PEOPLE'S DEMOCRATIC REPUBLIC OF ALGERIA
Ministry of Higher Education and Scientific Research

University Hadj Lakhdar -BATNA 1-
Faculty of Matter Sciences
Department of Physics

A DISSERTATION

Presented in fulfillment of the requirement of the degree of

Doctorate of Sciences

By:

TOUFIK BENTRCIA

Entitled:

Study and modeling of multigate nanoscale MOSFETs including quantum and traps effects: Application to the nanoscale devices design

---

Specialty: PHYSICS OF METAL MATERIALS AND SEMICONDUCTORS

Defended in 21/04/2019

Approved by:

Committee members:

President:	Dr. Mohamed CHAHDI	Professor	University of Batna 1
Advisor:	Dr. Fayçal DJEFFAL	Professor	University of Batna 2
Members:	Dr. El-djemaï BELBACHA	Professor	University of Batna 1
	Dr. Abdelhamid BENHAYA	Professor	University of Batna 2
	Dr. Abdelouahab BENTABET	Professor	University of Bordj Bou-Arréridj
	Dr. Ramdane MAHAMDI	Professor	University of Batna 2

RÉPUBLIQUE ALGÉRIENNE DÉMOCRATIQUE ET POPULAIRE
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

Université Hadj Lakhdar -BATNA 1-
Faculté des Sciences de la Matière
Département de Physique

THÈSE

Présentée en vue de l'obtention du

Diplôme de Doctorat

Par:

TOUFIK BENTRCIA

Thème:

Étude et modélisation des MOSFETs nanométriques à multi-grille incluant les effets quantiques et de pièges: Application à la conception des composants nanométriques

Spécialité: PHYSIQUE DES MATÉRIAUX MÉTALLIQUES ET SEMICONDUCTEURS

Soutenue le 21/04/2019

Devant le jury:

Président:	Dr. Mohamed CHAHDI	Professeur	Université de Batna 1
Rapporteur:	Dr. Fayçal DJEFFAL	Professeur	Université de Batna 2
Examinateurs:	Dr. El-djemaï BELBACHA	Professeur	Université de Batna 1
	Dr. Abdelhamid BENHAYA	Professeur	Université de Batna 2
	Dr. Abdelouahab BENTABET	Professeur	Université de Bordj Bou-Arréridj
	Dr. Ramdane MAHAMDI	Professeur	Université de Batna 2


وما بكم من نعمة فمن الله
١٣٤٨

Acknowledgments

This research work has been conducted within the LEPCM laboratory at the University of Batna -1-.

First of all, I would like to capture this opportunity to express the best gratitude to my advisor Professor Fayçal Djeffal. The current research outcome could not be reached without his deepest expertise and continuous guidance all along the past years. Sincerely, I would like to thank him a lot for the balance between rigor, intelligence and humor, which has made our lab an amazing research environment. His inspirations and pertinent orientations have always allowed shedding light and tackling encountered problems.

I would like to thank Professor Mohamed Chahdi from the University of Batna -1- for taking the heavy responsibility of presiding the examination committee despite his many duties. My thanks go also to the examination committee members namely Professor Eldjemï Belbacha from the University of Batna -1-, Professor Abdelhamid Benhaya from the University of Batna -2-, Professor Abdelouahab Bentabet from the University of Bordj Bou-Arréridj and Professor Ramdane Mahamdi from the University of Batna -2-, for taking valuable time to review my dissertation and provide relevant comments regarding my research work.

I appreciate the support of my friends during all stages of this dissertation through sound discussions and shared joyful moments. I apologize for not mentioning any noun but I am sure that genuine friends know they are concerned.

Finally, I want to express sincere gratitude to all my family and especially my parents, for whom I hide billions of sentences carved

in my heart not necessary to be written.

Dedication

To all whom I love, I dedicate this modest work.

List of contents

List of figures

List of tables

List of algorithms

General introduction

« *Who has begun has half done.*
Have the courage to be wise.
Begin! »

Quintus Horatius Flaccus

Motivation

There is no doubt that the electronics industry in our daily life concerns broad varieties of employed fabrication procedures and products. Among the most common production elements issued from such industry we have consumer goods including for instance cell phones and computers. However, these products are just the tip of the iceberg since the lion share of the electronic industry outcomes is dealing with a more basic level, where the elaboration of semiconductor devices and integrated circuits represent an obvious aspect of the electronics industry. In fact, colossal technological advances have been conducted since the late 1950s leading to a rapid progress in electronics field.

The invention of the first bipolar junction transistor at Bell Research Laboratories has moved the tendency at that time to a new era, where the revolutionized idea has allowed the fabrication of smaller electronic components with less power and improved reliability. These novel artifacts known as transistors have become a substitute of thermionic valves. The semiconductor material called germanium has been used initially for the fabrication purposes before the silicon, which is more abundant and accredited for nowadays applications. The tremendous downscaling of the metal oxide semiconductor field effect transistor (MOSFET) has allowed reaching a very high level of packing, where a modern microprocessor contains over 2 billion MOSFETs. However, additional efforts should be dedicated to remedy or at least reduce the severe drawbacks as the process technology scales beyond 100 nm feature sizes.

The multigate MOSFET paradigm has emerged in recent years as a potential candidate to push further the scaling capability and pursue the validity of Moore's law. In this kind of structures, the gate is wrapped around some facets of the channel body and the source/drain regions are shaped using ion implantation. Hence, the device performance is enhanced by the improvement of the gate's electrostatic control over carriers in the channel. Furthermore, the fabrication process of multigate MOSFETs shares many principal stages with the standard complementary metal-oxide-semiconductor (CMOS) process despite the need of some adjustments. At the deep nanoscale region, these devices are playing a crucial role in circumventing many parasitic effects. Nevertheless, the very low dimensions lead to a significant generation of defects and fluctuations at both device manufacturing and operation levels.

Motivated by the accumulated difficulties facing process engineers and device designers along the miniaturization roadmap, we are interested in this dissertation by the problem of developing reliable modeling tools and designs to tackle realistic constraints inherent for the double gate MOSFET devices. Indeed, under such conditions many assumptions and approximations adopted in the case of long devices are questionable. Therefore, it is our objective to put forward modeling approaches capable of handling various aspects of the device behavior on one hand. In a second part, we focus on the validation of improved designs compared to the conventional DG MOSFET in terms of several performance criteria. We believe that by adopting the joint exploitation of modeling and optimization methodologies, it is possible to alleviate many shortcomings and the amendment path can be tailored for more precise and robust findings.

Contribution of the dissertation

The research work presented in this dissertation deals essentially with the modeling and design at the nanoscale level of the DG MOSFET device including interface traps. Several categories of modeling strategies are implemented namely fuzzy based, evolutio-

nary based and numerical simulation based approaches. Our main contributions rely on the elaboration of novel modeling tools, introduction of structural adjustments to boost the conventional DG MOSFET and finally the assessment of the generated designs in terms of specific performance measures. Despite the fact that the aforementioned contributions are of theoretical order, they are worthy to be accredited in technology procedures in order to provide guiding recommendations before the practical fabrication. In what follows, the basic contributions of the dissertation are elucidated in detail :

Development of novel modeling techniques for the nanoscale DG MOSFET including interface traps :

This contribution is relative to the implementation of powerful modeling tools with the aim of providing a coherent framework that supports current technological and operational constraints. In other words, some novel techniques are proposed for the first time while other common methods are adapted to well defined situations. It is worth mentioning that the set of developed techniques consists of :

- Adaptive neuro-fuzzy inference system based approach ;
- Kriging metamodeling based approach ;
- Technology Computer Aided Design based approach.

Proposition and validation of new amended designs of the nanoscale DG MOSFET including interface traps :

This goal is achieved by the proposition of various enhanced designs through the introduction of structural modifications at the level of the DG MOSFET geometry. In the validation stage, we compare the obtained designs with the conventional structure in order to show any possible improvement.

Investigation of global performances associated to the nanoscale DG MOSFET including interface traps :

In this contribution, we conduct a global investigation of the proposed models and designs with respect to specific fields of application. So, several performance measures are extracted in this context to evaluate the specificity of these prototypes to digital and analog/RF applications.

Organization of the dissertation

The research work summarized in this dissertation is organized in four chapters. The first chapter is dedicated to the state of the art regarding the conventional MOSFET device since it constitutes the backbone of any further improvement. The three remaining chapters are basically dedicated to the elucidation of our contributions.

In the first chapter, we provide both an overview about the evolution of MOSFET devices in addition to the basic physics theory governing such structure. A special attention is devoted to various performance metrics due to their importance for evaluating any structural modification. Furthermore, different drain current modeling approaches are described. Besides the severe miniaturization drawbacks, some recognized remedying strategies are also detailed. We believe that the introduction of these concepts is a mandatory prerequisite since they form the cornerstone to build reliable models not only for the deduction of some performance criteria but also for handling new constraints.

We are interested in the second chapter by the proposition of an adaptive network fuzzy inference system to predict the degradation at nanoscale level of the DG MOSFET performance due to the hot carrier effect. The impact of both channel length and channel thickness on the degradation is addressed, where intensive numerical simulations are

generated to elaborate the training data set. Several membership functions are tested and the best one in terms of accuracy is selected. The predicted results fit with the numerical simulations, which allows the use of fuzzy models for analyzing the influence of interface traps jointly with the quantum confinement on the subthreshold behavior of the considered nanoscale device. Therefore, the proposed neuro-fuzzy based approach offers not only an easy to use but also a precise method to study the nanoscale devices, when subject to both hot carrier and quantum confinement effects.

In response to the high computational costs related to the development of physical based models, our aim in the third chapter is to investigate the efficiency of a new proposed framework, built upon kriging metamodeling and non-dominated sorting genetic algorithm, for the optimal design in terms of subthreshold and analog/RF criteria. The input variables of interest are limited to the geometrical parameters namely the channel length and the channel thickness. Data generated according to computer experiments are used to identify and adjust the kriging surrogate models. We emphasize that the obtained models can be used with a sufficient accuracy in a multi-objective context to offer several Pareto configurations. Therefore, a wide range of selection possibilities is available to the designer depending on the considered situation.

Since many innovative designs have been made feasible thanks to the availability of powerful TCAD tools, the idea behind the last chapter follows this vision. To this extent, we propose various modifications in order to boost the efficiency of scaled DG MOSFET device using ATLAS 2D simulator. We scrutinize the influence of several pertinent parameters including the gate work function, the molar fraction of silicon in a SiGe alloy based channel, the source/drain extensions and the channel doping profile. All these amendments are carried out in the context of digital, analog and RF applications. The carried numerical experiments reveal the key role played by TCAD tools in providing optimized circuits and opening up the door towards reliable technologic solutions under low costs.

Chapter 1

Overview of MOSFET devices : Evolution, properties and modeling

« Whether we are based on carbon or on silicon makes no fundamental difference ; we should each be treated with appropriate respect. »

Arthur C. Clarke

Abstract- *The accurate modeling of MOSFET devices for integrated circuit design has been driven for many years by the needs of digital circuit simulation. Conventional bulk CMOS technology is still prevalent in the microelectronics industry. Intel's 7 nm node has entered the development phase, where commercial mass production of chips is planned to begin during 2019. This technology will be replaced by 5 nm process around 2022. Hence, the purpose of this chapter is to provide both an overview about the evolution of MOSFET devices in addition to the basic physics theory. These aspects constitute the cornerstone to build reliable models not only for the deduction of some mandatory criteria but also for the support of new constraints.*

1.1 Introduction

In today's markets, we are experiencing an accelerated growth of mobile telecommunication/computing tools revealing the tendency towards efficient and reliable integrated circuits. The latter technology can be considered as the promising alternative to meet the mass production requirements. Indeed, this is justified by the low costs related to silicon manipulation besides the tiny scales of the obtained circuits. The emergence of the Complementary Metal Oxide Semiconductor (CMOS) paradigm in recent years instead of bipolar technology is motivated by the considerable reduction of the geometrical configuration of elementary components. The objective behind the miniaturization of CMOS devices is tightly linked to the boost of both speed and capacity of computation. The accreditation of CMOS technology has been well performed during the past progression but some drawbacks may arise by reaching the deep nanoscale level ([Taur et al., 1997] and [Isaac, 2000]).

Among the key components in building integrated circuits, Metal Oxide Semiconductor Field Effect Transistors (MOSFETs) are playing a vital role in reaching a level of flexibility and performance not seen previously. The most common investigation of MOSFET devices is practically unrestricted to a specific domain, where they are ranging from memories to portable calculators. The original proposition of the operating principle of such artifact dates back to the early 1930s. Despite the fact that the mature stage of MOSFETs has taken many decades to be reached, the subsequent developments and miniaturization as stated by Moore's law have been really magnificent. Such major interest in rendering devices smaller is due to the benefits gained by increasing the packing density, which yields improved ratio functionality/area ([Böer, 1992] and [Xie et al., 2012]).

This chapter is dedicated to the introduction and clarification of the main notions related to MOSFETs in the context of CMOS technology. The standard structure and operating mechanisms of a MOSFET in addition to its basic stages of fabrication are highlighted. A special attention is devoted to various performance metrics due to their importance for evaluating any structural modification. Furthermore, different approaches adopted in the drain current modeling of a MOSFET device are described. Besides the severe burdens generated by the downscaling process, some popular remedying strategies are also detailed. It is worthy to mention that this chapter can be regarded as the milestone for the three remaining chapters driven by our own contributions.

1.2 Historical evolution of CMOS technology

Complementary Metal Oxide Semiconductor, abbreviated as CMOS, can be considered as the basic technology adopted nowadays for realizing a wide range of integrated circuits. The widespread use of such technique can be justified by the excellent features of the obtained circuits including low power with very good signal/noise ratio on a small chip area. The fundamentals of the CMOS paradigm were proposed by Frank Wanlass while working for Fairchild Semiconductor and registered later as an US patent in 1967. Indeed, the idea stating the possibility of including discrete complementary devices (n-channel and p-channel transistors) to form a unique circuit was really innovative at that period characterized by the domination of the bipolar junction transistor substitute of the conventional vacuum tube ([Wanlass, 1967] and [Hoffmann, 2004]).

The fabrication of CMOS integrated circuits requires thin circular slices of silicon known as wafers, where each one is divided into a large number of elementary regions or dies. The size of a die is a critical parameter due not only to its impact on the device

complexity but also to the manufacturing revenue. Therefore, a wafer with a high level of dies number may include consequently a high number of integrated circuits. It should be mentioned that despite the availability of different wafer sizes, the standard wafer size today is 300 mm, which can hold about 100 of the same size die with an overall price of 6000 $. Because the degradation cost of equipments is approximately half the wafer production cost, targeting the largest available wafer sizes has been always the main concern of manufacturers when producing a new facility. The augmentation of a wafer manufacturing cost between two consecutive size generations is estimated to 30%, which results from the weak dependence of the needed resources and the wafer size [Nishi and Doering, 2007]. In Figure 1.1, we elucidate both the slicing process of the large cylindrical silicon ingot into wafers besides the final wafer product including the dies structures.

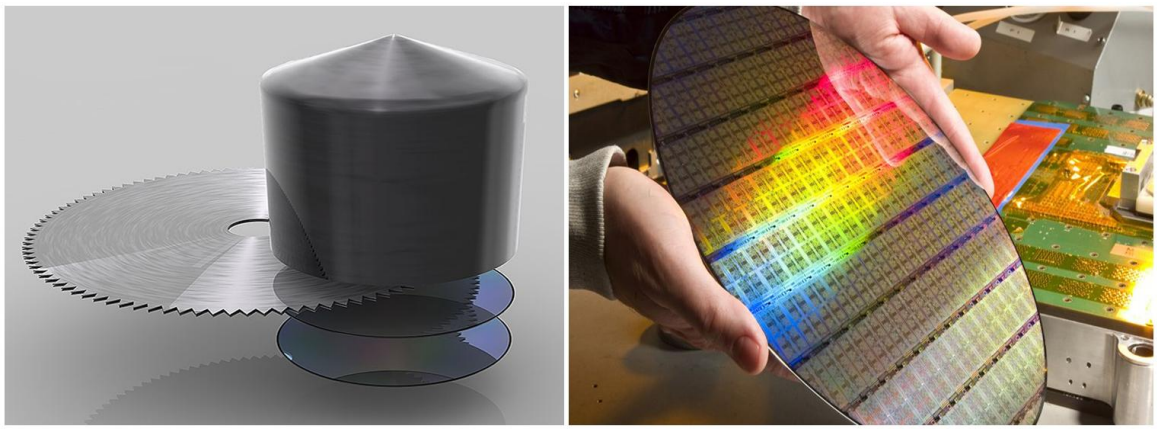

FIGURE 1.1 Visualization of silicon wafer engineering including slicing operation (picture on the right) and tracing of the dies structure (picture on the left).

The design process of CMOS circuits includes many elementary steps as depicted by Figure 1.2. The flexibility of the design procedure is a mandatory requirement, where the circuit specifications are not rigid and can be changed due to several factors for instance as a result of the trade-offs between the cost and the efficiency. However, it is not possible in almost all cases to approve major modifications on the circuit once the production process is launched [Baker, 2011].

In order to achieve higher performance in terms of the exploited chip area, the integration scale of different components has been investigated based on a modular methodology. The advent of such vision has allowed the classification of integrated circuits according to their complexity as illustrated in Table 1.1, [Fleischammer, 1996].

TABLE 1.1 Classification of logic integrated circuits by scale of integration.

Abbreviation	Designation	Gate count
SSI	Small Scale Integration	< 10
MSI	Medium Scale Integration	$< 10^2$
LSI	Large Scale Integration	$< 10^3$
VLSI	Very Large Scale Integration	$< 10^4$
ULSI	Ultra Large Scale Integration	$< 10^5$
GLSI	Giga Large Scale Integration	$< 10^6$

By moving from an integration scale to another, we obtain an enhancement of yields following a generic tendency. So, we notice that each new technology generation profits

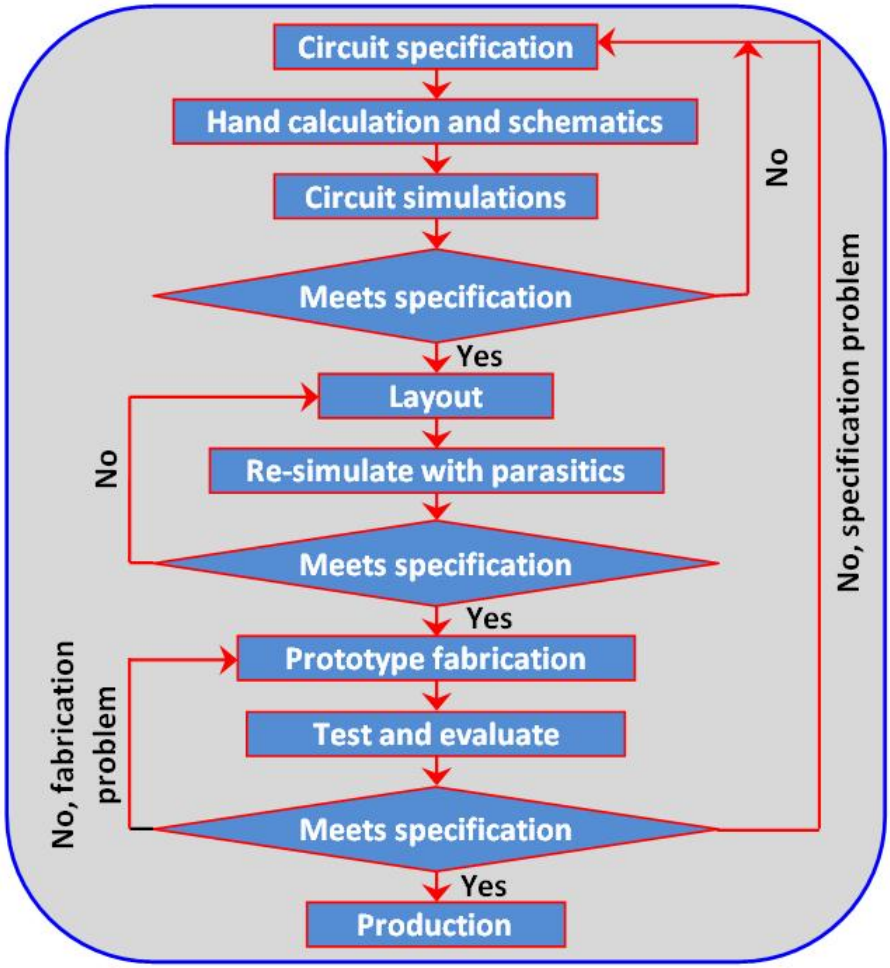

FIGURE 1.2 Illustration of different stages included within the design process of CMOS circuits.

from the advantages of elevate yield process in the previous generation. To preserve chip level components as an element for future silicon technologies, it is required that high yields be achieved at each technology generation. The simple Poisson model has been used to express the integrated circuit yield as illustrated below :

$$Y = Y_0 e^{\lambda A_c} \tag{1.1}$$

with Y_0 is the gross yield denoting the unusable wafer area near the wafer edge, λ is the defect density (defects/cm^2) and A_c represents the chip's active area.

Despite that such yield model provides a powerful mean to estimate the number of defective components as a function of the circuit area, it implicitly masks the mechanisms variability leading to defects. For this reason, the actual trend is accentuated on the employment of computer aided design and simulation allowing the yield assessment besides the sensitivity minimization of a proposed design against specific degradation effects [Tewksbury and Hornak, 1989].

The evolution of CMOS technology has been governed by Moore's law for many decades. The definition of such law assumes that the complexity level of a manufactured chip subject to minimal cost has an exponential tendency doubling in each period of time. The optimal component density for a period can be modeled mathematically as :

$$C_t = 2C_{t-1} \tag{1.2}$$

where C_t and C_{t-1} denote the component count in periods t and $t-1$, respectively. The notion of the period, known also as Moore's clock cycle, was initially set as doubling each year. Later, Moore updated this value in 1975 to show that the integration growth was slowing with a doubling period each two years, which remains constant between 1975 and 2006 [Huff, 2009]. The question regarding the ultimate limit of Moore's law remains open since it depends on the future innovations. Many believe that the downscaling process will continue indefinitely, but the physical constraints prohibit such trend. In fact, Heisenberg uncertainty or more precisely the Compton wavelength $\left(\lambda_C = \frac{\hbar}{cm_e}\right)$ imposes such threshold even with the boost of quantum computing. If electrons are considered as the smallest transistor components, the quantum limit year predicted by Moore's law is 2036. This value is in accordance with the Hawking prediction based on the speed of light and the atomic nature of matter [Powell, 2008]. Besides the quantum limit to Moore's law, the economical aspects are also an important factor that needs to be considered. The costs related to research, manufacturing and test within a new generation become more and more expensive, which is referred as Rock's law. The economic impact on the semiconductor business has been analyzed first by Karl Rupp and it was found that the slowing down due to economic reasons is likely to occur around 2020. Since any semiconductor chip is built upon a printed circuit board, more efforts have to be allocated for seeking higher bandwidths to meet the increasing performance demand of new applications. The data processing capability has pushed further the need for high speed communication across different components on the board. In summary, the immense flux of information interacting with the chip has motivated the replacement of simple input/output drivers with sophisticated high speed circuits depending in turn on reliable high bandwidth channels. So, this situation constitutes another limitation on Moore's law based on the properties of the printed circuit board [Wu et al., 2013].

An old joke has been widely quoted by experts, which states that the number of people predicting the end of Moore's law doubles every two years. This has been concretized by significant alterations affecting recently the integrated circuit industry such as the end of the International Technology Roadmap of Semiconductors (ITRS) published biannually from 1990s to 2016. The end of this roadmap has been followed by novel alternatives (for example the IEEE Rebooting Computing and the International Roadmap for Devices and Systems) dealing with enhancements at the level of architectures, devices and applications. Despite that the implemented nanoscale systems are characterized by a high computation in terms of floating point operations per second, they suffer from many harmful effects when used in biological environments like the human body due mainly to the heat dissipation. The systems interacting with body should take into account biocompatibility, autonomous operation in addition to limited instructions and energy employment from the outside. Thus, the end of Moore's law should initiate other fascinating opportunities dedicated to the investigation of nature inspired perspectives in order to balance the domination of computer science vision on today's human life [Peper, 2017].

A revolutionary vision has emerged recently concerning the use of analog transistor physics to realize neural computation by mimicking the currents in neuron ion channels, which guesses that an energy economization of about 10^4 can be obtained in comparison to traditional digital approaches. Nevertheless, the power dissipation in such neuromorphic systems did not take the scaling advantage, where the best circuits of nowadays range between 10 and 100 nW per computational cell. In the nano-CMOS era, the manufactu-

ring of large scale neuromorphic systems aimed at silicon cognition should be performed at a distinct scale of abstraction. Therefore, it is worthy to move to a more advanced level of abstraction using digital transistors rather than preserving analog devices to imitate the neuron biophysics. Such paradigm will allow the engineering of large scale, flexible silicon neural arrays by benefiting from the combination of high density digital memories and high speed digital communications interconnects [Cassidya et al., 2013].

1.3 Paradigm of MOSFET devices

Today's importance of the MOSFET technology is in a big part due to contribution in the emergence of the computer industry through its accreditation in implementing efficient integrated circuits. The need for miniaturization has highlighted the utility of integrated circuit concept, which in turn leads to a more focus on the downscaling of MOSFET device dimensions. Thus, it becomes an absolute reality that scaled devices will provide what computer applications need by more packaging densities and lower fabrication costs [Arns, 1998]. In what follows, we expose the main aspects related to the topology, operation and fabrication process associated to MOSFET devices.

1.3.1 Description and operation regimes of the standard MOSFET structure

The MOSFET device belongs to the class of devices controlled by voltage bias. The tremendous demand for digital circuits with high functionality and complexity has been among the main factors contributing to the development and success of the MOSFET design, which becomes more beneficial in terms of fabrication technology, chip area and reduced input current at the level of the control terminals. Therefore, the tendency was oriented to the MOSFETs rather than bipolar junction transistors for the fabrication of digital chips with high packing density besides better power efficiency. The basic material used in modern MOSFETs is Si jointly with SiO_2 in film form, where three terminals are included : source, drain and gate constituting the simple structure despite the availability in practice of other complex designs. The drain current modulation in the case of a MOSFET is carried out by the field effect of the gate electrode aiming at controlling the carriers in the inversion channel formed in the vicinity of the semiconductor/ insulator interface between the drain and the source terminals. More precisely, the carrier transport channel is two dimensional and formed parallel to the surface of the wafer. The operation of the device is governed by the lateral dimensions of the gate electrode especially the gate length. Since the gate oxide SiO_2 is subject to severe electronic quality constraints, it must support electric field values allowing the creation of the inversion layer at the Si/SiO_2 interface, which is equivalent to turning the device to the "ON" state. The magnitude of the interface electron states should be less than 10^{12} cm^{-2} eV^{-1} to ensure a reliable operation for the MOSFET device ([Huff and Gilmer, 2005] and [Valizadeh, 2016]). Figure 1.3 represents the standard design of an n-channel MOSFET, appropriate adaptations regarding the doping status of the different semiconductor regions or the polarity of bias conditions are possible.

The operation mechanism of MOSFETs is influenced by the characteristics of MOS capacitors. The availability of minority carriers in the proximity of the semiconductor side of the MOS capacitor allows the gate electrode to bias the MOS capacitor under the regimes of operation. There are three regimes to be considered namely, accumulation mode, de-

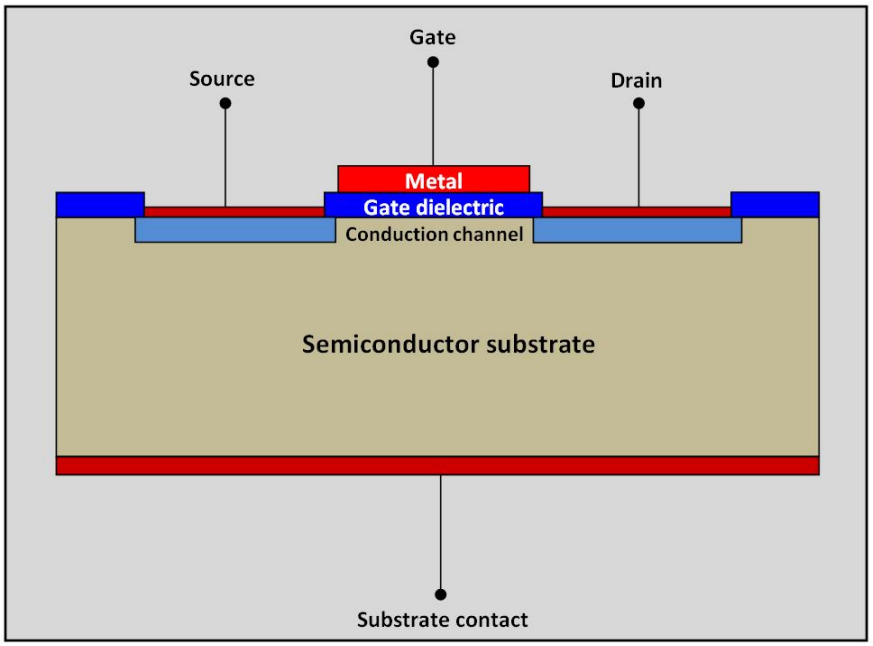

FIGURE 1.3 Schematic illustration of a typical MOSFET device.

pletion mode and inversion mode in the C-V characteristics of the MOS capacitor as detailed below [Bentarzi, 2011] :

- **Accumulation mode :** The application of an external voltage V_g to the silicon surface of the MOS capacitor leads to the change of the carrier densities in the surface region. In case of applying high negative bias to the gate, we obtain the formation of an accumulation layer due to the attraction of holes by the negative charges. The large concentration of holes will result in turn in the creation of a second electrode of a parallel plate capacitor with respect to the first gate electrode. Since the accumulation layer can be considered as an indirect ohmic contact with the p-type substrate, the capacitance of the device subject to the accumulation regime must be roughly equal to the capacitance of the oxide given by :

$$C_{ox} = \frac{\varepsilon_0 \varepsilon_{ox}}{t_{ox}} \tag{1.3}$$

  with ε_0 represents the permittivity of the free space, ε_{ox} is the relative permittivity of the oxide and t_{ox} refers to the oxide thickness. It should be mentioned that such capacitance remains constant with bias V_g during the maintain of the device in accumulation mode. Furthermore, it is independent of the frequency as long as the motion of the majority carriers, which is true when the applied small signal frequency is less than the reciprocal of the dielectric time constant of silicon (10^{11} Hz). The Fermi level near the silicon surface will approach under this condition the valance band edge.
- **Depletion mode :** Once the negative charges are eliminated from the gate, the holes leave the accumulation layer gradually until the entire silicon becomes neutral and the associated applied gate bias is known as the flat band voltage. The increase of positive bias on the gate with respect to flat band repels the holes, which generates at the surface a depletion region of carriers. Hence, under the depletion regime, the Fermi level close to the silicon surface will approach the center of the forbidden region. The increase of the positive voltage V_g results in the augmentation of

the surface depletion region width (X_D), while the capacitance from the gate to the substrate decreases. This behavior can be explained by the addition of the capacitance assigned to the surface depletion region in series to the capacitance across the oxide. The capacitance is independent of frequency because only majority carriers contribute to the substrate charge. The full capacitance per unit area under depletion conditions from the gate to substrate is expressed by :

$$C(V_g) = \left(\frac{1}{C_{ox}} - \frac{1}{C_s(V_g)} \right)^{-1} \tag{1.4}$$

with $C_s(V_g) = \frac{\varepsilon_0 \varepsilon_s}{X_D}$ indicates the silicon capacitance per unit area and $X_D = \sqrt{\frac{2\varepsilon_0 \varepsilon_s \psi_s}{qN_A}}$. The applied gate voltage V_g and the total band banding ψ_s are related by the formula :

$$V_g = \psi_s + \frac{\sqrt{2\varepsilon_0 \varepsilon_s q N_A \psi_s}}{C_{ox}} \tag{1.5}$$

- **Inversion mode :** As a result of increasingly applying positive voltage, the surface depletion region will get enlarged continuously until the observation of the surface inversion (n-type) onset characterizing the formation of the inversion layer. The Fermi level near the silicon surface will now be positioned close to the bottom of conduction band. The inversion layer is very thin of the order of few nanometers and is separated from the silicon bulk by the depletion layer. The inversion layer buildup is a threshold phenomenon marked by the equality of the minority carriers and doping concentrations. The depletion layer width attains a limit at the inversion beginning denoted by X_{DLim}. Under the inversion conditions, the capacitance will depend on the frequency because the charge density in the inversion layer may or may not be able to follow the AC variation of the applied gate voltage.

