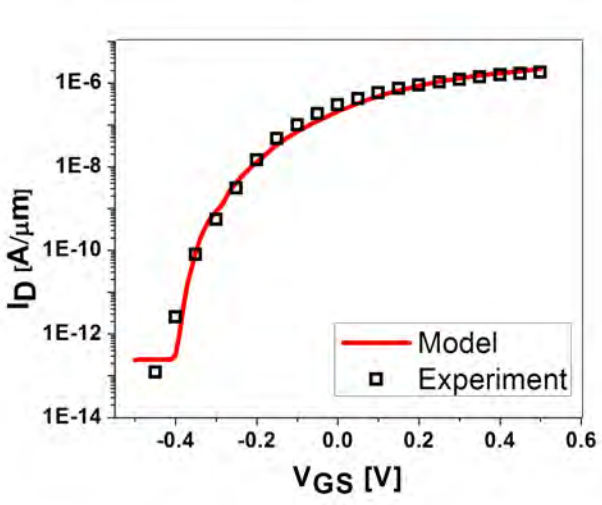
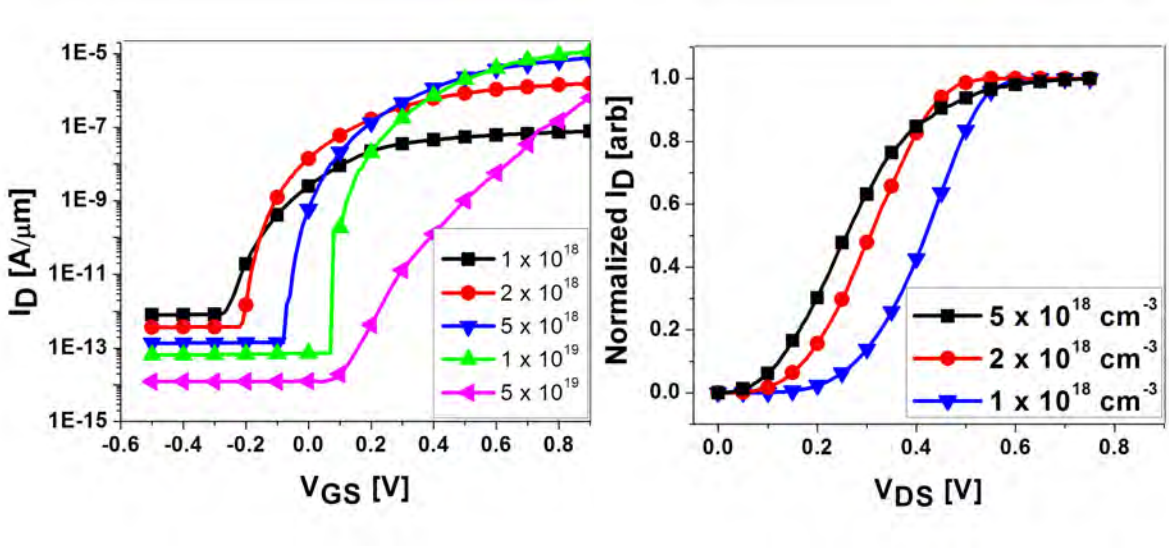


Figure 3.2: Measured experimental and simulated I_{DS}-V_{GS} data.

3.6 Design Optimization

Default Parameters			
Geometrical	**Value**	**Electrical**	**Value**
L_G	5 μm	ϕ_M	3.9 eV
$T_{OX\ OVER\ CHANNEL}$	3 nm	Q_F	6.5×10^{12} cm^{-2}
$T_{OX\ OVER\ SOURCE}$	2 nm	$N_{Channel}$	10^{15} cm^{-3}
T_{BOX}	200 nm	N_{Source}	10^{19} cm^{-3}
T_{Si}	70 nm	$N_{BulkTraps}$	0 cm^{-3}
T_{Ge}	21 nm	$N_{IntTraps}$	0 cm^{-2}
Ge/Si Interface	Vertical	E_{Traps}	$E_{V,Ge}$ + 0.1 eV
L_{Sp}	8 nm		
$L_{Overlap}$	18 nm		

Table 3.2: Default TFET device parameters for initial design optimization study.



(a) **(b)**

Figure 3.3: Simulated (a) I_D-V_{GS} and (b) I_D-V_{DS} data for varying source doping concentration N_{SOURCE} I_D-V_{DS} at V_{GT} = 0.5 V, normalized to I_D at V_{DS} = 0.5 V. N_{SOURCE} of 10^{19} cm^{-3} optimizes I_D-V_{GS} swing. Lower N_{SOURCE} gives less linear I_D-V_{DS} behavior.

Table 3.2 shows the default parameter values used for the optimization study. Since on-state tunneling occurs primarily within the source region for this TFET design, it is critical to optimize the source parameters in order to achieve maximum performance. Figure 3.3a shows that the source doping concentration (N_{SOURCE}) dramatically affects performance, and that an optimal value (~1×10^{19} cm^{-3}) exists. From Figure 3.3b, which shows normalized-I_D *vs.* V_D characteristics for V_G-$V_T \equiv V_{GT}$ = 0.5 V (where V_T is the tunneling turn-on voltage), it can be seen that the linearity of the I_D-V_{DS} characteristic at low V_{DS} improves with N_{SOURCE}.

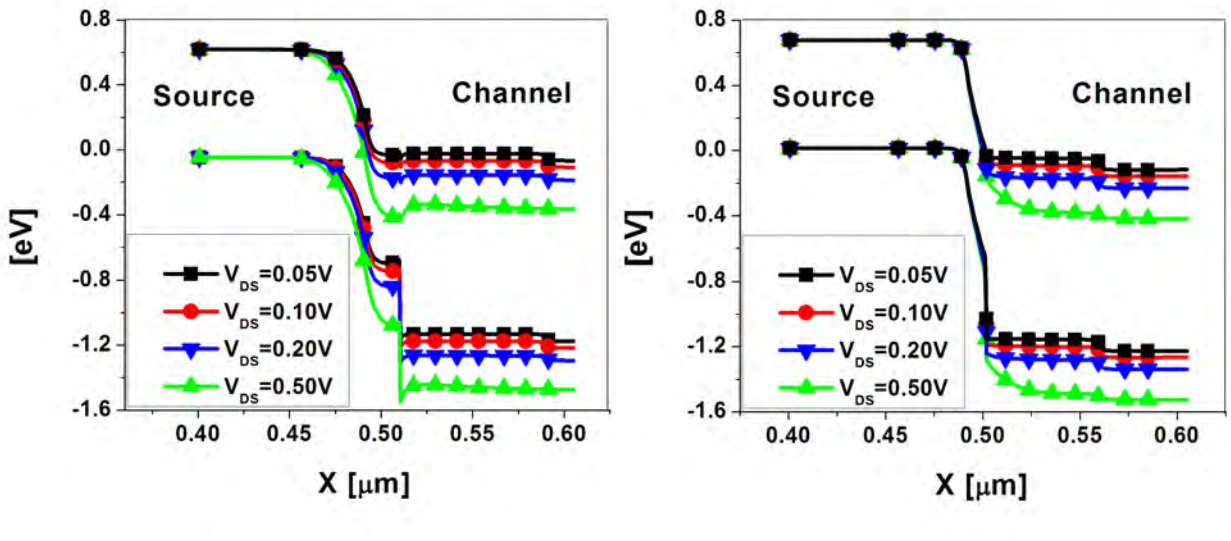


Figure 3.4: Energy band diagrams at the Source-Channel region for source doping N_{SOURCE} of (a) 10^{18} and (b) 10^{19} cm^{-3}.

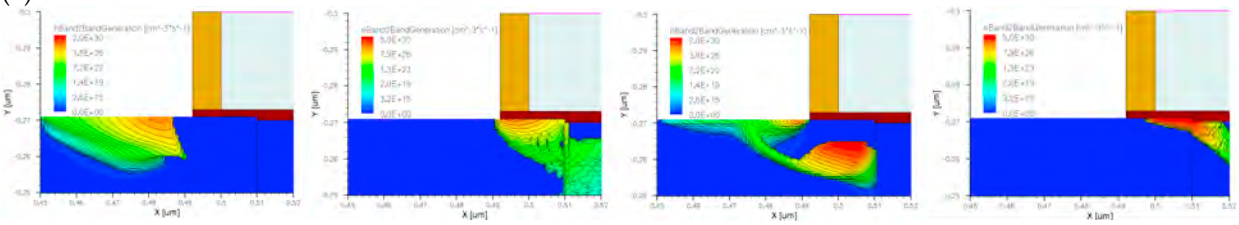


Figure 3.5: Carrier tunneling at Source–Channel interface in a device with source doping N_{SOURCE} of (a) 10^{18} and (b) 10^{19} cm^{-3}. Left and right figures show tunneling through valence and conduction band edges, respectively. In (a), note that the tunneling is mostly lateral. In (b), note that the tunneling is more vertical with a lateral component still present.

This can be understood by examining the energy-band diagrams in Figures 3.4a and 3.4b for N_{SOURCE} = 10^{18} cm^{-3} and N_{SOURCE} = 10^{19} cm^{-3}, respectively: The amount of overlap between valence-band states and conduction-band states at very low V_{DS} is significant for the case of higher N_{SOURCE}, so that V_{DS} does not "gate" the tunneling process. Band-to-band tunneling generation contour plots in Figures 3.5a and 3.5b show that tunneling occurs mostly laterally in the case of lower N_{SOURCE}, but mostly vertically in the case of higher N_{SOURCE}.

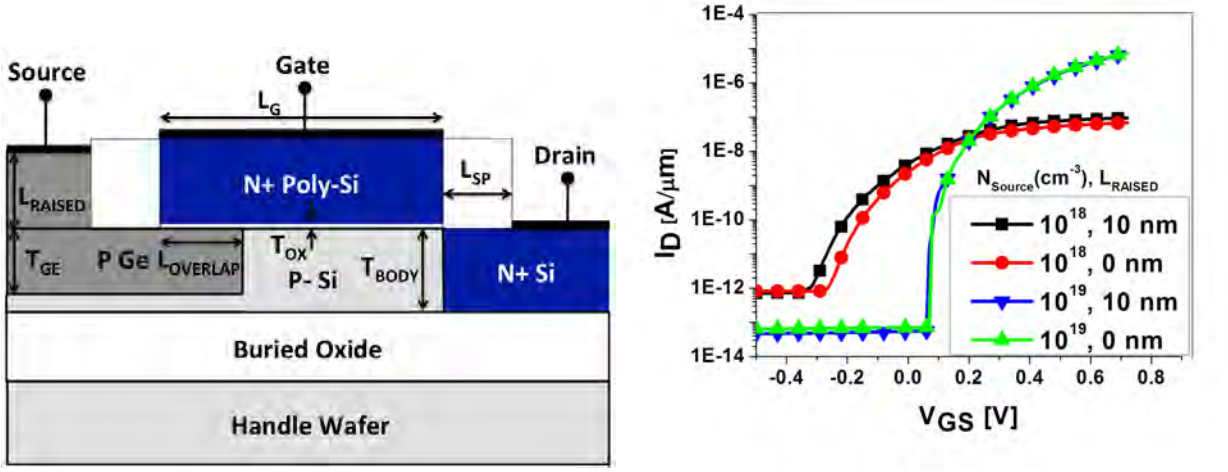


Figure 3.6: (a) Schematic cross-sectional view of Germanium-source tunneling transistor with a raised source. **(b)** Simulated I_{DS}-V_{GS} data for a raised source versus planar source. I_{DS} increases with raised source, particularly at lower N_{SOURCE}.

A raised source TFET cross-sectional schematic is shown in Figure 3.6a. The effect of a raised source region is shown in Figure 3.6b. A raised source structure is effective for increasing I_{ON}, moreso at lower doping levels. This structure utilizes a raised but not elevated source. An elevated raised source has been shown to improve performance more dramatically [22].

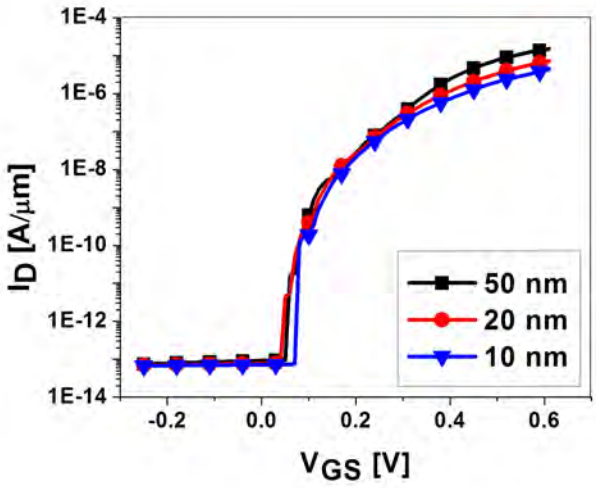


Figure 3.7: Simulated I_{DS}-V_{GS} data for varying Gate-Source overlap. Source doping N_{SOURCE} is 10^{19} cm^{-3}.

Figure 3.7 shows the impact of gate-to-source overlap. For the vertical tunneling design, a larger overlap increases the tunneling area and hence increases I_{ON}.

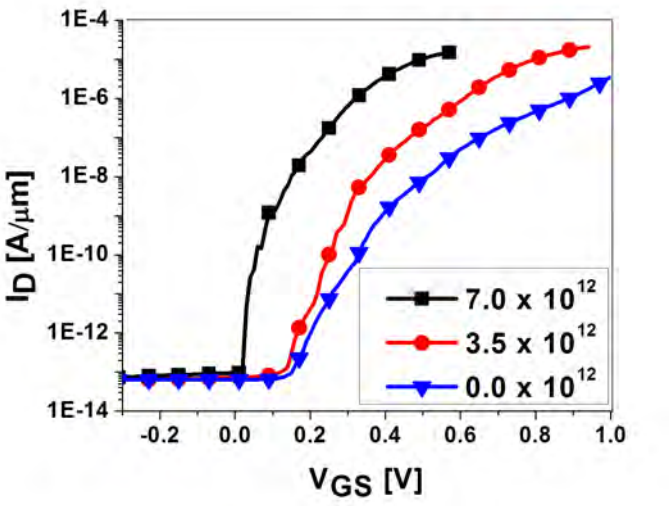


Figure 3.8: Simulated I_{DS}-V_{GS} data for varying Q_F interface charge in cm^{-2}.

Figure 3.8 shows the impact of fixed charge at the Ge-SiO_2 gate dielectric interface. Increasing Q_F shifts V_T negatively and increases I_{DS} for a given V_{GT}.

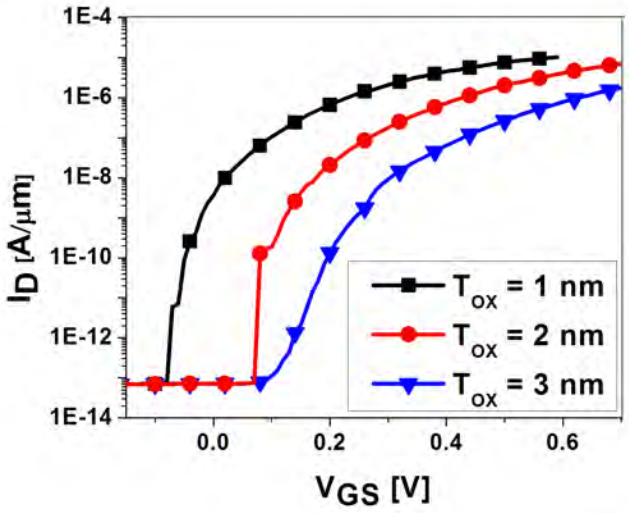
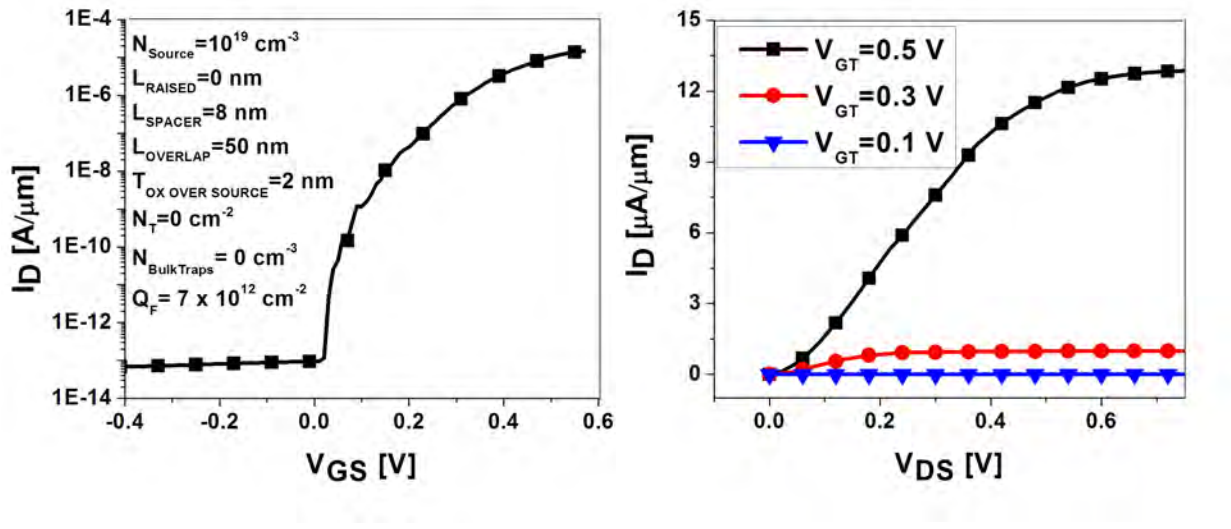


Figure 3.9: Simulated I_{DS}-V_{GS} data for varying T_{OX} over the Germanium source.

Figure 3.9 shows that scaling of the gate-oxide thickness over the source region is effective for improving I_{ON}, particularly in vertical tunneling devices.



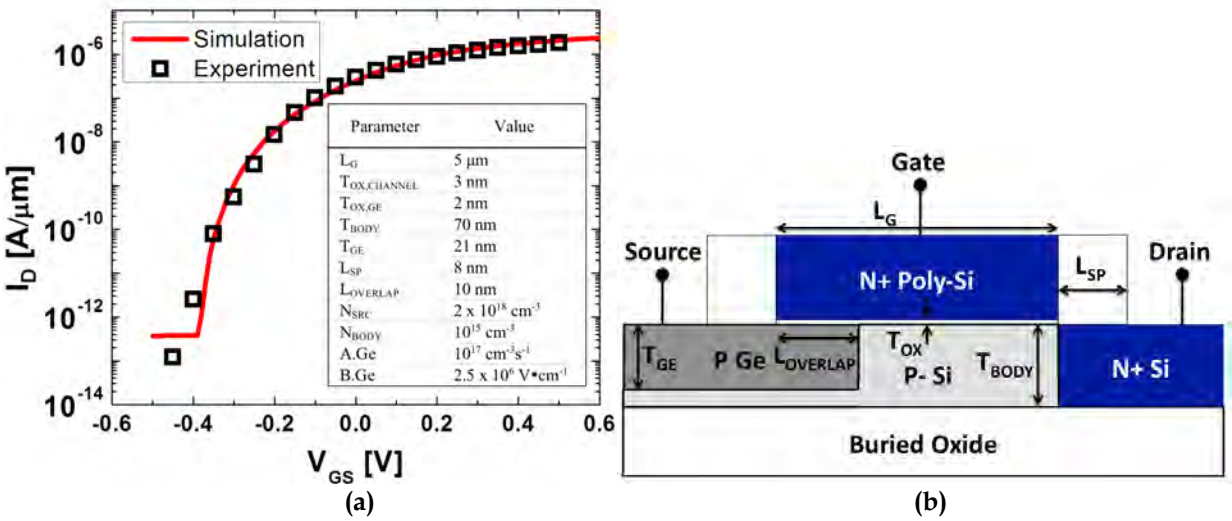
(a) **(b)**

Figure 3.10: An optimized device with I_{ON}/I_{OFF} of 10^7 and I_{ON} of over 10^{-5} A/μm at $V_{GT} = V_{DS} = 0.5$ V. I_D-V_{GS} and I_D-V_{DS} curves shown in (a) and (b) respectively.

