

Simulation of Double-Gate Silicon Tunnel FETs with a High-k Gate Dielectric

THÈSE N° 4729 (2010)

PRÉSENTÉE LE 6 JUILLET 2010

À LA FACULTÉ SCIENCES ET TECHNIQUES DE L'INGÉNIEUR
LABORATOIRE DES DISPOSITIFS NANOÉLECTRONIQUES
PROGRAMME DOCTORAL EN MICROSYSTÈMES ET MICROÉLECTRONIQUE

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Katherine BOUCART

acceptée sur proposition du jury:

Prof. J. Brugger, président du jury
Prof. M. A. Ionescu, Dr W. Riess, directeurs de thèse
Dr D. Antoniadis, rapporteur
Prof. C. Enz, rapporteur
Prof. J. Knoch, rapporteur



ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Suisse
2010

ACKNOWLEDGMENTS

Thanks first and foremost to Prof. Adrian Ionescu, who helped me find this interesting thesis topic, and who has been incredibly supportive and full of good ideas throughout. I appreciate his flexibility in letting me carry out the past few years of this thesis part-time from Zurich! It worked out well for everyone.

I would also like to thank my co-advisor, Dr. Walter Riess, for sharing his knowledge and working with me on several Tunnel FET studies and publications.

Thanks also to the jury members, Dr. Dimitri Antoniadis, Dr. Joachim Knoch, and Dr. ChristianENZ, for their comments and feedback, and to the president of the jury, Dr. Jürgen Brugger.

Being a few hours from EPFL during most of the thesis, I appreciated my contact and collaboration with other doctoral students in the group all the more. It was always enjoyable to work with Livio Lattanzio, Giovanni Salvatore, Mohammad Najmzadeh, and Luca De Michielis. Thanks also to former students, especially Vincent Pott and Kirsten Moselund, who helped me when I first arrived in the group in 2004.

Thanks to Marie Halm, Isabelle Buzzi, and Karin Jaymes, who helped with many, many bureaucratic details, and were always a pleasure to interact with.

Joseph Guzzardi and Raymond Sutter helped me enormously with computer and software issues. Since this thesis is based upon simulations in Silvaco Atlas, I must also thank my contacts at Silvaco, especially Alexandre Ferron, who was instrumental in helping me get the models working at the beginning and introducing me to the first non-local band-to-band tunneling model in Atlas. Thanks also to Ahmed Nejm at Silvaco who has helped me more recently.

Thanks to my family and friends, in particular Diana, who has been there for me through ups and downs. Thanks to my mom for always believing that I can do anything, even when she has no idea what I'm doing. Finally, I couldn't have finished this thesis without the support of my husband Julien, who even took time off work a few times to stay with the children so that I could go to meetings, workshops, and conferences, and who always believed in me.

TABLE OF CONTENTS

Table of Contents	1
Abstract	5
Résumé	7
List of Symbols	9
Chapter 1: Thesis overview	11
1.1 Objectives	12
1.2 Outline	12
Chapter 2: Introduction	15
2.1 The limitations faced by CMOS: the power crisis	16
2.2 Possible solutions to the power crisis	19
2.2.1 Circuit-level solutions	19
2.2.2 Device-level solutions	20
2.3 Small swing devices	22
2.4 Introduction to the Tunnel FET	23
2.4.1 Tunnel FET structure and operation	24
2.4.2 Band-to-band tunneling transmission	25
2.4.3 Subthreshold swing in Tunnel FETs	28
2.4.4 Tunnel FET temperature characteristics	30
2.5 Silvaco Atlas models	31
2.6 History and state-of-the-art of the tunneling transistor	31
2.7 Conclusion	36
2.8 Bibliography	37
Chapter 3: Tunnel FET Optimization	41
3.1 Tunnel FET parameter optimization	42
3.1.1 Device structure and operation	42
3.1.2 Double gate	42
3.1.3 Doping levels in the source, drain, and intrinsic regions	43
3.1.4 High-k gate dielectric	45
3.1.5 Thin film structure	47
3.2 2D Tunnel FET simulations	50
3.3 One final optimization: the band gap	53
3.4 Conclusion	59
3.5 Bibliography	60
Chapter 4: Threshold Voltages for Tunnel FETs	61
4.1 Why a new definition for Tunnel FET threshold voltage?	62
4.2 Simulation: device structure	62

4.3	Threshold voltages of Tunnel FETs	62
4.3.1	Gate threshold voltage, V_{TG}	62
4.3.2	Drain threshold voltage, V_{TD}	65
4.4	Gate length scaling effects on Tunnel FET threshold voltages	66
4.5	Discussion of V_{DS} and V_{GS} dependence of threshold voltages	68
4.6	The negative impact of V_{TD} on logic circuit design	72
4.6.1	V_{TD} in static operation	72
4.6.2	V_{TD} in dynamic operation	73
4.7	V_{TD} extraction from experimental data	73
4.8	Conclusion	76
4.9	Bibliography	77
Chapter 5: Tunnel FET Length Scaling		79
5.1	Scaling: conventional MOSFETs vs. Tunnel FETs	80
5.2	Simulation parameters	81
5.3	Length scaling and basic device characteristics	81
5.4	Scaling study in Tunnel FETs with high-k gate stack	86
5.5	Depletion, energy bands, and the off-state	88
5.6	Length scaling trend confirmation from experimental data	91
5.7	Scaling supply voltage	92
5.8	Conclusion	93
5.9	Bibliography	95
Chapter 6: Tunnel FET Parameter Variation		97
6.1	The importance of a parameter variation study	98
6.2	Performance boosters for silicon Tunnel FETs	98
6.3	Parameter fluctuation study	100
6.3.1	Dielectric permittivity and thickness	101
6.3.2	Junction width	102
6.3.3	Silicon body thickness	104
6.3.4	Gate oxide and gate contact alignment	105
6.3.5	Device length	109
6.3.6	Band gap reduction at the tunnel junction	110
6.3.7	Intrinsic region doping	111
6.4	Discussion of parameter fluctuation results	111
6.5	Comparison with conventional silicon MOSFETs	112
6.6	Conclusion	113
6.7	Bibliography	114
Chapter 7: Conclusion and Perspectives		117
7.1	Conclusion	118
7.2	Perspectives	119
7.3	Bibliography	120
Appendix: Models, Meshing, and Calibration		121
A.1	Silvaco Atlas models	121
A.1.1	Non-local band-to-band tunneling model	121
A.1.2	Bandgap narrowing	122

A.1.3	Quantum model	122
A.2	Device structure and meshing	123
A.3	IBM diodes for calibration	124
A.4	Fitting Tunnel FET data	125
A.5	Local band-to-band tunneling models	127
A.6	Bibliography	128

ABSTRACT

The down-scaling of conventional MOSFETs has led to an impending power crisis, in which static power consumption is becoming too high. In order to improve the energy-efficiency of electronic circuits, small swing switches are interesting candidates to replace or complement the MOSFETs used today. Tunnel FETs, which are gated p-i-n diodes whose on-current arises from band-to-band tunneling, are attractive new devices for low-power applications due to their low off-current and their potential for a small subthreshold swing.

The numerical simulations presented in this thesis have been carried out using a non-local band-to-band tunneling model in Silvaco Atlas. Numerical simulations based on correct underlying models are important for emerging devices, since they can provide insights about optimization before fabrication is carried out, can aid the understanding of device physics through 1D and 2D cross sections, and can be the basis for the formation of an accurate compact model.

In general, only CMOS-compatible materials and structures have been used in the Tunnel FET designs presented here. One goal of this thesis was to stay within the framework of what is possible in standard industrial nanoelectronics cleanrooms today, without requiring processes whose mastery lies many years in the future. For this reason, the focus of this thesis is on all-silicon devices, and heterostructures that incorporate other materials are only mentioned.

In chapter three, the optimization of the static characteristics of a Tunnel FET is carried out, looking at gate structure (single or double), doping levels of each device region, gate dielectric permittivity, and silicon body thickness. A study of the reduction of the band gap at the tunnel junction is also presented, showing the resulting improvement in on-current and subthreshold swing.

Chapter four introduces a new method for threshold voltage extraction in Tunnel FETs. This method has one key advantage over the commonly-used constant current threshold voltage extraction technique: it has a physical meaning. The transconductance method, which has already been used for conventional MOSFETs, pinpoints the Tunnel FET voltage at which the transition from strong control to weak control of the tunneling energy barrier width, and therefore the on-current, takes place. This is analogous to the threshold voltage in a conventional MOSFET which marks the transition from weak inversion to strong inversion at $\phi_s=2\phi_F$. It is found that Tunnel FETs have two threshold voltages, one in relation to the gate voltage, and the second in relation to the drain voltage, and each depends on the voltage applied at the opposite terminal.

A length scaling study is carried out in chapter five, demonstrating the scaling limits of Tunnel FETs at gate lengths on the order of 10-20 nm, due to p-i-n diode leakage current that degrades the off-current. Tunnel FETs designed to have better electrostatic control of the tunnel junction by the gate can scale further before they hit this diode leakage limit at some small gate length.

Chapter six presents an additive booster strategy for Tunnel FET optimization, and then uses the resulting optimized device as the basis of a parameter variation study. Here, one parameter is varied at a time, and the effects on the important characteristics (subthreshold swing, threshold voltage, and on-current) are evaluated. The parameters requiring the most control during fabrication are identified.

Since Tunnel FETs are emerging devices, the most important future work will be to fabricate fully-optimized n- and p-type devices, and to develop accurate compact models for their incorporation into circuits.

Keywords: TFET, Tunnel FET, tunneling transistor, band-to-band tunneling, high-k dielectric, double-gate, gated p-i-n diode, subthreshold swing

RÉSUMÉ

La miniaturisation des MOSFETs traditionnels a conduit à une impasse énergétique, dans laquelle la consommation électrique en régime statique est devenue trop grande. Afin d'améliorer l'efficacité en énergie des circuits électriques, des interrupteurs à faible inverse de pente sous seuil sont des candidats intéressants pour remplacer et compléter les MOSFETs traditionnels. Des transistors à effet de champ à jonction tunnel (Tunnel FETs) sont des diodes p-i-n à grille où le courant à l'état passant provient de porteurs passant de bande à bande par effet tunnel. Ils sont des dispositifs intéressants pour des applications à faible consommation de puissance car leur courant de fuite en état de veille est faible tandis qu'ils présentent un fort potentiel pour une faible inverse de pente sous seuil.

Les simulations numériques présentées dans ce travail de thèse, ont été réalisées par l'intermédiaire du logiciel Silvaco Atlas en utilisant un modèle d'effet tunnel de bande à bande non local. Des simulations numériques basées sur des modèles exacts sont importantes pour des dispositifs émergents puisqu'ils peuvent fournir des indications cruciales pour une optimisation précédant la fabrication. La simulation peut également aider à la compréhension des mécanismes physiques à travers des coupes 1D et 2D, et peuvent constituer la base sur laquelle un modèle compact peut être fondé.

En règle générale, seuls des matériaux et structures compatibles avec la fabrication des CMOS ont été utilisés dans les dispositifs présentés dans cette thèse. En effet, un des objectifs était de rester dans le cadre de ce qui est actuellement possible dans les salles blanches industrielles, sans exiger des procédés avancés dont la maîtrise se situe plusieurs années dans le futur. Pour cette raison, l'accent de ce travail est placé sur les dispositifs tout-silicium tandis que des hétérostructures qui incorporent d'autres matériaux ne sont que mentionnés.

Au chapitre trois, les caractéristiques statiques d'un FET à effet tunnel sont optimisées en étudiant la structure de grille (simple ou double), les niveaux de dopages dans chaque région du dispositif, et l'épaisseur de la couche de silicium. Une étude de la réduction de la largeur de bande interdite au niveau de la jonction tunnel est également présentée, montrant une amélioration du courant à l'état passant et de l'inverse de pente sous seuil.

Le chapitre quatre introduit une nouvelle méthode pour l'extraction de la tension de seuil d'un FET à effet tunnel. Cette méthode présente un avantage clé sur la technique traditionnelle d'extraction de la tension de seuil à courant constant dans le sens où elle a un sens physique. La méthode de transconductance, qui a déjà été utilisée pour des MOSFETs conventionnels, identifie la tension du FET à effet tunnel où on obtient la transition entre contrôle fort et contrôle faible de la largeur énergétique de la barrière à l'effet tunnel et donc du courant à l'état passant. En ce cas, on remarque l'analogie avec la tension de seuil d'un MOSFET conventionnel qui marque la transition entre faible et forte inversion à $\phi_s = 2\phi_F$. Nous présentons que les FETs à effet tunnel ont deux tensions de seuil, l'une liée à la tension de grille, l'autre liée à la tension de drain et chacune dépendante de la tension appliquée sur le terminal opposé.

Une étude du dimensionnement des grilles est présentée au chapitre cinq, démontrant les limites de la miniaturisation des grilles pour des FETs à effet tunnel pour des largeurs de l'ordre de 10-20nm, à cause du courant de fuite de la jonction p-i-n qui détériore le courant à l'état bloqué. Des FETs à effet tunnel qui présentent un meilleur contrôle électrostatique de la jonction tunnel par la grille peuvent être miniaturisés plus avant d'atteindre cette limite liée au courant de fuite.

Le chapitre six présente une stratégie booster pour l'optimisation des FETs à effet tunnel, et ensuite le dispositif optimisé est utilisé comme base pour une étude paramétrique. Un paramètre est alors varié à la fois, puis les effets sur les caractéristiques importantes (inverse de la pente sous seuil, tension de seuil, et courant à l'état passant) sont étudiés. Les paramètres exigeant le plus grand niveau de contrôle durant la fabrication sont ainsi identifiés.

Puisque les FETs à effet tunnel sont des dispositifs émergents, le travail à suivre le plus crucial sera de fabriquer des dispositifs complètement optimisés de type n et p et de développer des modèles compacts exacts pour leur incorporation dans des circuits.

Mots-clés: TFET, FET à effet tunnel, transistor à effet tunnel, effet tunnel bande à bande, diélectrique à fort k, double grille, diode p-i-n à grille, inverse de la pente sous seuil.

LIST OF SYMBOLS

A	area (m ²)
C _{dm}	bulk depletion capacitance at threshold (F)
C _g	gate capacitance (F)
C _L	total switched capacitive load (F)
C _{ox}	gate dielectric capacitance (F)
E	electron energy (eV)
E _F	Fermi energy (eV)
E _g	band gap energy (eV)
ΔE _g	amount by which band gap reduced at tunnel junction (eV)
f	frequency (Hz)
F	electric field (V/cm)
g _{ds}	drain-source conductance (A/V)
g _m	transconductance (A/V)
ħ	Planck's constant divided by 2π (eV.s)
I _{DS}	drain-source current (A or A/μm)
I _{leak}	leakage current in the off-state (A or A/μm)
I _{on}	drain-source current in on-state (A or A/μm)
I _{off}	drain-source current in off-state (A or A/μm)
k	Boltzmann constant (eV/K)
k(x)	quantum wave vector (1/m)
L, L _g	gate length (m)
L _{intrinsic}	intrinsic region length (m)
m	body-effect coefficient
m*	effective mass (kg)
P _{dynamic}	dynamic power consumption (W)
P _{static}	static power consumption (W)
PE	potential energy (eV)
q	magnitude of the electronic charge (C)
R _{channel}	channel resistance (Ohms)
R _{on}	resistance of device in on-state (Ohms)
S	subthreshold swing (mV/decade)
S _{avg}	average S from turn-on to threshold (mV/decade)
S _{pt}	point swing at steepest point on I-V curve (mV/decade)
t _{dielectric} , t _{ox}	gate dielectric thickness (m)
t _{Si}	silicon body thickness (m)
T	temperature (K)
T _t	band-to-band tunneling transmission probability
V _{DD}	supply voltage (V)
V _{DS}	potential difference between drain and source (V)
V _{eff}	bias at the tunnel junction (V)
V _{GS}	potential difference between gate and source (V)
V _T	threshold voltage (V)
V _{TD}	drain voltage at which threshold is reached (V)
V _{TG}	gate voltage at which threshold is reached (V)
w _b	band-to-band tunneling energy barrier width (m)
w _j	doping profile junction width, from high to low doping level (m)

$\Delta\Phi$	energy range over which tunneling can take place (eV)
ϵ_{ox}	dielectric constant of gate dielectric
ϵ_{Si}	dielectric constant of silicon
κ	device scaling factor
λ	screening length, natural length, or Debye length (m)
ϕ_{F}	potential in the neutral MOSFET body (V)
ϕ_{s}	surface potential (V)

Chapter 1

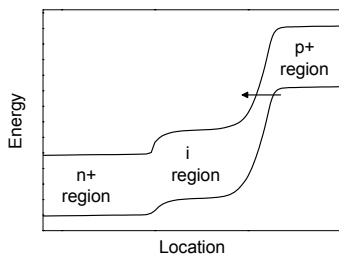
Thesis overview

1.1 Objectives

Numerical simulations of Tunnel FETs such as the ones presented in this thesis are important in several ways. They allow the investigation of the device physics, with the possibility to “see” inside a device through cross sections in 1D or 2D. Simulations also enable the thorough optimization of Tunnel FET design parameters, with the advantage that a parameter can be varied and the results obtained in half a day rather than after half a year in the cleanroom. In addition, when numerical simulations are backed up by correct physical models, the results can be useful to other researchers for the development of robust compact models. In a developing field where experimental results are still limited, these simulations can even be essential, since they allow the variation of a large number of parameters in a short amount of time. In this way, the work presented here can further our understanding of this emerging device, and can contribute to the progress made in future Tunnel FET fabrication and model development.

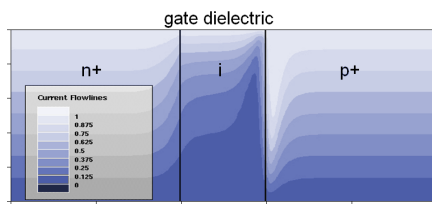
1.2 Outline

Chapter 2: Introduction



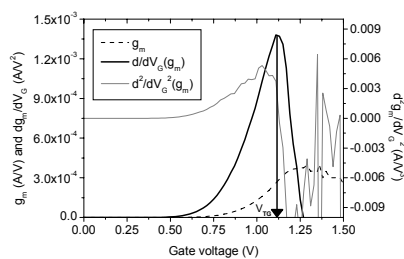
The motivation for the thesis work is presented, based upon the power crisis currently being experienced by conventional MOSFETs. Possible solutions to the crisis are mentioned, on both the circuit and device level. Small swing switches are introduced, and then in particular, the Tunnel FET is described and its operation explained. A band-to-band tunneling transmission equation derivation is shown, and the history and state-of-the-art of the Tunnel FET are given.

Chapter 3: Tunnel FET Optimization



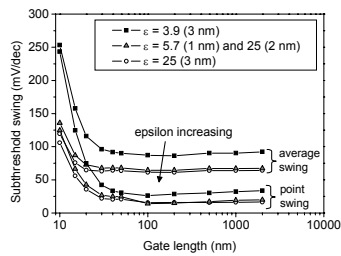
The following Tunnel FET parameters are varied in order to optimize the device: single or double gate, doping levels in the source, drain, and intrinsic regions, gate dielectric permittivity, silicon body thickness, and band gap at the tunnel junction. An asymmetrical lateral strain profile, or a heterojunction, would boost on-current by reducing the band gap only at the tunnel junction side of the device, while keeping off-current low with a large band gap at the drain side.

Chapter 4: Threshold Voltages for Tunnel FETs



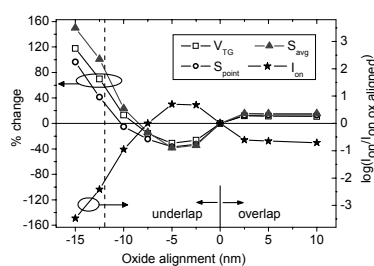
A new, more physical threshold voltage definition is proposed, in which the threshold would mark the transition between a quasi-exponential and a linear control of the drain current by an applied voltage. Two threshold voltages are identified -- one associated with the gate, and the second with the drain, and their interdependence is described. The deleterious effects of V_{TD} for circuits are mentioned, and then V_{TD} is extracted from measured Tunnel FET data.

Chapter 5: Tunnel FET Length Scaling



A detailed scaling length study is carried out on Tunnel FETs without varying other parameters simultaneously, in order to isolate and understand the effects of length scaling alone. The increased p-i-n leakage of short devices can be understood through a careful investigation of the band diagrams and depletion region width. The scaling of supply voltage is also briefly explored, and conclusions are drawn about what parameters should be simultaneously down-scaled in order to maintain good characteristics.

Chapter 6: Tunnel FET Parameter Variation



The additive booster technique of Tunnel FET optimization is explained and carried out on unoptimized devices. Then the effects of parameter fluctuations on Tunnel FET characteristics are studied, focusing specifically on gate threshold voltage, point and average subthreshold swing, and on-current. The parameters for which the characteristics are the most sensitive are identified as those requiring the tightest control during device design and fabrication.

Chapter 7: Conclusion and Perspectives

This chapter summarizes the material presented in the thesis, and gives suggestions for areas which need more study and investigation in the future. Perspectives are given, both related directly to this thesis work, and more generally.

Chapter 2

Introduction

This chapter presents the motivation for this thesis, and more generally for small swing devices, by explaining the power crisis currently faced by conventional MOSFETs, due to their ever-increasing static power consumption. The reasons behind this crisis are explained, and then some currently-used solutions are presented, both those on the circuit level and the device level. Small swing switches are introduced as the best device-level option, and then the most widely-researched examples are shown -- IMOS, MEM and NEM switches, and then finally, Tunnel FETs.

The introduction to the Tunnel FET follows, including its structure and operation. A derivation of a band-to-band tunneling transmission equation based on the WKB approximation is carried out, and then the non-constant subthreshold swing of the Tunnel FET is illustrated, and some definitions are given. The temperature-dependence of Tunnel FET characteristics is shown, and then the most important models used for the simulations in this thesis are presented. Finally, the history and state-of-the-art of tunneling transistors are reported.

2.1 The limitations faced by CMOS: the power crisis

The topic of this thesis is Tunnel FETs, but in order to understand why Tunnel FETs are interesting semiconductor switches, it is necessary to understand what is wrong with conventional MOSFETs. In order to understand that, we need to begin with an explanation of Dennard's scaling rules.

In 1974, R. Dennard, et al. published an article which has become very famous in the semiconductor device community, about how to scale a MOSFET while keeping the electric fields inside the device unchanged [1]. He recommended that all device dimensions be scaled by $1/\kappa$, while the doping of the source and drain regions should increase by a factor of κ . Applied voltages should also be scaled by $1/\kappa$. These rules have been roughly followed ever since, until rather recently.

The reason for which Dennard's scaling rules no longer work as well as they did in the past can be seen in Fig. 2.1, which shows the scaling trend from the 1.4 μm node to the 65 nm node. While the supply voltage V_{DD} decreased to about 20% of its original value, the threshold voltage V_T only went down to approximately half of its starting value. That threshold voltage decrease did not happen as a natural result of Dennard scaling. It had to come about in other ways, such as changing the doping of the channel region under the gate. Since the electric fields inside a MOSFET stay nearly constant when the scaling rules are followed correctly, the threshold voltage stays nearly constant as well, unless other changes are made.

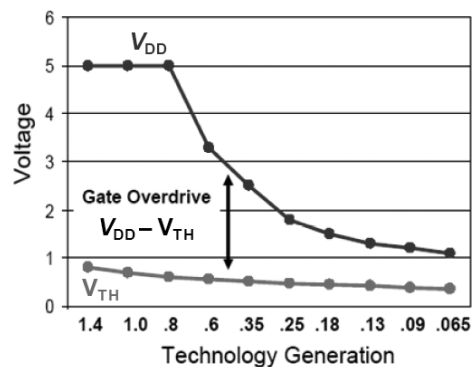


Figure 2.1: The trend of supply voltage and threshold voltage scaling vs. technology generation. V_{DD} decreases with device dimensions, but V_T does not. From [2].

The most important consequence of V_{DD} reducing during device scaling while V_T reduces significantly less, is that the gate overdrive, also shown in Fig. 2.1, goes down. When gate overdrive decreases, on-current decreases, which negatively affects device performance, the I_{on}/I_{off} ratio, and dynamic speed ($C_g V_{DD}/I_{on}$). There are two possible solutions to this problem of needing a high gate overdrive: either V_{DD} can stay higher than it should with constant field scaling, or V_T can be scaled down more aggressively. Both of these options, and their repercussions, will be discussed.

Fig. 2.2 shows that the formerly-followed scaling trends of $1/\kappa = 0.7$ every 2 or 3 years (bold and dashed lines at the top of the figure, for reference) no longer hold true for V_{DD} . In order to maintain acceptable levels of gate overdrive, V_{DD} scaling has slowed down drastically. When the supply voltage decreases along with device dimensions, then the power density $I_{on} V_{DD}/A$ (on-current times supply voltage divided by surface area) remains constant, which means that the energy needed to drive the chip, and the heat produced by the chip, remain constant. This assumes that when devices scale down, we don't see chip size decreasing, but rather, more complexity and functionality is added with each generation, and chip size remains more or less constant.

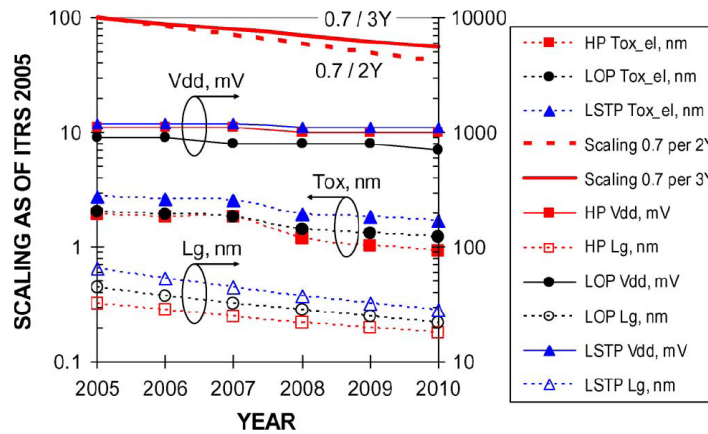


Figure 2.2: Scaling trends showing the decreases in t_{ox} and L_g , while V_{DD} stays almost unchanged. From [3].

When V_{DD} doesn't scale down, power density increases instead. For each MOSFET, the dynamic and static power consumption can be expressed as [4]

$$P_{dynamic} = fC_L V_{DD}^2 \tag{2.1}$$

where f is the frequency and C_L is the total switched capacitive load, and

$$P_{static} = I_{leak} V_{DD} \tag{2.2}$$

where I_{leak} is the sum of the leakage currents in the device when the MOSFET is in the off-state.

If V_{DD} does not decrease, and yet device dimensions decrease, and more devices are added to a chip such that chip size is not significantly reduced, then it can be expected that power consumption will rise considerably. The current trend of increasing power is illustrated in Fig. 2.3. The discussion up until now has not explained why static power would be increasing much faster than dynamic power, and that comes back to the second option for keeping a high gate overdrive: scaling down V_T .

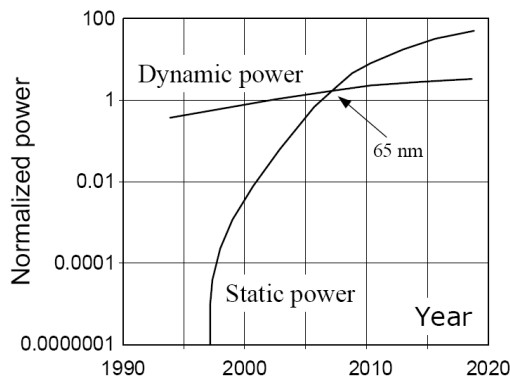


Figure 2.3: Trends of dynamic and static CMOS power, showing that static power consumption has become a greater problem than dynamic power consumption. From [4].

One characteristic of conventional MOSFETs is their fixed slope in subthreshold, when $I_{DS}-V_{GS}$ is plotted on a log-lin scale. This fixed slope means that once the device has been fully optimized in order to have the most abrupt possible turn-on with gate voltage, and the subthreshold swing $S = dV_{GS}/d(\log I_{DS})$ has hit its limit of 60 mV/decade at room temperature, then the only way to lower the threshold voltage further is to shift the $I_{DS}-V_{GS}$ characteristic horizontally on the x-axis, as illustrated in Fig. 2.4. If we want to shift V_T by 60 mV, then the price to pay is an increase of one

decade of off-current, and in turn, of static power. Further discussion of the immutability of the subthreshold swing in conventional MOSFETs will be presented in section 2.2.2.

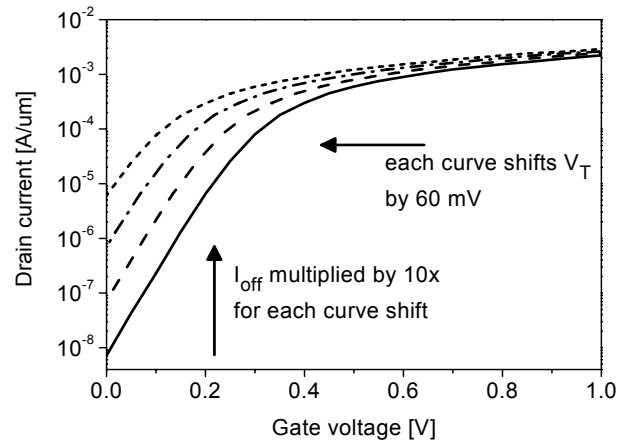


Figure 2.4: A typical I_{DS} - V_{GS} curve for a highly-optimized conventional MOSFET, showing the subthreshold swing limited to 60 mV/decade at room temperature. If we want to decrease V_T by shifting the curve left, we pay a price in leakage current. Solid curve's data is from an optimized asymmetrical double-gate conventional MOSFET in [5], which is then shifted three times. Such a shift could come from engineering the gate work function, for example.

Why is power dissipation such a problem? There are quite a few reasons for which circuits should use less energy, some of which will be mentioned here. The first can be seen on a global scale. We would like our computers, appliances, and gadgets to use less power because it's better for the environment. On a more personal level, it's less expensive to use less electricity. On a practical level, it's more convenient for battery-operated gadgets because their batteries will last longer before needing to be charged. And on a comfort level, it is better when laptops and hand-held gadgets have a lower power density and therefore produce less heat. Looking at Fig. 2.5, the trend of increasing power for Intel computer chips is shown. If we assume that chips tend to be on the order of 1 to 2 cm^2 , we can get a rough idea of the power density as well. According to [6], published in 2010, current power density is around 60-80 W/cm^2 . An ITRS presentation predicted that the power density for the 14 nm node would be greater than 100 W/cm^2 [7]. Fig. 2.5 also shows on its right axis that in order to cope with the increasing power density, the heatsink must grow in volume. This too has a limit, since we would like our appliances, computers, and gadgets to stay the same size or shrink, not get larger in order to accommodate a large heatsink required by the power-hungry chip inside.

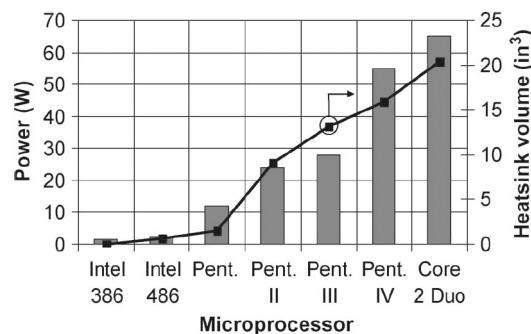


Figure 2.5: Computer chip power trends, along with the accompanying increase in heatsink volume. From [6].

2.2 Possible solutions to the power crisis

2.2.1 Circuit-level solutions

In order to try to reduce the power dissipation of CMOS, circuit engineers change their circuits, and device engineers change their devices. The bulk of this thesis is about devices, but in order to give a glimpse of some of the ways in which circuit designers have responded to the power crisis, a handful of currently-used circuit-level solutions will be presented in this section.

One efficient solution for reducing leakage is to cut off the power supply to idle circuit blocks by using sleep transistors. When sleep transistors control a circuit block, the supply voltage to those blocks is removed or reduced, so that the leakage from the transistors inside that block is eliminated or reduced drastically. One example of this idea is shown in Fig. 2.6, where sleep transistors control the switches connecting the virtual supply rails to the external supply voltage and the external ground. When the sleep circuit turns off those two switches, the arithmetic logic unit (ALU) is no longer powered.

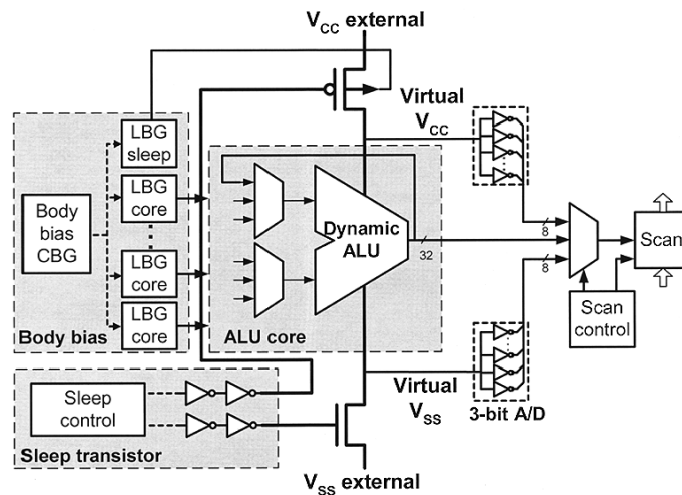


Figure 2.6: Circuit diagram of an ALU which will be turned off by sleep transistors. From [8].

Another way in which circuits can use less power is by incorporating two types of MOSFETs: one with a high threshold voltage, and the other with a low threshold voltage. These dual- V_T circuits use low- V_T CMOS on the critical paths, where speed is critical but leakage will be high, and high- V_T CMOS off the critical paths where leakage can be reduced without sacrificing performance [9].

Transistor stacks are another circuit-level technique that can reduce off-state leakage. Two stacked transistors as shown in Fig. 2.7 (right) have a lower off-current than one transistor alone (left). The reason for this is that the transistors act as a voltage divider, so the drain-source voltage is halved across each device. This has the effect of raising the potential barrier across the stacked transistors, as shown at the bottom of Fig. 2.7. As with high- V_T devices in a dual- V_T design, stacked devices can only be used on non-critical paths because they are slower [10].

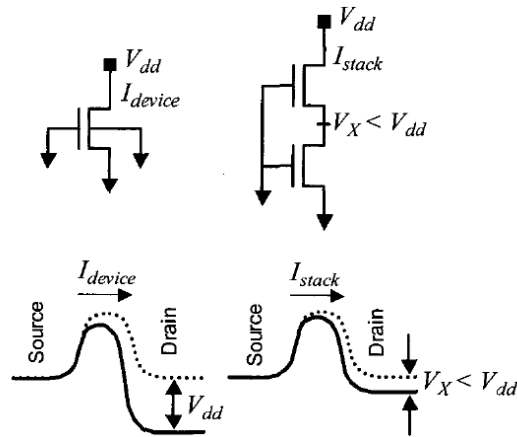


Figure 2.7: Left: one transistor alone, along with its band diagram. Right: two stacked transistors, with a lower leakage current due to the higher potential barrier shown in their band diagram. From [10].

Sub-threshold operation of MOSFETs is an additional energy-saving technique. Fig. 2.8 shows (left axis) that as V_{DD} scales down, there is a minimum level of energy per instruction that can be attained in sub-threshold [11]. The energy rises at higher V_{DD} due to active energy and at lower V_{DD} due to leakage energy. Operating with a lower supply voltage means that the frequency of operation must also decrease, as shown on the right axis of Fig. 2.8.

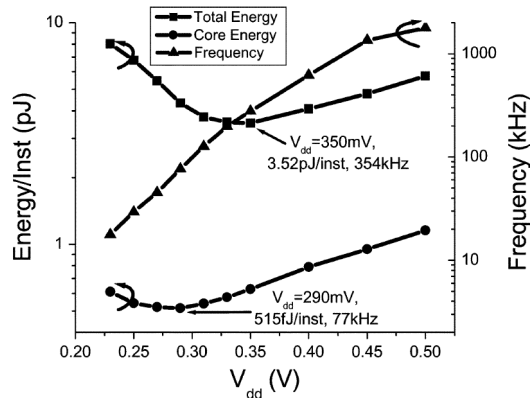


Figure 2.8: Energy per instruction vs V_{DD} (left axis) and frequency vs. V_{DD} (right axis), showing an energy minimum with V_{DD} . From [11].

A multi-core or many-core processor incorporates multiple smaller processors which can work in parallel in order to increase throughput without increasing frequency. Having multiple cores can decrease energy consumption, especially if the cores are asymmetric, where different circuits accelerate different processes. Asymmetric cores are not the best choice, however, if performance scalability is the goal. In that case, symmetric cores are a better option, but then the power savings will not be high [12].

2.2.2 Device-level solutions

Rather than coming up with circuit-level solutions to design around device problems, it would be better to change the devices themselves. So why not design conventional MOSFETs with $S < 60$ mV/decade at room temperature?

The answer to that question comes down to the way in which conventional MOSFETs produce their current in the subthreshold region. First, it is necessary to understand the Fermi-Dirac distribution function, which describes the probability of the occupation of energy levels by electrons. This function is presented in Fig. 2.9, plotted against $E-E_F$, so that the probability of energy level

occupation is 50% at $E-E_F = 0$ eV, or in other words, where the electron energy is equal to the Fermi level. At absolute zero, the Fermi-Dirac distribution function would be 100% for energies less than the Fermi level, and 0% for energies more than the Fermi level, and the function would be perfectly abrupt. As temperatures rise, the function becomes less abrupt, as can be seen in Fig. 2.9.

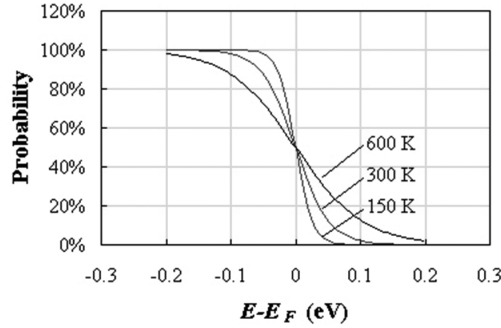


Figure 2.9: The Fermi-Dirac probability as a function of energy, for different temperatures. From [13].

The Fermi-Dirac distribution function can be represented mathematically as

$$f(E) = \frac{1}{1 + e^{(E-E_F)/kT}} \quad (2.3)$$

For conventional MOSFETs operating in the subthreshold region, as the gate voltage increases, the rate of increase of carriers in the channel is determined by this function, which is limited by kT/q . The Fermi-Dirac probability can be seen in the MOSFET subthreshold current equation [14]:

$$I_d \propto e^{\frac{(V_{GS}-V_T)}{m(kT/q)}} \quad (2.4)$$

where the kT/q term limits the rate of increase of the current with applied gate voltage. Solving for the swing, the kT/q term coming from the Fermi-Dirac distribution once again appears, and here serves as the determining factor for the well-known limit of 60 mV/decade at room temperature for conventional MOSFETs.

$$S = \frac{dV_g}{d(\log I_d)} = \ln(10) \frac{mkT}{q} \quad (2.5)$$

where m is the body-effect coefficient, whose value is close to 1 for a well-optimized device. To be more precise, $m = 1 + C_{dm}/C_{ox}$ [14], where C_{dm} is the bulk depletion capacitance at threshold, when $\phi_s = 2\phi_F$. In an optimized MOSFET with very good gate control, $C_{ox} \gg C_{dm}$, and m is slightly greater than 1. Then Eq. 2.5 becomes $S = \ln(10) \times 26$ mV, or about 60 mV/decade.

Since the subthreshold swing is inherently tied to the physical mechanism by which current is generated inside the device in subthreshold, in order to change this limit for swing, it is necessary to change the physical mechanism of the device. In the next section, several types of devices will be introduced that do not generate their subthreshold current in the same way, and that therefore are not confined by this limit on subthreshold swing.

2.3 Small swing devices

This section is an overview of a couple of the main technologies that compete with Tunnel FETs, but is by no means a state-of-the-art of all small swing devices and the progress made for each type. There are plenty of interesting ideas being studied that could reduce a switch's subthreshold swing to less than the 60 mV/decade limit of conventional MOSFETs at room temperature. Several ideas are so new that only one group is working on them, and they have no experimental confirmation. Here, just two types of devices will be mentioned, both of which have attracted international interest and have been explored and experimentally verified in multiple research groups: the IMOS, and MEMS/NEMS switches.

The IMOS is a gated p-i-n junction whose gate is offset from one of the junctions (see Fig. 2.10) such that a very high electric field exists in the non-gated portion of the i-region when the device is on, leading to avalanche breakdown. The impact ionization process means that the IMOS can have a very small subthreshold swing and high on-current [15]. Experimental transfer characteristics are presented in Fig. 2.10(b), showing subthreshold swings of about 4 and 9 mV/decade for n-channel and p-channel devices, respectively.

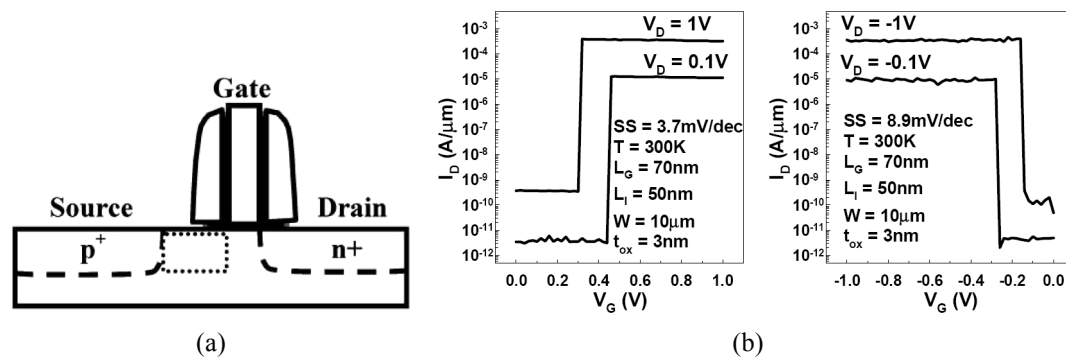


Figure 2.10: (a) A typical IMOS structure, in which a p-i-n diode is partially gated, with part of the intrinsic region left uncovered, from [16]. (b) Measured IMOS I_{DS} - V_{GS} characteristics for n-channel (left) and p-channel (right) devices, from [17], where the source was biased at -5.5 V .

The main problems with IMOS include scalability, since there must always be a gated and an ungated region between the source and drain, hot carrier degradation, since impact ionization necessarily creates hot carriers and these can go into the gate oxide, and the difficulty of low-voltage operation, since high voltages are required in order to induce breakdown.

The second type of small-swing switch that will be discussed in this section is the electro-mechanical relay, either on the micron scale (MEMS) or the nanometer scale (NEMS). While some MEMS switches seem to be just two-terminal devices, which might have extremely small values of swing [18] but are limited in terms of applications in circuits, others are three-terminal devices that could potentially replace conventional MOSFETs. There are many possible designs for this type of switch. One possibility is to fabricate a flexible cantilever beam connected electrically to the source terminal, which is activated by a gate electrode underneath, and pulled down to touch the drain electrode in the on-state [19]. Another possibility is to use a more typical MOSFET layout, with source, channel, and drain, as in the two devices shown in Fig. 2.11. The first (Fig. 2.11(a)), as we reported at IEDM 2005, has an air gap between the gate dielectric and the gate contact and the gate itself moves up (off-state) and down (on-state) [20]. The second (Fig. 2.11(b)) has source and drain regions on the substrate, and a moving gate/channel which can be pulled down into contact with the source and drain in the on-state [21].

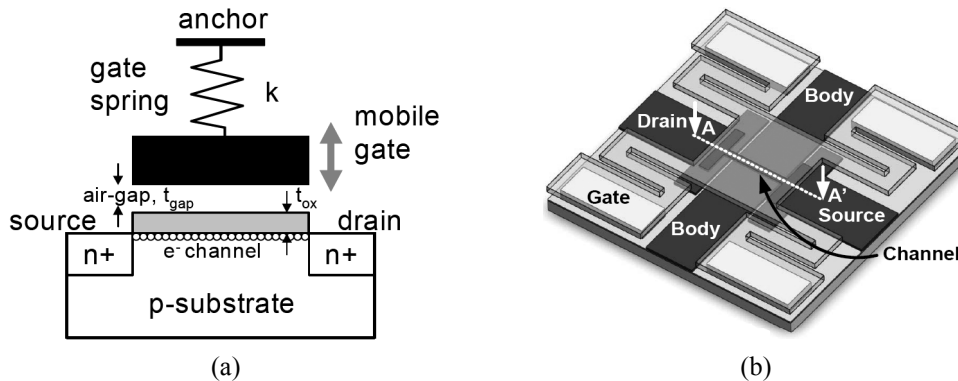


Figure 2.11: Two possible designs for MOSFET-like MEMS relays, with a source, drain, and gate. (a) From [20]. (b) From [21].

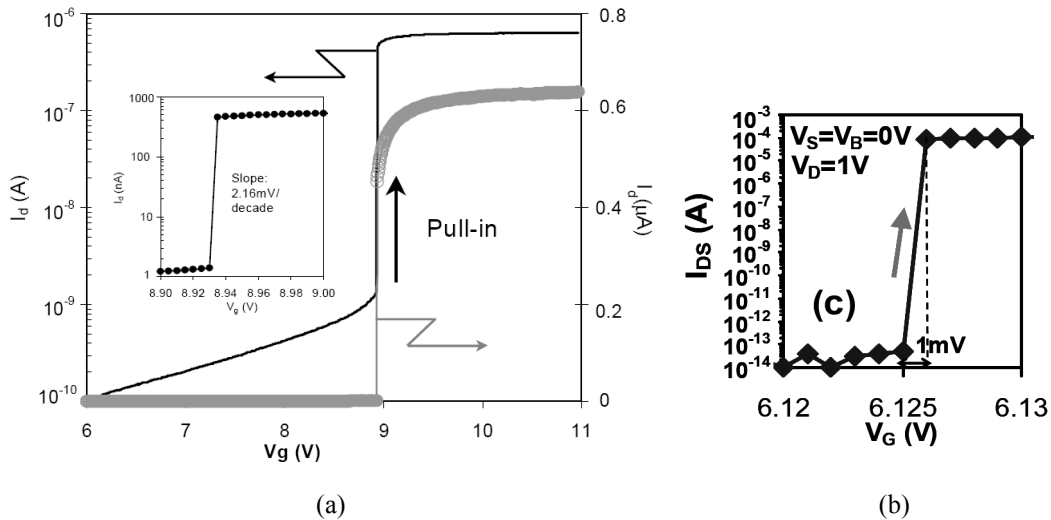


Figure 2.12: Measured I_{DS} - V_{GS} characteristics for the MEMS switches shown in Fig. 2.11(a) and (b), respectively, demonstrating extremely small subthreshold swings. The swing in this type of device is set by the voltage step size, has no fundamental limit, and can approach zero. (a) Junction leakage is still present for this relay design. From [20]. (b) Off-state current is much lower for a design in which the source and/or drain is physically separated from the MOSFET channel. From [21].

MEMS and NEMS switches are interesting due to their potentially high on-currents, very low off-currents, and small subthreshold swings. Their disadvantages include lower speed [19] and reliability problems due to their mechanical nature, such as structural damage where the two pieces need to touch each other and then come back apart thousands or millions of times [18].

A third type of small-swing switch is the focus of the rest of this thesis: the Tunnel FET. It will now be presented in great detail.

2.4 Introduction to the Tunnel FET

Tunnel FETs, also referred to as TFETs, Surface Tunnel Transistors (STTs) or Tunneling FETs, are promising devices to complement or even replace conventional MOSFETs for low-power applications. They offer the potential for a very low off-current and a small subthreshold swing.

Tunnel FETs are interesting as low-power devices because of their quantum tunneling barrier. When the devices are turned on, the carriers must tunnel through the barrier in order for current to flow

from source to drain. When the devices are off, the presence of the barrier keeps the off-current extremely low, several orders of magnitude lower than the off-current of a conventional MOSFET.

2.4.1 Tunnel FET structure and operation

Tunnel FETs are gated p-i-n diodes, or less commonly, gated p-n diodes. To switch the device on, the diode is reverse biased, and a voltage is applied to the gate. In order to be consistent with MOSFET technology, the names of the device terminals are chosen such that voltages are applied in a similar way for Tunnel FET operation. Since a reverse bias is needed across the p-i-n structure in order to create tunneling, and since an NMOS operates when positive voltages are applied to the drain and gate, the n-region of a Tunnel FET is referred to as its drain, and the p+ region as its source for an n-type device.

Fig. 2.13 shows the basic device structure for a typical p-i-n Tunnel FET. The structure shown is an n-type device, with a p+ source and an n+ drain. In a p-type Tunnel FET, the source would be doped n+ and the drain would be doped p+. All Tunnel FETs shown in this thesis have a metal gate with a work function of 4.5 eV.

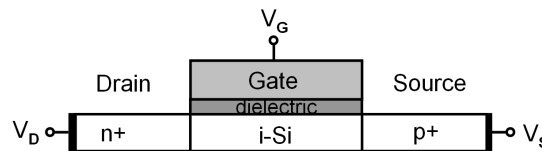


Figure 2.13: A simple Tunnel FET device structure, an n-i-p diode with one gate.

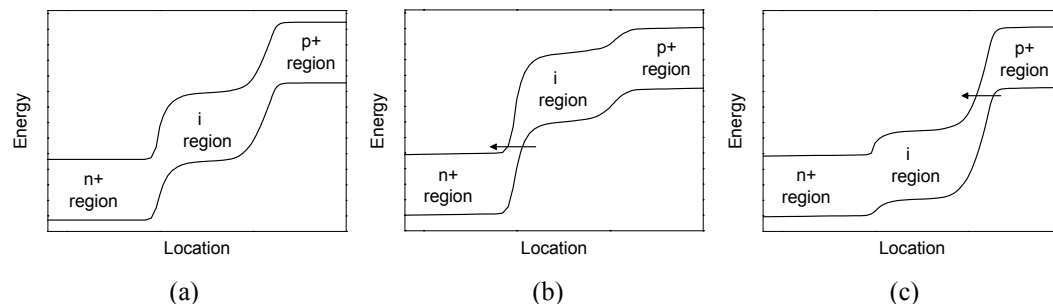


Figure 2.14: Energy band diagrams taken horizontally across the body of a Tunnel FET in (a) the off-state where the only current comes from p-i-n leakage, (b) the on-state with a negative bias on the gate leading to pFET-type behavior, and (c) the on-state with a positive bias on the gate leading to nFET-type behavior.

When a Tunnel FET is OFF, only p-i-n diode leakage current flows between the source and drain, and this current can be extremely low (less than a $\text{fA}/\mu\text{m}$). Fig. 2.14(a) shows the energy bands horizontally across the body of a Tunnel FET in the off-state, with a reverse bias applied across the p-i-n junction, but no voltage applied to the gate.

When a Tunnel FET is designed with symmetry between the n- and p-sides (similar doping levels, similar gate alignment, etc.), the device exhibits ambipolar behavior, whereby the transfer characteristics resemble those of a pFET when a negative voltage is applied to the gate, and those of an nFET when a positive voltage is applied to the gate. Fig. 2.14(b) shows the energy bands with a reverse bias applied across the device, and a negative voltage applied to the gate. The energy bands in the intrinsic region under the gate are lifted, and the energy barrier is now small enough for band-to-band tunneling to take place between the valence band of the intrinsic region and the conduction band of the n+-region. When a positive voltage is applied to the gate, on the other hand, the energy bands in the intrinsic region are pushed down, as in Fig. 2.14(c), and tunneling takes place between the valence band of the p+-region and the conduction band of the intrinsic region. The energy barrier width for band-to-band tunneling is the single most important factor that determines the amount of drain current through a Tunnel FET.

The on-current of an n-type Tunnel FET depends on the width of the energy barrier between the intrinsic and p+ regions, and the current increases exponentially with a reduction in this barrier width. Fig. 2.15(a) shows the dependence of the energy barrier width on the gate voltage for several different gate dielectric constants. The barrier width starts to saturate at high V_{GS} . Fig. 2.15(b) shows the exponential dependence of the simulated tunneling current on the barrier width. More details about the simulations will be presented in section 2.5.

