KATHOLIEKE UNIVERSITEIT
LEUVEN

Arenberg Doctoral School of Science, Engineering & Technology
Faculty of Engineering
Department of Electrical Engineering

DESIGN, FABRICATION AND CHARACTERIZATION OF TUNNEL FIELD EFFECT TRANSISTORS FOR ULTRA-LOW POWER CMOS APPLICATIONS

Daniele LEONELLI

Dissertation presented in
partial fulfillment of the
requirements for the degree
of Doctor of Engineering

October 2012

KU LEUVEN
Faculty of Engineering
Department of Electrical Engineering
ESAT – MICAS
Kasteelpark Arenberg 10, 3001, Leuven, Belgium

DESIGN, FABRICATION AND CHARACTERIZATION OF TUNNEL FIELD EFFECT TRANSISTORS FOR ULTRA-LOW POWER CMOS APPLICATIONS

PhD Jury:
Prof. Dr. Ir. Paul Van Houtte (Chairman)
Prof. Dr. Ir. Guido Groeseneken (Promotor)
Prof. Dr. Ir. Marc Heyns
Prof. Dr. Stefan De Gendt
Prof. Dr. Ir. Wim Dehaene (Secretary)
Dr. Ir. Anne Verhulst
Dr. Ir. Anne Vandooren
Dr. Bhuwalka Krishna Kumar (TSMC – Taiwan)

Dissertation presented in
partial fulfillment of the
requirements for the degree
of Doctor of Engineering

by

Daniele LEONELLI

In collaboration with
imec
Interuniversitair Micro-Electronica Centrum
Kapeldreef 75,
3001 Heverlee
Leuven, Belgium

October 2012

Printed By: *PROCOPIA NV*
Ambachtenlaan 29
3001 Heverlee
016/28.36.34
info@procopia.be
www.procopia.be

Kasteelpark Arenberg 10, B-3001 Heverlee (Belgium)

ISBN: 978-94-6018-540-3
Wettelijke depotnummer: D/2012/7515/74

To Jun • 叶世军

“Odi et amo. Quare id faciam,
fortasse requiris. Nescio, sed fieri
sentio et excrucior.” (Catullo)

Acknowledgments

PhD should be a work to demonstrate the capability of being able to manage a research project independently. Nevertheless, the research is not a result of one individual's work but a result of a coordinated team. For this reason, I owe my gratitude to all the people who directly and indirectly contributed to the success of this work.

First of all, I am grateful to Prof. Guido Groeseneken for giving me the opportunity to start my PhD at imec and for his trust during the difficult moments. In spite of my making tight time schedules, especially in the last part of the PhD work, he was always supportive and I really appreciate his sacrifice.

I need to thank Prof. Marc Heyns, who offered me very interesting discussions that broadened my knowledge by pushing me to think about new solutions and ideas. Thanks to him, I understood that research is very creative and people never stop dreaming about the future.

I am thankful to my daily advisor, Anne Vandooren, who started the TFET project and her work at imec just before the starting of my PhD. Her background on finFET architecture offered a good scientific guidance. In addition, I thank her for her support in reviewing this thesis and part of my previous work.

It was a tough time at the beginning since the TFET project was really at an embryonic stage. On the other hand, I was able to fabricate my finFET lot during the second year of my PhD thanks to the support and knowledge of Rita Rooyackers, who taught me most of the secrets behind the VLSI. Her long experience at imec was very helpful in finding the right people who quickly solved any processing problems.

I wish to express my gratitude to Anne Verhulst who supported the simulation part of the TFET and built a good foundation of this project and my PhD itself together with our previous team leader, Prof. Stefan De Gendt.

I would like to thank Prof. Tsu-Jae King Liu, who gave me the opportunity to collaborate with her at UC Berkeley. Unfortunately, the time was too short but it was a very important experience. In addition, I am grateful to Prof. Gerard Ghibaudo at INP (Grenoble, France) and Prof. Sven Rogge at TU Delft (Netherlands) who give me the opportunity to perform low temperature measurements on my experimental devices. Thanks to them I learned the importance and the difficulties involved in setting up an external collaboration.

I would like to thank the other team members and imec researchers who interacted with me and gave me a possibility to discuss critically about the concerns and difficulties I encountered during this work. In addition, I would like to thank all my friends who made my experience here colorful and gave me “sunny” days in this “cloudy” country. Listing them all is impossible and listing some of them is unfair so I think I should keep it blank.

I cannot forget the important people who indirectly contributed to this work. Firstly, I would like to thank my partner, 叶世军 (aka Jun), who was with me “in the pain and in the joy” since the beginning of my internship at imec Belgium in April 2007 and later PhD at imec. Thanks to her, I have discovered that teaching is much more complicated than learning. In fact, knowing the solution of a problem is as important as knowing the procedure to reach the solution. During this period I grew up and I believe I am a better person than before thanks to her. I hope we can continue to grow up together in future. She introduced me to the Natural Born Hikers (NBH) group. I have been a member of this group since 2007 and together with Tomas and other core members we were able to spread this experience to a wide range of people. This experience was not restricted only to discovering one’s own physical limits but it was also a vehicle to socialize with people from different cultures and become “a citizen of the world”.

Last but not least my parents. I owe them my life and gratitude and their support during the final phase of my PhD. They are good educators and they taught me that nothing is granted and our determination is the key to success. For this reason, I could find the strength to continue when I was feeling alone and depressed. I will always keep them in my heart.

Abstract

Silicon CMOS has emerged over the last 25 years as the predominant technology of the microelectronics industry. The concept of device scaling has been consistently applied over many technology generations, resulting in consistent improvement in both device density and performance. In the last decade, the shrinking of the transistor dimensions led to short channel effects (SCEs) which decreases the device performance. As a consequence, additional improvements were needed to maintain the performance improvements such as the introduction of SiGe S/D stressors to increase the carrier mobility due to the strain in the channel, the implementation of high-K and metal gate to reduce the gate leakage and, finally, the introduction of a new 3D architecture, finFET, to further suppress the SCEs. On the other hand, no solutions exist for the scaling of the dissipated power of the transistor. In fact, the scaling of the supply voltage (V_{DD}) is limited by the kT/q limit of the subthreshold slope which represents a physical limit for conventional MOSFETs. As a consequence, a new operation principle is needed.

In this context, the Tunnel FET (TFET) has been proposed as a potential candidate to replace the MOSFET because its carrier injection mechanism based on quantum mechanical tunneling of the electrons from valence band to the conduction band and it is not subjected to the kT/q limit. The basic embodiment of TFET is a gated p-i-n diode.

This thesis addresses the design, fabrication and characterization of TFETs following a CMOS compatible processing flow. The main goal is to understand the features of band to band tunneling from an experimental point of view and identify the best processing conditions and the best architecture for TFETs. The analysis starts from silicon homojunction gated p-i-n diodes to heterojunction devices where the source is replaced with SiGe with different germanium concentrations. Two different architectures are studied: finFETs and vertical nanowires. The finFETs are used as a test vehicle to study Si TFETs since the finFET processing is already mature. On the other hand, the vertical architecture is used to implement hetero junction TFET.

Finally, in view of the limitations for the basic TFET embodiments, a new architecture to boost the on current of TFETs is proposed and analyzed by TCAD simulations.

Korte samenvatting

Al 25 jaar wordt de micro-elektronica gedomineerd door silicium CMOS technologie. Dit was mogelijk dankzij het schalen van de transistorafmetingen. Zo kan bij elke nieuwe generatie de transistordichtheid verhoogd worden, en neemt ook de performantie toe. De laatste 10 jaar is het schalen van de transistor moeilijker geworden door een reeks korte kanaal-effecten, die de performantiewinst van de transistor in gedrang brachten. Hierdoor waren er een aantal radicale veranderingen aan de MOSFET noodzakelijk. De mobiliteit is verhoogd met silicium-germanium source/drain stressoren, de poort-lek wordt beperkt door het gebruik van een high-k dielectricum, en de architectuur is nu drie-dimensionaal met de introductie van de finFET.

Bij het schalen van de transistor, moet echter ook de voedingsspanning verlaagd worden om het vermogenverbruik onder controle te houden. Omdat de transistor grotendeels in het subtreshold regime werkt, is het niet mogelijk de spanning verder te verlagen zonder een groot verlies aan performantie. Dit komt door het thermische injectiemechanisme van de MOSFET, waar de subtreshold slope fysisch beperkt is tot kT/q. Er moet dus van werkingsprincipe veranderd worden.

De Tunnel-FET (TFET) is een mogelijke kandidaat om de MOSFET te vervangen, omdat het werkinsgmechanisme niet op thermische injectie berust, maar op kwantummechanische tunneling van electronen van de valentieband naar de conductieband (band-tot-band tunneling). De basisversie van de TFET is een p-i-n diode met een poort.

In deze thesis wordt het ontwerp, de fabricatie en de karakterisatie van TFET’s in een CMOS-compatibele flow besproken. Het doel is om vanuit een experimenteel standpunt het mechanisme van band-tot-band tunneling volledig te begrijpen, en zo fabricatieparameters te optimaliseren en de meest performantie architectuur te achterhalen. Er wordt vertrokken van een silicium homojunctie TFETs, en vervolgens

wordt de source vervangen door silicium-germanium. Twee verschillende architecturen worden besproken: finFET’s en verticale nanodraden. Vermits de fabricatie van finFET al vergevorderd is, kan de invloed van verschillende ontwerpparameters onderzocht worden. De tweede architectuur met verticale nanodraden wordt gebruikt om heterostructuur TFETs te onderzoeken. Uiteindelijk wordt er een derde architectuur voorgesteld en geanalyseerd met TCAD simulaties.

List of Symbols

Symbol	Unit	Description
C_d	F/m^2	Diffusion capacitance
C_{eff}	F/m^2	
C_{inv}	F/m^2	Gate capacitance in inversion regime
C_{ox}	F/m^2	Gate oxide capacitance
C_s	F/m^2	Semiconductor capacitance
C_{tot}	F/m^2	Total transistor capacitance
D_{it}	$cm^{-2}eV^{-1}$	Interface-trap density
D_p	cm^2/s	Diffusion coefficient for holes
E_C	eV	Conduction Band
E_F	eV	Fermi level
$E_{F,n}$	eV	Fermi level in the n region
$E_{F,p}$	eV	Fermi level in the p region
E_G	eV	Semiconductor bandgap at 300K
E_i	eV	Intrinsic Fermi level
E_T	eV	Trap energy level
E_V	eV	Valence band
F	V/cm	Electric Field
F_n	-	Fermi-Dirac distribution function for electron
F_p	-	Fermi-Dirac distribution function for holes
h	$eV\ s$	Planck constant = 4.1357×10^{-15}
$\hbar$	$eV\ s$	Reduced Planck constant = 6.5821×10^{-16}
I_D	$A/\mu m$	Drain current

Symbol	Unit	Description
I_{Dsat}	$A/\mu m$	Drain current of a MOSFET in saturation
I_{OFF}	$A/\mu m$	Off-state current
I_{ON}	$A/\mu m$	On-state current
k	eV/K	Boltzmann constant = 8.6174 x 10^{-5}
kT/q	V	Thermal voltage = 0.0259 (at 300K)
m^*	kg	Effective mass
m_C^*	kg	Electron effective mass in conduction band
m_V^*	kg	Light holes effective mass in valence band
m_r	kg	Reduced effective mass
N_A	cm^{-3}	Acceptor impurity concentration
N_C	cm^{-3}	Effective density of states in conduction band
N_V	cm^{-3}	Effective density of states in valence band
n_i	cm^{-3}	Intrinsic carrier concentration
q	C	Elementary charge = 1.60218 x 10^{-19}
T_{inv}	nm	Inversion layer thickness
V_{DD}	V	Supply voltage
V_{DS}	V	Voltage across drain and source
V_{GS}	V	Voltage across gate and source
V_{th}	V	Subthreshold voltage
W_D	nm	Depletion width
ε_0	F/cm	Dieletric permittivity in vacuum = 8.85418 x 10^{-14}
ε_{SiO_2}	F/cm	Dieletric permittivity of SiO_2
κ	cm^{-1}	Imaginary wave vector
τ_g	s	Carrier generation lifetime
τ_p	s	Carrier lifetime for holes

Symbol	**Unit**	**Description**
ϕ_m	V	Metal work function
$\phi_{B,n}$	V	Schottky barrier height for electrons
$\phi_{B,p}$	V	Schottky barrier height for holes
χ	V	Electron affinity
ψ_s	V	Surface potential

List of Acronyms

Acronym	**Description**
AC	Alternative Current
ALD	Atomic Layer Deposition
APF	Advanced Patterning Film
BEOL	Back End Of Line
BOX	Buried Oxide
BTBT	Band-to-Band Tunneling
CD	automatic Chromatic Dispersion measurements
CESL	Contact Etch Stop Layer
CET	Capacitance Equivalent Thickness
CMOS	Complementary Metal Oxide Semiconductor
CMP	Chemical Mechanical Planarization
CV	Capacitance Voltage
DC	Direct Current
DIBT	Drain Induced Barrier Thinning
DLTS	Deep Level Transient Spectroscopy
DIBL	Drain Induced Bias Lowering
DUV	Deep Ultraviolet lithography
EBL	Electron-Beam Lithography
EDS	Energy-Dispersive X-ray Spectroscopy
EI	Electrostatic Integrity
EOT	Equivalent Oxide Thickness
FD	Fully Depleted
FEOL	Front End Of Line
FGA	Forming Gas Anneal
FN	Fowler-Nordheim tunneling
GAA	Gate-All-Around FET

Acronym	**Description**
GIDL	Gate Induced Drain Leakage
HAADF	High Angle Annular Dark Field
HDD	Highly-Doped Drain
HDP	High Density Plasma
HF	Hydro-Fluoric acid
HM	Hard Mask
HP	High-Performance
HR-TEM	High-Resolution Transmission Electron Microscopy
IMOS	Impact-Ionization MOSFET
ITP	Intrinsic Transistor Performance
LA	Laser Anneal
LDD	Lightly-Doped Drain
LER	Line Edge Roughness
LOCOS	Local Oxidation of Si
LP	Low Power
LSTP	Low Standby Power
MOCVD	Metal-Organic Chemical Vapor Deposition
FET/MOSFET	Metal Oxide Semiconductor Field-Effect Transistor
MuGFET	Multiple-Gate FET
nMOS	n-type MOSFET
NDR	Negative Differential Resistance
NEMFET	Nano Electromechanical FET
OPC	Optical Photolithography corrections
pMOS	p-type MOSFET
PD	Partially Depleted
PE-ALD	Plasma-Enhanced ALD
PECVD	Plasma Enhanced Chemical Vapor Deposition
Poly	Polycrystalline Silicon
POR	Process Of Reference
RTA	Rapid Thermal Annealing

Acronym	Description
SBFET	Schottky Barrier FET
SCE	Short Channel Effect
SEG	Selective Epitaxial Growth
SEM	Scanning Electron Microscopy
SIMS	Secondary Ion Mass Spectroscopy
SOI	Silicon-On-Insulator
SCE	Short Channel Effect
SPER	Solid-Phase Epitaxial Regrowth
SRH	Shockley-Read-Hall
SRP	Spreading Resistance Probe
SS	Inverse Subthreshold Slope factor
SSRM	Scanning Spreading Resistance Measurement
STEM	Scanning Transmission Electron Microscopy
STI	Shallow Trench Isolation
STT	Surface Tunneling Transistor
TAT	Trap Assisted Tunneling
TCAD	Technology Computer-Aided Design
TFET	Tunnel Field Effect Transistor
VLS	Vapor Liquid Solid

Table of Contents

Chapter 1

Introduction

1.1 MOSFET Scaling Issues

The world of electronics has been revolutionized by the introduction of the integrated circuit (IC), an electronic circuit manufactured by patterning a semiconductor material and additional material to form interconnections between the semiconductor devices. The basic component of an integrated circuit is the transistor.

The most common transistor used in state-of-the-art electronics is the Metal Oxide Semiconductor Field Effect Transistor (MOSFET). The basic principle of a MOSFET was introduced and patented by Lilienfeld in 1925 [1] but it was first fabricated at BELL Labs by M.M.Atalla and D. Kahng only in 1960 [2]. The MOSFET consists of three terminals where the electrons or holes flow between two terminals, the drain and the source, is controlled by the gate terminal as show in Figure 1.1.

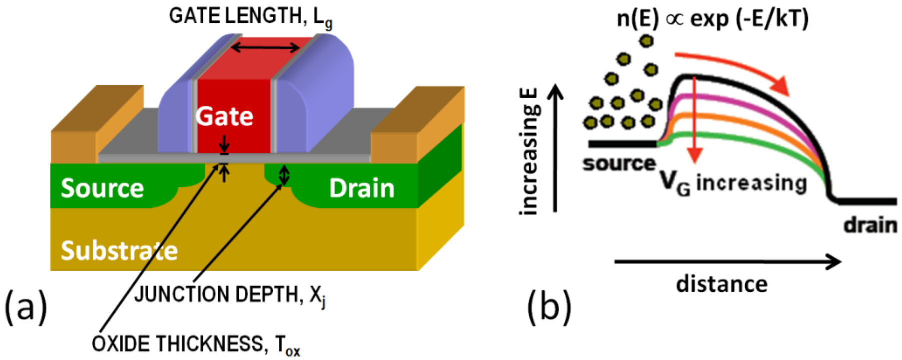


Figure 1.1: Schematic of (a) a bulk MOSFET and (b) modulation of the conduction band by the gate voltage in case of nMOS showing the basic carrier injection mechanism in MOSFET.

The simplicity of its principle of operation and design makes the MOSFET an extremely scalable device. By continuously reducing its dimensions and improving its performance, integrated circuits became smaller, faster and cheaper. This paradigm was already understood by Gordon Moore, co-founder of Intel, in 1960. It was translated into the well-known “Moore’s Law” [3] which predicted that the number of transistors on a chip would double approximately every two years (Figure 1.2). This statement together with Dennard’s scaling rules [4], which consists in a series of relationships where the electric field in the device is kept constant by reducing the device dimensions, were the main guidelines for the semiconductor industry for the past decades.

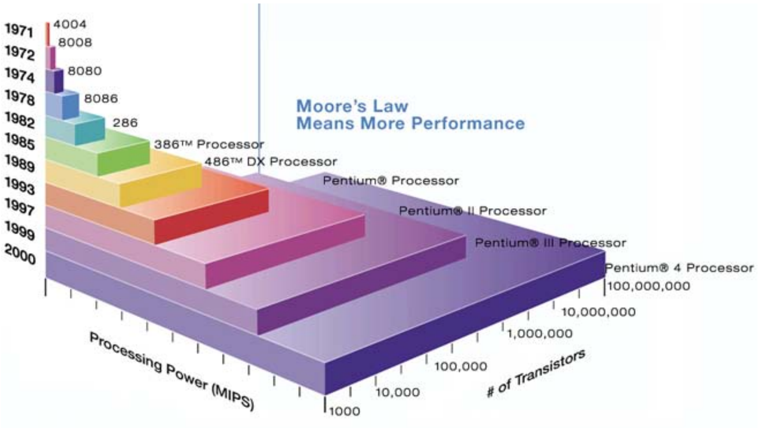
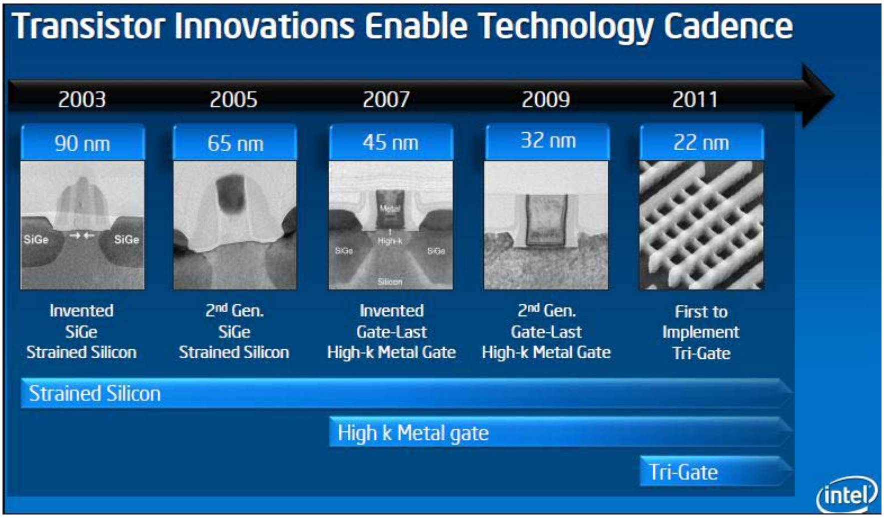


Figure 1.2: The growth of the complexity and processing power powered by the Moore’s law. The intel processors are used as a milestones for this process. (Source: Intel)

In the last decade, geometrical scaling was not sufficient to improve the device performance since the shrinking of the transistor would lead to short channel effects (SCEs). As a result, new materials and integration solutions offered new opportunities to improve the transistor performance and suppress SCEs. In Figure 1.3, some of the milestones of the evolution of the transistor are presented: the introduction of strain in the channel by SiGe S/D stressors to increase the carriers’ mobility, the implementation of high-K and metal gate to reduce the gate leakage and finally the introduction of a new architecture (finFET) to further suppress the SCEs and reduce the threshold voltage (V_{th}) roll-off.



Transistor Innovations Enable Technology Cadence

2003	2005	2007	2009	2011
90 nm	65 nm	45 nm	32 nm	22 nm
Invented SiGe Strained Silicon	2nd Gen. SiGe Strained Silicon	Invented Gate-Last High-k Metal Gate	2nd Gen. Gate-Last High-k Metal Gate	First to Implement Tri-Gate

Strained Silicon

High k Metal gate

Tri-Gate

Figure 1.3: The challenges of CMOS scaling beyond 90nm node. New materials and architectures need to be implemented to increase the performance. (source: Intel)

1.2 Power Crisis

The aggressive CMOS scaling increases not only the complexity of the new transistor but also the dissipated power. The total dissipated power can be divided in two components, the dynamic power and the static power, written as:

$$P_{TOT} = P_{dyn} + P_{static} \sim C_{TOT} V_{DD}^2 f + I_{OFF} V_{DD} \tag{1.1}$$

The dynamic power depends on the square of the supply voltage (V_{DD}), the switching frequency (f) and the total capacitance (C_{TOT}), while the static power depends on the supply voltage and off-state current (I_{OFF}). Historically, off-state power dissipation used to be a minor issue since it was a trivial task to minimize the leakage current, such as gate leakage or junction leakage, due to the large V_{th} and EOT. To lower the dissipated power, the supply voltage was reduced by reducing the threshold voltage (V_{th}) of the device and maintaining the same gate voltage overdrive. The introduction of deep submicron gate length and the difficulty to scale the supply voltage below 1V resulted into an increase of the leakage current and the static power followed along. This trend is demonstrated in Figure 1.4 where the two main components of the dissipated power are plotted for various logic technologies. If no measures are taken the total power will soon be dominated by the static power.

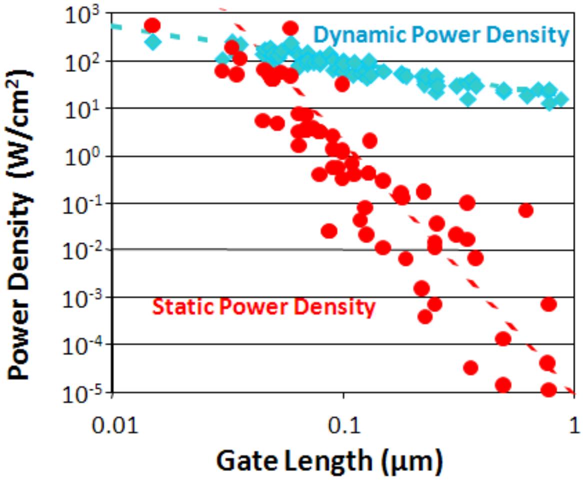


Figure 1.4: Power density versus gate length. The power density is divided into static and dynamic power [5].

Underlying the exponential increase in the static power is the concept of subthreshold slope (S). A practical interpretation of the subthreshold slope is the gate voltage required to change the drain current by one order of magnitude in the subthreshold regime. In case of MOSFET, the following expression can be written:

$$S = \frac{\partial V_g}{\partial (\log I_D)} = \frac{\partial V_g}{\partial \psi_S} \frac{\partial \psi_S}{\partial (\log I_D)} = \left(1 + \frac{C_s}{C_{ox}}\right) \frac{kT}{q} \ln 10 \qquad (1.2)$$

Where V_g is the gate voltage, I_d is the drain current, ψ_S is the surface potential, C_S is the semiconductor capacitance and C_{ox} is the gate oxide capacitance. In the hypothetical case of a perfect gate control, the ratio C_S/C_{ox} will be zero. The remaining kT/q term evaluates to roughly 60 mV/dec at room temperature, which is the wisely quoted MOSFET subthreshold slope barrier. Unfortunately, this subthreshold slope does not scale at all, since it is basically linked to the Boltzmann carrier distribution in the conduction band (or valence band) of the semiconductor. For this reason, the MOSFET leakage current increases exponentially as V_{th} reduces (Figure 1.5).

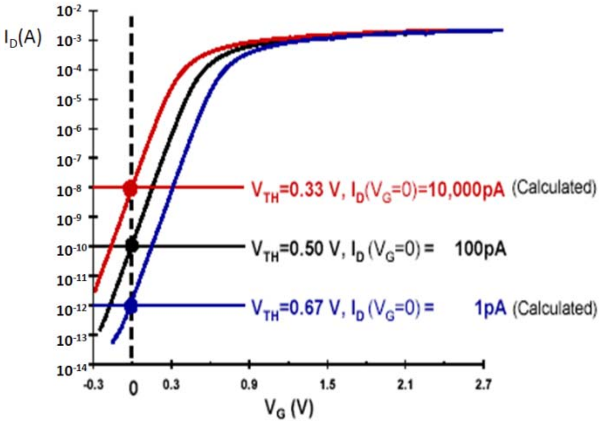


Figure 1.5: I_D versus V_G characteristic of an ideal MOSFET showing the increase of off current by reducing the V_{th}. [6]

If we want to solve the power crisis described above, we have to be able to break through the barrier of this subthreshold slope, so that we can further lower the threshold voltage and the power supply voltage without jeopardizing the off- and on-state current of the device. This quest for a new energy efficient switch is discussed in the next section.

1.3 The Quest for an Energy-Efficient Switch

The quest for a new digital switch is not new. Similar power constraints induced the development of bipolar devices, the first solid state transistors, to replace the vacuum tubes and mechanical switches in the late 1940s. Similarly, the current FET replaced bipolar transistors in the majority of semiconductor applications in the late 1980s.

As explained earlier, the MOSFET has a fundamental limit to the steepness of the subthreshold slope because of the temperature dependence of the thermionic injection of the carriers over an energy barrier [7]. The Holy Grail is to replace the MOSFET with a device able to mimic a nearly “ideal switch”, which is able to more abruptly reach the on current state with a subthreshold slope of almost zero.

The quest for an “ideal switch” should start by looking at the equation (1.2) of the subthreshold slope. Here, two different components can be extracted:

$$m = \frac{\partial V_g}{\partial \psi_S} \tag{1.3}$$

$$n = \frac{\partial \psi_S}{\partial (\log I_D)} \tag{1.4}$$

The term m is the body factor which is larger or equal to 1 in case of the MOSFETs and the term n that describes the change of drain current with the surface potential, reflecting the conduction mechanism in the channel. In case of MOSFET, the last term represents the diffusion term that limits the subthreshold slope for MOSFET to 60mV/dec at room temperature. As a consequence, two different approaches can be used to improve the subthreshold slope of the transistor: *scaling of the body factor* or introducing a *new principle of operation* to replace thermionic injection.

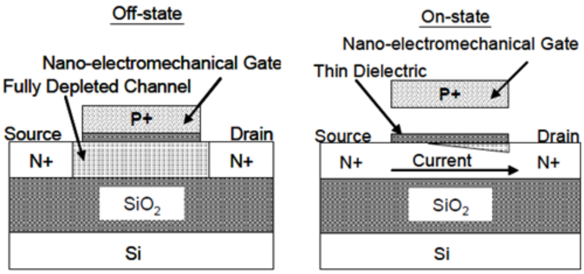


Figure 1.6: Schematic of the on- and off-state of a NEMFET [8].

The reduction of the body factor could be achieved by using Nano-electro-mechanical (NEM) switch [8-10] in which the on- and off-state are defined by the transition between two unstable mechanical states. The working principle of NEM-FET is to deplete the channel by placing the gate on the channel in off-state while the on-state is achieved by lifting the gate up (Figure 1.6). The drawbacks of this device are the high pull-up voltage and the low switching speed, limited by the mechanical component of the gate. Another concern is the difficulty to achieve high endurance because of surface wear and shock-induced failure. In conclusion, their implementation in conventional CMOS processing is extremely challenging by the presence of airgaps.

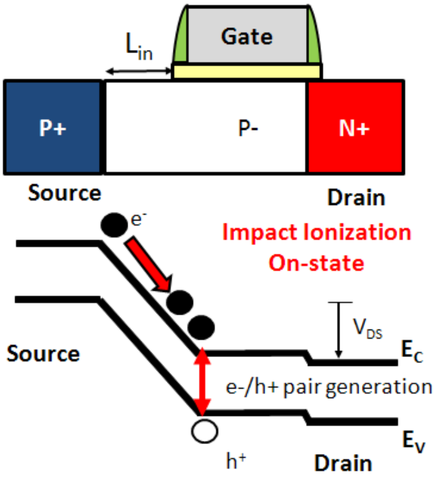


Figure 1.7: Basic device structure for the n-channel I-MOS and energy band diagram in on-state to explain the principle of operation.

Alternatively, the introduction of a new transport mechanism looks more promising and opens up a new research fields for the semiconductor industry. In the early stage, the impact ionization devices (IMOS) attracted a wide interest because of their extremely abrupt switching behavior [11-13]. The IMOS is a p-i-n diode in which the intrinsic area is partially covered by a gate (Figure 1.7). Impact ionization replaces the thermionic emission as the carrier injection mechanism and reduces the subthreshold swing down to few milliVolts per decade at room temperature. To achieve impact ionization, a strong electric field is required by reverse biasing the diode. For this reason, significant hot carrier effects cause threshold voltage instability and degradation in the subthreshold slope with repeated measurements. Nevertheless, the main drawback of this device is the low speed as shown in [14]. An alternative to IMOS is the Tunnel Field Effect Transistor (TFET), where the carrier injection mechanism is the band to band tunneling (BTBT). More details are discussed in the next section.

1.4 The TFET: from Concept to Device

The TFET consists of a gated reversed biased p-i-n diode where the gate is placed on the top of the intrinsic region. The concept of the TFET is as simple as shown in Figure 1.8 where the carrier injection mechanism of the conventional MOSFET is compared to the TFET. Contrary to the MOSFET where the carriers go over a barrier, in the TFET the carriers are tunneling through the barrier and, more precisely, electrons from the valence band are injected into the conduction band. The tunneling is controlled by the gate bias: when the gate is positive (or negative) the tunneling occurs at the p+ side (or n+ side). As a consequence, the conduction is ambipolar and p-mode and n-mode operations are possible on the same device.

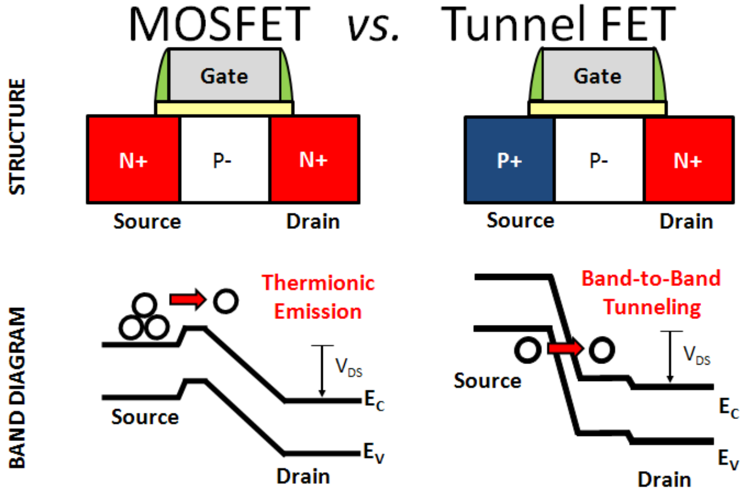


Figure 1.8: Schematic of MOSFET and TFET with relative band structure to explain the main carrier injection mechanism.

The gated p-i-n device was originally proposed to measure the subband splitting and transport properties of tunneling between bulk and a 2-D surface channel [15]. Later, it was studied for creating a multi-value logic in which the gate was used to control the negative differential resistance (NDR) and it was called Surface Tunneling Transistor (STT)[16-19]. The focus of the STT in this period was on the field control of the Esaki characteristic of the tunneling junction. Reddick and Amaratunga [20] introduced the first Si-based STT focusing their attention on the reverse biased characteristic, the Zener tunneling or TFET. Later, the discussion of the low subthreshold swing in the TFET held to several publications [21-27] which attracts an interest towards the study of this energy-efficient device.

At first glance, the complementary TFETs have an abrupt transition from 1 to 0 states. However, the transient response of a TFET inverter is worse compared to the conventional CMOS. In particular, Mookerjea et al. found that the gate-drain capacitance, C_{GD}, is larger than the gate capacitance, C_{GG}, because of the low resistance between channel and drain in contrast to the conventional MOSFET. For this reason, the effective load capacitance for unloaded TFET inverters can be more than twice the gate capacitance because of the Miller effect. This can degrade the delay time and generate overshoots. As a consequence, the TFETs are attractive for low-standby power applications. Looking at the International Technology Roadmap of Semiconductor (ITRS), it is clear that the conventional MOSFET needs to be replaced since it is not possible to reduce the saturation threshold voltage below 0.5V for low power applications. On the other hand, we expect the TFET to be able to operate at 0.3V of supply power with I_{ON}/I_{OFF} ratio of about 5 decades and the on current around 100 µA/µm. These requirements will make the TFET competitive since no solutions have been forecasted for conventional MOSFET. These specifications offer an overview of the target operation range of TFET. Contrary to the MOSFET, we use a different metric based on the variation of device current for a fixed voltage range as described in chapter 4.

1.5 Research Objective and Dissertation Outline

This dissertation proposes a critical study of the Tunnel Field Effect Transistor (TFET) as a new device concept and gives insights about its possible manufacturability and performance. The main objectives of this dissertation can be summarized as follows:

- to design and fabricate TFETs based on conventional gated diode architectures using both a planar and a vertical approach;

- to analyze the fabricated device through extensive electrical characterization and better understand the physics of its operation;
- to study and implement a hetero-junction TFET to boost its performance;
- to propose novel architectures for optimized TFET design.

The thesis is outlined as follows.

Chapter 2 presents the main transport mechanisms in p-i-n diodes including Band-to-Band Tunneling (BTBT). The understanding of additional carrier transport mechanisms present in the device is critical to properly design TFETs. In particular, temperature dependent mechanisms such as Trap-assisted Tunneling (TAT) can degrade the device performance. The complexity of predicting the BTBT is demonstrated by the mismatch between simulation results and experimental results. In fact, commercial simulator tools have several limitations and are not able to fully model quantum mechanical phenomena such as tunneling.

Chapter 3 presents the fabrication of TFET using CMOS compatible processing. This chapter is divided in two parts: (1) the fabrication of Finfet-based TFET, or Multiple Gate Tunnel Field Effect Transistor (MuGTFET) and (2) the fabrication of Gate-All-Around (GAA) TFET, based on vertical nanowires. This chapter also covers the work related to the new mask designs.

Chapter 4 introduces the measurement methodologies used to characterize TFET and, then, presents the TFET from a device point of view. In particular, the conventional TFET configuration is discussed together with its electrical characteristics and methods to identify the BTBT.

Chapter 5 deals with the electrical characterization of MuGTFETs. The main results on full Si TFET are summarized highlighting the impact of various geometrical and processing parameters. In addition, some considerations about the silicide engineering are presented.

Chapter 6 deals with the electrical characterization of nanowire-based vertical TFET. Both Si and SiGe source hetero junction TFETs are studied and compared

In chapter 7, a critical discussion is made to assess the future of TFET based on the conclusions of the two previous chapters. Based on TCAD simulations, a new configuration of the TFET is proposed: the hybrid TFET. The simulations predict a very steep subthreshold swing and an improvement of on-current for the hybrid TFET compared to the conventional TFET configuration analyzed in the previous chapters.

In chapter 8, the main conclusions from this work are summarized, including insights regarding the manufacturability of these devices. In addition, an outlook for further development of TFET is included with focus on the introduction of new materials to continue the CMOS roadmap beyond the 14nm node.

References

[1] J. E. Lilienfeld, "Method and apparatus for controlling electric current," Canada Patent CA272437, 1927, 1925.

[2] K. Dawon, "Electric Field Controlled Semiconductor Device," US Patent US3102230, 1963, 1960.

[3] G. E. Moore, "Lithography and the Future of Moore's Law," *Advances in Resist Technology and Processing XII,* vol. 2438, pp. 2-17, 1995.

[4] R. H. Dennard, *et al.*, "Design of ion-implanted MOSFET's with very small physical dimensions," *IEEE Journal of Solid-State Circuits*, vol. 9, pp. 256-268, 1974.

[5] B. Meyerson, in *(IBM) Semico Conf.*, 2004.

[6] D. Cox, " IRPS Tutorial ", 2004.

[7] S. M. Sze, *Physics of Semiconductor Devices: 3rd Edition*. New York: John Wiley & Sons, 2007.

[8] K. Hei, *et al.*, "A new nano-electro-mechanical field effect transistor (NEMFET) design for low-power electronics," in *IEEE International Electron Devices Meeting*, 2005, pp. 463-466.

[9] F. Chen et al., "Integrated circuit design with NEM relays," in *ICCAD*, 2008, pp. 750-757.

[10] K. A. e. al., "Design considerations for complementary nanoelectromechanical logic gates," in *IEDM Tech. Dig.*, 2007, pp. 299-302.

[11] K. Gopalakrishnan, *et al.*, "Impact ionization MOS (I-MOS) Part I: device and circuit simulations," *IEEE Transactions on Electron Devices,* vol. 52, pp. 69-76, 2005.

[12] K. Gopalakrishnan, *et al.*, "Impact ionization MOS (I-MOS) Part II: experimental results," *IEEE Transactions on Electron Devices*, vol. 52, pp. 77-84, 2005.

[13] F. Mayer, *et al.*, "Static and Dynamic TCAD Analysis of IMOS Performance: From the Single Device to the Circuit," *IEEE Transactions on Electron Devices*, vol. 53, pp. 1852-1857, 2006.

[14] C. Shen, *et al.*, "On the Performance Limit of Impact-Ionization Transistors," in *Electron Devices Meeting, 2007. IEDM 2007. IEEE International*, 2007, pp. 117-120.

[15] J. J. Quinn, *et al.*, "Subband spectroscopy by surface channel tunneling," *Surface Science,* vol. 73, pp. 190-196, 1978.

[16] T. Baba and T. Uemura, "GaAs surface tunnel transistors with a symmetric source-to-drain structure," *Compound Semiconductors* vol. 145, pp. 877-882, 1996.

[17] T. Uemura and T. Baba, "First demonstration of a planar-type surface tunnel transistor (STT): Lateral interband tunnel device," *Solid-State Electronics,* vol. 40, pp. 519-522, 1996.

[18] T. Uemura and T. Baba, "Surface tunnel transistor: Gate-controlled lateral interband tunneling device," *Japanese Journal of Applied Physics Part 1-Regular Papers Short Notes & Review Papers,* vol. 35, pp. 1668-1675, 1996.

[19] J. Koga and A. Toriumi, "Three-terminal silicon surface junction tunneling device for room temperature operation," *IEEE Electron Device Letters,* vol. 20, pp. 529-531, 1999.

[20] W. M. Reddick and G. A. J. Amaratunga, "Silicon Surface Tunnel Transistor," *Applied Physics Letters,* vol. 67, pp. 494-496, Jul 1995.

[21] C. Aydin, *et al.*, "Lateral interband tunneling transistor in silicon-on-insulator," *Applied Physics Letters,* vol. 84, pp. 1780-1782, 2004.

[22] K. K. Bhuwalka, *et al.*, "Scaling issues of n-channel vertical tunnel FET with δ–p^+ SiGe layer," in *Device Research Conference Digest*, 2004, pp. 215-216 vol.1.

[23] P. F. Wang, *et al.*, "Complementary tunneling transistor for low power application," *Solid-State Electronics*, vol. 48, pp. 2281-2286, 2004.

[24] K. K. Bhuwalka, *et al.*, "P-channel tunnel field-effect transistors down to sub-50 nm channel lengths," *Japanese Journal of Applied Physics Part 1-Regular Papers Brief Communications & Review Papers,* vol. 45, pp. 3106-3109, 2006.

[25] J. Appenzeller, *et al.*, "Comparing carbon nanotube transistors - The ideal choice: A novel tunneling device design," *IEEE Transactions on Electron Devices,* vol. 52, pp. 2568-2576, 2005.

[26] J. Knoch and J. Appenzeller, "A novel concept for field-effect transistors - the tunneling carbon nanotube FET," in *Device Research Conference Digest*, 2005, pp. 153-156.

[27] A. S. Verhulst, *et al.*, "Tunnel field-effect transistor without gate-drain overlap," *Applied Physics Letters*, vol. 91, pp. 053102-3, 2007.

Chapter 2

Physical Theories

2.1 Introduction

Quantum mechanical tunneling through a potential energy barrier is well known in the field of semiconductor devices. For instance, models for Fowler–Nordheim (FN) and direct tunneling through a gate insulator in MOSFETs are in good agreement with the experimental data [1-3]. Particularly, Band-to-Band Tunneling (BTBT) has been studied to characterize the Gate Induced Drain Leakage (GIDL) [4] in MOSFETs.

This chapter reviews the most general tunneling framework and focuses on the BTBT process, the main transport mechanism in Tunnel-FETs. However, other mechanisms such as Trap-Assisted Tunneling (TAT) and Shockley-Read-Hall (SRH) generation currents are also present and included in this chapter to offer a complete overview of the possible transport mechanisms present in TFETs.

Later, the discussion of BTBT in commercial Technology Computer-Aided Design (TCAD) simulators presents the limitations of the modeling of band-to-band tunneling and their domain of validity following a semiclassical approach.

2.2 Band-to-Band Tunneling

Tunneling is a quantum mechanical phenomenon where an electron moves though a potential barrier. Classically, an electron cannot move though a potential barrier because this process requires the particle to have a negative kinetic energy (i.e. conservation of the energy). In case of tunneling, this negative kinetic energy is expressed by an imaginary wave vector, κ, and results in a wave with an amplitude which decays exponentially [5]. The transmission probability, *T(E)*, through the barrier depends on both the width and the height of the barrier as depicted in Figure 2.1.

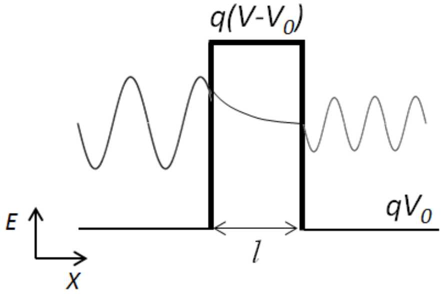


Figure 2.1: Schematic of the tunneling through a potential barrier. In case of a thin and finite potential barrier, the wave function will be able to move though the barrier with attenuated amplitude.

In the case of a homogenous pn junction, and similarly in the case of TFET, the barrier height is represented by the energy bandgap of the material while the barrier width, *l*, is called the tunneling path length. The tunneling through the forbidden band (Figure 2.2a) is formally equivalent to a particle tunneling through a potential barrier (Figure 2.2b), where the barrier can be approximated by a triangular barrier. Not all the electrons in the valence band of the p region can tunnel to the conduction band of the n region. The necessary conditions for an electron to tunnel from the valence band of the p side to the conduction band of the n+ side are:

- the presence of unoccupied states at the same energy level in the conduction band, in particular only electrons with an energy between the Fermi levels, E_{FP} and E_{FN}, are allowed to tunnel (when considering the approximation of the Fermi Dirac statistics at 0K),
- the barrier width and height should be small enough to have a non-zero tunneling probability,
- the momentum must be conserved in the tunneling process.

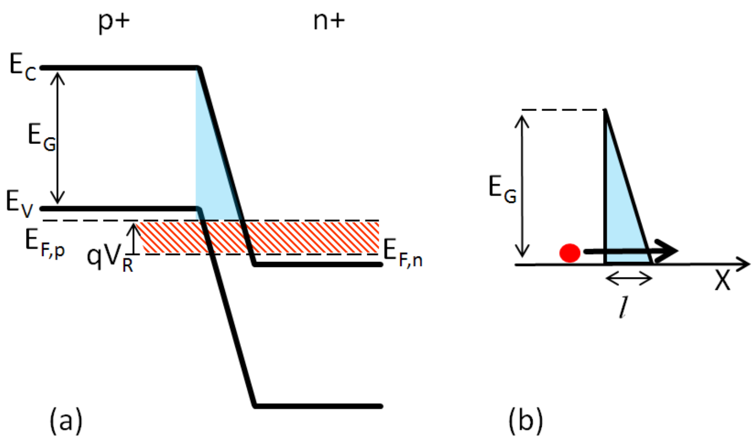


Figure 2.2: Energy bands of a degenerate pn junction when a small reverse bias is applied (a) and equivalent representation of the tunneling through a triangular barrier (b) where the barrier height is about the energy bandgap (E_G) and the width is represented by the tunneling distance, *l*. The shade area indicates the energy window where the tunneling occurs.

These statements can be generalized in the following equation describing the total tunneling current for the 1-dimensional case [6]:

$$I_{BTBT} = -\frac{2e}{h} W \int_{E_{F,n}}^{E_{F,p}} T(E)\,[F_n(E_{x=l}) - F_p(E_{x=0})]dE = \tag{2.1}$$

with

$$F_n(E_{x=l}) = \left(exp\left[\frac{E - E_{F,n}(l)}{kT}\right] + 1\right)^{-1} \tag{2.2}$$

$$F_p(E_{x=0}) = \left(exp\left[\frac{E - E_{F,p}(0)}{kT}\right] + 1\right)^{-1} \tag{2.3}$$

where W is the width of the device, F_n and F_p are the Fermi distribution function in the n region and p region, respectively. From equation (2.1) is clear that TFET can be seen as an energy filter because the carriers at the tails of the Fermi distribution are not allowed to tunnel through the barrier. As a consequence, a sub 60mV/dec subthreshold slope is possible.

2.2.1 The Kane's Model and its Limitations

One of the first expressions for the band to band generation rate was derived by Kane [7]. The complex mathematics involved in the derivation of the final expression can be reviewed in [7] and leads to this expression, for a direct semiconductor and a uniform electric field, *F*:

$$G_{Kane} = \frac{q^2\sqrt{m^*}F^2}{18\pi\hbar^2\sqrt{E_G}} exp\left(-\frac{\pi\sqrt{m^*}E_G^{3/2}}{2\sqrt{2}q\hbar F}\right) \tag{2.4}$$

where $\hbar$ is the reduced Plank's constant, m^* is the effective mass of the electrons, E_G is the energy bandgap and q is the electron charge. This equation shows that the tunneling has an exponential dependence on the electric field, *F*. In addition, there is no implicit temperature dependence in this expression. For instance, the dependence with the temperature of the energy bandgap is very small and it can often be neglected (Figure 2.3).

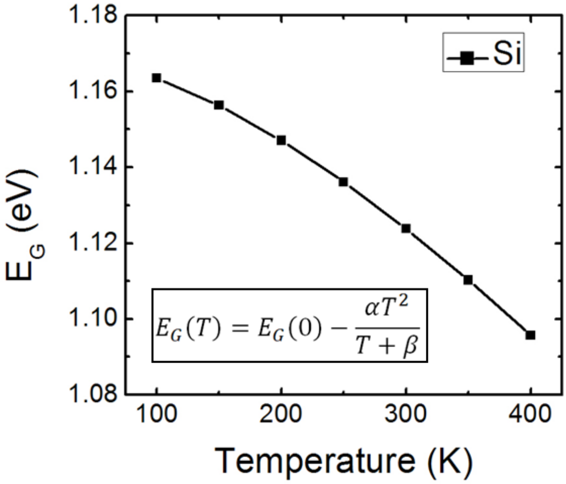


Figure 2.3: Variation of the energy bandgap of Silicon with temperature [8].

To describe the tunneling phenomenon, the basic expression of Kane’s formula expressed in equation (2.4) is often reduced to its functional form [9]:

$$G_{Kane} = AF^2 \exp\left(-\frac{B}{F}\right) \tag{2.5}$$

where A and B are parameters depending on the bandgap, E_G, and the effective masses of the conduction and valence bands. The validity of Kane’s formula is limited to uniform electric field along the tunneling path. For this reason, this formula is often referred to as a local approximation. BTBT is intrinsically a non local phenomenon since the tunneling occurs between two different locations in the semiconductor region. Local and non local models converge to similar results when an average field is used to express F.

2.2.2 Tunneling Probability and Imaginary Wave Vector

In case of non uniform electric field and when the potential varies slowly, an alternative approach can be used, such as the Wentzel-Kramers-Brillouin (WKB) approximation. Starting from the analytical solution of the Schrödinger equation for a rectangular barrier and using proper boundary conditions, the one-dimensional tunneling probability can be expressed as:

$$T_{WKB} \approx exp\left(-2\int_{x_1(E)}^{x_2(E)} \kappa(x,E)dx\right) \tag{2.6}$$

where x_1 and x_2 are the starting and the ending positions of the tunneling path, κ is the imaginary wave vector which characterizes the amount of wave function decay within the barrier. The exact shape of the barrier is not a concern but the imaginary wave vector dispersion relation, *E(k),* within the energy forbidden gap (i.e. the barrier) needs to be specified.

The dispersion relationship can be calculated in many different ways. In one dimensional space, the simplest expression for the energy dispersion relationship is the parabolic one or 1-band dispersion, where a parabola centered at the $k=0$ (Γ point) is considered [9] (assuming x as the direction of tunneling):

$$\kappa(E) = \frac{1}{\hbar}\sqrt{2\mathrm{m}^{*}\ (\mathrm{E}-\mathrm{E_V})} \tag{2.7}$$

The imaginary wave vector is supposed to be zero at the valence band and the conduction band as expected for a traveling electron wave, but equation (2.7) is zero only at the valence band. In addition, the tunneling effective mass might be different at the conduction band and valence band. To solve these issues, a 2-band asymmetric energy dispersion relationship is more suitable. Consequentially, the imaginary wave vector relation needs to be modified including the following equations [10]:

$$\kappa_V(E) = \frac{1}{\hbar}\sqrt{2\mathrm{m}_\mathrm{V}^{*}(E-E_V)} \tag{2.8}$$

$$\kappa_C(E) = \frac{1}{\hbar}\sqrt{2\mathrm{m}_\mathrm{C}^{*}(E_\mathrm{C}-E)} \tag{2.9}$$

$$\kappa(E) = \min\,(\kappa_V, \kappa_C) \tag{2.10}$$

where $\mathrm{m}_\mathrm{C}^{*}$ is the electron effective mass and $\mathrm{m}_\mathrm{V}^{*}$ is the light holes effective mass. The equation (2.10) allows the tunneling particle to move from κ_V to κ_C during the tunneling process.

In order to obtain a close form of the tunneling probability, we consider the following dispersion relationship describing a particle that moves through a potential barrier *U(x)* [9]:

$$\kappa(E) = \frac{1}{\hbar}\sqrt{2m_r(U(x)-E_X)} \tag{2.11}$$

where,

$$m_r = \left(\frac{1}{\mathrm{m}_\mathrm{V}^{*}} + \frac{1}{\mathrm{m}_\mathrm{C}^{*}}\right)^{-1} \tag{2.12}$$

In case of a bulk semiconductor in a three dimensional system, it is possible that the tunneling electron has part of its total momentum directed transverse to the tunneling direction. For this reason, the total energy of the tunneling electron consists of two components: E_X, energy along the tunneling direction and $E_\perp$, the transverse energy that is conserved during the tunneling process. In case of a triangular barrier, we can write:

$$U(x) = qFx + E = qFx + E_X + E_\perp \tag{2.13}$$

Considering that $l = {}^{E_G}\!/_{qF}$, we can combine the equations (2.6), (2.11) and (2.13) to reach a close formula of the tunneling probability as follows:

$$T(E, E_T) \cong \exp\left(-\frac{4}{3}\frac{\sqrt{2\mathrm{m_r} E_G^3}}{\mathrm{q}\hbar F}\right)\exp\left(-\frac{E_T}{\bar{E}}\right) \tag{2.14}$$

where
$$\bar{E} = \frac{\mathrm{q}\hbar F}{2\sqrt{2\mathrm{m_r} E_G}} \tag{2.15}$$

The factor $\bar{E}$ describes the impact of the transverse energy states in the total tunneling probability. The tunneling probability is enhanced by reducing the energy bandgap and the tunneling mass while increasing the electric field at the tunneling junction.

When the extrema of the tunneling path do not belong to the same point in the k space, indirect tunneling occurs. In this case, a phonon is needed to supply the momentum necessary for the indirect transition and the tunneling probability will be adapted accordingly to [11].

2.2.3 Concept of Effective Bandgap

So far we have discussed of tunneling from the valence band of one material to the conduction band of the same material. An interesting situation arises in case of tunneling across the interface between two different semiconductor materials. In particular, when the band offset of the valence band is staggered as the one depicted in Figure 2.4. In this case, the effective bandgap seen by the tunneling electron can be smaller than the one of either material, i.e. the source and channel material. As a consequence, we define the effective band gap as the different between the energy of the conduction band at the end of the tunneling path and the energy of the valence band at the beginning of the tunneling path in the off state. For instance, in Figure 2.4 the energy bandgap of the source, strained-Ge, is 0.58eV while an effective bandgap of 0.25eV is achieved when strained-Si is used as a channel material.

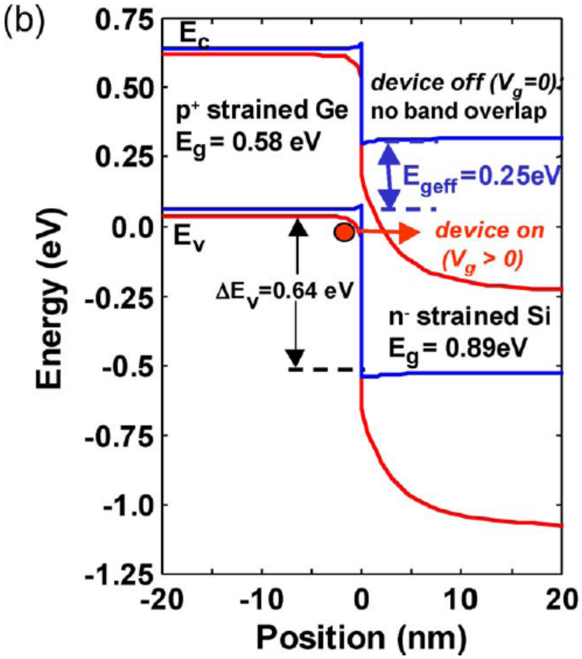


Figure 2.4: Energy band diagrams near the strained-Ge/strained-Si heterointerface of a simulated hetero junction TFET in ON and OFF state. [12]

The implementation of hetero-junction TFETs boosts the TFET performance because of the reduced energy band gap at the tunneling junction. In fact, the effective band gap represents the height of the effective tunneling barrier instead of the energy bandgap of the source or the channel.