The simulated I-V characteristics of an optimized Ge-source TFET are shown in Figure 3.10**.** I_{ON}/I_{OFF} exceeds 10^7 with I_{ON} > 10 μA/μm.

3.7 Advanced Calibration

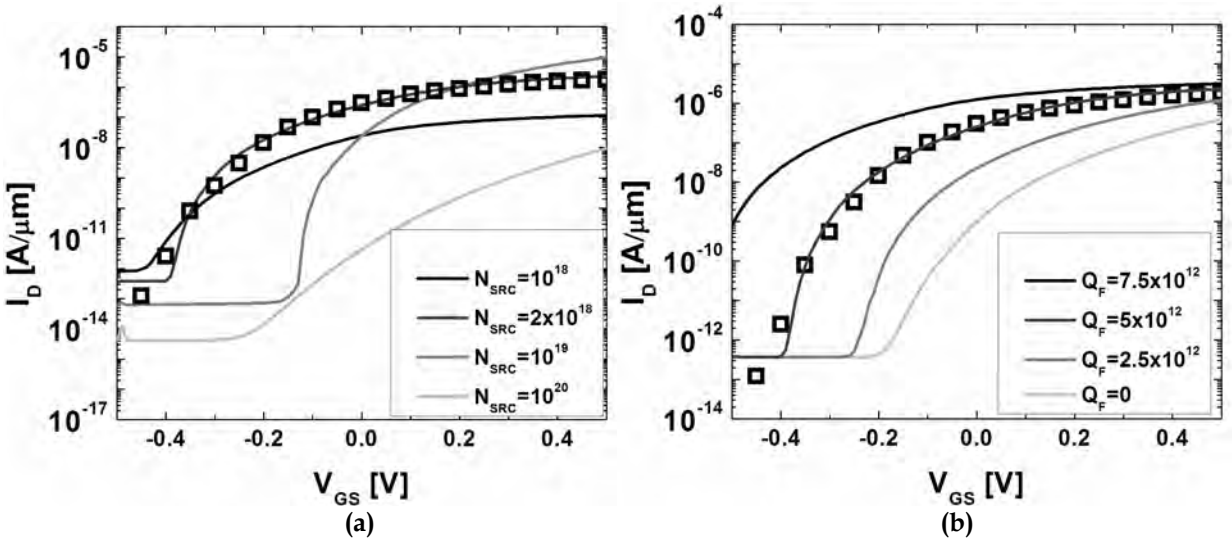
Following the initial calibration, several models were adjusted in the simulation. First, although the initial experimental device had a source doping concentration in the 10^{18} cm^{-3} range, the optimized device has a doping concentration in the 10^{19} cm^{-3}, necessitating the use of Fermi statistics. In addition, a study of reduced body thickness on SOI led to the use of the Density Gradient Model for quantization versus the previous use of the Modified Local Density Approximation (MLDA) model. The calibration was then performed again with these new models in place.



(a) (b)

Figure 3.11: (a) Simulated and measured I_D-V_{GS} characteristics of a Ge-source TFET for V_{DS} = 0.5 V. The device design parameter values used for the simulation, based on the device reported in Ref. [2], are shown in the inset. (b) Schematic cross-section of the Ge-source TFET structure studied in this work.

In calibrating the tunneling model, several device design parameters were adjusted. In each of the figures in this section, the device design parameters are as indicated in the inset of Figure 3.11a, unless otherwise noted. Figure 3.11a shows the simulated and measured experimental data are well calibrated. Figure 3.11b shows a cross-sectional schematic of the device simulated.



(a) (b)

Figure 3.12: (a) Simulated I_D-V_{GS} characteristics for various Ge source doping concentrations (cm^{-3}) and Q_F = 5x10^{12} q/cm^2. (b) Simulated I_D-V_{GS} characteristics for various values of oxide fixed charge (q/cm^2). Other device parameters are as indicated in Figure 3.11a and V_{DS} = 0.5 V. Measured data are shown with symbols, for reference.

Fig. 3.12a shows that the TFET I_D-V_{GS} characteristic is very sensitive to N_{SOURCE}. Fig. 3.12b shows the impact of oxide fixed charge density, which was adjusted to fit the experimental data. Subthreshold swing and turn-on voltage each decrease with

increasing Q_F. This is because positive fixed charge increases the transverse electric field and thereby enhances band-to-band tunneling within the source region for a given gate overdrive. In contrast, for TFETs designed for lateral tunneling (*i.e.* TFETs which have a very heavily doped source region), oxide fixed charge decreases source-to-channel tunneling [23].

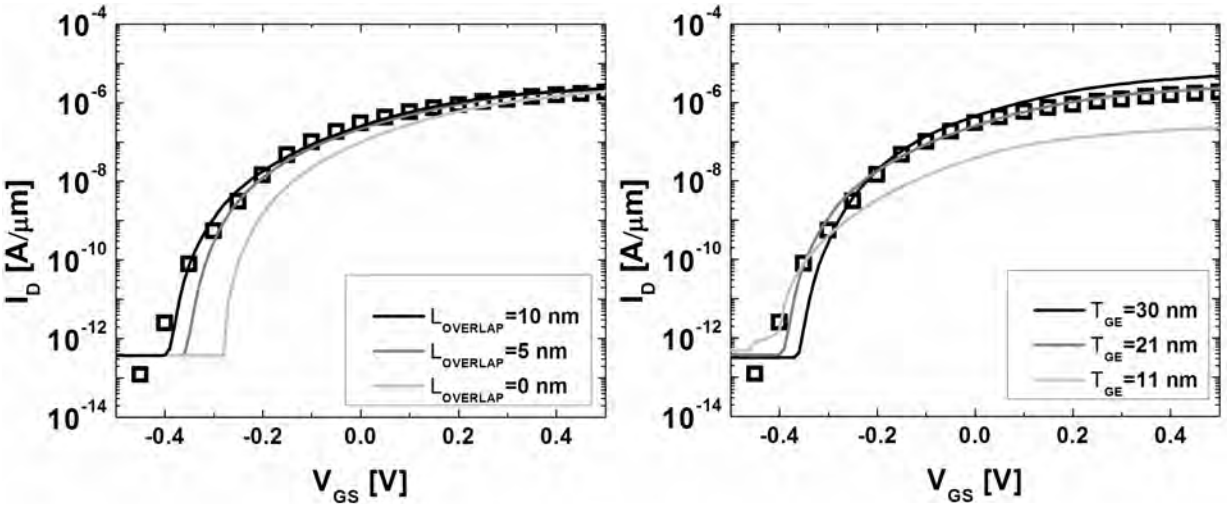


Figure 3.13: (a) Simulated I_D-V_{GS} characteristics for various gate-to-Ge-source overlap values. (b) Simulated I_D-V_{GS} characteristics for various Ge-source thickness values. Other device parameters are as indicated in Figure 3.11a and V_{DS} = 0.5 V. Measured data are shown with symbols, for reference.

The impacts of two geometrical design parameters, $L_{OVERLAP}$ and T_{GE}, are shown in Fig. 3.13. The primary effect of $L_{OVERLAP}$ is a shift in the device turn-on voltage. Increased T_{GE} is beneficial for improved performance, because it allows for a more vertical (orthogonal to the gate oxide/channel interface) tunneling across a shorter distance and over a larger area. The fit to measured data (21:10 vertical:lateral source dimension ratio) matches well with the experimentally observed 2:1 vertical:lateral etch-rate ratio for the etch process that was used to recess the silicon in the source region prior to selective deposition of Ge [2].

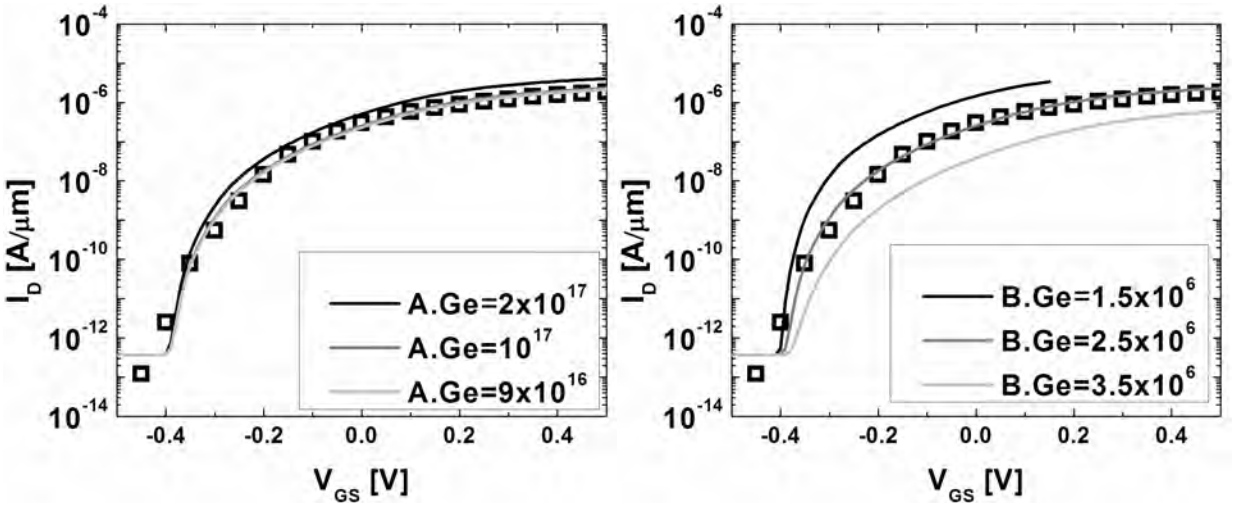


Figure 3.14: (a) Simulated I_D-V_{GS} characteristics for various Ge *A*-parameter values ($cm^{-3}s^{-1}$). (b) Simulated I_D-V_{GS} characteristics for various Ge *B*- parameter values (V/cm). Other device parameters are as indicated in Figure 3.11a and V_{DS} = 0.5 V. Measured data are shown with symbols, for reference. The starting parameter values derived using LEPM [14] are A.Ge = 9.11 x 10^{16} $cm^{-3}s^{-1}$ and B.Ge = 4.883 x 10^6 V/cm.

The values of the pre-exponential and exponential terms (*A* and *B* respectively) of the Keldysh model were also fitted to the measured data. The default values for Si and Ge, based on previous work using a local empirical pseudo-potential model (LEPM) [24], were used as a starting point. Modification of the *A* and *B* parameters for Si was found to have little effect on device performance. This is not surprising, since band-to-band tunneling occurs primarily within the Ge source region if it is not very heavily doped. The best-fit *A* and *B* values for Ge were found to be substantially different from the default values, as shown in Fig. 3.14. This is likely due to the fact that the Ge source is polycrystalline in the fabricated TFET. Polycrystalline Ge is known to have defect-related trap states that are energetically located approximately 0.1 eV above the valance band; these serve to effectively reduce the energy bandgap [25].

3.8 Advanced Design Optimization

Geometrical	Value	Electrical	Value
L_G	30 nm	ϕ_M	4.05 eV
T_{OX}	1 nm	Q_F	0 – 6.5 x 10^{12} cm^{-2}
T_{BODY}	50 nm	N_{BODY}	10^{18} cm^{-3}
T_{GE}	10 – 20 nm	N_{SRC}	10^{20}, 10^{19}, 10^{18} cm^{-3}
L_{SP}	16.5 nm	A.Ge	1.00 x 10^{17} $cm^{-3}s^{-1}$
$L_{OVERLAP}$	-2.5 – 10 nm	B.Ge	2.50 x 10^6 Vcm^{-1}

Table 3.3: TFET design optimization parameters

Table 3.3 lists the parameter values used for the source design optimization study in this section. The gate length (L_G) and equivalent gate oxide thickness (T_{OX})

were selected to be 30 nm and 1 nm, respectively, to be relevant for modern CMOS technology. The gate oxide is assumed to be HfO_2 (κ=23.4), with a physical thickness of 6 nm. Therefore, gate leakage is assumed to be negligible [26]. The thickness (T_{BODY}) and p-type dopant concentration (N_{BODY}) of the Si body were selected to be 50 nm and 10^{18} cm^{-3}, respectively, based on a previous body design optimization study [27]. For simplicity, the Si drain is assumed to be uniformly n-type doped (10^{20} cm^{-3}) and perfectly aligned to the edge of the gate electrode to avoid off-state gate-induced drain leakage. The source doping profile is assumed to be abrupt, since the *in-situ* boron doped poly-Ge is selectively deposited by a low-temperature (425°C) process [2]. A low operating voltage of 0.25 V, appropriate for ultra-low-power applications, is assumed.

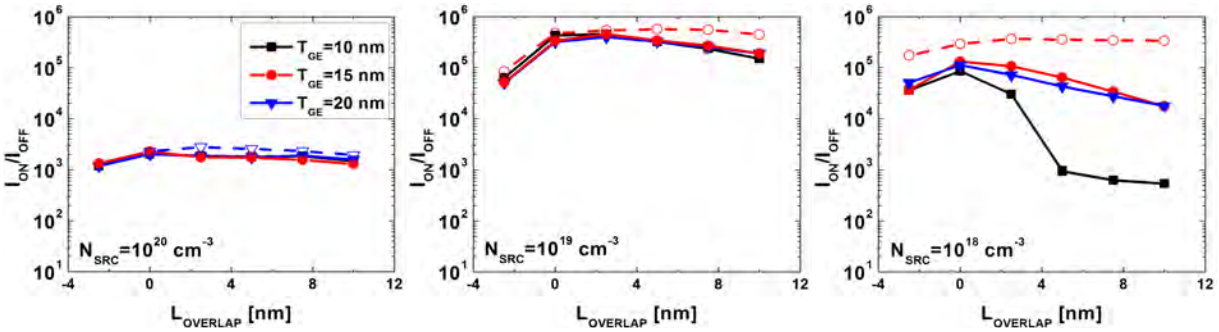
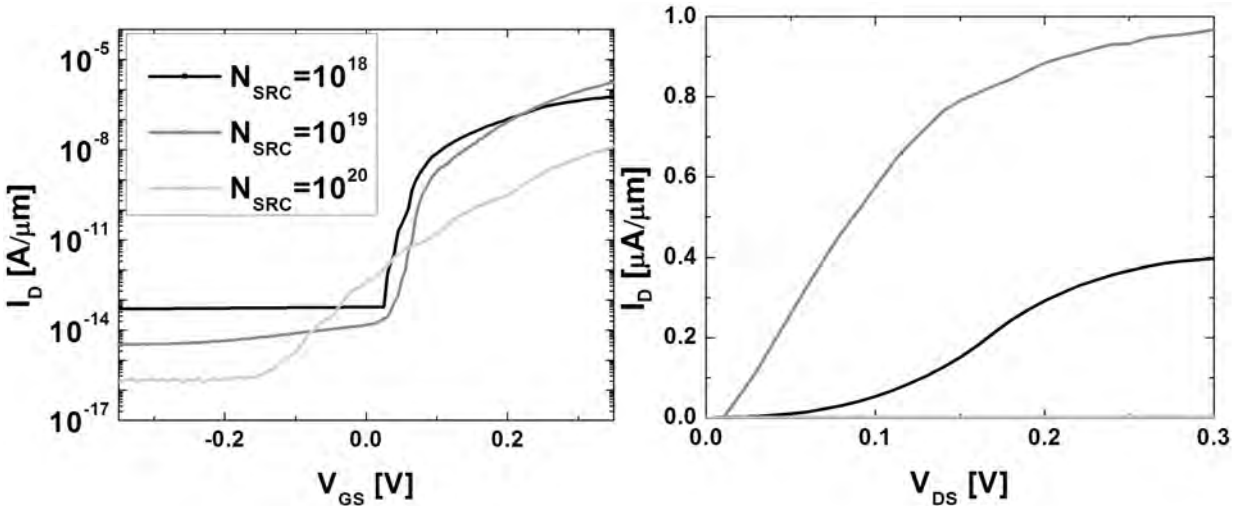


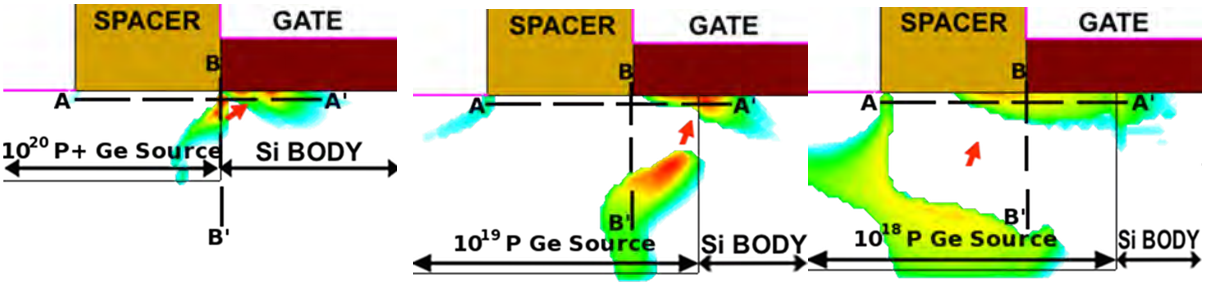
Figure 3.15: Simulated on/off current ratio as a function of gate-to-source overlap, for various Ge-source thickness values. I_{OFF} = 1 pA/mm and V_{DD} = 0.25V. (a) N_{SOURCE} = 10^{20} cm^{-3} (b) N_{SRC} = 10^{19} cm^{-3}, (c) 10^{18} cm^{-3}. Closed symbols with solid lines are for no fixed charge at the Ge/oxide interface; open symbols with dashed lines are for Q_F = 5×10^{12} q/cm^2. For simplicity, the impact of Q_F > 0 is shown only for the source thickness that yields the highest I_{ON}/I_{OFF}.

The gate voltage corresponding to a drain current (I_D) equal to the off-state leakage current (I_{OFF}) level of 1 pA/μm for V_{DS} = 0.25 V is defined to be the off voltage, V_{OFF}. I_{ON} is defined to be the drain current at V_G – V_{OFF} = 0.25 V, for V_{DS} = 0.25 V. Figure 3.15 shows how I_{ON}/I_{OFF} depends on $L_{OVERLAP}$ and T_{GE}, for different values of N_{SOURCE}. For very high N_{SOURCE} (10^{20} cm^{-3}), I_{ON}/I_{OFF} is rather low (< 7×10^3) and is relatively insensitive to $L_{OVERLAP}$ and T_{GE}. For moderate N_{SOURCE} (10^{19} cm^{-3}), I_{ON}/I_{OFF} is somewhat sensitive to $L_{OVERLAP}$ and relatively insensitive to T_{GE}; optimal performance (I_{ON}/I_{OFF} close to 10^6) is attained with $L_{OVERLAP}$ = 7.5 nm and T_{GE} = 20 nm. For low N_{SOURCE} (10^{18} cm^{-3}), I_{ON}/I_{OFF} is sensitive to $L_{OVERLAP}$ and T_{GE} if Q_F = 0, and it cannot match that which is achievable with N_{SOURCE} = 10^{19} cm^{-3}.



