Tunnel Field Effect Transistors: from Steep-Slope Electronic Switches to Energy Efficient Logic Applications

THÈSE N^{O} 6802 (2015)

PRÉSENTÉE LE 9 OCTOBRE 2015
À LA FACULTÉ DES SCIENCES ET TECHNIQUES DE L'INGÉNIEUR
LABORATOIRE DES DISPOSITIFS NANOÉLECTRONIQUES
PROGRAMME DOCTORAL EN MICROSYSTÈMES ET MICROÉLECTRONIQUE

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Arnab BISWAS

acceptée sur proposition du jury:

Prof. M. Kayal, président du jury
Prof. M. A. Ionescu, directeur de thèse
Prof. L. Selmi, rapporteur
Prof. E. Pop, rapporteur
Dr J. M. Sallese, rapporteur

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Suisse
2015

In a conflict between the heart and the brain,
follow your heart.
— Swami Vivekananda

Dedicated to my father and mother…

Acknowledgements

This thesis marks an end to an amazing journey of more than four years filled with wonderful and exciting experiences. This thesis has been written during my stay as a PhD student at the Ecole Polytechnique Fédérale de Lausanne(EPFL), and I would like to express my gratitude to all the people who supported me scientifically and emotionally in this journey.

First of all, I would like to express my deep appreciation for my thesis advisor, Prof. Adrian M. Ionescu, for his willingness to take me into his research group and for the wonderful opportunity he gave me to work on this very interesting topic. Thanks to his support, encouragement and constant discussions, my motivation for the project has always been very strong, also in the most difficult of times.

I would like to acknowledge the members of the jury for having accepted to review my PhD thesis, for the valuable feedback on my work and their time to disseminate it: Prof. Eric Pop from the Stanford University(USA), Prof. Luca Selmi from University of Udine(Italy), Dr. Jean-Michel Sallese from EPFL and Prof. Maher Kayal, president of the jury, from EPFL.

Special thanks to all scientific partners I have had the pleasure to interact with within the STEEPER and E2SWITCH European projects. Thank you for the useful discussion and brainstorming during the meetings on this challenging subject: IUNET, IBM Zurich, CEA Leti, IMEC Leuven, Cambridge CMOS sensors, FZJ Research Center Jülich and ETH Zurich.

I am also grateful to Luca De Michielis for his continuous help and support throughout my PhD. His knowledge in simulations and the insightful discussions that we have had on device physics, have been a great aid in writing this thesis. I also thank Livio for supervising my masters thesis in Nanolab and introducing me to the world of Tunnel FETs. I thank all my past and present Nanolab colleagues Erick, Mariazel, Ali, Clara, Cem, Jose, Francesco, Maneesha, Nilay, Sara, Ji, Seb, Alex, Giovanni, Donatello, Jyotshna, Anupama, Mohammad, Nenad, Matthieu and Montse for all the good times over the years.

I also want to thank Dr. Wladek Grabinsky and Dr. Antonios Bazigos, for all their help throughout my PhD, specially in setting up the characterization related projects. Thanks to all the engineers in CMI, EPFL who trained me in various equipments used in the fabrication process and for their help in solving problems related to clean room processing.

The awesome-twosome Nanolab secretaries Isabelle and Karin were always there to help with any bureaucratic formality. Thank you for all your help and kindness. I also would like to thank the doctoral secretaries Marie and Lucie for their help during the beginning and towards the end of my PhD. Thanks also to Joseph, Roland, Marc, Raymond and Cedric for all the

administrative help on the IT infrastructures.

A special thanks to my colleagues Hoel and Elizabeth with whom I have spent my formative years in Nanolab and shared a special bond. We even called ourselves as the "three musketeers". I will miss all the fun we had together specially during the conference trips. Special thanks to my best pals in town Wolfgang, Emanuele and Pankaj for being part of my journey through these years. Along with Francesco(Chicco) and Erick they have been like brothers to me. I will definitely miss all the good times over lunch, coffee time discussions and countless Friday and Saturday nights. My time in Nanolab would not have been the same without all the fun and camaraderie we have shared over the years. I will forever cherish the great times we had together.

Special thanks to fellow Bengalis and my good friends in campus Soumya(Bob) and Namrata. It was always cool hanging out with you guys in mother tongue goodness. I also would like to thank my officemates Nilay and Cem with whom I shared my office space ELB343 in Nanolab. Their friendship and support created an exceptional working environment. A special thanks goes to all my friends who made my stay in Lausanne so special and enjoyable: Gabriele, Clemens, Pietro, Enver, Oguz, Charly, Francesco(Chicco), Chloé, Annie, Massimo, Gianrocco, Loris, Rammohan, Anurag, Bakul, Aamani, Czuee, Chhavi and the list goes on.

I also acknowledge the help of all my master project and semester project students whose thesis work was directly or indirectly supervised by me during the course of my PhD: Saurabh, Francesco, Maneesha and Riccardo.

I would like to express my deepest appreciation for Irene who have hosted me in Lausanne for more than four years. Your love and kindness made me forget how far away from home I really was.

Last but not the least, there are no words to describe my gratitude towards my family for their unconditional love, inspiration and encouragement: my father, mother, sister and brother-in-law. They have been a pillar of support guiding me through thick and thin of times.

Lausanne, 27th July 2015 Arnab Biswas

Abstract

At the heart of the microelectronic revolution is the complimentary-metal-oxide semiconductor(CMOS) technology. Scaling of the MOSFET is governed by the need of higher speeds and package density. The main reason to make transistors smaller is to pack more and more devices in a given chip area. It is also expected that smaller transistors switch faster. Upon further decreasing the MOSFET device dimensions, the leakage current of the MOSFET increases,while the supply voltage can no longer be scaled down, both of which result in increased power consumption. By lowering V_{DD} from 500mV to 250mV while preserving the effective voltage, the leakage power has been shown to increase unacceptably by a factor of 275 in a 45-nm bulk CMOS technology. Another way of reducing the voltage supply without performance loss is to increase the turn-on steepness, which means decreasing the average sub-threshold swing, SS_{avg}, over 3 or 4 decades of current. Therefore, devices with a steep SS, called steep-slope switches, are expected to enable V_{DD} scaling.

The primary aim of this work has been the investigation of one such steep-slope switch: the Tunnel Field Effect Transistor, starting from a compact modelling perspective to its possible applications. We discuss the basic working principle of Tunnel FET and the main differences of a Tunnel FET with respect to a conventional MOSFET are pointed out and explained. We study in detail the band-to-band tunneling model in detail with reference to the Landauer's formulation and Kane's model. An overview of the all the band-to-band tunneling models available in commercial TCAD simulators are presented. A TCAD based simulation study is done to understand the main device characteristics. The importance of Miller effect specifically to Tunnel FETs and its impact on circuit behaviour is also discussed. We propose a new structure to mitigate this effect on the circuit performance of the Tunnel FET.

We developed a compact DC/AC model for the double gate Tunnel FET, which is capable of describing the I-V and C-V characteristics in all regimes of operation. The model takes in to account ambipolarity, drain side breakdown and all tunneling related physics. An indirect temperature dependence is also added to the model to study the temperature independent behaviour of tunneling. The model was further implemented in a Verilog-A based circuit simulator. Following calibration to experimental results of Silicon and strained-Silicon TFETs, the model has been also used to benchmark against a 28nm FDSOI CMOS node for digital and analog applications. Typical analog FOMs showed the benefits of a TFET over CMOS particularly in the low operating power regime. The circuits built with Tunnel FETs also showed interesting temperature behaviour which was superior to the compared CMOS node.

In the same work, we also explore and propose solutions for using TFETs for low power memory applications. Both volatile and non-volatile memory concepts are investigated and explored. The application of a Tunnel FET as a capacitor-less memory has been experimentally demonstrated for the first time, which opens a new field of applications for TFETs. We propose a new Tunnel FET based non-volatile memory concept, which uses vertical line tunneling to boost the memory operation and validated with simulations. We also developed process flows for the new structures for future realization in the clean room in EPFL.

Key words: Physics of semiconductor devices; modeling and simulation; Tunnel FET, band-to-band tunneling; compact model; Verilog-A; TCAD simulation; steep sub-threshold slope switches; analytical compact model; circuit and device level benchmarking; characterization; capacitor-less DRAM; SONOS; non-volatile memory.

Résumé

Au cœur de la révolution microélectronique est la technologie offerte métal oxyde semiconducteur (CMOS). Mise à l'échelle du MOSFET est régie par la nécessité de vitesses plus élevées et de la densité de l'emballage. La principale raison à réaliser des transistors plus petits est d'emballer plus en plus d'appareils dans une zone de la puce donnée. Il est également prévu que les petits transistors commutent rapidement. Lors de diminuer encore les dimensions du dispositif MOSFET, le courant de fuite du transistor MOSFET augmente, tandis que la tension d'alimentation ne peut plus être réduite, à la fois de l'augmentation qui se traduit par une consommation de courant. En abaissant V_{DD} de 500mV à 250mV, de la préservation de la tension tandis que efficace, la puissance de fuite a été démontré que l'augmentation inacceptable par un facteur de 275 dans une technologie CMOS en vrac de 45 nm. Une autre façon de réduire la tension d'alimentation sans perte de performance est d'augmenter le tour sur la pente, ce qui signifie diminuer le swing moyenne sous-seuil, SS_{avg}, sur 3 ou 4 décennies de courant. Par conséquent, les appareils dotés d'une pente raide SS, appelés commutateurs forte pente, sont attendus pour permettre V_{DD} mise à l'échelle.

Ces travaux de recherche ont porté sur les transistors à effet de champ, à effet tunnel (Tunnel FET or TFET) et à homo-jonction partant d'une perspective de modélisation compacte jusqu'à leurs applications possibles. En premier lieu, une étude basée sur une simulation TCAD est effectuée pour comprendre les caractéristiques principales du dispositif. Le principe de fonctionnement de base du Tunnel FET est abordé. Les principales différences d'un Tunnel FET par rapport à un MOSFET classique sont mises en exergue et expliquées. Le modèle de l'effet tunnel bande à bande a été étudié en détail en référence à la formule de Landauer et au modèle de Kane. Un aperçu de l'ensemble des modèles d'effet tunnel bande à bande disponibles dans les simulateurs commerciaux TCAD est présenté. L'importance de l'effet Miller en particulier pour les Tunnel FETs et son impact sur le comportement des circuits sont également traités.

Un modèle DC/AC compact, capable de décrire les caractéristiques I-V et C-V dans tous les régimes de fonctionnement a été développé. Le modèle prend en compte l'ambipolarité et toute la physique relative à l'effet tunnel. Une dépendance à la température est également ajoutée au modèle pour étudier le comportement indépendant de la température de l'effet tunnel. Le modèle a en outre été implémenté dans un simulateur de circuit en Verilog-A. Après étalonnage sur les résultats expérimentaux de TFETs sur silicium et silicium contraint, le modèle a également été utilisé comme référence pour comparaison avec un nœud CMOS FDSOI 28nm pour des applications numériques et analogiques. Les facteurs de mérite analogiques typiques ont démontré les avantages du TFET par rapport au CMOS, en particulier dans le régime de fonctionnement à faible puissance. Les circuits construits avec des Tunnel FETs ont également montré un comportement avec la température intéressant et supérieur au nœud CMOS comparé.

Dans cette recherche, nous explorons et proposons aussi des solutions pour l'utilisation des TFETs pour des applications de mémoire à faible puissance. Les deux concepts de mémoire volatile et non-volatile sont examinés et explorés. L'utilisation d'un Tunnel FET comme mémoire sans condensateur a été démontrée expérimentalement pour la première fois, ouvrant un nouveau champ d'applications pour les TFETs. Un nouveau concept de mémoire non volatile basée sur des TFETs et qui utilise un effet tunnel en ligne verticale pour améliorer le fonctionnement de la mémoire a été proposé et validé avec des simulations. De nouveaux concepts de dispositifs ont été proposés et leurs procédés de fabrication sont développés pour une réalisation future dans la salle blanche de l'EPFL.

Mots clefs : Physique des semiconducteurs ; modélisation et simulation ; Tunnel FET ; effet tunnel bande à bande ; modèle compact ; Verilog-A ; simulation TCAD ; commutateurs à pente sous seuil abrupte ; modèle analytique compact ; analyse comparative de dispositifs et circuits ; caractérisation ; mémoire vive dynamique (DRAM) sans condensateur ; SONOS ; mémoire non-volatile.

Zusammenfassung

Im Herzen des mikroelektronischen Umdrehung der komplementäre Metall-Oxid-Halbleiter (CMOS) -Technologie. Skalierung des MOSFET wird durch den Bedarf an höherer Geschwindigkeit und Packungsdichte bestimmt. Der Hauptgrund Transistoren kleiner zu machen ist, mehr und mehr Vorrichtungen in einer gegebenen Chipfläche zu verpacken. Es wird auch erwartet, dass kleinere Transistoren schneller schalten. Bei weiterer Verringerung der MOSFET Geräteabmessungen, den Leckstrom des MOSFET ansteigt, während die Versorgungsspannung nicht mehr verkleinert werden kann, die beide zu einem erhöhten Energieverbrauch. Durch die Absenkung V_{DD} von 500mV bis 250mV und gleichzeitig die effektive Spannung hat die Leckleistung wurde gezeigt, dass unannehmbar um einen Faktor von 275 in einem 45-nm-CMOS-Technologie Schütt erhöhen. Ein anderer Weg zum Reduzieren der Versorgungsspannung, ohne Leistungsverlust ist, um die Einschalt-Steilheit, die die durchschnittliche Unterschwellenschwenkverringerungseinrichtung zu erhöhen, SS_{avg}, mehr als 3 oder 4 Jahrzehnten Strom. Daher Geräte mit einem steilen SS, genannt Steillagen-Schalter, wird erwartet, dass V_{DD} Skalierung zu ermöglichen.

Das Ziel dieser Arbeit war die Untersuchung von Homoübergangs-Tunnel-Feldeffekt transistoren (TFET) ausgehend von der Perspektive einer kompakten Modellierung bis hin zu möglichen Anwendungen. Zu Beginn wurde eine TCAD basierte Simulationsstudie durchgeführt, um die wichtigsten Geräteeigenschaften zu verstehen. Das grundlegende Arbeitsprinzip eines Tunnel-FET wird diskutiert. Die Hauptunterschiede eines Tunnel-FET gegenüber eines herkömmlichen MOSFET werden aufgezeigt und die Unterschiede werden erläutert. Das Band-zu-Band-Tunnelmodell wurde im Detail mit Bezug auf die Landauer Formulierung und das Kanes Modell untersucht. Es wird eine Übersicht aller Band-zu-Band-Tunnelmodelle verfügbar in kommerziellen TCAD Simulatoren vorgestellt. Die Bedeutung des Miller-Effekts speziell für Tunnel-FETs und seine Auswirkung auf das Schaltungsverhalten wird ebenfalls diskutiert.

Eine kompaktes DC/AC-Modell wurde entwickelt, welches in der Lage ist, die Beschreibung der I-V und C-V Eigenschaften in allen Betriebsbereichen zu beschreiben. Das Modell berücksichtigt Ambipolarität, Abbau der Abflussseite und die gesamte tunnelbezogenen Physik. Außerdem wurde zu dem Modell eine Temperaturabhängigkeit hinzugefügt, um das temperaturunabhängige Verhalten des Tunnelns zu untersuchen. Zusätzlich wurde das Modell in einen Verilog-A basierten Schaltungssimulator implementiert. Nach der Kalibrierung mit experimentellen Ergebnissen von Silicon und Strained-Silicon-TFETs wurde das Modell ebenfalls

als Benchmark gegen einen 28-nm-CMOS FDSOI Knoten für digitale und analoge Anwendungen verwendet. Typische analoge FOMs zeigten die Vorteile eines TFET gegenüber CMOS insbesondere im niedrigen Betriebsleistungsbereich. Die mit Tunnel-FETs aufgebauten Schaltungen zeigten außerdem ein interessantes Temperaturverhalten, welches im Vergleich zum CMOS Knoten überlegen war.
In der gleichen Arbeit untersuchen wir auch die Verwendung von TFETs für niedrige Stromspeicheranwendungen und bieten Lösungsvorschläge an . Sowohl flüchtige als auch nichtflüchtige Speicherkonzepte wurden untersucht und erforscht. Die Anwendung eines Tunnel-FET als kondensatorloser Speicher wurde experimentell zum ersten Mal gezeigt, was neue Anwendungsbereiche für TFETs eröffnet. Ein neues TFET basierendes nicht-flüchtiges Speicherkonzept, welches vertikales Linientunneln verwendet, um den Speichervorgang zu steigern, wurde vorgeschlagen und mit Simulationen validiert. Neue Gerätekonzepte wurden vorgeschlagen und Prozessabläufe für dieselben wurden für die zukünftige Realisierung im Reinraum des EPFL entwickelt.

Stichwörter: Physik der halbleiterbauelemente; modellierung und simulation; Tunnel-FET; Band-zu-Band-Tunneling; kompaktes modell; Verilog-A; TCAD simulation; steiler unterschwellenknotenschalter; analytischen Kompakt modell; Schaltungs-und Geräteebenen Benchmarking; Charakterisierung; kondensatorlose DRAM; SONOS; nicht-flüchtiger Speicher.

Contents

Acknowledgements **i**

Abstract (English/Français/Deutsch) **iii**

List of figures **xiii**

List of tables **xxiii**

Acronyms **xxv**

1 Introduction **1**
- 1.1 The microelectronic revolution . . . 1
- 1.2 Evolution of nano-electronics . . . 2
- 1.3 The scaling problem . . . 3
  - 1.3.1 Challenges in MOSFET scaling . . . 4
  - 1.3.2 Beyond CMOS devices: an alternative ? . . . 6
- 1.4 Thesis overview . . . 8

2 Tunnel FET **11**
- 2.1 Band to band tunneling . . . 11
  - 2.1.1 Tunneling theory . . . 13
- 2.2 Overview of Simulation tools available . . . 15
  - 2.2.1 Sentaurus Local Tunneling model . . . 16
  - 2.2.2 Sentaurus Dynamic non-Local Tunneling model . . . 17
  - 2.2.3 Silvaco non-Local Tunneling model . . . 20
- 2.3 Basic working principle of a TFET . . . 22
  - 2.3.1 Point tunneling Tunnel FET . . . 22
  - 2.3.2 Line tunneling device . . . 26
  - 2.3.3 Temperature dependence . . . 27
- 2.4 FD-SOI Tunnel FET simulation and fitting to measurements . . . 27
  - 2.4.1 Calibration Methodology . . . 29
  - 2.4.2 Crystal orientation dependent tunneling generation rate . . . 31
- 2.5 Miller effect . . . 31
  - 2.5.1 Miller effect in Tunnel FETs . . . 33
  - 2.5.2 Proposed new structure to mitigate Miller effect in Tunnel FET . . . 34

2.6 Summary 37

3 Compact modelling of Tunnel FET 39

3.1 Overview 39
3.2 DC model version 1 (non-local) 40
  3.2.1 Model Description 40
  3.2.2 Model Evaluation 45
3.3 DC model version 2 (local) 48
  3.3.1 Model Description 49
  3.3.2 Surface potential 49
  3.3.3 Tunnelling path estimation 50
  3.3.4 Tunneling current calculation 53
  3.3.5 Comparison with simulations 53
3.4 Charge model of a DG-TFET 56
  3.4.1 CV measurements on FD-SOI TFETs from LETI 57
  3.4.2 Suitability of 100-0 (drain/source) charge distribution 57
  3.4.3 Description of charge model of a DG-TFET 58
  3.4.4 Comparison with TCAD simulation 59
3.5 Verilog-A implementation of full compact model 60
  3.5.1 Single stage inverter 62
  3.5.2 2-bit half adder 63
  3.5.3 15 stage ring oscillator 64
3.6 Model calibration and benchmarking study 65
  3.6.1 Description of fabricated devices 67
  3.6.2 28nm FDSOI CMOS design kit 67
  3.6.3 Calibration with experimental results 68
  3.6.4 Device level analog benchmarking 69
  3.6.5 Circuit level analog benchmarking 74
3.7 Summary 78

4 Tunnel FETs based Volatile Memories 79

4.1 Why should TFETs be used as a memory 79
4.2 TFET as a volatile memory: Capacitorless memory introduction 81
4.3 FIN-Tunnel FET as 1T/0C DRAM 84
  4.3.1 Principle of Operation 85
  4.3.2 Proposed Programming Scheme 86
  4.3.3 Experimental Results 88
  4.3.4 Temperature Dependence 89
  4.3.5 Transient Measurements 89
4.4 Doped pocket device simulations 96
  4.4.1 Principle of operation 96
  4.4.2 DRAM operation 98
4.5 Summary 100

5 Tunnel FET based Non Volatile Memories **101**
5.1 Flash Memory 101
5.1.1 Working Principle 102
5.1.2 Floating gate vs. SONOS 104
5.1.3 State of the art 105
5.2 Simulation methodology 107
5.2.1 Vertical p-type TFET architecture 111
5.3 Vertical TFET non-volatile memory device 113
5.3.1 Simulation of memory operation 114
5.3.2 Simulation of Retention characteristics 116
5.4 Fabrication of TFET non-volatile memory 117
5.4.1 Evaluation of gate-oxide stacks 118
5.4.2 Process flow 120
5.4.3 Results obtained 122
5.5 Summary 122

6 Conclusions **127**
6.1 Future perspectives 128

A Calculation of parameters in section 2.4 **131**

B Capacitance measurements **133**
B.1 Notes on trans-capacitance measurements in a FET 135

C Pulsed measurement details **137**

Bibliography **148**

My Publications **149**

Curriculum Vitae **151**

List of Figures

1.1 Evolution of microelectronic technology: An Osborne Executive portable computer, from 1982 with a Zilog Z80 4MHz CPU, and a 2007 Apple iPhone with a 412MHz ARM11 CPU; the Executive weighs 100 times as much, has nearly 500 times as much volume, cost approximately 10 times as much (adjusted for inflation), and has about 1/100th the clock frequency of the smart-phone. [source wikipedia] 1

1.2 Evolution of memory technology from 2005 to 2014. Over a span of 9 years the capacity of a micro SD card has increased by a factor of 1024. [source Wikipedia] 3

1.3 Plot of CPU transistor counts against dates of introduction; the line corresponds to exponential growth with transistor count doubling every two years. [source wikipedia] 4

1.4 (a)Power consumption trends with gate length scaling (b) comparison of dynamic and leakage power trend over the years. [courtesy IBM coorp.] and (c) Near-exponential increase of CPU power density in recent decades has flattened with the introduction of multi-core CPUs [4] 5

1.5 (a)Comparison of an n-type double gate MOSFET and Tunnel FET schematic. The different source doping is to be noted. (b)Qualitative comparison of a conventional MOSFET (green) and a Tunnel FET (blue), which has a steep off–on transition and the lowest I_{OFF}. 7

1.6 Comparison of the minimum switching energy, E_{min} , and the corresponding voltage supply, VDD_{min} , for a subthermal swing device (S < 60 mV decade^{-1} , green curve) and the ideal MOSFET (S = 60 mV decade^{-1} , red) at the same I_{ON}/I_{OFF}. (source [5]) 8

2.1 Generation of an electron and hole pair as a result of band-to-band tunneling. 11

2.2 Direct and indirect semiconductor band structure (a)GaAs (direct) (b)Silicon (indirect) [7] 12

2.3 Schematic showing energy window and ΔJ (current) [9] 13

2.4 Schematic showing the difference between local (left) and non-local tunneling model 15

2.5 Comparison of transfer characteristic of a 100nm DG-TFET local and nonlocal models in Sentaurus and the non-local model in Silvaco with all parameters set to default. 17

2.6 Comparison of generation rates in local and nonlocal models in Sentaurus (a) Schenk model (b) Kane's model (P=2) (c) Hurkx model 18
2.7 Comparison of electron and hole generation rates in the dynamic nonlocal model in Sentaurus. 19
2.8 Schematic of non-local band to band tunneling in reverse bias in Silvaco Atlas. [12] 21
2.9 (a)Schematic of a Tunnel FET with ntype biasing applied (b) nTFET mode with both OFF and ON conditions shown (b) pTFET mode with both OFF and ON conditions shown. [20] 23
2.10 Measured (a)transfer and (b)output characteristics of an all Silcon nanowire n-type Tunnel FET on fully depleted SOI wafers at room temperature. The ambipolarity and superlinear onset which are typical to Tunnel FETs are highlighted. 24
2.11 Examples of line tunneling structures (a) [29] (b) [30] where the primary tunneling mechanism is oriented to the direction of the gate electric field. The band-to-band generation is spread over a linear fashion, resembling a line, hence the name line tunneling. 26
2.12 Measured transfer characteristic of a all Silcon nanowire n-type Tunnel FET on fully depleted SOI wafers at elevated temperatures. [36] 27
2.13 (a) Basic structure and dimensions of TFET devices simulated. (b) SEM image showing the fabricated SOI TFET device. Inset shows the top view.[23] 28
2.14 (a)Impact of reduced mass on tunneling current shown for different value of reduced masses including the theoretical mass and the default mass extracted from default parameters and (b) Simulated and measured transfer characteristics. Symbols indicate experimental data and solid lines the corresponding simulation. 29
2.15 (a) Simulated and measured output characteristics. Symbols indicate measured data and solid lines the corresponding simulation and (b) Variation of band-to-band generation rate as a function of uniform electric field with different crystallographic directions. 31
2.16 (a) Amplifier with voltage gain $-A$, with an impedance Z connected from input to output and (b) Amplifier with feedback capacitor C_C. 32
2.17 (a) A simple MOS transistor circuit showing the main capacitance components and (b) Transient response characteristics of silicon TFET and MOSFET based inverters. [41] 33
2.18 Capacitance–voltage characteristics showing the gate (C_{gg}), gate-to-source (C_{gs}), and gate-to-drain (C_{gd}) capacitances as a function of gate-to-source voltage V_{GS} for (a) Si MOSFET and (b) Si TFET [41] 34
2.19 Proposed structure with $T_{Si} = 5nm$, $L_{g,eff} = 50nm$. Gate work-function is $3.9eV$. $N_{A,source} = 10^{20}cm^{-3}$; $N_{D,drain} = 10^{20}cm^{-3}$ while $N_{A,channel} = 10^{18}cm^{-3}$. $2.5nm$ of HfO_2 are used as gate-oxide; $t_{BOX} = 10nm$; $t_{bulk} = 1\mu m$ and the effective Gate length $L_{g,eff} = 50nm$. 35
2.20 Transfer (a) and output (b) characteristics of the proposed new structure. . . . 35

2.21 (a)Transfer characteristics comparison of the proposed new structure with a SOI TFET. Improved I_{ON}, better SS and faster device speed (Intrinsic delay = CV/I) is observed. (b)Variation of propagation delay and energy per switching cycle with supply voltage. The proposed TFET becomes comparable to CMOS at $V_{DD} = 0.3V$ and lower. 36

2.22 Capacitance-Voltage characteristics showing the drain-to-gate(C_{DG}) and gate(C_{GG}) capacitances as a function of V_{GS}. 37

2.23 (a)Transient response of the three technologies in a three-stage inverter configuration for an input period of $1\mu s$. (b)Voltage transfer characteristics of the three technologies reported in Fig. 2.20 & 2.23(a). Inset shows the inverter Gain = dV_{OUT}/dV_{IN}. 38

3.1 Potential profile $\psi_s(x)$ at the Silicon/Oxide for an n-type DG-MOSFET(green) and an n-type DG-TFET(red) with the same device parameters except the source doping (n+/p+). Inset shows the electron density of the two devices at the given bias. The cross sections are taken $1nm$ below the gateoxide-Silicon channel interface. 40

3.2 Comparison between the electron quasi Fermi levels of a TFET and a MOSFET at two different V_D. The TFET electron quasi Fermi level stays at $-qV_D$ for almost the entire gate bias range. The cross sections are taken $1nm$ below the gateoxide-Silicon channel interface. 41

3.3 (a) Surface potential variation with gate voltage for a DG-TFET. Solid lines represent TCAD simulations. Dotted lines indicate results from eqn.3.2. Symbols refer to eqn.3.3. Bottom: Simulated electron density with Fermi and Maxwell-Boltzmann (MB) statistics for different V_D (b) and gate dielectric materials (c). The cross sections are taken just below the gateoxide-Silicon channel interface. 42

3.4 Lateral potential profile of the channel in a DG-TFET at the surface (a) at weak inversion (b) at strong inversion. Lines indicate the model and symbols TCAD simulations. The cross sections are taken just below the gateoxide-Silicon channel interface. V_G = 0, 0.25, 0.5, 0.75, 1V 44

3.5 Two dimensional electrostatic potential solution of a double gate Tunnel FET 1nm below the gateoxide- channel interface from the conformal mapping technique. 45

3.6 (a) Transfer characteristics with different V_D and (b) output characteristics at different V_G. The dotted lines in the output characteristics show the curve with eqn.3.4. Rest of the plot uses eqn.3.5 with the transition function from eqn.3.8 . 46

3.7 (a) Transfer characteristics with varying gate dielectric constant. $\alpha = 0.8$ for HfO_2, 0.92 for Al_2O_3 and 1 for SiO_2 gate dielectric. (b) Transfer characteristics with varying source doping level. Low doping curves were corrected according to table 3.1. $L_g = 22nm$, $t_{Si} = 10nm$. 47

3.8 Conduction band energy profile at the source channel boundary for different source doping levels. The source depletion width increases with decreasing source doping. 48

3.9 Impact of source doping on the TFET output characteristics. The same transition function of eqn.3.8 is used here. 49

3.10 (a) Cross section of the device under study with typical dimensions marked (b) lateral potential profile of the channel in a DG-TFET at the oxide-Silicon interface at strong inversion for V_{GS}= 0,0.25,0.5,0.75,1V. Lines indicate the model and symbols TCAD simulation in all relevant plots. 50

3.11 The operating regimes of a TFET: ON state with $V_{DS} = V_{GS} = V_{DD}$, where supply voltage $V_{DD} = 0.75V$ 51

3.12 The operating regimes of a TFET: Ambipolar state with $V_{DS} = V_{DD}$ and $V_{GS} = -V_{DD}$, where supply voltage $V_{DD} = 0.75V$ 51

3.13 The operating regimes of a TFET: Breakdown state with $V_{DS} = V_{DD}$ and $V_{GS} = V_{DD}$, where supply voltage $V_{DD} = 0.75V$ 52

3.14 Schematic showing the location of the two tunnelling paths considered. The breakdown state biasing is used here to show both the tunnelling paths. 52

3.15 (a) Transfer characteristics for different V_{DS} (b) Output characteristics for different V_{GS}. 54

3.16 (a) Transfer curves with three different gate oxide material. (b) Transfer curves with gate length scaling with other parameters constant. 54

3.17 (a) Transfer curves with three different body thickness. (b) Transfer curves with source doping variation. 55

3.18 (a) Measured C_{GS}, C_{GD} and C_{GG} with respect to V_{GS}. Perfect symmetry is observed highlighting the ambipolar nature of TFETs. (b) Measured Gate-Drain capacitance for different V_{DS}. Inset shows measured C_{GD} curves with respect to V_{GS}. (c) Measured Gate-Source capacitance for different V_{DS}. Source contribution to inversion electrons is negligible even at $V_{DS} = 1.5V$ with device turned ON. Inset shows measured C_{GS} curves with respect to V_{GD}. Fully depleted SOI TFET with $L_G = 200nm$, $t_{Si} = 21nm$ and 6 nm SiO_2 dielectric 56

3.19 (top) TFET A simulated with artificially enhanced tunnelling injection to have 2 order of magnitude more current than TFET B. (bottom) TFET A shows a slightly higher C_{SG} and slightly lower C_{DG} compared to TFET B. C_{SG} remains negligible compared to C_{DG} in TFET A, verifying that tunnelling has little or negligible influence in the charge distribution of a TFET. 58

3.20 Modelled gate, drain and source charges of an n-type double gate tunnel FET (top) as a function of gate-source voltage and (bottom) as a function of drain-source voltage. 59

3.21 C_{GG}, C_{SG} and C_{DG} with respect to drain-source voltage at $V_{GS} = 1.5V$ (left) and gate-source voltage at $V_{DS} = 1.5V$ (right) computed by the simplified model and comparison with TCAD simulations. Symbols indicate simulations and solid lines indicate model for all relevant figures. 60

3.22 Modelled C_{GG}, C_{SG} and C_{DG} curves with respect to V_{GS} for three different drain voltages and comparison with TCAD simulations. Modelled C_{SG} curves with respect to V_{GS} although do not show a good match, predicts the trend correctly 61

3.23 Modelled output characteristic of a n type TFET showing both positive and negative drain voltages. Note the forward biased drain current has not gate dependence. 62

3.24 Schematics of circuits simulated (a) single stage inverter (b) 2-bit half adder (c) A ring oscillator (3 stage shown here) 63

3.25 (a) VTC of a TFET based single stage inverter with V_{DD} 0.1V to 1V. (b) Voltage gain of the inverter for the same V_{DD} range. Inset shows the simulated single stage inverter. (c) Transient response of a single stage TFET inverter at V_{DD}=1V with C_L=0fF and C_L=2fF. 64

3.26 All-Silicon 22nm TFET based 2-bit half adder with V_{DD}=1.2V (a) the two input signals V_A and V_B (inset shows the schematic of the simulated 2-bit half adder) (b) the sum and carry output signals. 65

3.27 (a) All-Silicon 22nm TFET based 15 stage ring oscillator output for V_{DD}=1V. A delay of τ_D=79.4ns is observed. (b) Leakage, dynamic and total energy with V_{DD} scaling (c) Switching energy-performance comparison. 66

3.28 (a) Schematic of the fabricated sSi NW TFET with TiN/HfO_2 gate stack (b) Highly doped $n+$ and $p+$ pockets at the silicide edges are formed after a low temperature anneal (c) sSi NW TFET fabrication process using tilted $B+$ and $P-$ ion implants into epitaxial $NiSi_2$ S/D contacts. (d) SEM image of single sSi NW TFET after fabrication. (e) SEM cross section along the NW showing the gate oxide and metal layers. [43] 67

3.29 SEM cross section along the NW showing the gate oxide and metal layers. [image courtesy: ST-Microelectronics] 68

3.30 (a) Measured transfer characteristics and calibrated model at different V_{DS} of the fabricated sSi NW TFET for (left) p-type and (right) n-type TFET (b) Measured output characteristics and calibrated model at different V_{GS} for (left) p-type (right) n-type TFET at room temperature. (Symbol: measurement, lines: model) 70

3.31 (a) Measured transfer characteristics with temperature (25°C to 100°C) with calibrated model; achieved I_{ON} is $20\mu A/\mu m$ and $2\mu A/\mu m$ for p- and n-type sSi homo-junction NW TFET at $V_{GS} = 1.5V$ and min point swings, $SS_{pt} \approx$ 70-100mV/dec at room temperature and $I_{OFF} < 1-10nA/\mu m$. (b)Measured transconductance and calibrated model at $|V_{DS}| = 0.1V$ with respect to temperature for p type (left) and n type (right) TFET. (Symbol: measurement, lines: model) 71

3.32 (a)(Left) experimentally extracted V_{TH} and average sub-threshold swing, SS_{AVG}, (over 4 decades of current) dependences, and, (right) I_{ON} and I_{OFF} dependences on temperature, from 25°C to 125°C. (b)Transfer characteristics of 28nm CMOS (calibrated simulation) and TFET (calibrated model) overlapped with respect to temperature. 72

3.33 (a)Transconductance per unit drain current for p type (left) and n type (right) with respect to VDS. (b)Intrinsic gain vs. gate voltage for p type (left) and n type (right). The calibrated model is used for the TFET plots. 73

3.34 (a)Measured (TFET) current gain cut-off frequency Vs current per unit width for p type (left) and n type (right) TFET (b)Measured (TFET) unity gain frequency per unit power vs. current per unit width for p type (left) and n type (right) TFET. 74

3.35 Schematics of investigated analog cells: (a) current mirror circuit (b) differential pair circuit and (c) a diode connected transistor circuit. 75

3.36 (a)TFET vs. CMOS current mirror: I_{OUT}/I_{IN} for temperature 25°C to 125°C (left) Differential pair output current as a function of temperature (25°C to 125°C) for both nTFET and nMOS (right) (b) Temperature dependence of V_{OUT} with I_{SS} as parameter; CMOS vs TFET for a diode connected circuit. TFET embodiment shows excellent linearity versus temperature, suggesting possible use a highly linear temperature sensor. 76

3.37 (left) Schematic of TFET differential stage circuit with an active load and (Right) gain (in dB) versus current, for CMOS and TFET and (right) showing the phase and the gain versus input current. The gain drops by 30 dB for the CMOS at room temperature. 77

4.1 Component power distribution, 2005 two-socket dual-core server.[84] 80

4.2 Component power distribution, 2008 four-socket quad-core server.[84] 81

4.3 Schematic cross-section of a partially depleted (PD) SOI nMOSFET used as a 1T-DRAM memory cell and the corresponding current transient for 0 state (undershoot) and 1 state (overshoot) resulting from the lack or excess of majority carriers (holes), respectively. Image courtesy [88] 82

4.4 Simulated potential profile at 1 nm below the gate-oxide/channel interface for both DG-MOSFET and DG-TFET. 83

4.5 (a) SEM image of a fabricated FD-SOI Tunnel FET showing the top gate partially covering the channel [23]. (b) Independent gate FinFET device structure for TFET based capacitorless DRAM. 84

4.6 2D cross-section of the Fin-TFET showing the hole density at the end of a write 1 operation (top). Hole density and electrostatic potential at a cross section 1nm above the Gate2 showing the formation of potential well/hole pocket (bottom) $V_{G1} = 0V$, $V_{G2} = -10V$, $V_S = 0.5V$. 86

4.7 (a) Simulated transfer curves at various Gate2 bias. (b) Simulated hysteresis curves observed with varying L_{IN} with same biasing conditions. Devices with longer L_{IN} shows stronger hysteresis. $V_{G1} = 2V, V_D = 1V$. 87

4.8 (a) Different relaxation times in the millisecond range are recorded when the back gate is biased at $V_{G2} = -10V$ (after writing 1), depending on the length of the L_{IN} region. The TFET with $L_{IN} = 500nm$ shows the longer discharge time while it is negligible in devices with $L_{IN} = 200nm$. (b) & (c) Measured drain current with respect to back gate voltage at fixed front gate and drain voltages, $V_{G1} = 4.5V, V_D = 4V$. 88

4.9 Simulated 2D hole densities after a WRITE 1 (a) and WRITE 0 (b) operation. (c) Simulated potential profile at 1nm below Gate1 showing the hole pocket. . . . 90

4.10 (a) Measured transfer characteristics with different Gate2 voltage at 25°C and 85°C temperature. As expected for a TFET the sub-threshold swing has a negligible temperature dependence. (b) Measured relaxation times as a function of temperature after a READ 1 operation. A minor dependence of the relaxation behavior on the temperature is experimentally demonstrated. 91

4.11 Timing diagram for different operations described in Table 4.3. Consecutive READ and WRITE operations for both states. The difference in READ current for states 1 and 0 is 10nA, as highlighted, a memory effect is observed. Programming voltages V_{READ}, V_{ERASE} and V_{PROG} are as defined in table 4.3 for TFET with $L_{G1} = 400nm, L_{IN} = 200nm$ and SiO_2 gate dielectric. 92

4.12 (a) Measured transfer characteristics, $ID - V_{FG}$ with back gate potential as a parameter, for $L_G = 400nm$, $L_{IN} = 200nm$. (b) and (c) Measured drain current with respect to back gate voltage at fixed front gate and drain voltages, $V_G = 4.5V$, $V_D = 4V$. A history effect (open hysteresis loop) is observed more prominently in the TFET device with $L_{IN} = 500nm$ 93

4.13 Consecutive ERASE/READ and WRITE/READ operations for a TFET (with $L_G = 400nm$, $L_{IN} = 200nm$ and HfO_2 gate dielectric) based capacitorless memory. (a)–(c) show the front gate, back gate, and the drain pulse scheme. (d) and (e) show the source current at two different frequencies sampled every $60\mu s$. Based on the previous WRITE/ERASE state, a difference of $20nA$ in the READ current (I_S) for states “1” and “0” is observed. 95

4.14 3D schematic of the proposed new structure (n-type) including a p+ doped pocket between the channel and the drain. 96

4.15 3D schematic of the proposed new structure (n-type) including a p+ doped pocket between the channel and the drain. 97

4.16 TCAD simulated 2D cross-section of the proposed new structure showing the hole density at a HOLD state after and WRITE "1"(top) and WRITE "0"(bottom) operation. 98

4.17 Continuous read/write cycle with hold time in between. A difference of $500nA$ is observed between the two memory states. 99

5.1 Comparison between the schematic cross-section of a floating gate device (a) and of a SONOS device (b). 102

5.2 (a) Programming via hot electron injection (b)erase via FN tunneling in a floating gate memory cell (source: Kiethley) 103

5.3 Storage capacity of prominent flash memory technologies over the years. [115] 105

5.4 SEM imaage of a close-up of V-NAND flash array from Intel [courtesy Chipwerks] 106

5.5 Schematic cross-section of the conventional MOSFET based SONOS structure. 107

5.6 Concentration of electrons during the program cycle that tunnel into the nitride region of the traditional SONOS device. 108

5.7 Trapped electron and hole charge in nitride region of SONOS device as a function of time during the program and erase cycles; corresponding applied gate bias is also shown. 109

5.8 ID -VG curves for program and erase states for n type MOFET and conventional lateral tunneling Tunnel FET. 110

5.9 Energy band diagram of the TFET-based memory. All the different layers are indicated. The desired hole-direct-tunneling process and the parasitic electron contribution are depicted. 111

5.10 Schematic cross section of the p-type vertical TFET architecture used in this study. 112

5.11 Transfer characteristic of the three vertical n & p-type TFETs with varying gate-source overlap length. The source is grounded and $V_D = 1V$. Note that the device with the longer source length offers better performances in terms of I_{ON} current, sub-threshold slope and ambipolar behavior 113

5.12 Schematic cross section of the final device. All the different layers are indicated (except for the oxide spacers) 113

5.13 TCAD simulation result showing in (a) the electron band-to-band generation at the end of the programming step and (b) the trapped electron concentration close to the interface between the tunneling oxide and the charge trapping layer. 114

5.14 Trapped electron and hole charge in nitride region of the memory device as a function of time during the program and erase cycles; corresponding applied gate bias is also shown (a). Hole band-to-band-tunneling generation and hole barrier-tunneling generation as a function of time during the programmed erase cycles (b). 116

5.15 $I_D - V_G$ curves for program and erase states for both n & p type vertical TFET and MOSFET non-volatile memory. 117

5.16 Comparison of retention characteristic of a n-type MOSFET and n-type Tunnel FET based SONOS flash memory. 117

5.17 (a)Simulated energy band diagram of a $SiO_2/HFO_2/Al_2O_3$ stack at $V_G = 10V$(b) Simulated energy band diagram of a $SiO_2/ZrO_2/Al_2O_3$ stack at $V_G = -9V$ . . . 118

5.18 (a)Schematic of cross-section of the fabricated MOS capacitors. (b) Measured capacitance voltage characteristics of a $SiO_2/HfO_2/Al_2O_3$ gateoxide stack at 1MHz frequency. A hysteresis of 10V is observed for a gate votage sweep of -20V to 20V. 120

5.19 Measured capacitance voltage characteristics of a $SiO_2/ZrO_2/Al_2O_3$ gate-oxide stack at 1MHz frequency. A reduction of the hysteresis window was observed on repeated measurements. 121
5.20 Process flow of proposed vertical tunneling TFET based non-volatile memory. (image not to scale) 124
5.21 Optical microscope view of the wafer surface after the two implantations. The n and p type implantations are clearly visible due to the difference in colour. . . 125

B.1 Four-Terminal Pair Measurement Principle in a LCR bridge. 133
B.2 Four-terminal pair configuration measurement contacts. 134

C.1 Connection schematic showing the three function generators, the parameter analyzer and the oscilloscope. 137

List of Tables

2.1 Non-Local Band-to-Band Tunneling Parameters for Sentaurus non-loca tunneling model. 20
2.2 Non-local band-to-band tunneling parameters for Silvaco TCAD. 21
2.3 Summary of calculated A_{path} and B_{path} values from mr. 31

3.1 Empirical conduction band shift for different doping levels 48
3.2 Model parameters available in Verilog-A 62
3.3 Truth table for an inverter 62
3.4 Truth table for a 2-bit half adder 63
3.5 Table of fitted parameter list for Verilog-A based model for both n & p type device at room temperature. 69

4.1 Components in 2005 and 2008 Test Systems [84] 80
4.2 Proposed programming scheme of Fin-TFET based capacitorless DRAM. 89
4.3 Programming conditions for indicated operations of capacitorless 1T TFET DRAM with $L_{G1} = 400nm$, $L_{IN} = 200nm$. 91
4.4 Programming conditions for indicated operations WRITE “1” and READ of capacitorless 1T TFET DRAM with $L_{G1} = 400nm$, $L_{IN} = 200nm$. The ERASE also corresponds to WRITE “0.” 93
4.5 Programming conditions for indicated operations of capacitorless 1T TFET DRAM with a doped pocket and $L_{G1} = 400nm$, $L_{IN} = 200nm$. 99
4.6 Parameters of non-local band-to-band tunneling model at 25°C, 55°C and 85°C fitted with measurements [105]. 100

5.1 Programming scheme for the proposed p-type memory cells. 108
5.2 Parameters of the device simulated. 114
5.3 Summary of programming scheme for n and p-type devices. 115

Acronyms

ALD Atomic Layer Deposition
BHF Buffered Hydro Fluoride
BOX Burried Oxide
BSIM Berkeley Short Channel IGFET Model
BTBT Band-to-Band Tunneling
BT Barrier Tunneling
CB Conduction Band
CCCS Current Controlled Current Source
CG Control Gate
CHE Channel Hot Electron
CMOS Complementary Metal-Oxide-Semiconductor
CPU Central Processing Unit
CTFET Complementary Tunnel-FET
CV Capacitance Voltage
DIBL Drain-Induced Barrier Lowering
DG Double Gate
DOS Density of States
DRAM Dynamic Random Access Memory
EOT Effective Oxide Thickness
eQF Electron Quasi Fermi energy
FD Fully-Depleted
FDSOI Fully-Depleted SOI
FEM Finite Element
FET Field-Effect Transistor
FG Floating Gate
FIB Focused Ion Beam
FN Fowler Nordheim
FoM Figure of Merit
GAA Gate-All-Around
HDD Highly Doped Drain
hQF Hole Quasi Fermi energy
LDD Lightly Doped Drain

LTO Low Temperature Oxide
MLDA Modified Local Density Approximation
MOSFET Metal Oxide Semiconductor Field Effect Transistor
NEM Nano Electro Mechanical
RAM Random Access Memory
RCA Radio Corporation of America
ROM Read Only Memory
RTA Rapid Thermal Anneal
SCE Short Channel Effect
SE Schrödinger Equation
SEM Scanning Electron Microscope
SiNW Silicon Nanowire
SONOS Silicon-Oxide-Nitride-Oxide-Silicon
SPICE Simulation Program with Integrated Circuit Emphasis
SRH Shockley-Read-Hall
SOI Silicon on Insulator
SS Subthreshold Swing
SSD Solid State Disk
TAT Trap-Assisted-Tunneling
TCAD Technology Computer-Aided Design
TFET Tunnel-FET
UTB Ultra Thin Body
VB Valence Band
VTC Voltage Transfer Characteristic
WKB Wenzel-Kramers-Brillouin

1 Introduction

1.1 The microelectronic revolution

Micro-electronics has brought about a second industrial revolution: In 1946, when the ENIAC (Electrical Numerical Integrator And Calculator) was announced, it was heralded in the press as a "Giant Brain". Following this, was the invention of the transistor, a small, low-power amplifier which also gave a boost to the computer industry due to its small size relative to that of a vacuum tube. However, since the layout of the components of the electronic devices on a single board using wiring produced some limitations whilst trying to downsize, Geoffrey W.A. Dummer (a radar scientist) decided to connect them on a single wafer which is referred to as an integrated circuit which was then built by Jack Kilby. This means that the whole circuit will be manufactured in the surface of a thin substrate of semiconductor material. This method of integration was an enormous improvement over the manual assembly of the circuits using electronic components.

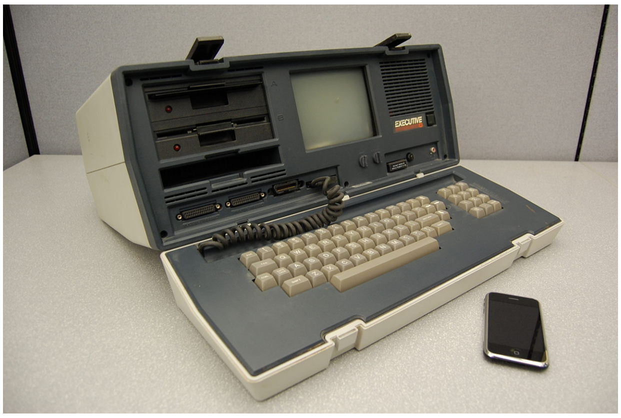


Figure 1.1 – Evolution of microelectronic technology: An Osborne Executive portable computer, from 1982 with a Zilog Z80 4MHz CPU, and a 2007 Apple iPhone with a 412MHz ARM11 CPU; the Executive weighs 100 times as much, has nearly 500 times as much volume, cost approximately 10 times as much (adjusted for inflation), and has about 1/100th the clock frequency of the smart-phone. [source wikipedia]

Integrated circuits were then however made possible by experimental discoveries which proved that such semiconductor devices could actually perform the functions of the vacuum tubes. Nowadays, integrated circuits are used in almost all of the electronic equipment in use and they have revolutionized the world of electronics. "Little did this group of onlookers know that Kilby's invention was about to revolutionize the electronics industry."

Microelectronics is nowadays considered to be the cornerstone of the computing revolution, the communications revolution and the consumer electronics revolution. Microprocessors are nowadays used in sewing machines, dish washers, washing machines, telephones, cookers, ovens, controllers for heat systems,televisions, cameras, CPUs, traffic lights, car ignition systems, accounting systems, cash terminals, banking terminals, library indexing systems and many more.

1.2 Evolution of nano-electronics

In recent years, the development of small and portable information and communication equipment has been remarkable; cell phones have evolved into mobile PCs, smart phones, and tablet PCs. The market size of such equipment is expected to double over the next few years. Silicon ultra large scale integration (ULSI) circuits are becoming increasingly important as core devices in the field of information and communication equipment, and they are expected to respond to various demands, such as high-speed operation, low power consumption, and a high level of functions. In addition, the contributions of ULSI-related technologies to the fields of environment, energy, medicine, and safety/security is required to establish a sustainable society and alleviate worldwide problems, which will lead to the development of new markets and the creation of new industries.

In 1965, Gordon Moore then R&D director at Fairchild Semiconductor and these days chairman emeritus of Intel Corp., Santa Clara, California, quantified the astounding growth of the new technology of semiconductors in a simple formula. Manufacturers, he said, had been doubling the density of components per integrated circuit at regular intervals, and they would continue to do so as far as the eye could see (Fig.1.3). This observation has since been dubbed "Moore's Law" [1] and is now enormously influential. Conventionally, improvement in the performance and functions of ULSI devices and reduction of their manufacturing cost have been realized by simultaneously achieving the high-speed operation and increased integration of devices through the downscaling of metal–oxide–semiconductor field-effect transistors (MOSFETs) according to the scaling rules [2]. Currently, transistors with a gate size of 32 nm or less are in practical use, and the downscaling technology will continue to be important as a ULSI technology.

However, in recent years, technical difficulties associated with microfabrication have increased; in addition, the costs of technological development and production have become huge. Furthermore, the downscaling of devices to the nanoscale decreases the on-state current and increases the off-state current of metal–oxide–semiconductor (MOS) transistors. These

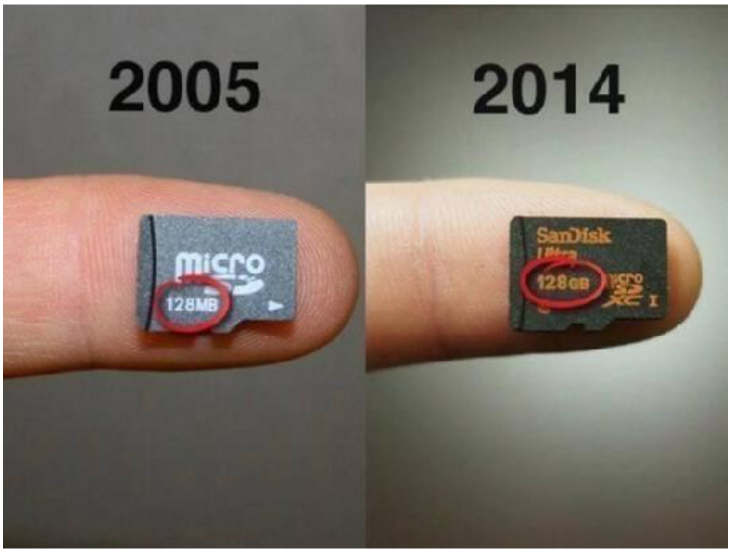


Figure 1.2 – Evolution of memory technology from 2005 to 2014. Over a span of 9 years the capacity of a micro SD card has increased by a factor of 1024. [source Wikipedia]

phenomena lead to a decrease in the current driving capability and an increase in the power consumption of complementary metal–oxide–semiconductor (CMOS) transistors, i.e., the basic logic gates of ULSI circuits, which means that the performance improvement associated with device scaling will no longer be possible.

Under such circumstances, a new direction for the technological development of ULSI devices is currently being examined. Concretely, in addition to technological development pursuing the scaling limit of ULSI devices, i.e., the conventional direction, a new direction of expanding the above-mentioned application fields of ULSI devices by combining different technologies and functions with the ULSI devices is being examined. The conventional direction is generally referred to as "more Moore", whereas the new direction is referred to as "more than Moore" in this field. Even with the "more than Moore" direction, high-performance Si platforms are used as the fundamental technology. Although the research and development of elementary devices with a new operation principle that can replace CMOS devices, called "beyond CMOS" devices, will be indispensable, such a logic gate with the desired low power consumption and high degree of integration has not yet been realized [3].

1.3 The scaling problem

At the heart of this microelectronic revolution is the complimentary-metal-oxide semiconductor(CMOS) technology. The advancements in the Silicon based CMOS technology has lead to a phenomenal growth of the semiconductor industry in the last three decades. Over the past decades, the MOSFET has continually been scaled down in size; typical MOSFET channel lengths were once several micrometres, but modern integrated circuits are incorporat-

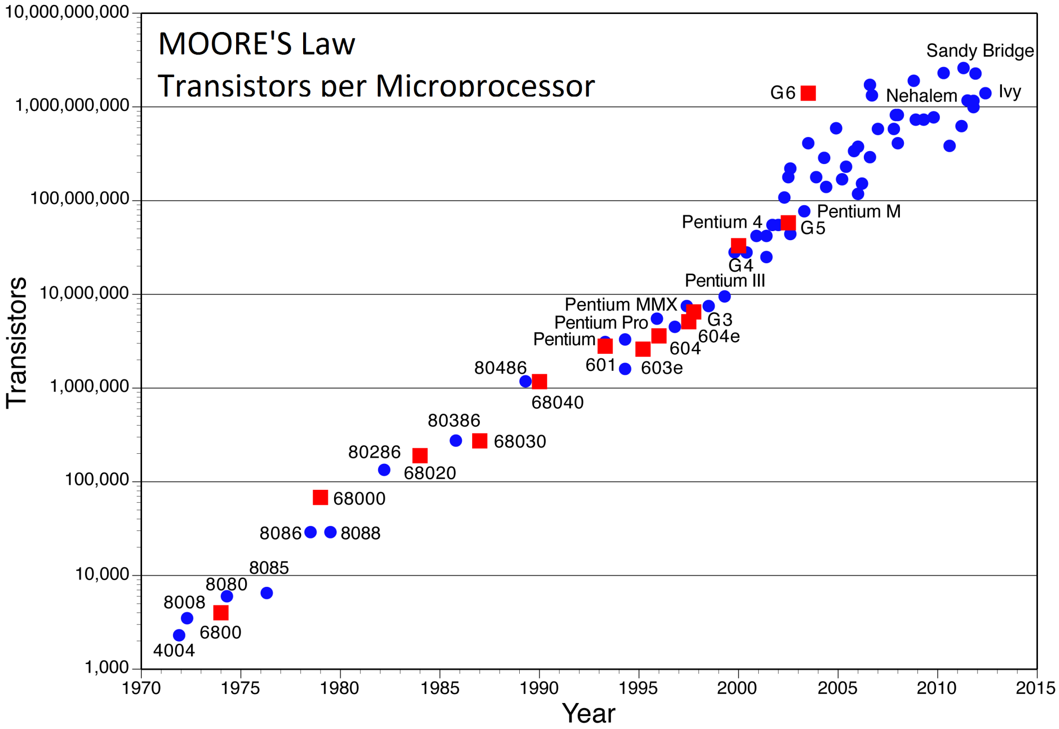


Figure 1.3 – Plot of CPU transistor counts against dates of introduction; the line corresponds to exponential growth with transistor count doubling every two years. [source wikipedia]

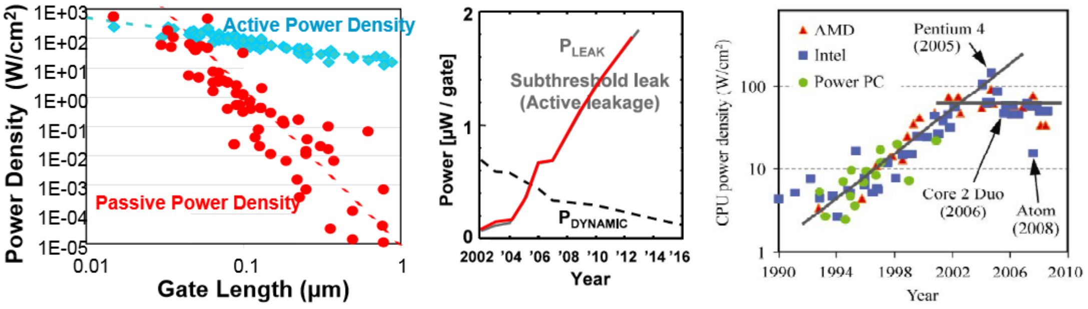
ing MOSFETs with channel lengths of tens of nanometers. Scaling of the MOSFET is governed by the need of higher speeds and package density(number of components per chip) and low cost per function on the chip. Smaller MOSFETs are desirable for several reasons. The main reason to make transistors smaller is to pack more and more devices in a given chip area. This results in a chip with the same functionality in a smaller area, or chips with more functionality in the same area. Since fabrication costs for a semiconductor wafer are relatively fixed, the cost per integrated circuits is mainly related to the number of chips that can be produced per wafer. Hence, smaller ICs allow more chips per wafer, reducing the price per chip.