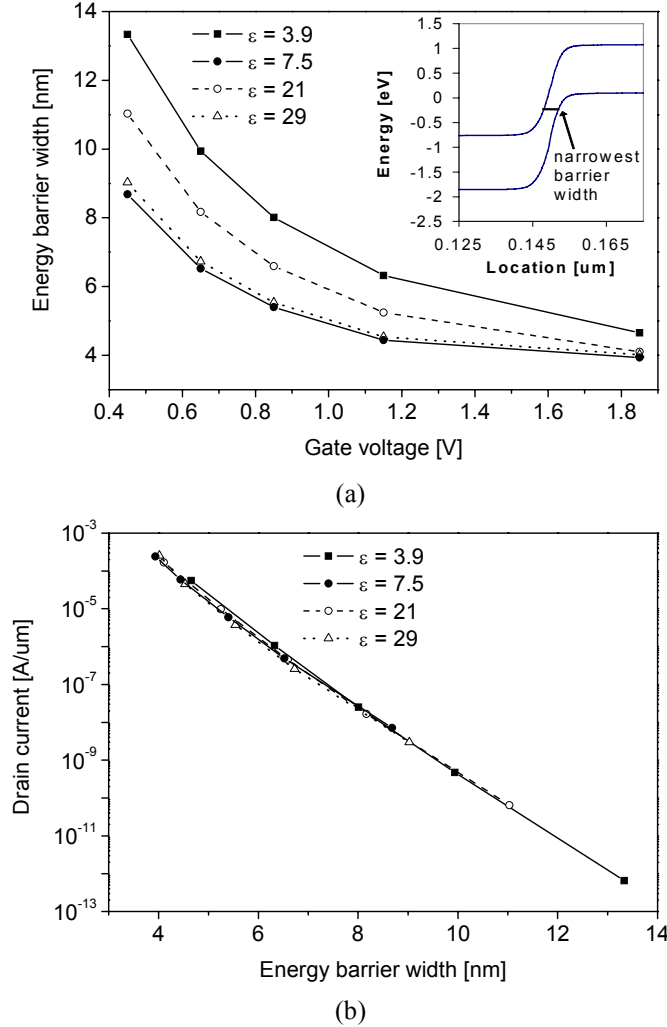


Figure 2.15: (a) Width of energy barrier for band-to-band tunneling, vs. V_{GS} , for different values of the gate dielectric constant. The value of the barrier width was extracted from the narrowest location on the simulated band diagrams at the tunnel junction (as shown in the inset), at a distance of 2.5 nm from the dielectric surface, with 1 V applied to the drain and the source grounded. (b) Drain current vs. energy barrier width for different values of the gate dielectric constant. Double-gate Tunnel FET with $L = 50$ nm, $t_{Si} = 10$ nm, $t_{dielectric} = 3$ nm, $V_{DS} = 1$ V.

2.4.2 Band-to-band tunneling transmission

An expression for the band-to-band tunneling current in Tunnel FETs can be found by using the WKB approximation and taking the tunnel barrier as a triangularly shaped potential barrier as shown in Fig. 2.16. With the WKB approximation, the band-to-band tunneling transmission is given by

$$T_t \approx \exp\left[-2\int_{-x_1}^{x_2} |k(x)| dx\right] \quad (2.6)$$

where $k(x)$ is the quantum wave vector of the electron inside the barrier. Inside a triangular barrier, the wave vector is

$$k(x) = \sqrt{\frac{2m^*}{\hbar^2}(PE - E)} \quad (2.7)$$

Here, PE is the potential energy, and E is the energy of the incoming electron. When the triangular barrier is drawn at the coordinates shown in Fig. 2.16, with the electron at the energy of the widest part of the triangle (at $E=0$), then the E term goes away, and PE can be replaced by the equation for the triangle: $E_g/2 - qFx$, where E_g is the band gap of the semiconductor material at the tunnel junction, and F is the electric field. Then,

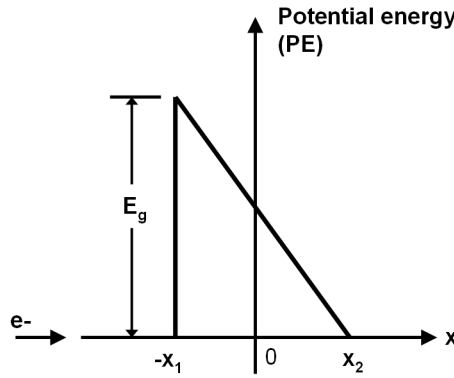


Figure 2.16: Band-to-band tunneling can be calculated by approximating the energy barrier width by a triangular potential energy barrier, where the electrons must tunnel through the widest distance at the base of the triangle. Redrawn from [22].

$$k(x) = \sqrt{\frac{2m^*}{\hbar^2} \left(\frac{E_g}{2} - qFx \right)} \quad (2.8)$$

Plugging this into Eq. 2.6 gives

$$T_t \approx \exp \left[-2 \int_{-x_1}^{x_2} \sqrt{\frac{2m^*}{\hbar^2} \left(\frac{E_g}{2} - qFx \right)} dx \right] \quad (2.9)$$

The next step is to carry out the integration.

$$T_t \approx \exp \left[\frac{4}{3} \frac{\sqrt{2m^*}}{q\epsilon\hbar} \left(\frac{E_g}{2} - qFx \right)^{3/2} \right]_{-x_1}^{x_2} \quad (2.10)$$

Looking back at the triangular barrier, we know that at $x = x_2$, $(E_g/2 - qFx) = 0$, and that at $x = -x_1$, $(E_g/2 - qFx) = E_g$, so

$$T_t \approx \exp \left(-\frac{4\sqrt{2m^*}E_g^{3/2}}{3q\hbar F} \right) \quad (2.11)$$

Eq. 2.11 is a general expression for band-to-band tunneling transmission. This equation can be improved slightly by making it more specific to tunneling transistors. Now referring to Fig. 2.17, the dimensions of the shaded triangular barrier are a height of $\Delta\Phi + E_g$, and a width of λ . The magnitude of the electric field corresponds to the slope of the energy bands, so we can replace the electric field F by $\Delta y/\Delta x = (\Delta\Phi + E_g)/\lambda$. Since electric field is measured in V/m, and the new term has the units eV/m, we must also cancel out an electron charge, which gives

$$I_{BTB} \propto T_t \approx \exp\left(-\frac{4\lambda\sqrt{2m^*E_g^{3/2}}}{3\hbar(\Delta\Phi + E_g)}\right) \quad (2.12)$$

(from [23]). $\Delta\Phi$ is the energy range over which tunneling can take place, E_g is the band gap at the tunnel junction, λ is a screening length, and m^* is the tunneling mass. There are four important conditions in order for band-to-band tunneling to take place: available states to tunnel from, available states to tunnel to, an energy barrier that is sufficiently narrow for tunneling to take place, and conservation of momentum [22]. In order for band-to-band tunneling to take place in materials with an indirect band gap such as silicon, crystal phonons are necessary in order to conserve momentum, and E_g in the numerator of Eq. 2.12 is replaced by $E_g - E_p$, where E_p is the phonon energy. The effective mass m^* must then change to m_{rx}^* , which is the reduced effective mass in the tunneling direction. If these changes are not made to Eq. 2.12, band-to-band tunneling current is overestimated for indirect materials.

The parameter λ deserves a bit more explanation. It has several different names, including screening length, natural length, and Debye length, and refers to the spatial extent of the electric field, or the length over which an electric charge has an influence before being screened out by the opposite charges around it [24]. It can be expressed in terms of the dielectric constants and thicknesses of the gate dielectric and semiconductor body of a device, and depends upon gate geometry. The expression for a double-gate device is [25]

$$\lambda = \sqrt{\frac{\epsilon_{Si}t_{Si}t_{ox}}{2\epsilon_{ox}}} \quad (2.13)$$

where ϵ_{Si} and t_{Si} are the dielectric permittivity and thickness of the silicon (or whatever semiconductor material is used to make the device), and ϵ_{ox} and t_{ox} are the dielectric permittivity and thickness of the gate dielectric. For a single-gate device, the factor of $(1/2)^{0.5}$ must be removed from the λ expression, and for a wrap-around gate, the expression becomes more complicated [25]. These equations for λ were created to describe conventional MOSFET behavior, but have also been used for Tunnel FETs [23], and it will be shown in Chapter 3 that they show the right trends for our Tunnel FET simulations (Figs. 3.9 and 3.15).

Fig. 2.17 shows how the band-to-band tunneling behavior of the Tunnel FET acts as a band pass filter that cuts off the low-energy and high-energy tails of the Fermi distribution of the n+-type source. The Fermi-Dirac distribution and Fermi level for the source are first drawn at the left within the source, and the low-energy tail of the distribution is crossed out because no carriers can exist at energies inside the band gap. Then on the channel side, the source Fermi-Dirac distribution is shown again, and this time the high-energy tail is crossed out since those energy levels can't exist inside the band gap of the channel. The result is the version of the distribution shown at the far right, in which only the electrons in the source within the energy range $\Delta\Phi$ are available for tunneling.

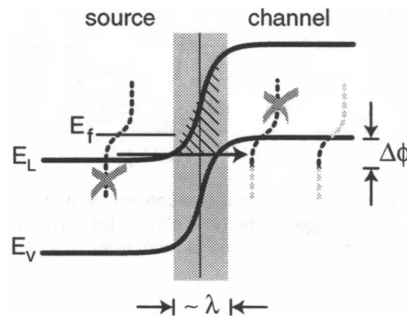


Figure 2.17: Energy band cross section of a Tunnel FET showing the triangular barrier approximation within the bands, $\Delta\Phi$, the screening length λ , and the filtering behavior of the device in the on-state, from [23].

2.4.3 Subthreshold swing in Tunnel FETs

The dependence of swing on gate voltage up to the threshold voltage (taken at $I_{DS} = 10^{-7}$ A/ μm) is shown in Fig. 2.18, demonstrating two important things. First, the subthreshold swing of Tunnel FETs is not constant, but rather is a function of gate voltage. And second, at low gate voltages, it is possible for Tunnel FETs to have a subthreshold swing less than the 60 mV/decade MOSFET limit at room temperature.

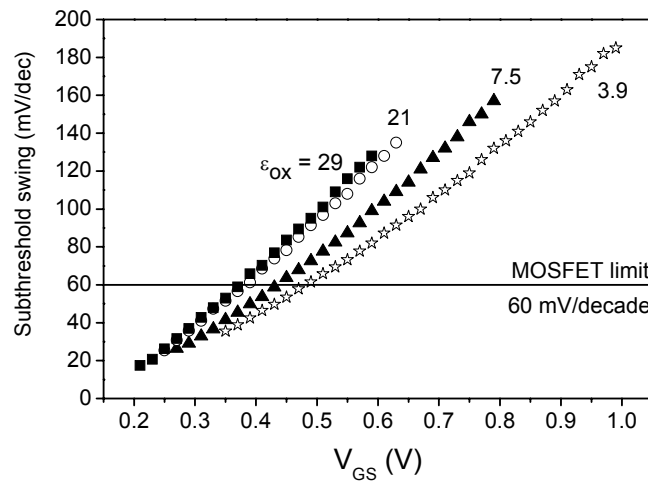


Figure 2.18: Dependence of the Tunnel FET subthreshold slope on gate voltage for different dielectric constants, from numerical simulation. Each curve goes up to the threshold voltage of that device. $L = 50$ nm, $t_{\text{dielectric}} = 3$ nm, $t_{\text{Si}} = 10$ nm, $V_{DS} = 1$ V. The points were generated by taking the swing value ($dV_{GS}/d(\log I_{DS})$) at each point on the I_{DS} - V_{GS} curves.

In order to derive an expression for the subthreshold swing of a band-to-band tunneling device, we can start with the expression given by Sze [26] for the tunneling current through a reverse-biased p-n junction:

$$I = aV_{eff}F \exp\left(-\frac{b}{F}\right) \quad (2.14)$$

where

$$a = Aq^3 \sqrt{2m^*/E_g} / 4\pi^2 \hbar^2, \quad (2.15)$$

and

$$b = 4\sqrt{m^*}E_g^{3/2} / 3q\hbar \quad (2.16)$$

V_{eff} is the bias at the tunnel junction and F is the electric field at the tunnel junction ([27], who took them from [26]). When the subthreshold swing is calculated as $S = dV_{GS}/d(\log I_{DS})$, the result [27] is

$$S = \ln 10 \left[\frac{1}{V_{eff}} \frac{dV_{eff}}{dV_{GS}} + \frac{F+b}{F^2} \frac{dF}{dV_{GS}} \right]^{-1} \quad (2.17)$$

Several conclusions can be drawn from this equation. First, as already illustrated in Fig. 2.18, it should be noted that in sharp contrast with a conventional MOSFET, the subthreshold swing is a function of V_{GS} . This means that the subthreshold region does not appear as a straight line when I_{DS} - V_{GS} is plotted on a log-lin scale, and the swing does not have one unique value. Swing is smallest at the lowest V_{GS} , and increases as V_{GS} increases. Fig. 2.19 shows a comparison of the I_{DS} - V_{GS} curves for a typical conventional MOSFET, and for a typical Tunnel FET.

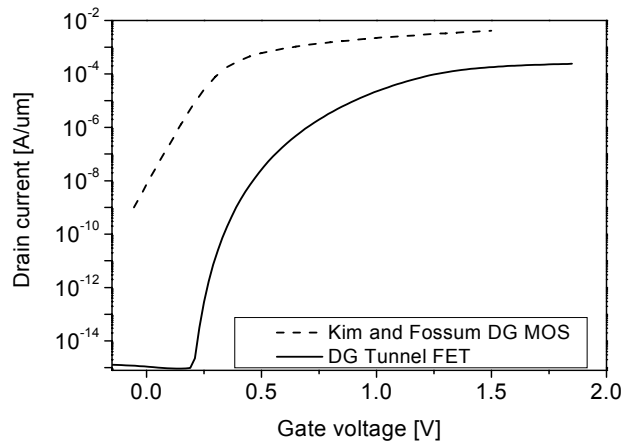


Figure 2.19: Qualitative comparison of I_{DS} - V_{GS} for a conventional MOSFET [5] and a Tunnel FET showing the non-constant subthreshold swing for the Tunnel FET. In general, Tunnel FETs have a much lower off-current and a lower on-current than conventional MOSFETs.

Due to the changing values of swing along the I_{DS} - V_{GS} curve, it is useful to define two different types of swing, point swing (S_{pt}) and average swing (S_{avg}). These are illustrated in Fig. 2.20. Point swing is the smallest value of the subthreshold swing anywhere on the I_{DS} - V_{GS} curve, typically found right as the device leaves the off-state and tunneling current starts to flow. Average swing is taken from the point where the device starts to turn on, up to threshold, often defined using the constant current technique. Average swing is the more useful value for circuit designers, though in order to truly utilize the average slope as shown in Fig. 2.20, the gate work function would need to be adjusted in order for the turn-on point to fall right at $V_{GS} = 0$ V.

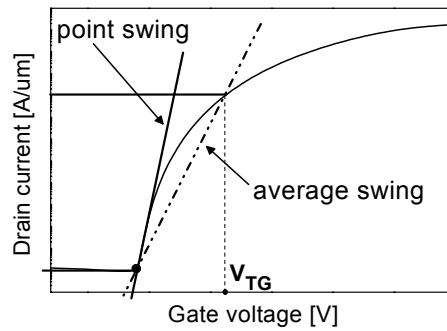


Figure 2.20: Visual definitions of point swing, taken at the steepest point of the I_{DS} - V_{GS} curve, and average swing, taken as the average from turn-on to threshold.

2.4.4 Tunnel FET temperature characteristics

The temperature dependence of silicon Tunnel FETs with an SiO_2 gate dielectric has been reported in [28] and [29]. Simulations of Tunnel FETs with a high-k dielectric show the same general trends: the off-current, caused by the generation of carriers in a reverse-biased junction, increases with temperature, while the on-current, coming from band-to-band tunneling, changes only slightly, as shown in Fig. 2.21. The inset of Fig. 2.21 shows that the subthreshold swing of the Tunnel FET for fixed values of V_{GS} is nearly constant as temperature increases, unlike that of a MOSFET, which degrades proportionally to the increase in temperature, as can be seen in Eq. 2.5. Due to rising off-current, the average subthreshold swing of Tunnel FETs will significantly degrade with increasing temperature, but beyond the leakage level, the current characteristics remain nearly unchanged.

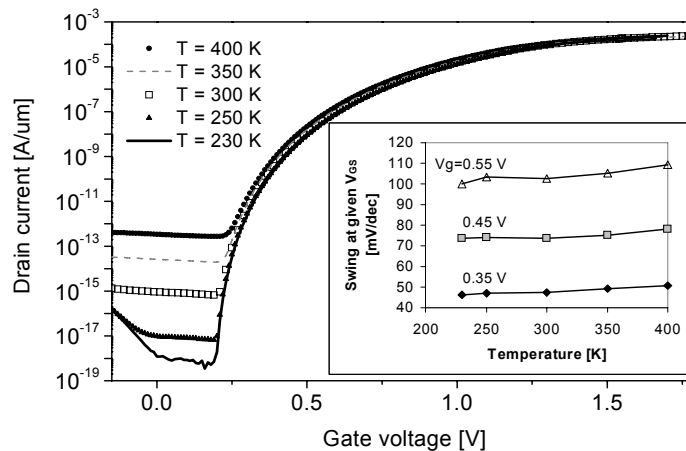


Figure 2.21: Simulated I_{DS} - V_{GS} characteristics for various temperatures for a double-gate Tunnel FET with $\epsilon_{\text{dielectric}} = 21$. $V_{DS} = 1$ V. As temperature increases, I_{off} increases, but I_{on} changes very little. Inset: Subthreshold swing at specific V_{GS} values, vs. temperature in Kelvin. The swing is only slightly affected by changes in temperature.

The use of a high-k dielectric rather than SiO_2 leads to a decrease in the threshold voltage shift caused by temperature. This is to be expected with the constant current method of V_T extraction, since with a higher dielectric constant, V_T falls on a steeper part of the I_{DS} - V_{GS} curve. While $\Delta V_T/\Delta T$ is in the range of 1-2 mV/K for Si/SiO_2 Tunnel FETs [29] and MOSFETs, we find that $\Delta V_T/\Delta T$ is 0.2-0.3 mV/K for Tunnel FETs with a gate dielectric constant of 21.

Although subthreshold swing doesn't degrade with increased temperature, when taken at a fixed value of V_{GS} , it must be kept in mind that circuit designers don't care about swing at a fixed value of gate voltage. They would be more interested in average swing, taken from turn-on up to threshold. Since an increase in temperature has a strong effect on I_{off} , as seen in Fig. 2.21, the steepest part of the curve is lost as temperature goes up, and so S_{avg} will be significantly degraded.

2.5 Silvaco Atlas models

The most important model for Tunnel FET simulations is the band-to-band tunneling (BTBT) model. There is a choice to be made between local models, which use simple equations in which the electric field is an important parameter, and non-local models, which have a more physical basis and don't depend on the electric field at the individual mesh points in the simulated device structure, but rather on band diagrams calculated along cross-sections through the device. The non-local BTBT model was used for all simulations in this thesis, except in the appendix when specified otherwise. The equations for all BTBT models available in Silvaco Atlas are given in the appendix.

In principle, all the right qualitative trends are captured by Atlas' non-local BTBT model, but the quantitative predictions are less reliable and need further calibration on data or even new model developments, as will be shown and discussed in more detail in the appendix. For this reason, the qualitative trends of Tunnel FET characteristics' dependence on device parameters is expected to be correct, but the quantitative values should be treated with great caution.

The energy bands for the non-local BTB tunneling model in Silvaco Atlas are calculated as shown in Fig. [30], where each energy in the allowed range shown in the figure has a corresponding contribution to the BTBT current. The tunneling probability is calculated using the WKB approximation.

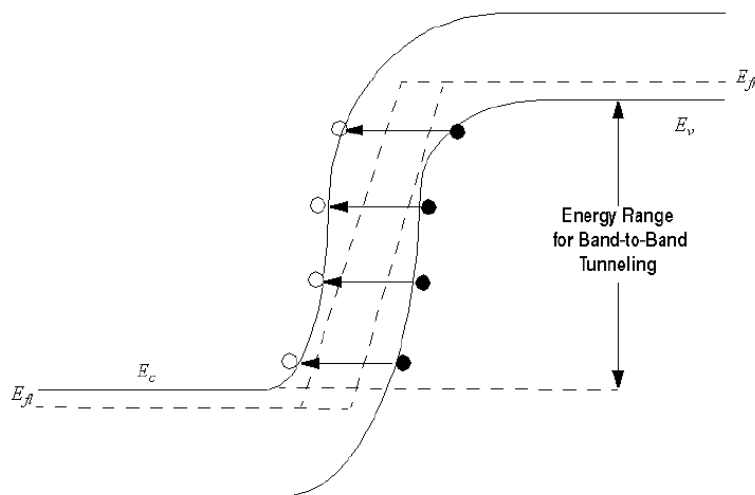


Figure 2.22: Example of energy bands used for calculating BTBT current in Silvaco Atlas. This figure will be presented again along with more explanation and equations in the appendix.

Two other important models that were applied during simulation are the bandgap narrowing model from Slotboom and de Graaf [31], which reduces the band gap when doping concentrations are high, and the quantum model, which uses a density gradient to simulate the effects of quantum confinement in thin material layers. Equations for these models are given in the appendix.

2.6 History and state-of-the-art of the tunneling transistor

Quinn et al. at Brown University [32] were the first to propose the gated p-i-n structure of a Tunnel FET in 1978, and suggested the usefulness of this device for spectroscopy. Banerjee et al. at Texas Instruments [33] studied the behavior of a three-terminal silicon tunnel device using a p⁻-region instead of an i-region under the gate. Takeda et al. at Hitachi [34] created a band-to-band tunneling MOS device on silicon that they called the B²T-MOSFET, and showed the lack of V_T rolloff when scaling, and the temperature dependence of the device characteristics. Baba at NEC [35] fabricated Tunnel FETs which he called Surface Tunnel Transistors, using MBE to create mesa structures in III-V materials. In 1995, Reddick and Amaratunga at Cambridge [36], seemingly unaware of all the previously mentioned work, published measured characteristics of silicon Surface Tunnel

Transistors. They were motivated by the desire for devices that would be faster than conventional MOSFETs, as tunneling devices are, and that could be scaled down more easily without running into problems such as punchthrough. They are sometimes erroneously given credit for being the first to make silicon Tunnel FETs. In 1997, Koga and Toriumi at Toshiba [37] proposed a post-CMOS three-terminal silicon tunneling device with the same structure as a Tunnel FET, though the experimental results which were presented showed a device that was forward-biased.

In 2000, Hansch et al. at the University of the German Federal Armed Forces in Munich [38] showed experimental results from a reverse-biased vertical silicon tunneling transistor made with MBE, with a highly-doped boron delta-layer to create an abrupt tunnel junction, and noted the saturation behavior in the I_D - V_G characteristics. Aydin and Zaslavsky at Brown Univ., along with their collaborators in New York and France [39] fabricated Lateral Interband Tunneling Transistors on SOI in 2004. These devices used a different Tunnel FET structure without an intrinsic region, instead placing the gate over a p-n junction, claiming that this would reduce gate capacitance and therefore increase speed. The authors also claim that there should be no current saturation for these devices. Similar devices on bulk silicon had already been investigated by Grove and Fitzgerald in 1965 [40].

In 2004, band-to-band tunneling was demonstrated in carbon nanotube (CNT) FETs by Appenzeller et al. [41], as illustrated in Fig. 2.23. In order to create the energy bands necessary for tunneling, a back gate and a top gate were used. The researchers claimed that the one-dimensionality of the CNTs led to extremely different band bending conditions than those in 3D semiconductors. A subthreshold swing smaller than the 60 mV/dec limit of conventional MOSFETs was reported for the first time, which was a momentous occasion, even if the low swing value was only between a couple of points at very low current values. A year later, Appenzeller et al. [42] published a comparison of several CNT transistors, and concluded that the Tunnel FET, now with only one gate, was the superior device and showed conventional-looking I_{DS} - V_{DS} output characteristics while still achieving a subthreshold swing of less than 60 mV/dec.

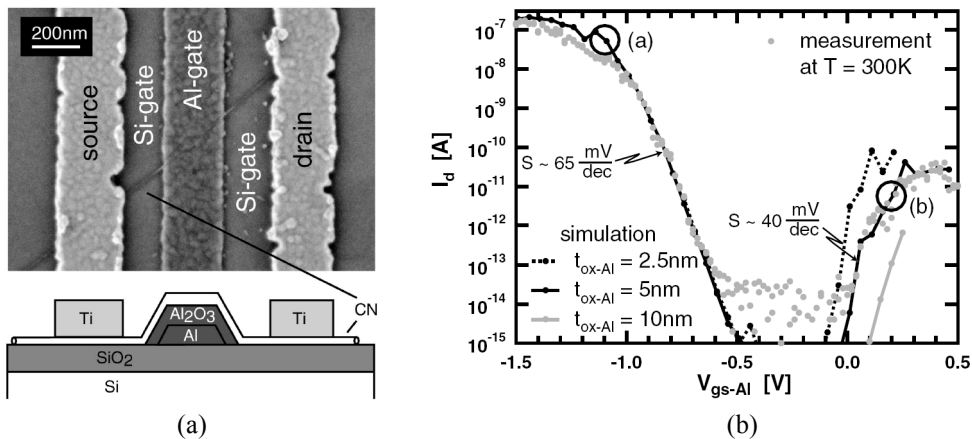


Figure 2.23: (a) 2004 CNT Tunnel FET with two tunnel junctions and back gate by Appenzeller and Knoch, showing the first realization of a tunneling device with a subthreshold swing of less than 60 mV/decade at room temperature. (b) I_{DS} - V_{GS} for this device, with the region of low swing marked. From [41].

In 2004, Bhuwarka et al. at the University of the German Federal Armed Forces in Munich [43] published the first of many articles about their vertical Tunnel FET on silicon with a SiGe delta layer, grown by MBE. The SiGe replaced the silicon delta layer already used by Hansch, and in theory, the smaller bandgap should have reduced the tunnel barrier width and increased tunneling current in the on-state as well as lowering the subthreshold swing. In 2006, the same group proposed a lateral Tunnel FET on SiGe on insulator [44], and showed through simulation that on-current would increase with the percentage of Ge in the SiGe. No experimental results have been published to date by this group for these devices.

2004 continued to be a busy year for the Tunnel FET, as Wang et al. at TUM (Munich) published experimental results for complementary silicon tunneling devices fabricated in a CMOS process with leakage currents of less than 10^{-11} A/ μm [45]. This group went on to report results for tunneling devices showing surprisingly high on-currents [46]. Sadly, these devices were not true Tunnel FETs since a parasitic conventional MOSFET dominated the device characteristics, and several of this group's articles (including [46]) had to be retracted [47].

In 2006, Zhang et al. at Notre Dame [48] remarked once again what others before them had noticed – that theoretically, it is indeed possible for Tunnel FETs to have a subthreshold swing lower than 60 mV/dec. The structure they studied was a gated p-n diode, but the general equations they put forth, and the band-to-band tunneling behavior, would be the same as for a gated p-i-n structure.

To put this thesis work into the context of all this Tunnel FET research activity, here is what we published around this same time period. We started publishing results of our Tunnel FET simulations in 2006, starting with ESSDERC 2006, and our first long article appeared in Transactions on Electron Devices in July of 2007 [49], just before many more groups became active on the topic and started publishing. The subject matter of those two articles is presented in Chapter 3. The same year, we published a Tunnel FET threshold voltage study [50], whose content will be seen in Chapter 4, and a length scaling study [51], whose content is presented in Chapter 5.

In 2007, Verhulst et al. at IMEC showed by simulation that shortening Tunnel FET gate length, so that the gate covers the source-side junction where tunneling takes place, but does not cover the majority of the intrinsic region, has the benefits of decreasing off-current (tunneling through the drain-side junction) and reducing speed, with a small or no reduction in the on-current, depending on the device design [52]. In the same year, Toh at the National University of Singapore published a study of double-gate Tunnel FET silicon body thickness optimization, in which he showed an optimal device thickness for maximum on-current [53]. Nagavarapu et al. at UCLA suggested a pnpn device design in 2008, in which a narrow region of the opposite doping is introduced into the Tunnel FET source just under the gate edge. This narrow region acts as a source of electrons, and increases the band bending and the electric field at the tunnel junction, thus increasing on-current [54].

The following publications all appeared in 2009. Schlosser et al. at the University of the German Federal Armed Forces in Munich studied the simulated advantages of putting an extremely high-k dielectric on a Tunnel FET, and the benefits of the fringing fields when the dielectric is only over the intrinsic region [55]. Vadizadeh et al. at the University of Tehran presented a simulation study of Tunnel FETs in which a high-k dielectric covered the tunnel junction, and the rest of the intrinsic region was covered by a low-k dielectric [56]. This technique led to little improvement in device characteristics, however; most of the shown improvement actually came from a shift in the gate work function which shifted their I-V curves along the voltage axis. Patel et al. at UC Berkeley simulated an interesting device whose band-to-band tunneling takes place perpendicularly to the gate dielectric surface, and showed that it would have a very small subthreshold swing and a high on-current [57]. Their device has an ultra-shallow n⁺ pocket at the surface of the p⁺ drain (for an n-type device), and the gate overlaps this pocket. When gate voltage is applied, carriers tunnel upward from the p⁺ source into the n⁺ pocket, and then drift to the drain.

There were also some fabricated Tunnel FET results in 2009. Sandow et al. from Forschungszentrum Jülich published experimental data for p-type Tunnel FETs on SOI, showing the effects of varying source and drain doping levels, gate dielectric thickness, and device length [58]. Kazazis et al., with his colleagues at Brown University and in France, fabricated Tunnel FETs on thin GeOI that showed very high leakage, with $I_{\text{on}}/I_{\text{off}} < 100$ [59]. Moselund et al. at IBM's Zurich Research Laboratory fabricated Tunnel FETs on silicon nanowires with a wrap-around gate, using two different gate dielectrics: SiO₂ and HfO₂ [60]. The nanowires were grown vertically and doped *in-situ*, and then deposited on a pre-patterned substrate where the gate dielectric was deposited and the drain, source, and gate contacts were made. Improvements in subthreshold swing and on-current were seen with the use of a high-k dielectric.

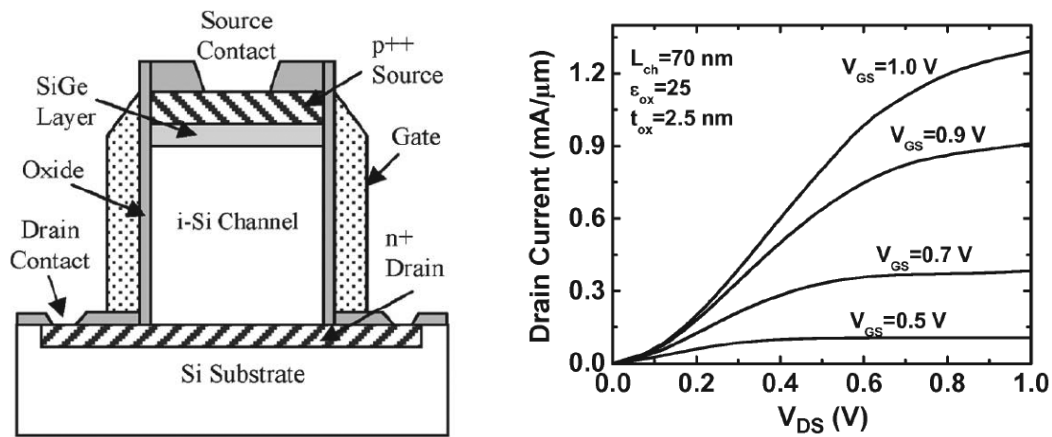


Figure 2.24: (a) N-type Tunnel FET device structure showing delta layer of SiGe at the tunnel junction, and (b) the resulting output characteristics for a device with 3 nm of SiGe. From [61].

Non-silicon Tunnel FETs

Starting in 2008, many research groups started to publish on the topic of alternative semiconductor materials that could be used in order to boost the low on-current typically seen for all-silicon Tunnel FETs. This section will only mention articles based upon Tunnel FETs comprised of one single material. Heterojunctions will be discussed in the next section.

Luisier et al. at Purdue University did an atomistic study of InAs Tunnel FETs, in which they found that subthreshold swings of less than 60 mV/decade at room temperature could only be attained if single-gate body thicknesses were less than 4 nm, and double-gate body thicknesses were less than 7 nm, or for nanowires of less than approximately 10 nm diameter [62]. Looking at more exotic material systems, Tunnel FETs could one day be fabricated on graphene nanoribbons, which are basically unrolled single-walled carbon nanotubes. The simulated transfer characteristics presented in [63] represent the extremely optimistic upper bounds of possible device performance and reach a simulated subthreshold swing of 0.19 mV/dec.

Mayer et al. at CEA-LETI fabricated Tunnel FETs on a buried oxide layer [64]. The semiconductor portion of the device was either silicon, Si_{1-x}Ge_x (x = 15% or 30%), or germanium. Both on-current and off-current increased significantly when the band gap of the material used was reduced from about 1.1 eV (silicon) to about 0.66 eV (germanium). Silicon nanowire Tunnel FETs have been transferred onto flexible plastic substrates [65], and though the devices were not optimized and had poor characteristics, they seemed to be insensitive to substrate bending within the tested range.

Heterostructure Tunnel FETs

With a heterostructure Tunnel FET, the materials are chosen such that the source material has a small band gap so that the energy barrier width at the source junction is reduced in the on-state, while the drain material has a large band gap, which creates the largest possible energy barrier width at the drain side in the off-state, to keep the off-current low. It is not enough to choose any small band gap material for the source and any large band gap material for the drain, however. The way in which the bands naturally line up with each other at the heterojunction, which depends on their electron affinities, is also crucial. This was illustrated by Verhulst et al. from IMEC in [66], showing that the best situation is a continuous valence band with a conduction band offset for p-type Tunnel FETs, and a continuous conduction band with a valence band offset for n-type Tunnel FETs. In this way, the energy barrier width in the on-state is minimized, and the highest possible on-current results.

Knoch at TU Dortmund University looked a bit more at band gap line-ups for III-V heterostructure Tunnel FETs, this time just for p-type devices [67]. In this simulation study, depending on the fractions of aluminium and gallium in the Al_xGa_{1-x}Sb nanowires on an InAs surface, the band diagrams at the tunnel junction varied from staggered to broken. The results showed that a minimum

swing could be attained using a heterojunction with a staggered band line-up that is nearly broken but not quite.

Strain-engineered Tunnel FETs

Heterojunction Tunnel FETs are interesting because they allow the independent manipulation of semiconductor properties at the two junctions of the device. Using a material with a small band gap at the source side increases the on-current. This effect can also be achieved with either a strained material at the source side, or with a lateral strain profile.

When low-band gap materials and strain are combined at the tunnel junction of a Tunnel FET, the band gap there is reduced by both effects, and the on-current is improved by both. A group at Stanford [68] presented experimental results from a heterostructure Tunnel FET whose strained germanium layer goes all the way across the device, comprising source, intrinsic region, and drain, as shown in Fig. 2.25(a). Because the germanium layer has a small band gap and is sandwiched between two silicon layers with a larger band gap (Fig. 2.25(b)), the current flows within the germanium layer. As a result, the on-current is high when the drain and gate voltages are 3 and 4 V, respectively, as seen in Fig. 2.25(c), but the device displays unwanted ambipolar behavior.

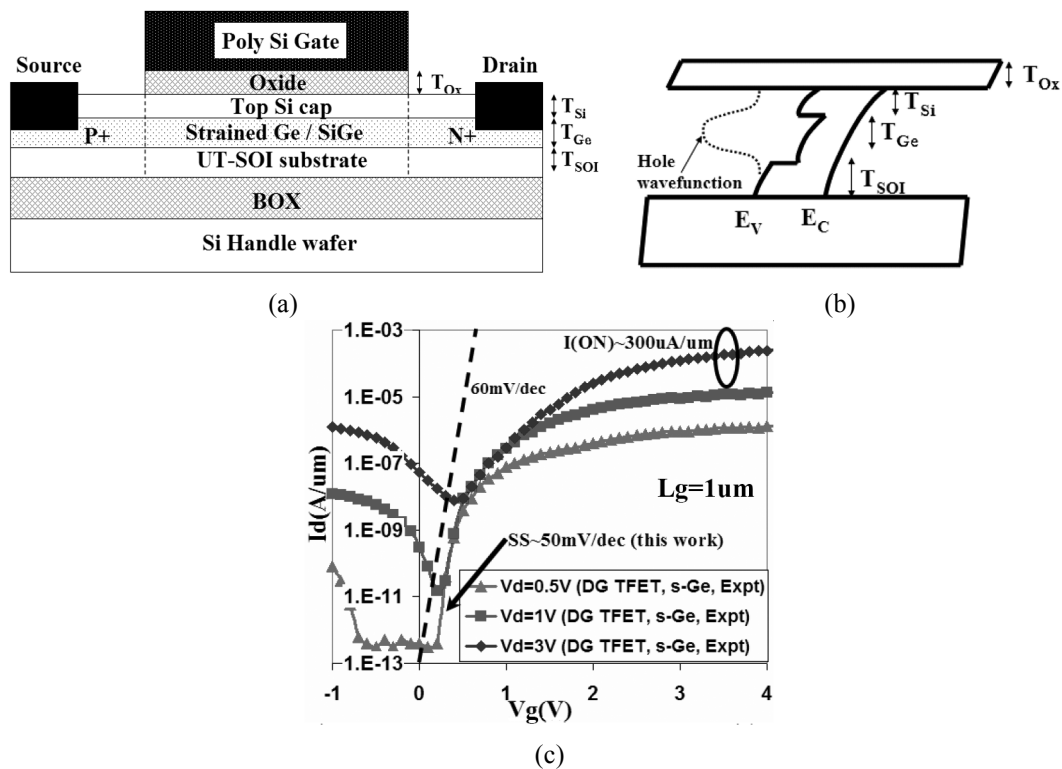


Figure 2.25: (a) Device schematic for fabricated Tunnel FETs, using a buried, strained Ge layer. (b) Band diagram for this device, showing the reduced band gap in the strained germanium layer. (c) Measured I_{DS} - V_{GS} characteristics, where a high on-current was obtained with $V_{DS} = 3$ V, and a low off-current and a small subthreshold swing were obtained with $V_{DS} = 0.5$ V. From [68].

Nayfeh et al. at MIT fabricated Tunnel FETs with a strained SiGe layer [69]. The devices were not designed to be Tunnel FETs but rather conventional p-type MOSFETs, but when the n-type bulk is used as a drain, Tunnel FET characteristics are the result. The devices show an increase in on-current by several orders of magnitude with increasing germanium content in the strained layer. Off-current increases as well, since the non-tunnel junction (at the drain side) also has a small band gap.

In 2009, we published a simulation study of Tunnel FETs with a lateral stress profile [70], demonstrating high on-current due to a reduced band gap at the tunnel junction, and small off-current and small subthreshold swing due to a large band gap at the drain junction. This material will be presented in Chapter 3.

Modeling of Tunnel FETs

Some recent work on the subject of Tunnel FET modeling will be mentioned in this section, even though this thesis does not go into the topic further. In 2007, Knoch et al. [71] analyzed the physics of dimensionality (1D and 3D) in Tunnel FETs. Vandenberghe and Verhulst, and their colleagues at IMEC have published several articles on the topic. They model two different types of tunneling current, happening in two locations in the device body: point and line tunneling [72]. They have also modeled characteristics for Tunnel FETs with various gate configurations [73]. More progress needs to be made before the models match the experimental data closely.

Good analytical and compact models for Tunnel FETs still need to be developed, both for static and dynamic behavior. These models would allow the extraction of figures of merit, and would enable benchmarking against conventional MOSFETs. They would also let circuit designers easily incorporate Tunnel FETs in their circuit simulations. None of this is currently possible. Such analytical and compact models will need to incorporate important device design variables such as gate oxide thickness, body thickness, and the band gaps at both the source-side and the drain-side junctions.

Tunnel FETs in circuits

The subject of Tunnel FETs in circuits is also outside the scope of this thesis, but once again, some recent work from 2008 and 2009 will be acknowledged. Fulde et al. at the Technical University of Munich analyzed the analog and digital device characteristics of p-type multiple-gate Tunnel FETs, found an intrinsic gain of more than 300, and showed how they could be used as part of a voltage reference circuit [74]. Kam et al. at UC Berkeley found that for slow applications ($f < \sim 500$ MHz), Tunnel FETs are more energy-efficient than conventional MOSFETs [75]. When I_{on} , and therefore V_{DD} , must be high, however, and speed is critical, then MOSFETs are the more suitable choice.

Koswatta et al. at Purdue University carried out an in-depth performance comparison between Tunnel FETs and conventional MOSFETs on a carbon nanotube platform [76]. They found that despite Tunnel FETs' limited on-currents, they had a fundamentally smaller switching energy than conventional MOSFETs, and could switch faster at higher I_{on}/I_{off} ratios. Kim et al. at the University of Michigan developed an SRAM cell based on heterojunction Tunnel FETs that showed greatly reduced leakage in comparison with conventional CMOS [77]. Finally, Mookerjee et al. at Pennsylvania State University explained the extraction of effective drive current and output capacitance for Tunnel FET delay calculations [78].

2.7 Conclusion

This chapter showed the reasons behind the search for a MOSFET-like switch with a subthreshold swing of less than 60 mV/decade at room temperature. The Tunnel FET is a strong contender in this category. As shown in the history and state-of-the-art, although the device was invented decades ago, it is a relatively young device, and still has a huge amount of unexplored potential. A fully optimized device has yet to be successfully fabricated and measured, and Tunnel FET modeling is still in its infancy, as is the exploration of its use in digital and analog circuits.

The following topics were presented in this chapter:

- *The limitations faced by CMOS: the power crisis* (Section 2.1)

The motivation for small swing switches goes back to the scaling of conventional MOSFETs, and the continuing increase in their static power consumption. This section explained that the leakage power increase comes from scaling down dimensions without scaling down the supply voltage, and from trying to keep gate overdrive high by reducing threshold voltage through shifting $I_{DS}-V_{GS}$ characteristics and thus increasing I_{off} .

- *Possible solutions to the power crisis* (Section 2.2)

Both circuit-level and device-level solutions were discussed. The use of sleep transistors, dual- V_T logic, stacked transistors, sub-threshold logic, and multi-core processors were mentioned as circuit-

level solutions. At the device level, it was shown that the subthreshold swing for conventional MOSFETs cannot scale below the 60 mV/decade limit at room temperature due to the physics of current generation in sub-threshold.

- *Small swing devices* (Section 2.3)

This section briefly presented the two “other” most widely investigated small swing switches: IMOS, and MEM / NEM switches, and described their operation principles. Their advantages and disadvantages were also outlined.

- *Introduction to the Tunnel FET* (Section 2.4)

This detailed section introduced the Tunnel FET structure and operation, and then showed the derivation of a basic band-to-band tunneling transmission equation. Subthreshold swing in Tunnel FETs was described, and definitions of point swing and average swing were illustrated. Tunnel FET temperature characteristics were presented.

- *Silvaco Atlas models and meshing* (Section 2.5)

Descriptions of the most important models used in the Silvaco Atlas simulations were given, including the models for non-local band-to-band tunneling, bandgap narrowing, and the quantum model.

- *History and state-of-the-art of the tunneling transistor* (Section 2.6)

A history of transistors that use band-to-band tunneling current in their on-state was given, starting from 1978 and continuing to the present day (2010), showing that the Tunnel FET is still an emerging device with much unexplored potential.

2.8 Bibliography

-
- [1] R. Dennard, F. Gaensslen, H.-N. Yu, V. Rideout, E. Bassous, and A. LeBlanc, “Design of Ion-Implanted MOSFETs with Very Small Physical Dimensions,” *IEEE J. Solid-State Circuits*, vol. SC-9, no. 5, pp. 256-268, Oct. 1974.
 - [2] P. Packan, Short Course, IEDM 2007.
 - [3] T. Skotnicki, C. Fenouillet-Beranger, C. Gallon, F. Boeuf, S. Monfray, F. Payet, A. Pouydebasque, M. Szczap, A. Farcy, F. Arnaud, S. Clerc, M. Sellier, A. Cathignol, J.-P. Schoellkopf, E. Perea, R. Ferrant, and H. Mingam, “Innovative Materials, Devices, and CMOS Technologies for Low-Power Mobile Multimedia,” *IEEE Trans. Electron Devices*, vol. 55, no. 1, Jan. 2008, pp. 96-130.
 - [4] P. Nilsson, “Arithmetic Reduction of the Static Power Consumption in Nanoscale CMOS,” *IEEE Int. Conf. on Electronics, Circuits and Systems*, 2006, pp. 656-659.
 - [5] K. Kim and J. Fossum, “Double-gate CMOS: Symmetrical- versus asymmetrical-gate devices,” *IEEE Trans. Electron Devices*, vol. 48, no. 2, Feb. 2001, pp. 294-299.
 - [6] B. Dang, M. Bakir, D. Sekar, C. King, Jr., and J. Meindl, “Integrated Microfluidic Cooling and Interconnects for 2D and 3D Chips,” *IEEE Trans. on Advanced Packaging*, vol. 33, no. 1, Feb. 2010, pp. 79-87.
 - [7] ITRS, “Assembly and Packaging” presentation, ITRS Spring Meeting, Annecy, France, 2007, http://www.itrs.net/Links/2007Spring/Presentations/07_Assembly_2007_Annecy.pdf.
 - [8] J. Tschanz, S. Narendra, Y. Ye, B. Bloechel, S. Borkar, and V. De, “Dynamic Sleep Transistor and Body Bias for Active Leakage Power Control of Microprocessors,” *IEEE J. Solid-State Circuits*, vol. 38, no. 11, Nov. 2003, pp. 1838-1845.
 - [9] V. Sundararajan and K. Parhi, “Low power synthesis of dual threshold voltage CMOS VLSI circuits,” in *IEEE Int. Symp. Low-Power Electronics and Design*, June 1999, pp. 139-144.
 - [10] S. Narendra, S. Borkar, V. De, D. Antoniadis, and A. Chandrakasan, “Scaling of stack effect and its application for leakage reduction,” in *Proc. Int. Symp. Low Power Electronic Design (ISLPED)*, 2001, pp. 195-200.
 - [11] S. Hanson, B. Zhai, M. Seok, B. Cline, K. Zhou, M. Singhal, M. Minuth, J. Olson, L. Nazhandali, T. Austin, D. Sylvester, and D. Blaauw, “Exploring Variability and Performance in a Sub-200-mV Processor,” *IEEE J. Solid-State Circuits*, vol. 43, no. 4, Apr. 2008, pp. 881-891.

- [12] T. Mori, Y. Ueda, N. Nonogaki, T. Terazawa, M. Sroka, T. Fujita, T. Kodaka, T. Mori, K. Morita, H. Arakida, T. Miura, Y. Okuda, T. Kizu, and Y. Tsuboi, "A Power, Performance Scalable Eight-Cores Media Processor for Mobile Multimedia Applications," *IEEE J. Solid-State Circuits*, vol. 44, no. 11, Nov. 2009, pp. 2957-2965.
- [13] B. Van Zeghbroeck, Principles of Semiconductor Devices, online at <http://ecee.colorado.edu/~bart/book/>, 2007.
- [14] Y. Taur and T. Ning, Fundamentals of modern VLSI devices, Cambridge University Press, New York, NY, 1998.
- [15] K. Gopalakrishnan, P. B. Griffin, and J. D. Plummer, "I-MOS: A novel semiconductor device with subthreshold slope lower than kT/q ," in *IEDM Tech. Dig.*, 2002, pp. 289-292.
- [16] E.-H. Toh, G. Wang, L. Chan, G.-Q. Lo, G. Samudra, and Y.-C. Yeo, "Strain and Materials Engineering for the I-MOS Transistor With an Elevated Impact-Ionization Region," *IEEE Trans. Electron Devices*, vol. 54, no. 10, Oct. 2007, pp. 2778-2785.
- [17] W. Choi, J. Song, J. Lee, Y. Park, and B.-G. Park, "70-nm Impact-Ionization Metal-Oxide-Semiconductor (I-MOS) Devices Integrated with Tunneling Field-Effect Transistors (TFETs)," in *IEDM Tech. Dig.*, 2005, pp. 955-958.
- [18] W. Jang, J. Lee, J.-B. Yoon, M.-S. Kim, J.-M. Lee, S.-M. Kim, K.-H. Cho, D.-W. Kim, D. Park, and W.-S. Lee, "Fabrication and characterization of a nanoelectromechanical switch with 15-nm-thick suspension air gap," *Appl. Phys. Lett.*, vol. 92, 2008, pp. 103110-1-103110-3.
- [19] K. Akarvardar, D. Elata, R. Parsa, G. C. Wan, K. Yoo, J. Provine, P. Peumans, R. T. Howe, and H.-S. P. Wong, "Design Considerations for Complementary Nanoelectromechanical Logic Gates," in *IEDM Tech. Dig.*, 2007, pp. 299-302.
- [20] N. Abel , R. Fritschi, K. Boucart, F. Casset, P. Ancey, and A. M. Ionescu, "Suspended-Gate MOSFET: bringing new MEMS functionality into solid-state transistor," in *IEDM Tech. Dig.*, 2005.
- [21] R. Nathanael, V. Pott, H. Kam, J. Jeon and T.-J. King Liu, "4-Terminal Relay Technology for Complementary Logic," in *IEDM Tech. Dig.*, 2009, pp. 223-226.
- [22] S.M. Sze, *Physics of Semiconductor Devices*, 2nd edn (New York: John Wiley & Sons, Inc., 1981).
- [23] J. Knoch and J. Appenzeller, "A novel concept for field-effect transistors – the tunneling carbon nanotube FET," *Device Research Conf. Digest*, 2005, pp. 153-156.
- [24] B. Streetman and S. Banerjee, *Solid State Electronic Devices*, 5th edn (New Jersey: Prentice Hall, Inc., 2000).
- [25] J.-P. Colinge, "Multiple-gate SOI MOSFETs," *Solid-State Elec.*, vol. 48, 2004, pp. 897-905.
- [26] S.M. Sze, *Physics of Semiconductor Devices*, 1st edn (New York: John Wiley & Sons, Inc., 1969).
- [27] Q. Zhang, W. Zhao, and A. Seabaugh, "Low-Subthreshold-Swing Tunnel Transistors," *IEEE Electron Device Lett.*, vol. 27, no. 4, 2006, pp. 297-300.
- [28] K. Bhuwarka, M. Born, M. Schindler, M. Schmidt, T. Sulima, and I. Eisele, "P-Channel Tunnel Field-Effect Transistors down to Sub-50 nm Channel Lengths," *Jap. J. Appl. Phys.*, vol. 45, no. 4B, 2006, pp. 3106-3109.
- [29] M. Born, K. Bhuwarka, M. Schindler, U. Abelein, M. Schmidt, T. Sulima, and I. Eisele, "Tunnel FET: A CMOS Device for High Temperature Applications," *15th International Conference on Microelectronics*, 2006, pp. 124-127.
- [30] *Atlas User's Manual*, Silvaco Int., Santa Clara, CA, May 26, 2006.
- [31] J. Slotboom and H. De Graaf, "Measurements of Bandgap Narrowing in Silicon Bipolar Transistors," *Solid State Electronics*, vol. 19, 1976, pp. 857-862.
- [32] J. Quinn, G. Kawamoto, and B. McCombe, "Subband Spectroscopy by Surface Channel Tunneling," *Surface Science*, vol. 73, 1978, pp. 190-196.
- [33] S. Banerjee, W. Richardson, J. Coleman, and A. Chatterjee, "A new three-terminal tunnel device," *IEEE Electron Device Lett.*, EDL-8, 1987, pp. 347-349.
- [34] E. Takeda, H. Matsuoka, Y. Igura, and S. Asai, "A band to band tunneling MOS device (B²T-MOSFET)," in *IEDM Tech. Dig.*, 1988, pp. 402-405.
- [35] T. Baba, "Proposal for Surface Tunnel Transistors," *Jpn. J. Appl. Phys.*, vol. 31, 1992, pp. L455-L457.
- [36] W. Reddick and G. Amaratunga, "Silicon surface tunnel transistor," *Appl. Phys. Lett.*, vol. 67, no. 4, 1995, pp. 494-496.