(a) (b)

Figure 3.16: (a) Simulated I_D-V_{GS} for V_{DS} = 0.25 V, for the optimal source design for various N_{SRC} values. (b) Corresponding I_D-V_{DS} for V_G-V_{OFF}= 0.25 V.



(a) (b) (c)

Figure 3.17: (a-c) Hole and electron tunneling contour plots at V_G-V_{OFF} = V_{DS} = 0.25 V for N_{SRC} = 10^{20}, 10^{19}, and 10^{18} cm^{-3}. The arrows are added to indicate the primary direction of tunneling.

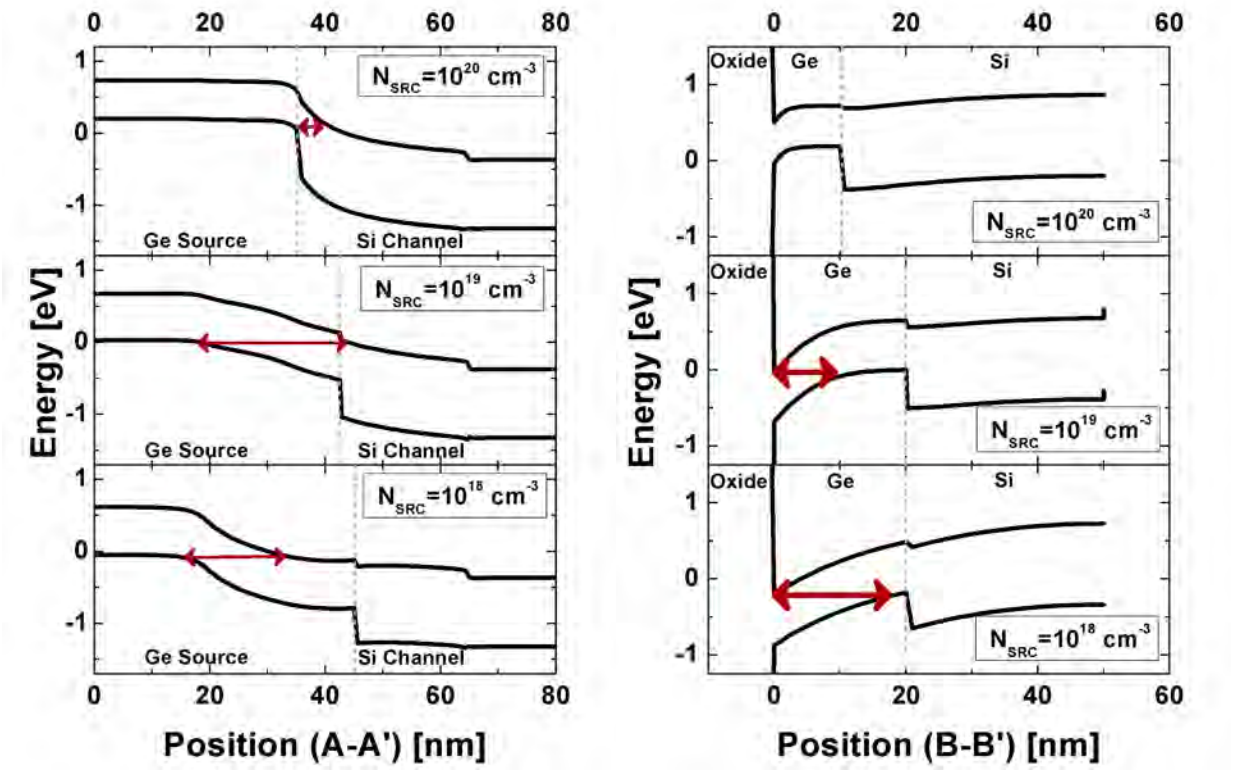


Figure 3.18: Simulated band diagrams along the (a) lateral and (b) vertical directions showing where tunneling occurs. Band diagrams were displayed for V_G-V_{OFF} = V_{DS} = 0.25 V.

To provide insight into the trends seen in Figure 3.15, simulated I_D-V_{GS} and I_D-V_{DS} characteristics, on-state tunneling contour plots, and band diagrams are shown in Figures 3.16, 3.17, and 3.18, respectively, for the optimal design (highest I_{ON}/I_{OFF}) for each value of N_{SRC}. For N_{SRC} = 10^{20} cm^{-3}, the source is not significantly depleted and tunneling occurs laterally from the Ge source into the Si channel, within a narrow region near to the gate-oxide interface. Since tunneling is primarily dependent on the inversion of the Si channel region, device performance is relatively insensitive to source design parameters. The subthreshold swing is large (see Fig. 3.12a and Fig. 3.17a) since Si has a lower tunneling rate due to its larger band gap.

For N_{SOURCE} = 10^{19} cm^{-3}, significant energy-band bending (*i.e.,* a potential drop) exists within the Ge-source so that tunneling can occur entirely within the Ge. By increasing T_{GE} to be larger than the depletion width (about 10 nm), tunneling within the Ge can occur in a more vertical direction, with a concomitant increase in the tunneling area. If tunneling occurs primarily vertically within the Ge (*i.e.,* for Q_F > 0), larger $L_{OVERLAP}$ is also beneficial for increasing the tunneling area (hence I_{ON}).

For N_{SOURCE} = 10^{18} cm^{-3}, the Ge-source region is more deeply depleted as compared to the higher N_{SOURCE} cases. Because the tunneling distance (*i.e.,* the width of the depletion region) is significantly larger, the tunneling probability is lower and hence I_{ON} is lower, although the tunneling area is larger.

From Fig. 3.12b it can be seen that a positive fixed charge at the Ge-dielectric interface can improve I_{ON}. This is because it serves to increase the vertical component of the electric field in the Ge source region, which helps to invert the source and thereby increases the area over which tunneling occurs.

These results indicate that there is an optimal value of N_{SOURCE}, ~10^{19} cm^{-3}, for which I_{ON}/I_{OFF} is maximized. This is in contrast to conventional Si homojunction and SiGe/Si heterojunction TFET designs, which have been shown by others to be optimized with very high source dopant concentration [11-14].

3.9 Conclusions

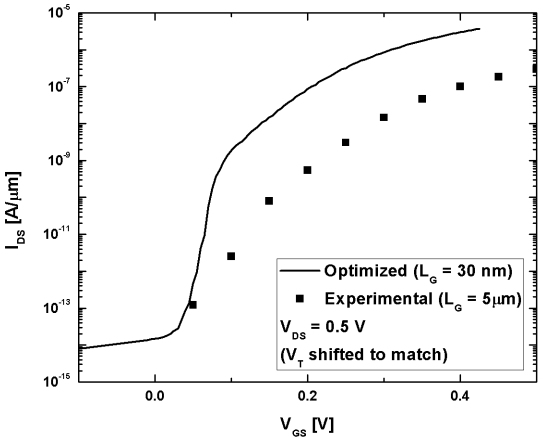


Figure 3.19: Simulated I_D-V_{GS} for V_{DS} = 0.5 V compared to the original experimental data, with turn-on voltages shifted to match.

A new non-local band-to-band tunneling simulation package calibrated to experimental data is used to conduct a source design optimization study for an n-channel TFET structure with a planar Ge source region. The optimal Ge source doping concentration is ~10^{19} cm^{-3}, and the Ge source should be at least 15 nm thick, to achieve I_{ON}/I_{OFF} > 10^5 for 0.25 V supply voltage with well-behaved output characteristics and tolerance for variability in $L_{OVERLAP}$ and T_{GE}. For this optimized design, tunneling occurs primarily within the Ge source, resulting in higher I_{ON}/I_{OFF} and improved transfer and output characteristics in comparison with other TFET designs. An optimized Ge-source TFET can well exceed the I_{ON}/I_{OFF} limit (10^4) of a MOSFET, so that it appears to be very promising for ultra-low-power applications. Figure 3.19 shows that compared to the experimental device [2], the optimized device has over one order of magnitude of increased on-state current.

3.10 References

1. P.F. Wang, K. Hilsenbeck, T. Nirschl, M. Oswald, C. Stepper, M. Weis, D. Schmitt-Landsiedel, and W. Hansch, "Complementary tunneling transistor for low power application," *Solid State Electron.*, vol. 48, no. 12, pp. 2281-2286, Dec. 2004.
2. S.H. Kim, H. Kam, C. Hu, and T.-J. K. Liu, "Germanium-Source Tunnel Field Effect Transistors with Record High I_{ON}/I_{OFF}," in *Symp. VLSI Tech. Dig.*, 2009, pp. 178-179.
3. S.E. Thompson, M. Armstrong, C. Auth, S. Cea, R. Chau, G. Glass, T. Hoffman, J. Klaus, Z. Ma, B. Mcintyre, A. Murthy, B. Obradovic, L. Shifren, S. Sivakumar, S. Tyagi, T. Ghani, K. Mistry, M. Bohr, and Y. El-Mansy., "A Logic Technology Featuring Strained-Silicon," *IEEE Electron Device Lett.*, vol. 25, no. 4, pp. 191-193, Apr. 2004.
4. F. Mayer, C. Le Royer, J.-F. Damlencourt, K. Romanjek, F. Andrieu, C. Tabone, B. Previtali, and S. Deleonibus, "Impact of SOI, $Si_{1-x}Ge_xOI$, and GeOI substrates on CMOS compatible Tunnel FET performance," *IEEE International Electron Devices Meeting*, pp. 163-166, 2008.
5. N. Patel, A. Ramesha, and S. Mahapatra, "Drive current boosting of n-type tunnel FET with strained SiGe layer at source," *Microelectronics Journal*, vol. 39, no. 12, pp. 1671-1674, Dec. 2008.
6. K. K. Bhuwalka, M. Born, M. Schindler, M. Schmidt, T. Sulima, and I. Eisele, "P-Channel Tunnel Field Effect Transistors down to Sub-50 nm Channel Lengths," *Japanese J. Appl. Phys.*, vol. 45, no. 4B, pp. 3106-3109, 2006.
7. K. K. Bhuwalka, J. Schulze, and I. Eisele, "Scaling the Vertical Tunnel FET With Tunnel Bandgap Modulation and Gate Workfunction Engineering," *IEEE Transactions on Electron Devices*, vol. 52, no. 5, pp. 909-917, May 2005.
8. Y. Khatami, and K. Banerjee, "Steep Subthreshold Slope n- and p-Type Tunnel-FET Devices for Low-Power and Energy-Efficient Digital Circuits," *IEEE Transactions on Electron Devices*, vol. 56, no. 11, pp. 2752-2761, Nov. 2009.
9. H. G. Virani, R. B. Rao, and A. Kottantharayil, "Investigation of Novel Si/SiGe Heterostructures and Gate Induced Source Tunneling for Improvement of p-Channel Tunnel Field-Effect Transistors," *Japanese J. Appl. Phys.*, vol. 49, pp. 04DC12, 2010.
10. W. G. Vandenberghe, A. S. Verhulst, G. Groeseneken, B. Sorée, and W. Magnus, "Analytical Model for Point and Line Tunneling in a Tunnel Field-Effect Transistor," in Proc. SISPAD, Sep. 2008, pp. 137-140.
11. C. Sandow, J. Knoch, C. Urban, Q.-T. Zhao, and S. Mantl. "Impact of electrostatics and doping concentration on the performance of silicon tunnel field-effect transistors," *Solid-State Electronics*, vol. 53, no. 10, pp.1126-1129, Oct. 2009.
12. O. M. Nayfeh, J. L. Hoyt, D. A. Antoniadis, "Strained-$Si_{1-x}Ge_x$ / Si Band-to-Band Tunneling Transistors: Impact of Tunnel-Junction Germanium Composition and Doping Concentration on Switching Behavior," *IEEE Transactions on Electron Devices,*

vol. 56, no. 10, pp. 2264-2269, Oct. 2009.
13. M. Fulde, A. Heigl, M. Weis, M. Wirnshofer, K. v. Arnim, Th. Nirschl, M. Sterkel, G. Knoblinger, W. Hansch, G. Wachutka, and D. Schmitt-Landsiedel, "Fabrication, Optimization and Application of Complimentary Multiple-Gate Tunneling FETs," *IEEE International Nanoelectronics Conference*, 2008, pp. 579-584.
14. P.-F. Wang, T. Nirschl, D. Schmitt-Landsiedel, and W. Hansch, "Simulation of the Esaki-tunneling FET," *Solid State Electron.*, vol. 47, no. 7, pp. 1187-1192, July 2003.
15. E. O. Kane, "Theory of Tunneling," *J. Appl. Phys.* vol. 32, no. 1, pp. 83-91, Jan. 1961.
16. L.V. Keldysh, *Sov. Phys. JETP*, vol. 6, pp. 665, 1958.
17. L.V. Keldysh, "Behaviour of Non-Metallic Crystals in Strong Electric Fields," *Sov. Phys. JETP*, vol. 6, pp. 763-770, 1958.
18. Medici User Guide, Synopsys Inc., 2010.
19. Sentaurus Device User Guide, Synopsys Inc., 2009.
20. A. Schenk, "Rigorous Theory and Simplified Model of the Band-to-Band Tunneling in Silicon," *Solid-State Electron.*, vol. 36, no. 1, pp. 19-34, 1993.
21. G. A. M. Hurkx, D. B. M. Klaassen, and M. P. G. Knuvers, "A New Recombination Model for Device Simulation Including Tunneling," *IEEE Transactions on Electron Devices,* vol. 39, no. 2, pp. 331–338, 1992.
22. S. H. Kim, S. Agarwal, Z.A. Jacobson, P. Matheu, C. Hu, and T.-J. K. Liu, "Tunnel Field Effect Transistor With Raised Germanium Source," *IEEE Electron Device Lett.*, vol. 31, no. 10, pp. 1107-1109, Oct. 2010.
23. X. Y. Huang, G. F. Jiao, W. Cao, D. Huang, H. Y. Yu, Z. X. Chen, N. Singh, G. Q. Lo, D. L. Kwong, and Ming-Fu Li. "Effect of Interface Traps and Oxide Charge on Drain Current Degradation in Tunneling Field-Effect Transistors," *IEEE Electron Device Lett.*, vol. 31, no. 8, pp.779-781, 2010.
24. D. Kim, T. Krishnamohan, L. Smith, H.-S. P. Wong, and K. C. Saraswat. "Band to Band Tunneling Study in High Mobility Materials: III-V, Si, Ge and strained SiGe," *Device Research Conference*, 2007, pp. 57-58.
25. H. Möller and V. Schlichting, Polycrystalline Semiconductors (Springer-Verlag), p. 331 (1989).
26. P. Sivasubramani, P. D. Kirsch, J. Huang, C. Park, Y. N. Tan, D. C. Gilmer, C. Young, K. Freeman, M. M. Hussain, R. Harris, S. C. Song, D. Heh, R. Choi, P. Majhi, G. Bersuker, P. Lysaght, B .H. Lee, H.-H. Tseng, J. S. Jur, D. J. Lichtenwalner, A. I. Kingon, and R. Jammy, "Aggressively Scaled High-k Gate Dielectric with Excellent Performance and High Temperature Stability for 32nm and Beyond," *IEEE International Electron Devices Meeting*, pp. 543-546, 2007.
27. S.H. Kim, Z.A. Jacobson, and T.-J. K. Liu, "Impact of Body Doping and Thickness on the Performance of Germanium-Source TFETs," *IEEE Transactions on Electron Devices.*, vol. 57, no. 7, pp.1710-1713, July 2010.

Chapter 4

Ultimate Scalability of the Raised Germanium Source TFET Design

4.1 Introduction

Tunnel field-effect transistors (TFETs) are of growing interest because of their potential energy-efficiency benefit for digital logic applications [1-3]. Various TFET designs have been shown theoretically and demonstrated experimentally to achieve sub-60 mV/dec subthreshold swing (S) at room temperature [4-7], which in principle allows them to achieve higher on/off current ratio (I_{ON}/I_{OFF}) than a MOSFET, for low operating voltage (V_{DD}). TFET designs can be broadly classified into two categories [8]: a "point tunneling" design, in which the source region is heavily doped and substantially aligned to the gate edge so that band-to-band tunneling (BTBT) in the on state occurs primarily from the source region to the channel region; and a "line tunneling" design, in which the source is moderately doped and substantially overlapped by the gate so that BTBT in the on state occurs primarily within the source region. Due to larger tunneling area, the line tunneling design provides for higher on-state drive current (I_{ON}) than the point tunneling design.

Since BTBT current increases exponentially with decreasing tunneling barrier height (*i.e.*, the effective energy band-gap), the use of a small-band-gap material in the tunneling region can substantially boost TFET I_{ON} [4,6,9]. Germanium (Ge) has a much smaller band-gap (E_g = 0.66 eV) than does silicon (E_g = 1.12 eV) and hence is advantageous as a tunneling material. Recently, a heterostructure TFET employing polycrystalline Ge (poly-Ge) in the moderately doped, gate-overlapped source region – a line tunneling design – was experimentally demonstrated with $I_{ON}/I_{OFF} \sim 3\times10^6$ and $I_{ON} \sim 1$ μA/μm, for 0.5 V operation [10]. Optimization of the planar Ge source dopant concentration and gate overlap distance is projected to yield >10× improvement in I_{ON}/I_{OFF}, so that Ge/Si heterostructure TFET technology can potentially provide for >10× lower energy consumption than MOSFET technology for digital logic circuits operating at clock frequencies up to ~100 MHz [11].

To be a compelling alternative switching device, the TFET should be at least as scalable as the MOSFET. The semiconductor industry technology roadmap projects that the minimum transistor gate length (L_G) will be less than 17 nm by the year 2015, and less than 8 nm by the year 2024 [12]. The impact of L_G scaling on TFET performance has been studied by several groups to date [13-17]. More importantly, however, the transistor length including contacts must continue to scale, in order to sustain the historic pace of increase in transistor density according to Moore's law [18]. For the planar line-tunneling TFET design, a reduction in the gate-to-source overlap with L_G scaling results in decreased tunneling area and hence decreased I_{ON}, which is undesirable. By elevating the Ge source region, as illustrated in Fig. 1, large gate-to-source overlap can be maintained with L_G scaling; also, this source design is projected to achieve higher I_{ON} and lower intrinsic delay than a planar source design [11]. In this work, the ultimate scalability of the raised-Ge-source heterostructure TFET is studied via two-dimensional (2-D) device simulation. DC and AC device performance are

projected for an optimized design with dimensions relevant to the end of the CMOS technology roadmap.