In Figure 1.4, we elucidate biasing condition, charge distribution and energy band diagram associated to the three regimes (accumulation mode, depletion mode and inversion mode).

If no gate voltage is applied, the source to drain electrodes can be identified to two $p-n$ junctions connected back to back. The reverse leakage current is the only current flowing from the source to the drain. The application of a sufficiently large positive bias to the gate results in the inversion of the MOS structure and the formation of a surface inversion channel between the highly doped (n^+) regions. A large current can flow through the formed conduction surface connecting the source to the drain, where the channel conductance can be adjusted by varying the applied gate voltage. The variation of the drain current as a function of the gate or drain voltage is characterized by different operating regions ([Sze and Lee, 2012] and [Wallmark and Carlstedt, 2013]) :

- **Subthreshold region :**When the semiconductor surface is weakly inverted and the applied gate voltage is less than the threshold voltage, the associated drain current is known as the subthreshold current. It is dominated by the diffusion instead of the drift component and is developed in the same manner as the collector current in a bipolar transistor with homogeneous base doping. Since the surface potential ψ_s is approximately equal to $V_g - V_{th}$, the drain current decreases exponentially when V_g becomes less than V_{th} according to the formula $I_d \sim e^{q(V_g - V_{th})/kT}$. The subthreshold region plays a vital role when the device is adopted as a low voltage, low power component, which is the case of a switch in digital logic and memory applications. Such

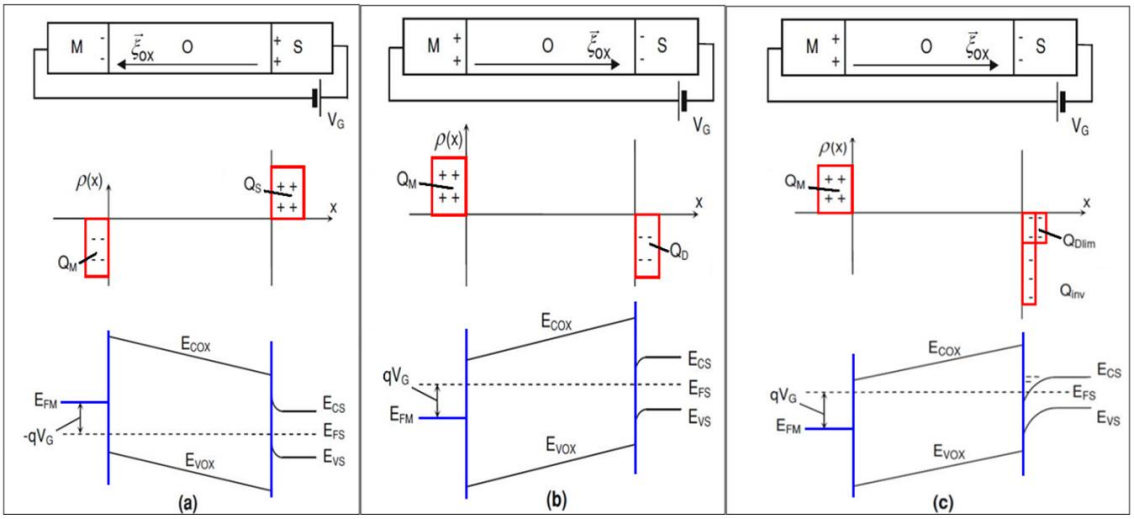

FIGURE 1.4 Representation of the structure features namely biasing condition, charge distribution and energy band diagram associated to (a) the accumulation mode (b) the depletion mode (c) the inversion mode.

importance is consolidated by the fact that the subthreshold region showcases the switching ability between the "ON" and "OFF" states.

- **Linear and Saturation regions :**We assume the existence of an inversion layer at the semiconductor surface due to the application of a voltage on the gate. For small values of the applied drain voltage, electrons will move from the source side to the drain side via the conducting channel. Therefore, the channel in this case will act as a resistor (constant-resistance) and the drain current is linearly proportional to the applied drain voltage, which refers to the linear region. By increasing the drain voltage gradually, it reaches a value V_{Dsat} for which the inversion layer thickness vanishes at a point called the pinch-off point near the drain side. Beyond this point, the drain current remains basically fixed because at the pinch-off point, the voltage V_{Dsat} remains the same even for $V_D > V_{Dsat}$. Hence, the number of carriers reaching this point from the source remains unchanged. Since the drain current is constant independently of the increase of the drain voltage, this region is known as the saturation region. The carriers are injected from the pinch-off point into the drain depletion region in an analogue manner to the carriers' injection from an emitter-base junction to the base-collector depletion region of a bipolar transistor. The summary of various regions is depicted in Figure 1.5, where the linear region is just around the origin, the region below the pinch-off is to the left of the broken curve and last, the saturation region is delimited to the right of the broken curve.

1.3.2 Fabrication process of MOSFET devices

We elucidate in this subsection the common elementary stages that are routinely used in combination during the fabrication process of a modern device or integrated circuit ([Pierret, 1996] and [Hu, 2009]) :

- **Oxidation of silicon phase :** The dioxide of silicon (SiO_2) has been widely used for various objectives such as serving to mask the introduction of doping into silicon and an isolator in the metal oxide semiconductor device. SiO_2 serves as both an

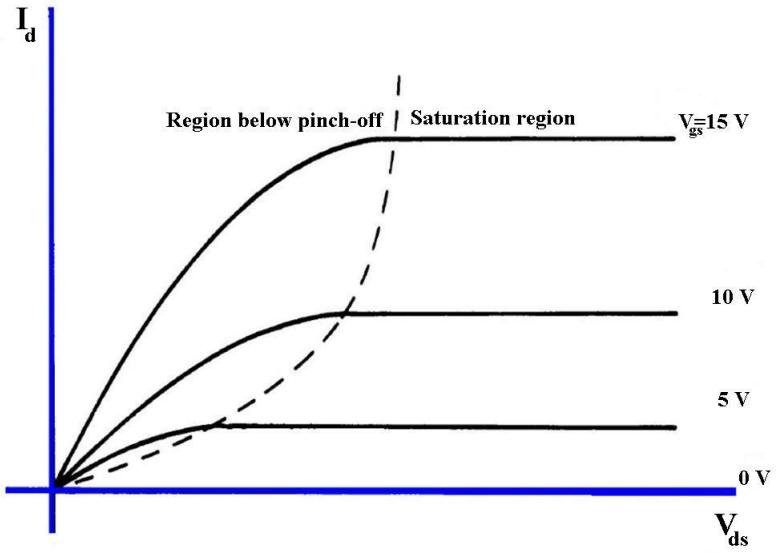

FIGURE 1.5 Variation of the characteristic curve attached to a MOSFET device for different applied gate voltages.

insulator for some device designs and as a barrier to diffusion during the fabrication procedure. The generation of SiO_2 layers with definite thickness can be realized through the reaction of Si with oxygen gas or water vapor (referred as dry and wet oxidation, respectively) at a high temperature. In both cases, the oxidizing materials may diffuse through the existing oxide to form more SiO_2 at the interface of Si and existing oxide. These reactions are governed by the equations given by :

$$\begin{cases} Si + O_2 \longrightarrow SiO_2 \\ Si + 2H_2O \longrightarrow SiO_2 + 2H_2 \end{cases} \tag{1.6}$$

The dry oxidation is used to form thin oxide films while the wet oxidation is preferred in obtaining thicker oxide layers, which can be justified by the fact that water vapor diffuses through SiO_2 faster than oxygen. The oxidation is carried out in a horizontal or a vertical heater at elevated temperatures of 700-1200 C produced by electrical resistance heating coils. It is worthy to mention that the thickness of the grown oxide is affected by many factors namely the heater temperature, the oxidation time, the ambient gas and the orientation of the Si surface. Furthermore, the wafers are mostly cut in the (100) plane for twofold reasons : the reduced interface trap density due to the low unsaturated bonds density in this plane in comparison to other planes and the high electron surface mobility.

- **Lithography phase :** The selective elimination of the oxide from regions in which dopant atoms are to be introduced is realized by a procedure known as photolithography or optical lithography. First, the wafer top surface is covered with a material called photoresist sensitive to ultraviolet (UV)light. A few drops of the liquid photoresist are deposited on the wafer, which is rotated at high speeds in order to produce a thin and homogeneous coating. A short bake at about 90 C after rotating is then achieved to drive solvent out of the resist. The next step consists in exposing the resist to ultraviolet light through a high precision reduction lens system and a photomask formed by a quartz photoplate including a copy of the patterns to be transferred to the SiO_2 film. While the opaque regions on the mask block the ultraviolet light, the photoresist regions exposed to the light are subject to chemical

reactions depending on the resist kind being used. We distinguish two types of resist : negative resist and positive resist. For the negative resist, the light strikes areas becoming polymerized and more difficult to dissolve in solvent. So, the polymerized regions remain once placed in a developer (solvent) whereas the unexposed regions dissolve and clean away. In case of the positive resist, a stabilizer is used to slow down the dissolution rate of the resist in the solvent. Such stabilizer clears up when exposed to light, giving a preferential removal of the exposed regions. Among the solutions to extend the resolution limit is to employ ultraviolet light with shorter and shorter wavelengths, where laser light sources of 248 and 193 nm are commonly adopted. But it is difficult to go beyond due to the lack of appropriate transparent materials for lenses and mask plates at this wavelength. With the aim of obtaining good optical resolution, only a limited surface (lithography field) of about 10 cm^2 in the wafer is exposed. The exposure phase is repeated iteratively for neighboring areas by moving the wafer until it is all exposed. It is mandatory to state that lithography is the most complex and costly procedure among all fabrication stages of the electronic devices, where a standard integrated circuit fabrication flow applies this technique more than twenty times using various photomasks.

- **Pattern transfer by etching phase :** Following the formation of the pattern in the resist by lithography, the generated resist pattern is often transferred to an underlying film such as SiO_2. The etching method based on the elimination of SiO_2 by Hydrogen Fluoride (HF) is called wet etching. The etched features are in general larger than the dimensions of the resist patterns because of the isotropy of the wet etching. In the other hand, the dry etching method (known also as plasma or reactive ion etching) can remedy this drawback and is the principal etching technology, where the wafer with patterned resist is exposed to plasma often created by a radio frequency electric field. A chemical reaction is obtained between the energetic species and the exposed regions of the material to be etched, whilst the ions in the plasma bombard the surface vertically and push away films of the reaction products on the wafer surface. It should be mentioned that this action is directional, which means that the etching is preferentially vertical due to the possibility of covering the vertical surfaces with the reaction products films. Consequently, it can be concluded that the etch rate is anisotropic. It is possible to get approximately vertical walls in the etched material if the reactor design and etching chemistry are properly chosen. The adoption of low pressures and highly one directional electric field leads to anisotropic etching. However, the dry etching can be guided towards isotropic or partially anisotropic features. Appropriate gases are needed in the etch chamber depending on the material to be etched. For instance, silicon and its compounds are etched by plasmas including fluorine, while aluminum can be etched by plasmas containing chlorine. Dealing with plasma during this phase can cause damage (known as plasma process induced damage or wafer charging damage) to the devices on the wafer. The principal damage mechanism is the charging of conductors by the ions in the plasma, giving an overly high voltage across a thin oxide and generating oxide breakdown. The sensitivity of such damage to the conductor size is known as the antenna effect.
- **Doping phase :** Generally, the density profile of the dopant atoms in the silicon is performed in two stages. The dopant atoms are first positioned on or near the surface of the wafer by one of the following techniques : ion implantation, gas-source doping, or solid-source diffusion. Due to the accurate control provided by the ion implantation, it can be considered as the paramount method. The introduction

of impurities into the semiconductor is achieved through ionizing and accelerating the impurities to high energies (from subkiloelectronvolt to megaelectronvolt) and then laterally bombarding the semiconductor surface. Since the implanted ions may result in some problems (displacement of the semiconductor atoms or wafer charging damage), a follow-up heating of the wafer and the introduction of electrons for charge neutralization is recommended. The gas-source doping is used in practice only for doping purpose of silicon with phosphorus, which is carried out in a furnace analogous to that used for oxidation. In solid-source diffusion technique, the Si surface is first coated with a thin layer (such as SiGe alloy) including dopant atoms as placed or because of subsequent dopants implant into this film. So, the dopants are diffused into the silicon and the employed SiGe thin film has to be removed later by wet etching.

- **Dopant diffusion phase :** In order to drive the dopants deeper into the semiconductor after dopant introduction via implantation or gaseous deposition, the diffusion stage is needed in contrast to the unwanted diffusion occurring at the post-implant annealing. The diffusion of dopant impurities is conducted with time at high temperature since it increases the diffusion rate. In case of opposite doping type between the diffusion dopant and the substrate, a line may be formed indicating the boundary where $N_a = N_d$. Such structure is called a $p-n$ junction with the thickness of the diffusion layer as the junction depth. The elaboration of very deep or shallow junctions is a mandatory requirement in some specific applications. The impurity concentration shape as a function of position inside the semiconductor can be observed to have a Gaussian profile given by :

$$N(x,t) = \frac{N_0}{\sqrt{\pi D t}} e^{-\left(\frac{x}{2\sqrt{Dt}}\right)^2} \tag{1.7}$$

where N_0 is the number of dopants per square centimeter, x denotes the distance with respect to the semiconductor surface, D represents the diffusivity for the considered impurity and heater temperature, and t designates the time for the diffusion stage. Diffusion temperatures vary from approximately 900 C to 1200 C. The surface doping concentrations generated may reach $10^{21}/cm^3$, which makes the surface region similar to the n- or p-type property of the diffusing atoms. The term diffusion can be used sometimes to indicate the joint process of gaseous dopant deposition and diffusion.

We depict in Figure 1.6 a recapitulative flowchart of the device fabrication process including the main phases namely (a) the formation of the SiO_2 film, (b) the selective removal of definite regions, (c) the introduction of dopant atoms into the wafer surface, and (d) the dopant diffusion into the semiconductor.

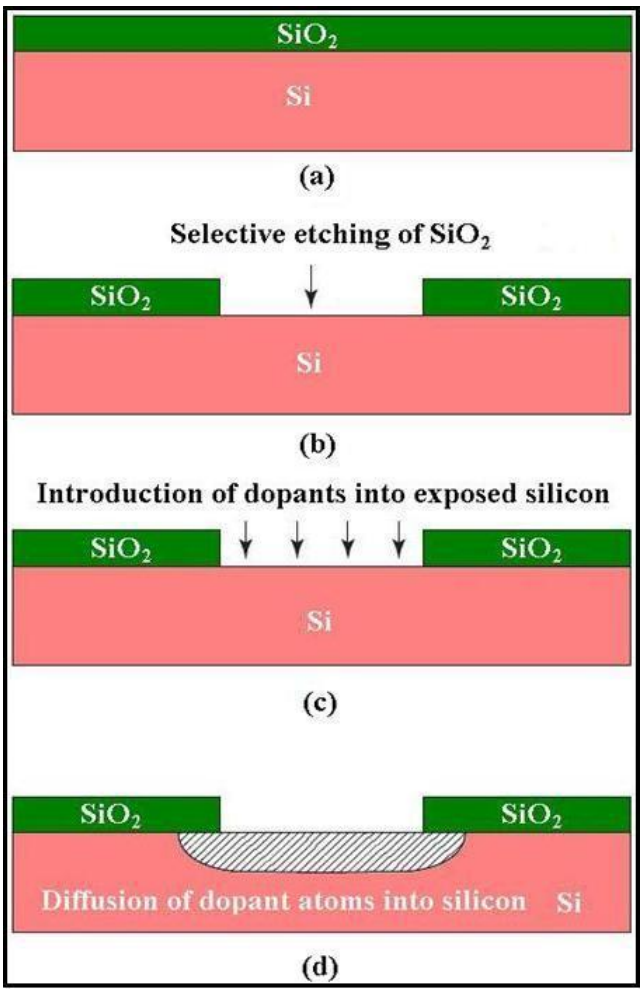

FIGURE 1.6 Summary of some principal steps during the fabrication of a silicon device (a) silicon oxidation (b) selective remove of the oxide (c) introduction of dopant atoms (d) dopant atoms diffusion into silicon.

1.4 MOSFET performance measures

In order to evaluate the performance of a MOSFET device when used in a well defined context, a set of metrics are needed not only in order to approve its validity under the given conditions but also to conduct comparative analyses with respect to other alternatives. In what follows, we provide some of the widely used criteria in MOSFET application fields.

1.4.1 Subthreshold parameters

Usually, the subthreshold parameters are recognized to be of a great importance during the design of digital circuits. Amongst the large number of the subthreshold parameters, we mention the following :

- **Threshold voltage :** The threshold voltage can be defined as the applied gate voltage needed to reach the threshold inversion configuration, which is considered as the condition for which the surface potential equals $\phi_s = 2\phi_{fp}$ and $\phi_s = 2\phi_{fn}$ for the p-type and the n-type channels, respectively. The control of the threshold voltage by the designer is a hard task since it depends on many factors mainly process and physical parameters [Neamen, 2011]. In this extent, the threshold voltage should satisfy

the compromise between low and high values in order to provide high current drive besides reducing the leakage current when the device is in the subthreshold regime. Due to the utilization of N-polysilicon as a gate for both types of MOS devices, the threshold voltage of a PMOS transistor is in general higher in magnitude than is desirable. In order to reduce the threshold voltage value, a boron implant is achieved in the channel region with enough high doses to overcompensate the N-well surface. The reduction of the effective channel length leads in turn to the reduction of the threshold voltage, where a fraction of the bulk depletion charge is created in the case of short channel MOS devices by the source and drain junctions. Therefore, smaller gate voltages are sufficient to trigger the surface inversion [Gu et al., 1996]. For submicron channel lengths, the threshold voltage may decrease by 0.3V. However, very low threshold voltages are not desirable since they blur the discrepancy between "OFF" and "ON" states, and make logic circuits more sensitive to noise [Annaratone, 1986]

- **Drain induced barrier lowering :** When the drain voltage is high and the gate voltage is less than the threshold voltage, an undesirable leakage current flowing below the surface of the channel can be detected and this additional leakage current is due to what is called Drain Induced Barrier Lowering (DIBL) [Gu et al., 1996]. The DIBL is related to short channel transistors and consists of source injection in the surface of the channel, which causes the threshold voltage lowering and the increase of the subthreshold current. Note that the p-channel device is more affected by DIBL than its n-channel counterpart. It is worthy to mention that long channel devices mark a constant potential peak under the channel from the drain to the source and such potential barrier is basically independent of the drain-to-source voltage. For short channel devices, the distance between the two diffused regions cannot accommodate the two depletion regions and the peak potential is no longer constant along the channel. By increasing the isolation between the two diffused regions using higher substrate concentration, we can obtain a reduction in the probability of DIBL due to the decrease of the likelihood of electric fields coupling from source to drain. Nevertheless, higher concentrations increase the electric field at the drain side resulting in hot electron injection. The same behavior is associated to the oxide thickness, where decreasing this parameter limits the magnitude of DIBL, but makes hot electron injection more pronounced. Thus, to avoid both these effects, conflicting constraints must be met [Annaratone, 1986].
- **Swing factor :** For a uniformly doped transistor, the variation of the drain current as a function of the gate voltage is approximately a straight line. The reciprocal slope of this line defines the gate switching ability required to decrease the subthreshold current value to any desired level. The swing factor, known also as the subthreshold swing (S), is the gate voltage swing required to reduce the drain current by one decade. The subthreshold swing is given in mV/decade by :

$$S = 2300\left(\frac{kT}{q}\right)\left[1+\frac{C_d}{C_{ox}}\right] \tag{1.8}$$

where C_d is the depletion layer capacitance of the source/drain junctions. For example, if S = 90 mV/decade, then the gate voltage must be reduced by 0.55V to bring the drain current from 1µA down to 1pA giving very low power dissipation in "OFF" state, which justifies the advantages of small swing factor values. The subthreshold swing increases with the scaling of the channel length. On the other hand,

reducing the oxide thickness and/or the substrate doping concentration results in the decrease of this criterion [Gu et al., 1996]. Furthermore, in order to get a fast turn-off, shallow implantations should be adopted [Sze, 1981].

- **Subthreshold current :** Under ideal conditions, the drain current flows only when the applied gate voltage exceeds the threshold voltage value, which is often the behavior of long channel devices. Though, even in this case there is a very weak drain current at gate voltages less than the threshold value. So, the subthreshold current flows primarily when the transistor is in the weak inversion regime. This subthreshold current becomes significant for short channel devices. The mechanism responsible for the subthreshold current differs between long channel and short channel devices. In long channel devices, the phenomenon is similar to that in a bipolar junction transistor, where the source, the drain and the p-type substrate play the roles of emitter, collector and base, respectively. The drain voltage drops quasi totally across the drain-substrate depletion region. Consequently, the electric field component parallel to the semiconductor/insulator interface is small, and the diffusion component of the subthreshold current is dominant [Shur, 1990].

1.4.2 Small signal parameters

On the other hand, the small signal parameters are needed when efforts are focused on the design of analog circuits. The two basic parameters employed for this purpose extensively in literature are explained below :

- **Transconductance :** Given that a MOSFET device in saturation regime generates a current due to the application of a gate-source overdrive voltage, it is interesting to define a figure of merit indicating the device efficiency in converting the voltage to the current. In particular, dealing with the changes of voltages and currents in processing signals motivates the definition of this figure in terms of the change in the drain current divided by the change in the gate-source voltage. This figure of merit is called the transconductance (g_m) and is formulated as :

$$g_m = \left.\frac{\partial \mathrm{I}_d}{\partial \mathrm{V}_g}\right|_{\mathrm{V}_{ds}=Cst} \tag{1.9}$$

  Hence, the transconductance is a sort of sensitivity measure of the device in the sense that for a high g_m, small alterations in the gate-source voltage provides large changes in the drain current. For this reason, it is widely common in analog design field to say that a MOSFET operates as a V/I converter to point out that it transforms a voltage change to a current change [Razavi, 2017].

- **Cut-off frequency :** The unity current gain frequency can be identified to the signal input frequency at which the extrapolated small signal current gain of the MOSFET equals one. The small signal current gain is defined as the amplitude of the small signal drain current to the small signal gate current. The unity current gain frequency is denoted by f_T and is read as the transit frequency or the cut-off frequency, which is used as a benchmark to assess the intrinsic device speed. While the performance of the complete device including the additional effects of the extrinsic parasitics, is always inferior. The expression of the cut-off frequency can be written as :

$$f_\mathrm{T} = \frac{g_m}{2\pi \mathrm{C}_{gs}\sqrt{1+2\frac{\mathrm{C}_{gd}}{\mathrm{C}_{gs}}}} \tag{1.10}$$

The basic decisive factors influencing the unity current gain frequency are the device transconductance g_m, the gate-to-source C_{gs} and the gate-to-drain C_{gd} capacitances. During the analog circuit design procedure, it is possible to trade power for speed by augmenting the device bias current and therefore g_m because of the direct proportionality of the cut-off frequency and the transconductance. MOSFET devices used at radio frequencies are usually operated in saturation since the transconductance is maximum for a defined device in the saturation regime. Despite that g_m can be enhanced by increasing the width of the device, this will not improve f_T because the parasitic capacitance C_g proportional to the width of the transistor balances any resulted enhancement ([Golio, 2002] and [Tsividis and McAndrew, 2012]).

1.4.3 Linearity parameters

The actual need of nowadays systems in terms of linearity is becoming highly demanded for wideband RF applications like mobile terminals. Despite the availability of system level strategies to boost the linearity aspect, but their deployment is based on complex circuitry. Thus, an analysis of the linearity behavior at the device level as a function of important design parameters is more suitable. In order to assess the RF linearity performance and intermodulation distortion of MOSFET devices, the aforementioned measures such as the threshold voltage or the subthreshold slope are not appropriate criteria to be accredited [Kumar et al., 2011]. The modeling of nonlinearity in a MOSFET with an AC biasing gate voltage V_g and producing a drain current I_d can be treated as a time variant nonlinear system. Therefore, the Taylor expansion of the drain current formula yields :

$$I_d = I_0 + g_{m1}V_g + g_{m2}\left(V_g\right)^2 + g_{m3}\left(V_g\right)^3 + ... \tag{1.11}$$

with the expansion factors are given by $g_{mn} = \frac{1}{n!}\frac{\partial^n I_d}{\left(\partial V_g\right)^n}$. It should be highlighted that even for a pure sinusoidal input, the MOSFET produces many harmonics besides a DC term, generating losses in useful output power. The presence of the first and the third transconductance derivatives constitutes a dominant nonlinear source, which is cumbersome for RF systems because it results in the distortion of the fundamental amplitude via signals in the adjacent bands [Ma and Kaya, 2004]. Among the numerous metrics used for this purpose, we have the extrapolated gate voltage amplitude at which the second harmonic becomes equal to the fundamental tone in the device's drain current (denoted by VIP2), the extrapolated gate voltage amplitude at which the third order harmonic becomes equal to the fundamental tone in the device's drain current (given by VIP3), the extrapolated input power at which the first and the third harmonic are equal (expressed by IIP3) and the third order intermodulation distortion (symbolized as IMD3). The analytical formulas of these measures are indicated below [Ghosh et al., 2012] :

$$\text{VIP2} = 4\frac{g_{m1}}{g_{m2}} \tag{1.12}$$

$$\text{VIP3} = \sqrt{24\frac{g_{m1}}{g_{m3}}} \tag{1.13}$$

$$\text{IIP3} = \frac{2}{3}\frac{g_{m1}}{g_{m3} \times R_s} \tag{1.14}$$

and

$$\mathrm{IMD3} = \left[\frac{9}{2} \times (\mathrm{VIP3})^2 \times g_{m3}\right]^2 \times \mathrm{R}_s \tag{1.15}$$

To gain high linearity and low distortion for the device operation, we need as high as possible values for VIP2, VIP3 and IIP3; and low value for IMD3. Furthermore, the singularity in VIP3 should take place at a lower gate voltage, where it is desirable to bias the amplifier at lower gate voltage for obtaining good features including high power gain, high linearity and low DC power consumption [Ghosh, 2013].

1.5 Various approaches of MOSFET modeling

The compactness and accuracy of MOSFET models are needed to satisfy present day requirements of circuit designs on semiconductors. A compact MOSFET model has been defined by the IEEE Standard Working Draft as a set of mathematical formulas including parameters used as inputs to a circuit simulator. The obtained model is assumed to be able to generate the device characteristics when subject to diverse configurations such as device dimensions, temperature ranges and a variety of operating conditions. To make these models acceptable by a wide category of users, they must match well defined application constraints. Meeting such specific needs of applications jointly with the respect of new technology nodes results is in many cases intractable for the model development procedure. For example, an accurate drain current prediction and charge conservation is the cornerstone in digital design. While for analog design, ensuring more severe conditions (continuity of current and its derivative in the entire region of operation) are mandatory. Furthermore, noise and distortion analysis are of interest for high frequency RF applications. There is no doubt that the precise analytical and compact device models are essential prerequisite for computer aided design activities such as design or optimization since they reflect with a high degree of precision the device response via a collection of modeling parameters adequate as inputs to circuit simulators. In this regard, different approaches have been developed to MOSFET modeling. The current available models are divided into three broad categories namely surface potential based models, charge based models and threshold voltage based models as detailed below [Bhattacharyya, 2010].

1.5.1 Threshold voltage based models

The threshold voltage based model is a piece-wise approach dealing with the MOSFET behavior in weak and strong inversion discriminated by the threshold voltage value. Such model has the advantage of introducing a major simplifying phase over the computationally intensive Pao-Sah model. However, the high number of model parameters is often considered as a key drawback for this approach despite it remains an accredited industry standard. Threshold voltage based models are elaborated on the local representation of the device operation in both strong and weak inversion. The approach belongs to the ones of the earliest models introduced into the SPICE circuit simulator. In this background, the resolution of the following equation is provided :

$$\mathrm{I}_{ds} = -\mu_{neff}\frac{\mathrm{W}}{\mathrm{L}}\int_0^{\mathrm{V}_{ds}} \mathrm{Q}_n^{'} d\mathrm{V}_{\mathrm{CS}} \tag{1.16}$$

where $Q_n^{'}$ is expressed in terms of the channel potential and the threshold voltage. The integrals in the drain current equations have been evaluated by assuming a charge sheet approximation leading to major simplifications. In general, the threshold voltage based modeling approaches are characterized by :

- The device operation is modeled with respect to the source,
- Interpolating functions are incorporated to guarantee continuity for current and its derivatives expressions at the transition regions of operation,
- The availability of these models for a wide range of technological modifications associated to MOSFETs.