It is also expected that smaller transistors switch faster. For example, one approach to size reduction is a scaling of the MOSFET that requires all device dimensions to reduce proportionally. The main device dimensions are the channel length, channel width, and oxide thickness. When they are scaled down by equal factors, the transistor channel resistance does not change, while gate capacitance is cut by that factor. Hence, the RC delay of the transistor scales with a similar factor.

1.3.1 Challenges in MOSFET scaling

Producing MOSFETs with channel lengths much smaller than a micrometre is a challenge, and the difficulties of semiconductor device fabrication are always a limiting factor in advancing integrated circuit technology. Though processes such as ALD have improved fabrication for small components, the small size of the MOSFET (less than a few tens of nanometers) has created operational problems. In fact, as the channel length goes down to values of tens of nanometers, tunneling of carriers can occur in the channel, resulting in an increase in the leakage current of the MOSFET transistor when switched off (I_{OFF}). This and other correlated phenomena which degrade the performance of a transistor with the reduction of the dimensions are commonly called short-channel effects (SCE).

Upon further decreasing the MOSFET device dimensions, the leakage current of the MOSFET increases, while the supply voltage can no longer be scaled down, both of which result in increased power consumption. The small dimensions also result in increased variability of the device performance, which compromises the chip design. It is therefore highly desirable to explore alternative device structures. As shown in Fig. 1.4(a), as the gate length reduces the passive power density approaches the active power density in magnitude. Also Fig. 1.4(c) the advancing nodes are predicted to have a much higher dynamic power than the preceding nodes.



(a) (b) (c)

Figure 1.4 – (a)Power consumption trends with gate length scaling (b) comparison of dynamic and leakage power trend over the years. [courtesy IBM coorp.] and (c) Near-exponential increase of CPU power density in recent decades has flattened with the introduction of multi-core CPUs [4]

As the transistor gate length is reduced, improved performance requires the supply voltage, V_{DD}, and simultaneously the threshold voltage, V_T, to be lowered to keep the overdrive factor ($V_{DD} - V_T$) high. As a consequence, the leakage current, I_{OFF} , increases exponentially because the SS of a MOSFET is not scalable but has a minimum value of 60 mV per decade (that is, it takes 60 mV to increase the current by one order of magnitude) at room temperature. Typical values of SS in advanced CMOS technology are close to 100 mV per decade; by lowering V_{DD} from 500 mV to 250 mV while preserving the overdrive, the leakage power has been shown to increase unacceptably by a factor of 275 in a 45-nm bulk CMOS technology. Another way of reducing the voltage supply without performance loss is to increase the turn-on steepness, which means decreasing the average sub-threshold swing, S_{avg} , defined as the average swing over 3 or 4 decades of current. Therefore, devices with a steep SS, called steep-slope switches, are expected to enable V_{DD} scaling.

1.3.2 Beyond CMOS devices: an alternative ?

In a MOSFET, the current-switching process involves the thermionic (temperature-dependent) injection of electrons, over an energy barrier. This sets a fundamental limit to the steepness of the transition slope from the off to the on state. The gate voltage required to change the drain current by one order of magnitude when the transistor is operated in the sub-threshold region is reflected in the expression of the sub-threshold swing, SS:

$$\begin{aligned} SS = \underbrace{\frac{dV_G}{d\Psi_S}}_{m}\underbrace{\frac{d\Psi_S}{d(logI_D)}}_{n} \simeq \left(1+\frac{C_d}{C_{ox}}ln10\frac{kT}{q}\right) \\ \longrightarrow \frac{kT}{q}ln10 \simeq 60mvdecade^{-1} = 300K \end{aligned} \tag{1.1}$$

where V_G is the gate voltage, I_D is the drain current, kT/q is the thermal voltage, and C_d and C_{ox} are the depletion and the oxide capacitances, respectively. The term m is the transistor body factor, and n is a factor that characterizes the change of the drain current with the surface potential, Ψ_S , reflecting the conduction mechanism in the channel. A sub-thermal SS would be less than $kT/qln10$ and could be obtained by using new physical principles rather than thermionic injection.

Many device innovations to lower SS below the MOSFET thermal limit, by decreasing the factors m and n in equation 1.1, have been proposed. The two main techniques are summarized below:

- m (body factor) less than 1 : active gate devices. This can be achieved by using the recently proposed negative-capacitance FET (NC-FET) or micro or nano-electromechanical (M/NEM) movable electrodes in M/NEM FET or NEM relay devices.
- n less than (kT/q)ln10 : new channel injection mechanisms other than thermionic. Reducing n to achieve a subthermal SS involves a modification of the carrier-injection mechanism. For example quantum-mechanical band-to-band tunnelling (BTBT) in Tunnel FETs or impact ionization in IFETs, both of which are capable to show sub 60 mV/decade sub-threshold slope.

In this thesis we will concentrate on the second option of achieving a sub-thermal SS i.e. devices with n less than (kT/q)ln10. Among the examples named, a promising candidate to replace the MOSFET in future technology nodes is the tunnel field-effect transistor (TFET). Due to its built-in tunnel barrier, the TFET does not suffer from short-channel effects which are deteriorating the off-current of MOSFETs. Down to channel lengths of 10 nm, low off-currents are expected for silicon TFETs. Another advantage of TFETs is that the sub-threshold slope can be smaller than 60 mV/ decade, the physical limit of MOSFETs, such that the supply voltage may be further reduced. At the same time, the structural similarity of the TFET with the MOSFET has resulted in TFET implementations with standard complementary metal-oxide semiconductor processing techniques. However, the on currents of TFETs are much smaller than that of MOSFETs, such that the TFET switching speed is smaller.

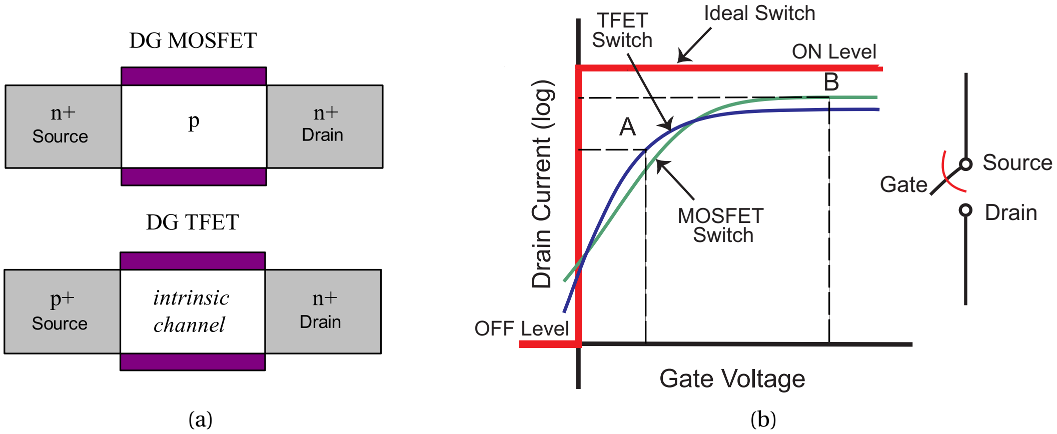


Figure 1.5 – (a)Comparison of an n-type double gate MOSFET and Tunnel FET schematic. The different source doping is to be noted. (b)Qualitative comparison of a conventional MOSFET (green) and a Tunnel FET (blue), which has a steep off–on transition and the lowest I_{OFF}.

Figure 1.5 shows a qualitative comparison of a MOSFET and a Tunnel FET that use quantum-mechanical tunnelling. At moderate performance requirements, such as operation point A, TFETs offer not only improved I_{ON}/I_{OFF} , but also superior performance (higher I_{ON} at the same voltage) or power savings at the same performance (lower voltage for the same I_{ON}) over MOSFETs. However, when a much higher performance is required, such as at operation point B, a MOSFET is the better solution. At operation point A, because of its sub-thermal sub-threshold swing, the TFET offers not only an improved I_{ON} / I_{OFF} but also a superior performance and a power saving at the same performance as a MOSFET. At operation point B, corresponding to higher performance, the MOSFET switch becomes the better solution.

The energy efficiency of a logic operation can be evaluated by analysing its switching energy diagram (Fig. 1.6), showing the balance of the dynamic, $E_{dynamic}$, and the leakage, $E_{leakage}$, components of the total switching energy, E, versus the V_{DD} :

$$E_{total} = E_{dynamic} + E_{leakage} \tag{1.2}$$

$$= \alpha L_d C V_{DD}^2 + L_d I_{OFF} V_{DD} \tau_{delay} \tag{1.3}$$

where L_d is the logic depth, C is the switched capacitance, τ_{delay} is the logic activity factor (typically ~ 0.01). From eqn. 1.3, it seems that CMOS logic has a lower limit in energy per operation, E_{min}, owing to the exponential increase of the sub-threshold leakage, I_{OFF}, with V_{DD} scaling. As can be seen from Fig.1.6, Tunnel FET have an inherent advantage of energy

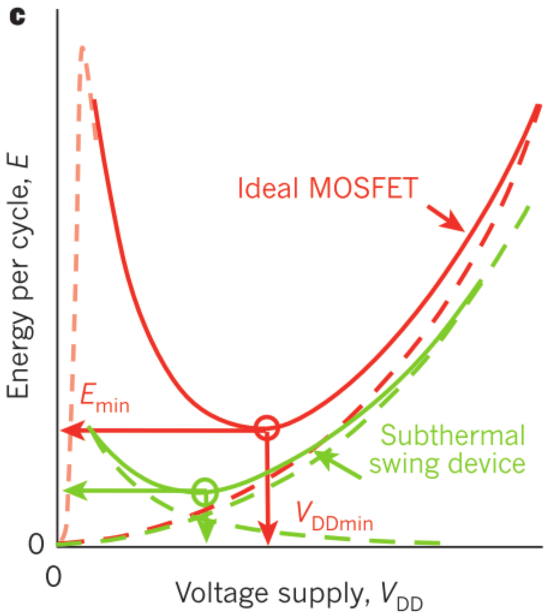


Figure 1.6 – Comparison of the minimum switching energy, E_{min} , and the corresponding voltage supply, VDD_{min} , for a subthermal swing device (S < 60 mV decade^{-1} , green curve) and the ideal MOSFET (S = 60 mV decade^{-1} , red) at the same I_{ON}/I_{OFF}. (source [5])

efficiency. This is primarily due to the lower leakage energy in the Tunnel FET. Also it should be noted that Tunnel FETs can be operated a much lower V_{DD} to a minimum energy point. The steep-swing TFET offers better energy efficiency at lower or moderate performance level.

1.4 Thesis overview

The aim of this work has been the investigation of homo-junction Tunnel Field Effect Transistors from a compact modelling perspective to its applications. A compact DC/AC model has been developed which is capable of describing the I-V characteristics in all regimes of operation. The model was further implemented in a Verilog-A based circuit simulator. Following calibration to experimental results of Silicon and strained-Silicon TFETs, the model has been also used to benchmark against a standard CMOS node for digital and analog applications. In the same work, we also explore and propose solutions for using TFETs for low power memory applications.

The work is structured in six chapters, including this introduction. Chapter 2 is devoted to the explanation of the principle of tunneling. An overview of the simulation tools available for simulation of ban-to-band tunneling are presented. A comparison of the two main types of models available in the simulations tools are also presented. The working principle of a Tunnel FET , for both n and p-type devices are presented next. A short overview of the calibration methodology of the non-local tunneling model in Sentaurus TCAD to experimental results is presented. Finally the Miller effect in Tunnel FETs are explained and ways to mitigate this effect are also discussed.

Chapter 3 deals with the compact modelling of the double gate Tunnel FET. Two versions are presented. The first being a non-local model adapted from the model described in Sentaurus TCAD. Secondly a fully analytical compact model was developed based on the Kane's model. This model can describe all tunneling related physics in all the modes of operation. The ambipolar behaviour, super-linear onset, drain side breakdown and temperature dependence was captured in this model. The model was then coded in Verilog-A to be used in a commercially available circuit simulator. Basic digital circuit blocks like an inverter, half-adders and a ring oscillator was simulated with this model. Finally the model was calibrated to measurements of a 350nm strained-Si nanowire Tunnel FETs on FD SOI wafers. The calibrated model was scaled down to 30 nm and then used to benchmark against a 28nm CMOS design kit from ST Microelectronics.

Chapter 4 deals with the application of Tunnel FETs as a capacitor-less DRAM, which is a form of volatile memory. The principle of operation is explained via TCAD simulations. The concept is verified with experimental results on all-Silicon TFETs in FD-SOI wafers. An improved structure in the form of a finfet with a doped pocket is also proposed which will enhance the memory operation. Finally the device concept was attempted to be fabricated in our clean room in CMi EPFL. A process flow for an independent gate FinTFET was developed to realize this.

Chapter 5 presents the application of Tunnel FETs as a non-volatile memory cell. It is shown that a vertical tunneling architecture with oxide-nitride-oxide stack boosts the memory window as compared to a lateral tunneling device or a conventional MOSFET device. The band-to-band tunneling generated carriers which are generated due to vertical tunneling contribute to the charges tunneling into the charge storage layer. Various gate-oxide stacks as an alternative to the O-N-O stack were fabricated as MOS capacitors and their hysteresis was compared. Finally a process flow was developed to fabricated the vertical tunneling device along with the high-k gate stack.

Chapter 6 summarizes the main contributions of the thesis and lists all the major achievements. Finally, some of the future perspectives related to possible applications of such nano-electronic devices and their usage in the near and far future are also briefly discussed. Enjoy the read.

2 Tunnel FET

In this chapter we introduce the underlying physics of a TFET. We discus in detail the tunneling phenomena as proposed originally by Kane [6] in 1961. An overview of all the simulation tools available to simulate band-to-band tunneling is presented. Further we discus the non-local B2B tunnelling model in detail. The basic working principle of Tunnel FET is discussed. Finally we discuss the importance of Miller effect specifically to Tunnel FETs and its impact on circuit behaviour. We also propose a new device architecture to mitigate this problem.

2.1 Band to band tunneling

In TFETs tunneling of interest is band-to-band tunneling. For band-to-band tunneling to occur, an electron in the valence band of semiconductor tunnels across the band gap to the conduction band without the assistance of traps. The band gap acts as the potential barrier that the particle tunnels across.

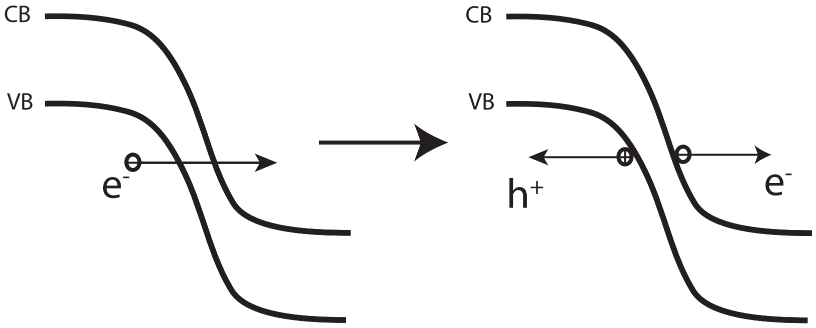


Figure 2.1 – Generation of an electron and hole pair as a result of band-to-band tunneling.

There are primarily two types of tunneling (a)Direct tunneling: Happens in direct semiconductors where the conduction band minima and the valence band maxima lie at the same k value. In direct tunneling, the k vector normal to the tunneling is conserved.

Tunneling occurs from the top of the valence band to the conduction band minima in the Γ valley. An electron travels from the valence band to the conduction band without the absorption or emission of photon in direct tunneling. For direct tunneling, the requirement for conservation of perpendicular momentum causes an increase in the tunneling. A particle with some perpendicular momentum in the valence band must tunnel to state with the same perpendicular momentum in the conduction band, which results in a longer tunneling path.

(b)Phonon assisted tunneling: A tunneling particle acquires a change in momentum by absorbing or emitting a phonon in the indirect tunneling process. In indirect semiconductors whose gamma-centered direct band gap $E_G(E_\Gamma)$, is much greater than their indirect band gap, E_G indirect tunneling is the main tunneling process. The direct tunneling process is negligible in indirect band gap materials like Silicon because the transmission probability decreases rapidly with increasing barrier height.

The electron tunneling through the band gap is akin to particle tunneling through a potential barrier, and the most probable tunneling path the smallest barrier. In the indirect tunneling process, the phonon does impart or absorb a change in the momentum of the particle. Therefore, the electron, phonon interaction of the indirect tunneling process decouples the perpendicular momentum of valence band and conduction band. An electron in the valence band can tunnel to any state in the conduction band such that energy and perpendicular momentum are conserved. 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.

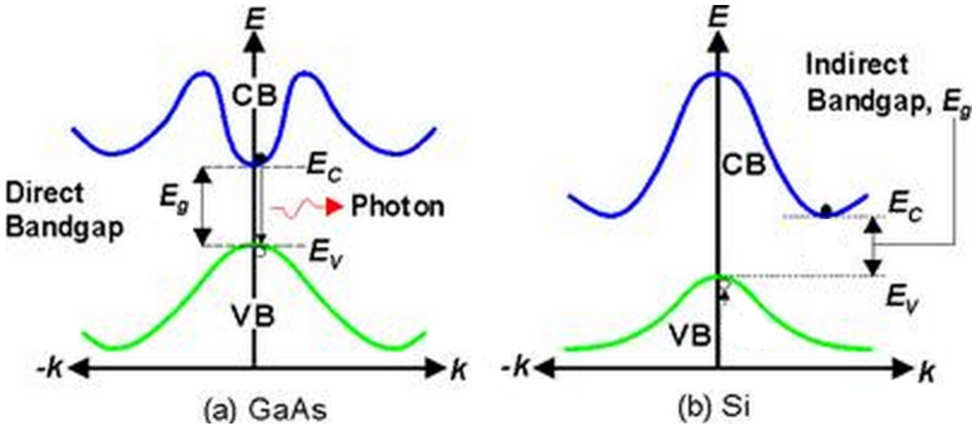


Figure 2.2 – Direct and indirect semiconductor band structure (a)GaAs (direct) (b)Silicon (indirect) [7]

2.1.1 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) [6]. It should be noted that Keldysh derived the same integral for the tunneling probability one year earlier [8], but it can be said that Kane's derivation is simpler while being more rigorous.

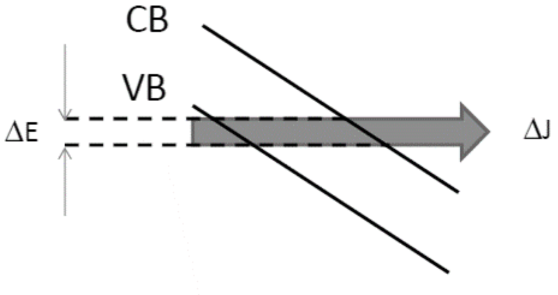


Figure 2.3 – Schematic showing energy window and ΔJ (current) [9]

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 we derive here Kane's equation for direct tunneling in a simpler way using Landauer formula and WKB approximation. We start with the Landauer formula [10]:

$$\begin{aligned} G_{B2B} &= |F|\frac{\Delta J}{\Delta E} \\ J &= \frac{2e}{2\pi\hbar A_{NORM}}\int\sum_{k_\perp} T(E,k_\perp)[f_v - f_c]dE \end{aligned} \tag{2.1}$$

where F is the electric field, E is the energy, G_{B2B} is the band to band generation, $T(E,k_\perp)$ is the tunneling probability, A_{NORM} is the normalizing area, f_c and f_v are the conduction and valence band occupancy factors and $k_\perp$ is the normal wave vector. As we consider field to be constant, we have the same ΔJ in each ΔE window. Also the electron and hole generation rates are both equal to G_{B2B}.

To simplify we can set $f_c = 0$ and $f_v = 1$. Since the structure is uniform, the transmission probability T does not depend on E converting the sum over $k_\perp$ to an integral, we get:

$$G_{B2B} = |F|\frac{\Delta J}{\Delta E} = \frac{e|F|}{\pi\hbar A_{NORM}}\sum_{k_\perp} T(E,k_\perp)dk_\perp \tag{2.2}$$

$$T = \frac{\pi^2}{9}exp\left(-2\int Im(k_x)dx\right) \tag{2.3}$$

The tunneling probability is computed under the WKB approximation as shown in eqn. 2.3. The integral of $Im(k_x)$ has to computed over a path conserving the total energy E and the normal wave vector $k_\perp$ [9]. To evaluate the integral of eqn. 2.3 we need an E-**k** relation valid in the band gap and linking the valence band states to the conduction band states. Kane's two band dispersion relation as shown in eqn. 2.4 is used for this purpose.

$$E_\pm = \frac{E_G}{2} + \frac{\hbar^2 k^2}{2m_0} \pm \frac{1}{2}\sqrt{E_G^2 + \frac{E_G\hbar^2 k^2}{2m_r}} \tag{2.4}$$

$$E = E_\pm - e|F|x \tag{2.5}$$

$$Im(k_x) = \sqrt{k^2 + (k_\perp)^2} \tag{2.6}$$

The total energy is given by the eqn. 2.5, the imaginary part of the wave vector are defined as in eqn. 2.6 [9]. Finally neglecting the second term in eqn. 2.4 , which is negligible in the band gap, setting $x = 0$ where $E = 0$ and combining eqns. 2.4, 2.5 and 2.6 we have the expression for $Im(k_x)$ as:

$$Im(k_x) = \sqrt{\frac{m_r}{E_G\hbar^2}}\sqrt{E_G^2 + E_G\frac{\hbar^2|k_\perp|^2}{m_r} - \left(e|F|x - \frac{E_G}{2}\right)^2} \tag{2.7}$$

$$T = \frac{\pi^2}{9}exp\left(-\pi\frac{\sqrt{m_r}E_G^{3/2}}{2\hbar e|F|}\right)exp\left(-\frac{\pi\hbar|k_\perp|^2}{2e|F|}\sqrt{\frac{E_G}{m_r}}\right) \tag{2.8}$$

Finally putting all equations together we have the expression for the band-to-band tunneling generation rate as derived by Kane and shown in eqn. 2.10 below:

$$G_{B2B} = \frac{e|F|}{4\pi^3\hbar}\int T(E,k_\perp)dk_\perp \tag{2.9}$$

$$G_{B2B} = \frac{e^2|F|^2\sqrt{m_r}}{18\pi\hbar^2\sqrt{E_G}}exp\left(-\frac{\pi\hbar|k_\perp|^2}{2e|F|}\sqrt{\frac{E_G}{m_r}}\right) \tag{2.10}$$

In this equation, F represents the (uniform) electric field strength, m_r represents the reduced mass and E_G is the band-gap energy. Commonly, this equation is reduced to the following to allow for a two variable calibration to experimental data, as shown in equation 2.11.

$$G_{B2B} = A_{path}|F|^2 exp\left(-\frac{B_{path}}{E_G}\right) \tag{2.11}$$

Keldysh showed that this equation was only relevant for direct tunneling (no phonon assistance), and that a slightly modified version is necessary to account for the phonon's effects in indirect tunneling [8]. Equation 2.12 shows this modified form, where the power of the field F is modified from 2 to 2.5.

$$G_{B2B} = A_{path}|F|^{2.5} exp\left(-\frac{B_{path}}{E_G}\right) \tag{2.12}$$

In both cases, the A_{path} is the linear parameter, and B_{path} is the exponential parameter, allowing relatively straightforward calibration as will be shown in section 2.4 and chapter 3.

2.2 Overview of Simulation tools available

TCAD simulation tools allow for calibration of models to experimental data, optimization of device designs, and variability analysis of designs. There are two main commercially available TCAD simulation tools: Synopsys Sentaurus [11] and Silvaco Atlas [12]. Both of these 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. There are two main classification of models as available for use in simulators: local and non local tunneling model.

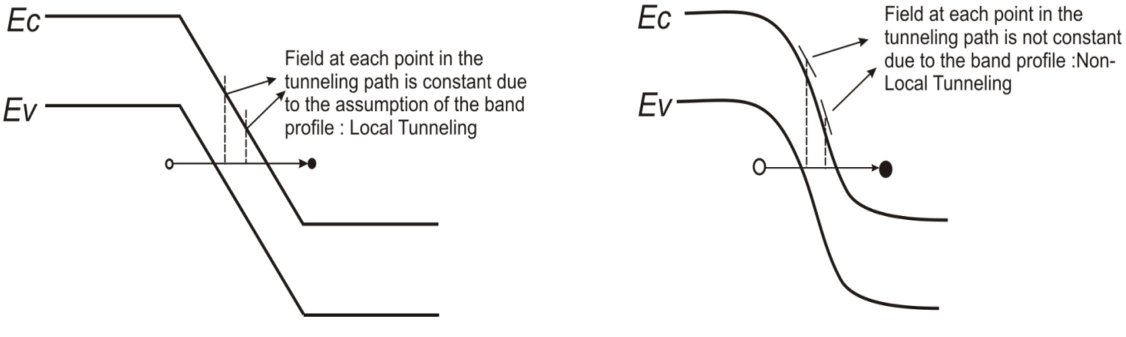

Ec
Ev
Field at each point in the tunneling path is constant due to the assumption of the band profile : Local Tunneling
Ec
Ev
Field at each point in the tunneling path is not constant due to the band profile :Non-Local Tunneling

Figure 2.4 – Schematic showing the difference between local (left) and non-local tunneling model

In local tunneling as shown in Fig.2.4(left) the energy barrier is approximated by a triangular barrier. Hence the electric field which is the gradient of the energy in the band diagram is constant at each point in the tunneling path. Non-local tunneling model does not make this approximation. In reality the electric field is far from being uniform and changes at each point in the tunneling path as can be seen in the schematic in Fig.2.4(right). Another important difference is that the local electron and hole generation profiles are the same, whereas in non-local model the holes are generated at the beginning of the tunneling path and electrons at the end of the tunneling path.

Sentaurus Device can simulate one, two, and three-dimensional structures [11]. It contains several tunneling models, including both local and nonlocal models. We show some examples of the band-to-band tunneling models available in commercially available TCAD simulators from Synopsys and Silvaco in the following section.

2.2.1 Sentaurus Local Tunneling model

The most basic local mode available in Sentaurus is the Kane's model. The generation rate is given by

$$G_{B2B} = A_{path}|F|^{P} exp\left(-\frac{B_{path}}{E_G}\right) \tag{2.13}$$

Where P is a parameter that takes the values 1, 1.5 or 2 depending on the value of the model. Since the band-to-band generation rate only depends on the electric field F, this model predicts current even at $V_{DS} = 0V$ which is not desirable. In order to correct this problem the occupancy function $(f_v - f_c)$ from the Landauer equation must be applied to the above equation. One example of such a modification is the Hurkx model [13] in Sentaurus. In this model the generation rate is gives as:

$$G_{net}^{B2B} = A_{path} D\left(\frac{F}{1V/cm}\right)^{P} exp\left(-\frac{B_{path}E_G(T)^{3/2}}{E_G^{3/2}F}\right) \tag{2.14}$$

where D is given by:

$$D = \frac{np - n_{i,eff}^2}{(n + n_{i,eff})(p + n_{i,eff})}(1 - |\alpha|) + \alpha \tag{2.15}$$

The Hurkx model also does not take into account any nonlocal dependence. Its advantage over the basic Kane model is that it incorporates temperature dependence in calculating band-gap. 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. More importantly in this model we can have both generation and recombination as well as zero net generation rate at equilibrium (zero current at $V_{DS} = 0V$)

Schenk model[14] is another local band-to-band tunneling model available in Sentaurus . No nonlocal dependence is used, meaning that no search for a region where the conduction band equals the valence band is performed. The primary limitation of local tunneling models is the consideration of a uniform electric field, due to which they overestimate the band-to-band generation by a large margin.

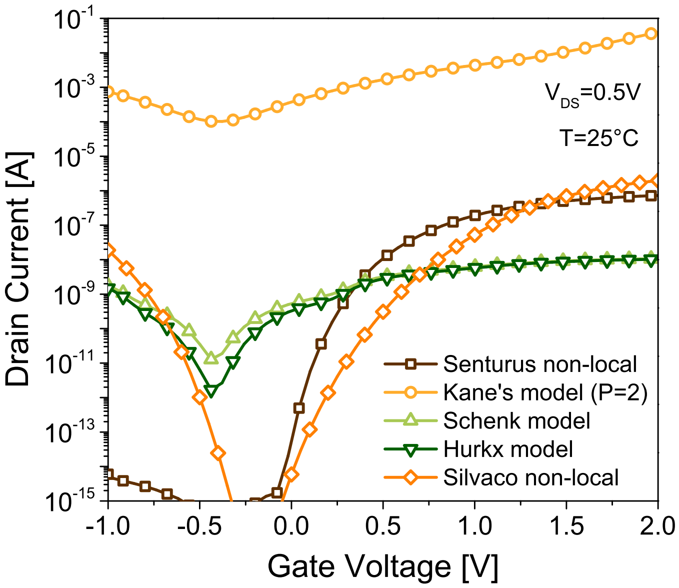


Figure 2.5 – Comparison of transfer characteristic of a 100nm DG-TFET local and nonlocal models in Sentaurus and the non-local model in Silvaco with all parameters set to default.

2.2.2 Sentaurus Dynamic non-Local Tunneling model

Local tunneling models proposed in the previous section by Kane[6], Hurkx [13] , Schenk [14] assumes either an average or maximum Electric field which is constant throughout the tunneling path. In the non-local band to band tunneling model, the tunneling current depends on the band edge profile along the entire path between the points connected by tunneling. Which means the electric field at each point in the tunneling path is dynamically changing. This makes tunneling a essentially a non-local process. The net hole recombination rate for a band-to-band tunneling process considering a non-local electric field can be written as:

$$R^P_{new} = |\nabla E_V(0)| C_P exp\left(-2\int_0^{x_o} \kappa_V dx - 2\int_{x_0}^{l} \kappa_C dx\right) \left[\left(exp\left[\frac{\varepsilon - E_{F,n}(l)}{kT(l)}\right]+1\right)^{-1} - \left(exp\left[\frac{\varepsilon - E_{F,p}(0)}{kT(0)}\right]+1\right)^{-1}\right] \tag{2.16}$$

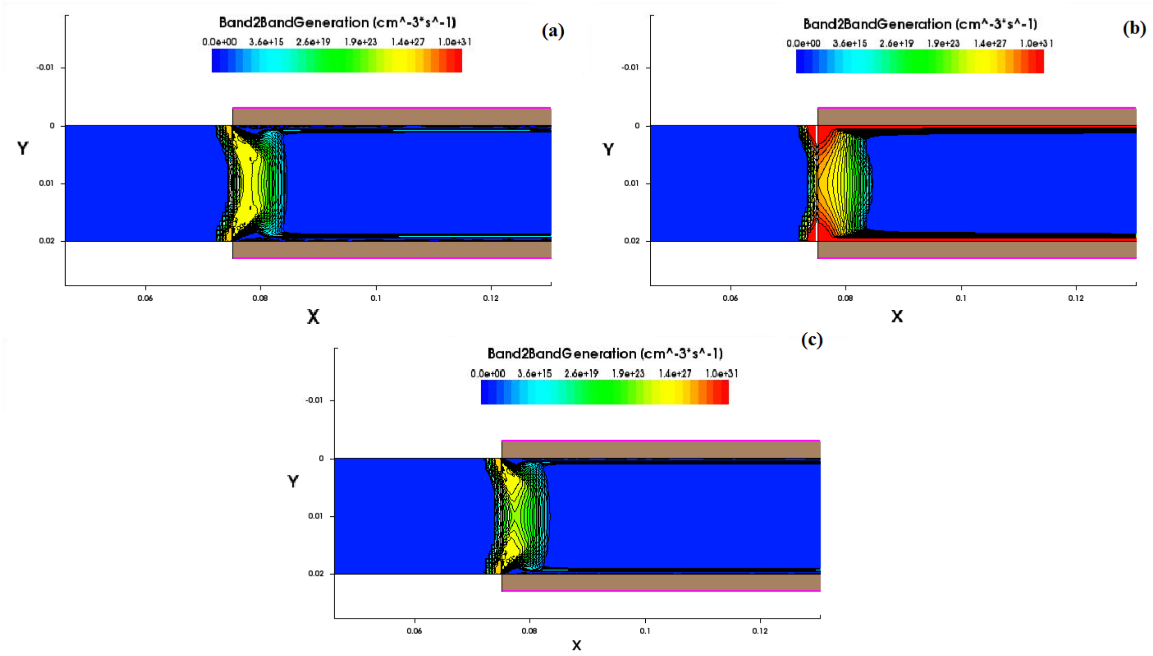


Figure 2.6 – Comparison of generation rates in local and nonlocal models in Sentaurus (a) Schenk model (b) Kane's model (P=2) (c) Hurkx model

where C_p here is a multiplicative factor for phonon assisted tunneling taking into account the phonon energy. This factor is different for direct tunneling. The C_P term along with the $exp(-2\int \kappa dx)$ term comes from the WKB approximation. The inverse exponential term is essentially the occupancy function assuming Fermi-Dirac statistics, which ensures zero current at zero V_{DS}.

Under the assumption of a uniform electric field the net recombination rate is reduced to the well-known Kane and Keldysh models assuming the form [11]

$$R_{net} = A\left(\frac{F}{F_0}\right)^P exp\left(-\frac{B}{F}\right) \tag{2.17}$$

where $F_0 = 1V/cm$ and $P = 2$ for direct tunneling and $P = 2.5$ for phonon assisted tunneling process. Since Silicon is an indirect band gap semiconductor, we concentrate here on the phonon assisted tunneling case. Ignoring band gap narrowing effect, the pre factor A_{path} and the exponential factor B_{path} for phonon-assisted tunneling can be expressed as [11] :

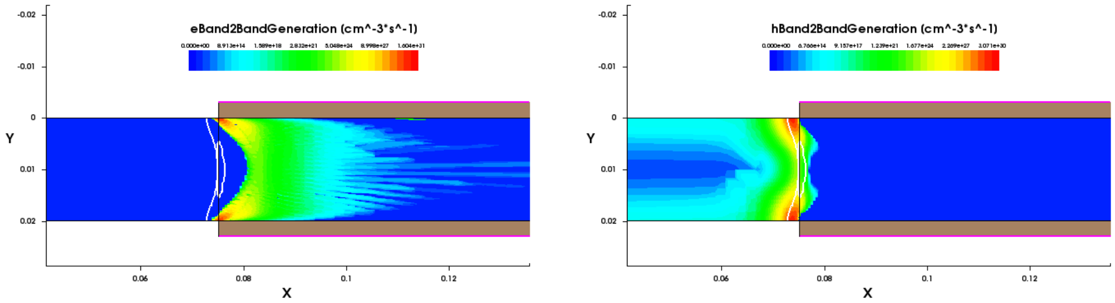


Figure 2.7 – Comparison of electron and hole generation rates in the dynamic nonlocal model in Sentaurus.

$$A_{path} = \frac{g(m_p m_n)^{3/2}\left[1+2N_{op}\right]D_{op}^2(qF_0)^{5/2}}{2^{21/4}\hbar^{5/2}m_r^{21/4}\rho\epsilon_{op}\left[E_g(300K)+\Delta C\right]^{7/4}} \tag{2.18}$$

$$B_{path} = \frac{2^{7/2}\pi m_r^{1/2}\left[E_g(300K)+\Delta C\right]^{3/2}}{3q\hbar} \tag{2.19}$$

Where g is a degeneracy factor; m_r is the reduced tunneling mass; q is the elementary charge; h is the Plank's constant; E_g is the indirect bang gap; $mc(mv)$ is the conduction (valence) band density of states effective mass; $N_{op} = [exp[(\epsilon_{op}/kT)-1]]^{-1}$ is the number of transverse optical (TO) phonons with energy ϵ_{op} at temperature T, where k is the Boltzmann constant; D_{op} is the deformation potential of TO phonons and ρ is the mass density of Silicon.

The degeneracy factor is given by $g = 2g_c g_v$ where g_c/g_v is the conduction/valence band valley degeneracy [15]. The factor 2 comes from spin degeneracy of the electrons. We only consider the lightest carriers as the lightest carriers have the dominant contribution to the tunneling current, g_v is equal to 1 (light hole band), and g_c depends on the tunneling direction ($g_c = 4$ for (100) direction) [15, 16]. In Kane's two-band dispersion relation [6], the effective mass in the conduction band m_c and the valence band m_v are related to the reduced mass m_r as:

$$\frac{1}{m_c} = \frac{1}{2m_r} + \frac{1}{m_0} \tag{2.20}$$

$$\frac{1}{m_v} = \frac{1}{2m_r} - \frac{1}{m_0} \tag{2.21}$$

There are two main phonon contributions to the tunneling current in Silicon, the TA (transverse acoustic) and TO (transverse optic) phonon [17][18]. Here we consider the process with TO phonons as the default parameters used in [11] are calibrated for TO phonons.

Figure 2.5 shows the comparison of the transfer characteristics of a $L_G = 100nm$

double gate TFET with all default parameters. The basic Kane's model greatly overestimates the current as shown. Whereas the Schenk and Hurkx model show very low I_{ON}/I_{OFF} ratio, the dynamic non-local model shows the most realistic results. Figure 2.6 on the other hand shows the generation rate scatter plots for local tunneling models. Figure 2.7 shows the electron(left) and hole(right) generation rates. Once again it has to be noted that all local models do not have separate electron and hole generation rates, whereas the non-local model has separate electron and hole generation plots as shown.

The main parameters in the model and their default values are listed in table 2.1. The calibration procedure is discussed in details in the next section.

Table 2.1 – Non-Local Band-to-Band Tunneling Parameters for Sentaurus non-loca tunneling model.

Parameter name	Default value	Units
A_{path}	$4X10^{14}$	$cm^{-3}s^{-1}$
B_{path}	$1.9X10^{7}$	Vcm^{-1}
C_{path}	0	$J^{2}cmkg^{-1}$
D_{path}	0	eV
P_{path}	0.037	eV
R_{path}	0	none

2.2.3 Silvaco non-Local Tunneling model

The non-local tunneling model in Silvaco Atlas is implemented in a similar way to Sentaurus and is based on the Landauer equation. The tunneling current is calculated, let us consider the energy band profile along each tunneling slice with reverse bias applied across the junction. The range of valence band electron energies for which tunneling is permitted is shown in the schematic of the energy band profile in Fig. 2.8. The highest energy at which an electron can tunnel is E_{upper} and the lowest is E_{lower}. The tunneling can be thought of being either the transfer of electrons or the transfer of holes across the junction. The rates for electrons and holes are equal and opposite because the tunneling results in the generation or recombination of electron-hole pairs.

The main parameters in the model and their default values are listed in table 2.2. To calibrate the model, the values of effective mass used in eqns. 2.25 and 2.26 are specified. The values of effective mass used in eqns. 2.25 and 2.26 using ME.TUNNEL and MH.TUNNEL are set in the command file. If ME.TUNNEL (MH.TUNNEL) is not specified, then the value specified for conduction and valence band effective masses MC (MV) is used. The tunneling current is most sensitive to the effective masses used in eqns. 2.25 and 2.26 because the tunneling probability depends exponentially on them.

Considering the tunneling process as a transfer of an electron across the junction the net

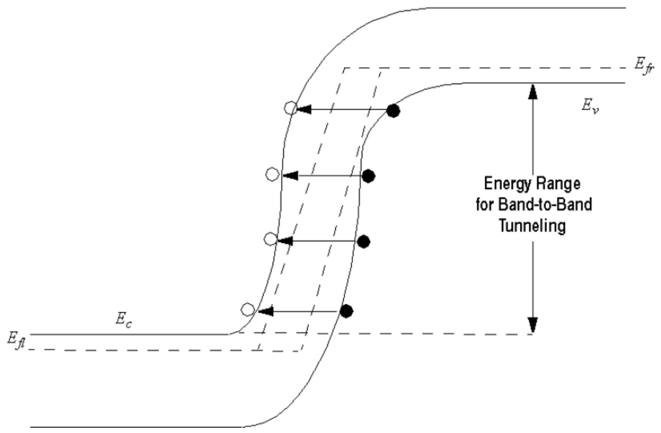


Figure 2.8 – Schematic of non-local band to band tunneling in reverse bias in Silvaco Atlas. [12]

Table 2.2 – Non-local band-to-band tunneling parameters for Silvaco TCAD.

Parameter name	Default value	Units
MC	0.16	None
MV	0.24	None
ME.TUNNEL	0.16	None
MH.TUNNEL	0.24	None

current per unit area for an electron with longitudinal energy E and transverse energy E_T is:

$$J(E) = \frac{e}{\pi\hbar}\int\int T(E)[f_l(E+E_T) - f_T(E+E_T)]\rho(E_T)dEdE_T \tag{2.22}$$

where $T(E)$ is the tunneling probability for an electron with longitudinal energy E. $\rho(E_T)$ is the 2-dimensional density of states corresponding to the 2 transverse wavevector components and f_l, f_T are the Fermi-Dirac function using the quasi Fermi-level on the left and right side of the junction respectively. The tunneling probability, $T(E)$, is then calculated using the WKB approximation

$$T(E) = exp\left(-2\int_{x_{start}}^{x_{end}} \kappa(x)dx\right) \tag{2.23}$$

As also seen in Fig. 2.8, the start and end points of the tunneling paths, x_{start} and x_{end}, depend on energy. Atlas calculates these start and end points for each value of E and calculates the evanescent wave-vector at points in between as:

$$\kappa(x) = \frac{\kappa_e\kappa_h}{\sqrt{\kappa_e^2+\kappa_h^2}} \tag{2.24}$$

where

$$\kappa_e(x) = \frac{1}{i\hbar}\sqrt{2m_0 m_e(x)(E - E_c(x))} \tag{2.25}$$

and

$$\kappa_h(x) = \frac{1}{i\hbar}\sqrt{2m_0 m_h(x)(E_v(x) - E)} \tag{2.26}$$

where $\hbar$ is Planck's constant, m_0 is the rest mass of an electron, $m_e(x)/m_h(x)$ is the electron/-hole effective mass and $E_c(x)/E_v(x)$ are the conduction and valence band energies. This values are put into eqn. 2.22 and 2.23 to give the tunneling current density at a given perpendicular energy, E, and the resulting current is injected into the simulation at x_{start} and x_{end}. This is repeated for all values of E between E_{lower} and E_{upper} and is done for every tunneling slice in the tunneling regions.

2.3 Basic working principle of a TFET

Tunnel FET use quantum-mechanical tunneling of electrons from the source to the channel as described in the previous section, as the primary carrier transport mechanism, allowing for sub-60 mV/dec subthreshold slopes. Tunnel FETs belongs to the family of so-called steep-slope devices that are currently being investigated for ultra-low-power electronic applications [5]. A key feature of the TFET, which is critical for low-power switching, is the possibility for an inverse sub threshold slope, SS, below the limit of 60 mV/dec for normal FETs in room temperature as discussed in chapter 1.

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 [5]. Ideally, these devices would have a very low off-state current (proportional to reverse biased diode leakage), a very low sub-threshold slope, and acceptable on-current. TFETs can be generally classified as point and/or line tunneling devices [19].

2.3.1 Point tunneling Tunnel FET

In a n-type point tunneling device as shown in Fig.2.9(a), the source is p-type doped whereas the drain is n-type doped. This is unlike a traditional MOSFET, where both the source and drain are doped alike. The biasing scheme is also shown in the Fig.2.9(a) which is again typical for a n-type device. In order to have a p-type device the source and drain doping are interchanged, and the biasing scheme also changes similar to a traditional MOSFET. Usually the channel region is oppositely doped as the source as well as to a much lower level ($\sim 10^{15} cm^{-3}$) than traditional MOSFETs. This is in order to get a larger band bending between the source and the channel.

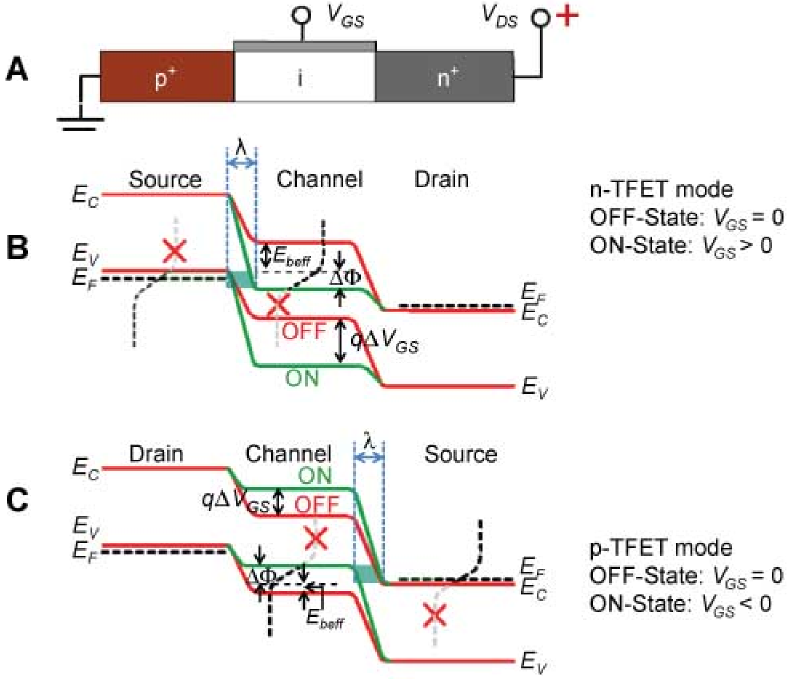
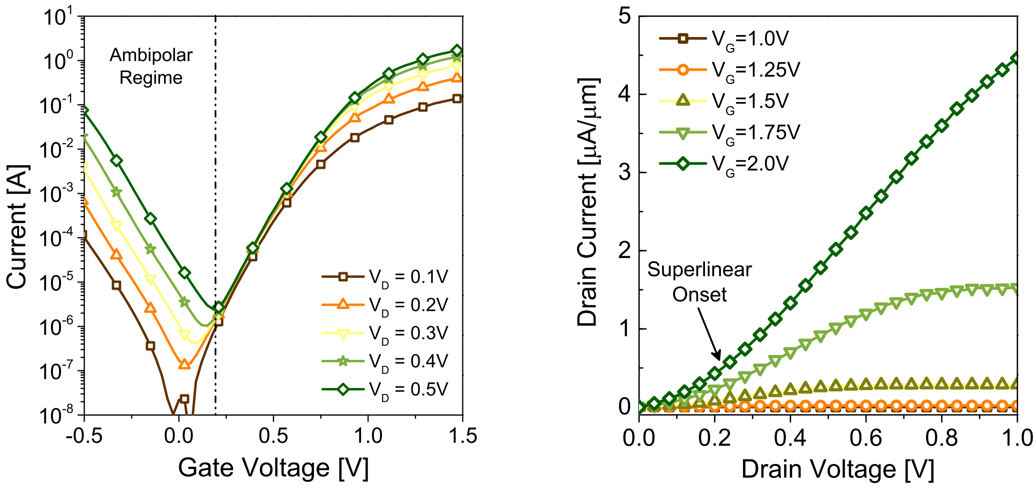

A
V_{GS}
V_{DS}
+
p+
i
n+
B
λ
Source
Channel
Drain
E_C
E_V
E_F
E_{beff}
ΔΦ
OFF
ON
$q\Delta V_{GS}$
E_F
E_C
E_V
n-TFET mode
OFF-State: $V_{GS} = 0$
ON-State: $V_{GS} > 0$
C
Drain
Channel
λ
Source
ON
OFF
$q\Delta V_{GS}$
E_C
E_V
E_F
ΔΦ
E_{beff}
E_F
E_C
E_V
p-TFET mode
OFF-State: $V_{GS} = 0$
ON-State: $V_{GS} < 0$

Figure 2.9 – (a)Schematic of a Tunnel FET with ntype biasing applied (b) nTFET mode with both OFF and ON conditions shown (b) pTFET mode with both OFF and ON conditions shown. [20]

The energy band diagrams of a n-type TFET is sown in Fig.2.9(b). In thermal equilibrium, or in OFF state no tunneling can take place due to the large potential barrier present (red curve). As the gate bias is increased to a positive value the energy bands starts bending reducing the potential barrier between the top of the valence band edge in the source and the top of the conduction band edge in the channel region. However just a positive gate bias is not sufficient to turn ON the device. As with just a positive gate bias the potential barrier is low, however there are no vacant states in the channel conduction band to where the electrons from the source valence band can tunnel to. The sates become available when a positive drain bias is applied as shown in the Fig. 2.9(b). This theory validates the Landauer's equation as shown in eqn. 2.1, which is essentially a product of the tunneling probability and the occupancy function (Fermi statistics). Hence for electrons to flow from source to the drain and constitute a current in the opposite direction, both tunneling porbability and occupancy function has be non-zero. As the gate bias increases, the source does not appreciably deplete, but the gate causes the channel region to invert, resulting in tunneling from the source to the channel (green curve).

Figure 2.9(c) shows the energy band diagrams of a p-type TFET. The operational principle is similar to n-type devices except now holes instead of electrons tunnel from the top of the conduction band in source to the available states in the bottom of the valence band in the channel region. Both figures 2.9(b) and (c) show the Fermi distribution function in the source. As indicated by the red cross mark, the high energy tail of the Fermi function is naturally blocked by the typical energy band formation in the ON state. This basically means the a Tunnel FET acts as a band pass filter, allowing only low energy electrons/holes to transfer from source to the drain. This results in the "cooling down" of the system and is one of the reason why Tunnel FETs exhibit sub 60 mv/decade sub-threshold slope in room temperature.



(a)

(b)

Figure 2.10 – Measured (a)transfer and (b)output characteristics of an all Silcon nanowire n-type Tunnel FET on fully depleted SOI wafers at room temperature. The ambipolarity and superlinear onset which are typical to Tunnel FETs are highlighted.

Figure 2.10(a) shows the measured transfer curves of a all Silcon nanowire n-type Tunnel FET on fully depleted SOI wafers. As can be seen from the figure there are two main differences here with respect to a MOSFET:

- Ambipolar conduction: As shown in Fig. 2.10(a), the Tunnel FET do not turn OFF in negative gate bias as opposed to a MOSFET which has negligible current in negetive gate bias for a n-type device. The reason for this behaviour can be seen in Fig. 2.9(b) and (c). In positive gate bias the band bending is such that electrons can tunnel from source to the channel. At zero gate bias, there is no tunneling and hence no current. However when gate bias is negative, the bands move in the opposite direction and the tunneling barrier near the drain-channel junction becomes favourable for tunneling. Hence in negative gate bias current still flows with a positive drain bias. There are ways to reduce this ambipolarity, the most common being reducing the drain doping and introducing an under-lap between the gate and the drain as shown in Fig. 2.13 [21, 22, 23, 24].

- Varying sub-threshold slope: Another difference of a Tunnel FET transfer curve with a MOSFET is that the sub-threshold slope is constantly changing with gate bias. This can be seen from the definition of sub-threshold slope in eqn. 2.27 and the drain current expression in section 2.2.1.

$$\begin{aligned} SS_{TFET} &= \left(\frac{d\log I_D}{dV_{GS}}\right)^{-1} \\ &\propto V_{GS}^2 \qquad\qquad if \;\; T_{WKB} \;\; is \;\; small \\ &\propto V_{GS} \qquad\qquad if \;\; T_{WKB} \approx 1 \end{aligned} \tag{2.27}$$

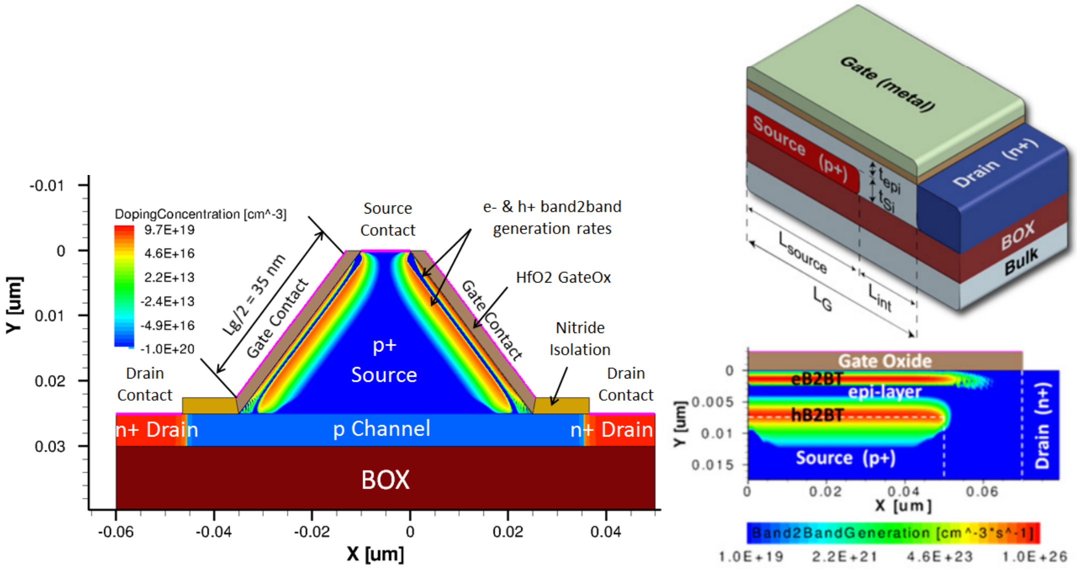
In conclusion, regardless of the assumptions made for simplification purposes, the sub-threshold slope of a TFET degrades with increasing gate-source potential. Hence, the sub-60mV/decade is only available at relatively low voltages which makes TFETs intrinsically better adapted for low voltage operation.

Figure 2.10(b) shows the output characteristics of a n-type Tunnel FET. The noticeable difference here with respect to a MOSFET output characteristic is the super-linear onset of the current. To explain this behaviour we have to understand the relation between the tunneling current and the Fermi-Dirac distribution as defined by the Landauer's equation. In [25], L. D. Michielis et al. has assumed the tunneling barrier is perfectly transparent inside the allowed tunneling energy window and zero if otherwise. Moreover, for a small tunneling window. In this case, the total tunneling current (resulting from the two contributions of current from the valence band to the empty states of the conduction band and from the conduction band to the empty states of the valence band [26],[11] can express as proportional to:

$$I_T \propto \int_{CB_{ch,min}}^{VB_{source,max}} [f_S(E) - f_{ch}(E)]\,dE \tag{2.28}$$

Hence according to [25], we can show that the modulation of the device current at the beginning of the drain sweep is mainly determined by a change in the occupancy functions and it is almost independent of the tunneling probability modulation. Thus, it can said that the main cause of the superlinear output onset is not the exponential dependence of the device current on the tunneling path. This also shows that, while the study of the tunneling barrier is indeed important when understanding the TFET behavior, an explanation based on the sole tunneling probability is not enough, and the analysis of the occupancy function dependence on the bias has to be necessarily taken into account.

Another important difference of a Tunnel FET output characteristic with that of a MOSFET is the uni-directional conduction. This is primarily because in the output characteristic, the TFET is biased as a revered biased diode and electrons are tunneling from source to the channel dependent on the drain bias. However when drain bias is negative, the p-i-n diode will be forward biased. Hence current flows just like in a diode with very little gate control. For this reason the output characteristic of Tunnel FET looks asymmetric around zero drain bias [24, 27, 28]. This is not a desirable effect and can have undesirable effects in circuit operation for example pass transistor in a SRAM cell.


DopingConcentration [cm^-3]
9.7E+19
4.6E+16
2.2E+13
-2.4E+13
-4.9E+16
-1.0E+20
Lg/2 = 35 nm
Gate Contact
Source Contact
e- & h+ band2band generation rates
HfO2 GateOx
Gate Contact
Nitride Isolation
Drain Contact
Drain Contact
p+ Source
n+ Drain
p Channel
n+ Drain
BOX
Y [um]
X [um]
-0.01
0
0.01
0.02
0.03
-0.06
-0.04
-0.02
0
0.02
0.04
Gate (metal)
Source (p+)
Drain (n+)
BOX
Bulk
t_{epi}
t_{Si}
L_{source}
L_G
L_{int}
Gate Oxide
eB2BT
epi-layer
hB2BT
Source (p+)
Drain (n+)
Y [um]
X [um]
0
0.005
0.01
0.015
0
0.02
0.04
0.06
Band2BandGeneration [cm^-3*s^-1]
1.0E+19
2.2E+21
4.6E+23
1.0E+26

(a) (b)

Figure 2.11 – Examples of line tunneling structures (a) [29] (b) [30] where the primary tunneling mechanism is oriented to the direction of the gate electric field. The band-to-band generation is spread over a linear fashion, resembling a line, hence the name line tunneling.

2.3.2 Line tunneling device

Wang in [31] discovered that the TFET current is composed of two components. The first component ”point tunneling” occurs at the source-channel interface and its dominant contribution is localized in a small area. The second component is located in the part of the source region if it is overlapped by the gate. Because the area where BTBT starts from resembles a line, this component is called ”line tunneling”. 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). Many line tunneling examples have been proposed in literature as simulation concept as well as experimental concepts. Some examples of Tunnel FET structures specially designed to have primarily "line-tunneling" are shown in the following figure.

In line tunneling the direction of the gate electric field is aligned to the tunneling direction, hence they are expected to show better sub-threshold slope and higher current than point tunneling alternatives. However in reality, due to quantization effects the experimental results have so far not been able to reproduce the simulation results.