4.2 Simulation Approach

4.2.1 Device Structure

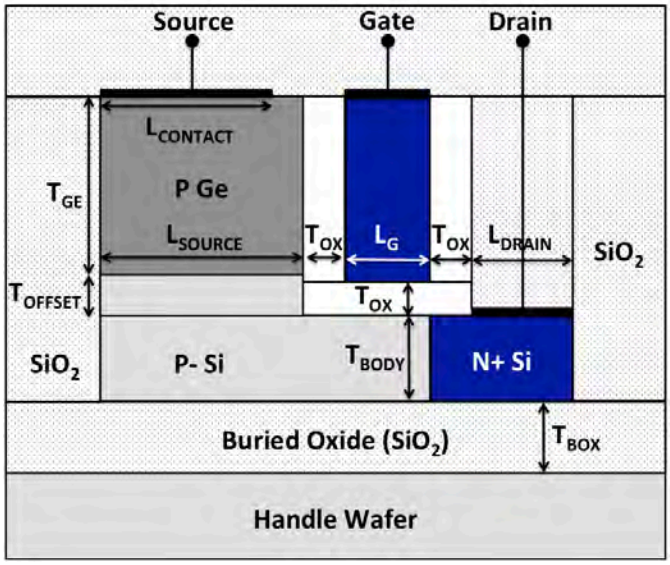


Figure 4.1: Schematic cross-section of the raised poly-Ge-source n-channel TFET design studied in this work.

Figure 4.1 shows a schematic cross-section of the n-channel TFET design investigated in this work, with various geometrical design parameters labeled. Note that the device length includes contacts and is defined as $L_{SOURCE} + 2\times T_{OX} + L_G + L_{DRAIN}$, where L_{SOURCE} and L_{DRAIN} are the lengths of the source and drain regions, respectively, and T_{OX} is the physical thickness of the gate dielectric. This structure can be fabricated by selectively doping the drain region n-type by masked ion implantation, then selectively growing in the source region a moderately doped p-type epitaxial silicon layer (of thickness T_{OFFSET}) followed by a more heavily doped p-type Ge layer (of thickness T_{GE} and dopant concentration N_{SOURCE}). The dopant concentration is assumed to change abruptly at the Ge/Si interface, since Ge can be grown at relatively low temperature (<450°C).

Parameter	Value
N_{SOURCE}	1.8 x 10^{19} cm^{-3}
N_{BODY}	2.7 x 10^{18} cm^{-3}
N_{DRAIN}	10^{20} cm^{-3}
L_{SOURCE}	14 nm
$L_{CONTACT}$	11 nm
T_{GE}	50 nm
T_{OFFSET}	2.5 nm
T_{BODY}	4.7 nm
T_{BOX}	10 nm
EOT	0.5 nm
L_G	4 nm
L_{DRAIN}	5 nm
$L_{DRAIN,\,OFFSET}$	0 nm

Table 4.1: Optimized TFET Design Parameter Values.

The dopant concentrations in the source, body, and drain regions are co-optimized for peak performance with geometrical design parameter values based on ITRS projections for fully depleted silicon-on-insulator (FDSOI) MOSFETs [12]; the optimized (default) values are summarized in Table 4.1. The gate oxide (of physical thickness T_{OX}) is assumed to have a dielectric constant of 23.4, roughly corresponding to HfO_2. The buried oxide thickness (T_{BOX}) is 10 nm to reduce drain-to-source coupling [19]. For simplicity, the drain region is assumed to be perfectly aligned to the gate edge; an offset-drain design is not necessary to achieve low off-state BTBT leakage current for the raised-source TFET design, in contrast to a planar-source TFET design [20-25]. Also, the top of the gate is assumed to be coplanar with the source contact, as simulations showed that a vertical offset between the gate and source contact has minor impact on device performance.

4.2.2 Models Used

Simulations were performed using Synopsys Sentaurus Device. Tunneling was modeled with the non-local BTBT model with dynamically determined tunneling paths [26]. The tunneling model includes direct and phonon-assisted tunneling processes for the calculation of carrier generation rates, and reduces to Kane's model [27] and Keldysh's model [28] in the uniform-electric-field limit. The tunneling model parameters were calibrated using measured current-*vs*.-voltage characteristics for the planar-Ge-source TFET in [10]: $A = 10^{17}$ $cm^{-3}s^{-1}$ and $B = 2.5\times10^{6}$ V-cm^{-1}.

Fermi statistics were assumed, and the doping dependence of Shockley-Read-Hall carrier recombination lifetime was included. For quantization effects, a density gradient quantization model was used because it is well suited for SOI devices and minimizes convergence issues commonly encountered when using the full Schrödinger model [28]. It should be noted that the quantization models only affect the carrier charge density, not the bandstructure, so that field-induced quantization effects [29] are not taken into account. (There may be some portion of the effect modeled since the

model parameters are calibrated to experimentally measured data). Airy function calculations show that there is a 0.14 eV offset due to quantum confinement; band-gap narrowing (by 0.06 eV) partially offsets this effect [30].

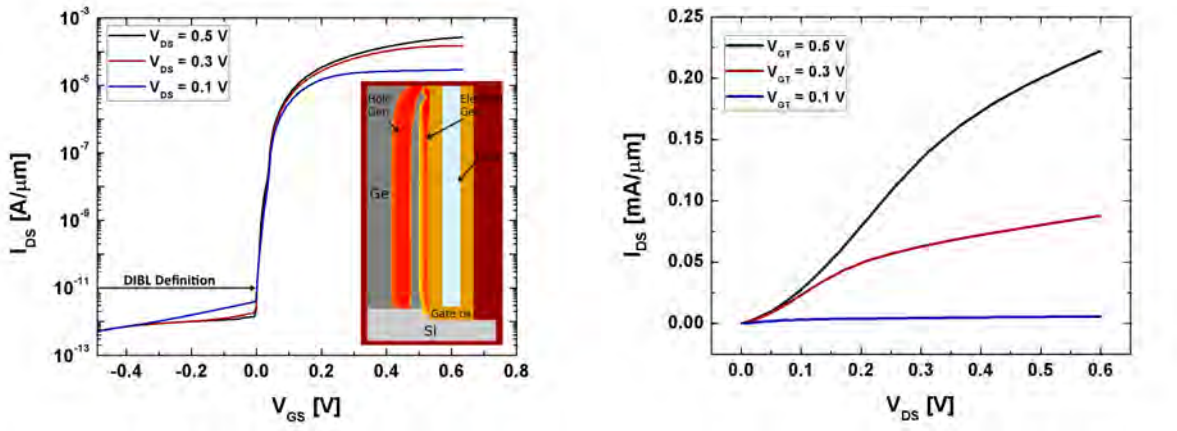
4.2.3 Modeling Assumptions

Gate leakage was assumed to be negligible, since high-permittivity dielectrics such as La_2O_3 that exhibit low direct tunneling gate current density (corresponding to < 1 pA/μm gate leakage current, for 0.5 nm equivalent oxide thickness) are anticipated to become available [31]. Positive fixed oxide charge (~8×10^{12} q/cm^2) is assumed to exist at each Ge/dielectric interface [32]. This charge serves to deplete the Ge near the isolation- and gate-dielectric interfaces, and thus limits the scalability of the source length, as discussed below. A distributed contact resistance of 10^{-8} Ω-cm^2 was assumed, consistent with ITRS projections [12].

The transistor saturation current I_{DSAT} (defined as the drain current I_D for $V_{GS} = V_{DS} = V_{DD}$) overestimates the effective drive current (I_{EFF}) for digital logic circuit operation, so that an estimation of inverter delay using I_{DSAT} is overly optimistic [33]. I_{EFF} for a TFET can be accurately calculated using a three-point model that takes into account drain-voltage overshoot due to large Miller capacitance, if the value of the total gate capacitance (C_{GG}) used to calculate inverter delay includes the Miller capacitance [34]. In this work, I_D for $V_{GS} = V_{DS} = 0.75\times V_{DD}$ was found to yield accurate estimations of inverter delay, when used together with the value of C_{GG} for $V_{GS} = V_{DS} = V_{DD}$, and is therefore taken to be I_{EFF} herein.

Finally, for capacitance calculations, mixed-mode simulations were used. Although the dynamic nonlocal tunneling model explicitly does not work with AC analysis (which was verified by seeing vastly different current values when run), these mixed-mode simulations were only used to calculate capacitance. This capacitance calculation is used with the DC simulation I-V values to calculate CV/I. By using the DC I-V values, the relative inaccuracies of the dynamic nonlocal tunneling model in AC analysis mode are reduced.

4.3 Results and Discussion



(a) **(b)**

Figure 4.2: Simulated (a) transfer and (b) output characteristics for the optimized raised-Ge-source TFET design (ref. Table 4.1). The inset in (a) shows BTBT-rate contours within the device, for V_{GS} = 0.5 V and V_{DS} = 0.5 V

Figure 4.2 shows the simulated transfer (I_D-V_{GS}) and output (I_D-V_{DS}) characteristics for the optimized raised-Ge-source TFET. Low off-state leakage current (approximately 1 pA/μm) and negligible effect of drain bias on the turn-on voltage are seen. The output characteristics show non-linear behavior at low V_{DS}, which is typical of TFETs [35].

4.3.1 Impact of Source Length and Source Doping

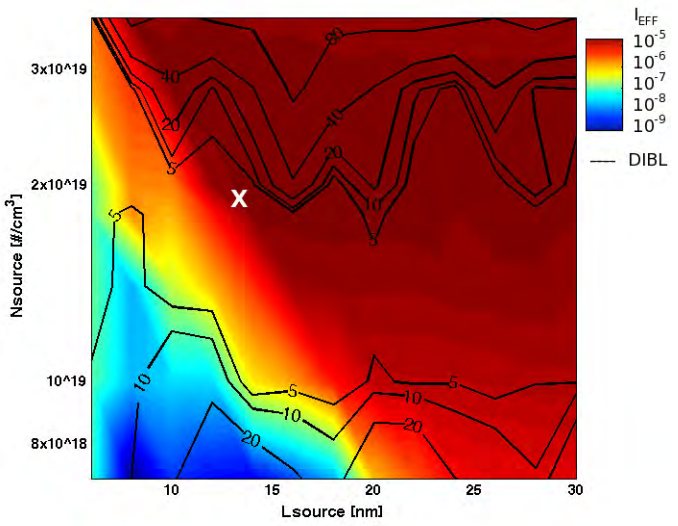


Figure 4.3: Contour plot showing the dependence of I_{EFF} (in A/μm) on source doping and source length, for V_{DD} = 0.25 V. Constant DIBL curves (in mV/V) are superimposed. Maximum I_{EFF} with DIBL < 5 mV/V is achieved at a minimum L_{SOURCE} = 14 nm and N_{SOURCE} = 1.8×10^{19} cm^{-3}.

To minimize the tunneling distance (for maximum on-state BTBT current), tunneling within the Ge source region should occur in a lateral direction that is largely transverse to the gate-dielectric interface, *i.e.* care should be taken to avoid fully depleting the Ge in the lateral direction. If L_{SOURCE} is scaled down such that the Ge becomes fully depleted in the on state, the effective drive current is reduced, as shown in Figure 4.3. Since the depletion width has an inverse square root dependence on N_{SOURCE}, L_{SOURCE} can be scaled down more aggressively for higher N_{SOURCE}.

Contour lines for constant drain-bias-induced voltage shift – analogous to drain-induced barrier lowering, or DIBL, in MOSFETs – are overlaid in Figure 4.3. DIBL is defined herein as the shift in gate voltage corresponding to an off-state leakage current of 10 pA/μm induced by a 1-Volt change in drain voltage, in units of mV/V. (DIBL is found by taking the gate voltage shift between simulated I_D-V_{GS} curves for V_{DS} = 0.05 V and V_{DS} = 0.55 V, and dividing by 0.5 V). For high values of N_{SOURCE}, the device turns on with point tunneling, which is more susceptible to the influence of the drain bias (because it affects the channel potential) and hence shows higher DIBL in the upper portion of Figure 4.3. Reverse-bias p/n-junction diode leakage increases with decreasing N_{SOURCE} and becomes dominant at low values of N_{SOURCE} (*e.g.* increased constant DIBL curves in the lower left region of Figure 4.3). A lower N_{SOURCE} increases the sensitivity of I_{EFF} on the drain bias, though with a lower net change in DIBL when compared to a high N_{SOURCE}. In both cases, at extremely low L_{SOURCE}, tunneling cannot occur within the source and DIBL drops dramatically. L_{SOURCE} = 14 nm with N_{SOURCE} =

1.8×10^{19} cm^{-3} is optimal for maximizing I_{EFF} while maintaining low DIBL. (In practice, a slightly larger value of L_{SOURCE} should be used to reduce variation in performance due to process-induced variations in L_{SOURCE}.)

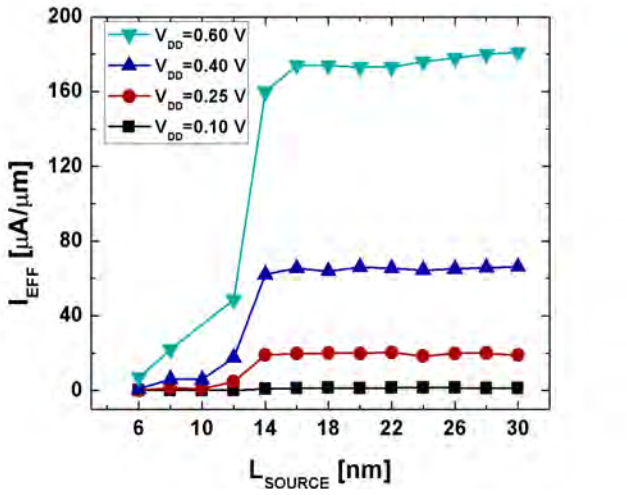


Figure 4.4: Dependence of I_{EFF} on L_{SOURCE}, for various values of V_{DD}. $N_{SOURCE} = 1.8\times10^{19}$ cm^{-3}. Note that I_{EFF} drops significantly for L_{SOURCE} below 14 nm.

Figure 4.4 confirms that the optimal minimum value of L_{SOURCE} does not change substantially across a wide range of V_{DD}.

4.3.2 Impact of Gate Length Scaling

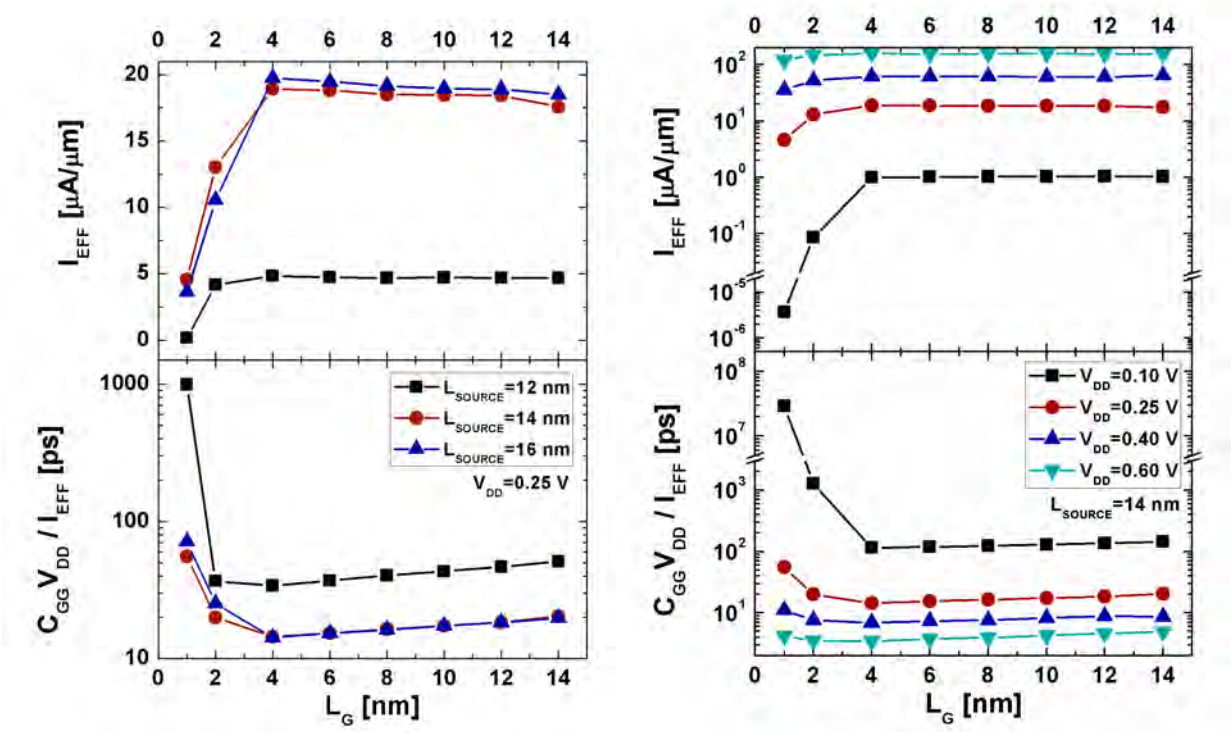


Figure 4.5: Gate-length dependence of I_{EFF} and $C_{GG}V_{DD}/I_{EFF}$ for (a) various values of L_{SOURCE} at V_{DD} = 0.25 V and (b) various values of V_{DD} at L_{SOURCE} = 14 nm. L_G scaling does not change the optimal L_{SOURCE} value or have a deleterious effect on performance for L_G > 4 nm.

Figure 4.5 shows the impact of the gate length on effective drive current and inverter delay, for various values of L_{SOURCE} and V_{DD}. Performance degrades when L_G is reduced to below 4 nm, because tunneling can occur directly from the source to drain in this regime.

4.3.3 Impact of Germanium Source Thickness and Vertical Offset

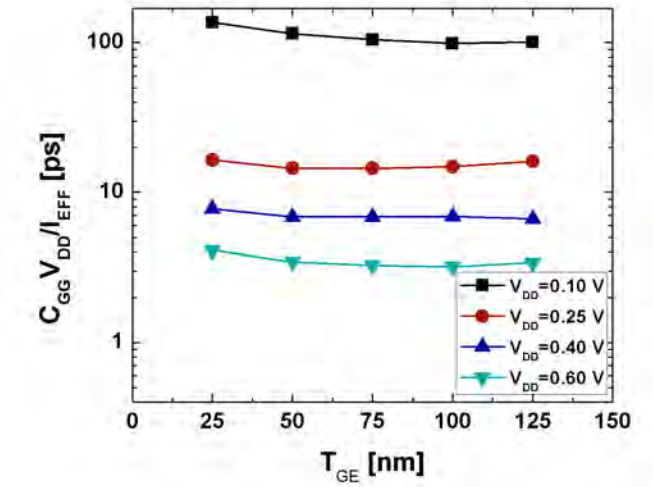
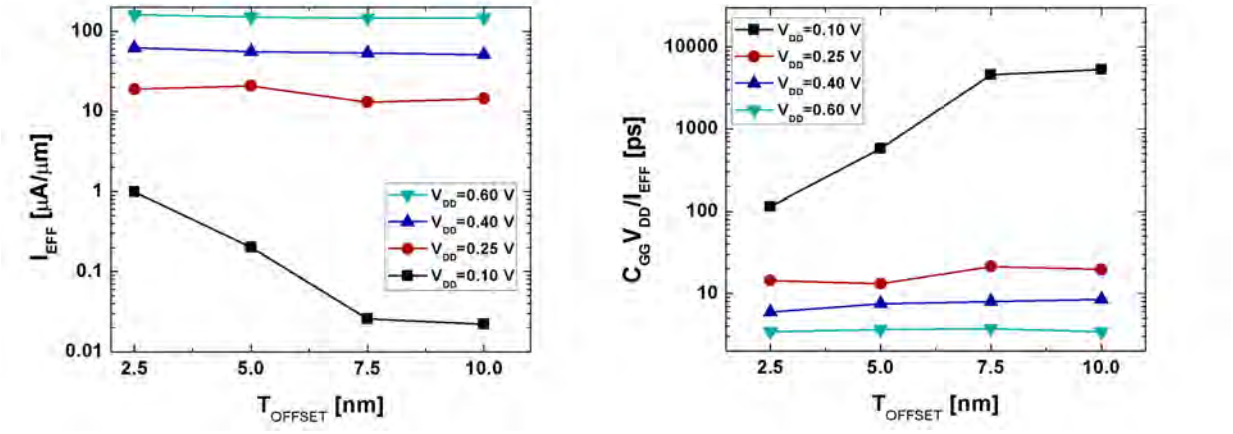


Figure 4.6: Dependence of $C_{GG}V_{DD}/I_{EFF}$ on raised-Ge-source thickness, for various values of V_{DD}. Delay generally reaches a minimum in the range 50 nm < T_{GE} < 100 nm, and increases rapidly as V_{DD} is lowered, due to the non-linear dependence on on-state resistance on V_{DS}.