- [37] J. Koga and A. Toriumi, "Negative differential conductance in three-terminal silicon tunneling device," *Appl. Phys. Lett.*, vol. 69, no. 10, 1996, pp. 1435-1437.
- [38] W. Hansch, C. Fink, J. Schulze, and I. Eisele, "A vertical MOS-gated Esaki tunneling transistor in silicon," *Thin Solid Films*, vol. 369, 2000, pp. 387-389.
- [39] C. Aydin, A. Zaslavsky, S. Luryi, S. Cristoloveanu, D. Mariolle, D. Fraboulet, and S. Deleonibus, "Lateral interband tunneling transistor in silicon-on-insulator," *Appl. Phys. Lett.*, vol. 84, no. 10, 2004, pp. 1780-1782.
- [40] A. Grove and D. Fitzgerald, "The Origin of Channel Currents Associated with P+ Regions in Silicon," *IEEE Trans. Electron Devices*, vol. ED-12, no. 12, 1965, pp. 619-626.
- [41] J. Appenzeller, Y.-M. Lin, J. Knoch, and Ph. Avouris, "Band-to-Band Tunneling in Carbon Nanotube Field-Effect Transistors," *Phys. Rev. Lett.*, vol. 93, no. 19, 2004, pp. 196805-1-4.
- [42] J. Appenzeller, Y. Lin, J. Knoch, Z. Chen, P. Avouris, "Comparing Carbon Nanotube Transistors – The Ideal Choice: A Novel Tunneling Device Design," *IEEE Trans. Electron Devices*, vol. 52, no. 12, 2005, pp. 2568-2576.
- [43] K. Bhuwarka, J. Schulze, I. Eisele, "Performance Enhancement of Vertical Tunnel Field-Effect Transistor with SiGe in the dp+ Layer," *Jap. J. Appl. Phys.*, vol. 43, no. 7A, 2004, pp. 4073-4078.
- [44] K. Bhuwarka, M. Born, M. Schindler, M. Schmidt, T. Sulima, and I. Eisele, "P-Channel Tunnel Field-Effect Transistors down to Sub-50 nm Channel Lengths," *Jap. J. Appl. Phys.*, vol. 45, no. 4B, 2006, pp. 3106-3109.
- [45] P.-F. Wang, K. Hilsenbeck, T. Nirschl, M. Oswald, C. Stepper, M. Weis, D. Schmitt-Landsiedel, W. Hansch, "Complementary tunneling transistor for low power application," *Solid-State Elec.*, vol. 48, 2004, pp. 2281-2286.
- [46] T. Nirschl, S. Henzler, J. Fischer, M. Fukle, A. Bargagli-Stoffi, M. Sterkel, J. Sedlmeir, C. Weber, R. Heinrich, U. Schaper, J. Einfeld, R. Neubert, U. Feklmann, K. Stahrenberg, E. Ruderer, G. Georgakos, A. Huber, R. Kakoschke, W. Hansch, D. Schmitt-Landsiedel, "Scaling properties of the tunneling field effect transistor (TFET): Device and circuit," *Solid-State Elec.*, vol. 50, 2006, pp. 44-51.
- [47] T. Nirschl, M. Weis, M. Fulde, D. Schmitt-Landsiedel, "Correction to 'Revision of Tunneling Field-Effect Transistor in Standard CMOS Technologies,'" *IEEE Electron Device Lett.*, vol. 28, no. 4, 2007, p. 315.
- [48] Q. Zhang, W. Zhao, and A. Seabaugh, "Analytic Expression and Approach for Low Subthreshold-Swing Tunnel Transistors," *Device Research Conference (DRC)*, June 20-22, 2005, Santa Barbara, CA.
- [49] K. Boucart, A. M. Ionescu, "Double-Gate Tunnel FET With High-k Gate Dielectric," *IEEE Trans. Electron Devices*, vol. 54, no. 7, July 2007, pp. 1725-1733.
- [50] K. Boucart, A. M. Ionescu, "Threshold voltage in Tunnel FETs: physical definition, extraction, scaling and impact on IC design," in *Proc. ESSDERC*, 2007, pp. 299-302.
- [51] K. Boucart, A. M. Ionescu, "Length scaling of the Double Gate Tunnel FET with a high-K gate dielectric," *Solid-State Elec.*, vol. 51, no. 11-12, Nov.-Dec. 2007, pp. 1500-1507.
- [52] A. Verhulst, W. Vandenberghe, K. Maex, and G. Groeseneken, "Tunnel field-effect transistor without gate-drain overlap," *Appl. Phys. Lett.*, vol. 91, 2007, pp. 053102-1-3.
- [53] E.-H. Toh, G. Wang, G. Samudra, and Y.-C. Yeo, "Device physics and design of double-gate tunneling field-effect transistor by silicon film thickness optimization," *Appl. Phys. Lett.*, vol. 90, 2007, pp. 263507-1-3.
- [54] V. Nagavarapu, R. Jhaveri, and J. Woo, "The Tunnel Source (PNPN) n-MOSFET: A Novel High Performance Transistor," *IEEE Trans. Electron Devices*, vol. 55, no. 4, 2008, pp. 1013-1019.
- [55] M. Schlosser, K. Bhuwarka, M. Sauter, T. Zilbauer, T. Sulima, and I. Eisele, "Fringing-Induced Drain Current Improvement in the Tunnel Field-Effect Transistor With High-k Gate Dielectrics," *IEEE Trans. Electron Devices*, vol. 56, no. 1, 2009, pp. 100-108.
- [56] M. Vadizadeh and M. Fathipour, "Using Low-k Oxide for Reduction of Leakage Current in Double Gate Tunnel FET," in *Proc. ULIS*, 2009, pp. 301-304.
- [57] P. Patel, K. Jeon, A. Bowonder, and C. Hu, "A Low Voltage Steep Turn-Off Tunnel Transistor Design," in *Proc. SISPAD*, 2009, pp. 1-4.
- [58] C. Sandow, J. Knoch, C. Urban, Q.-T. Zhao, and S. Mantl, "Impact of electrostatics and doping concentration on the performance of silicon tunnel field-effect transistors," *Solid-State Elec.*, vol. 53, 2009, pp. 1126-1129.

- [59] D. Kazazis, P. Jannaty, A. Zaslavsky, C. Le Royer, C. Tabone, L. Clavelier, and S. Cristoloveanu, "Tunneling field-effect transistor with epitaxial junction in thin germanium-on-insulator," *Appl. Phys. Lett.*, vol. 94, 2009, pp. 263508-1-3.
- [60] K. Moselund, H. Ghoneim, M. T. Björk, H. Schmid, S. Karg, E. Lörtscher, W. Riess, and H. Riel, "Comparison of VLS grown Si NW tunnel FETs with different gate stacks," in *Proc. ESSDERC*, 2009, pp. 448-451.
- [61] Y. Khatami and K. Banerjee, "Steep Subthreshold Slope n- and p-Type Tunnel-FET Devices for Low-Power and Energy-Efficient Digital Circuits," *IEEE Trans. Electron Devices*, vol. 56, no. 11, 2009, pp. 2752-2761.
- [62] M. Luisier and G. Klimeck, "Atomistic Full-Band Design Study of InAs Band-to-Band Tunneling Field-Effect Transistors," *Electron Device Lett.*, vol. 30, no. 6, 2009, pp. 602-604.
- [63] Q. Zhang, T. Fang, H. Xing, A. Seabaugh, and D. Jena, "Graphene Nanoribbon Tunnel Transistors," *Electron Device Lett.*, vol. 29, no. 12, 2008, pp. 1344-1346.
- [64] F. Mayer, C. Le Royer, J.-F. Damlencourt, K. Romanjek, F. Andrieu, C. Tabone, B. Previtali, and S. Deleonibus, "Impact of SOI, Si1-xGeOI and GeOI substrates on CMOS compatible Tunnel FET performance," in *IEDM Tech. Dig.*, 2008, pp. 163-166.
- [65] M. Lee, J. Koo, E.-A. Chung, D.-Y. Jeong, Y.-S. Koo, and S. Kim, "Silicon nanowire-based tunneling field-effect transistors on flexible plastic substrates," *Nanotechnology*, vol. 20, 2009, pp. 455201-1-6.
- [66] A. Verhulst, W. Vandenberghe, K. Maex, S. De Gendt, M. Heyns, and G. Groeseneken, "Complementary Silicon-Based Heterostructure Tunnel-FETs With High Tunnel Rates," *Electron Device Lett.*, vol. 29, no. 12, 2008, pp. 1398-1401.
- [67] J. Knoch, "Optimizing tunnel FET performance - Impact of device structure, transistor dimensions and choice of material," in *Proc. VLSI-TSA*, 2009, pp. 45-46.
- [68] T. Krishnamohan, D. Kim, S. Raghunathan, and K. Saraswat, "Double-Gate Strained-Ge Heterostructure Tunneling FET (TFET) With Record High Drive Currents and <60mV/dec Subthreshold Slope," in *IEDM Tech. Dig.*, 2008, pp. 947-949.
- [69] O. Nayfeh, J. Hoyt, and D. Antoniadis, "Strained SiGe/Si Band-to-Band Tunneling Transistors: Impact of Tunnel-Junction Germanium Composition and Doping Concentration on Switching Behavior," *IEEE Trans. Electron Devices*, vol. 56, no. 10, 2009, pp. 2264-2269.
- [70] K. Boucart, W. Riess, A. M. Ionescu, "Lateral Strain Profile as Key Technology Booster for All-Silicon Tunnel FETs," *Electron Device Lett.*, vol. 30, iss. 6, 2009, pp. 656 - 658.
- [71] J. Knoch, S. Mantl, J. Appenzeller, "Impact of the dimensionality on the performance of tunneling FETs: Bulk versus one-dimensional devices," *Solid-State Elec.*, vol. 51, no. 4, 2007, pp. 572-578.
- [72] W. Vandenberghe, A. Verhulst, G. Groeseneken, B. Sorée, and W. Magnus, "Analytical Model for Point and Line Tunneling in a Tunnel Field-Effect Transistor," in *Proc. SISPAD*, 2008, pp. 137-140.
- [73] A. Verhulst, B. Sorée, D. Leonelli, W. Vandenberghe, and G. Groeseneken, "Modeling the single-gate, double-gate, and gate-all-around tunnel field-effect transistor," *J. Appl. Phys.*, vol. 107, 2010, pp. 024518-1-8.
- [74] M. Fulde, A. Heigl, M. Weis, M. Wirnshofer, K. v. Arnim, T. Nirschl, M. Sterkel, G. Knoblinger, W. Hansch, G. Wachutka, and D. Schmitt-Landsiedel, "Fabrication, Optimization, and Application of Complementary Multiple-Gate Tunneling FETs," in *Proc. INEC*, 2008, pp. 579-584.
- [75] H. Kam, T.-J. King-Liu, E. Alon, and M. Horowitz, "Circuit-Level Requirements for MOSFET-Replacement Devices," in *IEDM Tech. Dig.*, 2008.
- [76] S. Koswatta, M. Lundstrom, and D. Nikonov, "Performance Comparison Between p-i-n Tunneling Transistors and Conventional MOSFETs," *IEEE Trans. Electron Devices*, vol. 56, no. 3, 2009, pp. 456-465.
- [77] D. Kim, Y. Lee, J. Cai, I. Lauer, L. Chang, S.J. Koester, D. Sylvester, D. Blaauw, "Low power circuit design based on heterojunction tunneling transistors (HETTs)," in *Proceedings of International Symposium on Low Power Electronics and Design*, 2009, pp. 219-224.
- [78] S. Mookerjee, R. Krishnan, S. Datta and V. Narayanan, "Effective Capacitance and Drive Current for Tunnel FET (TFET) CV/I Estimation," *IEEE Trans. Electron Devices*, vol. 56, no. 9, 2009, pp. 2092-2098.

Chapter 3

Tunnel FET Optimization

Before investigating the unique properties of Tunnel FETs or looking in detail at their behavior, it is important to carefully choose all device parameters such that characteristics are optimized, and to show why each choice was made as it was. In this chapter, Tunnel FET optimization is explored, considering the following parameters: single or double gate, doping levels in the source, drain, and intrinsic regions, different gate dielectric materials (SiO_2 and high- k), and silicon body thickness. 2D cross-sections of an optimized Tunnel FET are then presented for electric field, electric potential, and current flowlines, the last of which shows the two-dimensionality of current flow inside this type of device. Energy band diagrams in the y -direction (through the depth of the device) are also presented in order to aid understanding of the current flow. Finally, one last possibility for optimization is discussed, reducing the band gap at the tunnel junction in order to improve on-current. In order to keep off-current and subthreshold swing low, the band gap must remain large (that of unstrained silicon) at the drain-intrinsic region junction.

3.1 Tunnel FET parameter optimization

3.1.1 Device structure and operation

The Tunnel FETs optimized in this first section of the chapter are n-type devices whose dielectric covers the source, intrinsic region, and drain, as shown in Fig. 3.1. The single-gate and double-gate structures which will be compared in section 3.1.2 are shown, though the rest of the chapter will focus on the double-gate device. All simulations for this section were done in Silvaco Atlas, version 5.11.24.C. In this section, the doping profiles at the junctions are perfectly abrupt (0.5 nm, determined by the meshing) for all simulations. The energy band diagrams for this double-gate device in the off-state and on-state are shown in Fig. 3.2.

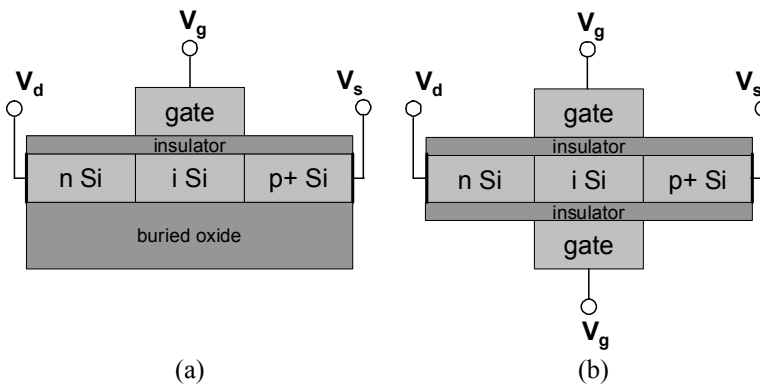


Figure 3.1: Structure of simulated n-type Tunnel FETs: (a) single-gate and (b) double-gate. SiO_2 and high-k gate dielectrics were studied. $t_{\text{dielectric}} = 3 \text{ nm}$ (physical) and $L_{\text{intrinsic}} = 50 \text{ nm}$. Drain doping (n) = $5 \times 10^{18} \text{ atoms/cm}^3$ and source doping ($p+$) = $10^{20} \text{ atoms/cm}^3$. Junctions were perfectly abrupt.

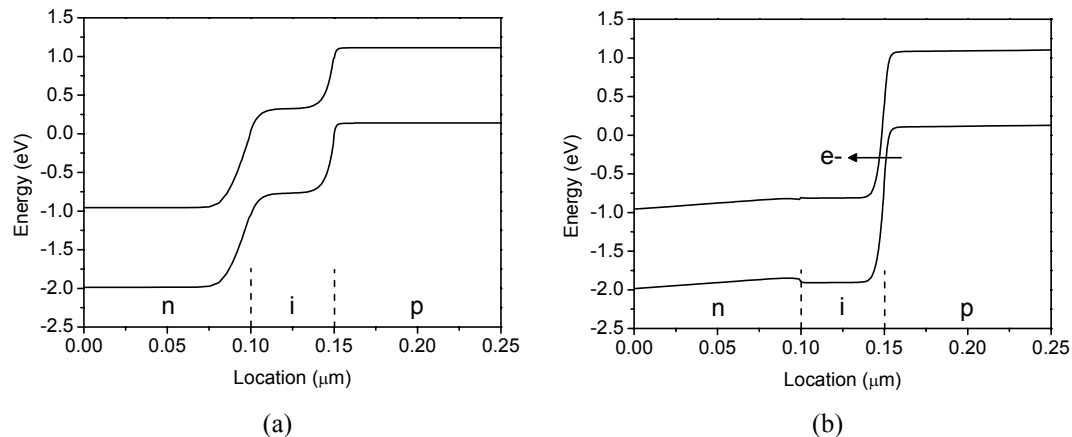


Figure 3.2: Energy band diagrams for the Tunnel FET structure shown in Fig. 1(b). (a) Off-state, $V_{\text{DS}} = 1 \text{ V}$ and $V_{\text{GS}} = 0 \text{ V}$. In this state, the only current is p-i-n diode leakage current. (b) On-state, $V_{\text{DS}} = 1 \text{ V}$ and $V_{\text{GS}} = 1.8 \text{ V}$. In this state, the energy barrier is thin enough that electrons can tunnel from the valance band of the $p+$ region to the conduction band of the intrinsic region.

3.1.2 Double gate

When changing from a single-gate design to a double-gate design, Tunnel FETs will benefit from the added gate, such that the current will be at least doubled. In this way, the on-current is boosted, while the off-current, still in the fA or pA range, increases by the same factor but remains extremely low. This can be seen in Fig. 3.3, for a device with $t_{\text{Si}} = 50 \text{ nm}$, simulated both as single-gate on SOI and as double-gate. The left axis shows a logarithmic scale, and the right axis a linear scale. The

solid curve shows the single-gate Tunnel FET's current multiplied by two, and the curve is coincident with the double-gate simulated results. When the device body is thin enough, however, the electrostatic control of the tunnel junction by the gate will be improved with a double-gate configuration, and the current will more than double, as shown in Fig. 3.4 for a device with $t_{\text{Si}} = 10$ nm. This solid line shows that the current for the double-gate device is much larger than that of the single-gate current doubled; the double gate gives an advantage of about two decades of current in sub-threshold, and about one decade for high V_{GS} (around 1.2 V). The broken curve corresponding to the right axis shows the ratio between the current for the double-gate device and the current for the single gate device, $I_{\text{DS,DG}}/I_{\text{DS,SG}}$. This effect of enhanced current is analogous to the volume inversion caused by the strong electrostatic control of the gates in ultrathin SOI MOSFETs [1]. The subthreshold swing and threshold voltage are also reduced with the double-gate configuration when $t_{\text{Si}} = 10$ nm.

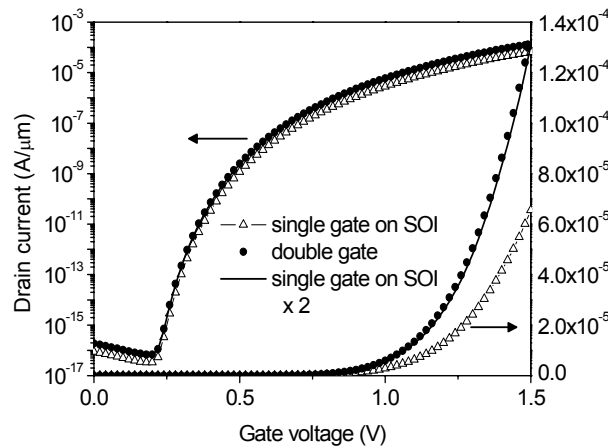


Figure 3.3: Transfer characteristics for Tunnel FETs with $t_{\text{Si}} = 50$ nm. The double-gate device has twice the current of the single-gate device on SOI. Left axis is log scale, and right axis shows the same data on a linear scale. $V_{\text{DS}} = 1$ V.

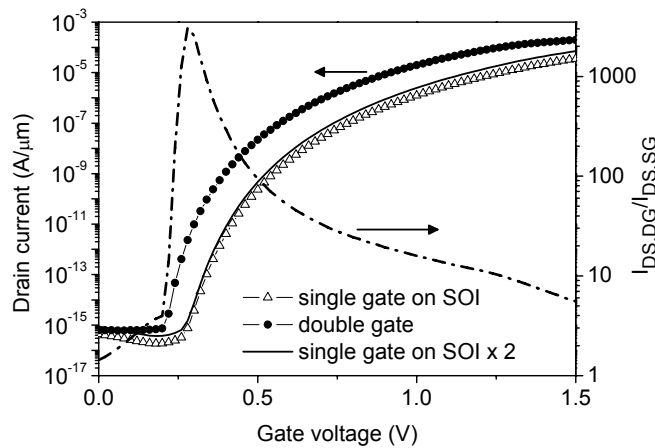


Figure 3.4: Transfer characteristics for Tunnel FETs with $t_{\text{Si}} = 10$ nm (left axis), and ratio between the double-gate and single-gate devices' current (right axis). $V_{\text{DS}} = 1$ V.

3.1.3 Doping levels in the source, drain, and intrinsic regions

The doping levels of the Tunnel FET source and drain must be carefully optimized in order to maximize on-current and minimize off-current, while still considering what is realistic for device fabrication. Different levels of p-doping for the source are shown in Fig. 3.5. The source doping has a large effect on the on-current level, since the tunneling takes place between the source and the intrinsic region. The highest-possible source doping level is desirable for optimized Tunnel FET behavior. In the rest of this chapter, a level of 1×10^{20} atoms/cm³ has been used, because it is a high

doping level, yet still realistic in terms of a design to be fabricated. Note that in all of chapter 3 except section 3.3, the Tunnel FETs had perfectly abrupt junctions, so the doping drops from the source level of 10^{20} atoms/cm³ p-type to the intrinsic level of 10^{17} atoms/cm³ n-type across one mesh point (mesh spacing = 0.5 nm).

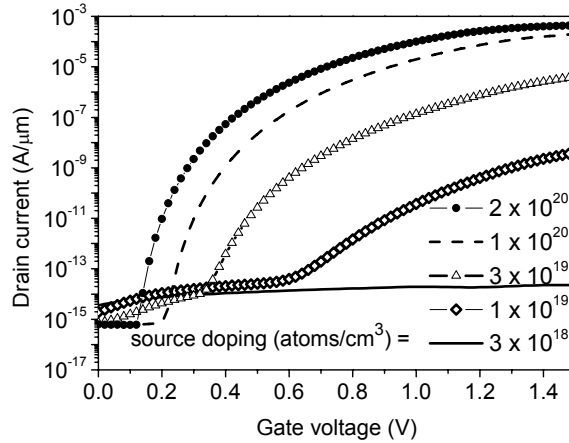


Figure 3.5: I_{DS} - V_{GS} for varying source doping levels. A very high source doping level is advantageous because it leads to a smaller swing and higher on-current. $V_{DS} = 1$ V.

The drain doping level is important for a different reason, and its effect is most clearly seen in the off-region of the I_{DS} - V_{GS} curve, as seen in Fig. 3.6. When the source and drain doping levels are equal (see the curve for drain doping = 1×10^{20} atoms/cm³), the Tunnel FET has ambipolar characteristics, which is not desirable for circuit designers. Lowering the drain doping not only suppresses the ambipolarity, but also lowers the off-current and “uncovers” a section of the I-V curve which has a lower subthreshold swing. A low drain doping is desirable, therefore, for an optimized device, but ideally, the doping should be high enough that contact formation is not too difficult. For the remainder of the simulated work shown in this thesis, a drain doping level of 5×10^{18} atoms/cm³ has been used. This doping level would require a silicide, or some similar technique, in order to create a good Ohmic contact on the drain side.

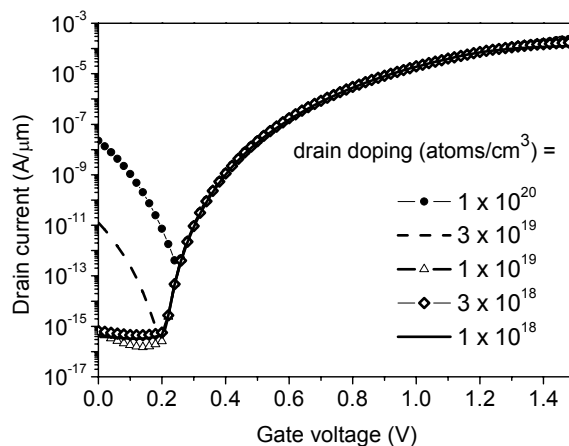


Figure 3.6: I_{DS} - V_{GS} for varying drain doping levels. The ideal drain doping level is low enough to suppress ambipolarity. $V_{DS} = 1$ V.

The last doping level to investigate is that of the intrinsic region. This doping level has little effect on Tunnel FET transfer characteristics until the doping level is high, as shown in Fig. 3.7. The only curve that is different from the others is for an n-type doping of 1×10^{18} atoms/cm³. As long as the doping level is less than 10^{17} atoms/cm³, the exact concentration and type is not important, and does not impact the gate’s ability to control the energy bands and the band-to-band tunneling at the tunnel junction. For the rest of this section, the intrinsic region doping is 10^{17} atoms/cm³.

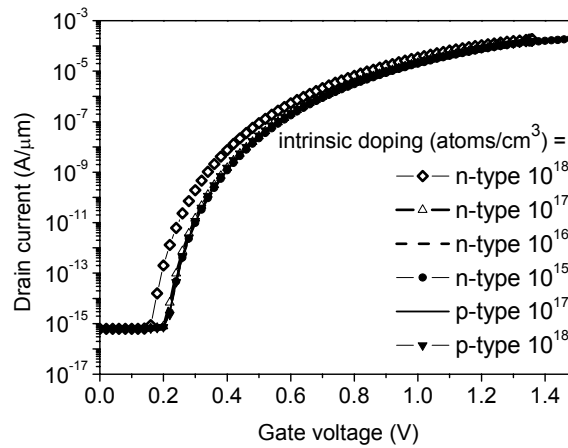
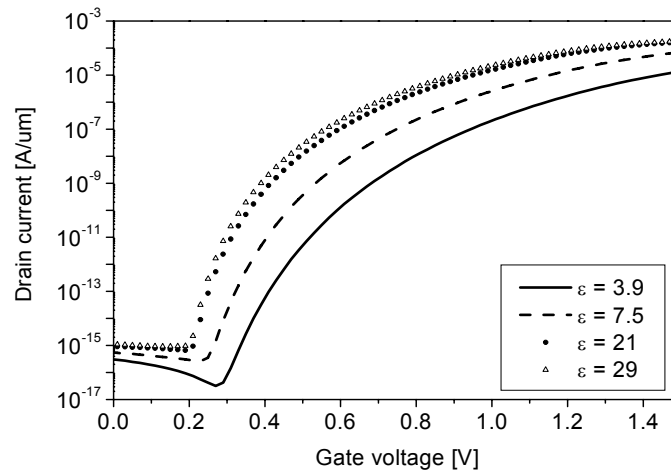


Figure 3.7: I_{DS} - V_{GS} for varying intrinsic-region doping levels. The doping type and level of this region plays little role in device characteristics until the doping levels become high. With an n-type doping of 10^{18} atoms/cm³, the transfer characteristics finally change, but now we have changed our device structure from a p-i-n gated diode to a p-n gated diode, which would be optimized differently and is outside the scope of this thesis. $V_{DS} = 1$ V.

3.1.4 High-k gate dielectric

An improved on-current and decreased subthreshold swing can be obtained by the careful choice of a gate dielectric. As shown in Fig. 3.8(a), current increases as the gate dielectric constant increases. Here, Si_3N_4 and two high-k dielectrics with dielectric constants of 21 and 29 are compared with SiO_2 , all with a physical thickness of 3 nm. In addition to improved I_{on} , both the point and average subthreshold swing improve as the result of the better gate coupling given by a high-k dielectric. The off-current is less than 1 fA for all materials.



(a)

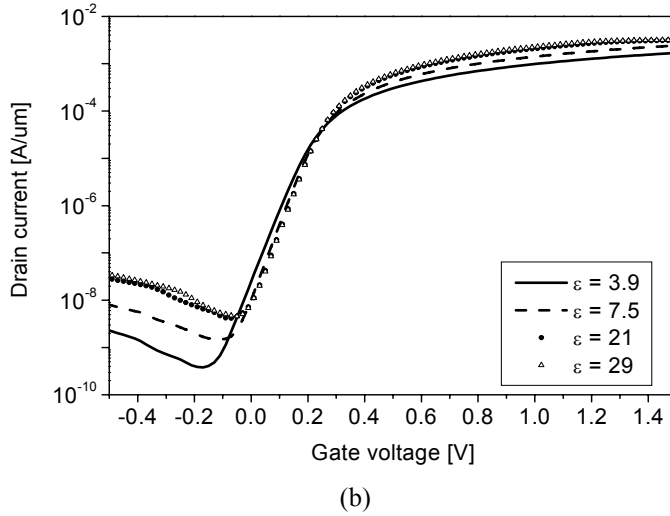


Figure 3.8: (a) DG Tunnel FET characteristics for various gate dielectrics. $\epsilon = 3.9$ corresponds to SiO_2 , $\epsilon = 7.5$ to Si_3N_4 , and $\epsilon = 21$ and $\epsilon = 29$ to high-k dielectrics such as HfO_2 and ZrO_2 . (b) The characteristics of a simplified double-gate conventional nMOSFET for various gate dielectrics. Junctions were ideally abrupt, as for the Tunnel FET, with source and drain doped to 10^{20} and the p-type body doped to 10^{17} . For both (a) and (b), $L = 50$ nm, $t_{\text{dielectric}} = 3$ nm, $t_{\text{Si}} = 10$ nm, and $V_{\text{DS}} = 1$ V.

The on-current of a Tunnel FET does not increase merely proportionally to the increase in the gate capacitance, as it would for a conventional MOSFET. A simplified conventional MOSFET 2D structure has been designed for numerical simulation in order to show the difference between the two (Fig. 3.8(b)). One striking difference between Figs. 3.8(a) and (b) is that the Tunnel FET's subthreshold swing continues to improve as the gate dielectric permittivity increases. In contrast, the swing for the MOSFET hits its 60 mV/decade limit at room temperature and cannot improve further.

Fig. 3.9 shows the linear relationship between $\log(I_{\text{DS}})$ and $\epsilon_{\text{ox}}^{-0.5}$, which supports what we expected to see about the dependence of the source-drain current on the gate dielectric permittivity, from Eqs. 2.12 and 2.13.

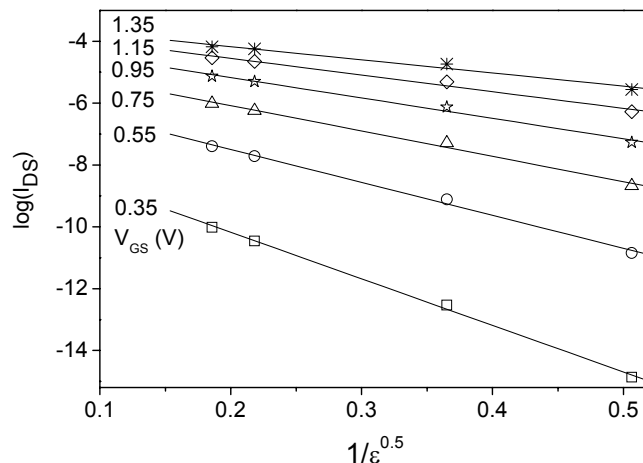


Figure 3.9: $\log(I_{\text{DS}})$ versus $1/\epsilon_{\text{ox}}^{0.5}$; the linearity of the plot is in good agreement with the proposed modeling in Eqs. 2.12 and 2.13. $V_{\text{DS}} = 1$ V.

While high-k dielectrics have advantages for device characteristics, when put directly in contact with a silicon channel, they can lead to defects at the semiconductor/dielectric interface. Although Tunnel FETs might be less sensitive to changes in channel mobility than MOSFETs since the transport through the tunnel junction dominates over any scattering in the channel, standard CMOS fabrication

techniques require an interfacial layer between the high-k dielectric and the silicon channel. A Tunnel FET with a more CMOS-compatible dielectric layer has been simulated, and the resulting I_{DS} - V_{GS} curve can be seen in Fig. 3.10. The simulated device had 1 nm of silicon oxynitride at the silicon surface, and 2 nm of a high-k with $\epsilon = 29$. It is clear that even with an interfacial layer, the subthreshold swing and on-current are very much improved over the device with an SiO_2 gate dielectric. In addition, Fig. 3.10 shows that with this device structure in which the gate dielectric covers the source and drain as well as the intrinsic region, as in Fig. 3.1, the improvement in characteristics with increasing dielectric permittivity saturates. Even going to a permittivity as high as 200, the on-current does not increase significantly over that of the device with $\epsilon = 29$, and the off-current and subthreshold swing are not quite as good. The situation is not the same when the gate dielectric is aligned with the intrinsic region, as will be seen in Chapter 6.

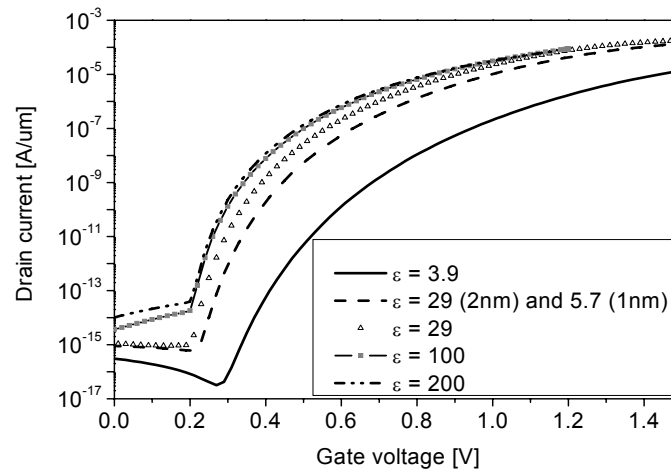


Figure 3.10: DG Tunnel FET characteristics for structures with 3 nm of SiO_2 , with 2 nm of high-k dielectric ($\epsilon_{\text{dielectric}} = 29$) and a 1 nm interfacial layer of silicon oxynitride ($\epsilon_{\text{dielectric}} = 5.7$), and with 3 nm of high-k dielectrics having $\epsilon_{\text{dielectric}} = 29, 100$, and 200 . $V_{DS} = 1$ V. The improvement in characteristics saturates with increasing dielectric permittivity, when the gate dielectric covers the drain, intrinsic region, and source regions.

High-k dielectrics bring additional challenges such as the limitations of soft and hard dielectric breakdown. Depending on the characteristics of fabricated high-k dielectric layers, it may be necessary to limit applied gate voltages more than what is reported here. For example, depending on whether the structure of a HfO_2 layer is more tetragonal or cubic, the breakdown field could be 3.9 or 6.7 MV/cm, leading to a breakdown voltage of $V_{GS} = 1.17$ V or 2.01 V [2]. Of course, these voltage numbers would change depending on the high-k dielectric thickness and/or the existence of an interfacial layer. Though the current simulations aim to stay within optimistic limits, a re-optimization of the design will be needed once the parameters of fabricated materials are known. In general, a high-quality gate dielectric with the highest possible permittivity is optimal for best Tunnel FET performance.

3.1.5 Thin film structure

Silicon-on-insulator (SOI) and silicon-on-nothing (SON) fabrication technology are two fabrication methods currently used to create double-gate devices. Both of these techniques are commonly used on thin films, down to several nanometers thick. Tunnel FETs are sensitive to this thickness, which influences the shape of its I_{DS} - V_{GS} curve, as shown in Fig. 3.11.

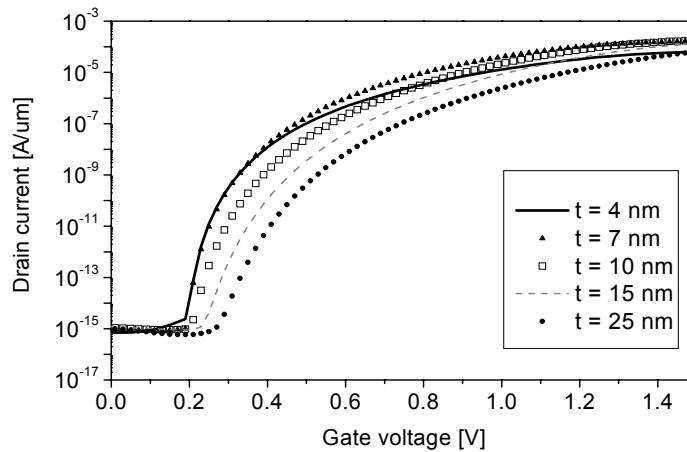


Figure 3.11: Double-gate Tunnel FET I_{DS} - V_{GS} characteristics for various silicon body thicknesses. $L_{intrinsic} = 30$ nm, $t_{dielectric} = 3$ nm, and $\epsilon_{dielectric} = 25$. $V_{DS} = 1$ V.

Several trends can be seen in this figure. First, the off-currents are practically independent of the thickness. As the film gets thinner than 7 nm, on-current starts to drop, possibly due to the reduced cross-sectional area available for current flow. This reduction of on-current happens in Tunnel FETs whose dielectric layer covers the source and drain, but not in those whose dielectric is aligned to the intrinsic region (at least not in the studied voltage range), as will be seen in Chapter 6. Toh, et al. [3] observed the same trends in their Synopsis simulations of double-gate Tunnel FETs, as shown in Fig. 3.12. For a full understanding of why this phenomenon takes place, more investigation is needed.

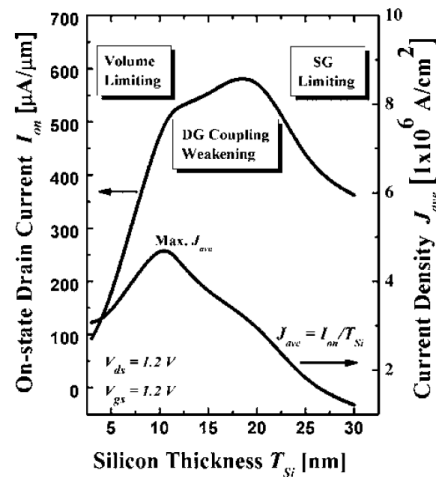


Figure 3.12: On-state drain current (left axis) and current density (right axis) vs. silicon body thickness of a double-gate Tunnel FET, from [3]. Maxima can be seen in both total current and current density. For the thickest devices, the effect is that seen in section 3.1.2, where a device above some critical thickness does not benefit from increased electrostatic gate control from a double gate. For the thinnest devices, the reduced cross-sectional area seems to limit the current flow.

Since the on-current has a maximum, the ratio I_{on}/I_{off} will also have a maximum when plotted against silicon layer thickness. Fig. 3.13 shows that for the particular device design studied here, this optimum value occurs when t_{Si} is between 7 and 8 nm, depending on the value chosen for V_{DD} , where I_{on} is taken at $V_{GS} = V_{DD}$. The maximum ratio is about 2×10^{11} , and as a comparison, the optimized asymmetrical DG MOSFET from [4] has an I_{on}/I_{off} ratio of 10^6 with I_{on} taken at $V_{GS} = 1.5$ V.

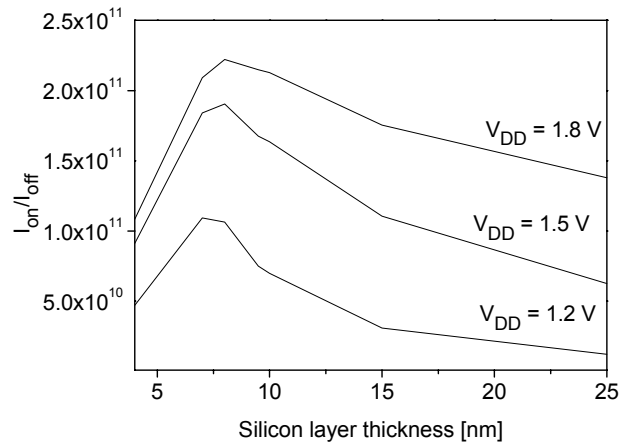


Figure 3.13: The ratio I_{on}/I_{off} as a function of the silicon layer thickness, for $\epsilon_{dielectric} = 29$. A maximum, the optimum point, occurs in the thickness range of 7 to 8 nm, depending on the value of V_{DD} used. 1 V was applied to the drain.

It is clear in Fig. 3.11 that both the average and point values of subthreshold swing decrease as thickness decreases, as shown in more detail in Fig. 3.14. The subthreshold swing also decreases as the gate dielectric permittivity increases. While the point swing is lower than the 60 mV/decade limit for MOSFETs for all dielectrics and thicknesses simulated, the average swing (as defined in Fig. 2.20) is lower than this limit only for a high-k dielectric with a constant of 29 and a silicon body thickness of less than 10 nm.

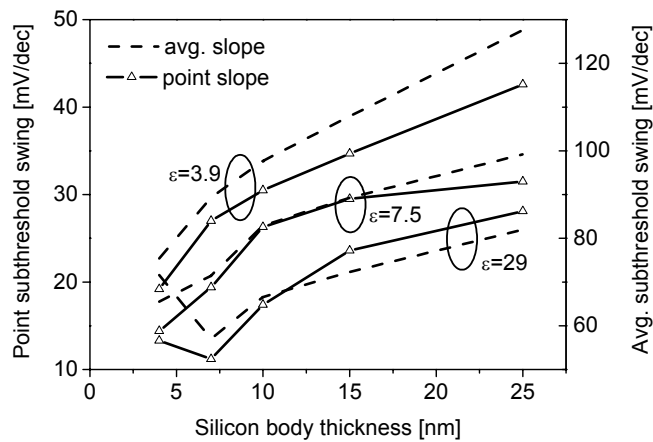


Figure 3.14: Point subthreshold swing (left y-axis) and average subthreshold swing (right y-axis) as a function of silicon layer thickness for different gate dielectric constants. $L_{intrinsic} = 50$ nm and $t_{dielectric} = 3$ nm. $V_{DS} = 1$ V.

Clearly, these values of the average swing depend upon our chosen definition for the threshold voltage. A lower constant-current value would lower the threshold voltage, and in turn would advantageously lower the average swing values. It is worth mentioning, however, that qualitatively, all the trends of the curves remain the same.

The relationship between the drain current and the body thickness can be seen in Chapter 2, Eqs. 2.12 and 2.13: $\log(I_{DS})$ is linearly dependent on $t_{Si}^{-0.5}$. Fig. 3.11 shows that this relation may hold true for devices thicker than about 7 nm with this device design where the gate dielectric covers the source and drain regions, but when the silicon body is too thin, the drain current could be limited by the reduced body thickness. Fig. 3.15 shows the extraction of the relationship between $\log(I_{DS})$ and $t_{Si}^{-0.5}$ for various values of the gate voltage.

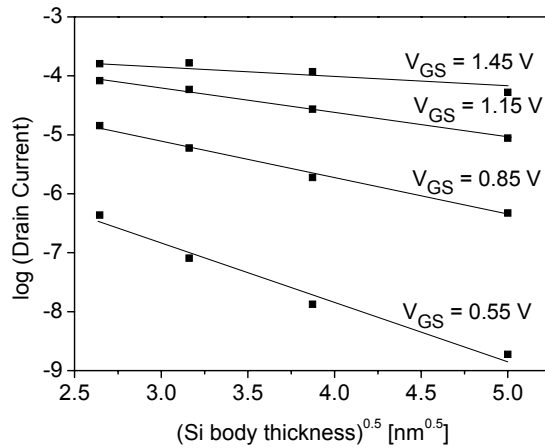


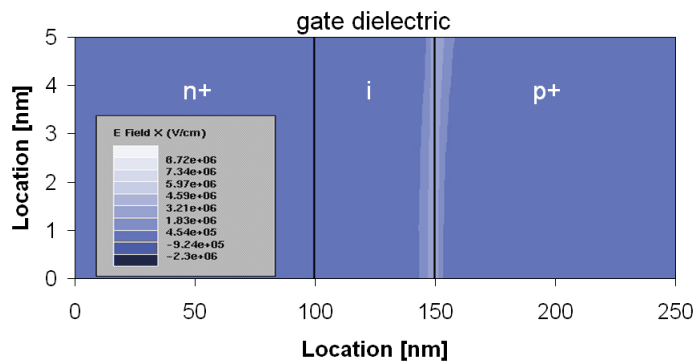
Figure 3.15: $\log(I_{DS})$ versus $t_{Si}^{0.5}$ at different gate voltages; the relationship is approximately linear, in agreement with the proposed modeling of Chapter 2, Eqs. 2.12 and 2.13. $V_{DS} = 1$ V.

The conclusion to be drawn from this study of the influence of Tunnel FET body thickness on device characteristics is that it is beneficial to use a thin body, especially for double-gate benefits as shown in section 3.1.2, but depending on the specific device design, it may be necessary to stay above some critical thickness so that current will not decrease, as previously discussed. The optimal thickness shown here, 7 nm, would vary somewhat depending on other device parameters. In subsequent simulations, a body thickness of $t_{Si} = 10$ nm is used.

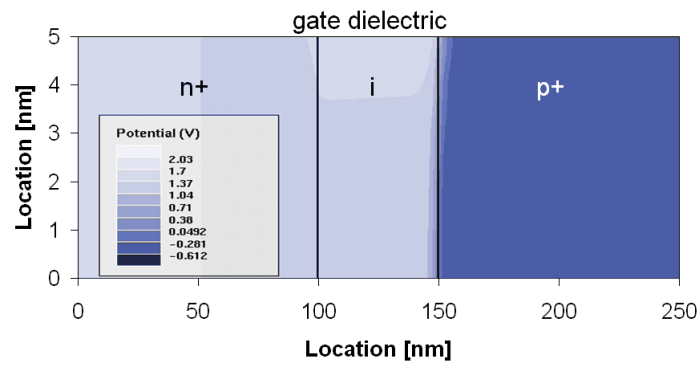
3.2 2D Tunnel FET simulations

All simulations carried out in Silvaco Atlas were 2D, and it is informative to look at vertical cross-sections of the energy bands of the Tunnel FET, as well as contour plots in two dimensions, in order to understand the functioning of the device. All diagrams are shown for one half of the double-gate devices, with the gate dielectric at the top and the center of the body at the bottom. The devices are always symmetrical around the center of the body.

Looking first at the x-direction component of the electric field across a device in the on-state (Fig. 3.16(a)), we see that the electric field is close to zero nearly everywhere. Between the intrinsic and p+ regions, where the tunneling takes place, we see a high positive field throughout the depth of the device. In the potential contour plot for this same device in the same state (Fig. 3.16(b)), we see that the potential drops abruptly at the tunnel junction, and once again, this holds true for the entire device depth, not just at the surface.



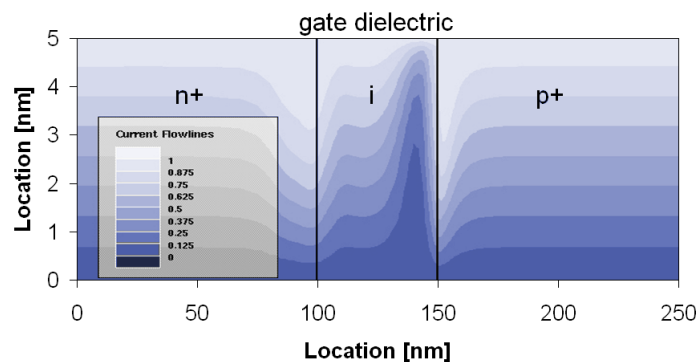
(a)



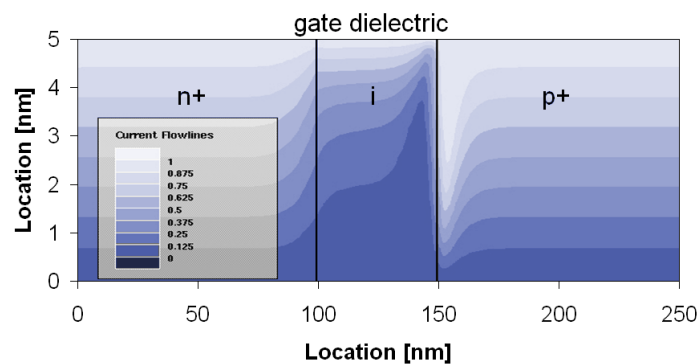
(b)

Figure 3.16: (a) Contour plot of the x-component of the electric field in one half of a double-gate Tunnel FET biased with $V_{DS} = 1$ V and $V_{GS} = 1.8$ V (in overdrive) with $t_{\text{dielectric}} = 3$ nm and $\epsilon_{\text{dielectric}} = 29$. (b) Contour plot of the potential of the same device under the same bias conditions.

In the diagrams of current flowlines, shown at the threshold voltage where $I_{DS} = 10^{-7}$ A/ μm (Fig. 3.17(a)) and in overdrive (Fig. 3.17(b)), it is clear that the current does not stay close to the gate dielectric as in a MOSFET. As the electrons move from right to left (source to drain) in the Tunnel FET, they move parallel to the interface through most of the source, then move away from the dielectric interface at about the location of the tunnel junction, and then, attracted by the positive voltage on the gate, flow closer to the interface before spreading back out and passing through the drain parallel to the interface, as they were in the source. (Electrical contacts are on the sides of the source and drain.)



(a)

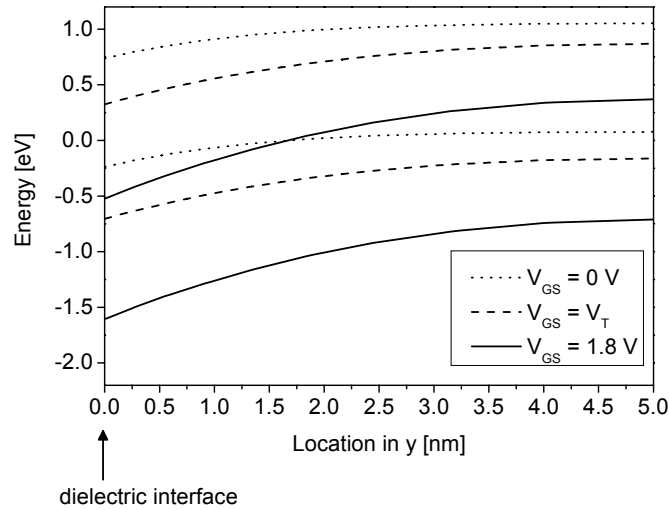


(b)

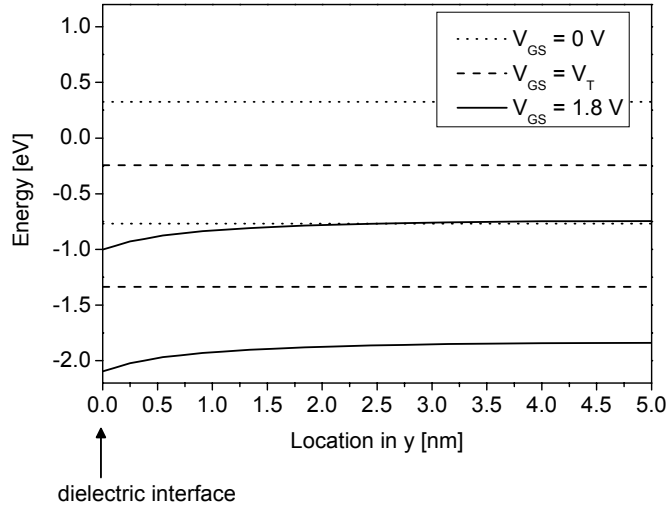
Figure 3.17: Current flowlines for the same Tunnel FET as in Fig. 3.16, with (a) $V_{GS} = V_T$ (obtained using the constant current technique) and (b) $V_{GS} = 1.8$ V. For both, $V_{DS} = 1$ V.

Inspection of some vertical energy band cross-sections can help in understanding this current flow. Fig. 3.18(a) shows the energy bands taken vertically at the junction between the source and intrinsic

regions, where the band-to-band tunneling takes place. The energy is lower at the dielectric surface than deep in the body, especially at high V_{GS} , as can also be noticed in the potential contours where we see that the tunnel junction (the abrupt change in potential) is more in the source near the dielectric, and more in the intrinsic region deeper in the body. So just at this junction, electrons will want to go toward the dielectric, to lower energy. Fig. 3.18(b) shows the energy bands cutting vertically through the very center of the intrinsic region. Here, the bands are nearly flat, and the electrons no longer stay as close to the gate dielectric, as evidenced by the current flowlines in Fig. 3.17.



(a)



(b)

Figure 3.18: Cross-sections of the conduction bands and valence bands of the same device as in Figs. 3.16 and 3.17, taken vertically from the dielectric interface through the body. (a) Cross-section at $x = 150$ nm, just at the junction between the intrinsic and p+ regions, where tunneling takes place. (b) Cross-section at $x = 125$ nm, at the center of the intrinsic region. For all figures, $V_{DS} = 1$ V.

The current flow pattern could give an advantage over conventional MOSFETs in terms of the effects of surface roughness on device characteristics. Tunnel FETs should be less affected by variations in mobility, due to the current being less confined to the surface under the gate.

3.3 One final optimization: the band gap

The optimizations carried out in section 3.1 were for an all-silicon device. Since the most frequent criticism of Tunnel FETs as a conventional MOSFET replacement is that their on-current is too low, the last optimization to be shown here is a very important one: the band gap at the tunnel junction. Reducing the band gap at the source-intrinsic region junction has the potential to raise the on-current by several orders of magnitude, thereby making them directly competitive with highly-scaled MOSFETs, as already mentioned in the state-of-the-art section in Chapter 2.

Many groups around the world are looking into band gap engineering for Tunnel FETs. Published experimental results for devices with reduced band gaps look promising. A team from Stanford University attained $300 \mu\text{A}/\mu\text{m}$ from their strained-germanium Tunnel FETs, which had a band gap of 0.4 eV, biased with $V_{\text{DS}} = 3 \text{ V}$ and $V_{\text{GS}} = 4 \text{ V}$ [5]. A group from CEA-LETI in France measured $4 \mu\text{A}/\mu\text{m}$ from p-type Tunnel FETs on GeOI, with $V_{\text{DS}} = -0.8 \text{ V}$ and $V_{\text{GS}} - V_{\text{BTBT}} = -2 \text{ V}$ [6]. Simulations also show promising characteristics from nanotubes and nanowires whose band gaps are smaller than that of silicon. M. Lundstrom's group at Purdue University predicts $10 \mu\text{A}$ from carbon nanotubes whose band gap is 0.667 eV, biased at the low voltages of $V_{\text{DS}} = 0.3 \text{ V}$ and $V_{\text{GS}} = 0.7 \text{ V}$ [7]. And finally, J. Knoch at TU Dortmund University in Germany published simulations that showed $1 \mu\text{A}$ from AlGaSb nanowire Tunnel FETs biased with $V_{\text{DS}} = -0.4 \text{ V}$ and $V_{\text{GS}} = -0.45 \text{ V}$ [8].

The Tunnel FETs presented in this section were simulated with Silvaco Atlas version 5.13.16.C with varying strain profiles all having their maxima at the tunnel junction, as shown in Fig. 3.19. They are doped with 5×10^{18} , 10^{16} , and 10^{20} atoms/cm³ in the n-type drain, slightly n-type intrinsic region, and p+-type source respectively, and the tunnel junction doping abruptness is 2 nm/decade. The simulated devices have a silicon thickness of 10 nm, a dielectric (SiO₂) thickness of 1.2 nm, and a gate length of 50 nm. The metal gates have a midgap work function. The nonlocal band-to-band tunneling model was used, along with band gap narrowing and the quantum model.

The key parameter of the devices investigated in this section is the asymmetrical uniaxial tensile strain which was simulated as a band gap narrowing at the source-intrinsic region junction, with 1 to 6 GPa of stress corresponding to 0.05 to 0.3 eV of band gap shrinkage [9]. Since this lateral strain profile, as illustrated in Fig. 3.19, is incorporated in simulations simply by changing the band gap along the intrinsic region of the device, the results could be generalized to apply to heterojunction Tunnel FETs with a reduced band gap at the tunnel junction. For that reason, this section can be seen not only as an exploration of how a lateral strain profile can improve Tunnel FET characteristics, but also as a final optimization to be carried out in order to have superior switch behavior. In this section, high stress and low band gap will be referred to interchangeably, and the reader must keep in mind that any conclusions drawn for Tunnel FETs with high stress at the tunnel junction can be extended to heterostructure Tunnel FETs with a low band gap material at the source side. While 6 GPa of stress is unrealistic for silicon (2 or 3 GPa is more realistic, as will be mentioned in the feasibility section later in this chapter), the corresponding band gap reduction of 0.3 eV is reasonable, or even small, when considering heterostructures.