2.3.3 Temperature dependence

From eqn. 2.9, 2.11 and 2.12 we have seen that tunneling phenomenon essentially does not depend on temperature. The primary temperature dependence in the measured characteristic in Fig.2.12 comes from the dependence of the energy band gap of Silicon to temperature [32, 33, 34, 35]. Essentially, as temperature increases, the band gap decreases and hence the current also increases (as tunneling current $\propto$ exponential of E_G^{-1}). TFETs thus show a significantly different temperature dependence than MOSFETs, whose operation essentially relies on the thermal energy of the carriers.

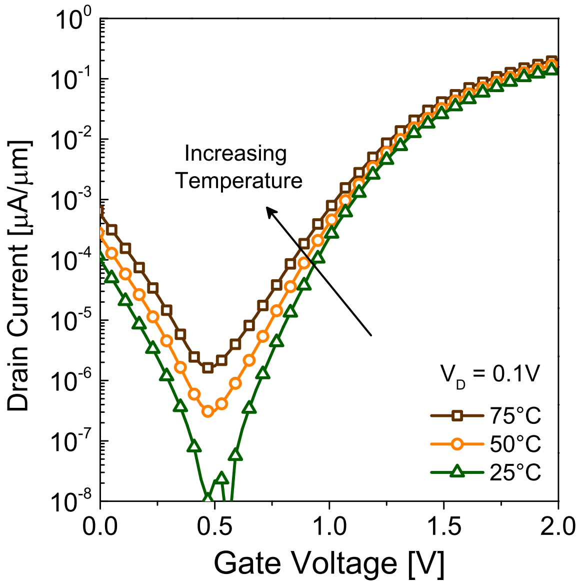


Figure 2.12 – Measured transfer characteristic of a all Silcon nanowire n-type Tunnel FET on fully depleted SOI wafers at elevated temperatures. [36]

The temperature dependence is more pronounced in the sub-threshold region. This is because Schokley-Read-Hall recombination is active in this regime which has a stronger temperature dependence than band-to-band tunneling.

2.4 FD-SOI Tunnel FET simulation and fitting to measurements

In this section we study the non-local tunneling model using a standard commercially available technology computer-aided design device simulator. Single gate Tunnel FET devices with 200nm gate length based on SOI technology are simulated and compared with measured data. A thorough description of the nonlocal Band-to-Band tunneling model implemented in Synopsys Sentaurus TCAD has been shown, demonstrating the importance of model parameters. By altering the default values of the model parameters we have obtained a physically meaningful fit with the measured data. Then the physical significance of all the model parameters have been discussed in details, with a complete physical explanation of the changes in the simulation results leading to the match with experiments. The method of parameter calibration presented in this work can be used to physically justify the modification of the various fitting parameters used in device simulation tools.

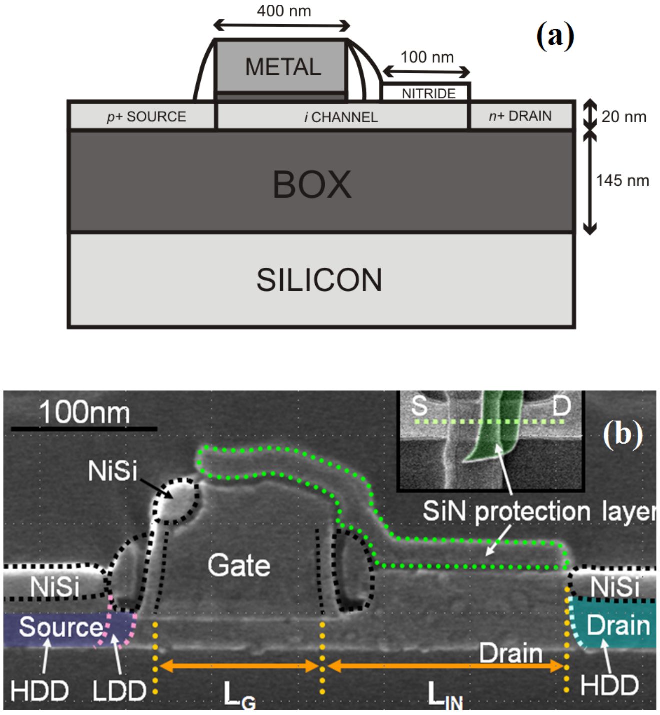


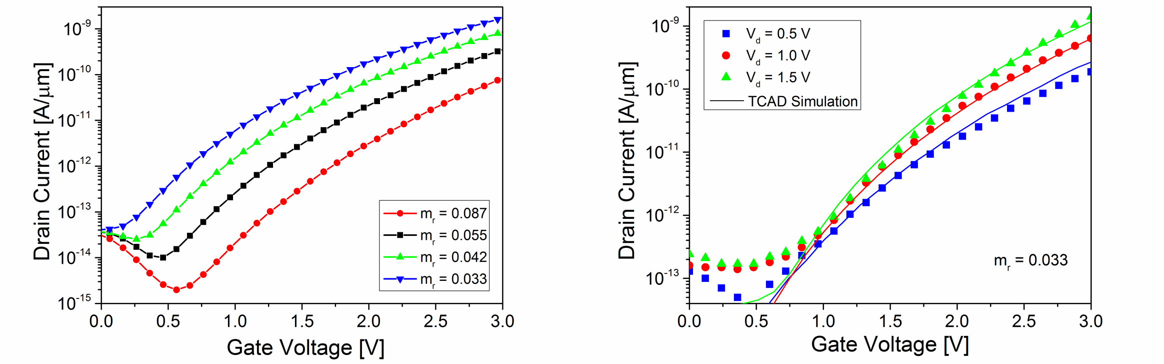
Figure 2.13 – (a) Basic structure and dimensions of TFET devices simulated. (b) SEM image showing the fabricated SOI TFET device. Inset shows the top view.[23]

UTB-SOI TFET measurements and simulation setup

In this section, we have used Sentaurus TCAD ver. 2010.12 which is a commercially available TCAD tool from Synopsys. The structure used for our simulations is as shown in Fig.2.13(a) corresponds to the devices fabricated by CEA-LETI [23]. Fig. 2.13(b) shows a SEM image of the gate stack and the channel region of the considered TFET devices. The SiN protection layer is to prevent the ambipolar nature of TFETs. The fabrication of UTB-SOI TFET devices (Fig.2.13) started from an SOI substrate with 145nm BOX and 20nm active Si layer using a MESA process [23]. The gate stack was composed of 3nm HfO_2 or 6nm SiO_2/TiN (10nm) / Poly Si (50nm). A $Si3N_4$ protection layer was deposited to introduce intrinsic regions adjacent to the drain side. This suppresses the source-drain leakage current by reducing the electric field at the tunnel junction.

In order to have a faithful model calibration we should have a realistic doping profile in our simulations. For this purpose conventional MOSFET devices were measured on the same wafer since all the devices (including the TFETs) were fabricated during the same process steps. The only difference was in the source and drain doping. These n- and p-MOSFET devices were simulated with standard drift/diffusion models. The source/drain doping, junction abruptness of the source-channel and drain-channel interface and series resistance were used to fit the simulation data to the measured values. The same values of doping and junction abruptness were then used as a starting point for the TFET simulations.

A source doping of 8x$10^{19} cm^{-3}$ (Boron) and drain doping of 5x$10^{19} cm^{-3}$ (Arsenic) were used for a n type TFET. The channel region was left undoped (1x$10^{15} cm^{-3}$). A junction abruptness value of 7nm/dec [37] was used in the source-channel as well as for the drain-channel junction. $6nm$ of SiO_2 dielectric was used for all simulations. The gate work function was fixed at $4.3eV$.



(a) (b)

Figure 2.14 – (a)Impact of reduced mass on tunneling current shown for different value of reduced masses including the theoretical mass and the default mass extracted from default parameters and (b) Simulated and measured transfer characteristics. Symbols indicate experimental data and solid lines the corresponding simulation.

2.4.1 Calibration Methodology

The non-local tunneling model as implemented in the Sentaurus TCAD has the default values for the above defined expressions specified in terms of the model parameters $A_{path} = A = 4x10^{14} cm^{-3} s^{-}1$, $B_{path} = B = 1.9x10^{7} V cm^{-1}$, $Rpath = mv/mc$ (ratio of valence band to conduction band masses) = 0, $P_{path} = \epsilon_{op}$ (phonon energy) = 0.037 eV and $D_{path} = \Delta C$ (conduction band offset) = 0 eV. The values of the constants are directly related to the physical parameters which influence the device characteristics.

Upto three different tunneling paths can be specified in the parameter file [11]. It is to be noted that the value of P_{path} decides whether a particular tunneling path will be direct or phonon assisted. A non-zero value for P_{path} dictates that tunneling path will be phonon assisted whereas a value of zero means direct tunneling. The default P_{path} value of 37meV value has no significance as the actual phonon energy used in the model is for TO phonons i.e 57.6 meV [17]; and any non-zero value in this field will have the same effect.

Parameters A_{path}, B_{path}, P_{path} and R_{path} determines gD^2_{op}/ρ, m_v and m_c. When $R_{path} = 0$; m_v/m_c are determined from eqns. 2.20) and 2.21). This extraction process is explained as an algorithm in the following :

I) First step is to determine the reduced mass mr from the default value of B_{path} according to the following relation:

$$m_r = A\left[\frac{3qhB_{path}}{2^{7/2}\pi(E_g + D_{path})^{7/2}}\right]^2 \tag{2.29}$$

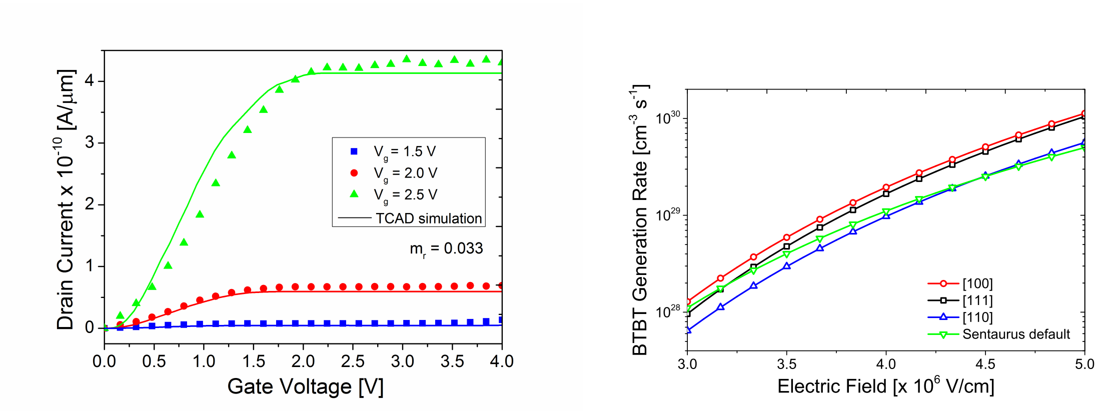
II) Next gD^2_{op}/ρ is determined using the value of m_r from eqn. 2.29 and m_c, m_v from eqn. 2.20 and eqn. 2.21

$$\frac{gD^2_{op}}{\rho} = \frac{2^{21/4}h^{5/2}m_r^{5/4}\epsilon_{op}[E_g + D_{path}]^{7/4}A_{path}}{(m_c m_v)^{3/2}[1 + N_{op}]q^{5/2}} \tag{2.30}$$

III) Now gD^2_{op}/ρ as calculated in the above step and plugged into the eqn. 2.18. So we have one set of values of mr and gD^2_{op}/ρ calculated from the default values of A_{path} and B_{path} at room temperature. The value of reduced mass m_r can be now used as a fitting parameter to calculate different values of A_{path} and B_{path}. These calculated values of A_{path} and B_{path} will thus be related by the reduced mass and hence more meaningful than what has been reported in some earlier works [38][39]. Details of the constants used and calculated parameter values are in the appendix A.

Summary of the best fit parameters

A series of simulations were performed to fit the experimental data using the parameters discussed above. Fig. 2.14(a) shows the transfer characteristics with varying reduced mass. The best fit parameter value of $m_r = 0.033 * m_0$ was then used to check the validity of the model in different bias conditions. The measured and simulated transfer characteristics of the device in Fig.2.13 with $L_g = 400nm$ and $L_{in} = 100nm$ is shown in Fig. 2.14(b). The fitted parameters are found to be valid for varying drain bias. Fig. 2.15(a) shows the comparison of simulated and experimental output characteristics at different gate voltages for the same device. Thus the calibrated model is also able to predict both transfer and output characteristics at different bias conditions with a reasonable accuracy. The values of the reduced mass and corresponding calculated values of A_{path} and B_{path} parameters which resulted in the best fit between the simulated and the measured data are summarized in table 2.3.



(a) (b)

Figure 2.15 – (a) Simulated and measured output characteristics. Symbols indicate measured data and solid lines the corresponding simulation and (b) Variation of band-to-band generation rate as a function of uniform electric field with different crystallographic directions.

Table 2.3 – Summary of calculated A_{path} and B_{path} values from mr.

Reduced mass (m_r)	$A_{PATH}(cm^{-3}sec-1)$	$B_{PATH}(V/cm)$	Remark
0.087	$9.25X10^{14}$	$2.4X10^{7}$	Theoritical
0.055	$4.0X10^{14}$	$1.9X10^{7}$	Default
0.042	$2.5X10^{14}$	$1.66X10^{7}$	-
0.033	$1.63X10^{14}$	$1.47X10^{7}$	Best fit

2.4.2 Crystal orientation dependent tunneling generation rate

In section 2.4.1, we have shown the calibration procedure for the default [100] crystallographic direction. However the values of A_{path} and B_{path} parameters can be calculated for different crystal directions if we know the values of the degeneracy factor g and the reduced mass m_r. These two parameters change values with different crystallographic directions [16]. It should be noted that the theoretical values of the reduced mass in different crystallographic directions as reported in literature are used [16] for these calculations. Fig. 2.15(b) shows the band-to-band generation rate as a function of the uniform Electric field with the theoretically calculated parameters. It is noticeable that the [100] direction has the highest whereas [110] direction has the lowest generation rate particularly at higher electric fields. Higher generation rate would lead to higher tunneling current. This fact could be utilized in the future to design TFETs to reduce ambipolarity effects with the source-channel junction in the [100] direction and the drain channel junction being in the [110] direction.

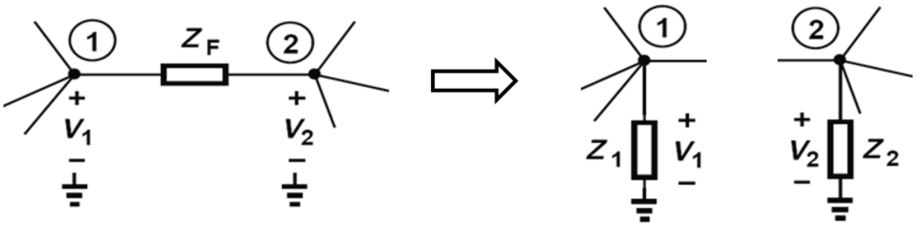
2.5 Miller effect

The effective input impedance of an amplifier depends on the impedance connected from input to output of the amplifier. The apparent scaling of this impedance often dominates the input impedance and frequency response of the amplifier. This effect, now commonly known as the Miller Effect, was first reported by John Miller [40].

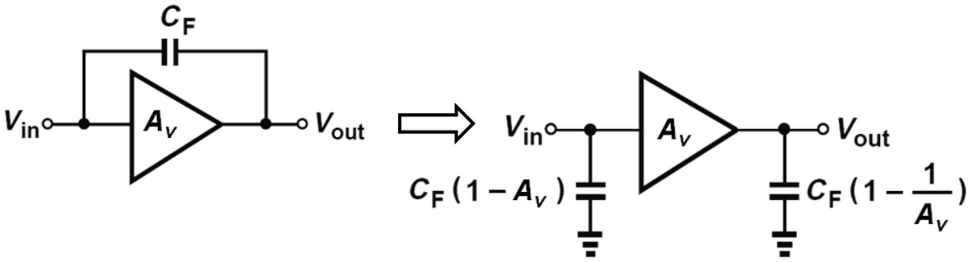
Consider the Fig. 2.16(a). If A_v is the voltage gain from node 1 to 2, then a floating impedance Z_F can be converted to two grounded impedances Z_1 and Z_2 as below:

$$\frac{V_1 - V_2}{Z_F} = \frac{V_1}{Z_1} \Rightarrow Z_1 = Z_F \frac{V_1}{V_1 - V_2} = Z_F \frac{1}{1 - A_V} \tag{2.31}$$

$$\frac{V_1 - V_2}{Z_F} = -\frac{V_2}{Z_2} \Rightarrow Z_2 = -Z_F \frac{V_2}{V_1 - V_2} = Z_F \frac{1}{1 - 1/A_V} \tag{2.32}$$



(a)



(b)

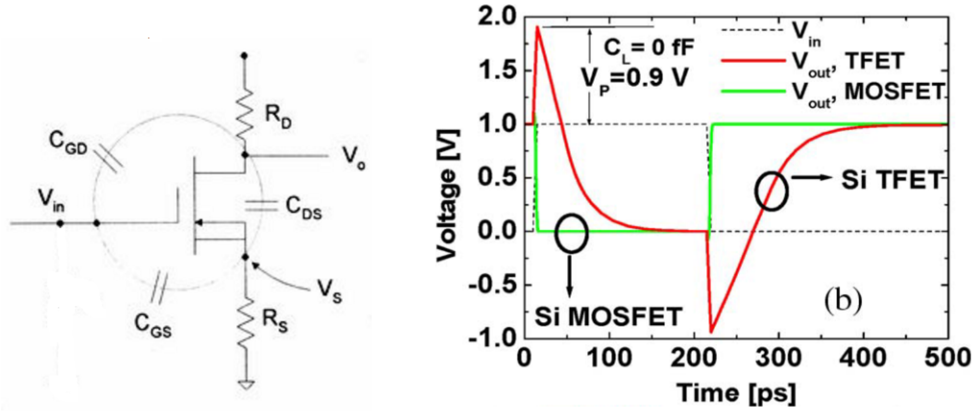
Figure 2.16 – (a) Amplifier with voltage gain $-A$, with an impedance Z connected from input to output and (b) Amplifier with feedback capacitor C_C.

Applying Miller's theorem, we can convert a floating capacitance between the input and output nodes of an amplifier into two grounded capacitances as shown below.

$$Z_1 = \frac{Z_F}{1 - A_V} = \frac{1/j\omega C_F}{1 - A_V} = \frac{1}{j\omega(1 - A_V)C_F} \tag{2.33}$$

$$Z_2 = \frac{Z_F}{1 - 1/A_V} = \frac{1/j\omega C_F}{1 - 1/A_V} = \frac{1}{j\omega(1 - 1/A_V)C_F} \tag{2.34}$$

Figure 2.16(b)(left) shows an example of Fig. 2.16(a) where the impedance coupling



(a) (b)

Figure 2.17 – (a) A simple MOS transistor circuit showing the main capacitance components and (b) Transient response characteristics of silicon TFET and MOSFET based inverters. [41]

the input to the output is the coupling capacitor C_F. The output impedance of the amplifier is considered low enough that the relationship $V_{out} = A_V V_{in}$ is presumed to hold. Figure 2.16(b)(right) shows a circuit electrically identical to Fig. 2.16(b)(left) using Miller's theorem. The coupling capacitor is replaced on the input side of the circuit by the Miller capacitance C_M, which draws the same current from the driver as the coupling capacitor in Fig. 2.16(b)(left). Therefore, the driver sees exactly the same loading in both circuits. On the output side, a capacitor $C_{Mo} = C_F(1 - 1/A_V)$ is present.

$$C_M = C_C\left(1 - \frac{V_o}{V_i}\right) = C_C(1 + A_v) \tag{2.35}$$

Thus the effective or Miller capacitance C_M is the physical C_F multiplied by the factor $(1 - A_V)$. The effective capacitance at their inputs is increased due to the Miller effect. This can reduce the bandwidth of the amplifier, restricting its range of operation to lower frequencies.

2.5.1 Miller effect in Tunnel FETs

From the above description, the C_F in case of a standard single transistor MOSFET/TFET circuit is the C_{gd} as shown in the Fig. 2.17(a). The Fig. 2.18 shows the comparison between the MOSFET and TFET capacitance curves. As shown, in a Tunnel FET the gate capacitance C_{gg} is dominated by the C_{gd} under all bias conditions. As a result of this large gate-drain capacitance, the inverter transient of a Tunnel exhibit large over and under shoots as shown in Fig. 2.17(b). The origin of this voltage overshoot is the high C_{gd} of the load TFET operating in the linear region. The output voltage of the TFET inverter stays at a very high value VMAX at the end of the input 0–1 transition. This continues to keep the pull-down

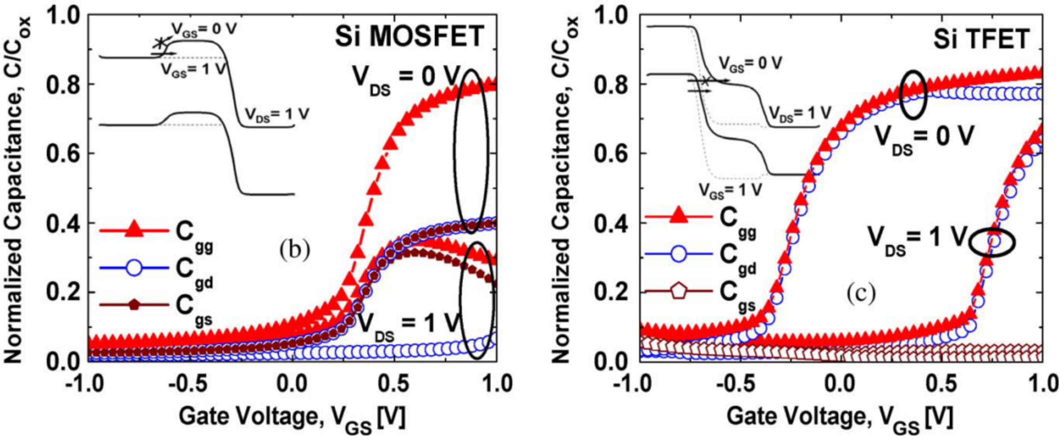


Figure 2.18 – Capacitance–voltage characteristics showing the gate (C_{gg}), gate-to-source (C_{gs}), and gate-to-drain (C_{gd}) capacitances as a function of gate-to-source voltage V_{GS} for (a) Si MOSFET and (b) Si TFET [41]

n-TFET in deep saturation (low C_{gd}) while the pull-up p-TFET remains in the linear region (high C_{gd}). Even as the output voltage starts transitioning from its peak overshoot value, the C_{gd} for p-TFET decreases while that for n-TFET increases, thereby maintaining a high Miller capacitance all throughout the transient.

Hence Miller effect is even more pronounced in Tunnel FETs due to the dominance of the drain-gate capacitance in the total gate capacitance as shown in the measured CV characteristics. This is due to fact that band-to-band tunneling has negligible contribution to the inversion charge density, atleast in the low injection levels of all-Silicon devices studied in this theses. Hence the source-gate capacitance remains negligible compared to the drain-gate capacitance. Miller effect is thus more amplified in TFET.

2.5.2 Proposed new structure to mitigate Miller effect in Tunnel FET

From the previous section, we know that the drain-to-gate capacitance in a TFET is almost equal to the gate capacitance in moderate and strong inversion regimes[41]. We will also see in section 3.5, that from the circuit point of view the device capacitances are equally important [41]. Due to enhanced Miller Effect, TFETs are known to exhibit large over/undershoot in transient operation as compared to CMOS. Therefore, the effort on improving I_{ON} should be simultaneous to an effort of reducing the Miller capacitance (C_{MILLER}). This subsection proposes a new architecture which addresses this issue.

In a TFET the gate modulation of the tunneling barrier is of primary importance. It is clear that in order to have a better gate modulation the tunneling direction should be aligned to the gate electric field. In this scenario, also known as “line tunneling”, the source is substantially overlapped by the gate and band-to-band (B2B) tunneling takes place from

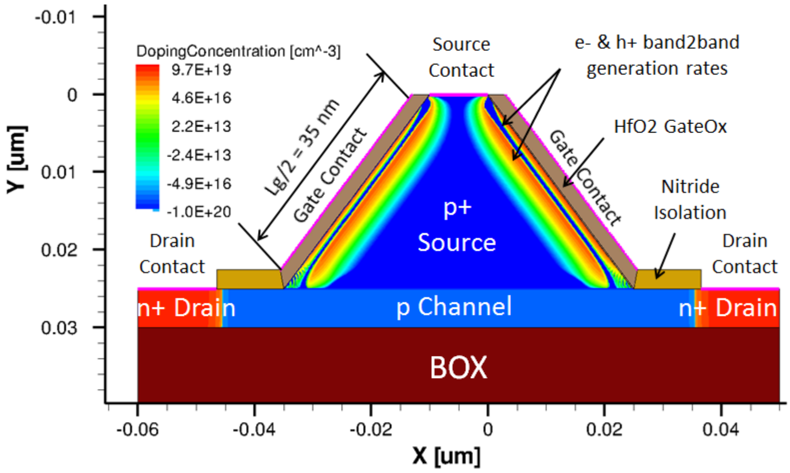
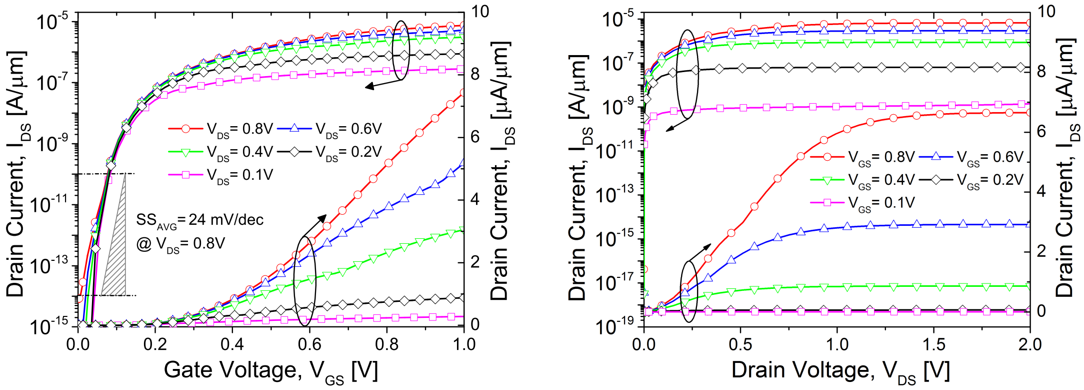

DopingConcentration [cm^-3]
9.7E+19
4.6E+16
2.2E+13
-2.4E+13
-4.9E+16
-1.0E+20
Lg/2 = 35 nm
Gate Contact
Source Contact
e- & h+ band2band generation rates
HfO2 GateOx
Gate Contact
Nitride Isolation
p+ Source
Drain Contact
Drain Contact
n+ Drain
p Channel
n+ Drain
BOX
Y [um]
X [um]
-0.01
0
0.01
0.02
0.03
-0.06
-0.04
-0.02
0
0.02
0.04

Figure 2.19 – Proposed structure with $T_{Si} = 5nm$, $L_{g,eff} = 50nm$. Gate work-function is $3.9eV$. $N_{A,source} = 10^{20}cm^{-3}$; $N_{D,drain} = 10^{20}cm^{-3}$ while $N_{A,channel} = 10^{18}cm^{-3}$. $2.5nm$ of HfO_2 are used as gate-oxide; $t_{BOX} = 10nm$; $t_{bulk} = 1\mu m$ and the effective Gate length $L_{g,eff} = 50nm$.

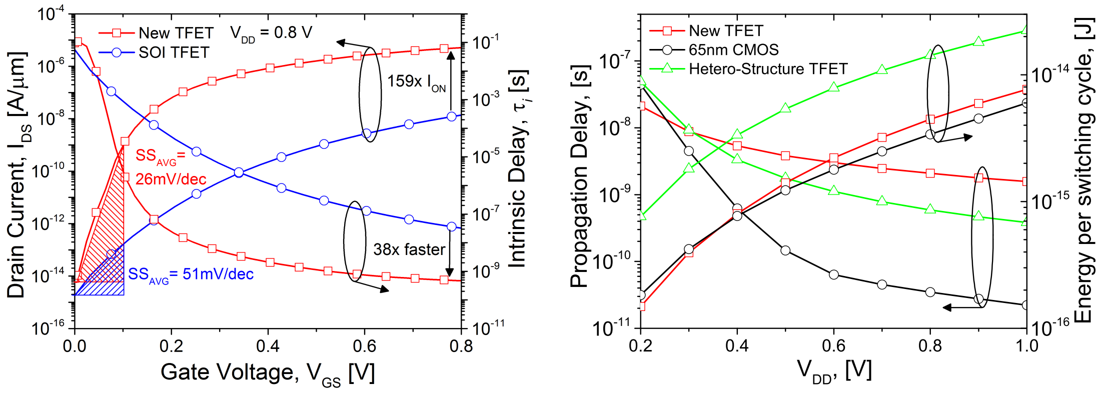


(a) (b)

Figure 2.20 – Transfer (a) and output (b) characteristics of the proposed new structure.

a gate-field induced inversion layer to the source underneath. This design can have a much larger current due to the larger tunneling area. Line tunneling structure as proposed by [42] shows improved current levels, however also has higher gate capacitances, which slightly overshadows the advantage gained by the improved on current level.

The proposed new all-Silicon structure, as shown in the Fig. 2.19 has an elevated source at an angle of 45°. The effective gate length (measured horizontally) is $50nm$. The entire source is overlapped by the gate electrode which aligns the tunneling direction to the gate-electric field. An ultra-thin BOX of $10nm$ is used to reduce the coupling between source and drain. Figure 2.19 also shows the e and hole B2B generation rates at $V_G = 0.8V$.



(a) (b)

Figure 2.21 – (a)Transfer characteristics comparison of the proposed new structure with a SOI TFET. Improved I_{ON}, better SS and faster device speed (Intrinsic delay = CV/I) is observed. (b)Variation of propagation delay and energy per switching cycle with supply voltage. The proposed TFET becomes comparable to CMOS at $V_{DD} = 0.3V$ and lower.

The simulations were performed using Synopsys TCAD ver. 2010.12 which uses a non-local B2B tunneling model to dynamically determine the tunneling path [11]. Physical models like Shockley-Read-Hall recombination, doping and field dependent mobility and Fermi statistics were also included in the simulations. For quantization effects, the MLDA (Modified Local Density Approximation) model was used. However it should be noted that quantum confinement effects are only partially taken into account [5], as the quantization model is not fully coupled with the B2B tunneling model.

Fig. 2.21(a)& (b) shows the transfer and output characteristic of the proposed device: an average sub-threshold slope (SS) of 24mV/decade over 4 decades and upto $10\mu A$ current level at $V_{DS} = 0.8V$ are observed. Fig. 2.21(a) compares the SS and intrinsic delay of the proposed structure to those of a conventional SOI TFET with 50nm gate length, 20nm T_{Si} and the rest of the parameters being similar to 2.19. Fig. 2.21(b) shows the propagation delay and switching energy per cycle for fan-out 1 inverters for three device technologies including hetero-structure (Ge-source n-type and InAs-source p-type) TFET like the ones simulated in [5] and SPICE simulated 65 nm $low - V_t$ CMOS, all the cases have been normalized with respect to cell areas. It is interesting to note that, the new structure can outperform CMOS as well as hetero-structure TFETs in terms of both propagation delay and switching energy for sub 0.25 V_{DD}.

In Fig. 2.22, at $V_{GS} = V_{DS} = 0.2V$ a C_{dg} lower than both other technologies is observed in the new structure. The higher C_{GG} of the new structure seen in Fig. 2.22 is due to the higher actual gate length of $70nm$ ($L_{g,eff} = 50nm$). This reduction in C_{DG} translates to the lowering of over/under shoot voltage of the new structure in a three-stage inverter configuration as can

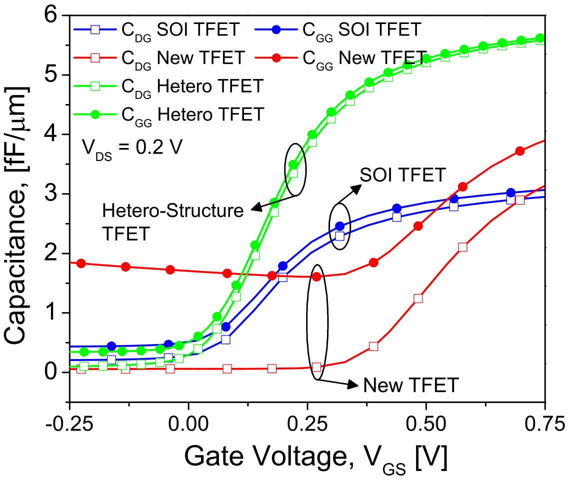


Figure 2.22 – Capacitance-Voltage characteristics showing the drain-to-gate(C_{DG}) and gate(C_{GG}) capacitances as a function of V_{GS}.

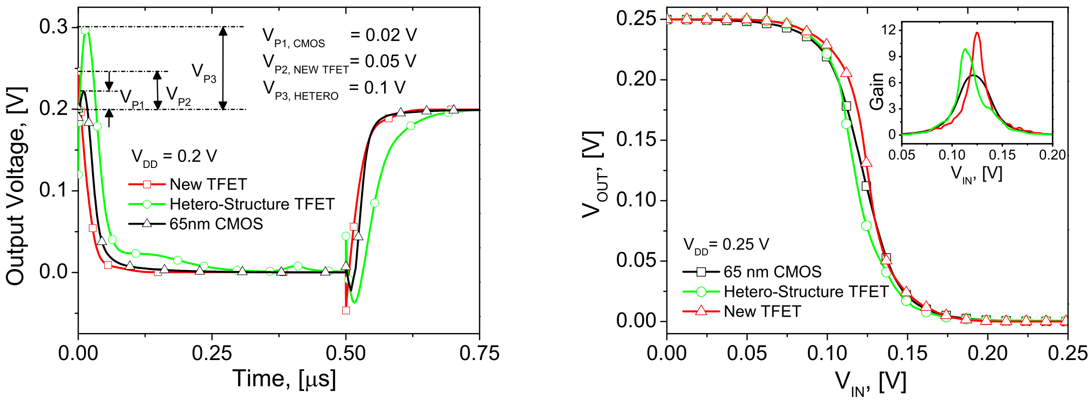
be seen from the transient behaviour of the three technologies in Fig. 2.23(a). Fig. 2.23(a) also shows lower rise and fall times for the new structure, confirming the results in Fig. 2.22.

Continuing the study at $V_{DD} = 0.25V$, Fig. 2.23(b) shows the voltage transfer characteristics of the three technologies in a single-stage inverter configuration. From the inset in Fig. 2.23(b), it is clear that the new structure has the highest gain amongst the three. Clearly the advantages of the proposed new structure are more pronounced for low supply voltages, where TFETs are intrinsically optimized for.

2.6 Summary

In this chapter, we have presented the basics of the tunneling theory. The various tunneling models available in commercial simulators are explained and the differences between them are highlighted. The principle of operation of a Tunnel FET is explained and its main difference with conventional MOSFETs are highlighted. An algorithm to calibrate the non-local tunneling model in a commercial TCAD device simulator to match measurement data has been shown. The model parameters have been adapted based on available measurement data of SOI TFET devices. Reported simulations with fitted parameter set have good level of agreement with measured data at different bias conditions for both transfer and output characteristics of the TFET devices. The dependence of the tunneling generation rate on different crystal directions was shown.

We have also proposed and simulated a new device architecture which provides higher current (159 times) as well as lower C_{MILLER} resulting in faster operation (38 times). Circuit level performance (for a inverter) is also remarkably good in low V_{DD} regime and even



(a) (b)

Figure 2.23 – (a)Transient response of the three technologies in a three-stage inverter configuration for an input period of 1μs. (b)Voltage transfer characteristics of the three technologies reported in Fig. 2.20 & 2.23(a). Inset shows the inverter Gain = dV_{OUT}/dV_{IN}.

better than CMOS at $V_{DD} = 0.2V$.

3 Compact modelling of Tunnel FET

In this chapter we focus on the compact modelling of a double gate Tunnel FET. A non-local model is first developed in section 3.1 which leads to a compact model in section 3.2. A dynamic model is shown in section 3.3. Finally the models described in section 3.2 & 3.3 is implemented in a Verilog-A based circuit simulator and some basic circuit blocks are simulated.

3.1 Overview

Tunnelling Field Effect Transistors (TFETs) are very promising devices to respond to the demanding requirements of future technology nodes [5]. The benefits of the TFET are specially linked to their potential for its sub-60mV/decade sub-threshold swing, a prerequisite for scaling the supply voltage well below 1V. Extensive experimental [23, 43] and modeling [44, 45, 25] works have been recently carried out on TFETs, but a predictive compact model that describes all the operating regimes of a TFET is still needed.

Some previous works such as [46, 47] are based on the simplified Kane's model [6] originally developed for two terminal diodes only. Kane's model does not properly account for the occupancy function dependence on the terminal bias for three terminal devices. For this reason compact models based Kane's approach might lead to inaccurate results such as non-zero drain current at zero V_D when applied to TFETs [48]. An approximate Taylor series solution for the channel potential is used in [47, 49] to calculate the tunneling current. Although this method accounts for 2D electrostatics, the results presented are not applicable to all operational regimes, as the effect of mobile charges in strong inversion condition is not taken into account. A recently presented model for double-gate tunnel FETs (DG-TFET)s [50], works well for low gate coupling but is not able to predict current levels accurately for stronger gate coupling with thinner and high-k gate dielectrics, which is essential for boosting the low current levels in TFETs [51]. A simplified model that captures the non-local nature of tunnelling and bias dependencies of the occupancy functions in all operation regimes is essential to study circuit and system level design with TFETs.

3.2 DC model version 1 (non-local)

In this work, the conformal mapping technique is applied to obtain an analytical closed form solution of the 2D Poisson's equation for a double-gate Tunnel FET. The generated band profiles are accurate in all regions of device operation. Furthermore, the current levels are estimated by implementing the non-local band-to-band model from Synopsys Sentaurus TCAD. A good agreement with simulations for varying device parameters is demonstrated and the advantages and limitations of the new modelling approach are investigated and discussed.

In this section a full 2D potential solution with conformal mapping, and non-local generation rate estimation: very accurate in estimating device characteristics with varying device parameters, however not suited to be used as a compact model in a circuit simulator due to its complex nature. Accuracy vs speed trade off: this model provides key understanding of the tunneling mechanism to further simplify the model in next section.

3.2.1 Model Description

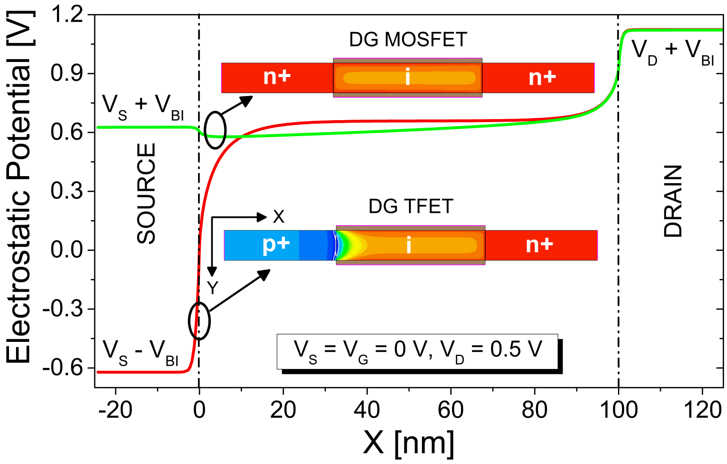


Figure 3.1 – Potential profile $\psi_s(x)$ at the Silicon/Oxide for an n-type DG-MOSFET(green) and an n-type DG-TFET(red) with the same device parameters except the source doping (n+/p+). Inset shows the electron density of the two devices at the given bias. The cross sections are taken $1nm$ below the gateoxide-Silicon channel interface.

Figure 3.1 shows the electrostatic potential at the surface of a DG-MOSFET compared to a DG-TFET at $V_S = 0V$ and $V_G = V_D = 0.5V$. The gate work function is 4 eV for both devices. The inset shows the electron density plots from numerical simulations. Also shown are the boundary conditions of the potential solution at source and drain end. There is a gradual potential drop across the channel for the MOSFET, while for the TFET, most of the potential drops across the source/channel tunnelling junction and the channel resistance is negligible as shown in Fig. 3.3. This also means that the Quasi Fermi potentials are different in the channel region of a TFET. The current in a TFET has an exponential dependence to the tunneling length [50] which depends on the band profile. For this reasons an accurate band profiling in a TFET is both challenging and essential.

In this work we consider a n-type DG lateral p-i-n TFET as shown in Fig. 3.1. A gate length $L_g = 100nm$, channel thickness of $t_{Si} = 20nm$, gate oxide thickness $t_{ox} = 3nm$ with HfO_2 dielectric, source (drain) doping of Boron (Arsenic) at $3x10^{20}cm^{-3}$ and channel doping of $1x10^{15}cm^{-3}$ were used for most of the simulations unless otherwise specified. For simplicity we consider abrupt doping profiles and ignore band gap narrowing effects in the model. Fermi statistics is included in the simulations in account of the high doping levels used. We consider tunnelling to be a 1D process, which is a reasonable approximation for well scaled devices [49]. The conformal mapping is applied only to the gated intrinsic Si-channel region of the device. As a consequence, we ignore any band bending inside the un-gated source and drain regions, meaning no depletion charges. This is an acceptable approximation for high source/drain doping levels as will be shown in section 3.3.2.

We now describe the model development step by step. The first step (A) is to have a good estimation of the surface potential (ψ_s), at the midpoint of the channel (i.e. $x = L_g/2$), which is essential for the conformal mapping step (B). The conformal mapping technique generates the band profiles from which the band-to-band (B2B) generation rate has to be computed non-locally at each spatial point (C).

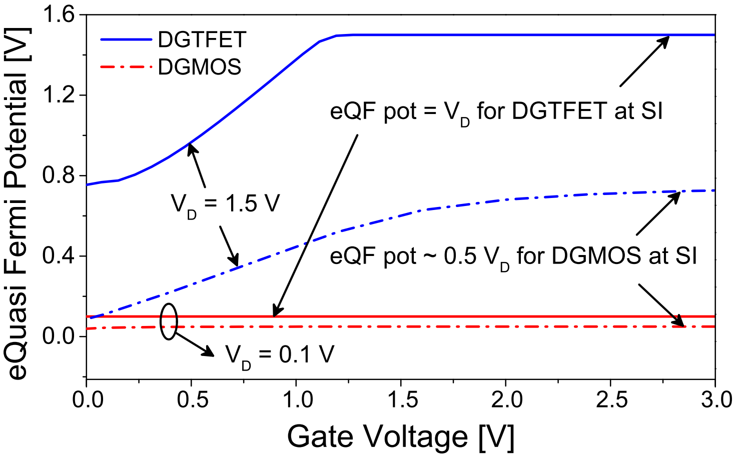


Figure 3.2 – Comparison between the electron quasi Fermi levels of a TFET and a MOSFET at two different V_D. The TFET electron quasi Fermi level stays at $-qV_D$ for almost the entire gate bias range. The cross sections are taken $1nm$ below the gateoxide-Silicon channel interface.

Surface Potential of DG-TFET

We start by checking the validity of the DG-MOSFET model reported by Sallese et al. [52] applied to a DG-TFET. We implement the simple charge based model given by:

$$V_G^* - E_{fn}(x) \approx \frac{Q_G}{C_{OX}} + U_T \log \frac{{Q_G}^2}{2\epsilon_{Si} q U_T n_i} + \frac{2Q_g}{q n_i t_{Si}} \tag{3.1}$$

$$V_G^* - \psi_S(L_g/2) = \frac{Q_G}{C_{OX}} \tag{3.2}$$

Where ${V_G}^*$ denotes the effective gate voltage ${V_G}^* = V_G - \Delta\phi_i$, $\Delta\phi_i$ is the work function difference between gate electrode and intrinsic Silicon channel, $\psi_S(L_g/2)$ is the surface potential at midpoint of the channel, Q_G is the gate charge, U_T is the thermal voltage. $E_{fn}(x)$ is the electron quasi-Fermi level which is assumed to be constant with $E_{fn}(x) = -qV_D$ throughout the channel region. This is a good approximation as can be see from the Fig.3.2 , the TFET electron quasi Fermi levels stays at $-qV_D$ for almost the entire gate bias. The gate charge is evaluated numerically from eqn.3.1 and $\psi_S(L_g/2)$ from eqn.3.2.

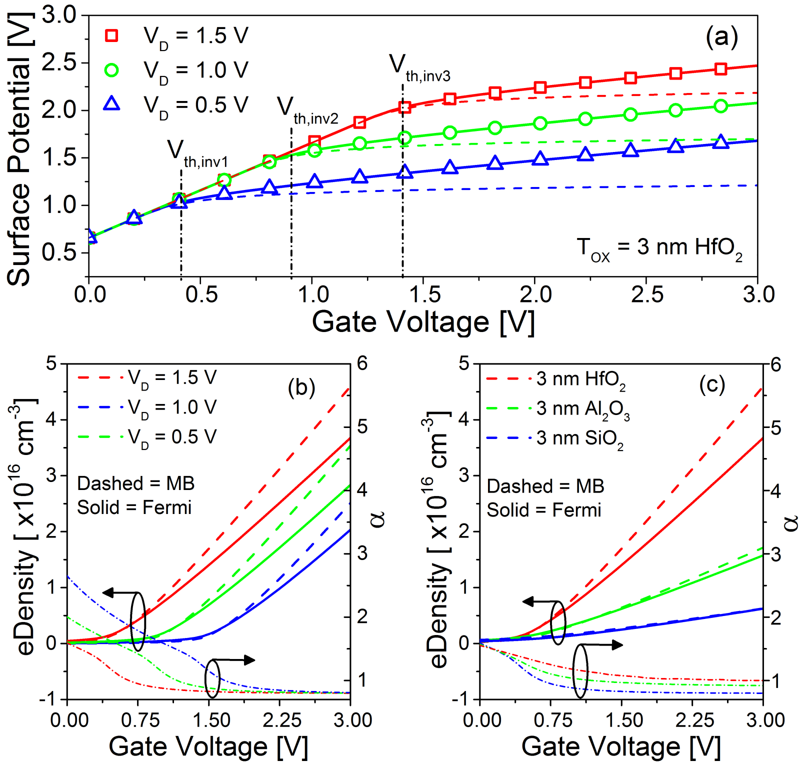


Figure 3.3 – (a) Surface potential variation with gate voltage for a DG-TFET. Solid lines represent TCAD simulations. Dotted lines indicate results from eqn.3.2. Symbols refer to eqn.3.3. Bottom: Simulated electron density with Fermi and Maxwell-Boltzmann (MB) statistics for different V_D (b) and gate dielectric materials (c). The cross sections are taken just below the gateoxide-Silicon channel interface.

Dashed lines in Fig. 3.3(a) shows the implementation of eqn.3.2 with parameters in section 3.2.1. However TCAD simulations with Fermi statistics show that TFET surface potential (solid lines) is larger than predicted by the model after strong inversion condition. This difference can be explained from Fig.3.3(b) and Fig. 3.3(c), where we compare the total space charge in the channel region with Fermi and Maxwell-Boltzmann (MB) statistics from TCAD simulations. We evaluate the ratio $\alpha = \frac{\rho_{Fermi}}{\rho_{MB}}$ where ρ is the space charge. Figure 3.3(b) shows that this factor is constant in the strong inversion regime for varying drain voltages. However as shown in Fig. 3.3(c), with varying gate dielectric material this ratio changes and approaches unity for lower gate coupling with 3 nm SiO_2. This ratio is then introduced in eqn.3.2 in to obtain eqn.3.3.

$$V_G^* - \psi_S(L_g/2) = \alpha \frac{Q_G}{C_{OX}} \tag{3.3}$$

With this observation eqn.3.3 can match the TFET surface potential as shown in Fig. 3.3(a) (symbols).

Conformal Mapping Technique

In this technique, the region wherein the 2D potential problem is solved is conformally mapped from a complex plane z into a plane w [53, 54, 55]. The function in eqn.3.4 maps a point w from the upper half of complex w-plane in complex z-plane

$$z = f(w) = 2\frac{\Delta y}{\pi} ln(\sqrt{w-1}\sqrt{w+1}) \tag{3.4}$$

Where $\Delta y = 2t_{OX} + t_{Si}$. The potential solution of any boundary value problem of first order on plane z is solved after mapping into a w-plane with the help of Poisson's integral [56].

In this section we try to adapt a recent DG-MOSFET model [57] by changing the source side boundary condition to $V_S – V_{BI}$ (TFET) instead of $V_S + V_{BI}$ (MOSFET). TCAD simulations show that TFET ϕ_S to be quite different from a MOSFET. So, an accurate surface potential is essential, to evaluate the potential in rest of the channel. By replacing the source side boundary condition and using surface potential from step (A) at $x = L_g/2$ we see that the model can accurately estimate the TFET potential profile in weak inversion as shown in the Fig. 3.4(a).

As shown in Fig. 3.3(b), after the inversion threshold voltage ($V_{th,inv}$) mobile charges starts to develop at the surface. This causes ψ_S to be almost pinned near the drain potential $V_{DS} + V_{BI}$ and only change weakly with gate bias. As explained in [58], at strong inversion condition the characteristic length (λ) decreases with increasing inversion charge. Thus λ will have a ψ_S dependence governed by eqn.3.5. Since the model developed in [57] uses a constant characteristic length, it is not applicable to strong inversion regime. For a complete study of

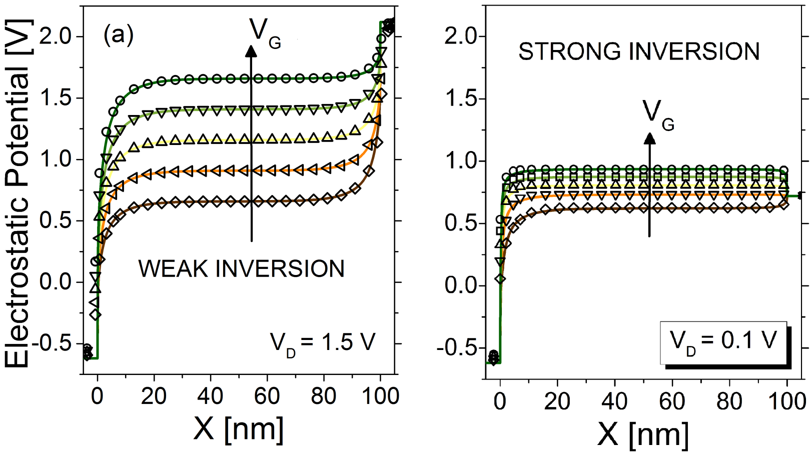


Figure 3.4 – Lateral potential profile of the channel in a DG-TFET at the surface (a) at weak inversion (b) at strong inversion. Lines indicate the model and symbols TCAD simulations. The cross sections are taken just below the gateoxide-Silicon channel interface. $V_G = 0$, 0.25, 0.5, 0.75, 1V

TFET characteristics it is important that the model also works in strong inversion.

$$1/\lambda^2 = 1/\lambda_0{}^2 + (\beta N_{inv}/\epsilon_{Si} t_{Si} \phi_S) \tag{3.5}$$

Where $\lambda_0 = \sqrt{\epsilon_{Si} t_{Si} t_{ox}/2\epsilon_{ox}}$ [59] is the characteristic length of a DG-MOSFET. N_{inv} is the inversion charge which varies with V_G and β is a parameter equal to 8 according to [58] for a DG configuration. From eqn.3.5 the value of λ is used to calculate the ratio $t_{ox}/\epsilon_{ox} = 2\lambda^2/t_{Si}\epsilon_{Si}$ at each V_G. This effective ratio then replaces t_{ox}/ϵ_{ox} in the conformal mapping model of [57]. With this modification we see that the model is accurately predicting the band profiles even in strong inversion regime as shown in Fig. 3.4(b). And since the Poisson's equation is solved in two dimensions using the technique, we can show the electrostatic potential in 2D as shown in Fig.3.5. This was not compared separately with TCAD simulations as the 1D cuts along the top of the channel where most of the generation tales place, match well the simulations.

Band-to-Band tunneling Model

With above mentioned modifications the conformal mapping step gives an accurate band profile not only at the surface but also for the entire channel domain resulting in a 2D potential distribution. The tunneling current is calculated after the band profiles are obtained from this potential. To evaluate the tunneling generation rate, we implement the phonon-assisted B2B tunneling model described in the Synopsys TCAD manual [11]. The model computes the transmission probability of holes and electrons using WKB approximation by calculating the imaginary wavevector through every point in the path. The hole generation

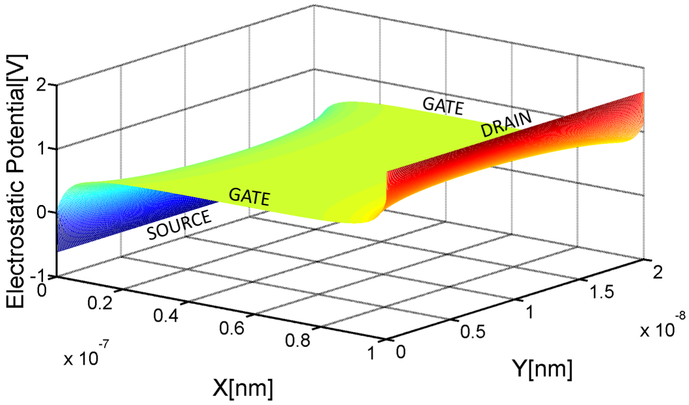


Figure 3.5 – Two dimensional electrostatic potential solution of a double gate Tunnel FET 1nm below the gateoxide- channel interface from the conformal mapping technique.

rate GB2B at a spatial point r, due to non-local B2B tunneling can be written as:

$$G_{B2B}(r) = |\Delta E_V(r)| C_P(r) T(r) f(r) \tag{3.6}$$

Where E_v is the valence band edge, C_p is a prefactor that takes into account the effect of electron-phonon interactions, T is the Wenzel-Kramers-Brillouin(WKB) integration factor describing the tunnelling probability, f is a factor that describes the electron and holes occupation probabilities.

The device is divided into horizontal slices, and in each slice, 1D B2B tunneling generation rate is obtained. The overall B2B generation is calculated by integrating the generation rates both horizontally and vertically.

$$q \int\int G_{B2B}(r) dxdy \tag{3.7}$$

The occupation probability of initial and final states is taken into account in the factor f of eqn.3.6 via the Fermi distribution functions evaluated at the initial and the final points of the tunnelling path. This ensures zero current at $V_D = 0V$.

3.2.2 Model Evaluation

The model is compared with TCAD simulations to check its validity with varying key device parameters such as gate/drain voltages, gate lengths, channel thickness, gate dielectric and source doping. Solid lines represent TCAD simulations and symbols indicate the model in the following figures. Figure 3.6(a) shows the transfer characteristics with varying V_D and Fig. 3.6(b) below shows the output characteristics for different V_G at constant source doping and gate oxide thickness. The results for $V_G{}^* \geq V_{th,inv}$ are improved by using a similar approach as in [60] of a transition function given by:

$$\beta = \beta_0[1 - c\exp(-(V_G{}^{*} - V_{th,inv} - \eta/\sigma)^2)] \tag{3.8}$$

Where $\beta_0 = 6$, $c = 0.27$, $\eta = 0.12$ and $\sigma = 0.48$. A good level of agreement is observed with TCAD simulations for all cases.

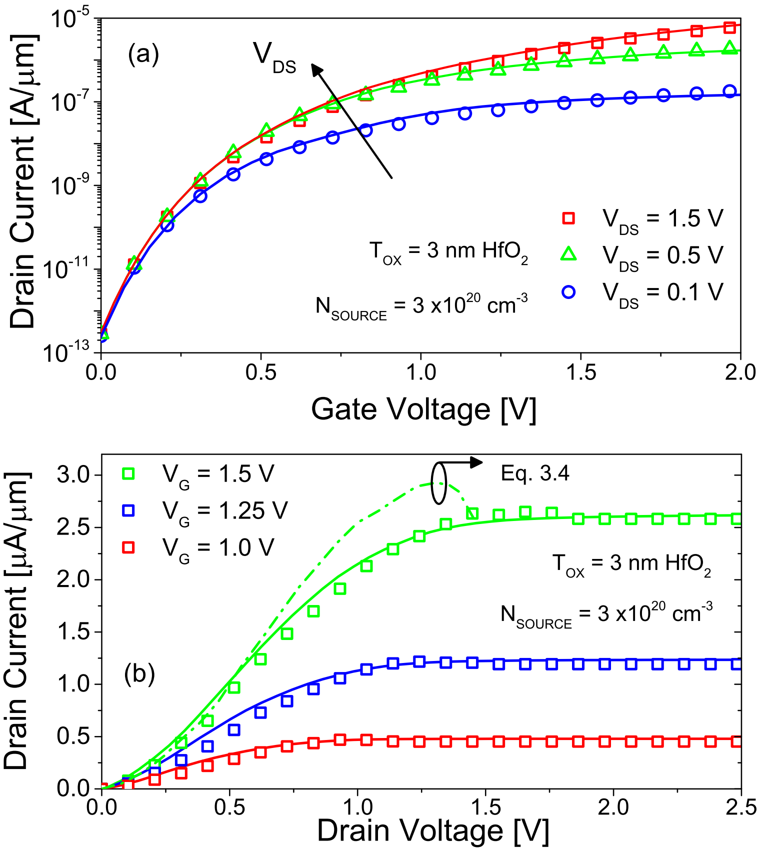


Figure 3.6 – (a) Transfer characteristics with different V_D and (b) output characteristics at different V_G. The dotted lines in the output characteristics show the curve with eqn.3.4. Rest of the plot uses eqn.3.5 with the transition function from eqn.3.8

The validity of the model is checked with varying gate dielectric constants. As explained in section II.A the parameter α was varied accordingly and a good level of agreement is observed as shown in Fig. 3.7(a).