Although the thickness of the raised Ge source and the thickness of the Si vertical offset region do not affect the scalability of the lateral dimensions of the TFET, they do affect the total gate capacitance, as well as the effective drive current. As can be seen from Figure 4.6, minimum delay is achieved in the range 50 nm < T_{GE} < 100 nm.



(a) **(b)**

Figure 4.7: Dependence of (a) I_{EFF} and (b) $C_{GG}V_{DD}/I_{EFF}$ on the thickness of the Si source-offset region, for various values of V_{DD}.

An increase in T_{OFFSET} results in increased channel resistance and hence reduced effective drive current, especially for low operating voltages, as shown in Figure 4.7.

Increased T_{OFFSET} also increases C_{GG} to compound the increase in delay. Decreased T_{OFFSET} reduces TFET performance due to increased influence of the drain bias resulting in more DIBL.

4.3.4 Impact of Equivalent Oxide Thickness

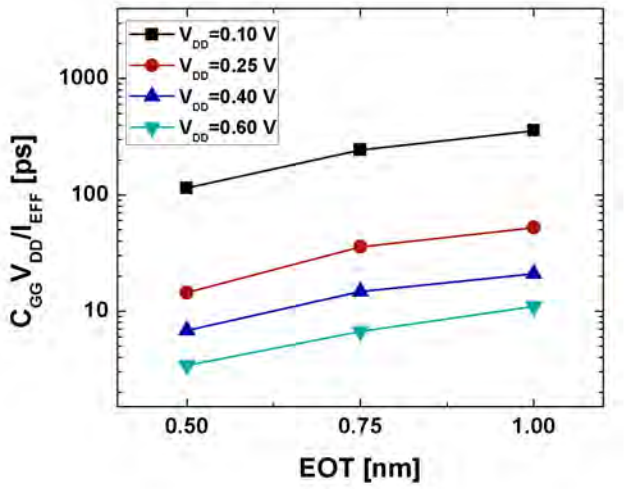


Figure 4.8: Dependence of $C_{GG}V_{DD}/I_{EFF}$ on equivalent gate-oxide thickness, for various values of V_{DD}.

Good capacitive coupling between the gate and the semiconductor at the gate-dielectric interface is critical to achieve steep switching behavior in a TFET, as previous simulation studies of T_{OX} dependence have shown [17]. For the line tunneling design, the voltage drop (*i.e.* the electric field) within the source must be maximized by minimizing EOT to achieve the highest BTBT rate in the on state. Figure 4.8 shows that TFET performance has an exponential dependence on EOT.

4.3.5 Impact of Body Thickness and Drain Parameters

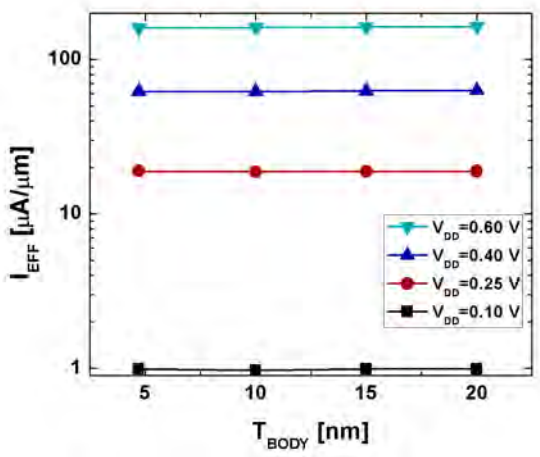
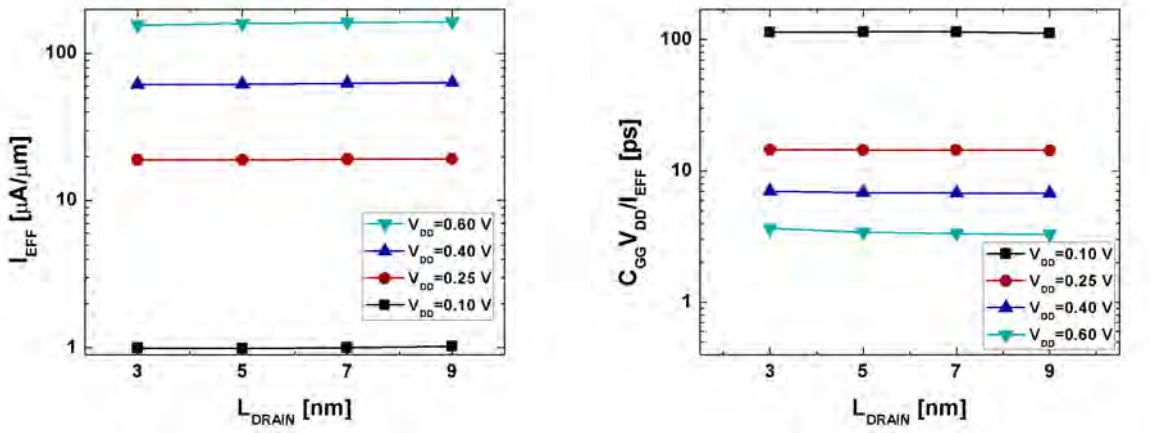


Figure 4.9: Dependence I_{EFF} on equivalent body thickness, for various values of V_{DD}.

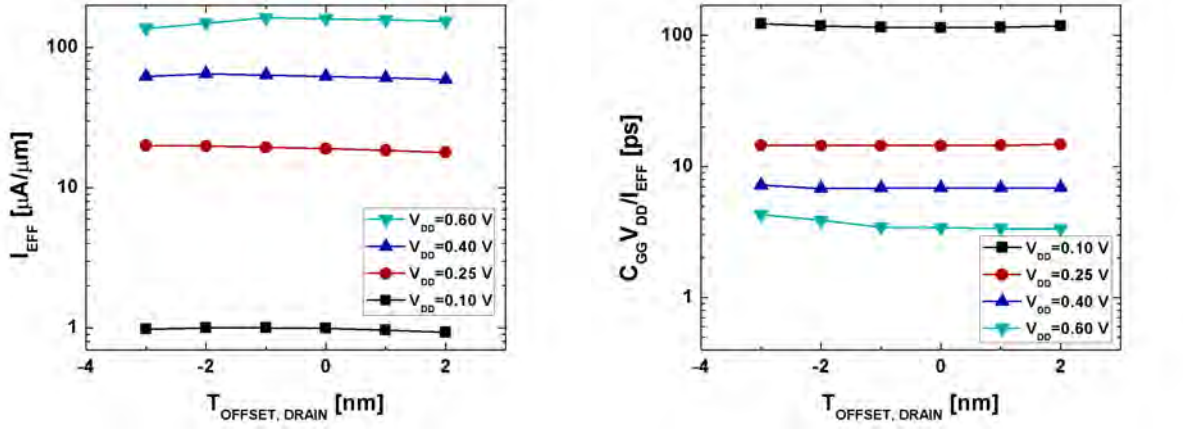
Further reduction in the body thickness was found to have little impact on device performance, because tunneling occurs completely within the Ge source region, as shown in Figure 4.9.



(a) **(b)**

Figure 4.10: Dependence of (a) I_{EFF} and (b) $C_{GG}V_{DD}/I_{EFF}$ on the length of the drain, for various values of V_{DD}.

As shown in Figure 4.10, the impact of drain length was also found to be minor: I_{EFF} decreases only slightly as L_{DRAIN} is further reduced, due to decreased drain contact area and hence increased drain contact resistance. L_{DRAIN} = 5 nm is taken to be a practical limit in this work.

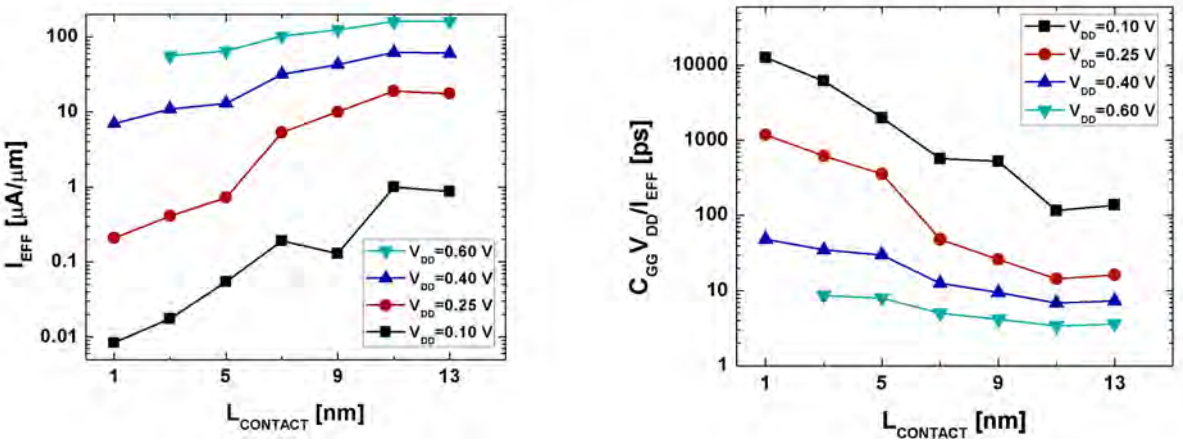


(a) (b)

Figure 4.11: Dependence of (a) I_{EFF} and (b) $C_{GG}V_{DD}/I_{EFF}$ on the drain offset, for various values of V_{DD}.

As shown in Figure 4.11, device performance was found to be insensitive to the gate-to-drain offset, except at high voltages where underlap of the gate can begin to lead to tunneling to the drain.

4.3.6 Impact of Source Contact Length



(a) (b)

Figure 4.12: Impact of source contact length on (a) I_{EFF} and (b) $C_{GG}V_{DD}/I_{EFF}$, for various values of V_{DD}. $L_{CONTACT}$ is optimized at 11 nm, *i.e.* the source contact ends 3 nm away from the (sidewall) gate-oxide interface.

To maximize the TFET drive current, BTBT should occur within the Ge source region (in the lateral direction) along the entire gate-dielectric interface, including at the topmost portion of the Ge source. If the contact were to be located along the entire top

surface of the source, the Ge would be unable to be inverted in its topmost region, resulting in decreased drive current. On the other hand, the source contact resistance increases as the length of the source contact ($L_{CONTACT}$) is reduced. As can be seen from Figure 4.12, $L_{CONTACT}$ = 11 nm is optimal for maximizing performance.

4.3.7 Energy-Delay Comparison

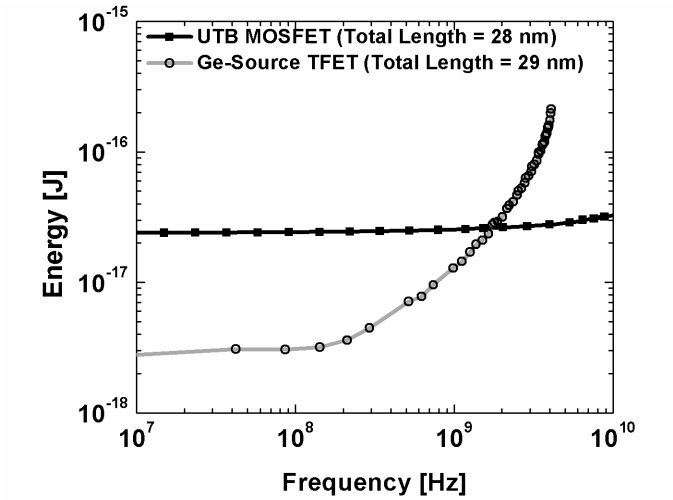


Figure 4.13: Comparison of total energy *vs*. frequency for 30-stage ring oscillators (activity factor α = 0.01, fan-out = 1) implemented with MOSFETs *vs*. TFETs. The cross-over point is at ~1.6 GHz.

Figure 4.13 compares the energy-delay performance of a Ge/Si TFET technology against that of an ultimately scaled FDSOI MOSFET technology (MOSFET device length with contacts = 28 nm) [ITRS], based on simulated DC current-*vs*.-voltage and AC capacitance-*vs*.-voltage characteristics [36]. The analysis assumed that p-channel devices for both TFET and FDSOI MOSFET are achievable with complementary characteristics. The TFET technology is projected to operate with >10× less energy at frequencies up to approximately 100 MHz due to its superior I_{ON}/I_{OFF} at low supply voltages. For operating frequencies above 1.6 GHz, the TFET technology loses its energy efficiency advantage because the MOSFET achieves higher I_{ON} at high supply voltages.

4.4 Conclusions

An optimized raised Ge-source TFET design can be scaled down to 29 nm device length with contacts without significant degradation in performance. As compared to a similarly sized ultra-thin-body MOSFET, the TFET is projected to be more energy efficient for digital logic circuits operating at frequencies up to 1.6 GHz.

4.5 References

1. P. F. Wang, K. Hilsenbeck, T. Nirschl, M. Oswald, C. Stepper, M. Weis, D. Schmitt-Landsiedel, and W. Hansch, "Complementary tunneling transistor for low power application," *Solid State Electronics*, vol. 48, no. 12, pp. 2281-2286, Dec. 2004.
2. Y. Khatami, and K. Banerjee, "Steep Subthreshold Slope n- and p-Type Tunnel-FET Devices for Low-Power and Energy-Efficient Digital Circuits," *IEEE Transactions on Electron Devices*, vol. 56, no. 11, pp. 2752-2761, Nov. 2009.
3. A. M. Ionescu and H. Riel, "Tunnel field-effect transistors as energy-efficient electronic switches," *Nature*, vol. 479, no. 7373, pp. 329-337, Nov. 2011.
4. F. Mayer, C. Le Royer, J.-F. Damlencourt, K. Romanjek, F. Andrieu, C. Tabone, B. Previtali, and S. Deleonibus, "Impact of SOI, $Si_{1-x}Ge_xOI$, and GeOI substrates on CMOS compatible Tunnel FET performance," *IEEE International Electron Devices Meeting*, pp. 163-166, 2008.
5. W. Y. Choi, B.-G. Park, J. D. Lee, and T.-J. K. Liu, "Tunneling Field-Effect Transistors (TFETs) With Subthreshold Swing (SS) Less Than 60 mV/dec," *IEEE Electron Devices Letters*, vol. 28, no. 8, pp. 743-745, Aug. 2007.
6. N. Patel, A. Ramesha, and S. Mahapatra, "Drive current boosting of n-type tunnel FET with strained SiGe layer at source," *Microelectronics Journal*, vol. 39, no. 12, pp. 1671-1674, Dec. 2008.
7. Y. Khatami, and K. Banerjee, "Steep Subthreshold Slope n- and p-Type Tunnel-FET Devices for Low-Power and Energy-Efficient Digital Circuits," *IEEE Transactions on Electron Devices*, vol. 56, no. 11, pp. 2752-2761, Nov. 2009.
8. W. G. Vandenberghe, A. S. Verhulst, G. Groeseneken, B. Sorée, and W. Magnus, "Analytical Model for Point and Line Tunneling in a Tunnel Field-Effect Transistor," in *Proc. SISPAD*, Sep. 2008, pp. 137-140.
9. O. M. Nayfeh, J. L. Hoyt, D. A. Antoniadis, "Strained-$Si_{1-x}Ge_x$ / Si Band-to-Band Tunneling Transistors: Impact of Tunnel-Junction Germanium Composition and Doping Concentration on Switching Behavior," *IEEE Transactions on Electron Devices,* vol. 56, no. 10, pp. 2264-2269, Oct. 2009.
10. S. H. Kim, H. Kam, C. Hu, and T.-J. K. Liu, "Germanium-Source Tunnel Field Effect Transistors with Record High I_{ON}/I_{OFF}," in *Symp. VLSI Tech. Dig.*, 2009, pp. 178-179.
11. S. H. Kim, S. Agarwal, Z.A. Jacobson, P. Matheu, C. Hu, and T.-J. K. Liu, "Tunnel Field Effect Transistor With Raised Germanium Source," *IEEE Electron Device Lett.*, vol. 31, no. 10, pp. 1107-1109, Oct. 2010.
12. International Technology Roadmap for Semiconductors (ITRS) 2009. http://public.itrs.net.
13. K. Boucart, A. M. Ionescu, "Length scaling of the Double Gate Tunnel FET with a high-K gate dielectric," *Solid-State Electronics*, vol. 51, no. 11-12, pp. 1500-1507, Nov.-Dec. 2007.