2.3 Leakage Mechanisms in TFET

In off-state, the TFET behaves as a two terminal reversed-biased p-i-n diode. A good understanding of the transport mechanisms in reversed biased pn (or pin) diodes is necessary to distinguish between the various leakage mechanisms in TFET. In this section, we describe two leakage mechanisms of interest: (1) Shockley–Read–Hall (SRH) recombination and (2) Trap-Assisted Tunneling (TAT). Another mechanism which increases the off-state current is the band-to-band tunneling at the drain side and it results in an ambipolar transfer characteristics. This phenomenon will be discussed later in chapter 4.

2.3.1 Theory of an Ideal Diode

The pn junction theory was established by Shockley [13-14] in the early 1950s. The ideal current-voltage characteristics are based on four assumptions:

- abrupt depletion region;
- Boltzmann approximation is valid;
- low carrier injection;
- no generation and recombination in the depletion region.

Following these approximations the total current of an ideal diode can be expressed by:

$$J_{TOT} = J_p + J_n = J_0 \left[exp\left(\frac{qV}{kT}\right) - 1 \right] \tag{2.16}$$

Where J_0 is the reverse saturation current, q is the electron charge, V is the voltage difference applied at two terminals, k is the Boltzmann’s constant and T is the temperature.

Unfortunately, the Shockley equation can only qualitatively describe experimental diodes. The departures from the ideal case are mainly due to: (1) presence of generation and recombination processes in the depletion layer, (2) surface effects and (3) high injection effects. Since the TFET is always reverse biased, our attention focuses on the transport mechanisms for a reverse biased pn junction.

2.3.2 SRH Component

Recombination is a process which involves the exchange of carriers between the conduction and the valence band. The recombination assisted by deep trap levels in the energy bandgap is called Shockley–Read–Hall (SRH) recombination. Taken into account the SRH statistics, a compact formula for the total reverse current of a pn diode can be derived. It can be expressed as the summation of the diffusion and the generation component in the depletion region (for one side abrupt junction):

$$J_R = q\sqrt{\frac{D_p}{\tau_p}\frac{n_i^2}{N_D}} + \frac{q n_i W_D}{\tau_g} \tag{2.17}$$

where,

$$n_i = \sqrt{N_C N_V} exp\left(-\frac{E_G}{2kT}\right) \tag{2.18}$$

where D_p is the diffusion coefficient, n_i is the intrinsic carrier concentration, τ_p is the minority carrier lifetime, τ_g is the generation lifetime, N_D is the donor concentration, W_D is the width of the depletion region, N_C and N_V are the effective density of states in conduction and valence band, respectively. This component is independent of the voltage applied and is mainly responsible for the lowest value of current in TFETs. In case of a material with a small energy bandgap, the diffusion component will become dominant because of the high value of n_i. For this reason, semiconductors such as germanium have a dominant diffusion component and the diode characteristics are closer to the ideal case. On the other hand, for semiconductors such as silicon (Si) the generation component is dominant.

2.3.3 Trap-Assisted Tunneling (TAT)

The TAT is a tunneling process which involves the capture and the emission of an electron (or hole) by a defect level in the energy bandgap with energy equal to E_T considering only a single trap event as shown in Figure 2.5. The emission of the electron can either be through tunneling or thermally assisted. The thermally-assisted TAT is mainly considered in our work because of its temperature dependence as shown below.

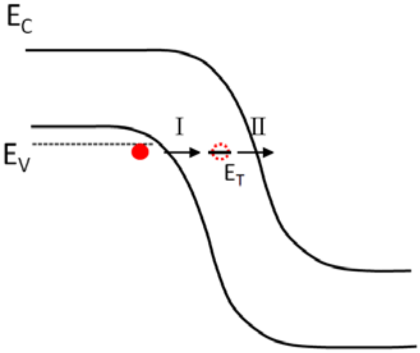


Figure 2.5: Schematic of a TAT process where an electron is captured (I) and is emitted (II) from a trap energy level equal to E_T.

Considering the thermally-assisted TAT and the Hurkx model [15-16], the SRH recombination is modified by the electric field enhancement factor Γ(x) according to [17]:

$$R_{trap}(x) = [1 + \Gamma(\mathrm{x})]R_{SRH}(x) \tag{2.19}$$

with
$$\Gamma(x) = \frac{\Delta E_T}{kT}\int_0^1 exp\left[\frac{\Delta E_T}{kT}u - \frac{4}{3}\frac{\sqrt{2m^*(\Delta E_T)^3}}{q\hbar|F|}u^{3/2}\right]du \tag{2.20}$$

where kT is the thermal energy, q the absolute value of the electron charge, $\hbar$ is the reduced Plank constant, $m*$ is the tunneling effective mass of the carriers, and ΔE_T is related to the trap level. The two terms inside the integral of the Eq. (2.20) indicates the thermal emission of the electron from the trap and the capture of the electron by tunneling. The first term is the one that determine the temperature dependence of the thermally-assisted TAT as a function of the energy level of the trap, ΔE_T. A non local approach might be preferred in case of TFET and it has been recently included in TCAD commercial simulators [18].

2.3.4 Arrhenius Plot

The Arrhenius plot is widely used to study the transport mechanisms in a diode, in particular to extract physical parameters [19-20], such as the activation energy, E_A. The general Arrhenius equation can be written in the following form:

$$J = A \cdot exp\left(-\frac{E_A}{kT}\right) \tag{2.21}$$

Diffusion and generation mechanisms in a reverse biased diode can be directly described noting that the two components described in equation (2.17) can be written in the following form:

$$J_{R,diff} \propto exp\left(-\frac{E_G}{kT}\right) \tag{2.22}$$

$$J_{R,gen} \propto exp\left(-\frac{E_G}{2kT}\right) \tag{2.23}$$

In case of TAT, the activation energy varies with the position of the energy trap level and, hence, it is impossible to describe the mechanism with a single value.

In case of silicon, typical value for the activation energy of SRH component is equal to half of the bandgap (~0.56 eV) since the generation component is dominant [equation (2.23)]. On the other end, the values for TAT are below midgap, in the range of 0.2~0.4eV for silicon [17]. In this work, we generally refer to BTBT as a process with activation energy below 0.1eV. This limit is purely arbitrary and it used to indicate a process with a weak temperature dependence such as the BTBT.

2.4 TCAD Simulation of TFETs

2.4.1 Introduction

In the past years, a lot of efforts have been made to improve the band-to-band tunneling models in commercial TCAD simulators. Several local models were introduced to quantify the GIDL component in sub-100nm MOSFETs but they were completely unable to describe correctly the BTBT in TFET since the field across the tunneling path is almost never constant. For instance, the electric field varies linearly in a region with constant doping concentration. As a consequence, the BTBT needs to be treated non locally.

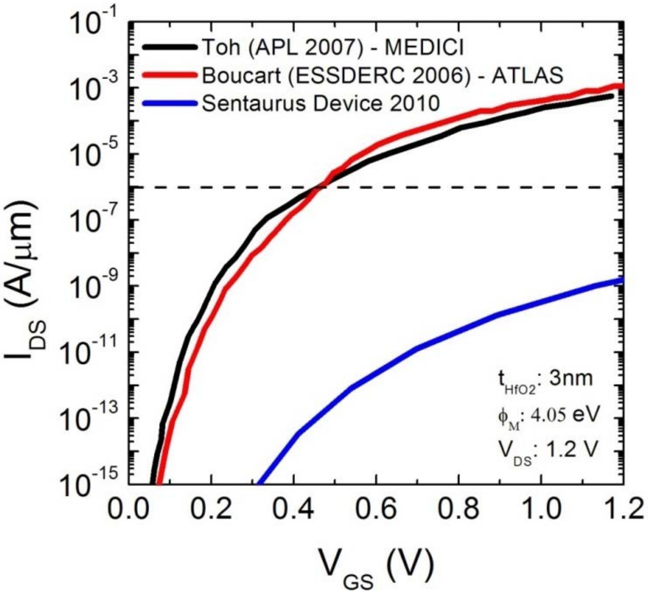


Figure 2.6: TCAD simulation of a TFET showing the different output in case of local direct and non local indirect (the Sentaurus Device curve) models. The device parameters are the same in all three simulations. The gate covers entirely the intrinsic region. [21-22]

Figure 2.6 shows three different simulations where local and non-local models have been implemented. The large disagreement between the curves is related to the inaccuracy of the local models and the fact that local direct models are used to describe tunneling in indirect material. The lack of a quantum mechanical solver in commercial simulator tools suggests to take with some cautions the outputs of the simulation. For this reason, a quantitative analysis is not straightforward and requires a calibration of the tunneling parameters.

2.4.2 BTBT in Sentaurus Synopsys

Synopsys Sentaurus was one of the first commercial TCAD tools able to offer a satisfactory self-consistent non-local BTBT model to compute the BTBT current in Si TFETs.

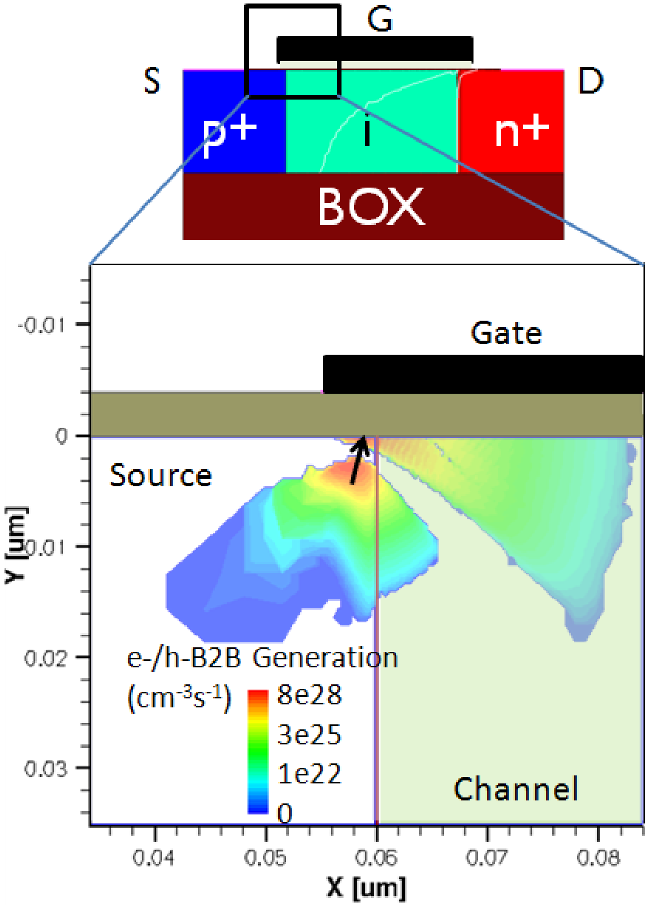


Figure 2.7: TCAD simulation output illustrating contours of BTBT generated holes and electrons at the ON state. The arrow indicates one of the possible tunneling paths. The direction of the tunneling is not entirely vertical or lateral because both lateral

and vertical field exists. The principal tunneling component originates a few nanometers below the oxide interface.

An advantage of this model is that the simulator dynamically searches for overlap between valence band and conduction band and then computes the generation rate along various tunneling paths, when calculating BTBT current (Figure 2.7) [18].

The tunneling path is a straight line with its direction opposite to the gradient of the valence band at the starting point while the tunneling energy is the difference between the valence band at the starting point and the conduction band at the ending position including an additional band offset as a parameter. When the tunneling path encounters Neumann boundaries or a semiconductor-insulator interface, it undergoes specular reflection if the option is not deactivated [18].

The dynamic calculation of the tunneling paths requires very fine mesh in the region where the tunneling occurs otherwise the results might be not accurate or have convergence issues. As a consequence, the mesh is automatically redefined according to the electric field calculated at each mesh points after evaluating the Poisson’s equation at ON state. This enables to exclude unnecessary mesh points and decreases the computation time. Nevertheless, the simulation time can still be important because the minimum mesh point is usually around 0.1~0.2nm. A brief introduction about the formalism of this model is given before introducing the tunneling parameters.

For a given tunneling path of length l starting at x = 0 until x = l, holes are generated at x=0 and electrons are generated at x = l. The net hole recombination rate at the starting point due to direct BTBT can be written as [18]:

$$\mathrm{R_{net}} = |\nabla \mathrm{E_V}(0)|\mathrm{C_d}\exp\left(-2\int_0^{\mathrm{l}} \mathrm{k}\, dx\right)\left[F_n(E_{x=l}) - F_p(E_{x=0})\right] \qquad (2.24)$$

The first term, the gradient of the valence band, defines the tunneling path while the last term describes the presence of empty states at the end of the tunneling path. The central term describes the tunneling probability as a result of the WKB approximation [see equation (2.6)] and it depends on the tunneling effective masses, the degeneracy factor, g, and the effective energy bandgap along the tunneling path. More details were given at the beginning of this chapter. For uniform electric field, equation (2.24) reduces to the well known Kane model [equation (2.5)].

The main fit parameters for the BTBT model are: the Kane's coefficient A, B, the energy offset for the conduction band, Δ_C, the phonon energy, ε_{op}, and the effective tunneling mass ratio, m_V/m_C. An important observation is that the terms A and B are used as input parameters to be consistent with the previous band-to-band tunneling models but they are only used to define the tunneling masses and the degeneracy factor used in equation (2.24) to calculate the tunneling generation rate. For direct transition, A and B are expressed as:

$$\mathrm{A} = \frac{\mathrm{g}\,\pi \mathrm{m}_\mathrm{r}^{1/2}(\mathrm{q}F_0)^2}{9h^2[\mathrm{E_G} + \Delta_\mathrm{C}]^{\frac{1}{2}}} \tag{2.25}$$

$$\mathrm{B} = \frac{\pi^2 \mathrm{m}_\mathrm{r}^{1/2}[\mathrm{E_G} + \Delta_\mathrm{C}]^{\frac{1}{2}}}{\mathrm{q}h} \tag{2.26}$$

In case of indirect transitions, the phonon energy is also considered in A and B. In this model, it is not possible to specify the tunneling masses directly but the reduced effective mass is calculated from the parameters A and B, then m_C and m_V, the tunneling masses of electrons and holes, are calculated from:

$$\frac{1}{m_C} = \frac{1}{2m_r} + \frac{1}{m_0} \tag{2.27}$$

$$\frac{1}{m_V} = \frac{1}{2m_r} - \frac{1}{m_0} \tag{2.28}$$

The presence of many fitting parameters indicates that an accurate calibration of the model is necessary.

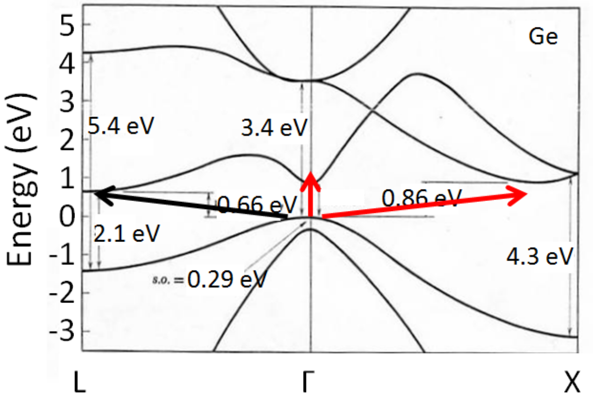


Figure 2.8: Energy band diagram of Germanium. The black arrow is indicating the common indirect transition while the red arrows indicate additional tunneling paths that can be implemented. The effective bandgap at X and Γ points are indicated.

The simulator tool allows multiple tunneling paths to consider tunneling at different valleys as illustrated in Figure 2.8. In this case, the value of Δ_C needs to be specified to consider the offset of the conduction band of the new location compared to the conduction band at the Γ, where the tunneling starts. As a consequence, $E_{G,tun} = E_G + \Delta_C$ is the effective band gap including the band offset.

2.4.3 Calibration of BTBT Models

The calibration of the parameters A and B is carried out by fitting the diode characteristics of pn junction composed by the target material. The latest version of Sentaurus provides parameters for Silicon and Germanium. The first are based on experimental data of pn diode [15-16] but for Germanium only theoretical results are available [23] which makes them less accurate. The verification of the Silicon parameters has been further proven by the analysis of pin diodes fabricated and characterized at imec [24]. Figure 2.9 shows a schematic of the fabricated diode. Diodes with different areas were implemented on 300mm wafers.

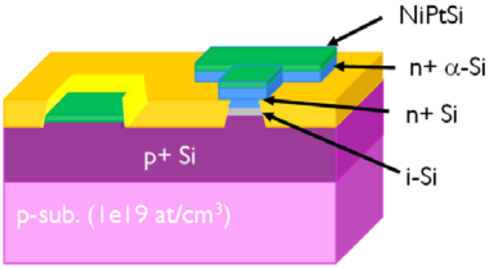


Figure 2.9: Schematic of the fabricated Si p-i-n diode. Only the p+ substrate is grown by blanket epitaxy. After trench isolation and partially etch back of the substrate, the rest of the layers are grown by epitaxy. Silicide is used as a contact material.

The presence of an “intrinsic” region simplifies the calibration process of TCAD models due to the presence of a nearly uniform electric field. Two of the main limitations of this approach are: (1) the control of the thickness of intrinsic layer and (2) the diffusion of the dopants in the intrinsic region. The electrical behavior of the diode is extremely sensitive to the variation of the thickness of the intrinsic region. In

Figure 2.10a, the electrical characteristics of diodes with the same dimensions are taken in several positions of the wafer. It is clear that the variability of the reverse diode currents is too large to pursuit a calibration. In

Figure 2.10b, a wafer map is created by extracting the diode currents at fixed reverse bias. A circular pattern can be recognized by observing the value of the current. This pattern is usually related to a non uniformity of the epixial growth which is on target in the center of the wafer but is thinner at the edge of the wafer. The calibration is still possible when SIMS analysis is performed next to the device selected for the calibration. In this way, the exact doping profile and thickness of intrinsic layer can be determined and used as an input for the simulations.

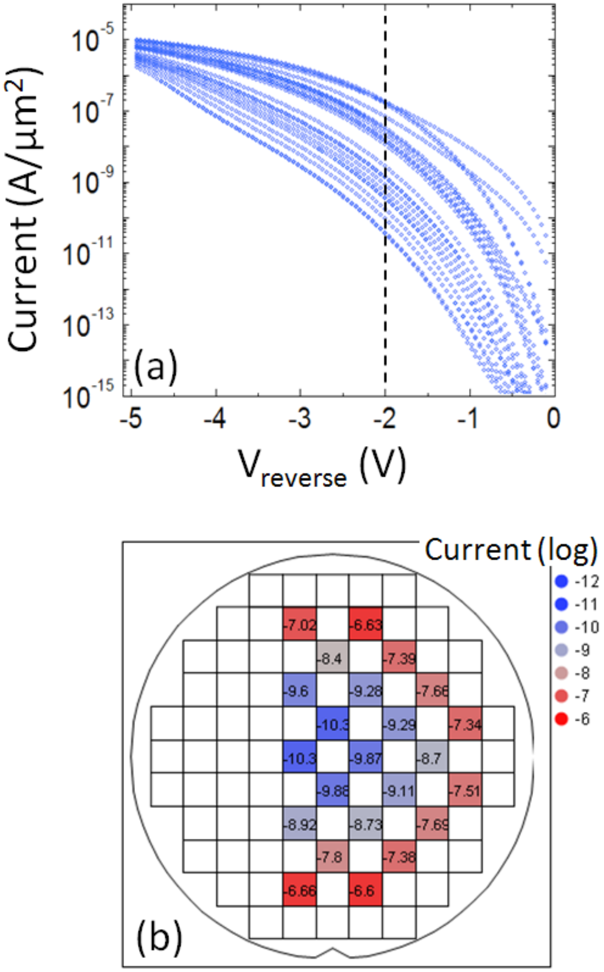


Figure 2.10: Reverse current of a pin diode (a) and wafermap of the current extracted at $V_{reverse}$ equals to -2V (b). The circular pattern is a clear indication of the variability of the thickness of the intrinsic layer which is epitaxially grown. The target thickness is 40nm for this sample. The diode dimensions are 50µm x 50µm.

2.5 Line and Point Tunneling

To describe the BTBT in TFETs both lateral and vertical components of the electric field needs to be taken into account. The lateral component is parallel to the gate while the vertical component is perpendicular to the gate. In this work, we prefer to use the definition of point and line tunneling (Figure 2.11) according to [25].

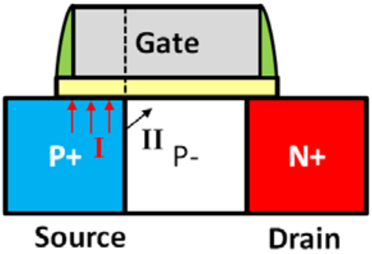


Figure 2.11: line (I) and point (II) tunneling components in TFET.

When the gate is placed mainly on the top of the channel region, the point tunneling is dominant and we define this configuration as "lateral" TFET as often found in literature. The "lateral" TFET is the most common configuration because of its simple fabricate with a conventional gate self-aligned process.

2.6 Summary

The band-to-band tunneling physics has been introduced together with their implementation in commercial simulation tools. The difficulty to describe this quantum mechanical mechanism with semiclassical models suggests that the calibration of the model parameters is mandatory. In case of hetero-junction TFETs, we would like to use quantum mechanical models proposed in the literature [26-32]. Nevertheless, these models present several limitations such as the 1D-dimensionality of the model and their assumptions such as extremely abrupt and defect-free hetero junction interfaces which are not always correct for experimental devices.

In this chapter, we also presented additional transport mechanisms active in TFETs which need to be considered when experimental data are analyzed. In particular, the impact of the TAT is responsible of the degradation of the swing as will be shown in Chapters 5 and 6.

References

[1] Y. Pan, "An experimental study of the Fowler-Nordheim tunneling induced degradation of LDD PMOSFET's," *IEEE Transactions on Electron Devices,* vol. 41, pp. 268-271, 1994.

[2] R. Stratton, "Theory of Field Emission from Semiconductors," *Physical Review,* vol. 125, pp. 67-82, 1962.

[3] W.-C. Lee and C. Hu, "Modeling CMOS tunneling currents through ultrathin gate oxide due to conduction- and valence-band electron and hole tunneling," *IEEE Transactions on Electron Devices,* vol. 48, pp. 1366-1373, 2001.

[4] J. Chen, *et al.*, "Sub-Breakdown drain leakage current in MOSFET," *IEEE Electron Device Letters,* vol. 8, pp. 515-517, Nov 1987.

[5] J. J. Sakurai, *Modern Quantum Mechanics*: Addison-Wesley Publishing Company, 1994.

[6] J. Knoch, *et al.*, "Impact of the dimensionality on the performance of tunneling FETs: Bulk versus one-dimensional devices," *Solid-State Electronics,* vol. 51, pp. 572-578, 2007.

[7] E. O. Kane, "Theory of Tunneling," *Journal of Applied Physics,* vol. 32, pp. 83-91, 1961.

[8] V. Alex, *et al.*, "Temperature dependence of the indirect energy gap in crystalline silicon," *Journal of Applied Physics,* vol. 79, pp. 6943-6946, 1996.

[9] S. M. Sze, *Physics of Semiconductor Devices: 3rd Edition*. New York: John Wiley & Sons, 2007.

[10] H. Flietner, "The E(k) Relation for a Two-Band Scheme of Semiconductors and the Application to the Metal-Semiconductor Contact," *physica status solidi (b),* vol. 54, pp. 201-208, 1972.

[11] L. V. Keldysh, "Behaviour of Non-Metallic Crystals in Strong Electric Fields," *Soviet Physics: J. Exptl. Theoret. Phys.,* vol. 33, pp. 994-1003, 1957.

[12] O. M. Nayfeh, *et al.*, "Design of Tunneling Field-Effect Transistors Using Strained-Silicon/Strained-Germanium Type-II Staggered Heterojunctions," *IEEE Electron Device Letters,* vol. 29, pp. 1074-1077, 2008.

[13] W. Shockley, "The theory of P-N junctions in semiconductors and P-N junction transistors," *Bell System Technical Journal,* vol. 28, pp. 435-489, 1949.

[14] W. Shockley, *Electrons and holes in semiconductors with applications to transistor electronics*: Van Nostrand, 1950.

[15] G. A. M. Hurkx, "On the modelling of tunnelling currents in reverse-biased p-n junctions," *Solid-State Electronics,* vol. 32, pp. 665-668, 1989.

[16] G. A. M. Hurkx*, et al.*, "A new recombination model describing heavy-doping effects and low-temperature behaviour," in *IEEE International Electron Devices Meeting* 1989, pp. 307-310.

[17] E. Simoen*, et al.*, "On the Temperature and Field Dependence of Trap-Assisted Tunneling Current in Ge p(+)n Junctions," *IEEE Electron Device Letters,* vol. 30, pp. 562-564, 2009.

[18] *Synopsys Sentaurus Package version F (September 2011)*

[19] A. Poyai*, et al.*, "Improved extraction of the activation energy of the leakage current in silicon p-n junction diodes," *Applied Physics Letters,* vol. 78, pp. 1997-1999, 2001.

[20] R. Scheer, "Activation energy of heterojunction diode currents in the limit of interface recombination," *Journal of Applied Physics,* vol. 105, 2009.

[21] E.-H. Toh*, et al.*, "Device physics and design of double-gate tunneling field-effect transistor by silicon film thickness optimization," *Applied Physics Letters,* vol. 90, p. 263507, 2007.

[22] K. Boucart and A. M. Ionescu, "Double Gate Tunnel FET with ultrathin silicon body and high-K gate dielectric," *the 36th European Solid-State Device Research Conference,* pp. 383-386 2006.

[23] M. V. Fischetti*, et al.*, "Theoretical Study of Some Physical Aspects of Electronic Transport in nMOSFETs at the 10-nm Gate-Length," *IEEE Transactions on Electron Devices,* vol. 54, pp. 2116-2136, 2007.

[24] S. Gupta, "Quantization of Band to Band Tunneling Models based on p-i-n Diodes," in *Master Thesis KUL*, ed, 2011.

[25] W. G. Vandenberghe*, et al.*, "Analytical model for a tunnel field-effect transistor," in *the 14th IEEE Mediterranean Electrotechnical Conference (MELECON)*, pp. 923-928, 2008.

[26] W. G. Vandenberghe, *et al.*, "Two-dimensional quantum mechanical modeling of band-to-band tunneling in indirect semiconductors," in *IEEE International Electron Devices Meeting*, 2011.

[27] S. O. Koswatta, *et al.*, "Nonequilibrium Green's Function Treatment of Phonon Scattering in Carbon-Nanotube Transistors," *IEEE Transactions on Electron Devices*, vol. 54, pp. 2339-2351, 2007.

[28] R. C. Bowen, *et al.*, "Efficient I-V simulation of quantum devices using full bandstructure models," in *IEEE/Cornell Conference on Advanced Concepts in High Speed Semiconductor Devices and Circuits*, 1995, pp. 435-439.

[29] G. Klimeck and M. Luisier, "From NEMO1D and NEMO3D to OMEN: Moving towards atomistic 3-D quantum transport in nano-scale semiconductors," in *IEEE International Electron Devices Meeting*,pp. 1-4, 2008.

[30] G. Klimeck, *et al.*, "Atomistic Simulation of Realistically Sized Nanodevices Using NEMO 3-D Part I: Models and Benchmarks," *IEEE Transactions on Electron Devices*, vol. 54, pp. 2079-2089, 2007.

[31] G. Klimeck, *et al.*, "Atomistic Simulation of Realistically Sized Nanodevices Using NEMO 3-D Part II: Applications," *IEEE Transactions on Electron Devices*, vol. 54, pp. 2090-2099, 2007.

[32] S. Steiger, *et al.*, "The nanoelectronic modeling tool NEMO 5: Capabilities, validation, and application to Sb-heterostructures," in *Device Research Conference Digest*, pp. 23-26, 2011.

Chapter 3

Fabrication of TFETs

3.1 Introduction

In spite of the optimistic prospective of the theory and the increasing number of publications with simulation results [1-3], the fabrication of TFETs still presents a series of challenges regarding the steepness of the junction profile, the required doping levels and the implementation of p-type and n-type doping on the same device.

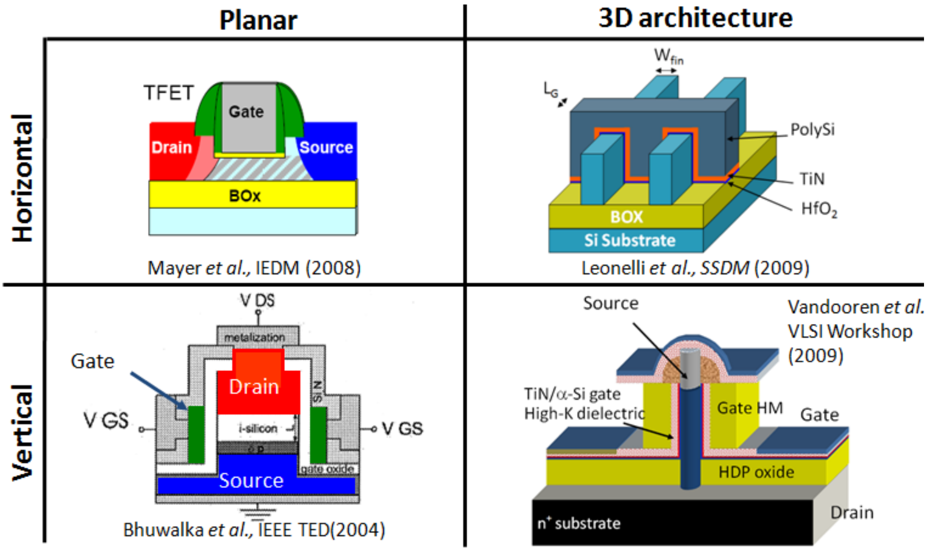


Figure 3.1: Possible implementations of a lateral TFET grouped by channel orientation (vertical or horizontal) and by dimensionality (planar or tridimensional).

Figure 3.1 shows four different types of architectures proposed in literature for TFET implementation: 1) vertical mesa structures [4-6], 2) planar-like devices fabricated on Silicon-On-Insulator (SOI) wafers

[7-9], 3) Multiple gate FET (MuGFET), or FinFET, fabricated on SOI wafers [10-11] and 4) gate-all-around devices [12-15].

The integration of vertical nanowire-based devices gives some important advantages:, a strong gate coupling to the channel [16], easier implementation of steep junctions and heterojunctions. The last is realized by growing an in-situ doped epitaxial layer [13-14] or through in-situ doping during the growth of the nanowire using VLS approach [17-18].

Contrarily, the planar-like devices suffer from the use of a single-gate configuration, and thus poor gate coupling from the fact that the junction is typically obtained by ion implantation and junction anneal, degrading the junction abruptness. In fact, Subramanian *et al* [19] compared a single gate and a double gate configuration where the backside was used as contact for the second gate and they have observed a great increase of performance, such as a reduced SS and an increased on current (I_{ON}) for the case of a double gate structure.

Implementation of TFETs in a planar architecture with improved gate coupling over the channel can be achieved using a MuGFET, or FinFET, technology [20]. The MugFET technology is a relatively mature technology compared to gate-all-around devices with nearly the same tight gate electrostatic control while still compatible with standard CMOS processing and available at imec.

For this reason, the MuGFET, or FinFET, is used in this work to initially study Si homo-junction TFETs. Silicon Fin-TFETs were used to obtain a first evaluation of the impact of different process and geometrical parameters on the BTBT conduction and as a test vehicle to validate the band to band model available in the simulation tools.

At the same time, a process flow for vertical nanowire-based TFETs was developed to demonstrate hetero-junction TFETs. In this case, a top-down approach was preferred because of its compatibility with CMOS processing and large scale production. Indeed, bottom-up approaches, in which the nanowire is grown by Vapor Liquid Solid (VLS) technique, are very interesting and have been used in academic research work [18]. However, its major drawback is the use of gold as a catalyst in case of Si NW growth, which is not compatible with CMOS processing as Au acts as a deep trap drastically reducing the carrier lifetime. In addition, placement of the nanowires and further processing are not trivial.

Since two different architectures have been studied in this work, this chapter is divided in two parts: the integration of FinFET-based and nanowire-based TFETs. The fabrication of FinFET devices is explained in details after the description of the available TFET devices. In addition, the similarities between the process flow for TFETs and MOSFETs allows the fabrication both devices on the same maskset. On the other hand, the integration of vertical nanowire is a completely new process. A dedicated maskset has been designed and the TFET structures are shortly explained. Furthermore, the final process flow is described as a result of an intense development. At the end, few alternatives to implement hetero junction TFET are proposed.

3.2 FinFET Integration of TFET

TFETs have been fabricated at imec using the 300mm line following a MugFET on SOI flow developed for the 65nm and 45nm technology nodes. The only difference is the presence of both NPLUS and PPLUS masks on the same device in order to implement source and drain regions with opposite types of dopants, as will be detailed further in this section.

3.2.1 Design of TFETs for Planar Integration

As explained in the previous paragraph, Multiple-Gate Tunneling Field Effect Transistors (MuGTFETs), or FinFET-based TFETs, have been implemented as a first attempt to study these type of devices. The motivations of this choice include:

- better gate electrostatic control over the channel;
- compatibility of the TFET design with the existing gate self-aligned design of MOSFET;
- processing knowhow already available at imec.

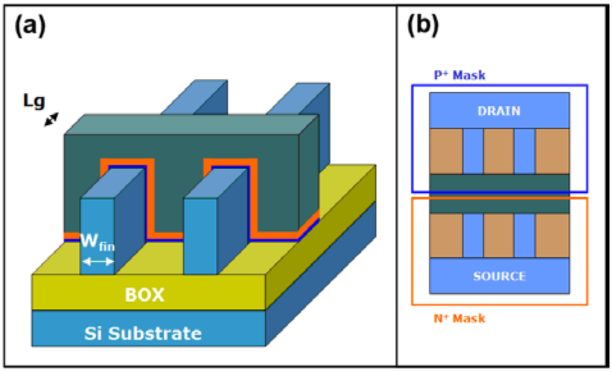


Figure 3.2: Schematic view of the tunneling MuGFET structure (a) and the applied implantation masks (b). This configuration is design for a pTFET.

Figure 3.2 shows the TFET implementation obtained by modifying the source/drain implantation mask. Contrary to the MOSFET case where only one type of doping is used to implant both the source and drain regions, the TFET requires a mask modification to implement two types of doping at the source and drain. The two masks NPLUS and PPLUS join in the middle of the gate providing an implant to the poly-Si gate on the top of the metal gate. This basic TFET design has a major drawback: the device scalability. Indeed the alignment accuracy of the source and drain mask to the gate level during photolithography is limiting the minimum device gate length. Contrary to the fin or gate mask layer, the implantation masks can only be aligned with an error of about 20~40nm. For this reason, it is not possible to implement very short gate configurations and most of the experimental results shown in this work are related to gate lengths of 150nm or greater. We believe that our observations for longer gate can be valid for shorter gate since the tunneling current does not change with the gate length because of the better control of the SCEs as explained in Chapter 3.

The introduction of a new design requires the drawing of different lithography masks. The cost of a mask set depends on the number of masks or reticles and on the minimum feature that needs to be patterned. In our case, narrow fins are required to have a good gate control of the channel. As a consequence, the cost of a dedicated mask can be prohibitive for this thesis. Fortunately, imec allows the possibility to participate in the design of shared maskset. In this way, several structures from different projects can be implemented according to the design rules. For instance, the area of the maskset dedicated to TFET devices is indicated by the red square in Figure 3.3a. One of the advantages is that conventional MOSFET and test structures are already available and it is possible to compare TFET and MOSFET that share the same processing steps. In this work, we have characterized device fabricated with the following masksets, the tape-out date is in parenthesis:

- Rhythm (Dec 2005) – 10 devices available for TFET;
- Salsa2 (Dec 2007) – 40 devices available for TFET;
- Salsa3 (Dec 2009) – New devices implemented.

Most of the results presented in this work are related to Salsa2 because this mask was available at the beginning of this work. Salsa3 was designed later based on the results from Salsa2. In Salsa2 and Rhythm, a few gated diodes were readily available for the purpose of reliability experiments on SOI substrate.

The devices are usually organized in arrays of 24 contact pads to facilitate the electrical characterization by full automatic measurement setup. The number of devices depends on the contacting scheme. A schematic of one of the four arrays available in Salsa2 and the design of a TFET is presented in Figure 3.3(b,c). Here, 10 devices are available and two contact pads are not connected. The devices are grouped into two columns with different gate length, 150nm and 250nm respectively. In each column five different fin widths are patterned. 10nm is the smallest feature present in this maskset. The Optical Photolithography corrections (OPC) and other lithography techniques to achieve nanometer features, such as double patterning, are not shown or described since they are automatically generated during the mask tape out process. A list of all the TFET device specifications for Salsa2 is shown in Figure 3.4.

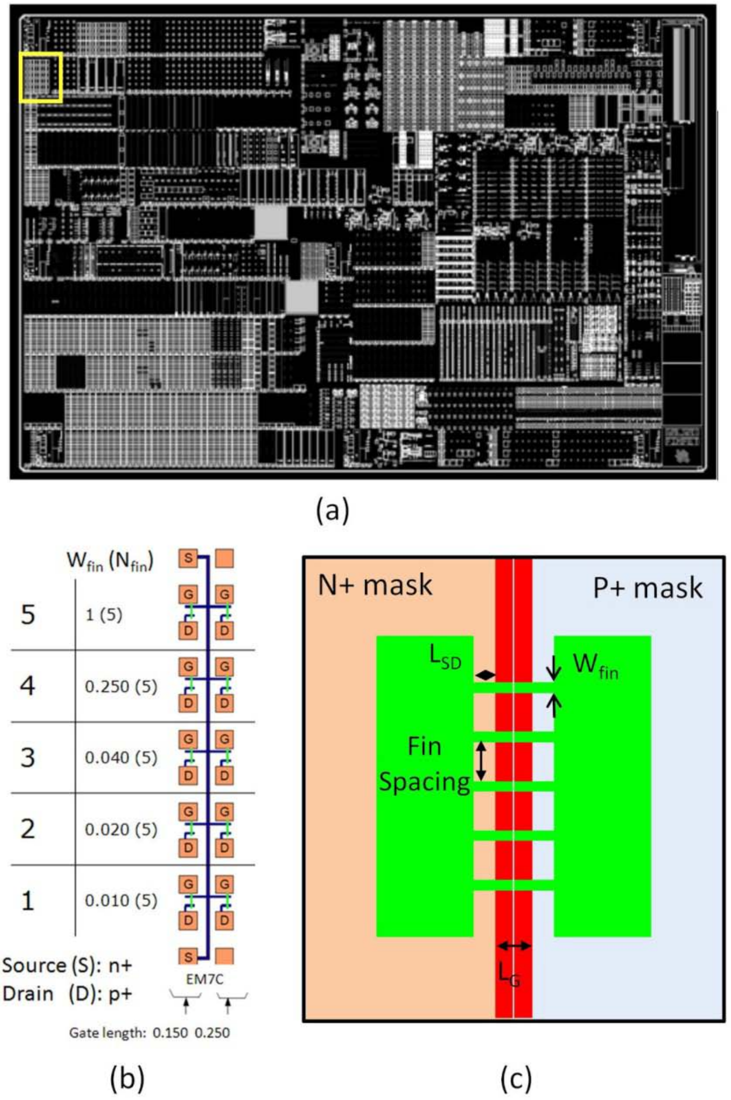
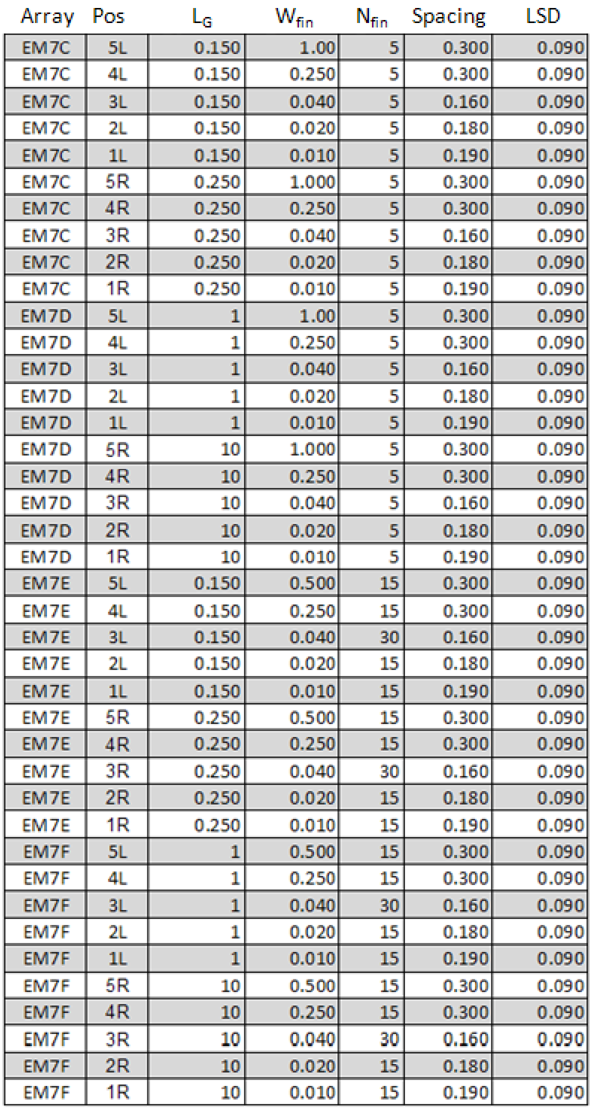


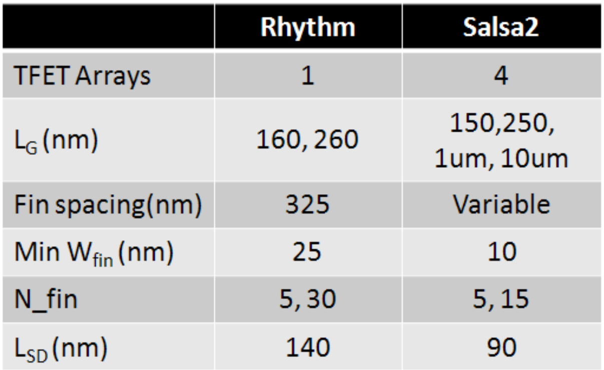
Figure 3.3: Schematic overview of the salsa2 maskset (a) where the yellow square indicates the location of TFET structures. Schematic representation of one of the array (b) and design of one of the TFET including the main geometrical parameters, such as fin width (W_{fin}), gate length (L_G), fin spacing, number of fins (N_{fin}) and source/drain spacing (L_{SD}) (c). In (b), the contact pads are marked with G for the gate contact, D for the drain contact and S for the source contact. Source is n-type because of the design definition. All the dimensions are in µm.



Array	Pos	L_G	W_{fin}	N_{fin}	Spacing	LSD
EM7C	5L	0.150	1.00	5	0.300	0.090
EM7C	4L	0.150	0.250	5	0.300	0.090
EM7C	3L	0.150	0.040	5	0.160	0.090
EM7C	2L	0.150	0.020	5	0.180	0.090
EM7C	1L	0.150	0.010	5	0.190	0.090
EM7C	5R	0.250	1.000	5	0.300	0.090
EM7C	4R	0.250	0.250	5	0.300	0.090
EM7C	3R	0.250	0.040	5	0.160	0.090
EM7C	2R	0.250	0.020	5	0.180	0.090
EM7C	1R	0.250	0.010	5	0.190	0.090
EM7D	5L	1	1.00	5	0.300	0.090
EM7D	4L	1	0.250	5	0.300	0.090
EM7D	3L	1	0.040	5	0.160	0.090
EM7D	2L	1	0.020	5	0.180	0.090
EM7D	1L	1	0.010	5	0.190	0.090
EM7D	5R	10	1.000	5	0.300	0.090
EM7D	4R	10	0.250	5	0.300	0.090
EM7D	3R	10	0.040	5	0.160	0.090
EM7D	2R	10	0.020	5	0.180	0.090
EM7D	1R	10	0.010	5	0.190	0.090
EM7E	5L	0.150	0.500	15	0.300	0.090
EM7E	4L	0.150	0.250	15	0.300	0.090
EM7E	3L	0.150	0.040	30	0.160	0.090
EM7E	2L	0.150	0.020	15	0.180	0.090
EM7E	1L	0.150	0.010	15	0.190	0.090
EM7E	5R	0.250	0.500	15	0.300	0.090
EM7E	4R	0.250	0.250	15	0.300	0.090
EM7E	3R	0.250	0.040	30	0.160	0.090
EM7E	2R	0.250	0.020	15	0.180	0.090
EM7E	1R	0.250	0.010	15	0.190	0.090
EM7F	5L	1	0.500	15	0.300	0.090
EM7F	4L	1	0.250	15	0.300	0.090
EM7F	3L	1	0.040	30	0.160	0.090
EM7F	2L	1	0.020	15	0.180	0.090
EM7F	1L	1	0.010	15	0.190	0.090
EM7F	5R	10	0.500	15	0.300	0.090
EM7F	4R	10	0.250	15	0.300	0.090
EM7F	3R	10	0.040	30	0.160	0.090
EM7F	2R	10	0.020	15	0.180	0.090
EM7F	1R	10	0.010	15	0.190	0.090

Figure 3.4: Description of all the TFET devices available in Salsa2. The main geometrical parameters are specified. All the dimensions are in µm.

A comparison between Rhythm and Salsa2 is presented in Table 3.1. It is clear that the Rhythm was not suitable to study TFETs because of the limited number of structures. In addition, an extreme trimming (~125nm for fin and ~90nm for the gate) was applied on this mask to achieve minimum feature size. As a consequence, the control of the fin width during the fin patterning is more critical and it constitutes a source of variability on the final devices. Furthermore, the additional distance between the source region and the gate (L_{SD}) increases the access resistance of the device.



	Rhythm	Salsa2
TFET Arrays	1	4
L_G (nm)	160, 260	150,250, 1um, 10um
Fin spacing(nm)	325	Variable
Min W_{fin} (nm)	25	10
N_fin	5, 30	5, 15
L_{SD} (nm)	140	90

Table 3.1: Comparison of the main geometrical differences between the maskset Rhythm and Salsa2. The dimensions specified are the target dimensions after fin patterning.

Salsa2 was a good starting material to study the TFETs even though is was not optimized for TFETs. A few limitations are listed here:

- Structures with single fin are not available;
- selection of fin widths was not optimum due to the large gap between the 40nm wide fin and 250nm-wide fin structures;
- absence of structures designed to study line and point tunneling.

For this reason, new structures were designed in the new Salsa3 maskset. In addition, an additional doping mask layer was ordered to turn existing MOSFET structures into TFETs in order to increase the number of devices with different fin widths. For instance, arrays designed for PMOS have been modified and TFET structures with different width of the fin, ranging from 5nm to 1um, and different gate lengths, from 20nm to 10um, have been implemented. The original contacting scheme was maintained to speed up the design of the new structures.

For the Salsa3 maskset two new architectures were designed, which are shown in Figure 3.5: the short gate and the shift gate.

The short gate concept aims to suppress the ambipolar behavior by increasing the distance between the drain region and the gate [21]. The variation of this parameter is obtained by shortening the gate length as is done in Figure 3.5a. In addition, it is possible to measure the effective distance of this gap by CD measurements during the processing.

The shift gate is designed to study the line tunneling by means of increasing the overlap between source and gate [22]. This concept requires the deposition of the gate after the source implantation since the N+ mask (source) is shifted to offer different gate/source overlaps. On the other hand, the P+ mask is not overlapping with the gate to reduce the ambipolar behavior.

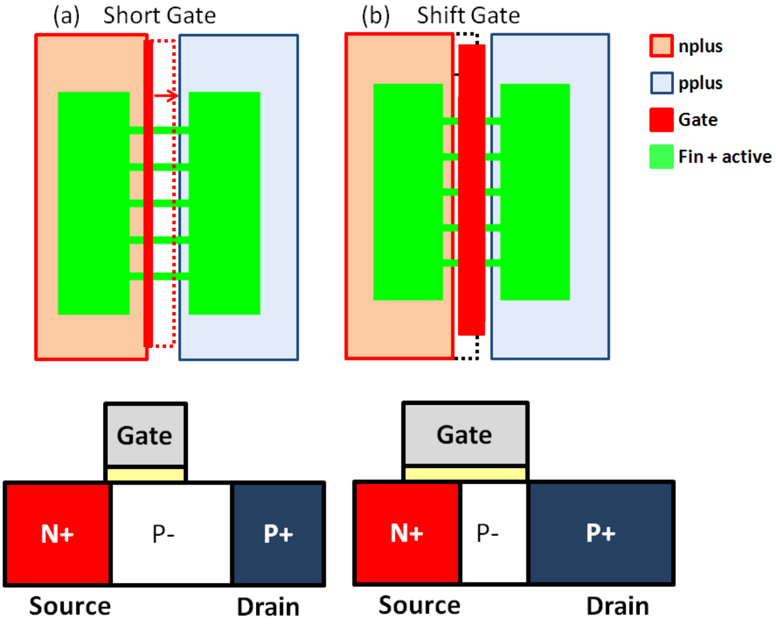


Figure 3.5: Implementation of short (a) and shift (b) gate concepts. Mask level for dummy structures and contact lines are not shown to simplify the drawing.

3.2.2 Fabrication of TFET for Planar Integration

As explained in the introduction, the FinFET architecture is used to implement TFETs because of its better electrostatic control. Narrow fins, or tall fins, are preferred over the wide, or planar-like fins, because of their better electrostatic control and the ability to fully deplete the channel. Compared to a planar integration process, the process steps for FinFET integration contain additional difficulties due to the tri-dimensional nature of the device.

Figure 3.6 shows the state of the art integration flow implemented for the fabrication of our MuGTFET [23]. The details of the processing are discussed below.

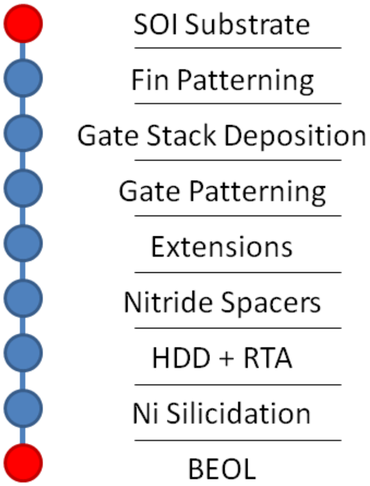


Figure 3.6: Schematic of a state of the art process flow for FinFET-based TFETs.

Substrate. In planar CMOS technologies, device isolation is achieved by well implants after the active area patterning. Since the TFET itself contains a pn junction, bulk technology will be responsible of additional leakage current. As a consequence, the SOI substrate represents a better choice since the BOX can provide optimal device isolation. In our experiment, the SOI substrate used consists of a standard (100) SOI wafer with 70nm Silicon on the top of 145nm

BOX. The doping of the Silicon layer is approximately 10^{14} cm^{-3} p-type.

Mesa isolation. A critical step in modern semiconductor integration is the isolation between adjacent transistors. In bulk technologies, Local Oxidation of Si (LOCOS) or Shallow Trench Isolation (STI) are commonly used. However, the suppression of the so-called bird's beak or the high aspect ratio of the trenches increases the complexity of the fabrication process. For this reason, the mesa isolation is an advantage in case of SOI substrates where the field regions are etched down to the Buried Oxide (BOX) and the remaining silicon active regions are completely isolated from each other. In case of FinFET, the mesa isolation is done during the fin patterning, described below.

Fin Patterning. This is the first process module in the FinFET fabrication. This module consists in the following steps:

- *Lithography step.* An Advanced Patterning Film (APF) is used as the fin Hard Mask to reduce the BOX recess after strip operation since the APF layer can be easily removed by Oxygen plasma instead of Hydro-Fluoric acid (HF). The 193nm lithography creates a resist pattern.
- *Silicon etching.* The resist pattern is transferred to the silicon by dry anisotropic etching and selective toward the silicon oxide.
- Wafer Cleaning. The cleaning step consists of an HF dip, which isotropically etches the BOX.

Two main challenges are faced during the fin patterning: the BOX recess and the damage from the dry etch of the sidewalls as shown in Figure 3.7a. The BOX recess consists of an undercut of the oxide below the fin. In case of narrow fins, the device can benefit from a GAA operation [24], on the other hand it leads to difficulties in further processing, for instance the gate patterning or a reduced mechanical stability of the fins. The reduction of the BOX recess requires special attention in implementing cleaning steps with HF.

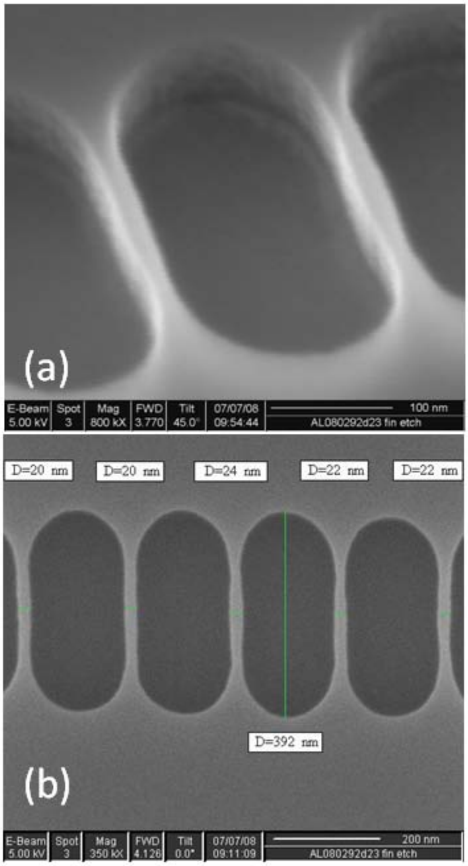


Figure 3.7: SEM inspection of fabricated FinFET after fin patterning: 45 tilt view (a) and top view (b). The tilt view shows the roughness of the sidewalls and the variation of width along the fin. For the top view the fin width is measured and reported on the top of each fin.

The patterning of narrow lines, such as the fins, is a challenge and a variability of the fin width is expected. Automatic Chromatic Dispersion (CD) measurements are carried out after processing to verify if the fin width is on target by analyzing the light contrast between the oxide and the Silicon (Figure 3.7b).

Corner Rounding. After the fin patterning, dry etch damage and Line Edge Roughness (LER) are present on the fin sidewalls (Figure 3.8). A high temperature step is required to reflow the Si and repair the etch damage. This step was performed at 840°C in H_2 ambient. An additional benefit of this process is the rounding of the corners on the top of the fins. This is particularly important in case of doped channels as sharp corners can result in double V_t behavior, as demonstrated in [25-26]. Sharp corners can also result in increased local electrical field and early gate dielectric breakdown [27].

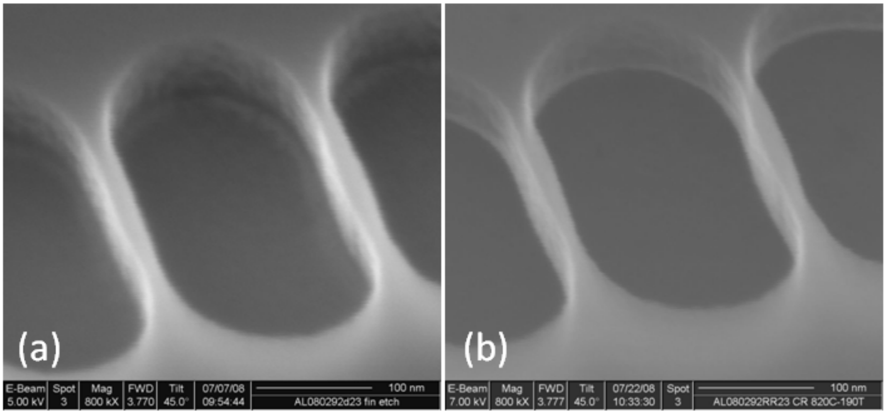


Figure 3.8: SEM view of FinFET before (a) and after (b) corner rounding.