We chose to only investigate the influence of band gap shrinkage on Tunnel FET characteristics, and for this reason, the carrier mobility was kept constant. This is a worst case scenario, since uniaxial tensile strain causes a reduction in carrier mass in silicon [10], so Tunnel FET on-current should increase even more than is shown in this section if uniaxial tensile strain is present at the tunnel junction.

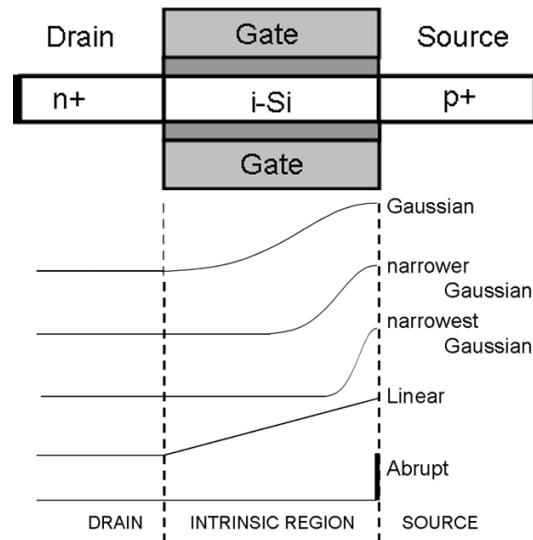


Figure 3.19: Cross section of simulated all-silicon double-gate Tunnel FET and lateral strain profiles (arbitrary units) applied on silicon films with a thickness of 10 nm, investigated in this section.

As already seen in Chapter 2, the critical factor that determines the on-current of a Tunnel FET is the tunneling barrier width at the tunnel junction in the on-state. In Fig. 3.20, the energy band diagram is shown along the channel for a reference silicon Tunnel FET (band gap = 1.1 eV) and for a Tunnel FET with a lateral strain profile in the channel (with a band gap minimum of 0.9 eV, corresponding to 4 GPa of stress, at the tunnel junction) in on-conditions. A narrowing of the barrier of 13% is seen in the case with the lateral profile, which increases tunneling probability and hence I_{on} .

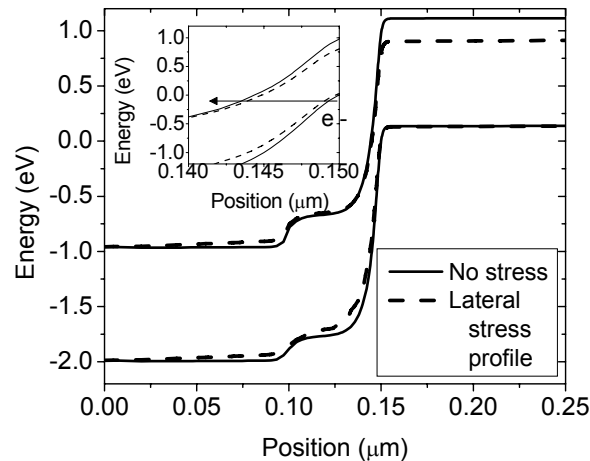


Figure 3.20: Band diagram profile along the channel for silicon Tunnel FET (“No stress” band gap = 1.1 eV) and strained channel (“Lateral profile” varying from 0.9 eV to 1.1 eV) in on-conditions: $V_{DS} = V_{GS} = 1$ V. Inset: barrier narrowing is 13% in the case with a lateral profile.

The optimization of silicon Tunnel FETs by using a reduced band gap only at the tunnel junction results in high I_{on} without the high I_{off} that is observed in all SiGe or GeOI Tunnel FETs that also have a small band gap at the drain-intrinsic region junction [6]. The improvement in Tunnel FETs with a lateral strain profile (maximum stress = 4 GPa) can be seen in Fig. 3.21, where we compare the $I_{DS}-V_{GS}$ characteristics with those of Tunnel FETs on unstrained silicon, and on uniformly strained silicon. The unstrained device has an I_{off} (at $V_{DS} = 1$ V and $V_{GS} = 0$ V) several decades lower than the uniformly strained device, due to the large band gap and therefore the wider tunneling barrier at the drain-side junction, but has comparatively low on-current. The uniformly strained device with 4 GPa of stress everywhere has an I_{on} (at $V_{DS} = V_{GS} = 1$ V) ten times higher than the unstrained device, thanks to the smaller band gap (0.9 eV) at the source-side junction, but has high

leakage. With the asymmetrically strain-engineered Tunnel FET, it is possible to take the best of both: high I_{on} , a decade higher than that of the unstrained device, and low I_{off} , about three decades lower than the device with strain everywhere.

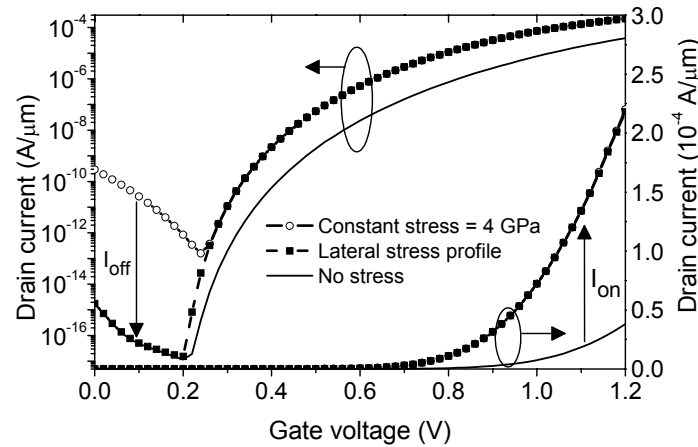


Figure 3.21: Drain current versus gate voltage for all-silicon Tunnel FETs with $L = 50$ nm and strain profile as a parameter. I_{off} is reduced for a lateral stress profile (stress peak 4 GPa), with respect to constant stress (4 GPa everywhere), and on-current is improved compared to the device without stress. $V_{DS} = 1$ V.

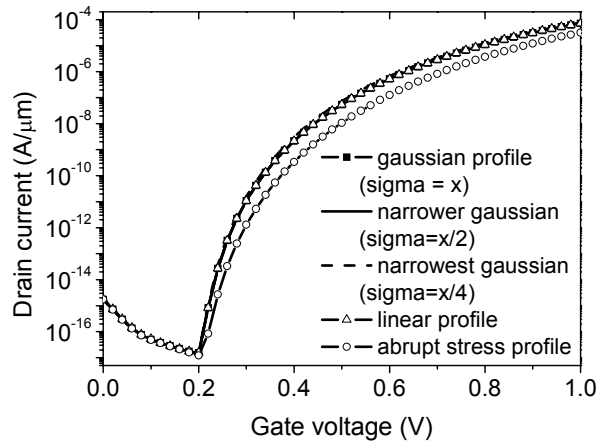


Figure 3.22: Influence of the shape of the lateral strain profile on all-silicon Tunnel FET characteristics, for the profiles shown in Fig. 3.19. The maximum band gap reduction was 0.2 eV for all profiles. $V_{DS} = 1$ V.

Fig. 3.22 reports simulations of I_{DS} - V_{GS} characteristics for the various lateral strain profiles shown in Fig. 3.19; we conclude that the exact profile shape is not very critical for the I_{on}/I_{off} improvement, as long as the band gap has its minimum at the source-intrinsic region junction, and its maximum at the drain-intrinsic region junction. The only curve which is not coincident with the others in Fig. 3.22 is for a perfectly abrupt stress profile, which is not physically realistic and only shown for reasons of comparison. In addition, the reason this curve is different than the others is due to the location of the abrupt profile, as can be seen in Fig. 3.19 -- right at the tunnel junction. If the abrupt change in band gap had instead been placed in the center of the intrinsic region, then the silicon region with the larger band gap would not have affected the tunneling barrier width, and the I_{DS} - V_{GS} curve would once again have been coincident with the other four.

Fig. 3.23 shows the transfer characteristics for Tunnel FETs with different amounts of maximum band gap shrinkage at the tunnel junction. For each incremental decrease of the band gap of 0.1 eV, the current increases by 3 to 4x. In addition, a remarkable improvement is found in the average subthreshold swing (S_{avg}), taken over more than seven decades from the turn-on point up to a current

of 10^{-7} A/ μm , presented in more detail in Fig. 3.24. S_{avg} ranges from 65 mV/dec. for the optimized Tunnel FET without strain, down to 36 mV/dec. for the device whose band gap shrinks to 0.8 eV at the tunnel junction.

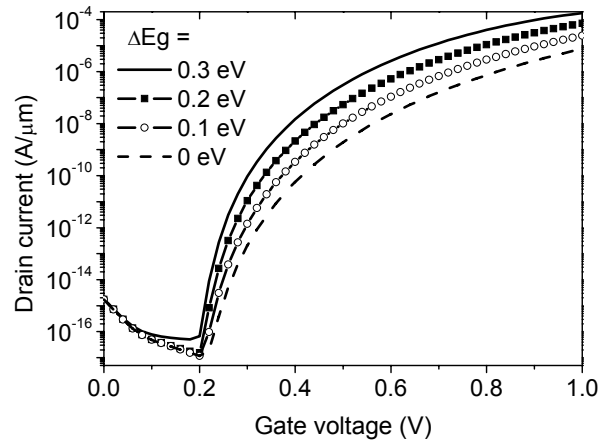


Figure 3.23: Simulation of $I_{\text{DS}}-V_{\text{GS}}$ characteristics' improvement with decreasing band gap at the source-intrinsic region junction. $L = 50$ nm, $V_{\text{DS}} = 1$ V.

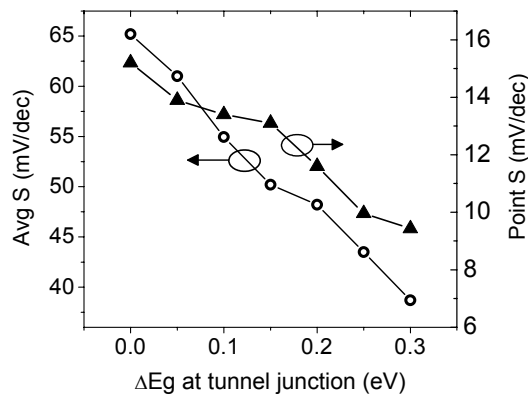


Figure 3.24: Average and point swings versus band gap reduction at the tunnel junction in the double-gate Tunnel FET with $L = 50$ nm. $V_{\text{DS}} = 1$ V.

The output characteristics presented in Fig. 3.25 show further important improvements in asymmetrically strained Tunnel FET behavior. First, I_{on} at $V_{\text{DS}} = V_{\text{GS}} = 1$ V is increased from $8 \mu\text{A}/\mu\text{m}$ in devices without strain to $72 \mu\text{A}/\mu\text{m}$ when a tensile strain of 4 GPa is applied on the source side (shrinking the band gap to 0.9 eV there). Second, Tunnel FETs, unlike conventional MOSFETs, have a drain threshold voltage as presented in [11] and discussed in detail in Chapter 4. The drain threshold voltage, V_{TD} , gives an unwanted high R_{on} resistance at low V_{DS} that is detrimental to the design and operation of Tunnel FETs as fast logic devices, seen in more detail in Chapter 4. If one decreases the band gap at the source tunnel, the drain threshold voltage decreases quasi-exponentially (Fig. 3.26), and both R_{on} (especially at low V_{DS}) and I_{on} are significantly improved. Fig. 3.26 also shows that the gate threshold voltage decreases almost linearly with band gap shrinkage at the source side (from 0.67 V for no strain, to 0.49 V for $\Delta E_{\text{g}} = 0.25$ eV), without affecting I_{off} , kept low by the non-strained drain junction.

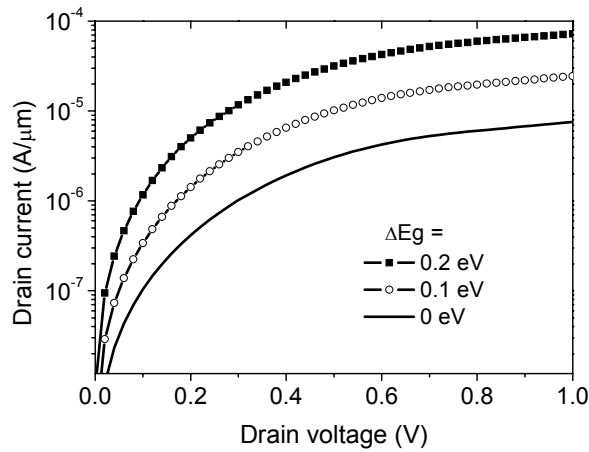


Figure 3.25: Drain current versus drain voltage for Tunnel FETs with different max. band gap shrinkage ΔE_g at the tunnel junction due to strain, for $V_{GS} = 1$ V. $\Delta E_g = 0.2$ eV corresponds to 4 GPa of stress.

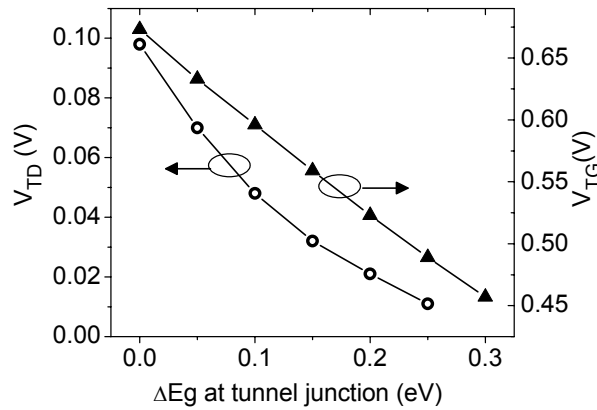


Figure 3.26: Drain and gate threshold voltages', V_{TD} and V_{TG} , dependence on ΔE_g at the tunnel junction; while a quasi-linear decrease is observed for V_{TG} , the V_{TD} decrease is much faster (quasi-exponentially). Both V_{TD} and V_{TG} are extracted at $I_{DS} = 10^{-7}$ A/ μm . 1 V was applied to the opposite terminal.

In Fig. 3.27 we present the relative improvement (%) of the drain current in all regions of operation, as induced by a band gap reduction of 0.2 eV at the source-intrinsic region junction. The Tunnel FET channel length is varied from 2 μm down to 50 nm. Typical I_{on} improvements at maximum applied drain and gate voltage (=1V) range from $\sim 400\%$ (in long channels) to $\sim 800\%$ (in short channels) for $\Delta E_g = 0.2$ eV. A significantly higher improvement is observed in channels where the intrinsic gate region is shorter than 100 nm; here the highest current increase ($> 1000\%$) is observed near the transistor turn-on point. In these short channel transistors, the reduced band gap has its strongest effect right where the current starts to increase and the subthreshold slope is steepest, and the current of the reference transistor is very low.

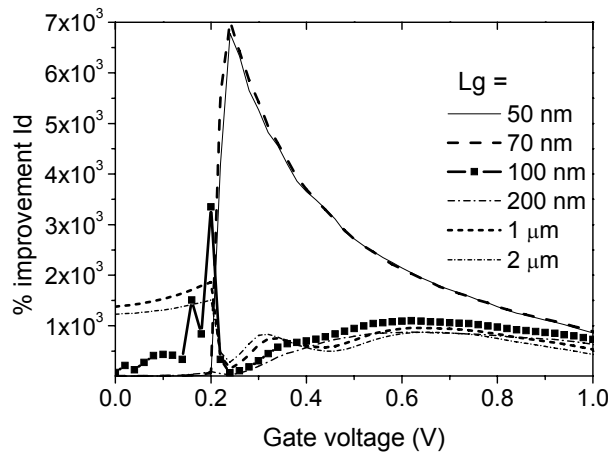


Figure 3.27: Percent improvement in drain current between a Tunnel FET without stress (reference), and devices with 4 GPa of stress ($\Delta E_g = 0.2$ eV) at the tunnel junction, for various intrinsic region lengths. $V_{DS} = 1$ V.

Feasibility of lateral strain profile formation in real devices

Several possibilities exist for the practical implementation of such a strain profile. Ref. [12] shows nanowires on strained-Si-on-insulator with stress levels up to between 2 and 3 GPa (see Fig. 3.28(a)). Strain relaxation would need to be implemented at the non-tunnel junction, and there are several possibilities for doing so. One example is the technique shown in [13] for relaxing uniaxial stress, using a compressive capping layer, pictured in Fig. 3.28(b). If the capping layer were only present over the drain junction, then the source junction would retain its uniaxial strain and therefore its low band gap. Another way of obtaining a lateral strain profile is by creating nanowires using a top-down process, where some step in the process produces strain within the wires. In [14], shown in Fig. 3.29(a), stress profiles from 0 (end of wire) to 2.6 GPa (center of wire) were engineered in silicon nanowires by oxidation. In [15], presented in 3.29(b), nanowires with stress of more than 4 GPa, resulting from metal gate deposition, were formed. Tunnel FETs could be fabricated on these wires with the tunnel junction on the wire where the stress level is high, and the other junction on the non-strained anchor where the band gap would be that of unstrained silicon. One further way to create a strain profile would be to employ a technique that produces uniaxial strain in a localized region. For example, a SiN capping layer could be used to create high tensile stress at the channel/source region, as shown in [16]. One of these techniques, or some combination, could lead to a solution for creating Tunnel FETs with the strain profile presented here.

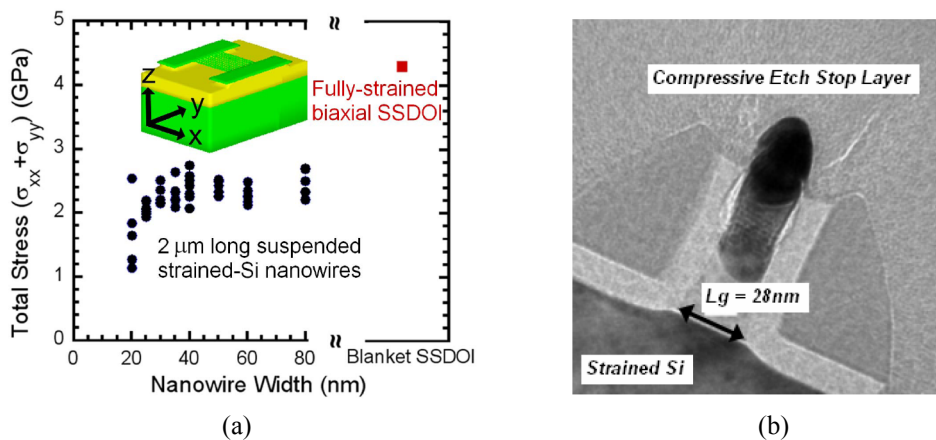


Figure 3.28: One possible approach to achieving a lateral strain profile in Tunnel FETs could use strained silicon nanowires as presented in part (a), from [12], and then relax the strain around the drain-intrinsic region junction using a technique such as that shown in (b), from [13], so that off-current stays low.

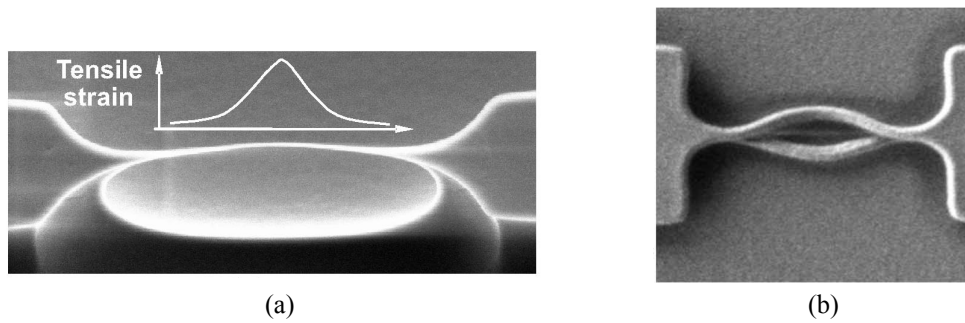


Figure 3.29: Another approach for creating Tunnel FETs with a lateral strain profile: build them on strained nanowires. (a) Silicon nanowires with a strain peak of 2.6 GPa at the center of the wires, from [14]. (b) Silicon nanowires with strain estimated to be more than 4 GPa, from [15]. In both cases, the source-intrinsic region junction could be placed in a high-strain region, near the center of the wires, while the drain-intrinsic region junction could be placed at the end of the wires or on the anchor where the band gap is the largest.

3.4 Conclusion

Before moving forward with other investigations in Tunnel FET behavior, it was necessary to start with an optimized device. Optimization was carried out on the following parameters: single or double gate, doping levels, gate oxide permittivity, silicon body thickness, and band gap at the tunnel junction. 2D simulations looked at device behavior in the on-state, since a solid understanding of what is happening in the device body is critical for optimization and further exploration of Tunnel FETs.

The following technical topics and contributions were presented in this chapter:

- *Tunnel FET parameter optimization* (Section 3.1)
This section shows the optimization of the fundamental Tunnel FET parameters: single or double gate, doping levels in the source, drain, and intrinsic regions, gate dielectric permittivity, and silicon body thickness. A double gate multiplies the on-current by two if the body thickness is larger than a certain value, and increases it by a larger factor for a very thin body. Increasing source doping strongly increases on-current, while decreasing drain doping can suppress ambipolar behavior. Intrinsic region doping has little effect until the doping level gets above 10^{18} atoms/cm³ (and is no longer intrinsic). Increasing gate dielectric permittivity greatly improves on-current and subthreshold swing. Finally, a maximum of on-current exists for some silicon body thickness -- 7 nm for the specific device design investigated here. Two other important device design parameters, gate oxide alignment and abruptness of the doping profile at the tunnel junction, were not investigated here but will be discussed in detail in Chapter 6.
- *2D Tunnel FET simulations* (Section 3.2)
Two-dimensional simulated cross-sections of optimized Tunnel FETs in the on-state show that the electric field is high at the tunnel junction down into the center of the body of a 10-nm thick double-gate device. Similarly, the potential changes abruptly at this junction all the way down into the body. Current flow lines show that the current flow is not confined to the surface, even in overdrive, but instead flows throughout the body thickness, then comes closer to the surface right at the band-to-band tunnel junction, before spreading out again. Energy band cross-sections in the y-direction (from the dielectric surface down to the center of the body) help explain this current flow pattern.
- *One final optimization: the band gap* (Section 3.3)
In order to improve Tunnel FET on-current, one more important optimization can be carried out. When the band gap is decreased at the tunnel junction (between the source and the intrinsic region), either by straining the silicon at that location or by using a material with a smaller band gap at the source side, on-current is increased. If the band gap remains large at the drain side of

the device, off-current and subthreshold swing remain low. Several possible techniques were suggested for the fabrication of Tunnel FETs with a lateral strain profile.

All results presented in this chapter, represent original contributions, except Fig. 3.12 from [3], and the material presented in the feasibility section at the end of section 3.3.

3.5 Bibliography

- [1] F. Balestra, S. Cristoloveanu, M. Benachir, J. Brini, and T. Elewa, "Double-gate silicon-on-insulator transistor with volume inversion: a new device with greatly enhanced performance," *IEEE Elec. Dev. Lett.*, vol. 8, no. 9, 1987, pp. 410-412.
- [2] J. McPherson, J-Y. Kim, A. Shanware, and H. Mogul, "Thermochemical description of dielectric breakdown in high dielectric constant materials," *Appl. Phys. Lett.*, vol. 82, no. 13, 2003, pp. 2121-2123.
- [3] E.-H. Toh, G. Wang, G. Samudra, and Y.-C. Yeo, "Device physics and design of double-gate tunneling field-effect transistor by silicon film thickness optimization," *Appl. Phys. Lett.*, vol. 90, 2007, pp. 263507-1-3.
- [4] K. Kim and J. Fossum "Double-Gate CMOS: Symmetrical- Versus Asymmetrical-Gate Devices," *IEEE Trans. Elec. Dev.*, vol. 48, no. 2, 2001, pp. 294-299.
- [5] T. Krishnamohan, D. Kim, S. Raghunathan, and K. Saraswat, "Double-Gate Strained-Ge Heterostructure Tunneling FET (TFET) With Record High Drive Currents and <60mV/dec Subthreshold Slope," *IEDM Tech. Dig.*, 2008, pp. 947-949.
- [6] F. Mayer, C. Le Royer, J.-F. Damlencourt, K. Romanjek, F. Andrieu, C. Tabone, B. Previtali, and S. Deleonibus, "Impact of SOI, Si_{1-x}Ge_xOI and GeOI substrates on CMOS compatible Tunnel FET performance," *IEDM Tech. Dig.*, 2008, pp. 163-166.
- [7] S. Koswatta, M. Lundstrom, and D. Nikonov, "Performance Comparison Between p-i-n Tunneling Transistors and Conventional MOSFETs," *IEEE Trans. Electron Devices*, vol. 56, no. 3, 2009, pp. 456-465.
- [8] J.Knoch, "Optimizing tunnel FET performance – Impact of device structure, transistor dimensions and choice of material," *Proc. VLSI-TSA*, 2009, pp. 45-46.
- [9] E.-H. Toh, G.H. Wang, L. Chan, G.-Q. Lo, G. Samudra, Y.-C. Yeo, "Strain and Materials Engineering for the I-MOS Transistor With an Elevated Impact-Ionization Region," *IEEE Trans. Electron Devices*, vol. 54, no. 10, pp. 2778-2785, Oct. 2007.
- [10] T. Krishnamohan, C. Jungemann, D. Kim, E. Ungersboeck, S. Selberherr, A.-T. Pham, B. Meinerzhagen, P. Wong, Y. Nishi, K. C. Saraswat, "High performance, uniaxially-strained, silicon and germanium, double-gate p-MOSFETs," *Microelectronic Engineering*, vol. 84, 2007, pp. 2063-2066.
- [11] K. Boucart, A. M. Ionescu, "A new definition of threshold voltage in Tunnel FETs," *Solid-State Elec.*, vol. 52, no. 9, Sept. 2008, pp. 1318-1323.
- [12] P. Hashemi, L. Gomez, M. Canonico, J.L. Hoyt, "Electron Transport in Gate-All-Around Uniaxial Tensile Strained-Si Nanowire n-MOSFETs," *IEDM Tech. Dig.*, 2008, pp. 865-868.
- [13] A.V.-Y. Thean, L. Prabhu, V. Vartanian, M. Ramon, B.-Y. Nguyen, T. White, H. Collard, Q.-H. Xie, S. Murphy, J. Cheek, S. Venkatesan, J. Mogab, C.H. Chang, Y.H. Chiu, H.C. Tuan, Y.C. See, M.S. Liang, Y.C. Sun, "Uniaxial-Biaxial Stress Hybridization For Super-Critical Strained-Si Directly On Insulator (SC-SSOI) PMOS With Different Channel Orientations," *IEDM Tech. Dig.*, 2005, pp. 509-512.
- [14] K. Moselund, P. Dobrosz, S. Olsen, V. Pott, L. De Michielis, D. Tsamados, D. Bouvet, A. O'Neill, A.M. Ionescu, "Bended Gate-All-Around Nanowire MOSFET: a device with enhanced carrier mobility due to oxidation-induced tensile stress," *IEDM Tech. Dig.*, 2007, pp. 191-194.
- [15] N. Singh, W.W. Fang, S.C. Rustage, K.D. Budharaju, S.H.G. Teo, S. Mohanraj, G.Q. Lo, N. Balasubramanian, D.-L. Kwong, "Observation of Metal-Layer Stress on Si Nanowires in Gate-All-Around High-k/Metal-Gate Device Structures," *IEEE Elec. Dev. Lett.*, vol. 28, no. 7, July 2007, pp. 558-561.
- [16] S. Pidin, T. Mori, R. Nakamura, T. Saiki, R. Tanabe, S. Satoh, M. Kase, K. Hashimoto, T. Sugii, "MOSFET current drive optimization using silicon nitride capping layer for 65-nm technology node," *Symp. on VLSI Technology*, 2004, pp. 54-55.

Chapter 4

Threshold Voltages for Tunnel FETs

This chapter explains the need for a new, physically-based definition of threshold voltage for Tunnel FETs. A new definition is proposed, based on the transition between a quasi-exponential dependence, and a linear dependence of the drain current on voltage. It is shown that the Tunnel FET has two threshold voltages: one in terms of gate voltage, V_{TG} and one in terms of drain voltage, V_{TD} . The extractions of V_{TG} and V_{TD} are performed based on the transconductance change method, and a systematic comparison with the constant current method is reported. The effect of gate length scaling on these Tunnel FET threshold voltages, as well as the dependence of V_{TG} on applied drain voltage and V_{TD} on applied gate voltage, are investigated. The negative impact of the drain threshold voltage on circuit design is briefly discussed, mentioning both static and dynamic operation. Finally, the threshold voltage extraction technique is carried out on a collaborator's measurement data, confirming the validity of the proposed method.

4.1 Why a new definition for Tunnel FET threshold voltage?

Although the physics determining the current flow inside a Tunnel FET is not the same as in a conventional MOSFET, the two devices' switching behavior is similar enough that it is tempting to analyze a Tunnel FET as if it were a MOSFET. Most typical CMOS benchmarking parameters are equally applicable to a Tunnel FET, so there is no problem when looking at off-current, on-current, $I_{\text{on}}/I_{\text{off}}$ ratio, transconductance, gate leakage, etc. There is one parameter, however, that is more connected to the nanoscale physics processes going on inside the device, and consequently deserves our particular attention: the threshold voltage.

The threshold voltage is one of the most important electrical parameters of a solid-state switch. For the conventional MOS transistor, there is a physical definition of the threshold voltage; it is the gate voltage marking the transition between weak inversion and strong inversion, when the density of carriers in the inversion channel at the surface equals the doping level in the substrate, $\phi_s = 2\phi_F$. Since there is no such identified mechanism in Tunnel FETs, until now the V_T has been extracted at a constant current, arbitrarily chosen at values suggested by CMOS experience.

In this chapter, a new way to extract a threshold voltage for a Tunnel FET is proposed, and the new V_T is a physically meaningful quantity. This new technique uses the second derivative of the I-V characteristics to find the transition point on the I-V curve where the current's dependence on the voltage changes from quasi-exponential (often just called "exponential" in this chapter for readability), to linear. This derivative method of threshold voltage extraction thus locates the voltage where there is a transition between strong and weak control of the tunneling energy barrier width. Beyond its physical meaning, this extraction method has the advantage that it does not depend upon the accuracy of a particular analytical model, but rather can be extracted directly from experimental data. The details of the technique will be explained in section 4.3.1.

4.2 Simulation: device structure

The double-gate n-type Tunnel FETs investigated here have been simulated with Silvaco Atlas, version 5.11.24.C. In all simulations, junctions were quasi-perfectly abrupt (junction width 0.5 nm), and the p-type source, intrinsic region, and n-type drain were doped at 1×10^{20} , 1×10^{17} , and 5×10^{18} atoms/cm³ respectively. The silicon body thickness is 10 nm.

The gate dielectric covers the drain, intrinsic region, and source in all simulations. Three different gate stacks were studied. The first uses 3 nm of SiO₂ as a gate dielectric ($\epsilon = 3.9$). The second employs 3 nm of a high-k dielectric ($\epsilon = 25$), which could be HfO₂ or ZrO₂. The third incorporates a gate dielectric with two components: a 1 nm interfacial layer of oxynitride ($\epsilon = 5.7$) and 2 nm of a high-k dielectric ($\epsilon = 25$), corresponding to a more realistic fabrication process. All Tunnel FETs simulated here use a midgap (metal) gate work function of 4.5 eV.

4.3 Threshold voltages of Tunnel FETs

4.3.1 Gate threshold voltage, V_{TG}

The easiest way to extract a gate threshold voltage for the Tunnel FET is to use the constant current method, and several researchers, including myself, have done so [1]-[5]. This technique has several advantages including simplicity of extraction and independence from the accuracy of a drain current model. Unfortunately, this method uses an arbitrary current value and has almost no physical meaning, despite its practical interest. The threshold current value used elsewhere in this thesis is $I_{DS} = 10^{-7}$ A/ μm , and the condition to which this corresponds inside a typical optimized Tunnel FET can be seen by studying the dependence of the drain current on w_b , the tunneling energy barrier width (Fig. 4.1). Here, the energy barrier width has been extracted by taking the energy bands across the Tunnel FET body and then measuring the narrowest point across the energy barrier on those

bands. This technique works well at applied gate voltages above several hundred mV due to the exponential dependence of tunneling probability on barrier width.

Fig. 4.1 shows that when the constant current method is applied to an optimized double-gate Tunnel FET, the threshold voltage corresponds to a w_b of about 7.3 nm, which is still in the region of a strong dependence of the barrier on the gate voltage, as shown by the dashed line in Fig. 4.2. This technique for extracting a threshold voltage is not based on a physical transition inside the device -- it is arbitrary.

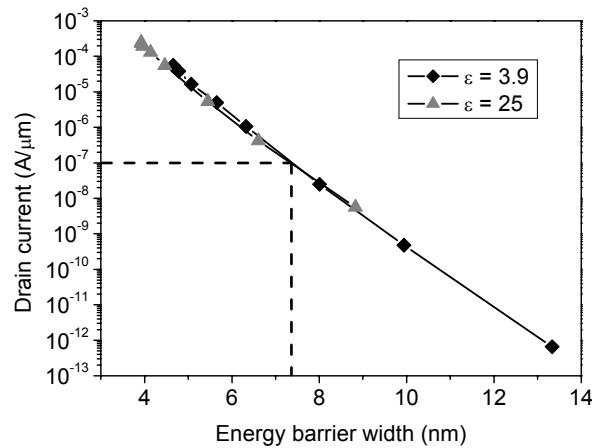


Figure 4.1: Drain current vs. energy barrier width w_b for different values of the gate dielectric permittivity. $V_{DS} = 1$ V.

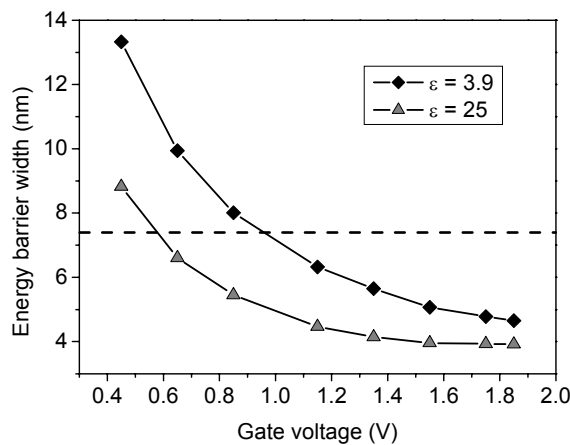


Figure 4.2: Energy barrier width for tunneling, w_b , vs. applied gate voltage of a double-gate Tunnel FET with two different gate dielectrics: SiO_2 and high-k. $V_{DS} = 1$ V.

The Tunnel FET's threshold voltage does not need to be arbitrary. It is possible to create a new, more physical definition of the gate threshold voltage V_{TG} for this device, related to the main mechanism that controls the current level -- the energy barrier width. Fig. 4.3 shows I_{DS} - V_{GS} curves for Tunnel FETs with three different gate dielectrics: SiO_2 , and high-k with and without an interfacial layer. The curves are shown on a logarithmic scale on the left axis, and on a linear scale on the right axis. Using both the logarithmic and linear scales, the shape of the curve can be understood. Up until about 1 V on the gate (at a slightly higher voltage for the device with SiO_2), the current has an exponential dependence on the gate voltage. Then the dependence becomes linear. In theory, the current should then saturate at high V_{GS} , since the energy barrier cannot decrease in width indefinitely, but unfortunately, that is beyond the abilities of Silvaco Atlas simulations to converge. It can be seen, however in several publications including by simulation in [6] and from experimental measurement in [7].

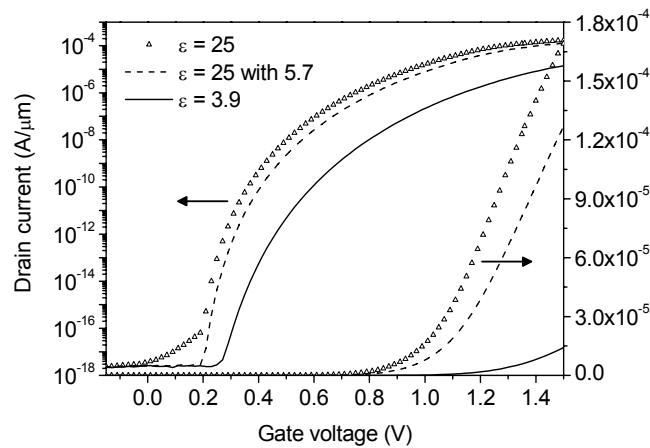


Figure 4.3: I_{DS} - V_{GS} in logarithmic (left axis) and linear (right axis) scales, $L = 100$ nm, for Tunnel FETs with three different gate dielectric stacks: 3 nm of high-k ($\epsilon_{\text{dielectric}} = 25$), 2 nm of high-k ($\epsilon_{\text{dielectric}} = 25$) over 1 nm of interfacial silicon oxynitride ($\epsilon_{\text{dielectric}} = 5.7$), and 3 nm of SiO_2 ($\epsilon_{\text{dielectric}} = 3.9$). $V_{DS} = 1$ V.

The new definition of threshold voltage is proposed as follows:

The Tunnel FET threshold voltage is the voltage marking the transition between an exponential dependence, and a linear dependence, of drain current on applied bias. This also marks the transition between the strong control and weak control of the tunneling energy barrier width at the tunnel junction by that voltage.

As will be shown in the next section, this will not only apply to V_{TG} , in relation to the bias applied to the gate, but also to V_{TD} , in relation to the bias applied to the drain. For this reason, the definition has been written in general terms and not tied specifically to gate voltage.

This new V_{TG} can be extracted using the transconductance change (TC) method (already validated for the MOS transistor [8]), which universally defines the threshold voltage of any non-linear device as the gate voltage corresponding to the maximum of the transconductance derivative, dg_m/dV_{GS} . As an example, starting with the curve corresponding to $\epsilon_{\text{dielectric}} = 25$ in Fig. 4.3, the drain current's 1st (g_m), 2nd (dg_m/dV_{GS}), and 3rd order derivatives (d^2g_m/dV_{GS}^2) are plotted in Fig. 4.4. An inflection point in the transconductance plot and a maximum of the transconductance derivative appear, which enables the extraction of V_{TG} as the gate voltage corresponding to $\max(dg_m/dV_{GS})$, after curve de-noising. This inflection point shows the transition point between a quasi-exponential dependence and a linear dependence, of current on V_{GS} . It is worth noting that this voltage point also corresponds to the zero of the 3rd order derivative; however, as shown in Fig. 4.4, this plot has much numerical derivative noise, which makes its practical use impossible.

The accuracy of the extraction by the derivative method will depend on the step size of the I_{DS} - V_{GS} measurement, in contrast with linear extrapolation methods where the use of many points could have a high overall accuracy. Since the exponential and linear regions can be clearly seen on the I-V curves graphed on a linear scale, as on the right axis of Fig. 4.3, it could be tempting to use an extrapolation method to extract V_T , since step size and noise would not be problematic. The disadvantage of such an extrapolation is that it wouldn't actually give the exact voltage where the current dependence changes from quasi-exponential to linear. Looking again at Fig. 4.3, if we extend the linear portion of one of the I-V curves back to the V_{GS} axis, the V_T extracted in that way would actually be at a voltage where the dependence of the current is still exponential, not at the transition point. If we want a truly physical definition of threshold voltage, we must use the maximum of the second derivative instead. In order to use the second derivative of current, measurements with small steps are needed. In this thesis, simulation step size was 20 mV, but for experimental data, smaller increments of about 5 mV would be beneficial, as would long integration times. Denoising methods are useful, and for this thesis, an extremely simple method was used,

where $dg_{i,\text{smoothed}} = (dg_{i-1} + 2dg_i + dg_{i+1})/4$. Experimental data, which is naturally more noisy, would probably require more advanced smoothing techniques.

Returning to Fig. 4.4, the extracted V_{TG} is 1.1 V, which is much larger than the threshold voltage defined by the constant current method ($V_{TG} = 0.6$ V at $I_{DS} = 10^{-7}$ A/ μm from Fig. 4.3), but this new threshold voltage has a physical significance since it marks the gate's transition from a quasi-exponential to a linear control of the drain current or from a strong to a weak control of the energy barrier width at the tunnel junction. As already mentioned, this definition of V_{TG} is analogous to a MOSFET threshold voltage for which a transition from weak to strong inversion is experienced at threshold (at $2\phi_F$). Since the energy barrier width depends not only on the gate voltage but also on the drain voltage, this technique also has the advantage that it can be used to find the Tunnel FET's second threshold voltage, V_{TD} .

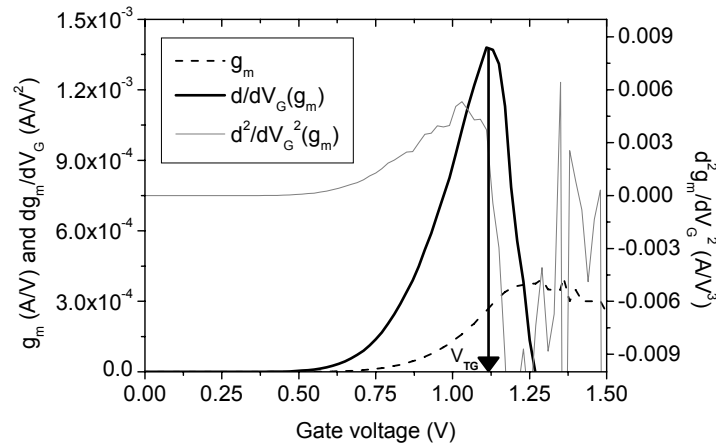


Figure 4.4: V_{TG} extraction: the three derivatives, g_m , dg_m/dV_{GS} , d^2g_m/dV_{GS}^2 for $L = 100$ nm and $\epsilon_{\text{dielectric}} = 25$. $V_{DS} = 1$ V.

4.3.2 Drain threshold voltage, V_{TD}

A unique characteristic of the Tunnel FET is that tunneling requires a certain minimum amount of drain voltage to turn the device on, whatever the applied V_{GS} . This is true because the energy barrier narrowing for tunneling on-current is a complex function of both V_{GS} and V_{DS} , experiencing a critical control by both of them. It follows that one can define a second threshold voltage, called here V_{TD} , in terms of the drain voltage. This feature is not shared by conventional MOS transistors.

Moreover, the tunneling current dependence on V_{DS} is physically similar to the one on V_{GS} (control of w_b by V_{DS}), and the same regions of dependence of current on voltage are seen -- first quasi-exponential at low V_{DS} , and then linear. One can therefore define the drain threshold voltage, V_{TD} , in the same way that the gate threshold voltage was defined, as the transition point at which the drain current dependence on the drain voltage changes from quasi-exponential to linear. Figs. 4.5(a) and (b) depict typical plots with the output current characteristics of a Tunnel FET and their 1st, 2nd and 3rd derivatives used for the extraction of the drain threshold voltage. Again, a maximum of dg_{ds}/dV_{DS} is observed, which enables the extraction of $V_{TD} = 0.5$ V for the device shown here. This value is roughly half of the V_{TG} extracted with the same technique.

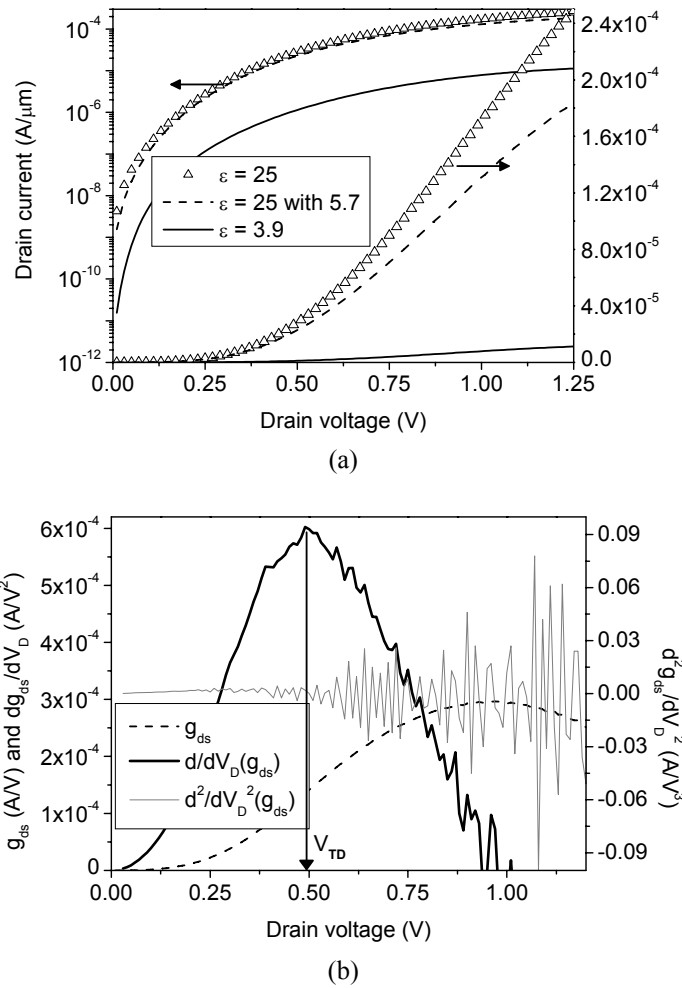


Figure 4.5: (a) I_{DS} - V_{DS} in log (left axis) and linear (right axis) scales for $L = 50$ nm. (b) V_{TD} extraction: g_{ds} , dg_{ds}/dV_{DS} , d^2g_{ds}/dV_{DS}^2 for $L = 50$ nm and $\epsilon_{\text{dielectric}} = 25$. $V_{GS} = 1.5$ V.

4.4 Gate length scaling effects on Tunnel FET threshold voltages

In this section, length scaling is carried out independently from the scaling of other parameters, in order to isolate the effects, as explained in much more detail in Chapter 5. Here, the focus is on the effects of gate length scaling on Tunnel FET threshold voltages and characteristics.

Scaling the gate length of the Tunnel FET is expected to have a limited effect on the physical threshold gate voltage V_{TG} since there is little effect on the transfer characteristics as long as the device is longer than some critical length at which too much p-i-n leakage current occurs in the off-state. Typical I_{DS} - V_{GS} and g_m - V_{GS} characteristics have been carefully simulated for Tunnel FETs with a high-k gate dielectric and channel lengths from $2 \mu m$ down to 20 nm (see Fig. 4.6). In Fig. 4.7 the derivatives of Tunnel FET transconductance for $V_{DS} = 1$ V are reported and the maxima of the plots, all within 20 mV of each other, are used for V_{TG} extraction. The extracted values are shown in Fig. 4.8.

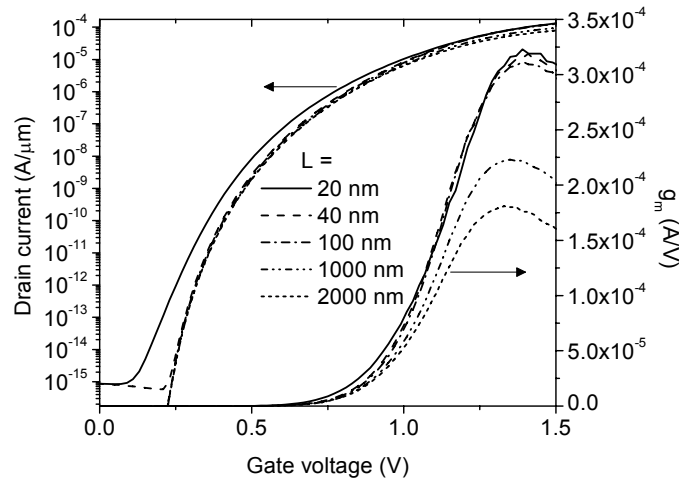


Figure 4.6: I_{DS} (left axis) and g_m (right axis) for Tunnel FETs with a gate dielectric stack of 1 nm silicon oxynitride interfacial layer ($\epsilon = 5.7$) and 2 nm high-k ($\epsilon = 25$). $V_{DS} = 1$ V.

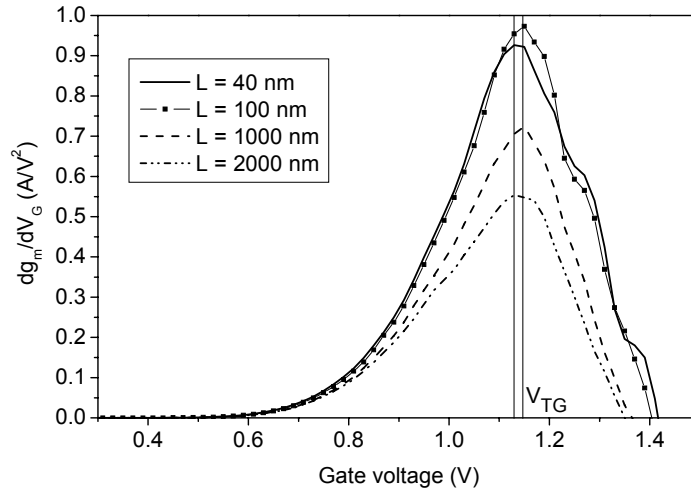


Figure 4.7: dg_m/dV_{GS} vs. V_{GS} for Tunnel FETs with a gate dielectric stack of 1 nm silicon oxynitride interfacial layer ($\epsilon = 5.7$) and 2 nm high-k ($\epsilon = 25$). $V_{DS} = 1$ V.

Looking at the length dependence of the threshold voltage V_{TG} in Fig. 4.8, reported for both the derivative method and the constant current method (as reference), a small V_{TG} roll-off is observed only for the constant current method, for channel lengths smaller than 40 nm. This V_T roll-off has a completely different cause than in a conventional MOSFET, coming from increased diode leakage (band-to-band tunneling that takes place at low V_{GS} and cannot be controlled by the gate) at smaller device lengths. As expected, the more physical transconductance change method for extracting V_{TG} in Fig. 4.8 is independent of this diode leakage, so no roll-off is observed. The physical gate threshold voltage is systematically higher (roughly the double) of the value extracted at 10^{-7} A constant current, for this particular case in which V_{TG} was extracted with $V_{DS} = 1$ V. This will be explained in more detail in the next section.

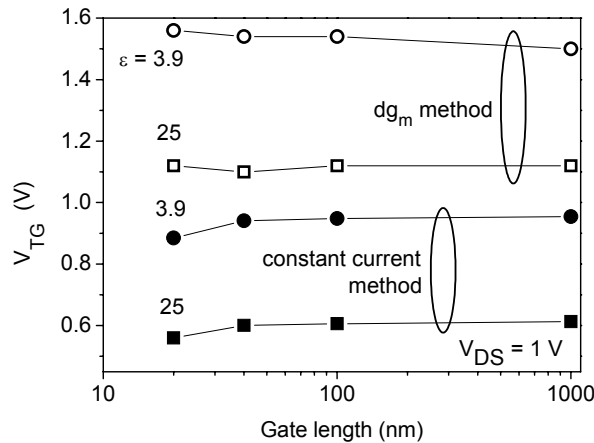


Figure 4.8: V_{TG} extracted by two methods: constant current and the dg_m/dV_{GS} approach, for $V_{DS} = 1$ V.

4.5 Discussion of V_{DS} and V_{GS} dependence of threshold voltages

The values of V_{TG} extracted with the dg_m/dV_{GS} technique reflect the combined control of gate and drain voltages on the tunneling barrier width, and depend on the applied drain-source voltage in an interesting way. With a higher drain voltage applied, the gate retains quasi-exponential control of the current over a larger voltage range. Fig. 4.9 shows that as a result, V_{TG} is higher with higher V_{DS} . The constant current method, on the other hand, is not sensitive to the change from a quasi-exponential to a linear dependence of current on V_{GS} , so threshold voltage values obtained in this way remain nearly constant for $V_{DS} > 1$ V, as will be seen near the end of this section.

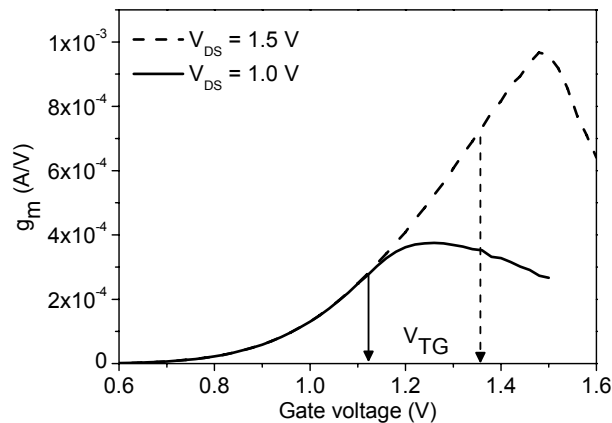


Figure 4.9: g_m vs. V_{GS} shows the dependence of the extracted V_{TG} value on the applied drain-source voltage. Here, the points of steepest slope are marked, corresponding to the maxima of dg_m/dV_{GS} , which are the extraction points for V_{TG} . $L = 100$ nm, gate dielectric permittivity = 25.