Next we validate our model with varying source doping. It is observed that the model works well for high source doping of $2X10^{20}cm^{-3}$ or higher. For lower source doping levels a significant difference is observed. This can be explained by our initial approximation of zero band bending inside the source and drain regions. It is clear from Fig. 3.8 that the approximation is only valid for high doping concentrations. This is an inherent drawback of the model since the conformal mapping technique we have implemented can be applied only

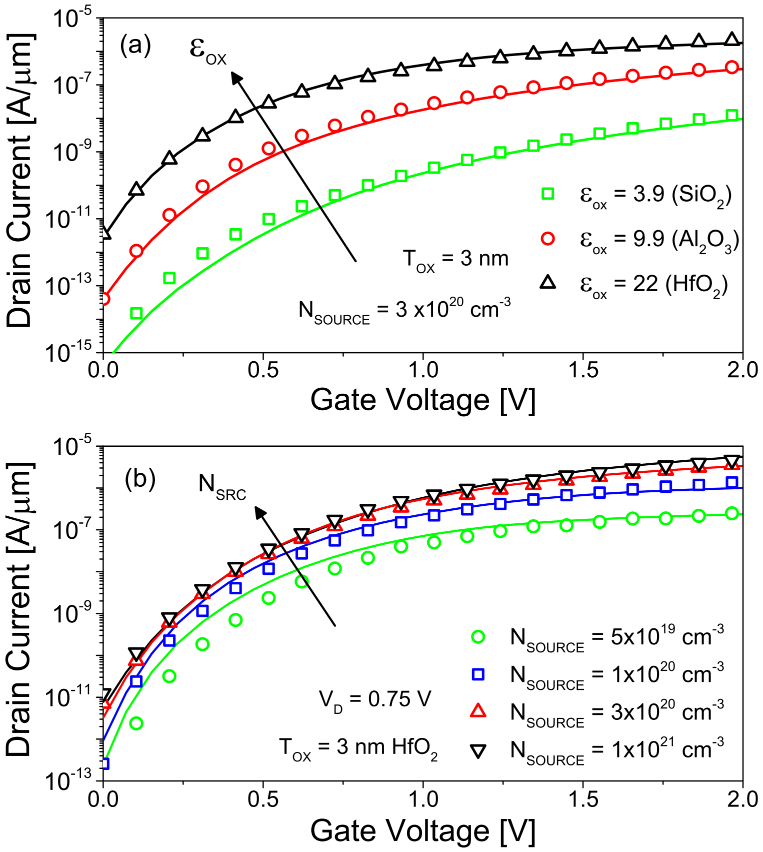


Figure 3.7 – (a) Transfer characteristics with varying gate dielectric constant. $\alpha = 0.8$ for HfO_2, 0.92 for Al_2O_3 and 1 for SiO_2 gate dielectric. (b) Transfer characteristics with varying source doping level. Low doping curves were corrected according to table 3.1. $L_g = 22nm$, $t_{Si} = 10nm$.

to the gated region; we cannot estimate the band bending in the non-gated regions of the device. This limitation can be overcome by introducing an empirical shift in the conduction band which will compensate for the underestimated tunnelling length. This factor will depend on the doping level with a larger shift for lower doping levels. As shown in Fig. 3.7(b), with this empirical correction the model can estimate the current levels even at lower source doping levels. Details of this empirical factor is listed in table 3.1 for different values of doping. No corrections were necessary for doping levels higher than $2X10^{20}cm^{-3}$.

Despite the fact that low doping levels need empirical correction, it should be noted that an excellent agreement is obtained without any corrections for high doping and high-K gate-dielectric that corresponds to the cases of major interest being associated to high currents [50]. Finally we validate our model for source doping dependence in output characteristics. The same corrections as in table 3.1 was applied and we observe a good match with simulations as shown in Fig. 3.9 below.

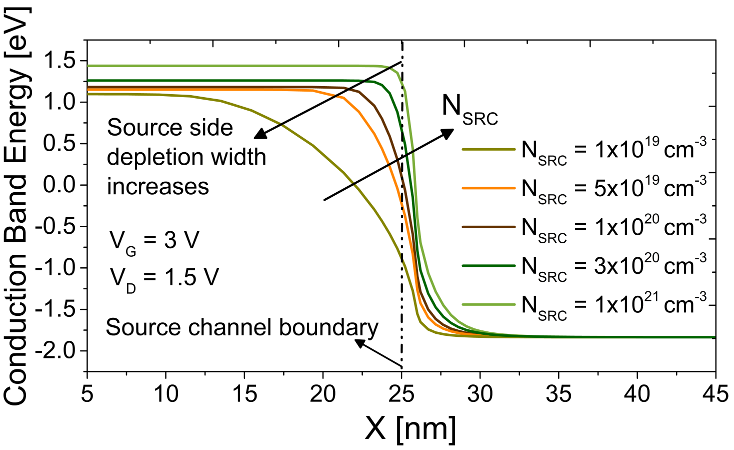


Figure 3.8 – Conduction band energy profile at the source channel boundary for different source doping levels. The source depletion width increases with decreasing source doping.

Table 3.1 – Empirical conduction band shift for different doping levels

Doping level (cm^{-3})	Correction (nm)
$5x10^{19}$	0.9
$1x10^{20}$	0.6
$3x10^{20}$	0.0
$5x10^{20}$	0.0

It is interesting to note that the model is able to capture the super-linear onset dependence [25] on source doping well, thanks to the accurate B2B tunnelling model featuring a realistic dependence of the occupancy functions.

3.3 DC model version 2 (local)

A recently presented model [61] for double-gate tunnel FETs (DG-TFET), was the first TFET compact model to be implemented in Verilog-A, however it did not describe the ambi-polar behavior typical of TFETs [23] or the drain side breakdown phenomena completely. A compact model that captures all tunneling related physics and bias dependencies of the occupancy functions in all operation regimes is essential to study circuit and system level design with TFETs. In this section, a compact model based on an analytical closed form solution of the 1D Poisson's equation for a double-gate Tunnel FET is derived. Furthermore, the current levels are estimated by implementing an algorithm based on the Kane's band-to-band tunnelling model. A good agreement with numerical simulations for varying device parameters is demonstrated and the advantages and limitations of the modelling approach are investigated and discussed. The model is implemented in a Verilog-A based circuit simulator and basic circuit blocks like an inverter, a 2-bit half adder and a 15 stage ring oscillator are

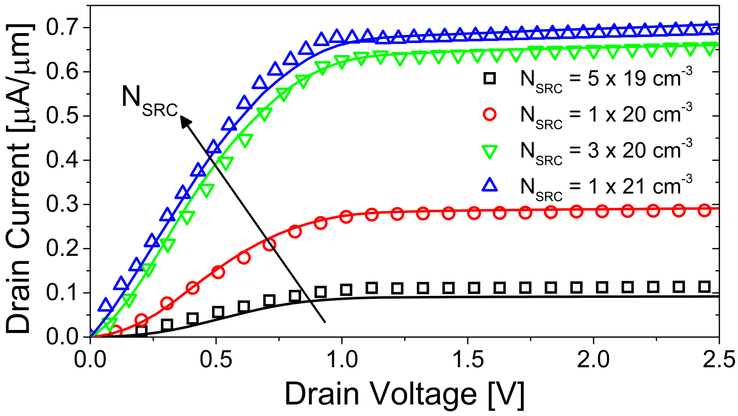


Figure 3.9 – Impact of source doping on the TFET output characteristics. The same transition function of eqn.3.8 is used here.

simulated to demonstrate the capabilities of the model. The switching energy of a Tunnel FET based circuit block is studied with V_{DD} scaling revealing interesting aspects of Tunnel FET circuit behaviour.

3.3.1 Model Description

The model development will be described step by step in this section. We consider an n-type all-Silicon DG *p-i-n* TFET as shown in Fig. 3.10(a). A gate length L_G=22nm, channel thickness of t_{Si}=10nm, gate oxide thickness t_{ox}=3nm with HfO_2 gate dielectric, source (drain) doping of Boron (Arsenic) at $10^{20} cm^{-3}$ and channel doping of $10^{15} cm^{-3}$ are used for all the simulations unless specified otherwise. For simplicity we consider abrupt doping profiles and ignore band gap narrowing effects in the model. Fermi statistics is included in the simulations to account for the high doping levels used. The Poisson's equation is solved in 1D only in the gated intrinsic body region of the device. As a consequence, we ignore any band bending inside the un-gated source and drain regions. This is an acceptable approximation for high source/drain doping levels as required for typical TFET specifications.

3.3.2 Surface potential

The 1D Poisson equation is solved in the lightly doped body of the DG-TFET. Under the assumption of a 1D electric field [62] in the gate-oxide material the Poisson equation is transformed into a simplified form as in Eqn.3.9.

$$\frac{\delta^2 \psi_S}{\delta_x^2} + \frac{\psi_{GS} - \psi_S}{\lambda^2} = \frac{qN_{ch}}{\epsilon_{Si}} \tag{3.9}$$

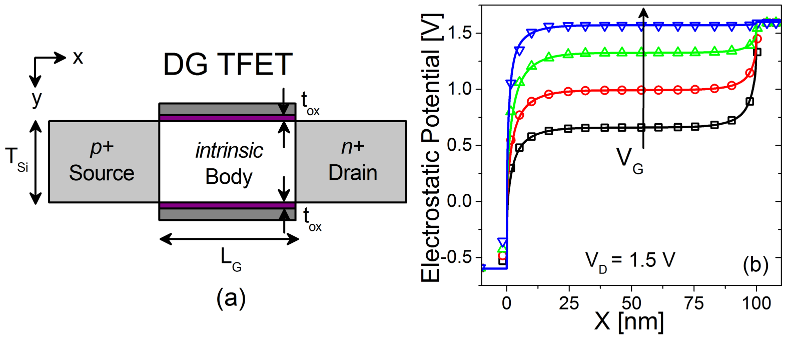


Figure 3.10 – (a) Cross section of the device under study with typical dimensions marked (b) lateral potential profile of the channel in a DG-TFET at the oxide-Silicon interface at strong inversion for V_{GS}= 0,0.25,0.5,0.75,1V. Lines indicate the model and symbols TCAD simulation in all relevant plots.

where $\psi_s(x)$ is the surface potential, ψ_{GS} is the gate potential ($= V_{GS} - V_{FB}$) with V_{FB} being the flat band voltage, ϵ_{Si} is the dielectric constant of Silicon and λ is the natural length of a double gate structure [62]. The source and drain side boundary conditions are defined as $\psi_S(0) = (V_S - V_{B_{SRC}})$ and $\psi_S(L_g) = (V_D + V_{B_{DRN}})$, with $V_{B_{SRC}}$ and $V_{B_{DRN}}$ being the built in potential of the source/drain-body junction respectively. The solution to eqn.3.9 under the defined boundary conditions is given as:

$$\begin{aligned}\psi_S &= \psi_S^0 \\ &+ (V_D - V_S + V_{BI_{DRN}} + V_{BI_{SRC}})\frac{\sinh(x/\lambda)}{\sinh(L_g/\lambda)} \\ &+ (V_S - V_{B_{SRC}} - \psi_S^0)\frac{\sinh(x/\lambda) + \sinh(L_g - x)/\lambda)}{\sinh(L_g/\lambda)}\end{aligned} \tag{3.10}$$

where ψ_S^0 is the long channel surface potential of a DG-TEFT from section 3.2.1 [29]. The results from eqn.3.10 is compared to TCAD simulations in Fig. 3.10(b) and show a good match.

3.3.3 Tunnelling path estimation

We know that the choice of the tunneling path is of utmost importance [63, 64]. In section, the optimal location of the tunneling path is explained. Using the potential solution described above and the model described in [29], the hole band-to-band generation (B2BG) rate and the Fermi occupation is evaluated non-locally for an n-type device to understand the main operation regimes of a TFET. The operation of a TFET can be broadly divided in four regimes (i) OFF state (ii) ON state (iii) ambipolar state (iv) breakdown state. Each of theses states will now be explained to understand the device operation.

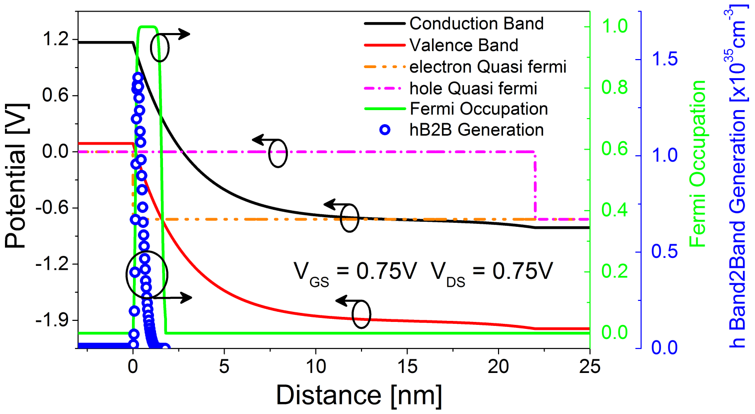


Figure 3.11 – The operating regimes of a TFET: ON state with $V_{DS} = V_{GS} = V_{DD}$, where supply voltage $V_{DD} = 0.75V$

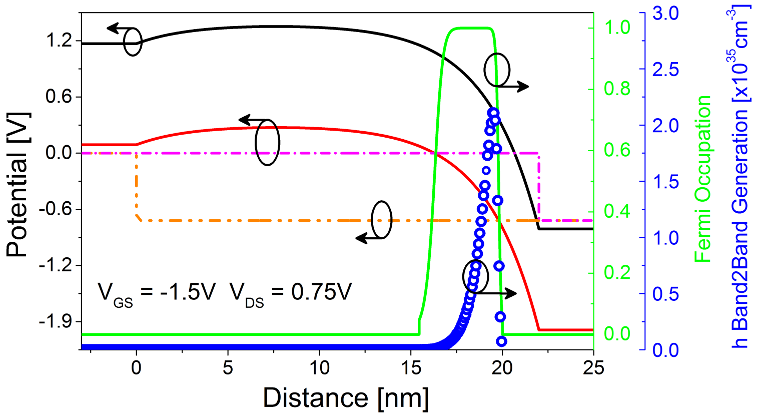


Figure 3.12 – The operating regimes of a TFET: Ambipolar state with $V_{DS} = V_{DD}$ and $V_{GS} = -V_{DD}$, where supply voltage $V_{DD} = 0.75V$

- Firstly, in the ON state (e.g. $V_{DS} = V_{GS} = V_{DD}$), the positive V_{DS} applied results in a small Fermi window close to the source as shown in Fig. 3.11. The hole B2BG (in blue symbols) is confined just to the Fermi window (green solid line) as expected.

- In the ambi-polar case (e.g. $V_{DS} = V_{DD}$ and $V_{GS} = -V_{DD}$), as shown in the Fig. 3.12 the Fermi window shifts towards the drain and all the holes are generated only close to the drain.

- Finally, in the breakdown state (e.g. $V_{DS} = V_{DD}$ and $V_{GS} = V_{DD}$), due to the large V_{DS} applied the Fermi window now extends across the entire body as shown in Fig. 3.13. Due to the very wide Fermi window, tunneling occurs near both source and drain junctions resulting in a large current.

- In the OFF state, V_{DS}=0V results in zero Fermi occupancy window and hence no band-

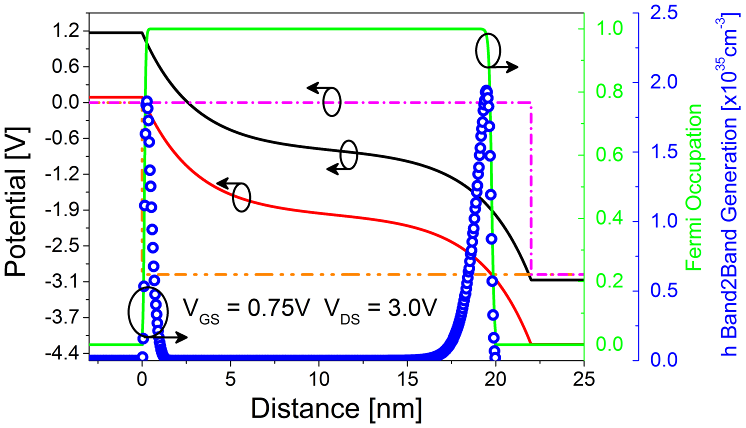


Figure 3.13 – The operating regimes of a TFET: Breakdown state with $V_{DS} = V_{DD}$ and $V_{GS} = V_{DD}$, where supply voltage $V_{DD} = 0.75V$

to-band generation rakes place, hence zero tunnelling current irrespective of the gate bias applied. The only current flowing is due to leakage mechanisms.

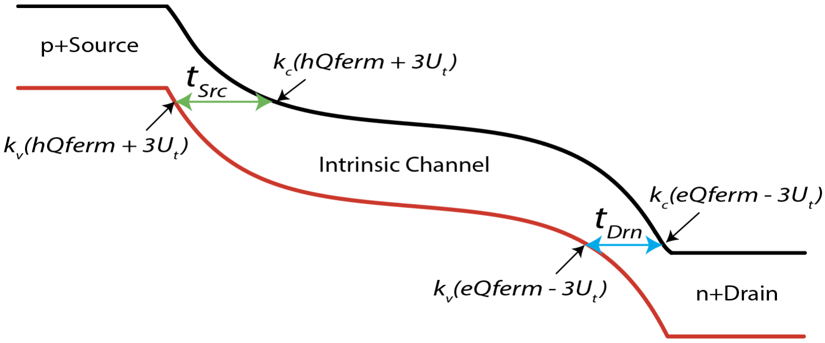


Figure 3.14 – Schematic showing the location of the two tunnelling paths considered. The breakdown state biasing is used here to show both the tunnelling paths.

It is important to notice the location of the peak of the hole B2BG in all the three conducting states of a TFET in Fig. 3.11, Fig. 3.12 and Fig. 3.13. On account of this observation, we can simplify the non-local tunnelling process to a local process with the tunnelling paths considered only at two locations: intersection of (i) $(hQferm + 3U_t)$ level with the valence band and (ii) $(eQferm - 3U_t)$ level with the valence band, where U_t is the thermal voltage. The $\pm 3U_t$ term is included to consider the tunneling at the edges of the Fermi occupation window where the tunneling widths are the smallest. Figure 3.14 shows the location of the $(e/hQferm \pm 3U_t)$ levels along with the two tunnelling paths. The above discussion is only for hole B2B generation. The same analysis holds true for electron B2B generation as well, which will be equal and opposite in magnitude to the hole B2B generation.

3.3.4 Tunneling current calculation

The tunnelling current consists of two components as explained in the section above. The Kane's model is used to estimate the tunnelling generation rates from the tunnelling width. The total tunnelling current I_{tunn} is then calculated from the sum of G_{src} and G_{drn} as shown in eqn.3.13.

$$G_{src} = A_{path}\left[\frac{E_g}{t_{src}}\right]^P \exp\left(\frac{-B_{path}t_{src}}{E_g}\right) \tag{3.11}$$

$$G_{drn} = A_{path}\left[\frac{E_g}{t_{drn}}\right]^P \exp\left(\frac{-B_{path}t_{drn}}{E_g}\right) \tag{3.12}$$

$$I_{tunn} = qW_{ch}t_{ch}(G_{src} + G_{drn}) \tag{3.13}$$

where A_{path} and B_{path} are fitting parameters described in [11], P is a constant =2.5 for indirect tunneling and =2.0 for direct tunnelling [11], E_g is the energy band-gap of Silicon, q the electronic charge, W_{ch} & t_{ch} are the body width and thickness. t_{src} and t_{drn} are the source and drain side tunnelling lengths defined as follows:

$$\begin{aligned} t_{src} &= \kappa_C[CB(hQf + 3U_t)] - \kappa_V[VB(hQf + 3U_t)] \\ t_{drn} &= \kappa_C[CB(eQf - 3U_t)] - \kappa_V[VB(eQf - 3U_t)] \end{aligned} \tag{3.14}$$

where $\kappa_C(E)$ and $\kappa_V(E)$ are the inverse function of eqn.3.10 which gives the values of position x for a given value of conduction band (CB) and valence band (VB) energy respectively and eQf and hQf are the electron Fermi level as shown in Fig. 3.14. As described in section IIB and in Fig. 3.14, the tunnelling lengths are evaluated by the difference in the position x, where the hole and electron quasi Fermi levels $\pm 3U_t$ intersects the VB and the CB. $\kappa(E)$ (evaluated in Matlab), gives us flexibility to choose a tunneling length at any point in the Fermi window. Hence, here we do not describe an expression for any specific tunnelling path. Finally, a Fermi factor as described in [11] is multiplied to eqn.3.13 to account for the available carriers at the beginning of the tunnelling path, and for the available states into which they can tunnel, at the end of the tunnelling path. This avoids non-zero current at V_{DS}=0V as well as results in a smooth and continuous current in all biasing conditions. This algorithm can be applied to any lateral tunnelling TFET architecture, provided $\kappa(E)$ exists for the surface potential solution used.

3.3.5 Comparison with simulations

The model is compared with TCAD [11] simulation results in this section. Figure 3.15(a) shows the comparison of the transfer curves at three different V_{DS}. It can be seen that the model describes all the operation TFET regimes properly, including the ambi-polar regime in this case. Fig. 3.15(b) shows the output characteristics at three different V_{GS}. We can see

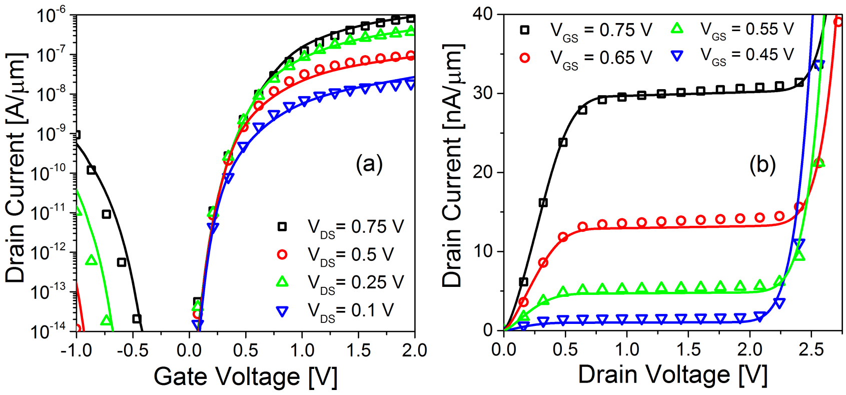


Figure 3.15 – (a) Transfer characteristics for different V_{DS} (b) Output characteristics for different V_{GS}.

from Fig. 3.15(b), that the model can describe the super-linear onset [25] and also the drain side breakdown of a TFET. This is possible due to physically meaningful assumptions and in particular the dual tunnelling path consideration in this model. This algorithm is similar to a previously reported work [45], however, in this work we consider only 2 tunnelling paths.

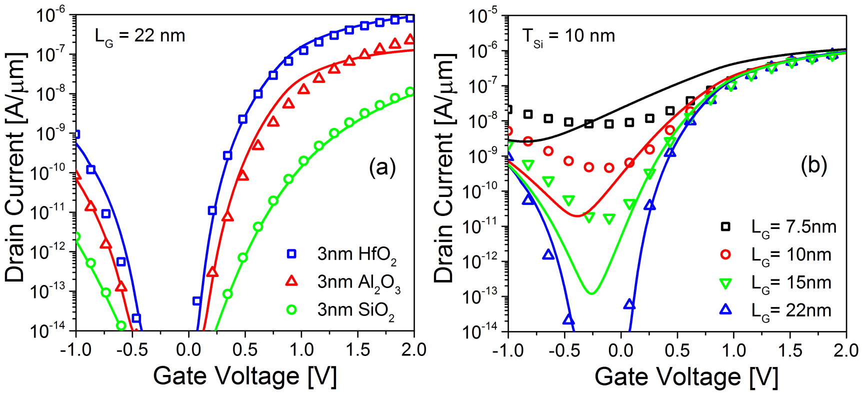


Figure 3.16 – (a) Transfer curves with three different gate oxide material. (b) Transfer curves with gate length scaling with other parameters constant.

In Fig. 3.16(a) we compare the transfer curves for three different gate-oxide material at 3nm thickness. The effect of gate length scaling on a TFET is shown in Fig. 3.16(b). As expected, TFET has a weak dependence on gate length scaling (V_{DS}=constant), except when the gate length reduces to about 7.5nm, where the electrostatic control of the gate is too weak to have a reliable FET operation (for t_{ch}=10nm). As no leakage mechanism is included in this model at this stage, the match here is not perfect. However, the trend is captured correctly.

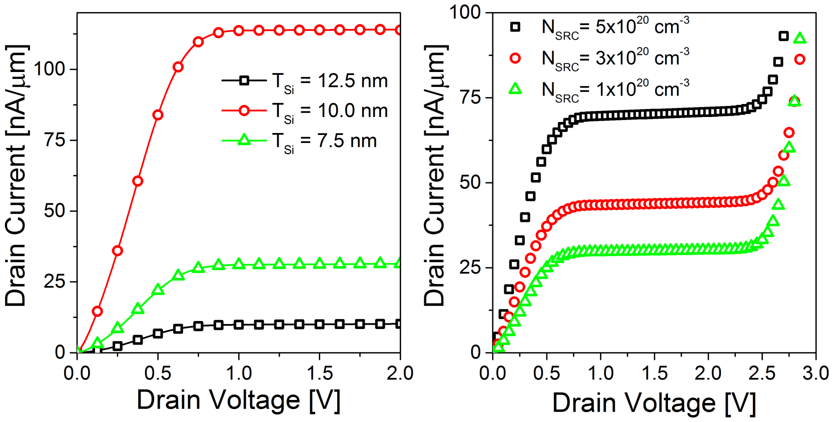


Figure 3.17 – (a) Transfer curves with three different body thickness. (b) Transfer curves with source doping variation.

3.4 Charge model of a DG-TFET

In this section we develop a simplified capacitance model for Double Gate TFETs. Previous efforts on capacitance model of Tunnel FET [65] uses a BSIM [66] solution to compute the inversion charges and hence the capacitances. [67] on the other hand talks mostly of fringing capacitances. A simplified model which uses the potential solution discussed in the previous section will be needed to get a complete DC/AC compact model.

To understand the underlying physics, capacitance-voltage measurements were done on all-Silicon SOI TFETs at different biasing schemes to support the model development. TCAD simulations [11] of DG-TFETs were used to validate the model.

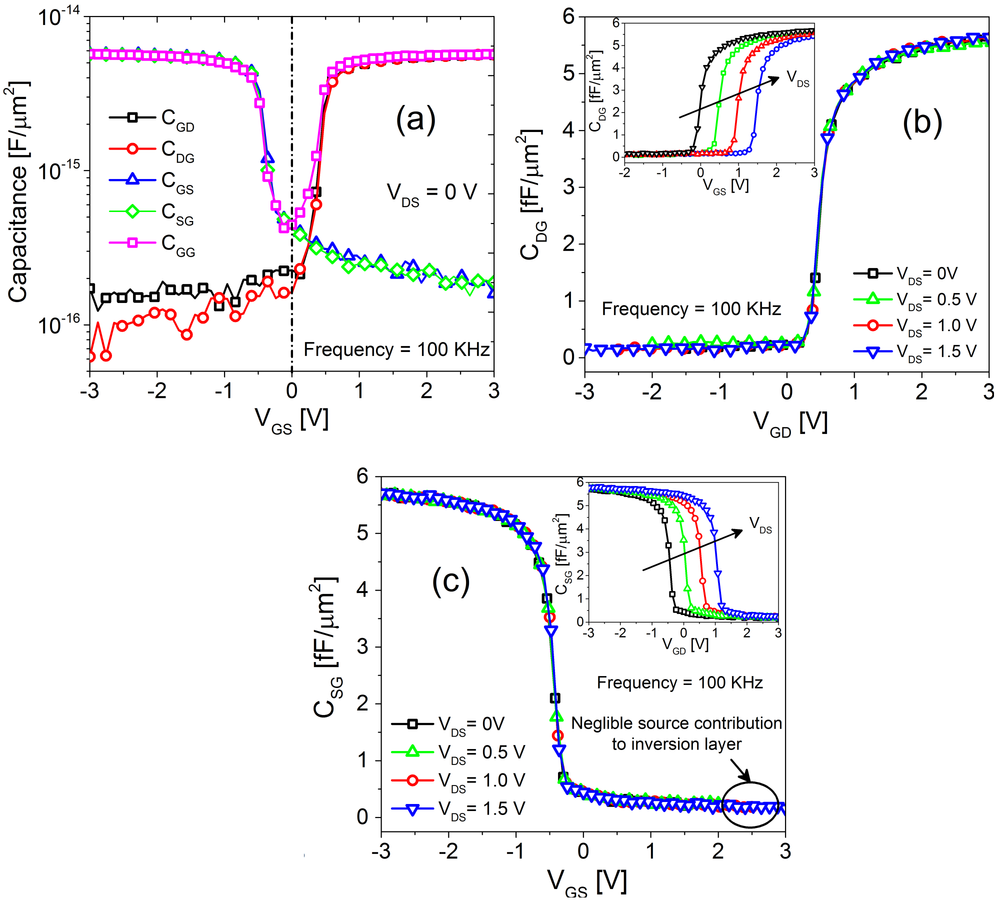


Figure 3.18 – (a) Measured C_{GS}, C_{GD} and C_{GG} with respect to V_{GS}. Perfect symmetry is observed highlighting the ambipolar nature of TFETs. (b) Measured Gate-Drain capacitance for different V_{DS}. Inset shows measured C_{GD} curves with respect to V_{GS}. (c) Measured Gate-Source capacitance for different V_{DS}. Source contribution to inversion electrons is negligible even at $V_{DS} = 1.5V$ with device turned ON. Inset shows measured C_{GS} curves with respect to V_{GD}. Fully depleted SOI TFET with $L_G = 200nm$, $t_{Si} = 21nm$ and 6 nm SiO_2 dielectric

3.4.1 CV measurements on FD-SOI TFETs from LETI

Capacitance-Voltage (CV) measurements were performed on all-Silicon SOI TFETs [23]. The HP 4284A precision LCR meter was used for all the CV measurements [68]. The DC bias was varied from -3 V to +3 V for most cases with a small AC signal of 30 mV. To measure C_{GS}, the low terminal of the setup was connected to the gate which monitors the current and the high terminal to the source where the bias is applied according to the equation 3.15.

$$C_{GS} = -\frac{\partial Q_G}{\partial V_S} | V_G, V_D, V_{SUB} \tag{3.15}$$

CV measurements on TFETs show perfectly symmetric nature due to ambi-polar characteristics [5] of TFETs as shown in Fig. 3.18. TFET capacitances were also found to be reciprocal in nature which means $C_{ij} = C_{ji}$ for $i, j = G, S, D$ at V_{DS}=0 V (Fig. 3.18(a)).

Drain-gate capacitance was measured at different values of V_{DS} in Fig. 3.18b. Similarly source-gate capacitance was measured at different values of V_{DS} in Fig. 3.18c. We observe that gate-source capacitance (C_{GS}) under strong inversion remains negligible compared to gate-gate capacitance (C_{GG}) even at high V_{DS} = 1.5 V (also shown in [69, 70]). So, it can be said that source has negligible contribution to the inversion charges, and that only depletion charges contribute to the source capacitance. Hence C_{GG} is dominated by the gate-drain capacitance (atleast for low injection levels smaller than the depletion charges, as in the studied devices). This is also the reason for the strong Miller effect in TFETs [41].

3.4.2 Suitability of 100-0 (drain/source) charge distribution

Further to the measurements, TCAD simulations were done to create a fictitious device TFET A with 2 orders of magnitude more current than a template 100 nm device in Fig. 3.19(top). As shown in the capacitance simulation in Fig. 3.19(bottom), the TFET A device with almost mA current level still has negligible C_{GS} compared to C_{GD} . TFET A shows a slightly higher C_{SG} and slightly lower C_{DG} compared to TFET B. This plot clearly shows that the additionally injected carriers from the source are still negligible compared to the inversion charge. So we can say that tunnelling generated carriers has little or negligible role to play in the charge distribution of a tunnel FET, atleast for the tunnelling injection levels studied. This makes the task of modelling the charge distribution in a TFET easier as we can now take following assumptions reliably:

- 100-0 charge partitioning scheme with 100% to drain.
- Source depletion charge entirely dictates the source-gate capacitance.

However it should be noted that in low bang-gap and low effective mass material like InAs based TFETs will show a much higher injection than all-Silicon TFETs. Also since InAs have a lower density of states, this limits the gate-drain capacitance and hence the impact of Miller effect. Hence the above two assumption will not hold true for such cases.

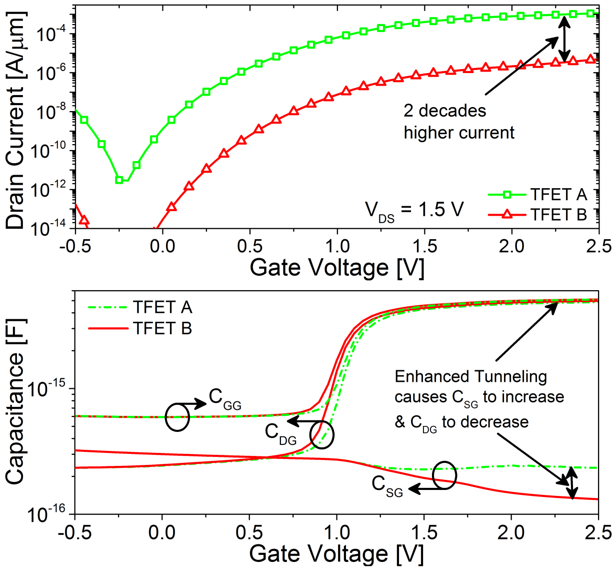


Figure 3.19 – (top) TFET A simulated with artificially enhanced tunnelling injection to have 2 order of magnitude more current than TFET B. (bottom) TFET A shows a slightly higher C_{SG} and slightly lower C_{DG} compared to TFET B. C_{SG} remains negligible compared to C_{DG} in TFET A, verifying that tunnelling has little or negligible influence in the charge distribution of a TFET.

3.4.3 Description of charge model of a DG-TFET

A 100nm double gate device with 3nm HfO_2 ($\epsilon_r = 21$) gate dielectric and 20nm Silicon body thickness was used in the simulations. Source/drain doping was $1X10^{20}cm^{-3}$ and abrupt junctions were assumed. As described in [29] and also in section 3.1, eqn.3.16 is used to compute the gate charge Q_G. The gate charge Q_G is then used to compute the surface potential ψ_S in eqn.3.17. The surface potential ψ_S used with Gauss law to approximate inversion charges in eqn. 3.18. The drain-body depletion charge under the gate is also estimated by Gauss law in eqn. 3.19. The total drain charge is computed in eqn.3.20 and finally the source charge is evaluated in eqn.3.21 by following charge conservation rule.

$$V_G^* - E_{fn}(x) \approx \frac{Q_G}{C_{OX}} + U_T \log \frac{{Q_G}^2}{2\epsilon_{Si} q U_T n_i} + \frac{2Q_g}{q n_i t_{Si}} \tag{3.16}$$

$$V_G^* - \psi_S(L_g/2) = \frac{Q_G}{C_{OX}} \tag{3.17}$$

$$Q_{D\_INV} = C_{OX}(V_{GS} - V_{FB} - \psi_S)(L_g W_g) \tag{3.18}$$

$$Q_{D\_DEPL} = \epsilon_{Si}(V_{GS} - V_{FB} - \psi_S)(t_{Si} W_g)/\lambda \tag{3.19}$$

$$Q_{DRAIN} = -(Q_{D\_INV}) + Q_{D\_DEPL} \tag{3.20}$$

$$Q_{SOURCE} = -(Q_{DRAIN} + Q_G) \tag{3.21}$$

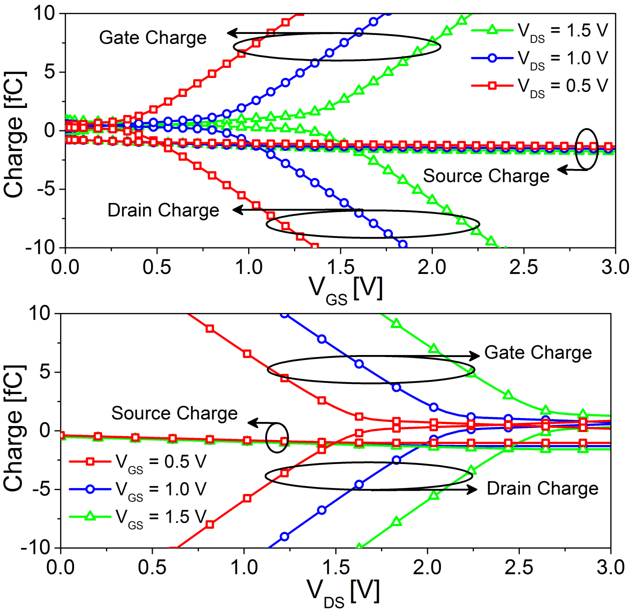


Figure 3.20 – Modelled gate, drain and source charges of an n-type double gate tunnel FET (top) as a function of gate-source voltage and (bottom) as a function of drain-source voltage.

Where λ (lambda) is the characteristic length of a double gate structure [59], V_{BI} corresponds to the built in potential in the source-channel pn junction, U_T is the thermal voltage, C_{OX} is the gate-oxide capacitance, V_{GS} is the gate-source bias, V_{FB} is the flat band voltage, L_g and W_g are the length and width of the gate, V_{DS} is the gate-source bias, t_{Si} is the Silicon layer thickness. Capacitances are then evaluated as derivatives of the computed charges. Earlier work on capacitance behaviour of a TFET such as [65] is based on BSIM, whereas [50] also compute charges but do not discuss capacitances voltage behaviour.

3.4.4 Comparison with TCAD simulation

The drain, source and gate charges with respect to V_{GS} and V_{DS} are shown in Fig. 3.20. As discussed earlier source charges remain negligible compared to drain charges even

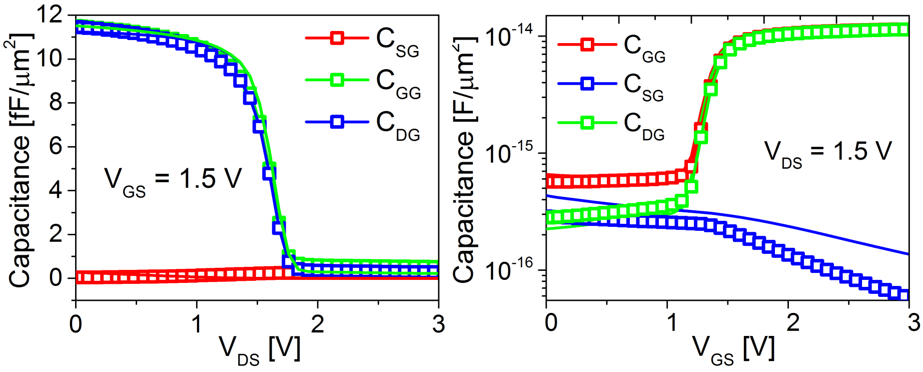


Figure 3.21 – C_{GG}, C_{SG} and C_{DG} with respect to drain-source voltage at $V_{GS} = 1.5V$ (left) and gate-source voltage at $V_{DS} = 1.5V$ (right) computed by the simplified model and comparison with TCAD simulations. Symbols indicate simulations and solid lines indicate model for all relevant figures.

at V_{DS}=1.5 V. The capacitance curves with respect to gate-source voltage (top) and drain-source voltage (bottom) are shown in Fig. 3.21. We see that in spite of the approximations introduced the model estimates the trans-capacitances with relative accuracy both above and below the inversion threshold. Finally as shown in Fig. 3.22, the model also works for varying drain voltages for C_{GG}, C_{SG} and C_{DG} @ V_{DS} = 0.5V, 1.0V, 1.5V. C_{SG} particularly shows large discrepancy due to the assumption of a constant depletion width for all t_{Si}. However the shape and the variation with V_{DS} and V_{GS} is captured in this model.

3.5 Verilog-A implementation of full compact model

The model described in section 3.2 and 3.3 has been implemented in Verilog-A. Simple circuit level simulations were carried out using the Cadence Virtuoso tool. A p-type device was emulated from the n-type device for this study. Tunnelling generation is assumed to be time independent and to have no impact on the charge distribution in a TFET. The dynamic behaviour of TFETs is modelled with a 100/0 charge partitioning scheme with 100% to drain as described in section 3.4.

To ensure continuity and smoothness of the model in all bias conditions, a basic diode model (eqn. 3.22) was added to take into account negative V_{DS}. No gate bias dependence is assumed for the forward bias diode current. This results in the typical uni-directional conduction feature of a TFET [61] as shown in Fig. 3.23. This unidirectional conduction has some interesting circuit implications and has been studied in literature [24, 27, 28]. In addition to reduce complexity of this level 1 model, no gate current model or any leakage mechanism were added to this version of the model. A leakage floor was added to avoid less than 1fA/μm current.

$$I_{diode} = I_S\left(exp(V_D/\eta V_T) - 1\right) \tag{3.22}$$

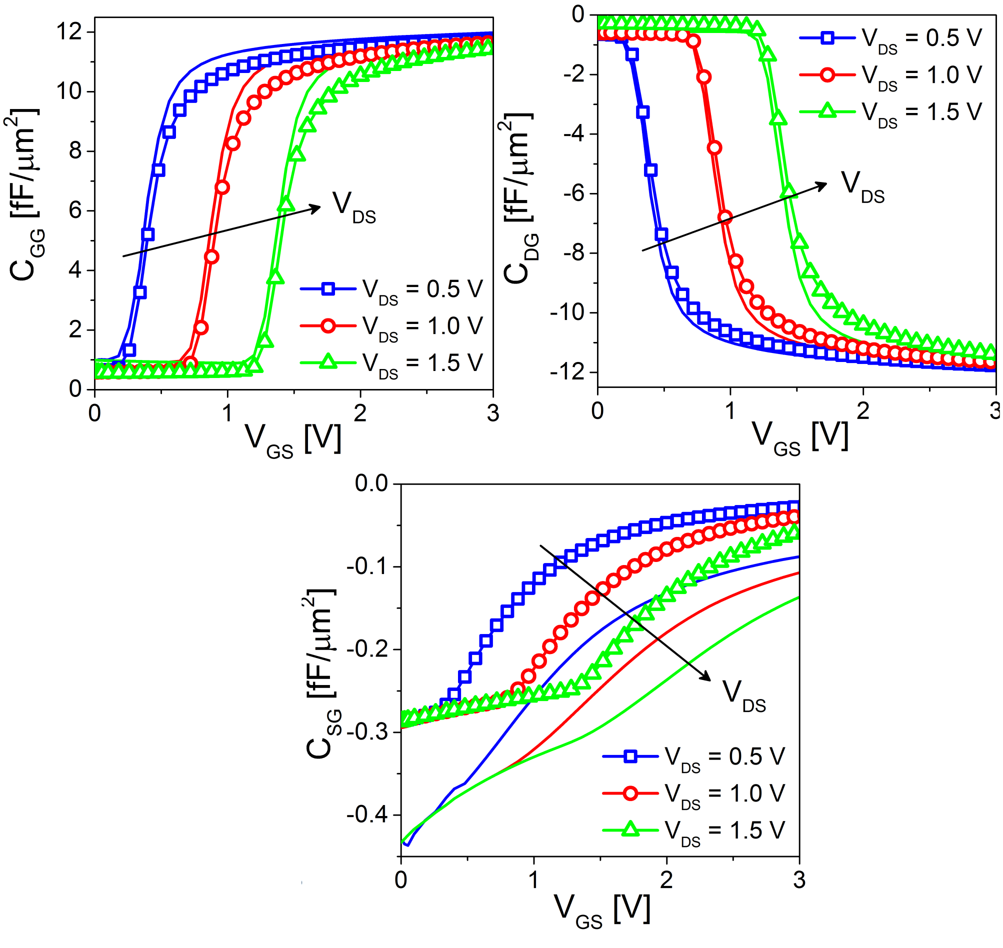


Figure 3.22 – Modelled C_{GG}, C_{SG} and C_{DG} curves with respect to V_{GS} for three different drain voltages and comparison with TCAD simulations. Modelled C_{SG} curves with respect to V_{GS} although do not show a good match, predicts the trend correctly

where I_S is the saturation current, V_D is the voltage across the diode, V_T is the thermal voltage and η is the ideality factor (=1). It should be noted that no analytical solution to the Poisson's equation of a DGTFET (eqn.3.9) exits. Hence we implemented a Newton Raphson algorithm to evaluate the charge in eqn.3.9 and determine the surface potential in eqn.3.10. The number of iterations were limited to 9 as convergence was achieved in 7-8 steps in all studied cases.

In this subsection we demonstrate the capabilities of the Verilog-A code based model to simulate multi transistor circuit blocks without any convergence or noise issues. Some basic circuit blocks as shown in Fig. 3.24 are simulated capturing the unique features of TFETs in a circuit. The model parameters and their values used in the following simulations are shown in table 3.2. The A_{path} and B_{path} are the main parameters used to tune the model to the simulation results as discussed in section 2.4.

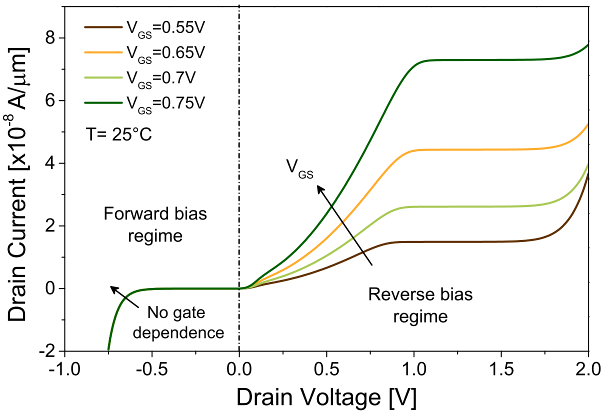


Figure 3.23 – Modelled output characteristic of a n type TFET showing both positive and negative drain voltages. Note the forward biased drain current has not gate dependence.

Table 3.2 – Model parameters available in Verilog-A

Parameter	Value	Parameter	Value
$A_{path}[cm^{-3}s^{-1}]$	$2.4 \cdot 10^{13}$	$N_{Src}/N_{Drn}[cm^{-3}]$	10^{20}
$B_{path}[Vcm^{-1}]$	$3.3 \cdot 10^{7}$	$t_{ch}[nm]$	10
$L_g[nm]$	22	$t_{ox}[nm]$	3
$W_{ch}[\mu m]$	1	ϵ_{ox}	21
$Type\ (n\,or\,p)$	1 or -1	Gate work function $[V]$	4.1

3.5.1 Single stage inverter

In digital logic, an inverter or NOT gate is a logic gate which implements logical negation. The truth table is shown below. An inverter circuit outputs a voltage representing the opposite logic-level to its input. Inverters can be constructed using a single NMOS transistor or a single PMOS transistor coupled with a resistor. Alternatively, inverters can be constructed using two complementary transistors in a CMOS configuration as the circuit simulated here. This configuration greatly reduces power consumption since one of the transistors is always off in both logic states. Processing speed can also be improved due to the relatively low resistance compared to the NMOS-only or PMOS-only type devices.

Table 3.3 – Truth table for an inverter

INPUT	OUTPUT
A	NOT A
0	1
1	0

First experimental demonstration of a TFET based inverter was shown by [71]. Here, a TFET

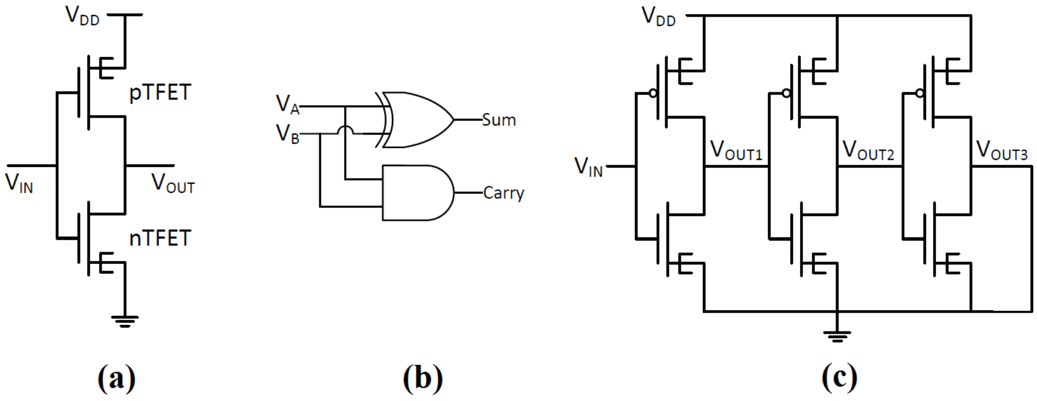


Figure 3.24 – Schematics of circuits simulated (a) single stage inverter (b) 2-bit half adder (c) A ring oscillator (3 stage shown here)

based single stage inverter is simulated with different operation voltages (V_{DD}). The voltage transfer characteristic (VTC) is plotted in Fig. 3.25(a) and Fig. 3.25(b) shows the voltage gain for V_{DD}=0.1 to 1V. A relatively high gain of 5 is observed for V_{DD} as low as 0.2V. However, as shown in Fig. 3.25(c), the inverter output during transient simulations shows large overshoots and undershoots with V_{DD}=1V and no load capacitance. This is due to the enhanced Miller capacitance originating from the 100/0 charge partitioning scheme of TFETs (as explained in section 2.5) [41]. The over/under shoot is reduced by adding a load capacitance $C_L = 2fF$, however at the cost of an increased propagation delay from 12ns to 18ns (Fig. 3.25(c)).

3.5.2 2-bit half adder

The half adder adds two single binary digits A and B. It has two outputs, sum (S) and carry (C). The carry signal represents an overflow into the next digit of a multi-digit addition. The value of the sum is (2C + S). The simplest half-adder design, pictured in the inset of Fig. 3.26(a), incorporates an XOR gate for S and an AND gate for C. With the addition of an OR gate to combine their carry outputs, two half adders can be combined to make a full adder. The truth table for a 2-bit half adder is shown in table 3.4.

Table 3.4 – Truth table for a 2-bit half adder

A	B	Sum	Carry
0	0	0	0
0	1	1	0
1	0	1	0
1	1	0	1

As benchmarking studies [72] commonly use 32 bit adders, we simulate here a 2-bit half adder. Figure 3.26(a) show the input signals and Fig. 3.26(b) show the output sum and carry signals

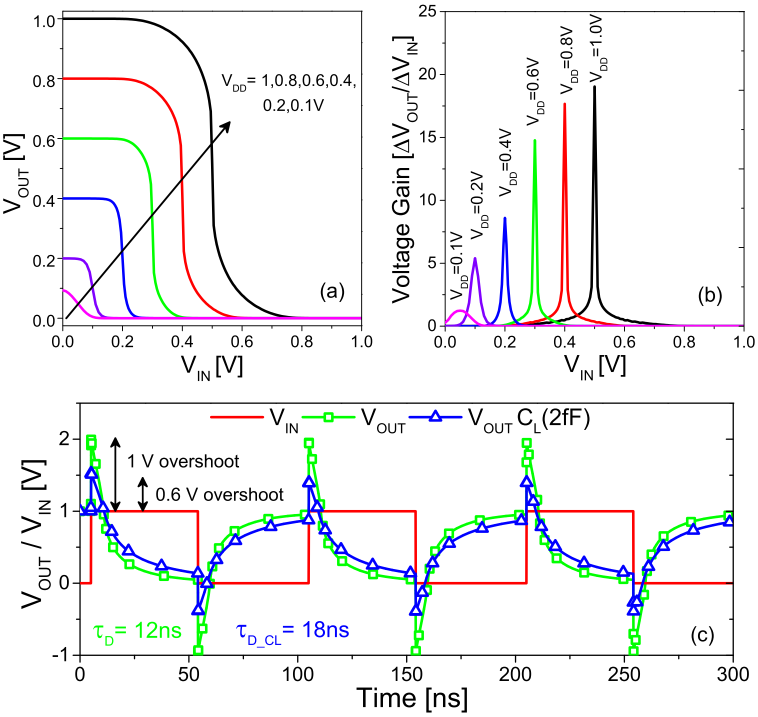


Figure 3.25 – (a) VTC of a TFET based single stage inverter with V_{DD} 0.1V to 1V. (b) Voltage gain of the inverter for the same V_{DD} range. Inset shows the simulated single stage inverter. (c) Transient response of a single stage TFET inverter at V_{DD}=1V with C_L=0fF and C_L=2fF.

of a 2-bit half adder comprising a XOR and an AND gate. As we can see the TFET based circuit faithfully reproduces the truth table.

3.5.3 15 stage ring oscillator

A ring oscillator is a device composed of an odd number of NOT gates whose output oscillates between two voltage levels, representing true and false. The NOT gates, or inverters, are attached in a chain; the output of the last inverter is fed back into the first. The ring oscillator uses an odd number of inverters to give the effect of a single inverting amplifier with a gain of greater than one. Rather than having a single delay element, each inverter contributes to the delay of the signal around the ring of inverters, hence the name ring oscillator. Adding pairs of inverters to the ring increases the total delay and thereby decreases the oscillator frequency. Changing the supply voltage changes the delay through each inverter, with higher voltages typically decreasing the delay and increasing the oscillator frequency.

A TFET based 15 stage ring oscillator is simulated for a V_{DD}=1V. The typical oscillation behaviour of the outputs of first five odd numbered stages are shown in Fig. 3.27(a). The high over and under shoots are present as expected for a typical TFET. An oscillation period

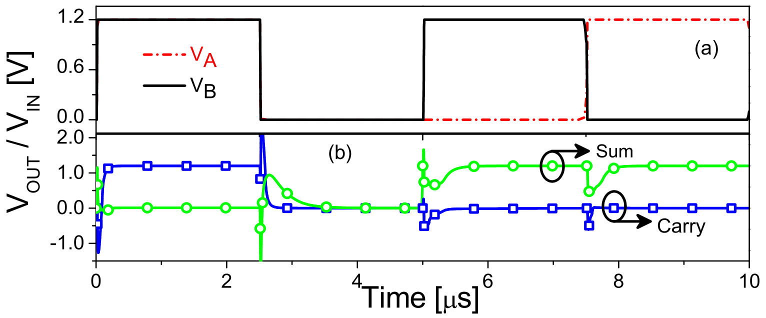


Figure 3.26 – All-Silicon 22nm TFET based 2-bit half adder with V_{DD}=1.2V (a) the two input signals V_A and V_B (inset shows the schematic of the simulated 2-bit half adder) (b) the sum and carry output signals.

of 2.38μs is observed which gives a delay of 79.4ns ($\tau_{delay} = period/2N$),where $N = 15$. The leakage, dynamic and the total energy are evaluated for this example according to eqn. 3.23. Figure 3.27(b) shows the leakage, dynamic and the total energy [5, 73] with V_{DD} scaling. At 1% activity factor (typical for low power applications), we can see that the leakage energy dominates for lower V_{DD}. Capacitance per stage of 2.5fF is used for this calculations. The V_{DD} at the point of minimum total energy, $V_{DD_{min}}$ is seen at a relatively low value of 0.3V. This value of $V_{DD_{min}}$ for a TFET based circuit is also expected to be lower than a comparable CMOS based solution.

$$\begin{aligned}
&E_{dynamic} = \alpha C V_{DD}^2\\
&E_{leakage} = I_{off} V_{DD} \tau_{delay}\\
&E_{total} = E_{dynamic} + E_{leakage}
\end{aligned} \tag{3.23}$$

Figure 3.27(c) shows the switching energy-performance comparison [5, 74]. The operation frequency in this plot is defined as $f = \frac{1}{\tau_{delay}}$. A 10$MHz$ operation frequency is observed at $V_{DD} = 1V$, which is much lower than what is seen in CMOS based equivalent circuit block. This is expected due to firstly the lower current levels in TFET and secondly the TFET used for this study was not optimized for any benchmarking. However, at lower performance range, TFETs are expected to be more energy efficient than comparable CMOS technology [74]. In this section, we do not perform any benchmarking with CMOS technology as the TFET used in the simulations is not optimized for a given technology node. A benchmarking study has been carried out in the next section.

3.6 Model calibration and benchmarking study

This section reports experimental calibration of the model described above on long channel (350nm) complementary n- and p-type sSi NW TFETs, the model is used to systematically benchmark the main analog figures of merit at device level: g_m/I_d, g_m/g_{ds}, f_T and

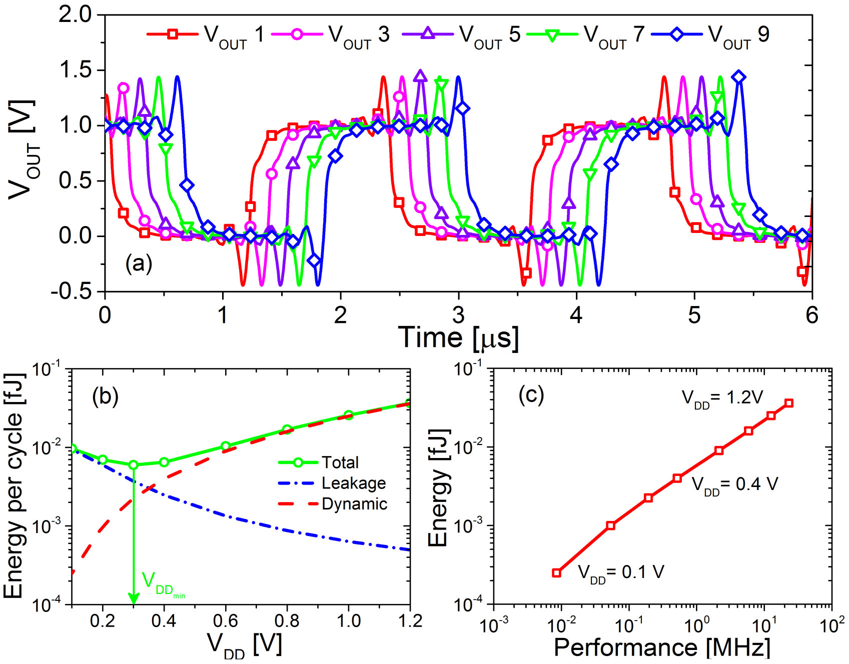


Figure 3.27 – (a) All-Silicon 22nm TFET based 15 stage ring oscillator output for V_{DD}=1V. A delay of τ_D=79.4ns is observed. (b) Leakage, dynamic and total energy with V_{DD} scaling (c) Switching energy-performance comparison.