Gate stack deposition. The deposition of the gate stack is a time critical operation, i.e. the deposition of the gate oxide, metal gate and poly-Si gate needs to be done in a sequence to assure a good quality of the gate stack. In this work, the gate stack is composed by $SiO_2/HfO_2/TiN/Poly\text{-}Si$. Atomic Layer Deposition (ALD) HfO_2 has been used because other solutions like Hf Silicate are characterized by large gate leakage; in addition, Hf Silicate requires additional steps such as a nitridation step to prevent phase separation after high temperature anneal. Prior the deposition of the HfO_2 and during the pre-clean, a chemical SiO_2 (~1nm), is formed. This oxide avoids the direct deposition the high-k on the silicon improving the stability of the high K and the interface quality.

After the gate oxide, Plasma Enhanced Atomic Layer Deposition (PE-ALD) is used to deposit 5nm of TiN used as a metal gate. The thickness of the metal gate was chosen to guarantee a mid-gap workfunction since we expect p- and n-type operation on the same device for TFET [28]. Later, a 200nm poly-Si layer is deposited and etched back down to 100nm to improve the topography prior the gate planarization step. A cross-section TEM image of a fin with $SiO_2/HfO_2/TiN/Poly\text{-}Si$ is shown in Figure 3.9.

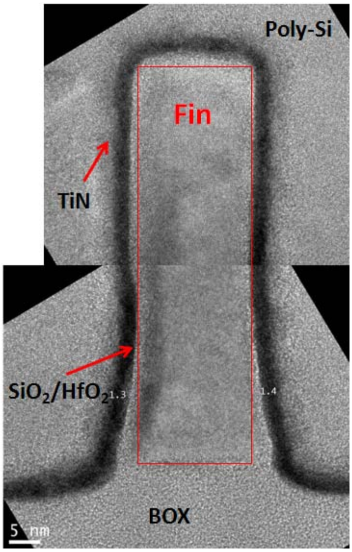


Figure 3.9: Cross-section TEM image of a fin after gate deposition. The gate stack is extremely conformal along the perimeter of the fin.

Gate Patterning. Contrary to the fin patterning where a flat starting substrate is present, the gate patterning is performed with a large topography due to the presence of fins underneath the gate. For this reason, it is necessary to minimize the topography to keep the control of the gate lithography and etch. For this reason, an etch back of the poly-Si, from 200nm to 100nm, is performed to reduce the topography before the lithography step and thus to obtain a local planarization.

The gate stack is etched in multiple steps. In our process flow, a two hardmasks (HMs) approach has been used by deposition of 100 nm of APF layer and 35 nm of SiOC. After the lithography step, the poly-Si is etched using a dry CF_4 etch chemistry in N_2 environment to avoid the formation of stringers (Figure 3.10a). The TiN dry etch is then performed in BCl_3/He chemistry. The following operation consists of the wet etch (HCl/HF chemistry) of the HfO_2, a time critical operation where extensive etch time might lead to gate undercuts as shown in Figure 3.10b.

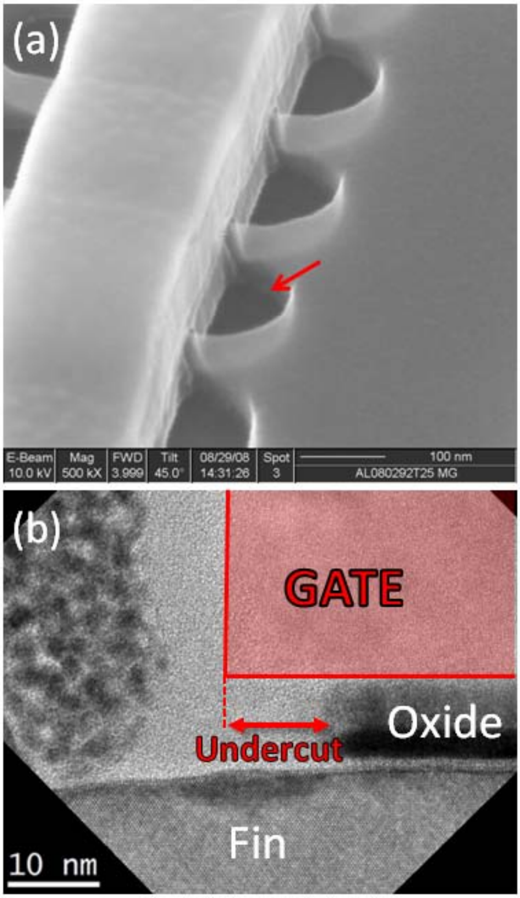


Figure 3.10: Top down SEM image after gate etch emphasizing the presence of stringers (a). These residues may lead to shorts between gate and source/drain regions after silicidation. TEM image of a cross section of a fin along the channel direction emphasizing the undercut due to a not optimized wet etch of the high-k oxide (b). The presence of undercut reduces the effective gate length of the device and the effective overlap with the source and drain.

Extension Implant. The aspect ratio of narrow fins suggests that most of the conduction is happening at the sidewalls. As a consequence, the extension, or Lightly Doped Drain (LDD), implants of the sidewalls are an important step in MuGFET fabrication. Using conventional ion beam implants with energy high enough to place the dopants all through the height of the fin (nominally 65 nm) will lead to a complete amorphization of the fin at the source and drain region, especially in case of heavy dopant species such as As or BF_2. For this reason, a tilt implant is preferred. Figure 3.11 shows the definition of tilt and twist angles. To avoid asymmetrical doping of the sidewalls two implantations are usually performed at opposite directions, conventionally called two quadrants (2Q) implants with half the dose used for each implantation, i.e. with a tilt of 45° and -45°. Two major limitations of the tilt implant are: the top of the fin receives a double dose compared to the sidewalls and the tilt angle is limited by the ratio between the fin height and the fin spacing due to shadowing effects. The twist angle is used to increase the overlap between the gate and the extensions but it may lead to a degradation of the oxide quality depending on the thickness of the oxide and on the energy/dose of the implant.

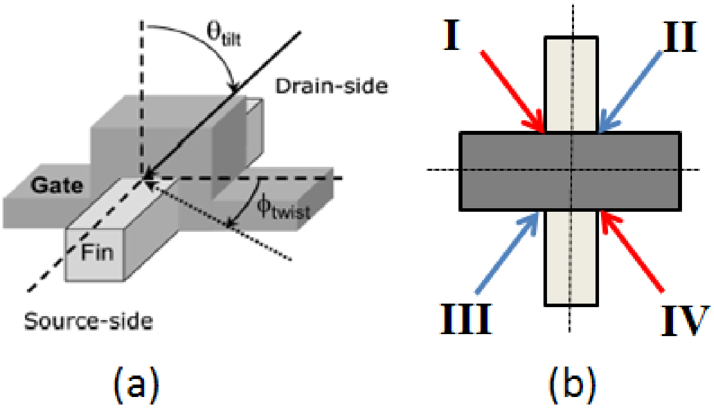


Figure 3.11: Schematic representation of the implantation angles (a) and top view to illustrate 2Q and 4Q implant (b). The tilt angle is the angle between the implantation vector and the direction perpendicular to the wafer plane while the twist angle is the angle between the projection of the implantation vector on the wafer plane and the direction parallel to the gate line. In case of 2Q, only the quadrant I and IV are implanted as there are opposite to each other. In 4Q, all the four quadrants are implanted and the implantation dose on the sidewalls will be a quarter of the total dose due to the shadowing of the gate.

In case of TFET, the Process Of Reference (POR) conditions have a tilt angle of 45° and a twist angle of 0 using a 2Q implant. The energy of the implantation was 10 KeV and a dose of 1×10^{15} atoms/cm^2 for As while an energy of 1 KeV and a dose of 1.5×10^{15} atoms/cm^2 was used for B. When the twist angle is different from zero, the 2Q implant cannot be used and two single 1Q implants are implemented at twist equal to α and (180- α).

The alignment of the LDD mask to the gate is extremely critical, especially for short gate TFETs, in order to avoid simultaneous implantation in the drain and the source regions. For this purpose the alignment is always verified by SEM inspection prior the implantations. In case of marginal alignment (Figure 3.12), a rework is necessary.

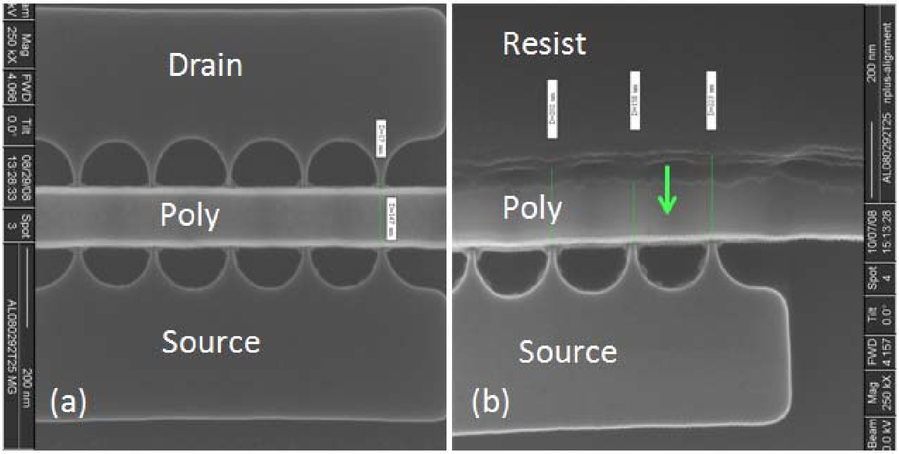


Figure 3.12: Top/down SEM images before (a) and after (b) NPLUS print. The resist is at the edge of the gate so a rework is necessary to shift the resist footing toward the middle of the gate.

Source/drain spacers and doping. After the LDD implants, source and drain spacers are deposited to avoid bridging during silicidation. The spacer formation consists of a deposition of 5nm oxide liner and 50nm nitride by Plasma Enhanced Chemical Vapor Deposition (PECVD) followed by an anisotropic etch. The nitride is dry etched using a dry CH_3F based chemistry as showed in Figure 3.13. In case of double spacers, the process of deposition and etch needs to be repeated.

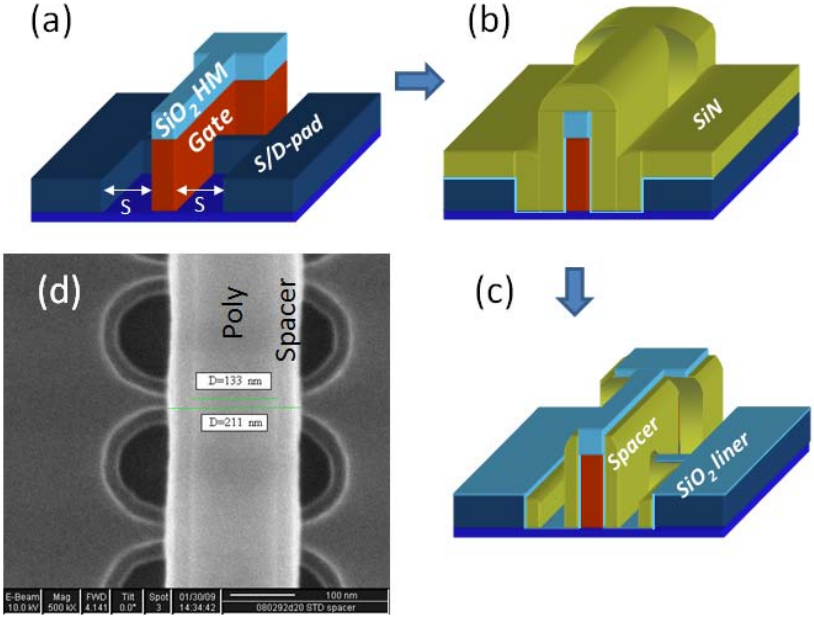


Figure 3.13: Formation of the spacer (a-b-c) and top/down SEM image after Spacer etch (d).

Heavily Doped Drain (HDD) implants are performed after the spacer formation. Contrary to the LDD, no tilt or twist is implemented and the energy of this implant is high enough to implant the dopants in the full volume of the fin. In our experiment, As and P n-type dopant species are used with an energy of 25 KeV and dose of 3×10^{15} atoms/cm^2 and 8 KeV and dose of 2×10^{15} atoms/cm^2, respectively. B is chosen as p-type dopant with energy of 3KeV and dose of 3×10^{15} atoms/cm^2.

Annealing. An annealing step is required to electrically activate the dopants and recrystallize the fin after the implantation damage during the HDD implants, in particular from As and P. The damage is more severe in case of narrow fins [29]. The activation process is conventionally carried out by a spike anneal which is a Rapid Thermal Anneal (RTA) where the wafer reaches a temperature of 1050°C for 1.5s. Since the abruptness of the junction is one of the requirements for TFETs, two alternative anneals were investigated: Laser Anneal (LA) and Solid-Phase Epitaxial Regrowth (SPER).

Laser Anneal (LA) is a diffusion-less activation technique readily available at imec [30]. The main concern about this process is the degree of activation and the ability to remove residual defects due to the implantations. This process was previously optimized at imec for FinFET processing, where imec demonstrated the possibility to reduce the defect density to levels similar to spike anneals [31]. It consists of moving a laser beam on the surface of the wafer; optimization parameters are the scanning speed, the pulse of the laser and its power as they set the temperature of the wafer. For our experiments, three different temperatures were used to optimize this process: 1200°C, 1250°C and 1300°C. This temperature is just a process parameter and it does not represent the temperature on the device.

Solid-Phase Epitaxial Regrowth (SPER) is another alternative explored to obtain a very shallow junction [32]. This method consists of a relatively low temperature anneal which should be sufficient to recrystallize and activate the dopants. In our case, the SPER anneal consists on a temperature step of one minute at 650°C.

Silicidation and back-end of line processing. To reduce the contact resistance between the contact plugs and the silicon source/drain/gate regions, a metal silicide is utilized. The so-called Self Aligned Silicidation (Salicidation) technique is implemented to form a silicide simultaneously on the source and drain regions and on the gate. The process is self aligned because the silicide only forms where silicon is present. The state-of-the-art silicide available at imec during the fabrication of this device was NiSi containing a small amount of Pt (around 10%) for a better thermal stability of the silicide [33-34]. Two steps process is used for the formation of the silicide:

- Deposition of the Ni and Pt. The thickness of the metal layer will determine the final silicide thickness.
- First thermal anneal at 270°C for 30s.
- Excess metal removal using aqua regia solution (a mixture of HCl/HNO_3).
- Second thermal anneal at 400°C for 30s.

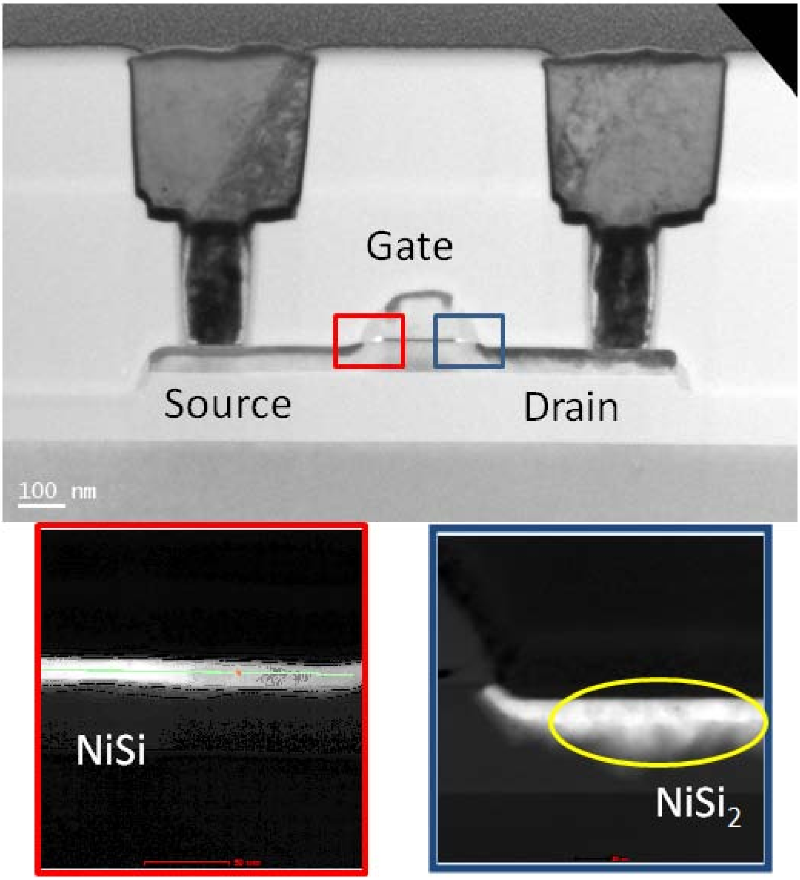


Figure 3.14: TEM image of a cross section of a fabricated TFET. In the enlargements, the High-Angle Annular Dark-Field (HAADF) images show the presence of two difference silicide phases.

The target is to produce a NiSi alloy which has a low resistivity. Nevertheless, there is a possibility to form a $NiSi_2$ phase which has an higher resistivity and worse uniformity characterized by pyramids following the (111) planes as shown in Figure 3.14.

After silicidation, a low temperature sintering in N_2 environment is performed to reduce the dangling bonds at the silicon/oxide interface. Then, the processing up to metal 1 level is continued using standard Back-End-Of-Line (BEOL) modules.

3.3 Integration of Vertical Nanowire-based TFET

The fabrication of vertical nanowire-based, or gate-all-around (GAA) TFETs can be divided in two main parts: (1) the formation of the nanowire itself and (2) the integration of the nanowire into the final device. At imec, two different approaches were considered for the nanowire formation (Figure 3.15): *top-down approach*, where the nanowires are directly patterned on the wafer using a lithography step, and *bottom-up approach*, where the nanowires are grown directly on the substrate by using VLS technique or SEG in a template.

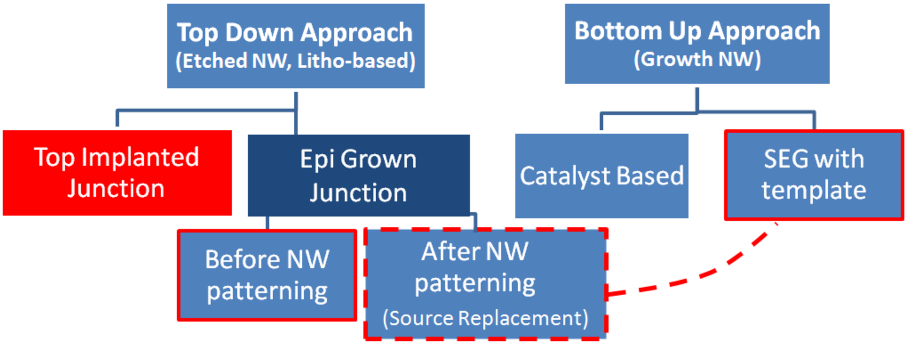


Figure 3.15: Schematic of possible approaches to integrate GAA TFETs.

The advantage of the bottom up approach is that the size of the nanowire does not depend on lithography and can potentially be aggressively scaled down. The work on growth of Si vertical nanowire was started using a VLS technique investigating alternative catalysts to gold to maintain CMOS compatibility [35]. Nevertheless, a wafer-scale controlled synthesis of nanowires for microelectronics manufacturing poses fundamental challenges such as the control of the wire diameter, orientation and placement of the nanowires. As a consequence, a seedless (catalyst-free) template-constrained growth was studied [36]. This experiment consists of the optimization of the epitaxial growth of SiGe in SiO_2 template (Figure 3.16). The presence of a nitride layer (around 20nm) below the SiO_2 is used to prevent or reduce the formation of facets during the epitaxial growth. Nevertheless, no further development was pursued because of the difficulty in obtaining facet-free growth, the variability of growth rate depending on the nanowire diameter and the slow speed of the growth due to the reduced exposed Si surface.

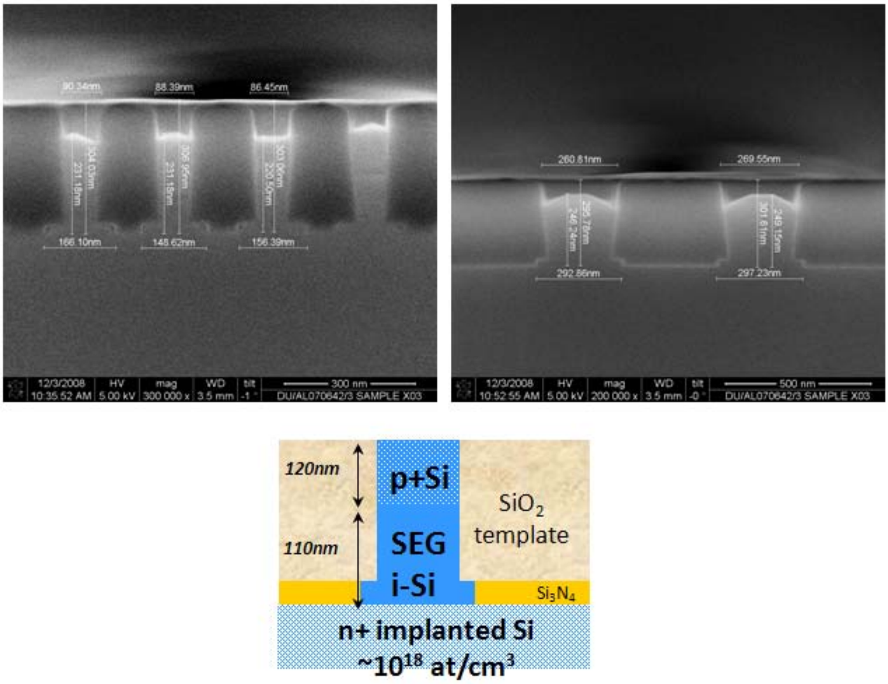


Figure 3.16: SEM image of a cross section of seedless grown nanowire in small (left) and large (right) oxide template. The opening at the foot of the nanowire is designed to reduce the faceting during the epitaxial growth.

The top down approach presents a more straightforward solution as it provides a good control of the nanowire size, orientation and position on the wafer. Nanowires are patterned into the substrate after the necessary layers are deposited to create the device. Regarding the formation of the source at the top of the nanowire, two different alternatives are available: *doping by implantation* and *epi growth* of an implanted region. The first option is done after the gate patterning and it is self-aligned. Unfortunately, the ion implantation is not able to provide the shallow junction needed for TFET operation. For this reason, the second option is preferred.

In case of epi grown source, two alternative approaches are proposed: *direct etch* or *source replacement*. In the direct etch approach, the channel layer and source layers are formed by blanket epitaxy on the substrate followed by the nanowire patterning while in the second approach it is characterized by the replacement of a dummy SiGe source by the final suitable source material after the gate patterning as shown in Figure 3.17. The main advantages of this concept are: (1) its compatibility for complementary TFETs since the dummy SiGe can be replaced by other semiconductor, (2) the source can be formed at the end of the device processing, reducing the thermal budget on the tunnel junction. Nevertheless, one of the main drawbacks is the alignment of the gate which depends on the control of etch of the gate HM and gate.

In our experimental Si nanowire TFETs, the source is either formed by implantation of the i-Si layer grown on a doped substrate or by in-situ doped blanket epitaxy on top of the i-Si channel. In our experimental SiGe-source nanowire TFETs, the source was formed by blanket epitaxy of SiGe on top of the i-Si channel.

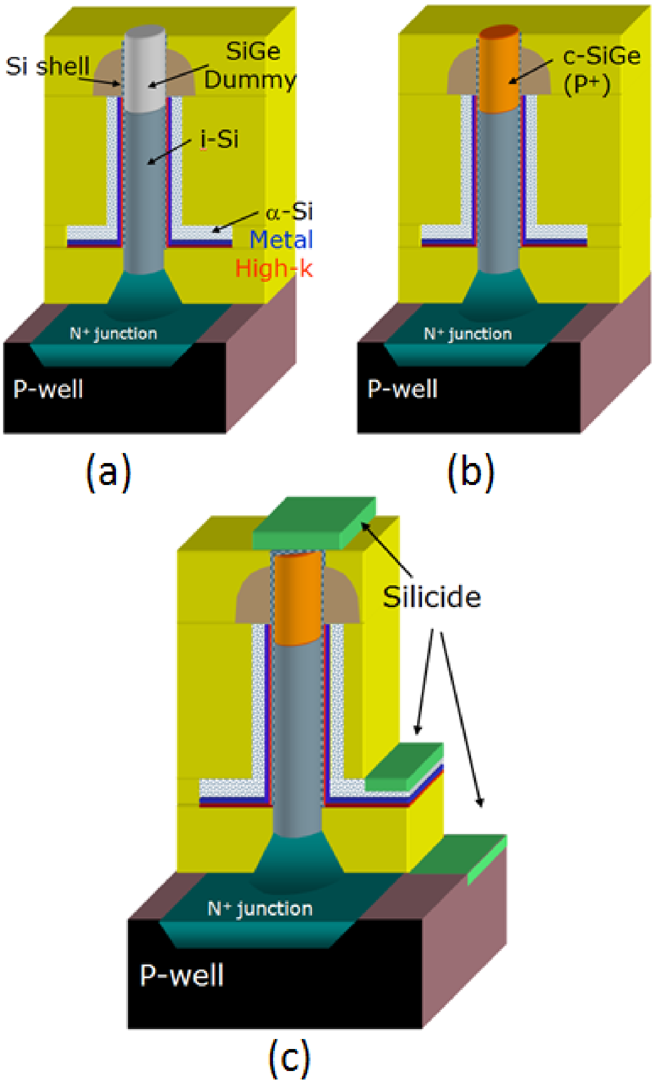


Figure 3.17: Schematic of the source replacement approach to integrate p- type TFETs on the silicon substrate. After the planarization step (a) the dummy SiGe is etched and the new source material is epitaxially grown (b) and then the contact area are opened by a sequence of etch steps and silicided (c). A Si shell is used to form a better quality gate oxide.

3.3.1 Design of vertical TFET

A transistor is basically composed by the drain, the gate and the source. Contrary to the planar designs where the three regions belong to the same level/plane, in the vertical structure these regions belong to different levels. For this reason, it requires a particular attention when one has to design vertical devices.

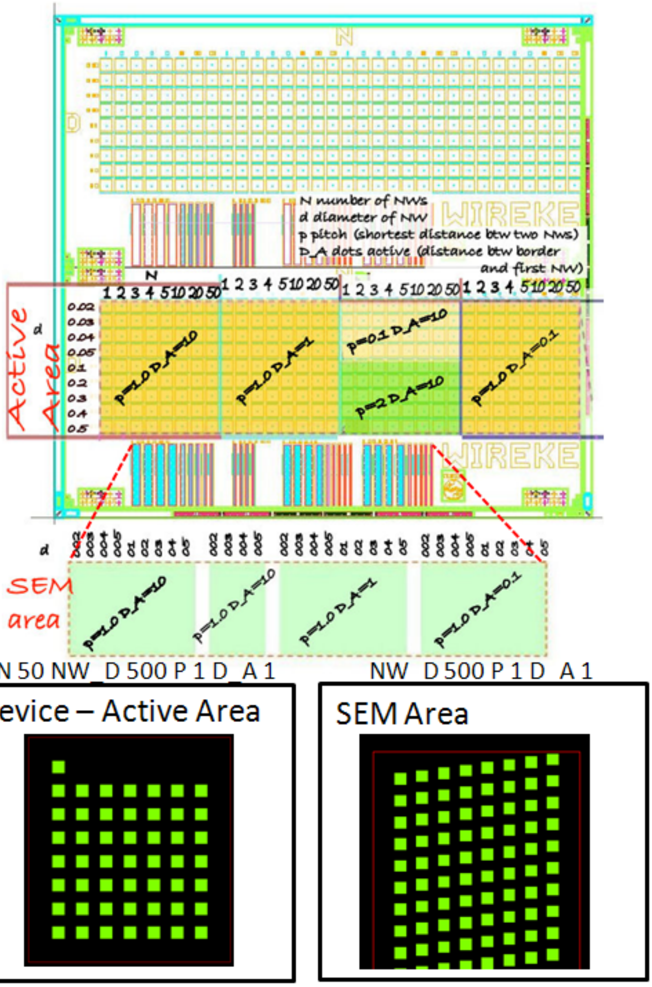


Figure 3.18: Wireke mask layout. In the inset, the design of the devices in the active Area and in the SEM area is shown. The green squares represent the nanowires while the red line defines the active area.

Two different masks have been used to fabricate GAA nanowire TFET: *Wireke* and *Tripel*. The *Wireke* mask was already available at the start of the project and was initially designed by NXP to study nanowires in two-terminal configurations (diodes or resistors). A gate contact was therefore not directly available and a new reticle (NW-Cap) had to be designed. Consequently, the electrical characterization of three-terminal devices was possible only with direct probing on the device by means of a nanoprober, a tool described in Chapter 4.

The Wireke layout is shown in Figure 3.18. All the design levels are exposed with optical lithography, except for the nanowire level where the Electron-Beam Lithography (EBL) is used because of the small features. Since EBL is extremely time-consuming, only the bottom-half of the die contains patterned nanowires. Two areas can be distinguished: (1) device area and (2) SEM area. Firstly, the device area contains arrays of nanowires, with different numbers of nanowires and different nanowire diameters. The design is very simple, using only three levels: two active levels including the nanowires EBL level and the NW-cap level, which is connecting the nanowires of the array together and is used as top contact. The back of the wafer is used as the bottom contact. Secondly, the SEM area is defined as long lines of nanowires. The nanowires are staggered in the horizontal direction to facilitate the cross sectional SEM inspection. In fact, the presence of an offset between nanowires increases the probability to successfully cut the nanowires close to the middle.

The Wireke maskset has however several limitations:

- *Nanowire definition by EBL* is a time consuming process and limits the number of patterned dies.
- *No CMP dummies*. Difficulty to control the uniformity after CMP steps, heavily used in vertical integration.
- *Maskset designed for 2-terminal device*. 3-terminal device are still possible but with increased processing complexity.
- *Electrical Characterization by nanoprober*. It is a time consuming technique with limited number of devices characterized (no statistical study possible), possible contacting issues and requiring high technical skills.

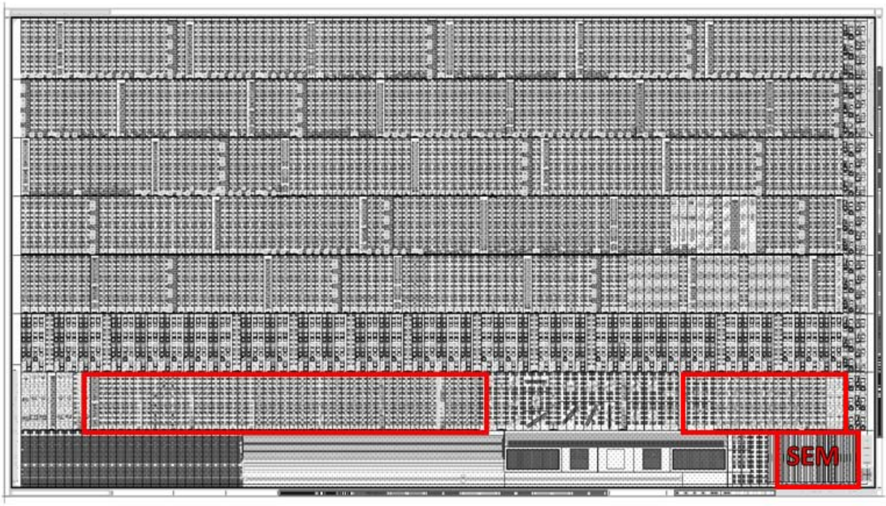


Figure 3.19: Tripel layout. The red boxes indicate the area dedicated to the TFET project. The SEM structures are indicated at the bottom right corner.

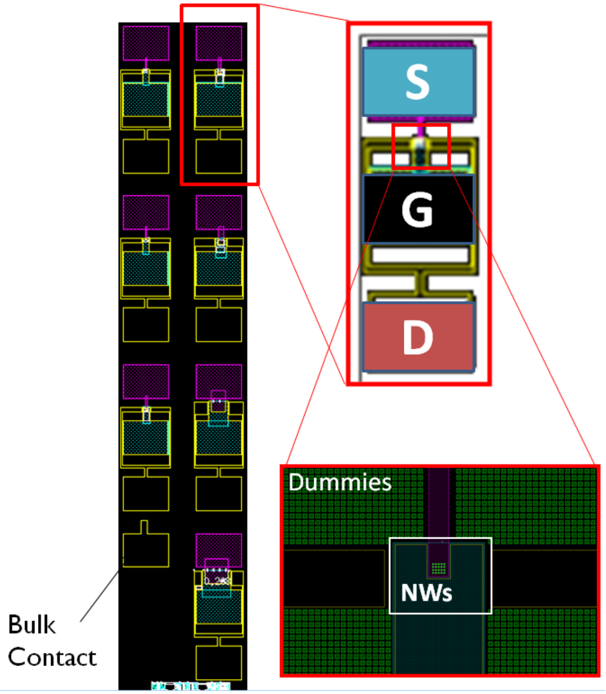


Figure 3.20: Design of a measurement array of vertical nanowire.

For this reason, a new shared maskset was co-designed for vertical TFET and III-V planar integration, *Tripel*. The maskset is 200mm compatible and the minimum features (nanowire diameter) are achieved with 193nm lithography. The layout and the areas allocated for TFETs are shown in Figure 3.19. Three main groups of structures are present: GAA TFETs for p- and n-type operation and SEM structures for physical characterization.

The GAA TFETs are grouped by nanowire spacing: 200, 300 and 500nm. For each group, TFETs with different diameters ranging from 90nm to 200nm on design, and different number of nanowires, from 1 to 5000, are present. Each GAA TFETs array contains seven devices (Figure 3.20). Each device has a set of three contacts: gate, source, drain and a common bulk contact. The bulk contact is optional and the drain contact can be replaced by the back side of the wafer depending on the processing. All structures are designed so that standard probing can be performed after silicidation. SEM structures are similar to the ones described for Wireke.

3.3.2 Fabrication of homojunction TFET

First, Si TFETs were fabricated, as silicon is a well known material widely used in microelectronics manufacturing, in which both p- and n-type TFET can be implemented. Contrary to the MuGFET case, a state-of-the-art process flow for vertical nanowire devices was not available at the start of this work, and several learning cycles were necessary to optimize the integration. The final process flow for Si n-TFETs is described in Figure 3.21. The details of the processing are discussed below.

Substrate. The substrate used for the integration is Si (100) with an n-type (As) doping concentration of 1×10^{19} at/cm^3. The highly doped substrate is not strictly necessary but it was implemented to reduce the drain resistance when the back side of the wafer is used as a drain terminal.

The blanket epitaxy of the intrinsic silicon region is performed on the full wafer. On top of the intrinsic silicon channel, an in-situ p-type doped source layer can be grown. Alternatively, an i-Si layer can be used as the source layer which will need to be doped by an implantation step later in the process.

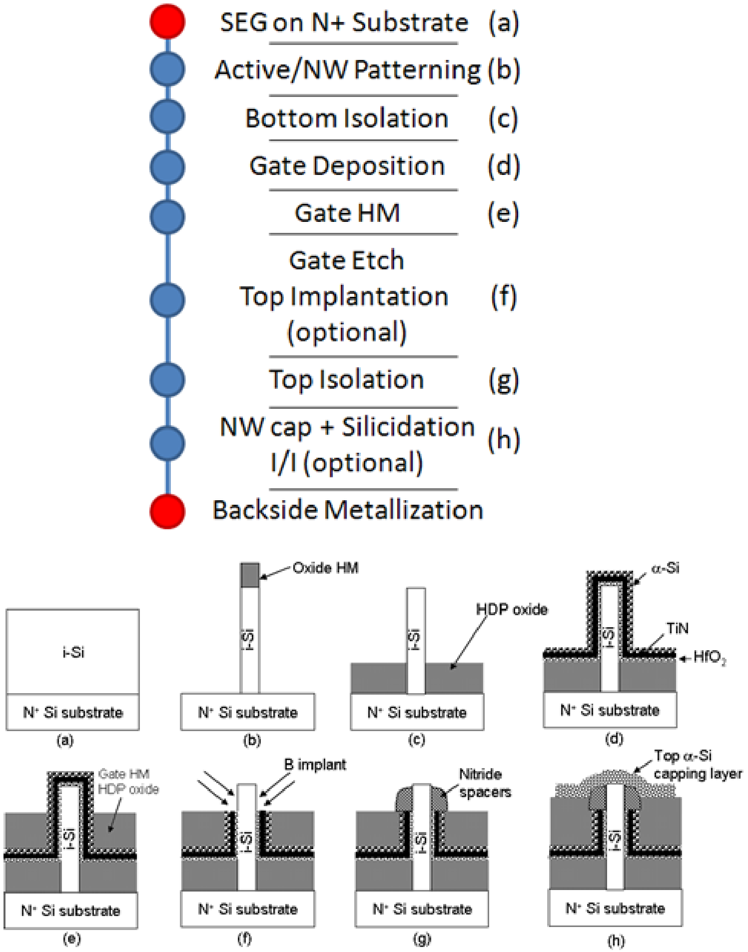
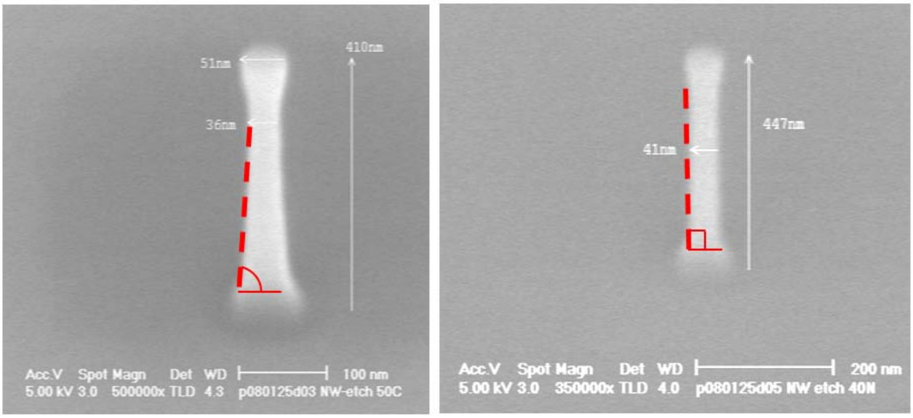


Figure 3.21: Main steps and schematic of the fabrication flow for vertical TFET. The top implantation step is not needed if the starting substrate has already a P+ region at the top.

Active/NW patterning. An oxide hardmask (HM) is deposited for the nanowire patterning. The active area and the nanowires are patterned with a combination of DUV and EBL in Wireke and with 193nm lithography in Tripel. Aggressive resist and HM trimming are necessary to reduce the nanowire diameter. For instance, the silicon nanowire with a 100 nm diameter on design will be reduced to 40 nm (in Tripel). The lithography pattern is transferred to the HM using dry CF_4 etch chemistry and to ~450 nm of silicon underneath using a dry HBr/O_2 etch chemistry. The nanowire patterning is extremely critical. The nanowire profile was optimized by tuning the power of the plasma, the chamber pressure and the O_2 flow taking into account that the nanowire profile depends on the maskset used (i.e. the amount of exposed silicon) as well as on the doping present in the silicon. Figure 3.22 shows the impact of the plasma power on the nanowire profile. In this experiment, the chamber pressure was 30mT and the O_2 flow was 8 sscm while the plasma power was increased from 70W (Figure 3.22a) to 90W (Figure 3.22b). After increasing the power a straight profile for the nanowire sidewalls was obtained.



(a) (b)

Figure 3.22: nanowire patterning before (a) and after (b) process optimization by increasing the plasma power.

Bottom isolation. This layer serves two purposes: (1) Reducing the gate/drain leakage and (2) reducing the ambipolar behavior (see *short gate concept* [21]). A High Density Plasma (HDP) oxide is deposited at 750C and a planarization step is performed before etching back of the oxide. The optimal thickness of the bottom oxide depends on the thermal budget during processing and the amount of As up-diffusion from the substrate.

Gate stack deposition. The gate stack deposition needs to be highly conformal similar to the MuGFET process flow because of the large nanowire topography. The gate stack is similar to the planar approach, using HfO_2 as the high-k gate dielectric and either MOCVD or ALD TiN as metal gate. A thin 30nm a-Si capping layer is deposited on top of the TiN to allow for gate silicidation as shown in the inset of Figure 3.23 and prevent contamination issues during the next processing step, HDP oxide deposition.

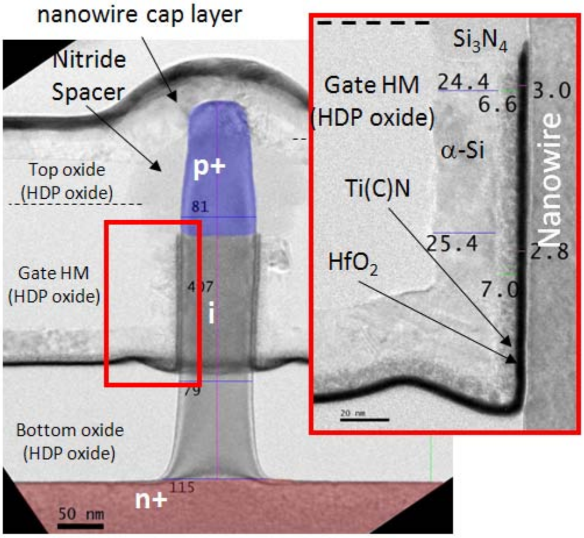


Figure 3.23: SEM image of a cross section of the fabricated TFETs. The top doping is implanted. In the inset, the gate stack is shown.

Gate HM. A HDP oxide is used again as a gate HM. Due to the high topography of the device, a CMP step is necessary after the deposition to control the wet etchback to expose the top of the nanowire prior the gate etch. The thickness of the gate HM defines the gate/source overlap in case of epi-grown source.

Top Gate etch. Once the top of the gated nanowire is exposed, a gate etch is performed to open the top source region and be able to contact the source. Similarly to the FinFET, a 2-step dry etch chemistry is used to remove the gate stack. A wet etch chemistry was adopted in the early stage of the development but it was difficult to control as shown in Figure 3.24.

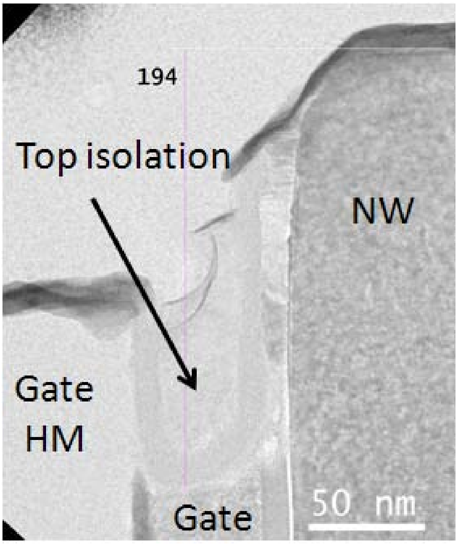


Figure 3.24: TEM image after top isolation. The top isolation results difficult and non homogenous because of the wet etch.

Top isolation. The gate is isolated from the top source region using a nitride spacer similar to the one used for planar integration as shown in Figure 3.23. The nitride spacer might be not sufficient to prevent source to gate leakage and thus an additional oxide layer (HDP oxide, planarization and etch-back) was implemented to increase the distance between the nanowire cap layer and the gate.

Nanowire cap. The nanowire cap serves two purposes: (1) connect multiple nanowires together and (2) prevent uncontrolled silicidation encroachment into the nanowires. The nanowire cap layer consists of a deposited undoped a-Si layer patterned to form the source contact area. This layer is then implanted and annealed to activate dopants and is then silicided. To further reduce the thermal budget during processing, the NWcap can be fully silicided which avoids implantation and annealing. In this case, with a in-situ doped source, no spike anneal is required and the tunnel junction abruptness after growth is maintained.

Prior the silicidation step, the gate and eventually the drain (only for Tripel) need to be exposed in order to silicide all contact regions with the top source (Figure 3.25). Etch of the gate HM and bottom isolation are performed using a combined dry and wet etch chemistry to avoid excessive undercut of these layers. Additional spacers at the gate edges are added when the top drain contact is used to avoid shorts during the silicidation as showed in Figure 3.26.

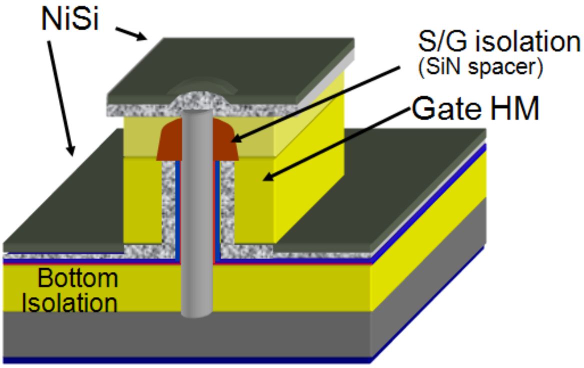


Figure 3.25: Schematic of the simultaneous silicidation of the gate and source contact.

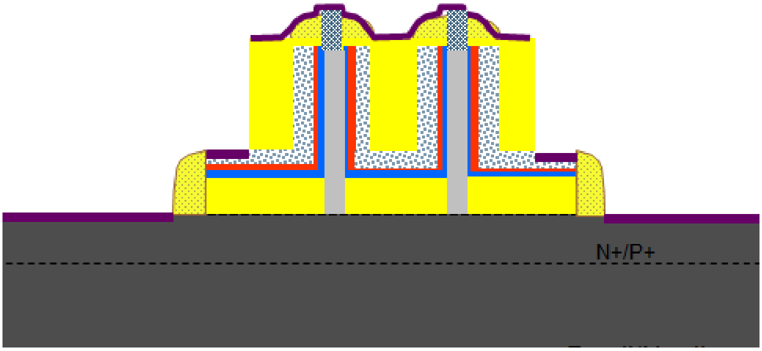


Figure 3.26: Schematic of the simultaneous silicidation of the gate, drain and source contact in case of Tripel. Additional spacers at the edge of the gate prevent shorts between the drain and the gate.

Silicidation

The silicidation process is a conventional two step process to obtain a NiSi. In case of vertical architecture, the presence of an amorphous capping layer requires a proper optimization of this process step. In fact, the full silicidation of the amorphous Si capping layer is required in order to avoid the doping and annealing step for the capping layer. In case the silicidation is too long, there is a possibility for the silicide to penetrate into the channel and create a Schottky contact or a short with the drain as shown in Figure 3.27.

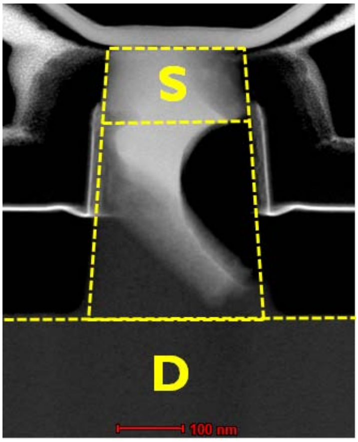


Figure 3.27: HAARD-STEM image of a cross-section of a vertical nanowire where the silicide went through the wire. The black area at the side of the nanowire represents a void created during the sample preparation. The three regions of the device (source, channel and drain) are delimited by dotted lines. The diameter of the nanowire is 200nm.

The optimized conditions for silicidation are: the first temperature step is 300°C for 30s while the second step is 450°C for 30s. Both the gate and the source are silicided together. If a front drain contact is provided, the drain is also silicided together with the source and gate.

Sintering & Backside metallization.

The last step of the process is the metallization of the backside to minimize the contact resistance and ensure a good ohmic contact during the electrical characterization. An alloy AlCu is used as a metal.

3.3.3 Fabrication of heterojunction TFET

The process flow for the fabrication of hetero junction TFETs is similar to the one used for homo Si junction. Because SiGe epitaxial growth is readily available at imec up to a Ge concentration of 45%, the first heterojunction TFET was implemented with SiGe sources. The defectivity of this layer depends on the Ge concentration and thickness of the layer which limits the processing window. SiGe can be doped in-situ during growth which provides a sharp doping profile at the interface with the undoped silicon (Figure 3.28) which can be maintained after processing thanks to the low thermal budget (<700°C), as no dopant activation is needed.

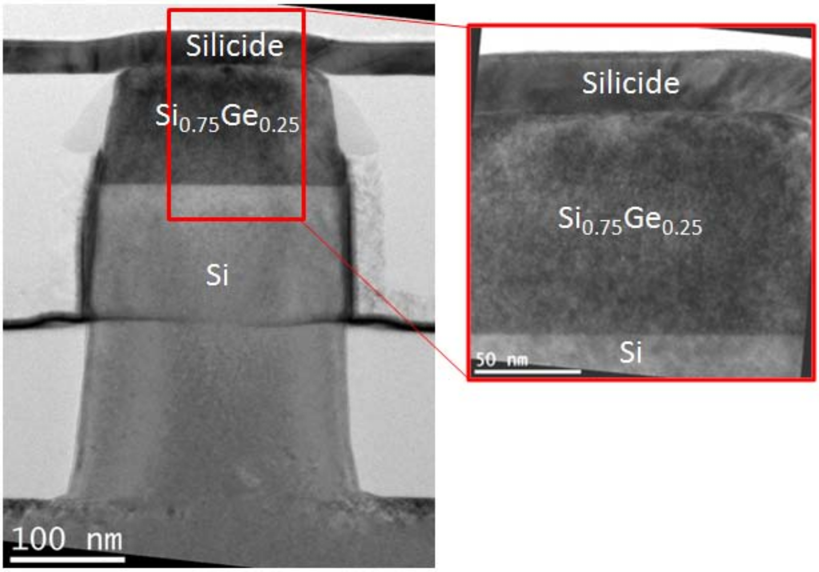


Figure 3.28: TEM image of a cross section of a 200nm wide nanowire. The sharp change in contract indicates the abruptness of the interface between the SiGe source and the Si channel region.

Only one modification in the process flow is required to implement a SiGe source: the blanket epitaxy of p+ SiGe on the top of the undoped Si. Optionally, a deposition of a Si capping layer (~nm) to encapsulate the nanowire before the gate stack deposition can be implemented to avoid the deposition of high-k on poor quality oxide at source/gate overlap, due to the formation of GeO_2. In this way, the compatibility with the homo junction process flow can be maintained (Figure 3.29). Moreover, SiGe and Si have different etch rates in case of dry etch and so the nanowire patterning had to be re-optimized. As a consequence, the top of the nanowire will be more tapered in case of SiGe as shown in Figure 3.30.

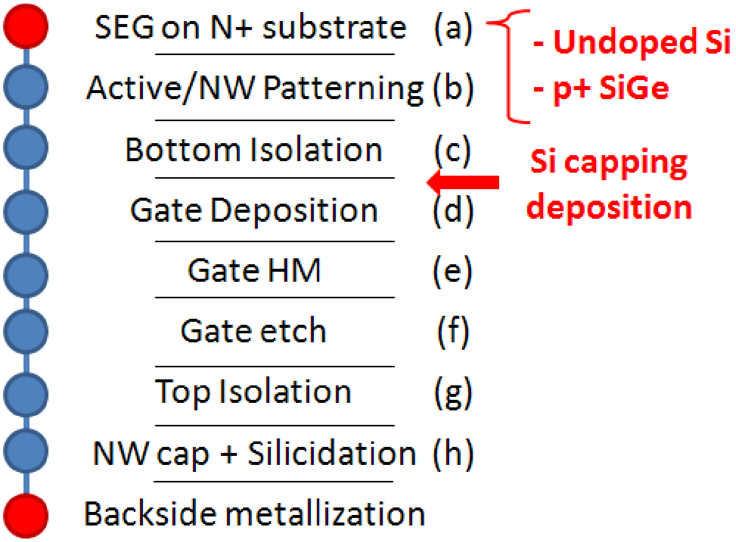


Figure 3.29: Schematic of the process flow for vertical hetero junction TFET. The additional steps compared to the homo junction integration are indicated in red.

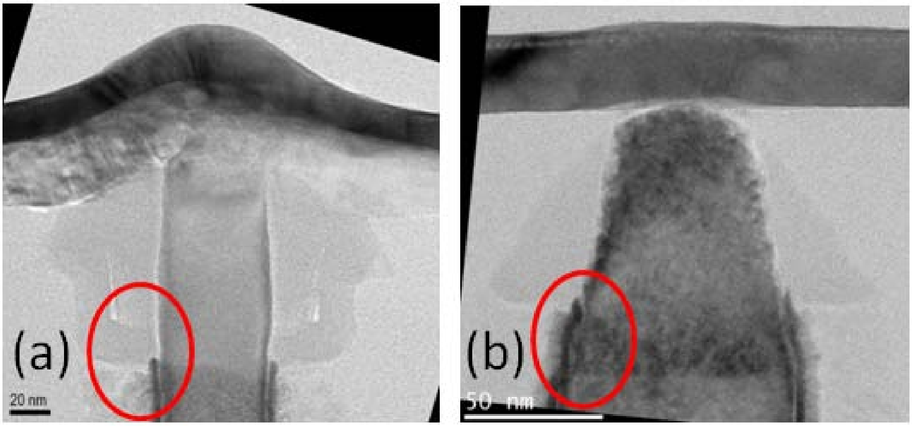


Figure 3.30: TEM profiles of the top of the nanowire in case of implanted source (a) or SiGe source (b).

3.4 Summary

In this chapter, we have discussed the design and the fabrication of TFET for two different architectures: finFET and vertical nanowire. For the finFET, an existing process flow to fabricate conventional MOSFET has been adapted for TFET. The p+ and n+ region has been created by ion implantation and the high-k/metal gate was used as a gate stack. In this process flow we have explored new annealing conditions to obtain very shallow junction. For instance, laser anneal and SPER anneal were explored.

In case of vertical nanowire, a new maskset has been designed to implement three terminal devices and allows the conventional electrical characterization (see chapter 4). Two different types of TFETs have been fabricated: Si homo-junction and SiGe/Si hetero-junction. In the first type, the tunneling junction was formed either by epitaxial growth or implantation. In case of hetero-junction, the source has been replaced with SiGe which was grown by blanket epitaxy before the nanowire patterning. In addition, a process flow where a dummy SiGe source is replaced after the gate patterning has also been proposed, i.e. source replacement approach. This technique allows the fabrication of p- and n-type TFETs with different source material on the same wafer.

References

[1] K. Boucart and A. M. Ionescu, "Double-gate tunnel FET with high-kappa gate dielectric," *IEEE Transactions on Electron Devices,* vol. 54, pp. 1725-1733, 2007.

[2] E.-H. Toh*, et al.,* "Device physics and design of germanium tunneling field-effect transistor with source and drain engineering for low power and high performance applications," *Journal of Applied Physics,* vol. 103, p. 104504, 2008.

[3] A. S. Verhulst*, et al.*, "Boosting the on-current of a n-channel nanowire tunnel field-effect transistor by source material optimization," *Journal of Applied Physics,* vol. 104, pp. 064514-10, 2008.

[4] K. K. Bhuwalka*, et al.,* "P-channel tunnel field-effect transistors down to sub-50 nm channel lengths," *Japanese Journal of Applied Physics Part 1-Regular Papers Brief Communications & Review Papers,* vol. 45, pp. 3106-3109, 2006.

[5] K. K. Bhuwalka*, et al.*, "Vertical tunnel field-effect transistor," *IEEE Transactions on Electron Devices,* vol. 51, pp. 279-282, 2004.