14. R. Woo, H.-Y. S. Koh, C. Onal, P. B. Griffin, and J. D. Plummer, "BTBT Transistor Scaling: Can they be Competitive with MOSFETs," *Device Research Conference*, 2008, pp. 75-76.
15. L. Liu and S. Datta, "Investigation of the scalability of ultra thin body (UTB) double gate tunnel FET using physics based 2D analytical model," *Device Research Conference*, 2010, pp. 15-16.
16. K. K. Bhuwalka, M. Born, M. Schindler, M. Schmidt, T. Sulima, and I. Eisele, "P-Channel Tunnel Field Effect Transistors down to Sub-50 nm Channel Lengths," *Japanese J. Appl. Phys.*, vol. 45, no. 4B, pp. 3106-3109, 2006.
17. K. K. Bhuwalka, J. Schulze, and I. Eisele, "Scaling the Vertical Tunnel FET With Tunnel Bandgap Modulation and Gate Workfunction Engineering," *IEEE Transactions on Electron Devices*, vol. 52, no. 5, pp. 909-917, May 2005.
18. G. G. Shahidi, "Device Scaling for 15 nm Node and Beyond," *Device Research Conference*, 2009, pp. 247-250.
19. T. Ernst, C. Tinella, C. Raynaud, and S. Cristoloveanu, "Fringing fields in sub-0.1 μm fully depleted SOI MOSFETs: Optimization of the device architecture," *Solid State Electron.*, vol. 46, pp. 373–378, Mar. 2002.
20. J. Wan, C. Le Royer, A. Zaslavsky, and S. Cristoloveanu, "SOI TFETs: Suppression of ambipolar leakage and low-frequency noise behavior," in *Proc. Of the European Solid-State Device Research Conference (ESSDERC)*, 2010, pp. 341-344.
21. C. Anghel, A. G. Hraziia, A. Amara, and A. Vladimirescu, "30-nm Tunnel FET With Improved Performance and Reduced Ambipolar Current," *IEEE Transactions on Electron Devices*, vol. 58, no. 6, pp. 1649-1654, June 2011.
22. W. Y. Choi, and W. Lee, "Hetero-Gate-Dielectric Tunneling Field-Effect Transistors," *IEEE Transactions on Electron Devices*, vol. 57, no. 9, pp. 2317-2319, Sept. 2010.
23. R. Jhaveri, V. Nagavarapu, and J. C. S. Woo, "Effect of Pocket Doping and Annealing Schemes on the Source-Pocket Tunnel Field-Effect Transistor," *IEEE Transactions on Electron Devices*, vol. 58, no. 1, pp. 80-86, Jan. 2011.
24. D. Kazazis, P. Jannaty, A. Zaslavsky, C. Le Royer, C. Tabone, L. Clavelier, and S. Cristoloveanu, "Tunneling field-effect transistor with epitaxial junction in thin germanium-on-insulator," *Appl. Phys. Lett.*, vol. 94, 263508, 2009.
25. T. Krishnamohan, D. Kim, S. Raghunathan, and S. Saraswat, "Double-Gate Strained-Ge Heterostructure Tunneling FET (TFET) With Record High Drive Currents and <60mV/dec Subthreshold Slope," in *IEEE International Electron Devices Meeting*, 2008.
26. Sentaurus Device User Guide, Synopsys Inc., 2009.
27. E. O. Kane, "Theory of Tunneling," *J. Appl. Phys.* vol. 32, no. 1, pp. 83-91, Jan. 1961.
28. L.V. Keldysh, "Behaviour of Non-Metallic Crystals in Strong Electric Fields," *Sov. Phys. JETP*, vol. 6, pp. 763-770, 1958.
29. W. G. Vandenberghe, B. Sorée, W. Magnus, G. Groeseneken, and M. V. Fischetti, "Impact of field-induced quantum confinement in tunneling field-effect devices," *Appl. Phys. Lett.*, vol. 98, 143503, 2011.
30. H. P. D. Lanyon and R.A. Tuft, "Bandgap Narrowing in Heavily Doped Silicon,"

IEEE International Electron Devices Meeting, pp. 316-319, 1978.
31. Y.-C. Yeo, T.-J. King, and C. Hu, "Direct tunneling leakage current and scalability of alternative gate dielectrics," *Appl. Phys. Lett.*, vol. 81, no. 11, pp. 2091-2093, Sept. 2002.
32. H. Shang, H. Okorn-Schimdt, J. Ott, P. Kozlowski, S. Steen, E.C. Jones, H.-S.P. Wong, and W. Hanesch, "Electrical Characterization of Germanium p-Channel MOSFETs," *IEEE Electron Device Lett.*, vol. 24, no. 4, pp. 242-244, Apr. 2003.
33. M. H. Na, E. J. Nowak, W. Haensch, and J. Cai, "The effective drive current in CMOS inverters," in *IEEE International Electron Devices Meeting*, pp. 121-124, Dec. 2002.
34. S. Mookerjea, R. Krishnan, S. Datta, and V. Narayanan, "Effective Capacitance and Drive Current for Tunnel FET (TFET) CV/I Estimation," *IEEE Transactions on Electron Devices*, vol. 56, no. 9, pp. 2092-2098, Sept. 2009.
35. A. Mallik, and A. Chattopadhyay, "Drain-Dependence of Tunnel Field-Effect Transistor Characteristics: The Role of the Channel," *IEEE Transactions on Electron Devices*, vol. 58, no. 12, pp. 4250-4257, Dec. 2011.
36. H. Kam, T.-J. K. Kiu, E. Alon, and M. Horowitz, "Circuit-level requirements for MOSFET-replacement devices," *IEEE International Electron Devices Meeting*, pp. 427-428, 2008.

Chapter 5

Comparison of Germanium Source Tunnel FET and Si MOSFET Technologies for Ultra-Low-Power Digital ICs

5.1 Introduction

Supply voltage (V_{DD}) scaling has not kept pace with transistor density scaling because metal-oxide-semiconductor field-effect transistors (MOSFETs) have off-state leakage current (I_{OFF}) that increases exponentially with decreasing threshold voltage (V_T), resulting in a growing tradeoff between circuit performance and power consumption [1]. High-performance and/or high-activity applications require large drive currents (high V_{DD} and/or low V_T) and hence must tolerate higher I_{OFF}, while low-power and/or low-activity applications require low I_{OFF} (high V_T and/or low V_{DD}) and hence must tolerate lower drive current (I_{ON}). As V_{DD} is scaled down to reduce the active energy consumed in performing a digital logic operation (proportional to the square of V_{DD}), a minimum value is reached for the total energy consumed per operation, due to increasing delay with decreasing V_{DD} (hence increasing static energy due to I_{OFF}) [2]. This minimum energy corresponds to the subthreshold regime of MOSFET operation ($V_{DD} < V_T$). To overcome this fundamental limit in energy efficiency, an alternative transistor design that can achieve lower I_{OFF} (hence lower static energy) than a MOSFET for a given I_{ON}, *i.e.* steeper subthreshold swing (S), is needed.

As discussed in previous chapters, the supply of carriers from the source region via band-to-band tunneling (BTBT) rather than thermionic emission (as in a MOSFET) has been proposed to provide for switching behavior that is steeper than the 60 mV/dec (at room temperature) limit of the MOSFET [3-4]. For example, an n-channel tunnel FET (TFET) has a p-type source region in which valence-band electrons tunnel into conduction-band states and subsequently drift to the drain under the influence of an applied drain-to-source voltage, in the on state (when the applied gate voltage is sufficiently high to cause the conduction band to overlap in energy with the valence band, with a narrow tunneling barrier width). In the off state, BTBT current is very low because the energy band-gap serves to cut off the high-energy tail of the valence electron distribution in the source. (Also, although the conduction band in the drain may overlap in energy with the valence band in the source, the barrier width is too large for significant BTBT to occur.) The steepness of the TFET transfer characteristic is limited by the steepness of the density-of-states distributions in the valence and conduction bands, and increases with tunneling area [5].

Many different TFET designs have been explored in recent years [6]. Since the tunneling rate is exponentially dependent on the tunneling barrier height (*i.e.* the band-gap energy E_G), semiconductor materials with a band-gap smaller than that of silicon (Si), such as germanium (Ge), and heterostructure material systems have been proposed to increase TFET I_{ON} [7-9]. In this work, the design and performance of integrated circuits comprising Si TFETs with an optimized Ge source design [10] are investigated via SPICE simulations using a calibrated Verilog-A model. To be a compelling alternative to the MOSFET, the TFET must be scalable down to sub-10 nm minimum feature size, *i.e.* end-of-roadmap [11] dimensions. Thus, this work compares TFETs

with MOSFETs at a device pitch of 29 nm.

Fundamental differences between a TFET and a MOSFET have implications for circuit design and performance. First, the inversion-layer channel in a TFET is directly coupled to the drain but not to the source, since the source is of opposite conductivity type. As a result, the gate-to-drain capacitance (*i.e.* Miller capacitance) is larger for the TFET [12]. Second, a TFET is asymmetric, with vastly different electrical characteristics when the source-drain p-n structure is reverse biased *vs.* forward biased. (When the source-drain structure is forward biased, gate control is limited.) Third, due to the very steep subthreshold swing, small variations in V_T can result in large variations in current. Finally, the optimal design for a heterostructure p-channel TFET may be very different than that for a heterostructure n-channel TFET, if the conduction-band offset not symmetric with the valence-band offset, as is the case for the Ge-Si material system.

The remainder of this chapter is organized as follows. First, the design optimization and modeling of Ge-source heterostructure TFETs is discussed. Next, the performance advantage of TFETs for digital logic circuits is demonstrated through ring oscillator simulations. Finally, the voltage scaling advantage of TFET technology for memory circuits is presented.

5.2 Germanium Source n-Channel TFET Design

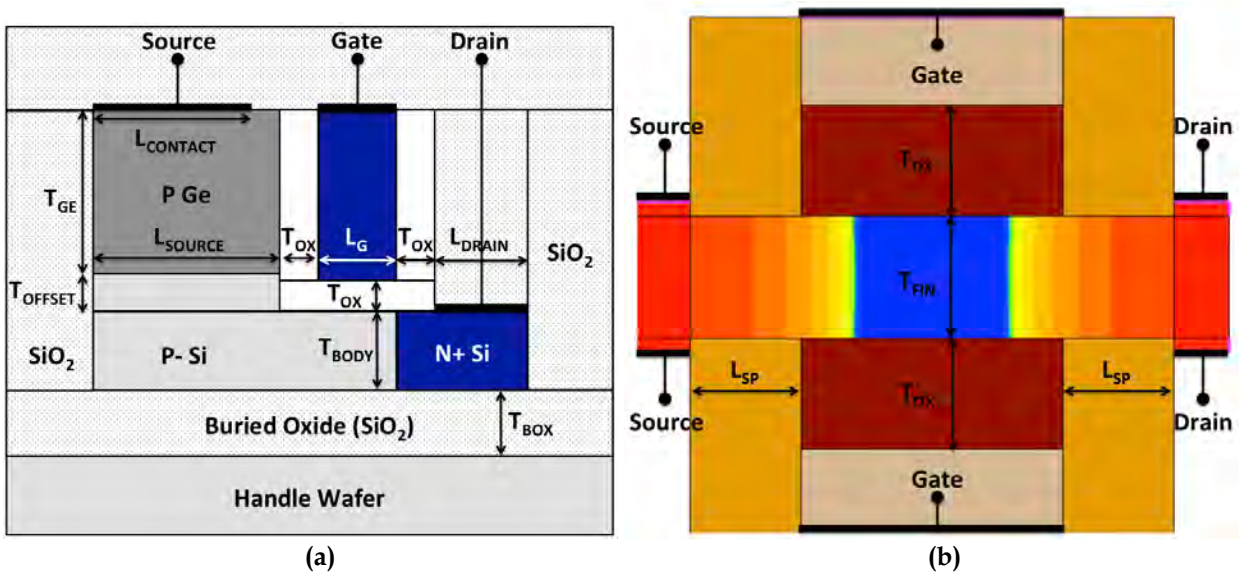
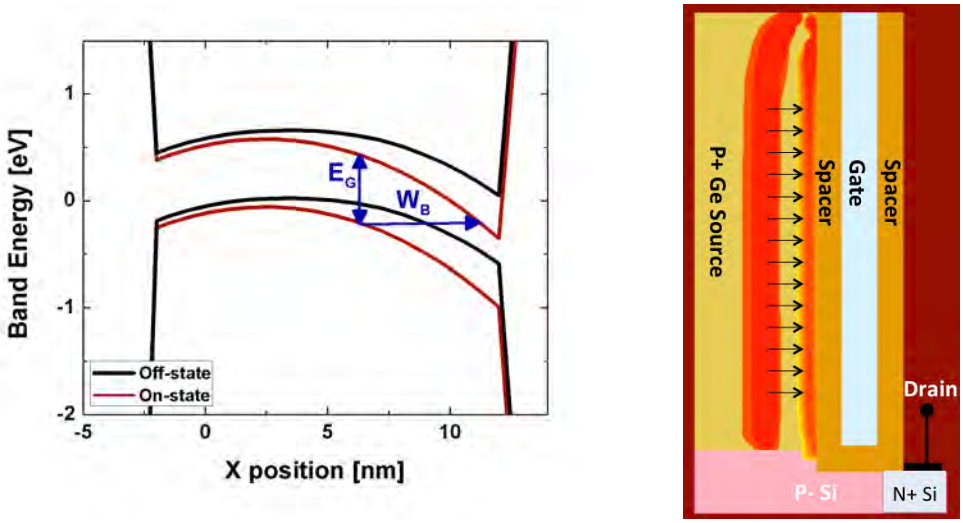


Figure 5.1: Schematic cross-section of the (a) Raised Ge-source TFET and (b) FinFET structure modeled in this study, with doping contours shown.

Figure 5.1 shows schematic cross-sectional views of the raised-Ge-source heterostructure TFET and FinFET (control device) structures studied in this work. The device designs were optimized for maximum performance. In the case of the FinFET, design parameters were taken from industry projections for the year 2018 [11]. The

device designs were each optimized for maximum performance and have the same device pitch. The TFET structure can in principle be fabricated using selective epitaxial growth of *in-situ*-doped Ge, which leverages embedded-source/drain stressor technology used to manufacture state-of-the-art p-channel MOSFETs today [13].



(a) (b)

Figure 5.2: (a) Energy band diagrams within the raised-Ge-source region of the Ge-source TFET. The energy band-gap (E_G) and barrier width (W_B) are key parameters for optimizing tunneling current. (b) On-state band-to-band tunneling rate contours showing that tunneling is confined within the Ge source region.

Figure 5.2(a) compares energy-band diagrams within the Ge source of the TFET, for the off state *vs.* the on state. The tunneling barrier energy (E_G) and barrier width (W_B) are labeled for the on state. Figure 5.2(b) is a contour plot of the tunneling rate for the on state, which shows that tunneling is confined to be within the (raised Ge) source region. The source tunneling design provides for larger on-state tunneling area and hence higher I_{ON} [5]. The raised source design allows a large overlap between the gate and the source to be achieved (for large on-state tunneling area) with a more compact device layout.

5.2.1 Device Modeling Approach

A look-up table is used in this work to perform the circuit analyses, since accurate analytical models for the raised-Ge-source heterostructure TFET are not yet available; the table is populated with data from technology computer-aided design (TCAD) simulation [14] of current-*vs.*-voltage (I-V) and capacitance-*vs.*-voltage (C-V) characteristics. This modeling approach has been shown to be simple but accurate for predicting circuit performance for exploratory devices [15].

Symbol	Description	Value
L_G	Gate length	4 nm
T_{OX}	Gate oxide thickness	3 nm
K	Gate oxide dielectric constant	23.4
$L_{CONTACT}$	Source contact length	11 nm
T_{GE}	Raised Ge thickness	50 nm
L_{SOURCE}	Source width	14 nm
N_{SOURCE}	Ge dopant concentration	1.8×10^{19} cm^{-3}
T_{OFFSET}	Si offset thickness	2.5 nm
T_{BODY}	Si body thickness	4.7 nm
N_{BODY}	Si body and offset dopant concentration	2.7×10^{18} cm^{-3}
L_{DRAIN}	Drain contact length	5 nm
N_{DRAIN}	Drain dopant concentration	10^{20} cm^{-3}
Q_F	Oxide fixed charge density at Ge interfaces	8×10^{12} q/cm^2
A.Ge	BTBT Ge tunneling coefficient	10^{17} $cm^{-3}s^{-1}$
B.Ge	BTBT Ge tunneling coefficient	2.5×10^{6} Vcm^{-1}

Table 5.1: Optimized TFET Design Parameter Values.

The TFET device simulations were conducted with Synopsys Sentaurus Device using the dynamic non-local BTBT simulation software package. The tunneling model parameters were calibrated for the Ge/Si heterostructure by fitting simulations to experimental data for a planar Ge-source TFET design, with geometrical design and material parameters derived from cross-sectional electron microscopy analyses, process data, and electrical measurements [10]. Positive fixed oxide charge is assumed to exist at the gate-oxide/Ge interfaces, as in [10]. Gate leakage current is assumed to be negligible, and depletion capacitance is assumed to be negligible because of the thin-body structures. The values of the TFET design parameters and the BTBT tunneling model parameters used in this work are summarized in Table 5.1 and are the same as the optimized parameters in Chapter 4.

5.2.2 Simulated Device Characteristics

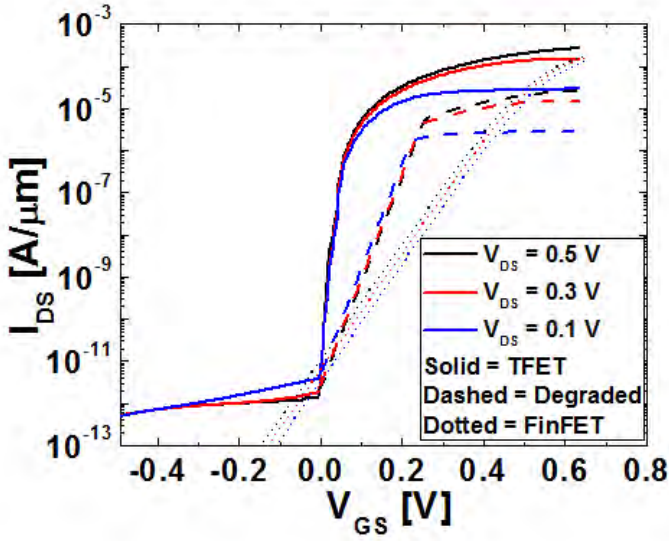


Figure 5.3: Simulated transfer characteristics for the raised Ge-source TFET, artificially degraded TFET, and FinFET structures, with V_{OFF} (at I_{OFF}=10 pA/ μ m) adjusted to be equal.

Figure 5.3 shows the transfer characteristics (I_{DS}-V_{GS}) of the TFET and FinFET devices. The TFET exhibits superior performance, with I_{ON} = 201 μA/μm at V_{DS} = 0.5 V, for I_{OFF} = 2 pA/μm. This is significantly better than published experimental results, as is the case for many simulation-based studies [6]. If full quantum simulation were used, the subthreshold swing likely would be degraded due to physical effects not included in the TCAD simulations, such as field-induced quantization and indirect phonon tunneling (see Chapter 6) [16]. To approximate the impact of these effects, an artificially degraded TFET that has a subthreshold swing of 40 mV/dec and I_{ON} that is only 10% of that for the optimized TFET is also considered in the circuit performance comparison. The transfer characteristics for the degraded TFET are shown with dashed lines in Figure 5.3(a).

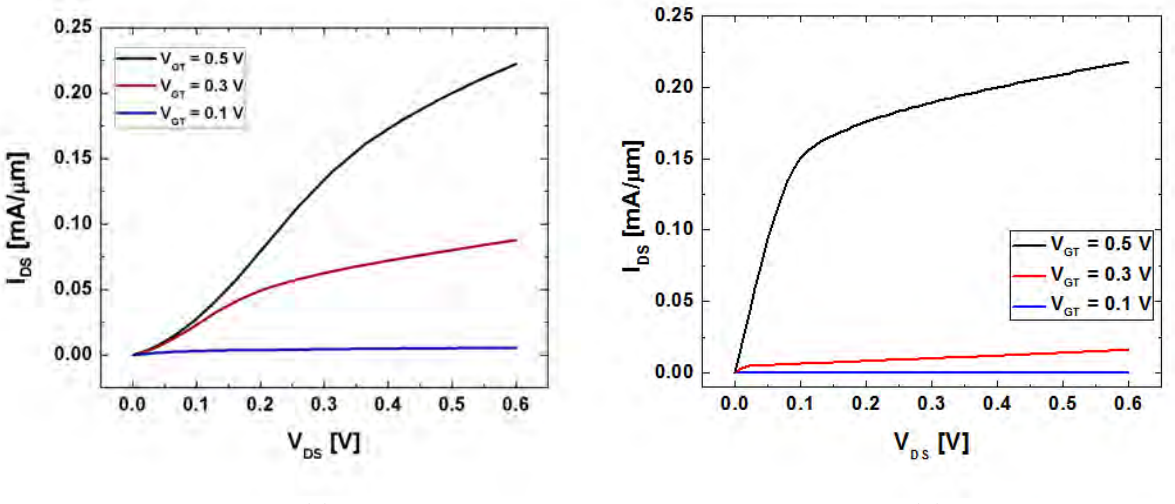


Figure 5.4: Simulated output characteristics of the (a) raised-Ge-source TFET and (b) FinFET structures.