$f_T/(I_d V_d)$ and their temperature dependence from 25°C to 125°C, allowing a direct comparison between 28nm Low Power (LP) FD-SOI CMOS node and 28nm DG-TFET. We demonstrate unique advantages of sSi DG TFET over CMOS, in terms of:

- Reduced temperature dependence of swing, SS, and (maximum) transconductance, g_{max},
- Higher transconductance per unit of current with peaks close to $40V^{-1}$, for currents lower than $10nA/\mu m$.
- Higher unity gain frequency per unit power for currents below $10nA/\mu m$ and
- Higher temperature stability of the analog characteristics.

Benchmarking at circuit level is shown, focusing on three main basic analog circuit cells: (a) current mirrors, (b) differential pairs, and, (c) diode-connected TFET circuits. The low power design space, analog performance and temperature dependence of these circuits is significantly improved by the TFETs. In [75], Trivedi et. al were the first to report on how the bandwidth-power trade-off in ubiquitous sensors with an ultra-low energy demand can be limited by excessive leakage and limited SS in MOSFET. Steep sub-threshold devices with ultra low off-current and higher g_m/I_{DS} are promising for such applications.

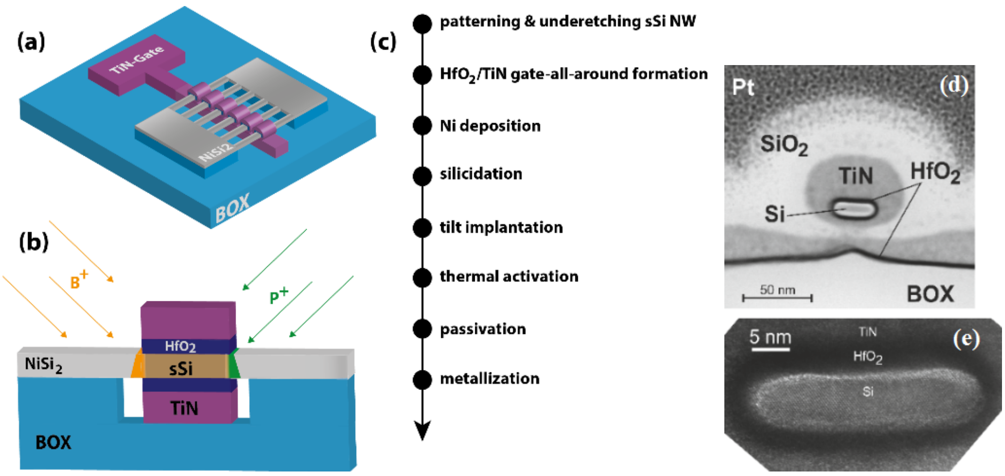

(a)
TiN-Gate
NiSi2
BOX
(b)
B+
P+
HfO2
NiSi2
sSi
TiN
BOX
(c)
patterning & underetching sSi NW
HfO2/TiN gate-all-around formation
Ni deposition
silicidation
tilt implantation
thermal activation
passivation
metallization
Pt
(d)
SiO2
TiN
HfO2
Si
50 nm
BOX
5 nm
TiN
HfO2
Si
(e)

Figure 3.28 – (a) Schematic of the fabricated sSi NW TFET with TiN/HfO_2 gate stack (b) Highly doped $n+$ and $p+$ pockets at the silicide edges are formed after a low temperature anneal (c) sSi NW TFET fabrication process using tilted $B+$ and $P-$ ion implants into epitaxial $NiSi_2$ S/D contacts. (d) SEM image of single sSi NW TFET after fabrication. (e) SEM cross section along the NW showing the gate oxide and metal layers. [43]

3.6.1 Description of fabricated devices

Strained Si (sSi) NW array TFETs as shown in the schematic of Fig.3.28(a) were fabricated with an top-down approach resulting in a nanowire cross section of 30x5nm^2 Fig.3.28(e) and a gate length of 350nm [43, 76]. A rough overview of the process steps is shown in Fig.3.28(c). The gate stack applied around the NW consists of 3nm HfO_2 as gate dielectric and 40nm TiN as gate metal (Fig.3.28(d)). $NiSi_2$ was formed at the entire Si area which were not covered by the gate. Depending on the type of TFET, the source/drain (S/D) junctions were formed by tilted $P-$ and $B+$ implantations into the silicide with subsequent out diffusion annealing at 500°C for 10s to form shallow doping pockets by dopant segregation right at the channel interface (Fig.3.28(b)).

3.6.2 28nm FDSOI CMOS design kit

A 28nm planar ultra thin body and box FD-SOI design kit from ST Microelectronics was used for this benchmarking study. A false colour cross section SEM image of the device is shown in Fig. 3.29. As shown in the figure the devices have a ultra thin Silicon body of 7nm thickness. Dual V_t configurations were available in the package namely LVT (low V_t) and HVT (high V_t). For this study the LVT flavour was used. A comparison of the transfer curves of this design kit with that of the measured TFET devices in shown in Fig. 3.32.

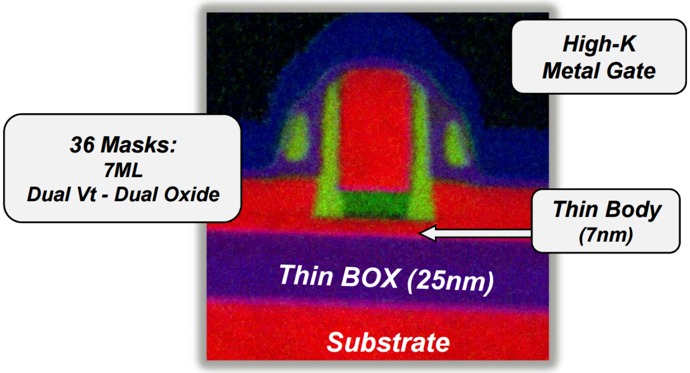


Figure 3.29 – SEM cross section along the NW showing the gate oxide and metal layers. [image courtesy: ST-Microelectronics]

3.6.3 Calibration with experimental results

The model described in the previous section has been extended with temperature dependence, SRH leakage current and then calibrated on the experimental sSi n- and p-type TFETs (Figs. 3.30(a) and 3.30(b)) and implemented in Verilog-A. The temperature dependence in the model primarily comes from the bandgap dependence of Silicon to temperature. The energy band-gap further influences the A_{path} and B_{path} parameters as shown below:

$$\begin{aligned} E_g(T) &= 1.166 - \frac{4.73X10^{-4}T^2}{T+636} \\ A_{path} &= \frac{g\pi {m_r}^{1/2}(qF_0)^2}{9h^2[E_g(300K)+\Delta_C]^{1/2}} \\ B_{path} &= \frac{\pi^2 {m_r}^{1/2}[E_g(300K)+\Delta_C]^{3/2}}{qh} \end{aligned} \tag{3.24}$$

In addition the electrostatic potential solution and the approximate Fermi integral used in the model also changes with the temperature. Combining all the above dependencies the model can predict temperature dependence of band-to-band tunneling with relative accuracy as shown in Fig. 3.31(a) for both n & p-type devices measured.

TFET transfer and output characteristics were measured from 25°C to 125°C and the model accordingly calibrated with very good accuracy. We show that for both p- and n-type TFET the sub-threshold swing (in the regions where BTBT dominates), the on-current, I_{ON}, and, the transconductance, g_m, have less dependence on temperature than CMOS (Fig.3.31(a) and (b)). Particularly remarkable is the reduced temperature dependence of g_{max} (Fig.3.31(b)) and of threshold voltage, V_T (Fig.3.32(a)) for sSi TFETs. The average SS was calculated over three decades of current below I_{VT} level shown in Fig. 3.31(a). The calibrated model does not take into consideration the trap-assisted tunneling (TAT), a fabrication artifact which partially affects the swing and its dependence on temperature due to defects at the interface HfO_2/Si

and at the junctions (following tilted ion implantation and limited temperature annealing). The model includes temperature dependent SRH leakage.

Table 3.5 – Table of fitted parameter list for Verilog-A based model for both n & p type device at room temperature.

Model Parameter	n-type	p-type
$A_{path}[cm^{-3}s^{-1}]$	7.2x10^{15}	6x10^{7}
$B_{path}[V/cm]$	8.4x10^{7}	9.3x10^{7}

3.6.4 Device level analog benchmarking

The calibrated model is scaled down and used to benchmark the analog figures of merit of 28nm complementary sSi TFET against a 28nm Low Power CMOS node; the transfer characteristics are depicted in Fig. 3.32(b), also shows higher temperature stability for TFET.

g_m/I_D and g_m/g_{ds}

The g_m/I_D ratio is a measure of the efficiency to translate current (hence power) into transconductance; i.e., the greater the g_m/I_D value, the greater the transconductance we obtain at a constant current value. Therefore, the g_m/I_D ratio is sometimes interpreted as a measure of the "transconductance generation efficiency". The choice of g_m/I_D is considered as a fundamental analog design tool and it is based on its relevance for the three following reasons:

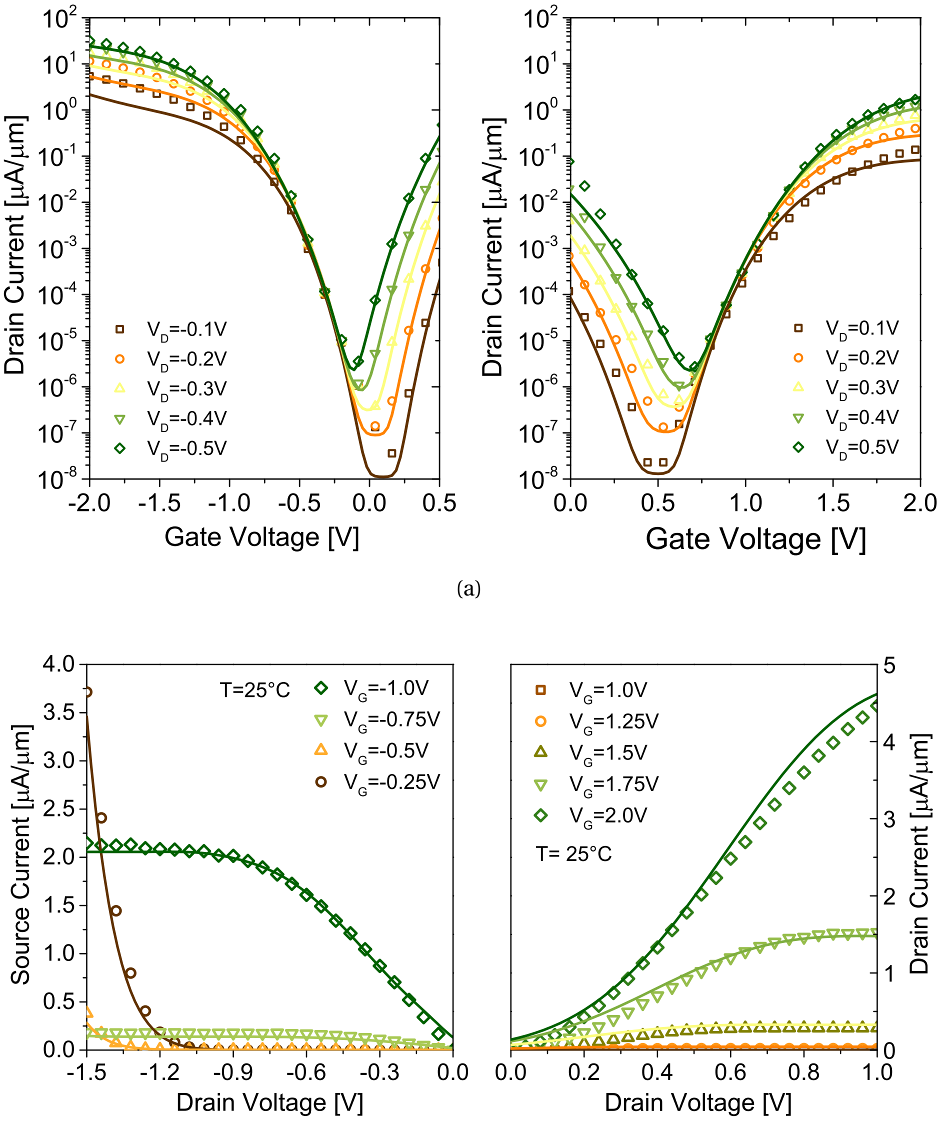
- It is strongly related to the performances of analog.
- It gives an indication of the device operating region.
- It provides a tool for calculating the transistors dimensions [77].

Another analog figure of merit, the intrinsic gain of MOS transistor is defined as:

$$A_{intrinsic} = \frac{g_m}{g_{ds}} \tag{3.25}$$

where g_m and g_{ds} are the transconductance and output conductance respectively. The intrinsic gain is the maximum gain a transistor can reach. It is the amplification ability of a MOS transistor. For analog circuits, the gain is the most important specification for an active device. From eqn.3.25, it can been seen that due to the channel length modulation effect, the output impedance is decreased with the channel length reduction. The intrinsic gain of a transistor is also reduced. To avoid the reduction of the transistor gain, the longer transistor channel length is normally used in amplifiers.

In Fig. 3.33, we show, transconductance per unit current, g_m/I_D (a), and DC gain, g_m/g_{ds} (b), respectively, as key analog figure of merit of n- and p-TFET versus CMOS. The

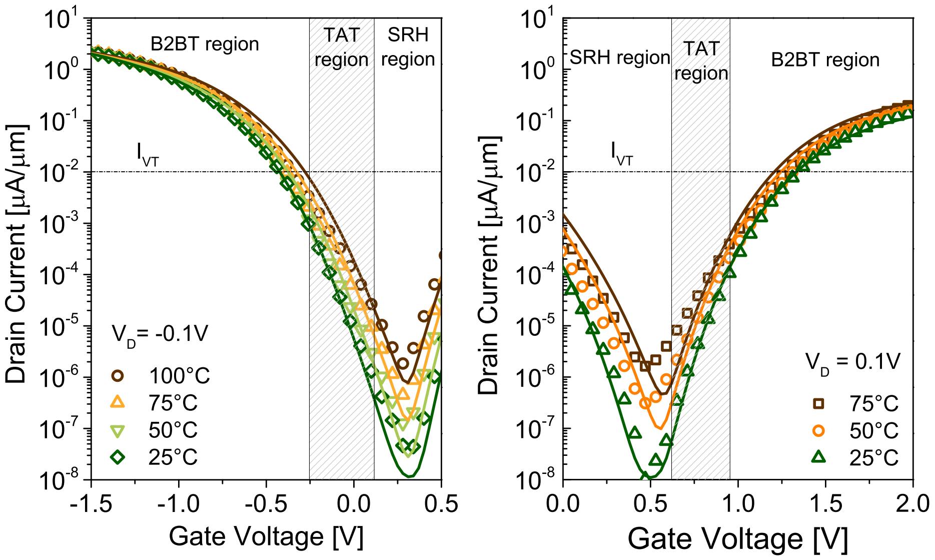


(a)

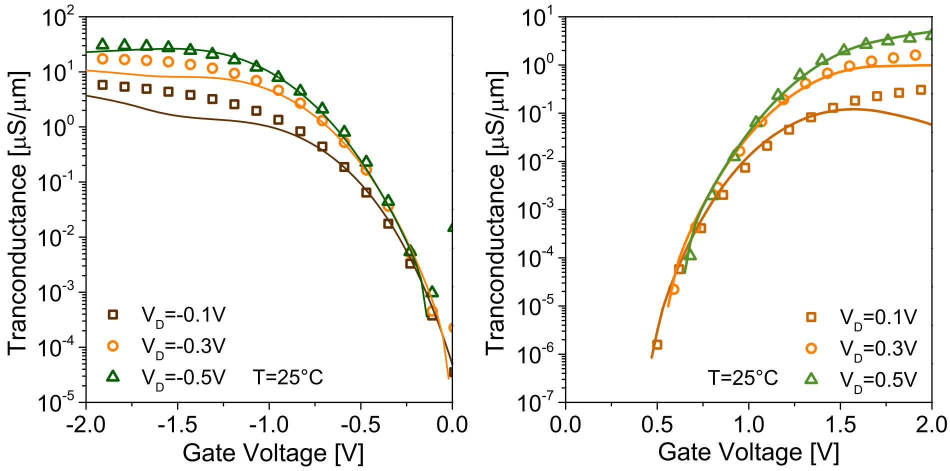
(b)

Figure 3.30 – (a) Measured transfer characteristics and calibrated model at different V_{DS} of the fabricated sSi NW TFET for (left) p-type and (right) n-type TFET (b) Measured output characteristics and calibrated model at different V_{GS} for (left) p-type (right) n-type TFET at room temperature. (Symbol: measurement, lines: model)

TFET is capable to achieve much higher g_m/I_d at very low currents ($< 10nA/\mu m$), close to the theoretical KT/q limit of $39V^{-1}$ [78] for the p-type device and has a higher g_m/g_{ds} in the $V_{GS} = 0-1V$ range, when BTBT dominates ($V_{DS} > 0.1V$). The intrinsic gain is plotted in Fig.3.33(b). The plots for TFET are generated by the calibrated model. Both n and p type TFET shows higher intrinsic gain compared to CMOS.



(a)

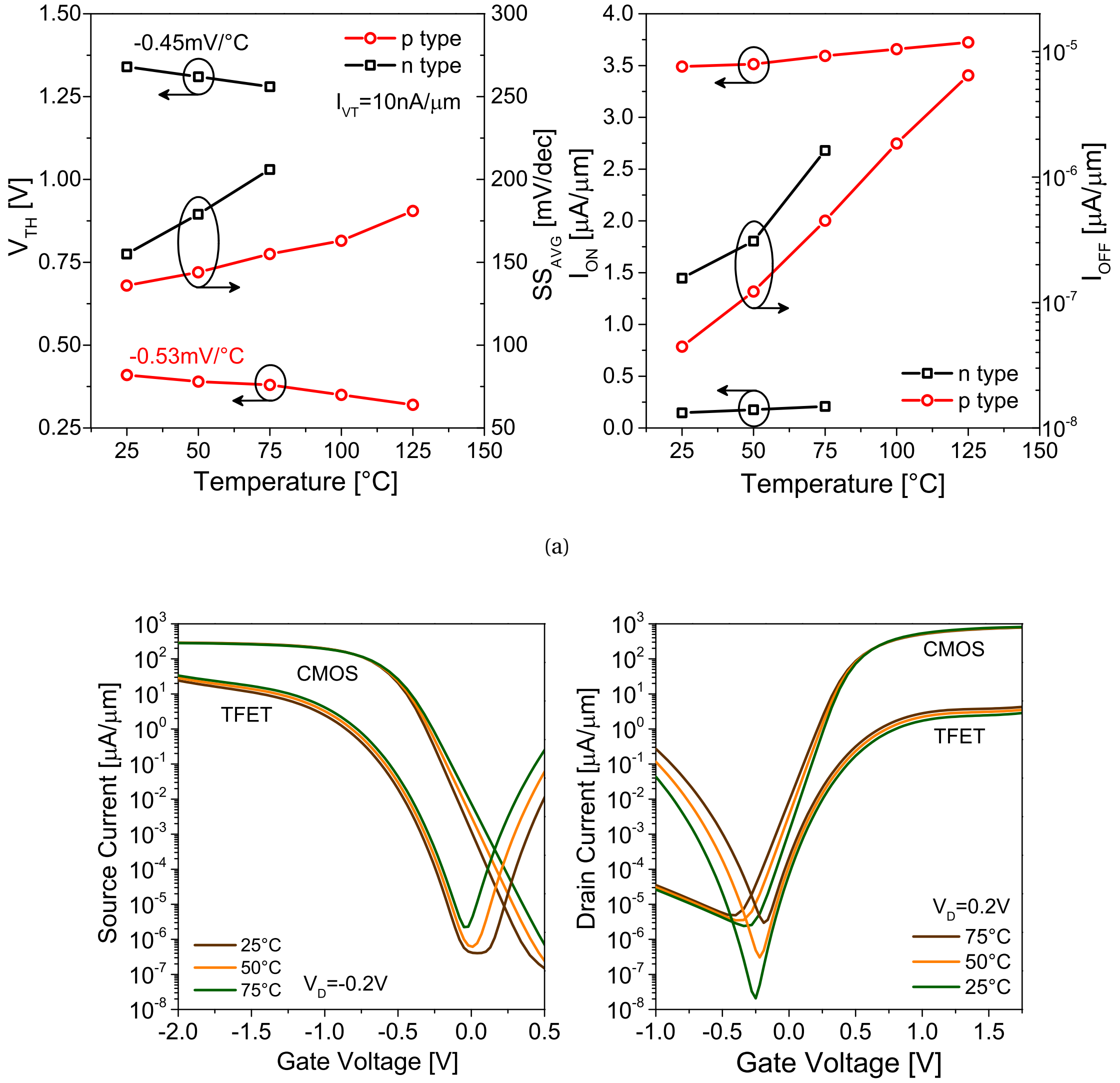


(b)

Figure 3.31 – (a) Measured transfer characteristics with temperature (25°C to 100°C) with calibrated model; achieved I_{ON} is $20\mu A/\mu m$ and $2\mu A/\mu m$ for p- and n-type sSi homojunction NW TFET at $V_{GS} = 1.5V$ and min point swings, $SS_{pt} \approx$ 70-100mV/dec at room temperature and $I_{OFF} < 1-10nA/\mu m$. (b)Measured transconductance and calibrated model at $|V_{DS}| = 0.1V$ with respect to temperature for p type (left) and n type (right) TFET. (Symbol: measurement, lines: model)

F_T and $F_T/(I_D V_D)$

A popular MOS transistor figure of merit is its unit gain frequency, which is defined as the signal frequency at which the magnitude of the short circuit current gain, I_{out}/I_{in}, is unity. A second figure of merit compares the unity gain frequency f_T per unit power vs.



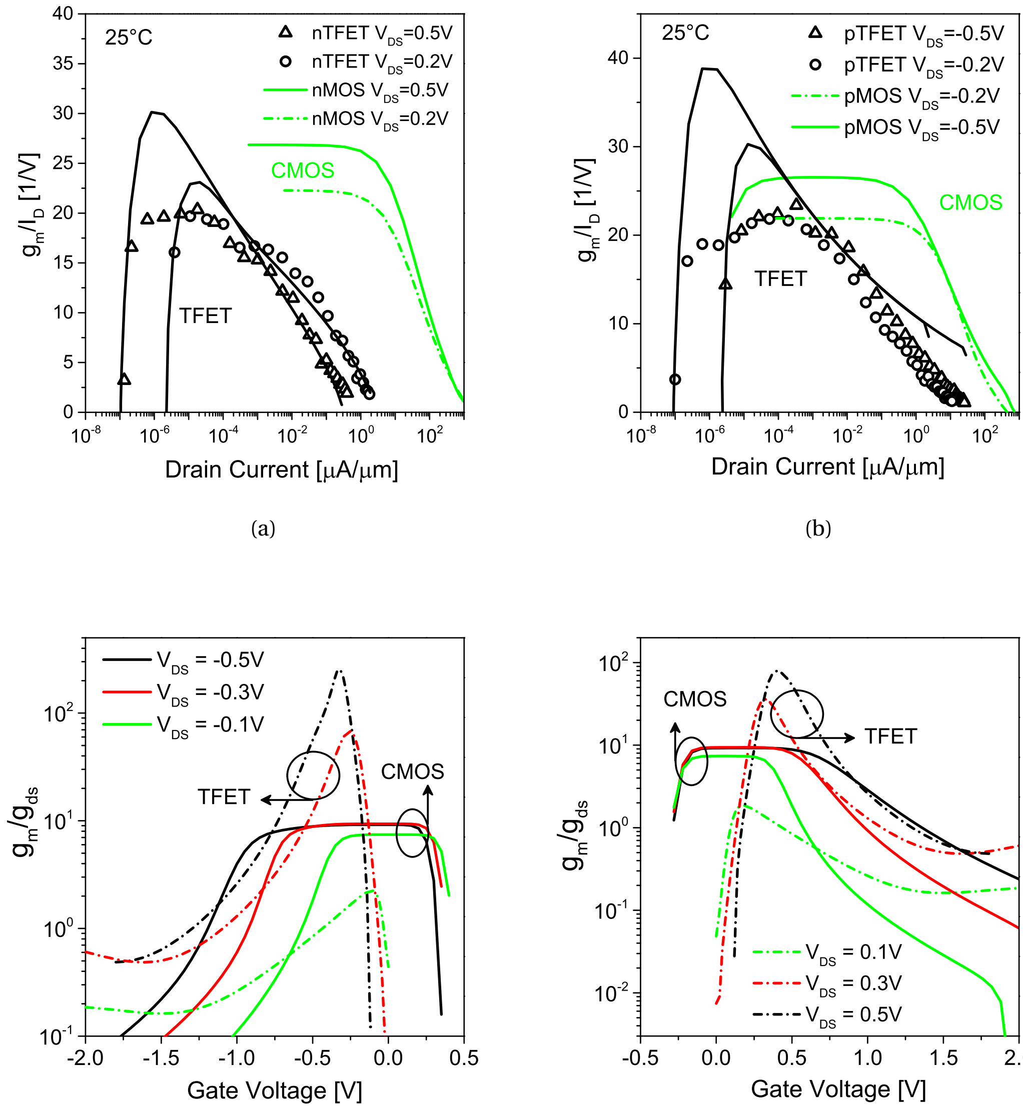
(a)

(b)

Figure 3.32 – (a)(Left) experimentally extracted V_{TH} and average sub-threshold swing, SS_{AVG}, (over 4 decades of current) dependences, and, (right) I_{ON} and I_{OFF} dependences on temperature, from 25°C to 125°C. (b)Transfer characteristics of 28nm CMOS (calibrated simulation) and TFET (calibrated model) overlapped with respect to temperature.

current/micron. This metric highlights the low power limits of CMOS and again shows the performance benefit of the TFET.

$$\begin{aligned} \omega_T &= 2\pi f_T \\ f_T &= \frac{g_m}{2\pi C_{gs}} \end{aligned} \tag{3.26}$$

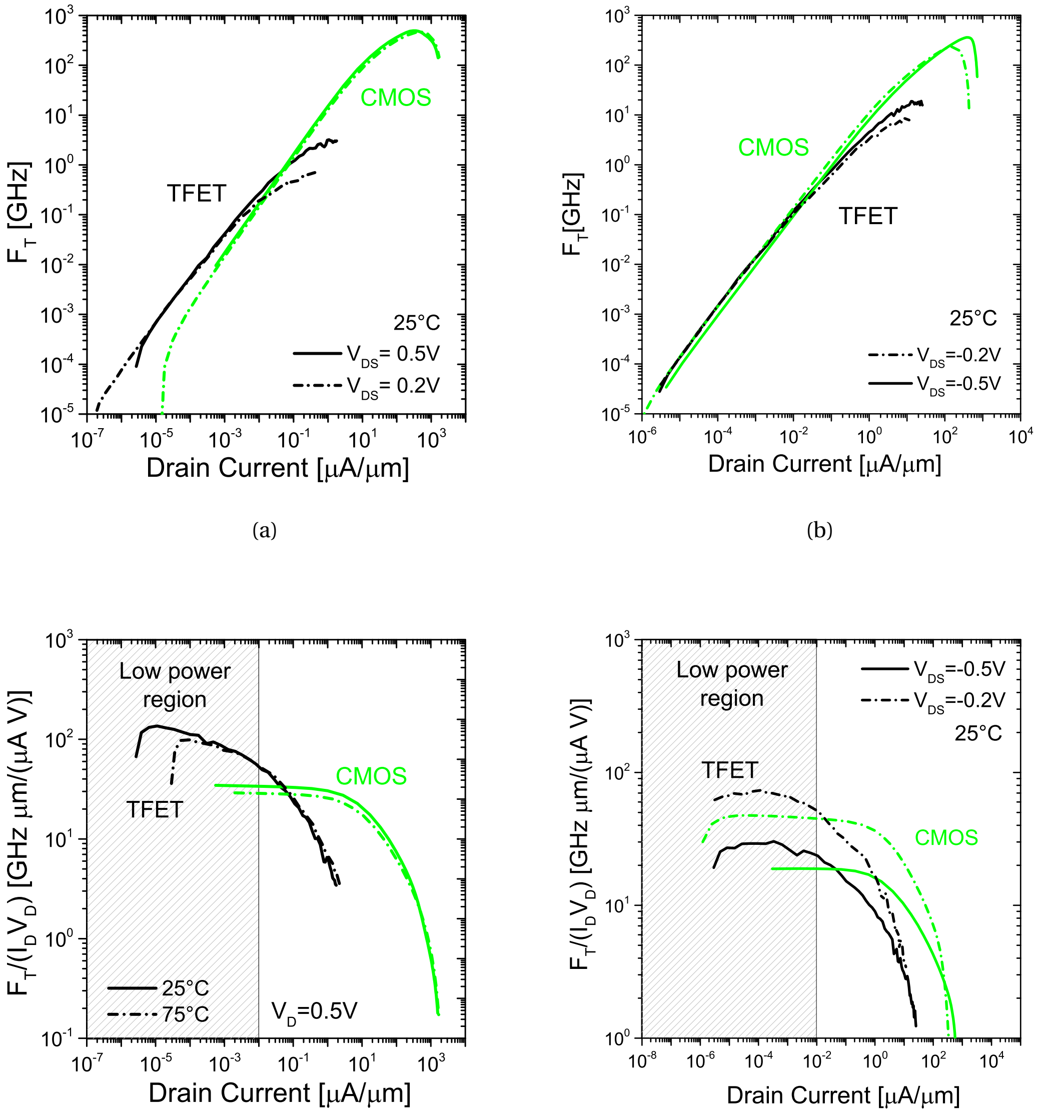


(a) (b)

(c) (d)

Figure 3.33 – (a)Transconductance per unit drain current for p type (left) and n type (right) with respect to VDS. (b)Intrinsic gain vs. gate voltage for p type (left) and n type (right). The calibrated model is used for the TFET plots.

where g_m is the transconductance and C_{gs} is the gate-source capacitance of the MOS transistor. The current gain cut-off frequency and unity gain frequency per unit power versus current of Figs. 3.34(a) and (b), shows the significant performance improvements at low levels of currents (below $10nA/\mu m$) offered by sSi TFET and especially the new analog design space that it offers. Moreover, the unity gain frequency per unit power of TFET has little temperature dependence and extends high values of $F_t/(I_d V_d)$ down to the pA region where CMOS is not capable to reach as shown in Fig.3.34(c) and (d).



(a) (b)

(c) (d)

Figure 3.34 – (a)Measured (TFET) current gain cut-off frequency Vs current per unit width for p type (left) and n type (right) TFET (b)Measured (TFET) unity gain frequency per unit power vs. current per unit width for p type (left) and n type (right) TFET.

3.6.5 Circuit level analog benchmarking

The benchmarking analog TFET IC cells: current mirror circuit, differential pair circuit and diode connected transistor, are depicted in Fig. 3.35.

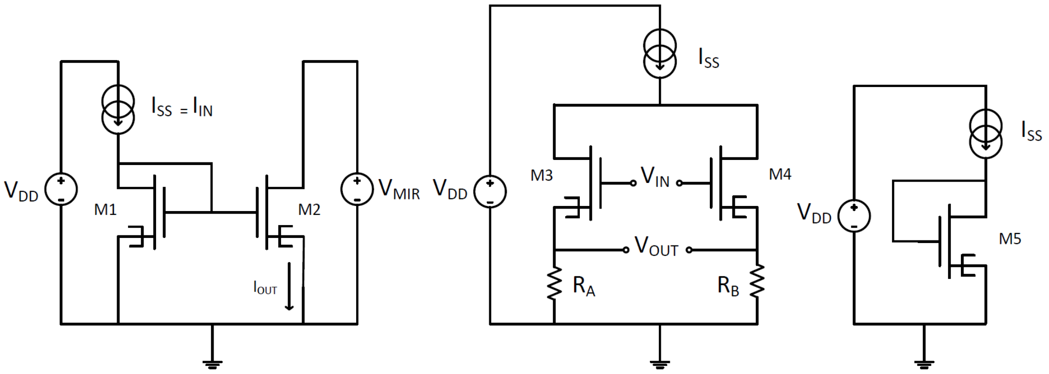


Figure 3.35 – Schematics of investigated analog cells: (a) current mirror circuit (b) differential pair circuit and (c) a diode connected transistor circuit.

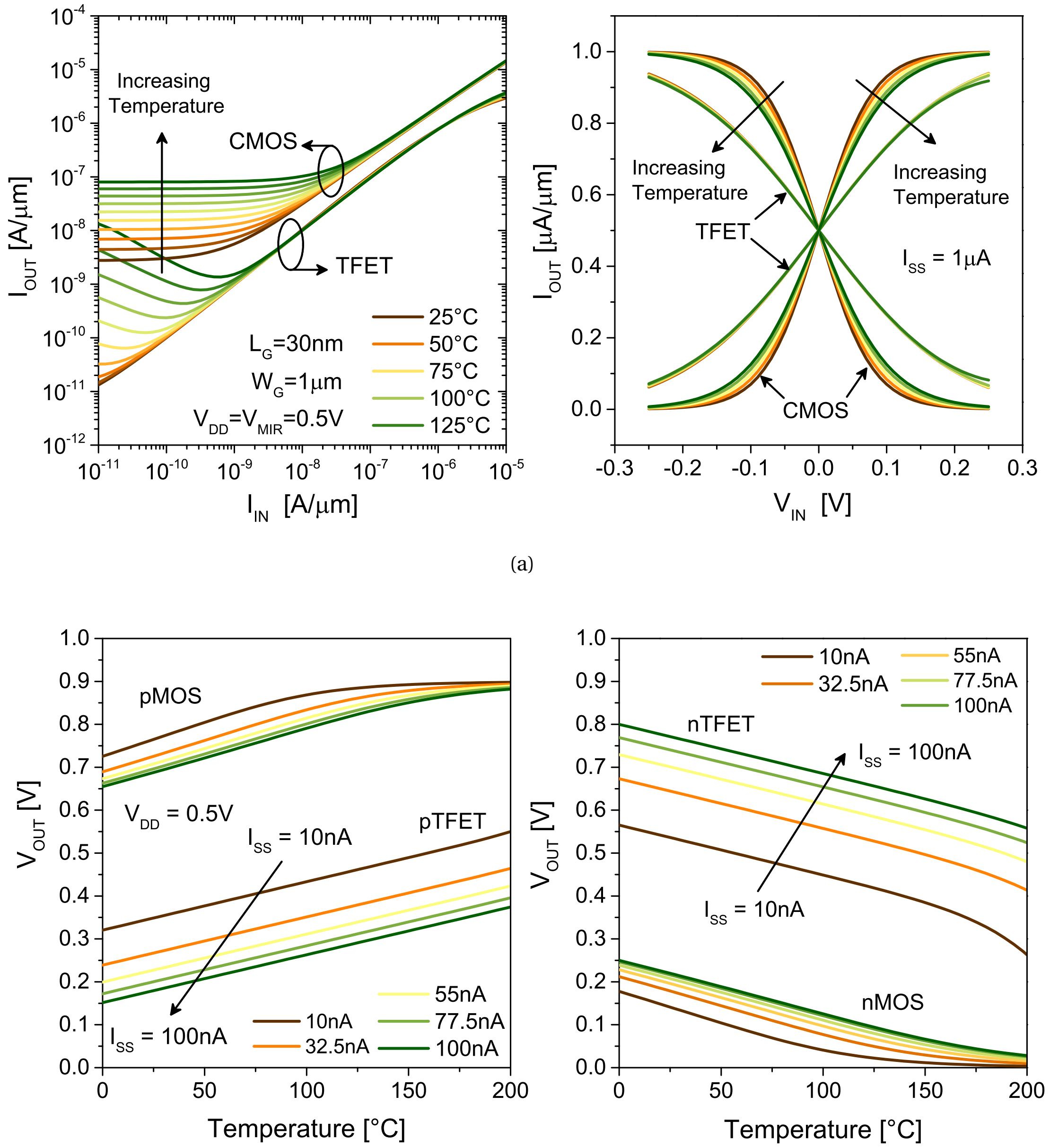
Current mirror

A current mirror is a circuit designed to copy a current through one active device by controlling the current in another active device of a circuit [79], keeping the output current constant regardless of loading. The current being 'copied' can be, and sometimes is, a varying signal current. Basically, an ideal current mirror is simply an ideal inverting current amplifier that reverses the current direction as well or it is a current-controlled current source (CCCS). The current mirror is used to provide bias currents and active loads to circuits.

The main specification that characterizes a current mirror is the transfer ratio (in the case of a current amplifier) or the output current magnitude (in the case of a constant current source CCS) [80]. There are also a number of secondary performance gauges with current mirrors, for example, temperature stability is also critical for certain applications. In this study case we show the transfer ratio between the output and input current and particularly the temperature stability as we have seen from the measurements in Fig.3.31, that TFETs are less temperature sensitive. Typical current mirror $I_{OUT} - I_{IN}$ plots are shown in Fig. 3.36(a)(left). The temperature stability is clearly evident from this figure. Also it is interesting to note that, TFET current mirrors can be operated at much lower current levels compared to CMOS in all temperature range.

Differential pair

A differential pair is two complementary transistors that transfer equal and opposite signals down their length [79], as shown in Fig.3.35(b). The name "differential" comes from its signalling scheme which is a method for electrically transmitting information using two complementary signals [81]. A differential pair is simulated to show the output current vs V_{IN} and V_{OUT} vs V_{IN} plots in Fig.3.36(a)(right). The current flowing through R_A and R_B and the voltage across them are symmetric about the zero volt V_{IN} as expected. Once again the, the temperature stability of TFET is evident as expected, even though the gain is lower than



(a)

(b)

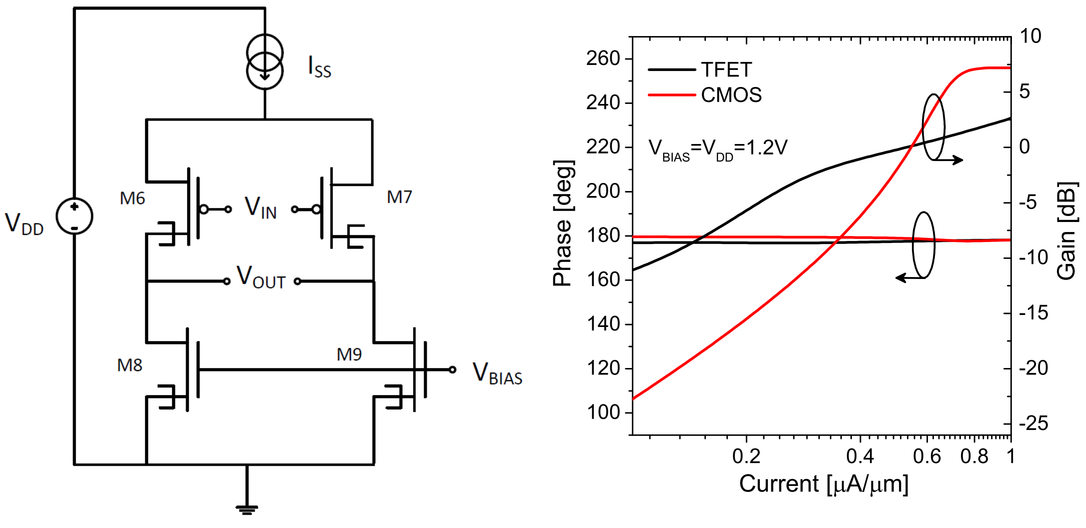
Figure 3.36 – (a)TFET vs. CMOS current mirror: I_{OUT}/I_{IN} for temperature 25°C to 125°C (left) Differential pair output current as a function of temperature (25°C to 125°C) for both nTFET and nMOS (right) (b) Temperature dependence of V_{OUT} with I_{SS} as parameter; CMOS vs TFET for a diode connected circuit. TFET embodiment shows excellent linearity versus temperature, suggesting possible use a highly linear temperature sensor.

CMOS.

Diode connected transistor

A diode-connected transistor is a method of creating a two-terminal rectifying device (a diode) out of a three-terminal transistor as shown in Fig.3.35(c). A characteristic of diode-connected transistors is that they are always in the saturation region for MOSFETs. Diode-connected transistors are used in current mirrors to provide a voltage drop that tracks that of the other transistor as temperature changes. They also have very low reverse leakage currents.

In Fig.3.36(b), comparison of pTFET(left) and nTFET(right) with CMOS shows that TFET shows very high linearity from 0 to 200°C. With CMOS, higher bias current is applied to improve the linearity at high temperature. However this comes as the expense of sensitivity. On the other hand, TFET sensitivity (i.e. $dV_D/dTemp$) remains same with increasing in bias current, thus sensitivity of temperature measurement is not compromised. In addition, TFET device require significantly lower bias current to achieve the linearity, thus improved "linearity" per bias current. Due to this improved "linearity", the TFET based diode connected transistor could potentially be used as a very good temperature sensor. Diode connected transistors have previously found applications in sensing and voltage regulators etc. [82, 83].



(a) (b)

Figure 3.37 – (left) Schematic of TFET differential stage circuit with an active load and (Right) gain (in dB) versus current, for CMOS and TFET and (right) showing the phase and the gain versus input current. The gain drops by 30 dB for the CMOS at room temperature.

Differential stage

Differential stage is basically a differential pair (as described above) with an active load as shown in Fig.3.37(a). Fig.3.37(b) reports a differential stage with TFET and FET L/W fixed at $30nm/1\mu m$ (M6 and M7 forms differential pair). We use an ideal constant current source $I_{SS} = 1uA$. The load transistor (M8 and M9) are biased in saturation region to ensure that maximum gain is achieved. Fig.3.37(b) shows that as bias current is reduced, the gain roll-off for TFET is significantly less compared to CMOS, which can be translated into improved energy efficiency.

3.7 Summary

Based on conformal mapping method, we have developed a simple physics-based DC-current model for DG-TFETs taking into account the non-local B2B tunnelling and in-series inversion MOS region. The limitations of the model for low source doping have been successfully addressed. The model agrees well with simulations for different device parameters and addresses the present accuracy challenges of Tunnel FET modelling in all regions of operation.

We have developed a physics-based compact model for DG-TFETs describing all tunnelling related characteristics and implemented it in Verilog-A for circuit simulations. The model agrees well with TCAD simulations in all regions of operation.

Based on capacitance measurements a first order capacitance model for DG-TFETs has been proposed. The model matches numerical simulations well and shows the same behaviour as measurements. Using the model to perform basic circuit simulations has revealed some unique features of a TFET.

Finally we have shown a systematic benchmarking study of analog figures of merit of 28nm sSi TFETs versus 28nm CMOS based on a universal compact model at both device and circuit levels. Our study suggests new design and performance space, especially due to the higher TFET gain at very low current levels ($1pA/\mu m$ to $10nA/\mu m$), and a higher temperature stability for key analog IC functions.

4 Tunnel FETs based Volatile Memories

In this chapter we propose and demonstrate the use of a Tunnel FET (TFET) as capacitorless DRAM cell based on TCAD simulations and experiments. We report more experimental results on Tunnel FETs implemented as a double-gate (DG) fully-depleted Silicon-On-Insulator (FD-SOI) devices. The Tunnel FET based DRAM cell has an asymmetric body and a partial overlap of the top gate (L_{G1}) with a total overlap of the back gate over the channel region (L_{G2}). A potential well is created by biasing the back gate (V_{G2}) in accumulation while the front gate (V_{G1}) is in inversion. Holes from the p+ source are injected by the forward-bias $p+i$ junction and stored in the electrically induced potential well. Programming conditions and related transients are reported and the role of temperature is investigated.

4.1 Why should TFETs be used as a memory

As device geometries shrink, the scaling of the conventional 1-Transistor/1-Capacitor (1T/1C) Dynamic RAM (DRAM) has become increasingly difficult, in particular due to the capacitor which has become harder to scale. The memory industry has achieved a lot of success in packing more and more bits per unit area in a silicon die. We are fast approaching the scaling limits for the physical capacitor element in a 1T/1C DRAM, and new materials and new device and integration approaches are required to meet the market demands.

In a Microsoft test lab [84], they connected a dual-socket, dual-core server that was shipped in 2005 (Server A) and a quad-socket, quad-core server that was shipped in 2008 (Server B) to power meters at the wall. They measured power consumption for the servers at idle and as components were removed. Components installed in each system are shown in table4.1.

With all components in place, idle power consumption was measured at 568 W on the 2005 system and 635 W on the 2008 system. Figure 4.1 and Fig. 4.2 detail the power consumption of individual components on these systems relative to overall system consumption.

In the newer server, Server B, the processors were not the largest consumers of power

Table 4.1 – Components in 2005 and 2008 Test Systems [84]

Component	Server A (shipped in 2005)	Server B (shipped in 2008)
Processors	2 dual-core processors	4 quad-core processors
Memory	32 2-GB double data rate (DDR) PC-2100	32 4-GB DDR-2 PC-5300 DIMMs
Network	1 PCI-X network adapter	2 PCIe network adapters
Disks	4 36 GB, 15k-RPM, 3.5-inch SCSI	4 72-GB, 15k-RPM, 2.5-inch SCSI

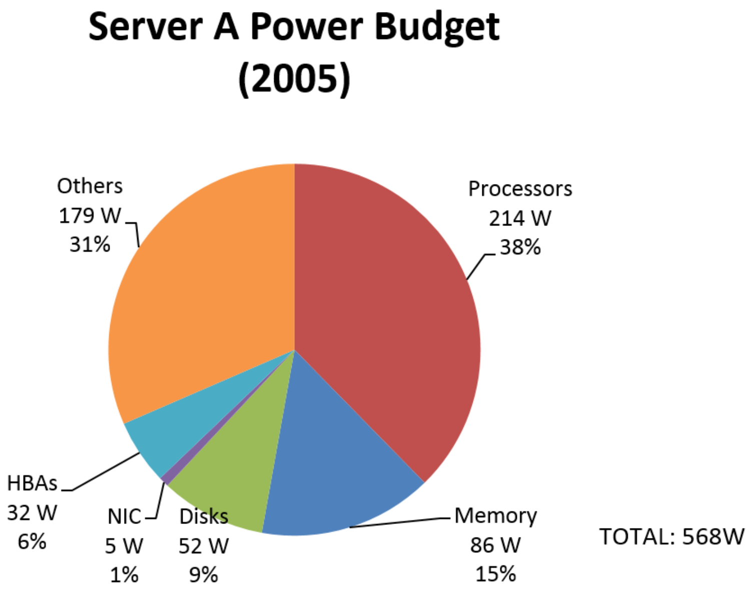


Figure 4.1 – Component power distribution, 2005 two-socket dual-core server.[84]

in the system. 32 sticks of memory can easily consume as much or more power than other system components at idle. The cause of the increase in memory power usage is explained in more detail in [84], but generally a doubling in bus speed or capacity can double the memory power consumption. Because both bus speed and capacity increased between systems, the fourfold increase in memory power consumption is understandable. Hence an energy efficient "Beyond CMOS" device like a Tunnel FET can dramatically reduce the power consumption in the future technology nodes.

An interesting property of a Tunnel FET is the fact that its sub-threshold slope is weakly dependent on temperature [85]. Tunnel FETs can thus be considered as very good candidates for operating at higher temperature, without compromising reliability. In section 4.3.4, we will see how the read-out operation of the memory cell is not degraded by temperature.

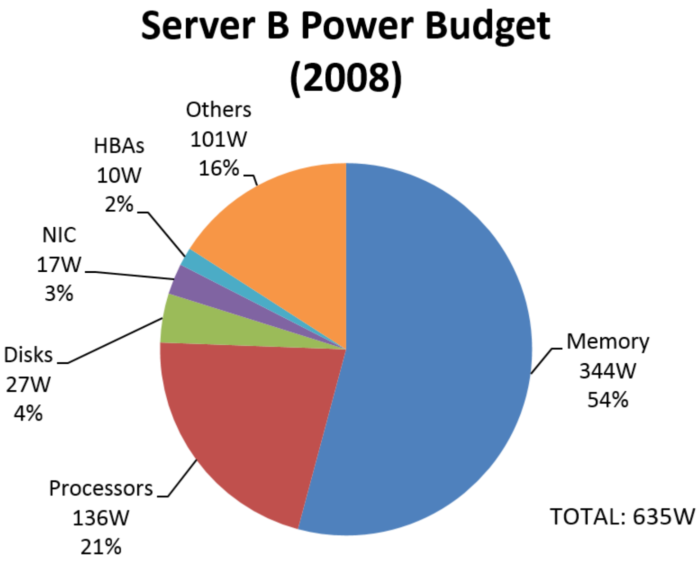


Figure 4.2 – Component power distribution, 2008 four-socket quad-core server.[84]

4.2 TFET as a volatile memory: Capacitorless memory introduction

The concept of capacitorless single-transistor (1T) DRAMs was first reported in the early 90's [86, 87], however they have attracted attention [30, 88, 89, 90] in more recent times. This is due to the ability of (1T) DRAM to achieve higher memory cell density and to solve the problems associated with the scaling of a physical capacitor. This memory cell uses the floating body of a single transistor to hold the information, i.e. to store the charge. In that way, the floating body (Fig. 4.3) of the SOI or SOI-like transistors provides naturally an ideal storage environment to achieve competitive performances within a small storage volume. In all 1T-DRAM variants, bit '1' (1-state) reflects the temporary generation of a majority carrier excess in the body which increases the potential and hence the drain current. Conversely, bit '0' (0-state) features a lower current due to the removal of majority carriers from the body. The majority carrier variation in the body can be sensed by measuring the current difference between the '0' and '1' states (Fig. 4.3).

During the 1-state programming, the amount of majority carriers (holes) in the body increases. This induces a reduction of the threshold voltage (V_{TH} -> V_{TH} - ΔV_{TH1}) and an increase in drain current I. The threshold voltage shift ΔV_{TH1} (or ΔV_{TH0}) results from the combined effects of the hole charge and the corresponding body potential variations. In the following paragraph, the body charging mechanisms are reviewed.

These memory cells can be classified in different groups according to the excess majority carrier generation method i.e. depending on how the 1-state is programmed. This extra charge can be achieved mainly by four ways [88]:

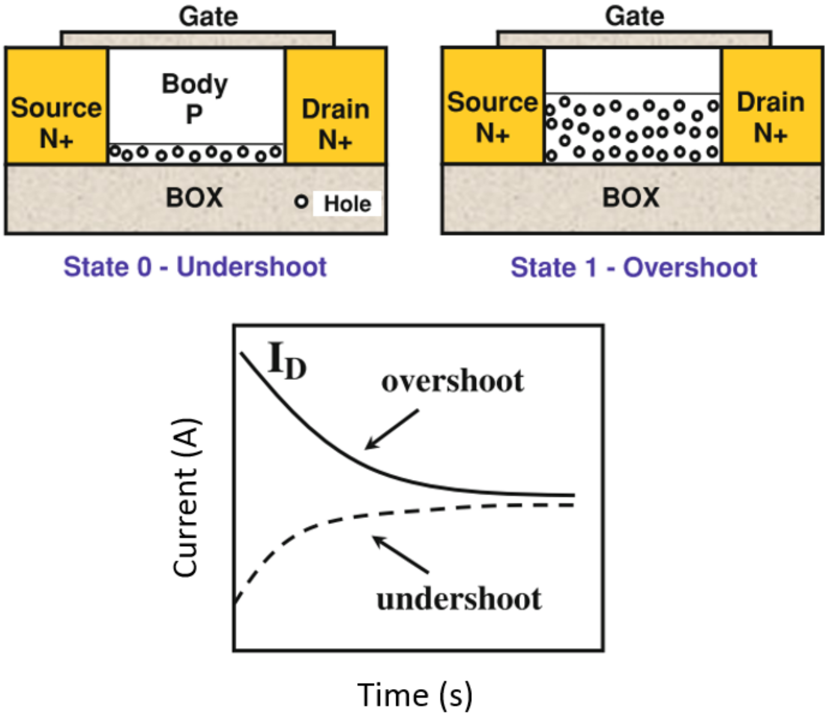


Figure 4.3 – Schematic cross-section of a partially depleted (PD) SOI nMOSFET used as a 1T-DRAM memory cell and the corresponding current transient for 0 state (undershoot) and 1 state (overshoot) resulting from the lack or excess of majority carriers (holes), respectively. Image courtesy [88]

- Impact ionization : The most frequently used method for the 1-state programming consists in generating holes inside the body is by impact ionization [91, 92, 93, 94, 95]. A relatively high positive drain voltage V_D is applied while the front interface is in inversion mode ($V_D > V_{GF} - V_{TH}|$). The holes generated at the pinch-off region, close to the drain, move and accumulate into the body.

- Bipolar junction transistor (BJT) effect: The second method for 1-state programming takes advantage of the intrinsic BJT effect which can be activated in the floating body of SOI MOSFETs [96, 97, 98]. In that case, the source (N+), body (P) and drain (N+) act as the emitter (or collector), base and collector (or emitter) of the BJT.

- Band-to-band (B2B) tunneling generation : In this method, the holes generated by B2B arise from the gate-to-drain (source) overlap region [99, 100, 101]. In order to produce the holes, the drain/oxide interface is in strong inversion; the band bending and the local electric field must be large enough to allow holes (electrons) tunneling from the conduction (valence) band into the valence (conduction) band. The holes are collected by the body and the electrons flow toward the drain contact.

- Gate tunneling current: The use of the parasitic direct tunneling current through the gate oxide in a body contacted PD SOI pMOSFET was proposed and demonstrated by Guegan et al. [89] for 1T-DRAM application. In this specific device, a N+ body contact is left floating and the polysilicon gate covering the body contact is N+ doped in order to allow a strong injection of electrons (here the majority carriers) by tunneling from the polysilicon conduction band into the body.

The (1T) DRAMs take advantage of the parasitic floating body effects in SOI transistors to store the information i.e. to hold the charge. In a nMOSFET based 1T-DRAM, the 1-state current level is triggered by the excess hole charge generated into the body by one of aforementioned techniques. However the read out operation is based on a MOSFET operation (drift/diffusion of carriers over a potential barrier) in all cases.

In Fig. 4.4 we see a comparison between the potential profiles at the surface of the channel for a DG-MOSFET and DG-Tunnel FET. The green line shows that, for the DG-MOSFET a potential well already exists which can hold the charges. The challenge in this case as discussed earlier is to generate the excess carriers. Whereas in case of a DG-Tunnel FET as shown by the red curve, there is no potential well by default due to the two different doping in the source and drain region. However we do have a reservoir of both types of carriers in source/drain region which can easily brought to the body area by forward biasing the source/channel junctions. The challenge in case of a Tunnel FET is thus to create and maintain a potential well which could hold these extra charges. In the following section we will discuss about how we can design a structure which would facilitate the formation of a potential well and still work as conventional Tunnel FET.

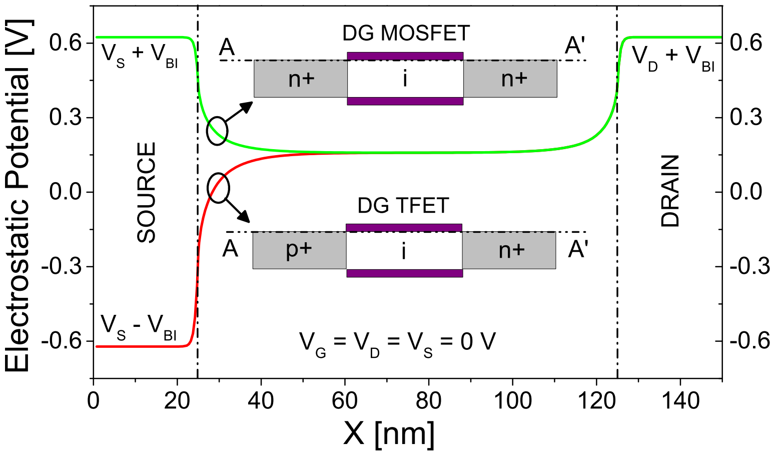


Figure 4.4 – Simulated potential profile at 1 nm below the gate-oxide/channel interface for both DG-MOSFET and DG-TFET.

4.3 FIN-Tunnel FET as 1T/0C DRAM

In this section, we demonstrate that all-Si double-gate (DG) Tunnel Field-Effect Transistor (TFET) can serve for building a class of devices: the capacitor-less Tunnel FET DRAM, where the very low I_{OFF} offers a major advantage for DRAM low power consumption/low refresh rate and the capacitor-less structure a very high potential for scalability. Also as the read-out operation will be based on band-to-band tunnelling, the read access time is not degraded by temperature due to weak temperature dependence of tunnelling current.

Figure 4.5(a) depicts a SEM cross section of the fabricated device on Fully Depleted Silicon-On-Insulator (FD-SOI). The fabrication of the TFET devices was performed on a SOI substrate with 145nm buried oxide and 20nm active Silicon layer using a MESA process [23]. The gate stack was composed of 3nm HfO_2/10nm TiN/50nm Polysilicon. A $Si3N_4$ protection layer was deposited and patterned to introduce intrinsic regions adjacent to the drain side. This reduces the ambipolar current by reducing the electric field at the drain side tunnel junction.11 Figure 4.5(b) shows a schematic view of the device with some important dimensions [85].

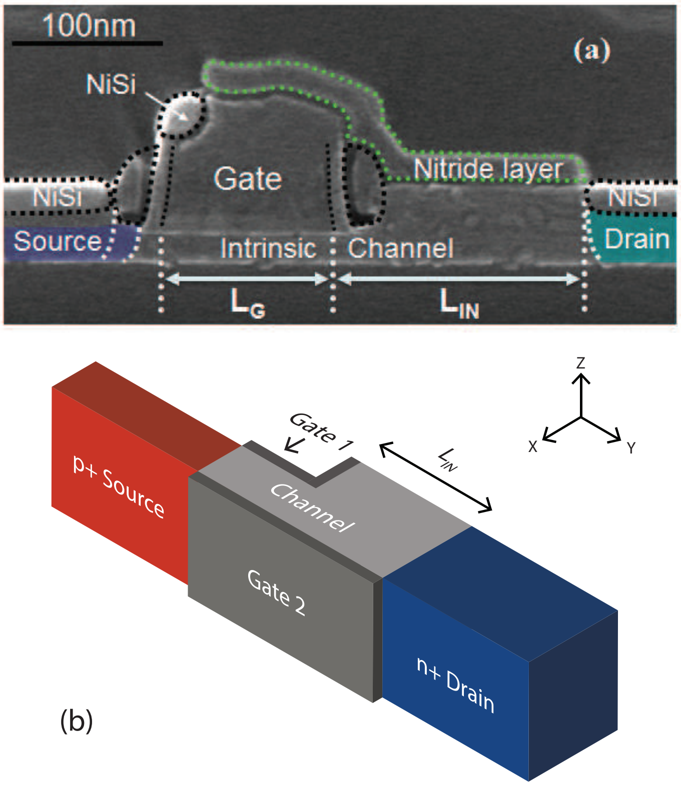


Figure 4.5 – (a) SEM image of a fabricated FD-SOI Tunnel FET showing the top gate partially covering the channel [23]. (b) Independent gate FinFET device structure for TFET based capacitorless DRAM.