[6] P. F. Wang*, et al.*, "Complementary tunneling transistor for low power application," *Solid-State Electronics,* vol. 48, pp. 2281-2286, 2004.

[7] W. Y. Choi*, et al.*, "Tunneling Field-Effect Transistors (TFETs) With Subthreshold Swing (SS) Less Than 60 mV/dec," *IEEE Electron Device Letters,* vol. 28, pp. 743-745, 2007.

[8] T. Krishnamohan, *et al.*, "Double-gate strained-Ge heterostructure tunneling FET (TFET) With record high drive currents and < 60 mV/dec subthreshold slope," in *IEEE International Electron Devices Meeting* San Francisco, CA,, 2008, p. 947.

[9] F. Mayer*, et al.*, "Impact of SOI, $Si_{1-x}Ge_x$OI and GeOI substrates on CMOS compatible Tunnel FET performance," in *IEEE International Electron Devices Meeting* New York, 2008, pp. 163-166.

[10] M. Fulde*, et al.*, "Fabrication, optimization and application of complementary Multiple-Gate Tunneling FETs," in *the 2nd IEEE International Nanoelectronics Conference* 2008, pp. 579-584.

[11] D. Leonelli, *et al.*, "Multiple-gate tunneling field effect transistors with sub-60mV/dec subthreshold slope" in *Solid State Devices and Materials Conference*, Sendai, Japan, 2009, pp. 767-768.

[12] A. Vandooren, *et al.*, "Impact of process and geometrical parameters on the electrical characteristics of vertical nanowire silicon n-TFETs," *Solid-State Electronics,* vol. 72, pp. 82-87, 2012.

[13] A. Vandooren, *et al.*, "A 35nm diameter vertical silicon nanowire short-gate tunnelFET," presented at the Silicon Nanoelectronics Workshop, Kyoto, 2009.

[14] Z. X. Chen, *et al.*, "Demonstration of Tunneling FETs Based on Highly Scalable Vertical Silicon Nanowires," *IEEE Electron Device Letters,* vol. 30, pp. 754-756, 2009.

[15] R. Gandhi, *et al.*, "CMOS-Compatible Vertical-Silicon-Nanowire Gate-All-Around p-Type Tunneling FETs With < 50-mV/decade Subthreshold Swing," *IEEE Electron Device Letters,* vol. 32, pp. 1504-1506, 2011.

[16] J. Knoch, *et al.*, "Impact of the dimensionality on the performance of tunneling FETs: Bulk versus one-dimensional devices," *Solid-State Electronics,* vol. 51, pp. 572-578, 2007.

[17] M. T. Bjork, *et al.*, "Silicon nanowire tunneling field-effect transistors," *Applied Physics Letters,* vol. 92, pp. 193504-3, 2008.

[18] K. E. Moselund, *et al.*, "Silicon Nanowire Tunnel FETs: Low-Temperature Operation and Influence of High-k Gate Dielectric " *IEEE Transactions on Electron Devices,* vol. 58, pp. 2911-2916, 2011.

[19] V. Subramanian, *et al.*, "Impact of fin width on digital and analog performances of n-FinFETs," 2007, pp. 551-559.

[20] D. Hisamoto, *et al.*, "FinFET-a self-aligned double-gate MOSFET scalable to 20 nm," *IEEE Transactions on Electron Devices,* vol. 47, pp. 2320-2325, 2000.

[21] A. S. Verhulst, *et al.*, "Tunnel field-effect transistor without gate-drain overlap," *Applied Physics Letters,* vol. 91, pp. 053102-3, 2007.

[22] W. G. Vandenberghe, *et al.*, "Analytical model for a tunnel field-effect transistor," in *the 14th IEEE Mediterranean Electrotechnical Conference (MELECON),* 2008, pp. 923-928.

[23] N. Collaert, *et al.*, "Performance enhancement of MUGFET devices using Super Critical Strained–SOI (SC-SSOI) and CESL," in *Symposium on VLSI Technology*, 2006, p. 52.

[24] J. P. Colinge, *et al.*, "Silicon-on-insulator `gate-all-around device'," in *IEEE International Electron Devices Meeting* 1990, pp. 595-598.

[25] M. P. Kumar, *et al.*, "Corner effects in SOI-Tri gate FinFET structure by using 3D Process and device simulations," in *the 3rd IEEE International Conference on Computer Science and Information Technology* 2010, pp. 704-707.

[26] X. Weize, *et al.*, "Improvement of FinFET electrical characteristics by hydrogen annealing," *IEEE Electron Device Letters,* vol. 25, pp. 541-543, 2004.

[27] G. Groeseneken, *et al.*, "Reliability issues in MuGFET nanodevices," in *IEEE International Reliability Physics Symposium*, 2008, pp. 52-60.

[28] R. Singanamalla, *et al.*, "On the impact of TiN film thickness variations on the effective work function of poly-Si/TiN/SiO_2 and poly-Si/TiN/HfSiON gate stacks," *IEEE Electron Device Letters,* vol. 27, pp. 332-334, 2006.

[29] R. Duffy, *et al.*, "Solid phase epitaxy versus random nucleation and growth in sub-20 nm wide fin field-effect transistors," *Applied Physics Letters,* vol. 90, p. 241912, 2007.

[30] S. Severi, *et al.*, "Channel engineering towards a full low temperature process solution for the 45 nm technology node NMOS transistors," *the 34th European Solid-State Device Research Conference,* pp. 225-228228, 2004 2004.

[31] C. Ortolland, *et al.*, "Laser-Annealed Junctions with Advanced CMOS Gate Stacks for 32nm Node: Perspectives on Device Performance and Manufacturability," in *Symposium on VLSI Technology*, Hawaii, 2008.

[32] T. Miyashita, *et al.*, "Feasibility Study of Solid-phase Epitaxial Regrowth (SPER) as an Ultra-shallow Junction (USJ) Technology for High-performance CMOS Devices," in *International Workshop on Junction Technology,* 2006, pp. 92-95.

[33] D. Z. Chi, *et al.*, "Comparative study of current-voltage characteristics of Ni and Ni(Pt)-alloy silicided p(+)/n diodes," *Applied Physics Letters,* vol. 78, pp. 3256-3258, May 2001.

[34] H. Iwai, *et al.*, "NiSi salicide technology for scaled CMOS," *Microelectronic Engineering,* vol. 60, pp. 157-169, 2002.

[35] F. Iacopi, *et al.*, "Size-dependent characteristics of indium-seeded Si nanowire growth," *Electrochemical and Solid State Letters,* vol. 11, pp. K98-K100, 2008.

[36] F. Iacopi, *et al.*, "Seedless Templated Growth of Hetero-Nanostructures for Novel Microelectronics Devices," in *MRS Proceedings,* 2009, pp. 1178-AA04-04.

Chapter 4

Basic Characterization of TFETs

4.1 Introduction

This chapter consists of two parts. First, a short description of the electrical characterization technique for TFET is presented. Conventional characterizations technique such as Direct Current (DC) and Capacitance Voltage (CV) techniques are illustrated together with state-of-the-art tool such as the nanoprober which was used to characterize the vertical nanowire in the early stage of this work. The physical analysis methods relevant to this work are discussed in the Appendix A.

In the second part of this chapter, the principle of operation of Tunnel FETs is explained using the most basic embodiment of this transistor. An important section focuses on the identification of the interband tunneling phenomenon. The main findings are also valid for TFETs with architectures [1-5] and configurations [6-9] other than the basic one described here. Later, the setup and methodologies to electrically characterize the TFET are briefly introduced.

4.2 Methods of Characterization of TFET

In general, the characterization of TFETs needs to be combined to the characterization of MOSFET structures in order to extract relevant parameters, such as EOT, which cannot be accurately extracted due to the asymmetrical doping in the TFETs. For this reason, the fabrication of MOSFET and TFET on the same maskset is important to extract the electrical parameters of the TFET.

4.2.1 Electrical Characterization Tools

The most common characterization technique is the DC measurement of current-voltage. Several measurement setups are available at imec from state-of-the-art Cascade manual probe stations to fully automated setups. This section offers an overview of the systems used, with a distinction between the nanoprober setup and the industrial probe station.

Nanoprober

The nanoprober consists of a set of micromanipulators installed on a designed support inside a Scanning Electron Microscope (SEM). An additional substage (for orthogonal positioning solutions) is mounted to enhance the accuracy and functionality of the standard microscope stage. A picture of the setup together with the micromanipulator and the sub-stage is shown in Figure 4.1. The micromanipolator is a three-axis manipulator device with an isolated tip for probing controlled by the NanoControl depicted in Figure 4.2. The NanoControl can control the horizontal and vertical movements of the micromanipulator together with its retraction and extension. The principle behind the micromanipulator is a piezoelectric motor, the Nanomotor®, consisting of a stator (piezo crystal) and a slider. The stator expands or contracts under electrical voltage allowing movements in the nanometer scale.

The nanoprober has extensively been used to characterize GAA TFETs with the Wireke layout because the mask was not designed for three terminals devices and, hence, direct probing on the device was necessary (see Chapter 3). After the development of the Tripel maskset, industrial probe stations were used for the electrical characterization of GAA TFETs. The nanoprober apparatus is not commonly used because it requires a complex sample preparation and contacting procedure. On the other hand, it allows more flexibility to directly characterize prototype devices without requiring a complete FEOL and/or BEOL processing or costly mask design. The device to be measured needs to have contact areas bigger than 100 nm^2, otherwise the realization of a good ohmic contact can be compromised.

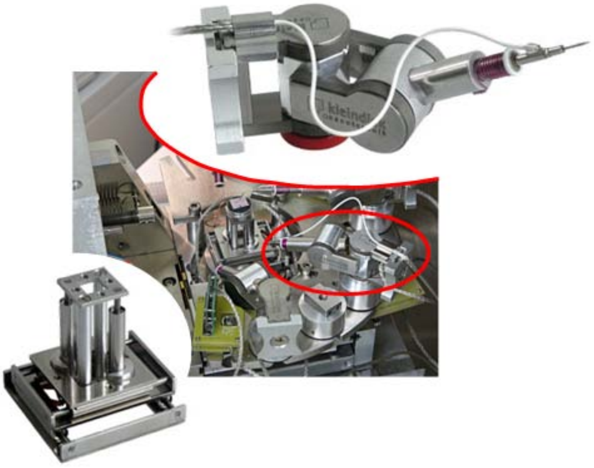


Figure 4.1: Internal view of the apparatus inside the SEM. In the inset, the enlargements of one of the micromanipulator (top) and the sub stage (bottom).

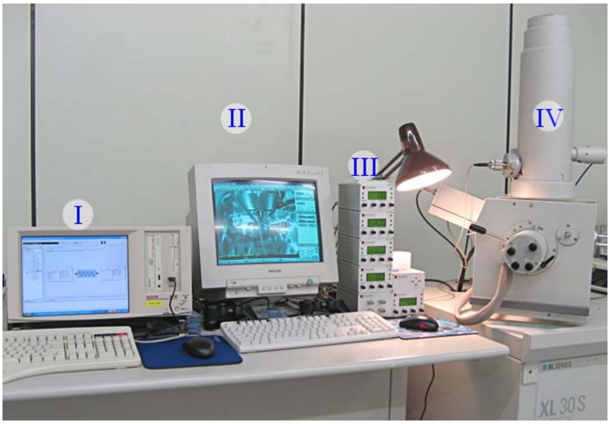
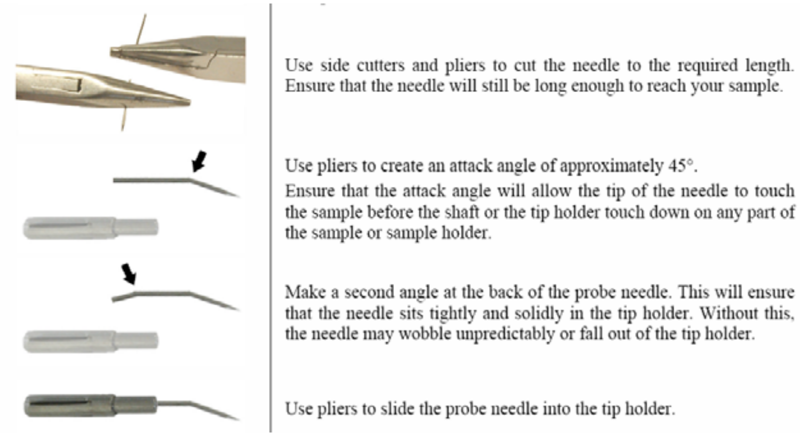


Figure 4.2: The nanoprober workstation is composed by: a semiconductor parametric analyser (Keithley 4200) (I), SEM PC control (II), Nanocontrols for micromanipulators and substage (III) and SEM (IV).

The micromanipulators of the nanoprober are connected by triaxial cables to the commercial semiconductor parametric analyzer (Keithley 4200). A picture of the complete workstation is presented in Figure 4.2.

The main challenges of the nanoprober are the preparation of the probe needles and the sample preparation. The probe needle is a tungsten wire of 10µm diameter, mounted on a thicker and longer (2 inches) tinned copper shaft. The tip is thin with a point radius smaller than 0.1µm, and flexible to minimize the damage on the structure. To be able to contact the device, the shaft needs to be bended as shown in Figure 4.3. The angle of the shaft is important to effectively transfer enough pressure on the contact pad and achieve a good contact for stable measurements.

The sample preparation consists in forming a back side contact since the sub stage is not connected to the parameter analyzer but it is only grounded to the SEM. A cleaning step with an HF solution (2% - 1min) is performed on the backside of the sample only to avoid damaging the devices at the front side. After cleaning, a tinned copper wire is attached to the backside with a copper tape. The other side of the copper wire is soldered to a tungsten shaft similar to the probe needle. The backside contact is connected to one of the micromanipulators. Additional InGa paste can be added on the backside of the sample to reduce the contact resistivity.



Use side cutters and pliers to cut the needle to the required length. Ensure that the needle will still be long enough to reach your sample.

Use pliers to create an attack angle of approximately 45°. Ensure that the attack angle will allow the tip of the needle to touch the sample before the shaft or the tip holder touch down on any part of the sample or sample holder.

Make a second angle at the back of the probe needle. This will ensure that the needle sits tightly and solidly in the tip holder. Without this, the needle may wobble unpredictably or fall out of the tip holder.

Use pliers to slide the probe needle into the tip holder.

Figure 4.3: Standard procedure for the probe needle preparation described in [10].

Figure 4.4 shows the measurement setup for the three terminal GAA TFET characterizations. Two probe needles are used to contact the gate and the source at the front side of the sample while the backside of the sample is used as drain contact. To avoid issues of contact resistance, the backside (substrate) is already doped and the expected device current is too low to be limited by spreading resistance of the wafer. The gate and the source on the other hand are silicided. The top contact (source) uses a capping layer to increase the contact area and connect multiple nanowires together (see chapter 4). Contacting a single nanowire is possible for nanowire diameters above 200nm but the characterization remains difficult because enough pressure needs to be exerted for proper contacting.

The procedure to contact the devices using the nanoprober is extremely time consuming. In fact, the wafer needs to be cleaved manually because the device is not encapsulated. The sample dimensions cannot exceed 2 cm^2 to allow enough freedom for the movements of the micromanipulators. For a good SEM imaging the working distance should be at least 5mm. The probe needle needs to be electrically initialized to remove the thin oxide layer created during the preparation and loading of the sample. This operation requires a current flow able to heat up the metal and evaporate the thin oxide layer. The procedure to approach the sample to the device is slow and tedious and it cannot be compared to a conventional measurement setup. Additional limitations of this setup are: AC measurements are not available and difficult to implement. Because of the long measurement time using the nanoprober, it is very difficult to measure many devices and/or dies on one wafer, and statistical analysis of the device and processing is not possible.

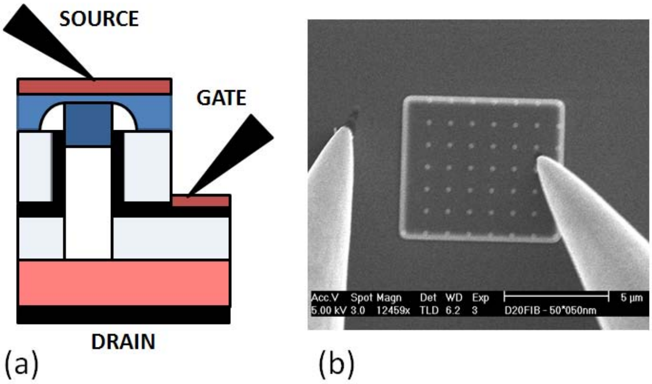


Figure 4.4: Contacting scheme (a) and SEM image of the contacting of the nanowires. Two probe needles are used to contact the gate, which is the field region in the picture, and the nanowire cap. Due to the topography of the structures the top of the nanowire is visible.

Conventional Probing Setup

The conventional probing setup available at imec includes a full automated system and a semi-automatic system capable to load and measure wafers up to 300mm diameter. The detailed description of these systems is not within the scope of this work, and only a general overview about their capabilities is presented in this work. The conventional probing setup is composed of:

- A probe station, where the wafer is loaded and the contact pads are connected by means of probe needles. The wafer holder allows vertical, horizontal and rotational movements in the xy plane to localize the target device and align the wafer. In addition, the temperature range for the measurements is comprised between -40°C and 300°C. In addition, the tool is isolated from mechanical vibration and electrical interference.
- Semiconductor parameter analyzer or additional equipments such as impedance analyzer, or pulse generator. The versatility of the setup allows several types of characterization. HP4156c and Agilent B1500 with high precision module (able to measure down to 10fA) are commonly used as semiconductor parameter analyzers.
- Support computer. It can be used to run script to control the difference instruments by external software through GPIB connection.

Low Temperature Measurement

Low temperature measurements are essential to study TFETs since band-to-band tunneling is mainly independent with the temperature contrary to most of the other conduction mechanisms. At low temperature, the tunneling mechanism is dominant and can be better studied. Unfortunately, a low temperature system, or cryo setup, on full wafers is extremely expensive and it is not available at imec. Two different liquid nitrogen (77 K) systems have been used in this work: a cryostat on small samples available at imec and a full 200mm wafer cryostat available at INP Grenoble. The first system requires cleaving of the wafer into small samples with an area less than 2cm^2. The semiconductor parametric analyzer used in both systems is similar to the conventional setup allowing a similar range of the measurements.

4.2.2 Methodologies and Parameters Extraction

Most of the characterization techniques used for MOSFETs can be used or adapted for the electrical characterization of the TFETs since it is basically a three terminal device where the device operation is controlled by a gate. Nevertheless, the figures-of-merit of band-to-band tunneling are different from the MOSFET as indicated in Chapter 3. Parameters such as threshold voltage need to be redefined.

DC Methods

The DC characterization is the basic and first technique to evaluate the TFET operation. In particular, I_{DS}-V_{GS} and I_{DS}-V_{DS}, are commonly used to extract parameters such as the average slope and maximum ON current at given supply voltage and benchmark TFETs. In addition, they are relevant to verify the uniformity over the wafer.

Gate Capacitance and T_{inv}

The gate capacitance, C_{GS} (V_{GS}), represents the capability to store a charge under a certain bias. In case of SOI device, C_{GS} is measured by grounding all terminals except for the gate which is biased. To extract the oxide capacitance, C_{ox}, a two elements model is used as illustrated in Figure 4.5 where an additional resistance component is added in parallel to take into account the possible gate leakage component. The presence of an additional capacitance, C_{it}, due to the interface states, in parallel with C_{ox} is neglected because of its minor impact when the interface trap density is low, as it is generally the case for silicon-silicon dioxide interfaces. In case of presence of additional parasitics, such as contact resistance, parasitic capacitance or high gate leakage current, the simple 2-element model might not be accurate and a better capacitance model need to be considered [11].

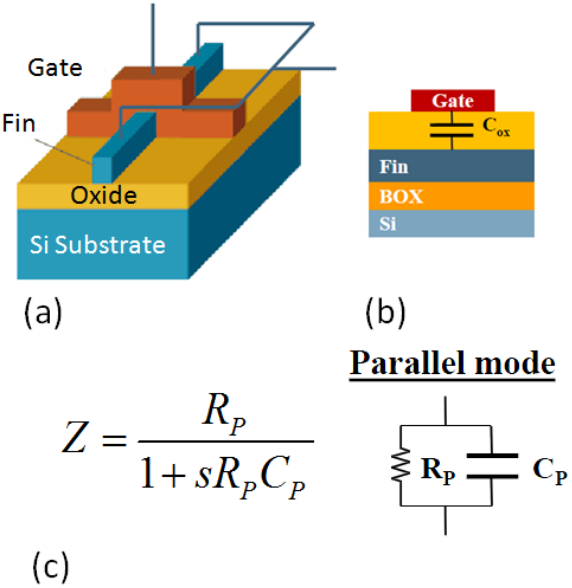


Figure 4.5: Tilted view of a SOI finfet (a), cross section of the fin showing the C_{OX} component (b) and schematic of the two elements model (c) used for CET extraction.

Due to the absence of a bulk contact two different parameters can be extracted in SOI FinFETs from a Capacitance-Voltage (CV) measurement. In case of nMOS, the capacitance extracted in inversion regime is used to calculate the CET_{inv} (Capacitance Equivalent Thickness in inversion) or EOT (Equivalent Oxide Thickness) of the device while the overlap capacitance can be extracted from the capacitance measured in the accumulation regime. In case of a TFET, both nMOS and pMOS operations are present in the same device. In this case, the inversion capacitance can be extracted in both inversion and accumulation regimes, since both types of carriers can be provided to the channel either from the source or from the drain regions of the device. The C-V characteristics of a TFET and MOSFETs with the same gate dielectric thickness are shown in Figure 4.6 and are superposed in the inversion regions. As a consequence, some electrical parameters, such as EOT, extracted from conventional MOSFETs can also be used to describe TFETs.

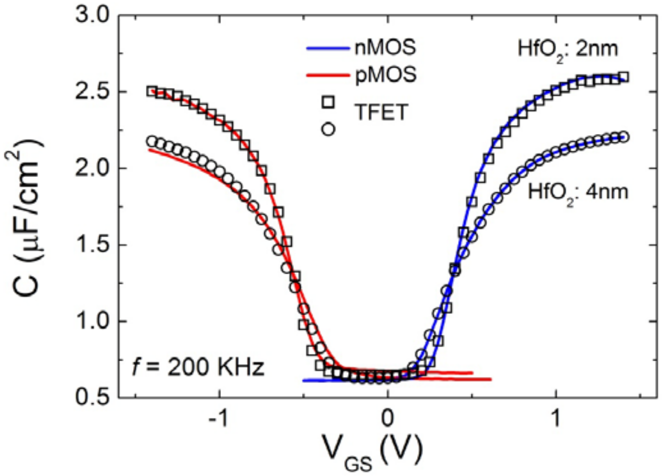


Figure 4.6: Capacitance-Voltage measurements of MOSFET and TFET with identical processing for two different HfO_2 thicknesses. The characteristic of the TFET is overlapping the two curves for nMOS and pMOS. The measured devices have a WL of 64µm^2.

The EOT indicates the thickness of an equivalent SiO_2 layer that would produce the same total measured capacitance. The extraction of this parameter can be cumbersome for SOI devices and an equivalent metric is used on this work. The inversion layer thickness is commonly used to compare high-k/metal gates. It is defined as follows:

$$T_{inv} = \varepsilon_0 \varepsilon_{SiO_2} \frac{A}{C_{\mathrm{inv}}} \tag{5.1}$$

where A is the gate area and C_{inv} is the inversion capacitance extracted from capacitance-voltage characteristics at V_{GS} = 1.5 V as shown in Figure 4.6. The value of T_{inv} is extracted from the slope of the maximum capacitance at 1MHz versus gate length assuming a constant gate width. Possible parasitics are removed by measuring devices with different gate lengths, i.e. different gate areas. EOT is obtained by substracting 4Å from the T_{inv} (or CET_{inv}) to take into account the effect of the quantum capacitance confinement.

Extraction of Interface Traps Density

Interface traps or states are attributed to dangling bonds at the semiconductor/insulator interface. Their density is commonly reduced by forming gas (H_2) anneal. Several methods have been developed to extract the interface traps density in MOSFETs, such as quasi-static method, conductance method [12], Deep Level Transient Spectroscopy (DLTS) [13] and charge pumping technique [14]. In particular, charge pumping method is suitable to small geometry devices instead of large diameter MOS capacitors; hence, the measurement can be performed directly on the devices without the use of dedicated structures.

The basic concept of this technique is explained in Figure 4.7 [15]. The MOSFET source and drain are tied together and slightly reverse biased with a voltage V_R while a pulse voltage with an amplitude large enough to drive the device between inversion and accumulation is applied at the gate. The shape of the pulse train can be square, triangular, trapezoidal, sinusoidal or trilevel. When the transistor is pulsed into inversion, the surface becomes fully depleted and the electrons flow from the source and drain region into the channel, where some of them fill the surface traps (Figure 4.7c). When the transistor is driven into accumulation, the mobile charge drifts back from the channel to the source and the drain but the charge trapped at the surface that has no time to be released from the states, recombines with majority carriers in the substrate (Figure 4.7d-f) and gives rise to a net substrate current flow, called charge pumping current, I_{CP} which is proportional to the interface traps density, D_{it}, according to the following equation:

$$I_{CP} = qfAD_{it}\Delta E \tag{5.2}$$

where q is the electron charge, f is the frequency, A is the area contribution of the Si-SiO_2 interface and ΔE is the energy region in the bandgap that contributes to the CP current. When an SOI substrate is used or when the device does not present a bulk contact, such as a vertical nanowire, the charge pumping experimental setup needs to be modified. In this case, a gated diode is used to extract the trap density [16] In particular, the P+ side acts as a collector contact of the I_{CP} current while the other N+ side provides the electrons to the channel (n-type conduction). Even though the basic charge pumping technique only gives an average value of D_{it} over the energy interval, ΔE, the accurate energy distribution of the interface traps can be obtained by varying the pulse rise and fall time [17].

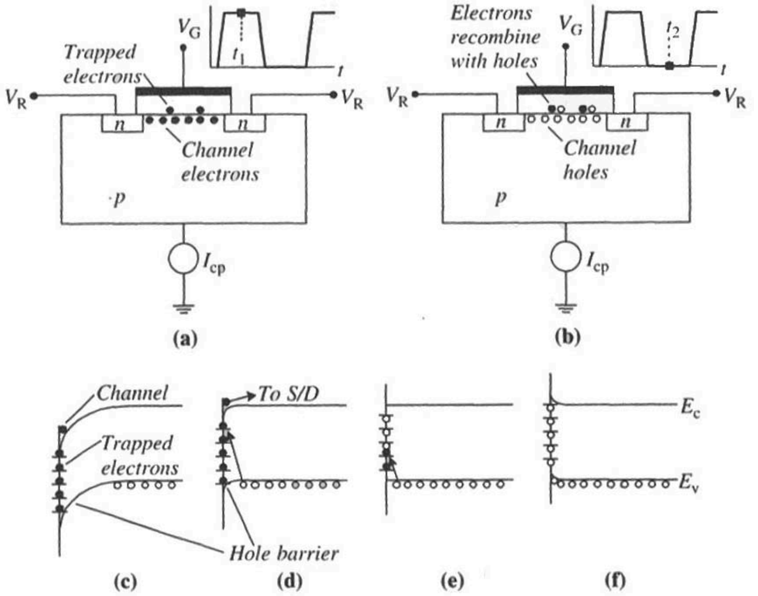


Figure 4.7: Device cross-section (a, b) and energy bands for charge pumping measurements (c-f). [15]

In this work, an average value of D_{it}, was extracted for finFET-based TFET. The extraction of the interface trap density, D_{it}, was carried out by applying a trapezoidal pulse to the gate electrode of the devices. The pulse is of constant amplitude and frequency and the measurement is obtained by sweeping the base voltage of the pulse as described in [17]. In addition, sidewalls and top interface components of the D_{it} were separated according to the methodology described in [18]. This technique has been used for FinFET-based TFETs and the results will be shown in Chapter 5. A similar extraction is not possible for vertical nanowire because of the presence of a large underlap at the drain side implemented to reduce the ambipolarity of the device.

4.3 Tunnel Field Effect Transistor (TFET)

It is not obvious to design and optimize a transistor that turns on with the tunneling process. Most of the tunneling mechanisms are considered detrimental for MOSFETs and major efforts are made to reduce the tunneling mechanisms instead of enhancing them. In this section, we avoid to speculate on the best design for TFET but we focus on the fundamental understanding of the interband tunneling in experimental devices starting from the basic design originally proposed in [19] and widely used to describe the Si TFET. The basic embodiment of a TFET is a gated reverse-biased diode as depicted in Figure 4.8. The gate voltage modulates the width of the tunneling barrier and not the height of the barrier as it is the case in a MOSFET. In the off-state, the width of the tunneling barrier is represented by the distance between the source and the drain while it becomes only a few nanometers when a positive or negative gate voltage is applied. As in the MOSFET, it is possible to distinguish two operation modes for a TFET:

- *p-type mode* when the gate bias is negative and the tunneling occurs between the n^+region and the channel (pTFET),
- *n-type mode* when the gate bias is positive and the tunneling occurs between p^+ region and the channel (nTFET).

These 2 modes of operation can occur in the same device. This is why the device is called ambipolar. In order to maintain a similar sign convention as in MOSFETs, the p+ region is called either source or drain in case of n-mode or p-mode, respectively. The source region, where the tunneling occurs, is always grounded. As a result, when different drain biases are applied, the gate-to-source bias is not modified. It is important to remember that the TFET is a one-directional device: the current always flows from the n^+ region to the p^+ region because the diode is always reverse biased.

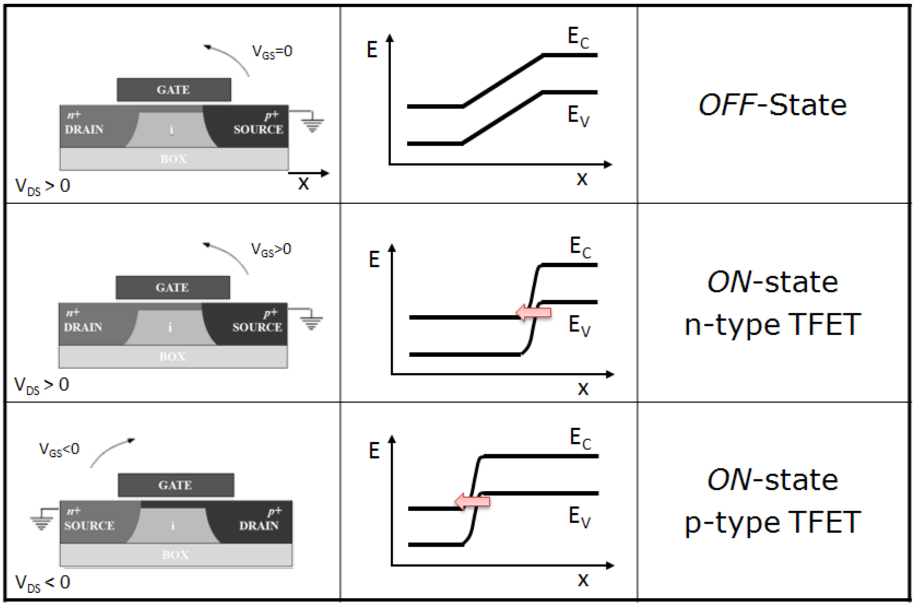


Figure 4.8: Schematic of the different operation modes in a TFET. N- and p- type operations are both possible on the same device.

4.3.1 Transfer Characteristics

Figure 4.9 shows the transfer characteristics of an experimental Si TFET for both p- and n-modes of operation, at three different drain voltages. More details about the fabrication of these devices were given in Chapter 3. The transfer characteristic is clearly different from the MOSFET one. Firstly, the ambipolarity is an intrinsic property of the asymmetric design of TFETs, as shown in Figure 4.8. The p- and n-mode operations are nearly symmetrical because the workfunction of the metal gate was chosen to be midgap and because electrons and holes tunnel with similar rates. This effect can be reduced by reducing the doping concentration at the drain side or by increasing the distance between the gate and the drain region [20]. Secondly, the subthreshold swing is strongly dependent on the gate bias and it is therefore not possible to extract a constant slope in the subthreshold regime, as in MOSFETs. In addition, the definition of threshold voltage is still under discussion and will be presented later. The device turn on is not as abrupt as expected because the dopants are implemented by ion implantation which does not allow to obtain a nearly abrupt junction. Epitaxy is preferred but is difficult to implement for this device configuration. In addition, the presence of TAT can degrade the swing at low gate voltages as will be shown in Chapter 5. For large gate bias the characteristics are nearly saturating because they are limited by the tunneling resistance. Finally, the impact of the drain voltage is negligible at low gate bias while a weak dependence is observed at large gate bias.

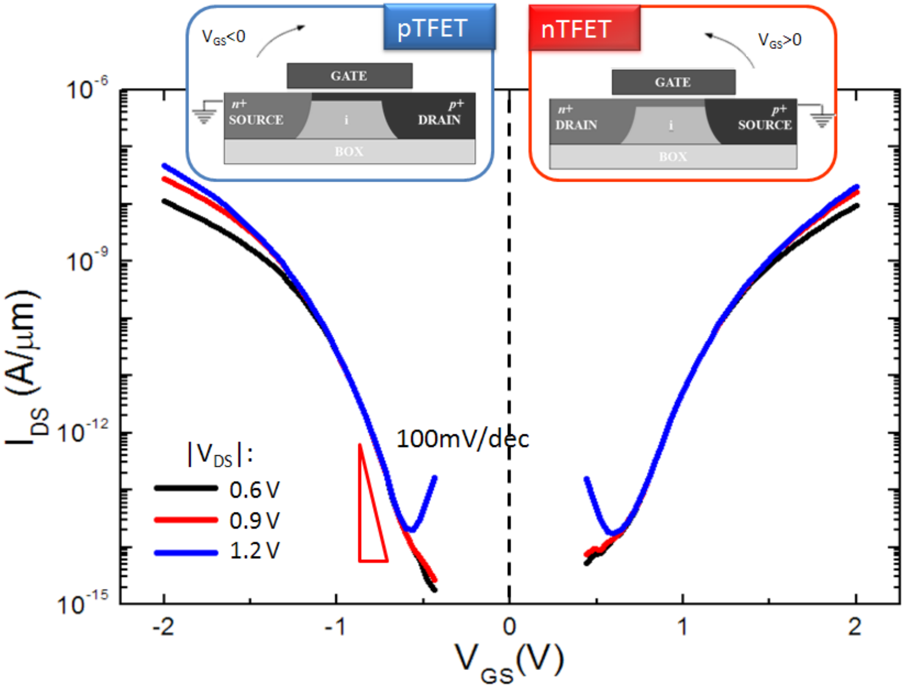


Figure 4.9: Transfer characteristic for a fabricated TFET at three different drain voltages showing the two operation modes. A schematic of the measurement setup is shown on the top of the figure. (W_{fin}: 250nm, L_G: 150nm)

4.3.2 Output Characteristics

Typical output characteristics (Figure 4.10a) are characterized by a super-linear onset at low drain bias and an excellent saturation of the current at higher bias. The presence of a super-linear onset represents one of main drawback of TFETs in circuit applications.

Its origin is related to the parabolic shape of the valence and conduction bands in the depletion region close to the source. In Figure 4.10b, a simulation of the energy bands along the channel direction is shown for different drain voltages. It is clear that increasing the drain voltage influences the amount of band bending near the source-channel interface where the tunneling occurs. This leads to a Drain Induced Barrier Thinning (DIBT) which is at the origin of the super-linear behavior at low drain bias as described in [21]. When larger drain bias are applied, the surface potential at the tunneling junction becomes independent from the drain voltage, i.e. the band edges of the conduction and valence band at the tunneling region are not modified anymore (Figure 4.10b). In this condition, the length of the tunneling path is constant as well the tunneling current, resulting in the saturation of the tunneling current.

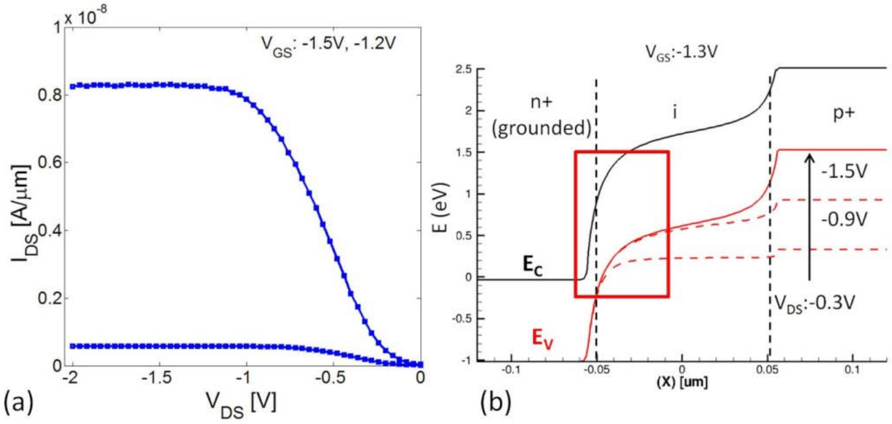


Figure 4.10: Output characteristics of experimental TFET at two different gate voltages (a) and simulated energy band diagram for different drain voltages for the valence band. The red area show the side where the tunneling occurs. (W_{fin}: 85nm, L_G: 150nm)

An alternative explanation of the super-linear onset involves the simple case of line tunneling [22] where the gate is overlapping the source region. In Figure 4.11a, a schematic of the energy band levels perpendicular to the gate is shown. The quasi Fermi level of the holes in the source region, $E_{F,p}$, is fixed by the source doping concentration while the quasi Fermi level for electrons in the channel, $E_{F,n}$, is assumed to be equal to V_{DS}. This assumption is valid only for small tunneling current, when the gradient of the $E_{F,n}$ along the channel direction is negligible. The tunneling is allowed only between $E_{F,n}$ and $E_{F,p}$ as indicated by the shaded area. For a fixed gate voltage, the edges of energy bands are fixed and only the $E_{F,n}$ is moving accordingly to the drain voltage. For low drain voltages, only the long tunneling paths are allowed, i.e. the one closer to $E_{F,p}$. For large drain bias, shorter tunneling paths become available. It results in the super-linear onset visible in Figure 4.11b. When $E_{F,n}$ crosses the bottom of the conduction band, the tunneling current saturates because the energy window for the tunneling remains constant and no more dependence with the drain bias is observed (Figure 4.11b). This effect is similar to the pinch-off effect in MOSFET with the only difference that in this case the inversion layer becomes thinner at the source and not at the drain.

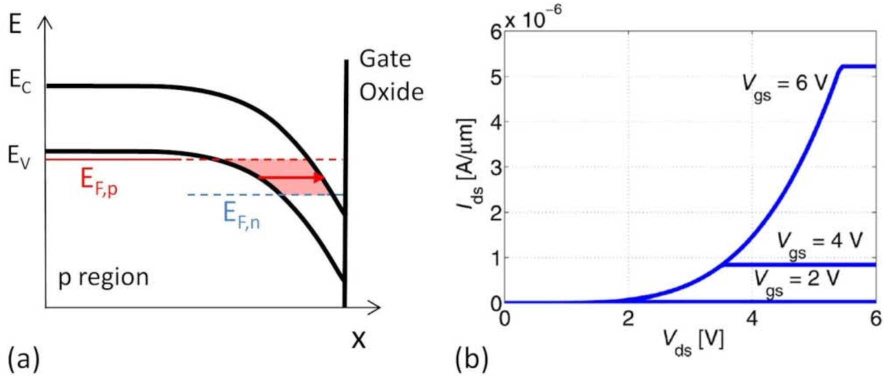


Figure 4.11: Schematic of the line tunneling configuration perpendicular to the gate (a) and output characteristics (b) as a result of the model presented in [22].

To reduce the undesired slow onset in the output characteristics, it is necessary to reach the saturation regime at lower drain voltage. Figure 4.12 shows the output characteristics for two different fin widths. In case of the narrow fin (W_{fin} 25nm), a reduction of the super-linear onset is observed because of the improved electrostatic control of the gate over the tunneling junction as it will be shown in chapter 5. In addition, a better linearity of the output characteristics can also be achieved by introducing an undoped or counter-doped layer [23] next or above the source material (beneath the gate oxide) In this way, the shape of the bands will be more linear or sub-linear rather than parabolic.

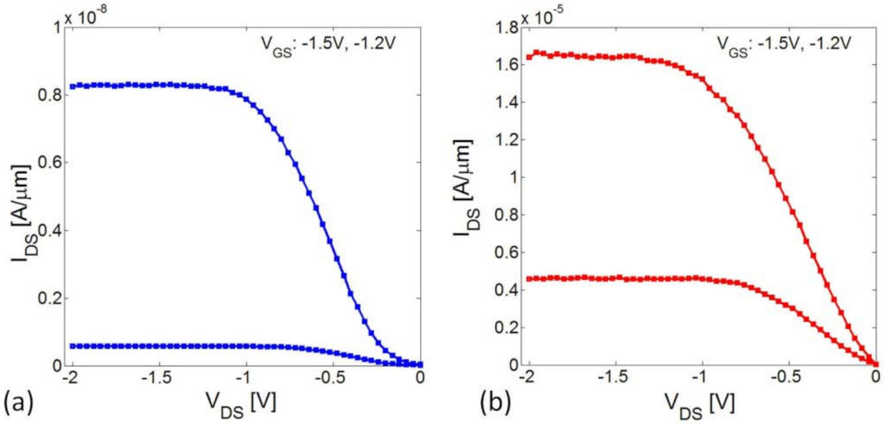


Figure 4.12: Output characteristics of experimental TFET with different fin width: 85nm (a) and 25nm(b) at two different gate voltages. (L_G: 150nm)

4.3.3 The subthreshold swing in TFET

Contrary to the MOSFET, the metric of the subthreshold slope is not applicable to TFET since it is not possible to extract a constant slope. In addition, the lack of the concept of threshold voltage complicates the definition of a universal metric for this device.

Several alternatives have been proposed in literature to benchmark TFETs. The most used method is to indicate the minimum point slope where the inverse slope of the I_{DS}-V_{GS} curve is extracted at the lowest current. Later, the definition of average slope was proposed [24-25] in which the threshold voltage is defined using the constant current method and the off current is the minimum current (Figure 4.13a). Alternatively, Boucart [17] proposed a new method to extract the threshold voltage based on the analysis of the transfer and output characteristics [26]. However, this definition is strongly dependent on the device geometry.

We believe that the metric of the subthreshold slope is not satisfactory to benchmark the TFET and we would like to propose a new metric based on the knowledge of the on-current and V_{DD} requirements for a given technology node (Figure 4.13b). In this case, the extraction of on-current is performed after defining the off-current level and the V_{DD} range.

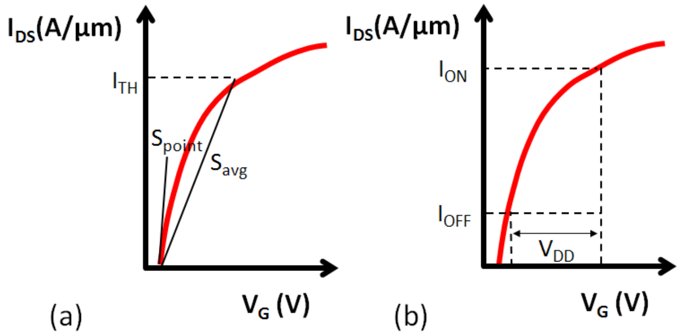


Figure 4.13: Schematic of the extraction of the point and average slope (a) and a new metric based on technology requirements (b).

4.3.4 How to Identify Band to Band Tunneling

From an experimental point of view it is important to verify if BTBT is the main transport mechanism in fabricated TFETs. As shown in the previous chapter, Kane's model can be used as a first approximation to describe the BTBT. In presence of an uniform electric field, the BTBT generation rate can be written as [27]:

$$G_{Kane} = A_{Kane}\left(E_{G,}m^*\right)F^2 \exp\left(\frac{B_{Kane}\left(E_{G,}m^*\right)}{F}\right) \tag{8.1}$$

where E_G is the energy band gap of the material where the tunneling occurs, m^* is the carrier effective mass, F is the electric field at the tunneling junction and A_{Kane} and B_{Kane} are fitting parameter. Equation (8.1) shows that tunneling is exponentially dependent on material properties, such as energy bandgap and effective masses, and on the electric field at the tunneling junction. The field at the tunneling junction is mainly controlled by the gate-source voltage in case of good electrostatic control. In addition, there is no explicit temperature dependence and the bandgap variation with temperature is expected to be small. BTBT transport can be identified in different ways: (1) good correlation (fitting) with Kane's model (2) weak temperature dependence and (3) independence of BTBT on gate length. We will next detail each of these properties.

Temperature Dependence

A feature of BTBT is its weak temperature dependence. Contrary to a MOSFET, the swing of a TFET should not change by varying the temperature unless parasitic mechanisms such as TAT and SRH are present. In order to reduce the impact of these parasitic components, measurements at low temperature are required. Figure 4.15 shows a typical temperature dependence of the transfer curves of a TFET. The current is slightly increasing by rising the temperature at large gate voltage because of the weak temperature dependence of the energy band gap. On the other hand, the strong temperature dependence observed at low gate bias is due to the presence of parasitic components such as SRH Generation and TAT as described in Chapter 2.

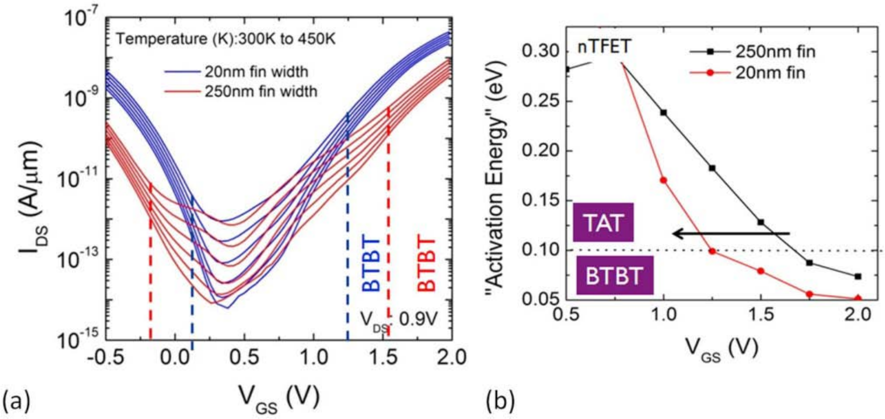


Figure 4.14: Transfer characteristics for experimental nTFET at different temperature (a) and Arrhenius plot (b). (L_G: 150nm)

From I_D-V_G Plot

The transfer characteristic of the TFET, or I_{DS}-V_{GS} plot, has an exponential dependence with the electric field at the tunnel junction. In lateral TFETs, the height of the tunneling barrier is defined by the energy bandgap, which is a material parameter. On the other hand, the tunneling length is mainly controlled by the gate bias. Assuming the [24]After adjusting equation (8.1) as suggested in [24] we obtain:

$$\log\left(\frac{I_{DS}}{V_{GS}^2}\right) = \log\left(\frac{A_{Kane}D^2}{E_g^{1/2}}\right) - \frac{B_{Kane}E_g^{3/2}}{DV_{GS}} \tag{8.2}$$

where D is a fitting parameter assuming that $F_{max} = DV_{GS}$ [24]. Thus, the plot of $\log(I_{DS}/V_{GS}^2)$ versus V_{GS}^{-1} should give a straight line indicating a good fit with Kane's model. A good linear fit of the experimental data with the Kane's model is shown in Figure 4.14. We do not attempt to determine the exact value of A_{Kane} and B_{Kane} because of the difficulty to analytically extract specific material parameters such as doping concentration. The calibration of the model requires specifically designed structures as described in chapter 2.

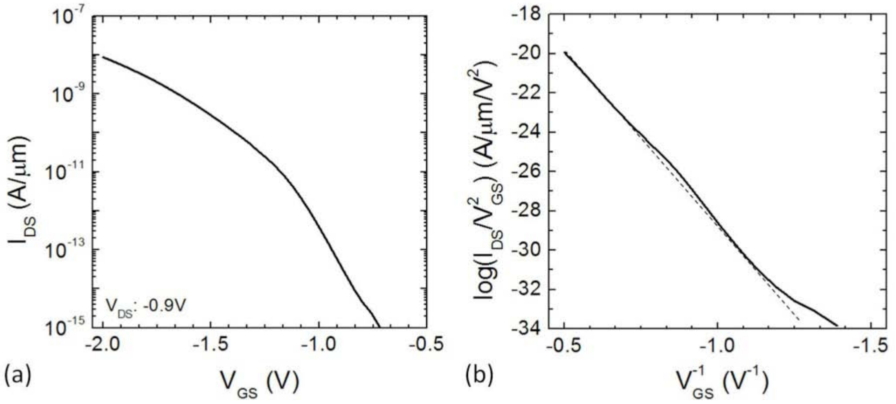


Figure 4.15: Transfer characteristic of an experimental Si pTFET (a) and the linear fit according to Kane's model. (W_{fin}: 250nm, L_G: 150nm)

An alternative method to show clearly the strong impact of the gate voltage on the tunneling current is to plot the subthreshold swing versus the gate voltage. Contrary to the MOSFET where a constant value of subthreshold slope can be extracted, the TFET is characterized by an increasing subthreshold swing with the gate voltage.

Gate Length Scaling

The third property of a TFET is that the on-current does not change when varying the gate length because band to band tunneling is extremely localized at the tunneling junction. It is therefore not affected by the gate length. Nevertheless, this property might not be valid for short gate lengths (below 100nm) where effects such as strain might change the performance of the device. Figure 4.16 shows experimental on-state current for different gate lengths down to 150nm. The current remains constant except for very long gate length for which the channel resistance becomes dominant.

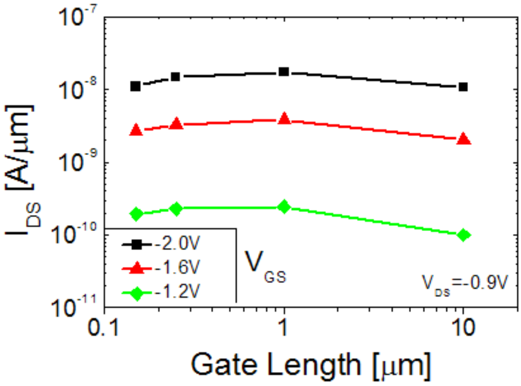


Figure 4.16: Gate length dependence in TFET. Data are extracted from finfet-based TFET at different drain voltages. The degradation observed for the 10um long device is related to channel resistance.

4.4 Summary

This chapter introduces to the characterization methods for TFET focus on the electrical characterization. Firstly, the measurements setup available is described including state of the art nanoprober and conventional industrial probing system. Later, some of the characterization techniques show how to extract device parameters when a bulk contact is not present. More information regarding the physical characterization techniques are presented in Appendix A.

In the second part of this chapter, the main features of TFET are listed. In particular, the strong dependence of the tunneling current with the gate voltage complicates the definition of threshold voltage and the benchmarks of TFET devices. A new metric is proposed based on V_{DD} and on-state current requirements. The transfer and output characteristics for TFET are briefly described. The non-linear onset present in TFET can be reduced by a proper device optimization such as increasing the gate coupling.

References

[1] A. Vandooren, *et al.*, "A 35nm diameter vertical silicon nanowire short-gate tunnelFET," presented at the Silicon Nanoelectronics Workshop, Kyoto, 2009.

[2] Z. X. Chen, *et al.*, "Demonstration of Tunneling FETs Based on Highly Scalable Vertical Silicon Nanowires," *IEEE Electron Device Letters,* vol. 30, pp. 754-756, 2009.

[3] S. Mookerjea, *et al.*, "Temperature-dependent I-V Characteristics of a Vertical $In_{0.53}/Ga_{0.47}/As$ Tunnel FET," *IEEE Electron Device Letters,* vol. 31, pp. 564-566, June 2010.

[4] D. Leonelli, *et al.*, "Performance Enhancement in Multi Gate Tunneling Field Effect Transistors by Scaling the Fin-Width," *Japanese Journal of Applied Physics,* vol. 49, p. 04DC10, 2010.

[5] M. Fulde, *et al.*, "Fabrication, optimization and application of complementary Multiple-Gate Tunneling FETs," in *the 2nd IEEE International Nanoelectronics Conference* 2008, pp. 579-584.

[6] K. Sung Hwan, *et al.*, "Tunnel Field Effect Transistor With Raised Germanium Source," *IEEE Electron Device Letters,* vol. 31, pp. 1107-1109, 2010.

[7] K. K. Bhuwalka, *et al.*, "Vertical tunnel field-effect transistor," *IEEE Transactions on Electron Devices,* vol. 51, pp. 279-282, 2004.

[8] D. Leonelli, *et al.*, "Novel architecture to boost the vertical tunneling in Tunnel Field Effect Transistors," in *IEEE International SOI Conference*, 2011, pp. 1-2.

[9] F. Mayer, *et al.*, "Impact of SOI, $Si_{1-x}Ge_xOI$ and GeOI substrates on CMOS compatible Tunnel FET performance," in *IEEE International Electron Devices Meeting* New York, 2008, pp. 163-166.

[10] Kleindiek, *Quick Reference: MM3A-EM System setup and quick reference guide*, 2006.

[11] K. J. Yang and H. Chenming, "MOS capacitance measurements for high-leakage thin dielectrics," *IEEE Transactions on Electron Devices,* vol. 46, pp. 1500-1501, 1999.

[12] E. H. Nicollian and A. Goetzberger, "The Si-SiO_2 Interface - Electrical Properties as Determined by the Metal-Insulator-

Silicon Conductance Technique," *Bell Syst. Tech. J.,* vol. 46, pp. 1055-1133, 1967.

[13] E. Rosencher and D. Bois, "Comparison of interface state density in MIS structure deduced from DLTS and Terman measurements," *Electronics Letters,* vol. 18, pp. 545-546, 1982.

[14] J. S. Brugler and P. G. A. Jespers, "Charge pumping in MOS devices," *IEEE Transactions on Electron Devices,* vol. 16, pp. 297-302, March 1969.

[15] D. K. Schroder, *Semiconductor Material and Device Characterization*, Third Edition ed.: Wiley Inter-Science, 2006.

[16] T. Ouisse, *et al.*, "Adaptation of the charge pumping technique to gated p-i-n diode fabricated on silicon on insulator," *IEEE Transactions on Electron Devices,* vol. 38, pp. 1432-1444, 1991.

[17] G. Groeseneken, *et al.*, "A reliable approach to charge-pumping measurements in MOS-transistors," *IEEE Transactions on Electron Devices,* vol. 31, pp. 42-53, 1984.

[18] G. Kapila, *et al.*, "Direct measurement of top and sidewall interface trap density in SOI FinFETs," *IEEE Electron Device Letters,* vol. 28, pp. 232-234, 2007.

[19] W. M. Reddick and G. A. J. Amaratunga, "Silicon Surface Tunnel Transistor," *Applied Physics Letters,* vol. 67, pp. 494-496, 1995.

[20] A. S. Verhulst, *et al.*, "Tunnel field-effect transistor without gate-drain overlap," *Applied Physics Letters,* vol. 91, pp. 053102-3, 2007.

[21] J. Knoch, *et al.*, "One-Dimensional Nanoelectronic Devices - Towards the Quantum Capacitance Limit," in *Device Research Conference Digest*, 2008, pp. 173-176.

[22] A. S. Verhulst, *et al.*, "Drain voltage dependent analytical model of tunnel field-effect transistors," *Journal of Applied Physics,* vol. 110, p. 024510, 2011.