The same analysis can be carried out with the drain threshold voltage V_{TD} . Fig. 4.10(a) shows V_{TD} vs. gate length for the two extraction techniques discussed, at $V_{GS} = 1$ V (left) and $V_{GS} = 1.5$ V (right). We see that there is not always a straightforward relationship between the threshold voltages extracted with the two techniques and that the dg_{ds}/dV_{DS} technique can sometimes give lower values, and sometimes give higher values, than those given by the constant current method. Fig. 4.10(b) and (c) explain the different threshold voltage levels visually. In Fig. 4.10(b) the constant current method is used for Tunnel FETs with SiO_2 and high-k dielectrics, at $V_{GS} = 1$ V and 1.5 V. Since the SiO_2 Tunnel FET biased with $V_{GS} = 1$ V is approaching I_{DS} - V_{DS} saturation at a current only a bit higher than the constant current cutoff of 10^{-7} A/ μm , its V_{TD} value extracted with this technique is much higher than for the other three curves shown here. The drain threshold voltages

extracted with the constant current technique decrease as higher gate voltages are applied, since less drain voltage is needed in order to hit the threshold current level. In contrast, the dg_{ds}/dV_{DS} method shows the opposite trend, analogous to that observed when extracting the gate threshold voltage with the dg_m/dV_{GS} method as was shown in Fig. 4.9. As seen in Fig. 4.10(c), with higher gate voltage applied, V_{TD} is higher because the drain retains its quasi-exponential control of the current over a larger range before the dependence becomes linear. One more remark is needed about the left figure of Fig. 4.10(a), since the V_{TD} values extracted with the TC method are almost the same, no matter what the gate dielectric. Keep in mind that the overall current values were always much higher for the device with a high-k dielectric, but at this particular gate bias, the shape of the curves was similar, and the transition from exponential to linear dependence happened around the same V_{DS} .

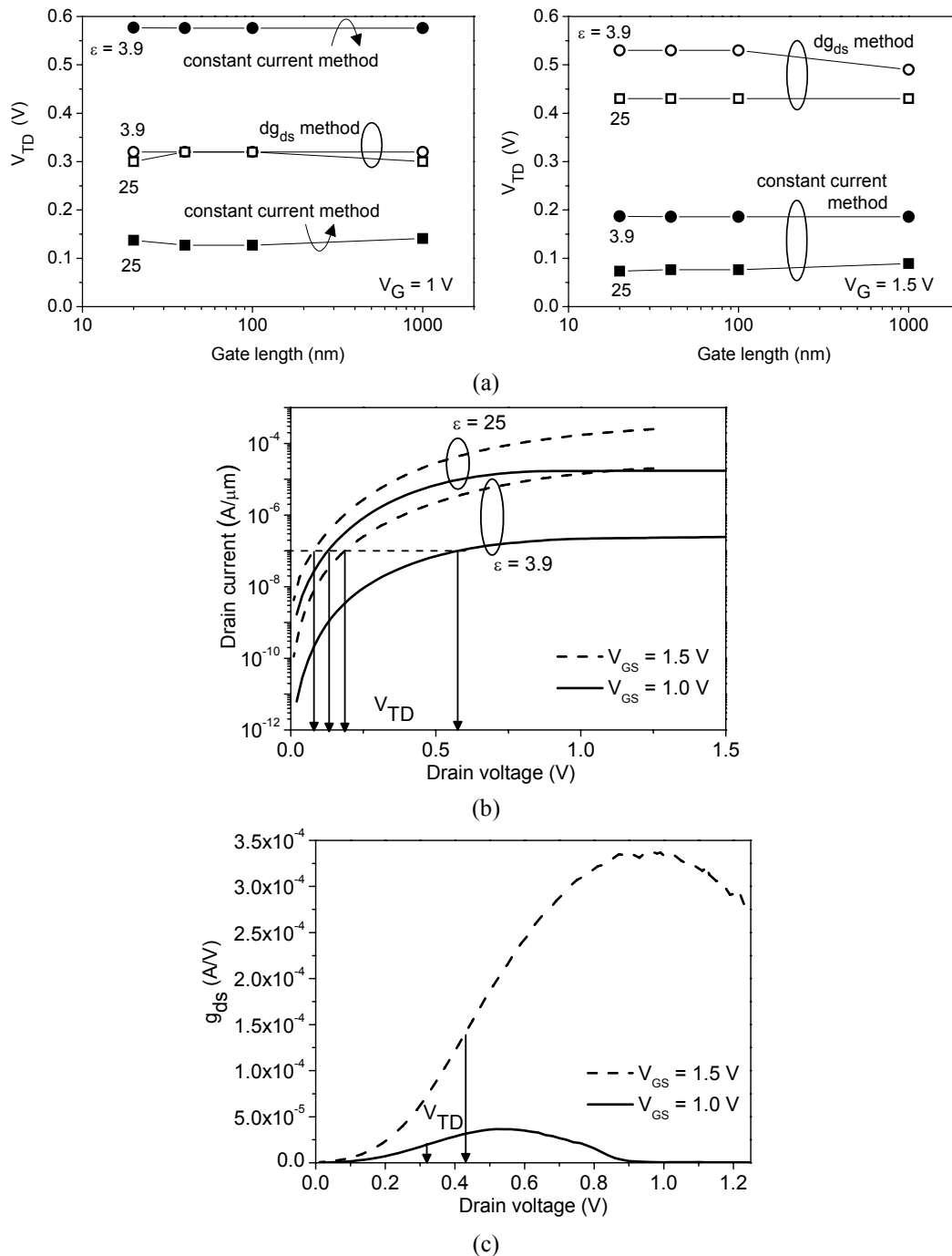
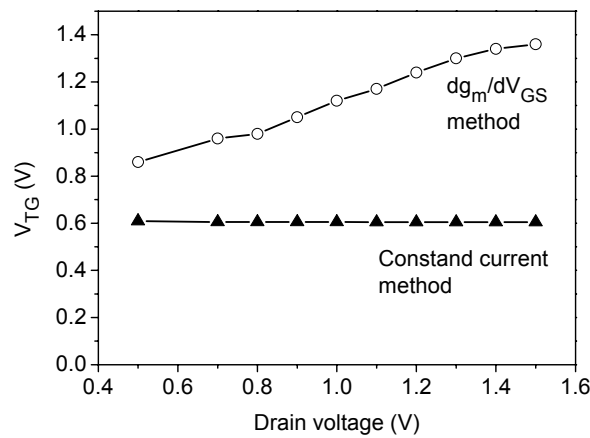


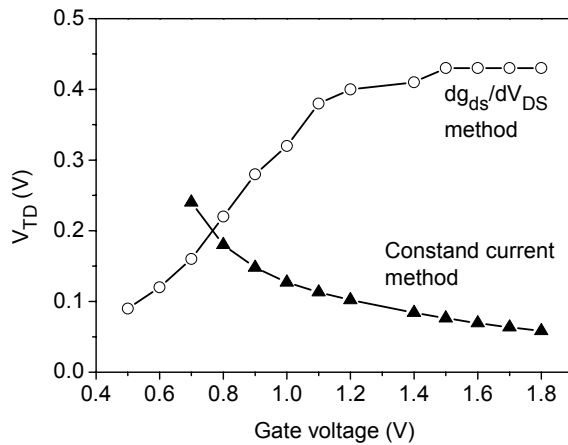
Figure 4.10: Illustration of the difference between the two methods of V_{TD} extraction. (a) V_{TD} for $V_{GS} = 1$ V (left) and 1.5 V (right). (b) The constant current method of V_{TD} extraction. $L = 40$ nm. (c) g_{ds} - V_{DS} for a Tunnel FET with $L = 40$ nm and dielectric permittivity = 25, showing the drain threshold voltages obtained with the dg_{ds}/dV_{DS} technique.

This trend, in which one voltage terminal (drain or gate) has a larger voltage range of quasi-exponential control over the drain current when higher voltage is applied on the other terminal (gate or drain), is further illustrated in Fig. 4.11. While the threshold voltages given by the constant current method stay constant (V_{TG} vs. V_{DS} in Fig. 4.11(a)) or decrease (V_{TD} vs. V_{GS} in Fig. 4.11(b)) with increasing voltage applied to the other terminal, the dg/dV method shows an increasing range of quasi-exponential control by one terminal as the voltage is increased on the other terminal.

Fig. 4.11(a) shows that the gate threshold voltage increases approximately linearly with drain voltage, but the data points suggest that it may saturate at high V_{DS} (currently not completely proven due to simulation convergence limitations). From this, we can conclude that in order for the gate to have the largest possible voltage range of quasi-exponential control over the drain current, and therefore strong control of the tunneling energy barrier width, the applied drain voltage should be high. In Fig. 4.11(b), the drain threshold voltage increases linearly with gate voltage and then saturates around $V_{GS} = 1.5$ V.



(a)



(b)

Figure 4.11: Threshold voltages, acquired by both the dg/dV and the constant current methods, for a Tunnel FET with gate dielectric permittivity = 25 and $L = 100$ nm. (a) V_{TG} vs. V_{DS} . (b) V_{TD} vs. V_{GS} .

The dependence of the gate threshold voltage on drain voltage, and the dependence of the drain threshold voltage on gate voltage bring up two questions. First, why should the range of exponential control of drain current depend on the bias applied to the other terminal? And second, why does the drain threshold voltage (and also possibly the gate threshold voltage at higher V_{DS} than that shown here) seem to saturate, as seen in Fig. 4.11?

Some understanding of the answers to these questions can be gained by remembering that the exponential region of the I-V curves is tied to the band-to-band tunneling transmission, which has a

quasi-exponential dependence on voltage. When we see an exponential dependence of current on voltage, this indicates that the voltage is directly affecting the energy barrier width at the tunnel junction, and band-to-band tunneling is increasing accordingly. When we see a linear dependence in the I-V curve, this indicates that the voltage is no longer strongly controlling the energy barrier width at the tunnel junction.

So why does the range of exponential control of the drain current depend on the “other” bias? One way of thinking about it is that one voltage enables the other to have more (exponential) control. One voltage alone, whether on the drain or the gate, does not create significant current flow through the device. And in the opposite situation, the higher the V_{DS} and the V_{GS} applied to the Tunnel FET, the higher the on-current through the device (neglecting the saturation region for now). The exponential region of the I-V curve has the fastest rate of increase of the current, and the linear part of the curve has a slower increase, and is actually the first step toward current saturation. So if one applied voltage is increased, then the other applied voltage is able to increase the current exponentially to a higher level before the linear dependence begins. In fact, the second applied voltage is able to control the tunneling energy barrier width over a longer voltage range. For I_{DS} - V_{GS} , this longer range of exponential control is desirable, since it leads to better characteristics, including a higher I_{on} . For I_{DS} - V_{DS} , this region of exponential control is, unfortunately, not desirable, since we want a MOSFET-like switch to have a linear dependence of current on drain voltage before going into saturation.

As for the second question, why the threshold voltage for one terminal shows saturation with the bias applied to the other terminal, the answer is tied to the saturation of control of the tunneling energy barrier width. Looking at Fig. 4.11(b) where the saturation is clear, we see that for high V_{GS} , the drain voltage has exponential control of the current up until $V_{DS} = 0.43$ V, and not beyond. The drain current has only a linear dependence on drain voltage after this point. The linear dependence lasts for several hundred mV, and then the current saturates, as shown by the curve for a Tunnel FET with a high-k dielectric and $V_{GS} = 1$ V in Fig. 4.10(c), and experimentally by [5],[7],[9]. With this particular Tunnel FET design, the on-current has at that point hit its maximum, since the tunneling energy barrier width does not decrease further.

The interdependence of gate and drain control of the tunneling energy barrier width, and therefore drain current, have one further important consequence for digital devices. The existence of a drain threshold voltage means that the I_{DS} - V_{GS} characteristics are not independent of V_{DS} , and this appears as something resembling the DIBL (drain-induced barrier lowering) effect in conventional MOSFETs, as can clearly be seen in the simulated curves in Fig. 4.12, as well as the experimental results presented in [9]-[12]. This effect could be renamed DIBN (drain-induced barrier narrowing) since the origin of the curve shift with increasing V_{DS} is the drain voltage’s ability to decrease the tunneling energy barrier width. Fig. 4.12(a), showing the shortest Tunnel FET, with $L = 20$ nm, shows more DIBN than the longer devices with $L = 100$ and 1000 nm. This is due to the increased p-i-n leakage of the short device at the highest value of V_{DS} .

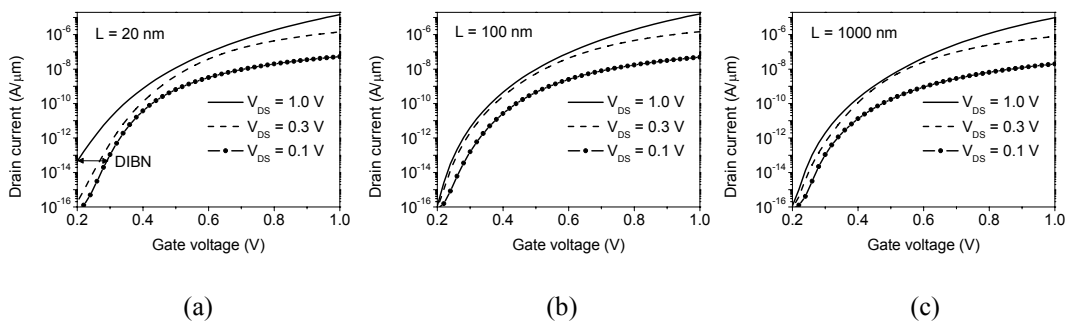


Figure 4.12: I_{DS} vs. V_{GS} for varying values of V_{DS} , showing a curve shift which resembles the DIBL effect in conventional MOSFETs. Three different device lengths shown are (a) 20 nm, (b) 100 nm, and (c) 1000 nm. p-i-n leakage on shortest device length increases the DIBN effect.

4.6 The negative impact of V_{TD} on logic circuit design

The existence of a second threshold voltage in Tunnel FETs can be problematic for circuit designers who are accustomed to designing with conventional MOSFETs, since it leads to several disadvantages, both in static and dynamic operation.

4.6.1 V_{TD} in static operation

The presence of a drain threshold voltage means that rather than the drain current starting to increase linearly as soon as a drain voltage is applied, the current remains extremely low for some mV, and then finally starts to increase linearly. This means that the device current is lower than it would be without V_{TD} . As a consequence, the main effect of the drain threshold voltage under static conditions is to increase the value of $R_{channel}$, defined as V_{DS}/I_{DS} . Figs. 4.13 and 4.14 show I_{DS} vs. V_{DS} and $R_{channel}$ vs. V_{DS} , respectively, for a Tunnel FETs with a varying band gap at the tunnel junction. Looking first at an all-silicon (unstrained) Tunnel FET, it is clear in Fig. 4.13 that more current must be applied to the drain before the current starts to rise linearly, when compared to the other devices. Then in Fig. 4.14, the all-silicon device has the highest $R_{channel}$. This channel resistance is bad for circuit designers because when it is high, losses are high, and V_{DD} is degraded with each successive transistor stage in a circuit, which puts limits on the fanout. One possible solution for decreasing V_{TD} , and thus lowering $R_{channel}$, is to decrease the band gap of the Tunnel FET at the tunnel junction, as discussed in detail in section 3.3. Figs. 4.13 and 4.14 show the effectiveness of this strategy.

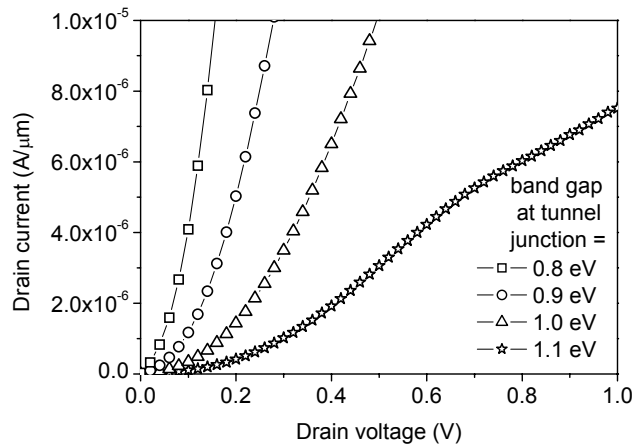


Figure 4.13: I_{DS} - V_{DS} for a double-gate Tunnel FET, for $V_{GS} = 1$ V, with changing band gap at the source-intrinsic region junction.

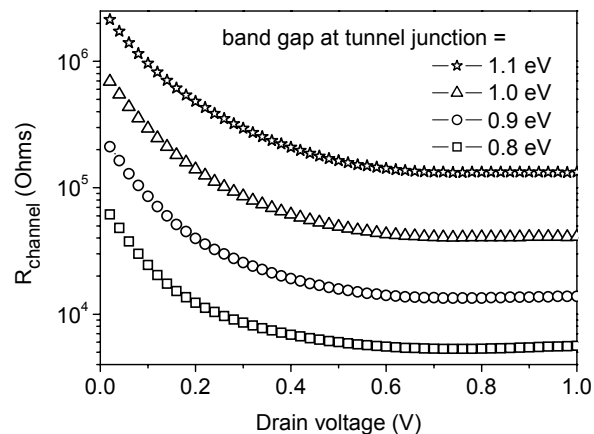


Figure 4.14: $R_{channel}$ vs. V_{DS} for a double-gate Tunnel FET with $L = 50$ nm, $t_{ox} = 1.2$ nm of SiO_2 , and $t_{Si} = 10$ nm. $V_{GS} = 1$ V.

R_{on} , defined as V_{DS}/I_{DS} when the Tunnel FET is in the on-state with $V_{DS} = V_{GS} = V_{DD}$, is especially a concern for low-power circuits where V_{DD} must not go above several hundred mV. Fig. 4.15 shows the extreme increase of the on-resistance when V_{DD} is reduced to 200 mV, for both conventional CMOS operated in subthreshold (from [13]), and for Tunnel FETs. Both types of devices experience drastically increasing values of R_{on} at low V_{DD} . An all-silicon Tunnel FET has an R_{on} value that is approximately 100x higher than that of the conventional CMOS at $V_{DD} = 1$ V, and about 10^5 x higher at $V_{DD} = 0.4$ V.

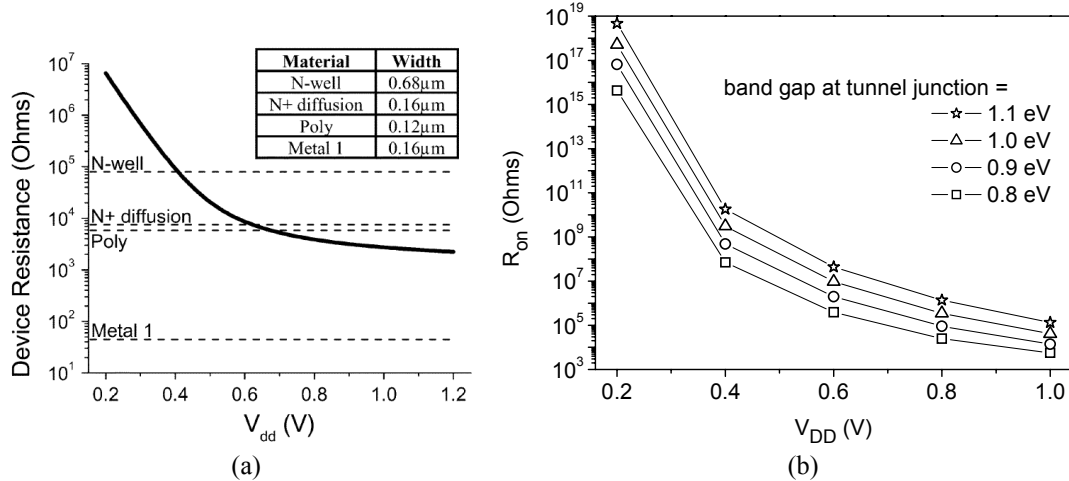


Figure 4.15: R_{on} , for $V_{GS} = V_{DS} = V_{DD}$ (a) for subthreshold conventional CMOS from [13], and (b) for Tunnel FETs with varying band gaps at the tunnel junction. In part (a), resistances of several types of metal lines, 100 μm long and with the given widths, are marked on the graph for reference.

Although changing the band gap at the tunnel junction will not lower R_{on} for Tunnel FETs to the (still high) levels shown for conventional CMOS in Fig. 4.15(a), it can help the situation somewhat. If the band gap is reduced, either by the choice of source material or with strain engineering, to 0.8 eV at the source-intrinsic region junction, R_{on} is reduced by about two orders of magnitude.

4.6.2 V_{TD} in dynamic operation

The main disadvantage of V_{TD} in dynamic operation is once again related to the high values of $R_{channel}$. Any higher resistance of a transistor leads to a higher RC product, meaning higher delay during switching. Fig. 4.15(a) shows resistances of 100- μm -long, minimum-width wires of different interconnect materials as a reference, which demonstrates how the device resistance can surpass the resistance of the interconnects, thus becoming the limiting factor in terms of delay in dynamic operation. For the conventional subthreshold CMOS logic in that figure, at a V_{DD} of 400 mV, the device resistance is more than 1000 times that of a metal wire. The optimized Tunnel FET with a reduced band gap of 0.8 eV at the tunnel junction has a resistance 10^6 times that of the metal wire at $V_{DD} = 400$ mV. Although quantum tunneling is a fast process, which should work as an advantage for Tunnel FET dynamic behavior, the high channel resistance caused by the drain threshold voltage will serve as an obstacle to high speed. Further analysis of the dynamic behavior of Tunnel FETs is beyond the scope of this thesis.

4.7 V_{TD} extraction from experimental data

The experimental results shown in [5] show clear exponential, linear, and saturation regions in the I_{DS} - V_{DS} curves of the investigated Tunnel FETs on SOI. Fortunately for this work, the author of that article was kind enough to provide experimental data, so that some V_{TD} extraction could be carried out on real measured curves.

The p-type Tunnel FETs whose schematic is shown in Fig. 4.16(a) were fabricated by C. Sandow, et al.[5], in Prof. Siegfried Mantl's group at Forschungszentrum Jülich. The I_{DS} - V_{DS} measurements in Fig. 4.16(b), carried out by C. Sandow, are for a device with source and drain doping levels of 4 x

10^{20} atoms/cm³ and a gate length of 6 μm . The transitions between exponential and linear dependence can be clearly seen, as can the saturation at high V_{DS} .

The output characteristics for a 2 μm long Tunnel FET, shown in Fig. 4.17(a) have been used to calculate first, second, and third derivatives, presented in Fig. 4.17(b). Since experimental data are inherently noisier than simulated results, it was necessary to apply a simple smoothing algorithm, mentioned in section 4.3.1, to the g_{ds} curve. The algorithm involved averaging, so with each iteration, points were lost at the low- V_{DS} end of the curve, which explains why the curves are not shown down to $V_{\text{DS}} = 0$ V. Also note that the curve has been flipped to resemble an n-type device curve, for ease of analysis and data extraction. A maximum on the $dg_{\text{ds}}/dV_{\text{DS}}$ curve allowed a threshold voltage extraction, giving $V_{\text{TD}} = 0.65$ V (which would actually be $V_{\text{TD}} = -0.65$ V for the p-type device). We can visually confirm this, since it corresponds with the transition we can see on the curve in Fig. 4.17(a).

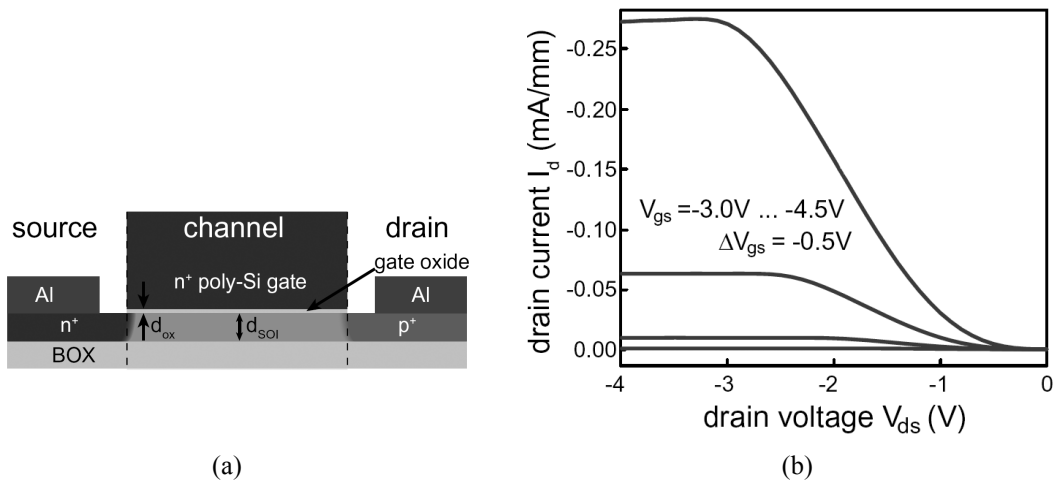
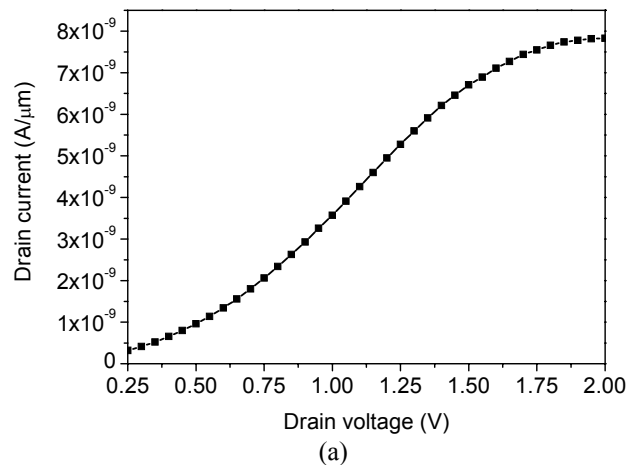


Figure 4.16: (a) Device schematic for p-type Tunnel FET on SOI, and (b) its $I_{\text{DS}}-V_{\text{DS}}$ characteristics, for S/D doping levels of 4×10^{20} atoms/cm³, $L_{\text{g}} = 6 \mu\text{m}$, and $t_{\text{ox}} = 4.5$ nm, from [5].



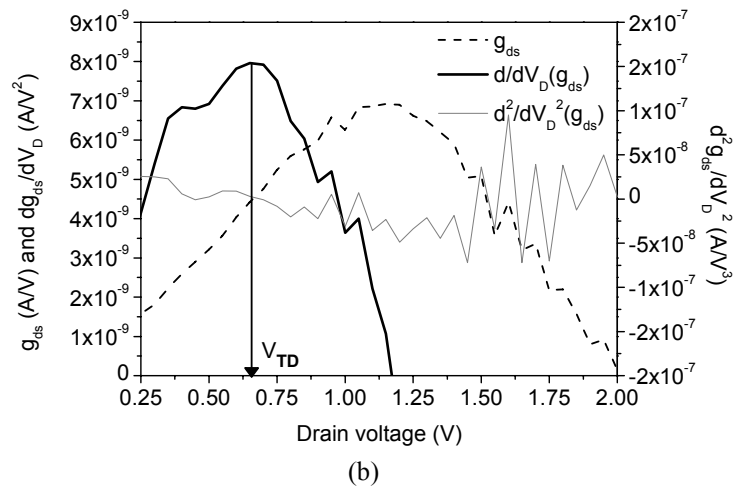


Figure 4.17: (a) Experimental I_{DS} - V_{DS} curve for a Tunnel FET with $L_g = 2 \mu\text{m}$, $t_{ox} = 4.5 \text{ nm}$, and S/D doping of $1.94 \times 10^{20} \text{ atoms/cm}^3$. $V_{GS} = -3 \text{ V}$. P-type device curve has been flipped to look like an n-type device curve, for ease of parameter extraction. (b) First, second, and third derivatives calculated for the I_{DS} - V_{DS} curve in part (a).

Fig. 4.18 shows the dependence of V_{TD} on V_{GS} , confirming the trend shown in Fig. 4.11(b) by numerical simulation. The provided data did not go to gate voltage levels where the saturation of V_{TD} could be seen. The data for V_{GS} values less than 2 V (actually -2 V for the p-type devices) were far too noisy for threshold voltage extraction, even after several smoothing iterations. Likewise, the I_{DS} - V_{GS} data was extremely noisy, and made V_{TG} extraction impossible. This is a drawback of this technique on experimental data. It is also possible, however, to extract the exponential-to-linear transition with a more visual technique directly on I-V curves.

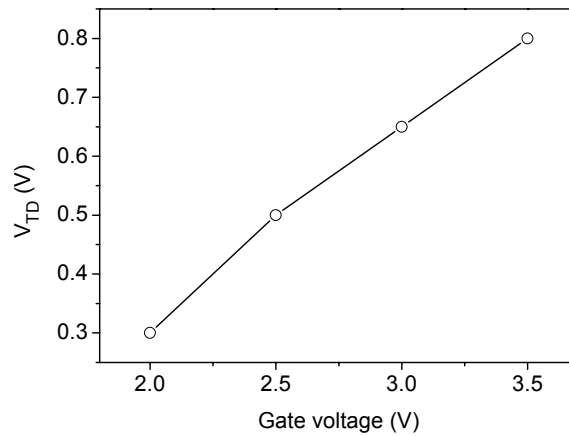


Figure 4.18: V_{TD} vs. V_{GS} , for the same device whose derivatives are shown in Fig. 4.17.

4.8 Conclusion

In this chapter, we have addressed the issue of threshold voltages in Tunnel FETs by proposing a new physical definition, related to the control of energy barrier narrowing, that can describe and quantify the two threshold voltages in band-to-band tunneling devices, V_{TG} and V_{TD} . We support our V_T -definition proposal by a careful analysis using numerical simulations and, finally, experimental data on fabricated SOI Tunnel FETs. Our definition proves to be of practical interest and less arbitrary than the constant current method.

The following technical topics and contributions were presented in this chapter:

- *Why a new definition for Tunnel FET threshold voltage?* (Section 4.1)
This section discussed the need for a new, physically-based definition of threshold voltage. Since a conventional MOSFET's threshold voltage marks the transition between weak and strong inversion, the Tunnel FET's threshold voltage should reflect an analogous change inside the device, rather than being extracted at an arbitrary constant current. The transconductance change method of V_T extraction locates the transition between the gate's strong (quasi-exponential) and weak (linear) control of the drain current, and therefore of the tunneling energy barrier width.
- *Threshold voltages of Tunnel FETs* (Section 4.3)
The insufficiency of the constant current method was explained in detail, and the transconductance change method was described and shown for both V_{TG} , the gate threshold voltage, and V_{TD} , the drain threshold voltage. This "second" threshold voltage, a unique property of Tunnel FETs, was introduced. Both quantities are important for benchmarking device performance.
- *Gate length scaling effects on Tunnel FET threshold voltages* (Section 4.4)
The results of gate threshold voltage extraction were presented for Tunnel FETs with lengths from 20 nm to 2 μm , showing an insensitivity of V_{TG} to gate length for both extraction techniques in the length range studied. The only exception to this insensitivity was the constant-current-derived V_{TG} for the shortest length, 20 nm, where p-i-n leakage caused a slight decrease of threshold voltage.
- *Discussion of V_{DS} and V_{GS} dependence of threshold voltages* (Section 4.5)
This section highlighted the interdependence of the drain and gate voltages' control of the tunneling energy barrier width. One result of this interdependence is that the value of each threshold voltage depends on the bias applied to the opposite terminal: V_{TG} depends on V_{DS} , and V_{TD} depends on V_{GS} . Another result is that I_{DS} - V_{GS} depends on the bias applied to V_{DS} , which creates drain-induced barrier narrowing (DIBN), a shift of the curve analogous to DIBL in conventional MOSFETs.
- *The negative impact of V_{TD} on logic circuit design* (Section 4.6)
This section presented the disadvantages of the drain threshold voltage for Tunnel FETs used in logic circuits. In static operation, the drain threshold voltage leads to increased channel resistance R_{ch} , which means more losses and less potential fanout. In dynamic operation, the high R_{ch} results in a higher RC product, and thus more delay during switching. One way to reduce V_{TD} and R_{ch} is by reducing Tunnel FET band gap at the tunnel junction.
- *V_{TD} extraction from experimental data* (Section 4.7)
Thanks to the measured Tunnel FET characteristics provided by a collaborator, the transconductance change technique could be carried out on actual data, confirming the validity of this method.

All results presented in this chapter, with the exception of the fabricated structure and the data used for threshold voltage extraction in section 4.7, represent original contributions.

4.9 Bibliography

- [1] K. Bhuiwarka, S. Sedimaier, A. Ludsteck, C. Tolksdorf, J. Schulze, and I. Eisele, "Vertical Tunnel Field-Effect Transistor," *IEEE Trans. Electron Devices*, vol. 51, no. 2, pp. 279-82, Feb. 2004.
- [2] K. Bhuiwarka, J. Schulze, and I. Eisele, "Scaling the Vertical Tunnel FET with Tunnel Bandgap Modulation and Gate Workfunction Engineering," *IEEE Trans. Electron Devices*, vol. 52, no. 5, pp. 909-917, May 2005.
- [3] K. Boucart and A. Ionescu, "Double Gate Tunnel FET with high-K gate dielectric," *IEEE Trans. Electron Devices*, vol. 54, no. 7, pp. 1725-1733, Jul. 2007.
- [4] K. Boucart, A. M. Ionescu, "A new definition of threshold voltage in Tunnel FETs", *Solid-State Electronics*, vol. 52, iss. 9, pp. 1318-1323, Sept. 2008.
- [5] C. Sandow, J. Knoch, C. Urban, Q.-T. Zhao, S. Mantl, "Impact of electrostatics and doping concentration on the performance of silicon tunnel field-effect transistors," *Solid State Electronics*, vol. 53, pp. 1126-1129, June 2009
- [6] A. Verhulst, W. Vandenberghe, K. Maex, and G. Groeseneken, "Tunnel field-effect transistor without gate-drain overlap," *Appl. Phys. Lett.*, vol. 91, pp. 053102-1-3, Jul. 2007.
- [7] F. Mayer, C. Le Royer, J.-F. Damlencourt, K. Romanjek, F. Andrieu, C. Tabone, B. Previtali, and S. Deleonibus, "Impact of SOI, Si_{1-x}Ge_xOI and GeOI substrates on CMOS compatible Tunnel FET performance," *IEDM Tech. Dig.*, pp. 163-166, 2008.
- [8] Booth R, White M, Wong H, Krutsick T, "The effect of channel implants on MOS transistor characterization," *IEEE Trans. Electron Devices*, vol. ED-34, pp. 2501-2509, 1987.
- [9] O. Nayfeh, J. Hoyt, and D. Antoniadis, "Strained-Si_{1-x}Ge_x/Si Band-to-Band Tunneling Transistors: Impact of Tunnel-Junction Germanium Composition and Doping Concentration on Switching Behavior," *IEEE Trans. Electron Devices*, vol. 56, no. 10, pp.2264-2269, Oct. 2009
- [10] W. Y. Choi, J. Y. Song, J. D. Lee, Y. J. Park, and B.-G. Park, "70-nm impact-ionization MOS devices integrated with tunneling FETs," *IEDM Tech. Dig.*, pp. 955-958, 2005.
- [11] P.F. Wang, K. Hilsenbeck, T. Nirschl, M. Oswald, C. Stepper, M. Weis, D. Schmitt-Landsiedel, and W. Hansch, "Complementary tunneling transistor for low power application," *Solid-State Electronics*, vol. 48, no. 12, pp. 2281-2286, Dec. 2004.
- [12] K. Moselund, H. Ghoneim, M. T. Björk, H. Schmid, S. Karg, E. Lörtscher, W. Riess, and H. Riel, "Comparison of VLS grown Si NW tunnel FETs with different gate stacks," *Proc. ESSDERC*, pp. 448-451, 2009.
- [13] S. Hanson, B. Zhai, M. Seok, B. Cline, K. Zhou, M. Singhal, M. Minuth, J. Olson, L. Nazhandali, T. Austin, D. Sylvester, and D. Blaauw, "Exploring Variability and Performance in a Sub-200-mV Processor," *IEEE Journal of Solid-State Circuits*, vol. 43, no. 4, pp. 881-891, Apr. 2008.

Chapter 5

Tunnel FET Length Scaling

In this chapter, the length scaling of the silicon double-gate Tunnel FET is studied. It is demonstrated that the scaling of the Tunnel FET is completely different than that of conventional MOS transistors, in terms of the modifications that must be made in order to keep a constant electric field, and in terms of performance gain. The effects of length scaling on device parameters are shown for devices with SiO₂ and high-k gate dielectrics, both with and without an interfacial layer. The simulations show that length scaling does not dramatically affect switch characteristics such as subthreshold swing, I_{on} , and I_{off} down to gate lengths of about 20 to 40 nm -- scaling limits are reached sooner by Tunnel FETs with an SiO₂ gate dielectric, while those with a high-k dielectric can be scaled further before characteristics are affected. A discussion of the length dependence of the transconductance, g_m , and output conductance, g_{ds} , of the Tunnel FET is also presented. Then the electron and hole concentrations and the band diagrams inside the silicon layer of the Tunnel FET in the off-state at different gate lengths are examined. Next, some experimental results published by another group are cited to support the simulated predictions. Finally, supply voltage scaling is briefly investigated, and some design suggestions given to boost the on-current degraded by V_{DD} reduction.

5.1 Scaling: conventional MOSFETs vs. Tunnel FETs

For any future logic device, be it the conventional MOSFET or a device that could replace or complement it such as the Tunnel FET, we would like to be able to scale the device down, for reasons of lower cost and faster speed. It is important when scaling Tunnel FETs, however, to not blindly follow the same scaling conventions that we have used for conventional MOSFETs. Bhuwalka, et al. have discussed scaling rules for silicon Tunnel FETs, and have pointed out that constant field scaling does not apply [1]. Nevertheless, much of their investigation into scaling decreases gate length and oxide thickness at the same time [1],[2], according to Dennard's scaling rules [3]. When studying Tunnel FET behavior, it is more useful to scale one parameter at a time, in order to understand the effects of each change and not confound them.

In this chapter, we look at length scaling independent from the scaling of other parameters in order to isolate the effects. This approach is motivated by the fact that the conduction mechanism in the Tunnel FET is completely different from that in the MOSFET; I_{on} is not dictated by the gate-controlled intrinsic MOS part (i-region) of the Tunnel FET, but rather by the tunnel junction between the source and the intrinsic region.

Dennard's scaling rules are based on constant-field scaling, meaning that the electric fields inside the device have the same magnitude, even when all the device dimensions change. This is accomplished by scaling down all device dimensions in x , y , and z , including gate length, device width, gate dielectric thickness, and device thickness, by a factor of $1/\kappa$. At the same time, the supply voltage V_{DD} is scaled by the same factor, $1/\kappa$. The doping must increase by a factor of κ in order to scale down the depletion widths by $1/\kappa$. When scaling is carried out in this way, the internal electric fields of the conventional MOSFET stay constant, as do the carrier velocity, the inversion-layer charge density, the channel resistance, and the power density. The drift current, which is dominant in the on-state, scales down by $1/\kappa$, while the diffusion current, dominant in sub-threshold, increases by a factor of κ [4].

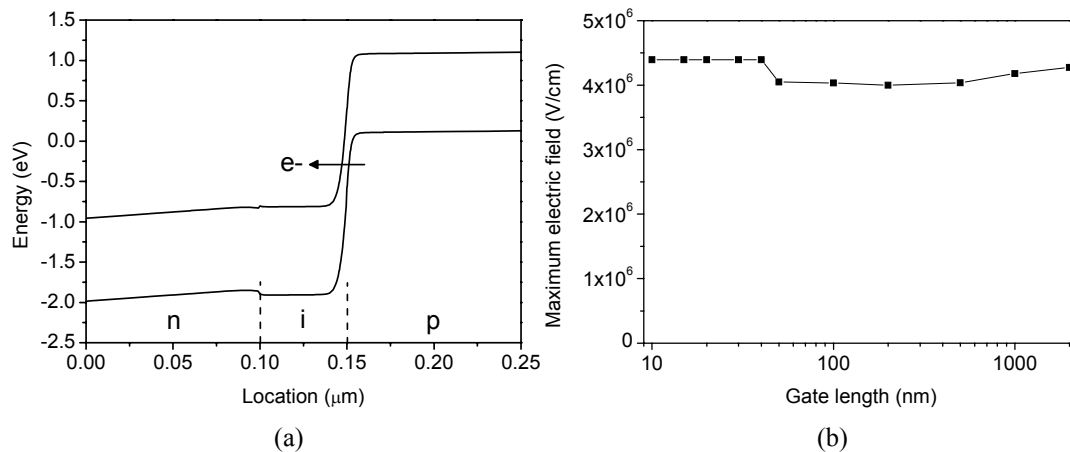


Figure 5.1: (a) Energy band diagram taken horizontally across the body of the Tunnel FET in the on-state, 2.5 nm from the dielectric interface. $\epsilon_{\text{dielectric}} = 25$, $L = 50$ nm, $V_{DS} = 1$ V, $V_{GS} = 1.5$ V. The electric field is high at the source junction and close to zero elsewhere. (b) Maximum electric field for Tunnel FETs in the on-state, with lengths from 10 nm to 2 μm . The maximum is always at the tunnel junction, and is independent of device length.

We can draw several conclusions pertaining to Dennard scaling of MOSFETs, and then make a comparison with Tunnel FETs. First, while scaling all dimensions simultaneously is the correct way in which to maintain a constant electric field in conventional MOSFETs, such is not the case in Tunnel FETs. The energy bands shown in Fig. 5.1(a) show that a high electric field (where the bands have a steep slope) is present only at the source junction in the on-state. Elsewhere, where the bands are flat or nearly flat, the electric field is close to zero. This is the case for any device length, and Fig. 5.1(b) shows that the maximum electric field vs. gate length is nearly constant. The maximum electric field is always at the tunnel junction, and the variations seen in Fig. 5.1(b) come from

unavoidable differences in the meshes for different device lengths. This means that device length can be scaled independently of other dimensions, doping levels, and supply voltage. The second big difference between MOSFET and Tunnel FET scaling is that off-current does not increase by the scaling factor κ for Tunnel FETs. On the other hand, p-i-n diode leakage becomes very important for device lengths below some critical length L_{crit} , as will be discussed in detail later in the chapter.

One more important point about device scaling must be made here. One crucial part of conventional MOSFET scaling has been to decrease V_{DD} . As discussed in Chapter 2, dynamic power consumption depends on V_{DD}^2 , so when constant field scaling is carried out properly, the dynamic power consumption of each device decreases, and the dynamic power density of a circuit or chip remains constant. In addition, MOSFET drift current per unit width stays constant. (Chapter 2 showed that the scaling rules are no longer being followed, and V_{DD} is no longer decreasing with device dimensions, but for the purposes of this paragraph, that is irrelevant.)

The situation for Tunnel FETs is not at all the same. Since device characteristics remain nearly identical for a large range of device lengths, the drop in current level cannot be compensated by scaling down length. If we want to scale down V_{DD} in order to decrease dynamic power consumption per device, and if we want to simultaneously keep on-current constant, then it is necessary to also scale other important device dimensions, such as t_{ox} and tunnel junction abruptness. This will be investigated in more detail in section 5.7.

5.2 Simulation parameters

The Tunnel FETs presented here were simulated using Silvaco Atlas, version 5.11.24.C. A non-local band-to-band tunneling model was used, band-gap narrowing was applied, and gate leakage was ignored. In all simulations, the source, intrinsic region, and drain were doped at 1×10^{20} , 1×10^{17} , and 5×10^{18} atoms/cm³ respectively. Junctions were quasi-perfectly abrupt (junction width 0.5 nm) in sections 5.3, 5.4, and 5.5. Three different gate stacks will be shown in this chapter. The first uses a 3 nm thick SiO₂ gate dielectric, with dielectric constant $\epsilon = 3.9$. The second employs 3 nm of a high-k dielectric with $\epsilon = 25$, corresponding to HfO₂. The third incorporates a gate dielectric with two components: 1 nm of an interfacial layer of oxynitride ($\epsilon = 5.7$) and 2 nm of a high-k dielectric ($\epsilon = 25$), which corresponds to a more realistic fabrication process. All Tunnel FETs simulated here use a metal gate with a work function of 4.5 eV. Simulations were carried out for Tunnel FETs with gate lengths ranging from 10 nm to 2 μm .

5.3 Length scaling and basic device characteristics

In this section, transfer and output characteristics will be presented for Tunnel FETs with two different types of gate dielectrics: SiO₂, and a high-k dielectric with a permittivity of 25. From these characteristics, the transconductance, g_m , and the output conductance, g_{ds} , will be derived and discussed.

The $I_{\text{DS}}-V_{\text{GS}}$ characteristics for n-type devices with a high-k gate dielectric are shown in Fig. 5.2(a), with those for Tunnel FETs with an SiO₂ gate dielectric shown in Fig. 5.2(b) for comparison. The high-k dielectric provides the gate with better capacitive control of the tunnel junction, and as has already been shown in [5] and [6] and Chapter 3, this leads to better characteristics as the device turns on – both a lower subthreshold swing and a higher on-current. We see from Fig. 5.2 that the high-k dielectric also gives the gate more control to turn the device off, especially at shorter gate lengths. This advantage will be discussed more in section 5.5.

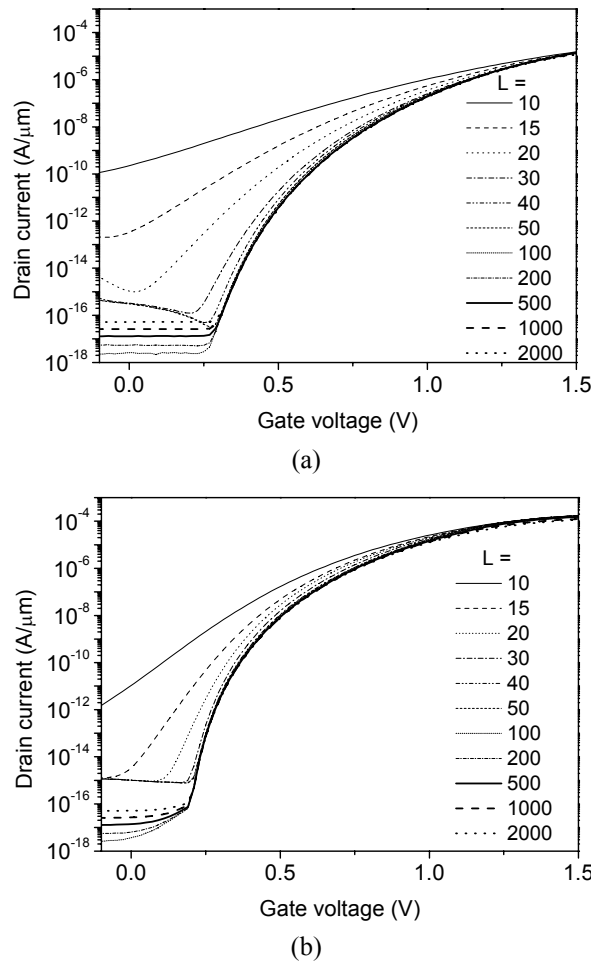


Figure 5.2: I_{DS} - V_{GS} characteristics with $V_{DS} = 1$ V for Tunnel FETs of varying lengths for (a) $\epsilon_{\text{dielectric}} = 3.9$ and (b) $\epsilon_{\text{dielectric}} = 25$. Tunnel FETs with a high-k dielectric can be scaled to shorter gate lengths before the gate loses control of the tunnel junction in the off-state.

The same I_{DS} - V_{GS} characteristics are shown in Fig. 5.3(a) and (b) on a linear scale (left y-axis), along with the transconductance (right y-axis). While it was almost not visible on a log scale in the previous figures, here we can see that for gate lengths greater than about 50 nm, the on-current starts to drop slightly. This will be shown in more detail in the following section. This reduction in I_{on} is related to the higher resistance of the longer intrinsic region, in series with the tunnel junction. When the source-to-drain current is high enough, which only happens in a well-optimized Tunnel FET, it implies that the resistance of the tunnel junction has become so low that it is comparable to the resistance of the intrinsic region. Only in this situation will the length of the intrinsic region play a role in the value of I_{on} , and cause a current decrease for long devices.

The transconductance shown on the right y-axes of Fig. 5.3(a) and (b) shows that, as expected, the transconductance of high-k devices is about an order of magnitude higher than that of SiO_2 devices. In contrast with the conventional MOSFET, the Tunnel FET transconductance does not significantly decrease after hitting a maximum value as seen in Fig. 5.3(b), which suggests that it is not dictated by the value of the carrier mobility in the intrinsic MOSFET-like intrinsic region degrading at higher transverse electric fields. Moreover, the effect of length scaling is both qualitatively and quantitatively different than that of the conventional MOSFET; a much higher dependence of g_m on the device length is observed in the subthreshold region compared to the above-threshold region, where the influence of length scaling is quite limited. This shows that even for a well-optimized Tunnel FET, the resistance of the intrinsic region (which only plays a role in the above-threshold region when L is long) is less important than the leakage which degrades the off-state when L is short.

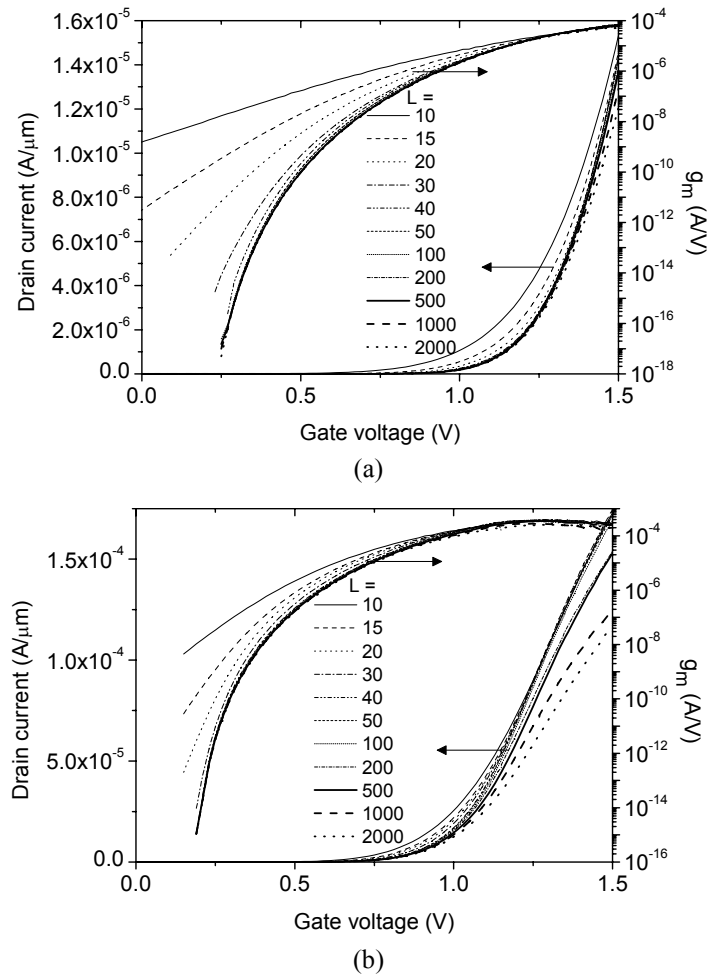


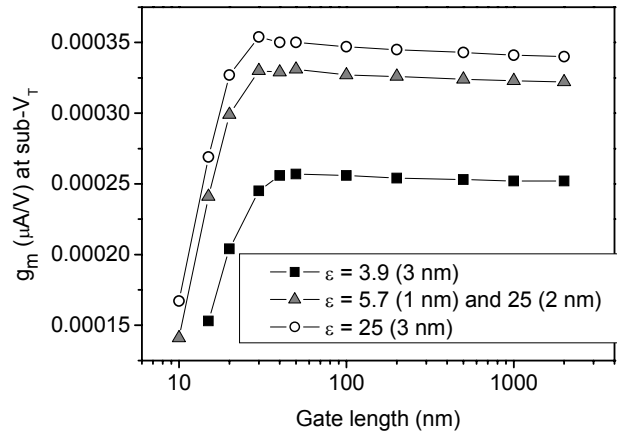
Figure 5.3: Linear I_{DS} - V_{GS} characteristics (left y-axis) and transconductance (right y-axis) with $V_{DS} = 1$ V for Tunnel FETs of varying lengths for (a) $\epsilon_{\text{dielectric}} = 3.9$ and (b) $\epsilon_{\text{dielectric}} = 25$.

Fig. 5.3 showed the transconductance, g_m , for the entire range of gate voltages. Now looking at g_m at specific values of V_{GS} on the I_{DS} - V_{GS} curves, the improved scalability of high-k devices is once again apparent. Fig. 5.4 shows the transconductance in the subthreshold region (Fig. 5.4(a)), at threshold (Fig. 5.4(b)), and at its maximum (Fig. 5.4(c)). In subthreshold, and even more markedly right at V_T , we see that g_m remains nearly constant for longer devices, but then drops for shorter devices due to diode leakage current at low V_{GS} . g_m scales better for high-k devices than for SiO_2 Tunnel FETs, since there is less diode leakage for the devices with a high-k gate dielectric at any given length. Though Figs. 5.4(a) and (b) show the same general trend, the percentage by which g_m drops for short lengths is much greater in the subthreshold region, shown in (a). This is once again due to diode leakage current. The benefits given by a high-k dielectric are best seen in Fig. 5.4(b), where devices with an SiO_2 dielectric have a reduced transconductance for gate lengths less than 100 nm, while high-k devices can be scaled to 20 nm.