1.5.2 Charge based models

This approach is based on the formulation of the drain current as a function of the inversion charge density at the source and drain sides of the channel. It is a very appropriate methodology when dealing with low power designs. Moreover, such class of models is physics based with a limited requirement of empirical fitting and a relatively easy extraction procedure of parameters. In the charge based modeling approach, the focus is on the simplified but precise description of the inversion charge, where the general drain current expression is written in terms of the inversion charge density as the variable in the integral :

$$I_{ds} = -\mu_{neff} \frac{W}{L} \int_{Q_{nS}^{'}}^{Q_{nD}^{'}} Q_n^{'} \frac{dV_C}{dQ_n^{'}} dQ_n^{'} \tag{1.17}$$

Among the features of this class of models we have :

- Most of the charge based models are developed with the bulk as the reference, which makes it consistent with the physical symmetry of the device,
- The continuity of both current and its derivatives in transition from different operation modes is usually ensured through mathematical interpolation functions,
- An analytical compact expression for small signal parameters in any operating region can be obtained by the inversion charge density linearization, which is mainly suitable for analog design,
- Some charge based models use numerical algorithms to accurately model the channel charge.

1.5.3 Surface potential based models

The drain current expression is developed based on the surface potential at the source and drain ends of the channel. The computation burdens related to the surface potential calculation has been remedied by efficient analytical or numerical algorithms. In addition, this method does not adopt a local modeling approach, where it considers drift-diffusion current transport providing a single piece formulation for the current offering derivatives continuity over the whole operating regimes. Due to their physics based nature, the surface potential based models have showcased many benefits in RF circuit simulation and have supported well the scaling phenomena. The surface potential approach starts by computing the surface potential at the source and the drain ends of the

channel based on the applied terminal voltages. Then, the channel charge depends on the surface potential at a given channel point from which expressions of current are obtained. So, the integrating variable (channel potential V_C) is replaced by the surface potential ϕ_s as follows :

$$I_{ds} = -\mu_{neff} \frac{W}{L} \int_{\phi_{sS}}^{\phi_{sD}} Q_n'(\phi_s) \frac{dV_C}{d\phi_s} d\phi_s \tag{1.18}$$

The surface potential based models have the given properties :

- Most of surface potential based models consider the substrate/bulk as the reference terminal ensuring symmetric device operation,
- The surface potentials at the source and drain ends are calculated either analytically or numerically for given terminal voltages,
- The approach has the capability to support various phenomena resulting from the technology scaling,
- The remarkable benefit of this modeling approach is its inherent potential to accurately account for the continuity of the drain current and its higher derivatives along the full range of the device operation.

It should be highlighted that the integrals in the drain current equations for the three modeling approaches are evaluated by supposing a charge sheet approximation, which leads to major simplifications. In Figure 1.7, different approaches implemented for the MOSFET modeling are summarized.

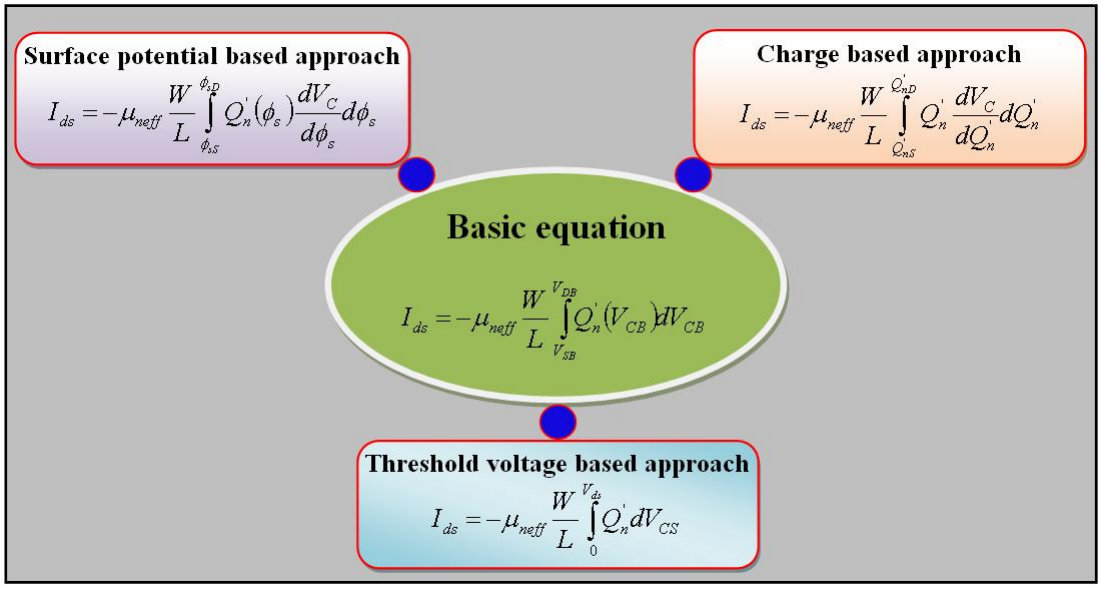

FIGURE 1.7 Dependence of various approaches for the drain current modeling on the transport equation.

1.6 Downscaling of MOSFET devices and degradation effects

An obvious advantage of reducing a MOSFET device concerns the increase of components density of packing. However, such miniaturization is not offered for free but nume-

rous parasitic effects arise, which if not treated appropriately will result in the degradation of the device performance. We present below different parasitic effects related mainly to the reduction of the geometrical configuration of the transistor.

- **Gate leakage :** Since the gate oxide thickness has been extremely reduced in recent MOSFET structures, the direct tunneling becomes a vital concern for designers due to its dependence on both "OFF" and "ON" operation states. Indeed, this phenomenon may result in a significant standby leakage power dissipation, which motivates the development of design procedures for reducing the total leakage power. The impact of both gate induced drain leakage and band to band tunneling is negligible in comparison to the gate tunneling current at bias voltages within the practical ranges of miniaturized MOSFET devices. The gate tunneling current includes three basic components namely the gate to bulk leakage current, the current between the gate and the source-drain extension overlap and the gate to channel tunneling partitioned among the source and the drain terminals. The origin and magnitude of these components differ according to the operation mode and type of the device, where they are controlled by carrier conduction processes. It should be highlighted that various mechanisms such as electron tunneling from conduction band, electron or hole tunneling from valence band can be preponderant relying on the device type and supply voltages [Drazdziulis and Larsson-Edefors, 2003].
- **Parasitic capacitances :** The influence of parasitic capacitances on the MOSFET operation at high frequencies cannot be neglected, where these capacitances are divided into two categories : extrinsic capacitances and intrinsic capacitances. Extrinsic capacitances expressed using seven small signal capacitances between each pair of transistor terminals are assigned to different regions of the device's geometry. On the other hand, the intrinsic MOSFET capacitances are considerably more complicated since they depend strongly not only on the applied terminal voltages but also on the field distribution within the device. These terminal capacitances can be calculated by considering the ratio of the charge change associated with each terminal to a change in the voltage at another terminal with fixed voltages at all remaining terminals. It should be stated that these capacitances are assessed in terms of the device operating regime depending in turn on the terminal voltages, where simplified expressions can be obtained for the linear and saturation regions [Golio, 2002].
- **Random dopants fluctuation :** When the scaling process of MOSFET devices reaches sub 32 nm technology nodes, the doping concentration in the channel approaches 10^{18} cm^{-3} or even higher. A heavily doped channel region results in intrinsic random fluctuations for miniaturized MOSFETs to the deep nanometer scale. Since the mean number of dopant atoms in a channel region of a micron-scale sized device is in the order of thousands, the variation effects in the number of dopants are not considerable. However, there are only a few tens of dopant atoms in the channel region of nanoscaled MOSFETs. For instance, let's consider a cube with a side of 30 nm and a doping concentration of 10^{18} cm^{-3}, which yields a number of dopants in the cube 10^{18} $cm^{-3} \times$ 30 nm^3=27. Such simple example counting the number of impurity atoms in the channel region confirms that the device performance in this case can be tightly dependent on the limited number of dopants. Hence, the device becomes sensitive to the fluctuation in the accurate number of dopants. Despite that MOSFETs have in general identical number of dopants but the discrepancy in terms of the spatial position of dopants in the channel region may result in device performance variation too. Based on the analytical expression

of the MOSFET threshold voltage, it is possible to conclude that the random dopant fluctuation certainly contributes to the threshold voltage alteration of the device [Shin, 2016].

- **Velocity saturation :** It is well recognized that the geometrical dimensions of MOSFETs are continuously reduced to boost the speed and packing of components in modern integrated circuits. The applied voltages cannot be proportionally reduced since the operating voltages should be kept above the threshold voltage. Consequently, a short channel device operates under high lateral and vertical electric fields in the channel. It happens that the saturation of the current occurs at a drain-to-source voltage smaller than the voltage causing channel pinch-off at the drain end, which can be justified by the proportionality of the current magnitude to the drift velocity of the carriers. In fact, the latter velocity has a linear tendency in terms of the lateral electric field in the channel and saturates due to the carrier velocity saturation if the field is augmented beyond a specific level. Therefore, the lateral electric field is stronger than the critical velocity saturation value, but not being stronger than the vertical electric field at the drain end of the channel i.e. the channel is not pinched off. Despite the discrepancy between the current saturation and the pinch-off mechanisms, the MOSFET structure can be employed as a voltage controlled current source [Dimitrijev, 2012].
- **Mobility degradation :** The assumption that the mobility of carriers is constant should be revised because the value of such parameter emanates in reality from many physical parameters related mainly to the carrier scattering mechanisms. The carrier scattering in the channel is significantly affected by the vertical electric field due to the applied gate voltage, where carriers attracted to the boundary between the semiconductor and the gate dielectric material have higher collision rate with lattice defects. As a result, the carriers' mobility will be a function of the applied gate voltage value. Semi-empirical formulations have been adopted to model the mobility degradation because of the vertical fields in MOSFET structures. It is proved that such alteration can be expressed as a series resistance in the source of the device. The mobility is further decreased when the carriers in the channel are subject to high lateral electric field, which is more manifesting in short channel devices. Particularly, at high electric fields, the drift velocity of carriers deviates from the linear trend and even saturates. Such velocity bounding can be responsible for the saturation in the drain current since a device can reach the active region before the drain voltage attend the voltage associated to the channel pinch-off. Hence, the standard mobility equation must be adjusted in order to account for the carrier saturation velocity phenomenon [Alioto and Palumbo, 2005].
- **Impact ionization :** The impact ionization can be seen as a key generation mechanism, where it takes place in numerous semiconductor devices causing harmful parasitic effects. The impact ionization can only happen when the carrier gains energy from the electric field equals or higher than a threshold value. Based on the energy conservation and momentum laws at a collision event, it can be concluded that a minimum energy at least equals one half the bandgap is needed. The accurate determination of the ionization rate, defined as the number of electron-hole pairs generated by a carrier per unit distance traveled, is an intractable task. Generally, the solution is obtained using local field approximation for very high and lower electric fields. A more rigorous model valid for both high and low electric fields has been proposed but due to its analytical complexity, the original models are still widely used. Furthermore, the high field region width, where the impact ionization arises,

is one of the principal parameters in short channel transistors since it governs many degradation phenomena including the drain breakdown voltage, the substrate current and the hot electron generation ([Maes et al., 1990]and [Wong, 2000]).

- **Hot carrier effect :** Carriers moving from the source to the drain may receive a sufficient amount of energy between two scattering events under intense electric fields (>10^4 V/cm). These carriers are known as hot carriers since their kinetic energy can reach higher values than the lattice temperature. Despite that the majority of the hot carriers continue towards the drain, a small portion holds enough energy to generate electrons and holes by impact ionization. For an NMOS device, the generated electrons contribute to the drain current while almost all generated holes are accumulated in the substrate and give rise to the substrate current. The latter provides an indirect measure of the drain electric field and implicitly the vulnerability of a device to hot carriers. Moreover, the photon emission taking place during the hot carrier effect in the drain side constitutes also an intensity indication of the hot carrier generation. Some of the hot carriers gain higher energies (about 3.2 eV for electrons and 4.7 eV for holes) allowing them to get through the energy barrier at the Si/SiO_2 boundary and be injected into the oxide as a gate current. Some injected carriers may remain trapped in the oxide or lead to the breaking of Si-H weak bonds. The resulted traps and defects can modify the electrical behavior of the MOSFET as expressed by the threshold voltage or the subthreshold slope, which in turn may ultimately influence the good operation of the overall integrated circuit [Acovic et al., 1996].
- **Quantum limitation :** The quantum limitation imposed by the geometrical reduction of the device includes both confinement and tunneling effects. The quantum confinement effect arises when the wavefunctions associated to carriers are detained in narrow spaces between barriers. In MOSFETs with very thin bodies (less than 5 nm), such phenomenon occurs in the channel core since the carriers are squeezed between the gate insulator and the built in field of the body. The quantum confinement leads to the raise of the carriers' ground state energy, which augments the threshold voltage and shifts the carriers' location a bit far from the Si/SiO_2 boundary. This shift reduces the impact of the gate insulator scaling by extending the effective oxide thickness by 0.5 to 1.5 nm. Thus, the threshold voltage varies inversely as the square of the silicon film thickness. The quantum mechanical tunneling is commonly more aggressive with scaling in comparison to the confinement effect. The tunneling occurs when carriers tunnel throughout the device barriers generating leakage current. The continuous downscaling induces unacceptable power dissipation, where the leakage may lead to the suppression of the logic state in some types of dynamic logic circuits. We distinguish two kinds of tunneling leakage namely tunneling current through the gate insulator and tunneling current through the drain to body junction [Oda and Ferry, 2005].

1.7 Emergent techniques for enhancing the MOSFET performance

With the aim of circumventing the undesirable downscaling parasitic effects in MOSFET devices, several novel alternatives have been reported in literature. In this section, the most eminent strategies adopted by both research and engineering communities are discussed.

1.7.1 Channel engineering and gate engineering technologies

It is possible to modify both the electric field distribution and the potential contours by adjusting the doping profile inside the channel, which aims at minimizing the "OFF" leakage current while maximizing the linear and saturated drive currents. This trend has been materialized through non-uniformly doped MOSFET designs in vertical and lateral directions called channel engineering. The vertical non-uniformity can be adopted for instance to control the threshold voltage or avoid punch through. Regarding the lateral non-uniformity, it may be due to the intended pocket implantation. On the other hand, the gate engineering technology is used to reduce the hot carrier effect and to provide higher gate controllability over the channel conductance. The dual material gate design was proposed by amalgamating two different metal gates having different work functions, which results in a step in the surface potential profile. The work function of the metal gate near the source side is higher than the one associated to the metal gate near the drain side $(\phi_1 > \phi_2)$. The dual material gate design offers enhancements in terms of short channel effect reduction because the step shape in the surface potential profile not only protects the device channel from any drain effect but also increases the average electric field and consequently the carrier transport efficiency. The metal gate near the drain side absorbs additional drain-source voltage above saturation which covers the channel region at the vicinity of the source from any drain potential variation. The logical extension of the dual material gate paradigm has been proposed and evaluated recently as showcased by the triple material double-gate MOSFET structure [Sarkar, 2012]. Figure 1.8 provides a schematic view of a triple material gate MOSFET device where the channel is Gaussian laterally doped.

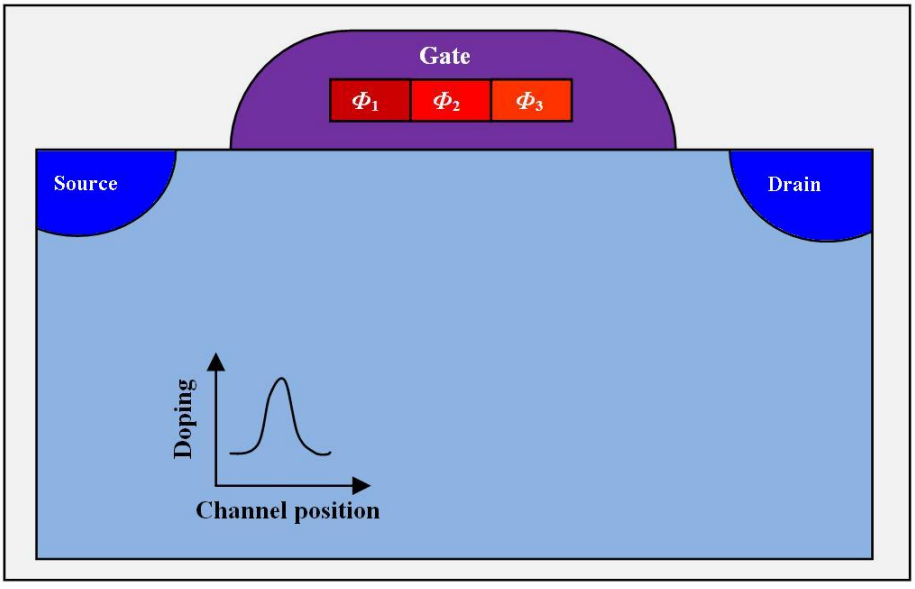

FIGURE 1.8 Schematic view of a triple material gate MOSFET device with Gaussian lateral doping.

1.7.2 New channel materials and alloys

Due to the increasing complexity and cost related to the downscaling of silicon CMOS, significant interests are actually agglomerated regarding the utilization of new channel materials. As a first alternative, it would be beneficial to reduce the degeneracy and to

lower the carrier effective mass through strained silicon, which yields improved carrier mobility and drive current of the device. Despite that a certain strain level can be process-induced, higher strains were achieved by growing silicon on a relaxed $Si_{1-x}Ge_x$ virtual substrate. Due to the dependence of the transport properties of carriers in the strained quantum well on the quality of buffer layers, proper $Si_{1-x}Ge_x$ buffers with a smooth surface morphology, low threading dislocations density on the surface and a high degree of relaxation are strongly advised. It is worthy to highlight that a practical Ge content in the SiGe buffer layer for subsequent growth of tensile strained silicon with high electron mobility ranges from 10% up to 40%. On the contrary, formation of a compressive strained Ge with high hole mobility is more complicated since it necessitates higher Ge composition SiGe buffers varying from 60% up to 90%. Furthermore, the virtual substrate development permits the creation of strained germanium layers for p-channel operation for which the hole mobility can surpass the silicon electron mobility. The reduction of self heating effect is made possible by reducing the SiGe buffer layer thickness, which is justified by the poor thermal conductivity of SiGe layers in comparison to their counterparts based on silicon. The hole mobility in devices including high-κ materials demonstrates a significant boost with respect to unstrained germanium and the universal silicon values subject to weak effective field. However, it seems that the device performance is actually affected by the high-κ dielectric to channel interface rather than the Ge channel [Balestra, 2008].

1.7.3 Multigate MOSFET prototypes

The multigate MOSFET framework investigates a third dimension to remedy short channel effects, where an electrode is wrapped around many facets of the channel body. The first occurrence of the double gate MOSFET appellation was in a report published in 1984 and stating the possibility of improving the immunity against the short channel effects compared to the single case. The fabrication of the double gate MOSFET device has been performed after five years as the fully depleted lean-channel transistor. The tri-gate MOSFET consists of a thin silicon film with the gate covering three sides. In the gate all around reported for the first time in 1990, the gate electrode is wrapped around all the channel region sides. The efficiency of this type of gate for the channel electrostatic control render the fabrication of a MOSFET device without junctions between the source, the channel region and the drain possible. Hence, the fabrication procedure can be greatly simplified when dealing with these junctionless multigate transistors at the nanoscale level. Additionally, it is feasible to introduce electron-trap layers or nanocrystals in the gate dielectric in order to elaborate nanowire memory transistors. For instance, among the shortest MOSFETs implemented in practice we find a tri-gate structure with a 3.8 nm gate length. Such device has a subthreshold slope of 92 mV/decade and a DIBL of 148 mV/V. The multigate devices are generally composed of several parallel nanowire known as fingers and sharing a common gate electrode. By doing so, the current drive of the structure can be enhanced just by increasing the fingers' number. Roughly speaking, the fabrication process of the multigate designs is built upon the same elementary phases of other complementary metal-oxide-semiconductor process but including some fine tunings. So, the enhancement strategies used in the case of the standard silicon CMOS technology can straightforwardly be adopted in the context of multigate topologies such as the use of silicon epitaxy with the goal of reducing the resistance of the source, the drain and the gate electrodes [Ferain et al., 2011]. In Figure 1.9, both double gate and surrounding gate MOSFET topologies are visualized.

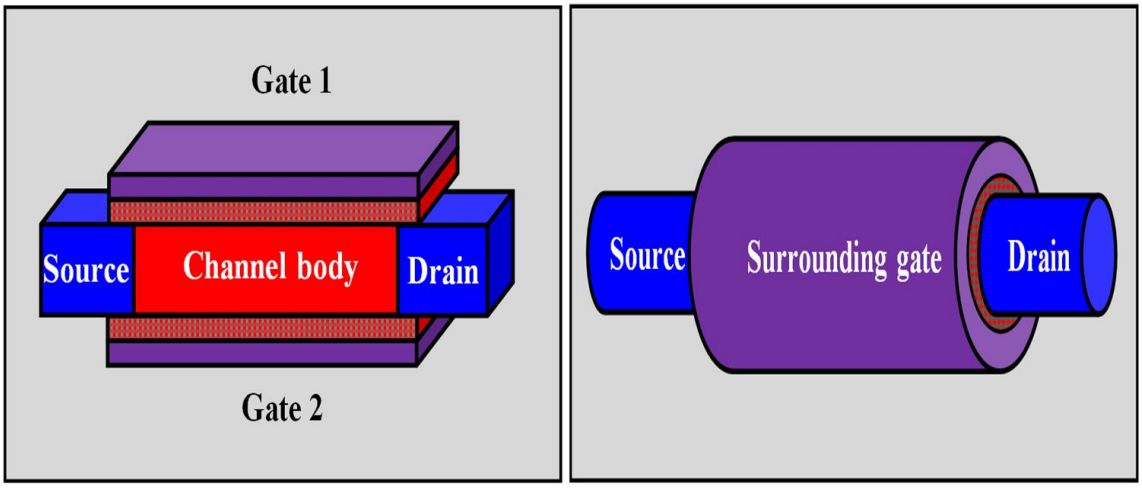

FIGURE 1.9 Elucidation of the structural configuration of the double gate MOSFET device (picture on the right) and the surrounding gate MOSFET device (picture on the left).

1.7.4 Dielectric gate development

The principal defy to the downscaling of gate dielectric resides in the drastic augmentation of the tunneling current with reducing the film thickness. For example, the leakage current of a thin film with a thickness of 2 nm can reach a magnitude of 1 to 10 A/cm^2, which represents a serious drawback mainly for portable machines. Thus, the reduction of the leakage current can be considered as the main reason behind focusing on the replacement of SiO_2 dielectric. In fact, the gate leakage current has an exponential dependence with the dielectric thickness while the capacitance linearly depends on the thickness. However, another exponential dependent term exists in the tunneling current which is the barrier height between the cathode and the conduction band of the insulator and for a large number of dielectrics, the tunnel current exponentially depends upon such barrier height. Consequently, three quality criteria of a dielectric should be considered namely higher dielectric constant, suitably large bandgap, and barrier height. Many materials have been proposed to replace SiO_2 dielectric but without success due to crystallization induced leakage currents and poor reliability in thick equivalent oxide thickness regions. Researches relative to high-κ gate dielectrics have been intensified during the 1990s and the tendency has been rapidly oriented towards the ZrO_2 and HfO_2 classes with larger bandgaps especially with their good thermal stability. By using hafnium based gate dielectric, mobility amelioration has been observed and for a thickness less than 2 nm the mobility degradation due to transient charging effects can be suppressed. The integration of metal electrodes with high-κ gate dielectrics is motivated by their thermal stability in contact with poly silicon gate. The difficulty of altering the effective work function of a poly-Si gate when integrated with high-κ dielectrics can be attributed to the fact that the effective work function of a poly-Si/high-κ dielectric stack is defined by the silicon/hafnium bonding and not Fermi level of the poly-Si gate. For some devices with a poly-Si/high-κ stack, the threshold voltage may be high and exceeds the practical range, which imposes a serious shortcoming in the implementation of high-κ dielectrics. In order to gain applications with good performances, it is necessary to employ dual work function metal electrodes with effective work functions approaching the conduction and valence bands [He and Sun, 2012]. The gate stack MOSFET design includes a high-κ layer placed on the silicon oxide film as indicated by Figure 1.10.

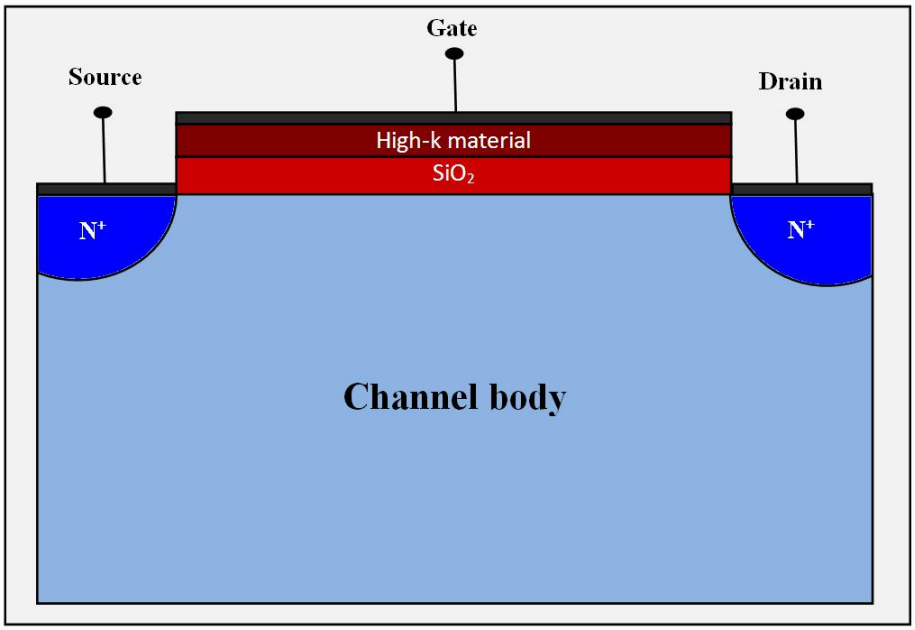

FIGURE 1.10 Cross sectional view of a gate stack MOSFET device.

1.8 Conclusion

It is strongly believed that the MOSFET devices are the workhorses leading the actual technological revolution including the advanced information processing tools. The continuous downscaling of these devices has allowed making it possible not only very intensified levels of integration but also enhanced functionality options. Nevertheless, going beyond 42 nm node has resulted in many parasitic effects which may degrade drastically the operating performance. Instead of using trial and error practical experiments, analytical modeling based on physics laws has been adopted mainly for long channel structures for which some approximation schemas are verified. For short channel devices, many assumptions are not valid and the accurate development of the drain current expression, which constitutes the initial step towards elaborating other criteria, is a cumbersome task needing the resolution of Schrödinger/Poisson equations self-consistently. By consequence, the development of more competitive approaches offering a compromise between flexibility and accuracy is required.