For this study the Tunnel FET architecture used is shown in Fig. 4.5 (b). It is basically an independent gate Fin-Tunnel FET implementation of the FD-SOI devices shown in Fig. 4.5(a) [23]. It is known that for efficient Tunnel FET operation a fully depleted body is required. However for charge storage in the body partial depletion is preferable. In order to meet both conditions a step like fin is used, which manages the trade-off between the electrostatic control of the Gate1 and the ability to store charges by Gate2. Simulations showed that 50-60nm fin widths for the partially depleted part were optimum. Gate1 partially covers the channel region and will be used as the main control gate for Tunnel FET operation. Gate2 on the other hand covers the entire channel region and as explained below will be used to induce a potential well in the channel region. This potential well will then be used to store charges for memory operation.

4.3.1 Principle of Operation

To understand the principle of operation let us take a 2D cut of the device shown in Fig. 4.5(b) along XY plane. Simulations were done in Synopsys Sentaurus TCAD [11]. The devices simulated had fin width of 25nm in the source side and 50nm in the drain side, fin height= 65nm; Gate1 length (L_{G1}) = 80nm; Gate2 length (L_{G2}) = 200nm and a pocket length $L_{IN} = 120nm$, unless otherwise specified. A source/drain doping of $1x10^{20}cm_{-3}$ of Boron/Arsenic was used. The channel was intrinsically doped. 3nm of SiO_2 gate oxide was used for both Gate1 and Gate2. The same metal work-function was used for both the gates. Fig. 4.6 (a) shows the 2D hole-density plot with $V_{G1} = 2V$, $V_{G2} = -1V, V_D = 0V, V_S = 0V$. The memory operation is based on the creation of an induced potential well achieved by biasing Gate2 with a negative potential. A small positive bias on the source can help to flood the body with excess holes. These excess holes are then trapped in the potential well close to Gate2 as shown in Fig. 4.6 (top). Fig. 4.6(b) shows the cross-section at 1nm above the Gate2. We can clearly see the build-up of holes in the hole-density plot. The figure also shows an induced potential well where the holes can be stored. Hence, as the device turns on, the presence or absence of the excess carriers will affect the threshold voltage of the device indicating two different memory states.

The simulated transfer curves of the device in Fig. 4.5(b) is shown in Fig. 4.7(a). This is similar to what has been measured in [102] for FD-SOI devices. The principle of charge storage can be further verified by the hysteresis curves showed in Fig. 4.7(b). With $V_{G1} = 2V$ and $V_D = 1V$, as the back gate bias is swept back and forth the charge storage in the partially depleted region results in two different current paths. When the V_{G2} goes from -1V to 2V in the first sweep, holes start to accumulate in the induced potential well in the body affecting the body potential, resulting in the lower branch on current in the plot. As V_{G2} increases to +2V, the accumulated holes are evacuated to the drain at this state. Now, as V_{G2} goes back to -1.5V in the following sweep, the drain current follows the top branch due to the lack of the excess charges in the body. Based on this observation, we can set the design a programming scheme of a memory cell with read potential at Gate2 at 1.5V. A fast sweep rate of 1mV/μsec was used in this transient simulation for convergence issues.

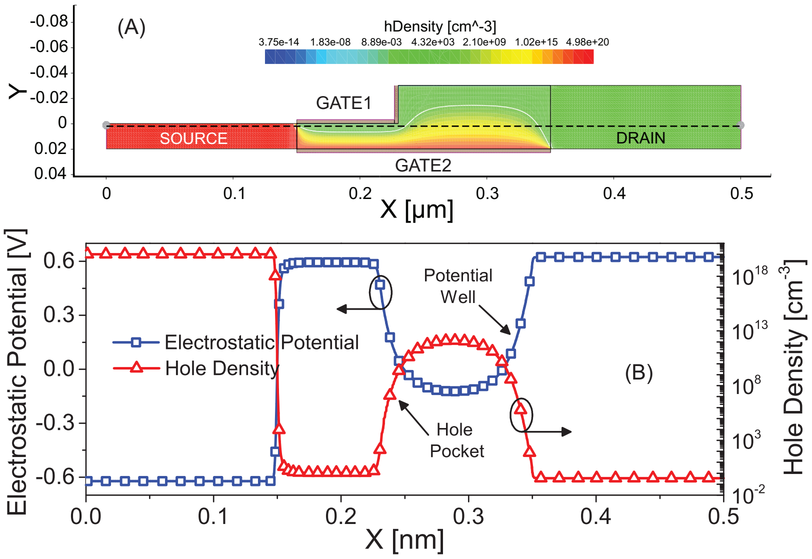


Figure 4.6 – 2D cross-section of the Fin-TFET showing the hole density at the end of a write 1 operation (top). Hole density and electrostatic potential at a cross section 1nm above the Gate2 showing the formation of potential well/hole pocket (bottom) $V_{G1} = 0V$, $V_{G2} = -10V$, $V_S = 0.5V$.

It is also interesting to note that the hysteresis is stronger for devices with longer L_{IN}. This is expected as length of the partially depleted region greatly influences the charge storage as also explained in [102]. So in this particular structure, the memory cell will loose its retention characteristics with gate length scaling. There will be a minimum L_{IN} required for the memory cell to work satisfactorily.

The dependence of charge storage on L_{IN} is also observed in measurements from the devices described in [102]. The relaxation time is observed to be more prominent for devices with longer L_{IN}. Fig. 4.8(a) shows that depending on the length of the intrinsic region (L_{IN}) the discharge time varies. The device with longer $L_{IN} = 500nm$ has a slower discharge than the device with $L_{IN} = 200nm$. This means that longer the L_{IN} the more charge it stores and hence it takes longer to discharge. This experiment also demonstrates the importance of charge storage in the un-gated region. The discharge time can be related to the retention time of a conventional capacitorless DRAM, as it indicates how long the device can hold the charge once the write cycle is complete. In this context we observe a discharge time in the order of few milliseconds for devices with longer L_{IN}. The same principle can also be observed with steady state measurements in Fig. 4.8(b) and (c). The device with a longer L_{IN} clearly shows a stronger history effect. This observation validates the simulation results shown in Fig. 4.7.

4.3.2 Proposed Programming Scheme

The proposed programming scheme for the device in Fig. 4.6(a) is shown in table 4.2. Each step is explained in more details in the following points:

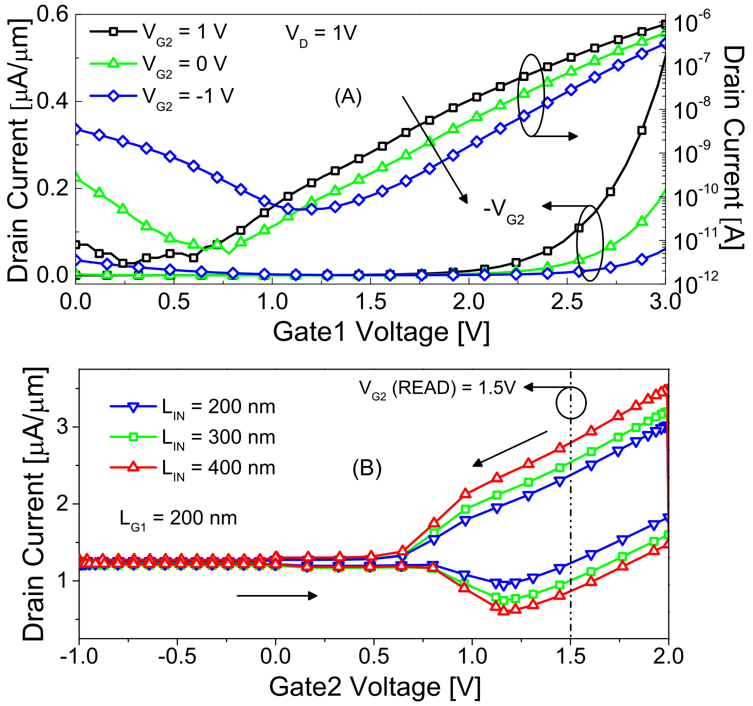


Figure 4.7 – (a) Simulated transfer curves at various Gate2 bias. (b) Simulated hysteresis curves observed with varying L_{IN} with same biasing conditions. Devices with longer L_{IN} shows stronger hysteresis. $V_{G1} = 2V, V_D = 1V$.

(i) WRITE 1: The write 1 step involves biasing the source with a small positive voltage, together with a negative bias on Gate2. This would push holes into the body which will then be trapped in the induced potential well caused by Gate2 as shown in Fig. 4.6(b). Bias values used in simulation, $V_{G1} = 0V, V_{G2} = -1V, V_S = 0.25V, V_D = 0V$.

(ii) WRITE 0/ERASE: The write 0 step would simply mean putting a positive bias to the Gate2 ($\approx 2V$). This will remove any induced potential well from the previous state and the holes in the body diffuse back to the source or recombine to drain. Bias values used in simulation, $V_{G1} = 0V, V_{G2} = 2V, V_S = -0.25V, V_D = 0V$.

(iii) HOLD: After the write 1 or 0 step the source goes back to zero and a small positive bias on Gate2 holds on to excess charges (if any) in the body. Bias values used in simulation, $V_{G1} = 0V, V_{G2} = 1.0V, V_S = 0V, V_D = 0V$.

(iv) READ: The readout operation is carried out via Tunnel FET operation with BTBT from p+ source to the intrinsic channel. For this, the drain is biased at V_{DD} and V_{G1} > threshold voltage (V_{TH}) of the device. Source remains at zero bias for the read operation. Gate1 controls the read out current. The threshold voltage of the device is affected by the presence or absence of excess carriers in the body. Bias values used in simulation, $V_{G1} = 2V, V_{G2} = 1.5V, V_S = 0V, V_D = 1V$.

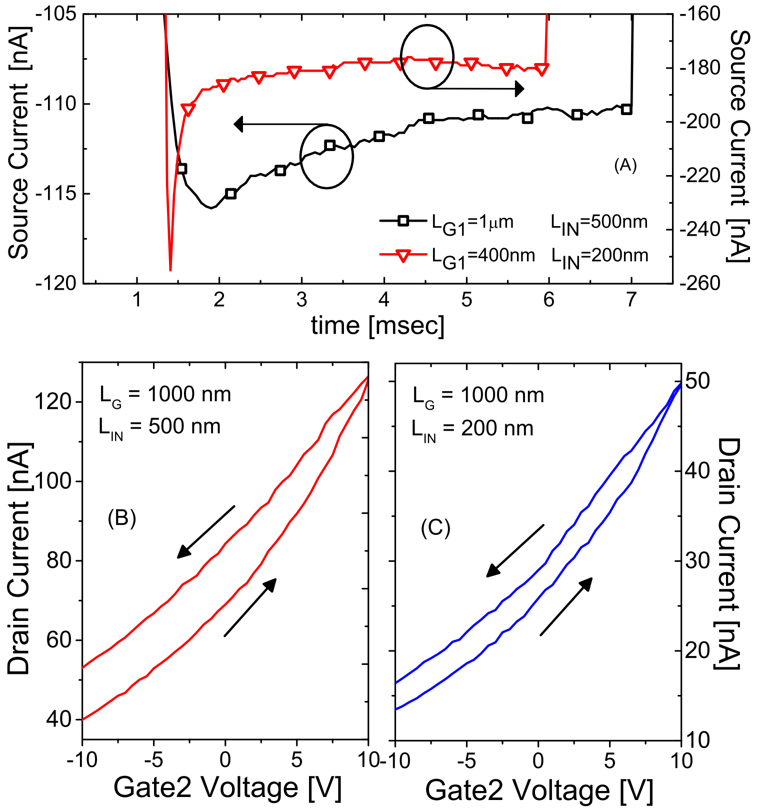


Figure 4.8 – (a) Different relaxation times in the millisecond range are recorded when the back gate is biased at $V_{G2} = -10V$ (after writing 1), depending on the length of the L_{IN} region. The TFET with $L_{IN} = 500nm$ shows the longer discharge time while it is negligible in devices with $L_{IN} = 200nm$. (b) & (c) Measured drain current with respect to back gate voltage at fixed front gate and drain voltages, $V_{G1} = 4.5V, V_D = 4V$.

Transient simulations were done to observe the hole density in the body in HOLD mode preceded by WRITE 1 and WRITE 0 operations. The results are given in Fig. 4.9. A clear difference is present showing that more holes are stored under the front gate after WRITE 1 compared to after WRITE 0 state. The same principle is also observed in the potential profiles at 1nm below the Gate1. After a WRITE 1 operation, there is a hole pocket present in the channel region. Similarly after a WRITE 0 operation, the hole pocket is much smaller. This hole pocket will primarily act as a resistance to current flow. Now as the device is turned on, the presence or absence of this excess carriers will affect drive current of the device indicating two different memory states.

4.3.3 Experimental Results

The basic connection schematic used for the pulsed transient measurements is show in the Fig.C.1. The Agilent 4156C semiconductor parameter analyser was used in sampling

Table 4.2 – Proposed programming scheme of Fin-TFET based capacitorless DRAM.

State	$V_D[V]$	$V_{G1}[V]$	$V_{G2}[V]$	$V_S[V]$
WRITE 1	0	0	Negative bias to induce potential well/accumulation of holes	Small positive bias to push holes into the body
WRITE 0	0	0	Zero or small positive bias	Small negative bias to aid the erase process
HOLD	0	0	Small positive bias to maintain the potential well	0
READ	V_{DD}	$> V_{TH}$	Value at which maximum hysteresis observed (1.5V)	0

mode for the transient measurements. Three function generators(FG) were used to pulse the two gates and the drain terminals. The source current was recorded at every $60\mu s$ (limited by equipment specifications). The function generator outputs were also monitored in a digital oscilloscope as shown in Fig.C.1 in appendix C.

4.3.4 Temperature Dependence

The transfer characteristics of the reported Tunnel FET devices have been measured from $25°C$ to $85°C$ and they show almost unchanged sub-threshold swing over the entire range as highlighted in the inset of Fig. 4.10(a) and low I_{OFF}. This is expected as band-to-band tunneling current is only weakly dependent on temperature [85]. Most importantly in Fig. 4.10(b), the discharge cycle (as highlighted in Fig. 4.11(b)) of the source current after a READ 1 operation, remains unaffected (at $T = 25°C, 55°C, 85°C$). This reflects the retention ability of the memory cell at high temperatures, in contrast with results reported on MOSFET-based solutions [96, 93], wherein the retention ability of the devices degrade with increasing temperature.

4.3.5 Transient Measurements

SiO_2 gate dielectric devices

Transient measurements were carried out FD-SOI Tunnel FET devices [23] with 20 nm Silicon layer thickness, 145 nm of BOX and 6nm of SiO_2, according to the programming scheme is depicted in Table 4.3. This scheme is different from what was reported in [102] in order to better understand the principle of operation. The timing diagrams of READ and WRITE operations are shown in Fig. 4.11. The Gate1, Gate2 and drain were pulsed according to Fig. 4.11(a) below.

During the WRITE cycle the Gate1/drain cycle is 50 % phase shifted to the Gate2 pulse. This was to ensure that induced potential well stays as the device is turning ON after a

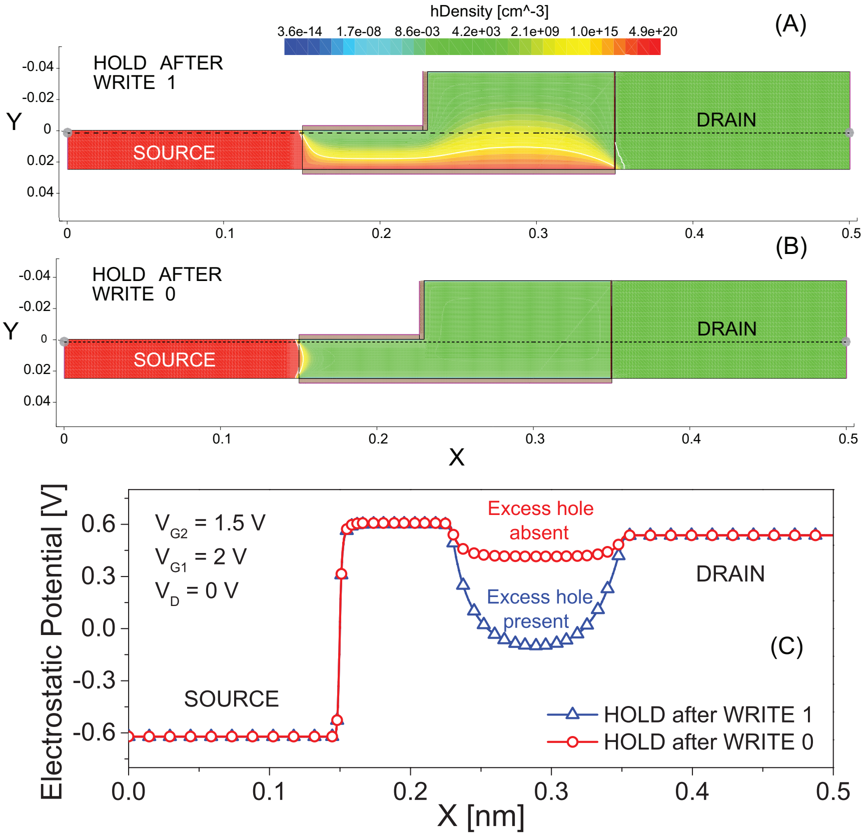
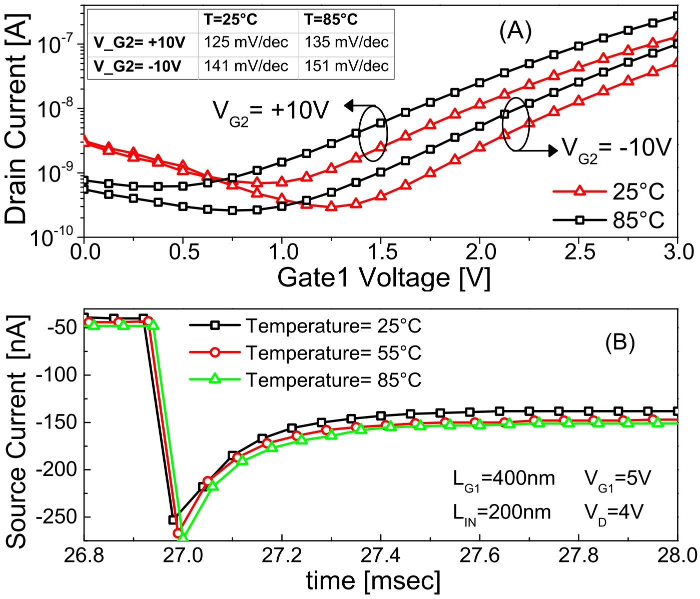


Figure 4.9 – Simulated 2D hole densities after a WRITE 1 (a) and WRITE 0 (b) operation. (c) Simulated potential profile at 1nm below Gate1 showing the hole pocket.

WRITE 1. As highlighted in Fig. 4.11(b), after a WRITE 1 operation the Gate2 switches from -10V to +10V, with the device still ON the potential well ceases to exist and the stored charges can now escape. In other words the capacitor discharges as the Gate2 switches back to +10V. This capacitor discharge can be observed in the slow transient in Fig. 4.11(b). This experiment demonstrates the importance of the charge storage in the hole pocket outside the front gate overlap, controlled by V_{G2}. As shown in Fig. 4.11(d) a memory effect is thus observed.

During the ERASE cycle the Gate1/drain pulse is in opposite phase to the Gate2 pulse. Finally the READ cycle has the same biasing conditions (most negative Gate2 bias) for both WRITE and ERASE. The device turns ON at the most negative Gate2 bias i.e. -10V. Depending on the previous state, a clear difference in READ (source) current levels (ΔI_S) of 10nA is observed (Fig. 4.11(d)).

In the experimental setup, the biasing conditions are slightly different than what has been suggested in section 4.3.2. Most importantly a HOLD state could not be experimentally demonstrated, because of the very low ON current drive ($\approx 10nA$) of the devices measured. The reported high values of V_{G2} in this section as compared to section 4.3.2 is due to the fact that the measured device was a planar device on a SOI substrate with a 145nm BOX. Hence, a high



	T=25°C	T=85°C
V_G2= +10V	125 mV/dec	135 mV/dec
V_G2= -10V	141 mV/dec	151 mV/dec

Figure 4.10 – (a) Measured transfer characteristics with different Gate2 voltage at 25°C and 85°C temperature. As expected for a TFET the sub-threshold swing has a negligible temperature dependence. (b) Measured relaxation times as a function of temperature after a READ 1 operation. A minor dependence of the relaxation behavior on the temperature is experimentally demonstrated.

Table 4.3 – Programming conditions for indicated operations of capacitorless 1T TFET DRAM with $L_{G1} = 400nm$, $L_{IN} = 200nm$.

State	$V_D[V]$	$V_{G1}[V]$	$V_{G2}[V]$	$V_S[V]$
WRITE 1	0.5	0	-10	0
WRITE 0	0	0	+10	0
READ	4	4	-10	0

V_{G2} was required to have the same effect as -2V (V_{G2}) in the simulated FinFET structure. Also it was not possible to measure the source current and at the same time pulse the source potential (V_S kept at a constant potential) due to the equipment limitations in the measurement setup. In erase operation, the influence of positive source is partially compensated by using highly positive gate bias at Gate2 (V_{G2}). In read operation drain and Gate1 potentials are chosen such that a desired ON current drive is achieved. Another difference in the biasing scheme can be observed in V_D for write operation. This has been done to stop the holes to recombine at the drain side. Also the Gate2 bias for read is different based on the observed history effect. At last Gate1 (V_{G1}) is biased positively in all the operation modes again due to the setup limitations. The reason of the low frequency choice is based on the ON current levels of the fabricated devices and the high parasitic capacitance contributed by non optimized test pads and the test instruments.

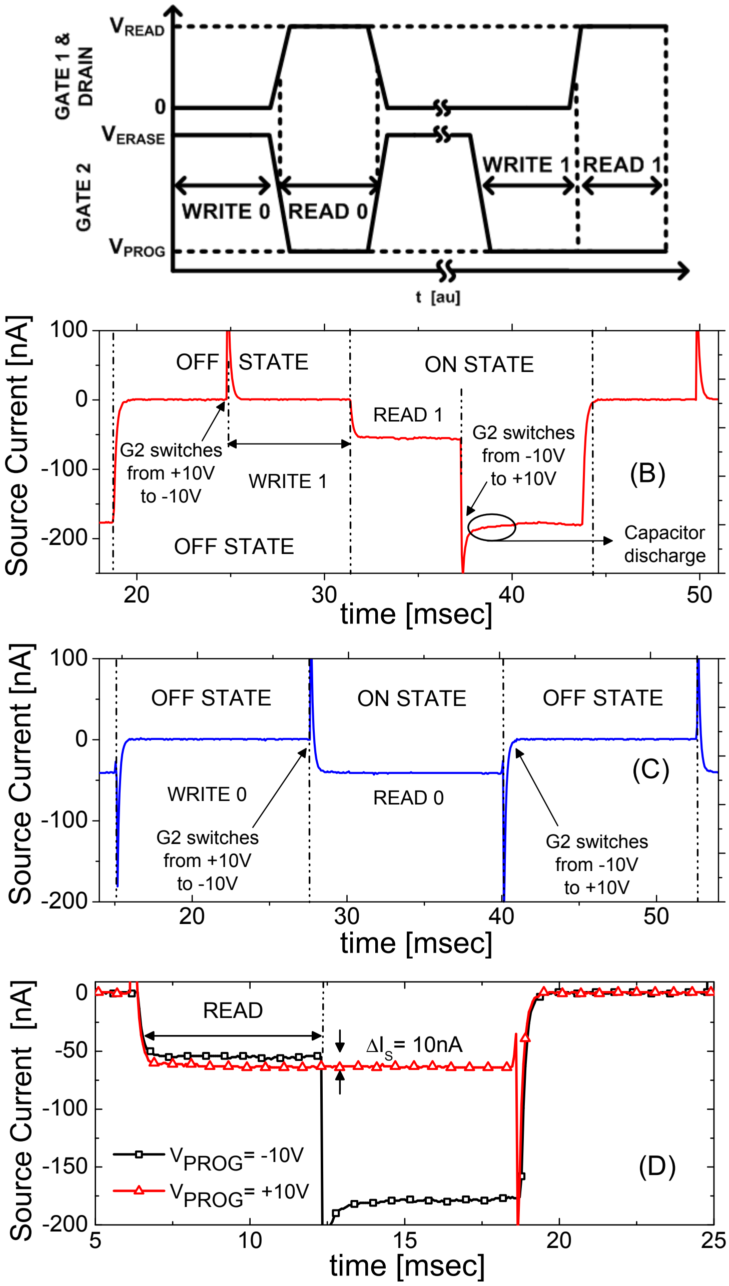


Figure 4.11 – Timing diagram for different operations described in Table 4.3. Consecutive READ and WRITE operations for both states. The difference in READ current for states 1 and 0 is 10nA, as highlighted, a memory effect is observed. Programming voltages V_{READ}, V_{ERASE} and V_{PROG} are as defined in table 4.3 for TFET with $L_{G1} = 400nm, L_{IN} = 200nm$ and SiO_2 gate dielectric.

Table 4.4 – Programming conditions for indicated operations WRITE “1” and READ of capacitorless 1T TFET DRAM with $L_{G1} = 400nm$, $L_{IN} = 200nm$. The ERASE also corresponds to WRITE “0.”

State	$V_D[V]$	$V_{G1}[V]$	$V_{G2}[V]$	$V_S[V]$
WRITE 1	4.5	0	-10	0.5
WRITE 0	4.5	0	+10	0.5
READ	4	4.5	-10	0.5

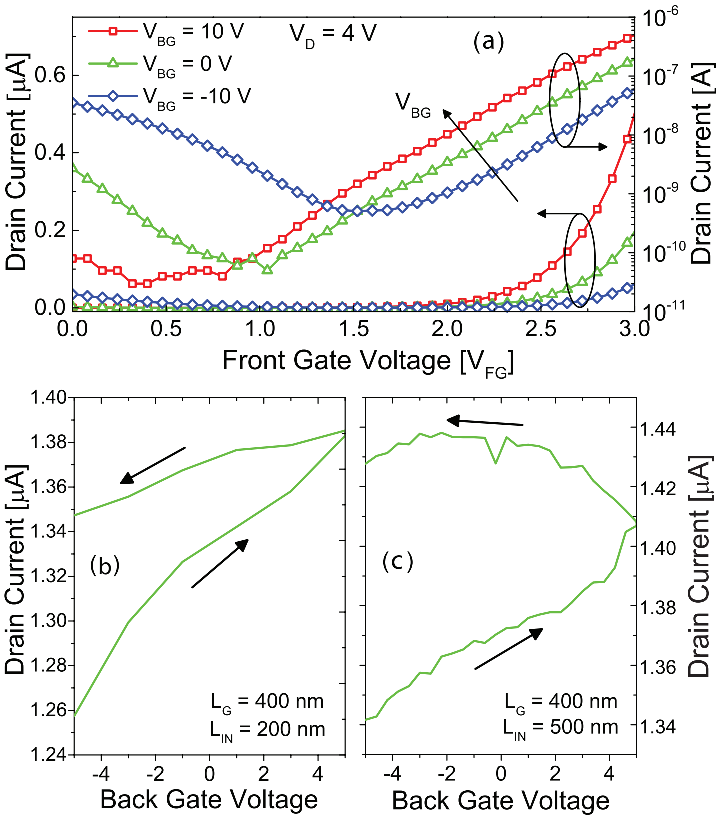


Figure 4.12 – (a) Measured transfer characteristics, $ID - V_{FG}$ with back gate potential as a parameter, for $L_G = 400nm$, $L_{IN} = 200nm$. (b) and (c) Measured drain current with respect to back gate voltage at fixed front gate and drain voltages, $V_G = 4.5V$, $V_D = 4V$. A history effect (open hysteresis loop) is observed more prominently in the TFET device with $L_{IN} = 500nm$

HfO_2 gate dielectric devices

Additionally transient measurements were carried out FD-SOI Tunnel FET devices [23] with 20nm Silicon layer thickness and 145nm of BOX and 3nm of HfO_2 gate dielectric, according to the programming scheme is depicted in Table 4.4. This scheme was reported in [102]. The timing diagrams of READ and WRITE operations are shown in Fig. 4.11. The Gate1, Gate2 and drain were pulsed according to Fig. 4.11(a) below.

The programming scheme is depicted in Table 4.4 and the timing diagrams for READ and WRITE operations are shown in Fig.4.13. The front gate, back gate, and drain were pulsed according to Fig.4.13. The source current was recorded every $60\mu s$ (limited by measurement setup). A READ operation is performed immediately after a WRITE ("1"/"0") step. The read out current was recorded at the most negative back gate bias (-10 V). Depending on the previous stored state, a clear difference in READ (source) current levels (ΔI_S) is observed (Fig.4.13). This difference, ΔI_S, is more prominent for devices with longer L_{IN} and is larger than 10–20nA for devices that have larger levels of I_{ON} current (close to $1\mu A$ at the maximum applied voltages). This experiment demonstrates the importance of the charge storage in the hole pocket in the region outside the front gate overlap, controlled by the back gate voltage amplitude. As highlighted, a memory effect is thus observed.

Consecutive ERASE/READ and WRITE/READ operations for a TFET (with $L_{G1} = 400nm$, $L_{IN} = 200nm$) based capacitorless memory. Fig.4.13(a)–(c) show the front gate, back gate, and the drain pulse scheme. Fig.4.13(d) and (e) show the source current at two different frequencies sampled every $60\mu s$. Based on the previous WRITE/ERASE state, a difference of $20nA$ in the READ current (I_S) for states "1" and "0" is observed. This difference, ΔI_S, is more prominent for devices with longer L_{IN} and is larger than $10-20nA$ for devices that have larger levels of I_{ON} current (close to $1\mu A$ at the maximum applied voltages). This experiment demonstrates the importance of the charge storage in the hole pocket in the region outside the front gate overlap, controlled by the back gate voltage amplitude. As highlighted, a memory effect is thus observed.

The history effect in Figures 4.12(b) and 4.12(c) shows a ΔI_S of $\sim 100nA$ at $V_{BG} = -5V$; however, in the transient measurements, only $20nA$ of ΔI_S is observed. This is because the quasi-static measurements of Figure 4.12 are much slower than transient measurements. The transient measurements were done at two different frequencies $50Hz$ and $200Hz$. Memory effect was not observed for frequencies higher than $500Hz$. The validation is performed at low frequency due to the low level of currents in TFETs and un-adapted pads for high frequency measurements.

It is worth noting that TCAD simulations have shown that a more effective WRITE "1" operation could be performed by biasing the source with a small positive voltage only during the WRITE "1" cycle. However, the limitations in the measurement setup required the terminal that was being sensed (Source in our case) could not be pulsed simultaneously. Hence, we could not verify this result experimentally.

If we can have an independent control of the back gate, then the proposed memory cell can be integrated with other cells in an array fashion with the control of 4 signals. As the write and erase operations are managed by drain and back gate terminals, for write/erase selectivity, the cells which share the source line should not be connected to the same back gate terminal line. For read operation, similarly, cells sharing the front gate line should be connected to different drain lines. Hence, the unit cell area can be estimated as $16F^2$ [103].

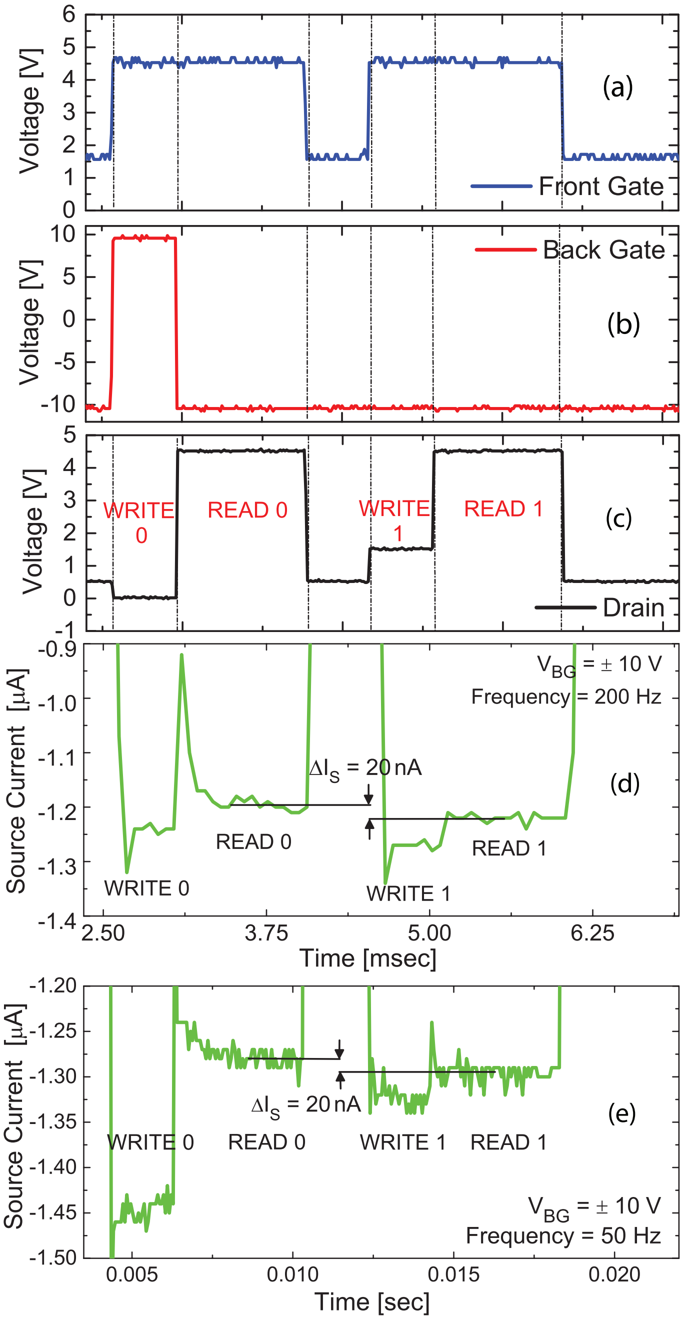


Figure 4.13 – Consecutive ERASE/READ and WRITE/READ operations for a TFET (with $L_G = 400nm$, $L_{IN} = 200nm$ and HfO_2 gate dielectric) based capacitorless memory. (a)–(c) show the front gate, back gate, and the drain pulse scheme. (d) and (e) show the source current at two different frequencies sampled every $60\mu s$. Based on the previous WRITE/ERASE state, a difference of $20nA$ in the READ current (I_S) for states "1" and "0" is observed.

4.4 Doped pocket device simulations

In this section we propose and validate by experimentally calibrated simulations a silicon Tunnel FET (TFET) based capacitor-less DRAM cell, implemented as a fully-depleted FinFET with CMOS compatible process. The devices have a conventional FinFET structure except for a $p+$ (for n-type TFET) doped pocket of length L_{PKT} and doping N_{PKT} between the intrinsic channel and the $(n++)$ drain. This doped pocket creates a necessary condition to store holes injected from the source-to-body junction. In [102], there was a need to induce a potential well in order to store the excess charges; whereas in the present case a potential well is permanently present due to the doped pocket. The drain voltage is used as a control voltage to either fill the potential well with carriers (WRITE “1”) by attracting holes from the $p++$ source or repel them to empty the well of carriers (WRITE “0”). In contrast with the SOI Z-RAM® [96, 98] there is no need of impact ionization to create/inject the hole charge in the device body, the holes being injected by the forward-bias $p+i$ junction, which significantly improves the device reliability as no impact ionization is involved.

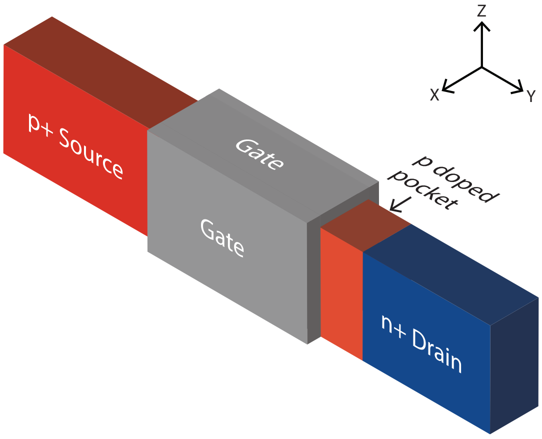


Figure 4.14 – 3D schematic of the proposed new structure (n-type) including a p+ doped pocket between the channel and the drain.

4.4.1 Principle of operation

In this section, we further build upon [102] that showed that all-Si double-gate Tunnel FET can serve for building a new class of devices: the capacitorless Tunnel FET DRAM, where the very low I_{OFF} is offering a major advantage for DRAM low power consumption/low refresh rate and the zero-capacitor structure a very high potential for scalability. The proposed new structure as shown in Fig. 4.14. Fig. 4.15(b) shows the energy band diagram of the device in Fig. 4.14. As shown in the image, the doped pocket acts as a resistance to the flow of carriers between the source and the drain and the drain bias is used to control this resistance to the current flow.

The potential well/pocket is of length $L_{PKT} = 25nm$ and is p-type doped at a value $N_{PKT} = 5x10^{18}cm^{-3}$ to enable optimum functioning of the memory cell. A higher pocket doping results in a deeper potential well as well as a higher resistance for carriers resulting in lower currents. A lower doping on the other hand requires a very high drain bias for memory operation, which is not feasible for a $L_G = 100nm$ gate length device. In the calculations to verify the memory operation of the device, abrupt doping profiles were assumed and the calculations/simulations were carried out for a 2D cross-section of the device of Fig.4.14. It should be noted that in the present structure the pocket length is not scalable. This is because smaller the pocket length, smaller will be the charge storage capacity. Hence as the pocket length is scaled down, after a certain length charge storage will no longer be possible.

Measurements on FDSOI TFET devices as reported in [29][23] were performed at elevated temperatures and used to calibrate the non-local band-to-band (B2B) tunnelling model in Sentaurus TCAD [11]. The charge relaxation characteristics of the proposed memory cell is simulated at an elevated temperature of 85°C and is shown to be not degrading at higher temperature as is the case in conventional capacitorless DRAMs [104].

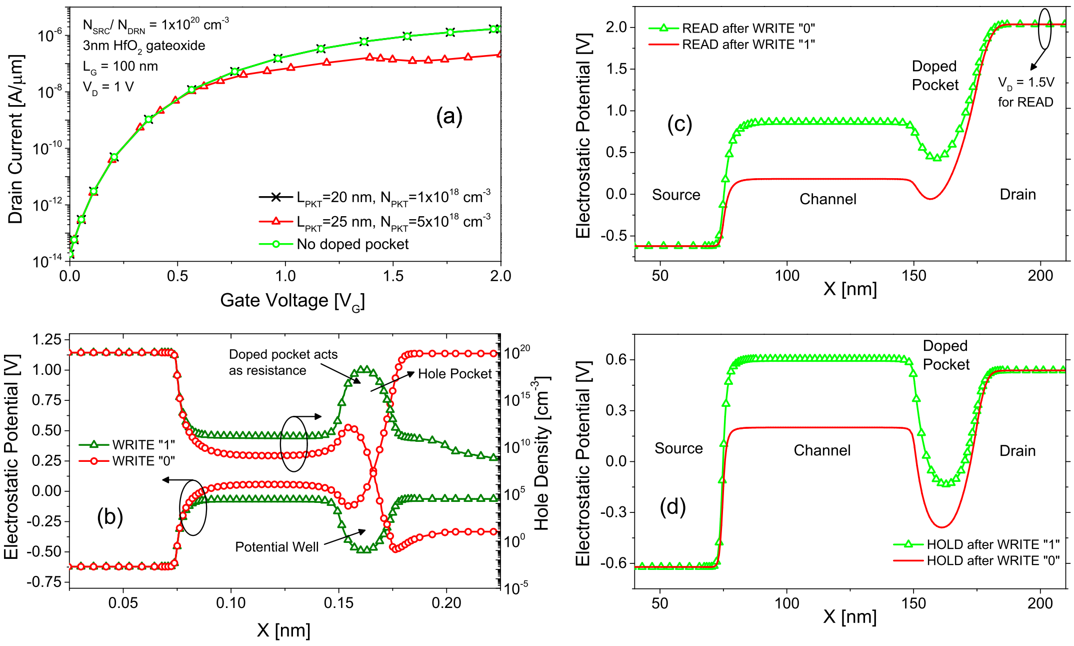


Figure 4.15 – 3D schematic of the proposed new structure (n-type) including a p+ doped pocket between the channel and the drain.

Transfer curves of the device in Fig. 4.14 was simulated for different L_{PKT} and N_{PKT} as shown in Fig. 4.15(a). The values $L_{PKT} = 25nm$ and $N_{PKT} = 5x10^{18}cm^{-3}$ were chosen to enable optimum functioning of the memory cell. Higher pocket doping results in a deeper potential well as well as a higher resistance to carrier resulting in lower currents. Lower doping on the other hand would require very high drain bias for memory operation, which is not feasible for a $L_G = 100nm$ device. Fig. 4.15(b) shows the simulated hole densities at the end of a WRITE “1” and WRITE “0” state. It is quite evident that the hole density in the pocket area after a WRITE “1” state is quite higher than after WRITE “0” state (also shown in Fig. 4.16). This is reflected in the potential profiles in the same figure as well. Fig. 4.15(c,d) again shows the simulated potential profiles during HOLD state (top) and READ (bottom) after both WRITE “1” and WRITE “0” states. As the device turns ON after these two states a difference in drive current is expected, which forms the basis of the memory operation. It is to be noted that the memory state reading operation is carried out using band to band tunneling of carriers from the source region to the intrinsic channel region in this case as well.

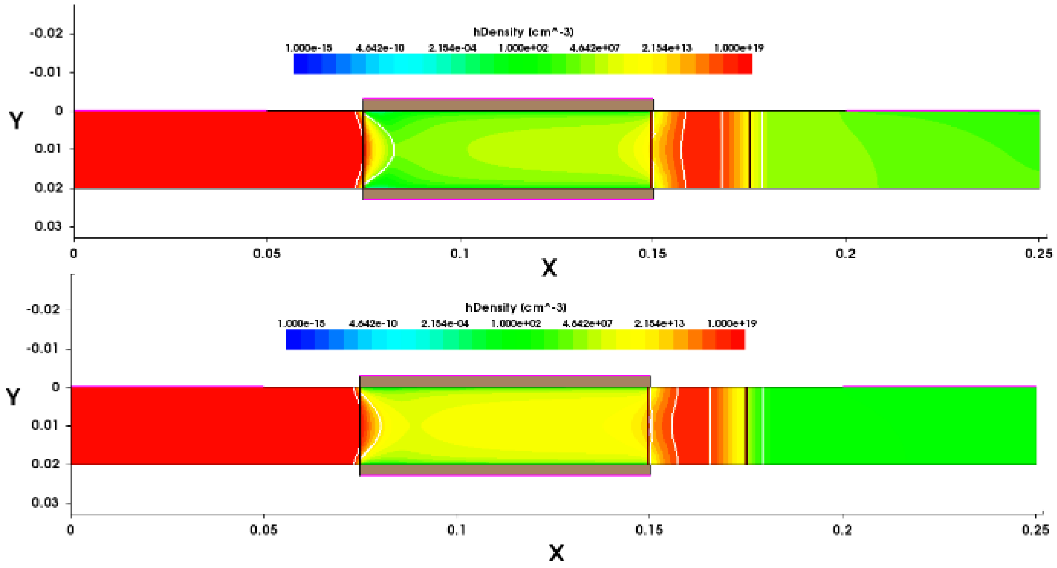


Figure 4.16 – TCAD simulated 2D cross-section of the proposed new structure showing the hole density at a HOLD state after and WRITE "1"(top) and WRITE "0"(bottom) operation.

4.4.2 DRAM operation

The programming scheme is depicted in table 4.5. The effect of a back and forth sweep of the control drain bias is used to study the hysteresis effect. As shown in Fig. 4.17(a), a significant difference of $900nA$ is observed at $V_D = 1.5V$ between the two READ currents. Fig. 4.17(a) also shows the same curves at elevated temperatures of $55°C$ and $85°C$. Temperature dependent simulations were carried out using the fitted parameter list calculated using [105]. A minimal temperature dependence was observed.

The timing diagrams for READ, HOLD and WRITE operations are shown in Fig. 4.17(c) and the corresponding simulated drain current is shown in Fig.4.17(d). Depending on the previous stored state, a clear difference in READ current levels ($\Delta I_D = 500nA$) is observed. Fig. 4.17(b) shows the charge relaxation characteristic of the proposed new memory cell. A charge relaxation time in the order of $100\mu sec$ was observed for the simulated device. The same figure also shows the charge relaxation characteristic at an elevated temperature of $85°C$.

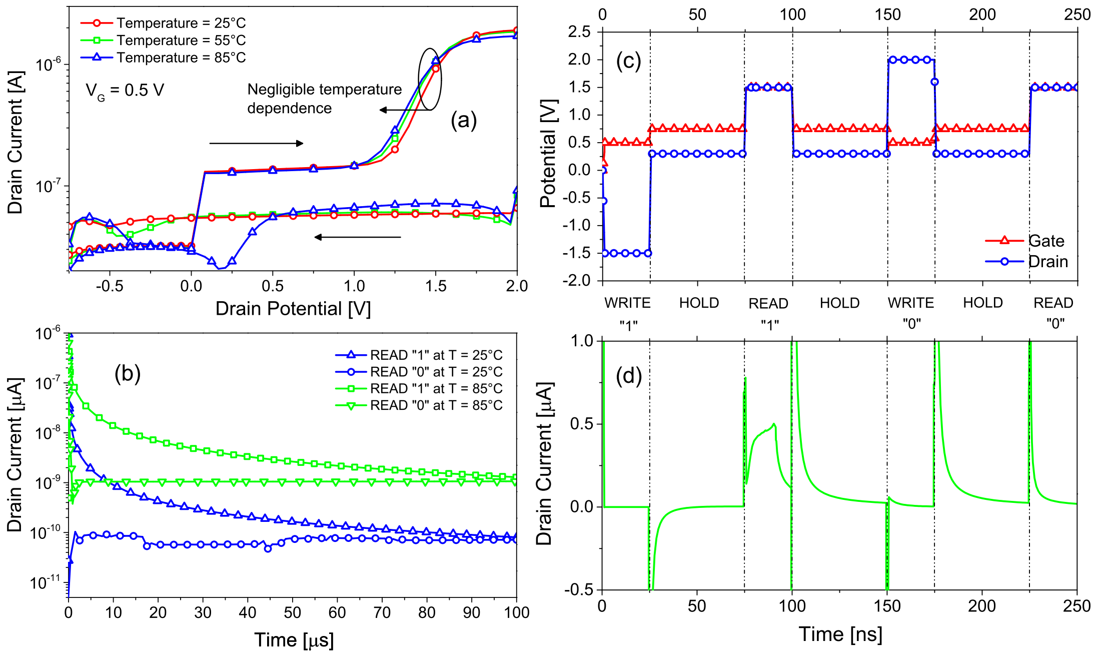


Figure 4.17 – Continuous read/write cycle with hold time in between. A difference of $500nA$ is observed between the two memory states.

Table 4.5 – Programming conditions for indicated operations of capacitorless 1T TFET DRAM with a doped pocket and $L_{G1} = 400nm$, $L_{IN} = 200nm$.

State	$V_D[V]$	$V_G[V]$	$V_S[V]$
WRITE 1	-0.75	0.5	0
WRITE 0	2.0	0.5	0
READ	1.5	1.5	0

It is to be noted that the charge relaxation characteristic is not degrading with increasing temperature, which is related to the TFET sub-threshold swing temperature stability [30][85]. Also to be noted is that the read '1' current has to be read within the first 20ns of the read pulse. This is a limitation of the present structure as the read '1' is a destructive read operation.

In summary we have validated a new scalable 1T capacitorless DRAM cell based on DG fin Tunnel FET structure with a doped pocket, which opens a new field of applications for TFETs. A clear DRAM memory operation has been proposed and verified with simulations. Charge relaxation times in the order $100\mu sec$ is shown as not degrading at high temperature ($85°C$) which is one of the advantages for TFETs based memory cells.

Table 4.6 – Parameters of non-local band-to-band tunneling model at 25°C, 55°C and 85°C fitted with measurements [105].

Temperature	$A_{PATH}(cm^{-3}sec{-}1)$	$B_{PATH}(V/cm)$
25°C	$4X10^{14}$	$1.9X10^{7}$
55°C	$4.23X10^{14}$	$1.881X10^{7}$
85°C	$4.51X10^{14}$	$1.861X10^{7}$

4.5 Summary

In summary we have presented a new application of Tunnel FETs as a memory cell. Simulation results on a new asymmetric TFET structure that can be used as a capacitorless DRAM. We have also experimentally demonstrated a scalable implementations of such 1T capacitorless DRAM cell based on DG Tunnel FET with a δI_S of $10nA$ and with retention times in the order of 100's of microseconds to few milliseconds, at room temperature in devices with channel length varying from $400nm$ to $1\mu m$. The discharge cycle was shown to be unaffected when the temperature varies between 25°C to 85°C.

We have also validated a new scalable 1T capacitorless DRAM cell based on DG fin Tunnel FET structure with a doped pocket, which opens a new field of applications for TFETs. A clear DRAM memory operation has been proposed and verified with simulations. Charge relaxation times in the order $100\mu sec$ is shown as not degrading at high temperature (85°C) which is one of the advantages for TFETs based memory cells.

5 Tunnel FET based Non Volatile Memories

In this chapter we will see how Tunnel FETs can be used as non-volatile memory cells. Non-volatile memory (NVM) or non-volatile storage is computer memory that can retrieve stored information even after having been power cycled. While [106] was the first to buld a non-volatile memory with a Tunnel FET, they have used a conventional lateral tunneling nano-wire structure. In this chapter we will show how vertical tunneling Tunnel FETs can be utilized in a way to boost the primary memory operation.

5.1 Flash Memory

Flash memory is an electronic non-volatile computer storage medium that can be electrically erased and reprogrammed. Flash memory was invented by Dr.Fujio Masuoka in 1980 at Toshiba, developed from EEPROM (electrically erasable programmable read-only memory). Flash memory can be divided into NOR and NAND-based memory [107]. EPROMs had to be completely erased before being rewritten. NOR-based flash memory provides high read performance and enables full address and data bus access which allows applications to run directly from the flash memory instead of reading the program into the system RAM first. The disadvantages are extremely slow write-and-erase cycles and a bigger cell size compared to NAND-based flash memory, which makes it cost effective in low-capacity data storage which rarely needs to be updated, like in computer BIOS or the firmware of set-top boxes.

NAND flash memory has about half the cell size of NOR flash memory and is an ideal solution for high-capacity data storage. It offers fast read and write performance, but lacks the easy memory access of NOR flash memory. Data must be read serially in blocks. Typically block sizes range from hundreds to thousands of bits. This feature disables the use of NAND flash memory as a drop-in replacement for program Read Only Memorys (ROMs), because most microcontrollers and microprocessors need byte-level access. Therefore, NAND flash memory is used in the category of other secondary storage devices like hard disks or optical media (e.g. CD, DVD). It is utilized as mass storage such as memory cards and USB flash drives. Due to the extremely high packing density it was possible to release a new generation of memory card formats exhibiting extremely small feature size. For instance, the microSD (TM) card has an area of about $1.5cm^2$, with a thickness of less than $1mm$ and offers presently up to 128GB of storage capacity at the same time.

5.1.1 Working Principle

Flash memory is made out of memory cells which are placed in an array. Every memory cell contains one floating gate transistor allowing to store at least one bit (Fig. 5.1(a)). Flash memory stores information in an array of memory cells made from floating-gate transistors. In traditional single-level cell (SLC) devices, each cell stores only one bit of information. Some newer flash memory, known as multi-level cell (MLC) devices, including triple-level cell (TLC) devices, can store more than one bit per cell by choosing between multiple levels of electrical charge to apply to the floating gates of its cells.

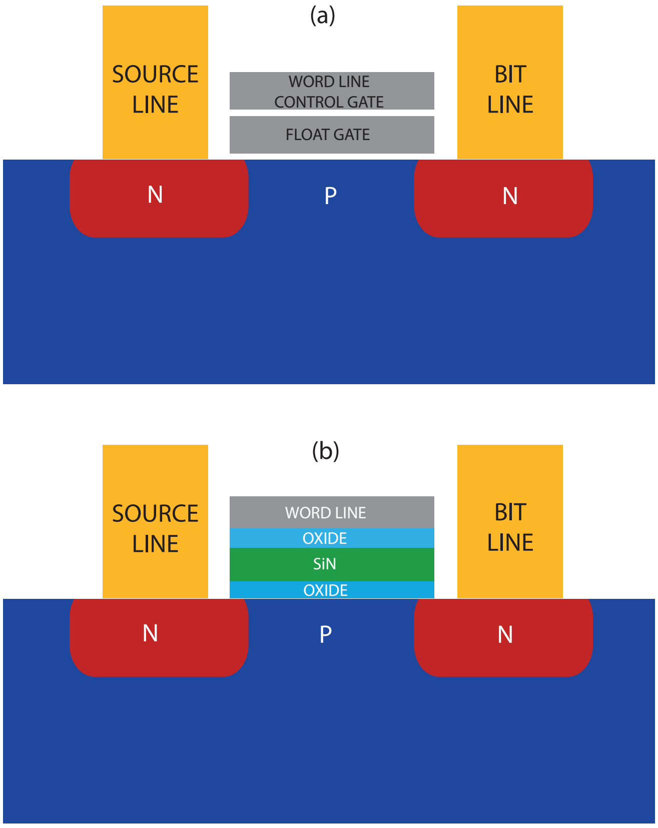


Figure 5.1 – Comparison between the schematic cross-section of a floating gate device (a) and of a SONOS device (b).

Floating-gate transistor

In flash memory, each memory cell resembles a standard MOSFET, except the transistor has two gates instead of one. On top is the control gate (CG), as in other MOS transistors, but below this there is a floating gate (FG) insulated all around by an oxide layer. The FG is interposed between the CG and the MOSFET channel. Because the FG is electrically isolated by its insulating layer, electrons placed on it are trapped until they are removed by another application of electric field.

Programming and erasing

In a floating gate type flash memory Channel Hot Electron (CHE) injection is employed for programming and Fowler Nordheim (FN) tunneling is employed for erasing the device as shown in Fig.5.2. Applying a large lateral field along the channel of the devices causes electrons to accelerate until they are scattered by the lattice or another carrier. The appropriate vertical field will result in "lucky" electrons, which will accumulate enough energy to cross the $Si - SiO_2$ barrier and inject themselves into the gate dielectric where they are stored in traps. Once the FG is charged, the electrons in it screen (partially cancel) the electric field from the CG, thus, increasing the threshold voltage (VT1) of the cell. This means that a higher voltage must be applied to the CG to make the channel conductive.

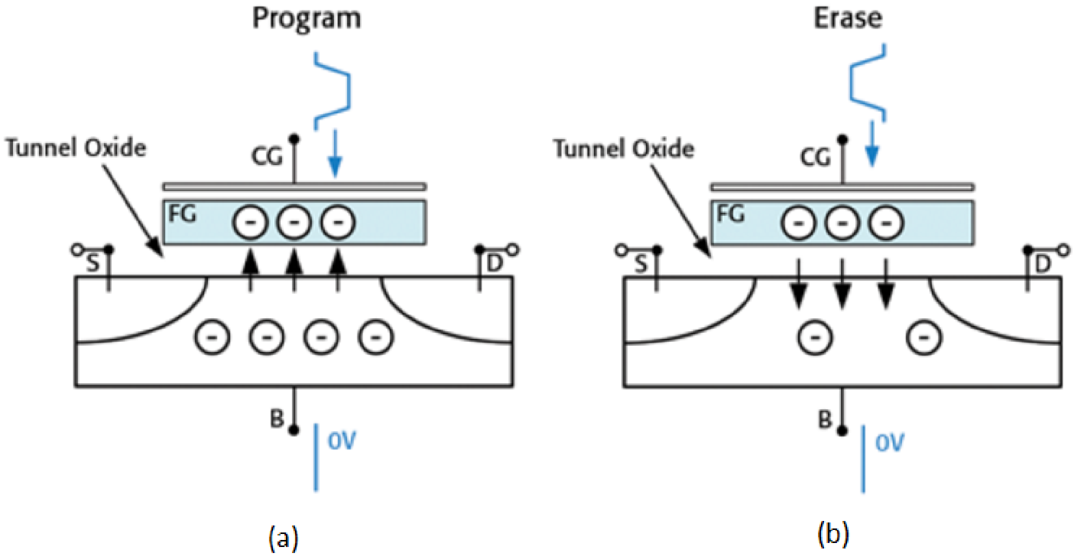


Figure 5.2 – (a) Programming via hot electron injection (b)erase via FN tunneling in a floating gate memory cell (source: Kiethley)

In order to read a value from the transistor an intermediate voltage between the threshold voltages (VT1 & VT2) is applied to the CG. If the channel conducts at this intermediate voltage, the FG must not be charged (if it were, we would not get conduction because the intermediate voltage is less than VT2), and hence, a logical "1" is stored in the gate. If the channel does not conduct at the intermediate voltage, it indicates that the FG is charged, and hence, a logical "0" is stored in the gate. The presence of a logical "0" or "1" is sensed by determining whether there is current flowing through the transistor when the intermediate voltage is asserted on the CG. In a multi-level cell device, which stores more than one bit per cell, the amount of current flow is sensed (rather than simply its presence or absence), in order to determine more precisely the level of charge on the FG.