[23] H. Chenming, *et al.*, "Prospect of tunneling green transistor for 0.1V CMOS," in *IEEE International Electron Devices Meeting* 2010, pp. 16.1.1-16.1.4.

[24] K. K. Bhuwalka, *et al.*, "A simulation approach to optimize the electrical parameters of a vertical tunnel FET," *IEEE Transactions on Electron Devices,* vol. 52, pp. 1541-1547, 2005.

[25] K. Boucart and A. M. Ionescu, "Double-gate tunnel FET with high-kappa gate dielectric," *IEEE Transactions on Electron Devices,* vol. 54, pp. 1725-1733, 2007.
[26] K. Boucart and A. M. Ionescu, "A new definition of threshold voltage in Tunnel FETs," *Solid-State Electronics,* vol. 52, pp. 1318-1323, Sep 2008.
[27] E. O. Kane, "Zener tunneling in semiconductors," *Journal of Physics and Chemistry of Solids,* vol. 12, pp. 181-188, 1960.

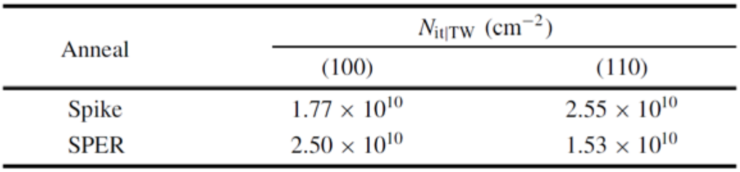
Chapter 5

FinFET-based Si TFETs

5.1 TFET at imec

At the early stage of the thesis, silicon material has been selected as a starting material to study the band-to-band tunneling because of its easy integration and compatibility with CMOS processing. The finFET configuration was used as a first test vehicle to study the homo-junction Si TFET because of its more mature technology compared to the GAA technology while still providing nearly the same tight gate control and most integration steps readily available at imec.

The integration flow used for finFET-based TFET has been extensively discussed in chapter 3. Figure 5.1 reports the main process steps for the Process Of Reference (POR). This process flow is considered as the process reference for homo-junction Si TFETs presented in this chapter.



Anneal	$N_{it\|TW}$ (cm^{-2})	
	(100)	(110)
Spike	1.77×10^{10}	2.55×10^{10}
SPER	2.50×10^{10}	1.53×10^{10}

Table 5.1: Summary of the extracted density of interface traps. SPER anneal is compared with spike anneal. Fin topwall with (100) orientation and fin sidewall with (110) orientation are reported.

The good quality of the silicon/oxide interface was verified by performing charge pumping measurements (as described in Chapter 4). A low density of interface traps close to 10^{10} cm^{-2} for both top and sidewall contributions was obtained. No dependence on anneals is observed (Table 5.1).

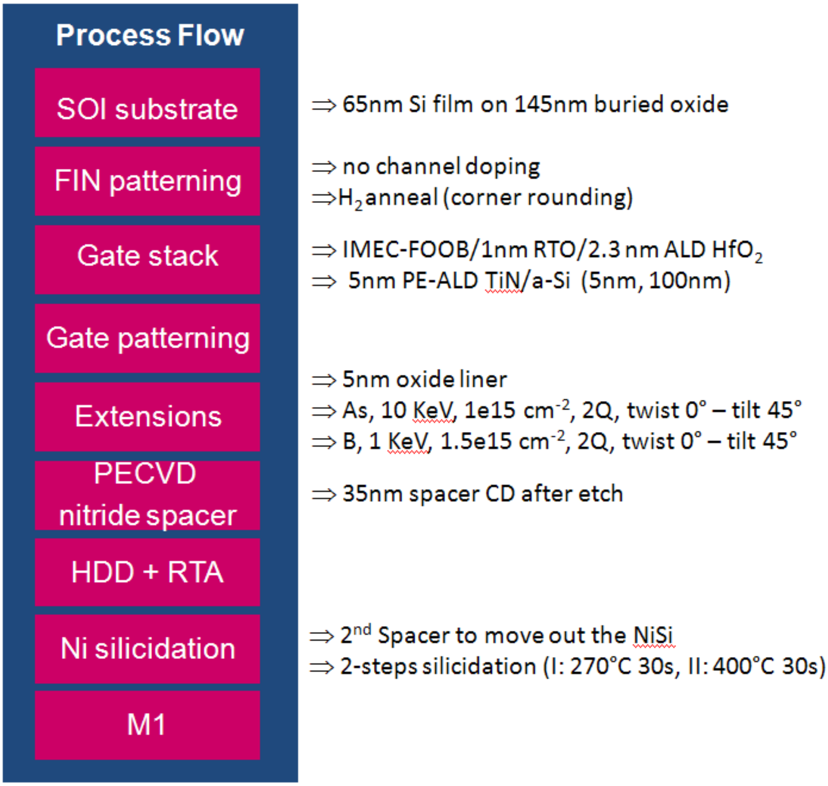


Figure 5.1: Process flow used as a process of reference (POR) for TFETs.

The electrical results of the TFETs processed using the POR conditions are presented in Figure 5.2. The p- and n-type operations are measured on the same device because of the ambipolarity of the TFET. The on-current of this device is below 100 nA/µm at V_{GS} of 2 V and V_{DS} of 1.2V while the subthreshold swing is far to be close to 60 mV/dec. It is clear that a process optimization is needed to boost the on-current and reduce the swing.

In this chapter, an extensive study of several geometrical and process parameters has been carried out to find their impact on the performance. The variation of the gate oxide thickness, the fin width and the dopant profile is analyzed. All these parameters should have an impact on the electric field at the tunneling junction. In particular, several methods to achieve abrupt doping profiles are proposed: optimization of the implantation conditions, modification of the activation anneal and dopant segregation by means of silicidation. Most of the learning can be transferred to new materials and offers a few guidelines for further development of TFET devices.

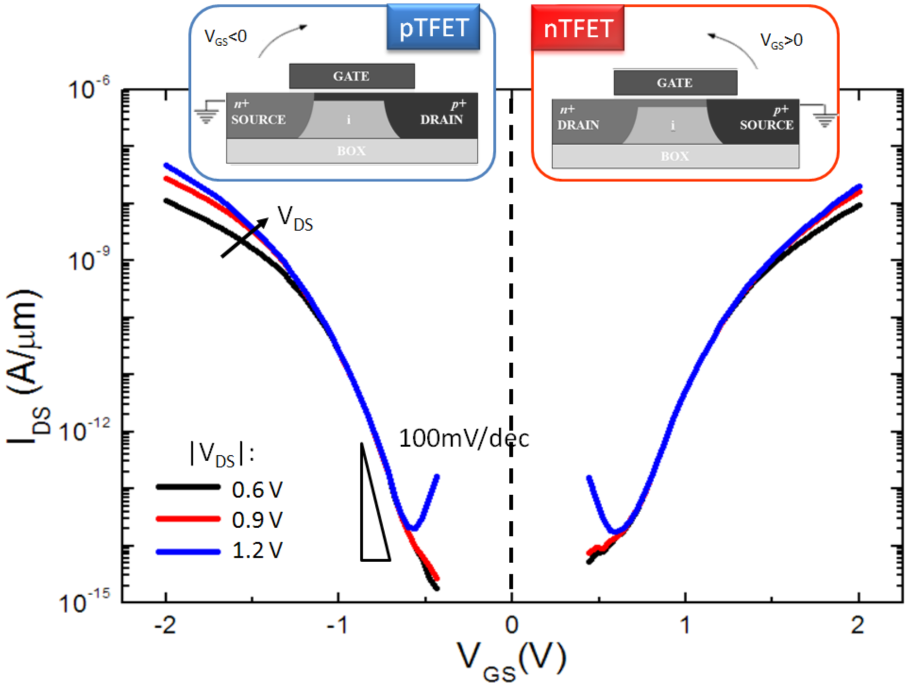


Figure 5.2: Transfer characteristics of a fabricated TFET measured in n- and p-type operation at different drain voltage. The side where the tunneling occurs is named conventionally source and kept grounded. (W_{fin} 20nm, N_{fin} 15, L_G 150nm)

5.2 Impact of Fin Width in TFETs

The finFET architecture, originally proposed by Hisamoto [1], aims to increase the gate control over the channel by partially surrounding the channel with a gate. The improved Electrostatic Integrity (EI) of this configuration is able to reduce SCEs in conventional MOSFETs because of the reduction of the influence of the electric field lines between the source and the drain. A better EI is expected to be beneficial to TFET as well because the tunneling current is exponentially dependent on the electric field at the tunneling junction which is controlled by the gate voltage.

This section studies the impact of improved gate control over the channel on the performance of fabricated Si MuGTFETs [2]. For narrow fins, the devices are considered as Fully Depleted (FD) double gate devices and the sidewall components are dominant over the top component. In this case, the silicon body thickness is represented by the fin width itself. On the other hand, for wide fins, the dominant component is the top one instead of the sidewalls and, consequently, the device can be considered as Partially Depleted (PD) devices where the body thickness is the fin height which is constant and equal to 65 nm.

5.2.1 Gate Geometry and Electrostatic Integrity

In conventional planar SOI transistor the EI can be improved by decreasing the silicon film thickness (or the junction depth in case of a bulk device). The motivation behind this phenomenon can be understood by solving the Poisson’s equation using the depletion approximation:

$$\nabla^2\Phi(x,y,z) = \frac{qN_a}{\varepsilon_{Si}} \tag{5.1}$$

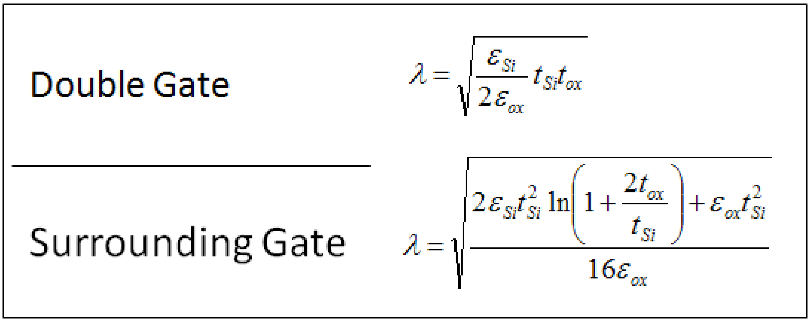
For single-gate SOI MOSFETs, the equation(5.1) can be re-written as follows [3]:

$$\frac{d^2\varphi(x)}{dx^2} - \frac{\varphi(x)}{\lambda^2} = 0 \tag{5.2}$$

where,

$$\lambda = \sqrt{\frac{\varepsilon_{Si}}{\varepsilon_{ox}} t_{ox} t_{Si}} \tag{5.3}$$

λ is called the natural length and is a parameter that represents the spread of the electric potential along the channel. It depends on the silicon film thickness (t_{Si}) and the oxide thickness (t_{OX}), two parameters which have a major impact of the EI of the device. In the same way, λ can be extracted for other configurations (Table 5.2). The natural length is typically used to estimate the maximum silicon film thickness and oxide thickness that can be used to avoid the SCEs at a given transistor length.



Double Gate	$\lambda = \sqrt{\frac{\varepsilon_{Si}}{2\varepsilon_{ox}} t_{Si} t_{ox}}$
Surrounding Gate	$\lambda = \sqrt{\frac{2\varepsilon_{Si} t_{Si}^2 \ln\left(1+\frac{2t_{ox}}{t_{Si}}\right) + \varepsilon_{ox} t_{Si}^2}{16\varepsilon_{ox}}}$

Table 5.2: Natural length for double gate and surrounding gate.

5.2.2 Impact of Gate Geometry on TFET

The concept of natural length, or screening length, has been extended to TFET by Knoch and Appenzeller [4-5], where it is used to describe the electrostatic of the tunneling junction assuming that the electric field is concentrated at the tunneling junction as shown in Figure 5.3. In this picture, λ represents the length of the tunneling path while $\Delta\Phi$ is the energy range where the tunneling occurs.

As a consequence, the electric field can be expressed by:

$$F = \frac{E_G + \Delta\phi}{\lambda} \tag{5.1}$$

And the tunneling probability after WKB approximation can be written as:

$$T(E) \propto \exp\left(-\frac{4\sqrt{2m^*}E_G^{3/2}}{3|e|\hbar(E_G + \Delta\phi)}\lambda\right) \tag{5.2}$$

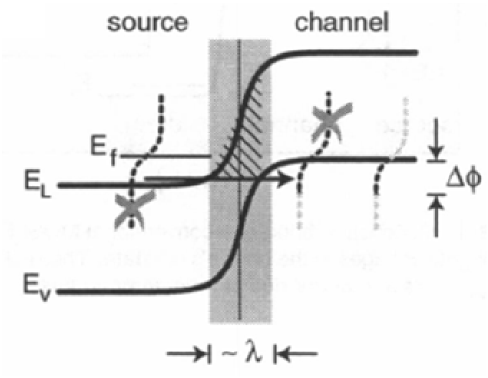


Figure 5.3: Conduction and valence bands in a tunneling. The Fermi level filter action of the band gap is shown. [4]

Following a similar approach, a linear dependence of the logarithmic of the tunneling current with the square root of the silicon film thickness has been qualitatively shown by a commercial TCAD simulator tool for double gate structures [6]. In particular, an improvement of the subthreshold swing has been observed when decreasing the silicon thickness as well as increasing the gate dielectric constant. Unfortunately, a degradation of the device performance is also reported for a silicon thickness below 10nm, possibly due to the reduced cross-sectional area available for current flow (i.e. channel resistance).

In this work, the reported linear trend has been verified experimentally on finFET-based TFETs, fabricated with the Salsa2 maskset featuring a minimum fin width of 10 nm. The device fabrication follows the POR conditions described in chapter 3 and the transfer characteristics are shown in Figure 5.4a. In Figure 5.4b, the plot of the tunneling current versus square root of the fin width is shown. At large fin widths, the finFET configuration effectively behaves as a single-gate configuration. At smaller fin widths, the coupling of the gates at the sidewalls is stronger and a rather linear dependence in observed. This linear correlation with the square root of the fin width is compared with the analytical modeling reported in [7] designed to search for a more quantitative agreement. The difficulty to obtain a quantitative agreement indicates that additional parameters such as the doping profile and concentration may also vary with the fin width and affect the band-to-band tunneling.

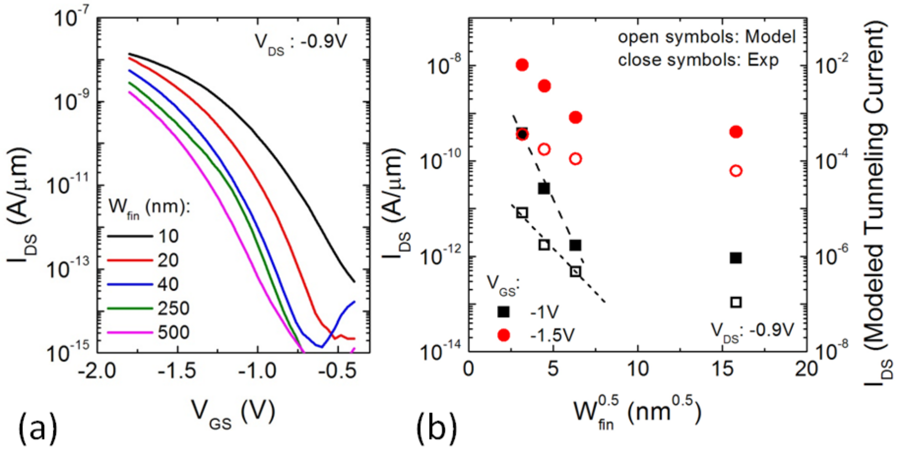


Figure 5.4: Transfer characteristics of 5 devices with different fin widths (a) and plot of the device current versus square root of the fin width showing the linear trend only for narrow fins (W_{fin} < 100nm) (b).

A 3D TCAD process simulation [8] is performed to show the expected doping distribution after the HDD implantation and annealing (Figure 5.5). For narrow fins, the doping profile along the channel direction (Figure 5.5b) results in a more uniform doping concentration compared to wide fins (Figure 5.5d), where the middle of the fin in the extension region remains undoped. The difference in doping profile for the two different fin dimensions leads to a different electric field distribution and this variation should be taken into account for the modeling. Unfortunately, it is extremely difficult to extract the effective active dopant concentration in case of narrow fins limiting the analysis to a qualitatively approach.

In conclusion, the dependence with the fin width can be explained by combining (1) the improved EI of the double gate architecture compared to the planar devices and (2) the better doping uniformity for narrow fins. The scaling of the fin width will be limited by quantum mechanical effects for dimensions below 10nm in case of Si TFET.

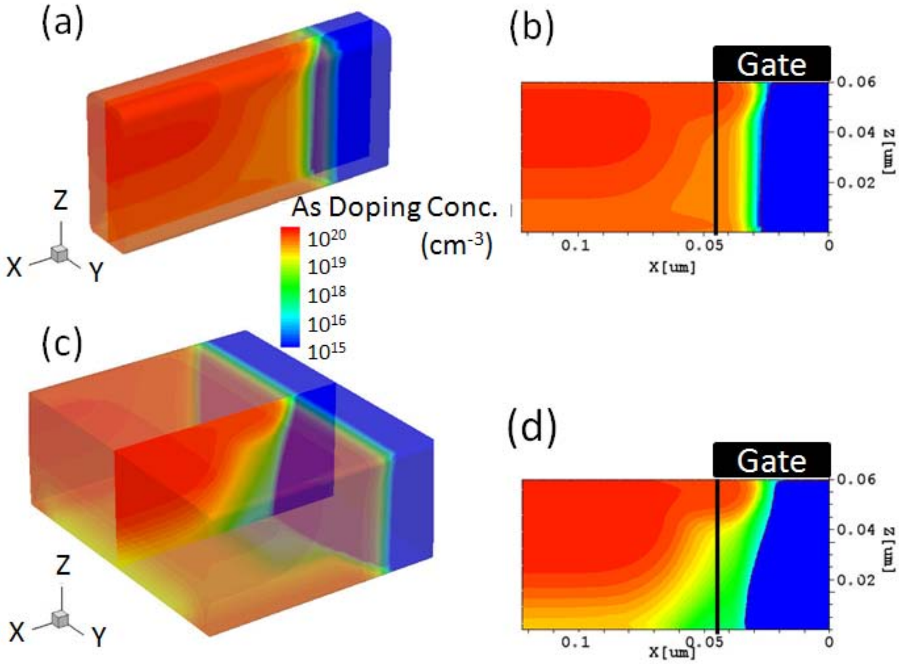


Figure 5.5: TCAD simulations show the expected doping profile for narrow (a) and wide (b) fin. On the right a cross section of the fin along the channel direction is shown for narrow (b) and wide (d) fins.

5.3 Oxide Scaling and TFET Performance

The EI can be improved by reducing the silicon film thickness, or the fin width, but also by reducing the gate oxide thickness or by increasing the permittivity constant (k) of the gate dielectric as shown in [6-7]. The oxide scaling offers a larger room for improvement compared to the scaling of the fin width because the reduction of the fin width will soon be limited by the increasing of the channel resistance and the geometrical quantization.

In our experiment, the interfacial oxide layer (IL) remained constant and the thickness of the high-k oxide was modified. Two thicknesses of HfO_2 were implemented: 2nm and 4nm on top of a 1nm IL. Larger HfO_2 thicknesses are not implemented because of their oxide quality degradation [9]. C-V measurements on conventional SOI MOSFET, fabricated together with the TFET devices, were used to calculate the EOT by extracting T_{inv}. A T_{inv} of 1.9 nm, corresponding to an EOT of 1.5nm, was extracted for the reference sample of 2nm HfO_2 while a value of T_{inv} of 2.6 nm (i.e. EOT of 2.2 nm) was extracted for the thicker gate dielectric.

Figure 5.6a shows the I_{DS}-V_{GS} characteristic of the measured devices (W_{fin}: 250nm) for two different gate oxide thicknesses. The impact of the gate oxide thickness is marginal compared to the simulation results, where a strong increase of the on-current together with an improvement of the swing was suggested [6, 10]. In the experimental devices with larger oxide thickness, the low onset voltage might be partially caused by a shift of flatband voltage for high-K oxides, for instance, the presence of fixed charges due to the instability of thick HfO_2 layer [11]. The same behavior was also observed for the threshold voltage in MOS devices measured on the same wafer. In addition, the gate leakage component is larger for thinner gate oxide, as expected, and becomes dominant at low gate and drain biases (Figure 5.6b).

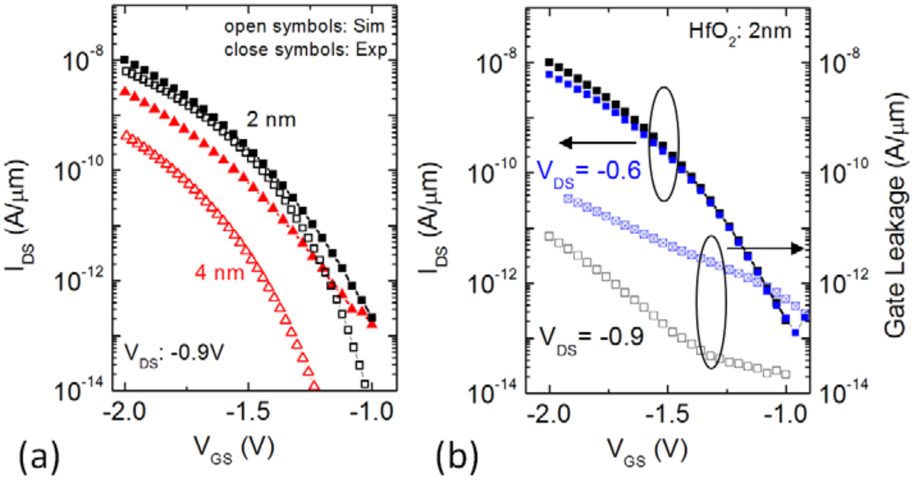
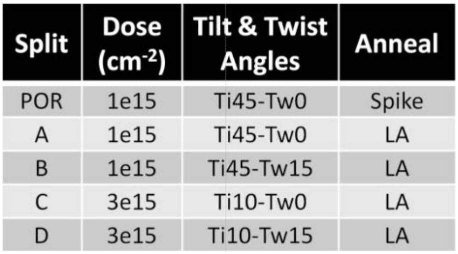


Figure 5.6: Transfer characteristics for pTFETs with 2nm and 4nm HfO_2 at V_{DS} = -0.9V. The close symbols represent the experimental data while the open symbols the simulation results obtained for abrupt profile and for equivalent EOT values (a). Transfer characteristics for pTFETs with 2nm HfO_2 at two different drain voltages (b). The data points are obtained by averaging several devices. (W_{fin} 250nm, L_G 150nm and N_{fin} 5)

5.4 Tunneling Junction Engineering through Doping

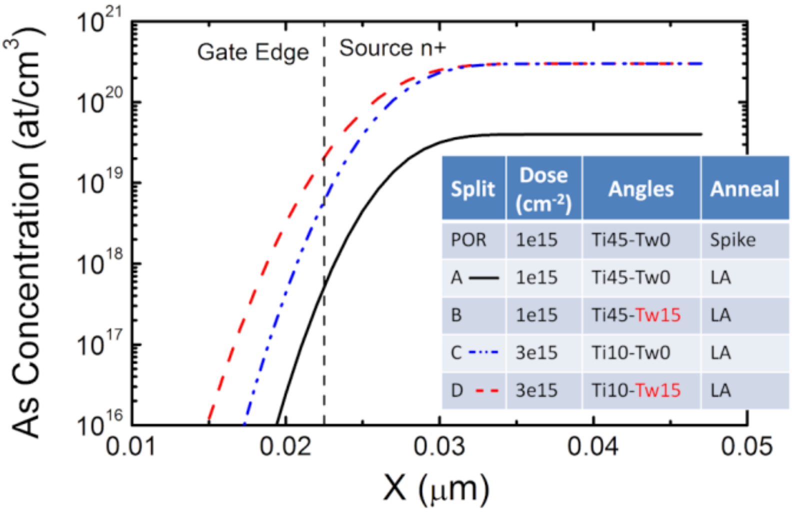
In section 5.2, it was shown that the current improvement resulted from the improved device geometry but also from differences in doping profile of the narrow fin compared to the wide fin. Junction doping engineering is an alternative method to boost the band to band tunneling current. The strategy used to tune the doping profile is based on engineering of the implantation angle and dose at the source junction.

The different implantations conditions investigated are listed in Table 5.3. The implantation is done in two quadrants and symmetric to the channel direction to provide a symmetrical doping profile along the fin. In addition, some wafers received a sub-ms laser as anneal activation step instead of the standard spike anneal. The purpose of this laser anneal is to increase the dopant activation rate and the steepness of the junction as reported in [12]. The focus is to optimize the tunneling at the n^+ side only. The tunneling at the p+ side is not considered because of the high diffusion of boron, used as p-type dopant.



Split	Dose (cm^{-2})	Tilt & Twist Angles	Anneal
POR	1e15	Ti45-Tw0	Spike
A	1e15	Ti45-Tw0	LA
B	1e15	Ti45-Tw15	LA
C	3e15	Ti10-Tw0	LA
D	3e15	Ti10-Tw15	LA

Table 5.3: Extension implantation conditions implemented for the dopant optimization at the n+ side. Sub-ms laser anneal is used as activation anneal after the HDD implant.



Split	Dose (cm^{-2})	Angles	Anneal
POR	1e15	Ti45-Tw0	Spike
A —	1e15	Ti45-Tw0	LA
B	1e15	Ti45-Tw15	LA
C -··-	3e15	Ti10-Tw0	LA
D - -	3e15	Ti10-Tw15	LA

Figure 5.7: Simulated doping profiles after the extension implantation with As. The energy used is 10keV for all the splits. Cut is performed along the fin direction, in the middle of the fin and 5 nm from the top of the fin. The doping splits are described in the inset.

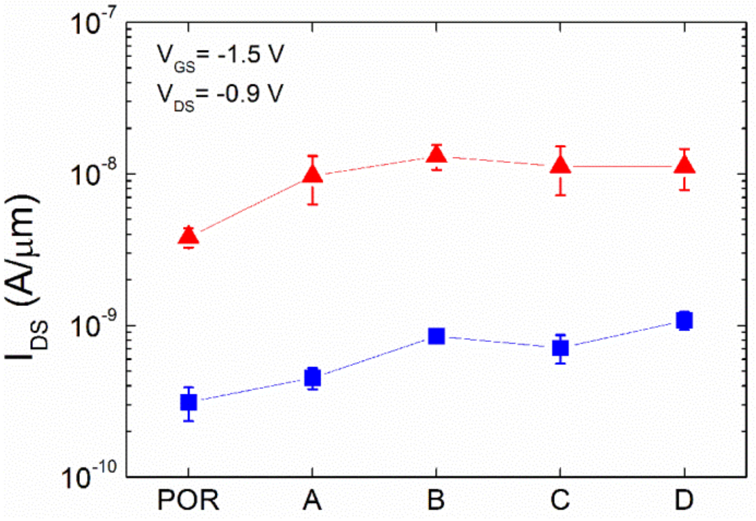


Figure 5.8: Average currents extracted for pTFET for different As extension implants. Trends for narrow fins (red), W_{fin} 20nm, and wide fins (blue), W_{fin} 250nm, are plotted. The values are extracted at V_{GS} equal to -1.5V.

The POR condition is used as a reference and two different doses and tilt angles are used to evaluate the impact of the total implantation dose. In addition, two different twist angles are implemented to increase the overlap of the source region with the gate. Figure 5.7 shows the simulated doping concentrations for three different conditions A ,C and D prior to the annealing.

The electrical results obtained for the different splits are summarized in Figure 5.8 for wide and narrow fins. Error bars are added to each value to show the dispersion of the results. First, a clear improvement is observed for the sub-ms laser anneal over the spike (POR) condition likely due to a steeper junction or to the higher dopant activation as explained at the beginning of this section. Secondly, comparing splits A and C an additional improvement of current for wide fins occurs due to the increased doping concentration. Thirdly, further increase of current is present in case of a larger twist angle during implantation because of a higher dopant concentration present under the gate. On the other hand, for narrow fins, the impact of the implantation conditions is much less outspoken. The reason is related to the implantation process in finFET devices. As explained at the beginning of this section, two quadrants implantation is used to assure a uniform implantation of the sidewalls of the fin. In the case of narrow fins, the implantation depths join in the middle of the fin (Figure 5.5). The profile achieved in narrow fins is therefore less sensitive to the implantation angles. In addition, we observe that the variability for narrow fins is larger than for wide fins. This might be due to fin width variability after etch for narrow fins and/or the full amorphization of the fins during the implantation, affecting the stability of these devices especially for higher doping dose.

For wide fins, the impact of the doping conditions remains relatively marginal. We believe that doping by implantation is not able to guarantee a profile steep enough at the tunneling junction to boost the tunneling rate. The impact of a graded junction on the tunneling characteristics has never been discussed in literature so far. In general, only abrupt and box-like profiles are considered. In this work, we now report the results of 2D-TCAD simulations [8] to qualitatively understand the impact of four different types of profiles at the tunneling junction on the tunneling performance in the case of wide fins.

The simulated structure is an SOI gated p-i-n diode with gate length of 120nm, channel thickness of 65nm and the buried oxide thickness of 145nm. The drain is p-type with concentration of $1\times10^{20}\text{cm}^{-3}$ and the channel concentration is $1\times10^{15}\text{cm}^{-3}$ p-type. Four different doping profiles are implemented for the n-type source with the following specifications:

- Graded profile with a peak concentration of $1\times10^{20}\text{cm}^{-3}$,
- Abrupt profile with a doping concentration of $1\times10^{20}\text{cm}^{-3}$,
- Graded profile with a peak concentration of $1\times10^{19}\text{cm}^{-3}$,
- Abrupt profile with a doping concentration of $1\times10^{19}\text{cm}^{-3}$.

In graded junctions, the peak concentration is located 5nm from the gate edge under the gate. In both cases, the doping gradient follows a Gaussian distribution as shown in Figure 5.9c.

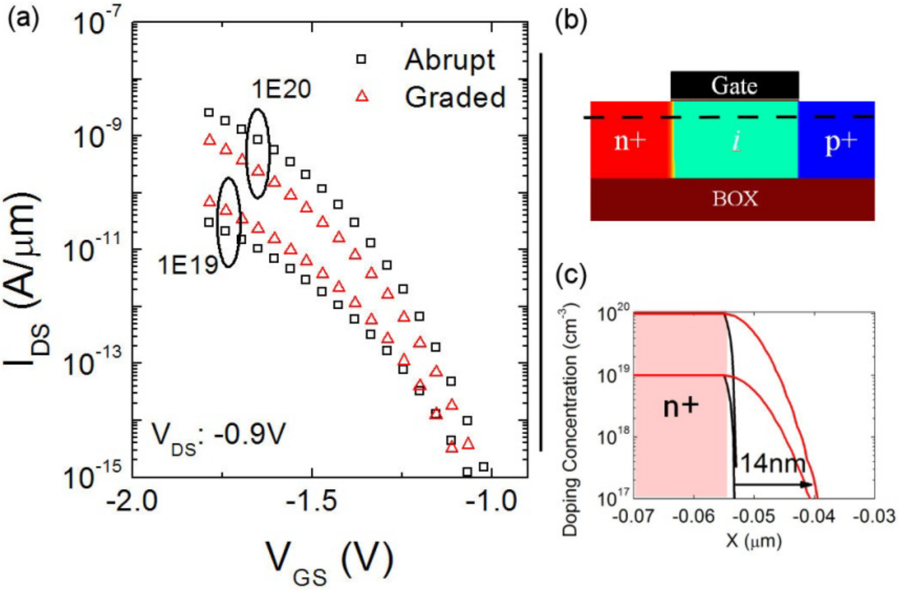


Figure 5.9: Simulated input characteristics for abrupt profile and for 1D Gaussian profile (Graded) along the channel as well as for two different doping concentrations (a). Schematic of the simulated device (b) showing the cut line where the doping profiles are extracted (c).

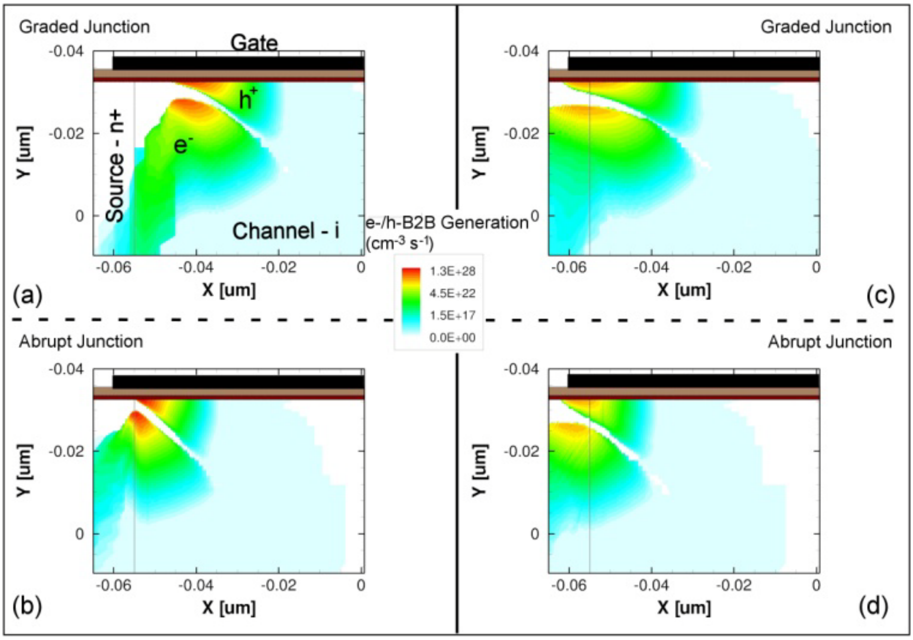


Figure 5.10: Band to band generation rate for electrons and holes extracted from 2D-TCAD simulation. The maximum donor concentration in the source is 1×10^{20} cm^{-3}(a,b) and 1×10^{19} cm^{-3} (c,d). Abrupt (b,d) and graded (a,c) junction profiles are shown. Position of the gate, source and channel are indicated in (a). Drain is not visible because of the plot range. The simulated device has a gate length of 120nm and the silicon channel thickness is 65nm.

Figure 5.9a shows the electrical results of the simulated structures. In case of abrupt junctions, a lower doping concentration results in a performance degradation as expected [13]. In case of graded junctions, a similar behavior is obtained but with a much reduced degradation. To better understand the reason behind this behavior, the band-to-band generation rate of electrons and holes is plotted in Figure 5.10 for the different structures. The band-to-band generation rate of electrons and holes is directly proportional to the tunneling current. Devices with a lower doping concentration (Figure 5.10c, d) have a lower carrier generation compared to the ones with higher doping concentration (Figure 5.10a, b). On the other hand, the graded junction in Figure 5.10c has a bigger tunneling area compared to the abrupt junction in Figure 5.10d enhancing the line (vertical) tunneling over the point (lateral) tunneling [14]. As a consequence, for lower doping concentration a higher tunneling current is reported for the graded junction. In case of the higher doping concentration, this effect is not observed because the tunneling distance, the distance between the two generation regions, increases (Figure 5.10a, b).

5.5 Impact of the Annealing Conditions

The annealing step is performed after the highly doped drain (HDD) implantation and it is responsible for the activation of the dopants and the recrystallization of implantation damage. This annealing also results in dopant diffusion and, hence, defines the final junction abruptness, critical for TFETs. The annealing conditions used in the experiments are:

- conventional spike anneal (POR),
- sub-ms laser anneal (LA),
- low temperature anneal for Solid Phase Epitaxy Regrowth (SPER) able to activate the dopants [15-16].

The spike anneal is performed at high temperature (1050°C) as described in chapter 3. On the other hand, laser anneal creates shallower junctions than the spike anneal because of the extremely fast pulse of the laser, while still activating the dopants. Three different conditions are used for the laser anneal to optimize the process (1200°C, 1250°C, 1300°C) while keeping constant the other process parameters such as the scan speed. The third anneal condition (SPER) is performed at 650C for 10min. This anneal is able to recrystallize the amorphized silicon during the implantation process while activating dopants and maintaining an ultra-shallow junction. However, end-of-range defects are expected to remain due to the low temperature of the process.

5.5.1 The Trend in MOSFET

The impact of the annealing steps on MOSFET is analyzed in this section. An estimation of the lateral depth of the junction is obtained by extracting the gate overlap capacitance from C-V measurements performed on test structures made of large area MOSFETs. Figure 5.11 summarizes the extracted lateral depth, L_{ov}, for the three different anneal conditions at the n-doped region and the p-doped region. The outcome is not very surprising: at the n+ side, the overlap length (L_{ov}) is comparable for all the splits due to the limited diffusion of As atoms. In contrast, at the p+ side B atoms diffused more than two times in the case of spike anneal compared to the case of LA or SPER anneals due to the larger diffusivity of Boron.

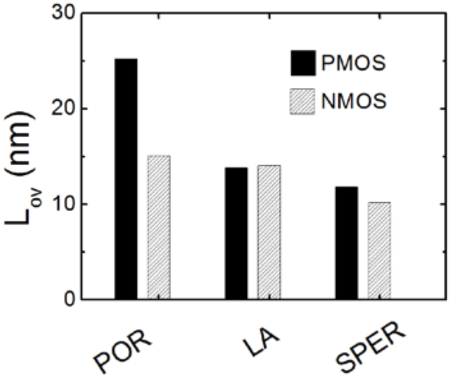


Figure 5.11: Overlap lengths extracted from overlap capacitance measurements for nMOS and pMOS for the different annealing conditions. The LA at 1250°C is displayed in this plot.

Electrical results on MOSFETs show a strong correlation between the gate overlap lengths and the transistor performance. In particular, an improvement of the pMOS drive current is recorded for LA (Figure 5.12) due to the shallower doping profile. In the case of nMOS devices there is no clear improvement as expected. On the other hand, an anomalous increase of off-current for nMOS is observed in case of the SPER. To understand the origin of this behavior, I_{DS}-V_{GS} plots for nMOS are shown in Figure 5.13 for LA and SPER anneals.

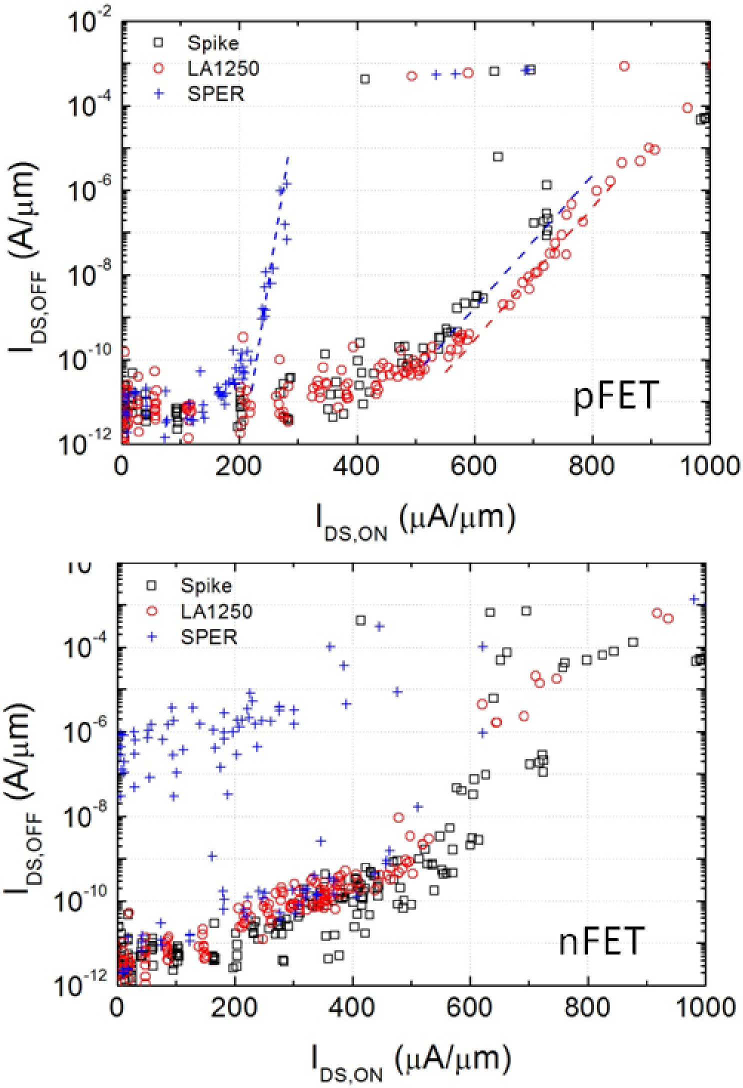


Figure 5.12: ITP extract at V_{DD} = 1V (nMOS), -1V (pMOS) for narrow fins (W_{fin} 20nm).

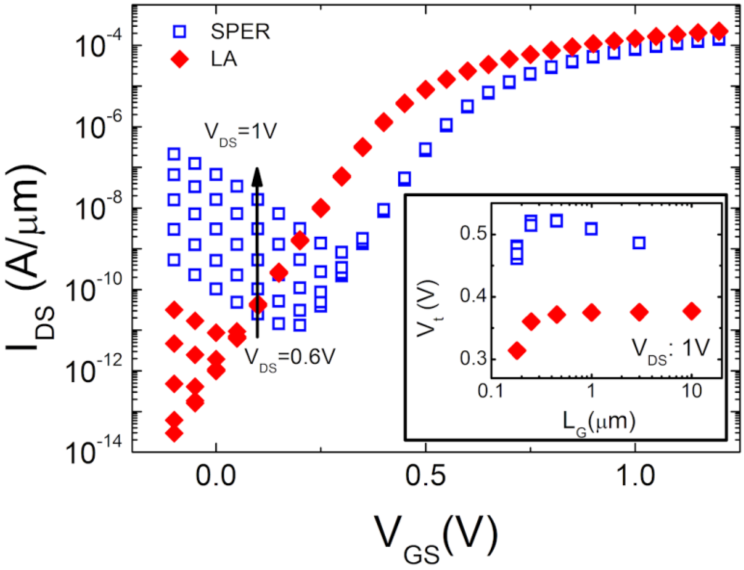


Figure 5.13: nMOS I_{DS}-V_{GS} characteristics for SPER and LA anneals. In the inset, the extracted V_t are plotted as a function of the gate length, L_G. (W_{fin} 250nm, L_G: 180nm)

In case of SPER, the leakage current in nMOSFETs is likely due to Gate-Induced Drain Leakage (GIDL) enhanced by the presence of silicide at the gate edge confirmed by TEM analysis (Figure 5.15) as previously reported in [17]. In Figure 5.14, a schematic shows how the presence of the silicide close to the gate edge is responsible for an enhanced GIDL component. The presence of silicide is a consequence of the nickel piping occurring during the silicidation process [18] enhanced by the presence of implantation damage not completely cured after the activation anneal. Contrarily to MOSFETs, the silicide-induced high electric field could be beneficial for TFETs.

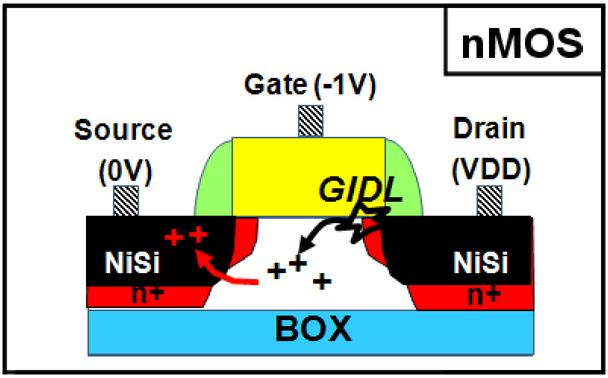


Figure 5.14: Schematic view of the enhanced GIDL for nMOS.

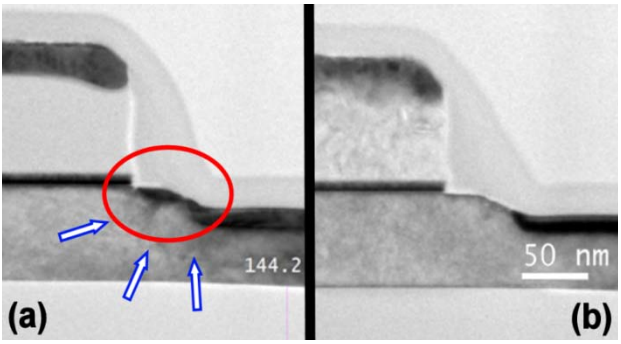


Figure 5.15: Cross-sectional TEM image of the n+ side along the channel for the SPER (a) and LA (b) in wide fins. (W_{fin} 250nm)

5.5.2 The Trend in TFET

Figure 5.16 shows the experimental I_{DS}-V_{GS} characteristics for the TFETs with a fin width of 250nm. For pTFETs, the characteristic with LA shows slightly better performance than in the case of spike anneal. The SPER split shows an anomalous behavior characterized by an increment of more than 3 decades of the drain current together with a large reduction of the tunneling onset voltage for both narrow and wide fins. The good reproducibility of the data is demonstrated by the cumulative probability plot of the device current in Figure 5.17. In particular, the tight distribution for the narrow fins (SPER split) suggests the formation of a homogenous silicide.

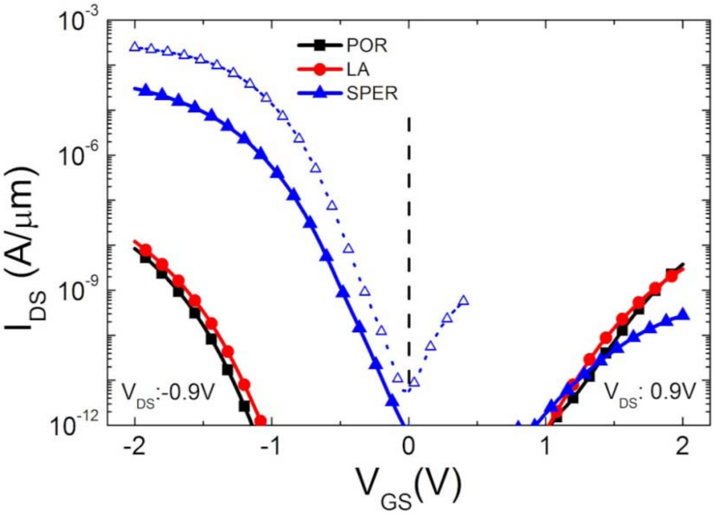


Figure 5.16: Transfer characteristics for pTFET and nTFET, extracted at $|V_{DS}|$ = 0.9V and for three different annealing processes. The device has W_{fin} 250nm (solid), L_G 150nm and N_{fin} 5. The device with the best performance (dashed) has W_{fin} 10nm.

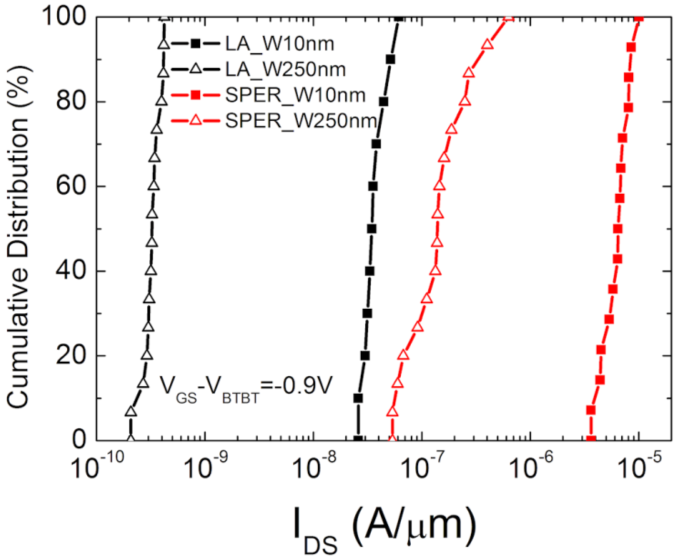


Figure 5.17: Cumulative distribution of the current extract at V_{DS} of -0.9V and same drive current for two different splits including narrow (W_{fin}: 10nm) and wide (W_{fin}: 250nm) fins. V_{BTBT} represents the lowest current recorded for the given drain voltage.

The anomalous behavior for the SPER split may not only be due to the position of the silicide close to the gate as observed in Figure 5.15 but also to the presence of end-of-range defects and dopant pile-up which can also contribute to boost the tunneling current [19]. To verify the presence of segregated dopants, SSRM measurements have been performed on dedicated structures. SSRM is able to characterize active dopants in two dimensions and, hence, it is preferred over SIMS analysis to obtain information on the lateral junction profile. The SSRM results show that a higher concentration of dopants is present in the proximity of the silicide along the channel direction (Figure 5.18).

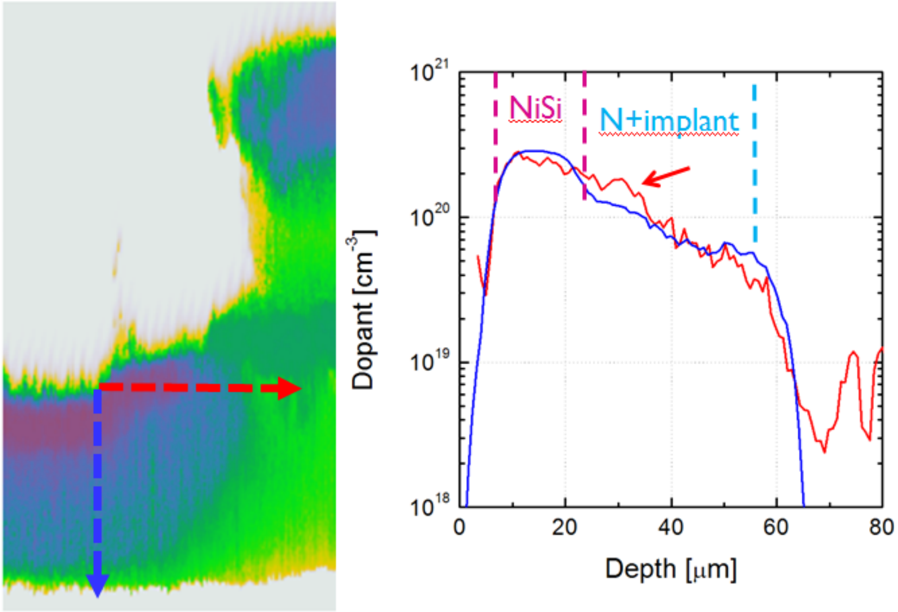


Figure 5.18: SSRM results and plot of the dopants concentration versus depth. Two curves correspond to two different directions: the red is along the channel and the blue is perpendicular to the substrate. Along the channel, a higher concentration of n+ dopants is recorded and indicated with a red arrow.

The presence of defects/traps at the source is confirmed by the I_{DS}-V_{GS} characteristics at different temperatures in Figure 5.19a where three different regions are identified. In Figure 5.19b, the extraction of an activation energy around 0.4V confirms the presence of TAT at low gate voltages (see chapter 4) which is responsible for the degradation of the subthreshold swing [19].

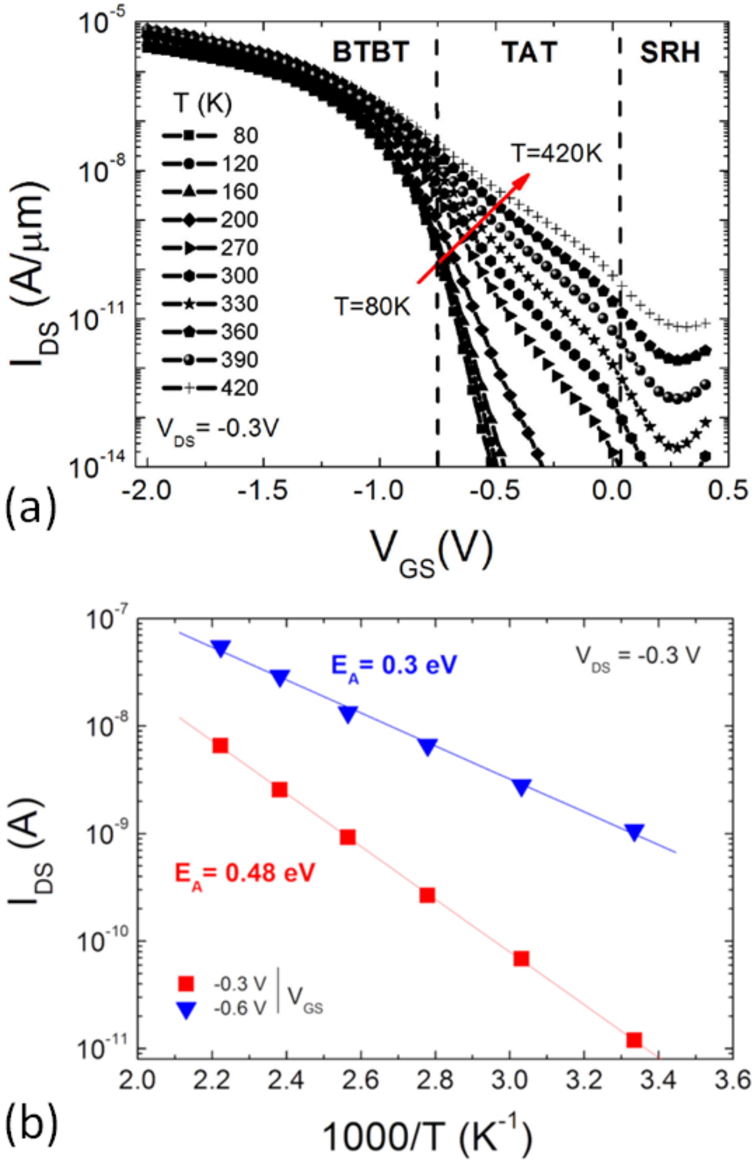


Figure 5.19: Transfer characteristics of experimental TFET for different temperatures (a) where three different transport mechanisms are identified: BTBT, TAT and SRH. Arrhenius plot extracted for two different gate voltages in the TAT region. (W_{fin}: 250nm)

Additional devices have been fabricated to investigate further the impact of the silicide close to the tunneling junction and the results are discussed in the next section. On the other hand, a similar current increase with SPER anneal is not observed in nTFETs where the tunneling occurs at the P-doped region. As a consequence, this behavior seems to be related to the implantation damage at the source (n+) side using As dopants with much higher amorphizing power than B atoms.

In conclusion, the effort to reduce the doping gradient has marginal effect except for SPER where a structural modification of the source region is responsible for the current increase. More details will be presented in the next section.

5.6 Silicide Engineering by Dopant Segregation

Silicide Engineering is not novel in CMOS integration. In the past years, there has been a large interest for Schottky barrier (SB) MOSFETs to replace conventional MOSFETs in advanced technology nodes with the promise for better control of the SCEs by shallower junctions. The fabrication of a SB MOSFET is similar to the conventional processing for MOSFET except for the absence of implantation. Unfortunately, the carrier injection is not sufficient to compete with the conventional MOSFET unless a very low Schottky barrier is implemented. The barrier height for electrons, ϕ_{Bn}, is material dependent and, in the ideal case, can be written as:

$$q\phi_{Bn0} = q(\phi_m - \chi) \tag{5.3}$$

where q is the electron charge, ϕ_m is the metal workfunction and χ is the electron affinity of the semiconductor.

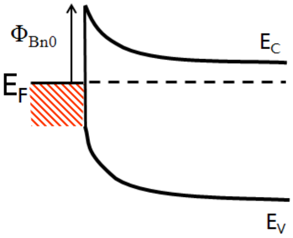


Figure 5.20: Schematic of the energy bands of a Schottky contact with n-type semiconductor.