1.9 References

Acovic, A., La Rosa, G., and Sun, Y. C. (1996). A review of hot-carrier degradation mechanisms in MOSFETs. *Microelectronics Reliability*, 36(7-8) :845–869. 29

Alioto, M. and Palumbo, G. (2005). *Model and Design of Bipolar and MOS Current-Mode Logic : CML, ECL and SCL Digital Circuits.* Springer Press. 28

Annaratone, M. (1986). *Digital CMOS Circuit Design.* Kluwer Academic Press. 21

Arns, R. G. (1998). The other transistor : Early history of the metal-oxide semiconductor field-effect transistor. *Engineering Science and Education Journal*, 07(05) :233–240. 13

Baker, R. J. (2011). *CMOS : Circuit Design, Layout, and Simulation.* Wiley Press. 10

Balestra, F. (2008). New semiconductor devices. *Acta Physica Polonica A*, 114(05) :945–974. 31

Bentarzi, H. (2011). *Transport in Metal-Oxide-Semiconductor Structures : Mobile Ions Effects on the Oxide Properties*. Springer Press. 14

Bhattacharyya, A. B. (2010). *Compact MOSFET Models for VLSI Design*. Wiley Press. 24

Böer, K. W. (1992). *Survey of Semiconductor Physics Volume II Barriers, Junctions, Surfaces, and Devices*. Springer Press. 9

Cassidya, A. S., Andreas, J. G., and Andreou, G. (2013). Design of silicon brains in the nano-CMOS era : Spiking neurons, learning synapses and neural architecture optimization. *Neural Networks*, 45 :04–26. 13

Dimitrijev, S. (2012). *Principles of Semiconductor Devices*. Oxford University Press. 28

Drazdziulis, M. and Larsson-Edefors, P. (2003). A gate leakage reduction strategy for future CMOS circuits. In *The* 29^{th} *European Solid-State Circuits Conference, ESSCIRC'03*, Estoril, Portugal. 27

Ferain, I., Colinge, C. A., and Colinge, J. P. (2011). Multigate transistors as the future of classical metal-oxide-semiconductor field-effect transistors. *Nature*, 479 :310–316. 31

Fleischammer, W. (1996). *Quality by Design for Electronics*. Springer Press. 10

Ghosh, P. (2013). *Analytical Modeling and Simulation of Dual Material Gate Stack Architecture Cylindrical Surrounded Gate Mosfets*. PhD thesis, University of Delhi. 24

Ghosh, P., Haldar, S., Gupta, R. S., and Gupta, M. (2012). An investigation of linearity performance and intermodulation distortion of GME CGT MOSFET for RFIC design. *IEEE Transactions on Electron Devices*, 59(12) :3263–3268. 23

Golio, M. (2002). *RF and Microwave Semiconductor Device Handbook*. CRC Press. 23, 27

Gu, R. X., Sharaf, K. M., and Elmasry, M. I. (1996). *High-Performance Digital VLSI Circuit Design*. Springer Press. 21, 22

He, G. and Sun, Z. (2012). *High-k Gate Dielectrics for CMOS Technology*. Wiley Press. 32

Hoffmann, K. (2004). *System Integration : From Transistor Design to Large Scale Integrated Circuits*. Wiley Press. 9

Hu, C. C. (2009). *Modern Semiconductor Devices for Integrated Circuits*. Pearson Press. 16

Huff, H. (2009). *Into the Nano Era : Moore's Law Beyond Planar Silicon CMOS*. Springer Press. 12

Huff, H. and Gilmer, D. (2005). *High Dielectric Constant Materials VLSI MOSFET Applications*. Springer Press. 13

Isaac, R. D. (2000). The future of CMOS technology. *IBM Journal of Research and Development*, 44(03) :486–504. 9

Kumar, S. P., Agrawal, A., Chaujar, R., and Gupta, R. S. Gupta, M. (2011). Device linearity and intermodulation distortion comparison of dual material gate and conventional AlGaN/GaN high electron mobility transistor. *Microelectronics Reliability*, 51(03) :587–596. 23

Ma, W. and Kaya, S. (2004). Impact of device physics on DG and SOI MOSFET linearity. *Solid-State Electronics*, 48(10-11) :1741–1746. 23

Maes, W., Demeyer, K., and Van Overstraeten, R. (1990). Impact ionization in silicon : A review and update. *Solid-State Electronics*, 33(06) :705–718. 29

Neamen, D. A. (2011). *Semiconductor Physics and Devices : Basic Principles*. McGraw-Hill Press. 20

Nishi, Y. and Doering, R. (2007). *Handbook of Semiconductor Manufacturing Technology*. CRC Press. 10

Oda, S. and Ferry, D. (2005). *Silicon Nanoelectronics*. CRC Press. 29

Peper, F. (2017). The end of Moore's law : Opportunities for natural computing?. *New Generation Computing*, 35(03) :253–269. 12

Pierret, R. F. (1996). *Semiconductor Device Fundamentals*. Addison Wesley Press. 16

Powell, J. R. (2008). The quantum limit to Moore's law. *Proceedings of the IEEE*, 96(08) :1247–1248. 12

Razavi, B. (2017). *Design of Analog CMOS Integrated Circuit*. McGraw-Hill Press. 22

Sarkar, A. (2012). *Study of Short Channel Effects in Sub 100 nm MOSFETs their Remedies*. PhD thesis, Jadavpur University. 30

Shin, C. (2016). *Variation-Aware Advanced CMOS Devices and SRAM*. Springer Press. 28

Shur, M. (1990). *Physics of Semiconductor Devices*. Pearson Press. 22

Sze, S. M. (1981). *Physics of Semiconductor Devices*. Wiley Press. 22

Sze, S. M. and Lee, M. K. (2012). *Semiconductor Devices : Physics and Technology*. Wiley Press. 15

Taur, Y., Buchanan, D. A., Chen, W., Frank, D. J., Ismail, K. E., Lo, S. H., Sai-Halasz, G. A., Viswanathan, R. G., Wann, H. J. C., Wind, S. J., and Wong, H. S. (1997). CMOS scaling into the nanometer regime. *Proceedings of the IEEE*, 85(04) :486–504. 9

Tewksbury, S. K. and Hornak, L. A. (1989). Wafer level system integration : A review. *IEEE Circuits and Devices Magazine*, 05(05) :22–30. 11

Tsividis, Y. and McAndrew, C. (2012). *Operation and Modeling of the MOS Transistor*. Oxford University Press. 23

Valizadeh, P. (2016). *Field Effect Transistors, a Comprehensive Overview : From Basic Concepts to Novel Technologies*. Wiley Press. 13

Wallmark, J. T. and Carlstedt, L. G. (2013). *Field-Effect Transistors in Integrated Circuits*. Springer Press. 15

Wanlass, F. M. (1967). Low stand-by power complementary field effect circuitry. US Patent US28878663A. 9

Wong, H. (2000). Drain breakdown in submicron MOSFETs : A review. *Microelectronics Reliability*, 40(01) :3–15. 29

Wu, J., Shen, Y. L., Reinhardt, K., Szu, H., and Dong, B. (2013). A nanotechnology enhancement to Moore's law. *Applied Computational Intelligence and Soft Computing*, pages 1–13. 12

Xie, Q., Xu, J., and Taur, Y. (2012). Review and critique of analytic models of MOSFET short-channel effects in subthreshold. *IEEE Transactions on Electron Devices*, 59(06) :1569–1579. 9

Chapter 2

An ANFIS based approach for the subthreshold behavior modeling of nanoscale DG MOSFETs including interface traps

« *Artificial Intelligence can be a supplement to human insight, not substitute.* »

Abhijit Naskar

Abstract- *A fuzzy framework based on an adaptive network fuzzy inference system is proposed to evaluate the relative degradation of the basic subthreshold parameters due to the hot carrier effects for nanoscale thin film double gate MOSFETs. The impact of the channel length and thickness on the resulted degradation is addressed and the 2-D numerical simulations are used for the elaboration of the training database. Several membership function shapes are tested and the best one in terms of accuracy is selected. The predicted results agree well with the 2-D numerical simulations and can be efficiently used to investigate the impact of the the interface fixed charges jointly with the quantum confinement on the subthreshold behavior of the nanoscale double gate MOSFET. Therefore, the proposed neuro-fuzzy based approach not only offers a simple but also an accurate technique to study the nanoscale devices, when subject to both hot carrier and quantum confinement effects.*

2.1 Introduction

The actual trend in the microelectronics industry requires the fabrication of components with very small dimensions. The Double Gate Metal Oxide Semiconductor Field Effect Transistor (DG MOSFET) architecture has been considered as a promising device for nanoscale CMOS based applications. Such a device, when subject to a continuous downscaling, has resulted in many challenging drawbacks that are reflected by the increase of some undesirable quantities such as the leakage current [Sverdlov et al., 2003]. In addition, as the channel length is scaled down to increase both the operation speed and the integration density, it can reach the same order of magnitude as the depletion layer widths of the source and the drain junctions and at this level, the so-called short channel effects occur. This situation mainly induces the deterioration of the gate controllability efficiency over the channel. Therefore, more attentions should be focused on the proposition and evaluation of novel architectures ([Abdi et al., 2009], [Bentrcia et al., 2013b] and [Chebaki et al., 2012]). As stated by the International Technology Roadmap of Semiconductors (ITRS), a 10 nm gate length with fully depleted silicon on insulator technology and a 7 nm gate length with double gate devices are predicted in the near future. Various enhancements have been also conducted at the level of fabrication materials, where the use of high-κ materials and strained silicon has been demonstrated to be beneficial in improving performances of the multigate MOSFETs [Association, 2009]. In these devices, the gate electrode is wrapped around a silicon nanowire, forming a multigate structure with excellent control of the channel potential allowing the full depletion of the channel region. Roughly speaking, a wide range of applied modifications at both design and practical implementation levels, where new materials, alloys and structural improvements have been adopted in order to obtain higher performance ([Kaur et al., 2007], [Amat et al., 2009] and [Ghosh et al., 2012]).

However, even with all these efforts, many of the proposed improvements will be put into question if the severe constraints imposed by the actual trends in downscaling and high packaging of electronic components are not treated appropriately [Association, 2009]. In fact, by reducing the channel length of the device, its performance is affected, and such correlation is known as the short channel effect (SCE). The situation becomes worse when reaching the sub 100 nm technology nodes, where SCEs are aggressively amplified and various device characteristics are degraded considerably [Difrenza et al., 2003]. In addition, with the continuous scaling of the geometrical dimensions of MOSFET devices, their accurate modeling under various conditions of the working regime is becoming not only a critical domain, but also a necessary subject for understanding the miniaturization limits. The analytical modeling that has succeeded in some simple situations is in this case an intractable task due to the many mathematical and computational drawbacks encountered, especially when manipulating the joint Schrödinger/Poisson equations.

In addition to the short channel effects, another phenomenon that is closely linked to the reduction of the channel length consists in the formation of interface traps under the hot carrier injection effect leading to the alteration of the device operation [Bentrcia and Djeffal, 2011]. In fact, such reliability concern is the result of the interface trap generation because of the increase of the maximum electric field experienced by carriers in the channel near the drain side, since the associated power supply voltage is scaled at a slower rate compared to the channel length [Pagey, 2003]. An impact ionization process is initiated when carriers moving from the source to the drain sides acquire enough kinetic energy in the high field region of the drain junction. A fraction of carriers penetrates the Si-

SiO_2 potential barrier and becomes trapped in the gate oxide. From a phenomenological viewpoint, hot carriers can be identified by two main characteristics : (i) violation of the thermal equilibrium with the lattice and (ii) energy gain higher than the thermal energy. With the accumulation of injected carriers along the Si/SiO_2 boundary, an interface trap buildup and the trapping of carriers in the dielectric occur, initiating in turn the degradation of many device parameters such as the transconductance [Djeffal et al., 2011c]. At advanced stages of the process, the energy gained by the carriers in the high field region of the silicon substrate induces the break of bonds associated with extrinsic or intrinsic defects in the oxide, and such rearrangement in its atomic structure is the generator factor of the device instabilities observed during the hot carrier injection. The hot carrier induced damage in MOSFET devices has two harmful consequences : trapping of carriers in the defect sites in the oxide region, and the creation of interface traps at the silicon/oxide interface. As a result, the most obvious incident of this aging phenomenon is that various parameters of the device become dependent on the operating duration. For this reason, the carrier trapping in gate oxides is considered as the principal cause of instability for short channel MOSFETs, and has become one of the most important concerns for future MOS-integrated circuit development.

In order to have an accurate description of the DG MOSFET behavior with both an ultra thin silicon body and short channel length, the classical semiconductor equation should be extended to account for the quantum effects becoming dominant at the nanoscale level. The modeling framework based on quantum mechanics tools has received less interest due to the intractability or even the impossibility of getting a self-consistent solution from the Schrödinger/Poisson equations in terms of complexity and computational cost [Prégaldiny et al., 2004]. The resolution methodologies are based mostly on numerical methods, which are highly time consuming due to the intrinsic nature of the physical effects governing the electrical properties under such constraints. The development of compact analytical formulations that are both supporting quantum aspects and valid in all operating regions is really a very hard drawback to tackle [Wen and Singh, 2010]. Hence, other alternative approaches have been used to study the immunity of DG MOSFETs based on the analysis of subthreshold or small signal parameters before and after stressing. Compact models are widely used for this purpose. These models are built from some simplifying assumptions that are required to reduce the complexity of the resolution approach without significantly altering the quality of the obtained results ([Ghoggali et al., 2010], [Bendib and Djeffal, 2011], [Djeffal et al., 2011b] and [Bentrcia et al., 2012a]). But since this is not always the case, soft computing based approaches have been proposed to include short channel and quantum effects in the same framework for DG MOSFET performance modeling ([Djeffal et al., 2007], [Bentrcia et al., 2013c], [Bentrcia et al., 2013a] and [Bentrcia et al., 2012b]).

In this chapter, we investigate the capabilities of an Adaptive Network based Fuzzy Inference System (ANFIS) to predict the degradation behavior of nanoscale thin film DG MOSFETs in the subthreshold regime. This degradation is a complex phenomenon for thin film MOSFETs, which mainly depends on the hot carrier effect and quantum confinement. Two geometrical parameters are selected as the input variables to the fuzzy system (the channel thickness and the channel length) in order to forecast the change in the subthreshold parameters : threshold voltage, drain induced barrier lowering and OFF current denoted by V_{th}, DIBL and I_{OFF}, respectively. In order to get an optimized ANFIS structure, we try some types of membership functions (MFs) during the ANFIS training phase, so that the best one in terms of accuracy could be selected. We have also explored more in details the framework for modeling more accurately the relative degradation in

terms of the threshold voltage, where several types of membership functions are tested. For this measure, the efficiency of the proposed approach is validated through comparison with an Artificial Neural Network (ANN), where superior performances are recorded. The swing factor behavior is modeled and used to study the electrical behavior of a DG MOSFET based amplifier.

The remainder of this chapter is organized as follows. In Section 2.2, we outline the hot carrier degradation problem, in addition to some of the relevant literature. In Section 2.3, we present the main concepts relative to the ANFIS architecture and its hybrid learning algorithm. Section 2.4 depicts in detail the elaboration procedure of the numerical database and the configuration of various device parameters. The computational results, associated in general to the subthreshold parameters and more particularly to the threshold voltage and swing factor, are discussed in Section 2.5. Finally, we achieve this work with an outlook on the main concluding remarks in Section 2.6.

2.2 Parasitic effects and impact on the threshold voltage

In response to the actual advances in semiconductor technology, besides the need for faster integrated circuits, the associated MOSFET sizes are continuously downscaled and therefore approaching their physical limits. On the other hand, an extreme difficulty is encountered when scaling some pertinent parameters, which is the case of the supply voltage used to operate these devices. This discrepancy in the velocity of downscaling geometrical and electrical parameters is the origin of a major reliability problem for short channel DG MOSFETs [Djeffal et al., 2011c].

During the operation of the device, high electric fields are created in the silicon substrate near the drain-substrate junction due to the formation of a pinch-off region. The intensity of the electric fields increases rapidly with the reduction in the channel length, as predicted by Poisson's equation. When high drain voltages are applied, carriers in the channel and pinch-off regions reach non-equilibrium energy distribution. The energetic carriers qualified as hot carriers trigger various mechanisms at the silicon/oxide interface, and a fraction of the injected hot carriers lose their energy via interaction [Tyaginov et al., 2010]. As the amount of injected hot carriers into the oxide depends strongly on the magnitude of the electric field inside the silicon substrate, good estimation of the hot carrier degradation requires the accurate determination of the electric field distribution in silicon. Many of the approaches proposed for the modeling of hot carrier distribution are based on the assumption of linear proportionality between the average carrier energy $< \mathrm{E} >$ and the local electric field E_{loc} :

$$\langle \mathrm{E} \rangle = q \langle l \rangle \mathrm{E}_{loc} \tag{2.1}$$

where l is the free path of a carrier.

According to the above formula, the maximum energy gained by a carrier is at most equal to that of an elementary charge moving across the potential difference between the drain and source electrodes of the device. If this assertion is true, no carrier injection should take place when the drain-source voltage is below the potential barrier at the Si/SiO_2 interface. However, significant substrate and gate currents are well detected at drain biases below this value, which constitutes a weak point for this model.

In order to remedy this paradoxical aspect, a more general framework should be used. Under the quantum theory of barrier penetration, carriers going across the local energy

barrier at the Si/SiO_2 interface are considered as particles incident upon a barrier potential. So, for each carrier there are two probability measures : a reflection probability and a transmission probability [Eisberg and Renick, 1969]. Both probabilities are related to the opacity of the potential barrier, which depends in turn on many parameters such as the carrier energy. Thus, the carrier injection process is a complex function of the applied biases, the position along the channel, and the type of carrier being injected among other factors, which cannot be modeled analytically in short channel devices.

After the long operating duration of a MOSFET device, a fraction of the injected carriers is trapped at defect sites in the oxide or at sites very close to the Si/SiO_2 boundary. Interfacially trapped holes that are positively charged have a high probability of being occupied by the electrons injected into the oxide. An amount of energy is released when the interfacially trapped hole is neutralized by electron capture, and such energy is sufficient to disrupt the chemical bonds near the interface to create an interface trap [Djeffal et al., 2011a]. The damage generated by the hot carrier effect is localized near the drain side since the electric field is maximal at this location. One way to look at this effect is to divide the MOSFET channel into two regions : the fresh region and the damaged region [Ho et al., 2005], which results in a threshold voltage that is larger than the fresh part of the channel. Thus, the general trend of a damaged device is that the threshold voltage increases and the drive current decreases as negative trapped charge builds up or interface trap generation proceeds [Mahapatra et al., 2000]. Based on analytical models, such effect is more pronounced at large drain voltages for larger interface charge region lengths, and the device is severely affected by the presence of negative or positive traps. The threshold voltage near the drain end increases drastically due to the presence of negative interface charge density and reduces gradually for positive interface charge density [Singh, 2008].

It is common to characterize the device degradation by assessing the shift before and after the device is stressed. Therefore, the relative variation in a device parameter can be used to estimate the hot carrier induced failure time, which indicates that the circuit may no longer operate properly, where separate equations including empirical constants are developed for gate voltage regions. For example, it is observed that the variation in threshold voltage with stressing time follows a power law as [Naseh et al., 2006] :

$$\frac{\Delta V_{th}}{V_{th0}} \approx K_1 t^{n_1} \tag{2.2}$$

where both fitting parameters K_1 and n_1 are dependent on the biasing applied voltages (V_{ds} and V_{gs}).

The threshold voltage of a MOSFET device defines the applied gate voltage corresponding to the switch of the device to the "ON" state. Such value is strongly modified by various miniaturization effects [Abdelhamid, 2007]. Accurate expressions for the modeling of the threshold voltage are highly recommended so that the correct behavior of new designed circuits could be predicted appropriately. In what follows, we depict succinctly how some parasitic or shrinking effects may influence the threshold voltage variation.

- **Hot carrier effect :** The MOSFET device aging due to the long operating duration is reflected by damages in the form of interface traps created by hot carriers at the interface between the channel and the oxide regions [Bentrcia and Djeffal, 2011]. Many studies have been reserved to the analysis of MOSFET performances when subject to the interface trap degradation. Various analytical compact models have been proposed (see for example [Djeffal et al., 2011b] and [Bentrcia et al., 2009]). The existence of interface charge densities leads to an increase or a decrease in the

threshold voltage depending on the sign of the trapped charges. An additional term in the model of the fresh device is included as defined in [Naseh et al., 2006] :

$$\Delta V_{th} = -\frac{Q_0 + Q_{it}}{C_{ox}} \tag{2.3}$$

where Q_0 is the trapped charge density in the oxide, Q_{it} is the interface trap charge density and C_{ox} is the oxide capacitance. It can be concluded that the device becomes less sensitive to biasing and higher gate voltage values are needed to make the device switching from the "OFF" state to the "ON" state.

- **Reduction of geometrical parameters effect :** In the case of dimensional parameters, experimental measurements indicate that the threshold voltage tends to decrease by reducing the channel length. This monotone reduction becomes more noticeable when the channel length becomes comparable to the source and the drain depletion widths. The change in the threshold voltage due to the short channel effect is given by [Arora, 2007] :

$$\Delta V_{th} = \frac{Q_b}{C_{ox}} \frac{X_j}{L} \left(\sqrt{1 + \frac{2X_{dm}}{X_j}} - 1 \right) \tag{2.4}$$

where Q_b is the bulk charge per unit area, X_j is the junction depth and X_{dm} is the maximum depletion width.

For the more general case where both device parameters (the channel width and length) are of the same order as the depletion width (small geometry device), the change in the threshold voltage caused by the small geometry effect can be estimated as the sum of the short channel and narrow width effects as [Galup-Montoro and Schneider, 2007] :

$$\Delta V_{th} \approx \Delta V_{th,L} + \Delta V_{th,W} \tag{2.5}$$

- **Quantum confinement effect :** If we consider the quantization of energy levels, the distribution of accumulated or inverted carriers at the interface is different from the classical prediction [Balaguer et al., 2012]. The model of any surface potential based parameter such as the threshold voltage or the drain induced barrier lowering will be modified. Since the quantum effects are significant for oxide and channel thicknesses less than 1.5 and 5 nanometers respectively, the inclusion of quantum corrections in the available up-to-date DG MOSFET models cannot be ignored. The threshold voltage shift resulted from the quantum confinement effects can be expressed by [Morris and Abebe, 2010] :

$$\Delta V_{th} = \frac{\beta t_{ox} N_A}{2\varepsilon_{ox}} \left(\frac{kT\varepsilon_{si}}{n_i} \right)^{1/2} \tag{2.6}$$

where β is a fitting parameter, t_{ox} is the oxide thickness, N_A is the channel doping concentration and n_i represents the intrinsic carrier concentration.

2.3 Background of ANFIS based approach

In almost artificial intelligence based models, the cornerstone of the elaborated approaches consists of the development of an interpolative input-output mapping, where an equivalent easy to use scheme is introduced with a set of configurable parameters to be adjusted so that a good match between the predicted and real outputs is obtained. In general, a training algorithm is used to minimize such discrepancy between both outputs. It is worth noting that there is no guarantee that the execution of such algorithms will lead to a global optimum, and therefore a validation phase for these models is a mandatory requirement.

The general structure of a fuzzy inference system includes transformations based on membership functions and a fuzzy extended version of the implication rule that maps the input characteristics to a single valued output [Zadeh, 1973]. In many cases, the type of membership functions and the rule structure is essentially predetermined by the user's interpretation of the variables' characteristics in the model. However, in some situations, simply looking at the data cannot specify these aspects and the need for more robust strategies arises.

The ANFIS structure can be considered as a multilayered feed forward network, where neural network learning algorithms and fuzzy reasoning cooperate in the same framework to map nonlinear relationships within complex systems. In this formalism, the input space is divided into many local subregions, which means that several of them can be activated simultaneously by a single input, and then a simple local model given by a linear function is attached to each region [Jang, 1993]. The ability of ANFIS based approaches, to combine the advantages of the fuzzy computation with the numerical prediction provided by the neural networks, has motivated their successful application in various engineering fields such as energy management, machining process and medical diagnosis ([Lo, 2003], [Ying and Pan, 2008] and [Akdemïr et al., 2010]).

In order to highlight the ANFIS architecture and function, we consider a system with two inputs identified in this chapter as the channel length and the channel thickness, because they are the principal parameters having a strong influence on the short channel and the quantum confinement effects. The number of fuzzy sets attached to each input is denoted by n and m, respectively. Hence, the total number of Takagi-Sugeno-Kang fuzzy IF-THEN rules is $n \times m$. The generic expression of such rules is given by :

$$\text{Rule } k : \text{IF L is } \mathrm{A}_i \text{ and } t_{si} \text{ is } \mathrm{B}_j \text{ THEN } f_k = p_k \times \mathrm{L} + q_k \times t_{si} + r_k \text{ for } i = \overline{1,m}\ j = \overline{1,n}\ k = \overline{1,m \times n} \tag{2.7}$$

where A_i and B_j are the fuzzy sets associated with the geometrical input parameters, and p_k, q_k and r_k are the linear parameters in the consequent part of the rule. The corresponding ANFIS architecture is composed of interconnected nodes organized in five functional layers as follows [Singh, 2010] :

i) **Input nodes :** Each node of this layer deduces the membership grades of the crisp inputs based on the appropriate fuzzy sets using membership functions :

$$\begin{cases} \mathrm{O}_{1,s} = \mu_{\mathrm{A}_i}(\mathrm{L}) & \text{for } s = \overline{1,m} \\ \mathrm{O}_{1,s} = \mu_{\mathrm{B}_j}(t_{si}) & \text{for } s = \overline{m+1,m+n} \end{cases} \tag{2.8}$$

If we assume that we have a generic membership function, then its representation can be given by the following expression :

$$\mu(x) = \mathrm{F}(x; a_{v=1:h}) \tag{2.9}$$

where x is a generic variable denoting the channel length or the channel thickness, and $\{a_{v=1:h}\}$ indicates the parameters' set of the membership function in the premise part of the fuzzy rule that fixes the shape of the membership function;

ii) **Rule nodes :** Nodes in this layer determine the fulfillment degree of the premise part of the rules by input values, using node functions as :

$$\mathrm{O}_{2,k} = t\left(\mu_{\mathrm{A}_i}(\mathrm{L}), \mu_{\mathrm{B}_j}(t_{si})\right) = \mu_{\mathrm{A}_i}(\mathrm{L}), \mu_{\mathrm{B}_j}(t_{si}) = \omega_k \quad \text{for } k = \overline{1, m \times n} \tag{2.10}$$

where t is a t-norm operator such as multiplication or minimum. The ω_k output indicates the firing strength of the k^{th} rule;

iii) **Average nodes :** The role of this third layer is to calculate the ratio of each rule's firing strength to the sum of all the rule's firing strength. So, $\overline{\omega_k}$ is taken as the normalized firing strength :

$$\mathrm{O}_{3,k} = \overline{\omega_k} = \frac{\omega_k}{\sum_{l=1}^{m \times n} \omega_l} \quad \text{for } k = \overline{1, m \times n} \tag{2.11}$$

iv) **Consequent nodes :** Each node in this layer computes the normalized output of a rule for current outputs L and t_{si} based on the node function :

$$\mathrm{O}_{4,k} = \overline{\omega_k} f_k = \overline{\omega_k}\left(p_k \times \mathrm{L} + q_k \times t_{si} + r_k\right) \tag{2.12}$$

where $\{p_k, q_k, r_k\}$ is the consequent parameter set of rule k;

v) **Output nodes :** This layer contains only a single node that sums the outputs of the previous layer nodes to obtain the overall network output :

$$\mathrm{O}_{5,1} = \sum_{k=1}^{m \times n} \overline{\omega_k} f_k = \frac{\sum_{k=1}^{m \times n} \omega_k f_k}{\sum_{k=1}^{m \times n} \omega_k} \tag{2.13}$$

The deposition of these layers is illustrated in Figure 2.1.

The least squared errors criterion is used to tune up the ANFIS architecture configuration and is given by :

$$\mathrm{SE} = \sum_{o=1}^{\mathrm{N}} \mathrm{SE}_o = \sum_{o=1}^{\mathrm{N}} \left(\mathrm{T}_o - f_{outo}\right)^2 \tag{2.14}$$

where SE_o is the error measure for the o^{th} entry of the given training data set, T_o is the desired output of the o^{th} entry and f_{outo} is the output of ANFIS using the o^{th} entry. If

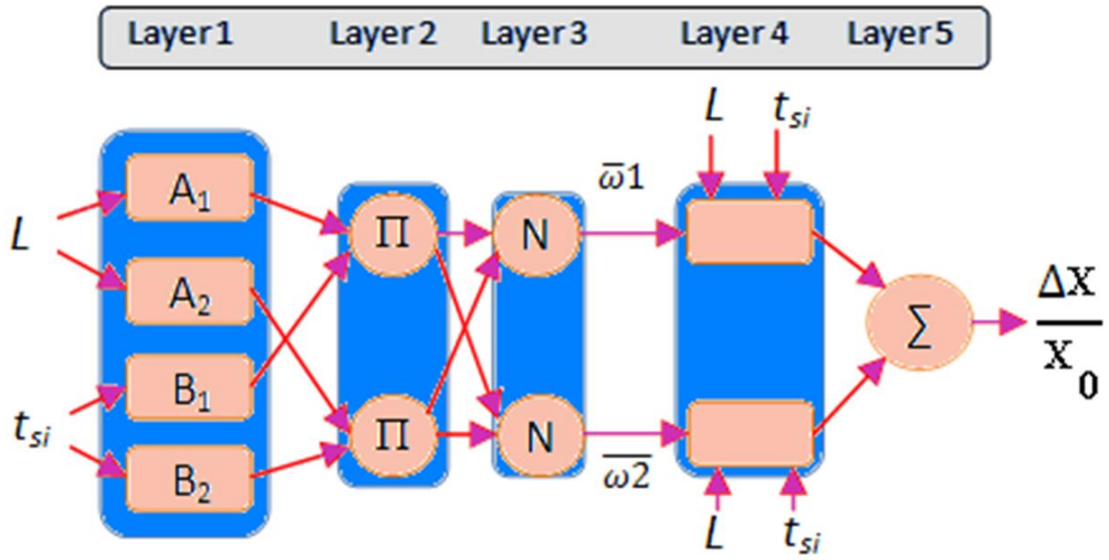

FIGURE 2.1 Illustrative representation of an ANFIS framework with two input parameters and two membership functions.

the premise parameters $\{a_{v=1:h}\}$ are fixed, then the output of the whole system can be expressed as a linear combination of the consequent parameters $\{p_k, q_k, r_k\}$ as :

$$f = \mathrm{B}\theta \tag{2.15}$$

where

$$f = \begin{bmatrix} f_{out1} \\ . \\ . \\ f_{outN} \end{bmatrix}$$

$$\theta = \begin{bmatrix} p_1 \\ q_1 \\ r_1 \\ . \\ . \\ p_{m\times n} \\ q_{m\times n} \\ r_{m\times n} \end{bmatrix}$$

and

$$\mathrm{B} = \begin{bmatrix} \overline{\omega_1}\mathrm{L}_1 & \overline{\omega_1}t_{si1} & \overline{\omega_1} & . & . & \overline{\omega_{m\times n}}\mathrm{L}_1 & \overline{\omega_{m\times n}}t_{si1} & \overline{\omega_{m\times n}} \\ \overline{\omega_1}\mathrm{L}_2 & \overline{\omega_1}t_{si2} & \overline{\omega_1} & . & . & \overline{\omega_{m\times n}}\mathrm{L}_2 & \overline{\omega_{m\times n}}t_{si2} & \overline{\omega_{m\times n}} \\ . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . \\ \overline{\omega_1}\mathrm{L}_N & \overline{\omega_1}t_{siN} & \overline{\omega_1} & . & . & \overline{\omega_{m\times n}}\mathrm{L}_N & \overline{\omega_{m\times n}}t_{siN} & \overline{\omega_{m\times n}} \end{bmatrix}$$

Since this equation is a standard linear least square problem, its algebraic solution (the estimator of the unknown matrix θ) can be written as [Lawson and Hanson, 1974] :

$$\theta^* = \left(B^T B\right)^{-1} B^T f \tag{2.16}$$

where B^T and B^{-1} are the transpose and the inverse of matrix B, respectively.

The hybrid learning algorithm of the ANFIS combines the gradient method with the least square method to update the parameters, where a generic parameter α formed by the union of the premise and the consequent parameters is updated using the formula :

$$\Delta\alpha = -\eta \frac{\partial SE}{\partial \alpha} \tag{2.17}$$

with SE is the overall error and η a learning rate calculated according to :

$$\eta = \frac{\delta}{\sqrt{\sum_\alpha \left(\frac{\partial SE}{\partial \alpha}\right)^2}} \tag{2.18}$$

where δ is the step size.

During the training phase, α is updated at each training epoch in a hybrid manner. More specifically, the consequent parameters of α are updated firstly using a least square algorithm, and the premise parameters are then adjusted by back propagating the errors. However, other alternative methods exist for these optimization processes in learning, such as the genetic algorithm (GA) and particle swarm optimization (PSO) based approaches. But the hybrid learning algorithm remains the most commonly used one in various applications ([Nariman-Zadeh et al., 2004] and [Jiang et al., 2012]).

2.4 Elaboration of the numerical database

Since the accuracy of a properly trained ANFIS depends on the accuracy and the effective representation of the data used for its training, the database employed for this purpose should be carefully elaborated. The 3-D schematic cross sectional view of the DG MOSFET device used in this work is depicted in Figure 2.2, where an interface trap density is assumed to exist near the drain side as a result of the hot carrier injection. In the elaboration procedure of the training database, different channel lengths and channel thicknesses are considered for evaluating the range of effects that may be more significant for analysis.