5.1.2 Floating gate vs. SONOS

SONOS, short for "Silicon-Oxide-Nitride-Oxide-Silicon" is closely related to the floating gate flash memory concept described above. The floating gate device stores charge in a Polysilicon gate electrode as free carriers in the conduction band. The SONOS device stores charge in spatially isolated deep level traps. SONOS device is basically a MOSFET where the gate has been replaced by an ONO (Oxide-Nitride-Oxide) dielectric. Charges, holes or electrons, are injected into the nitride layer using direct tunneling through the tunnel oxide bottom layer [108].

As we can see from Fig.5.1(b), SONOS is very similar to standard floating gate devices, but it offers better performances in term of charge storage, scalability and level of programming voltages. Floating gate devices scaling is seriously limited by this fact. SONOS, on the other hand, requires a very thin layer of tunneling oxide in order to work and this opens to the possibility of strong scaling of both dimensions and power consumption [109]. The following points break down the fundamental differences between the SONOS and Floating Gate technologies:

- In memory operations, SONOS devices employ both electrons and holes. Floating gate devices use only electrons.

- SONOS structure is linearly scalable with programming voltage. Floating gate structure is not easily scaled.

- SONOS structures are immune to single defect memory loss because charge is stored in traps distributed throughout the dielectric. Floating gate structures are susceptible to single defect memory loss where a single defect in the tunnel oxide allows free charge to escape into the bulk.

- Floating gate has thicker tunnel oxide than SONOS devices, generally around the range of 7-10 nm. This leads to better retention for floating gate devices at the expense of higher programming voltages.

- SONOS devices operate with reduced electric fields in the tunnel oxide (Modified Fowler-Nordheim and Direct tunneling) as opposed to floating gate devices which use exclusively Fowler-Nordheim tunneling.

5.1.3 State of the art

There is a lot of promise of high-density cell integration in NAND flash memory, thus it gets much attention as a mass storage device [110, 111, 112, 113]. Currently, the cell integration density of the NAND flash memory is increasing very rapidly due to its simple structure suitable for high-resolution lithography [114]. The integration density of NAND flash memory is further enhanced by a factor of two by adopting multilevel cell operation.

As the cell integration density is increased, however, flash memory cell suffers from increased parasitic capacitance between the cells, and it has been have noticed that it generates serious problems in multilevel cell operation. We continue to see progressive scaling in embedded SRAM, DRAM, and floating-gate based Flash for very broad applications. However, due to the major scaling challenges in all mainstream memory technologies, we see a continued increase in the use of smart algorithms and error-correction techniques to compensate for increased device variability. In further response to these challenges, we see logic processes adopting FinFET devices along with read- and write-assist circuits in SRAMs. Emerging memory technologies are making steady progress towards product introductions, including PCRAM and ReRAM, while STT-MRAM is beginning to become a strong candidate for both standalone and embedded applications.

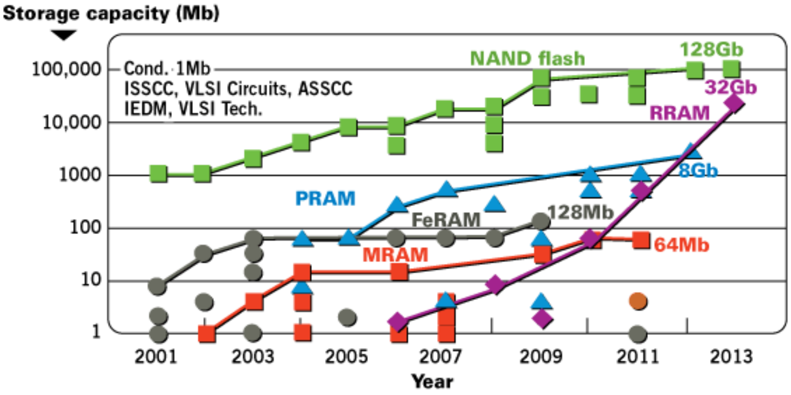


Figure 5.3 – Storage capacity of prominent flash memory technologies over the years. [115]

In the past decade significant focus has been put on the emerging memories field to find a possible alternative to floating gate nonvolatile memory (NVM). The emerging NVMs, such as phase-change memory (PRAM), ferroelectric RAM (FeRAM), magnetic spin-torque-transfer RAM (STT-MRAM), and resistive memory (ReRAM), are showing potential to achieve high cycling capability and lower power per bit for both read and write operations. Some commercial applications, such as cellular phones, have recently started to use PRAM, demonstrating that reliability and cost competitiveness in emerging memories is becoming a reality. Fast write speed and low read-access time are being achieved in many of these emerging memories.

NAND Flash memory continues to advance towards higher density and lower power, resulting in low-cost storage solutions that are enabling the replacement of traditional hard-disk storage with solid-state disks (SSD.s). Multiple bits per cell has proven to be effective in increasing the density. Figure 5.3 shows the observed trend in NAND Flash capacities presented at leading conferences over the past 2 decades with scaling, device variability and error rates increase, requiring system designers to develop sophisticated algorithms to offset this trend. For improved overall reliability. Possible future scenarios include 3D stacked NAND vertical gates as a solution to further increase the NAND density.

3D NAND flash

It's been clear for several years that three-dimensional NAND die stacking, in which chip layers are oriented vertically as opposed to horizontal planar structures was the way forward for next-generation chip designs. Until now, Samsung has been the only company to take that plunge, but that is going to change in 2015 with the launch of Intel's own solution in 2015. Intel, like Samsung, is expected to announce that it uses a much larger process node for its 3D NAND. In Samsung's case, it uses a 40nm process for 3D NAND, despite the fact that its working on 14nm planar technology for both logic and DRAM devices. Intel and Micron have already launched 16nm 2D NAND (Fig.5.4), but the fundamental characteristics of flash mean that device reliability decreases as process nodes shrink.

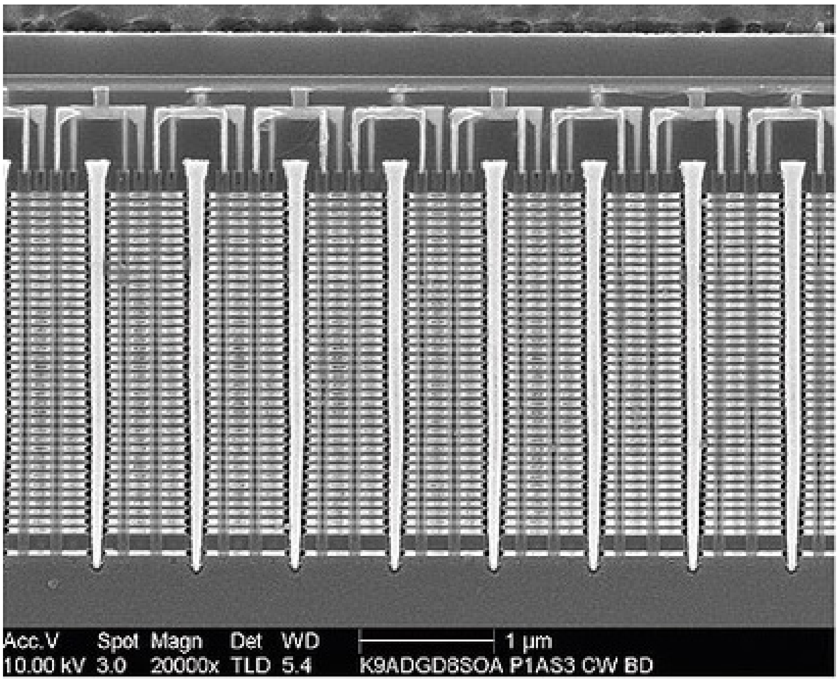


Figure 5.4 – SEM imaage of a close-up of V-NAND flash array from Intel [courtesy Chipwerks]

In case of a vertical V NAND memory, the SONOS stack is then oriented vertically, using a polysilicon cylinder as the substrate silicon, and wrapping the other layers around the central cylinder. The wordlines (control gates) become a horizontal layer, and the bitlines are connected to the top of the poly-Silicon cylinder; the select gates are formed by the top and bottom conductive layers [116].

5.2 Simulation methodology

The conventional SONOS structure and the TFET-based non-volatile memory have been created using Sentaurus Structure Editor. In both cases the gate stack consists of a 1.8nm SiO_2 layer, an 8.0nm Si_3N_4 layer, and a 4.0nm SiO_2 layer between the channel area and the poly-silicon gate (Fig.5.5). In the case of the TFET-based structure the silicon substrate has been substituted with a thick buried oxide layer and a 20 nm thick silicon epitaxial layer (SOI TFET) in order to avoid perturbations to the TFET current levels because of the substrate contact. Source and drain in the MOS based structure are doped with a Gaussian profile of Arsenic ($10^{20} cm^{-3}$), while the substrate is doped uniformly with boron ($10^{16} cm^{-3}$); in the case of the TFET's source the Arsenic is substituted with Boron.

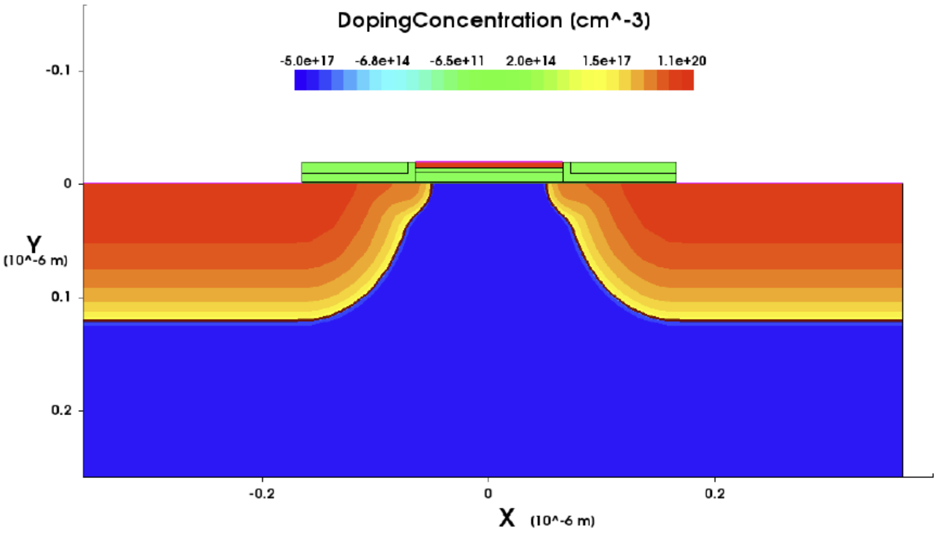


Figure 5.5 – Schematic cross-section of the conventional MOSFET based SONOS structure.

Sentaurus Device is also used to simulate the program and erase cycle of both the devices and to simulate the ID–VG curves during the read cycle. The single programming cycle is composed of a write phase, an hold phase of $2.5 msec$, an erase phase and an hold phase of $5 msec$. In the write phase for a p-type device a voltage of 9 V is applied to the gate contact for $.5 msec$, in the erase phase a voltage of -8 V is applied for $7.5 msec$ and in the hold phases no voltage is applied. The programming voltages of a n-type device will be of opposite polarity and discussed later. The programming scheme is also shown in Table 5.1.

The program and erase operations are conducted within a single transient simulation by varying the gate voltage, then the final program and erase states are saved at the end of a

Table 5.1 – Programming scheme for the proposed p-type memory cells.

State	$V_D[V]$	$V_G[V]$	$V_S[V]$
WRITE '1'	0	9	0
WRITE '0'	0	-8	0
HOLD	0	0	0

series of programming cycles to obtain the steady-state solution. In order to facilitate the convergence of the simulation (especially in case of the TFET-based memory), the meshing strategy in the gate stack has been designed to result in a high-quality mesh without excessive node counts. The rest of the meshing strategy has been defined using regular refinement boxes, which specify the allowed minimum and maximum mesh spacing within a specified area. For more details on the simulation models used refer to the example project in Sentaurus Workbench [11] "SONOS Memory".

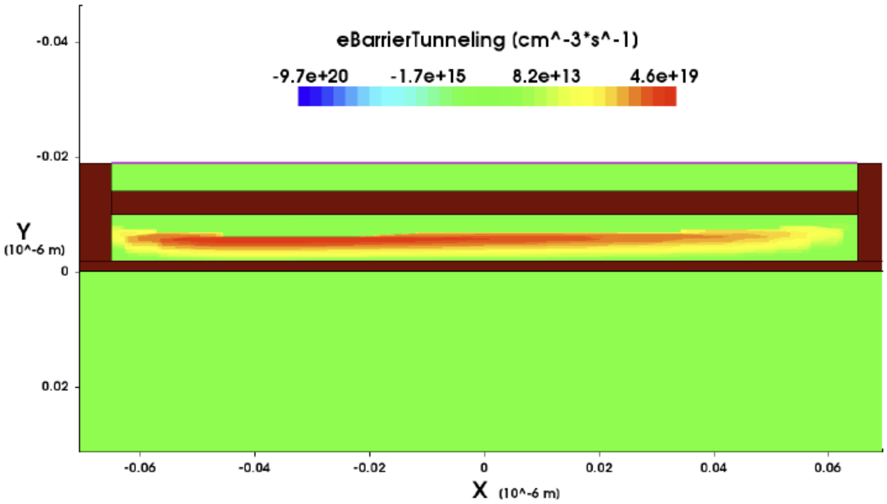


Figure 5.6 – Concentration of electrons during the program cycle that tunnel into the nitride region of the traditional SONOS device.

Throughout the program and erase cycles, the electrons and holes can tunnel between the nitride and the channel area and between the polysilicon gate and the nitride. As example, the plot of the density of electrons tunneling during the program and erase cycle is shown in Fig.5.6. The results of the simulation for both devices are shown in Fig.5.7. In both cases it takes approximately 4–5 cycles before the charge in the nitride reaches a steady-state value in the program and erase states.

As can be seen from Fig.5.6, the numerical values of the trapped charges are very similar, but the number of trapped electrons in the second case is lower than in the former case: this is due to the fact that in the MOSFET structure we have two reservoirs (source and drain) of free electrons available for tunneling through the bottom oxide, while in the

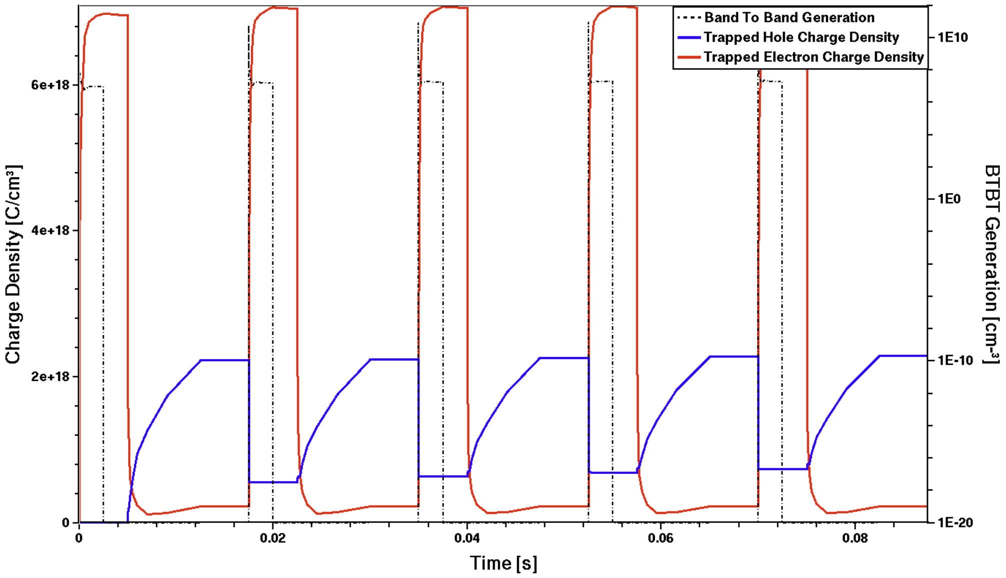


Figure 5.7 – Trapped electron and hole charge in nitride region of SONOS device as a function of time during the program and erase cycles; corresponding applied gate bias is also shown.

TFET device the p-doped source can not give the same number of electrons of the n-doped drain. At the same time the second plot suggests a possible solution to this drawback. In fact during each programming phase ($V_G = -9V$) there are some extra electrons which are tunneling (direct tunneling) into the silicon nitride layer (dot-dashed black line in Fig.5.7) coming from the source-channel interface (band to band tunneling): it could be interesting to exploit this interband tunneling mechanism to boost the total number of trapped electrons. It is important to notice that in this simulation the absolute value of the BTBT generation is quite low because of two main reasons.

- Firstly the horizontal TFET structure is not a suitable configuration for generating a big amount of free electrons because intraband tunneling can take place only along the 20 nm-long source/channel interface and because the tunneling junction is not designed having the same orientation of the component of the electric field modulated by the gate.
- Secondly, since during the programming phase the drain voltage is biased at 0V (Table 5.1), the on-state regime and the BTBT generation are strongly limited. Unfortunately it is difficult to increase the BTBT generation acting on the drain voltage: in fact any positive drain bias would result also in an increase of the drain current, and consequently of the power consumption, during the programming cycle.

In sections 5.2 and 5.3, we will show an optimized structure which will try to solve the issue related to the horizontal structure. As previously said, the program and erase states are saved to a file during the write cycle and are then used as the initial structure to sweep the gate bias and to plot the transfer curve with the source grounded and the drain fixed at 1.0 V: this is the reading phase. The comparison between the two devices is shown in Fig.5.8.

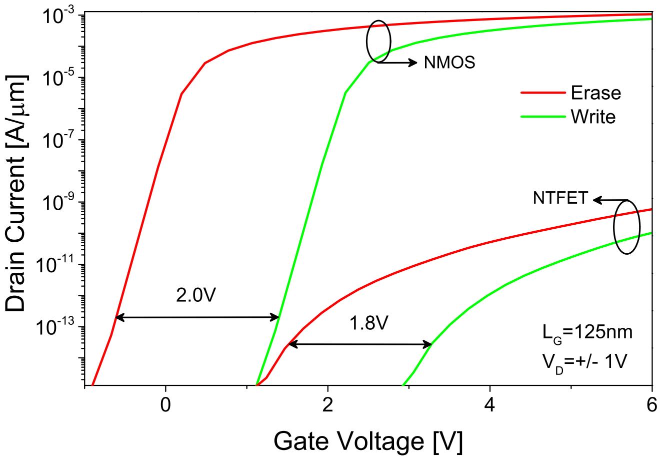


Figure 5.8 – ID -VG curves for program and erase states for n type MOFET and conventional lateral tunneling Tunnel FET.

As can be seen the traditional SONOS structure shows better performances, especially in term of the difference between the 1-state and 0-state current levels. However it is important to remark that also the TFET-based structure presents memory behaviour with a threshold voltage shift similar to the MOSFET case. The most important issues to face for the TFET case are the low values of the current levels and the degradation of the subthreshold swing with respect to the ideal case. Another evident drawback is the shift of the program curve towards high gate voltages, which forces to work at higher V_G during the read phase.

It is important to comment on a peculiar characteristic which is common to both the programming cycles in Fig.5.7. During the erase phases the red curves representing the trapped electrons present an unexpected behaviour: after a first decrease, due to the recombination with the holes injected from the silicon into the nitride layer, they start to slowly increase. This is undesirable because it reduces the number of positive trapped charge inside the nitride layer during the erase phases and consequently the memory window of our devices. Physically this phenomenon is associated with a parasitic tunneling flux of electrons coming from the polysilicon gate to the trapping layer. This can be understood looking at the band diagram of the devices along the y-direction, which is shown in Fig.5.9 for the TFET-based memory. The height of the potential barrier for the electrons in the polysilicon gate is not high enough to prevent a tunneling mechanism into the silicon nitride, but it is interesting to notice that the barrier height for the holes is higher enough to prevent an undesired holes flux during the programming phase. The best solution to this problem would be the substitution of the top oxide silicon dioxide layer with another material; this material should be characterized by an higher band gap (with respect to SiO_2) in order to provide an higher barrier to the passage

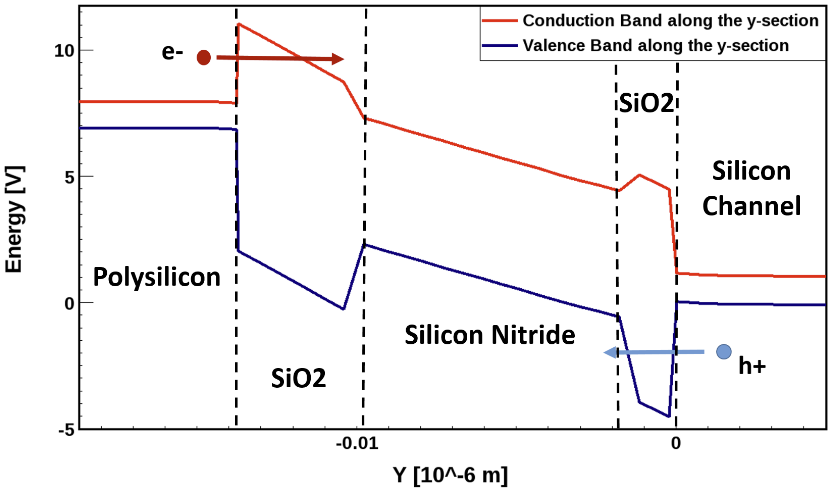


Figure 5.9 – Energy band diagram of the TFET-based memory. All the different layers are indicated. The desired hole-direct-tunneling process and the parasitic electron contribution are depicted.

of electrons from the polysilicon gate to the nitride layer.

Because of the limited number of gate-oxide material, in the next sections it has been decided to keep the same gate stack, but to move from an n-type to a p-type configuration. In this way, during the erase cycle, the parasitic tunneling of holes from the polysilicon to the nitride is completely negligible thanks to the high energy barrier for holes in SiO_2. On the other hand, during the programming phase, the big amount of holes injected in the nitride from the channel makes negligible the parasitic electron contribution from the polysilicon gate.

5.2.1 Vertical p-type TFET architecture

The goal of this section is to optimize the TFET used in the memory device, starting from the results obtained in section 5.2. As discussed previously, p-type device will be ideal for the O-N-O gate stack simulated. However the most significant modification with respect to the previous part has been the replacement of the horizontal TFET with a vertical structure [117] with embedded source architecture as shown in Fig.5.10

Since the tunneling direction is is now aligned to the component of the electric field modulated by the gate, this device offers an optimal control of the carrier injection mechanism and a strong modulation on the drain current levels. This solution has already proved to show steep sub-threshold swing and relatively high on-current levels [117]. To further improve the I_{ON} current and the sub-threshold swing an hetero-junction has been chosen for the source/channel interface. Contrary to what is usually proposed for the horizontal configuration, in this case the source is made of silicon and the part of the channel in contact with

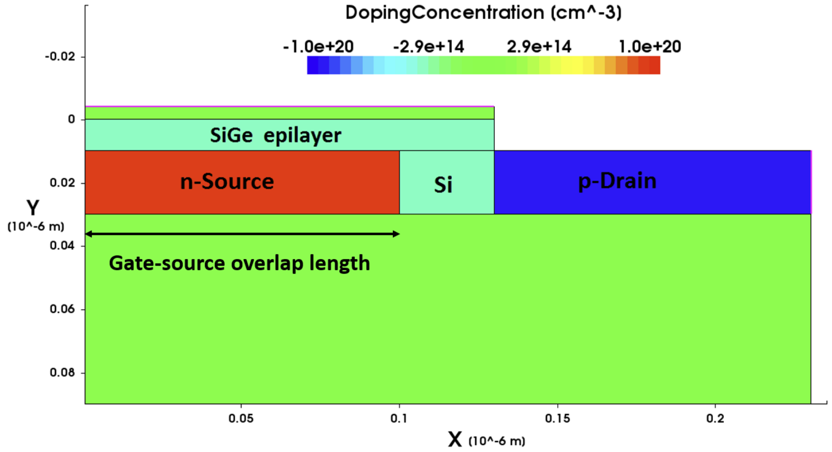


Figure 5.10 – Schematic cross section of the p-type vertical TFET architecture used in this study.

the source is made of Silicon Germanium($Si_{0.7}Ge_{0.3}$). This idea was proposed by Kathami et Banerjee [118] and represents an interesting way to increase both the I_{ON} current and the steepness of the subthreshold swing; by taking advantage of the small bandgap of SiGe.

The source doping was kept at $10^{20} cm^{-3}$, the channel area and the SiGe part was kept undoped ($10^{15} cm^{-3}$,) and the drain was lowly doped at $10^{19} cm^{-3}$. The choice of a lower drain doping is needed to reduce the unwanted ambipolar behavior of TFET. Moreover this configuration offers an important advantage with respect to the traditional SiGe/Silicon hetero-junction device. The fabrication flow is much easier because the SiGe epitaxial layer can be grown directly on top of the Silicon surface.

A set of simulations were focused to investigate the role of the doped source-gate overlap length as shown in Fig.5.10. Three different overlap lengths have been simulated using Sentaurus TCAD for both n and p-type devices. The simulation results (Fig.5.11) proves that maximizing the source length under the gate can boost device performance in terms of I_{ON}, sub-threshold slope and ambipolar behaviour. It is easy to understand that the higher saturation current is due to the larger overlap area between the silicon source and the SiGe channel. It is interesting to notice that the I_{ON} is directly proportional to the source length: doubling the source length will double the saturation current. The longer overlap area also explains the steeper sub-threshold swing of the device. The reduced ambipolar behaviour can be explained observing that if the source region is closer to the channel/drain interface then the channel modulation for positive gate bias is made more difficult by the source bias: the fixed energy bands of the source region contrast the channel modulation induced by the varying gate bias.

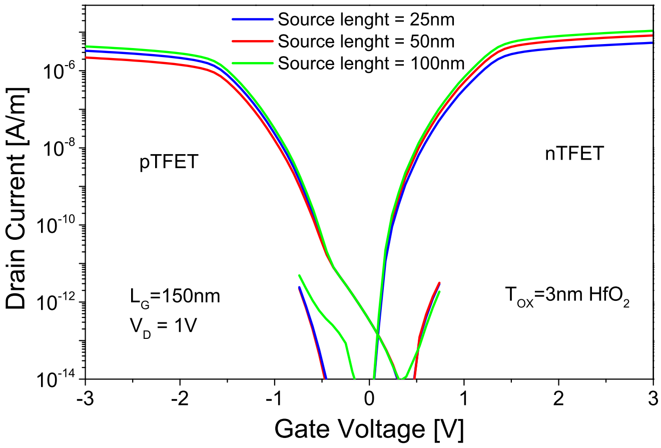


Figure 5.11 – Transfer characteristic of the three vertical n & p-type TFETs with varying gate-source overlap length. The source is grounded and $V_D = 1V$. Note that the device with the longer source length offers better performances in terms of I_{ON} current, sub-threshold slope and ambipolar behavior

5.3 Vertical TFET non-volatile memory device

In section 5.4, the vertical p-type TFET architecture and the SONOS structure are combined together to create an optimized non-volatile memory device. The final structure, built using Sentaurus Structure Editor [11], is shown in Fig.5.12.

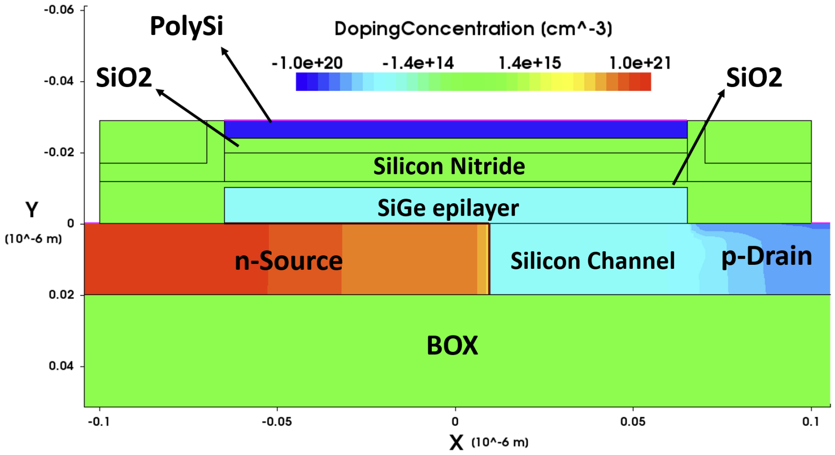

PolySi
DopingConcentration (cm^-3)
-1.0e+20
-1.4e+14
1.4e+15
1.0e+21
SiO2
SiO2
Silicon Nitride
SiGe epilayer
n-Source
Silicon Channel
p-Drain
BOX
Y
(10^-6 m)
-0.06
-0.04
-0.02
0
0.02
0.04
-0.1
-0.05
0
0.05
0.1
X (10^-6 m)

Figure 5.12 – Schematic cross section of the final device. All the different layers are indicated (except for the oxide spacers)

The doping schemes are kept same as discussed in the previous section. In both cases a Gaussian profile has been to have more realistic results. The gate stack is exactly the same discussed in section 5.3. In principle, since the device has a p-type configuration, the programming scheme of the memory should be changed with respect to the n-type device: the programming voltage should be $V_P = -9V$ and the erase voltage should be $V_E = 8V$. During the programming phase, electrons are trapped inside the nitride layer instead of holes and the programmed ID-VG curve is shifted towards low negative gate voltages. Thus the reading phase can be carried out at lower voltages (lower power consumption) compared to the nTFET case presented in section 5.3.

5.3.1 Simulation of memory operation

Table 5.2 – Parameters of the device simulated.

Parameter	Value	Parameter	value
L_G [nm]	130	T_{SiGe} [nm]	10
N_{source} $[cm^{-3}]$	10^{20}	$L_{overlap}$ [nm]	100
N_{drain} $[cm^{-3}]$	10^{19}	T_{tunn_ox} [nm]	2
N_{PolySi} $[cm^{-3}]$	10^{20}	T_{trap_layer} [nm]	10
T_{Si} [nm]	20	$T_{block_{l}ayer}$ [nm]	4

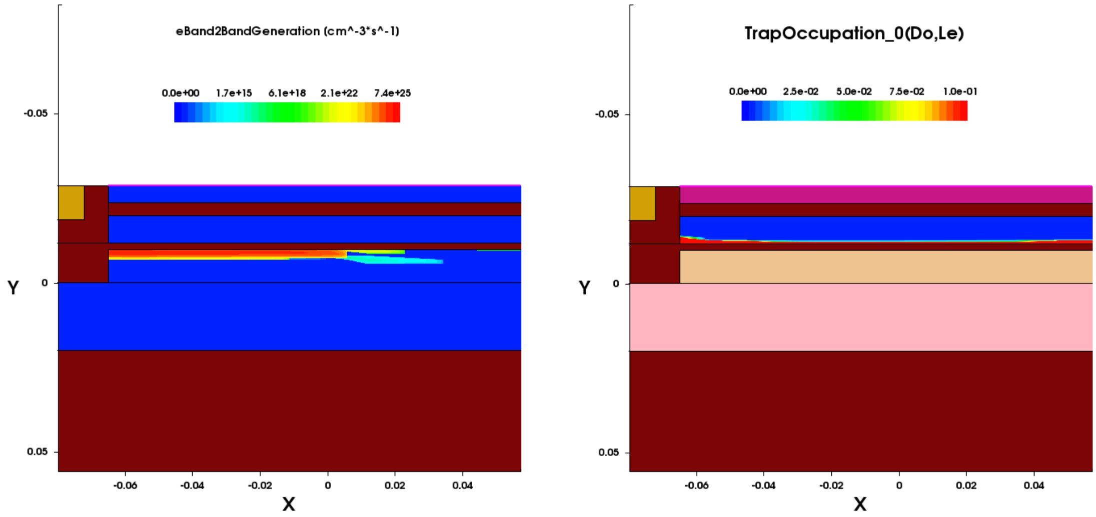


Figure 5.13 – TCAD simulation result showing in (a) the electron band-to-band generation at the end of the programming step and (b) the trapped electron concentration close to the interface between the tunneling oxide and the charge trapping layer.

The simulation of two programming cycles for a pTFET as well as a pMOS is shown in Fig.5.14(a) and (b). It is interesting to notice in Fig.5.14(a), the increased hole trapped charges in the case of pTFET at the same programming voltage, due to the boost by band-to-band tunneling. In Fig.5.14(b), we see how the hole band-to-band tunneling generation is boosting the total number of holes injected in the trapping layer: in fact the red and the green lines have similar trend. From Fig.5.13(a) we can see that the band-to-band generated electrons at the end of the programming state is just close to the tunneling oxide and hence can easily tunnel through. Also it can be seen from Fig.5.13(b), how the trapped charges are more where the band-to-band generation is higher, clearly showing that the B2B generated carriers are boosting the memory operation.

Table 5.3 – Summary of programming scheme for n and p-type devices.

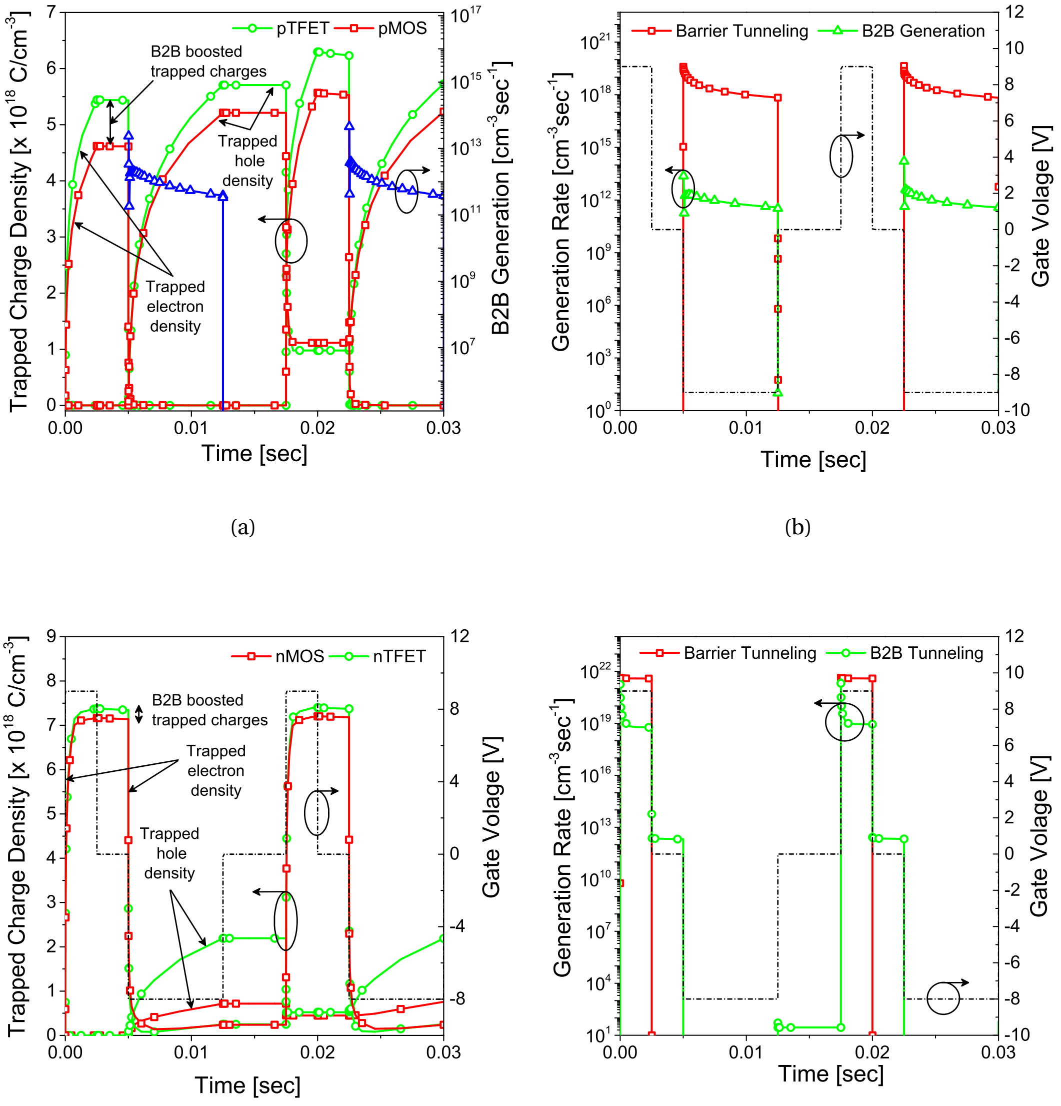
n-type	p-type
$V_{prog} = 9V$	$V_{prog} = -8V$
$V_{erase} = -9V$	$V_{erase} = 9V$
$V_{hold} = 0V$	$V_{hold} = 0V$

A similar trend is seen in the case of nTFETs and nMOS as shown in Fig.5.14(c) and (d). Once again it is clearly evident how the band-to-band tunneling in the TFET is boosting the trapped electron density at the same programming voltage. This increased trapped charges ultimately translates into a bigger memory window, when the memory cell read after a write or erase operation. When the gate voltage is swept from 0 V to $+8(-9)V$, the two curves present a spike which is due the opening of a wide B2B tunneling window at the source/channel interface. The large number of electrons(holes), generated in the SiGe via band-to-band tunneling, is available to tunnel through the bottom oxide into the nitride layer. As soon as the concentration of trapped electrons(holes) increases, the negative(positive) charge inside the nitride layer has the effect of screening the positive(negative) voltage applied to the gate contact and consequently reduces the tunneling window both at the source/channel interface and at the channel/bottom oxide interface. This explains the decreasing trend for both the barrier tunneling curves even if the gate voltage is fixed at $+9(-8)V$.

During the reading phase, the gate bias is swept from zero 0 V to +(-)6 V for n(p)-type, the source is grounded and the drain is fixed at +(-)1.0 V. The I_D–V_G curves of program and erase states are plotted in Fig.5.15. As can be seen, the performances of the new device are improved with respect the the TFET-based memory of section 5.3. Thanks to the vertical architecture enhanced with the hetero-structure Si/SiGe, the saturation current is increased of more than three orders of magnitude: the horizontal TFET allowed a saturation current $I_{ON} \geq 10-10$A, the vertical presents $I_{ON} \geq 10-7$A. It is important to notice that the MOSFET structure is characterized by a saturation $I_{ON} \geq 10-5$.

On the other hand, the sub-threshold swing is partially improved: while the horizontal TFET configuration had SS ≈ 350 mv/dec, the sub-threshold slope has now an absolute value SS ≈ 115 mV /dec. So, it was not possible to go below the MOSFET limit of 60 mV/dec with the TFET based memory and that the traditional SONOS showed better sub-threshold swing (SS ≈ 105 mV /dec).

The most significant result is that the modified programming scheme and the bigger number of trapped holes due to the band-to-band-tunneling generation (boosting effect) cause a threshold voltage shift of 2.6V, which is higher of the corresponding value (2V) for the traditional MOSFET based SONOS flash memory.



(a) (b)

(c) (d)

Figure 5.14 – Trapped electron and hole charge in nitride region of the memory device as a function of time during the program and erase cycles; corresponding applied gate bias is also shown (a). Hole band-to-band-tunneling generation and hole barrier-tunneling generation as a function of time during the programmed erase cycles (b).

5.3.2 Simulation of Retention characteristics

We can also simulate the retention capabilities of the presented vertical tunneling Tunnel FET based flash memory and compare it to a conventional MOSFET based solution. Figure 5.16 shows this comparison. As expected from Fig.5.14 (a) & (c), as the Tunnel FET has

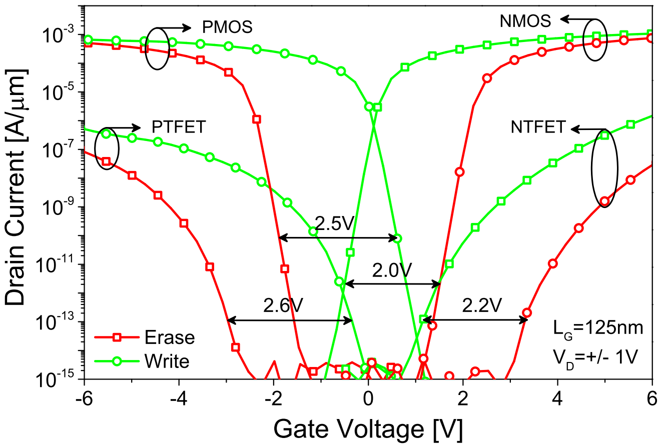


Figure 5.15 – $I_D - V_G$ curves for program and erase states for both n & p type vertical TFET and MOSFET non-volatile memory.

more trapped charge at the same programming voltage (due to the B2B tunneling boost), the retention period is expectedly longer compared to MOSFET as well. This is another possible advantage of using a vertical tunneling Tunnel FET solution for flash memories

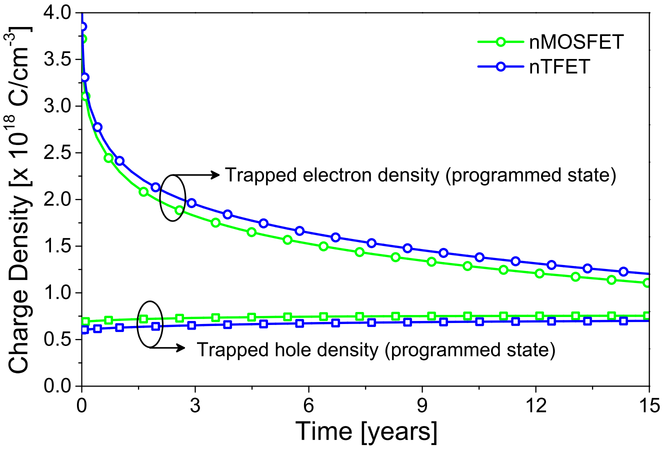


Figure 5.16 – Comparison of retention characteristic of a n-type MOSFET and n-type Tunnel FET based SONOS flash memory.

5.4 Fabrication of TFET non-volatile memory

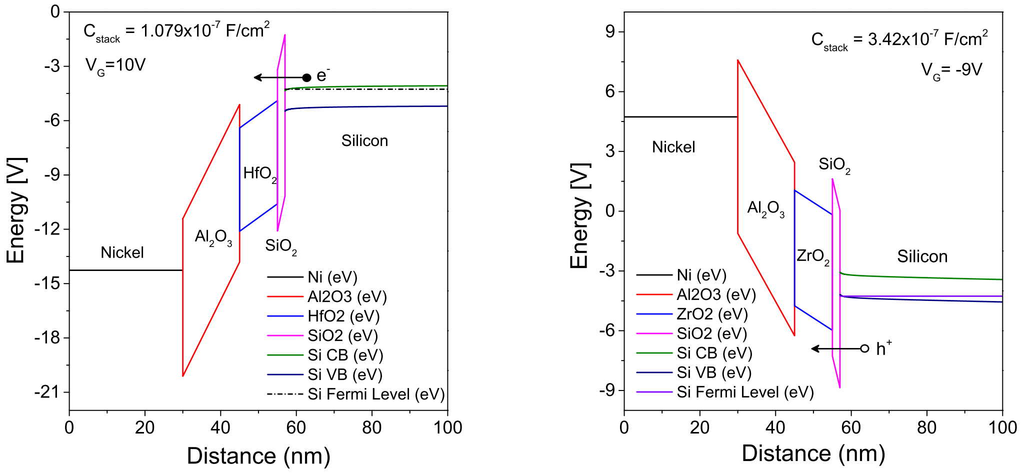
The above described concept was fabricated in the CMi clean room in EPFL. Due to the vertical nature of the design, the lateral doping profiles do not play a important role in the saturation current and sub-threshold slope of the device. The most important tunneling junction is the interface between the Silicon source and the Silicon-Germanium channel region. As the epitaxially grown after the source is doped, so we should have a shaper junction at the interface if the thermal budget is kept low after the doping.

5.4.1 Evaluation of gate-oxide stacks

The first step to have a working non-volatile memory is to first evaluate the gate-oxide stack to be used on top of the active device. For this project it was decided to use high-k gate oxide stacks (Fig.5.18(a)) instead of traditional O-N-O stack due to the following reasons:

- High-k dielectrics like HfO_2 ($\epsilon_r = 20$), ZrO_2($\epsilon_r = 25$) and TiO_2 ($\epsilon_r = 80$) have higher charge trapping capability than a Silicon Nitride layer. (references)
- Using a high-K dielectric instead of a Silicon Nitride layer ($\epsilon_r = 7$) will boost the sub-threshold slope of the final device to possible sub-60 mv/decade.

In order to evaluate the order of the oxides to be used and their required thickness, firstly a tool was used to simulated the energy band diagrams. Secondly standard MOS-capacitors were also fabricated with the desired stack. Capacitance-voltage measurements were performed on the fabricated MOS-caps and the results are presented below.



(a)

(b)

Figure 5.17 – (a)Simulated energy band diagram of a $SiO_2/HFO_2/Al_2O_3$ stack at $V_G = 10V$(b) Simulated energy band diagram of a $SiO_2/ZrO_2/Al_2O_3$ stack at $V_G = -9V$

Simulation of gate-oxide stack

A software tool was used to simulate the energy band diagram of the desired gate stacks. As shown in Fig.5.17(a), the first stack was simulate at $V_G = 10V$. At this biasing condition, which mimics the programming condition in a n-type device, electrons from the Silicon conduction band can tunnel into the charge trapping layer. In should be noted that, only some particular order of the gate oxides can function as a non-volatile flash memory. The energy band gap of the charge trapping layer should be the least and the energy band gap of the blocking layer should be the largest. Also important is band edge alignment between the three oxides. Fig. 5.17(b), shows the second stack at $V_G = -9V$. In this condition, which mimics the erase condition in a n-type device, holes from the Silicon valence band can tunnel through the tunneling oxide into the charge trapping layer.

Fabricated MOS capacitors

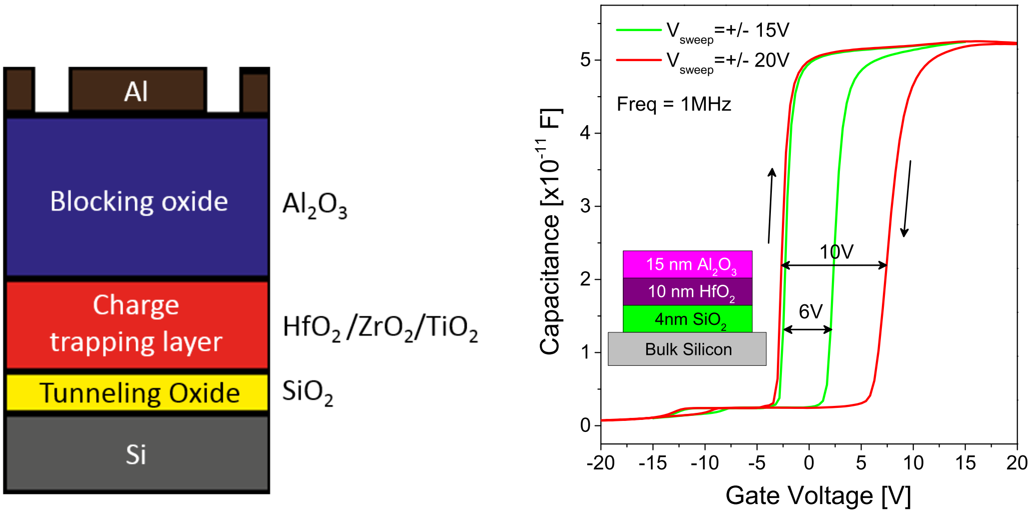
MOS capacitors were fabricated starting from bulk Silicon wafers. The gate oxide stack was deposited on bulk wafers via Atomic Layer Deposition (ALD) after a standard RCA cleaning step. Finally a 300nm of Aluminium was deposited and patterned using optical lithography to form 100μmx100μm square pads. The post lithography Aluminium etch was carried out with reactive ion etching (RIE). Three different stacks were fabricated for a comparison:

- 4nm SiO_2 /10nm HFO_2 /15nm Al_2O_3
- 4nm SiO_2 /10nm ZrO_2 /15nm Al_2O_3
- 4nm SiO_2 /10nm TiO_2 /15nm Al_2O_3

The fabricated MOS capacitors were were measured with a LCR bridge to determine the hysteresis. As shown in Fig.5.18(b) for a $SiO_2/HfO_2/Al_2O_3$ gate oxide stack, the hysteresis clearly depends on the sweep voltages. For a sweep voltage of +15V to -15V, a 6V hysteresis was observed. On increasing the sweep voltages to +20V to -20V, the hysteresis increased to 10V. A higher sweep voltage causes more tunneling to occur, hence more charges are trapped in the charge trapping layer causing a bigger hysteresis.

Further measurements were done on wafers with $SiO_2/ZrO_2/Al_2O_3$ oxide stack. Since ZrO_2 and HfO_2 have similar relative permittivity ($\epsilon_r = 25$) and energy band gap values ($5.8eV$), we get a similar hysteresis value of 6V for a sweep voltage of +15V to -15V as shown in Fig.5.19. A series of measurements were done on the same stack keeping all conditions same. It was observed that the hysteresis window was reduced to 4V at the end of the sixth run. This could be due to the following reasons.

- Trapped charges in the oxide interfaces. However the capacitance measurements were done at a relatively high frequency of 1MHz, so its unlikely that the traps were active during the measurements.



(a) (b)

Figure 5.18 – (a)Schematic of cross-section of the fabricated MOS capacitors. (b) Measured capacitance voltage characteristics of a $SiO_2/HfO_2/Al_2O_3$ gateoxide stack at 1MHz frequency. A hystersis of 10V is observed for a gate votage sweep of -20V to 20V.

- After the lithography step to define the MOS capacitors, only the metal layer was etched, keeping the entire oxide stack blanket throughout the wafer. Hence it is possible that the erase operations were not entirely driving out the trapped charges.

To verify this, the wafers were further etched down to the bulk Silicon layer. Further repeated measurements were done and the hysteresis was not found to change any more. This confirmed the need to etch down the entire gate stack, to achieve a reliable erase operation when the final device is made.

5.4.2 Process flow

The process flow was designed keeping in mind the observations made during the MOS capacitor measurements. Also the advantage of fabricating the vertical Tunnel FET device of Fig.5.12 is that the lateral doping decay is not critical. As we know band-to-band tunneling is very sensitive to the doping profiles [37], with less than 10nm/decade required for optimal device design in case of lateral devices. Having a vertical structure, eliminates this problem as the channel in grown epitaxially over a doped source. However epitaxial growth over a highly doped surface poses its own problems as we will see later. The process flow is shown in Fig.5.20 below.

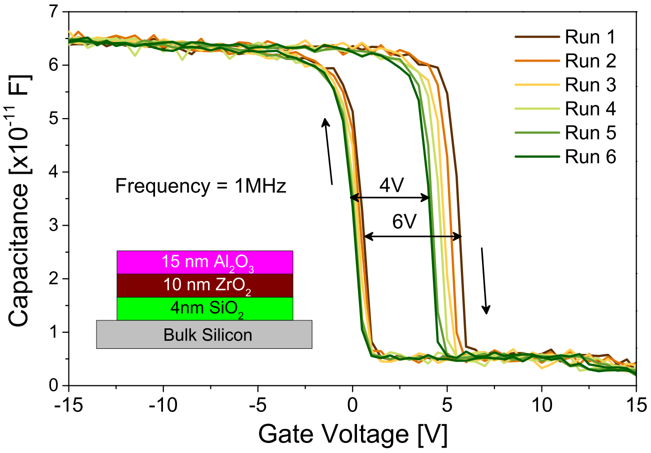


Figure 5.19 – Measured capacitance voltage characteristics of a $SiO_2/ZrO_2/Al_2O_3$ gate-oxide stack at 1MHz frequency. A reduction of the hysteresis window was observed on repeated measurements.

Wafer preparation

We can start from a 100nm SOI wafers (SIMBOND) with 400nm BOX layer. The first step is to thin down the Silicon active layer to a desired thickness of 40nm (T_{Si}). For this purpose sacrificial dry oxidation was employed as shown in Fig.5.20(b)

First implantation

After the wafer has been thinned down, the first lithography step will be done. Photolithography will be employed to to open 100μmx100μm implantation windows in the thermally grown oxide. The patterned oxide will be removed by a BHF dip. Ion implantation will then be performed with Phosphorus ions at 30KeV energy and 1x10^{15}$at.cm^{-2}$ dosage through a 15nm screening LTO layer.

Second implantation

Since for a tunnel FET, we need both n and p implanted regions, a second implantation needs to be done. For this firstly the implanted wafer has to be stripped off the remaining oxide. Then, a further 200nm LTO will be grown on top of the Silicon to act as the hard mask for the second implantation. Ion implantation will again be performed with Boron ions at 10KeV energy and 1x10^{15}$at.cm^{-2}$ dosage also through a 15nm screening LTO layer.

SiGe epitaxial growth

After the second implantation the remaining LTO will again be stripped with a BHF dip. At this stage the wafers needs to be shipped abroad for a outsourced SiGe epitaxial step. The wafers will be cleaned at 1120°C in H_2 to remove any native oxide so that we have an atomically clean surface for the epitaxial growth to be uniform.

Active area definition

Next the active area of individual devices will be defined as shown in Fig.5.20(k). Ion beam etching was employed to etch down to the SiO_2 BOX. Next the desired gate-oxide stack will be deposited on the wafer by Atomic layer deposition (ALD).

Contact hole formation

Next the contact holes needs to be formed in the active areas with ion beam etching as shown in Fig.5.20(m). 300nm of Nickel will then be deposited with electron beam evaporation.

Metal area definition

Finally the metal layer will be patterned with another lithography step and etched with ion beam etching to define the gate, source and drain contact pads as shown in Fig.5.20(p). To improve the contact between the metal and Silicon in the contact areas a low temperature anneal will be performed at 450°C for 1min. This will lead to the formation of $NiSi$ at the contacts.

5.4.3 Results obtained

We started processing 6 identical SOI wafers at the same time. Both implantations were carried as described above till the step shown in Fig.5.20(g). At this stage the top Silicon surface looked like as shown in Fig.5.21.

Further processing of this wafers is still under way. More tests are needed to be done to have a atomically clean surface before the epitaxial step.

5.5 Summary

A brief overview of flash memories have been presented. The differences between a floating gate type and SONOS flash memory and the later's advantages have been highlighted. The operating principle of floating gate memories are also discussed.

The simulation methodology employed to simulate the write and erase operations in TCAD are discussed. Extensive TCAD simulations were done firstly to optimize a vertical tunneling TFET. This optimized Tunnel FET was then simulated with a oxide-nitride-oxide stack to demonstrate the working of a flash memory. Simulations were also done on a conventional MOSFET structure to compare the performance with respect to the Tunel FET.

It was shown that the vertical Tunneling FET showed a larger memory window compared to the conventional MOSFET with the same programming and erase voltages. It was also shown the contribution of band-to-band generated carriers boosting the trapped charges in the charge trapping layer.

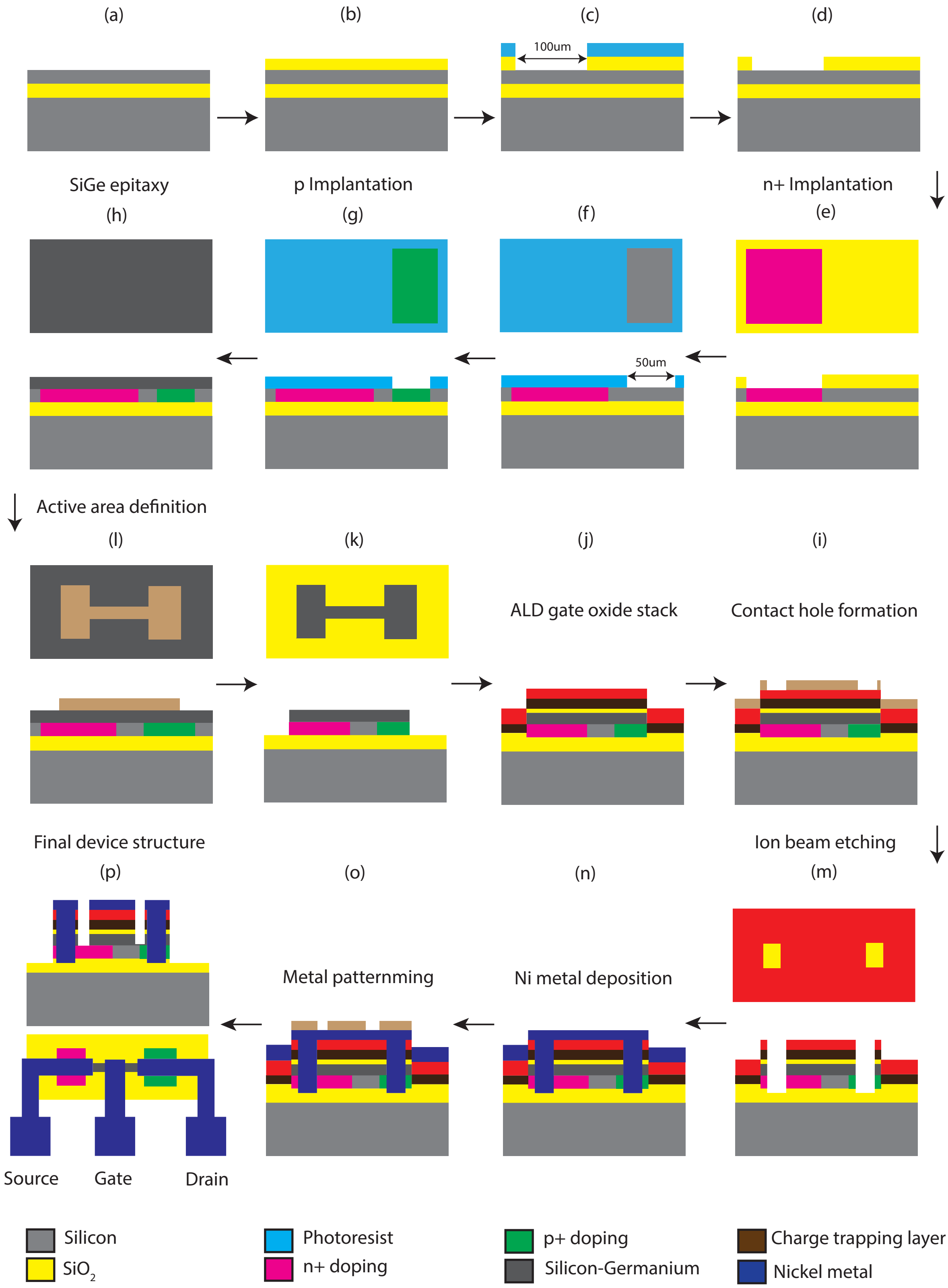

(a)
(b)
(c)
(d)
100um
SiGe epitaxy
p Implantation
n+ Implantation
(h)
(g)
(f)
(e)
50um
Active area definition
(l)
(k)
(j)
(i)
ALD gate oxide stack
Contact hole formation
Final device structure
Ion beam etching
(p)
(o)
(n)
(m)
Metal patternming
Ni metal deposition
Source
Gate
Drain
Silicon
SiO_2
Photoresist
n+ doping
p+ doping
Silicon-Germanium
Charge trapping layer
Nickel metal

Figure 5.20 – Process flow of proposed vertical tunneling TFET based non-volatile memory. (image not to scale)

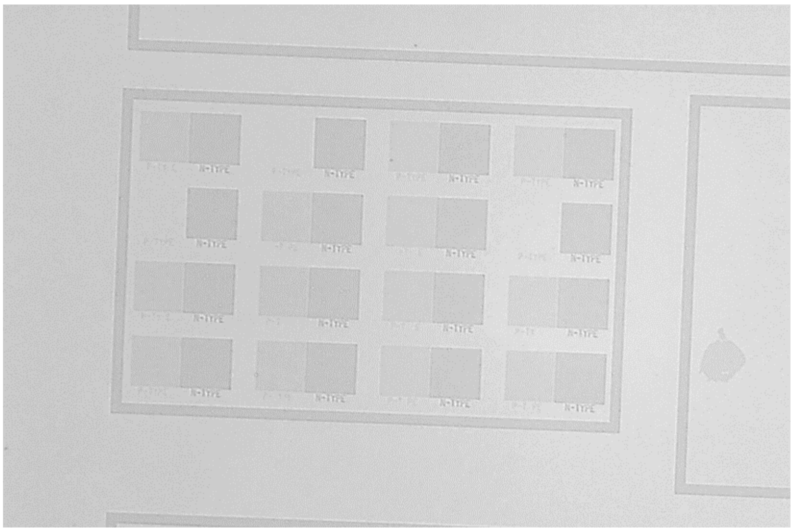


Figure 5.21 – Optical microscope view of the wafer surface after the two implantations. The n and p type implantations are clearly visible due to the difference in colour.