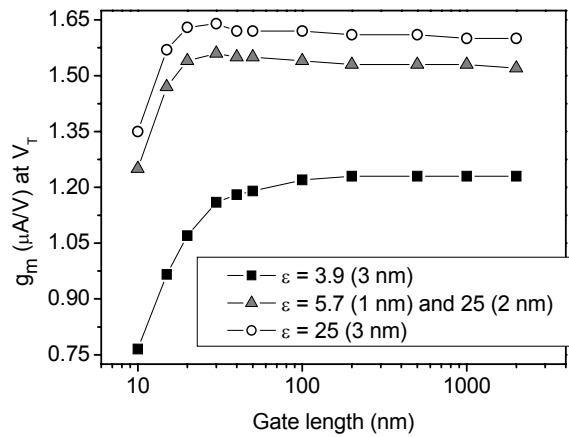
For a transconductance value taken at $V_{G\text{max}} = 1.5$ V, as seen in Fig. 5.4(c), on the other hand, the degradation of $g_{m,\text{max}}$ is related to an effect of series resistance in the intrinsic region, which is more pronounced for longer gate lengths. While the maximum transconductance of a conventional MOSFET is proportional to $1/L$, we see clearly that this is not the case for the Tunnel FET. As mentioned earlier, the resistance in the intrinsic region plays a role in I_{on} degradation only when the device is optimized enough to have a high on-current, showing that the resistance of the tunnel junction is low, and becomes comparable with the resistance of the intrinsic region. This is demonstrated by Fig. 5.4(c), in which the device with 3 nm of a high-k dielectric shows a degradation of 34% in its g_m when the device length is increased from 50 nm to 2 μm . The device with 3 nm of SiO_2 , on the other hand, is less well-optimized, and has a lower on-current. The

intrinsic region's resistance is therefore less important for long devices, and there is only 22% degradation of g_m when the intrinsic region length is increased from 50 nm to 2 μm .

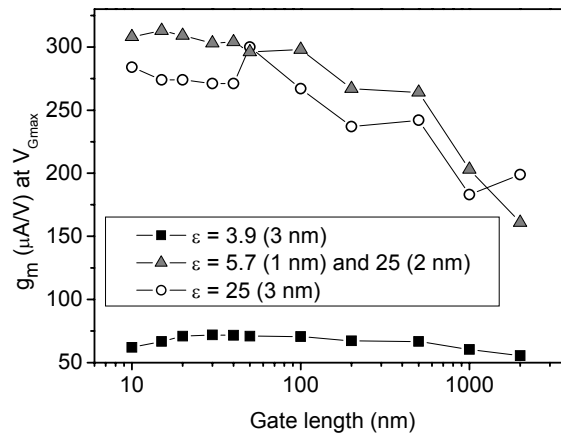
The noise in the curves in Fig. 5.4(c) is an artifact of simulation, from necessary meshing differences in devices of different intrinsic region lengths.



(a)



(b)



(c)

Figure 5.4: Tunnel FET transconductance vs. L (a) in the sub-threshold region (at $I_{DS} = 10^{-11}$ A), (b) at the threshold voltage (determined by $I_{DS} = 10^{-7}$ A), and (c) at the maximum voltage $V_{Gmax} = 1.5$ V. $V_{DS} = 1$ V. Tunnel FETs with high- k dielectrics can be scaled to shorter lengths before the transconductance in sub-threshold and at threshold drops.

In Fig. 5.5(a) and (b), the output characteristics (I_{DS} vs. V_{DS}) are shown for Tunnel FETs with SiO_2 and high-k gate dielectrics. The saturation mechanism of I_{DS} - V_{DS} characteristics is completely different from that of the conventional MOSFET; it is essentially related to the saturation (or lack thereof) of the tunneling barrier width with the applied V_{DS} voltage at a given V_{GS} voltage; this explains why in Figs. 5.5(a) and (b), we still see a slight continuous increase of I_{DS} at high V_{DS} . Fig. 5.5(c) depicts the output conductance, g_{ds} versus V_{DS} . The high-k Tunnel FETs have a drain-source conductance about an order of magnitude higher than those with SiO_2 . Both types of devices show a g_{ds} that increases with V_{DS} and then reaches a maximum before degrading due to series resistance at high V_{DS} . This g_{ds} is higher at shorter gate lengths, as was g_m , because there is less series resistance of the intrinsic region.

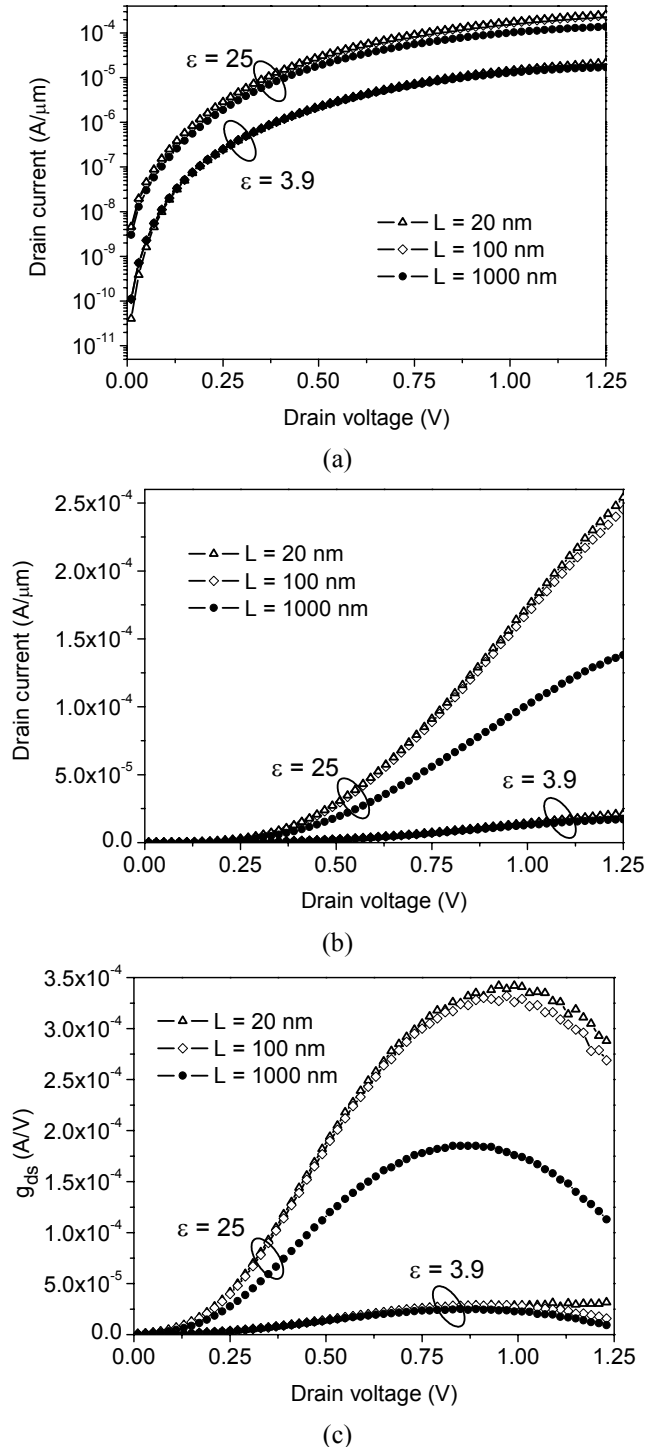


Figure 5.5: (a) I_D vs. V_D on a log scale, (b) I_D vs. V_D on a linear scale, and (c) $g_{ds} = dI_D/dV_D$ vs. V_D , all for $L = 20, 100,$ and 1000 nm, $\epsilon_{\text{dielectric}} = 3.9$ and 25 . $V_G = 1.5$ V.

5.4 Scaling study in Tunnel FETs with high-k gate stack

For conventional MOSFET fabrication, the incorporation of an interfacial dielectric layer is crucial when using a high-k gate dielectric, due to the deleterious effects of high-k materials on the Coulomb scattering component of the mobility in the MOSFET channel [7]. Because the simulated Tunnel FETs studied here should be compatible with standard CMOS fabrication techniques, it is important to look at the effects of length scaling on devices incorporating an interfacial layer under the high-k dielectric. In this section, we investigate the length dependence of the following device characteristics: (i) I_{Dmax} , (ii) I_{on}/I_{off} , (iii) V_T , and (iv) subthreshold swing, for Tunnel FETs with three different configurations of gate dielectric. Two are the same as in the preceding section: 3 nm of SiO_2 , and 3 nm of a high-k dielectric with a permittivity of 25. The third has a 1 nm interfacial oxynitride layer with $\epsilon_{dielectric} = 5.7$, topped by 2 nm of the high-k dielectric with $\epsilon_{dielectric} = 25$.

The values of $\%I_{Dmax}$, the maximum drain current of each device extracted at $V_{DS} = 1$ V and $V_{GS} = 1.5$ V, normalized by I_{Dmax} for the shortest length studied here ($L = 20$ nm) are shown in Fig. 5.6 for Tunnel FETs with the three different types of dielectric stacks. Starting with the longest devices on Fig. 5.6 ($L = 2$ μ m) and then scaling down, the on-current increases considerably until about $L = 50$ nm, and then remains nearly constant when scaling from 50 nm down to 20 nm. We conclude that the dependence of I_{Dmax} in the Tunnel FET is completely different than that of a simplified conventional MOSFET, shown for reference (and calculated very simply as $I_{Dmax} \propto 1/L$) in Fig. 5.6. The maximum drain current of a Tunnel FET does not scale inversely proportionally to the gate length as it does for a MOSFET; its improvement is rather a consequence of the reduction of the in-series resistance of the intrinsic region. At lengths less than about 100 nm, when this resistance becomes quasi-negligible in comparison with the resistance of the tunnel junction, the current gain versus L saturates.

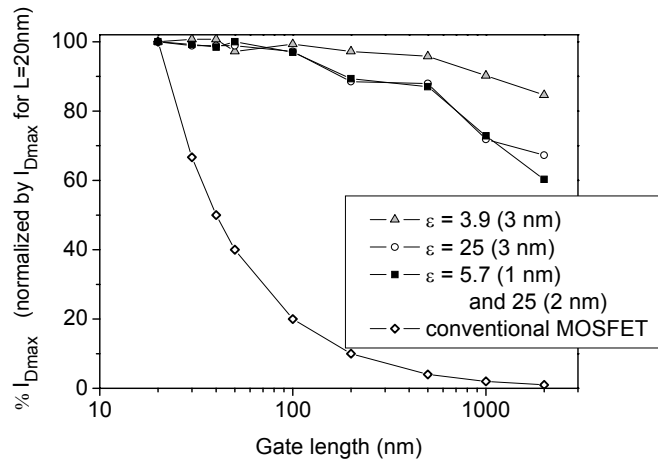


Figure 5.6: $\% I_{Dmax}$ (taken at $V_{DS} = 1$ V, $V_{GS} = 1.5$ V) vs. L . Here, $\% I_{Dmax} = 100 \cdot (I_{Dmax} / I_{Dmax} \text{ for } L=20 \text{ nm})$. On-current remains nearly constant for devices 50 nm long and less, I_{Dmax} for a 2 μ m device is about 70% that of the 20 nm device (rather than 1%, as it would be for a conventional MOSFET). The trend for the conventional MOSFET assumes $I_{Dmax} \propto 1/L$ and ignores any other scaling effects in the deep submicron range.

In Fig. 5.7, we see the I_{on}/I_{off} ratio for Tunnel FETs with the three dielectric stacks studied. Individual data points should not be over-interpreted, since meshing differences between devices of different gate lengths play a role (and might explain the abrupt change between $L = 50$ nm and $L = 100$ nm). The general trends are clear, however. Since the off-current for devices 100 nm long and longer is lower than for shorter devices (see Fig. 5.2), the I_{on}/I_{off} ratio is higher for those long devices. When scaling, however, the real region of interest is 50 nm and less. Here we see that for the devices with SiO_2 , the ratio is constant from 50 nm down to 30 nm, and then starts to fall off. For high-k devices, however, the ratio remains constant down to 20 nm before falling off at around a 15 nm gate length. So for this important parameter, scaling can be pushed further for high-k dielectric devices than for those with SiO_2 .

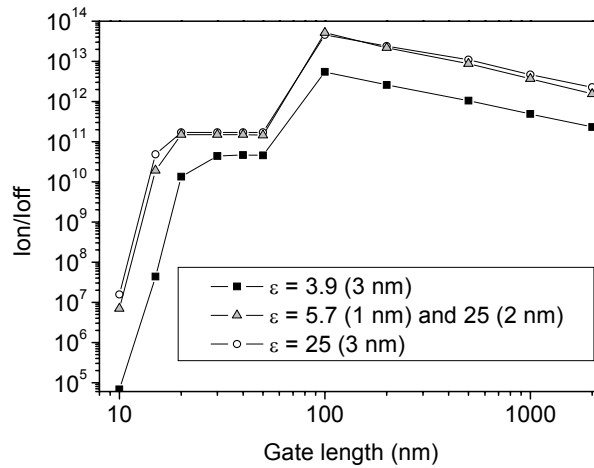


Figure 5.7: $I_{\text{on}}/I_{\text{off}}$ vs. L . The higher ratio for gate lengths greater than 50 nm are due to the lower off-current in those devices. When scaling down from 50 nm, the high-k devices retain their good ratios down to shorter lengths. For I_{off} , $V_D = 1$ V and $V_G = 0$ V. For I_{on} , $V_D = 1$ V and $V_G = 1.5$ V.

Looking at the dependence of threshold voltage on length as shown in Fig. 5.8, and taking Tunnel FETs with a gate length of 50 nm as the reference, V_T rolloff is observed. No matter what the gate dielectric, from 50 nm to 30 nm, the threshold voltage falls by about 3% from its original value, and from 50 nm to 20 nm, by about 7%. This V_T rolloff has a completely different cause than in a conventional MOSFET, coming from diode leakage (band-to-band tunneling that takes place at low V_G and cannot be controlled by the gate). The only way to control the rolloff is to have better electrostatic control of the tunnel junction, which will keep V_T constant down to shorter gate lengths. This can be carried out by any standard technique to improve electrostatic control, several of which were already investigated or mentioned during the discussion of optimization in Chapter 3, for example, reducing the gate dielectric thickness or the body thickness, or by moving to a gate-all-around nanowire platform. The magnitude of the threshold voltage can, of course, be adjusted by the choice of gate material, and is defined here using the constant current method, with threshold set at $I_{\text{DS}} = 10^{-7}$ A/ μm , when $V_{\text{DS}} = 1$ V.

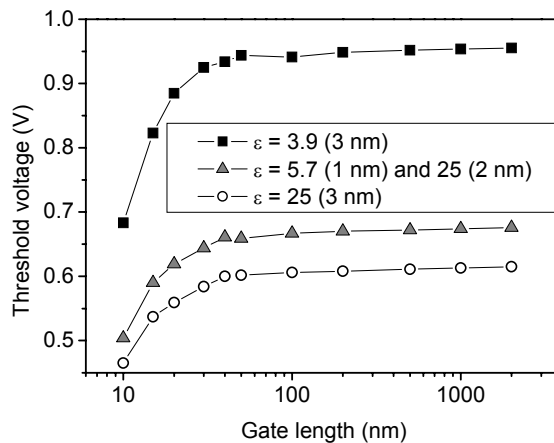


Figure 5.8: Threshold voltage vs. L . V_T changes at approximately the same rate for all dielectric values when scaling down. $V_{\text{DS}} = 1$ V.

Continuing to look at the effects of length on important device parameters, Fig. 5.9 shows the changes in both point and average subthreshold swing. Average swing is calculated between the point where the Tunnel FET starts to turn on, and threshold, while point swing is taken at the point on the $I_{\text{DS}}-V_{\text{GS}}$ curve where slope is steepest (refer to Chapter 1). As with $I_{\text{on}}/I_{\text{off}}$ and g_m at V_T , the high-k dielectrics show a clear advantage for subthreshold swing when scaling. Point swing deteriorates less with high-k than with SiO_2 down to a gate length of 30 nm. Even more important,

however, is average swing (defined in Fig. 2.20), since it is more relevant for the use of Tunnel FETs in circuit applications. The average swing of high-k Tunnel FETs is nearly constant down to a gate length of 20 nm. The scaling trends for swing, g_m at V_T , I_{on}/I_{off} , and V_T show that below a certain gate length, the Tunnel FET is no longer a functioning device, and is well on its way to becoming a simple diode (two-terminal device) rather than a gated diode (three-terminal device).

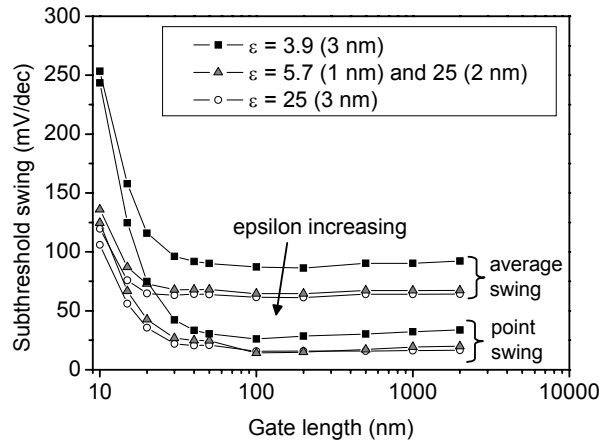


Figure 5.9: Point and average subthreshold swing vs. L . For both types of swing, Tunnel FETs with a high-k dielectric scale better before the swing deteriorates: to 30 nm for the point swing, and to 20 nm for the average swing. $V_{DS} = 1$ V.

5.5 Depletion, energy bands, and the off-state

The degradation of most of the device parameters presented here – I_{on}/I_{off} , V_T , g_m at V_T , and subthreshold swing – can be explained by the decreased ability of the gate to turn the device off at $V_{GS} = 0$. Bhuwalka, et al. have already explained this [1],[2],[8] as the effect of Zener breakdown, meaning that electron tunneling still takes place when it shouldn't, in the off-state. There are several ways of thinking about and looking at Tunnel FET behavior in the off-state, and we will examine two of those here.

In 1992, Toshio Baba predicted that it would be possible to scale Tunnel FETs (or Surface Tunnel Transistors, as he called them) to a gate length equal to the “depletion layer width where electron tunneling is suppressed” [9], on the order of 10 nm. In a p-i-n diode, the depletion region includes the entire intrinsic region in addition to the depleted areas of the p- and n-regions. Examining the electron and hole concentrations in a horizontal cross-section of the Tunnel FET body reveals the differences in abruptness at the edges of the intrinsic region for devices with SiO_2 and a high-k dielectric. Fig. 5.10 shows the electron and hole concentrations for Tunnel FETs with SiO_2 (left), and a high-k dielectric (right), for the gate lengths (a) $L = 40$ nm, (b) $L = 20$ nm, and (c) $L = 10$ nm. The electron concentration profiles are similar for the different dielectrics, but the hole concentration profiles at the source-intrinsic region junction are better-defined for the high-k devices, thanks to the better capacitive control by the gate. This is especially noticeable for the Tunnel FETs with gate lengths of 40 and 20 nm. In the devices with SiO_2 the gate is already losing control of the off-state at these gate lengths, and the depleted regions are narrower and less defined. At 10 nm, both types of devices are experiencing Zener breakdown (electron tunneling) in the off-state, though the high-k device shows the small advantage of still having a lower carrier concentration in the intrinsic region under the gate.

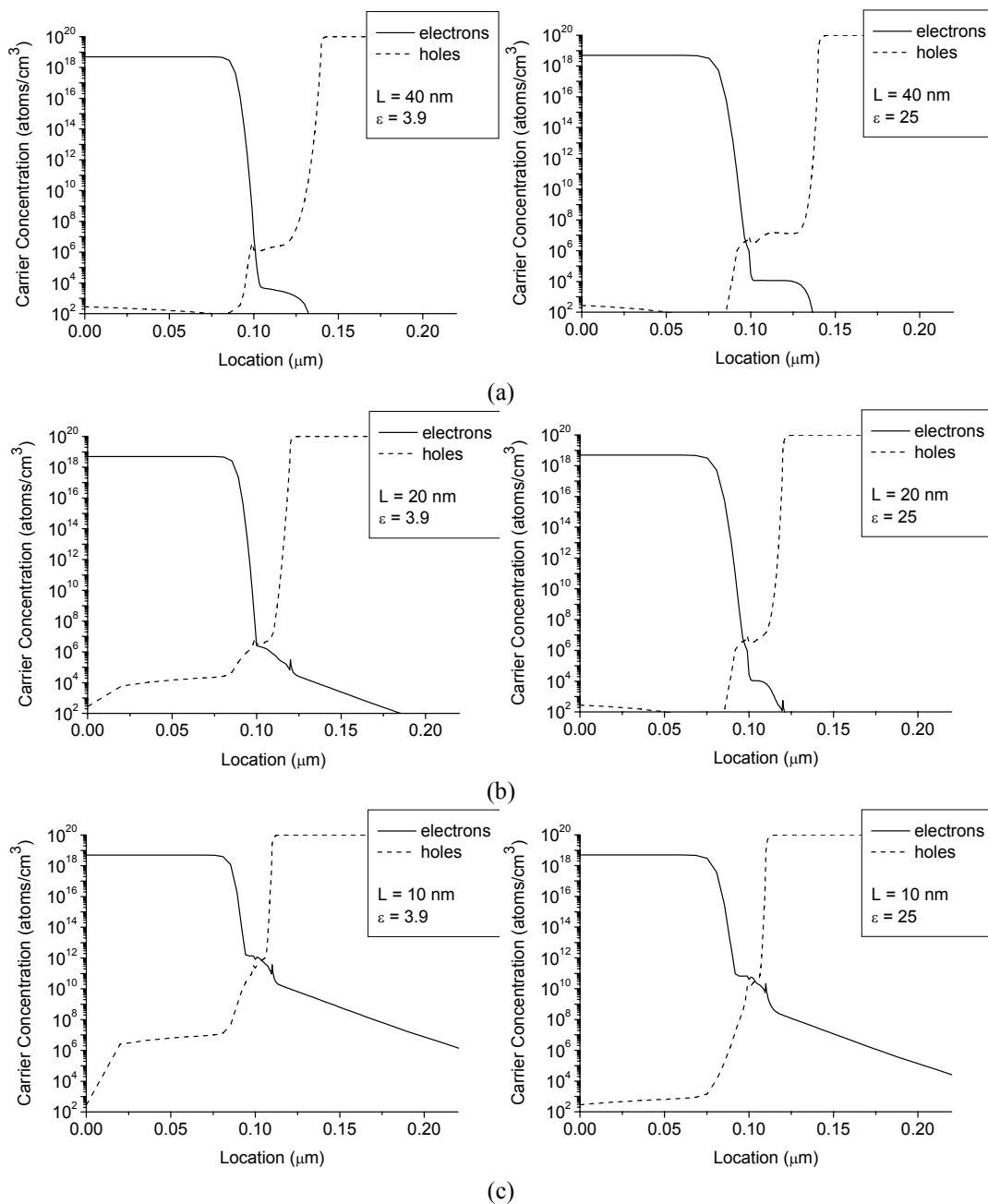


Figure 5.10: Electron and hole concentrations in a horizontal cross section of the Tunnel FET body, 2.5 nm from the dielectric surface, for two different gate dielectrics: $\epsilon_{\text{dielectric}} = 3.9$ (left), and $\epsilon_{\text{dielectric}} = 25$ (right), in the off-state ($V_{DS} = 1$ V and $V_{GS} = 0$ V). (a) $L = 40$ nm, (b) $L = 20$ nm, (c) $L = 10$ nm. The high-k Tunnel FETs show a wider depletion region for the same applied voltages.

The second way of looking at Tunnel FET behavior in the off-state is to look at the extent of the depletion region with energy band diagrams across the Tunnel FET body. The depletion region corresponds to the region between the flat bands in the n-region at the left and the flat bands in the p-region at the right. It is, however, easier to understand at what point tunneling will begin to take place in the OFF-state by examining the width of the band-to-band energy barrier at the i-p junction; a rule of thumb is that “significant” tunneling begins at barrier widths of less than about 10 nm. Fig. 5.11 shows cross-sections of Tunnel FETs with SiO_2 (left) and a high-k dielectric (right), having gate lengths of 40, 20, and 10 nm. The gates of the 40 nm devices still manage to turn the devices off, and we see that the intrinsic region bands are well-defined (especially for the high-k device) and create wide energy barriers between the source and intrinsic, and intrinsic and drain regions. At 20 nm, the SiO_2 Tunnel FET is experiencing Zener breakdown in the off-state. The 10 nm devices are

both experiencing band-to-band electron tunneling at $V_{GS} = 0$ V, and we see that the intrinsic regions are no longer wide enough to be seen as distinct regions in the band diagrams, nor to block tunneling current from flowing. These short devices can no longer turn off, as already seen in the I_{DS} - V_{GS} curves in Fig. 5.2.

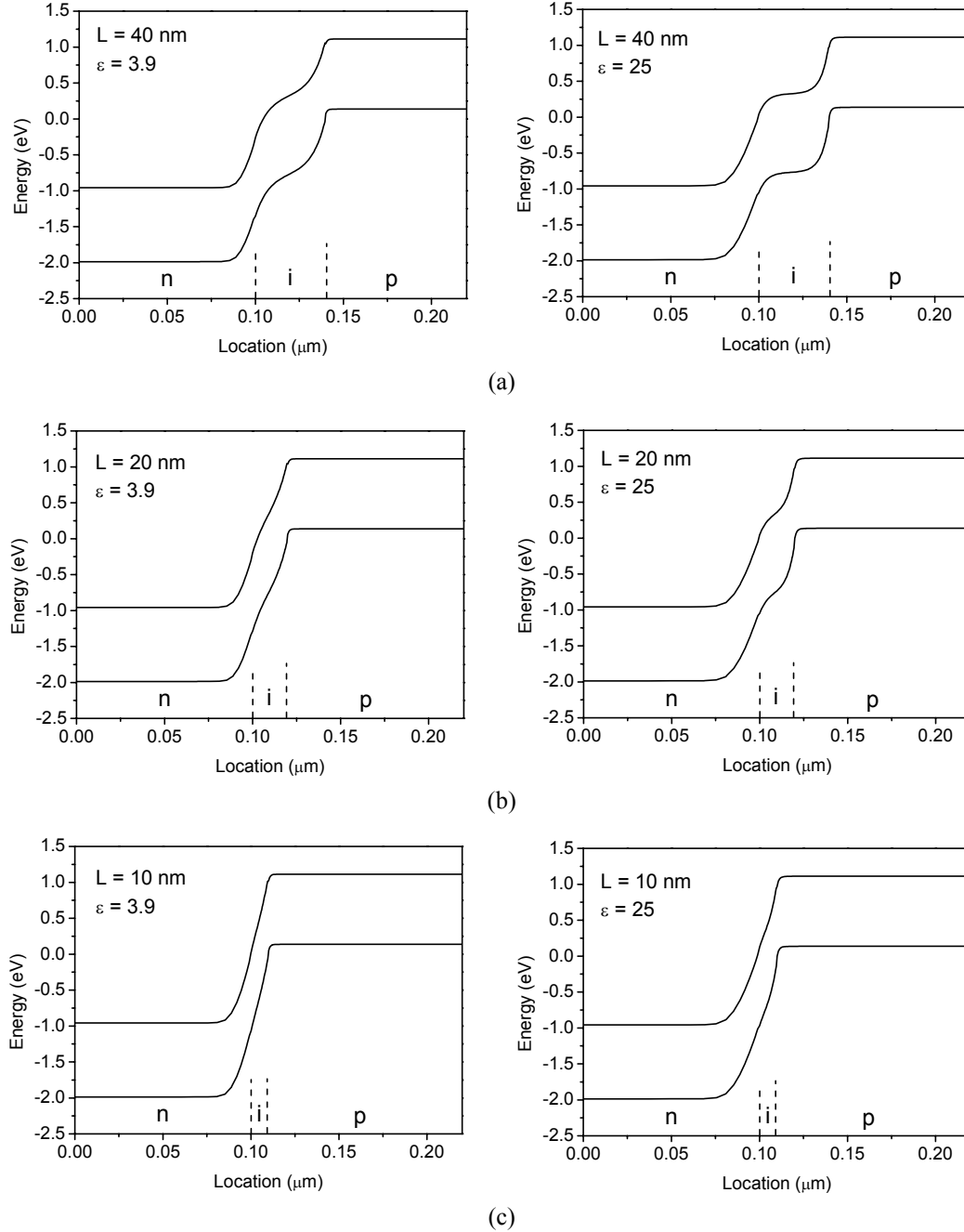


Figure 5.11: Energy band diagrams in a horizontal cross section of the Tunnel FET body, 2.5 nm from the dielectric surface, for three different gate lengths: (a) 40 nm, (b) 20 nm, and (c) 10 nm, in the off-state ($V_{DS} = 1$ V and $V_{GS} = 0$ V). $\epsilon_{\text{dielectric}} = 3.9$ (left) and $\epsilon_{\text{dielectric}} = 25$ (right). At shorter gate lengths, the energy barrier between the intrinsic and p-regions is not sufficiently wide to block band-to-band tunneling.

5.6 Length scaling trend confirmation from experimental data

Before making the transition from simulated n-type Tunnel FETs to experimental p-type Tunnel FETs that will confirm our predictions from simulation, it is important to present some simulated p-type device characteristics. In general, experimentalists prefer to fabricate p-type Tunnel FETs due to the importance of the extremely abrupt doping profile at the tunnel junction, and the fact that it's easier to attain on the n+ side (with arsenic, for example) than on the p+ side (with boron, which diffuses easily). Fig. 5.12 shows simulated transfer characteristics for both n- and p-type double-gate Tunnel FETs, both with SiO₂ and with a high-k dielectric. The absolute value of the current is shown, for easy comparison of the two types of curves. V_{BTBT} , the voltage at which band-to-band tunneling begins, is 0.3 V and 0.2 V for the n-type devices and -0.82 V and -0.72 V for the p-type devices, for SiO₂ and high-k dielectrics, respectively. The curves are clearly symmetric, and p-type Tunnel FETs are expected to show all the same trends with length scaling as the n-type devices discussed up until this point in the chapter.

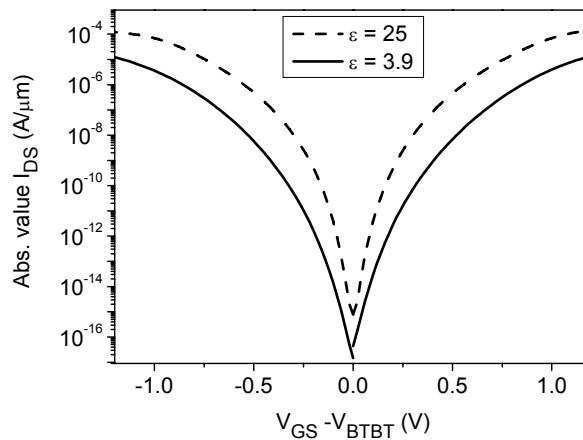


Figure 5.12: Simulated n- and p-type DG Tunnel FETs, showing their symmetric characteristics. $t_{ox} = \text{nm}$, $t_{Si} = 10 \text{ nm}$, $L = 50 \text{ nm}$. $V_{DS} = 1 \text{ V}$ for n-type and $V_{DS} = -1 \text{ V}$ for p-type devices.

In 2009, C. Sandow et al. published some experimental results for p-type Tunnel FETs, which confirm the lack of dependence of on-current on Tunnel FET device length [10]. Fig. 5.13(a) shows a schematic for their single-gate devices fabricated on SOI, and Fig. 5.13(b) presents the saturation current measured at $V_{DS} = -3.5 \text{ V}$ for devices ranging in length from 2 to 12 μm . On a log scale, there is no visible current degradation for long devices, showing that the resistance of the tunnel junction dominates over the comparatively negligible resistance of the intrinsic region in this length range.

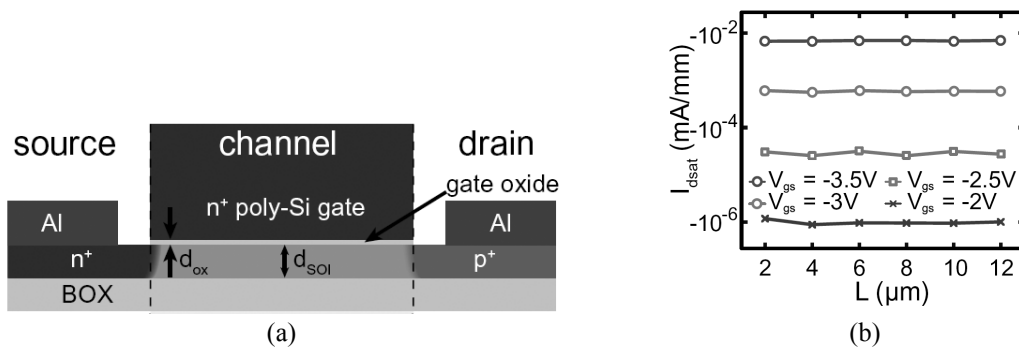


Figure 5.13: (a) Schematic of fabricated p-type Tunnel FET on SOI and (b) its reported characteristics showing the insensitivity of I_{dsat} to intrinsic region length, from [10].

5.7 Scaling supply voltage

Up until now, this scaling study has focused purely on length scaling, in order to isolate and understand its effects and limits. When scaling down conventional MOSFETs, however, V_{DD} scaling is essential, in order to keep dynamic power density constant. Operation at low V_{DD} is also desirable for Tunnel FETs, since one of their most likely applications is in low-power circuits. Since scaling down Tunnel FET length does not make up for the current level lost when the supply voltage is reduced, other techniques for boosting I_{on} must be explored.

Fig. 5.14 shows the effect on the transfer characteristics when V_{DD} is reduced from 1 V to 0.4 V. The simulated Tunnel FET had a relaxed junction abruptness of 20 nm for 4 decades of doping, and its gate dielectric is 3 nm of SiO_2 . In the discussion that follows, V_{DD} is reduced to 0.6 V. In terms of ITRS 2009 [3], this is an extreme reduction, since V_{DD} is only scaled down to 0.75 V in 2021 for low standby power devices, and to 0.6 V in 2021 for low operating power devices.

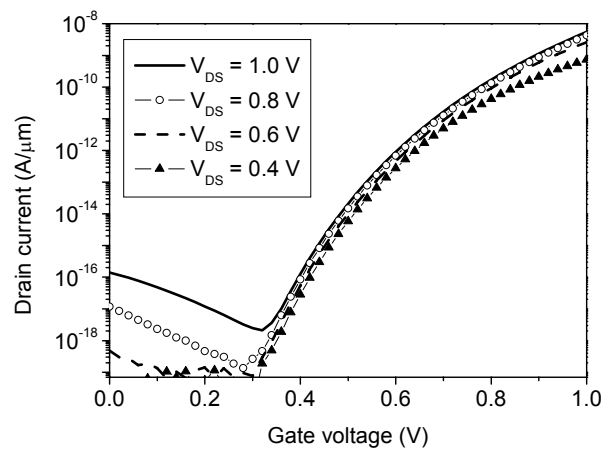


Figure 5.14: I_{DS} - V_{GS} when scaling V_{DD} . Here, V_{DS} is reduced from 1 V to 0.4 V by 0.2 V increments. The resulting loss in I_{on} , subthreshold swing, and V_{TG} needs to be compensated by other parameter adjustments.

There are several possibilities for improving the characteristics that have been degraded by V_{DD} reduction. Two are investigated in Fig. 5.15. One possibility is to improve the doping profile abruptness at the tunnel junction, as in Fig. 5.15(a). With a steeper doping profile, the Tunnel FET turns on at a lower value of V_{GS} , and on-current is increased, but not to the needed level (shown by the solid curve). In addition, the achievable junction abruptness has a physical limit, and so this parameter cannot be exploited past a certain point. A better possibility for recovering the current lost during V_{DD} reduction is to reduce the EOT (effective oxide thickness) of the gate dielectric. This is shown in Fig. 5.15(b). This technique is well-known, since it is already part of conventional MOSFET constant field scaling. With high- k dielectric innovation, the limits of this scaling possibility should not be an immediate obstacle. Fig. 5.15(b) shows that EOT reduction should be able to make up the current lost when reducing V_{DD} , and should give the additional advantages of improving the subthreshold swing and lowering the threshold voltage. Other scaling possibilities, not shown here, could be to reduce the silicon body thickness of a double-gate device, to scale down the diameter of a device built on a nanowire, or to reduce the band gap at the tunnel junction by using another material for the source or by adding tensile strain. All of those ideas were explored during optimization, in Chapter 3.

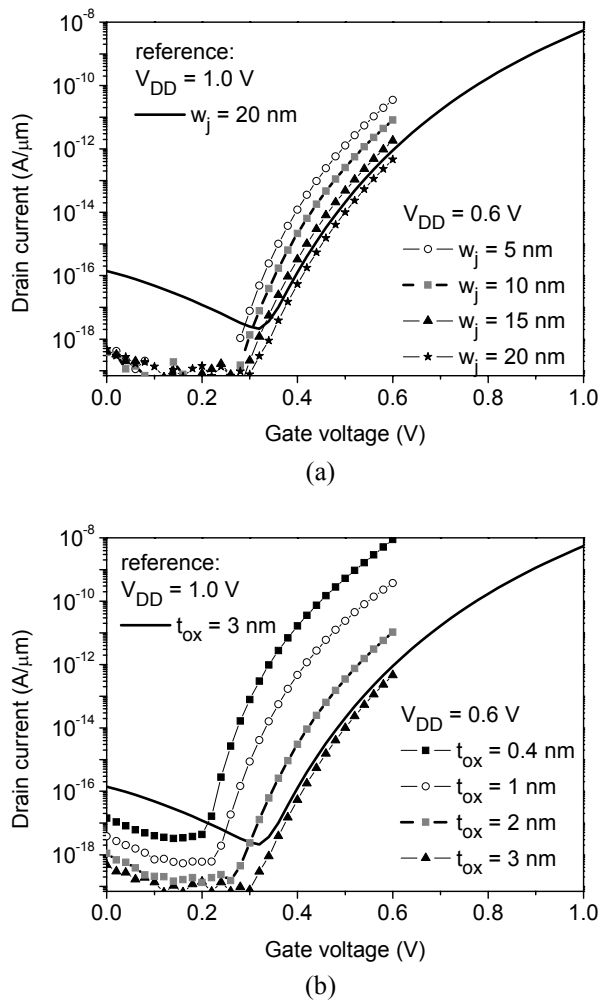


Figure 5.15: Two possibilities for recuperating some of the on-current lost when scaling V_{DD} to 0.6 V. (a) The width of the doping profile at the tunnel junction, expressed in nanometers for four decades of doping change (from 10^{20} to 10^{16} atoms/cm³). (b) The EOT (effective oxide thickness) of the gate dielectric.

5.8 Conclusion

For any new device that could replace or complement existing CMOS technology, a clear understanding of how it scales down is essential. Because the characteristics of a Tunnel FET do not change in the same way as those of a conventional MOSFET when its length is decreased, we changed this parameter alone in order to fully understand the effects. Double-gate Tunnel FETs with a thin silicon body (10 nm) and a high- k dielectric (3 nm, $\epsilon = 25$) should be able to be scaled down to about 20 nm without deleterious effects on their characteristics. A less optimized device, with a thicker body, a larger EOT, or only one gate, will not be able to scale as far before experiencing unacceptably high levels of leakage in subthreshold, and better-optimized devices, with better electrostatic control by the gate of the junction, should be able to scale even closer to Baba's 10-nm predicted limit [9].

The following technical topics and contributions were presented in this chapter:

- *Scaling of Tunnel FETs vs. conventional MOSFETs* (Section 5.1)
This section discussed the way in which conventional MOSFETs are usually scaled in order to keep electric field constant, and why that does not work for Tunnel FETs. It was suggested that changing each parameter separately is better for Tunnel FET studies, in order to understand and quantize the effects.

- *Length scaling and basic device characteristics* (Section 5.3)
The transfer and output characteristics of the simulated Tunnel FETs, ranging in length from 10 nm to 2 μm , were shown, along with their derivatives, g_m and g_{ds} . It was demonstrated that characteristics are insensitive to length variation, with two exceptions:
 - At short gate/intrinsic region lengths less than some critical length L_{crit} , p-i-n leakage starts to dominate the off-state and subthreshold region of the $I_{DS}-V_{GS}$ curves. This happens “sooner” (at longer lengths) for Tunnel FETs with lower gate dielectric permittivity.
 - At long gate/intrinsic region lengths, series resistance from the intrinsic region starts to be of the same magnitude as the resistance of the tunnel junction, and to slightly decrease Tunnel FET current at high V_{GS} .
- *Scaling study in Tunnel FETs with high-k gate stack* (Section 5.4)
Several derived device characteristics were presented in this section: I_{Dmax} , I_{on}/I_{off} , threshold voltage, and subthreshold swing, all shown as a function of gate/intrinsic region length. Along with the two gate dielectrics shown in the previous section: 3 nm of SiO_2 and 3 nm of a high-k dielectric with $\epsilon_{dielectric} = 25$, a third dielectric stack was investigated, with a 1 nm layer of oxynitride ($\epsilon = 5.7$), followed by 2 nm of the high-k dielectric with $\epsilon = 25$. The device characteristics reflected the impact of the two sources of degradation mentioned above, p-i-n leakage in the off-state for short devices, and some small series resistance effect for long devices.
- *Depletion, energy bands, and the off-state* (Section 5.5)
P-i-n leakage in the off-state can be understood in several different ways, two of which have been shown in this section: depletion region width, and energy barrier width/definition. From both these metrics, it is clear that the devices simulated here experience high leakage with $L = 10$ nm, no matter what the gate dielectric, and are not experiencing leakage in the off-state with $L = 40$ nm. The difference between Tunnel FETs with a SiO_2 gate dielectric and a high-k gate dielectric are seen clearly when $L = 20$ nm. In this case, the high-k Tunnel FET still has a wider depletion region, and better-defined energy bands in which the energy barrier is wide enough to prevent unwanted band-to-band tunneling in the off-state. Such is not the case with the SiO_2 Tunnel FET.
- *Length scaling trend confirmation from experimental data* (Section 5.6)
This section presented the only information from another source. Experimental data from Sandow, et al. confirms the length-insensitivity of the on-current for p-type Tunnel FETs on SOI in the length range from 2 μm to 12 μm .
- *Scaling supply voltage* (Section 5.7)
An alternative view of scaling, the reduction of V_{DD} , was explored, showing the drop in Tunnel FET on-current due to the decrease of V_{DS} and V_{GSmax} . Junction abruptness and effective oxide thickness (EOT) were suggested as parameters to scale in order to recuperate the current level needed in the on-state. Junction abruptness does not succeed when V_{DD} is reduced from 1 V to 0.4 V, but EOT does, and is more scalable in the long term.

All results presented in this chapter, with the exception of Section 5.6, represent novel contributions. To the best of my knowledge, no one else has carried out (and published) an in-depth study on length scaling as presented in this chapter. T. Baba [9] made an educated estimation that Tunnel FETs should be able to be scaled down to 10 nm. K. Bhuwalka studied some scaling trends by simulation, always scaling L and t_{ox} together [1],[2] and published experimental transfer characteristics for two device lengths: 25 and 70 nm [8]. C. Sandow [10] and F. Mayer [11] published length scaling data based on experimental results, more than a year after our scaling study [12] was released.

5.9 Bibliography

- [1] K. Bhuiwarka, M. Born, S. Sedlmaier, J. Schulze, and I. Eisele, "Scaling Parameters for Tunnel Field-Effect Transistors," *Conf on Ultimate Integration of Si*, pp. 135-138, 2005.
- [2] K. Bhuiwarka, J. Schulze, and I. Eisele, "Scaling the Vertical Tunnel FET with Tunnel Bandgap Modulation and Gate Workfunction Engineering," *IEEE Trans. Electron Devices*, vol. 52, no. 5, pp. 909-917, May 2005.
- [3] ITRS Roadmap, 2009, available online at <http://www.itrs.net>.
- [4] Y. Taur and T.H. Ning, *Fundamentals of Modern VLSI Devices*. New York, NY: Cambridge University Press, 1998.
- [5] K. Boucart and A. Ionescu, "Double Gate Tunnel FET with ultrathin silicon body and high-k gate dielectric," *Proc. ESSDERC*, pp. 383-386, 2006.
- [6] K. Boucart and A. Ionescu, "Double Gate Tunnel FET with high-K gate dielectric," *IEEE Trans. Electron Devices*, vol. 54, no. 7, pp. 1725-1733, Jul. 2007.
- [7] W. Zhu, J. Han, and T. Ma, "Mobility Measurement and Degradation Mechanisms of MOSFETs Made with Ultrathin High-k Dielectrics," *IEEE Trans. Electron Devices*, vol. 51, no. 1, pp. 98-105, Jan. 2004.
- [8] K. Bhuiwarka, M. Born, M. Schindler, M. Schmidt, T. Sulima, and I. Eisele, "P-Channel Tunnel Field-Effect Transistors down to Sub-50 nm Channel Lengths," *Jpn. J. Appl. Phys.*, vol. 45, no. 4B, pp. 3106-3109, Apr. 2006.
- [9] T. Baba, "Proposal for Surface Tunnel Transistors," *Jpn. J. Appl. Phys.*, vol. 31, no. 4B, pp. L455-L457, Apr. 1992.
- [10] C. Sandow, J. Knoch, C. Urban, Q.-T. Zhao, and S. Mantl, "Impact of electrostatics and doping concentration on the performance of silicon tunnel field-effect transistors," *Solid State Electron.*, vol. 53, pp. 1126-1129, June 2009.
- [11] F. Mayer, C. Le Royer, J.-F. Damlencourt, K. Romanjek, F. Andrieu, C. Tabone, B. Previtali, and S. Deleonibus, "Impact of SOI, Si_{1-x}Ge_xOI and GeOI substrates on CMOS compatible Tunnel FET performance," *IEDM Tech. Dig.*, pp. 163-166, 2008.
- [12] K. Boucart, A. M. Ionescu, "Length scaling of the Double Gate Tunnel FET with a high-k gate dielectric", *Solid-State Electronics*, vol. 51, iss. 11, pp. 1500-1507, Nov. 2007.

Chapter 6

Tunnel FET Parameter Variation

This chapter has three main parts: the simulated optimization of a silicon Tunnel FET by an additive booster technique, a study of parameter fluctuations on the optimized static characteristics, and finally, a comparison with the sensitivity of conventional MOSFETs. First, I_{on} , subthreshold swing, and gate threshold voltage are optimized by the addition of these technology boosters: high-k gate dielectric, more abrupt doping profile at the tunnel junction, thinner silicon body, higher source doping, double gate, gate dielectric aligned to the intrinsic region, and a shorter device length. Then beginning from an optimized device, a parameter variation study is carried out, in which one parameter is varied at a time, and the resulting fluctuations of the device characteristics are analyzed quantitatively and qualitatively. Gate dielectric permittivity and thickness, and doping junction width are pinpointed as the parameters requiring the tightest control during Tunnel FET fabrication in order to limit characteristic fluctuations. Body thickness, gate dielectric alignment with the tunnel junction, and intrinsic region length may also need tight control depending on their target values. When compared with conventional MOSFETs, Tunnel FET characteristics are found to be much more sensitive to variations in body thickness (especially on-current), and in effective oxide thickness (threshold voltage, subthreshold swing, and on-current).

6.1 The importance of a parameter variation study

While the basic parameters to optimize in order to have good Tunnel FET characteristics are known – an abrupt doping profile at the tunnel junction, high capacitive coupling from the gate to the tunnel junction, etc. – it is crucial to also understand the sensitivity of device characteristics to parameter fluctuations. This is well-understood for conventional MOSFETs, and the principal sources of fluctuation, including random discrete dopants, line edge roughness, polysilicon granularity, and oxide thickness fluctuations, have been thoroughly studied [1],[2]. The same needs to be accomplished with Tunnel FETs, so that the main sources of characteristic fluctuation can be pinpointed, and then controlled.

The goals of this chapter are twofold: to optimize the device characteristics of a Tunnel FET using the additive booster technique, and then to investigate the influence of the parameter fluctuations which are unavoidable in device fabrication and their impact on device characteristics. The critical parameters that will need to be the most tightly controlled when fabricating these devices are identified, and the parameter variation is quantified.

All simulations were carried out in Silvaco Atlas version 5.13.16.C with a non-local band-to-band tunneling model.

6.2 Performance boosters for silicon Tunnel FETs

While Tunnel FETs have a low off-current and the potential for a small subthreshold swing, they generally have a lower on-current than conventional MOSFETs. Many solutions have been proposed [3]-[14], but performance boosters are often suggested independently – just adding a high-k dielectric, or just using a thin body (or nanowires), or just decreasing the band gap at the tunnel junction, and only a few are typically applied at one time. These techniques, though they increase on-current and reduce subthreshold swing, only have limited power to improve device characteristics. In order to achieve superior characteristics for Tunnel FETs, it is essential to apply as many boosters as possible to the same device, so that the improvements in device performance are cumulative. Then with the fully optimized device, parameter fluctuations can be investigated in great detail.

Before looking at the simulated results of additive booster technology illustrated in Fig. 6.2 and quantified in Table 1, it is important to go back to a fundamental analytical expression for band-to-band tunneling transmission, to see what can be changed and how. With the WKB approximation and taking the tunnel barrier as a triangularly shaped potential barrier [15], as demonstrated in Chapter 2, the band-to-band tunneling transmission is given by

$$I_{BTB} \propto T_t \approx \exp\left(-\frac{4\lambda\sqrt{2m^*}E_g^{3/2}}{3\hbar(\Delta\Phi + E_g)}\right) \quad (6.1)$$

where $\Delta\Phi$ is the energy range over which tunneling can take place, E_g is the band gap at the tunnel junction, λ is a screening length, and m^* is the tunneling mass. m^* and E_g can be reduced by changing the materials used or by introducing strain at the tunnel junction, thereby increasing tunneling current. The parameter λ , which is a measure of the spatial extent of the electric field, was given for a double-gate device in Eq. 2.13. The general dependencies in this equation (on the thickness and the dielectric constant of both the gate dielectric and the semiconductor material) are correct, as verified by simulation in Figs. 3.9 and 3.15, but a more complete expression for λ could also include its dependence on other fabrication choices such as tunnel junction doping abruptness, source doping (included in an equation for λ in [7]), and gate dielectric alignment. By manipulating device design in order to minimize λ , band-to-band tunneling is significantly improved in the on-state, and good device characteristics are the result.

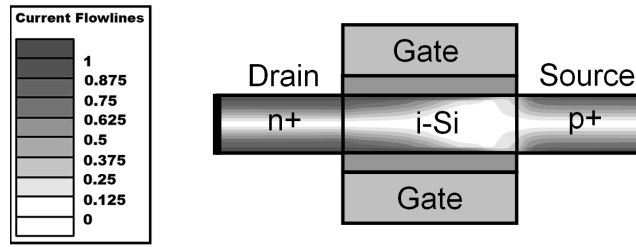


Figure 6.1: Silicon Tunnel FET device structure after the optimizations of Table 1. Current flowlines for $V_{DS} = V_{GS} = 1$ V.

The starting point of our additive booster investigation is a silicon Tunnel FET with an asymmetrically-doped n-i-p structure. The gate dielectric thickness is 3 nm, and the drain and intrinsic region dopings are 5×10^{18} and 10^{15} atoms/cm³, respectively. The other parameters for the base device are shown in Table 1, column 2.

The initial optimization was carried out in the following way. Starting with the base device in the “Before” column of Table 1, each parameter was changed, one at a time, to a level compatible with state-of-the-art CMOS fabrication. The goal was to optimize the parameters in such a way that they would be greatly improved over the base device levels, realistically attainable with modern fabrication processes, and would still have a bit of room for improvement so that the parameter could be varied both up and down during the parameter variation study. For example, the gate dielectric was SiO₂ in the base device, and was improved to the widely-used high-k dielectric HfO₂, which can have a dielectric constant of about 25. HfO₂ is a state-of-the-art material for CMOS fabrication today, but there are high-k dielectrics with even larger permittivities, so it is not the ultimate limit, and leaves some room for variation.

Table 1: Tunnel FET parameters before and after optimization.

Parameter	Before	After optimization
Gate dielectric ϵ	3.9	25
Junction width	47 nm / 5 decades	12 nm / 5 decades
Body thickness	50 nm	10 nm
Source doping	8×10^{19} atoms/cm ³	1.5×10^{20} atoms/cm ³
Gates	Single	Double
Oxide alignment	Over all	Over intrinsic
Device/gate length	90 nm	30 nm

After each parameter had been improved individually, the simulation results were analyzed and the parameters were ranked, from those that gave the most to the least improvement in Tunnel FET on-current. Then, beginning with the base device once again, the parameter optimizations were carried out one at a time in the same order. The first seven boosters of Fig. 6.2 applied to the base device were as follows: high-k gate dielectric, more abrupt doping profile at the tunnel junction, thinner device body, higher source doping, double gate, gate oxide aligned to the intrinsic region, and finally, shorter intrinsic region (and gate) length. Curve H in Fig. 6.2 shows the optimized characteristics for an all-silicon Tunnel FET.