From the simulated device structure, it can be seen that the source/drain regions are highly doped in comparison to the channel. Various subthreshold characteristics (I_{OFF}, V_{th} and DIBL) associated to a DG MOSFET structure with different channel length and channel thickness values are obtained numerically using the ATLAS 2D simulator. In order to account for the quantum mechanical confinement of carriers, we use the Bohm Quantum Potential (BQP) model, which calculates a position dependent potential energy to correct the carriers' distribution. Two fitting parameters are included within this model to allow the approximation of the quantum behavior for many devices and materials. The description of the generation/recombination mechanisms is achieved through the Shockley-Read-Hall (SRH) model assuming that the transition of carriers between bands takes place via a single trap energy level localized deeply in the gap. Hence, the simulation

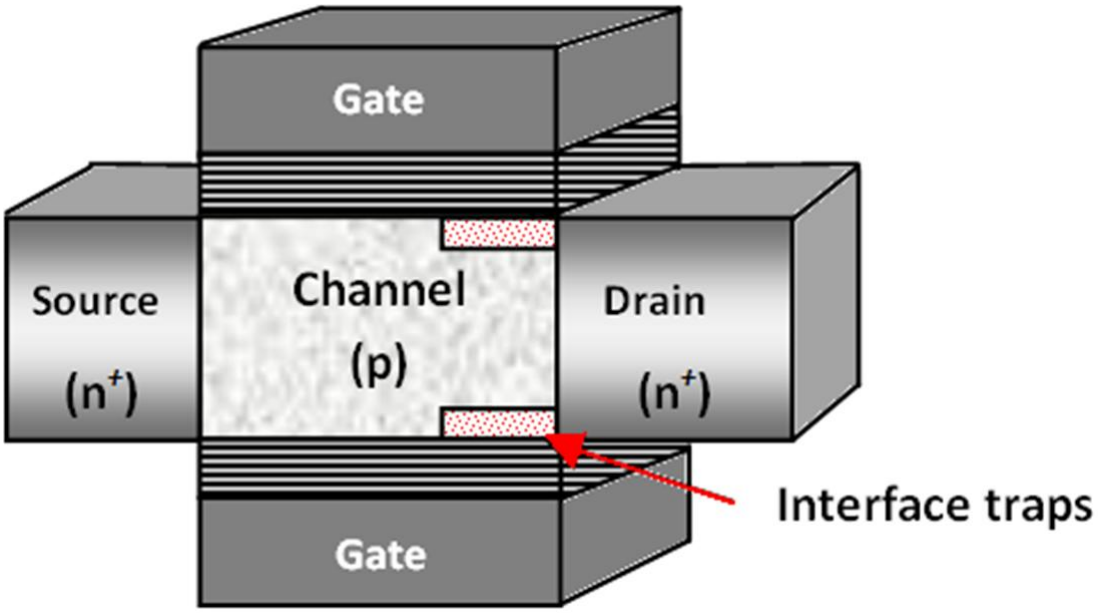

FIGURE 2.2 Cross sectional view (3D) of the symmetrical DG MOSGET including uniform interface traps near the drain side.

of the leakage currents is rendered possible. We introduce also the Fermi-Dirac statistics so that our numerical model can support properties of very highly doped materials. The field dependent mobility (fldmob) model is needed to take into consideration any type of velocity saturation effect. By enabling the energy balance transport (HCTE) model, it is possible to accurately simulate non-local carrier heating, which is of paramount importance for deeply scaled devices [Silvaco, 2012]. A total of 91 configurations are generated by parsing the channel length range with a step of 5 nm and the channel thickness range with a step of 0.5 nm. The set of geometrical and electrical parameters conserved during the running of all simulations is given in Table 2.1.

TABLE 2.1 Summary of the device design and simulation parameters.

Parameter	**Notation**	**Value**
Oxide thickness	t_{ox}	1.5 nm
Drain/source doping	$N_{D/S}$	10^{20} cm^{-3}
Channel doping	N_A	10^{15} cm^{-3}
Work function	ϕ_{MS}	4.55 eV
Interface trap density	N_F	5×10^{12} cm^{-3}
Drain voltage	V_{ds}	0.1 V
Gate voltage	V_{gs}	0.7 V

In our work, the subthreshold behavior of the nanoscale DG MOSFET is based on the extraction of some pertinent parameters describing the function of the device. The threshold voltage expresses the gate voltage at which an inversion charge connecting the source/drain electrodes is established. Figure 2.3 compares the threshold voltage curves for fresh and damaged cases as obtained using numerical simulation. For both cases, the threshold voltage increases with the channel length to reach a constant value, which is the long channel threshold voltage value. As it is predicted, the damaged device has a higher

threshold voltage because a higher applied gate voltage is needed to make the device turn on. Another interesting feature is related to the recorded results for small channel lengths, where both curves become closer in this case. This is can be explained by the strong correlation between the hot-carrier degradation and the short channel effect.

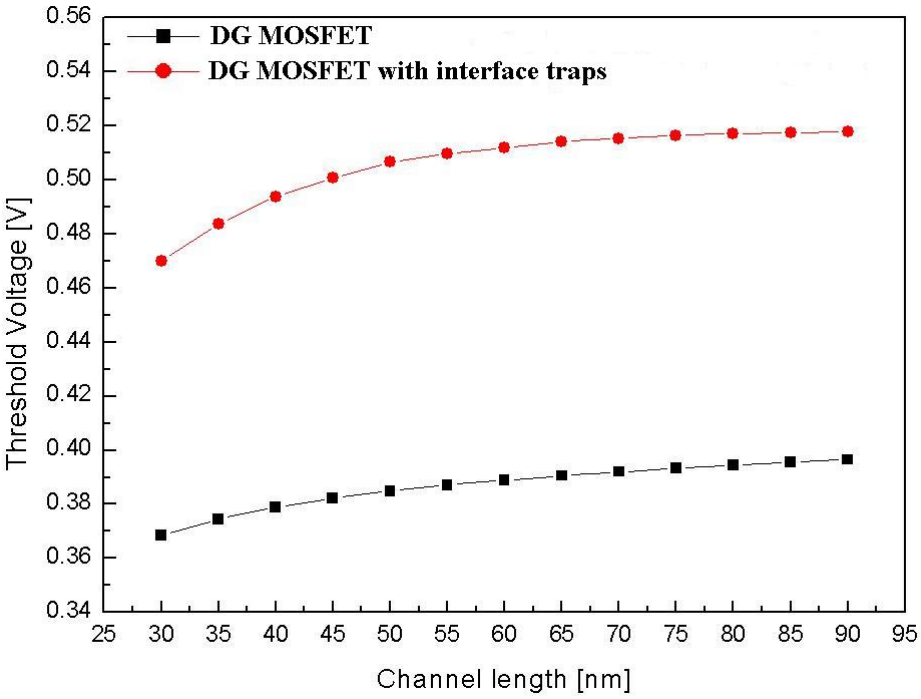

FIGURE 2.3 Variation of the DG MOSFET threshold voltage, with and without traps, as a function of the channel length (t_{si}=2 nm).

In Figure 2.4, the variation in the threshold voltage as a function of the channel thickness is presented with a fixed value of the channel length.

The drain induced barrier lowering, which reduces the threshold voltage for an increase in the drain voltage, is calculated from the difference between the threshold voltages at low and high values of the drain-source voltage, and these are equal to 0.1 and 0.3 V, respectively. As shown in Figure 2.5, the DIBL effect is more pronounced for the short channel length devices with and without interface traps.

The OFF-current arises due to the fact that the DG MOSFET is not an ideal switch since there is still some leakage current that flows through it in the "OFF" state. For nanoscale devices, both short channel and hot carrier degradation effects can induce a remarkable change in such drain controllability in the subthreshold regime.

Figure 2.6 shows the variation of the subthreshold current as a function of the channel length. From this figure, it can be noted that the subthreshold current is pronounced for short channel lengths in both cases (fresh and damaged devices). Another remarkable feature that is clearly shown resides in the diminution of the subthreshold current when interface traps are considered, which can be due to the trapping of a portion of carriers forming the subthreshold current in addition to the increase in the leakage current for nanoscale devices.

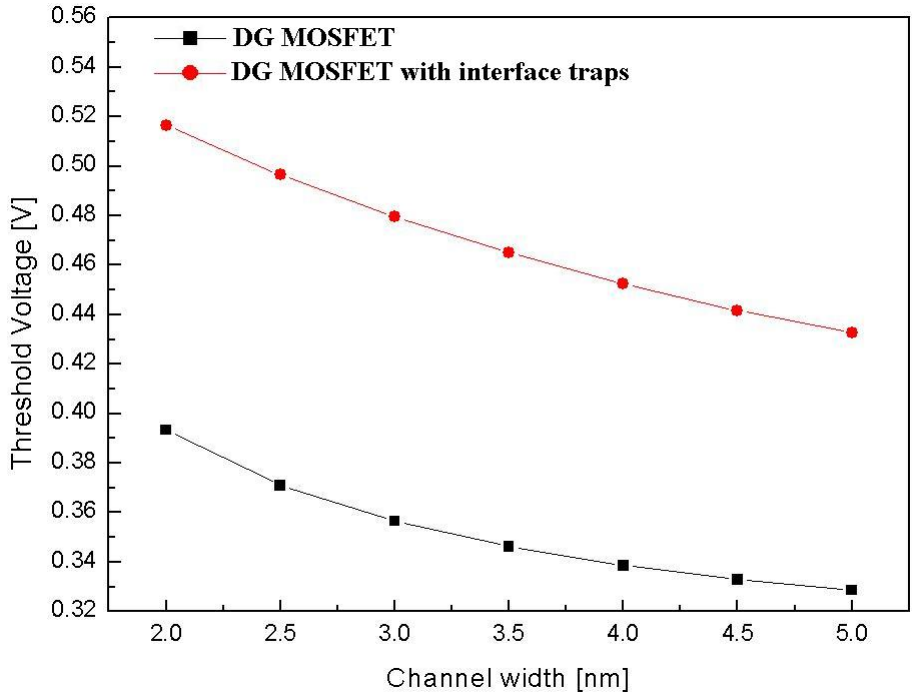

FIGURE 2.4 Variation of the DG MOSFET threshold voltage, with and without traps, as a function of the channel thickness (L=75 nm).

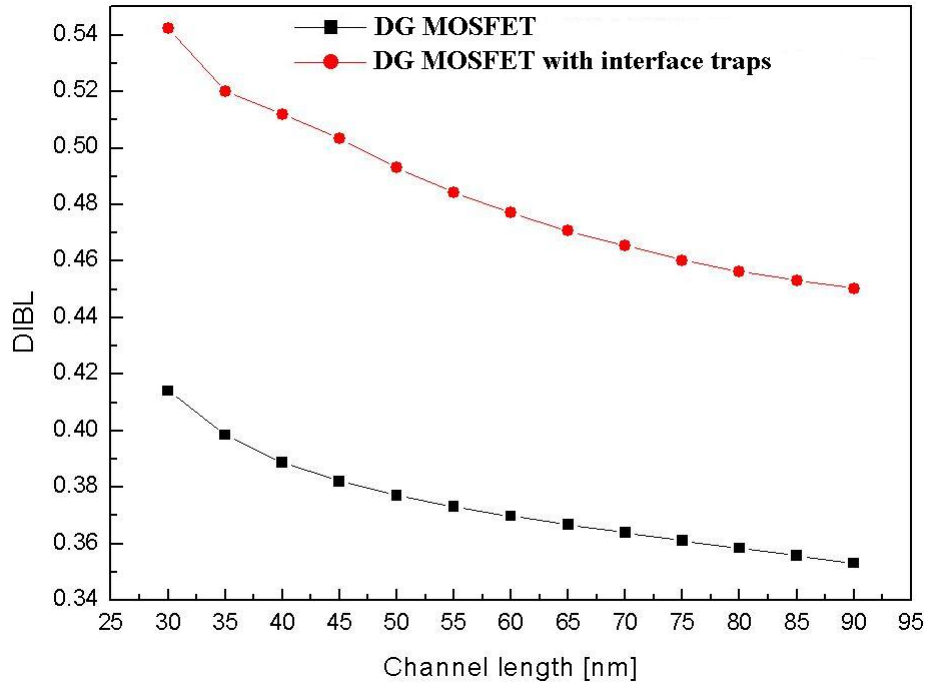

FIGURE 2.5 Variation of the DG MOSFET DIBL, with and without traps, as a function of the channel length (t_{si}=3 nm).

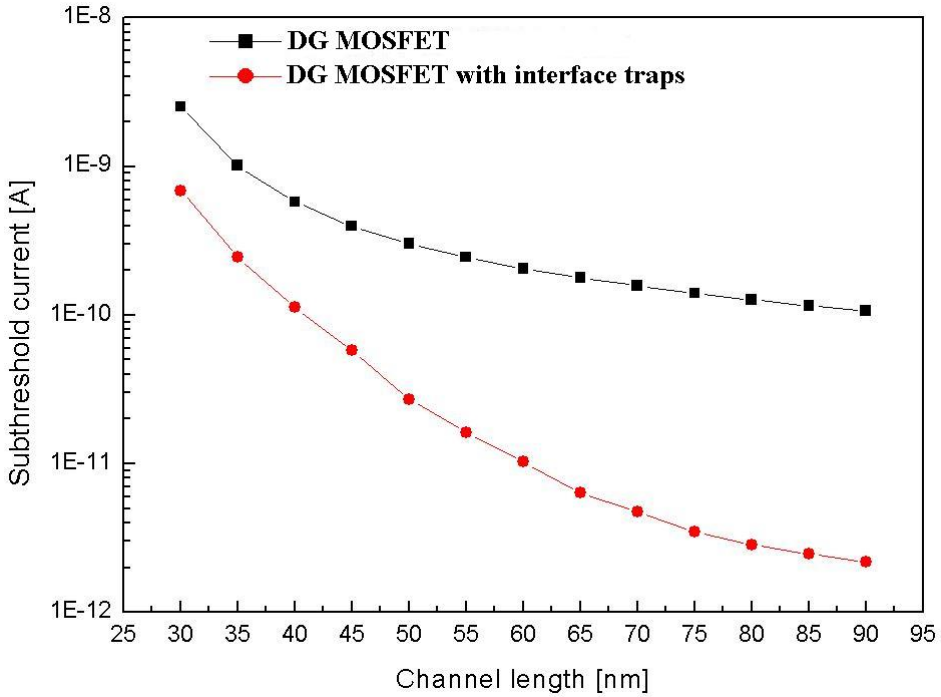

FIGURE 2.6 Variation of the DG MOSFET subthreshold current, with and without traps, as a function of the channel length (t_{si}=4 nm).

2.5 Results and discussion

In this section, we showcase in three parts the main results obtained through the application of the proposed framework. The first part includes the outcomes resulted by a global application of ANFIS for the subthreshold parameters (I_{OFF}, V_{th} and DIBL). The second part is dedicated to a deeper analysis of the threshold voltage parameter, where a large set of membership functions is tested. Furthermore, a checking set is considered in the training procedure to avoid the overfitting problem. So, we get an enhancement of the obtained ANFIS as consolidated by the comparison with an ANN based approach. In the last part, we develop an ANFIS model for the swing factor, which will be used later to study the electrical performance of a single DG MOSFET based amplifier.

2.5.1 Global modeling of the subthreshold behavior

Since the performance of a well trained ANFIS is strongly affected not only by the exhaustiveness of the training data on the considered interval but also by the type and the number of MFs used to characterize the input variables, the training data set is obtained numerically using the ATLAS 2D simulator. The specification of the type and number of MFs is determined by the trial and error approach because there exists no well established method that can achieve this task optimally. So, the MFs are selected heuristically and verified empirically.

In this work, the number of MFs associated with the input variables is determined by exhaustively searching a fixed interval, and we find that the best results are located within the interval $[2-15]\times[2-15]$. Thus, the number of fuzzy IF-THEN rules for ANFIS is bounded by the following interval $[4-225]$. The obtained results seem to suggest that using Gaussian combination, generalized bell-shaped and Π-shaped MFs for the threshold voltage, the drain-induced barrier lowering and subthreshold current, respectively, gives superior performance compared to other types of MFs. The analytical expressions of the three adopted MFs are indicated below.
The Gaussian combination-shaped MF is defined as :

$$\mu(x;\sigma_1,\sigma_2,c_1,c_2) = e^{-\frac{(x-c_1)^2}{2\sigma_1^2}} + e^{-\frac{(x-c_2)^2}{2\sigma_2^2}} \tag{2.19}$$

where $\sigma_{i=1,2}$ and $c_{i=1,2}$ are standard deviations and means associated to the Gaussian functions. Whereas the generalized bell-shaped MF is given by :

$$\mu(x;a,b,c) = \frac{1}{1+\left|\frac{x-c}{a}\right|^{2b}} \tag{2.20}$$

and the Π-shaped MF can be formulated as :

$$\mu(x;a,b,c,d) = \begin{cases} 0 & \text{if} \quad x \le a \\ 2\left(\frac{x-a}{b-a}\right)^2 & \text{if} \quad a \le x \le \frac{a+b}{2} \\ 1-2\left(\frac{x-b}{b-a}\right)^2 & \text{if} \quad \frac{a+b}{2} \le x \le b \\ 1-2\left(\frac{x-c}{d-c}\right)^2 & \text{if} \quad c \le x \le \frac{c+d}{2} \\ 2\left(\frac{x-d}{d-c}\right)^2 & \text{if} \quad \frac{c+d}{2} \le x \le d \\ 0 & \text{if} \quad x \ge d \end{cases} \tag{2.21}$$

The best membership function configurations for the channel length and channel thickness in the case of the prediction of the threshold voltage relative degradation is illustrated in Figures 2.7 and 2.8.

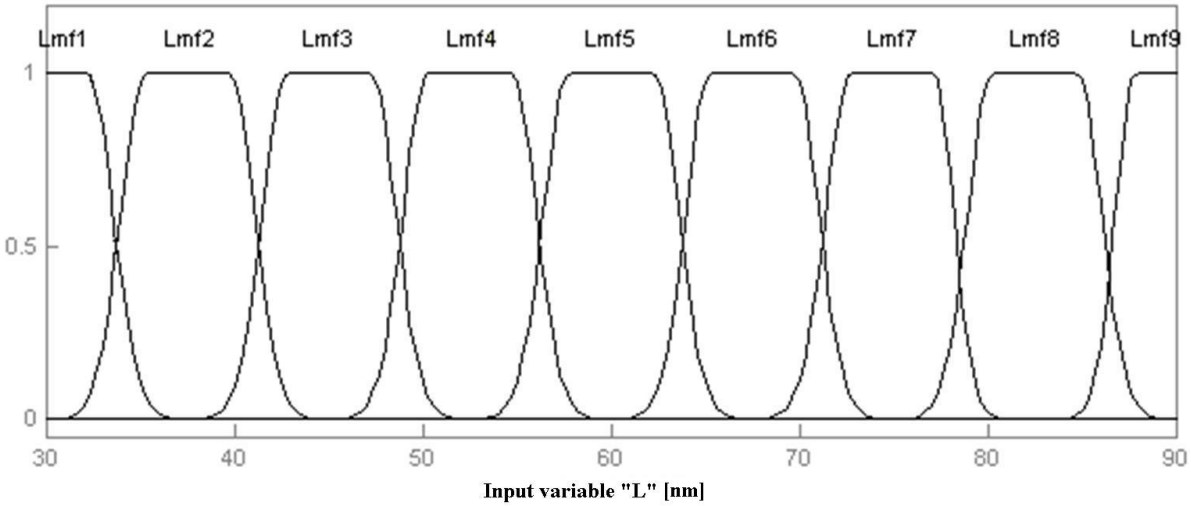

FIGURE 2.7 Partition of the channel length input parameter range using Gaussian combination membership functions in the case of the threshold voltage degradation.

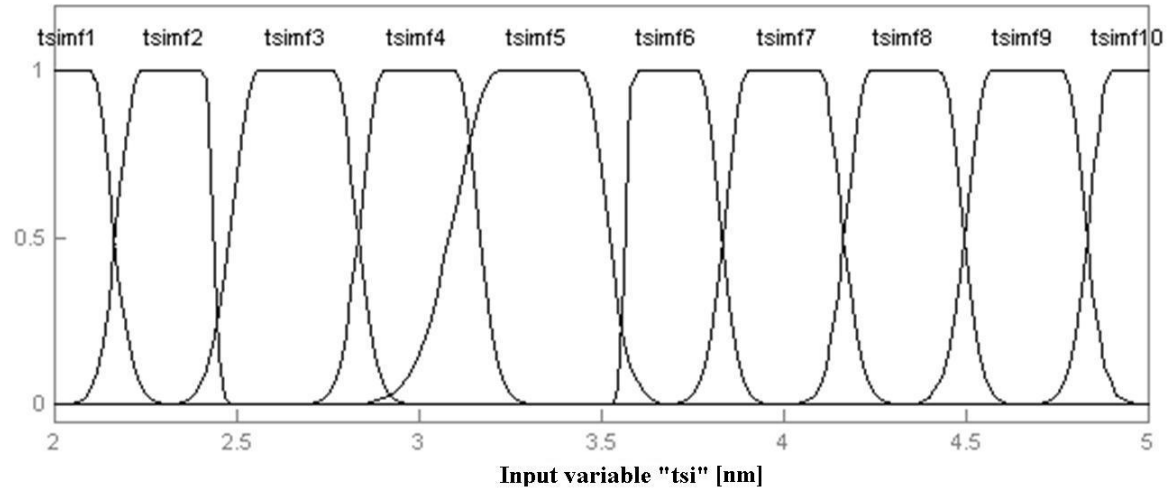

FIGURE 2.8 Partition of the channel thickness input parameter using Gaussian combination membership functions in the case of the threshold voltage degradation.

In order to show the obtained fuzzy rule responses, Figures 2.9, 2.10 and 2.11 depict the ANFIS controller rule surface for the variation of the relative degradation of the threshold voltage, drain-induced barrier lowering and the subthreshold current, respectively, as a function of both channel length and channel thickness. Since the threshold voltage and the drain-induced barrier lowering are correlated, they present a similar behavior characterized by the presence of many local optima, which expresses the modeling complexity of both parameters in contrast to the subthreshold current, which presents a

more shaped behavior. A similar situation has been deduced using analytical compact modeling for the case without introducing quantum effects, where the analytical models associated with the threshold voltage are more complicated and requires more elaborated methods for their resolution compared to those associated with the subthreshold current ([Bentrcia et al., 2009], [Djeffal et al., 2009a], [Djeffal et al., 2009b] and [Bentrcia et al., 2010]).

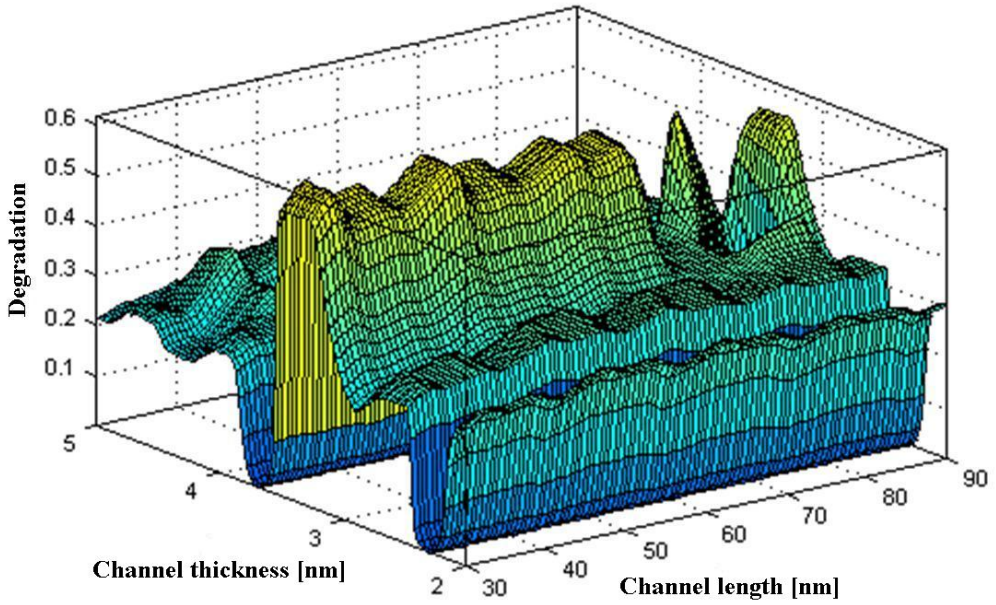

FIGURE 2.9 ANFIS controller rule surface for the threshold voltage relative degradation.

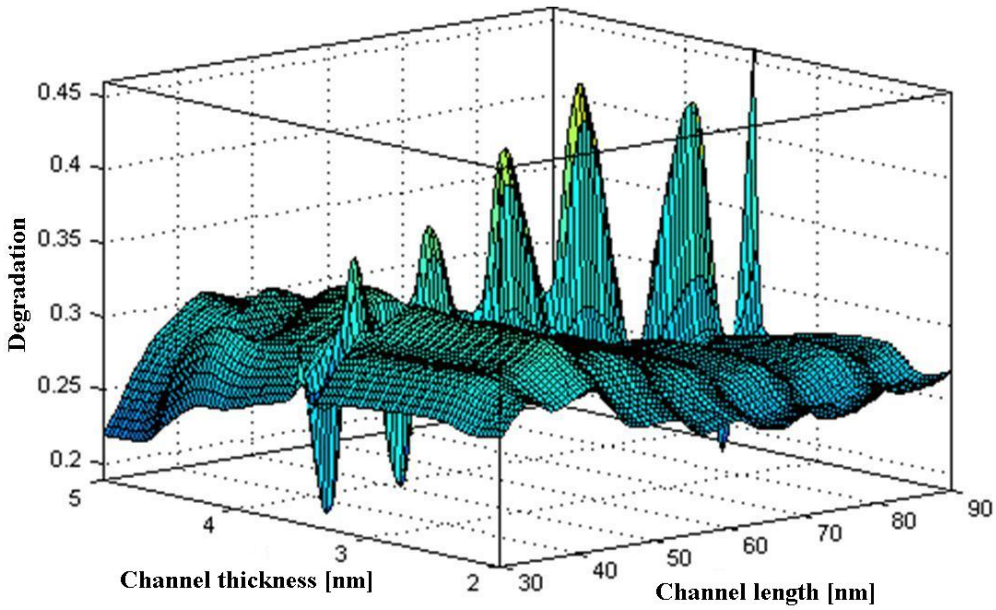

FIGURE 2.10 ANFIS controller rule surface for the DIBL relative degradation.

Figures 2.12 and 2.13 represent the regression curves of the threshold voltage relative degradation for the training and testing phases. From these figures, it is easy to note that

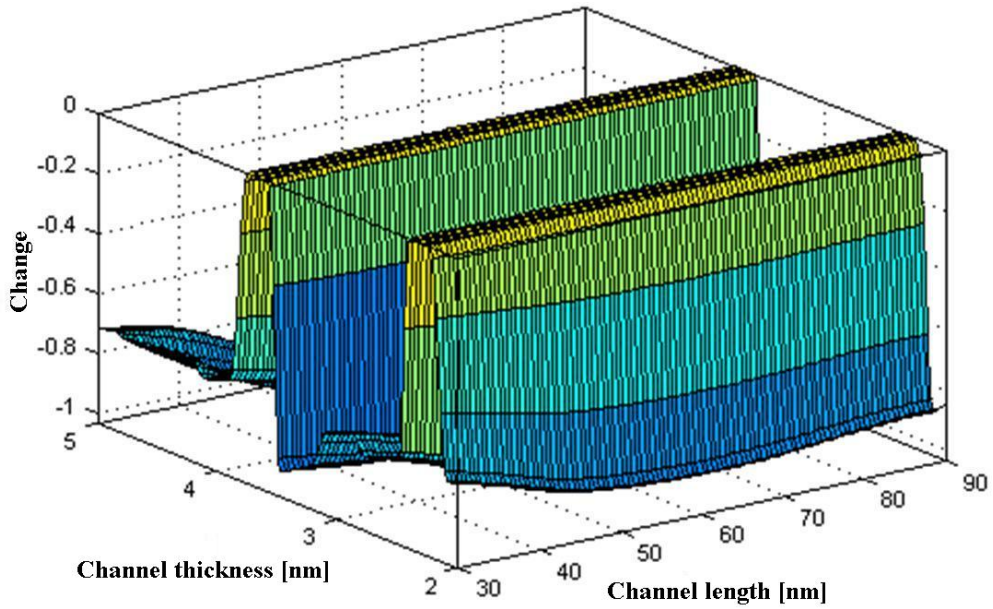

FIGURE 2.11 ANFIS controller rule surface for the subthreshold current relative degradation.

a positive correlation exists between the predicted and numerical values of the relative degradation in the case of the testing data set, which means that a sufficient agreement is satisfied between the predicted and numerical results.

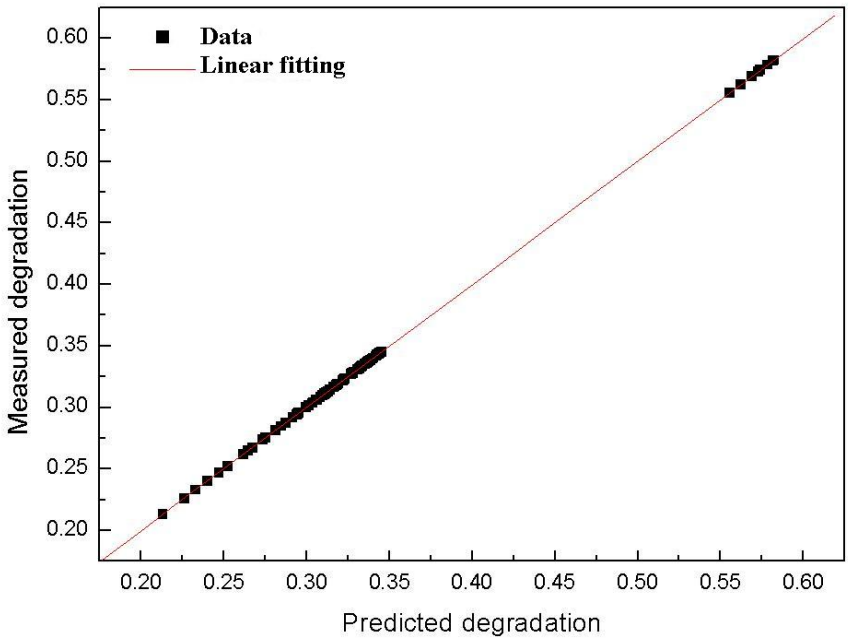

FIGURE 2.12 Regression line of the threshold voltage relative degradation for the training data.