The reduction of the Schottky barrier is achieved by modifying the work function of the silicide. In case of n-type, the ϕ_{Bn0} can be reduced from 0.47 eV (NiSi) to 0.28 eV ($ErSi_{1.7}$)[20-22]. Alternatively, dopant segregation (DS) can be implemented [23-24] to further lower $q\phi_{Bn}$. In case of NiSi, a reduction of the electron Schottky barrier to 0.074eV by forming silicide (NiSi) on n-Si using Antimony (Sb) was reported [25].

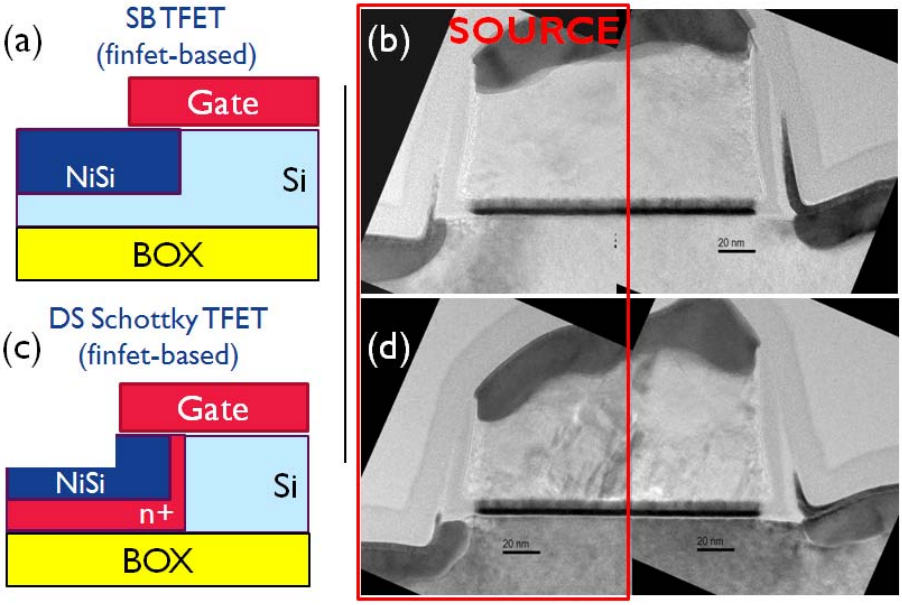


Figure 5.21: Schematics and TEM images of the fabricated devices, SBTFET (a,b) and DS Schottky TFET (c,d). The lateral silicide thickness is larger in case of the n+ layer. The silicide at the drain side is not encroaching. (W_{fin}: 500nm)

In case of TFETs, we need to increase the effective Schottky barrier to suppress the thermionic current which cannot result in a subthreshold swing below 60mV/dec at 300K. To understand how to fabricate and optimize such a device, two different devices are compared: the Schottky Barrier (SB) TFET and Dopant Segregated (DS) Schottky TFET. A schematic of the two devices together with a TEM images are shown in Figure 5.21. Starting from the understanding of the SB TFET in section 5.6.1, section 5.6.2 will explain how BTBT is possible in DS Schottky TFET. Contrary to the POR condition, the high-k thickness used is 3nm because we expect an increase of the gate leakage component due to the proximity of the silicide to the gate.

5.6.1 SB TFET

In a Schottky Barrier (SB) TFET [26], the source region is replaced by silicide while the drain is kept highly doped as in the conventional device. The main feature of this device is the presence of gate-controlled Schottky barrier tunneling at the source. Here, we study the p-type SB TFET which has a p^+ region at the drain and, consequently, we focus our attention on the hole electron barrier, ϕ_{Bp}, at the source side. In Figure 5.22, simulation results of the SB TFET structure, described in the inset, show the presence of three different regions where different transport mechanisms are dominant: Thermionic Emission (TE), Thermionic Field Emission (TFE) and Field Emission (FE) which is the tunneling through the Schottky barrier. The TE region is present at low gate bias and represents the current due to the carriers able to go over the Schottky barrier. As a consequence, a linear dependence with temperature was extracted for the subthreshold swing (kT/q limit). When the gate bias increases, the barrier becomes thinner and tunneling through the barrier starts. Initially, the current can be dominated by TAT or FN tunneling and this explains the temperature dependence in the TFE region. When the Fermi level of the metal crosses the valence band of the semiconductor, the tunneling through the barrier (FE) becomes dominant which results in a weak temperature dependence. Experimental results (Figure 5.23) are in strong agreement with these simulation results. Here, the improvement of the current for narrow fins is attributed to the better electrostatic control of the gate.

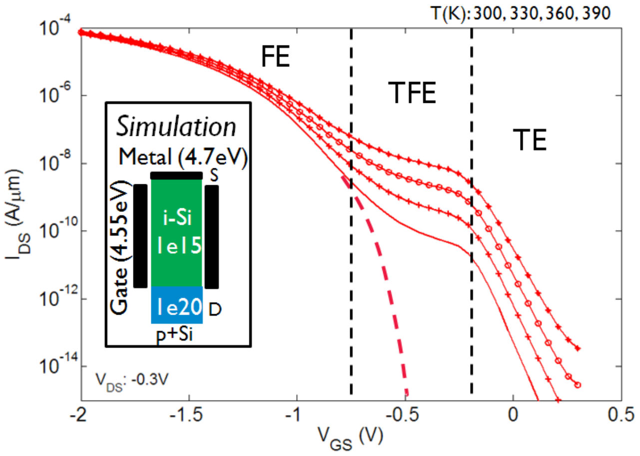


Figure 5.22: Simulation results of a SB TFET for different temperatures. In the inset, the schematic of the simulated device is shown. Three different regions are identified (TE, TFET and FE). The dotted line corresponds to the case in which only FE is present.

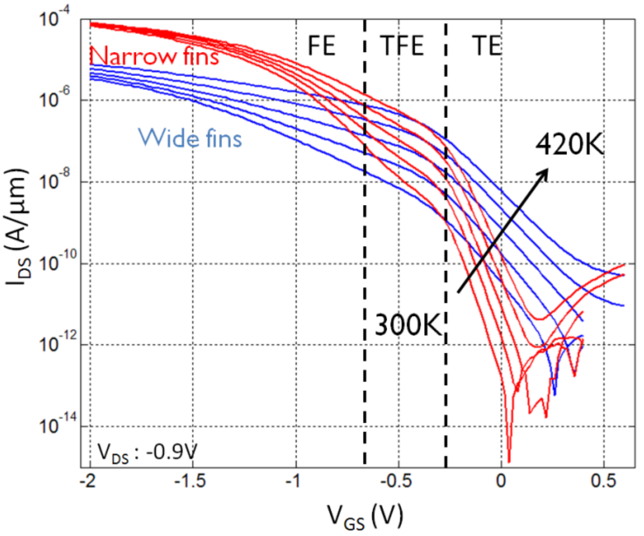


Figure 5.23: I_{DS}-V_{GS} plots of fabricated SB TFET at different temperatures. The results for narrow (W_{fin} 20nm) and wide fins (W_{fin} 500nm) are shown and closely match the simulation results. The presence of three different regions is indicated only for the narrow fins.

The subthreshold swing of the SB TFET can never be below 60mV/dec because of the presence of the TE component. The introduction of a n^+ pocket is designed to increase the hole Schottky barrier and suppress the TE component. In this way, BTBT might be the dominant transport mechanism at low gate bias as explained in the next section.

5.6.2 Dopant-Segregated (DS) Schottky TFET

The fabrication of DS Schottky TFET follows the standard process flow of the POR described at the beginning except for the recess of the source prior silicidation. The dopant segregation is obtained by increasing the thickness of the silicide (Figure 5.21c).

The formation of a highly doped layer next to the silicide induces a band bending in the region close to the Schottky junction, yielding a thinner tunneling barrier for electrons and increasing the effective hole Schottky barrier (Figure 5.24). For this reason, BTBT might be the dominant transport mechanism at low gate bias as observed in [27-28].

Figure 5.25 shows the transfer characteristics of the fabricated SB TFET and DS Schottky TFET. For comparison, a control device is also indicated as BTBT_ref. It is clear that the presence of the n^+ layer reduces the TE component. This phenomenon is even more visible when analyzing its dependence with the temperature in Figure 5.26. At low gate voltage, the linear dependence of the inverse slope cannot be extracted (Figure 5.26d). Nevertheless, a temperature dependence is still present, most likely due to TFE or TAT components. The barrier tuning was not sufficient to improve the subthreshold swing at low gate bias. Possible improvements can be obtained by improving the dopants segregation technique. For instance, An *et al.* [29] proposed to perform the implantation after the metal deposition before the silicidation process in order to push the dopants toward the silicide/semiconductor interface during the silicidation process. In this way, the detrimental effect of the implantation damage can be avoided.

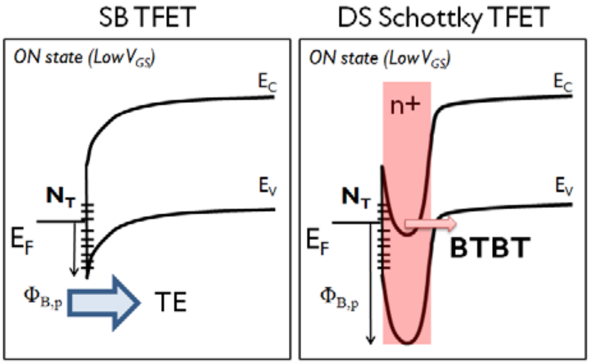


Figure 5.24: Schematic of the energy band diagram of SB TFET and DS Schottky TFET when low gate bias is applied.

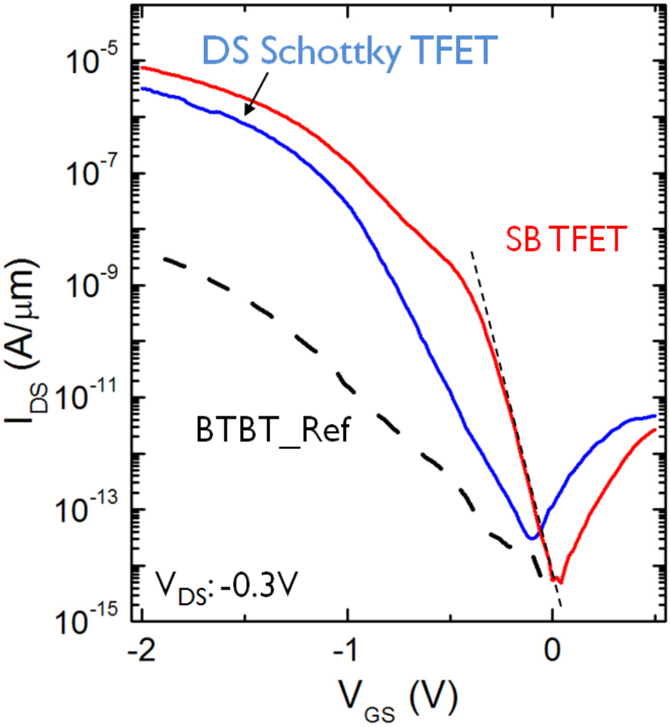


Figure 5.25: I_{DS}-V_{GS} characteristics of the fabricated SB TFET and DS Schottky TFET for narrow fin (W_{fin}: 20nm). The dotted line indicates the POR device with high-k thickness of 3nm. (L_G: 150nm)

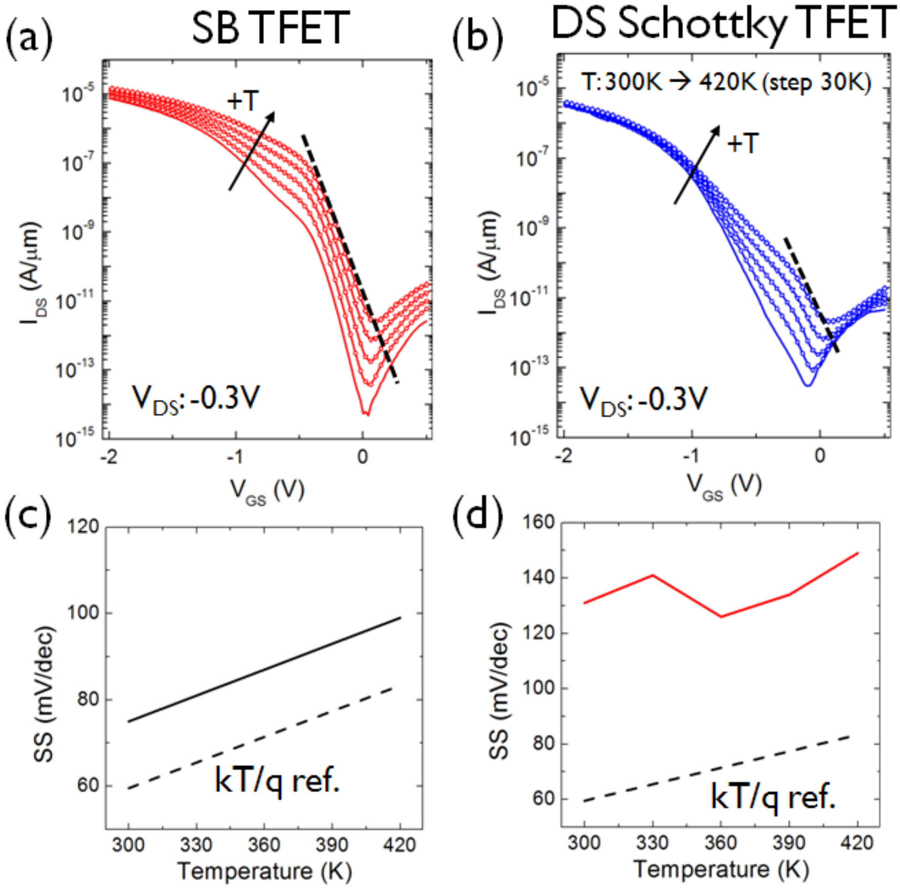


Figure 5.26: Transfer characteristics and extract of the inverse slope at low gate voltage for SB TFET (a,c) and DS Schottky TFET (b,d). (W_{fin}: 20nm)

On the other hand, DS Schottky TFETs have several limitations which undermine further optimization of this device:

- the tuning of the barrier requires the introduction of different silicides for n- and p-type operations which increases the process complexity;
- the large hole Schottky barrier (for p-type operation) might become a bottleneck for on-current requirements;
- semiconductor materials with large bandgap are required to minimize the thermionic emission; hence, the BTBT cannot be improved;

Alternative implementations of a Schottky barrier to boost the TFET on-current are investigated in literature. Lattanzio *et al.* proposed to add a thin oxide layer in-between the metal and the semiconductor in the source region. In their simulations, the device shows the sub-60mV/dec subthreshold swing at low gate bias because of the induced BTBT while large on-current is observed at larger gate voltages as expected for SB FET. An alternative implementation was also suggested in [30], where a SB FET and a TFET were fabricated in parallel using a self aligned process. All these solutions have however issues regarding the complexity of the fabrication, the scalability and the reproducibility. For this reason, this approach is not further developed in this work.

5.7 Performance of TFET

Figure 5.27 shows the transfer characteristics of some of the fabricated Si TFETs compared with a conventional MOSFET fabricated on the same layout. It is clear that none of the fabricated TFETs outperform the MOSFET. In particular, conventional Si TFET cannot even reach 1µA/µm on current at an effective V_{DD} of 0.9V. The SPER split looks promising but the scaling of the supply voltage is limited by the presence of TAT at low gate voltages.

In conclusion, the TFET needs additional performance boost, for instance by changing the source material and replacing it with a lower bandgap semiconductor as explained in the next chapter where the heterojunction TFETs are introduced.

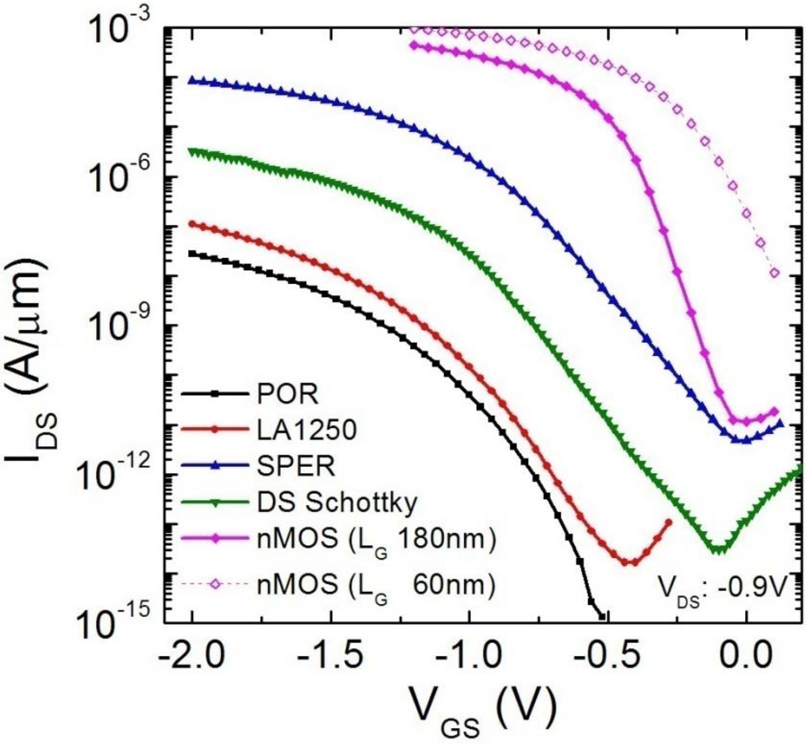


Figure 5.27: Transfer characteristics of Si pTFET fabricated in this work compared with conventional pMOS. In case of pMOS, the curve of a short channel device is added to show the impact of SCEs. (W_{fin}: 20nm, L_G: 150nm).

5.8 Summary

In this chapter, the impact of several geometrical and processing parameters of finFET-based Si TFETs has been analyzed in detail. An improvement of the on-current and subthreshold swing has been observed by reducing the gate oxide thickness and the fin width. As expected, the better gate coupling is important to optimize this device. The doping engineering is, however, more challenging when ion implantation techniques are used because of the large implantation damage, the high temperature step necessary to activate the dopants and the difficulty to obtain an abrupt doping profile.

The strong temperature dependence of the current with the temperature at low gate voltage suggests that transport mechanisms other than BTBT are dominant for both narrow and wide fins (see chapter 4). In narrow fins, the degradation is less visible due to a lower onset of the BTBT component due to the better gate coupling.

The implementation of dopant-segregated SB TFET was explored to boost the on-current. An improvement of the on-current of about 4 decades was achieved but BTBT was still not dominant at low gate voltages and, hence, the subthreshold swing resulted degraded. Further development of this approach was abandoned because the engineering of the Schottky barrier height is strictly material dependent and the difficulty to efficiently suppress the temperature dependent mechanisms at low voltage bias.

None of the fabricated TFETs achieved the specified targets because of the large energy bandgap of silicon. In order to improve the tunneling performance, a low energy bandgap material such as germanium for p-type or III/V for n-type TFETs needs to be used [31-32].

References

[1] D. Hisamoto, *et al.*, "FinFET-a self-aligned double-gate MOSFET scalable to 20 nm," *IEEE Transactions on Electron Devices,* vol. 47, pp. 2320-2325, 2000.

[2] D. Leonelli, *et al.*, "Performance Enhancement in Multi Gate Tunneling Field Effect Transistors by Scaling the Fin-Width," *Japanese Journal of Applied Physics,* vol. 49, p. 04DC10, 2010.

[3] R. H. Yan, *et al.*, "Scaling the Si MOSFET: from bulk to SOI to bulk," *IEEE Transactions on Electron Devices,* vol. 39, pp. 1704-1710, 1992.

[4] J. Knoch and J. Appenzeller, "A novel concept for field-effect transistors - the tunneling carbon nanotube FET," in *Device Research Conference Digest,* 2005, pp. 153-156.

[5] J. Knoch, *et al.*, "Impact of the dimensionality on the performance of tunneling FETs: Bulk versus one-dimensional devices," *Solid-State Electronics,* vol. 51, pp. 572-578, 2007.

[6] K. Boucart and A. M. Ionescu, "Double-gate tunnel FET with high-kappa gate dielectric," *IEEE Transactions on Electron Devices,* vol. 54, pp. 1725-1733, Jul 2007.

[7] A. S. Verhulst, *et al.*, "Modeling the single-gate, double-gate, and gate-all-around tunnel field-effect transistor," *Journal of Applied Physics,* vol. 107, p. 024518, 2010.

[8] *Synopsys Sentaurus Package version C (June 2009).*

[9] H.-W. You and W.-J. Cho, "Charge trapping properties of the HfO[sub 2] layer with various thicknesses for charge trap flash memory applications," *Applied Physics Letters,* vol. 96, pp. 093506-3, 2010.

[10] E.-H. Toh, *et al.*, "Device physics and design of germanium tunneling field-effect transistor with source and drain engineering for low power and high performance applications," *Journal of Applied Physics,* vol. 103, p. 104504, 2008.

[11] K. Iwamoto, *et al.*, "Experimental evidence for the flatband voltage shift of high-k metal-oxide-semiconductor devices due to the dipole formation at the high-k/SiO_2 interface," *Applied Physics Letters,* vol. 92, pp. 132907-3, 2008.

[12] T. Noda, *et al.*, "Analysis of dopant diffusion and defect evolution during sub-millisecond non-melt laser annealing based

on an atomistic kinetic Monte Carlo approach," in *IEEE International Electron Devices Meeting,* San Francisco, CA,, 2006, p. 4

[13] C. Sandow*, et al.*, "Improving the performance of band-to-band tunneling transistors by tuning the gate oxide and the dopant concentration," in *Device Research Conference Digest*, 2008, pp. 79-80.

[14] W. Vandenberghe*, et al.*, "Analytical model for point and line tunneling in a tunnel field-effect transistor," in *International Conference on Simulation of Semiconductor Processes and Devices*, 2008, pp. 137-140.

[15] L. Capello*, et al.*, "Solid-phase epitaxial regrowth of a shallow amorphised Si layer studied by X-ray and medium energy ion scattering," *Materials Science and Engineering: B,* vol. 124-125, pp. 200-204, 2005.

[16] K. Funk*, et al.*, "Maximizing Boron Activation in Solid Phase Epitaxy A Case of Implant Choice and RTP Processing," in *International Workshop on Junction Technology*, 2006, pp. 40-43.

[17] T. Hoffmann*, et al.*, "GIDL (Gate-Induced Drain Leakage) and parasitic schottky barrier leakage elimination in aggressively scaled HfO_2/TiN FinFET devices," *IEEE International Electron Devices Meeting* pp. 743-746, 2005.

[18] A. Lauwers*, et al.*, "Ni based silicides for 45 nm CMOS and beyond," *Materials Science and Engineering B-Solid State Materials for Advanced Technology,* vol. 114, pp. 29-41, Dec 2004.

[19] D. Leonelli*, et al.*, "Silicide Engineering to Boost Si Tunnel Transistor Drive Current," *Japanese Journal of Applied Physics,* vol. in press, 2011.

[20] M. Jang*, et al.*, "A 50-nm-gate-length erbium-silicided n-type Schottky barrier metal-oxide-semiconductor field-effect transistor," *Applied Physics Letters,* vol. 84, pp. 741-743, 2004.

[21] A. Cavallini*, et al.*, "Deep levels in Er-doped liquid phase epitaxy grown silicon," *Applied Physics Letters,* vol. 72, pp. 468-470, 1998.

[22] A. Cavallini, *et al.*, "Electrical and optical characterization of Er-doped silicon grown by liquid phase epitaxy," *Journal of Applied Physics*, vol. 85, pp. 1582-1586, 1999.

[23] T. Yi, *et al.*, "Schottky barrier tuning at NiSi/Si interface using pre-silicide aluminum and sulfur co-implant," in *the 10th IEEE International Conference on Solid-State and Integrated Circuit Technology*, 2010, pp. 1021-1023.

[24] A. Kinoshita, *et al.*, "High-performance 50-nm-gate-length Schottky-source/drain MOSFETs with dopant-segregation junctions," in *Symposium on VLSI Technology*, 2005, pp. 158-159.

[25] W. Hoong-Shing, *et al.*, "Sub-0.1-eV Effective Schottky-Barrier Height for NiSi on n-Type Si (100) Using Antimony Segregation," *IEEE Electron Device Letters*, vol. 28, pp. 703-705, 2007.

[26] R. Jhaveri, *et al.*, "Asymmetric Schottky Tunneling Source SOI MOSFET Design for Mixed-Mode Applications," *IEEE Transactions on Electron Devices*, vol. 56, pp. 93-99, 2009.

[27] K. Jeon, *et al.*, "Si tunnel transistors with a novel silicided source and 46mV/dec swing," in *Symposium on VLSI Technology*, 2010.

[28] R. Gandhi, *et al.*, "CMOS-Compatible Vertical-Silicon-Nanowire Gate-All-Around p-Type Tunneling FETs With <= 50-mV/decade Subthreshold Swing," *IEEE Electron Device Letters*, vol. 32, pp. 1504-1506, 2011.

[29] X. An, *et al.*, "The modulation of Schottky barrier height of NiSi/n-Si Schottky diodes by silicide as diffusion source technique," *Chinese Physics B*, vol. 18, pp. 4465-4469, Oct 2009.

[30] H. Qianqian, *et al.*, "Self-depleted T-gate Schottky barrier tunneling FET with low average subthreshold slope and high I_{ON}/I_{OFF} by gate configuration and barrier modulation," in *IEEE International Electron Devices Meeting* 2011, pp. 16.2.1-16.2.4.

[31] S. Mookerjea, *et al.*, "Experimental Demonstration of 100nm Channel Length $In_{0.53}Ga_{0.47}As$-based Vertical Inter-band Tunnel Field Effect Transistors (TFETs) for Ultra Low-Power Logic and SRAM Applications," in *IEEE International Electron Devices Meeting*, San Francisco, CA, 2009, pp. 949-951.

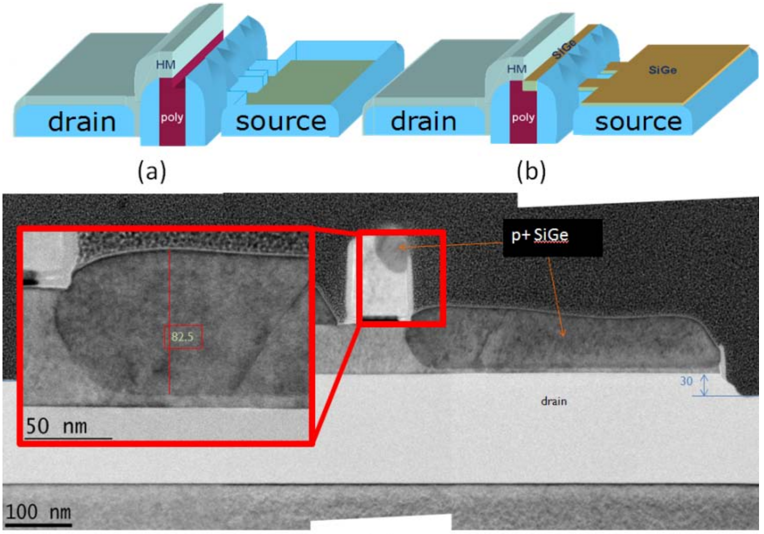
[32] S. Mookerjea, *et al.*, "Temperature-dependent I-V Characteristics of a Vertical $In_{0.53}$/$Ga_{0.47}$/As Tunnel FET," *IEEE Electron Device Letters*, vol. 31, pp. 564-566, 2010.

Chapter 6

Vertical Nanowire-based TFETs for Heterojunction Integration

6.1 Introduction

As introduced in the previous chapter, the integration of heterojunction TFET is required to boost the tunneling current. Therefore, the integration of heterojunction using SiGe source was explored in Fin-TFET architectures where a Si etch-back followed by a SEG of B-doped $Si_{0.75}Ge_{0.25}$ is performed at the source while protecting the drain region using an oxide HM (Figure 5.29 a and b). As illustrated in Figure 6.1, a good regrowth of the B-doped SiGe source was achieved after optimization of the cleaning conditions.



(a) (b)

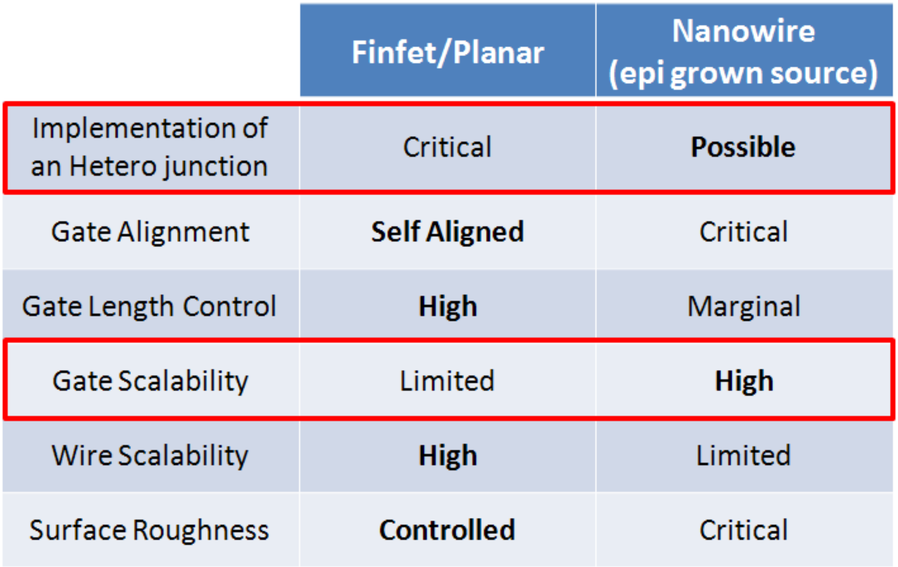
(c)

Figure 6.1: Schematic of the Si etch back (a) and SiGe regrowth (b) used to implement hetero junction TFET on SOI. TEM image shows a good quality of the SiGe layer (c). Unfortunately, no overlap with the gate is observed.

However, as soon as the SiGe source features an overlap with the gate, no grow or uncontrolled growth was observed. The reason is that the HF pre-clean prior to SEG attacks the gate oxide and exposes the TiN which affects the SEG process. In addition, the gate oxide removal would lead to TiN gate and source shorts resulting in device failure. The only alternative to avoid the pre-clean issue is to perform an in-situ etch back of the source and SEG of the SiGe material. This solution was however not pursued because considered as very challenging and the heterojunction integration was focused on a vertical TFET configuration. In addition, the performances of Si TFETs are not competitive with conventional MOSFETs. In particular, efforts to boost the tunneling current by optimization of the process and geometrical parameters are not effective enough. As a consequence, it was suggested [1] that the implementation of heterojunction TFETs can strongly boost the tunneling current while maintaining a low off-current. The improvement of the tunneling current is due to the lowering of the effective energy bandgap between the source and the channel as explained in chapter 2.

In this chapter the implementation of heterojunction TFETs in a vertical device configuration is investigated. The reason behind the choice of vertical nanowires as a new test vehicle for TFET is multifold. Firstly, the vertical architecture offers the advantage to implement a hetero tunneling junction by epitaxial growth of the two different materials needed in the source and the channel. Secondly, a gate-all-around configuration provides the best electrostatic control of the gate on the channel.

The vertical nanowires TFETs, or GAA TFETs, are fabricated using state-of-the-art standard CMOS processing steps, described in Chapter 4. Table 6.1 summarizes the main differences between the vertical and the planar processes. The implementation of a heterojunction is the main advantage of the nanowire together with its better scalability since the gate position is vertical. On the other hand, the gate alignment to the tunnel junction and the wire scalability might be critical because the process flow is not gate self-aligned and because of lithography limitations for nanowire patterning and the nanowire etching with high aspect ratio. In addition, etch-induced surface roughness is also critical because of the deep patterning of the 450nm tall nanowires, much taller than what is used for fins. Contrary to finfet-based TFETs, the fabrication of the vertical devices focuses on nTFETs instead of pTFETs since the targeted source material is $Si_{1-x}Ge_x$ with different Ge concentration (up to 100%). The use of III/V materials for the source region of pTFETs is not mature enough at imec, especially on full wafers, and is not considered in this work.



	Finfet/Planar	Nanowire (epi grown source)
Implementation of an Hetero junction	Critical	**Possible**
Gate Alignment	**Self Aligned**	Critical
Gate Length Control	**High**	Marginal
Gate Scalability	Limited	**High**
Wire Scalability	**High**	Limited
Surface Roughness	**Controlled**	Critical

Table 6.1: Advantages and challenges to implement lateral TFET with two different archtectures: finfet and nanowire. The red squares indicate the key features able to make the TFET competitive with conventional MOSFET.

This chapter is mainly divided in three parts. In the first part, the Si TFETs are presented as a reference device where the impact of some geometrical and process parameters is analyzed. Later, the heterojunction TFETs are introduced focusing on tuning of the Ge concentration at the source to boost the tunneling current. At the end, the performances of vertical TFETs are summarized with emphasis on the mechanisms responsible for the degradation of the subthreshold swing such as TAT.

6.2 Vertical-based Si TFET

The Si TFET is used as a reference to benchmark the performance improvement obtained with a heterojunction. Figure 6.2(a-c) shows a schematic and TEM images of a vertical device with some typical dimensions labeled. The nanowire height is around 400nm. This nanowire height is necessary in order to allow: (1) the presence of a bottom isolation (2), a top isolation and (3) a gate length around 150nm.

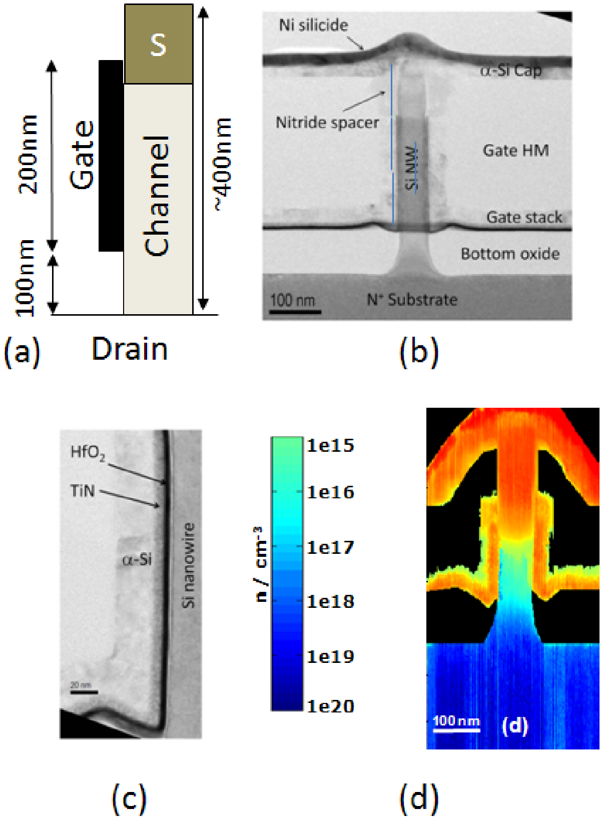


Figure 6.2: Schematic of a fabricated GAA TFET (a), TEM image of a cross section of a 40nm-wide Nanowire (b) and enlargement of the gate stack (c). SSRM image with the doping quantification in case of implanted source (d).

The bottom isolation is used to (1) suppress the ambipolar behavior in the device already observed in the finfet structure (Chapter 5) and (2) to reduce the gate to drain leakage current and parasitic capacitance. The thickness of the bottom isolation needs to be optimized taking into account the diffusion of the dopants from the substrate into the intrinsic region during the epitaxial growth of the diode regions and during the device processing as shown by the SSRM measurements in Figure 6.2d. On the other hand, the top isolation avoids shorts between the gate and the source. It consists of nitride spacers and an oxide layer as described in chapter 3. In addition, the amount of gate/source overlap can be adjusted during the gate HM etching.

Typical transfer and output characteristics of vertical Si nTFETs are shown in Figure 6.3. In this case, the source is either implanted or epitaxially grown. The devices fabricated with the wireke maskset are measured using the nanoprober setup described in chapter 4. The measurements are performed with the p^+ source (top contact) grounded and a reverse bias applied to the n^+ drain (substrate). The device current is measured at the source and normalized according to the total perimeter:

$$\mathrm{W_{tot} = N_{NW}\pi D_{NW}} \tag{6.1}$$

where N_{NW} is the number of nanowires and D_{NW} is the nanowires diameter. This approach is valid for narrow nanowires because of their circular section while the section of nanowires with a diameter exceeding 100 nm is closer to a square shape. In this case, the normalization slightly overestimates the tunneling current for large nanowires. A very low I_{OFF} of 0.3 pA/µm is obtained with an I_{ON}/I_{OFF} ratio ~10^4, for a supply voltage of 1.5 V. The subthreshold slope reported here (average ~300 mV/dec) is in line with other published data [2-3]. A point subthreshold slope lower than 60 mV/dec, typically measured at very low gate voltage and low drain bias [4-7] could not be measured due to the non-negligible gate leakage relative to the tunneling current in this bias range. Both device characteristics, i.e. the one obtained using the grown source integration flow or the one using an implanted source, are very similar. The reason for this is that a spike anneal is performed in both samples to activate the dopants implanted in the capping layer. As a consequence, the abruptness of the tunneling junction in the grown source device is not much improved compared to the implanted source device and the electrical performance is equivalent in both devices.

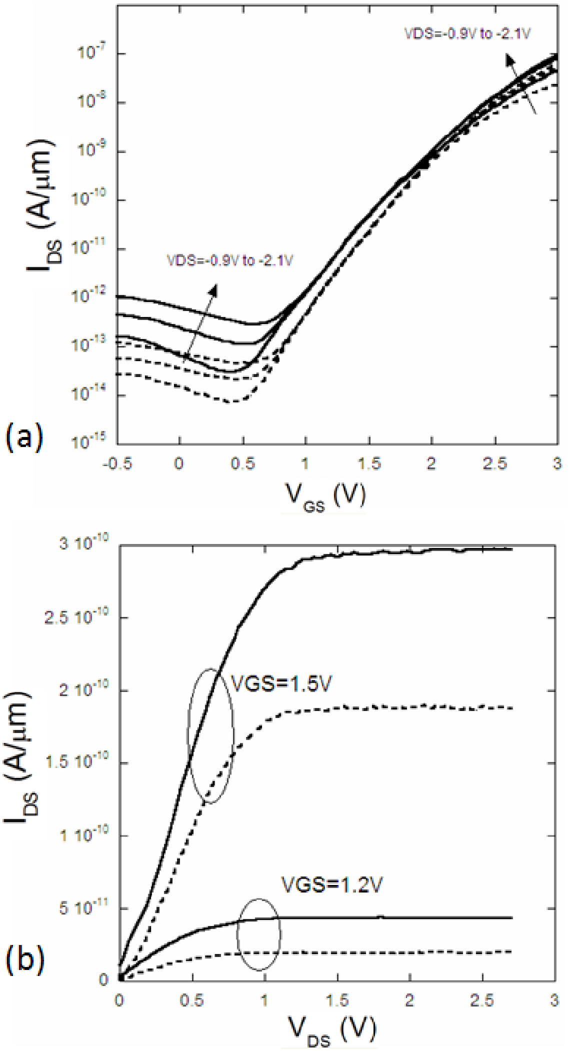


Figure 6.3: Transfer (a) and output (b) characteristics of a vertical SiNW n-TFET with implanted (solid lines) and epitaxially grown (dashed lines) P+ source for different drain biases. The output characteristics show a perfect saturation of the TFET device and the saturation voltage depends on the gate voltage similar to a MOSFET. (NW_D 40nm, L_G 150nm, $L_{G/S}$ 20nm and N_{NW} 50)

6.2.1 Impact of the Nanowire Diameter

The nanowire diameter ranges from 40nm to 200nm in the vertical TFETs. As for the fin width dependence in fin-based TFETs (chapter 5), a nanowire diameter dependence for the narrow nanowires is expected [8-10]. Using a similar approach as in [9], the natural length which describes the electrostatic integrity of the gate-all-around devices can be expressed by equation (6.2):

$$\lambda = \sqrt{\frac{2\,\epsilon_{\mathrm{si}} t_{Si}^{2}\, ln\left(1+\frac{2\mathrm{t_{ox}}}{\mathrm{t_{Si}}}\right)+\epsilon_{\mathrm{ox}} t_{Si}^{2}}{16\epsilon_{\mathrm{ox}}}} \tag{6.2}$$

In this equation, the silicon thickness, t_{Si}, represents the nanowire diameter, and has a stronger impact on the natural length compared to the double gate configuration presented in the previous chapter. This statement is also confirmed by the analytical modeling presented in [10].

Two sets of data are analyzed: implanted source and epi-doped source. In case of implanted source, the source is self-aligned to the gate (as described in Chapter 4). Therefore, the overlap is mainly controlled by the implantation energy and does not vary much with the nanowire size [11]. In fact, the overlap between the gate and the source doping profile in similar for different nanowire diameters as illustrated in SSRM measurements in Figure 6.4. The transfer characteristics of devices for different nanowire diameters are illustrated in Figure 6.5. A clear improvement of the tunneling current is observed for narrow nanowires. Unfortunately, the number of measured devices is limited due to nanoprobing and not sufficient to extract a statistical trend as it was done for the fin-based TFETs. When the diameter exceeds 100 nm, the current shows no dependence and the advantage of the GAA architecture is lost.

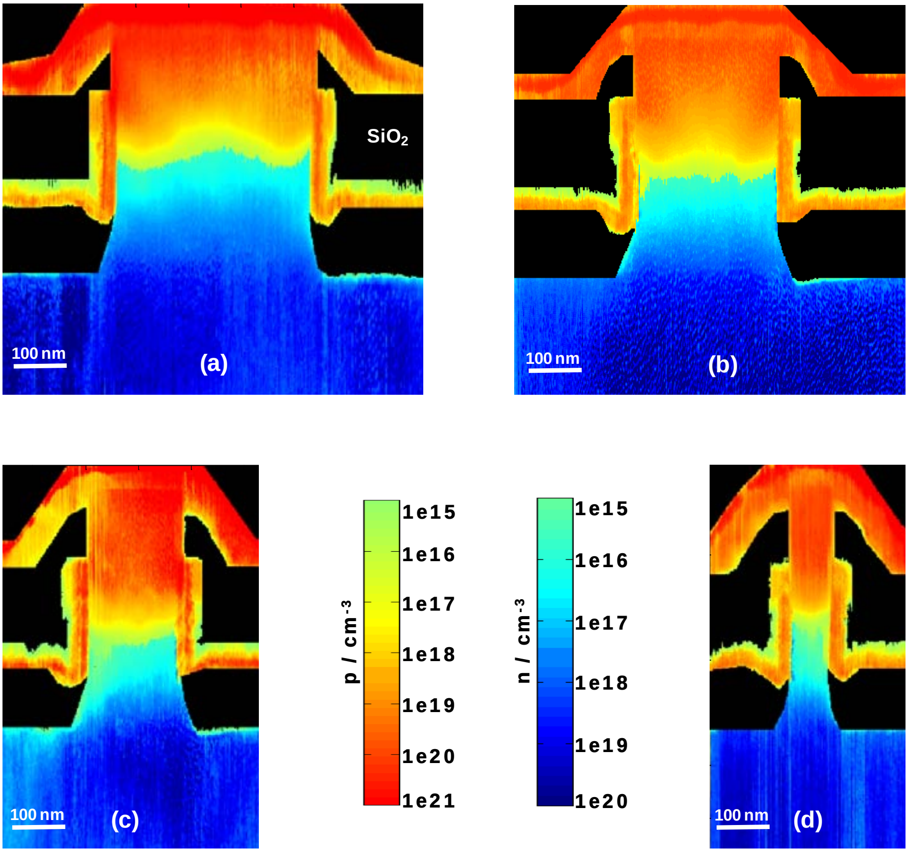


Figure 6.4: SSRM measurements of nanowires with different diameters where the source was implanted. The overlap between the gate and the source doping profile is very similar for the different nanowire diameters.

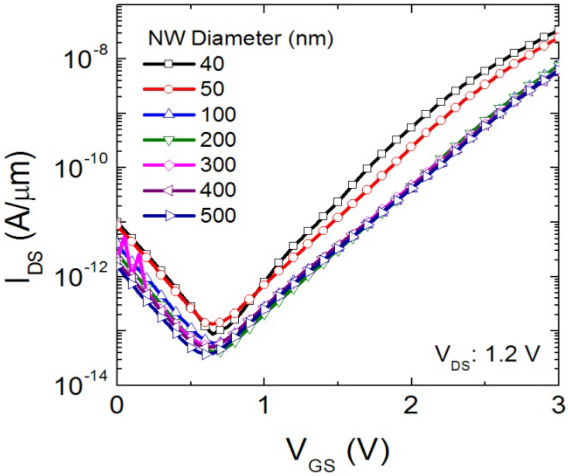


Figure 6.5: Input characteristics of vertical nTFETs featuring different nanowire diameters in the case of the implanted-source device. In this plot, the device features a 12nm $L_{G/D}$.

In case of devices with epi-doped source, the gate is not self-aligned with the source and the overlap at the source can inherently vary with processing resulting in more ambiguous results. For instance, the input characteristics with different nanowire diameters of epi-doped source TFETs are illustrated in Figure 6.6 for two different processes for which isolation layers were formed either by CMP and oxide etchback (process A: CMP-based) or directly by wet etch-back (process B: CMP-free). The trend with the nanowire size is clearly opposite in the two process flows. This phenomenon can be explained by the TEM images in Figure 6.7, comparing wide and narrow nanowires for process B. The narrow wires clearly exhibit a smaller gate-source overlap than the large wires, resulting in the performance degradation as shown in the section 6.2.2. For process A, the gate overlap is larger for the small nanowires, and the trend is reversed. Depending on the processing, an “artificial” nanowire size dependence can be induced due to modification of other diameter-dependent device parameters, such as the gate-source overlap.

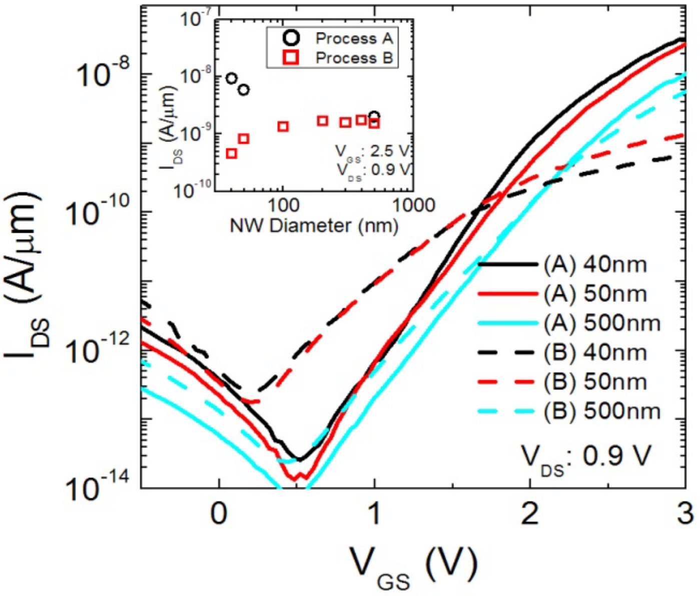


Figure 6.6: Input characteristics of vertical nTFETs (epi-doped source) featuring different nanowire diameters for processes A and B.

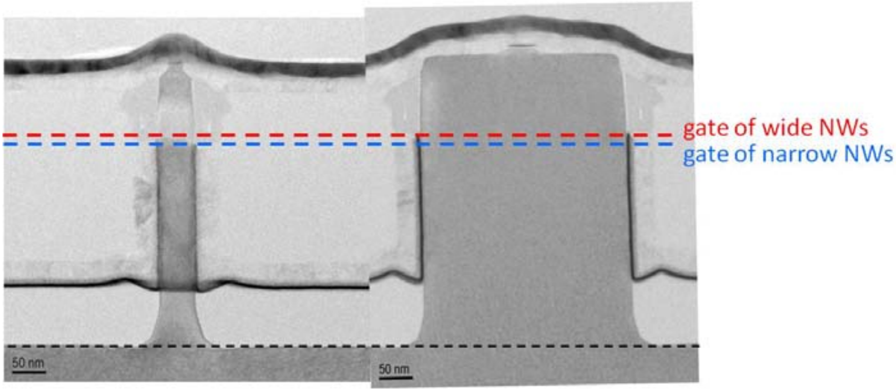


Figure 6.7: TEM cross-section images of the vertical TFETs comparing two different nanowire diameters (300 nm and 50 nm) for process B.

6.2.2 Impact of the Gate-Source & Gate-Drain overlaps

The impact of gate-source overlap has been studied by varying intentionally the thickness of the gate hardmask. In this case, epi-doped source TFETs are used and the analysis is carried out for wide fins to avoid data interpretation issues illustrated in the previous subsection. The gate-source overlap, $L_{G/S}$, was calculated based on thickness measurement by ellipsometry of the bottom oxide, gate stack and gate hardmask. This calculation does not include the amount of boron diffusion due to the thermal budget during processing. The "effective" gate-source overlap is therefore larger than the one calculated, but the relative difference in gate-source overlap between wafers remains, as all wafers received identical processing conditions.

In Figure 6.8a, the input characteristics show a strong degradation of the nTFET on-state current when decreasing the gate-source overlap. This effect can be intuitively understood by considering the doping profile at the source obtained from SIMS analysis (Figure 6.8b). As the gate position is moved away from the source, increasingly lower doping levels are controlled by the gate and, consequently, a degradation of the on-state current is observed due to a lower tunneling probability.

Figure 6.8a also shows the effective suppression of the ambipolarity by increasing the underlap between the gate and the drain, defined by the thickness of the bottom isolation. The drain-gate underlap, $L_{G/D}$, is calculated in a similar way as for the gate-source overlap and does not take into account the up-diffusion of arsenic dopants from the substrate during processing. The optimum bottom oxide thickness depends on the amount of As up diffusion and, therefore, the thermal budget during processing.

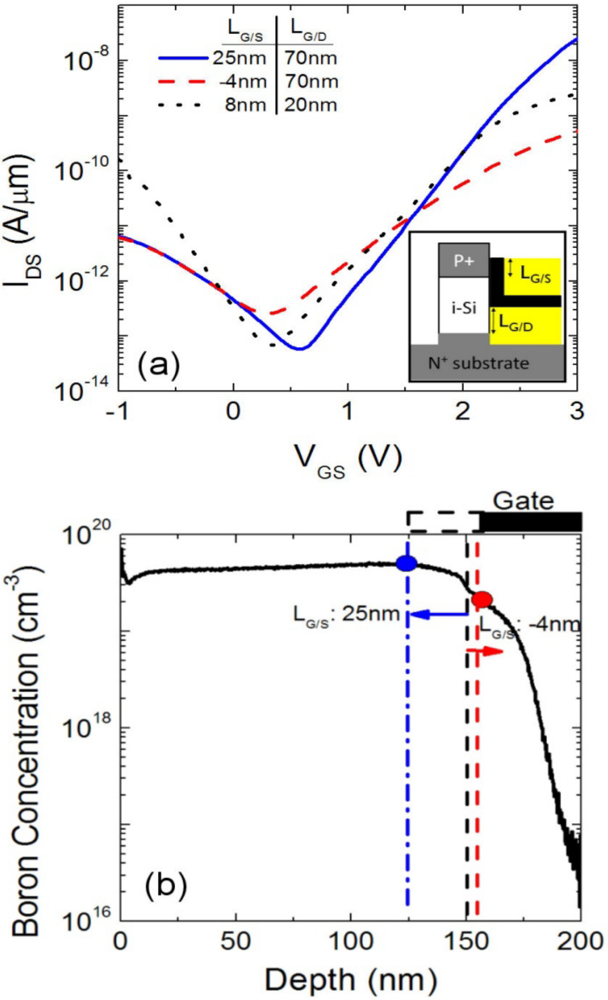


Figure 6.8: Transfer characteristics of vertical TFETs characterized by a different gate-source overlaps and gate-drain underlaps (V_{DS}=1.2V) (a). The nanowire has a diameter of 500nm and the source is doped during SEG. Boron profile at the top source junction obtained from SIMS analysis (b). The markers indicate the expected position of the gate edge for two measured devices: 4nm G/S underlap, 25nm G/S overlap. The black dotted line is used as a reference to indicate the position of the junction after epitaxial growth.

6.3 Heterojunction Vertical nTFET

6.3.1 Introduction

Germanium growth on Si is hampered by the large lattice mismatch (4%) between Ge and Si which results in growth that is dominated by misfit dislocations or non uniform layers. The growth of a good quality strained $Si_{1-x}Ge_x$ layer on Si is limited by its critical thickness (Figure 6.9) which decreases when the Ge content increases [12-13]. For this reason, $Si_{1-x}Ge_x$ source (with different Ge concentration) has been initially implemented to assess the impact of the reduced energy bandgap of the source on the tunneling current [14].

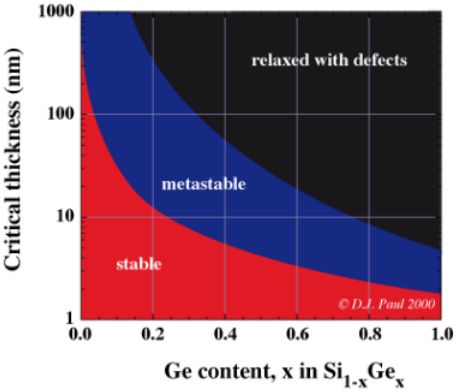


Figure 6.9: Critical Thickness versus Ge content for $Si_{1-x}Ge_x$ grown on bulk Si [15].

6.3.2 Impact of the Ge Concentration and δ-layer

The reduction of the critical thickness of the SiGe layer with an increasing Ge content might limit the processing window since the thickness of the source layer should be at least 50nm-thick to assure a good source isolation. For this reason, we implement a p-doped $Si_{0.73}Ge_{0.27}$ layer on top of a strained SiGe doped δ-layer with higher Ge content. In fact, the introduction of a doped δ-layer between the source and the Silicon channel is meant to provide a higher Ge content at the tunneling junction and lead to the pinning of the Fermi level at the source edge as in [16-17]. For instance, a δ-layer 5nm thick allows to have a stable phase of SiGe with a Ge content up to 45%.

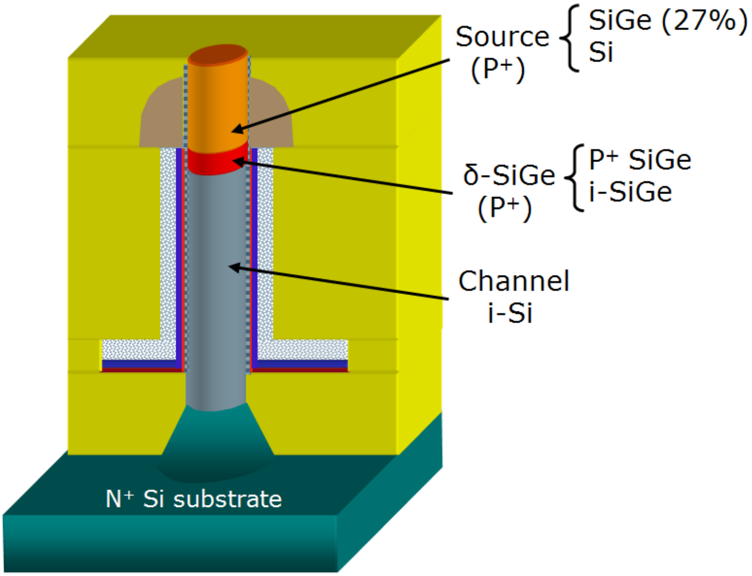
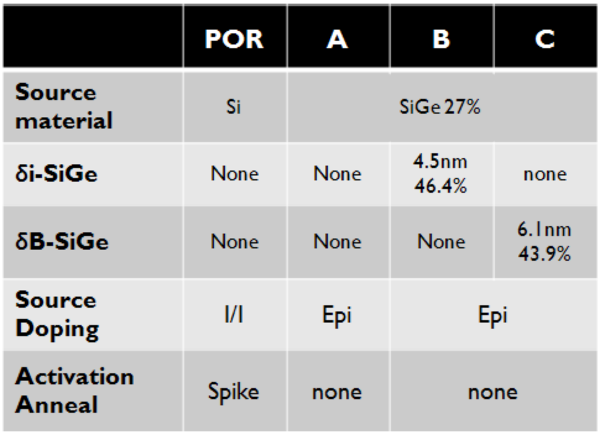


Figure 6.10: Schematic of the fabricated Si and SiGe/Si TFETs. In case of homojunction the δ-SiGe is not present.