Figure 5.4 shows the output characteristics (I_{DS}-V_{DS}) of the TFET and FinFET devices, respectively. For both devices, V_T is equal to V_{OFF} where I_{OFF} = 1 nA/μm. The TFET exhibits sublinear behavior at low V_{DS} with poor saturation, but with significantly higher current at low V_{GT}.

5.2.3 Circuit Design Considerations

5.2.3.1 Gate Capacitance

The inversion-layer charge in the channel region of a MOSFET can be supplied by either the source region or the drain region, so that the gate-to-channel capacitance is divided between the gate-to-source capacitance and the gate-to-drain capacitance. In contrast, the inversion-layer charge in the channel region of TFET is supplied only by the drain region, so that the gate-to-drain capacitance ("Miller capacitance") includes all of the gate-to-channel capacitance. A larger Miller capacitance results not only in longer propagation delay but also in greater transient voltage overshoot and undershoot for digital logic operation.

A raised-source design has greater gate-to-source capacitance (which increases with the thickness of the raised Ge source, T_{GE}) than does a planar source structure [10]. Thus, a design tradeoff for T_{GE} exists, since I_{ON} and gate capacitance both increase with T_{GE}. This tradeoff was found to be optimized for T_{GE} = 50 nm to minimize inverter propagation delay.

5.2.3.2 Asymmetric Current Flow

The TFET is designed to allow current flow in only one direction (charge carrier flow from the source to the drain) under the control of the gate voltage. Thus, a TFET cannot serve well as a pass transistor, *e.g.* an access device in a conventional 6-transistor (6T) static memory (SRAM) cell design.

5.2.3.3 Forward-biased Diode Current

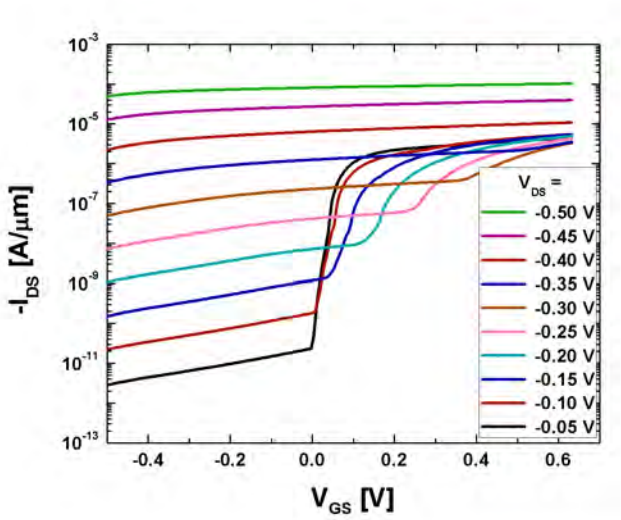


Figure 5.5: TFET transfer characteristics with negative drain voltage. When the p-type source is biased at a voltage higher than the n-type drain, the diode intrinsic to the TFET turns on. For sufficiently large V_{DS}, in this case, the gate cannot control the current flow.

Figure 5.5 shows the transfer characteristics of the n-channel Ge-source TFET for negative values of V_{DS} (p-type source biased at higher voltage than the n-type drain): for small values of V_{DS}, the forward-biased diode current can be suppressed by the gate voltage; however, for large values of V_{DS}, it cannot. Clearly, a TFET-based circuit should avoid the possibility of forward-biasing the source-drain diode structure.

5.2.3.4 V_T Variation

Because the TFET is designed to have a steep I_{DS}-V_{GS} characteristic, small variations in V_T can result in large variations in I_{ON} and I_{OFF} and hence propagation delay, noise margin and power consumption. Since S is steeper at lower values of current [17], these performance variations become more significant as V_{DD} is scaled down.

5.2.3.5 P-Channel Devices

Digital circuits require n-channel and p-channel transistors that operate in a perfectly complementary manner to minimize static power consumption. Due to the large valence-band offset between Ge and Si, holes cannot flow easily from Ge into Si. Thus, it may be difficult in practice to achieve a p-channel TFET with performance comparable to the n-channel Ge-source heterostructure TFET. Therefore, multiple scenarios are considered in this work for the p-channel device, including the use of p-channel Si MOSFETs for the pull-up devices.

5.3 Ring Oscillators

To assess the benefit of TFET technology for digital logic applications, a comparison of the performance *vs.* power tradeoff for 31-stage ring oscillators (ROs) with minimum-sized inverters and fan-out of 3 is conducted. For this analysis, p-channel device performance is assumed to mirror n-channel device performance for both TFET and MOSFETs. By varying V_{DD} and V_T (by adjusting the gate work function), the delay *vs.* switching energy tradeoff can be studied across a wide range of leakage power.

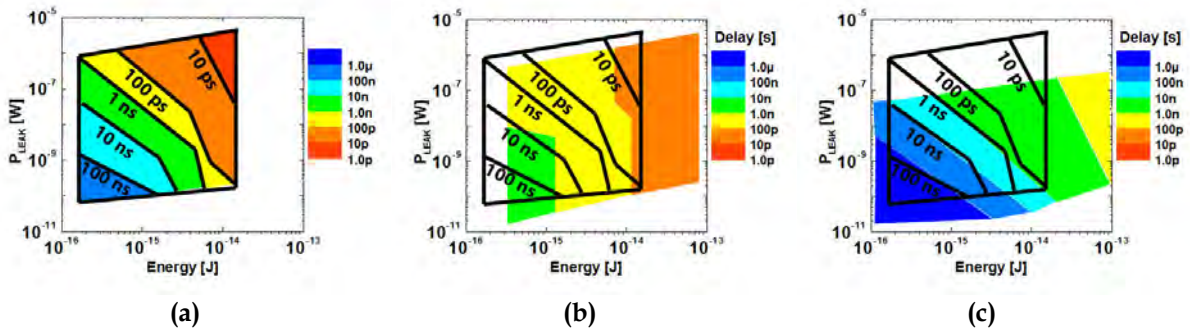


Figure 5.6: Contour plots of stage delay as a function of the leakage power and switching energy for ring oscillators implemented with (a) FinFETs, (b) raised-Ge-source TFETs, and (c) degraded TFETs.

Figure 5.6 shows contour plots of RO stage delay as a function of the leakage power and switching energy. The leakage power is equal to I_{OFF} x V_{DD} x Stages. The energy consumed is calculated by finding the current drawn per switching event and multiplying by the delay time and supply voltage. The delay is the stage delay. Consistent with the findings in [10], optimized Ge-source heterostructure TFETs are projected to provide for more energy-efficient operation (by more than 7×) than FinFETs for delays greater than 100 ps. However, the degraded TFET technology (Figure 5.6(c)) does not offer any advantage over FinFET technology, which indicates

that TFETs in practice must attain performance better than 10% of that for ideal Ge-source heterostructure TFETs, to be attractive for digital logic applications.

5.4 Memory Elements

5.4.1 SRAM Cells

5.4.1.1 6T SRAM Cell Design

The conventional 6T SRAM cell design comprises two cross-coupled inverters (each comprising an n-channel pull-down transistor and a p-channel pull-up transistor) and two pass transistors that control bit-line access to the complementary storage nodes for read and write operations. To read data, the bit-lines are pre-charged to V_{DD} and then the access transistors are turned on. If the storage node is at 0 V, then the voltage on the bit-line drops substantially; if the storage node is at the cell supply voltage (V_{SUPPLY}), then the voltage on the bit-line does not drop substantially. To reduce the possibility of upsetting the stored data during a read operation, the access transistor should not be stronger (*i.e.* have higher drive current) than the pull-down transistor; thus, the pull-down transistor is typically wider than the access transistor. To write data, the bit-lines are programmed and then the access transistors are turned on. A bit-line at 0 V will cause its storage node to be pulled down, if its access transistor is stronger than the pull-up transistor.

The unidirectional current flow characteristic of the TFET poses a problem for this cell topology because the access transistor must be able to conduct current from the bit-line to the storage node to pull down the bit-line voltage substantially during a read-'0' operation, and it also must be able to conduct current from the storage node to the bit-line to pull down the storage-node voltage during a write-'0' operation. In addition, forward-biased diode current can lead to disturbance of the stored data. For example, if an access TFET is oriented "inward" for current flow into the storage node (*i.e.* for writing a '0' into the storage node), then if a '1' is stored (*i.e.* the storage node is at V_{SUPPLY}) and the bit-line is at 0 V (*e.g.* to write a '0' into another cell sharing the same bit-line) the storage node will be discharged through the access TFET; if the storage node is pulled down sufficiently, the stored data can flip to '0'. For the opposite example in which the access TFET is oriented "outward" for current flow out of the storage node (*i.e.* for reading a '0'), if a '0' is stored and the bit-line is at V_{SUPPLY} (*e.g.* to write a '1' into another cell sharing the same bit-line) the storage node will be charged up through the access TFET; if the storage node is pulled up sufficiently, the stored data can flip to '1'.

5.4.1.2 Alternative SRAM Cell Designs

Alternative TFET-based cell designs have been proposed to overcome the

problems with the conventional 6T cell design. A 6T design using a virtual ground was proposed by Singh *et al.* [18]. The virtual ground assists write operation, but the read noise margins are very low unless the bit-lines are pre-charged to only one-half of V_{SUPPLY}. This cell design is not considered further herein.

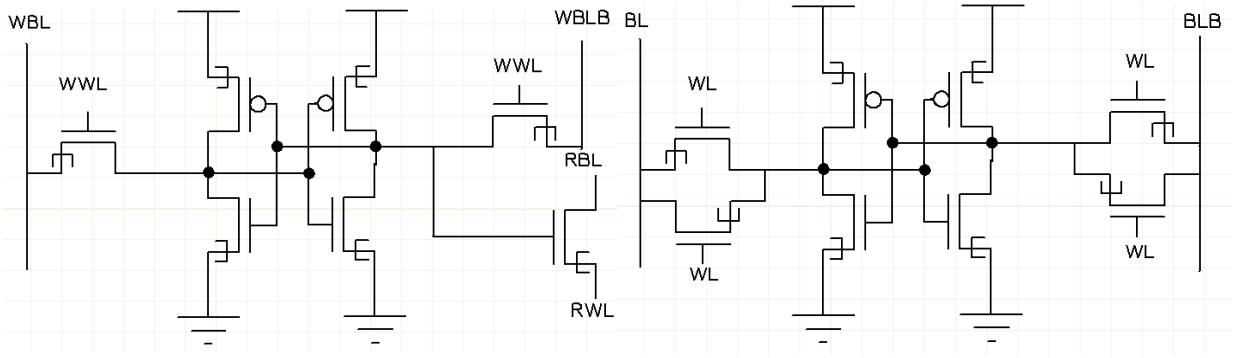


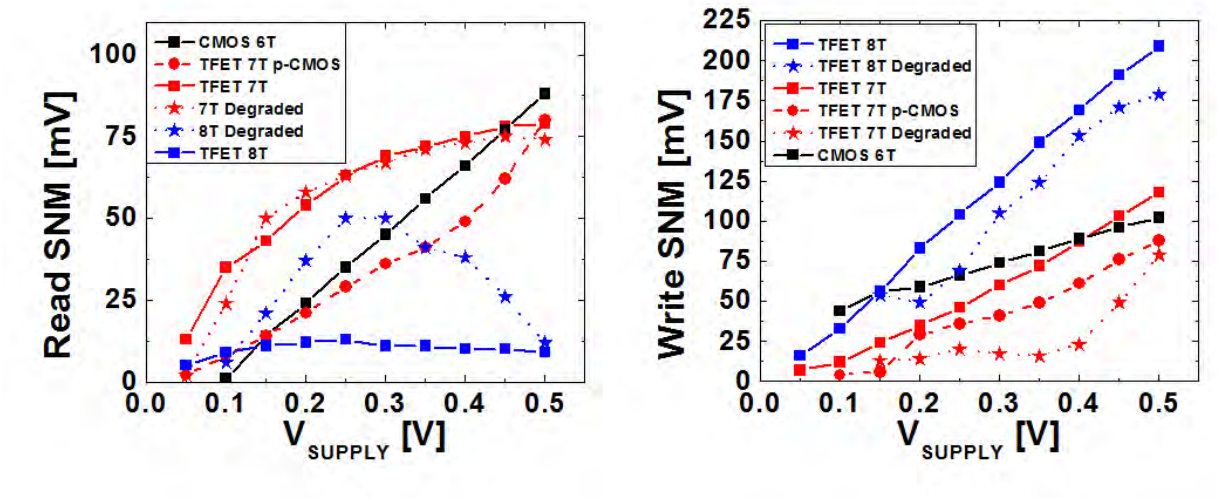
Figure 5.7: TFET-based SRAM cell designs: (a) 7T and (b) 8T.

Figure 5.7(a) shows a 7T cell design proposed by Kim *et al.* [19], in which the read operation is isolated from the write operation through the use of an additional transistor. Because the risk of a read disturb is eliminated, the access transistors can be sized for better writeability.

Figure 5.7(b) shows an 8T cell design, which employs a pair of inward-oriented and outward-oriented TFETs, connected in parallel to form a transmission gate, for each storage node.

5.4.1.3 Comparison of Cell Designs

SPICE simulations were conducted to compare the read and write noise margins of the various SRAM cell designs. Read noise margin (RNM) is calculated graphically from a butterfly plot, and is the length of a side of the largest square that can fit in the smaller of the two lobes. Write noise margin (WNM) is the difference between V_{SUPPLY} and the minimum bit-line voltage required to flip the data storage node [18].

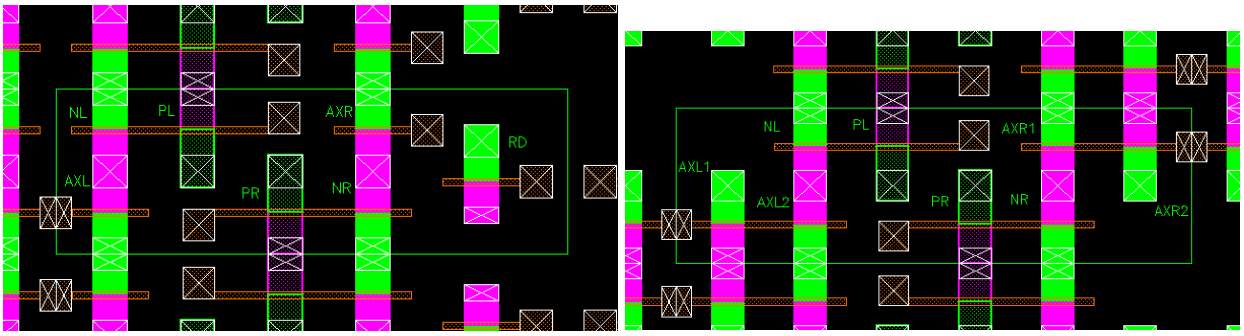


(a) (b)

Figure 5.8: (a) Read Static Noise Margin and (b) Write Static Noise Margin vs. Cell Supply Voltage for various SRAM cell designs.

Figure 5.8(a) compares RNM *vs.* V_{SUPPLY} for the 3 TFET-based cell designs, as well as a conventional complementary MOSFET (CMOS) 6T design and another design that uses TFETs for the n-channel devices and MOSFETs for the p-channel (pull-up) devices. The CMOS 6T cell has better RNM for V_{SUPPLY} greater than 0.5 V because the MOSFET has superior characteristics at higher operating voltages. The 7T design has better RNM than the 8T design because of the isolated nature of the read operation. For the 8T SRAM cell design, degraded TFET performance provides for better RNM at moderate voltages because the forward-biased diode current in the access TFET serves more effectively to assist a read-'0' operation in this case.

Figure 5.8(b) compares WNM for the various SRAM cell designs. Cell designs that have greater access transistor current achieve higher WNM values. The TFET 8T cell design has the greatest access transistor current, and therefore has the best writeability.



(a) (b)

Figure 5.9: SRAM cell layouts for (a) 7T TFET design and (b) 8T TFET design. The rectangular outline delineates a unit cell. Metal lines not shown for clarity.

To examine the tradeoff between cell size and performance, the TFET-based SRAM cells were laid out as shown in Figure 5.9. Design rules for the 22 nm node were scaled proportionately for the 14 nm node (the equivalent node for the 2018 ITRS FinFET design [20]). The 6T CMOS cell requires 106.4 nm × 220.64 nm, or 0.023476 μm^2. The 7T TFET design requires 86.4 nm × 269.2 nm, or 0.023259 μm^2 (within 1% of the CMOS cell area). The 8T TFET design requires 86.4 nm × 287.2 nm, or 0.024814 μm^2, ~6% more area than the CMOS cell.

5.4.2 Register-File Cells

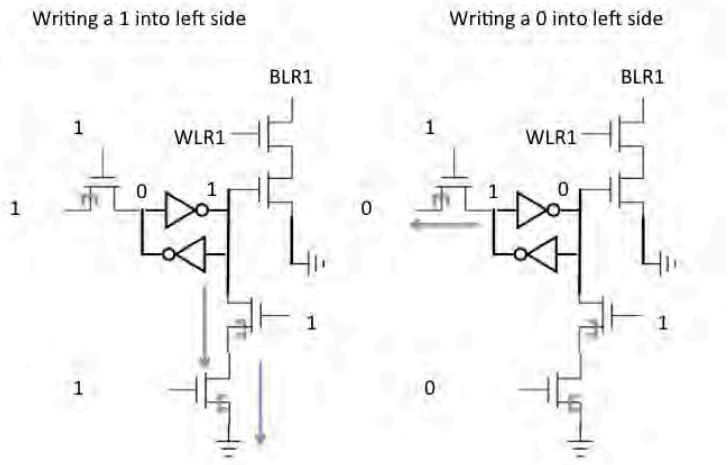


Figure 5.10: Two-sided Register File write operations.