When the boosters were applied in this order, the resulting improvements are roughly the same as those that they gave when applied one at a time to the base device, with one exception that stands out from Fig. 6.2. Looking only at curves A through H, there is generally a decreasing amount of improvement given by each consecutive booster, except for curve D, where the silicon body has been thinned from 50 nm to 10 nm. Here, the improvement was less drastic than when applied alone to the

base device, since some optimizations had already been carried out, and they changed the value of λ before the body thinning was applied. This implies that λ may not have a straightforward analytical expression that incorporates all of the parameters added in curves B through H of Fig. 6.2, but that instead, the parameters might have complicated inter-dependencies.

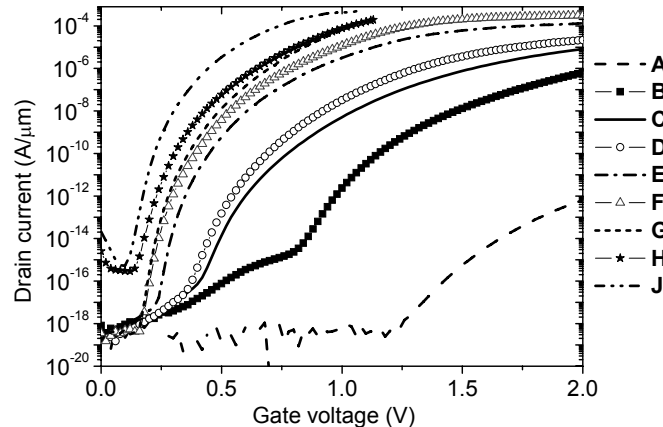


Figure 6.2: Additive boosters. A: base device (“Before” column of Table 1). B: Like A with high- k dielectric. C: Like B with narrower junction. D: Like C with thinner body. E: Like D with higher source doping. F: Like E with double gate. G: Like F with oxide only over intrinsic region. H: Like G with shorter length. J: Like H with band gap = 0.8 eV at the tunnel junction (which could correspond to strained Si [14], or a SiGe source [3],[6],[13]). $V_{DS} = 1$ V.

The last booster, curve J in Fig. 6.2, consists of a smaller band gap applied at the tunnel junction, which can represent an improvement resulting from a lateral strain profile within the device [14] or a heterostructure in which a lower band gap material is used for the Tunnel FET source [3],[6],[13].

The unoptimized device (curve A in Fig. 6.2) shows poor performance, having a low on-current even at $V_{GS} = 2$ V, but Tunnel FETs have a huge potential when designed well. It is clear from Fig. 6.2 that several device parameters have a very strong influence on Tunnel FET on-current, threshold voltage, and subthreshold swing. As boosters are applied, V_{TG} is clearly shifted to lower gate voltages, and accompanying this is a significant increase in on-current and a decrease in swing. While the unoptimized Tunnel FET in curve A has a point swing of 69.5 mV/dec (and the average swing from turn-on to threshold can’t be calculated because it never reaches the threshold current of 10^{-7} A/ μm), the optimized device in curve H has a point swing of 22.9 mV/dec and an average swing of 59.7 mV/dec. The device with a reduced band gap of 0.8 eV at the tunnel junction in curve J has further improved swings of 18.1 mV/dec (point) and 40.7 mV/dec (average). Off-state leakage is higher for the devices of curves H and J due to the shorter intrinsic region and gate length, discussed more later in section 6.3.5, but is still in the fA/ μm range. This gives a static power reduction on the order of 10^4 times compared with today’s advanced CMOS nodes.

From the viewpoint of device optimization, it is beneficial that these devices are very responsive to changes in certain parameters. But seen from the viewpoint of manufacturability, it can be extremely disadvantageous, or at least challenging, to work with a device that is so sensitive to parameter fluctuations.

6.3 Parameter fluctuation study

Since the physics of Tunnel FETs, governed by the band-to-band tunneling rate, is different from that of conventional MOSFETs, the sensitivity of device characteristics to parameter fluctuations is expected to be different as well, and to potentially raise new technology challenges. In the following sections, the optimized all-silicon Tunnel FET (right column of Table 1, curve H of Fig. 6.2) is the starting point, and we investigate its characteristics’ sensitivity to the variation of the following parameters: dielectric permittivity and thickness, doping profile width at the tunnel junction, body thickness, gate contact and dielectric alignment, gate / intrinsic region length, band gap at the tunnel

junction, and intrinsic region doping. As already mentioned in section 6.2 in relation to curve D in Fig. 6.2, the fluctuation of Tunnel FET characteristics will depend somewhat on how optimized the device is. For this reason, a fully optimized Tunnel FET is used for this parameter variation study, since only such a device would be incorporated into a large-scale production effort in which characteristic fluctuations become important, and their control necessary.

6.3.1 Dielectric permittivity and thickness

Capacitive coupling between the gate and channel is a critical parameter for highly-scaled conventional MOSFETs, so oxide thickness variations have been a major focus of studies on parameter fluctuations for this type of device [1]. Tunnel FETs are even more sensitive to changes in gate capacitance than conventional MOSFETs (see Fig. 3.8), thus it is no surprise that changes in dielectric permittivity and thickness can cause major fluctuations in Tunnel FET characteristics.

The I_{DS} - V_{GS} of an optimized Tunnel FET with varying gate dielectric permittivity is shown in Fig. 6.3. The physical gate dielectric thickness is always 3 nm unless specified otherwise. On-current and subthreshold swing are clearly both very sensitive to changes in the gate dielectric constant. In Fig. 6.4, $\epsilon = 25$ is taken as the target, and the percent change in V_{TG} (gate threshold voltage), S_{point} (the lowest subthreshold swing value on the I_{DS} - V_{GS} curve), and S_{avg} (the average swing from turn-on to threshold at 10^{-7} A/ μ m) are shown on the left axis. The change in I_{on} (at $V_{DS} = V_{GS} = 1$ V) is easier to see on an exponential scale, so it is plotted as $\log(I_{on}/I_{on,\epsilon=25})$ on the right axis of Fig. 6.4. The variation of the permittivity has a drastic influence on all characteristics, especially when changing from a relatively low permittivity such as that of SiO₂, to high-k. For example, when changing from SiO₂ to HfO₂ ($\epsilon \sim 25$), threshold voltage and subthreshold swing approximately cut in half, and the on-current goes up by three orders of magnitude.

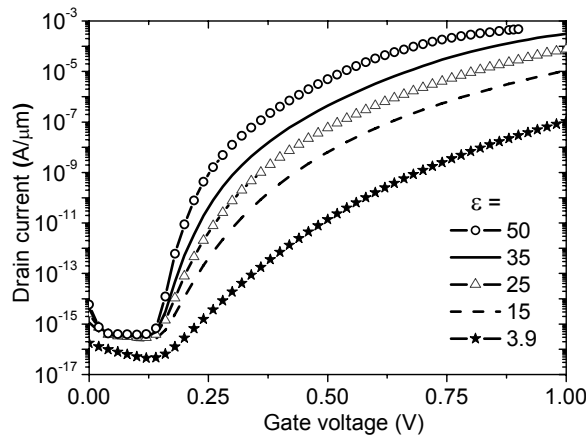


Figure 6.3: I_{DS} - V_{GS} for different gate dielectric permittivities. $V_{DS} = 1$ V.

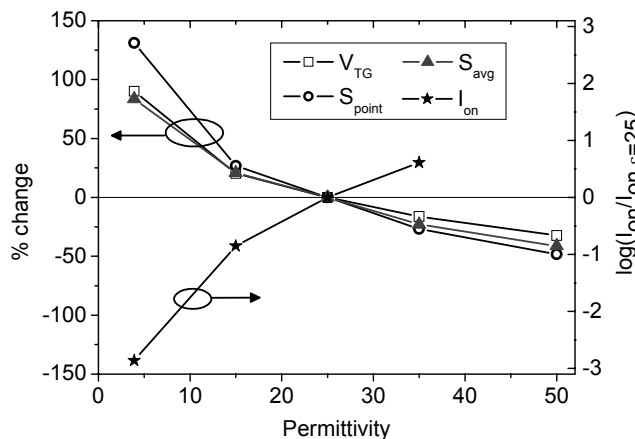


Figure 6.4: (Left) % change in S_{point} , S_{avg} , and V_{TG} and (right) orders of magnitude by which I_{on} changes when gate dielectric permittivity changes. I_{on} is the most sensitive characteristic. $V_{DS} = 1$ V.

The variations presented in Fig. 6.4 can be applied to complicated gate stacks with more than one gate dielectric material, by calculating an effective permittivity for the stack. Since it is sometimes more useful to look at fluctuations of the gate stack expressed as variations of an effective thickness, the sensitivity of I_{on} and V_{TG} to t_{ox} variations is shown in Fig. 6.5. The gate dielectric constant is 25. Although it may look as if the fluctuations become less extreme with increasing t_{ox} , this is because we are looking at units that have not been normalized to the I_{on} or the V_{TG} . If we look at fluctuations as percentages, they are significant everywhere in the studied range of dielectric thickness.

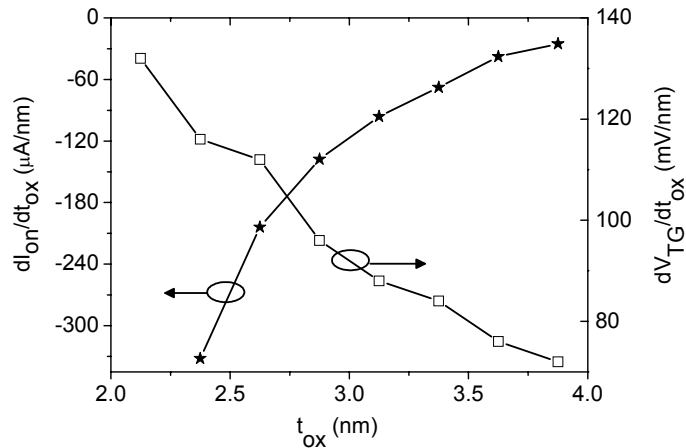


Figure 6.5: I_{on} and V_{TG} sensitivity as a function of t_{ox} . Gate dielectric permittivity = 25. $V_{DS} = 1$ V.

6.3.2 Junction width

Junction width (w_j), or the distance across which the doping falls from its high level in the source to its low level in the intrinsic region, is an important factor that determines the steepness of the energy bands at the tunnel junction, and therefore the minimum possible tunnel barrier width in a Tunnel FET in the on-state.

Studied parameter range: The range of junction widths investigated here was from nearly perfectly abrupt, 0.5 nm for 5 decades of doping, which is 0.1 nm/decade, to 30 nm for 5 decades of doping, which is 6 nm/decade. The state-of-the-art for junction abruptness falls within this range. Surdeanu et al. [16] used implantation and laser annealing to create junctions with an abruptness of 1.8 nm/decade for Sb and 4.5 nm/decade for B (see Fig. 6.6(a)). Nguyen et al. [17] employed vapor phase doping with laser annealing, and attained a junction abruptness of 3 nm/decade for As, and about 4 nm/decade for B. The ITRS requirements for high-performance logic will require a junction depth of 7.3 nm in 2015, and if we assume that junction to be over approximately two decades of doping, that gives a required abruptness of about 3.6 nm/decade. As a comparison, our simulated 21-nm junction, used as the target w_j below, has an abruptness of 4.2 nm/decade, shown in Fig. 6.6(b).

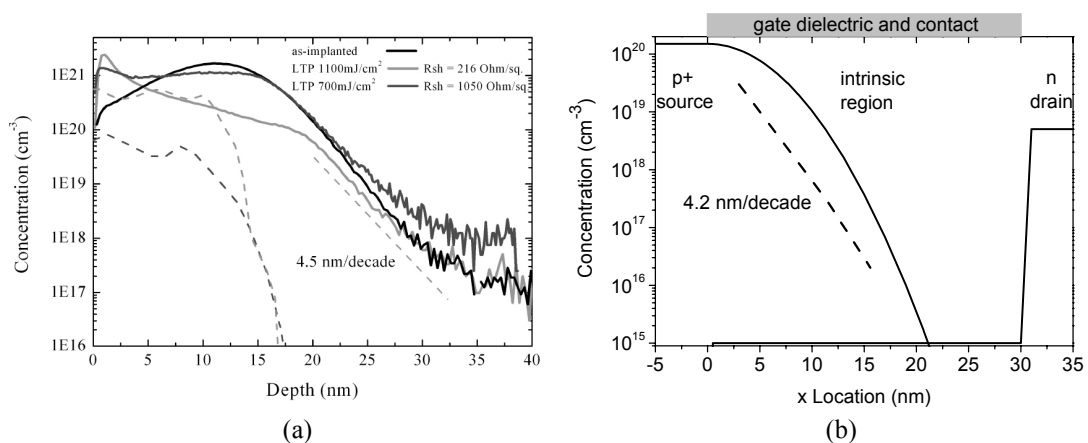


Figure 6.6: (a) Fabricated junctions created by the implantation and laser annealing of boron, from [16]. (b) Our p-type source-intrinsic region junction, $w_j = 21$ nm, gate alignment shown.

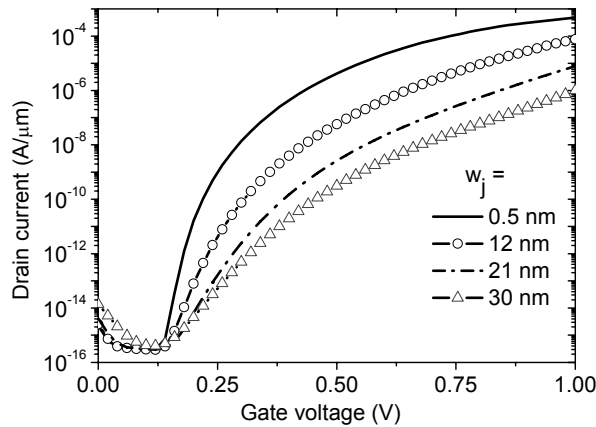


Figure 6.7: I_{DS} - V_{GS} for varying junction widths, over 5 decades of doping from the source at 1.5×10^{20} atoms/cm³ to the intrinsic region at 10^{15} atoms/cm³. 0.5 nm is the mesh spacing, and thus corresponds to the most abrupt junction that can be simulated. $V_{DS} = 1$ V.

The dependence of I_{DS} - V_{GS} on junction abruptness is seen in Fig. 6.7. In Fig. 6.8(a), $w_j = 21$ nm is the reference point, and all four characteristics show sensitivity to fluctuations, with no flat areas on the curves to suggest a better target. Fig. 6.8(b) shows the sensitivity of V_{TG} and I_{on} to junction width. Note that the values here are not normalized, so what seems to be a smaller I_{on} sensitivity is related to the fact that I_{on} is much lower with an increasing junction width. Fig. 6.9 allows us to extract the standard deviation (σ) of V_{TG} and S_{avg} when the σ of w_j is known for a particular fabrication process.

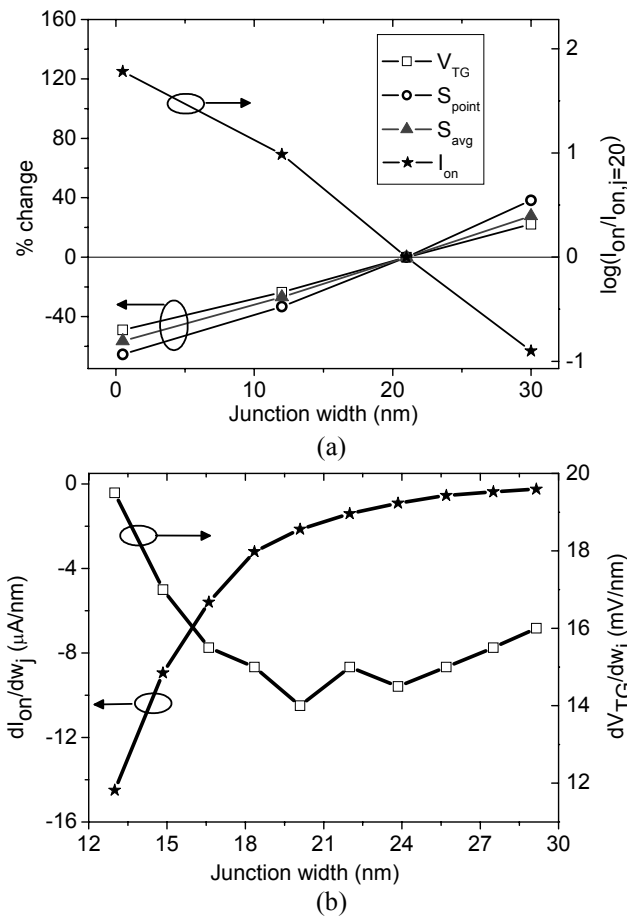


Figure 6.8: (a) (Left) % change in S_{point} , S_{avg} , and V_{TG} and (right) orders of magnitude by which I_{on} changes when w_j changes. Again I_{on} is more sensitive than the other characteristics. (b) I_{on} and V_{TG} sensitivity as a function of junction width w_j . $V_{DS} = 1$ V.

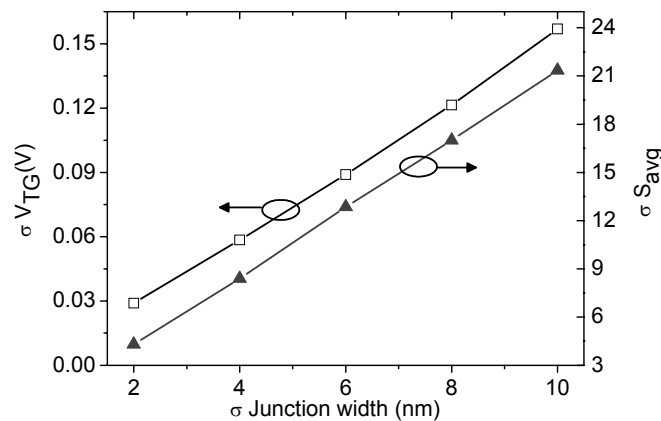


Figure 6.9: Standard deviation of threshold voltage (σV_{TG} on left axis) and average subthreshold swing (σS_{avg} on right axis) vs. standard deviation of junction width for a Tunnel FET whose target junction width is 21 nm for 5 decades of doping difference. $V_{DS} = 1$ V.

6.3.3 Silicon body thickness

Device thickness is a parameter that is becoming more important as more devices are built on thin films and incorporate double or multi gates. For achieving ultimate performance, the device thickness is a crucial parameter because it leads to better gate control and the reduction of bulk capacitive effects.

Studied parameter range: The range of silicon body thickness studied here is from 5 to 50 nm. Comparing with the state-of-the-art that has been published for ultra-thin body (UTB) MOSFETs, this is a reasonable thickness bracket. In 2002, Uchida et al. investigated devices that were from 2.3 to 8 nm thick, and estimated that the thickness fluctuation was around four atomic layers [19]. In 2007, Shimizu et al. looked at the mobility enhancement in UTB single-gate and double-gate MOSFETs with an SOI thickness as low as 1.8 nm [20], and Gomez et al. studied the electron transport in MOSFETs made with strained Si directly on insulator where the thinnest device had a silicon thickness of 2 nm with a standard deviation of 0.5 nm [21]. In 2009, Schmidt et al. fabricated UTB MOSFETs ranging from 0.9 nm to 25 nm thick [22]. The thinnest device was fully functional, and only four atomic layers thick. In terms of ITRS requirements, the UTB silicon layer thickness in 2013 should be 7 nm for high-performance logic [18].

In Chapter 3 during optimization, it was shown that for a Tunnel FET whose gate dielectric covers the source and drain regions as well as the intrinsic region, there is an optimal body thickness for a double-gate device. As body thickness becomes thinner, first characteristics improve and on-current increases, and then a peak is reached, after which on-current decreases for thinner silicon body layers. The same effect is not seen when the gate dielectric is aligned to the intrinsic region, as shown by the $I_{DS}-V_{GS}$ curves in Fig. 6.10, at least not within the range of gate voltages and body thicknesses studied here. With an aligned gate dielectric, the improved electrostatics lead to an increase of on-current, and lower subthreshold swing and threshold voltage, when the silicon body is made thinner. It is possible that for the curve in Fig. 6.10 for the device with $t_{Si} = 5$ nm, that at higher V_{GS} around 1.25 V, the I-V curve might flatten out enough that it would cross the curves for thicker devices, similar to what was seen in Chapter 3 for Tunnel FETs with a gate dielectric everywhere. Unfortunately, due to Silvaco Atlas convergence limitations, no quantitative statement can be made.

One very important trend shown in Fig. 6.10 is that for a body thickness less than 20 nm, the characteristics improve with decreasing t_{Si} , while for a body thickness of 20 nm or more, characteristics are all nearly the same, with the exception of some small differences in the off-region. This means that the choice of an optimal Tunnel FET body thickness necessitates a trade-off, as shown in Fig. 6.11. Here, 10 nm is chosen as the target thickness, but the characteristics are very sensitive around this target point, especially toward thinner layers. Although swing, V_{TG} , and I_{on} can be greatly improved by decreasing body thickness (or nanowire diameter), there is the drawback

of an increase in the fluctuation of characteristics smaller than some critical body thickness, here 20 nm. In order to fully take advantage of double-gate control of the device, the body thickness should be under this limit, so tight control of this parameter will be necessary in order to minimize the fluctuations in device behavior.

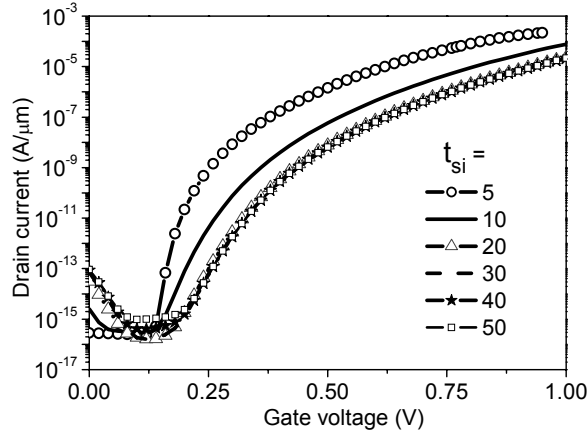


Figure 6.10: I_{DS} - V_{GS} showing that for a double-gate Tunnel FET with the gate dielectric aligned to the intrinsic region, on-current, gate threshold voltage, and subthreshold swing all improve with decreasing body thickness t_{Si} . $V_{DS} = 1$ V.

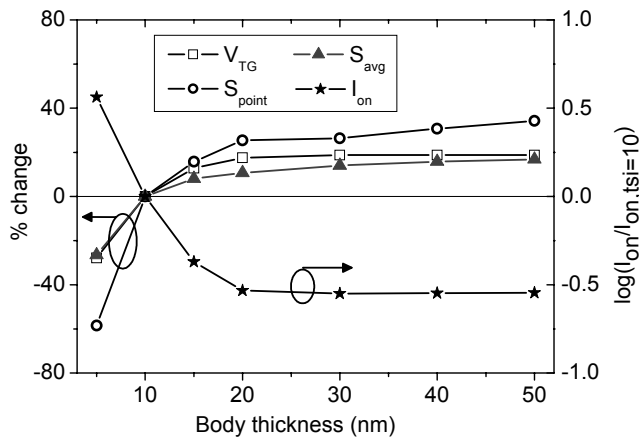


Figure 6.11: (Left) % change in S_{point} , S_{avg} , and V_{TG} and (right) orders of magnitude by which I_{on} changes when body thickness changes. The flat regions suggest a design space with minimized fluctuation, but all characteristics improve at smaller t_{Si} where fluctuation is higher. $V_{DS} = 1$ V.

6.3.4 Gate oxide and gate contact alignment

Gate alignment is another important booster, and there have been recommendations to align the gate dielectric with the tunnel junction in order to take advantage of fringing [12], and to shorten the gate on the drain end in order to increase device speed [22]. It is important to understand how such design choices would influence the magnitude of characteristic fluctuations. Here, we present a detailed investigation of gate alignment, including underlap and overlap.

Fig. 6.12(a) shows a device schematic to clarify how gate alignment is being changed in this section – the gate contact and the oxide alignment change together, with negative alignment values indicating underlap, and positive alignment values indicating overlap. It is important to understand our definition of alignment for the source-intrinsic region junction. Similar to what other groups have shown [7],[12], alignment = 0 is defined as that which results from a self-aligned process, in which there is dopant diffusion under the gate. The actual electrical junction (where the n-doping curve and the p-doping curve cross) is 12 nm into the intrinsic region.

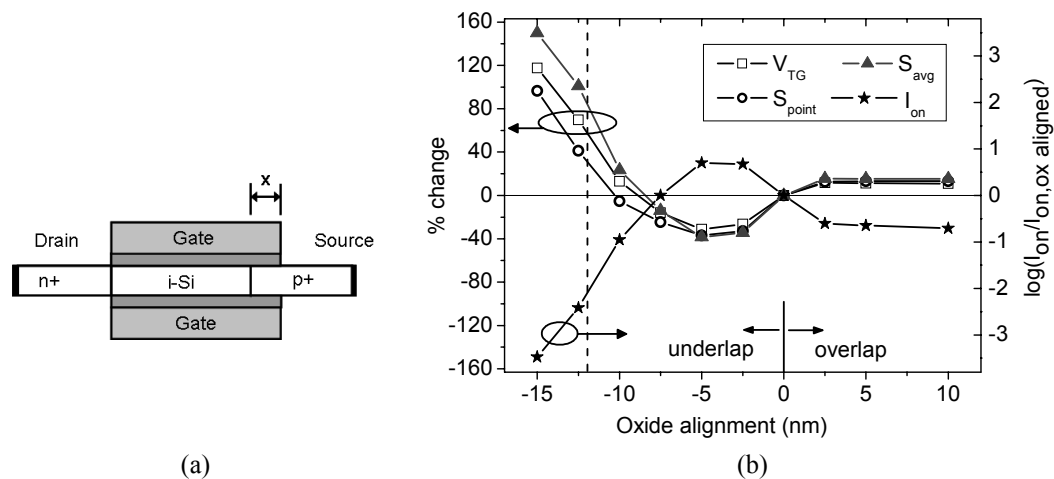


Figure 6.12: (a) Device schematic showing the oxide and gate contact alignment, x , with the tunnel junction. (b) (Left) % change in S_{point} , S_{avg} , and V_{TG} and (right) orders of magnitude by which I_{on} changes when gate alignment changes. The vertical dashed line at $x = -12$ nm shows the location of the n/p electrical junction. An improvement in device characteristics is seen for a slight gate underlap (around 5 nm), and then the characteristics worsen quickly for increased underlap. A region of stable characteristics is seen for an overlap of about 2.5 nm or more. $V_{DS} = 1$ V.

Fig. 6.12(b) shows that the alignment of the gate oxide has a large influence on characteristics, and one that is not necessarily intuitive at first glance. For a gate that is overlapped by several nanometers or more, characteristics are stable. This is the case even beyond what is shown in Fig. 6.12, with only a slight performance degradation for a gate that covers the source region entirely (100 nm for the device shown here). As the gate alignment moves in the underlapped direction, toward the electrical junction marked by the dashed line, the characteristics hit a peak in which the on-current at $V_{DS} = V_{GS} = 1$ V goes up by a factor of 4.5, and the threshold voltage and subthreshold swing hit a low point. This surprising improvement is illustrated by the I_{DS} - V_{GS} curves in Fig. 6.13. While the curve for a gate underlap of 10 nm has the characteristics that could be expected for a gate that no longer has good control of the tunnel junction, the curve for a gate underlap of 5 nm shows improvements in on-current, threshold voltage, and swing. The aligned gate (alignment = 0) has the highest current at gate voltages higher than those shown in Fig. 6.13, but the 5-nm-underlapped gate has better characteristics in the important low-voltage region, critical for low-power devices.

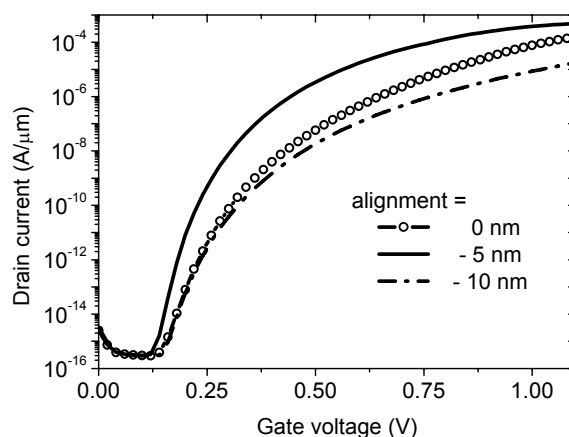


Figure 6.13: I_{DS} - V_{GS} for changing gate dielectric and contact alignment as illustrated in Fig. 6.12(a). When the gate edge is underlapped by 5 nm from the standard self-aligned location, an improvement in swing and on-current are seen. At higher gate voltages, the aligned gate has a higher current. $V_{DS} = 1$ V.

The 2D cross-sections in Fig. 6.14 show the potential distributions in the on-state at $V_{DS} = V_{GS} = 1$ V, for the region under the gate edge at the tunnel junction, for devices with gate alignments of (b) -15

nm, (c) -10 nm, (d) -5 nm, (e) 0, and (f) 5 nm. Cross-sections only show half of the 10 nm body thickness, since the distributions are identical (mirrored) from the center of the body down to the bottom gate. The important thing to see in these figures is the width of the potential gradient. It is clear that with an alignment of -5 nm (Fig. 6.14(d)), the potential gradient is compressed near the dielectric/silicon surface, which reflects the steeper energy bands in this case, leading to a reduced tunnel barrier width and therefore higher I_{on} , smaller subthreshold swing, and reduced gate threshold voltage. In Fig. 6.14(b), on the other hand, the potential gradient is spread out, indicating that the energy bands are not as steep and the tunnel barrier is not as narrow as in (d). At the other extreme, Fig. 6.14(f) shows that the potential gradient is spread out and even crosses into the p+ region due to gate overlap. Again for this case, the energy bands are less steep than for (d), and the energy barrier is wider, leading to lower current.

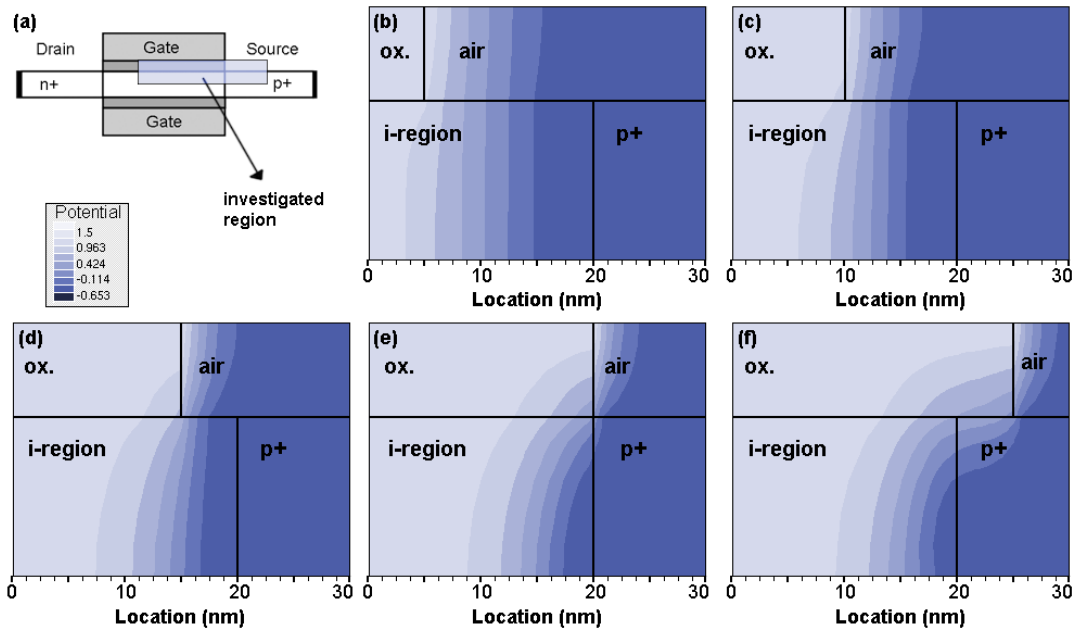


Figure 6.14: Potential distribution cross sections of Tunnel FETs in the on-state ($V_{DS} = V_{GS} = 1$ V) when the gate contact and oxide alignment are changed together. (a) Device schematic with studied cross-section indicated. Only half of the silicon body thickness is shown because the distributions are always perfectly symmetric down to the bottom gate. In order, the cross-sections show a (b) 15 nm gate underlap, (c) 10 nm gate underlap, (d) 5 nm gate underlap, (e) aligned gate, and (f) 5 nm gate overlap. The potential change is the most “squeezed” in the third cross-section (5 nm gate underlap), which means that the electric field is the highest, the energy bands are the steepest, and the tunneling energy barrier width is the narrowest in this case.

Fig. 6.15 shows the total current density qualitatively, in the same area marked in Fig. 6.14(a), again in the on-state. Fig. 6.15 (a) to (c) corresponds to Fig. 6.14 (c) to (e). Several interesting features are visible in these diagrams. First, we see the energy barrier where carriers are forbidden as a dark zone with no current. Next, the energy barrier width is clearly narrower in Fig. 6.15(b), where the gate underlap is 5 nm. This is consistent with the potential cross-sections, and with the higher current level. Finally, the energy barrier has a minimum width a bit less than 1 nm from the dielectric surface in all three devices shown, and then extends straight down to the center of the device when the gate is underlapped by 10 or 5 nm, but not for the aligned case where the dark zone shifts to the left under the gate toward the center of the body ($y = 0$ here). The potential gradient also shifted to the left under the gate for the aligned case, seen in Fig. 6.14(e), whereas the potential gradients for underlapped gates in Figs. 6.14(b), (c) and (d) extended more vertically to the center of the device body. Keep in mind that the actual electrical junction is located 12 nm into the intrinsic region, which would be at location $x = 8$ nm in these cross-sections.

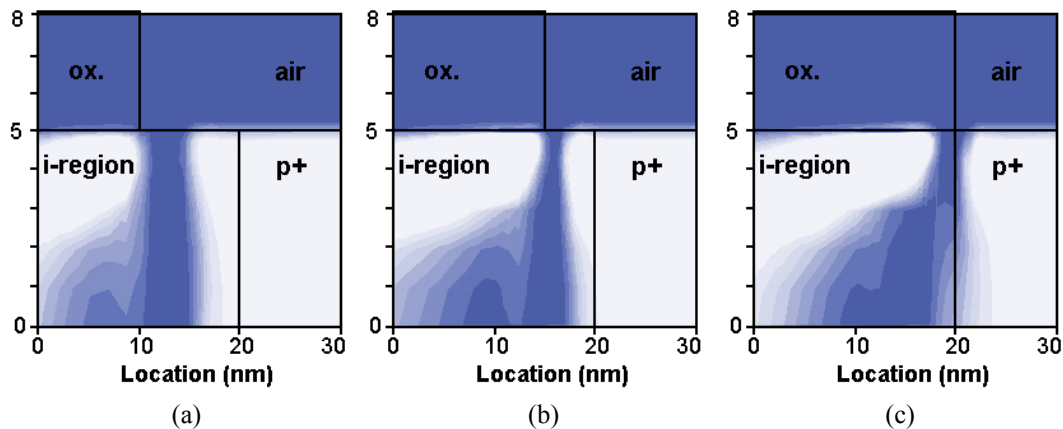
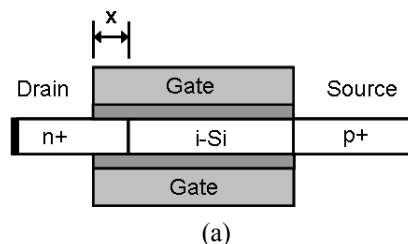


Figure 6.15: Total current density cross-sections for Tunnel FETs in the on-state ($V_{DS} = V_{GS} = 1$ V) for gate alignments (oxide and contact) of (a) -10 nm, (b) -5 nm, and (c) 0 nm. The device shown in (b) has the highest on-current, and the lowest V_{TG} and S values. Dark regions have no current flow. The lightest region corresponds to a current of (a) 8×10^4 A/cm², (b) 3.5×10^6 A/cm², and (c) 7×10^5 A/cm². Light regions indicating current flow inside the dielectric or air are visual artefacts resulting from interpolation between mesh points.

A second case of gate misalignment was studied, in which the gate contact remained aligned with the intrinsic region, and the gate dielectric was underlapped and overlapped. The results were extremely similar to those shown in Figs. 6.12 through 6.15, indicating that the placement of the gate contact is not critical, but the placement of the gate dielectric is.

Now looking at the other end of the gate, it is crucial to understand the importance of the alignment of the gate contact and dielectric to the drain junction, while leaving the gate aligned to the tunnel junction on the source side, as illustrated in Fig. 6.16(a). Verhulst et al. looked at devices with an intrinsic region length of 100 nm and showed characteristics that were nearly unchanged for currents $< 10^5$ A/ μ m when the gate was made smaller on the drain side, apart from a beneficial reduction in I_{off} [22]. Similar simulation results are shown in Fig. 6.16(b) for $L_{intrinsic} = 100$ nm, both with and without a shortened gate (10 nm underlap). For devices near the leakage limit, however, the effect of gate-shortening is different. Fig. 6.16(c) shows $I_{DS} - V_{DS}$ for optimized Tunnel FETs with $L_{intrinsic} = 30$ nm, and leakage has increased for the shortened gate.

The bands in Fig. 6.16(d) and (e) explain the leakage in a short device with a gate that is underlapped by 10 nm. For longer devices ($L_{intrinsic} = 100$ nm), the intrinsic region bands are well-defined and the energy barrier is too wide everywhere for significant tunneling to take place. For shorter devices ($L_{intrinsic} = 30$ nm), the intrinsic region bands are no longer well defined. The reduced gate length shortens the effective length of the intrinsic region, and leads to more p-i-n diode leakage in the off-state. It is worth mentioning, however, that the effective intrinsic region length is not equal to the new shortened gate length, as can be shown by a quantitative example. When $L_{intrinsic} = L_{gate} = 30$ nm, I_{off} is around 10^{-15} A/ μ m. For $L_{intrinsic} = L_{gate} = 10$ nm, I_{off} is about 10^{-8} A/ μ m. In the case of a shortened gate, when $L_{intrinsic} = 30$ nm and $L_{gate} = 10$ nm, I_{off} is approximately 10^{-13} A/ μ m.



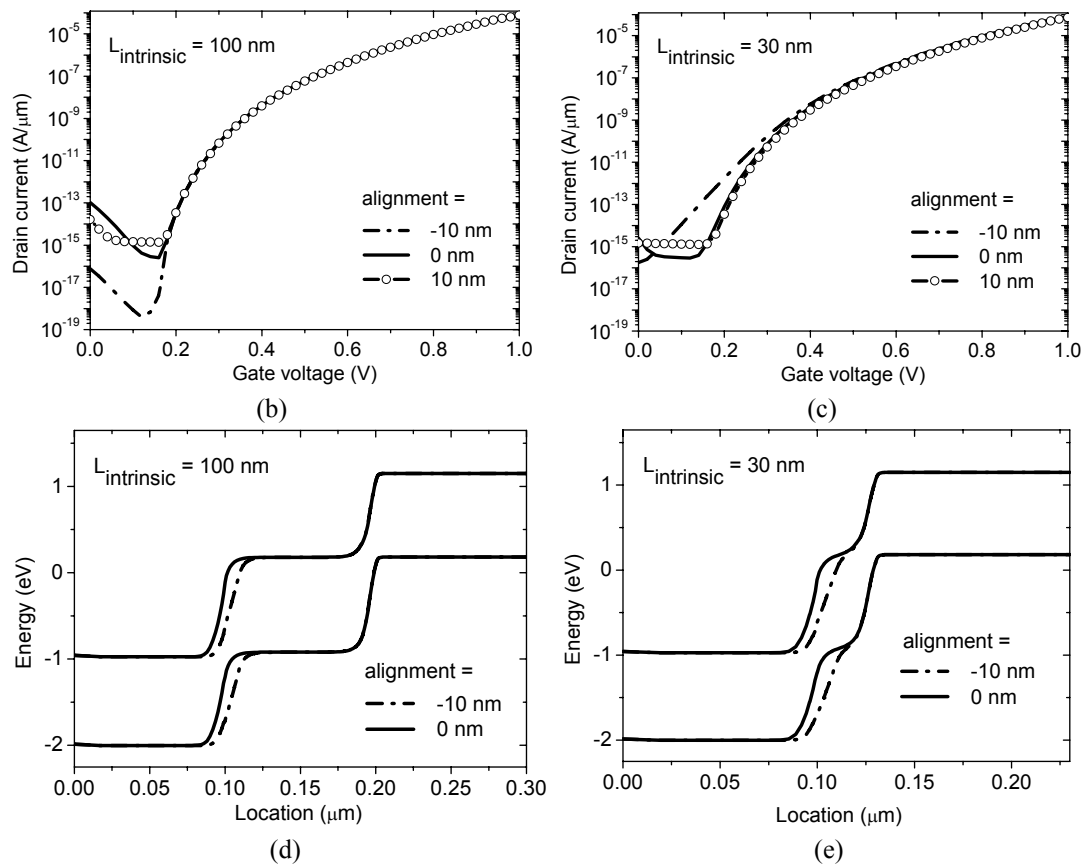


Figure 6.16: (a) Device schematic showing the drain-side gate alignment, marked x in the figure. (b), (c) I_{DS} - V_{GS} and (d), (e) energy bands for an underlap of 10 nm, and a perfect alignment of the gate at the drain side. Leakage results when gate is shortened by 10 nm on the drain side for devices with $L = 30$ nm. For the band diagrams, $V_{DS} = 1$ V, $V_{GS} = 0.2$ V.

6.3.5 Device length

Device length scaling is a key factor driving conventional MOSFET progress, and is carried out following a set of predictable guidelines. For Tunnel FETs, due to the different underlying physics, intrinsic region and gate scaling can be carried out without accompanying changes in other dimensions (as discussed in detail in Chapter 5, and in [23]). In terms of characteristic fluctuations, the choice of Tunnel FET i-region / gate length requires another trade-off decision. As Fig. 6.17 shows, Tunnel FET scaling is limited by p-i-n leakage to a critical length (L_{crit}) of about 25 nm for our optimized devices.

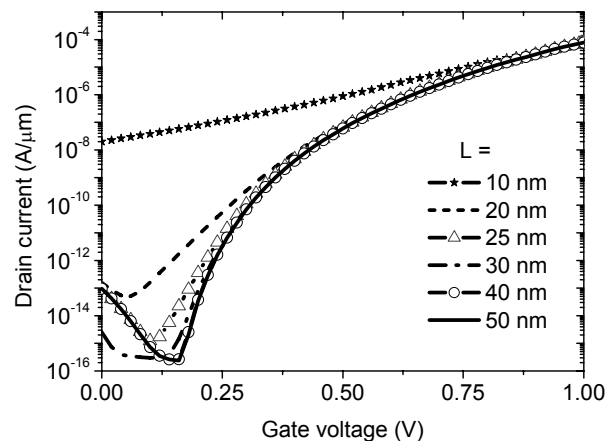


Figure 6.17: I_{DS} - V_{GS} for changing intrinsic region (gate) lengths. Off-current is minimal down to $L = 30$ nm and then increases due to p-i-n leakage. $V_{DS} = 1$ V.

Although shorter devices are attractive for reasons of cost and speed, Fig. 6.18 reveals the downside of designing Tunnel FETs too close to the leakage limit: unacceptably high fluctuation of V_{TG} , S_{point} , and S_{avg} . The standard deviation of the line width, line edge roughness, and other fabrication process-specific sources of variation need to be kept in mind when choosing how close to place the target Tunnel FET length to L_{crit} . Better gate control (decreasing λ in Eq. 1) could decrease L_{crit} slightly, for example by increasing gate capacitance, decreasing body thickness, or using nanowires.

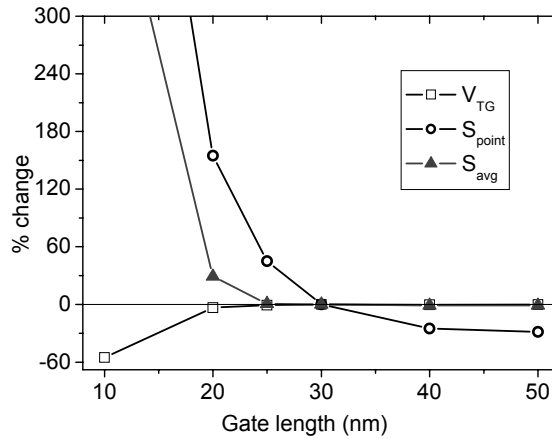


Figure 6.18: % change in S_{point} , S_{avg} , and V_{TG} when gate length changes. The characteristics show low sensitivity to length fluctuations for $L > 25$ nm. $V_{DS} = 1$ V.

6.3.6 Band gap reduction at the tunnel junction

Band gap reduction at the tunnel junction has been proposed to improve on-currents in Tunnel FETs. In principle, there are two primary possibilities: a heterojunction at the source side of the device [3],[6],[13], or a lateral strain profile with a strain maximum at the tunnel junction [14]. This last performance booster (curve J in Fig. 6.2) can be added to the optimized Tunnel FET described in the right column of Table 1: a smaller band gap of 0.8 eV at the tunnel junction, while keeping the relatively large silicon band gap at the drain-side junction to suppress leakage and maintain low off-current. Fig. 6.19 shows the impact of the variation of this parameter.

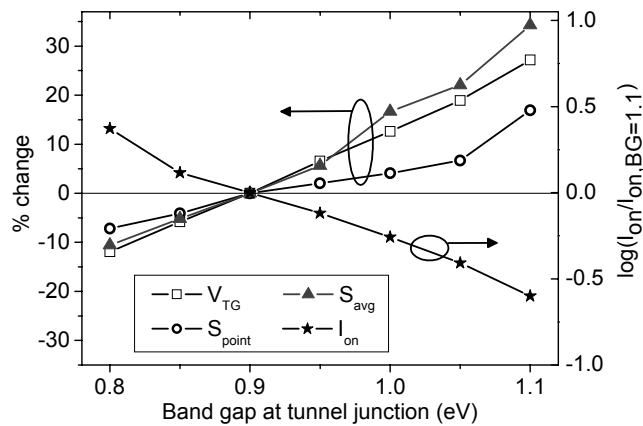


Figure 6.19: (Left) % change in S_{point} , S_{avg} , and V_{TG} and (right) orders of magnitude by which I_{on} changes when band gap at the intrinsic / source junction changes. $V_{DS} = 1$ V.

It is also important to understand the effect of changes in band gap on Tunnel FET characteristics for devices built on extremely scaled nanowires, since the band gap of silicon will start to increase when nanowires are scaled to diameters less than about 5 nm [24]. The increased band gap will cause a reduction in on-current which could potentially cancel out the beneficial effects of further diameter decrease. This point needs more investigation.

6.3.7 Intrinsic region doping

In contrast to silicon CMOS where intrinsic region doping is a critical parameter, the fluctuation of this quantity does not influence Tunnel FET characteristics. This is reported in Table 2, where a variation of \pm one decade of body doping has no impact on the electrically relevant parameters. This is an advantage of Tunnel FETs with a low-doped or intrinsic region under the gate over conventional CMOS, since random discrete dopants in the channel are not a concern in terms of fluctuations in characteristics.

6.4 Discussion of parameter fluctuation results

Tunnel FETs experience fluctuations in their characteristics that depend on the parameter target values chosen, and which are often not symmetric when decreasing or increasing the parameter value. Table 2 summarizes the results of our simulations by reporting the quantitative parameter variation of V_{TG} , I_{on} , S_{point} , and S_{avg} as percentages, for the Tunnel FET with the target values listed in the first column. Parameters were varied in both the negative and positive directions. After looking at their effects on characteristics, parameters can be separated into three categories: (i) those with a large influence on Tunnel FET characteristics everywhere in the studied parameter range, (ii) those with a large influence on characteristics in some part of the studied range, and a small influence elsewhere, and (iii) those with little influence on Tunnel FET characteristics.

Table 2: Summary of the results of parameter fluctuations on key characteristics.

Parameter and target value	Parameter variation		V_{TG} % variation		I_{on} % variation		S_{point} % variation		S_{avg} % variation	
Gate dielectric permittivity = 25	-10	+10	20	-16	-86	309	27	-27	21	-23
Gate dielectric thickness = 3 nm	-0.75 nm	+0.75 nm	-15	12	220	-66	-29	17	-16	12
Junction width = 21 nm / 5 decades doping	-5.5 nm/dec	+5.5 nm/dec	-13	13	275	-72	-19	19	-18	15
Silicon body thickness = 10 nm	-5 nm	+5 nm	-28	13	266	-57	-58	16	-27	8
Gate alignment at src. junction = 0 nm	-2.5 nm	+2.5 nm	-27	12	371	-75	-33	13	-34	16
Intrinsic region/gate length = 30 nm	-5 nm	+5 nm	-0.8	1.1	1.7	-2.2	45	-17	0.7	0
Band gap at tunnel junction = 0.9 eV	-0.05 eV	+0.05 eV	-6	7	30	-24	-4	2	-5	6
Intrinsic region doping = 10^{16} atoms/cm ³	-one dec	+one dec	>.1	>.1	>.1	>.1	>.1	>.1	>.1	>.1

In order to minimize the fluctuations of device characteristics, it is the most critical to tightly control the parameters in the first category: the dielectric permittivity and thickness (gate capacitance), and the junction width (doping diffusion) at the tunnel junction. Since there is no range of parameter values that leads to low characteristic sensitivity, the only option is good control during fabrication. For all these parameters, the on-current is the most sensitive parameter, and it changes by a factor of 2 or 3 for the target and variation values shown in Table 2.

Band gap at the tunnel junction also fits into this category, though its effect is slightly less extreme. When the band gap at the source junction changes from 0.9 to 0.85 or 0.95 eV, the on-current varies by about 20 to 30%.

The second category is comprised of the parameters whose variation causes a large fluctuation in characteristics in some range, and little to no change in another range. In these cases, a trade-off must be made. The parameters of this type are body thickness, gate alignment with the tunnel junction, and intrinsic region length.

If a double-gate or multi-gate structure is used, a thin device is desirable in order to maximize the electrostatic control of the gate on the tunnel junction. When this design choice is made, the characteristics will be sensitive to body thickness variations, and once again, tight control of this parameter will be crucial. When 10 nm is the target silicon thickness for the optimized devices

studied here, a variation of body thickness by 5 nm causes a change of more than a factor of 2 for on-current. If a thin Tunnel FET body is not necessary to a certain design, the devices can be made in the non-sensitive part of the body thickness range (in the case shown here, for $t_{Si} > 20$ nm), and control can be relaxed.

The next parameter in the second category is gate alignment with the tunnel junction. Once again, there is a choice to be made between attaining the best characteristics possible by finding the precise alignment of the gate that produces the narrowest tunnel barrier in the on-state, and stabilizing characteristics by overlapping the gate to the source region. If maximum optimization is chosen, there are two parameters to control precisely – dopant diffusion (to control the placement of the electrical junction), and the location of the gate edge. Even without maximum optimization, if the gate is used for a self-aligned process in which dopants are implanted and then diffuse under the gate in an annealing step, the gate edge is automatically located in a region of high sensitivity. For the devices shown here, a 2.5 nm variation in the self-aligned gate edge placement causes the on-current to vary by a factor of 3 or 4. To reduce, and even practically eliminate this fluctuation, the gate dielectric (with or without the gate contact) should overlap the source region.

The third parameter requiring a trade-off is device length. The design choice here is a bit more straightforward because avoiding the region of characteristic fluctuations doesn't require giving up higher I_{on} and lower swing. In order to keep I_{off} low, however, aggressive scaling must be avoided. If very small Tunnel FETs are to have a very high I_{on}/I_{off} ratio, it will be necessary to maximize electrostatic control, by building gate-all-around devices on nanowires, for example. Even so, there will be some critical device length L_{crit} below which I_{off} and swing will be too high due to diode leakage, and Tunnel FETs must not be designed too close to this critical length in order to avoid characteristic fluctuations. As shown in Table 2, fluctuations are small for variations of ± 5 nm around a device length of 30 nm. If $L_{intrinsic}$ were 25 nm, however, and varied by 5 nm to give a minimum length of 20 nm, the minimum current would vary by a factor of 50, S_{avg} by 28%, and S_{point} by 76%.

The alignment of the gate to the drain junction is a design choice linked to the minimum device length. Shortening the gate effectively shortens the intrinsic region, as discussed in section 6.3.4.

The parameters in the third category lead to little or no variation in characteristics. Tunnel FETs show insensitivity to intrinsic region doping concentration, even when it varies by an order of magnitude. Random discrete dopants in the channel are not a concern, which is an advantage over conventional MOSFETs on bulk, and a similarity with ultra-thin body (UTB) MOSFETs.

6.5 Comparison with conventional silicon MOSFETs

The magnitudes of the characteristic variations for Tunnel FETs can be compared to those for conventional silicon MOSFETs, as presented in a process variation sensitivity study by Xiong and Bokor [25] for 20-nm long, 5-nm thick FinFETs simulated in ISE. Their channel region was intrinsic, t_{ox} was 1 nm, and V_{dd} was 0.8 V. The Tunnel FETs with which we will compare their results in this section had $L = 30$ nm, $t_{Si} = 10$ nm, effective oxide thickness (EOT) = 0.47 nm, and $V_{dd} = 1$ V. Xiong and Bokor looked at the variation of three key parameters: gate length, oxide thickness, and silicon body thickness (in their case, fin width).