Figure 6.10 shows the schematic of the fabricated devices while additional details about the splitmatrix are summarized in Table 6.2. Si TFET is used as reference to highlight the improvement obtained with SiGe source TFETs where three different SiGe splits are designed to investigate the impact of the δ-layer. In particular, the case without δ-layer and the case when this layer is highly doped (δB-SiGe) or intrinsic (δi-SiGe) are investigated. For the fabricated devices, an overlap $L_{G/S}$ of 30nm is implemented to reduce the variability in case of marginal gate to source overlap shown in section 6.2.2. The ambipolar behavior is reduced by defining a bottom isolation of 150nm. The physical gate length is around 150nm including the overlap with the source.



	POR	**A**	**B**	**C**
Source material	Si	SiGe 27%		
δi-SiGe	None	None	4.5nm 46.4%	none
δB-SiGe	None	None	None	6.1nm 43.9%
Source Doping	I/I	Epi	Epi	
Activation Anneal	Spike	none	none	

Table 6.2: Splitmatrix for vertical nanowire nTFETs. Thickness and Ge contents reported are measured by XRD measurements.

In Figure 6.11, the SIMS measurements show: (1) the presence of the δ-layer as a bump for the germanium contents (splits B and C) and (2) the limited diffusion of the boron in the silicon channel. The maximum concentration of boron is not the active concentration and additional micro four-point-probe measurements indicate that only 1 x 10^{20}at/cm^3 is the concentration of the electrically activated dopants. For the underlying layers, this characterization is not possible and we assume that the Boron concentration below 1 x 10^{20}at/cm^3 is fully active since the doping was performed in-situ during the epitaxial growth. After this assumption, we can observe that the doping level is extremely high in the SiGe region for all the splits indicating the i-SiGe layer is also highly doped.

The transfer characteristics for the vertical nTFETs are displayed in Figure 6.12. The impact of the SiGe source is clear: the drive current improved together with the subthreshold swing as expected. When the δ-SiGe layer (splits B and C) is implemented a further reduction of the tunneling onset is observed compared to split A due to the higher germanium concentration.

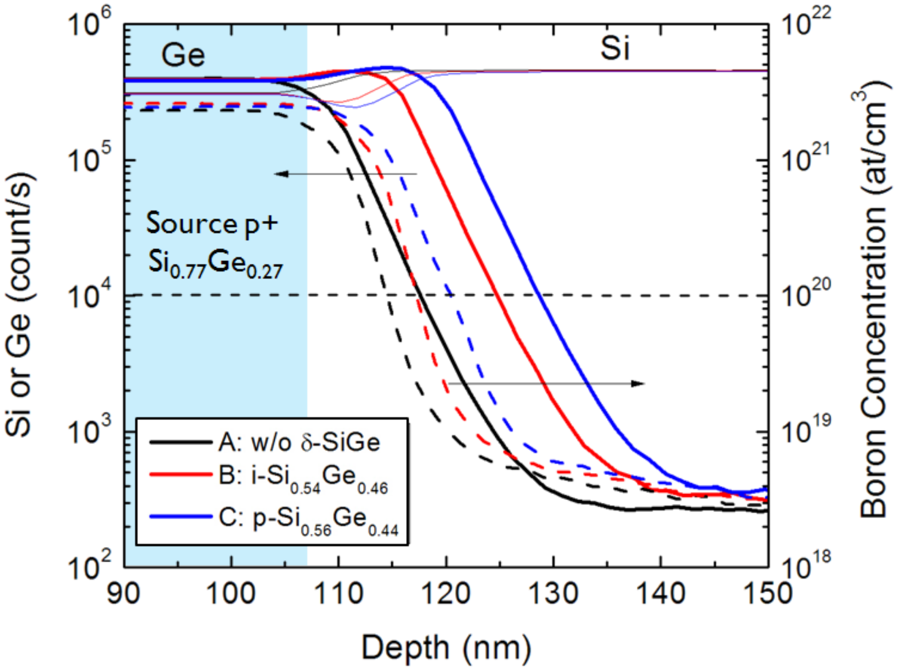


Figure 6.11: SIMS profile for the SiGe splits showing the contents of Si and Ge and the quantification of the boron concentration in the sample. The data are obtained from test wafers which have the same thermal budget of the actual processing. The data are aligned to the source ($Si_{0.77}Ge_{0.27}$) edge.

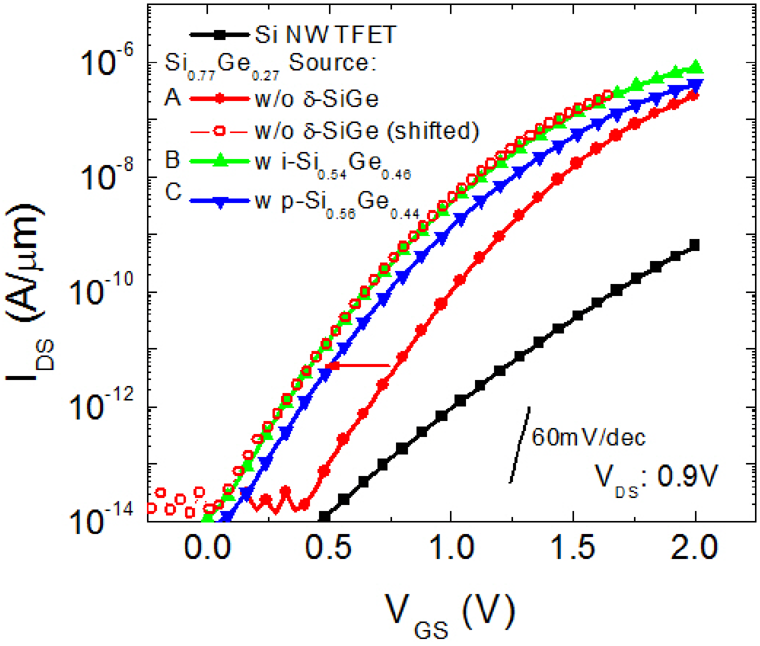


Figure 6.12: Transfers characteristics of vertical nTFET. The curves are obtained by averaging data of several devices. The nanowire diameter is 200nm.

After shifting the characteristic of split A, it is clear that no improvement of the subthreshold swing is present for the three SiGe splits at the low gate voltages. The reason should be attributed to the presence of traps/defects at the SiGe/Si interface or in the SiGe region. This assumption is confirmed by TCAD simulations in Figure 6.13 (for split C). In this simulation, non local TAT and BTBT models with a new set of calculated BTBT parameters [18] have been used and no strain is considered. In particular, three different characteristics have been simulated: the first includes non local BTBT and TAT models, the second includes only BTBT and the last includes only the TAT. In addition, the carrier lifetime has been reduced by a factor 50 to match the experimental results and to emulate the presence of defects in the device. It is clear that the earlier onset of the TAT component is dominant over the BTBT component at low gate voltages. On the other hand, the BTBT component is dominant in the saturation regime. The presence of TAT is responsible of the degradation of the sub threshold swing.

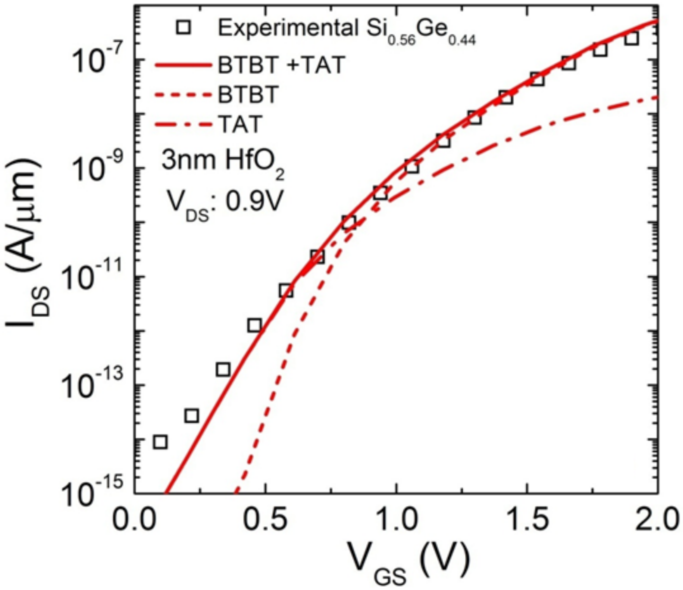


Figure 6.13: Results of a TCAD simulation of vertical SiGe nanowire with a $Si_{0.56}Ge_{0.44}$ doped source similar to split C.

6.3.3 *Impact of the Si Capping Layer*

As shown in Chapter 4, the integration of complementary TFETs with different materials might be compromised by the poor quality of the interface between the nanowire and the gate oxide. To offer a better quality interface a Si capping could be deposited before the gate oxide deposition. The impact of the Si capping layer is explored in this section.

Preliminary experiments show how critical is the control of this thin Si layer. In Figure 6.14, a Si capping layer of 18 monolayers (~2.3nm) has been implemented but this layer was fully removed at the end of the processing; most likely, the growth rate of this layer at the sidewalls is slower than the one on horizontal surfaces and/or the cleans prior to gate stack deposition are enough to remove a large portion of the capping layer. As a consequence, no clear impact on the tunneling current is observed in Figure 6.15. More studies are needed to assess its impact on the device performance.

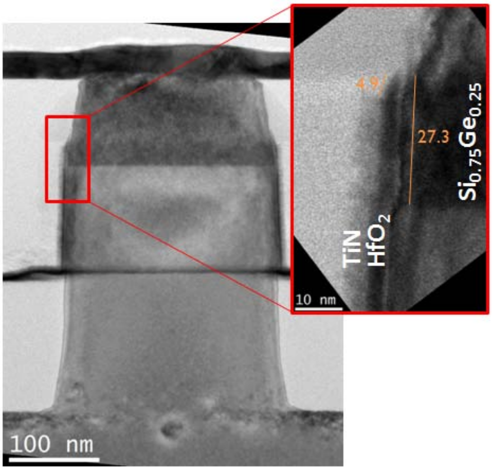


Figure 6.14: TEM image of a vertical nanowire TFET with a diameter of 200nm. A clear contrast between the SiGe source and the intrinsic Si channel indicates the sharpness of this interface. The black line in the middle of the nanowire represents the projection of the gate meaning the cleaving is not exactly in the middle of the specimen. The enlargement shows a presence of a thin oxide layer between the HfO_2 and the source but it is difficult to determine if this layer belongs to the thermal oxide or the Si capping layer.

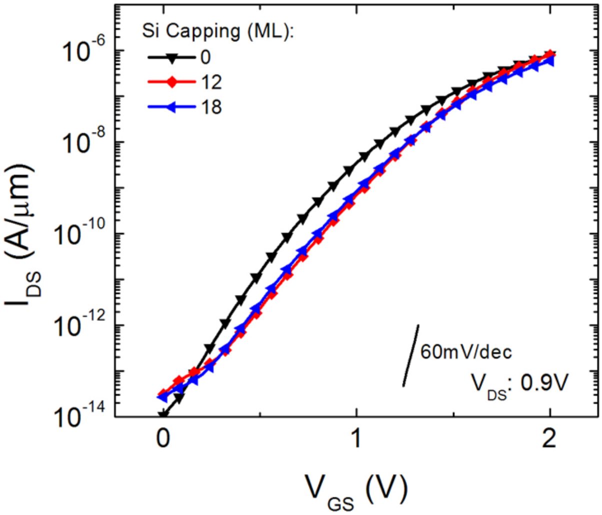


Figure 6.15: Transfer characteristics for vertical nTFET (split C) for different thicknesses of the Si capping layer measured at V_{DS} = 0.9V. A shift of the tunneling onset is observed without a clear improvement of the subthreshold swing. (D_{NW}: 200nm N: 20)

6.4 Analysis of the TFET Performance

We have demonstrated that the implementation of heterojunction TFETs can effectively boost the on-current. However, it has little impact on the subthreshold swing which remains too low. In this section the origin of this degradation is investigated for Si nanowire as well as for the heterojunction TFETs by analyzing the interface trap density and the temperature dependence of the input characteristics.

6.4.1 Extraction of Interface Traps

Vertical Si nMOSFETs were also fabricated following the same integration scheme as the Si TFETs but using arsenic implantation for the top source doping. The position of the bottom (drain) junction was tuned to ensure proper gate-drain overlap. Figure 6.16 shows the transfer characteristics of the vertical MOSFETs for 100nm and 50nm wide nanowires.

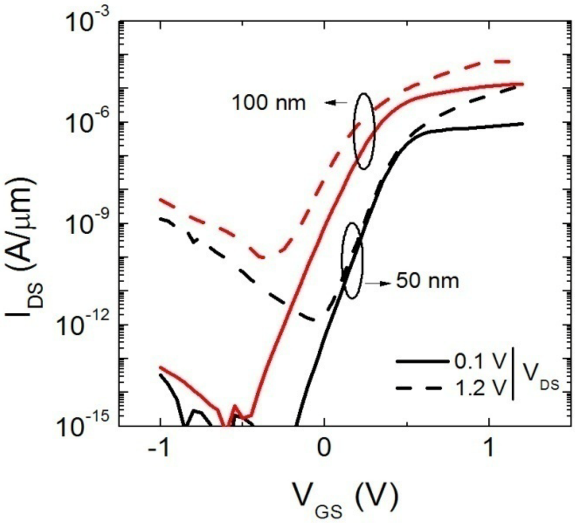


Figure 6.16: Input characteristics of vertical nMOSFETs with 50nm & 100nm nanowire diameter (dashed line: Vds=0.1V, solid line: Vds=1.2V).

For the 50nm nanowires, DIBL and short channel effects are fully suppressed and a perfect control of the gate on the channel is achieved.

Therefore, one can use the expression of the subthreshold slope in gate-all-around devices [19] to extract the interface trap density, D_{it}:

$$S = \frac{kT}{q}\ln 10\left(1+\frac{C_{it}}{C_{ox}}\right) \cong \frac{kT}{q}\ln 10\left(1+\frac{qD_{it}}{C_{ox}}\right) \tag{6.3}$$

The 75mV/dec extracted slope results in an interface trap density close to $2\text{x}10^{12}\text{eV}^{-1}\text{cm}^{-2}$. This concentration is much higher than the one extracted for finfet in the previous chapter and reflects the roughness of the sidewalls after nanowire etching. The presence of traps is expected to have a strong impact on the tunneling characteristics as previously reported for III-V channel TFETs [20-21], by the presence of additional tunneling mechanisms through traps. Temperature measurements are carried out to further analyze this phenomenon in Si TFETS, as well as in the case of heterojunction devices.

6.4.2 Temperature Dependence

The impossibility to achieve sub 60mV/dec swing suggests that additional mechanisms are dominant at low gate voltage. For this reason, the devices have been characterized at different temperatures to identify other transport mechanisms.

In Figure 6.17, the vertical Si TFET transfer characteristics show a clear temperature dependence suggesting the presence of TAT [22]. Results of TCAD simulations are able to match the experimental data after reducing the carrier lifetime by a factor 10 (Figure 6.18). TCAD simulation includes the non local BTBT model and TAT model based on Schenk [23]. In this case the TAT is dominant over the BTBT up to a gate voltage of 2V. The presence of TAT with a high density of interface traps and the low source concentration ($2\text{x}10^{19}$ at/cm^3) explains the poor performance of Si TFETs.

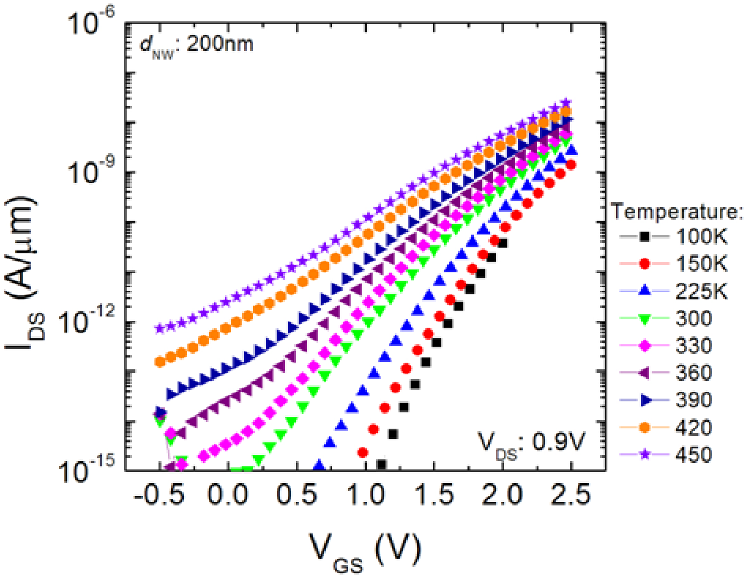


Figure 6.17: Transfer characteristics of Si TFET for different temperatures measured at drain voltage of 0.9V. (D_{NW}: 200nm)

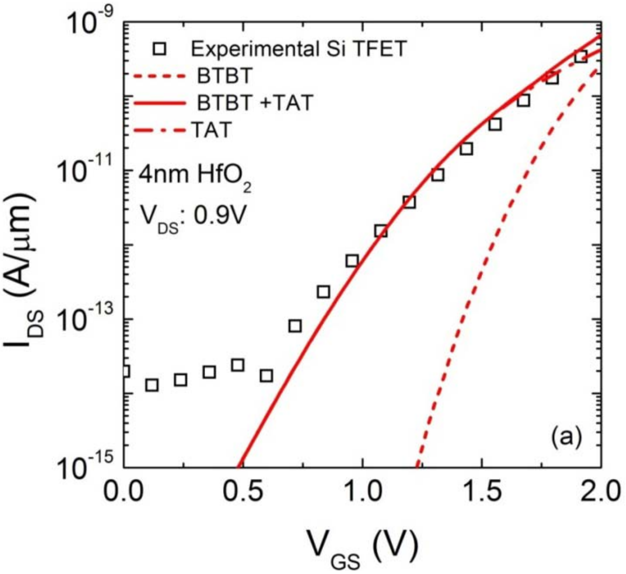


Figure 6.18: Results of a TCAD simulation showing that TAT is dominant over BTBT in case of Si TFET.

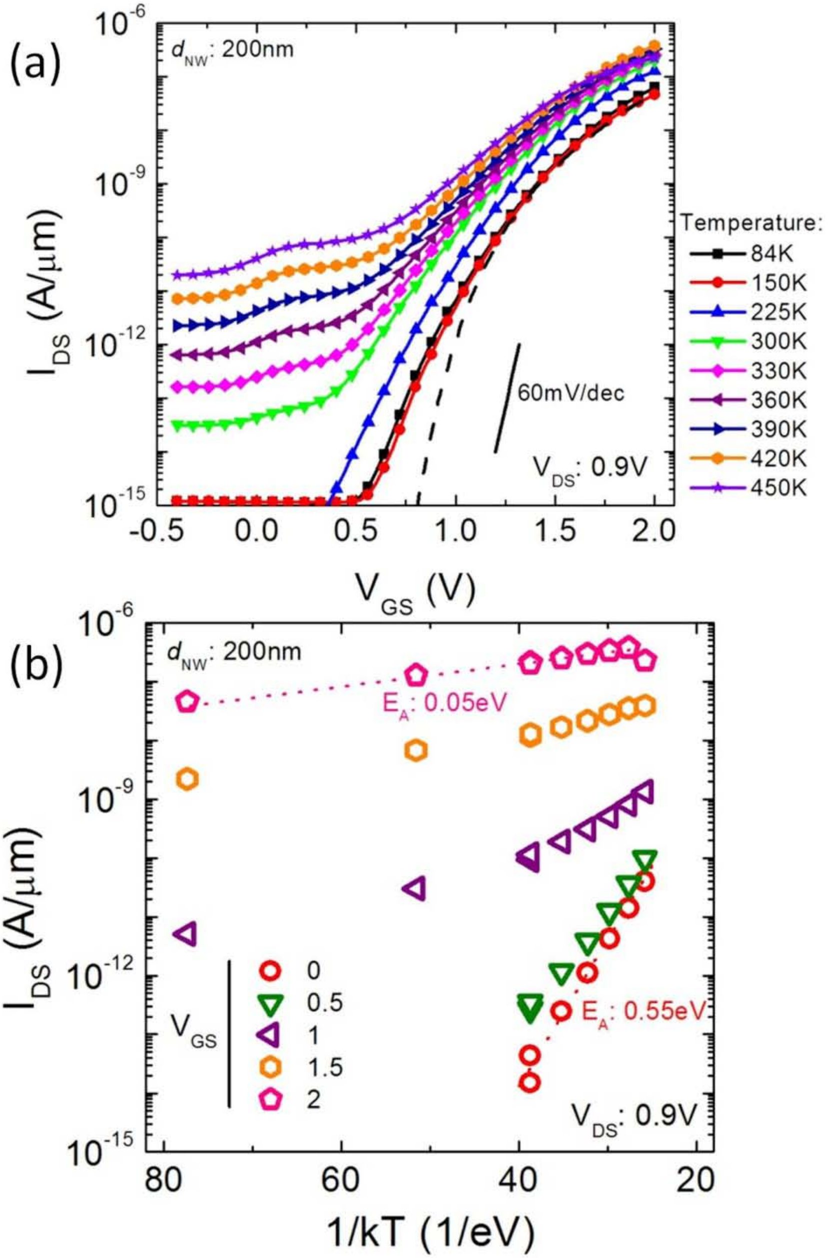


Figure 6.19: Transfer characteristics of SiGe (split A) for different temperatures measured at drain voltage of 0.9 V (a). The dash line is the result of a TCAD simulation where only BTBT is included. Extraction of activation energy at different gate voltages (b). (D_{NW}: 200nm)

In case of the SiGe source (split A), the temperature dependence shown in Figure 6.19a suggests that the device current is mainly dominated by BTBT tunneling due to the small positive variation with the temperature. The positive dependence of the band-to-band tunneling current with temperature is related to the small reduction of the source energy bandgap with temperature. In Figure 6.19b, the activation energy at different gate voltages is extracted confirming that the main leakage mechanism at low gate voltage is SRH in the Si channel since the activation is about half of the Si energy bandgap. Nevertheless, we cannot exclude the presence of the TAT at low gate voltage as shown in Figure 6.13. In particular, the pure tunneling characteristic which should be visible below 150K shows a swing still larger than expected.

The extracted activation energy as a function of gate voltage is shown in Figure 6.20 where a SiGe device (split A) is compared to a Si TFET. It reveals that band-to-band tunneling featuring an activation energy less than 0.1eV is only occurring at gate voltages above 2.5V in case of vertical Si TFET contrary to the finfet-based Si TFET. At lower voltages, the larger activation energy corresponds to trap-assisted tunneling. Finally, the OFF state of the device is characterized by a generation current (E_A: ~0.56eV in Si) typical for reverse-biased diode or a surface recombination current (E_A: ~0.7eV) [24]. In the case of the SiGe source TFET, the E_A is less than 0.1eV featuring BTBT for gate voltages above 1.5V similarly to the results obtained with finfet-based Si TFETs where TAT is still present but only at very small currents or over a small range at low gate bias.

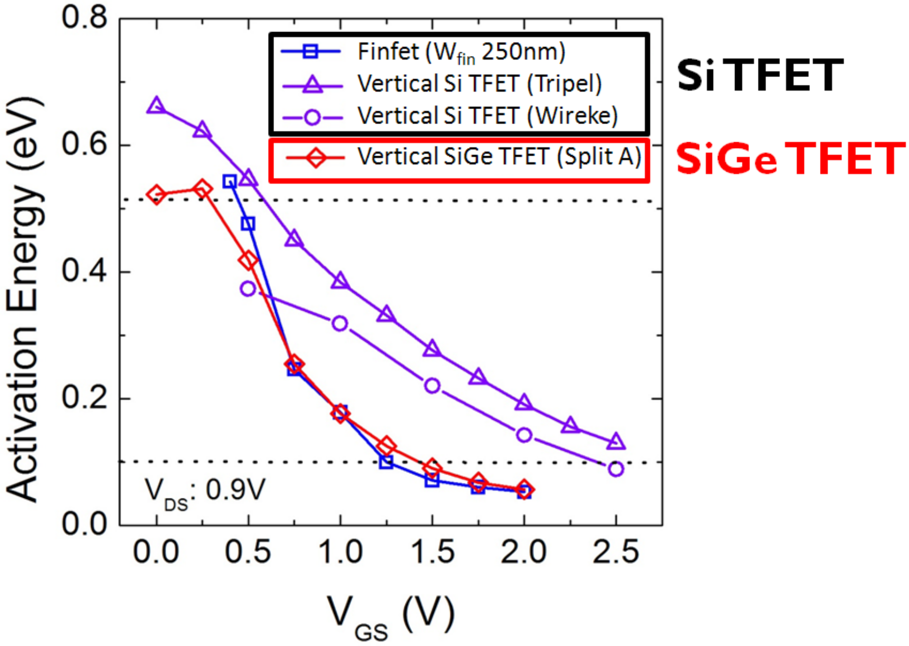


Figure 6.20: Arrhenius plot of vertical TFETs compared with the finfet-based TFET extracted at V_{DS} of 0.9V. (W_{fin} 250, L_G 150)

6.5 Summary

The vertical architecture is a preferred architecture for the implementation of heterojunction TFETs but faces several processing challenges, such as gate alignment and nanowire contacting. In spite of the processing complexity, SiGe TFETs were successfully integrated with improved drive current as shown in Figure 6.21.

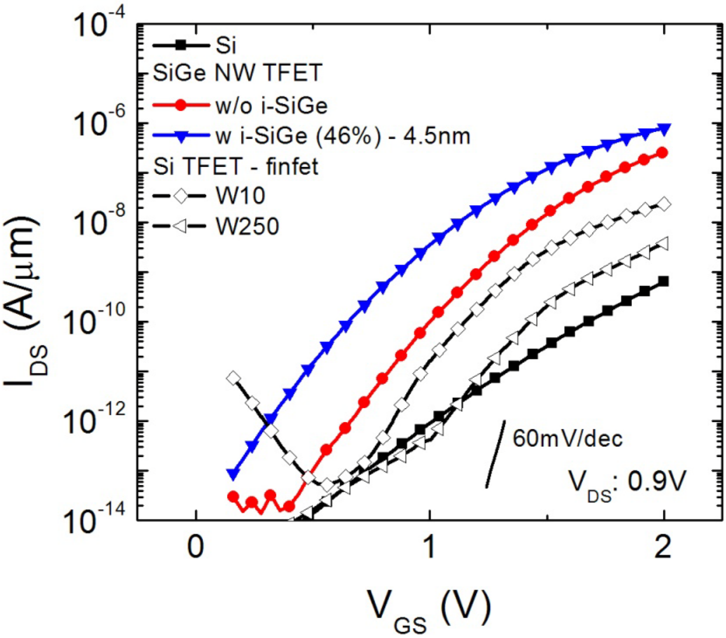


Figure 6.21: Transfer characteristics of vertical TFET compared to Si TFET (POR conditions) for wide and narrow fins (L_G: 150 nm). The nanowire diameter is 200nm, $L_{G/S}$ is 25 nm while the L_G is 150 nm and bottom isolation is 100 nm. The 60 mV/dec slope is indicated as a reference. All the curves are measured at V_{DS} = 0.9 V.

In the planar case, the characteristics are degraded by TAT at low voltage bias. On the other hand, the characteristics of the vertical Si TFETs are completely dominated by TAT while the BTBT component is reduced due to the low source doping. The traps might be located inside the source material or at the nanowire sidewalls as suggested by the extraction of the interface traps for Si devices. It is there important to further improve the surface quality of the etched nanowires by optimized etch chemistry and efficient passivation of dangling bonds, as well as maintaining a low defectivity in the source material, especially in case of heterojunctions where defects are more likely to be present. We performed different re-oxidations following the Si nanowire etch. The oxidations were performed either at 550°C with an oxide thickness of 1.3nm repeated twice or five times, or at 1070°C with an oxide of 10nm. However, no significant change was observed electrically. In view of these results, these oxidations appear still insufficient to fully cure the damaged nanowire surface.

In conclusion, we believe that the implementation of heterojunction may not be enough to achieve the intended device specifications and that further modifications of the TFET architecture are needed. In particular, it is necessary to further boost the vertical tunneling and reduce the point tunneling component to achieve a reasonable value of drive current as explained in the next chapter.

References

[1] A. S. Verhulst, *et al.*, "Boosting the on-current of a n-channel nanowire tunnel field-effect transistor by source material optimization," *Journal of Applied Physics,* vol. 104, pp. 064514-10, 2008.

[2] K. K. Bhuwalka, *et al.*, "Vertical tunnel field-effect transistor," *IEEE Transactions on Electron Devices,* vol. 51, pp. 279-282, 2004.

[3] C. Sandow, *et al.*, "Impact of electrostatics and doping concentration on the performance of silicon tunnel field-effect transistors," *Solid-State Electronics,* vol. 53, pp. 1126-1129, 2009.

[4] W. Y. Choi, *et al.*, "Tunneling Field-Effect Transistors (TFETs) With Subthreshold Swing (SS) Less Than 60 mV/dec," *IEEE Electron Device Letters,* vol. 28, pp. 743-745, 2007.

[5] R. Gandhi, *et al.*, "CMOS-Compatible Vertical-Silicon-Nanowire Gate-All-Around p-Type Tunneling FETs With < 50-mV/decade Subthreshold Swing," *IEEE Electron Device Letters,* vol. 32, pp. 1504-1506, 2011.

[6] D. Leonelli, *et al.*, "Performance Enhancement in Multi Gate Tunneling Field Effect Transistors by Scaling the Fin-Width," *Japanese Journal of Applied Physics,* vol. 49, p. 04DC10, 2010.

[7] F. Mayer, *et al.*, "Impact of SOI, $Si_{1-x}Ge_x$OI and GeOI substrates on CMOS compatible Tunnel FET performance," in *IEEE International Electron Devices Meeting* New York, 2008, pp. 163-166.

[8] K. Boucart and A. M. Ionescu, "Double-gate tunnel FET with high-kappa gate dielectric," *IEEE Transactions on Electron Devices,* vol. 54, pp. 1725-1733, Jul 2007.

[9] J. Knoch, *et al.*, "Impact of the dimensionality on the performance of tunneling FETs: Bulk versus one-dimensional devices," *Solid-State Electronics,* vol. 51, pp. 572-578, 2007.

[10] A. S. Verhulst, *et al.*, "Modeling the single-gate, double-gate, and gate-all-around tunnel field-effect transistor," *Journal of Applied Physics,* vol. 107, p. 024518, 2010.

[11] T. H. A. Schulze, P. Eyben, A.S. Verhulst, R. Rooyackers, A. Vandooren, J. Mody, A. Nazir, D. leonelli, W. Vandervorst,,

"Observation of diameter dependent carrier distribution in nanowire-based transistors," *Nanotechnology,* vol. 22, p. 185701, 2011.

[12] K. Varga, *et al.,* "Critical layer thickness in Stranski–Krastanow growth of Ge on Si(0 0 1)," *Surface Science,* vol. 562, pp. L225-L230, 2004.

[13] A. Sammak, *et al.,* "High-quality p+ n Ge diodes selectively grown on Si with a sub-300nm transition region," in *European Solid-State Device Research Conference,* 2011, pp. 359-362.

[14] R. Braunstein, *et al.,* "Intrinsic Optical Absorption in Germanium-Silicon Alloys," *Physical Review,* vol. 109, pp. 695-710, 1958.

[15]http://www.sp.phy.cam.ac.uk/~SiGe/Strain%20in%20Silicon%20Germanium%20%28SiGe%29.html.

[16] K. K. Bhuwalka, *et al.,* "Scaling issues of n-channel vertical tunnel FET with δ–p^+ SiGe layer," in *Device Research Conference Digest,* 2004, pp. 215-216 vol.1.

[17] G. Han, *et al.,* "Source Engineering for Tunnel Field-Effect Transistor: Elevated Source with Vertical Silicon-Germanium/Germanium Heterostructure," *Japanese Journal of Applied Physics,* vol. 50, p. 04DJ07, 2011.

[18] K. Kuo-Hsing, *et al.,* "Direct and Indirect Band-to-Band Tunneling in Germanium-Based TFETs," *IEEE Transactions on Electron Devices,* vol. 59, pp. 292-301, 2012.

[19] D. J. Wouters, *et al.,* "Subthreshold slope in thin-film SOI MOSFETs," *IEEE Transactions on Electron Devices,* vol. 37, pp. 2022-2033, 1990.

[20] S. Mookerjea, *et al.,* "Temperature-dependent I-V Characteristics of a Vertical $In_{0.53}Ga_{0.47}As$ Tunnel FET," *IEEE Electron Device Letters,* vol. 31, pp. 564-566, 2010.

[21] W. C. Kao, *et al.,* "Effect of interface states on sub-threshold response of III–V MOSFETs, MOS HEMTs and tunnel FETs," *Solid-State Electronics,* vol. 54, pp. 1665-1668, 2010.

[22] E. Simoen, *et al.,* "On the Temperature and Field Dependence of Trap-Assisted Tunneling Current in Ge p(+)n Junctions," *IEEE Electron Device Letters,* vol. 30, pp. 562-564, May 2009.

[23] A. Schenk, "A model for the field and temperature dependence of Shockley-Read-Hall lifetimes in silicon," *Solid-State Electronics,* vol. 35, pp. 1585-1596, 1992.

[24] A. B. Sproul and M. A. Green, "Intrinsic carrier concentration and minority carrier mobility of silicon from 77K to 300K," *Journal of Applied Physics,* vol. 73, pp. 1214-1225, Feb 1993.

Chapter 7

Hybrid TFET: A New Architecture

7.1 Introduction

In this chapter, the limitations of the lateral TFET configuration are discussed and a new architecture is proposed: the hybrid TFET. In this way, we aim to boost the tunneling current and achieve a steeper subthreshold swing by boosting the line tunneling component and by suppressing the point tunneling which is responsible of the slow turn-on of the lateral TFET devices.

This work is mainly based on TCAD simulations and it focuses on the implementation of heterojunction SiGe/Si TFET. Nevertheless, a process flow to implement the simulated architecture is proposed but unfortunately electrical results were not available before the competition of this thesis. The hybrid architecture can be used for heterojunction TFET with III/V material but this topic is not treated in this work because of the limitations of the available simulation tools.

7.2 The Limitations of the Lateral TFET

In the previous chapters, the performance of the TFET has been analyzed for two different architectures: finFET and vertical nanowire as depicted in Figure 7.1. In case of Si TFETs, we have argued that the expected sub 60 mV/dec was not achieved because of the presence of a doping gradient due to the implantation (section 5.4). In fact, we can imagine to divide the source region in several regions longitudinal to the channel. Each region has a different doping concentration from highly doped to lightly doped close to the channel. At low electric field, the tunneling is starting in the lowly doped region. When the electric field increases, the other regions are gradually switching on. As a consequence, the overall tunneling current and subthreshold swing result to be degraded.

To solve this issue, we have implemented heterojunction TFETs where the source is replaced with epitaxially grown SiGe (with germanium content up to 46%). The improved abruptness of the junction, however, did not break the kT-limit of the experimental TFETs.

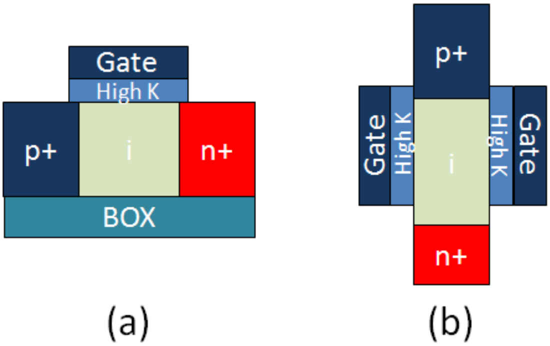


Figure 7.1: Schematic of the two lateral TFET configurations implemented in this work for finFET (a) and vertical nanowire (b).

We believe that the gradual turn on is an intrinsic limitation of the lateral TFET design. In the lateral configuration, the point tunneling is dominant because we assume only a marginal overlap between the source and the gate. Figure 7.2 shows a simulated I_{DS}-V_{GS} plot for a lateral Si TFET together with the plots of the generation rate of electrons and holes to better understand when the tunneling starts. At low bias, the band-to-band tunneling is occurring at the edge of the source and the tunneling path is mainly directed along the channel (Figure 7.2b). Increasing the gate voltage moves the tunneling direction more perpendicular to the gate and the tunneling path is getting shorter which indicates a larger tunneling current (Figure 7.2c). It is clear that if the lateral component which has longer tunneling paths is suppressed, the turn on of the device can be improved.

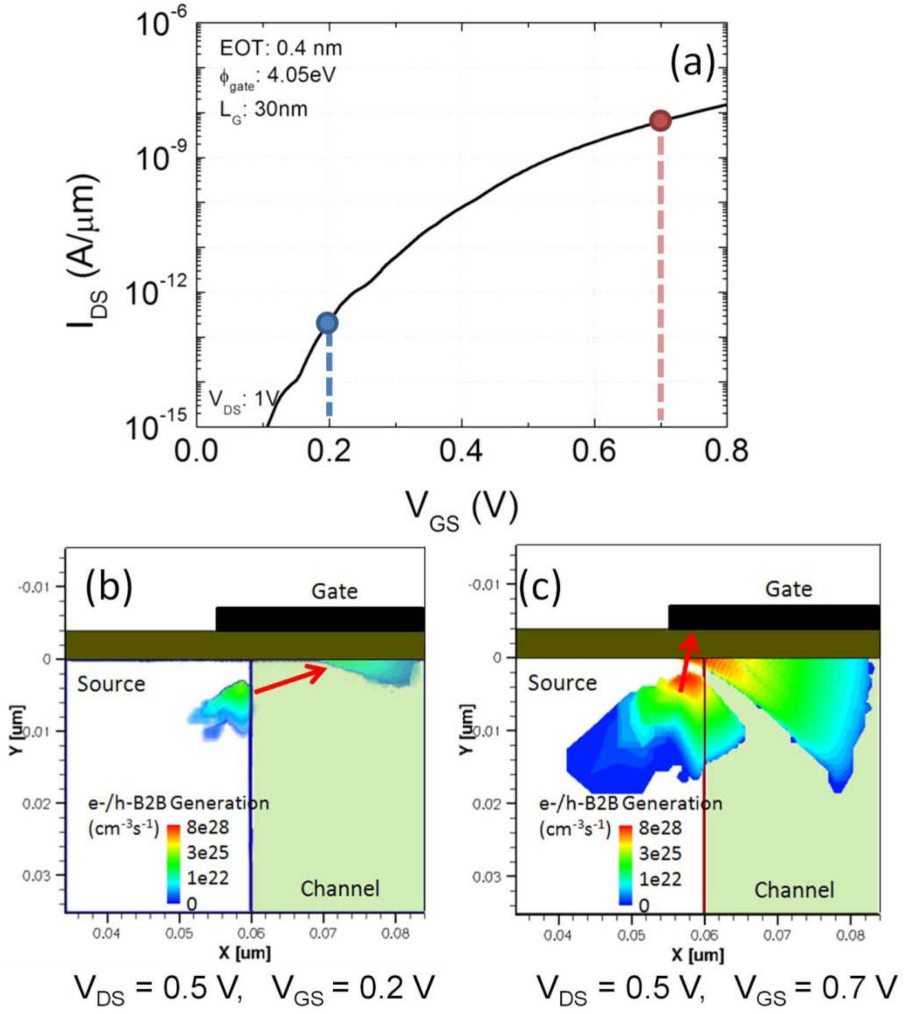


Figure 7.2: TCAD simulation of a lateral Si TFET (a) and band to band generation plots for electron and holes in proximity of the tunneling region for low gate voltage (b) and high gate voltage (c).

Increasing the gate-to-source overlap is not sufficient. Indeed, the vertical component of the tunneling, or line tunneling, will be enhanced as shown in Figure 7.3. Nevertheless the lateral electric field is not completely suppressed because of the presence of fringing fields.

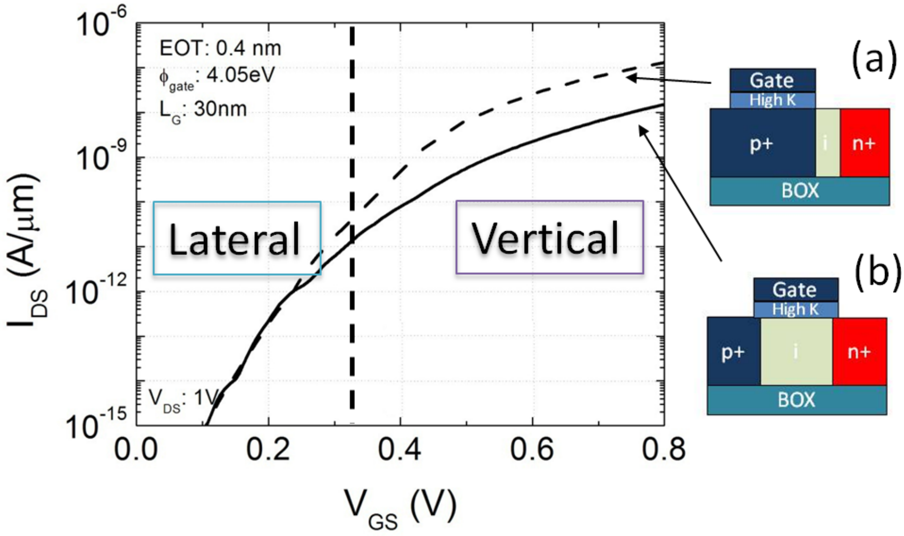


Figure 7.3: Transfer characteristics of a TFET where the gate fully overlap the source (a) and a lateral TFET configuration (b).

Alternatively, the source doping can be decreased maintaining a large gate/source overlap. Experimental results with sub60mV/dec were demonstrated in [1] but the lower doping affects drastically the ON current of this device.

Another option is to increase the vertical electric field by introducing a counter-doped pocket between the source and the gate [2]. Figure 7.4 shows how the generation rate looks like when such a highly doped pn junction is implemented. In this case, most of the tunneling component is vertical (i.e. line tunneling) but the doping concentration and position are extremely critical parameters which increase the process complexity.

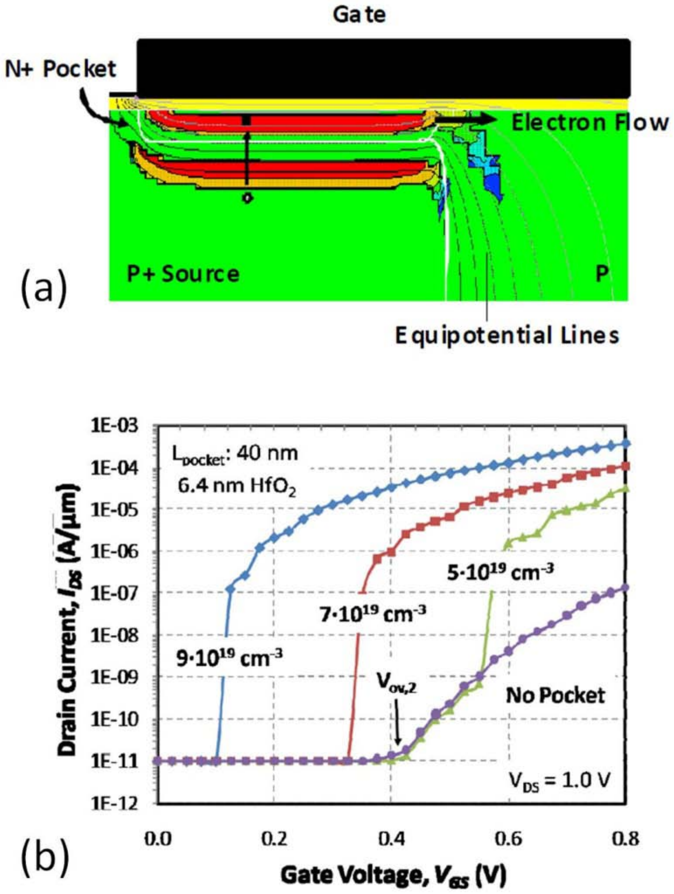


Figure 7.4: Carrier generation plot showing the main tunneling direction (a) and transfer characteristics showing how sensitive is the tunneling current with the doping concentration of the counter-doped pocket [3].

Another solution is to add an intrinsic or lowly doped thin layer between the gate and the source and maintaining a complete overlap between the gate and the source as describe in the next section.

7.3 The Hybrid TFET

In the “hybrid” TFET [4], the source region is fully overlapped with the gate and a thin channel region (i.e. pocket) is placed in-between the source and the gate oxide as shown [5] in Figure 7.5. Here the device is a Si TFET where the Si source and drain region are highly doped p-type (10^{20} cm^{-3}) and n-type (10^{20} cm^{-3}), respectively while the Si channel region is lowly n-type doped (10^{15} cm^{-3}).

This novel architecture is able to concentrate most of the electric field in the channel region between the gate and the source. As a consequence, the vertical component of tunneling is dominant over the lateral component as shown in Figure 7.6. Here, it is clear that the tunneling paths are orthogonal to the gate oxide and have the same tunneling lengths for a given gate voltage because of the presence of a uniform electric field induced by the gate voltage. In case of an optimized pocket, the saturation of the ON current is only occuring for large gate voltage, outside the targeted V_{DD} range, i.e. 0.5V or below.

The planar configuration proposed is only one of many possible implementations. More details can be found in [patent pending].

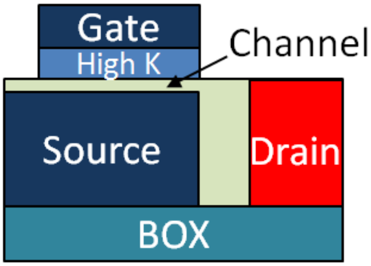


Figure 7.5: Schematic of the hybrid TFET implemented on SOI substrate for planar architectures.

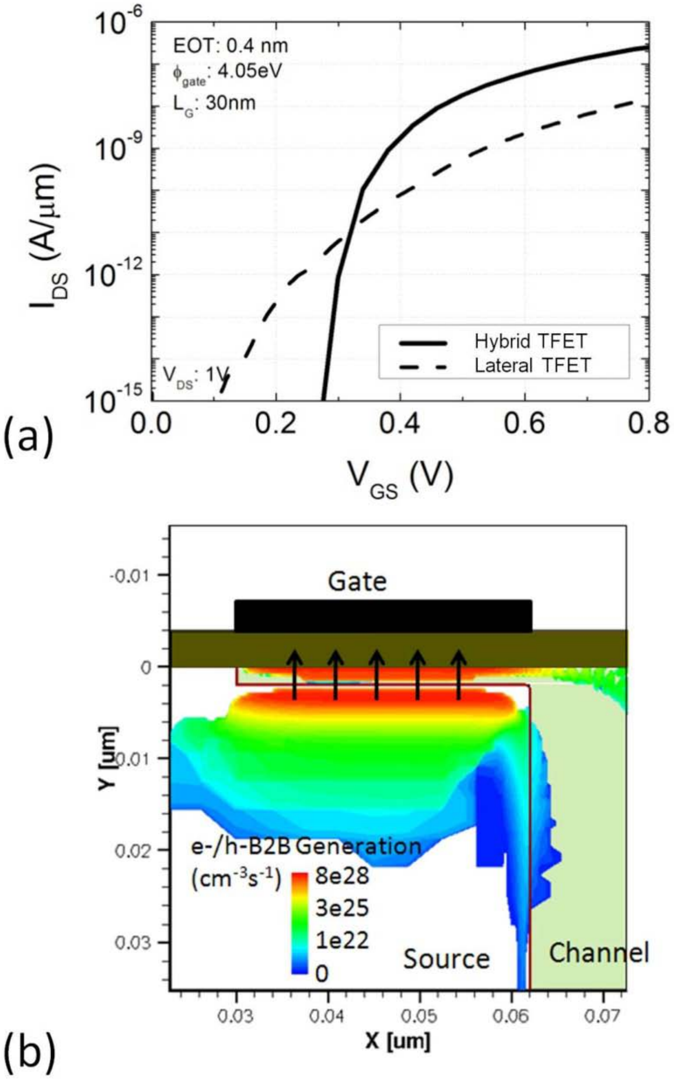


Figure 7.6: Transfer characteristics of the hybrid TFET compared to the lateral TFET (a) and band to band tunneling generation rate for electrons and holes in case of the hybrid TFET at ON state (b). Non local BTBT, Schockley-Read-Hall recombination and drift diffusion models were used to model the carrier transport. Fermi statistics is assumed.

7.3.1 Impact of Geometrical Parameters

The main geometrical parameters are indicated in Figure 7.7. The source is represented on the top of the BOX but in reality an additional layer can be present. For instance, in case of heterojunction TFET the Ge is grown on the undoped Si. The thickness of the source, t_{source}, is kept equal to 60 nm in most of our simulations but it can be reduced without performance degradation down to 30nm if the source doping is equal or higher than 10^{19} cm^{-3}. The channel thickness, t_{ch}, is an important parameter but it is not a critical process variable since the epitaxial growth of this layer can be well controlled. The misalignment of the position of the source edge with respect to the gate is quantified by the parameter *X2*, which is positive when the source is extending farther than the gate edge as shown in Figure 7.7.

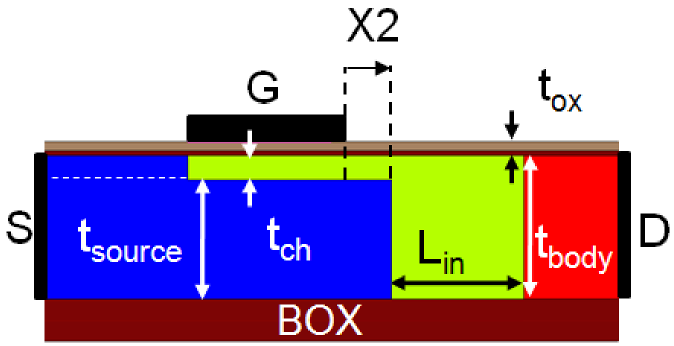


Figure 7.7: Schematic of the hybrid configuration. Gate, source and drain are labelled with G, D and S respectively. Geometrical parameters are also indicated.

The initial simulations consider a homojunction Si TFET since the Si parameters have been already validated by experimental calibration as explained in Chapter 2. Most of the finding for homojunctions can be partially transferred to the case of the heterojunction TFETs.

Impact of the channel thickness

The channel thickness limits the minimum tunneling path if we consider a highly doped source since the channel is lowly doped. The variability with the channel thickness is shown in Figure 7.8. A wider channel is characterized by a reduced saturation current because the tunneling barrier is limited by the physical channel thickness. For this reason we need to minimize the channel thickness. The quantization effects are not taken into account in this simulation because of the limitations of the simulation tool. Nevertheless, quantization is an important parameter which can shift the device characteristics as shown in [6].

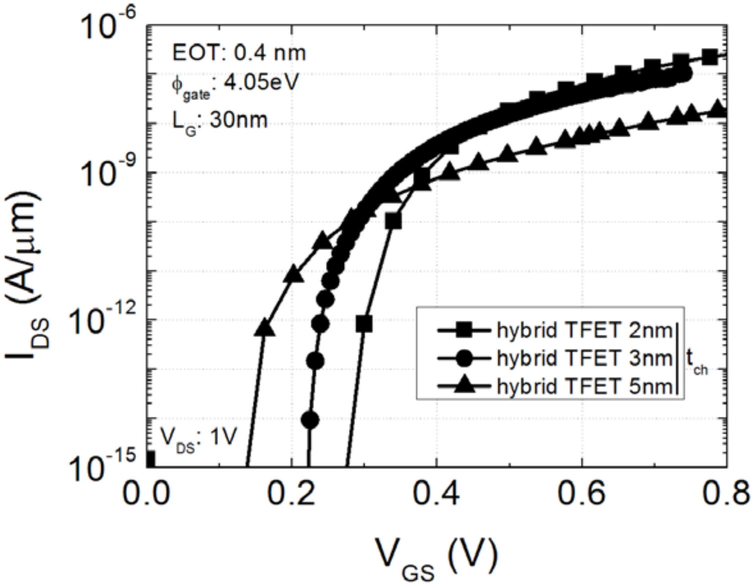


Figure 7.8: TCAD simulations showing the impact of the channel thickness on the performance of the Si hybrid TFET.

Impact of the Gate Misalignment

As shown in Figure 7.9, the gate to source misalignment has a strong impact on the device swing in case of negative values of X2 because of the corner effects. In fact, most of the electric field is concentrated on the source corner for low gate voltages and as a result parasitic tunneling occurs. At higher gate voltage, the main tunneling component is between the channel and the source so no degradation is observed compared to the case where the gate is perfectly aligned with the source edge. In case of positive variation of the X2 parameter, less degradation of the tunneling onset is observed but a degradation of the current is visible. This degradation is related to the fact that the tunneling generation in the region comprised between the drain spacer and the source is low because the absence of gate field. This region will be the bottleneck for the carrier generation when the device is on, resulting in a lower current.

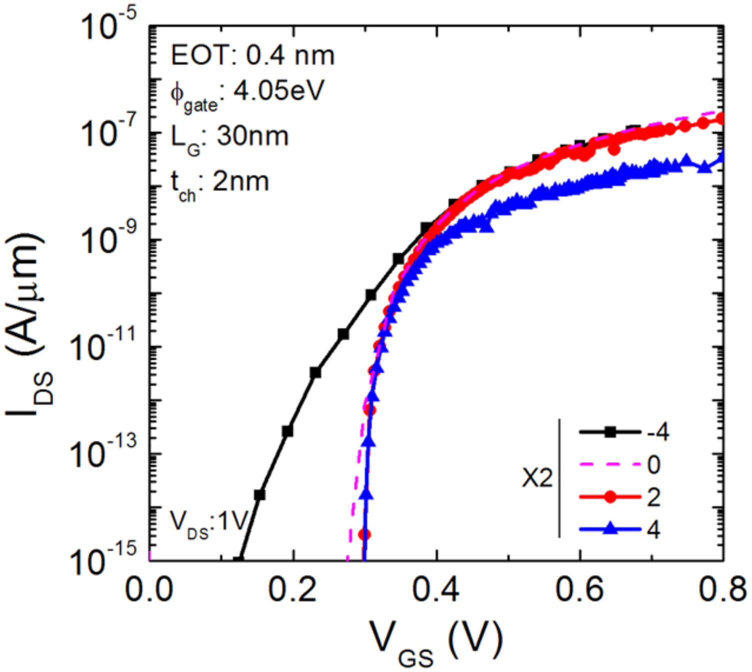


Figure 7.9: TCAD simulations showing the impact of the gate to source misalignment on the performance of the Si hybrid TFET.

The parameter, X2, is one of the most critical parameters for this configuration. This variability can be reduced in two ways: (1) including a counter-doped pocket between the gate and the source similarly to the pnpn configuration presented in section 7.2 or (2) reducing the doping concentration at the source. The first option reduces any edge effect but results in a large degradation of the onset due to variation in the doping concentration (already shown in Figure 7.4). In addition, Random Dopants Fluctuations (RDF) should also be taken into account for thin channels. The second option is preferred but it is only able to reduce the degradation without removing it. From now on, X2 is fixed to 2nm and is used as default parameter in the next simulations.