Figures 2.14 and 2.15 represent the regression plots of the drain-induced barrier lowering and the subthreshold current relative degradations for the testing phase. In Figure 2.14, the correlation between the predicted and numerical values is negative in contrast

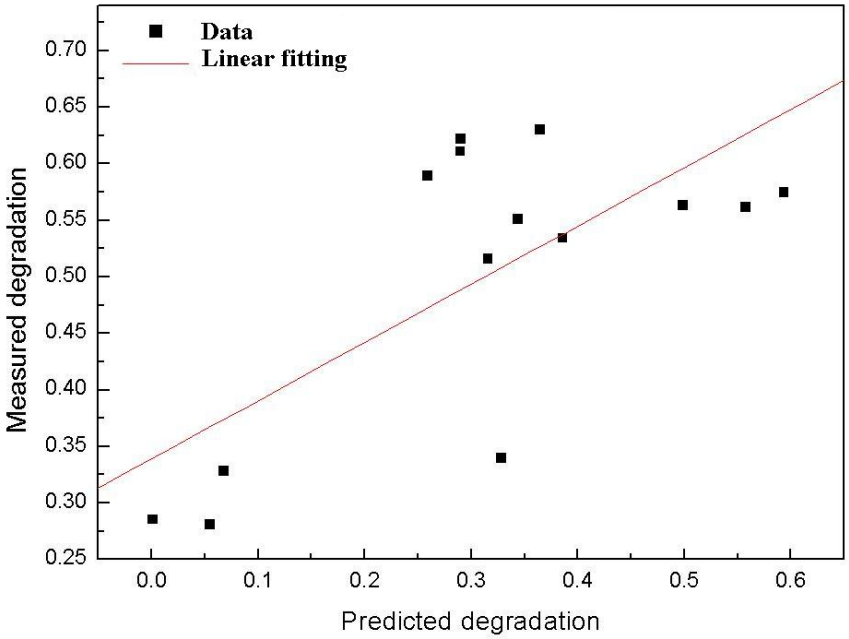

FIGURE 2.13 Regression line of the threshold voltage relative degradation for the testing data.

to Figure 2.15, which shows a positive correlation.

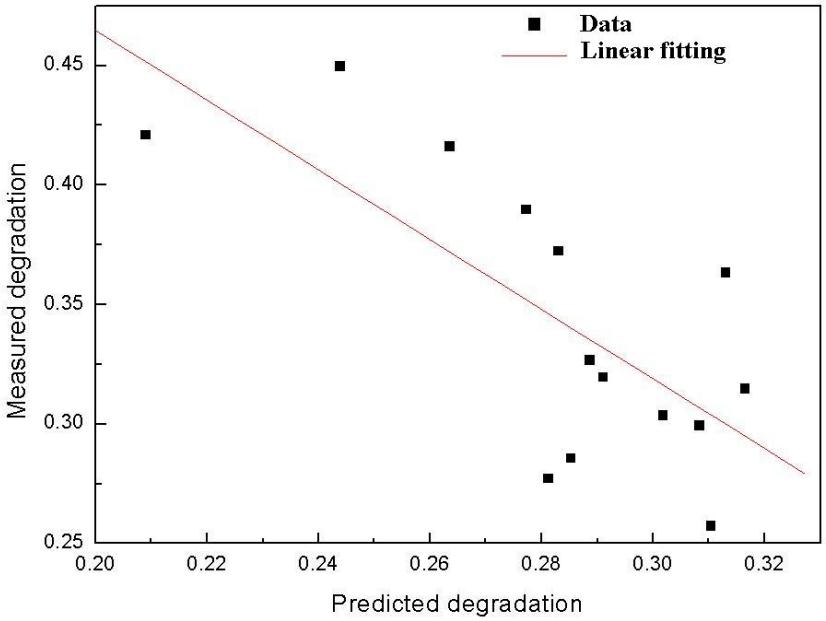

FIGURE 2.14 Regression line of the DIBL relative degradation for the testing data.

From these figures, it can be noticed that for different output variables, the elaborated fuzzy models were able to describe efficiently and with high accuracy, the general tendency behavior of the relative degradation in the threshold voltage, DIBL and subthreshold current. Hence, the adoption of fuzzy logic based frameworks can be of crucial importance in dealing with nanoscale devices, especially with the inclusion of the parasitic effects occurring at this level and making the analytical modeling an intractable or even

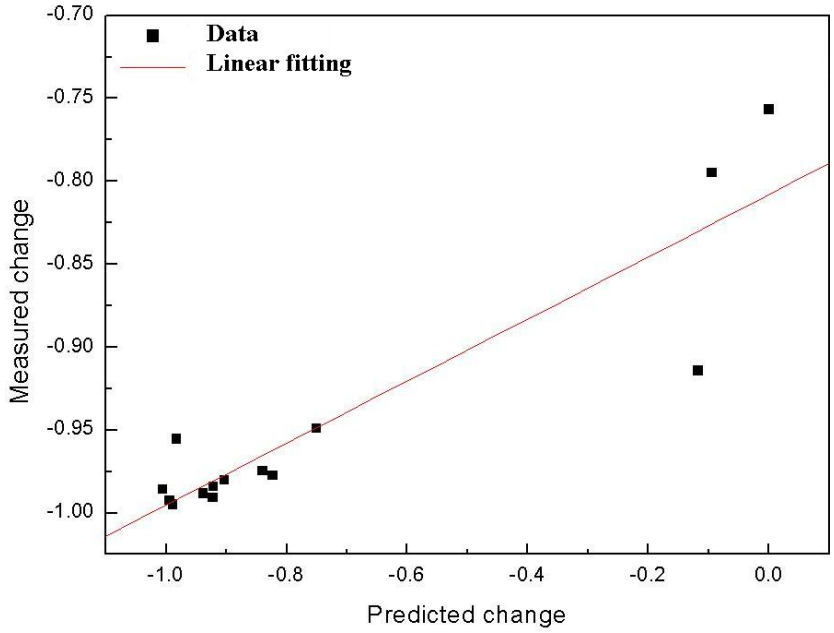

FIGURE 2.15 Regression line of the subthreshold current relative degradation for the testing data.

an impossible task to achieve. A summary of the main obtained results for both phases (training and testing phases) is provided in Table 2.2 for different subthreshold behavior parameters, where the statistical criteria are calculated and based on the following expressions :

$$\mathrm{MSE} = \frac{1}{\mathrm{N}_{obs}} \sum_{1}^{\mathrm{N}_{obs}} \left[\left(\frac{\Delta \mathrm{Y}}{\mathrm{Y}} \right)_{Num} - \left(\frac{\Delta \mathrm{Y}}{\mathrm{Y}} \right)_{Calc} \right]^2 \tag{2.22}$$

$$\mathrm{R} = \frac{Cov\left[\left(\frac{\Delta \mathrm{Y}}{\mathrm{Y}} \right)_{Num}, \left(\frac{\Delta \mathrm{Y}}{\mathrm{Y}} \right)_{Calc} \right]}{\sigma_{\left(\frac{\Delta \mathrm{Y}}{\mathrm{Y}} \right)_{Num}} \sigma_{\left(\frac{\Delta \mathrm{Y}}{\mathrm{Y}} \right)_{Calc}}} \tag{2.23}$$

TABLE 2.2 Summary of the main obtained results.

	MF		Training phase		Testing phase	
Parameter	**Type**	**Number**	**MSE(%)**	**R**	**MSE(%)**	**R**
V_{th}	Gaussian Combination	5×10 5×10	2.38×10^{-12}	1	0.05	0.71
DIBL	Generalized Bell-shaped	9×10 7×5	1.61×10^{-6}	1	0.01	-0.72
I_{sub}	Π-shaped	2×10	3.96×10^{-4}	0.99	0.13	0.91

2.5.2 Modeling of the threshold voltage behavior

As the accuracy of the trained ANFIS depends on the correctness and the effective representation of the data used during the learning procedure, a total of 91 observations are obtained by sampling the channel length and the channel thickness ranges with a step of 5 nm and 0.5 nm, respectively. These observations are divided into the training and the checking sets (65 and 14 observations for each set, respectively), where the latter is used to avoid the overfitting problem leading to bad prediction performances. The testing set is composed of 12 observations used to validate the prediction ability of the resulted fuzzy system. Because of the strong influence of membership function types on the quality of the obtained decision system, it is indispensable to select these functions in an optimal manner. The only solution is the trial and error approach due to the absence of any deterministic method that permits the specification of membership functions in the ANFIS. In this chapter, six different types of widely used membership functions are tested (Gaussian, Gaussian combination, sigmoid difference, generalized bell, Pi, and sigmoid product shaped MFs). We find that the best membership function leading to superior results in the calculation of the threshold voltage relative degradation is the Gaussian combination shaped membership function. The criteria of selection are based on the best values of the mean square error and the correlation coefficient relative to the training phase.

The number of membership functions for both parameters has been tested over values comprised between 2 and 15. The optimal values are found to be equal to 6 for the channel length and to 10 for the channel thickness. The number of the resulted fuzzy IF-THEN rules for the inference system is equal to 60 (6×10). The examination of the prediction ability of these fuzzy rules for testing confirms well that the use of the Gaussian combination shaped membership overpasses other shaped MFs in terms of accuracy. The partition of the input intervals according to the selected number and type of the membership functions for the channel length and thickness is presented in Figures 2.16 and 2.17, respectively.

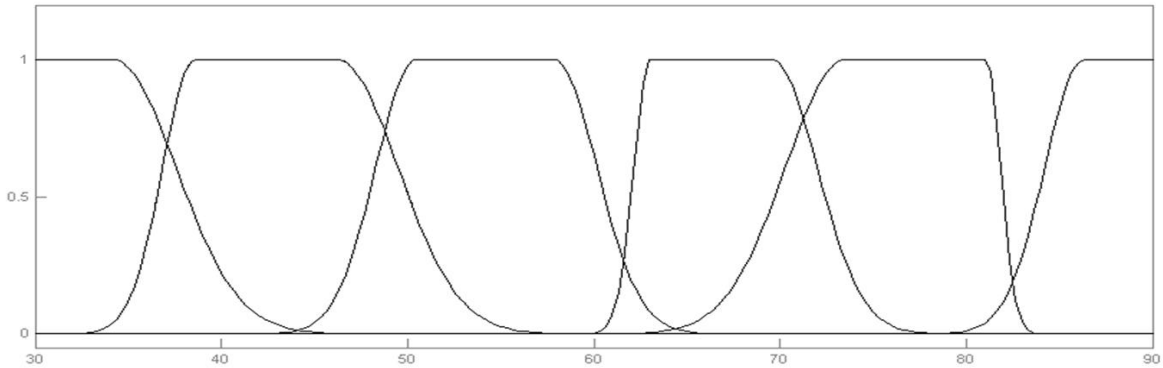

FIGURE 2.16 Partition of the channel length range using Gaussian combination shaped membership functions.

The output surface of the obtained ANFIS is visualized in Figure 2.18, where the predicted relative degradation of the threshold voltage along the variation ranges of inputs is plotted as a function of the channel length and channel thickness. As deduced from this graph, the highest degradation is located at the vicinity of a value equals to 3.5 nm for the channel thickness. The variation of the relative degradation is characterized by the presence of many local optima, making the analysis of the device immunity against the hot

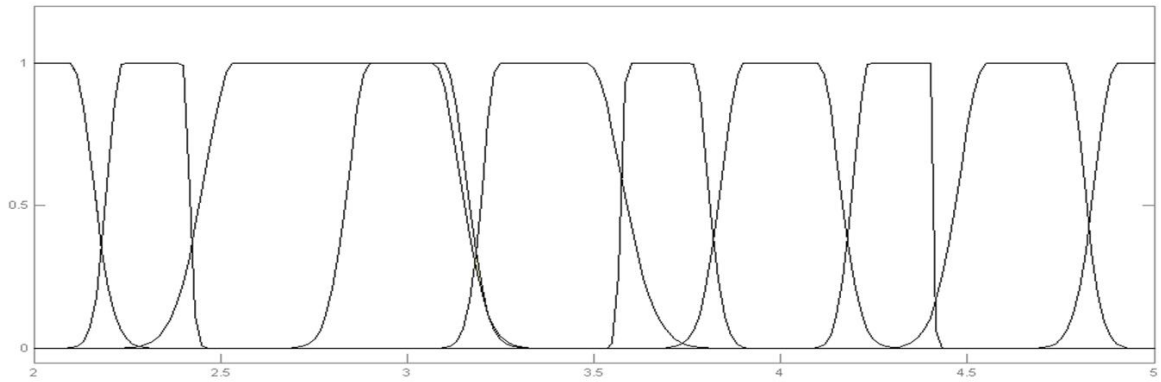

FIGURE 2.17 Partition of the channel thickness range using Gaussian combination shaped membership functions.

carrier degradation a very hard task especially when additional constraints are considered within the device.

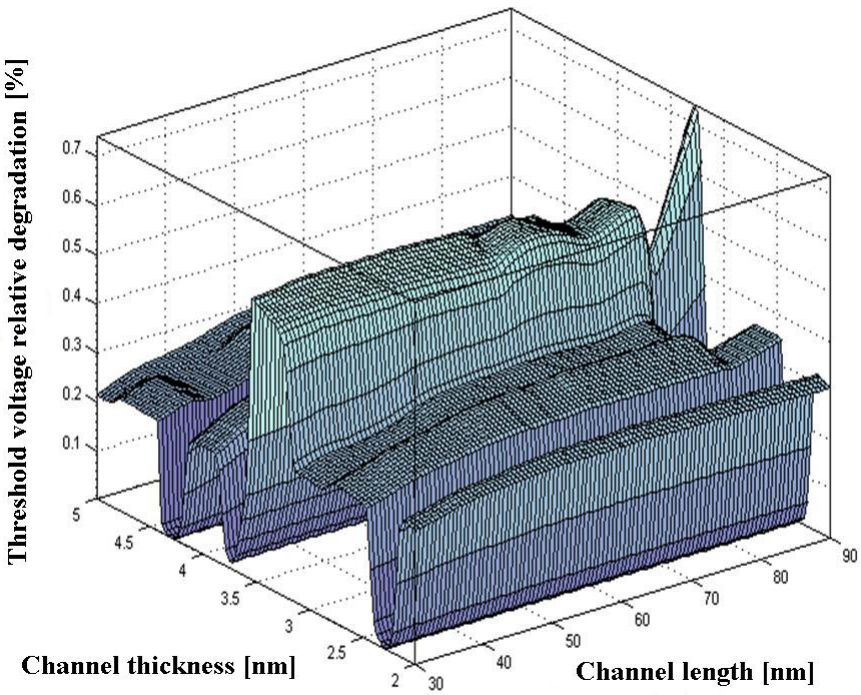

FIGURE 2.18 Response surface of the obtained ANFIS over the input parameter ranges.

The regression curves of the numerical and predicted data of the threshold voltage relative degradation are shown in Figures 2.19 and 2.20 for the training and testing data sets. It is easy to note that a sufficient agreement is satisfied between the predicted and numerical results especially in the training stage.

A more rigorous validation methodology can be conducted by comparing the performance criteria of our ANFIS based approach to other frameworks. A two layer feed-forward artificial neural network with sigmoid hidden and linear output layers has been used for this purpose. The number of neurons in the hidden layer is fixed to a value of 50 neurons, whereas the learning algorithm used to train the network is Levenberg-

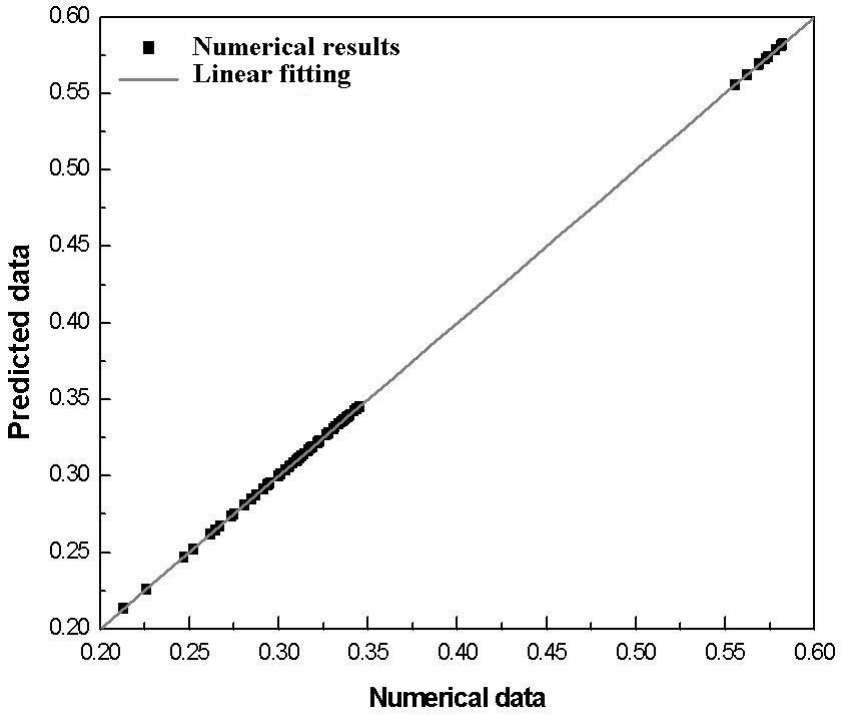

FIGURE 2.19 Regression plot of the threshold voltage relative degradation in the training data set.

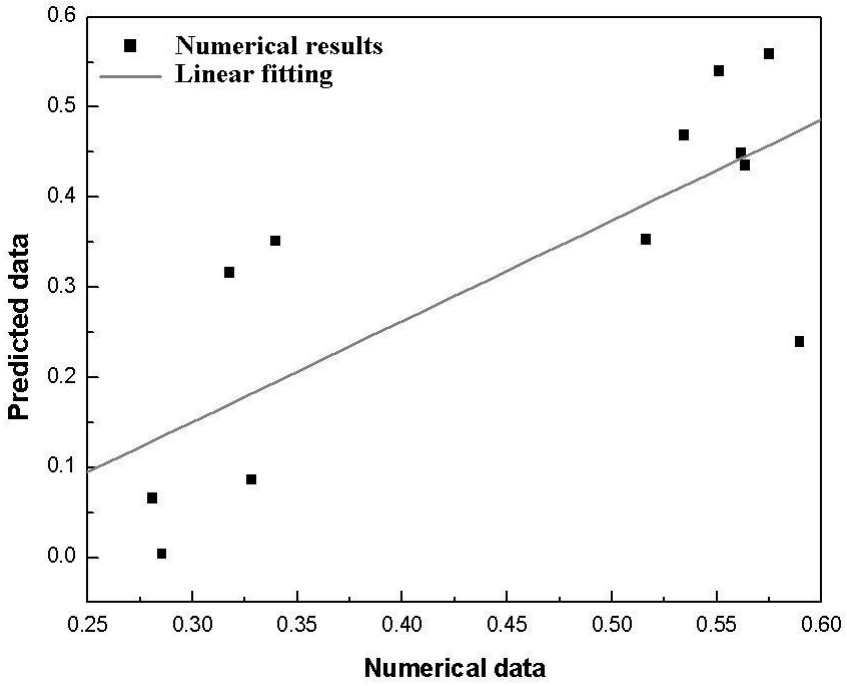

FIGURE 2.20 Regression plot of the threshold voltage relative degradation in the testing data set.

Marquard backpropagation algorithm. The statistical criteria for both formalisms are calculated and compared based on the following expressions :

$$\mathrm{MSE} = \frac{1}{\mathrm{N}_{obs}} \sum_{1}^{\mathrm{N}_{obs}} \left[\left(\frac{\Delta \mathrm{V}_{th}}{\mathrm{V}_{th}} \right)_{Num} - \left(\frac{\Delta \mathrm{V}_{th}}{\mathrm{V}_{th}} \right)_{Calc} \right]^2 \tag{2.24}$$

$$\mathrm{R} = \frac{Cov\left[\left(\frac{\Delta \mathrm{V}_{th}}{\mathrm{V}_{th}} \right)_{Num}, \left(\frac{\Delta \mathrm{V}_{th}}{\mathrm{V}_{th}} \right)_{Calc} \right]}{\sigma_{\left(\frac{\Delta \mathrm{V}_{th}}{\mathrm{V}_{th}} \right)_{Num}} \sigma_{\left(\frac{\Delta \mathrm{V}_{th}}{\mathrm{V}_{th}} \right)_{Calc}}} \tag{2.25}$$

A summary of the main results is provided in Table 2.3. It is clearly indicated that our proposed approach offers higher efficiency compared to the employed artificial neural network. This can be seen as a natural consequence of integrating hybrid learning algorithms with fuzzy logic tools.

TABLE 2.3 Comparison of ANN and ANFIS performance criteria.

	ANN		**ANFIS**	
Criterion	**Training**	**Testing**	**Training**	**Testing**
Mean squared errors	3×10^{-3}	0.13	2.43×10^{-8}	3×10^{-2}
Correlation coefficient	1	0.67	1	0.77

Furthermore, since the correlation coefficient for the testing set in the case of ANFIS is close to 0.8, we can say that we have a strong correlation between the numerical and the predicted relative degradation values of the threshold voltage. Consequently, it can be claimed that the performances of our fuzzy system are satisfactory and can be adopted for further analysis of more complicated integrated circuits.

2.5.3 Modeling of the Swing factor behavior

Despite that the transconductance is basically extracted in saturation regime since it is closely related to the device function in this regime, it is possible to deduce an analytical expression for such parameter based on the general subthreshold swing model given by :

$$\mathrm{S} = \left(\frac{\partial \mathrm{V}_{gs}}{\partial \log \mathrm{I}_{ds}} \right) = \left(\frac{1}{\ln(10)\,\mathrm{I}_{ds}} \times \frac{\partial \mathrm{I}_{ds}}{\partial \mathrm{V}_{gs}} \right)^{-1} = \ln(10) \left(\frac{\mathrm{I}_{ds}}{g_m} \right) \tag{2.26}$$

where V_{gs} is the gate voltage, I_{ds} the drain current and g_m the transconductance, which is defined in our case for the DG MOSFET device in the subthreshold regime. From the latter equation, we can obtain a compact model of the transconductance for the fresh and damaged devices as follows :

$$g_m = \ln(10) \left(\frac{\mathrm{I}_{ds}}{\mathrm{S}} \right) \tag{2.27}$$

In order to study the electrical behavior of nanoscale circuits, we propose the simulation of the nanoscale CMOS amplifier, which is considered as one of the most basic elements of digital VLSI circuits. Therefore, the main results obtained in the previous phases are used to evaluate the influence of the interface traps and quantum effects on the circuits' design, where the amplifier consisting of a nanoscale DG MOSFET device and a constant current source is considered. The gain of the circuit which is a basic parameter can be used to estimate the degradation of the amplifier operation and it is given by the formula :

$$\mathrm{G} = \left(\frac{g_m}{g_{ds}} \right) \tag{2.28}$$

where g_{ds} denotes the output conductance.

We determine the number of membership functions associated with the input parameters by exhaustively searching the product interval $[2-6] \times [2-6]$. Thus, the number of fuzzy IF-THEN rules for ANFIS is bounded by the following interval $[4-36]$. In this study, we suggest using Triangular, Generalized bell and Gaussian-shaped membership functions for the subthreshold swing parameter modeling.

After the running of the learning algorithm where the number of epoch is set to 200 and the comparison of the results generated by these membership functions, we find that the triangular function has higher rate of accuracy. Therefore, the number of membership functions attached to the input parameters is equal to $[6-4]$ in the case of the fresh device and to $[2-3]$ in the case of the damaged device. A summary of the main obtained results for the training and testing phases is provided for different device status in Table 2.4.

TABLE 2.4 Summary of the main obtained results.

	Training phase		**Testing phase**	
Device state	**MSE**	**R**	**MSE**	**R**
Fresh device	4.68×10^{-11}	1	1.98×10^{-9}	0.99
Damaged device	1.20×10^{-6}	0.98	7.48×10^{-6}	0.91

Once the training and testing phases' evaluation is achieved, it is possible to use obtained ANFIS models to extend our initial data set, which can be used in turn to calculate the transconductance values for the fresh and damaged devices. Based on these values, the relative degradation of the transconductance is illustrated in Figure 2.21 as a function of the channel length. It is shown that a high degradation is recorded around 40 nm. This observation can be explained by the complex mechanism caused by both the hot carrier and quantum effects. In addition, the degradation of the transconductance is reduced as a function of the channel length, for L more than 40 nm. In this case, the quantum effects can be neglected, and therefore, the transconductance degradation becomes less pronounced for long channel devices. This phenomenon can be explained by the fact that the degradation of the transistor parameters is caused only by the hot carrier effect.

Since the gain is among the principal parameters reflecting the efficiency of a single transistor based amplifier, it is also modeled under stress condition as a function of the channel length. The amplifier is composed of a nanoscale DG MOSFET and a constant current source as shown in Figure 2.22. The constant current source was set at $\mathrm{I}_{bias} = 10^{-3}\mathrm{A}$.

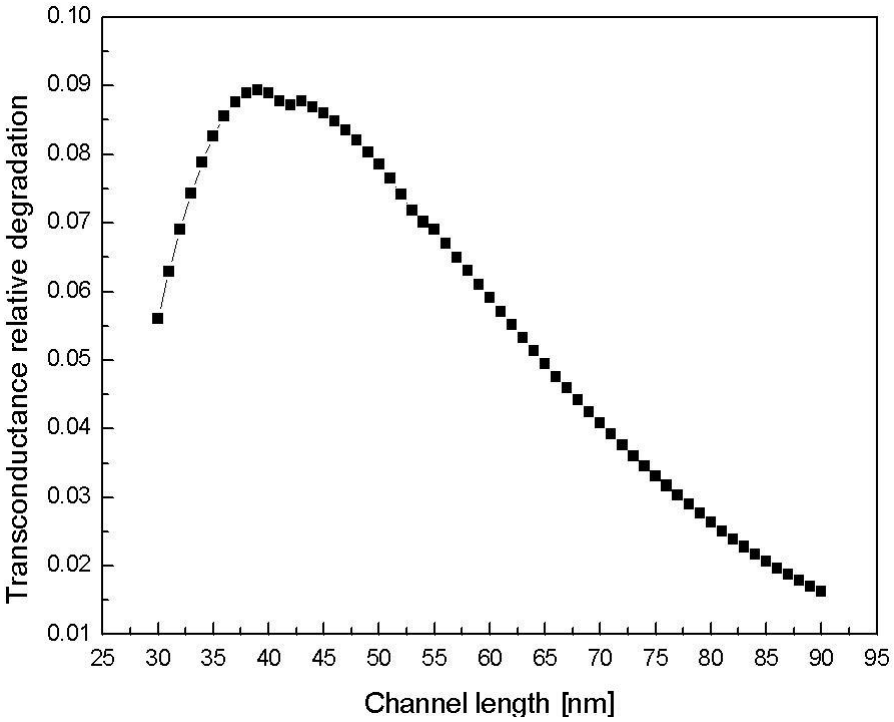

FIGURE 2.21 Transconductance relative degradtion as a function of the channel length for a channel thickness equals to 3 nm.

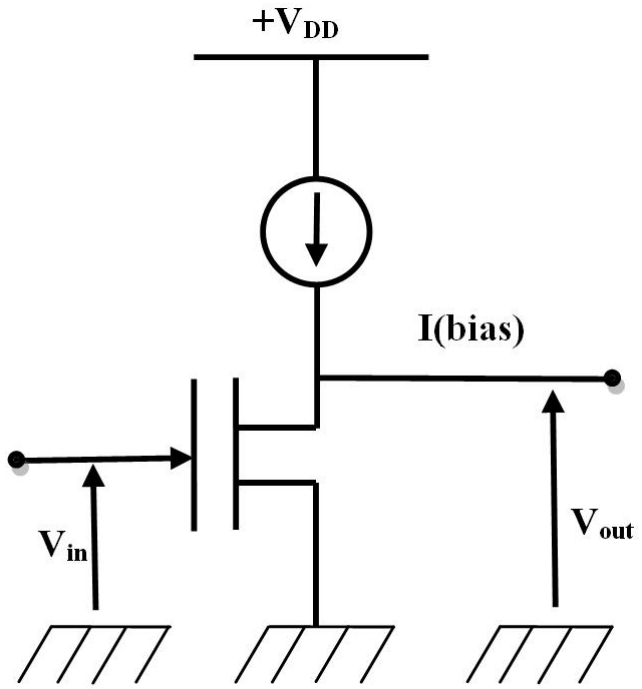

FIGURE 2.22 Circuit schema of a single DG MOSFET based amplifier.

The variation of the gain relative degradations with and without traps is illustrated in Figure 2.23. It is clear that the incorporation of the hot carrier and quantum confinement effects leads to a reduction of the performances of the amplifier. In this context, the gain for fresh device with the channel length equals to 30 nm, was about 34. However, the gain

is degraded to about 32, for an interface charge density equals to 5×10^{12} cm^{-2}. In our study, we assume that the output conductance of the amplifier is considered as constant in subthreshold regime, where it is fixed to 10^{-3}S.

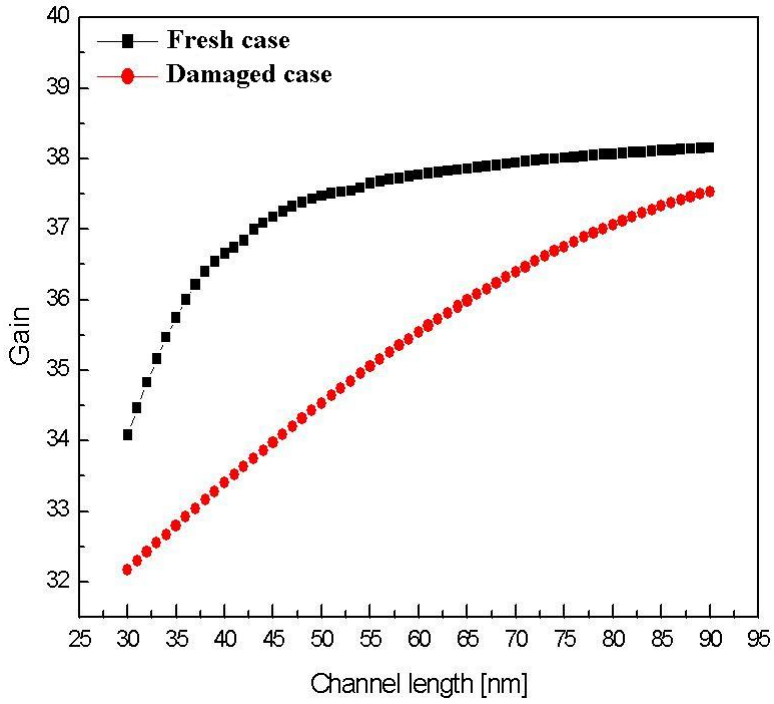

FIGURE 2.23 Gain of the DG MOSFET amplifier as a function of the channel length for a channel thickness equals to 3 nm.