Their calculated 3σ value for EOT was 1 Å, while our Tunnel FET EOT variation was 1.2 Å, as shown in Table 2. For their n-type FinFETs, the changes in characteristics for this change in oxide thickness were $\Delta V_T = 5.7$ mV, $\Delta S = 1.4$ mV/dec, and $\Delta I_{on} = 4.0$ %. For our Tunnel FETs, reporting the changes for -1.2 Å and +1.2 Å separately since they were not symmetric, the corresponding changes were $\Delta V_T = -81$ and 62 mV, $\Delta S_{avg} = -8.9$ and 6.8 mV/dec, and $\Delta I_{on} = 220$ and -66 %. For each characteristic, the variations for Tunnel FETs were 1 to 2 orders of magnitude higher than that of conventional MOSFETs.

The next parameter varied by Xiong and Bokor was gate length, whose variation has a much stronger effect on conventional MOSFETs than on Tunnel FETs. For gate lengths of 20 nm +/- 2 nm, they

saw changes of $\Delta V_T = 18$ mV, $\Delta S = 3.6$ mV/dec, and $\Delta I_{on} = 3.6$ %. As long as a Tunnel FET is designed far enough above L_{crit} , a 2 nm length variation will have a negligible effect on characteristics.

Finally, Xiong and Bokor looked at the variation of characteristics due to changes in fin width, corresponding to our Tunnel FET silicon thickness t_{Si} . For 5 nm +/- 1 nm, they saw changes of $\Delta V_T = 33$ mV, $\Delta S = 4.0$ mV/dec, and $\Delta I_{on} = 5.4$ %. For Tunnel FETs, we looked at the very sensitive range of 10 nm +/- 2 nm, and again reporting the characteristic changes for +/- separately since they were not symmetric, the variations were $\Delta V_T = -83$ and 39 mV, $\Delta S_{avg} = -8.9$ and 4.0 mV/dec, and $\Delta I_{on} = 192$ and -47 %. For very thin silicon layers, the changes in threshold voltage and subthreshold swing are comparable for conventional MOSFETs and Tunnel FETs, while the fluctuation of on-current for Tunnel FETs is 1 to 2 orders of magnitude higher for Tunnel FETs.

6.6 Conclusion

We have reported the first systematic simulation-based study of the sensitivity of the static characteristics of optimized silicon Tunnel FETs to parameter fluctuations. We predict a much reduced sensitivity to doping fluctuations and gate length scaling (when devices are designed with $L > L_{crit}$) compared to conventional CMOS. A trade-off will need to be made when choosing the desired alignment of the gate dielectric at the source end, between improved characteristics with a design requiring tight control of the gate edge, and non-optimal characteristics with an overlapped gate whose edge variations would not lead to fluctuations in characteristics. Finally, this study suggests that the control of the high-k gate process, junction doping width at the tunnel junction, and film thickness in UTB SOI, with less parameter fluctuation than that required by CMOS, is crucial for future high-performance Tunnel FETs with reproducible characteristics.

The following technical topics and contributions were presented in this chapter:

- *The importance of a parameter variation study* (Section 6.1)

This section introduced the topic and explained that in order to know what parameters need to be the most tightly controlled during Tunnel FET fabrication, a detailed study of parameter variation is essential.

- *Performance boosters for silicon Tunnel FETs* (Section 6.2)

One parameter at a time was improved, in a step-by-step technique of Tunnel FET optimization. It was shown that superior characteristics can be obtained when all parameters are optimized in the same device design. The resulting device was used for the parameter fluctuation study.

- *Parameter fluctuation study* (Section 6.3)

The characteristic fluctuations were investigated for variation of the following Tunnel FET parameters: gate dielectric permittivity and thickness, tunnel junction doping profile abruptness, silicon body thickness, gate oxide and contact alignment, device length, band gap at the tunnel junction, and intrinsic region doping. For all but the last parameter, the changes in on-current, subthreshold swing, and gate threshold voltage were reported for variations around a target value. The on-current values changed more drastically than swing and V_T .

- *Discussion of parameter fluctuation results* (Section 6.4)

This section analyzed the magnitude of characteristic fluctuations due to parameter variations, and put the parameters into categories depending on the size and nature of the sensitivity of the characteristics. In the first category, the parameters that always had a large influence on characteristics were the dielectric permittivity and thickness, the junction width at the tunnel junction, and band gap at the tunnel junction. In the second category, the parameters that had a large influence on characteristics in some range of values, and little influence in another range of values, were oxide alignment at the tunnel junction, silicon body thickness, and gate length. In the third category, the parameter that had little effect on characteristics was intrinsic region doping level.

The demonstration of additive boosters, and the parameter variation study and its analysis represent original contributions.

- *Comparison with conventional silicon MOSFETs* (Section 6.5)

Comparing the sensitivity of Tunnel FETs to parameter variation with that of conventional MOSFETs, it was found that Tunnel FETs show higher sensitivity to changes in gate oxide thickness and body thickness. For comparable variations in body thickness, Tunnel FET on-current varied 1 to 2 orders of magnitude more than MOSFET on-current. For comparable variations in effective oxide thickness, Tunnel FET threshold voltage, subthreshold swing, and on-current all showed 1 to 2 orders of magnitude more sensitivity than those same parameters for conventional MOSFETs. These two parameters will thus require much tighter control during Tunnel FET fabrication than for conventional CMOS fabrication.

6.7 Bibliography

- [1] G. Roy, A. Brown, F. Adamu-Lema, S. Roy, and A. Asenov, "Simulation Study of Individual and Combined Sources of Intrinsic Parameter Fluctuations in Conventional Nano-MOSFETs," *IEEE Trans. Electron Devices*, vol. 53, no. 12, pp.3063-3070, Dec. 2006.
- [2] H. P. Tuinhout, "Impact of parametric mismatch and fluctuations on performance and yield of deep-submicron CMOS technologies," in *Proc. ESSDERC*, Florence, Italy, 2002, pp. 95-101.
- [3] O. Nayfeh, J. Hoyt, and D. Antoniadis, "Strained-Si_{1-x}Ge_x/Si Band-to-Band Tunneling Transistors: Impact of Tunnel-Junction Germanium Composition and Doping Concentration on Switching Behavior," *IEEE Trans. Electron Devices*, vol. 56, no. 10, pp.2264-2269, Oct. 2009.
- [4] K. Moselund, H. Ghoneim, M. T. Björk, H. Schmid, S. Karg, E. Lörtscher, W. Riess, and H. Riel, "Comparison of VLS grown Si NW tunnel FETs with different gate stacks," in *Proc. ESSDERC*, 2009, pp. 448-451.
- [5] C. Le Royer and F. Mayer, "Exhaustive experimental study of tunnel field effect transistors (TFETs): From materials to architecture," in *Proc. ULIS*, Aachen, March 18-20, 2009, pp. 53-56.
- [6] T. Krishnamohan, D. Kim, S. Raghunathan, and K. Saraswat, "Double-Gate Strained-Ge Heterostructure Tunneling FET (TFET) With Record High Drive Currents and <60mV/dec Subthreshold Slope," in *IEDM Tech. Dig.*, Dec. 15-17, 2008, pp. 947-949.
- [7] C. Sandow, J. Knoch, C. Urban, Q.-T. Zhao, S. Mantl, "Impact of electrostatics and doping concentration on the performance of silicon tunnel field-effect transistors," *Solid State Electron.*, vol. 53, pp. 1126-1129, 2009.
- [8] M. Fulde, A. Heigl, M. Weis, M. Wirnshofer, K. v. Arnim, T. Nirschl, M. Sterkel, G. Knoblinger, W. Hansch, G. Wachutka, and D. Schmitt-Landsiedel, "Fabrication, Optimization, and Application of Complementary Multiple-Gate Tunneling FETs," in *Proc. INEC*, 2008, pp. 579-584.
- [9] K. Bhuwarka, J. Schulze, and I. Eisele, "Scaling the vertical tunnel FET with tunnel bandgap modulation and gate workfunction engineering," *IEEE Trans. Electron Devices*, vol. 52, no. 5, pp. 909-917, May 2005.
- [10] K. Boucart and A. M. Ionescu, "Double-Gate Tunnel FET With High-k Gate Dielectric," *IEEE Trans. Electron Devices*, vol. 54, no. 7, p.1725-1733, July 2007.
- [11] J. Knoch, "Optimizing tunnel FET performance – Impact of device structure, transistor dimensions and choice of material," in *Proc. VLSI-TSA*, 2009, pp.45-46.
- [12] M. Schlosser, K. Bhuwarka, M. Sauter, T. Zilbauer, T. Sulima, and I. Eisele, "Fringing-Induced Drain Current Improvement in the Tunnel Field-Effect Transistor With High-k Gate Dielectrics," *IEEE Trans. Electron Devices*, vol. 56, no. 1, pp.100-108, Jan. 2009.
- [13] A. Verhulst, W. Vandenberghe, K. Maex, and G. Groeseneken, "Boosting the on-current of a n-channel nanowire tunnel field-effect transistor by source material optimization," *J. Appl. Phys.*, vol. 104, pp. 064514-1-10, 2008.
- [14] K. Boucart, W. Riess, and A. M. Ionescu, "Lateral Strain Profile as Key Technology Booster for All-Silicon Tunnel FETs," *IEEE Electron Device Lett.*, vol. 30, no. 6, pp.656-658, June 2009.
- [15] J. Knoch, S. Mantl, and J. Appenzeller, "Impact of the dimensionality on the performance of tunneling FETs: Bulk versus one-dimensional devices," *Solid-State Electronics*, vol. 51, iss. 4, April 2007, pp. 572-578.

-
- [16] R. Surdeanu, Y.V. Ponomarev, R. Cerutti, B.J. Pawlak, C.J.J. Dachs, P.A. Stolk, M.A. Verheijen, M. Kaiser, M.J.P. Hopstaken, J.G.M. van Berkum, F. Roozeboom, L.K. Nanver, I. Hofliijk, R. Lindsay, "Laser Annealing for Ultra-shallow Junction Formation in Advanced CMOS," *Electrochemical Society Meeting*, 2002, p. 413.
- [17] N. D. Nguyen, E. Rosseel, S. Takeuchi, J.-L. Everaert, R. Loo, J. Goossens, A. Moussa, T. Clarysse, M. Caymax, and W. Vandervorst, "Vapor phase doping and sub-melt laser anneal for the fabrication of Si-based ultra-shallow junctions in sub-32 nm CMOS technology," in Proc. *ISDRS*, 2009.
- [18] Semiconductor Industry Association (SIA), International Technology Roadmap for Semiconductors (ITRS), 2009. [Online]. Available: <http://www.itrs.net/>.
- [19] K. Uchida, H. Watanabe, A. Kinoshita, J. Koga, T. Numata, and S. Takagi, "Experimental Study on Carrier Transport Mechanism in Ultrathin-body SOI n- and p-MOSFETs with SOI Thickness less than 5 nm," in *IEDM Tech. Dig.*, 2002, pp. 47-50.
- [20] Shimizu, K., Hiramoto, T., "Mobility Enhancement in Uniaxially Strained (110) Oriented Ultra-Thin Body Single- and Double-Gate MOSFETs with SOI Thickness of Less Than 4 nm," in *IEDM Tech. Dig.*, 2007, pp. 715-718.
- [21] L. Gomez, I. Åberg, and J. Hoyt, "Electron Transport in Strained-Silicon Directly on Insulator Ultrathin-Body n-MOSFETs With Body Thickness Ranging From 2 to 25 nm," *IEEE Electron Device Lett.*, vol. 28, no. 4, 2007, pp. 285-287.
- [22] Schmidt, M., Lemme, M.C., Gottlob, H.D.B., Kurz, H., Driussi, F., Selmi, L., "Mobility extraction of UTB n-MOSFETs down to 0.9 nm SOI thickness," in Proc. *ULIS*, 2009, pp. 27-30.
- [22] A. Verhulst, W. Vandenberghe, K. Maex, and G. Groeseneken, "Tunnel field-effect transistor without gate-drain overlap," *Appl. Phys. Lett.*, vol. 91, 2007, pp. 053102-1-3.
- [23] K. Boucart, A. M. Ionescu, "Length scaling of the Double Gate Tunnel FET with a high-K gate dielectric", *Solid-State Electronics*, vol. 51, iss. 11-12, Nov.-Dec. 2007, pp. 1500-1507.
- [24] D. Ma, C. Lee, F. Au, S. Tong, and S. Lee, "Small-Diameter Silicon Nanowire Surface," *Science*, vol. 299, 2003, pp. 1874-1877.
- [25] S. Xiong and J. Bokor, "Sensitivity of Double-Gate and FinFET Devices to Process Variations," *IEEE Trans. Electron Devices*, vol. 50, no. 11, 2003, pp.2255-2261.

Chapter 7

Conclusion and Perspectives

7.1 Conclusion

Numerical simulations have proven to be an effective means to investigate Tunnel FET behavior and the dependence of its static characteristics on changes in dimensions, doping, and other parameters. The work presented here can be useful to other researchers who will be designing and fabricating Tunnel FETs, and developing analytical and compact models for these devices.

The main accomplishments of this work can be summarized as follows:

- The optimization of the static characteristics of Tunnel FETs by the variation of gate structure (single or double), source, drain, and intrinsic region doping levels, gate dielectric material, and silicon body thickness was carried out. The resulting optimized device had a double gate, a high source doping and lower drain doping to suppress ambipolar behavior, a high-k dielectric, and a silicon body thickness of 10 nm.
- An investigation of band gap reduction at the tunnel junction was performed. The best characteristics were attained when band gap was small at the junction where tunneling takes place so that on-current was improved, and large at the other (drain-side) junction so that off-current stayed low.
- A new threshold voltage extraction technique for Tunnel FETs was proposed, in which the transconductance method pinpoints the voltage at which a transition takes place between strong and weak control of the tunneling energy barrier width, and therefore the on-current, by the applied bias. Using this technique, the threshold voltage has a physical meaning, and it marks the transition between a quasi-exponential and a linear dependence of drain current on applied voltage.
- Two threshold voltages were identified in Tunnel FETs, one in relation to the gate voltage, and the second in relation to the drain voltage. It was found that each of these threshold voltages depends upon the bias applied to the opposite terminal; drain threshold voltage depends upon the applied gate-source voltage, and gate threshold voltage depends upon the applied drain-source voltage.
- A detailed study of the length scaling of Tunnel FETs was conducted, in which all other parameters stayed constant. It was shown that Tunnel FETs are immune to the effects of length scaling, with two exceptions. The first is not of large concern for advanced device designers; with long devices (around 1 μm or more) on-current is degraded by a few percent by the resistance of the intrinsic region. The second is very important for devices that need to be compatible with highly-scaled CMOS. The leakage current in highly-scaled Tunnel FETs is a serious issue, and it seems that a critical gate length on the order of 10 to 20 nm exists, below which p-i-n leakage degrades the off-current significantly.
- Supply voltage scaling was investigated. One important finding was that the scaling of the effective oxide thickness was a possibility for keeping on-current constant when reducing V_{DD} .
- Finally, a parameter variation study was carried out, looking at the effects on important static characteristics: on-current, threshold voltage, and subthreshold swing. The most important finding was that gate dielectric thickness and dielectric constant, silicon body thickness, and tunnel junction doping profile abruptness are the parameters whose tight control is the most critical during Tunnel FET fabrication. In addition, above the critical gate length at which p-i-n leakage becomes a concern, gate length variations cause almost no fluctuation in characteristics, unlike the case for conventional MOSFETs.

7.2 Perspectives

Since the beginning of this thesis work in early 2005, the progress made with Tunnel FETs has come a long way. Referring back to the state-of-the-art in Chapter 2, it can be seen that the majority of Tunnel FET work has been done in the past few years. I spent the first few years of my thesis work waiting for fabricated devices and experimental results. Now many have started to appear in the literature, but there is still much progress to be made in terms of optimization to achieve superior device characteristics.

This perspectives section is broken down into two parts: first more generally, the progress still to be made on the subject of Tunnel FETs, and then more specifically, further work that could be carried out as a continuation of what I have begun here.

General Tunnel FET future work: As already mentioned, the biggest future challenge is to successfully design and fabricate fully-optimized Tunnel FETs of both n-type and p-type, that show low off-currents beyond what is possible for conventional MOSFETs, high on-currents, and average subthreshold swings of less than 60 mV/decade at room temperature. I believe that all-silicon Tunnel FETs will only be appropriate for specialized applications in which on-current can be low, due to the band-to-band tunneling barrier which, in series with a MOSFET-like channel, will always limit the on-current. If Tunnel FETs are to be used as a conventional MOSFET replacement, the on-current must be boosted to ITRS roadmap levels by some technique that reduces the band gap at the tunnel junction. A lateral strain profile, as presented in Chapter 3, is one possibility, but the maximum reported stress levels (see Chapter 3 bibliography) only correspond to a band gap reduction to about 0.9 eV, whereas germanium has a band gap of around 0.67 eV. Although the fabrication of low-defect heterojunctions will offer a new set of challenges to device engineers, I believe that their development is essential to the future of high-performance Tunnel FETs.

Further work will also be necessary in order to develop accurate analytical and compact models for Tunnel FETs. Although some work has been done in these areas, the theoretical framework which will allow experimental data to be fitted has not yet been developed. As will be shown in the appendix, even the non-local models of well-known device simulation tools such as Silvaco Atlas need to be improved if they are to be used to accurately fit experimental Tunnel FET data. More calibration and tuning of the models is required, and will become possible once more experimental data is available.

As a next step, a better understanding of the dynamic characteristics of Tunnel FETs are also necessary. AC simulations were originally intended to be part of this thesis work, but unfortunately, this was not possible due to software limitations.

Continuation of this work: There are several topics presented in this thesis which still have some open points, or certain aspects that are not fully understood. These points need further study.

One of these topics is the dependence of static characteristics on very small silicon film thickness, as in ultra-thin body (UTB) devices. In Chapter 3, the on-current was seen to have a maximum at some small body thickness on the order of 7 nm. The reason for the decrease in on-current with smaller t_{Si} was hypothesized to be the limited volume available for current flow, but this should be studied further to be fully understood. Once again, more experimental data is necessary in order to allow the calibration of the device simulation tools. Then it will be possible to discover whether the optimum silicon film thickness is an artifact coming from the limitations of the simulator, or a real physical phenomenon in ultra-thin structures.

In Chapter 6, it was mentioned that in very small nanowires, the electrostatic gate control is increased, which should improve device characteristics, but that the small dimensions lead to a larger band gap which would have a negative effect on Tunnel FET on-current. The interaction of these advantageous and disadvantageous effects in small nanowires deserves more investigation, especially if the goal is to build 1D Tunnel FETs that operate at the quantum capacitance limit, as described in [1] and recommended in [2].

Finally, one last important point is the current state of device simulators. When I began this thesis, there was no non-local band-to-band tunneling model in Silvaco Atlas. It was added in 2006, and continues to be improved. For novel devices such as Tunnel FETs, it would be extremely advantageous to have a close interaction between the developers of the simulation tools, and the researchers producing the experimental data. Then it will become possible for the simulators to go beyond predicting trends, and give accurate estimations of on-current and other important device characteristics.

I am convinced that once Tunnel FETs have been investigated and developed more fully, and once p- and n-type Tunnel FETs with highly optimized characteristics have been fabricated, that these promising devices will live up to their potential. I am optimistic that these devices, or some variation upon them, will bring lower power consumption and better energy-efficiency to computers, appliances, and devices everywhere.

7.3 Bibliography

- [1] A. Rahman, J. Guo, S. Datta, and M. Lundstrom, "Theory of Ballistic Nanotransistors," *IEEE Trans. Electron Devices*, vol. 50, no. 9, 2003, pp. 1853–1864.
- [2] J. Knoch, M. Björk, H. Riel, H. Schmid and W. Riess, "One-Dimensional Nanoelectronic Devices - Towards the Quantum Capacitance Limit," in *DRC Technical Digest*, 2008, pp. 173-176.

APPENDIX: MODELS, MESHING, AND CALIBRATION

This appendix describes the models and meshing used in the simulations presented in this thesis, and the efforts made to fit experimental data using the non-local band-to-band tunneling (BTBT) model in Silvaco Atlas.

A.1 Silvaco Atlas models

All equations cited in this section come directly from [1].

A.1.1 Non-local band-to-band tunneling model

All band-to-band tunneling was simulated using the non-local BTB tunneling model in Silvaco Atlas. This model requires a special mesh to be applied around the area where the tunneling can take place. The energy bands are calculated as shown in Fig. A.1, and each energy in the allowed range shown in the figure is considered, and start- and end-positions for tunneling are determined for each energy. For an electron that has an energy between $E - \Delta E/2$ and $E + \Delta E/2$ (where ΔE is a small energy increment), the contribution to the BTBT current is

$$J(E) = \frac{qkTm^*}{2\pi^2\hbar^3} T(E) \ln \left\{ \frac{1 + \exp[E_{F1} - E]/kT}{1 + \exp[E_{Fr} - E]/kT} \right\} \Delta E \quad (2.1)$$

Here, $T(E)$ is the tunneling probability, E_{F1} and E_{Fr} are the Fermi levels shown in Fig. A.1, and

$$m^* = m_0 \sqrt{m_e(x_{end})m_h(x_{beg})} \quad (2.2)$$

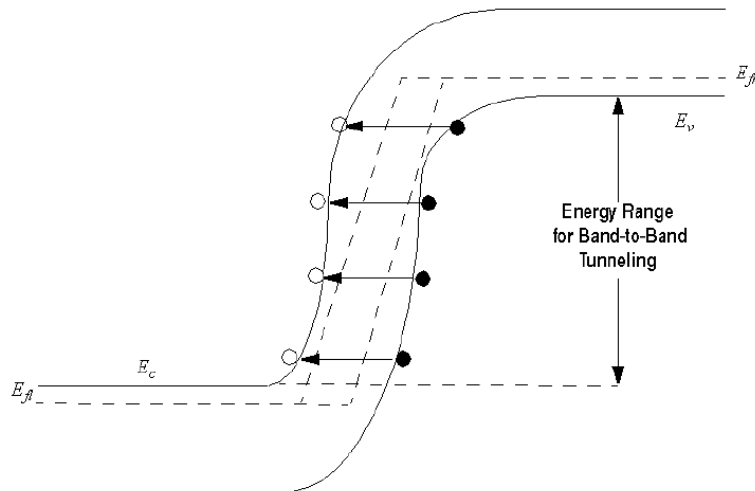


Figure A.1: Band diagram for calculating band-to-band tunneling current, from [1].

The tunneling probability $T(E)$ is calculated with a two-band approximation for the evanescent wavevector, given by

$$k = \frac{k_e k_h}{\sqrt{k_e^2 + k_h^2}} \quad (2.3)$$

with

$$k_e = \frac{1}{i\hbar} \sqrt{2m_0 m_e (E - E_c)} \quad (2.4)$$

and

$$k_h = \frac{1}{i\hbar} \sqrt{2m_0 m_h (E_v - E)}. \quad (2.5)$$

A carefully-applied WKB method is used to calculate the tunneling probability. The density of states effective masses are used for the conduction and valence bands. The electron and hole masses used everywhere in this thesis were the values given by Silvaco in example band-to-band tunneling files: $m_e = 0.20$ and $m_h = 0.34$. These masses were changed (made slightly heavier) only here in this appendix when attempting to fit experimental data.

Section A.4 of this appendix will show that the magnitude of the band-to-band tunneling current for this non-local model is not well-calibrated to experimental data and is probably too high. The general I_{DS} - V_{GS} and I_{DS} - V_{DS} curve shapes are correct, however, as seen by comparisons with experimental data. The relative trends and dependencies on variations in device dimensions and parameters also appear to be dependable, as confirmed in part by agreement with general band-to-band tunneling equations, as was shown in Figs. 3.9 and 3.15.

A.1.2 Bandgap narrowing

Silvaco Atlas uses a bandgap narrowing model from Slotboom and de Graaf [2] when doping concentrations are high, which reduces the band gap by lowering the conduction band and raising the valence band by the same amount.

The equation used is

$$\Delta E_g = BGN.E \left\{ \ln \frac{N}{BGN.N} + \left[\left(\ln \frac{N}{BGN.N} \right)^2 + BGN.C \right]^{0.5} \right\} \quad (2.1)$$

for which all simulations presented here used the default values for the variables: $BGN.E = 9.0 \times 10^{-3}$, $BGN.N = 1.0 \times 10^{17}$, and $BGN.C = 0.9$.

A.1.3 Quantum model

This model uses a density gradient to simulate the effects of quantum confinement in thin material layers. When this model is used, the standard equation for electron current,

$$\vec{J}_n = qD_n \nabla n - qn\mu_n \nabla \psi - \mu_n n (kT_L \nabla (\ln n_{ie})) \quad (2.1)$$

is slightly modified to become

$$\vec{J}_n = qD_n \nabla n - qn\mu_n \nabla(\psi - \Lambda) - \mu_n n(kT_L \nabla(\ln n_{ie})) \quad (2.2)$$

where Λ is a quantum correction potential. The same modification is made to the hole current equation.

Quantum confinement should also increase the band gap of the material, but this effect is not included in Silvaco Atlas' Quantum model.

A.2 Device structure and meshing

Fig. A.2 shows an annotated version of a typical device structure simulated in Silvaco Atlas, along with its mesh lines. The device shown here has an intrinsic region length of 50 nm. The gate dielectric covers the drain, i-region, and source in the example shown here, and the gate contact is aligned to the intrinsic region. Although only one gate dielectric and one gate contact are shown here, this simulation is for a double-gate device, since the bottom boundary of the device is treated as reflective, as it would be in a perfectly symmetric structure where the bottom boundary of Fig. A.2 is actually the horizontal center line in a double-gate Tunnel FET. (Multiple simulations confirmed that the results are the same for this device, mirrored and with current multiplied by two, as for a fully-simulated double-gate device.)

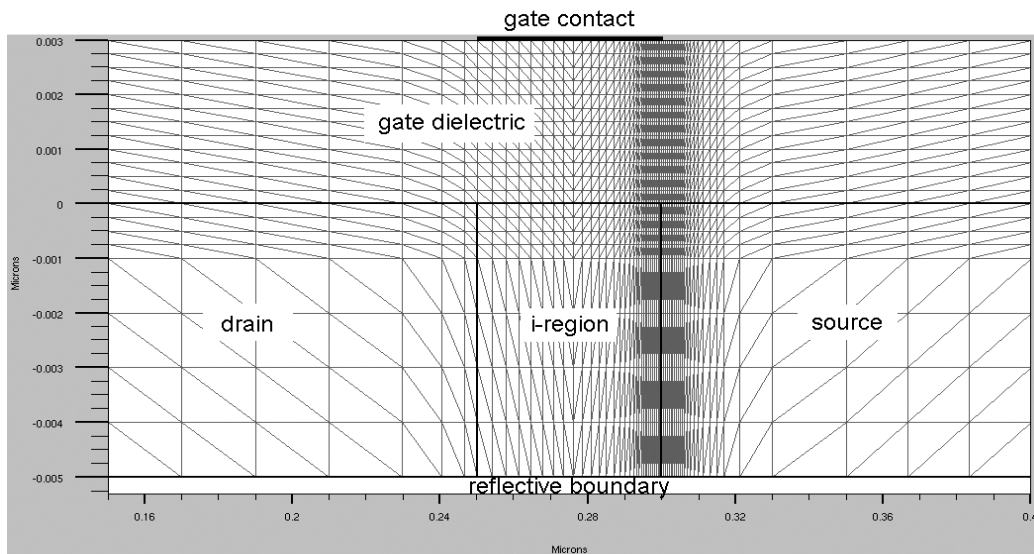


Figure A.2: One example of the Tunnel FET device structure simulated in Silvaco Atlas for this thesis, with its mesh lines. Half of a double-gate device was simulated to save on simulation time. For this device, $L = 50$ nm, $t_{Si} = 10$ nm (the upper 5 nm of which are shown here), and $t_{ox} = 3$ nm.

Fig. A.2 shows that the meshing was fine at the tunnel junction (0.5 nm, vertically) and in the gate dielectric (0.25, horizontally). If the mesh were changed to be less fine in those two critical areas, the resulting characteristics would be affected, as shown in Fig. A.3, where the dashed curves only have a meshing spacing of 1 nm at the tunnel junction and within the gate dielectric. The logarithmic scale (left axis) shows that some of the resolution in the off-region is lost with a less fine mesh, and the linear scale (right axis) shows that on-current is underestimated with a less fine mesh.

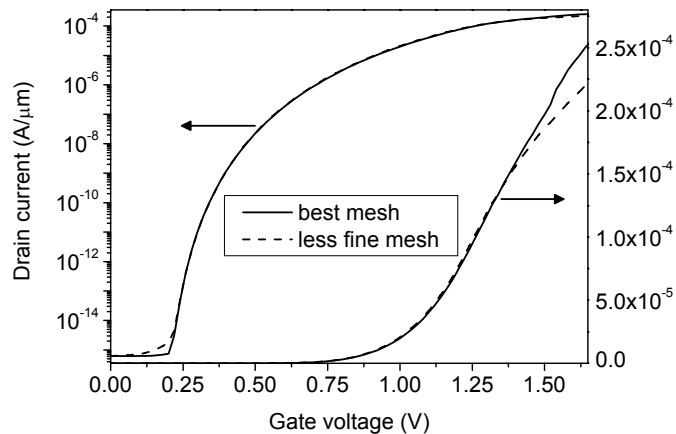


Figure A.3: Transfer characteristics of the Tunnel FET shown in Fig. 2, with the standard mesh (solid lines) and a less fine mesh where the minimum mesh spacing is 1 nm (dashed lines), showing the effects of meshing on off-current and on-current. $V_{DS} = 1$ V.

A.3 IBM diodes for calibration

The first attempt made to check the accuracy of Silvaco Atlas' non-local band-to-band tunneling model was with p-n diodes fabricated by IBM in New York [3]. They had published a graph with current vs. barrier width, which should in theory make calibration easy, since it is possible to measure the barrier width in simulated Tunnel FETs using device cross sections at particular applied voltages, and to know what drain current flows at those voltages. Fig. A.4 shows the result, where the symbols near the top right are the Silvaco Atlas simulation results using the non-local BTBT model, and the bold black symbols and lines are from the article.

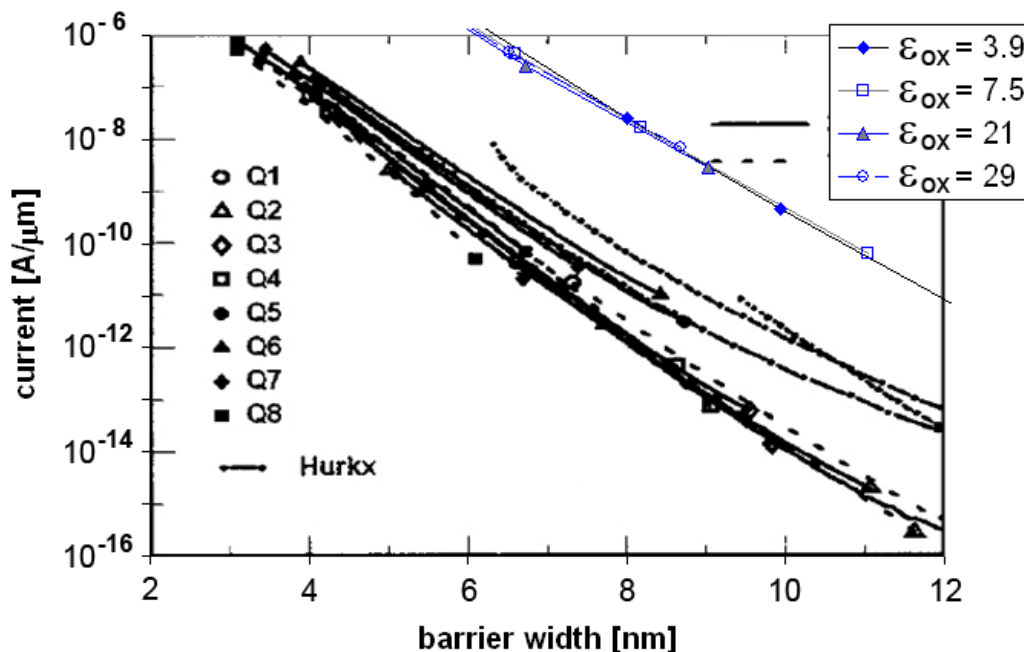


Figure A.4: Current vs. tunneling energy barrier width for IBM diodes from [3], and from Silvaco Atlas simulations. The simulations show the results for Tunnel FETs with four different gate dielectric constants.

From Fig. A.4, Silvaco Atlas seems to overestimate the band-to-band tunneling current by two to four orders of magnitude. There were, however, several problems with this comparison, making it hard to know by exactly how much the current is overestimated. First, the article reports the current

as a current density in A/cm^2 , and in their devices, tunneling took place perpendicularly to the silicon wafer surface. Their device schematic is shown in Fig. A.5. In the simulations, the current is in $\text{A}/\mu\text{m}$, and tunneling takes place in a small area under the gate dielectric in a direction parallel to the wafer surface. Some of the problems were summarized in an email from Stephen Wilson, an engineer at Silvaco:

“It seems that you have used $t_{\text{Si}} = 10 \text{ nm}$ for converting your data. But a look at the actual injected tunnel current densities shows it to be localised near to the oxide. Because the process is highly non-linear, expecting averages to agree is not a realistic assumption. In addition, although Solomon claims that the current density for large currents is not dependent on the details of the doping process, a p-i-n doping profile like yours may give a different result to the p-n profiles in the reference. This is because the tunnelling current is calculated from the product of statistical factor and transmission probability over the range of energies. The different profile may mean that the shortest tunnelling distance may have a different statistical factor, and so a comparison is not really possible.”

Since a direct comparison is not easy to make, it is not possible to know by how much the non-local BTBT model over-estimates Tunnel FET on-current. As mentioned in Chapter 7, this points again to the fact that it is necessary for simulator developers and experimentalists to work together closely in order to create calibrated simulation models.

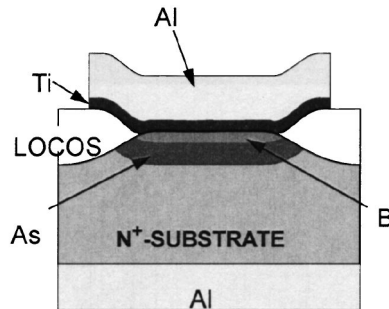


Figure A.5: Device schematic for diodes measured in [3].

A.4 Fitting Tunnel FET data

Knowing that the current tended to be overestimated, the electron and hole masses were changed for the simulations reported in this section. Instead of using the masses given in section A.1.1, the values $m_e = 0.265$ and $m_h = 0.563$ were used, which are the default electron and hole masses in the Sentaurus device simulator, and are slightly larger than those used elsewhere in this thesis.

The first attempt to fit Tunnel FET data used the results from Sandow et al. [4]. The results of this attempt, using the non-local BTBT model as described in section A.1.1, were not better than those for the diode fitting. The all-silicon p-type Tunnel FETs were in a 20-nm thick SOI layer, with a source and drain doped at $1.94 \times 10^{20} \text{ atoms}/\text{cm}^3$, and a doping abruptness of 3 nm/decade at the tunnel junction. One simplification was made in order to facilitate simulation convergence: a device length of 200 nm rather than 2 μm was simulated, which is expected to lower on-current at V_{Gmax} , but should otherwise not change device characteristics, as shown in Chapter 5.

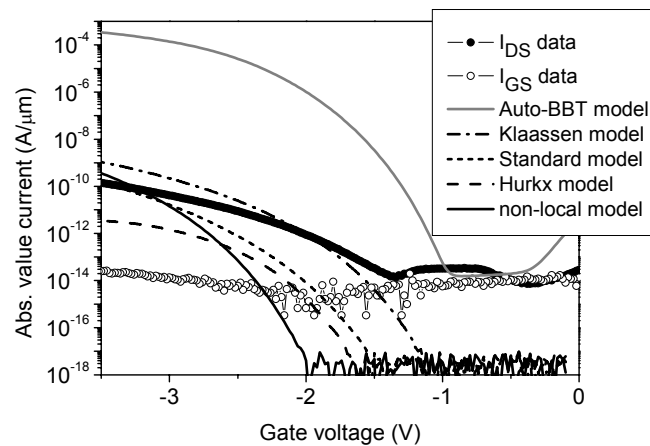


Figure A.6: Absolute value of drain (and gate) current, vs. gate voltage, from experimental devices in [4], and using five different band-to-band tunneling models in Silvaco Atlas.

The experimental (dots) and simulated (lines) I_{DS} - V_{GS} curves are shown in Fig. A.6, using not only the non-local model used in the rest of this thesis, but also four other local models provided in Silvaco Atlas, described in section A.5. Since tunneling is inherently a non-local process which depends upon the energy barrier width and the available energy states on either side of that barrier, the local models which depend instead upon the magnitude of the electric field at each mesh point should not give physically realistic results. Probably by chance then, the Klaassen local model gave the closest fit to the experimental results, while the non-local model had a turn-on which happened at a much more negative gate voltage, and showed a subthreshold swing which was much better than the experimental one. Coincidentally, the maximum on-current value shown in Fig. A.6 was similar to that of the experimental data, but only because of the voltage range presented here. A second attempt to fit experimental data was made for the all-silicon p-type Tunnel FETs from LETI, presented in [5]. The results are shown in Fig. A.7. Once again, the non-local model had a turn-on point at a much more negative gate voltage, and the subthreshold swing was again too small, which resulted in an on-current (at $V_{GS} = -2.5$ V) that was more than three orders of magnitude too high.

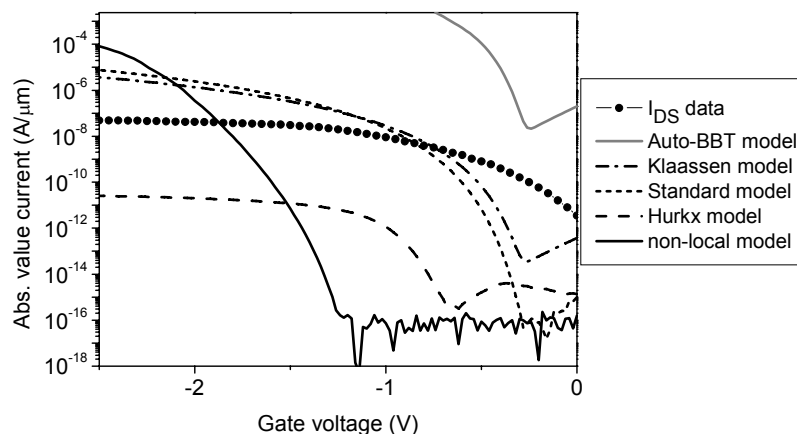


Figure A.7: Absolute value of drain current, vs. gate voltage, for experimental p-type Tunnel FETs from [5], and from five different BTBT models in Silvaco Atlas.

These failed attempts to fit experimental data show the shortcomings of the non-local band-to-band tunneling model in Silvaco Atlas. The only parameters which can be adjusted in the model are the electron and hole masses, and when they are kept within a realistic range, the simulation results do not fit the experimental data any better than what has been presented in this appendix. It would be advantageous to have good all-silicon n-type Tunnel FET experimental data to fit, but such is not the case at the present time. P-type Tunnel FETs are easier to fabricate because it is easier to make a very abrupt junction with arsenic than with boron, due to the high diffusion rates of boron, and as a

consequence, good all-silicon n-type Tunnel FET data has not been found. The late turn-on is seen only for p-type devices with Silvaco Atlas' non-local model. As can be seen elsewhere in this thesis, n-type Tunnel FETs turn on between about 0 and 0.5 V, in general, depending on the device design.

After the experience of using Silvaco Atlas for five years, I believe that the non-local BTBT model gives good relative results when parameters are varied. For example, when the gate dielectric constant is changed, we can get an idea of how much the on-current will change relative to some reference device. On the other hand, the absolute values of on-current, subthreshold swing, and threshold voltage, are not calibrated. The on-current could be mis-estimated by two or more orders of magnitude. This is an understandable situation when simulating emerging devices, that for the moment still lack the adequate experimental data necessary to calibrate these tools.

A.5 Local band-to-band tunneling models

These local models were used only in this appendix, and nowhere else in the thesis. They were used during attempts to fit experimental data with the band-to-band tunneling models provided in Silvaco Atlas. In general, local BTBT models are not the best choice for simulations, since band-to-band tunneling is inherently a non-local process which depends upon the tunneling barrier width and the availability of states on either side of the barrier. Local models do not take barrier width and available states into consideration, but instead depend upon an approximation of tunneling current based on the electric field at each mesh point.

There are four local band-to-band tunneling models in Silvaco Atlas. The general equation, used by all four of these local models, is

$$G_{BBT} = BB.A \times E^{BB.GAMMA} \times \exp\left(-\frac{BB.B}{E}\right)$$

where E is the electric field. The other variables are defined depending upon which model is chosen.

For the standard model, BBT.STD, the variables are $BB.A = 9.6615 \times 10^{18}$, $BB.B = 3.0 \times 10^7$ V/cm, and $BB.GAMMA = 2.0$. This model should be used for materials with a direct band gap.

For the Klaassen model, BBT.KL, the variables are $BB.A = 4.00 \times 10^{14}$, $BB.B = 1.9 \times 10^7$ V/cm, and $BB.GAMMA = 2.5$. This model should be used for materials with an indirect band gap.

For the AUTOBBT model, the variables are calculated as

$$BB.A = \frac{q^2 \sqrt{(2 \times MASS.TUNNEL \times m_0)}}{h^2 \sqrt{EG300}}$$

and

$$BB.B = \frac{\pi^2 EG300^{3/2} \sqrt{\frac{MASS.TUNNEL \times m_0}{2}}}{qh}$$

and then $BB.GAMMA = 2$. Here, q is the charge of an electron, h is Planck's constant, EG300 is the band gap at room temperature, m_0 is the rest mass of the electron, and MASS.TUNNEL is the effective mass of the electron.

Another variation of the basic equation is possible using the Hurkx model, BBT.HURKX. Then the general model has one additional factor called D [6], expressed as

$$D = \left(\exp[\phi_p - q\psi / kT] + 1 \right)^{-1} - \left(\exp[\phi_n - q\psi / kT] + 1 \right)^{-1}$$

where ϕ_n and ϕ_p are the quasi-Fermi levels for the electrons and the holes, respectively, and ψ is the electrostatic potential.

A.6 Bibliography

- [1] *Atlas User's Manual*, Silvaco Int., Santa Clara, CA, May 26, 2006.
- [2] J. Slotboom and H. De Graaf, "Measurements of Bandgap Narrowing in Silicon Bipolar Transistors," *Solid State Electronics*, vol. 19, 1976, pp. 857-862.
- [3] P. Solomon, J. Jopling, D. Frank, C. D'Emic, O. Dokumaci, P. Ronsheim, W. Haensch, "Universal tunneling behavior in technologically relevant P/N junction diodes," *J. Appl. Phys.*, vol. 95, no. 10, 2004, pp. 5800-5812.
- [4] C. Sandow, J. Knoch, C. Urban, Q.-T. Zhao, and S. Mantl, "Impact of electrostatics and doping concentration on the performance of silicon tunnel field-effect transistors," *Solid-State Elec.*, vol. 53, 2009, pp. 1126-1129.
- [5] C. Le Royer and F. Mayer, "Exhaustive Experimental Study of Tunnel Field Effect Transistors (TFETs): From Materials to Architecture," in *Proc. ULIS*, 2009, pp. 53-56.
- [6] G. Hurkx, D. Klaassen, and M. Knuvers, "A new recombination model for device simulation including tunneling," *IEEE Trans. Electron Devices*, vol. 39, no. 2, 1992, pp. 331-338.

Curriculum vitae

KATHY BOUCART

Limmattalstrasse 266
8049 Zürich
Phone: +41 43 333 44 02
Email: kathy.boucart@yahoo.com

Date of birth: 15.09.1971
Citizenship: USA
B permit



ENGINEERING EXPERIENCE

PhD work, Nanolab, EPFL (Ecole Polytechnique Fédérale de Lausanne) 2005-2010

- Thesis title: "Simulation of Double-Gate Silicon Tunnel FETs With a High-k Gate Dielectric"
- Initiated project to study Tunnel FET physics and characteristics.
- Optimized Tunnel FETs by simulation with Silvaco Atlas.
- Carried out and published studies of length scaling, a new physically-based threshold voltage, and parameter variation for Tunnel FETs.

Researcher, Nanolab, EPFL (Ecole Polytechnique Fédérale de Lausanne) 2004-2005

- Led research project on the design and fabrication of lateral and vertical MEMS resonators on thin SOI (Silicon-on-Insulator) substrates.
- Designed fabrication process and masks for MEMS resonators. Designed and carried out process development tests in cleanroom, analyzed results, and set up final fabrication process.
- Modelled resonator geometry and found modes and frequencies of resonance in Ansys.
- Wrote Matlab program to find resonator output characteristics (current, capacitance, resistance) given dimensions and quality factor.

Graduate Student Researcher, Device Group, UC Berkeley 2002-2003

- Investigated the use of thermally grown ultra-thin (10 Å) silicon nitride films for ultra-shallow and ultra-abrupt junction formation in CMOS applications.
- Designed experiments to demonstrate such films and studied their electrical and boron diffusion barrier properties.

Quality and Reliability Intern, Intel Corp, Santa Clara, CA March-Aug. 2002

- Researched degradation of PMOSFETs due to high bias and high temperature.
- Designed and executed tests on devices to understand impacts of stress and determine trends.
- Modeled transistor behavior dependence on defects, and modeled contributions of degradation parameters (threshold voltage shift and mobility degradation) to drain current shift.

Staff Researcher, Dept. of EECS, UC Berkeley 2001-2002

- Designed and built laser interference optics system to create sub-micron patterns on GaAs substrates, for the growth of quantum dots.
- Carried out processing for grating fabrication and characterization.

Technical Intern, Intel Corp, Hillsboro, OR summer 2000

- Designed assembly language programs to test for faults (short circuits and open circuits) within the floating point multiplier of the Pentium IV processor.
- Studied the micro-architecture, logic, and schematics in order to target untested paths. Excluded untestable faults.
- Achieved the project goal of 98% coverage for the multiplier.

AWARDS AND FELLOWSHIPS

- Young Scientist Award for ESSDERC presentation in 2006.
- Recipient of National Science Foundation Fellowship during Masters at UC Berkeley
- Recipient of Microelectronics and Computer Development Fellowship, and UT College of Engineering Graduate Fellowship, during engineering studies at UT Austin.

OTHER PROFESSIONAL EXPERIENCE

3D Artist / Web Developer, Microsoft Corp., Redmond, WA 1996-1998

- Designed and modeled 3D virtual environments.
- Scripted multimedia multi-user virtual environments.
- Built corporate web site for V-Chat, a virtual chat product.
- Learned C, C++, and VBScript.
- Managed contractors.

3D Artist / Scripter, Zombie Virtual Reality Entertainment 1995

Digital Video Editor and 3D Artist, Pacific Interactive 1994-1995

- Edited digital video.
- Designed, 3D modeled and animated environments and sequences for games.

TECHNICAL SKILLS

- Scientific software: Silvaco Atlas and Tonyplot, Origin, Matlab, Mathematica, Ansys.
- Office software: Word, Excel, PowerPoint.
- Testing laboratory: I-V, C-V, and charge pumping for device characterization. Ellipsometry.
- Clean room: Oxide growth, ultra-thin thermal nitride growth, optical and laser direct-writing lithography, reactive ion and chemical wet etch, chemical vapor deposition, chemical mechanical polishing, profilometry, SEM.
- Optics: Design and alignment of optical systems. Internal alignment of multi-mirror laser systems.

EDUCATION

EPFL (Ecole Polytechnique Fédérale de Lausanne) June 2010

- PhD from the Institute of Electrical and Electronics Engineering (IEL), under the supervision of Dr. Adrian Ionescu.

University of California, Berkeley (GPA: 3.92/4.00) Dec. 2003

- Masters degree in Electrical Engineering, under the supervision of Dr. Tsu-Jae King Liu.

The University of Texas at Austin (GPA: 4.00/4.00) Dec. 2001

- Completed all undergraduate work to begin a graduate program in Solid State Engineering.
- Taught two discussion sections for the Introduction to Computing class.

Northwestern University (GPA: 3.86/4.00) June 1994

- Bachelor of Science in Radio/TV/Film. Focus within major was computer graphics.
- Graduated with Highest Distinction.

LANGUAGES

- English: mother tongue
- French: fluent
- German: basic

PERSONAL INTERESTS

Spending time with my children, hiking, running, jewelry-making, reading, studying German, playing the piano.

LIST OF PUBLICATIONS

K. Boucart, W. Riess, A. M. Ionescu, "Asymmetrically Strained All-Silicon Tunnel Fets Featuring 1V Operation," *ESSDERC 2009*, Athens, Greece, 14-18 September 2009.

K. Boucart, W. Riess, A. M. Ionescu, "Lateral Strain Profile as Key Technology Booster for All-Silicon Tunnel FETs," *IEEE Electron Device Letters*, vol. 30, iss. 6, pp. 656 - 658, June 2009.

K. Boucart, A. M. Ionescu, "A new definition of threshold voltage in Tunnel FETs," *Solid-State Electronics*, vol. 52, iss. 9, pp. 1318-1323, September 2008.

K. Boucart, A. M. Ionescu, "Threshold voltage in Tunnel FETs: physical definition, extraction, scaling and impact on IC design," *ESSDERC 2007*, pp. 299-302, Munich, Germany, 11-13 Sept. 2007.

K. Boucart, A. M. Ionescu, "Double-Gate Tunnel FET With High-k Gate Dielectric," *IEEE Transactions on Electron Devices*, vol. 54, no. 7, pp. 1725-1733, July 2007.

K. Boucart, A. M. Ionescu, "Length scaling of the Double Gate Tunnel FET with a high-K gate dielectric," *Solid-State Electronics*, vol. 51, iss. 11-12, pp. 1500-1507, Nov.-Dec. 2007.

K. Boucart, A. M. Ionescu, "Double gate tunnel FET with ultrathin silicon body and high-k dielectric," *ESSDERC 2006*, pp. 383-386, Montreux, Switzerland, 19-21 Sept. 2006. (Young Scientist Award)

K. Buchheit (K. Boucart's maiden name), N. Abelé, A.M. Ionescu, "Design and Fabrication Issues in Ultra-Thin Film SOI MEMS resonators," *EuroSOI 2005*.

K. Buchheit (K. Boucart's maiden name), H. Takeuchi, T.-J. King, "Properties of Ultra-Thin Thermal Silicon Nitride," *MRS Conference Proceedings*, 2003.

Co-author:

L. Lattanzio, L. De Michielis, G. A. Salvatore, D. Bouvet, K. Boucart, A. M. Ionescu, "Ferroelectric Tunnel FET with a SiO₂/Al₂O₃/P(VDF-TrFE) gate stack", *ULIS 2010*, Glasgow, Scotland, UK, 17-19 March 2010.

M. Najmzadeh, K. Boucart, W. Riess, A. M. Ionescu, "Asymmetrically strained all-silicon multi-gate n-Tunnel FETs", article in press, *Solid State Electronics*, 2010.

A. M. Ionescu, K. Boucart, K. E. Moselund, V. Pott, D. Tsamados, "Small slope micro/nano-electronic switches," *International Semiconductor Conference, CAS 2007*, Sinaia, Romania, 15-17 Oct. 2007.

N. Abelé, K. Séguéni, K. Boucart, F. Casset, L. Buchaillet, P. Ancey, A. M. Ionescu, "Ultra-low voltage MEMS resonator based on RSG-MOSFET," *MEMS 2006*, pp. 882-885.

N. Abelé, R. Fritschi, K. Boucart, F. Casset, P. Ancey, A. M. Ionescu, "Suspended-Gate MOSFET: bringing new MEMS functionality into solid-state transistor," *IEDM 2005*, Washington DC, USA, December 2005. (late news)

N. Abelé, V. Pott, K. Boucart, F. Casset, K. Séguéni, P. Ancey, A.M. Ionescu, "Electro-Mechanical Modeling of MEMS Resonators with MOSFET detection," *MSM 2005 (Nanotech)*, Vol. 3, pp. 553-556, 2005.

N. Abelé, V. Pott, K. Boucart, F. Casset, K. Segueni, P. Ancey, A.M. Ionescu, "Comparison of RSG-MOSFET and capacitive MEMS resonator detection," *IEE Electronics Letters*, Volume: 41, Issue 5, pp. 242 - 244, March 2005.

A. M. Ionescu, K. Buchheit (K. Boucart's maiden name), P. Dainesi, S. Ecoffey, S. Mahapatra, V. Pott, "Nanowires: a Realistic Approach for Future Hybrid Nanoelectronics?" 3th NID Workshop, *NID 2004*, Athens, February 2004. (invited)

A. M. Ionescu, V. Pott, S. Ecoffey, S. Mahapatra, K. Moselund, P. Dainesi, K. Buchheit (K. Boucart's maiden name), M. Mazza, "Emerging nanoelectronics: multi-functional nanowires," International Semiconductor Conference, *CAS 2004*, Sinaia, Romania, October 2004. (invited)

P.C. Ku, K. Buchheit (K. Boucart's maiden name), C.J. Chang-Hasnain, J.A. Hernandez, "Buried selectively-oxidized AlGaAs structures grown on nonplanar substrates for device applications," *CLEO 2002*, pp. 137-138.