7.3.2 Optimization of the Hybrid TFET

Although the hybrid configuration is able to boost the ON current in Si TFET, the improvement is not large enough to be competitive at operation at V_{DD} = 0.5 V or below. In addition, the scalability of this configuration is limited by a minimal distance between the source and drain region necessary to reduce the diode leakage which degrade the I_{ON}/I_{OFF} ratio. To improve further the drive current and device scalability, we introduce a configuration with raised drain and replace the silicon source with germanium [7].

Since the default germanium parameters available in Sentaurus device [5] are not calibrated with experimental data, a new set of parameters calibrated with the experimental results in [1] are used. For the following simulations an additional model is included to consider the confined carrier distribution that occurs near the Si-SiO_2 interface: the Modified Local-Density Approximation (MLDA) model [8].

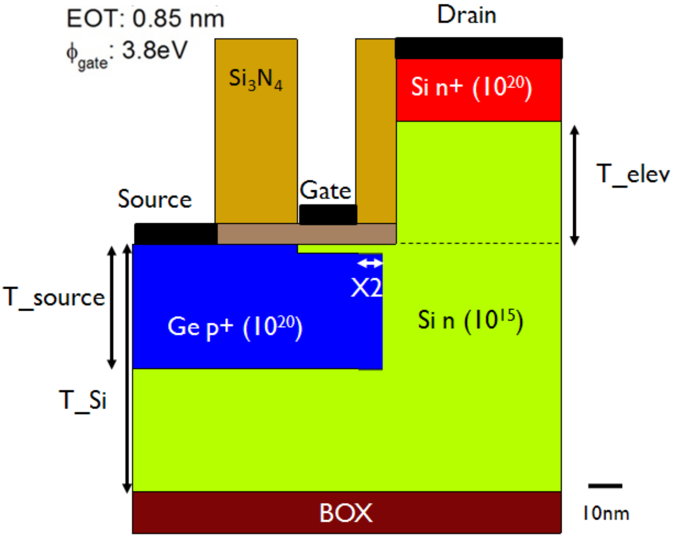


Figure 7.10: Schematic of the new architecture used in TCAD simulation.

The new device architecture is shown in Figure 7.10. T_elev indicates the elevation of the drain. For simplicity and better convergency an L-shaped channel is implemented. In case SEG is used to obtain this layer, it will result in a more sloped profile because of the formation of facets during the epitaxial growth. The gate length has been reduced to 14nm and the V_{DD} to 0.5V. The oxide EOT is set to 1nm and no gate leakage is considered in the simulation. The source is Ge highly doped p-type (10^{20} at/cm^{-3}) and later replaced by Ge with a lower doping (10^{19} at/cm^{-3}) because of the improved performance as shown later. The spacer thickness which represents the distance between the gate edge and the elevated drain edge can be adjusted.

Impact of the Source Doping

In highly doped p-type germanium, the position of the Fermi level is deeper in the valence band compared to silicon because of the lower density of states (equation (7.1)).

$$E_F - E_V = kT\ln\left(\frac{N_V}{N_A}\right) \tag{7.1}$$

where N_V is the effective density of states in the valence band and N_A is the doping concentration. As a consequence, if we lower the doping concentration the position of the Fermi level will be closer to the valence band improving the tunneling current as observed in Figure 7.11. In addition, lower doping concentration shifts the tunneling path entirely in the germanium source. On the other hand, if the doping level is too low the Fermi level will be higher than the conduction band and we will observe a degradation of the tunneling current. The transfer characteristics for different doping concentration are shown in Figure 7.12. The best results are achieved for low doping concentrations (1 x 10^{19} or 2 x 10^{19} at/cm^{3}). The excellence performance for the Ge-source hybrid TFET suggests that this device might be a potential candidate for low power application (below V_{DD} = 0.5 V).

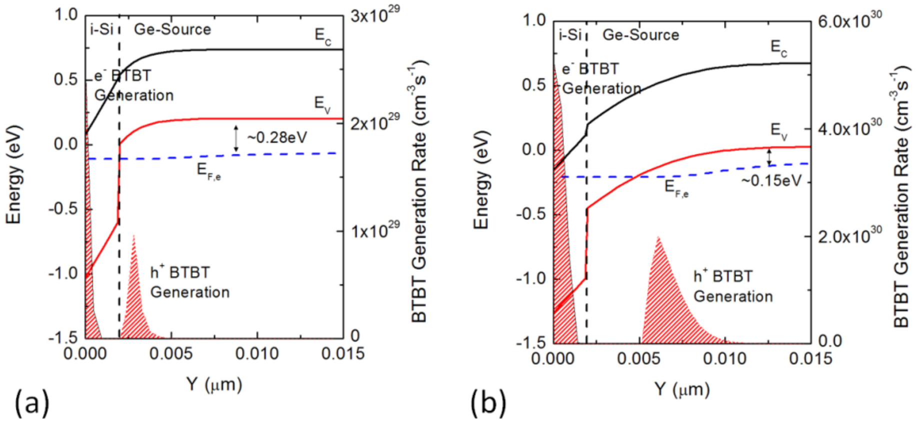


Figure 7.11: Energy bands and electron and holes BTBT generation for two different source doping concentration: 1 x 10^{20} at/cm^3 (a) and 1 x 10^{19} at/cm^3 (b). The cut is performed perpendicular to the gate and in the middle of the gate. (V_{DS}: 0.3 V)

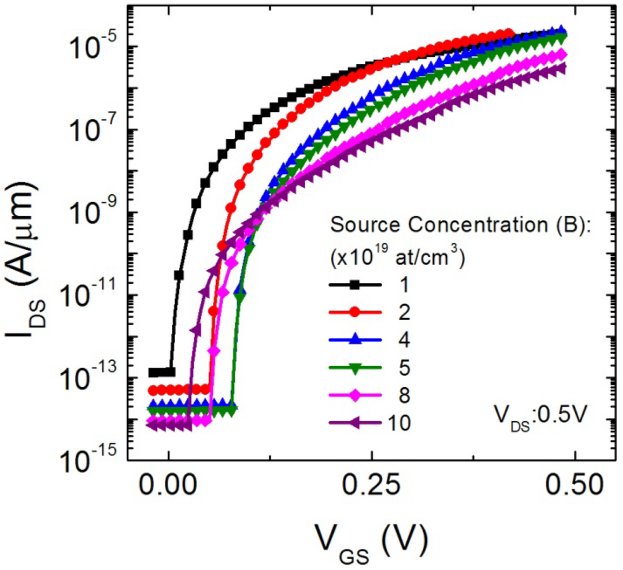


Figure 7.12: Transfer characteristics of hybrid TFET for different doping concentrations. (L_G: 14nm)

The implementation of III/V materials has not been done because of the lack of parameters for TCAD simulations. Nevertheless, we believe that this concept can be easily adapted by a proper choice of materials as demonstrated by the several promising experimental results appearing recently in the literature [9-10].

Impact of the Raised Drain

The combined contribution of the drain spacer width and T_elev is equivalent to the parameter L_{in} specified earlier in Figure 7.7. As a consequence, a degradation of the I_{ON}/I_{OFF} ratio is expected for small values of these parameters, as shown in Figure 7.13. T_elev of 30nm and a spacer width of 10nm are used as a standard reference in the following simulations.

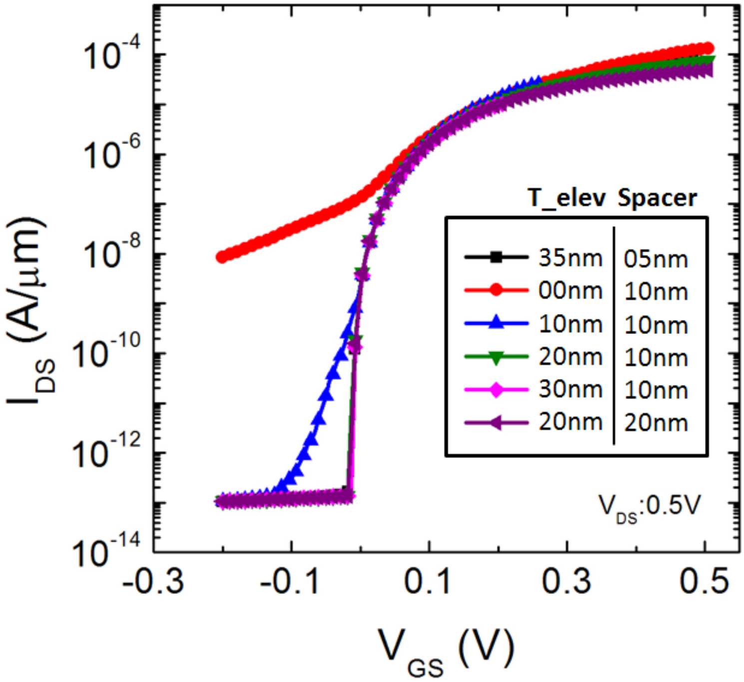


Figure 7.13: TCAD simulations showing the impact of T_elev and the drain spacer width for Ge-source hybrid TFET (N_{source} = $1x10^{19}cm^{-3}$).

7.3.3 The Fabrication of the Hybrid TFET

The fabrication of the planar hybrid TFET was started but a final prototype device did not reach process completion at the time this manuscript was written. For this reason, no experimental result will be presented but the main steps of the process flow are summarized in this section (see Figure 7.14).

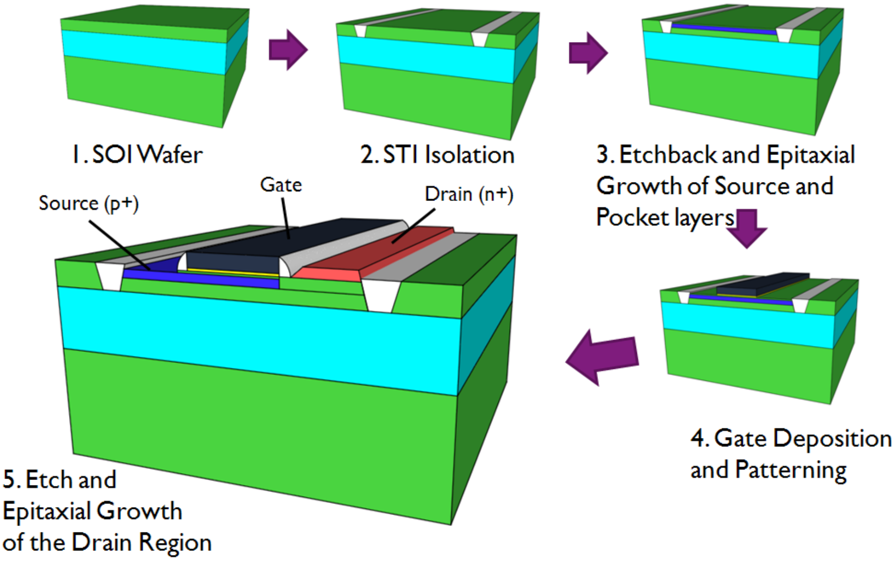


Figure 7.14: Schematic view of the proposed process flow to integrate the hybrid concept.

The starting substrate is a SOI wafer similar to the one used for the finfet processing. Contrary to the finfet, STI isolation is needed to define the active area of the device. After the STI, the Si is etched back and the SiGe source and the Si channel (pocket) are epitaxially grown. Different splits in Ge concentration and doping level are being implemented. The initial experiments use SiGe as source material. The gate is then deposited and patterned. After the spacer formation, a source protection layer is deposited to protect the source region from the controlled etch-back to etch the source material present at the drain region. The gate to source misalignment is defined during this etch together with the spacer thickness. Afterward, undoped Si is grown on the top of the residual Si at the drain side followed by doped silicon epitaxial growth. After the removal of the source protection layer, the device is then ready for silicidation and back-end processing.

7.4 Summary

In this chapter, a new device concept has been introduced: the hybrid TFET. This device is able to provide a better swing and boost the on current of the lateral TFET design by overlapping the source with the gate and concentrating the tunneling along a direction normal to the gate. In this way, all the possible tunneling paths have the same length and the gradual turn on of the device observed in the lateral configuration is not present.

In addition, the impact of different processing and geometrical parameters show that the main source of variability is the gate misalignment. In fact, in case the source does not fully overlap the gate, a parasitic point tunneling component arises and degrades the swing at low voltage bias. The control of this parameter can be done if a self aligned process is implemented as described in the previous section.

In this work, the impact of the quantum confinement of the carriers is not taken into account because of limitations of the simulator tool. The impact of the quantum confinement was predicted by theoretical calculations for Si TFET in [6], where a strong shift of the tunneling onset was recorded. The presence of the intrinsic layer between the source and the gate might reduce the impact of the confinement but this effect needs to be taken into account in future simulations.

This configuration is suitable to implement heterojunction TFET but only the case with Ge source is shown in this chapter because of the impossibility to simulate the III/V material with commercial TCAD simulators.

References

[1] S. H. Kim, et al., "Germanium-source tunnel field effect transistors with record high I_{ON}/I_{OFF}," in Symposium on VLSI Technology, Kyoto, Japan, 2009, pp. 178-179.

[2] V. Nagavarapu, et al., "The tunnel source (PNPN) n-MOSFET: A novel high performance transistor," IEEE Transactions on Electron Devices, vol. 55, pp. 1013-1019, 2008.

[3] P. Patel, et al., "A Low Voltage Steep Turn-Off Tunnel Transistor Design," in International Conference on Simulation of Semiconductor Processes and Devices (SISPAD), 2009, pp. 1-4.

[4] D. Leonelli, et al., "Novel architecture to boost the vertical tunneling in Tunnel Field Effect Transistors," in IEEE International SOI Conference, 2011, pp. 1-2.

[5] Synopsys Sentaurus Package version D (March 2010)

[6] W. G. Vandenberghe, et al., "Impact of field-induced quantum confinement in tunneling field-effect devices," Applied Physics Letters, vol. 98, pp. 143503-3, 2011.

[7] A. S. Verhulst, et al., "Boosting the on-current of a n-channel nanowire tunnel field-effect transistor by source material optimization," Journal of Applied Physics, vol. 104, pp. 064514-10, 2008.

[8] G. Paasch and H. Übensee, "A Modified Local Density Approximation. Electron Density in Inversion Layers," physica status solidi (b), vol. 113, pp. 165-178, 1982.

[9] G. Zhou, et al., "InGaAs/InP Tunnel FETs With a Subthreshold Swing of 93 mV/dec and I_{ON}/I_{OFF} Ratio Near 10^6," IEEE Electron Device Letters, vol. PP, pp. 1-3, 2012.

[10] Z. Guangle, et al., "Vertical InGaAs/InP Tunnel FETs With Tunneling Normal to the Gate," IEEE Electron Device Letters, vol. 32, pp. 1516-1518, 2011.

Chapter 8

Conclusions, Future Work and Outlook

8.1 General Conclusions

The general purpose of this thesis dissertation was to propose a critical study of the Tunnel FET as a potential replacement for conventional MOSFETs beyond the 14nm node and offer insights about its possible manufacturability. We would like to remember that the carrier injection mechanism in TFETs is based on band-to-band tunneling, a quantum mechanical phenomenon which is totally different from the classical thermal injection over a barrier characteristic of a conventional MOSFET. The novelty of this device brings a lot of incognita's about the impact of several processing and geometrical parameters and the choice of a suitable device architecture.

The approach of this study was mainly experimental and was focusing on a particular configuration: the lateral TFET. It consists of a gated p-i-n diode and represents the basic embodiment of a TFET. TFETs have been fabricated according to a CMOS compatible process flow and two architectures have been implemented:

- FinFET on SOI. A mature technology used for a fast learning for homojunction Si TFET. The use of SOI substrate is mandatory for a good device isolation and to reduce junction leakage currents [1].
- Vertical nanowires. An exploratory architecture suitable to implement heterojunction TFETs which should boost the tunneling current without increasing the off-current. In particular, SiGe-source TFETs with various Germanium concentrations were implemented.

These two architectures were chosen for their excellent control of the electrostatic potential thanks to the good coupling of the gate electrode to the channel. The transition from a homojunction device to a heterojunction device is necessary to reduce the effective tunneling bandgap without increasing the off-current [2]. In fact, the off-state current is limited by the large energy band gap of the channel material.

In chapter 2, simple semi-classical models such as the Kane’s model [3] qualitatively predict that the band to band tunneling generation rate (proportional to the tunneling current) is strongly dependent on the electric field at the tunneling junction, the tunneling effective masses and the energy band gap, or effective energy band gap. As a consequence, good electrostatic control and low energy bandgap materials are needed to implement TFETs. We have also pointed out the difficulty to describe quantitatively this quantum mechanical mechanism with semi-classical models. The necessity of non local models together with an accurate calibration of the model parameters is mandatory. Nevertheless, an ad-hoc quantum mechanical approach [4-10] is suggested to verify the limitations of the semi classical approach. For instance, it has been theoretically demonstrated that effects such as the quantum confinement can strongly impact the tunneling behavior [11].

In chapter 2, we have also introduced transport mechanisms other than band-to-band tunneling which are often not included during the simulation but are present in experimental devices. In particular, the presence of TAT represents one of the major bottlenecks to achieve a sub-60mV/dec swing in TFETs. This phenomenon is more severe in vertical devices because of the severe damage during the nanowire patterning, the lower source doping and the presence of large amounts of interface traps (verify only in case of Si TFET). For this reason, it is of utmost importance to understand not only the band-to-band tunneling but also all the possible leakage mechanisms present which can affect the device performance and how to suppress them.

In chapter 3, the design and fabrication of TFETs have been described. This chapter was divided in two main parts because of the two different architectures studied: (1) the finFET and (2) vertical nanowire.

The design and process flow for finFET-based Si TFETs is very similar to the process flow used for conventional MOSFETs. Contrary to the MOSFET case where only one type of doping is used to implant both the source and drain regions, the TFET requires a mask modification to implement two types of doping at the source and drain. The two masks NPLUS and PPLUS join in the middle of the gate providing an implant to the poly-Si gate. For this reason, we could not verify the device scalability when the gate length is below 150nm. To avoid silicide encroachment, additional spacers have been deposited after the HDD implants. A standard BEOL processing allows the electrical characterization in industrial probing stations. Furthermore, a possible implementation of planar heterojunction TFET was explored. The integration of heterojunction is realized by a Si etch back followed by a SEG of B-doped $Si_{0.75}Ge_{0.25}$ in the source, similarly to the implementation of SiGe stressors used in CMOS technology [12]. It has been observed that there is either no growth nor uncontrolled growth in case of overlap with the gate because of contamination issues due to the exposure of the gate stack to the pre-cleaning prior the SEG (chapter 5). As a consequence, we suggested moving toward vertical architectures when implementing heterojunction TFETs.

As explained earlier, vertical nanowire-based TFETs have been designed and fabricated to implement heterojunction TFET. Contrary to the finFET approach, the integration of vertical nanowires as TFETs required the complete development of a new process flow and the design of a dedicated maskset. For this, several learning cycles were necessary before a finalized process flow was reached. The main challenges were:

- *nanowire patterning*. To achieve a device with 150nm gate length and proper top and bottom isolation, the height of the etched nanowire was around 400nm. The tuning of the dry etch was critical to achieve a nearly straight profile of the nanowire sidewalls.
- *gate alignment strategy for epi grown source*. The gate alignment is critical because it depends on the uniformity of the CMP step and the control of the wet etch-back of the oxide which covers the top of the nanowire.
- *source replacement* for heterojunction TFET.

In chapter 4, an overview of several electrical characterization techniques used in this study have been described. In particular, the extraction of device parameters such as EOT or D_{it} has been explained shortly. Later, an overview of the TFET operation modes is presented with an emphasis on the method used to identify the band-to-band tunneling over other transport mechanisms. Some interesting features of the TFET are its weak temperature dependence and its insensitivity to gate length scaling. In case of very short gates, we expect an increase of the off-state current due to the reverse biased diode leakage which might affect the I_{ON}/I_{OFF} ratio. This effect is not studied in this work because of limitations in the device fabrication of very short gates. The transfer and the output characteristics of TFETs have been shown and discussed for the lateral TFET configuration focusing on the main difference with the conventional MOSFETs. In particular, the difficulty to find a proper metric to benchmark the TFET is discussed. It is indeed not possible to define a clear threshold voltage in TFETs similarly to MOSFETs.

In chapter 5, the performance of homojunction Si TFETs has been analyzed. The POR device was characterized by an average slope of 100mV/dec and an I_{ON}/I_{OFF} ratio around 7 decades considering a V_{DD} range of 1V. This poor performance is related to the large energy bandgap of silicon. Nevertheless, the impact of several geometrical and process parameters to improve the tunneling performance has been investigated. The main findings can be listed as follows:

- *impact of the dimensionality.* A clear improvement of the device performance has been observed in case of narrow fins. The best gate coupling together with a more uniform doping of the fin is able to reduce the tunneling onset. This beneficial shift allows the device to be more robust against TAT and a better subthreshold swing is recorded at low gate voltages
- *doping profile.* The impact of the doping concentration and profile was studied by implementing different implantation conditions during the LDD. The limited impact on the device performance is explained by the presence of diffusion due to the thermal activation steps. As a

consequence, we suggest that an in-situ doped source is necessary to achieve the required abruptness.

- *oxide thickness*. TCAD simulation predicts that the scaling of EOT has a great impact on the TFET performance. In our experiments the EOT was varied from 2.2 nm to 1.5 nm using 2 nm or 4 nm of HfO_2. The thinner oxide showed a performance improvement but the gate leakage current increased as well and further EOT scaling was not possible. As a consequence, a boost of on-current is required before scaling the EOT. This boost of the on-current requires the implementation of a new source material.
- *silicide Engineering*. A novel configuration has been introduced to boost the tunneling current, the DS Schottky TFET. In this case, the silicide is encroaching under the gate creating a highly doped dopant segregated region which is supposed to boost the tunneling current. Despite of the large increase of the on-current and reduction of the tunneling onset, this device has intrinsic limitations in achieving a stable sub-60mV/dec regime.

In chapter 6, the performance of vertical nanowires is discussed. At the first stage, the homojunction Si TFETs have been implemented to study the impact of geometrical parameters which differ from a planar integration scheme, such as the gate alignment in case of an epi-grown source. Contrary to the finFETs, the vertical nanowires have more interface traps and larger interface roughness which degrade the device performance. Nevertheless, the epi-grown source and a large source-to-gate overlap have helped device performance. In addition, the presence of a bottom isolation of 50nm is able to suppress the ambipolarity in case of low thermal budget processing which limits the up-diffusion of the dopants of the drain region into the channel. In this chapter, a preliminary study of heterojunction TFETs is discussed as well. In particular, SiGe with Germanium concentration up to 45% has been implemented by growing a thin layer with the high Germanium concentration followed by a layer with a lower Germanium concentration. In this case, a great improvement of the on-current is observed together with a reduction of the tunneling onset. As observed in the narrow fins, the reduction of the tunneling onset reduces the impact of the TAT. Nevertheless, TAT is still dominant at low gate voltage hiding the possible sub-60mV/dec swing. Further studies are required to assess the impact of higher Ge concentrations.

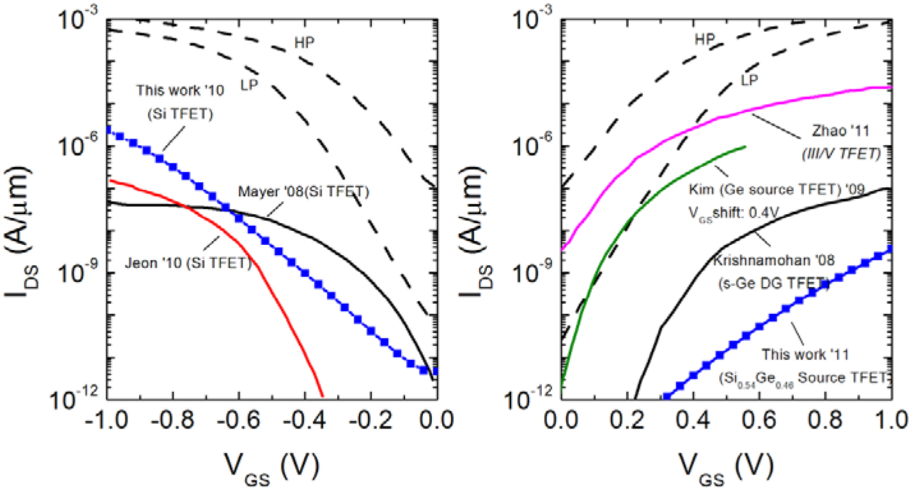


Figure 8.1: Comparison of published TFET I_{DS}-V_{GS} characteristics for experimental p-channel (a) [13-14] and n-channel (b) [15-19] TFETs extracted at V_{DD} of 1V. The dash lines are measured characteristics for HP and LP 32-nm node MOSFET technology [20]. The blue squares represent the experimental data obtained during this work. All the measurements were reported at room temperature.

An overview about recent state-of-the-art TFET performance [13, 15, 17, 19, 21-23] together with the best results obtained in this work is shown in Figure 8.1. The TFET characteristics are compared with 32-nm MOSFET designed for High Power (HP) or ultra Low Power (LP) applications. The high power applications MOSFET are designed to reduce the clock delay which is proportional to $C_{eff}*V_{DD}/I_{Dsat}$ and represent the upper boundary. The ultra low power, or Low Standby Power, applications are designed to minimize the static power dissipation and represent the lower boundary. So far, it is clear that the Si- and Ge-based TFET are not in a current range of interest for either HP or LP applications as the on current is well below 100µA/µm and the supply voltage is still close to 1V. In addition, the experimental data in this work agree with the best results available in literature for equivalent material system. On the other end, the introduction of new materials together with the implementation of heterojunction TFET [2] may provide the required boost in on-current to place as a promising candidate for sub 0.5 V logic operation.

Chapter 7 anticipates the outlook of this work. Here, a new device architecture, the hybrid TFET, is proposed since we believe that the lack of a 60mV/dec swing is related to some limitations of the lateral TFET configuration in which the point tunneling is the dominant tunneling component. The hybrid TFET consists of a gated p-i-n diode where the gate is fully overlapping the source and an additional pocket similar to the intrinsic channel is added between the gate stack and the source. The preliminary simulation experiments shows on-currents of 100µA/µm at V_{DD} of 0.5V with Ge source and a gate length of 45nm. In this case, the subthreshold swing is below the kT-limit for more than 7 decades. The weaknesses of this device are: the degradation of the tunneling onset when the gate does not fully overlap the source and the strong dependence of the tunneling current on the thickness of the pocket. These parameters can be controlled during the fabrication of the device, in particular the thickness of the pocket because of the selectivity of the epitaxial growth of this layer. In conclusion, a process flow has been designed but the results are not available yet and, hence, they are not presented in this work.

8.2 Future Work and Outlook

In this thesis, a in depth study of the potentialities of TFETs has been presented but there is certainly a lot of room for improvements. Because of the lack of time and delays in the processing of experimental devices, a certain number of studies could not be performed and three of them should receive more consideration in the future:

- Impact of the Si capping layer in vertical heterojunction TFETs. The growth on the nanowire sidewalls needs to be controlled and the surface roughness should be improved.
- Decouple the point and line tunneling in heterojunction TFETs by changing the doping of the source or the gate-to-source overlap or inserting a wide bandgap pocket between the source and the channel.
- Ge-source TFETs should boost the tunneling current but the source replacement approach is needed because of the

impossibility to grow a strained Germanium layer on a blanket wafer. This approach consists of replacing a poly-SiGe dummy source with a thin layer of Ge followed by the deposition of amorphous germanium to fill the top region. If the processing is successful we should be able to show whether nTFETs with Ge-source can replace conventional nFET.

When this PhD work started the question was: what is the ideal tunnel FET configuration? At the beginning, the answer was: the one with the lower bandgap, the most abrupt junction and the highest doping concentrations. Is it still true? After four years, our knowledge about TFETs substantially improved and we are now able to answer that question in a different way. For this reason, we would like to give some recommendations for future research in this field.

What we are looking for is basically: the best tunneling barrier and the best TFET architecture. Regarding the tunneling barrier, it is known that the only materials where bandgap engineering is possible are III/V materials. Among the III/V compounds, only the one with compatible lattice parameters can be chosen to achieve a good quality of the heterojunction interface. The implementation of broken bandgap heterojunctions is very promising [24]. In this way, it is possible to implement materials with a bandgap which is not too small and does not degrade the density of states. The doping does not need to be maximized, ideally the Fermi level should be aligned with the valence band (in case of n-type operation) to make the tunneling more efficient. My suggestion is to define two research paths: a material approach and an integration approach. The former should aim to characterize the properties of new broken bandgap materials such as InAs/GaSb with emphasis on the interface quality and properties. In fact, we do not know how abrupt will be the interface and this aspect is not taken into account in the simulations, although it is very important experimentally. Regarding the integration approach, we suggest to start some learning on the passivation of III/V materials and gate stack options. Recent publications show good results with InGaAs system [25] which is a well known III/V compound.

Regarding the best TFET architecture, we have already discussed about it in chapter 7: we are looking for an architecture where the tunneling is normal to the gate to maximize the control of the gate over the tunneling barrier. The hybrid TFET concept presented in chapter 7 needs to be implemented and evaluated. Nevertheless, we should be aware of the scalability of this solution since TFET should replace MOSFET beyond 14nm. Certainly, the planar approach proposed in chapter 7 is not extremely scalable. For this reason, vertical architecture is more suitable for extreme scaling since the restriction on the gate length is more relaxed. A vertical implementation of this concept has been proposed in a pending patent.

References

[1] C. Woo Young, *et al.*, "70-nm impact-ionization metal-oxide-semiconductor (I-MOS) devices integrated with tunneling field-effect transistors (TFETs)," in *IEEE International Electron Devices Meeting* 2005, pp. 955-958.

[2] A. S. Verhulst, *et al.*, "Boosting the on-current of a n-channel nanowire tunnel field-effect transistor by source material optimization," *Journal of Applied Physics*, vol. 104, pp. 064514-10, 2008.

[3] E. O. Kane, "Theory of Tunneling," *Journal of Applied Physics*, vol. 32, pp. 83-91, 1961.

[4] W. G. Vandenberghe, *et al.*, "Two-dimensional quantum mechanical modeling of band-to-band tunneling in indirect semiconductors," in *IEEE International Electron Devices Meeting* 2011, pp. 5.3.1-5.3.4.

[5] S. O. Koswatta, *et al.*, "Nonequilibrium Green's Function Treatment of Phonon Scattering in Carbon-Nanotube Transistors," *IEEE Transactions on Electron Devices*, vol. 54, pp. 2339-2351, 2007.

[6] R. C. Bowen, *et al.*, "Efficient I-V simulation of quantum devices using full bandstructure models," in *IEEE/Cornell Conference on Advanced Concepts in High Speed Semiconductor Devices and Circuits*, 1995, pp. 435-439.

[7] G. Klimeck and M. Luisier, "From NEMO1D and NEMO3D to OMEN: Moving towards atomistic 3-D quantum transport in nano-scale semiconductors," in *IEEE International Electron Devices Meeting* 2008, pp. 1-4.

[8] G. Klimeck, *et al.*, "Atomistic Simulation of Realistically Sized Nanodevices Using NEMO 3-D Part I: Models and Benchmarks," *IEEE Transactions on Electron Devices*, vol. 54, pp. 2079-2089, 2007.

[9] G. Klimeck, *et al.*, "Atomistic Simulation of Realistically Sized Nanodevices Using NEMO 3-D Part II: Applications," *IEEE Transactions on Electron Devices*, vol. 54, pp. 2090-2099, 2007.

[10] S. Steiger, *et al.*, "The nanoelectronic modeling tool NEMO 5: Capabilities, validation, and application to Sb-heterostructures," in *Device Research Conference Digest*, 2011, pp. 23-26.

[11] W. G. Vandenberghe, *et al.*, "Impact of field-induced quantum confinement in tunneling field-effect devices," *Applied Physics Letters*, vol. 98, pp. 143503-3, 2011.

[12] Z. Da, *et al.*, "Embedded Source/Drain SiGe Stressor Devices on SOI: Integrations, Performance, and Analyses," *IEEE Transactions on Electron Devices*, vol. 53, pp. 3020-3024, 2006.

[13] F. Mayer, *et al.*, "Impact of SOI, $Si_{1-x}Ge_xOI$ and GeOI substrates on CMOS compatible Tunnel FET performance," in *IEEE International Electron Devices Meeting* New York, 2008, pp. 163-166.

[14] K. Jeon, *et al.*, "Si tunnel transistors with a novel silicided source and 46mV/dec swing," in *Symposium on VLSI Technology*, 2010.

[15] W. Y. Choi, *et al.*, "Tunneling Field-Effect Transistors (TFETs) With Subthreshold Swing (SS) Less Than 60 mV/dec," *IEEE Electron Device Letters*, vol. 28, pp. 743-745, 2007.

[16] S. H. Kim, *et al.*, "Germanium-Source Tunnel Field Effect Transistors with Record High I_{ON}/I_{OFF}," in *Symposium on VLSI Technology*, Tokyo, 2009, pp. 178-179.

[17] S. Mookerjea, *et al.*, "Experimental Demonstration of 100nm Channel Length $In_{0.53}Ga_{0.47}As$-based Vertical Inter-band Tunnel Field Effect Transistors (TFETs) for Ultra Low-Power Logic and SRAM Applications," in *IEEE International Electron Devices Meeting*, San Francisco, CA,, 2009, pp. 949-951

[18] H. Zhao, *et al.*, "InGaAs Tunneling Field-Effect-Transistors With Atomic-Layer-Deposited Gate Oxides," *IEEE Transactions on Electron Devices*, vol. 58, pp. 2990-2995, 2011.

[19] T. Krishnamohan, *et al.*, "Double-gate strained-Ge heterostructure tunneling FET (TFET) With record high drive currents and < 60 mV/dec subthreshold slope," in *IEEE International Electron Devices Meeting* San Francisco, CA,, 2008, p. 947.

[20] C. H. Jan, *et al.*, "A 32nm SoC platform technology with 2nd generation high-k/metal gate transistors optimized for ultra low power, high performance, and high density product applications," in *IEEE International Electron Devices Meeting*, 2009, pp. 1-4.

[21] D. Leonelli, *et al.*, "Drive current enhancement in p-tunnel FETs by optimization of the process conditions," *Solid-State Electronics*, vol. 65–66, pp. 28-32, 2011.

[22] A. Vandooren, *et al.*, "Impact of process and geometrical parameters on the electrical characteristics of vertical nanowire silicon n-TFETs," *Solid-State Electronics,* vol. 72, pp. 82-87, 2012.

[23] V. Nagavarapu, *et al.*, "The tunnel source (PNPN) n-MOSFET: A novel high performance transistor," *IEEE Transactions on Electron Devices*, vol. 55, pp. 1013-1019, 2008.

[24] M. Luisier and G. Klimeck, "Performance comparisons of tunneling field-effect transistors made of InSb, Carbon, and GaSb-InAs broken gap heterostructures," in *IEEE International Electron Devices Meeting* 2009, pp. 1-4.

[25] G. Dewey, *et al.*, "Fabrication, characterization, and physics of III/V heterojunction tunneling Field Effect Transistors (H-TFET) for steep sub-threshold swing," in *IEEE International Electron Devices Meeting* 2011, pp. 33.6.1-33.6.4.

Appendix

Physical Characterization Methods

During device processing, top SEM inspection of the device is routinely performed in order to verify the general morphology of the device. However, to obtain more detailed information about the morphology of the device such as the uniformity and composition of a silicide or dopant profiles, additional techniques, such as Transmission Electron Microscope (TEM), Energy-Dispersive X-ray Spectroscopy (EDS) or secondary ion mass spectrometry (SIMS) are required. In this work, several techniques have been used and a short description of each of them is presented in this section.

Transmission Electron Microscopy (TEM)

In a Transmission Electron Microscope (TEM), high energy electrons are transmitted through a thin foil of material. The purpose is to form an image of the sample structure or to make local chemical analysis of the material. The technique has sub-nanometer lateral resolution for imaging and nanometer resolution for chemical analysis. As very thin samples are needed, special preparation of the specimens for TEM analysis is needed. In this work, the TEM was mainly used to assess the quality and physical dimensions of the gate stack, observe implantation damages and other morphological features (Figure 1).

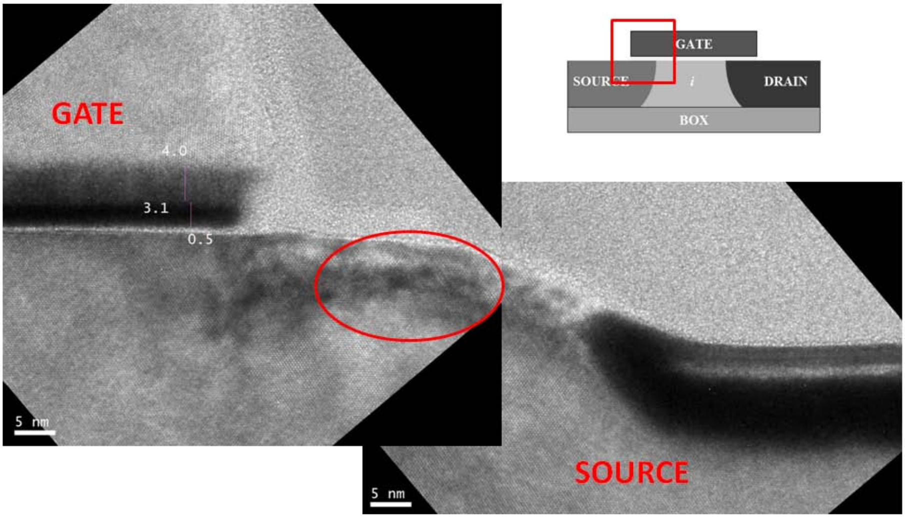


Figure 1: TEM Cross sectional view of a fabricated TFET (W_{fin} 500nm). A quantification of the gate stack is illustrated together with the strain Si present under the nitride spacer. The inset shows the region shown in the TEM image.

Another imaging technique often used is the High Angle Annular Dark Field (HAADF) Scanning Transmission Electron Microscopy (STEM) able to generate a Z-contrast imaging without diffraction effects. This contrast image has an intensity that is approximately proportional to the square of the atomic number, Z^2 and is therefore strongly composition dependent. For instance, along the [110] direction the bright spots correspond to two atomic columns for Silicon. Another possibility is to detect medium angle scattered electrons (called dark field STEM, DF-STEM). In that case diffraction effects are important and lattice defects will show up more clearly.

Another technique available with TEM is the EDS which measures the energy distribution of X-rays generated by interaction of the electron beam with the thin foil. It allows qualitative/quantitative analysis of the elements present in the material at selected points or along a line on the specimen. In principle all elements heavier than B can be measured. Peak overlaps may limit the sensitivity depending on the combination of elements present. To have a good spatial resolution, EDS is generally acquired in STEM mode. EDS is useful to characterize the composition of a metal gate, the type of silicide or the presence of contaminants in the device.

Spreading Resistance Profiling (SRP)

The Spreading Resistance Probe (SRP) is used to obtain resistivity depth profiles on Silicon or Germanium doped samples (with determination of the electrical junction depths). The carrier depth profiles can be computed based on specific mobility models. The sample needs to be cleaved and the beveled surface is measured as illustrated in

Figure 2. Contrary to the SSRM, the measurement cannot be executed on the fabricated device. This technique was used in this work to extract the profile of the active dopant in case of vertical devices and analyze the impact on the profile of different annealing steps. This technique is characterized by a simple sample preparation and a short measurement time. Unfortunately, there are issues in presence of junction due to carrier depletion effects.

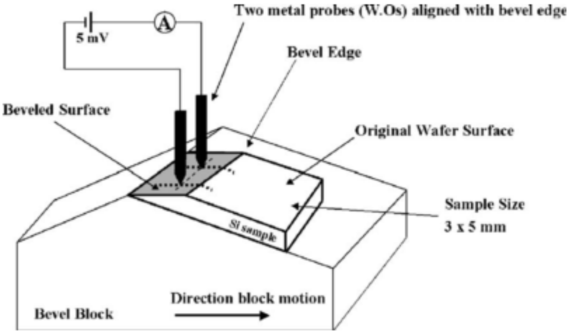


Figure 2: Schematic of the apparatus for SRP

Secondary Ion Mass Spectrometry (SIMS)

Secondary Ion Mass Spectrometry (SIMS) is suited for the detection of trace elements, such as dopants or impurities, along the depth of a sample with very high sensitivity. SIMS results can be quantified in concentration only if the trace element is present in low concentration (less than few percent) and if the matrix is known, for instance: Si, SiGe, Ge, GaAs, InGaAs, InP, GaN.

Since depth scales are determined using crater depth measurements, a constant sputter rate need to be considered during the whole SIMS depth profile. For this reason, SIMS is suitable for single layer samples and the spatial accuracy is not at the nanometer scale as in SSRM. Contrary to the SSRM and SRP, SIMS cannot distinguish between dopants which are electrically active and not.

Scanning Spreading Resistance Microscopy (SSRM)

Scanning spreading resistance microscopy (SSRM) is used to obtain information on the two-dimensional carrier distribution in semiconductor materials with nm-spatial resolution. SSRM uses a conductive probe (diamond) to evaluate the spreading resistance of the material in between the probe and a back contact (Figure 3). This resistance can be converted to an active doping concentration after a proper calibration.

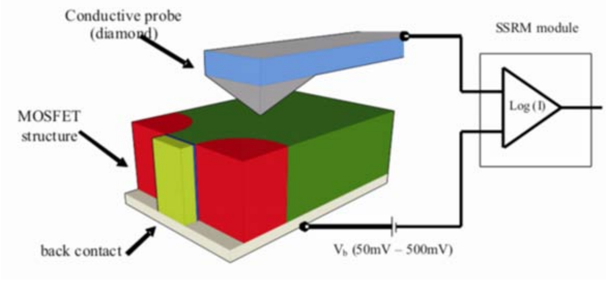


Figure 3: Schematic of the apparatus for SSRM.

SSRM is very sensitive to dopant concentration variations and outperforms all other 2D-techniques such as SRP and SIMS. Within one structure, extremely small variations of resistivity (10-20%) can be monitored and interpreted quantitatively. However small variations in doping concentrations from wafers to wafers or between different chips on a single wafer cannot be monitored (<20%). SSRM is currently a mature technique on silicon (Si). Similarly to SRP, SSRM benefits in theory from an extremely large dynamic range (10^{14} to 10^{21} at/cm^3). The SSRM is extremely important for the carrier profiling of vertical nanowire because it is possible to probe directly on the measured device in case of 200nm wide nanowire.

List of Publications

Patents

- R. Rooyackers, **D. Leonelli**, A. Vandooren, A. Verhulst, R. Loo, S. De Gendt. Method of Manufacturing a Vertical TFET. *EP2378557A1/US2011-0253981A1*
- **D. Leonelli**, A. Vandooren, M. Heyns, C. Huyghebaert. Hybrid TFET. *Provisional filed in May 2011*
- M. Heyns, C. Huyghebaert, A. Verhulst, **D. Leonelli**, R. Rooyackers, W. Dehaene. Tunnel Transistor, logical gate comprising the transistor, static random access memory using the logical gate and method for making such a tunnel transistor. *Provisional filed in July 2011*

Publications in journals with peer review

[1] Verhulst, B. Soree, **D. Leonelli**, W. Vandenberghe and G. Groeseneken, “Modeling the single-gate, double-gate and gate-all-around tunnel field-effect transistor”, Journal of Applied Physics, Vol. 107 (2) 024518, 2010.

[2] **D. Leonelli**, A. Vandooren, R. Rooyackers, A. Verhulst, S. De Gendt, M. Heyns and G. Groeseneken, “Performance enhancement for multi gate tunneling field effect transistors by scaling the fin-width”, Japanese Journal of Applied Physics, Vol. 49 (4) 04DC10, 2010.

[3] A. Schulze, T. Hantschel, P. Eyben, A. Verhulst, R. Rooyackers, A. Vandooren, J. Mody, A. Nazir, **D. Leonelli** and W. Vandervorst “Observation of diameter dependent carrier distribution in nanowire-based transistors”, Nanotechnology, Vol. 22 (18), 2011.

[4] **D. Leonelli**, A. Vandooren, R. Rooyackers, A. Verhulst, S. De Gendt, M. Heyns and G. Groeseneken “Silicide Engineering to Boost Si Tunnel Transistor Drive Current”. Japanese Journal of Applied Physics, Vol.50, 04DC05, 2011.

[5] **D. Leonelli,** A. Vandooren, R. Rooyackers, S. De Gendt, M. Heyns, and G. Groeseneken. “Drive Current Enhancement in p-Tunnel FETs by

Optimization of the Process Conditions". Solid State Electronics, Vol. 65-66, pp. 28-32, 2011.

[6] A. Verhulst, **D. Leonelli**, R. Rooyackers and G. Groeseneken, "Drain voltage dependent analytical model of tunnel field-effect transistors", Journal of Applied Physics, Vol. 110 (2), 024510, 2011.

[7] A. Vandooren, **D. Leonelli**, R. Rooyackers, K. Arstila, G. Groeseneken and C. Huyghebaert, "Impact of process and geometrical parameters on the electrical characteristics of vertical nanowire silicon n-TFETs", Solid State Electronics, Vol. 72, pp. 82-87, 2012.

[8] P. G. Der Agopian, M. D. V. Martino, S. G. S. Filho, J. A. Martino, R. Rooyackers, **D. Leonelli** and C. Claeys, "Temperature impact on the tunnel FET off-state current components", Solid State Electronics, (accepted), 2012.

[9] K.-H. Kao, A. Verhulst, W. Vandenberghe, B. Soree, W. magnus, **D. Leonelli**, G. Groeseneken and K. De Meyer, "Optimization of gate-on-source-only Tunnel FETs with counter-doped pockets", Transactions on Electronic Devices, (accepted), 2012.

Conference Proceedings and Abstracts

[1] A. Vandooren, R. Rooyackers, **D. Leonelli**, F. Iacopi, S. De Gendt, A. Verhulst, M. Heyns, E. Kunnen, D. Nguyen, M. Demand, P. Ong, W. Lee, J. Moonens, O. Richard, W. Vandenberghe and G. Groeseneken, "A 35nm diameter vertical silicon nanowire short-gate tunnel FET", in Nanotechnology Workshop, 2009. (13-14 June 2009; Kyoto, Japan.)

[2] A. Verhulst, W. Vandenberghe, **D. Leonelli**, R. Rooyackers, A. Vandooren, S. De Gendt, M. Heyns and G. Groeseneken, "Tunnel field-effect transistors for future low-power nano-electronics", in ECS Transactions; Vol. 25; Issue 7, pp.455-462, 2009.

[3] M. Heyns, F. Bellenger, G. Brammertz, M. Caymax, S. De Gendt, B. De Jaeger, A. Delabie, G. Eneman, G. Groeseneken, M. Houssa, **D. Leonelli**, D. Lin, C. Merckling, M. Meuris, J. Mitard, J. Penaud, G. Pourtois, R. Rooyackers, M. Scarrozza, E. Simoen, S. Van Elshocht, W. Vandenberghe, A. Vandooren and A. Verhulst, "Ge and III/V devices on Si for advanced CMOS", in 5th Handai Nanoscience and Nanotechnology International Symposium, 2009. (1-3 September 2009; Osaka University, Japan.)

[4] M. Heyns, F. Bellenger, G. Brammertz, M. Cantoro, M. Caymax, S. De Gendt, B. De Jaeger, A. Delabie, G. Eneman, G. Groeseneken, M. Houssa, F. Iacopi, **D. Leonelli**, D. Lin, W. Magnus, K. Martens, C. Merckling, M. Meuris, J. Mitard, J. Penaud, G. Pourtois, M. Scarrozza,

E. Simoen, S. Sioncke, B. Soree, S. Van Elshocht, W. Vandenberghe, A. Vandooren, A. Verhulst, and W. Wang, “Scaling of nanoelectronics beyond the Si roadmap”, in 1st International Workshop on Si based Nano-Electronics and –Photonics, 2009. (20-23 September 2009; Vigo, Spain.)

[5] M. Heyns, F. Bellenger, G. Brammertz, M. Caymax, S. De Gendt, B. De Jaeger, A. Delabie, G. Eneman, G. Groeseneken, M. Houssa, **D. Leonelli**, D. Lin, K. Martens, C. Merckling, M. Meuris, J. Mitard, J. Penaud, G. Pourtois, M. Scarrozza, E. Simoen, S. Van Elshocht, W. Vandenberghe, A. Vandooren, A. Verhulst, and W. Wang, “High mobility channel materials and novel devices for scaling of nanoelectronics beyond the Si roadmap”, in MRS Fall Meeting Symposium, 2009. (30 November-4 December 2009; Boston, MA, USA.)

[6] **D. Leonelli**, A. Vandooren, R. Rooyackers, A. Verhulst, S. De Gendt, M. Heyns, and G. Groeseneken, “Multiple-gate tunneling field effect transistors with sub-60mV/dec subthreshold slope”, in Extended Abstracts Solid State Devices and Materials Conference (SSDM), pp.767-768, 2009. (7-9 October 2009; Sendai, Japan.)

[7] F. Iacopi, R. Rooyackers, A. Vandooren, W. Vanherle, S. Takeuchi, A. Hikavyy, R. Loo, A. Milenin, **D. Leonelli**, K. Arstila, H. Bender, M. Caymax, S. De Gendt, and M. Heyns, “Hetero Ge/Si and $Si_{1-x}Ge_x$/Si nanowires for vertical microelectronics devices”, in 19th Symposium of MRS-Japan, 2009. (7-9 December 2009; Yokohama, Japan.)

[8] A. Verhulst, W. Vandenberghe, **D. Leonelli**, R. Rooyackers, A. Vandooren, S. De Gendt, M. Heyns and G. Groeseneken, “Tunnel field-effect transistors for future low-power nano-electronics”, in 216th ECS Meeting, pp.2402, 2009. (4-9 October 2009; Vienna, Austria.)

[9] C. Claeys, **D. Leonelli**, R. Rooyackers, A. Vandooren, A. Verhulst, M. Heyns, G. Groeseneken and S. De Gendt, “Fabrication and characterization of Si and hetero-junction tunnel field effect transistors”, in Global COE International Symposium Silicon Nano Devices in 2030: Prospects by World’s Leading Scientists, 2009. (13-14 October 2009; Tokyo, Japan.)

[10] M. Heyns, F. Bellenger, G. Brammertz, M. Caymax, M. Cantoro, S. De Gendt, B. De Jaeger, A. Delabie, G. Eneman, G. Groeseneken, G. Hellings, M. Houssa, F. Iacopi, **D. Leonelli**, D. Lin, W. Magnus, K. Martens, C. Merckling, M. Meuris, J. Mitard, J. Penaud, G. Pourtois, M. Scarrozza, E. Simoen, B. Soree, S. Van Elshocht, W. Vandenberghe, A. Vandooren, P. Vereecken, A. Verhulst and W. Wang, “Shaping the future of nanoelectronics beyond the Si roadmap with new materials and

devices”, in Optical Microlithography XXIII, pp.764003, 2010. (21-25 February 2010; San Jose, CA, USA.)

[11] A. Verhulst, W. Vandenberghe, **D. Leonelli**, R. Rooyackers, A. Vandooren, G. Pourtois, S. De Gendt, M. Heyns, and G. Groeseneken, “Boosting the on-current of Si-based tunnel field-effect transistors”, in 218th ECS Meeting, pp.1893, 2010. (10-15 October 2010; Las Vegas, NV, USA.)

[12] **D. Leonelli**, A. Vandooren, R. Rooyackers, S. De Gendt, M. Heyns and G. Groeseneken, “Optimization of tunnel FETs: Impact of gate oxide thickness, implantation and annealing conditions”, in 40th European Solid-State Device Research Conference (ESSDERC), pp.170-173, 2010. (13-17 September 2010; Sevilla, Spain.)

[13] **D. Leonelli**, A. Vandooren, R. Rooyackers, A. Verhulst, S. De Gendt, M. Heyns and G. Groeseneken, “Drive current improvement in Si tunnel field effect transistors by means of silicide engineering”, in International Conference on Solid State Devices and Materials (SSDM), pp.693-694, 2010. (22-24 September 2010; Tokyo, Japan.)

[14] A. Verhulst, W. Vandenberghe, **D. Leonelli**, R. Rooyackers, A. Vandooren, G. Pourtois, S. De Gendt, M. Heyns and G. Groeseneken, “Boosting the on-current of Si-based tunnel field-effect transistors”, in SiGe, Ge, and Related Compounds 4: Materials, Processing and Devices, pp.363-372, 2010. (10-15 October 2010; Las Vegas, NV, USA.)

[15] A. Verhulst, W. Vandenberghe, **D. Leonelli**, R. Rooyackers, A. Vandooren, J. Zhuge, K.-H. Kao, B. Soree, W. Magnus, M. Fischetti, G. Pourtois, C. Huyghebaert, R. Huang, Y. Wang, K. De Meyer, W. Dehaene, M. Heyns and G. Groeseneken, “Si-based Tunnel Field-Effect Transistors for Low-Power Nano-Electronics”, in Device Research Conference, 2011. (20-22 June 2011; Santa Barbara, USA.)

[16] C. Claeys, **D. Leonelli**, R. Rooyackers, A. Vandooren, A. Verhulst, M. Heyns, G. Groeseneken and S. De Gendt, “Trends and challenges in Si and hetero-junction tunnel field effect transistors”. ECS Transactions, Vol. 35 (5), 2011.

[17] A. Vandooren, **D. Leonelli**, R. Rooyackers, K. Arstila, G. Groeseneken and C. Huyghebaert “Electrical Results of Vertical Si N-Tunnel FETs”, in European Solid-State Device Research Conference (ESSDERC), 2011.

[18] **D. Leonelli**, A. Vandooren, R. Rooyackers, A. Verhulst, C. Huyghebaert, S. De Gendt, M. Heyns and G. Groeseneken, “Novel Architecture to Boost the Vertical Tunneling in Tunnel Field Effect Transistors”, in IEEE International SOI conference, 2011. (3-6 October 2011; Tempe, United States.)

Curriculum Vitae

Daniele Leonelli was born in Milan, Italy on February 02, 1983. He received the B.S. and M.S. degree in Physics Engineering from Politecnico di Milano, Italy in 2005 and 2007, respectively. His Master thesis was entitled “Growth of Si nanowires and novel integration architectures for advanced Si process technologies” and it was developed in collaboration with imec. Since 2008, he was working towards a Ph.D. degree at the electrical Engineering Department (ESAT), Microelectronics sensors division (MICAS) of KU Leuven, and CMOS Technology Department at imec, Belgium. The main topic of his work was the design, fabrication and characterization of Tunnel FET for ultra low power applications. From 2008 to 2012, he received a Ph.D. fellowship from KU Leuven. He was in Grenoble in September 2010 as visiting student under the supervision of Prof. Gerand Ghibaudo. In 2011, he was visiting scholar in UC Berkely, California under the guidance of Prof. Tsu-Jae King Liu and he collaborated to design new architectures for Tunnel FET. He has authored or co-authored approximately 20 technical papers for publication in journals and presentations at conferences.

Arenberg Doctoral School of Science, Engineering & Technology
Faculteit Bio-ingenieurswetenschappen
Departement Biosystemen
Onderzoeksafdeling Gentechnologie
Kasteelpark Arenberg 21 bus 2462
B-3001 Leuven

KATHOLIEKE UNIVERSITEIT LEUVEN

ASSOCIATIE K.U.LEUVEN