2.6 Conclusion

In this chapter, an ANFIS based approach has been proposed to efficiently estimate the relative degradation of the threshold voltage, the drain-induced barrier lowering and the subthreshold current, which are considered the principal parameters characterizing the CMOS based devices in the subthreshold regime. The input parameters have been limited to the channel length and the channel thickness since they have the major impact on parasitic effects. The elaboration of the training, validation and testing benchmarks has been made easy thanks to ATLAS 2D simulator, where a hybrid learning algorithm has been used to configure the different premises and consequent parameters included within the IF-THEN rules. A comparison with an artificial neural network simulation has been conducted in terms of the threshold voltage criterion. The obtained performance has demonstrated that the developed fuzzy logic based approach offers higher efficiency encouraging its implementation in electronic device simulators to study the nanoscale CMOS circuits including the aging phenomenon. The interesting features of the ANFIS capabilities in dealing with the complexity and uncertainty related to the phenomenological description of the device behavior at a nanoscale level which are statistical in nature, make our proposed approach an adequate tool for future reliability prediction purposes. It should be noted also that additional device parameters such as the oxide thickness parameter can be included to extend the proposed ANFIS based approach so that more accuracy can be gained in the device behavior modeling.

2.7 References

Abdelhamid, H. M. (2007). *Compact Modeling of Multiple Gate MOS Devices*. PhD thesis, University of Rovira i Virgili. 42

Abdi, M. A., Djeffal, F., Meguellati, M., and Arar, D. (2009). Two-dimensional analytical threshold voltage model for nanoscale graded channel gate stack DG MOSFETs. In 16^{th} *IEEE International Conference on Electronics, Circuits, and Systems, Hammamet, Tunisia*. 39

Akdemïr, B., Güneş, S., Oran, B., and Karaaslan, S. (2010). Prediction of cardiac end-systolic and end-diastolic diameters in m-mode values using adaptive neural fuzzy inference system. *Expert Systems with Applications*, 37(08) :5720–5727. 44

Amat, E., Rodríguez, R., Nafría, M., and Aymerich, X. (2009). Channel hot-carrier degradation under AC stress in short channel nMOS devices with high-κ gate stacks. *Microelectronic Engineering*, 86(7-9) :1908–1910. 39

Arora, N. (2007). *MOSFET Modeling for VLSI Simulation Theory and Practice*. World Scientific Press. 43

Association, S. I. (2009). International technology roadmap for semiconductors. Technical report. 39

Balaguer, M., Roldan, J. B., Donetti, L., and Gamiz, F. (2012). Inversion charge modeling in n-type and p-type double-gate MOSFETs including quantum effects : The role of crystallographic orientation. *Solid-State Electronics*, 67(01) :30–37. 43

Bendib, T. and Djeffal, F. (2011). Electrical performance optimization of nanoscale double-gate MOSFETs using multi-objective genetic algorithms. *IEEE Transactions on Electron Devices*, 58(11) :3743–3750. 40

Bentrcia, T. and Djeffal, F. (2011). *Compact modeling of multi-gate MOSFET including hot-carrier effects*, chapter 04. in M. J. Kwon (eds.), CMOS Technology, Nova Science Press. 39, 42

Bentrcia, T., Djeffal, F., Abdi, M. A., Chahdi, M., and Boukhennoufa, N. (2009). An accurate two dimensional threshold voltage model for nanoscale GCGS DGMOSFET including traps effects. In 3^{rd} *IEEE International Conference on Signals, Circuits and Systems, Djerba, Tunisia*. 42, 54

Bentrcia, T., Djeffal, F., and Arar, D. (2010). An analytical two dimensional subthreshold current model for nanoscale GCGS DG MOSFET including interfacial traps effects. *Revue des Sciences et de la Technologie*, 1(2) :103–112. 54

Bentrcia, T., Djeffal, F., Arar, D., and Meguellati, M. (2013a). ANFIS-based computation to study the nanoscale circuit including the hot-carrier and quantum confinement effects. In *The* 5^{th} *International Conference on Modeling, Simulation and Applied Optimization, Hammamet, Tunisia*. 40

Bentrcia, T., Djeffal, F., and Benhaya, A. (2012a). Continuous analytic IV model for GS DG MOSFETs including hot-carrier degradation effects. *Journal of Semiconductors*, 33(01) :1–6. 40

Bentrcia, T., Djeffal, F., and Chahdi, M. (2013b). An analytical two dimensional subthreshold behavior model to study the nanoscale GCGS DG Si MOSFET including interfacial trap effects. *Microelectronics Reliability*, 53(04) :520–527. 39

Bentrcia, T., Djeffal, F., and Chebaki, E. (2012b). ANFIS-based approach to studying subthreshold behavior including the traps effect for nanoscale thin-film DG MOSFETs. *Journal of Semiconductors*, 34(08) :1–9. 40

Bentrcia, T., Djeffal, F., Meguellati, M., and Arar, D. (2013c). New approach based on ANFIS computation to study the threshold voltage behavior including trap effects for nanoscale DG MOSFETs. In *The World Congress on Engineering, London, UK*. 40

Chebaki, E., Djeffal, F., and Bentrcia, T. (2012). Two-dimensional numerical analysis of nanoscale junctionless and double gate MOSFETs including the effect of interfacial traps. *Physica Status Solidi C*, 09(10-11) :2041–2044. 39

Difrenza, R., Linares, P., and Ghibaudo, G. (2003). The impact of short channel and quantum effects on the MOS transistor mismatch. *Solid-State Electronics*, 47(07) :1161–1165. 39

Djeffal, F., , Bendib, T., and Abdi, M. A. (2011a). A two-dimensional semi-analytical analysis of the subthreshold-swing behavior including free carriers and interfacial traps effects for nanoscale double-gate MOSFETs. *Microelectronics Journal*, 42(12) :1391–1395. 42

Djeffal, F., Bentrcia, T., Abdi, M. A., and Bendib, T. (2011b). Drain current model for undoped gate stack double gate (GSDG) MOSFETs including the hot-carrier degradation effects. *Microelectronics Reliability*, 51(03) :550–555. 40, 42

Djeffal, F., Bentrcia, T., and Bendib, T. (2011c). An analytical drain current model for undoped GSDG MOSFETs including interfacial hot-carrier effects. *Physica Status Solidi (c)*, 08(03) :907–910. 40, 41

Djeffal, F., Dibi, Z., Hafiane, M. L., and Arar, D. (2007). Design and simulation of a nanoelectronic DG MOSFET current source using artificial neural networks. *Materials Science and Engineering : C*, 25(7-8) :1111–1116. 40

Djeffal, F., Ghoggali, Z., Dibi, Z., and Lakhdar, N. (2009a). Analytical analysis of nanoscale multiple gate MOSFETs including effects of hot-carrier induced interface charges. *Microelectronics Reliability*, 49(04) :377–381. 54

Djeffal, F., Meguellati, M., and Benhaya, A. (2009b). A two-dimensional analytical analysis of subthreshold behavior to study the scaling capability of nanoscale graded channel gate stack DG MOSFETs. *Physica E : Low-Dimensional Systems and Nanostructures*, 41(10) :1872–1877. 54

Eisberg, R. M. and Renick, R. (1969). *Notes on Modern Physics*. Wiley Press. 42

Galup-Montoro, C. and Schneider, M. H. (2007). *MOSFET Modeling for Circuit Analysis and Design*. World Scientific Press. 43

Ghoggali, Z., Djeffal, F., and Lakhdar, N. (2010). Analytical analysis of nanoscale double-gate MOSFETs including the hot-carrier degradation effects. *International Journal of Electronics*, 97(02) :119–127. 40

Ghosh, P., Haldar, S., Gupta, R. S., and Gupta, M. (2012). An analytical drain current model for dual material engineered cylindrical/surrounded gate MOSFET. *Microelectronics Journal*, 43(01) :17–24. 39

Ho, C. S., Huang, K. Y., Tanga, M., and Liou, J. J. (2005). An analytical threshold voltage model of NMOSFETs with hot-carrier induced interface charge effect. *Microelectronics Reliability*, 45(7-8) :1144–1149. 42

Jang, J. S. R. (1993). ANFIS : Adaptive-network-based fuzzy inference system. *IEEE Transactions on Systems, Man, and Cybernetics*, 23(03) :665–685. 44

Jiang, H. M., Kwong, C. K., Ip, W. H., and Wong, T. C. (2012). Modeling customer satisfaction for new product development using a PSO-based ANFIS approach. *Applied Soft Computing*, 12(02) :726–734. 47

Kaur, H., Kabra, S., Haldar, S., and Gupta, R. S. (2007). An analytical drain current model for graded channel cylindrical/surrounding gate MOSFET. *Microelectronics Journal*, 38(03) :352–359. 39

Lawson, C. L. and Hanson, R. J. (1974). *Solving Least Squares Problems*. Prentice-Hall Press. 47

Lo, S. P. (2003). An adaptive-network based fuzzy inference system for prediction of workpiece surface roughness in end milling. *Journal of Materials Processing Technology*, 142(03) :665–675. 44

Mahapatra, S., Parikh, C. D., Ramgopal Rao, V., Viswanathan, C. R., and Vasi, J. (2000). A comprehensive study of hot-carrier induced interface and oxide trap distributions in MOSFETs using a novel charge pumping technique. *IEEE Transactions on Electron Devices*, 47(01) :171–177. 42

Morris, H. C. and Abebe, H. (2010). A compact quantum surface potential model for a MOSFET device. *Mathematical and Computer Modelling*, 51(7-8) :893–900. 43

Nariman-Zadeh, N., Darvizeh, A., and Dadfarmai, M. H. (2004). Design of ANFIS networks using hybrid genetic and SVD methods for the modelling of explosive cutting process. *Journal of Materials Processing Technology*, 155-156 :1415–1421. 47

Naseh, S., Jamal Deen, M., and Chen, C. H. (2006). Hot-carrier reliability of submicron NMOSFETs and integrated NMOS low noise amplifiers. *Microelectronics Reliability*, 46(2-4) :201–212. 42, 43

Pagey, M. P. (2003). *Hot-Carrier Reliability Simulation in Aggressively Scaled MOS Transistors*. PhD thesis, Vanderbilt University. 39

Prégaldiny, F., Lallement, C., and Mathiot, D. (2004). Accounting for quantum mechanical effects from accumulation to inversion, in a fully analytical surface-potential-based MOSFET model. *Solid-State Electronics*, 48(05) :781–787. 40

Silvaco (2012). *Atlas User Manual : Device Simulation Software*. 48

Singh, A. K. (2008). An analytical study of hot-carrier degradation effects in sub-micron MOS devices. *The European Physical Journal Applied Physics*, 42(02) :87–94. 42

Singh, M. (2010). *Adaptive Network-Based Fuzzy Inference Systems for Sensorless Control of PMSG Based Wind Turbine with Power Quality Improvement Features*. PhD thesis, University of Quebec. 44

Sverdlov, V. A., Walls, T. J., and Likharev, K. K. (2003). Nanoscale silicon MOSFETs : A theoretical study. *IEEE Transactions on Electron Devices*, 50(09) :1926–1933. 39

Tyaginov, S. E., Starkov, I. A., Triebl, O., Cervenka, J., Jungemann, C., Carniello, S., Park, J. M., Enichlmair, H., Karner, M., Kernstock, C., Seebacher, E., Minixhofer, R., Ceric, H., and Grasser, T. (2010). Interface traps density-of-states as a vital component for hot-carrier degradation modeling. *Microelectronics Reliability*, 50(09-11) :1267–1272. 41

Wen, T. P. and Singh, A. K. (2010). A comprehensive analytical study of an undoped symmetrical double-gate MOSFET after considering quantum confinement parameter. *Microelectronics Journal*, 41(2-3) :162–170. 40

Ying, L. C. and Pan, M. C. (2008). Using adaptive network based fuzzy inference system to forecast regional electricity loads. *Energy Conversion and Management*, 49(02) :205–211. 44

Zadeh, L. A. (1973). Outline of a new approach to the analysis of complex systems and decision processes. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-3(1) :28–44. 44

Chapter 3

A surrogate based approach for the modeling and design of nanoscale DG MOSFETs including interface traps

« The discovery of what matters is important. If you figure it out, you know WHAT to optimize, and WHERE. One huge step closer to more winning tests with bigger impact. »

Peep Laja

Abstract- *In response to the high computational costs related to the development of physical based models for Double Gate MOSFETs including short channel, quantum confinement and hot carrier effects, more flexible alternatives have been proposed for the prediction of device performances. Our aim in this chapter is to investigate the efficiency of a new proposed framework, built upon kriging metamodeling and Non-dominated Sorting Genetic Algorithm version II, for the optimal design in terms of subthreshold and analog/RF criteria. The input variables of interest are limited to the geometrical parameters namely the channel length and the channel thickness. Data generated according to computer experiments, based on ATLAS 2D simulator, are used to identify and adjust kriging surrogate models. It is emphasized that the obtained models can be used accurately in a multi-objective context to offer several Pareto configurations. Therefore, a wide range of selection possibilities is available to the designer depending on situations under consideration.*

3.1 Introduction

Within the last few years, the interest in miniature electronic artifacts is growing rapidly because of the strong demand implied by today's applications. Such shrinking trend has motivated the semiconductor industry to further invest in the development of novel device designs jointly with implementation techniques [Geng, 2005]. This is conducted in order to follow Moore's law predictions down to the deep nanoscale realm. In fact, nanoscale sized devices have become the cornerstone for building integrated circuits with dimensions and performance limits well beyond our optimistic estimations. However, according to the International Technology Roadmap for Semiconductors (ITRS), Complementary Metal Oxide Semiconductor (CMOS) scaling technology is approaching the fundamental physical limits in the near future as imposed by the properties and dimensions of the employed semiconductor material layers [http ://itrs.net, 2013].

Even though bulk Metal Oxide Semiconductor Field Effect Transistor (MOSFET) is still of value for the 100 nm node, the reduction of the channel length leads to some parasitic degradation effects like variability problems resulted mainly from the random discrete doping ([Bindu et al., 2010] and [Asenov et al., 2009]). The Double Gate Metal Oxide Semiconductor Field Effect Transistors (DG MOSFET) have emerged as a competitive option to boost the performance of CMOS devices. The presence of two gates in the design permits to get a more robust electrostatic control over carriers inside the channel body, which remedies significantly the harmful parasitic effects. Motivated by all these benefits from economical and technological viewpoints, the use of DG MOSFET devices seems to be of paramount utility for the next generations of electronic circuits [Bentrcia et al., 2014]. In fact, many challenging burdens result from the continuous downscaling of such devices into deep nanometer domain. For instance, the short channel effect leads to less control over the potential distribution in addition to the overlap of the depletion regions associated with both channel and source/drain junctions [Djeffal et al., 2008]. Due to the quantum mechanical effect, the reduction of the channel thickness leads to the confinement of carriers in a thin layer, which affects the transport and scattering properties [Wagner et al., 2006]. Moreover, the intensive applied voltages allow carriers to be accumulated in the oxide near the drain side. This aging phenomenon known as the hot carrier effect contributes to the degradation of the overall performance of the device [Bentrcia and Djeffal, 2011].

The key factors that limit how far a DG MOSFET can be downscaled depend on three crucial parameters namely the channel length, the channel thickness and the gate oxide thickness. For a gate oxide thickness less than two nanometers, gate leakage currents induced by tunneling phenomena cross through the dielectric region. Moreover, this unwanted effect leads to an increase in power dissipation in addition to the alteration of the circuit stability [Darbandy et al., 2012]. Roughly speaking, it is possible to categorize various degradation mechanisms altering the performance of DG MOSFET devices based on the three geometrical aforementioned parameters. The short channel effect arises because of the channel length reduction, where the close proximity between source and drain affects the gate control over the potential distribution and current flowing inside the channel in addition to the depletion region in the channel due to the gate overlap with that associated to the source/drain junctions. As a result, the effective charge controlled by gates becomes smaller because of such overlapping of fields. The maximum channel barrier at the source side region might be lowered by the lateral drain field to produce in turn a reduced threshold voltage [Bentrcia and Djeffal, 2014]. In the quantum mechanical effect, the diminution of the silicon layer thickness gives rise to a high degree of geo-

metrical confinement, which leads to the creation of high electric fields at silicon/oxide boundaries besides a well between the oxide and the silicon potentials. Hence, under inversion conditions, carriers are confined in this potential well where they occupy only discrete (quantized) energy levels. The quantum confinement is characterized by two physical phenomena; the threshold voltage in this case is higher compared to its conventional counterpart. The second aspect concerns the charge distribution, which behaves in a different way since it does not reach its maximum at the oxide/semiconductor interface, but inside the silicon film, whereas it virtually vanishes right at the interface ([Mohammadi and Afzali-Kusha, 2010] and [Wang et al., 2010]). The last category of the undesirable effects occurring in nanoscale DG MOSFET devices is the hot carrier effect considered as a major reliability concern. Basically, the application of high gate voltages above the threshold value in deeply scaled channels results in extremely elevated electric fields, which allows carriers to acquire high kinetic energies. After long duration of the device operation, these carriers, known as hot carriers, are accumulated near the drain junction of the transistor and generate a pinch-off region having different properties with the pileup of localized non-uniform interface traps [Bentrcia and Djeffal, 2011]. Since both carrier types (electrons and holes) contribute to the hot carrier injection mechanism, it should be highlighted that the electrons are mostly collected at the drain, while holes are driven back from the gate/drain and flow towards the substrate. The obtained substrate current may be used as a measure of the hot carrier generation, where it is possible to deduce the most damaging biasing conditions by determining gate and drain voltages associated to the occurrence of the substrate current peaks [Amat et al., 2013].

The mathematical interpretation of all previous degradation mechanisms can be formulated using equations that reflect the physical behavior of the device. These equations or models may be valid along all regimes or only over a limited region of the DG MOSFET operation and allow computing the device performance measures as a function of the electrical or geometrical parameters ([Bentrcia et al., 2012a] and [Taur, 2000]). Since the modeling frameworks can be viewed as a passage from the integrated circuit design stage to the mature process technology development stage, the development of a model must respect the compromise between accuracy, computational cost and simplicity of deployment to meet the challenging requirements of nowadays circuit designs.

Despite the fact that the double gate MOSFET is possibly the most studied since the pioneering work of Taur [Taur, 2000], there is still a lack of reliable compact description of the nanoscale device responses in above threshold regime. The lion's share of published works is dedicated to the analytical or numerical modeling of such devices with undoped channel core and symmetric biasing conditions. It is to mention that these works feature complicated formulations, which need approximation hypotheses for treatment. Besides, the developed expressions in some cases show continuity or region validity problems between different operating regimes ([Abdelhamid, 2007] and [Chaudhry, 2013]). Intelligent frameworks such as artificial neural networks or fuzzy inference systems have been also exploited for this purpose but with less efficiency in providing closed forms due to their consideration as black box routines ([Djeffal et al., 2007] and [Bentrcia et al., 2012b]). In response to these drawbacks, empirical models have become a powerful tool offering a wide range of possibilities for the designer. We distinguish two procedures to adjust an empirical model coefficients : (i) either using a dedicated TCAD software simulator, or (ii) from electrical measurements on a number of test devices with various configurations [Arora, 2007].

From a general classification perspective, it is appropriate to divide existing modeling tools into two basic families, namely numerical simulation based models and compact

models. Numerical simulators known also as device simulators provide rich functionality toolboxes for the investigation of different properties of a device for instance : the electrical, optical and thermal behavior. Without exaggeration, numerical methods can be qualified as exact modeling approaches due to their very high degree of precision. The justification behind is that the full solution of the charge distribution in the channel is obtained by two or three dimensional solving of the coupled Poisson's, Schrödinger, and the transport equations self-consistently. On the other hand, the requirements of numerical methods in terms of intensive computation and enormous storage memory size prohibit their application for circuit simulation but they can be used as a reference or to consolidate the efficiency of other modeling strategies [Camiola et al., 2012]. Regarding the compact modeling paradigm, it is aimed at providing an easy to use representation, which is as accurate as possible, of different device characteristics. Compact models are considered of a big deal when used for the performance evaluation of integrated circuits including a large number, sometimes several thousands, of components. The DG MOSFET compact models fall into three subtypes : physical models, table lookup models and empirical models. Physical device models profit from the deep knowledge available on the device geometry and carrier transport equations. The parameters of the model obtained in this case have physical meaning and the prediction ability is supported by simply modifying the device configuration values. However, the compact modeling of modern VLSI devices with their innovative enhancements needs significant time and often requires major modification of the existing models. The table lookup models contain device data stored in a tabular format for different bias configurations and device geometries. Such database, collected from experimental tests or generated from device level simulators, is accessed for design or analysis purposes once the table lookup model is executed. The disadvantage of such modeling procedure is that it gives no physical insights regarding the device behavior in addition to the uncertainty of the model validity outside the picked data range. The empirical models are expressed by equations based entirely on data fitting. The accuracy of prediction in this case is fully dependent on the sample size values being used that do not have a well defined physical meaning. It should be stated out that the number of the fitting parameters augments with the complexity of the device structure under study. Two sources of data may be used to tune up the model parameters either from a device level numerical simulator or from electrical measurements on a number of test devices having distinct dimensions [Gildenblat, 2010].

The prevailing advance in dedicated Technology Computer Aided Design (TCAD) software packages has stimulated the replacement of physical experiments, conducted for sensitivity analysis, with computer experiments. The principal difference between both kinds of experiments dwells on the absence of random errors in computer based experiments due to the deterministic nature of implemented programs. Thus, the unique source of uncertainty is related to the lack of knowledge about the relationship formula between the input factors and responses [Schonlau, 1997]. Typically, numerical simulations are the only practical tool suitable for dealing with complicated DG MOSFET structures essentially when subject to critical conditions. Because of the expensive computational burdens related to a numerical simulation model, the assessment of the device performance measures is based on metamodeling technique, in which a mathematical model surrogate of the device performance is used to approximate the relationship between the system performance measure and the set of design parameters. Metamodeling techniques have emerged tremendously in manufacturing industries as a framework for the establishment of surrogate compact models of expensive processes. A metamodeling technique is based on a limited number of real or numerical experiments and allows replacing the original

simulation model with less computational costs and satisfactory precision. The investigation of the metamodeling technique is a phase posterior to numerical simulations in order to build surrogate models, which can be used either for forecasting a system performance for unknown configurations or for adjusting the system configuration to optimize the overall performance. Hence, the generated surrogate models may serve for prediction or optimization tasks of the design once they are proved to be valid ([Gary Wang and Shan, 2006] and [Zhang et al., 2014]). A generic metamodeling procedure is composed of two basic stages. Firstly, a set of sample points in the design parameter space is selected. Then, adequate statistical of the considered performance measure is fitted to the sample points [Sacks et al., 1989]. Therefore, the accuracy of a surrogate model is significantly influenced by the experimental design used to select data points, the size of the design space and the accuracy of the simulation at each data point. Once a consistent model is accredited, it may serve for prediction or optimization tasks of the modeled DG MOSFET device.

In this chapter, we propose a new approach for the investigation of nanoscale DG MOSFET subthreshold and analog/RF performances, which fully accounts for realistic constraints. This is realized by integrating computer based experiments, kriging metamodeling and multi-objective optimization in the same framework. The adoption of ATLAS 2D simulator for generating the computer sample data permits to include the most important physical aspects at nanoscale dimensions such as short channel, quantum mechanical confinement and hot carrier injection effects. The output responses are expressed by the following parameters : OFF-current, threshold voltage, swing factor, transconductance and cut-off frequency. While the input variables are identified to the channel length and the channel thickness. This choice for input/output parameters is approved by twofold reasons : geometrical input variables are the key factors responsible for various downscaling parasitic effects such as the short channel and the quantum mechanical effects. The transconductance and cut-off frequency parameters play a vital role in the design of devices dedicated to analog/RF applications. These parameters constitute the cornerstone for deducing more complicated analog/RF performance measures such as the transconductance frequency product (TFP) ([Grabinski et al., 2006] and [Pradhan et al., 2014]). Despite the striking simplicity of the suggested approach in comparison to the existing pure physical based models, it takes the advantages of numerical simulation accuracy in addition to the compactness of mathematical formulations. A good agreement is obtained for the metamodel predicted responses compared to the 2-D device numerical simulations. Hence, the developed compact models are used in the context of a multi-objective genetic algorithm to provide for the designer a set of Pareto configurations. This offers more flexibility in selecting adequate combinations of geometrical parameters that are well adjusted for the application field.

This chapter proceeds as follows. We briefly present in Section 3.2 the basic aspects attached to various parasitic effects occurring at nanoscale dimensions. Then, we introduce in Section 3.3 a detailed view of our integrated framework for the modeling and optimization of nanoscale DG MOSFET devices. The results obtained throughout different stages are reported in Section 3.4. By calculating some statistical measures, we demonstrate that the developed models are accurate enough to be used as a surrogate of the investigated design. Furthermore, the application of the multi-objective genetic algorithm results in non-dominated solutions in terms of Pareto optimality. Finally, we address, on the light of this study, the main concluding remarks in Section 3.5.

3.2 Classification of various parasitic mechanisms in MOSFETs

Although the DG MOSFET structure has been recognized as a promising electronic component in the last few years, physical constraints still impose strict limitations on the operation of deeply scaled technologies. In practice, the scaling of the supply voltage by the same proportion as the device physical dimensions is too restrictive since it leads to a reduced drive current and this situation is not recommended for high speed circuits. The doping density and profile is also a key factor that should be considered due to its impact on the scalability of DG MOSFETs as confirmed by many studies. The actual parasitic phenomena occurring in nanoscale devices can be classified mainly into the short channel effects, the quantum mechanical confinement effects and the hot carrier injection effects. These challenging barriers are the most critical aspects that need to be treated first before launching the manufacturing process of any advanced CMOS technology. In what follows, we provide some features characterizing each degradation class.

3.2.1 Short channel effects

Upon the reduction of the DG MOSFET channel length, the created electric fields rise considerably and lead to a modification in various responses with respect to the long channel device. The short channel effect becomes more pronounced when the channel length is of the order of the depletion layer thickness. Usually, the inclusion of the short channel effect in compact models proceeds by substituting the threshold voltage with an effective threshold voltage depending on the transistor geometry and bias conditions. A similar methodology is adopted for the surface potential based models, where the gate bias is replaced by an effective gate bias. It should be noted that the general structure of the basic compact model is not altered since extensions are only introduced at the level of some elementary parameters. In this subsection, we review some of the short channel harmful consequences on the good operation of the device.

3.2.1.1 Drain induced barrier lowering

The drain induced barrier lowering is a critical effect for short channel devices operating near threshold values. With reducing the channel length and increasing the drain-source voltage, the drain depletion region moves closer to the source depletion region. The reverse bias of the drain junction creates a field pattern that can lower the potential separating the source from the drain. This in turn results in an increased injection of electrons by the source over the reduced channel barrier, which gives rise to both increased drain current and static dissipation power. Among the strategies used to reduce the drain induced barrier lowering, high values for the doping concentration in the substrate are investigated to screen the effect of the drain voltage over the channel electrostatics [Arora, 2007].

3.2.1.2 Channel length modulation

The channel length modulation plays a crucial role for analog applications operating basically in saturation regime, where the gradual channel approximation is less valid especially near the drain side. As the channel length is reduced, this effect is intensified due to the increase of the carriers' number flowing from the source resulting in an increase

of the drain current. The current in this case can be calculated by conserving the same formula developed under the gradual channel approximation. But the effective channel length in this case is reduced by the amount of the drain section where two dimensional effects are accounted for [Galup-Montoro and Schneider, 2007]. It is clear that the effective channel length depends on the voltage drop, which necessitates its inclusion during the computation of the effective drain to source voltage for more accurate results.

3.2.1.3 Carrier mobility reduction

The response ability of carriers with respect to the applied electric field in a semiconductor material is known as mobility. This parameter has a paramount impact on DG MOSFET devices since it affects the drain current. For short channel devices, the mobility of carriers depends on electrical fields in both lateral and transversal directions. Furthermore, large power dissipation magnitudes can trigger an important temperature increase in lattice and carriers. Both parameters (electric field and temperature) degrade the mobility through various scattering mechanisms governing the surface inversion carriers [Weber et al., 2006]. The presence of the source and drain resistances can be the origin of the mobility reduction in deeply scaled devices [Song et al., 2011]. In order to attenuate such mobility anomalies, many enhancement strategies have been proposed such as the adoption of group III-V materials or pure germanium (Ge) during the channel manufacturing process [Tachi et al., 2011].

3.2.2 Quantum mechanical confinement effects

The downscaling of DG MOSFET devices is accompanied by thinner oxide thicknesses and more heavily doped channels in order to preserve the controllability of the channel with high drive current values. However, very high electric fields in the silicon/oxide interface and a potential well between the oxide field and the Si potentials are created in this case. These wells are formed by the oxide barrier and the silicon conduction band, which bends down steeply toward the surface due to the applied gate field. Therefore, carriers in the inversion layer are confined in a narrow potential well, where quantum effects have to be accounted. The exact treatment of quantum mechanical effects requires the resolution of the coupled Schrödinger/Poisson/Transport equations self-consistently, which excludes or makes the possibility of obtaining compact models a very intractable task. For the weak inversion mode, the influence of various sub-bands on the calculation of the inversion carrier density may be neglected because the splitting of the energy levels is small compared with the thermal voltage. For the strong inversion mode, the Poisson-Boltzmann equation can be adapted to include some quantum consequences such as the surface potential increase or the inversion capacitance reduction using semi-empirical modifications. Some insights about the behavior of nanoscale DG MOSFETs when subject to quantum effect are given below.

3.2.2.1 Inversion charge displacement

High substrate doping is needed to avoid the short channel effects at nanoscale level especially with the scaling of the gate oxide thickness down to values less than or equal to two nanometers. The energy levels are quantized during the inversion conditions and hence carriers occupy only distinct states, which differ from their classical counterpart by some fixed values of energy [Chaudhry and Roy, 2010a]. Such quantization has a significant modification on the inversion charge properties. At room ambient temperature (300

K), the inversion charge thickness is about 50 A° and depends on the substrate doping concentration besides to the applied gate voltage. When calculating the charge distribution based on the classical and the quantum formalisms, we distinguish the following differences. Under classical assumptions, the carriers' density attains its maximal value at the interface between the oxide/silicon layers and decreases steadily with the penetration into the bulk. In the quantum case, the carriers' density is approximately zero at the interface, reaches its maximal value at a location about the quarter of the channel thickness and then decreases with the distance from the surface. The second difference is related to the crystal orientation for which the carriers' distribution is dependent only in the quantum mechanical case.