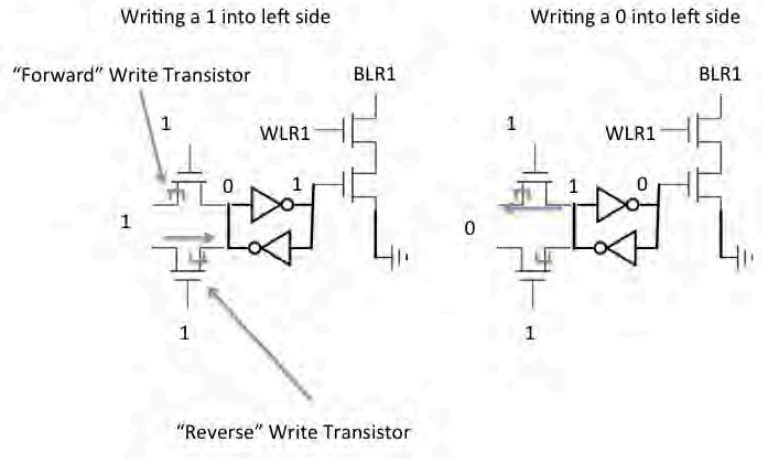
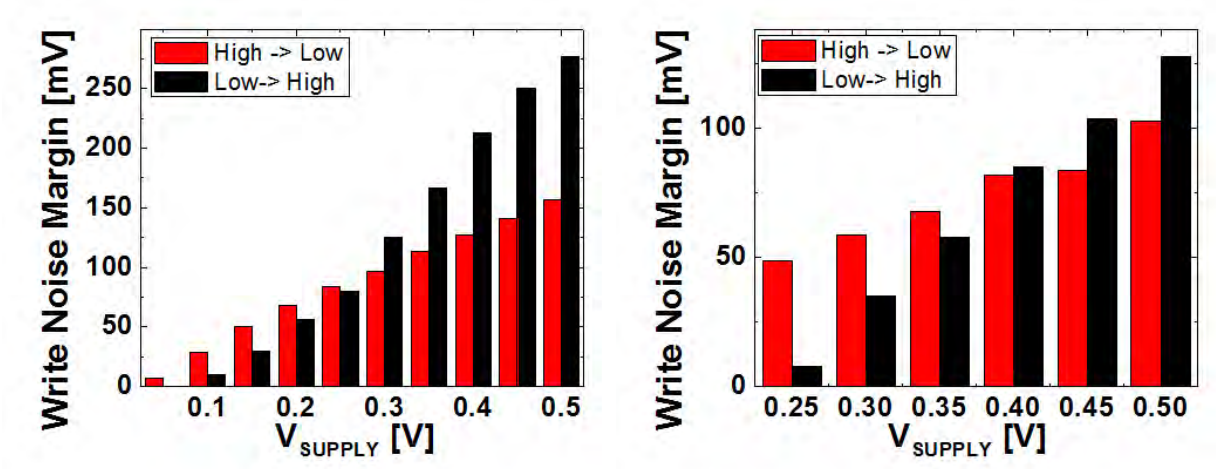


Figure 5.11: One-sided Register File write operations.

Register-file cells have multiple ports for read and write access to the storage nodes and can be one- or two-sided, as shown in Figures 5.10 and 5.11. For the two-

sided design, writing a '1' and writing a '0' to the storage node are accomplished through two separate current paths. The one-sided design uses a 2-TFET transmission gate to write to the cell, and employs one less transistor per write port than the two-sided design.



(a) **(b)**

Figure 5.12: Write Noise Margin vs. Supply Voltage for Register File (a) two-sided and (b) one-sided cells.

To evaluate the performance of the register-file cells, write margin and read delay are measured from SPICE simulations. Write margin is defined as the difference between the desired voltage to be placed on the node (V_{SUPPLY} or 0 V) and the write voltage trip point, and is plotted in Figure 5.12. The one-sided cell design cannot be written for V_{SUPPLY} below 0.25 V. (The access transistors can not flip the storage node, unless they are sized to be more than 3 times larger than the inverter transistors.) The two-sided cell design has higher write margins because of the separate paths used to write a '1' *vs.* a '0'.

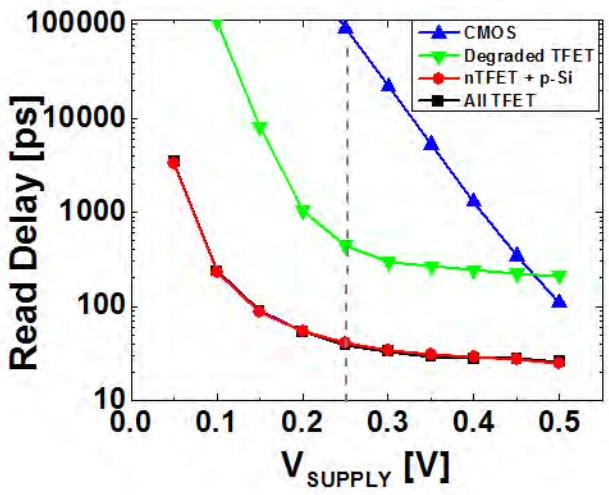


Figure 5.13: Comparison of read delay *vs.* supply voltage for register-file cells. Two-sided and one-sided cells have identical read schemes and hence the same delay.

The read delay is defined as the time required for the read bit-line (BLR1) voltage to drop by 50% after the read word line voltage (WLR1) is pulsed high. Because the read delay depends mostly on the current flow of the read access transistor, it is ~10× longer for a (10×) degraded TFET technology, as shown in Figure 5.13.

5.5 Conclusions

At 14-nm node dimensions, an optimized Ge-source heterostructure TFET technology is projected to provide for greater than 7× improvement in energy efficiency (at the same leakage power and stage delay) as compared to CMOS technology, for digital logic with stage delay slower than 100 ps. For static memory, a 7-TFET cell design is projected to have higher read noise margin and comparable write noise margin than a conventional 6-FinFET-MOSFET cell design of comparable layout area, for supply voltages in the range from 0.25 V to 0.5 V. For register files, a two-sided TFET-based cell design is necessary for writeability.

5.6 References

1. E. J. Nowak, "Maintaining the benefits of CMOS scaling when scaling bogs down," *IBM Journal of Research and Development*, Vol. 46, pp. 169-180, 2002.
2. S. Hanson, B. Zhai, K. Bernstein, D. Blaauw, A. Bryant, L. Chang, K. K. Das, W. Haensch, E. J. Nowak, and D. M. Sylvester, "Ultra low voltage, minimum-energy CMOS," *IBM Journal of Research and Development*, Vol. 50, pp. 469-490, 2006.

3. J. Appenzeller, Y.-M. Lin, J. Knoch, and P. Avouris, "Band-to-band tunneling in carbon nanotube field-effect transistors," *Physical Review Letters*, vol. 93, pp. 196805.1-196805.4, 2004.
4. P.-F. Wang, K. Hilsenbeck, T. Nirschl, M. Oswald, C. Stepper, M. Weis, D. Schmitt-Landsiedel, and W. Hansch, "Complementary tunneling transistor for low power application," *Solid-State Electronics*, vol. 48, no. 12, pp. 2281-2286, 2004.
5. W. G. Vandenberghe, A. S. Verhulst, G. Groeseneken, B. Sorée, and W. Magnus, "Analytical model for point and line tunneling in a tunnel field-effect transistor," in *Proceedings of SISPAD*, pp. 137-140, 2008.
6. A. C. Seabaugh, and Q. Zhang, "Low-voltage tunnel transistors for beyond CMOS logic," *Proceedings of the IEEE*, vol. 98, no. 12, pp. 2095-2110, 2010.
7. K. K. Bhuwalka, J. Schulze, and I. Eisele, "Performance enhancement of vertical tunnel field-effect transistor with SiGe in the dp^+ layer," *Japanese Journal of Applied Physics*, vol. 43, pp. 4073-4078, 2004.
8. F. Mayer, C. Le Royer, J.-F. Damlencourt, K. Romanjek, F. Andrieu, C. Tabone, B. Previtali, and S. Deleonibus, "Impact of SOI, $Si_{1-x}Ge_xOI$, and GeOI substrates on CMOS compatible tunnel FET performance," *IEEE International Electron Devices Meeting Technical Digest*, pp. 163-166, 2008.
9. S. H. Kim, H. Kam, C. Hu, and T.-J. K. Liu, "Germanium-source tunnel field effect transistors with record high I_{ON}/I_{OFF}," in *Symposium on VLSI Technology Conference Digest*, pp. 178-179, 2009.
10. S. H. Kim, S. Agarwal, Z.A. Jacobson, P. Matheu, C. Hu, and T.-J. K. Liu, "Tunnel field effect transistor with raised germanium source," *IEEE Electron Device Letters*, vol. 31, pp. 1107-1109, 2010.
11. International Technology Roadmap for Semiconductors (ITRS) 2011 Edition. http://public.itrs.net.
12. S. Mookerjea, R. Krishnan, S. Datta, and V. Narayanan, "On enhanced Miller capacitance effect in interband tunnel transistors," *IEEE Electron Device Letters*, vol. 30, no. 10, pp. 1102-1104, 2009.
13. T. Ghani *et al.*, "A 90nm high volume manufacturing logic technology feature novel 45nm gate length strained silicon CMOS transistors," *IEEE International Electron Devices Meeting Technical Digest*, pp. 978-980, 2003.
14. Sentaurus Device User Guide, Synopsys Inc., 2009.
15. J. Lin, E. H. Toh, C. Shen, D. Sylvester, C. H. Heng, G. Samurdra, and Y. C. Yeo, "Compact HSPICE model for IMOS device," *Electronics Letters*, vol. 44, no. 2, pp. 91-92, 2008.
16. W. G. Vandenberghe, B. Sorée, W. Magnus, G. Groeseneken, and M. V. Fischetti, "Impact of field-induced quantum confinement in tunneling field-effect devices," *Appl. Phys. Lett.*, vol. 98, 143503, 2011.
17. Q. Zhang, W. Zhao and A. Seabaugh, "Low-subthreshold-swing tunnel transistors," *IEEE Electron Device Letters*, vol. 27, pp. 297-300, 2006.
18. J. Singh, K. Ramakrishnan, S. Mookerjea, S. Datta, N. Vijaykrishnan, and D.

Pradham, "A novel Si-tunnel FET based SRAM design for ultra low-power 0.3V V_{DD} applications," in *Proc. ASP-DAC*, pp. 181-186, 2010.
19. D. Kim, Y. Lee, J. Cai, I. Lauer, L. Chang, S. J. Koester, D. Sylvester, and D. Blaauw, "Low power circuit design based on heterojunction tunneling transistors (HETTs)," in *Proc. ISLPED*, pp.219-224, 2009.
20. B. S. Haran *et al*., "22 nm technology compatible fully functional 0.1 μm^2 6T-SRAM cell," *IEEE International Electron Devices Meeting Technical Digest*, pp. 625-628, 2008.

Chapter 6

Conclusion

6.1 Summary of Work

This dissertation examined the challenges transistor scaling will face in future years. To enable future improvements in technology (as has been the case for the past five decades thanks to Moore's Law [1]), radical changes will be necessary. These changes will need to be made in both the structures of switching devices and the heterogeneity of the devices needed. It is unlikely that one device will be able to meet all of the requirements of future devices, but understanding the advantages and disadvantages of different device structures will be of greater importance.

In Chapter 2, an analysis was conducted of prospective devices that could replace or supplement CMOS in the future. The definitive history was recorded and an understanding of device principles was made. Current work in the field was examined and benchmarked in six areas: I_{ON}, I_{OFF}, Energy, Manufacturing Complexity, Cost, and Area Used. Analysis showed that no one device has demonstrated characteristics that are completely superior to existing CMOS, but many surpass CMOS in one or more areas and are promising candidates for future research.

One of these candidates is Tunneling FETs (TFETs). Chapter 3 examines the current state of tunneling models in industry standard TCAD tools. Using a Germanium source TFET fabricated in UC Berkeley's Microfabrication laboratory, a new tunneling model with dynamic nonlocal path detection and support for indirect and heterostructure tunneling is calibrated and validated [2]. While past works showed that high source doping concentrations were ideal for TFET design, this work showed that a doping concentration in the 10^{19} cm^{-3} range allows for tunneling to occur underneath the gate, allowing the tunneling area and hence on-state current to increase [3-7]. Further simulations optimized the planar Germanium source's thickness and gate overlap parameters.

Chapter 3 demonstrated that a raised Germanium source holds an advantage over a planar design. Later work optimized both the raised source design and the body thickness and doping of Germanium source TFETs [8-9]. Chapter 4 builds on this by scaling the device down to a 4 nm gate length and finding the impacts of various parameter changes. Ultimately, the optimal source doping concentration of 1.8×10^{19} cm^{-3} is found to allow for TFET scaling to 29 nm of device length (isolation to isolation).

Chapter 5 extended the analysis of TFETs to the circuit level. Using a Verilog-A look-up table implemented in SPICE, ring oscillators, SRAM cells, and Register Files were studied and compared to a FinFET at an equal device length (29 nm at the 2018 ITRS node) [10]. TFET ring oscillators were found to perform better than FinFETs at stage delays slower than approximately 1 ns (corresponding to ~1 GHz). However, a degraded TFET model was also developed to account for the potential that TFETs can not be fabricated with as high performance as has been simulated, due to unforeseen physical phenomenon. (TFET simulations often show better characteristics than fabricated devices [11].) The degraded TFET does not perform better than FinFETs.

SRAM cells showed that TFETs perform better than FinFETs at voltages below 0.45 V, but require slightly more area (approximately 1% more per cell). Finally register file simulations show that two-sided register file cells perform better than FinFET designs below 0.45 V as well.

In conclusion, TFETs are a potential supplement to CMOS technologies. Their applications are limited to device lengths above 29 nm, voltages below 0.45 V, and frequencies below 1 GHz. One application for TFETs may be in mobile processors. Many mobile devices, such as smartphones and tablets, need to be on constantly, but are only actively in use for short periods of time. During the inactive use periods, the phone processor performs noncritical tasks such as syncing e-mail, playing music files, and awaiting calls. To increase standby battery life, the heterogeneous integration of TFETs with high performance CMOS would allow devices to have high active performance when using the high performance cores but low standby power consumption when only using the low power cores, increasing standby battery life. The nVidia Tegra 3 mobile processor for applications in smartphones and tablets contains five cores: one low power "companion" core and four high performance cores [13]. All of the cores in this device are CMOS, but in the future, the "companion" core could be fabricated with TFET devices. This demonstrates that mobile processor vendors are willing to sacrifice some die size to increase functionality (in this case, standby power consumption). Other applications for TFETs could include devices such as music players, watches, health and fitness tracking devices, and any other device requiring long battery life and moderate performance. Self-powered devices, such as those utilizing solar or energy harvesting power sources, would also be an application for TFETs.

6.2 Future Directions

There are several areas where future work would be useful in understanding and verifying TFET technology as a potential CMOS supplement:

Examine other potential devices: In Chapter 2, the scope was limited to charge-based devices that did not require large architecture shifts. Some devices that could be included in future analyses include single electron transistors, spin-based devices, and devices for reversible computing. ITRS and some other authors have examined the potential of these devices, but using different metrics from those developed in Chapter 2 [10, 12].

Additional model calibration: In Chapter 3, Sentaurus Device's Dynamic nonlocal tunneling model was calibrated to a fabricated device [2]. However, this calibration would be more robust if other devise were available to verify the calibration, particularly at different doping concentrations.

Field-induced confinement: One physical effect not included in the simulations

performed is field-induced quantum confinement. Field-induced quantum confinement increases the effective conduction band energy in the source of TFETs, changing the threshold voltage and potentially changing the behavior of devices that utilize line tunneling (such as those studied in this work) [14]. Some work has been done to include this effect manually in simulators, but should be solved self-consistently with the other Sentaurus Device quantization models.

Improve tunneling models: Recently, several works have shown that a new method of using the tunneling models may provide improved accuracy between simulated and fabricated characteristics [15-16]. These works separate indirect and direct tunneling. Trap-assisted tunneling is likely present in the polycrystalline Germanium source region of the devices studied in this work, and should also be included in future simulations.

Fabricate devices: Ideally, the parameters optimized in Chapter 4 would be used to fabricate devices to verify the simulated characteristics.

System-level analysis: Circuit level analysis of several basic circuits was performed in Chapter 5, but system-level analysis, as was done in [17], would be needed before TFETs could be implemented in industry.

6.3 References

1. G. E. Moore, "Cramming more components onto integrated circuits," *Electronics*, vol. 38, pp. 114-117, 1965.
2. S. H. Kim, H. Kam, C. Hu, and T.-J. K. Liu, "Germanium-source tunnel field effect transistors with record high I_{ON}/I_{OFF}," in *Symposium on VLSI Technology Conference Digest*, pp. 178-179, 2009.
3. C. Sandow, J. Knoch, C. Urban, Q.-T. Zhao, and S. Mantl, "Impact of electrostatics and doping concentration on the performance of silicon tunnel field-effect transistors," *Solid-State Electronics*, vol. 53, no. 10, pp. 1126-1129, Oct. 2009.
4. O. M. Nayfeh, J. L. Hoyt, D. A. Antoniadis, "Strained-$Si_{1-x}Ge_x$ / Si Band-to-Band Tunneling Transistors: Impact of Tunnel-Junction Germanium Composition and Doping Concentration on Switching Behavior," *IEEE Transactions on Electron Devices*, vol. 56, no. 10, pp. 2264-2269, Oct. 2009.
5. M. Fulde, A. Heigl, M. Weis, M. Wirnshofer, K. v. Arnim, Th. Nirschl, M. Sterkel, G. Knoblinger, W. Hansch, G. Wachutka, and D. Schmitt-Landsiedel, "Fabrication, Optimization and Application of Complimentary Multiple-Gate Tunneling FETs," *IEEE International Nanoelectronics Conference*, 2008, pp. 579-584.
6. P.-F. Wang, T. Nirschl, D. Schmitt-Landsiedel, and W. Kansch, "Simulation of the Esaki-tunneling FET," *Solid-State Electronics*, vol. 47, no. 7, pp. 1187-1192, July 2003.
7. W. G. Vandenberghe, A. S. Verhulst, G. Groeseneken, B. Sorée, and W. Magnus, "Analytical model for point and line tunneling in a tunnel field-effect transistor," in *Proceedings of SISPAD*, pp. 137-140, 2008.

8. S. H. Kim, S. Agarwal, Z.A. Jacobson, P. Matheu, C. Hu, and T.-J. K. Liu, "Tunnel field effect transistor with raised germanium source," *IEEE Electron Device Letters*, vol. 31, pp. 1107-1109, 2010.
9. S. H Kim, Z. A. Jacobson, and T.-J. K. Liu, "Impact of Body Doping and Thickness on the Performance of Germanium-Source TFETs," *IEEE Transactions on Electron Devices*, vol. 57, no. 7, pp. 1710-1713, July 2010.
10. International Technology Roadmap for Semiconductors (ITRS) 2011 Edition. http://public.itrs.net.
11. A. C. Seabaugh, and Q. Zhang, "Low-voltage tunnel transistors for beyond CMOS logic," *Proceedings of the IEEE*, vol. 98, no. 12, pp. 2095-2110, 2010.
12. K. Bernstein, "Device and Architecture Outlook for Beyond CMOS Switches," *Proceedings of the IEEE*, vol. 98, no. 12, pp. 2169-2184, 2010.
13. "Variable SMP – A Multi-Core CPU Architecture for Low Power and High Performance," white paper, NVIDIA, 2011.
14. W. G. Vandenberghe, B. Sorée, W. Magnus, G. Groeseneken, and M. V. Fischetti, "Impact of field-induced quantum confinement in tunneling field-effect devices," *Applied Physics Letters*, vol. 98, 143503, 2011.
15. K.-H. Kao, A. S. Verhulst, W. G. Vandenberghe, B. Sorée, G. Groeseneken, and K. De Meyer, "Direct and Indirect Band-to-Band Tunneling in Germanium-Based TFETs," *IEEE Transactions on Electron Devices*, vol. 59, no. 2, pp. 292-301, 2012.
16. C. Kampen, A. Burenkov, and J. Lorenz, "Challenges in TCAD Simulations of Tunneling Field Effect Transistors," in *Proceedings of the European Solid-State Device Research Conference (ESSDERC) 2011*, pp. 139-142, 2011.
17. V. Saripalli, A. Mishra, S. Datta, and V. Narayanan, "An energy-efficient heterogeneous CMP based on hybrid TFET-CMOS cores," in *2011 48th Design Automation Conference (DAC)*, pp. 729-734, 2011.