6 Conclusions

In chapter 1 we have seen a brief overview of the current microelectronic technology, the problems associated with scaling the conventional MOSFET and possible alternatives that can circumvent the problem. In chapter 2, we have presented the basics of the tunneling theory. The various tunneling models available in commercial simulators are explained and the differences between them are highlighted. An algorithm to calibrate the non-local tunneling model in a commercial TCAD device simulator to match measurement data has been shown. The model parameters have been adapted based on available measurement data of SOI TFET devices. Reported simulations with fitted parameter set have good level of agreement with measured data at different bias conditions for both transfer and output characteristics of the TFET devices. The dependence of the tunneling generation rate on different crystal directions was shown. We have also proposed and simulated a new device architecture which provides higher current (159 times) as well as lower C_{MILLER} resulting in faster operation (38 times).

In chapter 3, Based on conformal mapping method, we have developed a simple physics-based DC-current model for DG-TFETs taking into account the non-local B2B tunnelling and in-series inversion MOS region. The limitations of the model for low source doping have been successfully addressed. The model agrees well with simulations for different device parameters and addresses the present accuracy challenges of Tunnel FET modelling in all regions of operation. We have developed a physics-based compact model for DG-TFETs describing all tunnelling related characteristics and implemented it in Verilog-A for circuit simulations. The model agrees well with TCAD simulations in all regions of operation. Based on capacitance measurements a first order capacitance model for DG-TFETs has been proposed. The model matches numerical simulations well and shows the same behaviour as measurements. Using the model to perform basic circuit simulations has revealed some unique features of a TFET. Finally I have shown a systematic benchmarking study of analog figures of merit of 28nm sSi TFETs versus 28nm CMOS based on a universal compact model at both device and circuit levels. Our study suggests new design and performance space, especially due to the higher TFET gain at very low current levels ($1pA/\mu m$ to $10nA/\mu m$), and a higher temperature stability for key analog IC functions.

In chapter 4, we have presented a new application of Tunnel FETs as a memory cell. Simulation results on a new asymmetric TFET structure that can be used as a capacitorless DRAM. We have also experimentally demonstrated a scalable implementations of such 1T capacitorless DRAM cell based on DG Tunnel FET with a ΔI_S of $10nA$ and with retention times in the order of 100's of microseconds to few milliseconds, at room temperature in devices with channel length varying from $400nm$ to $1\mu m$. The discharge cycle was shown to be unaffected when the temperature varies between 25°C to 85°C. We have also validated a new scalable 1T capacitorless DRAM cell based on DG fin Tunnel FET structure with a doped pocket, which opens a new field of applications for TFETs. A clear DRAM memory operation has been proposed and verified with simulations. Charge relaxation times in the order $100\mu sec$ is shown as not degrading at high temperature (85°C) which is one of the advantages for TFETs based memory cells. Finally a process flow for fabrication of independent gate fin tunnel FETs on SOI substrate is presented. We demonstrate successfully the use of e-beam resist as a implantation mask for a p-type species. We see three orders magnitude difference between the implanted and un-implanted regions of the implanted wafer. Also the difference to the un-implanted region and un-implanted wafer is an order of magnitude and could be due to interfering bulk-conduction or due to resistivity difference on different parts of Silicon prime grade wafers.

In chapter 5, we presented another application of Tunnel FET as a non-volatile memory. A vertical tunneling TFET structure was used to demonstrate via simulations that band-to-band generated carriers can boost the non-volatile memory operation. Also a process flow was designed to fabricate this device in CMi EPFL. MOS caps with various gate stacks were fabricated and the hysteresis was measured in order to find the right combination of gate-oxides.

6.1 Future perspectives

It has been shown both theoretically and experimentally that tunnel FETs are among the most promising candidate amongst emerging devices in the roadmap of energy efficient beyond CMOS devices. The major technological challenge with Tunnel FETs has been the lack of a sub 60mV/dec slope at room temperature and sufficiently high on current simultaneously. Some Tunnel FET specific characteristics which degrade the circuit performance of TFET-based circuits requires further consideration in device level optimization. So far, heterojunction TFETs (III-V with Silicon) appears to be most promising for high on current and small sub-threshold swing attributes are also advantageous as they allow independent optimization of drain and source junctions and also reduces Miller effect. In order for TFETs to become attractive alternatives to MOSFET or atleast co-exist with MOSFET in hybrid-logic, the technological challenges have to be overcome.

There is plenty of room for improvement for Tunnel FETs in technological field as discussed. Concerning the modelling work done in this theses, the following points can be pointed out.

- From a compact modelling point of view, there is definitely work needed to further improve upon what is presented in this theses. Firstly, the presented model is suited only for homo-junction devices. However as pointed out, hetero-junction devices are the most promising from a technological point of view. So the model has to be suitably modified to take into account hetero-junctions.
- The Verilog-A code based compact model developed as a part of this theses also needs work. In its current state the model can be called a "level-zero" model as it takes into account only the tunneling related physics. More functionality like leakage mechanisms, multiple geometry support, a noise model could be added step by step. Also a comprehensive algorithm for calibration to measurements or simulations is needed.
- The capacitance-voltage model presented in this thesis is only valid for low injection devices like in Silicon. As already pointed out in chapter 3, the approximation of the inversion charge being entirely coming from the drain will fall apart in case of higher injection levels like in III-V hetero-junction devices. Hence the model has to be suitably modified to accommodate the increase in contribution from source to the inversion charges in case of higher injection devices.

From the application oriented results discussed in this thesis, we can think of the following improvements that can further improve the state of the art. Firstly for the Tunnel FET based 1T-DRAM concept:

- The TFET based 1T-DRAM cell can be further optimized to boost the tunneling current and the difference between the two memory states (ΔI_S).
- The programming conditions can be further optimized as well to boost the ΔI_S and the retention time. A detailed study of the memory cell disturb conditions in an array configuration and its variation with temperature is needed.
- The doped pocket based 1T-DRAM can be studied in more details to determine the optimized doping in the pocket, the position of the pocket with respect to the gate and the length of the pocket. Also the scaling needs to be studied if the principle still works at much smaller gate length.
- Regarding the fabrication of the independent gate finTFET devices many obstacles were faced in order to realize the desired gate separation. The process flow needs to scrutinised from scratch in order to circumvent some of the problems faced in the fabrication.

Regarding the design of the vertical tunneling based non-volatile memory concept the following can be said on its future perspectives:

- Similar to the capacitor-less memory concept, here as well the device concept can be optimized through TCAD simulations in order to maximize the tunneling current as well as the band-to-band tunneling generated carrier's contribution to the charges tunneling through the tunneling oxide to the charge trapping layer. The goal is to achieve a larger memory window with the lowest possible programming voltage.
- The gate stack needed for a flash memory are different for n and p type devices. Further experimental results are needed to choose the right combination of high-k gate dielectrics and their thickness.
- From a fabrication point of view, again we faced many challenges to grow a SiGe epitaxial layer on top of highly doped Silicon layer. Further tests are needed in order to achieve an atomically clean surface before the epitaxial step.

A Calculation of parameters in section 2.4

Eqn. 2.29 is written as :

$$m_r = A\left[\frac{3qhB_{path}}{2^{7/2}\pi(E_g + D_{path})^{7/2}}\right]^2 \tag{A.1}$$

Using $B_{path} = 1.9x10^7 Vcm^{-1}$, $D_{path} = 0eV$, $E_g = 1.12eV$ at $300K$, we get $m_r = 0.05485 * m_0$ (where m_0 is the electron rest mass)

Now eqn.2.30 can be written as:

$$\frac{gD_{op}^2}{\rho} = \frac{2^{21/4}h^{5/2}m_r^{5/4}\epsilon_{op}[E_g + D_{path}]^{7/4}A_{path}}{(m_c m_v)^{3/2}[1+N_{op}]q^{5/2}} \tag{A.2}$$

Using $A_{path} = 4$ x $10^{14} cm^{-3}s^{-1}$, $\epsilon_{op} = 57.6meV$, $E_g = 1.12eVat300K$ $D_{path} = 0eV$
m_c is calculated from $1/m_c = 1/(2m_r) + 1/m_0$ with $m_r = 0.05485 * m_0$
m_v is calculated from $1/m_c = 1/(2m_r) - 1/m_0$ with $m_r = 0.05485 * m_0$
N_{op} is calculated from $N_{op} = [exp(\epsilon_{op}/kT) - 1]]^{-1}$ at $300K$ with $\epsilon_{op} = 57.6meV$

Finally using the value of g for [100] = $2 * g_c * g_v = 2 * 4 * 1 = 8$ and $\rho = 2329Kg/m^3$ for Silicon. Putting all values we extract the deformation potential $D_{op} = 1.28$ x $10^{11} eV/m$ for TO phonons in Silicon for the default values of A_{path} and B_{path}.

B Capacitance measurements

During capacitance measurements, any mutual inductance, interference of the measurement signals, and unwanted parasitic factors in the connection method will have significant effects on the measurements, especially at high frequency. The HP 4284A employs the four-terminal pair measurement configuration which permits easy, stable, and accurate measurements and avoids the measurement limitations inherent to such factors. Fig. B.1 shows the four-terminal pair measurement principle. The UNKNOWN terminals consists of four coaxial connectors.

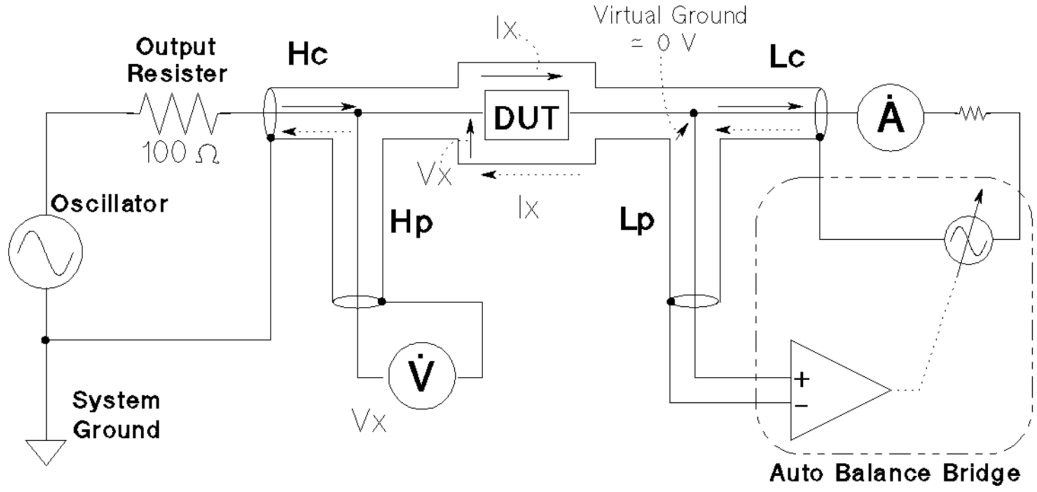

Output
Resister
100 Ω
Oscillator
System
Ground
Hc
Ix
Virtual Ground
≃ 0 V
Lc
DUT
Vx
Ix
Hp
Lp
Ȧ
V̇
Vx
+
−
Auto Balance Bridge

Figure B.1 – Four-Terminal Pair Measurement Principle in a LCR bridge.

HCUR : High current
HPOT : High potential
LPOT : Low potential
LCUR : Low current

The four-terminal pair measurement method has the advantage in both low and high impedance measurements. The outer shield conductors work as the return path for the measurement signal current (they are not grounded). The same current flows through both the center conductors and outer shield conductors (in opposite directions), but no external magnetic fields are generated around the conductors (the magnetic fields produced by the inner and outer currents completely cancel each other). Because the measurement signal current does not develop an inductive magnetic field, test leads do not contribute additional errors due to self or mutual inductance between the individual leads.

This paragraph gives general notes and techniques for using the four-terminal pair configuration efficiently. To realize accurate measurements using the four-terminal pair measurement technique, the following are required to make measurement contacts (the number labels in the following description corresponds to the numbers in Fig.B.1).

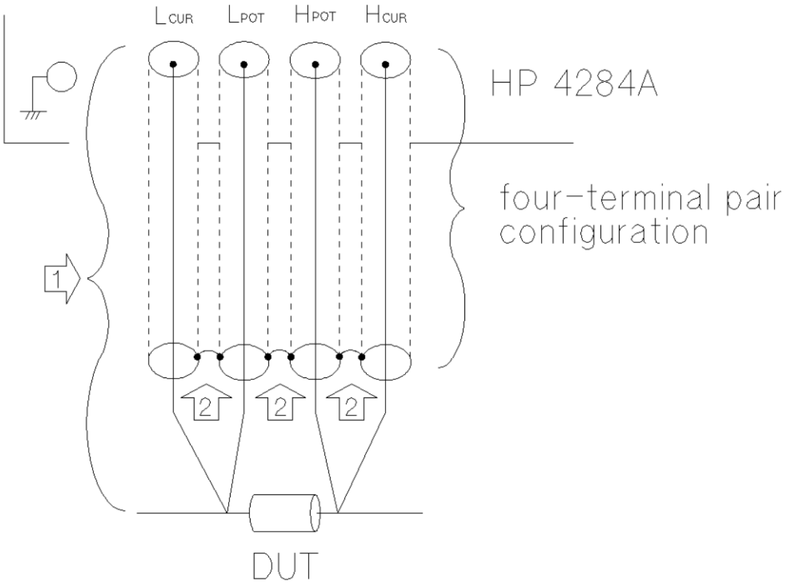


Figure B.2 – Four-terminal pair configuration measurement contacts.

- The signal path between the HP 4284A and DUT should be as short as possible.

- To construct the four-terminal pair measurement circuit configuration, the outer shields of HCUR and HPOT, LCUR and LPOT terminals must be respectively connected together at the point as near as possible to the point at which the DUT will connected.

- Keep connections between the point at which the shielding ends and DUT as short as possible.

B.1 Notes on trans-capacitance measurements in a FET

To measure trans-capacitance in a FET like for example C_{GS}, the low terminal of the setup shown in Fig.B.1 was connected to the gate which monitors the current and the high terminal to the source where the bias is applied according to the equation B.1 below:

$$C_{GS} = -\frac{\partial Q_G}{\partial V_S} | V_G, V_D, V_{SUB} \tag{B.1}$$

Hence in the above example the two low terminals are connected to the source terminal, whereas the two high terminals are connected to the gate to give us C_{GS}. However the definition also says that V_{DS} should remain constant. To ensure this the drain bias should be synchronized to the source bias in order to get a constant V_{DS} and hence preventing any current flow during capacitance measurements. An easier alternative would be to measure C_{SG} instead. Rewriting the above definition for C_{SG}:

$$C_{SG} = -\frac{\partial Q_S}{\partial V_G} | V_S, V_D, V_{SUB} \tag{B.2}$$

In this case as the source potential is not changing, the drain need to be synchronised to get $V_{DS} = 0V$.

C Pulsed measurement details

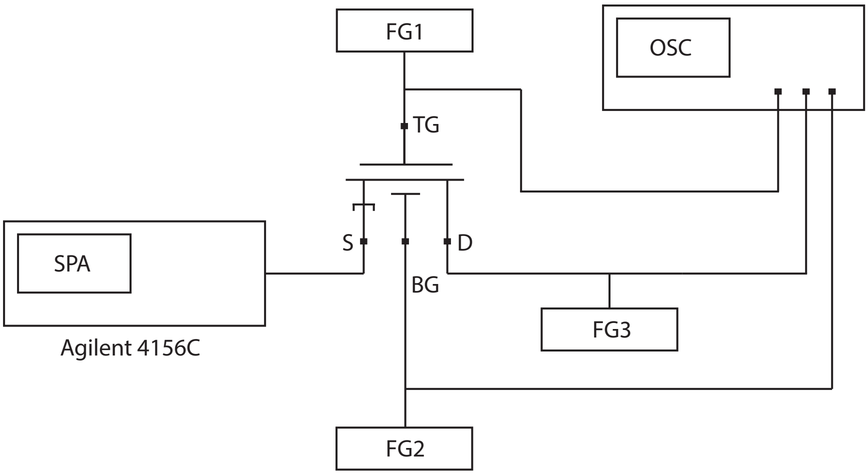


Figure C.1 – Connection schematic showing the three function generators, the parameter analyzer and the oscilloscope.

For the pulsed measurements done on the volatile memories in chapter 4, the above connection schematic was used. The three function generators were programmed individually to a have custom output pulse to replicated the write, erase and read conditions described in chapter 4. Also all the three function generators were synchronised to have the same pulse edge. The oscilloscope was used to monitor the output of all the three function generators. The current was monitored at the source terminal of the device at an interval of $60 \mu secs$ (limitation of the equipment). The parameter analyser was used in sampling mode for this purpose.

Bibliography

[1] G. E. Moore, “Cramming more components onto integrated circuits,” *Electronics*, vol. 38, no. 114, 1965.

[2] R. Dennard, F. Gaensslen, V. Rideout, E. Bassous, and A. LeBlanc, “Design of ion-implanted MOSFET’s with very small physical dimensions,” *IEEE J. Solid-State Circuits*, vol. 9, pp. 256–268, Oct. 1974.

[3] S. Zaima, “Technology Evolution for Silicon Nanoelectronics: Postscaling Technology,” *Jpn. J. Appl. Phys.*, vol. 52, p. 030001, Mar. 2013.

[4] E. Pop, “Energy dissipation and transport in nanoscale devices,” *Nano Res.*, vol. 3, pp. 147–169, Mar. 2010.

[5] A. M. Ionescu and H. Riel, “Tunnel field-effect transistors as energy-efficient electronic switches,” *Nature*, vol. 479, pp. 329–337, Nov. 2011.

[6] E. Kane, “Theory of tunneling,” *JAP*, vol. 32, no. 1, pp. 83–91, 1961.

[7] M. Fujita, “Silicon photonics: Nanocavity brightens silicon,” *Nat. Photonics*, vol. 7, pp. 264–265, Mar. 2013.

[8] L. Keldysh, “Behavior of non-metallic crystals in strong electric fields,” *Sov. J. Exp. Theor. Phys.*, vol. 6, no. 4, p. 763, 1958.

[9] P. Palestri, “Semiclassical and quantum mechanical modeling of tunnel FET devices,” tech. rep., DIEGM University of Udine, 2012.

[10] R. Landauer, “Spatial Variation of Currents and Fields Due to Localized Scatterers in Metallic Conduction,” *IBM J. Res. Dev.*, vol. 1, pp. 223–231, July 1957.

[11] Synopsys, “Sentaurus TCAD user guide Ver: I2014.09,” tech. rep.

[12] Silvaco, “Atlas user’s manual,” tech. rep., 2014.

[13] G. Hurkx, “On the modelling of tunnelling currents in reverse-biased p-n junctions,” *Solid. State. Electron.*, vol. 32, pp. 665–668, Aug. 1989.

[14] 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.

[15] S. M. Sze, *Physics of Semiconductor Devices.* Wiley, 1981.

[16] K.-H. Kao, A. S. Verhulst, W. G. Vandenberghe, B. Soree, G. Groeseneken, and K. De Meyer, “Direct and Indirect Band-to-Band Tunneling in Germanium-Based TFETs,” *IEEE Trans. Electron Devices*, vol. 59, pp. 292–301, Feb. 2012.

[17] A. Chynoweth and R. Logan, “Phonon-assisted tunneling in silicon and germanium Esaki junctions,” *Phys. Rev.*, vol. 125, no. 1959, 1962.

[18] C. Rivas, R. Lake, G. Klimeck, W. R. Frensley, M. V. Fischetti, P. E. Thompson, S. L. Rommel, and P. R. Berger, “Full-band simulation of indirect phonon assisted tunneling in a silicon tunnel diode with delta-doped contacts,” *Appl. Phys. Lett.*, vol. 78, no. 6, pp. 814–816, 2001.

[19] W. Vandenberghe, A. Verhulst, G. Groeseneken, B. Soree, and W. Magnus, “Analytical model for point and line tunneling in a tunnel field-effect transistor,” in *Int. Conf. Simul. Semicond. Process. Devices, SISPAD 2008.*, vol. i, pp. 137–140, IEEE, 2008.

[20] Y. Zhu and M. K. Hudait, “Low-power tunnel field effect transistors using mixed As and Sb based heterostructures,” *Nanotechnol. Rev.*, vol. 2, Jan. 2013.

[21] D. B. Abdi and M. Jagadesh Kumar, “Controlling Ambipolar Current in Tunneling FETs Using Overlapping Gate-on-Drain,” *IEEE J. Electron Devices Soc.*, vol. 2, pp. 187–190, Nov. 2014.

[22] G. B. Beneventi, E. Gnani, A. Gnudi, S. Reggiani, and G. Baccarani, “Can interface traps suppress TFET ambipolarity?,” *IEEE Electron Device Lett.*, vol. 34, no. 12, pp. 1557–1559, 2013.

[23] F. Mayer, C. Le Royer, J. F. Damlencourt, K. Romanjek, F. Andrieu, C. Tabone, B. Previtali, and S. Deleonibus, “Impact of SOI, Si1-xGexOI and GeOI substrates on CMOS compatible tunnel FET performance,” in *Tech. Dig. - IEDM*, vol. 4, pp. 1–5, IEEE, Dec. 2008.

[24] J.-S. Jang and W.-Y. Choi, “Ambipolarity Factor of Tunneling Field-Effect Transistors (TFETs),” *JSTSJournal Semicond. Technol. Sci.*, vol. 11, pp. 272–277, Dec. 2011.

[25] L. De Michielis, L. Lattanzio, and A. M. Ionescu, “Understanding the Superlinear Onset of Tunnel-FET Output Characteristic,” *IEEE EDL*, vol. 33, pp. 1523–1525, Nov. 2012.

[26] L. Esaki, “New Phenomenon in Narrow Germanium Junctions,” *Phys. Rev.*, vol. 109, pp. 603–604, Jan. 1958.

[27] M. Cotter, Huichu Liu, S. Datta, and V. Narayanan, “Evaluation of tunnel FET-based flip-flop designs for low power, high performance applications,” in *Int. Symp. Qual. Electron. Des.*, pp. 430–437, IEEE, Mar. 2013.

[28] H. Liu, S. Datta, and V. Narayanan, “Steep Switching Tunnel FET : A Promise to Extend the Energy Efficient Roadmap for Post- CMOS Digital and Analog / RF Application s Electrical Engineering Dept ., 3 Computer Science Engineering Dept ., the Pennsylvania State University,” in *Proc. 2013 Int. Symp. Low Power Electron. Des.*, vol. 2, pp. 145–150, 2013.

[29] A. Biswas, L. De Michielis, C. Alper, and A. M. Ionescu, “Conformal mapping based DC current model for double gate tunnel FETs,” in *Conf. Dig. ULIS*, pp. 85–88, 2014.

[30] P. Agopian and M. Martino, “Experimental analog performance of pTFETs as a function of temperature,” in *SOI Conf.*, 2012.

[31] P.-f. Wang, *Complementary Tunneling-FETs (CTFET) in CMOS Technology.* PhD thesis, Technischen Universität München, 2003.

[32] M. Martino, P. D. Agopian, and S. D. S. Filho, “Temperature Influence on Tunnel Field Effect Transistors (TFETs) with Low Ambipolar Currents,” *ECS Trans.*, vol. 39, no. 1, pp. 77–84, 2011.

[33] S. Mookerjea, D. Mohata, T. Mayer, V. Narayanan, and S. Datta, “Temperature-Dependent I-V Characteristics of a Vertical InGaAs Tunnel FET,” *IEEE Electron Device Lett.*, vol. 31, pp. 564–566, June 2010.

[34] F. S. Neves, M. D. V. Martino, P. G. D. Agopian, J. a. Martino, R. Rooyackers, D. Leonelli, and C. Claeys, “Temperature Influence on Nanowire Tunnel Field Effect Transistors,” *ECS Trans.*, vol. 49, pp. 223–230, Aug. 2012.

[35] T. Nirschl, P. P. Wang, W. Hansch, and D. Schmitt-Landsiedel, “The tunnelling field effect transistors (TFET): the temperature dependence, the simulation model, and its application,” in *ISCAS*, vol. 3, pp. III–713, IEEE, 2004.

[36] G. V. Luong, L. Knoll, M. J. Süess, H. Sigg, A. Schäfer, S. Trellenkamp, K. K. Bourdelle, D. Buca, Q. T. Zhao, and S. Mantl, “High on-currents with highly strained Si nanowire MOSFETs,” in *ULIS 2014 - 2014 15th Int. Conf. Ultim. Integr. Silicon*, pp. 73–76, 2014.

[37] S. S. Dan, A. Biswas, C. L. Royer, W. Grabinski, and A. M. Ionescu, “A novel extraction method and compact model for the steepness estimation of FDSOI TFET lateral junction,” *IEEE Electron Device Lett.*, vol. 33, no. 2, pp. 140–142, 2012.

[38] C. Kampen, A. Burenkov, and J. Lorenz, “Challenges in TCAD simulations of tunneling field effect transistors,” in *Solid-State Device Res. Conf. (ESSDERC), 2011 Proc. Eur.*, pp. 139–142, IEEE, 2011.

[39] M. Hermle, G. Letay, and S. Philipps, “Numerical simulation of tunnel diodes for multi-junction solar cells,” in *IEEE Photovolt. Spec. Conf.*, pp. 10–13, 2008.

[40] J. M. Miller, “Dependence of the input impedance of a three-electrode vacuum tube upon the load in the plate circuit,” *Sci. Pap. Bur. Stand.*, vol. 15, no. 351, pp. 367–385, 1920.

[41] S. Mookerjea, R. Krishnan, S. Datta, and V. Narayanan, “On enhanced Miller capacitance effect in interband tunnel transistors,” *IEEE EDL*, vol. 30, no. 10, pp. 1102–1104, 2009.

[42] S. Kim, S. Agarwal, Z. Jacobson, P. Matheu, C. Hu, and T. Liu, “Tunnel field effect transistor with raised germanium source,” *IEEE Electron Device Lett.*, vol. 31, no. 10, pp. 1107–1109, 2010.

[43] Q.-T. Zhao, S. Richter, C. Schulte-Braucks, L. Knoll, S. Blaeser, G. V. Luong, S. Trellenkamp, A. Schaefer, A. Tiedemann, K. Bourdelle, J.-M. Hartmann, and S. Mantl, “Strained Si and SiGe Nanowire Tunnel FETs for Logic and Analog Applications,” *IEEE J. Electron Devices Soc.*, vol. 6734, no. XXX, pp. 1–1, 2015.

[44] A. S. Verhulst, B. Soree, D. Leonelli, W. G. Vandenberghe, and G. Groeseneken, “Modeling the single-gate, double-gate, and gate-all-around tunnel field-effect transistor,” *JAP*, vol. 107, no. 2, p. 024518, 2010.

[45] C. Wu, R. Huang, Q. Huang, C. Wang, J. Wang, and Y. Wang, “An analytical surface potential model accounting for the dual-modulation effects in tunnel FETs,” *IEEE TED*, vol. 61, pp. 2690–2696, 2014.

[46] M. J. Lee and W. Y. Choi, “Analytical model of single-gate silicon-on-insulator (SOI) tunneling field-effect transistors (TFETs),” *Solid. State. Electron.*, vol. 63, pp. 110–114, Sept. 2011.

[47] M. G. Bardon, H. P. Neves, R. Puers, and C. Van Hoof, “Pseudo-2D Model for Double-Gate Tunnel FETs Considering the Junctions Depletion Regions,” *IEEE TED*, vol. 57, pp. 827–834, Apr. 2010.

[48] L. De Michielis, N. Dagtekin, A. Biswas, L. Lattanzio, L. Selmi, M. Luisier, H. Riel, and a. M. Ionescu, “An innovative band-to-band tunneling analytical model and implications in compact modeling of tunneling-based devices,” *Appl. Phys. Lett.*, vol. 103, no. 12, p. 123509, 2013.

[49] Y. Lu, A. Seabaugh, P. Fay, S. J. Koester, S. E. Laux, T. W. Haensch, and S. O. Koswatta, “Geometry dependent tunnel FET performance - Dilemma of electrostatics vs. quantum confinement,” in *Conf. Dig. DRC*, pp. 17–18, 2010.

[50] L. Zhang, X. Lin, J. He, and M. Chan, “An Analytical Charge Model for Double-Gate Tunnel FETs,” *IEEE Trans. Electron Devices*, vol. 59, pp. 3217–3223, Dec. 2012.

[51] K. Boucart and A. M. Ionescu, "Double-Gate Tunnel FET With High K Gate Dielectric," *IEEE Trans. Electron Devices*, vol. 54, no. 7, pp. 1725–1733, 2007.

[52] J.-M. Sallese, F. Krummenacher, F. Prégaldiny, C. Lallement, A. Roy, and C. Enz, "A design oriented charge-based current model for symmetric DG MOSFET and its correlation with the EKV formalism," *Solid. State. Electron.*, vol. 49, pp. 485–489, Mar. 2005.

[53] M. Schwarz, T. Holtij, A. Kloes, and B. Iniguez, "2D analytical framework for compact modeling of the electrostatics in undoped DG MOSFETs," *Proc. 18th Int. Conf. Mix. Des. Integr. Circuits Syst. - Mix. 2011*, no. 3, pp. 405–410, 2011.

[54] T. Holtij, M. Schwarz, A. Kloes, and B. Iñíguez, "Threshold voltage, and 2D potential modeling within short-channel junctionless DG MOSFETs in subthreshold region," *Solid. State. Electron.*, vol. 90, pp. 107–115, 2013.

[55] A. Kloes, M. Schwarz, and T. Holtij, " A New Physics-Based Explicit Compact Model for Lightly Doped Short-Channel Triple-Gate SOI MOSFETs," *IEEE Trans. Electron Devices*, vol. 59, pp. 349–358, Feb. 2012.

[56] Weber E, *Electromagnetic fields. Mapping of fields.* New York: John Wiley, vol. 1 ed., 1950.

[57] M. Schwarz, T. Holtij, A. Kloes, and B. Iñíguez, "Analytical compact modeling framework for the 2D electrostatics in lightly doped double-gate MOSFETs," *Solid. State. Electron.*, vol. 69, pp. 72–84, Mar. 2012.

[58] Chen Shen, Sern-Long Ong, Chun-Huat Heng, Ganesh Samudra and Y.-C. Yeo, "A Variational Approach to the Two-Dimensional Nonlinear Poisson's Equation for the Modeling of Tunneling Transistors," *IEEE Electron Device Lett.*, vol. 29, pp. 1252–1255, Nov. 2008.

[59] K. Suzuki, T. Tanaka, and Y. Tosaka, "Scaling theory for double-gate SOI MOSFET's," *IEEE Trans. Electron Devices*, 1993.

[60] A. Pan and C. O. Chui, "A Quasi-Analytical Model for Double-Gate Tunneling Field-Effect Transistors," *IEEE Electron Device Lett.*, vol. 33, pp. 1468–1470, Oct. 2012.

[61] L. Zhang, J. He, and M. Chan, "A compact model for double-gate tunneling field-effect-transistors and its implications on circuit behaviors," in *Tech. Dig. - IEDM*, no. 5, pp. 6.8.1–6.8.4, IEEE, Dec. 2012.

[62] S. Biesemans, S. Kubicek, and K. D. Meyer, "New current-defined threshold voltage model from 2D potential distribution calculations in MOSFETs," *Solid. State. Electron.*, vol. 39, no. I, pp. 43–48, 1996.

[63] L. De Michielis, M. Iellina, P. Palestri, A. M. Ionescu, and L. Selmi, "Effect of the choice of the tunnelling path on semi-classical numerical simulations of TFET devices," *Solid. State. Electron.*, vol. 71, pp. 7–12, May 2012.

[64] L. De Michielis, M. Iellina, P. Palestri, A. M. Ionescu, and L. Selmi, “Tunneling path impact on semi-classical numerical simulations of TFET devices,” *2011 12th Int. Conf. Ultim. Integr. Silicon, ULIS 2011*, pp. 146–149, 2011.

[65] Y. Yang, X. Tong, L.-t. Yang, P.-f. Guo, L. Fan, and Y.-c. Yeo, “Tunneling Field-Effect Transistor: Capacitance Components and Modeling,” *IEEE Electron Device Lett.*, vol. 31, pp. 752–754, July 2010.

[66] B. Sheu, D. Scharfetter, P.-K. Ko, and M.-C. Jeng, “BSIM: Berkeley short-channel IGFET model for MOS transistors,” *IEEE J. Solid-State Circuits*, vol. 22, pp. 558–566, Aug. 1987.

[67] B. Bhowmick, “An Analytical Model for Fringing Capacitance in Double gate Hetero Tunnel FET and Analysis of effect of Traps and Oxide charges on Fringing Capacitance,” *Int. J. VLSI Des. Commun. Syst.*, vol. 3, pp. 1–11, Feb. 2012.

[68] R. Baek, C. Baek, S. Lee, and S. Suk, “C–V Characteristics in Undoped Gate-All-Around Nanowire FET Array,” *Electron Device Lett. IEEE*, vol. 32, no. 2, pp. 116–118, 2011.

[69] A. Mallik and A. Chattopadhyay, “Tunnel Field-Effect Transistors for Analog/Mixed-Signal System-on-Chip Applications,” *IEEE Trans. Electron Devices*, vol. 59, pp. 888–894, Apr. 2012.

[70] S. Mookerjea, R. Krishnan, S. Datta, and V. Narayanan, “Effective Capacitance and Drive Current for Tunnel FET (TFET) CV/I Estimation,” *IEEE Trans. Electron Devices*, vol. 56, pp. 2092–2098, Sept. 2009.

[71] L. Knoll, Q.-t. Zhao, A. Nichau, S. Trellenkamp, S. Richter, A. Schafer, D. Esseni, L. Selmi, K. K. Bourdelle, and S. Mantl, “Inverters With Strained Si Nanowire Complementary Tunnel Field-Effect Transistors,” *IEEE Electron Device Lett.*, vol. 34, pp. 813–815, June 2013.

[72] D. E. Nikonov and I. a. Young, “Overview of beyond-CMOS devices and a uniform methodology for their benchmarking,” *Proc. IEEE*, vol. 101, no. 12, pp. 2498–2533, 2013.

[73] S. Hanson, M. Seok, D. Sylvester, and D. Blaauw, “Nanometer device scaling in subthreshold logic and SRAM,” *IEEE TED*, vol. 55, no. 1, pp. 175–185, 2008.

[74] H. Kam, T. King-Liu, E. Alon, and M. Horowitz, “Circuit-level requirements for MOSFET-replacement devices,” *Tech. Dig. - IEDM*, vol. 11, no. 6, p. 2008, 2008.

[75] A. Trivedi, S. Carlo, and S. Mukhopadhyay, “Exploring tunnel-FET for ultra low power analog applications: a case study on operational transconductance amplifier,” *Des. Autom. Conf. (DAC), 2013.*, no. V, 2013.

[76] L. Knoll, S. Richter, A. Nichau, S. Trellenkamp, A. Schäfer, K. Bourdelle, J. Hartmann, Q. Zhao, and S. Mantl, “Strained Si and SiGe tunnel-FETs and complementary tunnel-FET inverters with minimum gate lengths of 50nm,” *Solid. State. Electron.*, vol. 97, pp. 76–81, July 2014.

[77] F. Silveira, D. Flandre, and P. Jespers, “A gm/ID based methodology for the design of CMOS analog circuits and its application to the synthesis of a silicon-on-insulator micropower OTA,” *IEEE J. Solid-State Circuits*, vol. 31, no. 9, pp. 1314–1319, 1996.

[78] B. Senale-Rodriguez, Yeqing Lu, P. Fay, D. Jena, A. Seabaugh, H. G. Xing, L. Barboni, and F. Silveira, “Perspectives of TFETs for low power analog ICs,” in *2012 IEEE Subthreshold Microelectron. Conf. SubVT 2012*, pp. 8–10, 2012.

[79] S. Gray, Paul R and Hurst, Paul and Meyer, Robert G and Lewis, *Analysis and design of analog integrated circuits.* Wiley, 2001.

[80] P. Crawley and G. Roberts, “High-swing MOS current mirror with arbitrarily high output resistance,” *Electron. Lett.*, vol. 28, no. 4, p. 361, 1992.

[81] Keng Leong Fong and R. Meyer, “High-frequency nonlinearity analysis of common-emitter and differential-pair transconductance stages,” *IEEE J. Solid-State Circuits*, vol. 33, pp. 548–555, Apr. 1998.

[82] T. Brayshaw, F and Roberts, “Physiological pressure sensing and telemetry means employing a diode connected transistor transducer,” 1973.

[83] H. Van Tran, “CMOS voltage regulator with diode-connected transistor divider circuit,” 1998.

[84] M. corporation, “Power In, Dollars Out: How to Stem the Flow in the Data Center,” tech. rep., 2010.

[85] M. Born and K. Bhuwalka, “Tunnel FET: A CMOS device for high temperature applications,” in *Int. Conf. MIicroelectronics*, no. Miel, pp. 14–17, 2006.

[86] M. Tack and M. Gao, “The multistable charge-controlled memory effect in SOI MOS transistors at low temperatures,” *IEEE Trans. Electron Devices*, vol. 31, no. 5, pp. 1373–1382, 1990.

[87] H.-J. W. H.-J. H. Wann and C. H. C. Hu, “A capacitorless DRAM cell on SOI substrate,” in *Proc. IEEE Int. Electron Devices Meet.*, pp. 635–638, IEEE, 1993.

[88] M. Bawedin, S. Cristoloveanu, and A. Hubert, “Floating-Body SOI Memory: The Scaling Tournament,” *…-On-Insulator Mater. …*, no. Umr 5130, 2011.

[89] G. Guegan, P. Touret, G. Molas, C. Raynaud, and J. Pretet, “A novel capacitor-less 1T-DRAM on partially depleted SOI pMOSFET based on direct-tunneling current in the partial n+ poly gate,” *…Ext. Abs. Solid State Devices …*, 2010.

[90] C. Hu and T. King, “A capacitorless double-gate DRAM cell,” *Electron Device Lett. IEEE*, vol. 23, no. 6, pp. 345–347, 2002.

[91] S. Okhonin, M. Nagoga, J. Sallese, and P. Fazan, “A SOI capacitor-less 1T-DRAM concept,” *IEEE Int. SOI Conf.*, 2001.

[92] S. Okhonin, M. Nagoga, J. Sallese, and P. Fazan, “A capacitor-less 1T-DRAM cell,” *IEEE Electron Device Lett.*, vol. 23, pp. 85–87, Feb. 2002.

[93] C. Kuo, T. King, and C. Hu, “A capacitorless double-gate DRAM cell design for high density applications,” *IEDm Int. electron devices Meet.*, pp. 843–846, 2002.

[94] T. Shino, T. Ohsawa, T. Higashi, K. Fujita, N. Kusunoki, Y. Minami, M. Morikado, H. Nakajima, K. Inoh, T. Hamamoto, and A. Nitayama, “Operation Voltage Dependence of Memory Cell Characteristics in Fully Depleted Floating-Body Cell,” *IEEE Trans. Electron Devices*, vol. 52, pp. 2220–2226, Oct. 2005.

[95] I. Ban, U. E. Avci, U. Shah, C. E. Barns, D. L. Kencke, and P. Chang, “Floating Body Cell with Independently-Controlled Double Gates for High Density Memory,” in *2006 Int. Electron Devices Meet.*, pp. 1–4, IEEE, 2006.

[96] S. Okhonin, M. Nagoga, E. Carman, R. Beffa, and E. Faraoni, “New Generation of Z-RAM,” in *2007 IEEE Int. Electron Devices Meet.*, pp. 925–928, IEEE, 2007.

[97] Ki-Whan Song, Hoon Jeong, Jae-Wook Lee, Sung In Hong, Nam-Kyun Tak, Young-Tae Kim, Yong Lack Choi, Han Sung Joo, Sung Hwan Kim, Ho Ju Song, Yong Chul Oh, Woo-Seop Kim, Yeong-Taek Lee, Kyungseok Oh, and Changhyun Kim, “55 nm capacitor-less 1T DRAM cell transistor with non-overlap structure,” in *2008 IEEE Int. Electron Devices Meet.*, pp. 1–4, IEEE, Dec. 2008.

[98] S. Okhonin, M. Nagoga, C.-W. Lee, J.-P. Colinge, A. Afzalian, R. Yan, N. Dehdashti Akhavan, W. Xiong, V. Sverdlov, S. Selberherr, and C. Mazure, “Ultra-scaled Z-RAM cell,” in *2008 IEEE Int. SOI Conf.*, pp. 157–158, IEEE, Oct. 2008.

[99] E. Yoshida and T. Tanaka, “A capacitorless 1T-DRAM technology using gate-induced drain-leakage (GIDL) current for low-power and high-speed embedded memory,” *IEEE Trans. Electron Devices*, vol. 53, pp. 692–697, Apr. 2006.

[100] M. Bawedin, S. Cristoloveanu, and D. Flandre, “A Capacitorless 1T-DRAM on SOI Based on Dynamic Coupling and Double-Gate Operation,” *IEEE Electron Device Lett.*, vol. 29, pp. 795–798, July 2008.

[101] E. Y. T. Tanaka, “Scalability study on a capacitorless 1T-DRAM from single-gate PD-SOI to double-gate FinDRAM,” in *IEDM Tech. Dig. IEEE Int. Electron Devices Meet. 2004.*, pp. 919–922, IEEE, 2004.

[102] A. Biswas, N. Dagtekin, W. Grabinski, A. Bazigos, C. Le Royer, J.-M. Hartmann, C. Tabone, M. Vinet, and A. M. Ionescu, “Investigation of tunnel field-effect transistors as a capacitor-less memory cell,” *Appl. Phys. Lett.*, vol. 104, p. 092108, Mar. 2014.

[103] N. Dagtekın, *Practical Applications of Tunnel Field Effect Transistors.* PhD thesis, École polytechnique fédérale de Lausanne, 2015.

[104] J. Liu, B. Jaiyen, Y. Kim, C. Wilkerson, and O. Mutlu, “An experimental study of data retention behavior in modern DRAM devices,” in *Proc. 40th Annu. Int. Symp. Comput. Archit. - ISCA ’13*, vol. 41, (New York, New York, USA), p. 60, ACM Press, 2013.

[105] A. Biswas, S. S. Dan, C. L. Royer, W. Grabinski, and A. M. Ionescu, “TCAD simulation of SOI TFETs and calibration of non-local band-to-band tunneling model,” in *Microelectron. Eng.*, vol. 98, pp. 334–337, Elsevier B.V., July 2012.

[106] Y. R. Jhan, Y. C. Wu, H. Y. Lin, M. F. Hung, Y. H. Chen, and M. S. Yeh, “High performance of fin-shaped tunnel field-effect transistor SONOS nonvolatile memory with all programming mechanisms in single device,” *IEEE Trans. Electron Devices*, vol. 61, no. 7, pp. 2364–2370, 2014.

[107] Jesung Kim, Jong Min Kim, S. Noh, Sang Lyul Min, and Yookun Cho, “A space-efficient flash translation layer for CompactFlash systems,” *IEEE Trans. Consum. Electron.*, vol. 48, pp. 366–375, May 2002.

[108] S. J. Wrazien, *Retention reliability in polysilicon-oxide-nitrideoxide-silicon (SONOS) nonvolatile semiconductor memory devices.* PhD thesis, Lehigh University, 2002.

[109] J. Bu, *Polysilicon-Oxide-Nitride-Oxide-Silicon (SONOS) Nonvolatile Semiconductor Memory Devices Design, Fabrication, and Characterization.* PhD thesis, Sherman Fairchild Center, 2001.

[110] S. Aritome, S. Satoh, T. Maruyama, H. Watanabe, S. Shuto, G. Hemink, R. Shirota, S. Watanabe, and F. Masuoka, “A 0.67 μm/sup 2/ self-aligned shallow trench isolation cell (SA-STI cell) for 3 V-only 256 Mbit NAND EEPROMs,” in *Proc. 1994 IEEE Int. Electron Devices Meet.*, pp. 61–64, IEEE.

[111] K. Suh and Y. Suh, “A 3.3 V 32Mb NAND Flash Memory with incremental Step Pulse Programming Scheme,” *ISSCC*, pp. 32–34, 1995.

[112] S. Satoh, H. Hagiwara, T. Tanzawa, K. Takeuchi, and R. Shirota, “A novel isolation-scaling technology for NAND EEPROMs with the minimized program disturbance,” in *Int. Electron Devices Meet. IEDM Tech. Dig.*, pp. 291–294, IEEE, 1997.

[113] K. Takeuchi, S. Satoh, T. Tanaka, K. Imamiya, and K. Sakui, “A negative Vth cell architecture for highly scalable, excellently noise immune and highly reliable NAND flash memories,” in *1998 Symp. VLSI Circuits. Dig. Tech. Pap. (Cat. No.98CH36215)*, pp. 234–235, IEEE, 1998.

[114] Jung-Dal Choi, Seong-Soon Cho, Yong-Sik Yim, Jae-Duk Lee, Hong-Soo Kim, Kyung-Joong Joo, Sung-Hoi Hur, Heung-Soo Im, Joon Kim, Jeong-Woo Lee, Kang-Ill Seo, Man-Sug Kang, Kyung-Hyun Kim, Jeong-Lim Nam, Kyu-Charn Park, and Moon-Yong Lee,

"Highly manufacturable 1 Gb NAND flash using 0.12 μm process technology," in *Int. Electron Devices Meet. Tech. Dig. (Cat. No.01CH37224)*, vol. 3, pp. 2.1.1–2.1.4, IEEE, 2001.

[115] O. Kevin Zhang, Intel, "ISSCC 2013 Tech Trends," tech. rep., 2013.

[116] J. J. J. Jang, H.-S. K. H.-S. Kim, W. C. W. Cho, H. C. H. Cho, J. K. J. Kim, S. I. S. S. I. Shim, Y. J. Y. Jang, J.-H. J. J.-H. Jeong, B.-K. S. B.-K. Son, D. W. K. D. W. Kim, K. Kihyun, J.-J. S. J.-J. Shim, J. S. L. J. S. Lim, K.-H. K. K.-H. Kim, S. Y. Y. S. Y. Yi, J.-Y. L. J.-Y. Lim, D. C. D. Chung, H.-C. M. H.-C. Moon, S. H. S. Hwang, J.-W. L. J.-W. Lee, Y.-H. S. Y.-H. Son, U.-I. C. U.-I. Chung, and W.-S. L. W.-S. Lee, "Vertical cell array using TCAT(Terabit Cell Array Transistor) technology for ultra high density NAND flash memory," in *2009 Symp. VLSI Technol.*, pp. 14–15, 2009.

[117] L. De Michielis, L. Lattanzio, P. Palestri, L. Selmi, and A. M. Ionescu, "Tunnel-FET architecture with improved performance due to enhanced gate modulation of the tunneling barrier," in *69th Device Res. Conf.*, pp. 111–112, IEEE, June 2011.

[118] Y. Khatami and K. Banerjee, "Steep subthreshold slope n- and p-type Tunnel-FET devices for low-power and energy-efficient digital circuits," *IEEE Trans. Electron Devices*, vol. 56, no. 11, pp. 2752–2761, 2009.

My Publications

Conference proceedings:

A. Biswas and A. M. Ionescu. Compact modeling of DG-Tunnel FET for Verilog-A Implementation, in 45th European Solid-State Device Research Conference (ESSDERC), Graz, Austria, September 14-18, 2015.

A. Biswas and A. M. Ionescu. A Capacitance-Voltage model for DG-TFET, in Proceedings of the IEEE Silicon Nanoelectronics Workshop, Kyoto, Japan, June 14-15, 2015.

A. Biswas and M. A. Ionescu. Study of Fin-Tunnel FETs with doped pocket as Capacitor-less 1T DRAM, in SOI-3D-Subthreshold Microelectronics Technology Unified Conference (S3S), San Francisco, California, USA, October 6-9, 2014.

A. Biswas, N. Dagtekin, C. Alper, L. De Michielis and A. Bazigos et al. Compact Modeling of Homojunction Tunnel FETs, in Proceedings of the 21st International Conference on Mixed Design of Integrated Circuits and Systems (MIXDES), Lublin, Poland, June 19-21, 2014.

A. Biswas, L. De Michielis, C. Alper and M. A. Ionescu. Conformal Mapping Based DC Current Model for Double Gate Tunnel FETs, in 15th International Conference on Ultimate Integration on Silicon (ULIS), Stockholm, Sweden, April 7-9, 2014.

A. Biswas, N. Dagtekin, W. Grabinski, A. Bazigos and C. Le Royer et al. A study on 1T Capacitor-less Tunnel FET DRAM Exploiting Ungated Body, in ISDRS 2013, Washington DC, USA, December 11-13, 2013.

A. Biswas, C. Alper, L. D. Michielis and A. M. Ionescu. New Tunnel-FET Architecture with Enhanced Ion and Improved Miller Effect for Energy Efficient Switching, in 70th Device Research Conference, State College, PA, USA, June 18-20, 2012.

L. Lattanzio, L. De Michielis, A. Biswas and A. M. Ionescu. Abrupt switch based on internally combined Band-To-Band and Barrier Tunneling mechanisms, in 40th European Solid State Device Research Conference (ESSDERC), Sevilla, Spain, September 14-16, 2010.

Journal papers:

A. Biswas and A. M. Ionescu. 1T Capacitor-less DRAM cell based on asymmetric Tunnel FET design, in IEEE Journal of the Electron Devices Society, vol.3, no.3, pp.217-222, May 2015.

A. Biswas, N. Dagtekin, W. Grabinski, A. Bazigos and C. L. Royer et al. Investigation of Tunnel Field-Effect Transistors as a Capacitor-less Memory Cell, in Applied Physics Letters, vol. 104, num. 9, p. 092108, 2014.

L. De Michielis, N. Dağtekin, A. Biswas, L Lattanzio, L Selmi, M Luisier, H Riel, AM Ionescu. An innovative band-to-band tunneling analytical model and implications in compact modeling of tunneling-based devices, in Applied Physics Letters, vol. 103, num. 12, p. 123509, 2013.

A. Biswas, S. S. Dan, C. Le Royer, W. Grabinski and M. A. Ionescu. TCAD Simulation of SOI TFETs and Calibration of Non-local Band-to-Band Tunneling Model, in Microelectronic Engineering, vol. 98, p. 334–337, 2012.

S. S. Dan, A. Biswas, C. L. Royer, W. Grabinski and A. M. Ionescu. A Novel Extraction Method and Compact Model for the Steepness Estimation of FDSOI TFET Lateral Junction, in IEEE Electron Device Letters, vol. 33, num. 2, p. 140-142, 2012.

L. Lattanzio, A. Biswas, L. De Michielis and A. M. Ionescu. A tunneling field-effect transistor exploiting internally combined band-to-band and barrier tunneling mechanisms, in Applied Physics Letters, vol. 98, 2011.

L. Lattanzio, A. Biswas, L. De Michielis and A. M. Ionescu. Abrupt switch based on internally combined Band-to-Band and Barrier Tunneling mechanisms, in Solid-State Electronics 11/2011; 65:234-239.

Workshop and posters:

R. Rhyner, C. Bessire, L. De Michielis, N. Dagtekin, A. Biswas, A. Schenk, M. Björk, K. Moselund, H. Schmid, H. Riel, and A. M. Ionescu, "Enabling energy efficient Tunnel FET-CMOS co-design by compact modeling and simulation", NANOTERA Plenary annual meeting, 2012.

Patent application:

Title of the invention: "Memory device and operation method of the said device".
Inventors: Arnab Biswas, Nilay Dagtekin, Adrian Mihai Ionescu.
Publication number: US 2015/79800 A1
Publication date: June 25th, 2015
Filed on: December 6, 2013.

Arnab BISWAS

ELB 343, STI-IEL-Nanolab, EPFL
Station 11, CH-1015 Lausanne, Switzerland

Email arnab.biswas@outlook.com
Nationality Indian

Key Points

- PhD in Nanoelectronics from EPFL, Switzerland (September 2015).
- Skilled in the use of proprietary and commercial tools for semiconductor device simulation (TCAD) and understanding of semiconductor device physics.
- Experience in developing process flows and understanding of fabrication process for advanced CMOS devices.

Education Summary

PhD, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, (2010 to 2015)

Master of Technology, Vellore Institute of Technology), India, (2008-2010)
Major : Nanotechnology
Grade : 9.44/10 (CGPA)

Bachelor of Technology, Sikkim Manipal Institute of Technology, India, (2002-2006)
Major : Electronics and Communication Engineering

PhD Training

- Research objectives :
    - Work focused on physics based compact model development of Tunnel FET.
    - Proposal and implementation of new device architectures.
    - Experimental characterization of new low-power electronic transistors.
    - Benchmarking of Tunnel FET device and circuit level figure of merits against standard CMOS nodes.
    - Proposal of new energy efficient applications for Tunnel FETs.

- Research experience with the European STEEPER and E2SWITCH project (2011-2015) -- in a multinational team with colleagues from:
    - Industry (IBM Research, Intel MC and Cambridge CMOS Sensors).
    - Research institutes (CEA-LETI, IMEC and Forschungszentrum Jülich)
    - Academia (ETH Zurich, University of Bologna and University of Udine).

- Publication of research in more than 15 international journals, conferences and workshops. One US patent under process.
    - Google scholar H-index of 4 as of July 2015.

Masters Training

- Design and fabrication of Schottky Tunnel FET on thin SOI substrates, at NANOLAB, Ecole Polytechnique Federale de Lausanne (EPFL), Switzerland. (October 2009 to March 2010)
- Analytical model and simulation of Subthreshold current and slope of a surrounding gate MOSFET, at Nanotechnology Lab, VIT University, Vellore, India. (February 2009 to April 2009)

Technical Skills, Competences & Tools

Compact modelling	Expert in semiconductor physics, compact model development of advanced electronic devices.
CAD and other tools	• Deep knowledge of Synopsys Sentaurus TCAD simulation tools. Experience with Silvaco Atlas. • Extensive use of Matlab, Verilog-A, Cadence Virtuoso, Agilent IC-CAP. Experience in LabVIEW, OCEAN script, C/C++, Python.
Characterization tools	Extensive experience in DC/AC electrical characterization tools for semiconductor devices. • Semi-automated probe station (Cascade Microtech, Summit 1200B) • Semiconductor parameter analyzer, switching matrix (Agilent) • LCR bridge, CV meter (Agilent, Kiethley)
Clean room experience	Experience in working in class-100 clean room at CMi, EPFL. • Photolithography: coating, exposure and development of micron sized features. • Thin film deposition: Electron beam evaporation, sputtering. • Metrology tools: SEM, FIB, AFM, Ellipsometry.

Employment History

Tata Consultancy Services Ltd., Mumbai, India (September 2006 to August 2008)

Role : Assistant Systems Engineer in an IBM-Mainframes support project. Responsibilities involved responding immediately technical issues related to production support.

Languages

- English: Full professional proficiency
- Bengali/Hindi: Mother tongue/native speaker
- French: Intermediate level
- German: Beginner level

Awards & Achievements

- Awarded merit scholarship for securing first rank in masters from VIT University, India for the year 2008-09 and 2009-10.
- Secured a full score of 10.00 (GPA) in 1st semester Nanotechnology master’s program at VIT University, India in 2008-09.

Other Information

- Supervisor of two semester and two master student projects in NANOLAB, EPFL.
- Core member in Association Des Étudiants Indiens De Lausanne “YUVA”. (2013-2014)
- Webmaster for the nanolab.epfl.ch (NANOLAB) group web page. (2010-2015)
- Included in IEEE TED/EDL golden reviewers list. (2012-2014)