3.2.2.2 Poly-silicon gate depletion

In order to control the work function difference in nanometer devices, poly-silicon gates are employed instead of metal gates. The flat band voltage and implicitly the threshold voltage are reduced. The poly-silicon gate/SiO_2 interface is depleted as the gate voltages are applied, which is known as the poly-silicon depletion effect. The effective gate capacitance is decreased as a result of the augmentation of the equivalent oxide thickness. Moreover, the quantization of energy levels at the SiO_2/gate interface initiates a delay in the inversion charge formation in addition to the shift in the poly-silicon gate potential with the applied gate voltages [Chaudhry and Roy, 2010b].

3.2.2.3 Threshold voltage shift

The classical definition of the threshold voltage corresponding to a semiconductor surface potential band bending equals to two times the Fermi potential becomes invalid for devices including the quantum mechanical confinement effect. This can be justified by the fact that the electron concentration near the silicon/oxide interface and inside the channel body is modified. Hence, the increase in the surface potential due to the inversion layer quantization shifts upward the threshold voltage. The threshold voltage value getting higher can be expressed by introducing in the classical formulation a correcting term including quantum and fitting parameters in such a manner that it vanishes at long channel lengths [Chaudhry, 2013].

3.2.3 Hot carrier effects

The principal cause of the hot carrier effect is due to the sharp peak in the pinch-off region exhibited by the lateral electric field during the transistor operation. Carriers traversing this region are subject to non-equilibrium energy distribution under the application of high drain biases. This effect is a reliability concern where carriers having enough energy are injected from the semiconductor into the surrounding dielectric films. Over prolonged periods of operation, the mobile carriers are accumulated in the oxide and generate the buildup of interface traps. The performance of the DG MOSFET device is consequently degraded as reflected by the reduction in some criteria like the transconductance or the cut-off frequency for analog applications. One of the most successful techniques to cope with the hot carrier effect resides in the inclusion of a field reducing region (lightly or moderately doped drain) in the transistor architecture. These regions reduce the amount of resulted damage so that the device expected lifetime can be promoted. Two basic monitors used mainly to identify the device status when subject to such

degradation are the substrate current and the gate current. In what follows, we describe some aspects relative to the hot carrier effect.

3.2.3.1 Impact ionization

As carriers in the silicon body gain energies beyond a definite threshold value, they can generate electron-hole pairs through the impact ionization mechanism. With some probability, an electron in the conduction band excites an electron from the valence band resulting in two electrons in the conduction band and a hole in the valence band. Both the total energy and momentum quantities are conserved during the impact process [Miyano et al., 1992]. In general, the impact ionization process itself has no dependence on the electric field. The characterization of the hot carrier effect using the substrate current generated by impact ionization has been extensively investigated. In fact, the exponential dependence of the substrate current on the maximum lateral electric field in the channel allows its precise measure and detection [Chau, 2014].

3.2.3.2 Carrier injection

Carriers acquiring energies higher than the local energy barrier at the Si/SiO_2 interface have a high probability of getting injected into the oxide layer. The carriers that cross the interface encounter a potential well between the interface and the location of peak oxide potential. At this boundary, carriers are partitioned into three classes according to the associated mechanisms : reflection, tunneling and transmission. The band alignment at the Si/SiO_2 interface results in an energy barrier of about 3.1 eV for electrons and 4.8 eV for holes. As a result of the large difference between the energy barriers for electrons and holes, under similar conditions, electrons are injected into SiO_2 in much larger quantities than holes [Pagey, 2002]. Due to these differences between electron and hole energy barriers, the hot carrier degradation mechanisms are noticeably different in $n-$ and p-channel MOSFETs.

3.2.3.3 Interface trap formation

The energetic carriers getting injected into the oxide layer have been recognized to produce instabilities in the device characteristics for both $n-$ and p-channel MOSFETs. These instabilities occur through the creation of electrically active defects in the oxide and at the Si/SiO_2 interface. These interface defects or traps introduce energy states in the Si band gap at the interface vicinity, where the occupancy of these interface states depends on the local surface potential [Schroder, 2006]. Once the device is turned on, the interface active traps get populated through the removal of minority carriers from the inversion layer resulting in the creation of localized defects near the drain side. It should be stated that these defects have a detriment effect on the device operation through the generation of parasitic currents and contribution to the aging phenomena [Maricau et al., 2008].

3.3 Presentation of the proposed modeling framework

Despite that several studies dealing with the integration of surrogate models and global optimization algorithms are reported in literature, the application of such hybrid approaches has not been up to our knowledge investigated in the field of nanoscale device

design. In the following, we first present the general flowchart of our proposed framework (see Figure 3.1) and then we discuss in detail the adopted elementary methods.

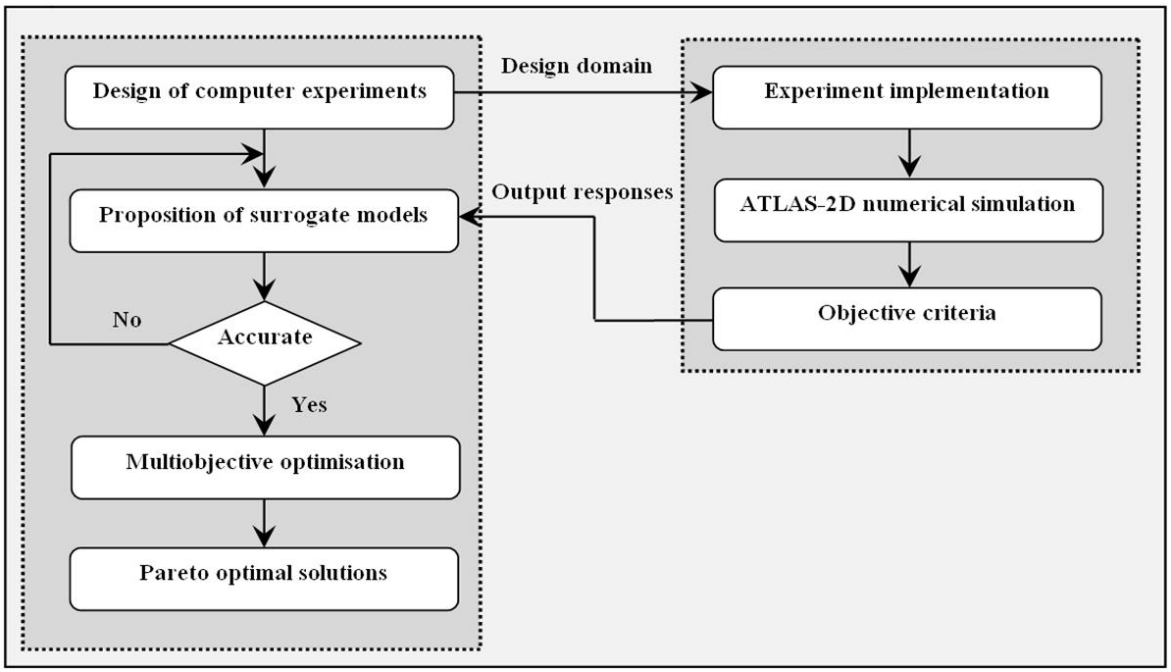

FIGURE 3.1 Flowchart of the proposed framework adopted for the optimal design of nanoscale DG MOSFET device.

3.3.1 Design of computer experiments

Instead of the direct manipulation of complex systems through controlled physical experiments, computer simulations are becoming a usual cheap solution for information collection or behavior analysis. For this reason, it is common in many industrial sectors to focus on software simulations rather than physical prototyping, where numerical experiments are conducted with the aim of optimizing tooling or components ordering. By doing so, manufacturing plants can avoid time consuming and costly physical experimentations [Kleijnen, 2015].

The design of experiments is a powerful paradigm dealing with the better conception of an experimentation procedure of interest. The elaboration of an experiment design is initiated with the definition of the problem in addition to the selection of input variables known as factors. The design space is delimited by assigning a variability range to each factor. The number of values taken by a variable is known as the variable levels and is obtained by discretizing the continuous variable within its associated range. Usually, the number of levels is the same for all variables, but some designs of experiment techniques permit distinct number of levels among variables. The idea behind the design of experiments resides in maximizing the information gained from a given number of experiments whilst using minimum resources [Mason et al., 2003]. Based on the set of realized experiments, it is possible to develop mathematical models to predict the mapping function relating the factors and the responses in a given system.

The class of optimal measures has been introduced to facilitate the design of nonlinear models by using a low number of experiments and without being restricted with particular operating areas. Different sets of samples are checked to find the one minimizing a well

defined objective function. This is achieved using an iterative method involving tedious computations to be accomplished with a set of samples spread over the whole design space. Statistically, the information content of measurements can be evaluated by the Fisher matrix, a real valued and symmetric matrix of the estimated parameters [Telen et al., 2012]. The optimization of the Fisher information matrix can be conducted using a scalar criterion denoted by J. Basically, four measures of the information content are established. The A-optimal design is seldom used because it can generate non-informative experiments with a covariance matrix not positive definite. This criterion attempts to maximize the Fisher information matrix trace. The D-optimal design is targeted to maximize the Fisher information matrix determinant, which makes it equivalent to the minimization of the geometric mean of the errors in the parameters. The largest error is minimized in the case of E-optimal design corresponding to the maximization of the smallest eigenvalue of the Fisher information matrix. The V-optimal design seeks to minimize the average prediction variance over a set of specific points, resulting in minimum discrepancy between the predicted and the real responses [De Aguiar et al., 1995]. So, a V-optimal design is equivalent to the selection of a set of experiments with the lowest average prediction variance as provided by the criterion :

$$J_V = \left(\frac{1}{N_{exp}} \sum_{i=1}^{N_{exp}} \chi_i^T \left(X^T X \right)^{-1} \chi_i \right) \tag{3.1}$$

where X refers to the matrix of inputs and possibly their transformations, N_{exp} is the number of experiments, χ_i is a single candidate experiment and T presents the transpose operation. The three remaining optimal design criteria deduced from the Fisher information matrix can be formulated mathematically as given below [Hametner et al., 2013] :

$$J_A = Tr\left(\left(X^T X \right)^{-1} \right) \tag{3.2}$$

$$J_D = \left| X^T X \right| \tag{3.3}$$

$$J_E = \lambda_{min}\left(X^T X \right) \tag{3.4}$$

with the notation |.| stands for the determinant and the Fisher information matrix is expressed by the product $\left(X^T X\right)$.

With the increase of the computer memory size and availability of highly specialized simulation packages in the recent years, the design of computer experiments has emerged as a central branch of the conventional design of experiments. This type of experiments is devoted to the design and analysis of deterministic computer codes, which produce the same responses every time the same input values are supplied. Because the random error is not present for such situations, the conventional experiment practices such as randomization, replication, and blocking are not necessary in the context of computer based experiments [Chen et al., 2003].

3.3.2 Development of the surrogate models

Metamodeling, called also surrogate modeling, is the process of constructing a statistical model that approximates the response of a system. The data used to deduce the model is generated from a complex deterministic simulation tool in which the random

variation existing in the real system is not represented. The use of metamodeling has grown in large part due to the increased utilization of computationally expensive computer simulations, where it is employed for the purposes of prediction, optimization, or model tuning. Depending on the statistical form of the metamodel, it may provide more concrete and explainable knowledge of the simulation response. Similarly to the classical design of experiments, the development of the metamodel requires careful consideration of the experimental runs used to build the metamodel and the outline of the suggested analytical form [Forrester et al., 2008].

Kriging, known as spatial correlation modeling, was developed initially for geostatistical modeling but has been extended to deterministic computer simulations and experiments. This method is traditionally a deterministic modeling technique that provides an exact interpolation and does not support the experimental error. A kriging prediction creates interpolated surfaces using weighted combinations of surrounding points, where the weighs are based on the distance between points and their specific locations. Kriging assumes that the closer the inputs are, the more positively correlated the outputs are. Note that in the deterministic simulation, kriging is an exact interpolator compared to linear regression analysis in the sense that the predicted values at observed input values are exactly equal to the simulated output values. The main concepts of kriging metamodeling are presented below [Kleijnen, 2015].
Given an initial design of experiments $\mathrm{X} = \left[x_1, x_2, ..., x_{\mathrm{N}_0}\right]$, with $x_i \in \mathbb{R}^n (i = 1, 2, ..., \mathrm{N}_0)$ the i^{th} experiment, and $y = \left[y(x_1), y(x_2), ..., y\left(x_{\mathrm{N}_0}\right)\right]$, with $y(x_i) \in \mathbb{R}$ the corresponding response to X. The kriging model consists of polynomial and random parts that can be denoted, for any input vector $x_i \in \mathbb{R}^n$, as follows :

$$\hat{y}(x) = \sum_{j=1}^{p} \beta_j f_j(x) + z(x) \tag{3.5}$$

where the term $\sum_{j=1}^{p} \beta_j f_j(x)$ is a regression model obtained by the linear combination of p chosen polynomial functions f_j. The coefficients β_j are the regression factors and z is a random error function assumed to have mean and covariance as follows :

$$\mu(z(x)) = 0 \tag{3.6}$$

$$Cov\left(z(x), z\left(x'\right)\right) = \sigma^2 \mathrm{R}_f\left(\theta, x, x'\right) \tag{3.7}$$

where σ^2 is the process variance and R_f is a parametric correlation function. In spite of the availability of many forms of correlation functions in the literature (such as the exponential, linear and Gaussian correlation functions), we use in this work the anisotropic Gaussian correlation function. This choice is justified because such function provides a process with infinitely differentiable paths. Moreover, the random variables in reliability studies are often of different nature and their average converges to the normal distribution :

$$\mathrm{R}_f\left(\theta, x, x'\right) = \prod_{k=1}^{k=s} e^{\left[-\theta_k\left(x_k - x_k'\right)^2\right]} \tag{3.8}$$

with θ a vector of parameters that define the inverse of the correlation length in each

direction, x_k and $x_k^{'}$ are the k^{th} coordinates of points x and x', s is the number of the coordinates of both points.

It can be observed that the kriging interpolation model depends on the vector of regression coefficients β, the vector of correlation parameters θ and the variance σ^2 of the stationary Gaussian process. These parameters are estimated as the solution of a generalized least squares problem by fitting the kriging model to a sample of N support points generated based on an experiment design procedure [Gaspar et al., 2014]. First, for a specified set of support points, the vector of correlation parameters θ is deduced as the solution of minimizing the maximum likelihood function as :

$$\hat{\theta} = \arg\min_{\theta} \left[\left| \mathrm{R}_f(\theta) \right|^{\frac{1}{\mathrm{N}}} \hat{\sigma}^2(\theta) \right] \tag{3.9}$$

Hence, the vector of regression coefficients and the process variance can be obtained as :

$$\hat{\beta} = \left(\mathrm{F}^{\mathrm{T}} \mathrm{R}_f^{-1} \mathrm{F} \right)^{-1} \mathrm{F}^{\mathrm{T}} \mathrm{R}_f^{-1} y \tag{3.10}$$

$$\hat{\sigma}^2 = \frac{1}{\mathrm{N}} \left(y - \mathrm{F}\hat{\beta} \right)^{\mathrm{T}} \mathrm{R}_f^{-1} \left(y - \mathrm{F}\hat{\beta} \right) \tag{3.11}$$

where F is the regression matrix for which the lines are the vectors $f\left(x^{(d)}\right) = \left(f_1\left(x^{(d)}\right), ..., f_p\left(x^{(d)}\right)\right)^{\mathrm{T}}$ of basis functions evaluated at the N realizations $x^{(d)}$ $(d = 1, 2, ..., \mathrm{N})$ of the vector of basic random variables and y is a vector with the corresponding true values $y^{(k)}$ of the limit state function.

In comparison to other traditional interpolation approaches, the kriging paradigm demonstrates several promising features. For instance, it uses a reduced set of known points during the forecasting process. In addition, the kriging models are not subject to any hypothesis about the underlying functional form as assumed in the case of polynomial regression models. Therefore, the kriging formalism can predict the behavior of arbitrary functions with high accuracy for both global as well as local approximations. The main advantages of kriging method over other spatial interpolation techniques (inverse distance weighting, splines, polynomial regression,...) involves not only the support of geometric characteristics of configurations, but also the consideration of the spatial correlation structure, which yields more reliable predictions [Montero et al., 2015]. For more details about the comparison of kriging metamodeling with respect to alternative approaches, the reader may refer to the following works ([Lucifredi et al., 2000], [Matta et al., 2012] and [Gaspar et al., 2014]).

3.3.3 Multi-objective optimization procedure

Multi-objective optimization refers to the process of considering several different and competing objectives simultaneously in the same framework. This type of optimization problems is widely encountered in engineering fields. Generally, a multiobjective minimization problem includes a set of design (or decision) variables, a set of M objective functions and a set of constraints. The optimization goal is given by :

$$\min_{x \in \Omega} f(x) = \min_{x \in \Omega} \left(f_1(x), f_2(x), ..., f_{\mathrm{M}}(x) \right) \tag{3.12}$$

with Ω is a subset of the design space ($\Omega \subseteq \mathrm{X}$) known as the feasible region that satisfies some specific constraints and it is defined as :

$$\Omega = \left\{ x \in \mathrm{X} \middle| \left(g_{i=1,u}(x) \geq 0, h_{j=1,v}(x) = 0, \mathrm{L}_{k=1,w} \leq x_{k=1,w} \leq \mathrm{U}_{k=1,w} \right) \right\} \tag{3.13}$$

where u is the number of inequality constraints, v is the number of equality constraints and w is the number of boundary constraints associated with the design variables.

In contrast to single objective optimization problems where the feasible region is completely ordered with respect to the objective function, the situation is changed in the context of several objectives since only a partial order can be established among the elements of the feasible region [Durillo et al., 2010]. A point x^* belonging to the feasible region Ω is said to be Pareto minimal if the following relation holds :

$$\forall i \in \{1,2,...,\mathrm{M}\} : f(x) = f\left(x^*\right) \wedge \exists i \in \{1,2,...,\mathrm{M}\}\ f_i\left(x^*\right) < f_i(x) \tag{3.14}$$

The Pareto optimality principle used to rank distinct solutions states that it would be not possible to decrease some criterion without increasing at least another one.

The Non-dominated Sorting Genetic Algorithm in its second version, abbreviated as NSGA II, is one of the most recognized algorithms based on Pareto dominance rule for the solving of multi-objective problems. Compared to its preceding version, the NSGA II has reduced computational complexity of the order of $\mathrm{O}\left(\mathrm{M} \times w^2\right)$, integrates elitism, enhances the diversity of Pareto optimal solutions and allows the handling of any existing constraints [Li et al., 2010].

NSGA II starts by generating a random parent population P_0 of size n_p. During several consecutive generations $t \geq 0$, the population is ranked based on the non-domination sorting procedure to create Pareto fronts. Consequently, every chromosome in the population will receive two attributes, including the non-domination rank and crowding distance. The binary tournament selection strategy repeatedly selects one winner with respect to the crowded comparison operator, where two members are first selected among the population. Then, the member with the superior crowding distance is selected if they share an equal rank. Otherwise, the member with the lower rank is chosen until all n_p parents have been selected. Each pair of parent individuals is recombined with a crossover probability p_c and using the crossover operator to create two child individuals, which are inserted in the offspring population Q_t or the two parents are retained in Q_t if the crossover operation is not performed. Individuals of Q_t are subject to the mutation operator with mutation probability p_m. The resulted offspring and the original parent populations are merged to form the population $\mathrm{R}_t = \mathrm{P}_t \cup \mathrm{Q}_t$ with size $2n_p$. Once, the population R_t is sorted in turn into nondominated fronts and the crowding distance is computed for each individual, the best half of R_t is transferred to the next generation. It should be highlighted that elitism is automatically guaranteed in this case because all parent and offspring individuals are involved. The combined population is considered as the current parent population and is used to create the next new offspring population with the execution of the above steps in order. The whole evolutionary process is repeated until some predefined stopping condition is met [Deb, 2001]. Different steps of the NSGA II implementation are outlined in Algorithm 3.1.

Algorithm 3.1 Presentation of the NSGA II approach

Input : Maximum generation number, population size, crossover rate, mutation rate
Begin
Initialize all algorithmic variables and set the counter $t \leftarrow 0$;
Generate initial population P_0 randomly;
Evaluate population P_0;
Calculate (rank; crowding distance) for each individual i in P_0;
for ($k = 1$ to the maximum number of iterations) **do**
  Select parent individuals from population P_t;
  Apply crossover operator to parents and create population Q_t;
  Apply mutation operator to each individual in Q_t;
  Evaluate Q_t;
  Combine Q_t and P_t to create a new population R_t;
  Calculate (rank; crowding distance) for each individual i in R_t;
  Select better half of R_t to generate P_{t+1};
  Increment the counter t ($t \leftarrow t+1$);
end for
End
Output : Best non-dominated solutions of the optimization problem.

3.4 Simulation and results

In what follows, we present in two subsections the main results related to the modeling and optimization of both subthreshold measures (I_{OFF}, V_{th} and DIBL) and analog/RF measures (g_m and f_T) associated to the nanoscale structure. The validity of the obtained models is consolidated through the evaluation of several statistical criteria before their adoption in the context of a multiobjective optimization.

3.4.1 Modeling and optimization of the subthreshold performances

Based on the description of various parasitic effects occurring in DG MOSFETs at nanoscale level as shown in Section 3.2, two geometrical parameters for the transistor device modeling are selected namely the channel length and the channel thickness denoted by L and t_{si}, respectively. The input parameters are independent and have uniform distributed values over the ranges 30 nm to 90 nm for the length and 2 nm to 5 nm for the thickness. Thus, the input parameter vector can be defined as (L, t_{si}) $\in [30nm, 90nm] \times [2nm, 5nm]$. While for the response parameters : the OFF-current, the swing factor and the threshold voltage are selected due to their widespread use for digital circuit applications.

The collection of sample data requires the knowledge of input configurations at which to run the deterministic computer code in order to cover efficiently the design space. For the design and analysis of computer experiment purposes, the Latin hypercube sampling is frequently used. However, in our work, optimal design strategies are more suitable since the input parameters are not equally meshed and to allow further extension if any constraints are to be defined on the design space. By using three optimality criteria (A, D and V), a set of 91 sample points are obtained as the input data of kriging metamodels, with each configuration represents a fixed geometrical dimension of the device. The computation procedure of the proposed optimal designs is done iteratively using the coordinate exchange method. In this context, each value of a specified factor in the design is examined at each iteration to test if the change leads to an improvement of the optima-

lity criterion and in this case the old value is updated with the new one. The examination process is continued until no updating is done for a complete iteration. This is reiterated several times using arbitrary initial conditions, so that the possibility to be trapped in local optima during the search can be significantly reduced. The remaining 16 sample points (in blue) resulted from the discretization of input parameter ranges are used to validate the accuracy of the kriging metamodels. The 75 sample points (in red) retained for the fitting of kriging models are depicted in Figure 3.2.

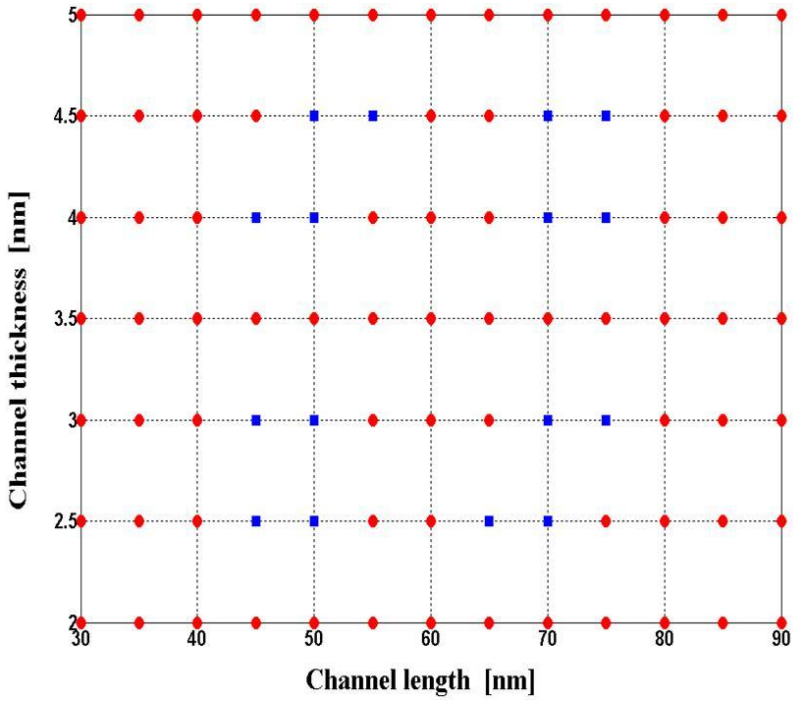

FIGURE 3.2 Scatter plot of D-optimal design including both fitting (red circles) and testing (blue square) sample points used for kriging metamodels development.

To get the output responses of the nanoscale DG MOSFET device associated to the sample configurations as retained by the optimal design of experiments, SILVACO environment is used to model the subthreshold behavior under the influence of various parasitic effects. This powerful tool is able to manipulate two and three dimensional geometries and provides a rich library of facilities [Silvaco, 2012]. A schematic three dimensional view of the nanoscale DG MOSFET device under analysis is illustrated in Figure 3.3. The p-type doping concentration is uniform in the silicon channel characterized by a length L and a thickness t_{si}. Both ranges of length and thickness values are discretized with steps of 5 nm and 0.5 nm respectively, which results initially in 91 combinations of input parameters. The gate dielectric thickness is denoted by t_{ox} with both gates are subject to the same gate voltage (symmetric). At the vicinity of the threshold voltage value, it can be assumed that all interface traps are occupied with electrons. Thus, it is possible to use interface-trap density and interface-trap charge density terms interchangeably. The investigation of hot carrier induced impact on various subthreshold parameters can be achieved by considering a fixed interface charge density near the drain side, thus the body channel can be seen as formed from fresh and damaged regions connected in series.

The set of all geometrical and electrical parameters remaining fixed during the running of simulations is given in Table 3.1.

In order to achieve pertinent modeling of the quantum mechanical confinement, we

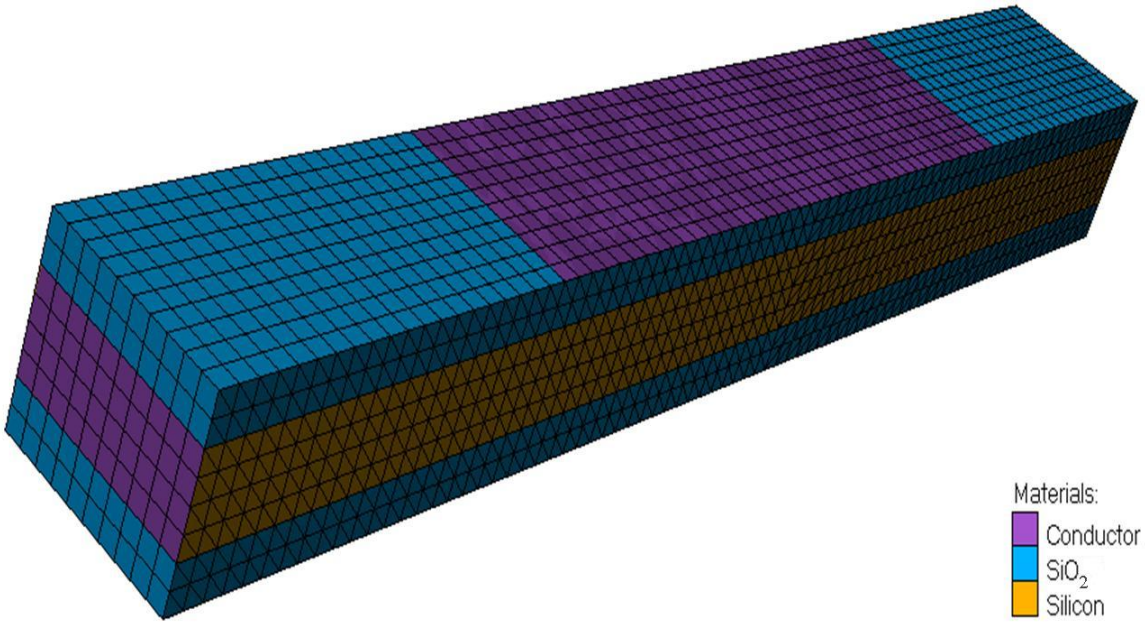

FIGURE 3.3 Schematic 3D visualization of the nanoscale DG MOSFET structure under study.

TABLE 3.1 Geometrical and electrical configuration of the nanoscale DG MOSFET structure under study.

Parameter	**Notation**	**Value**
Oxide thickness	t_{ox}	1.5 nm
Drain/source doping	$N_{D/S}$	10^{20} cm^{-3}
Channel doping	N_A	10^{15} cm^{-3}
Work function	ϕ_{MS}	4.55 eV
Interface trap density	N_F	5×10^{12} cm^{-3}
Drain voltage	V_{ds}	0.1 V
Gate voltage	V_{gs}	0.7 V

adopt Bohm Quantum Potential (BQP) formalism. The use of this approach offers many advantages such as better convergence properties in addition to the flexible calibration against the Schrödinger-Poisson equation.

ATLAS 2D simulator uses an iterative scheme to solve the non-linear BQP equation jointly with a set of semi-classical equations, where both models are treated to convergence repeatedly until self-consistency is obtained between the generated solutions. In this model, a position dependent quantum potential Q is added to the potential energy of carriers and it is given by the following formula [Iannaccone et al., 2004] :

$$Q = \frac{-h^2}{2} \frac{\gamma \nabla \left(M_t^{-1} \nabla \left(n^{\alpha} \right) \right)}{n^{\alpha}} \tag{3.15}$$

with α and γ are fitting parameters, M_t^{-1} is the inverse effective mass tensor and n is the carrier density.

Since the device is subject to continual biasing voltages whereby electrons and holes concentrations are disturbed from the equilibrium values, it is required to include some

types of generation-recombination mechanisms. The Shockley-Read-Hall recombination described first by Shockley and Read and extended later by Hall [Goudon et al., 2007] is taken into account in the numerical simulation framework. Another important effect to be account for is the saturation of the carriers' velocity when accelerated in an electric field. This is reflected by the reduction of the effective mobility because of the dependence between the drift velocity with respect to the mobility and the electric field component in the direction of the current flow. The field dependent mobility can be implemented using Caughey and Thomas formula [Caughey and Thomas, 1967], which results in a smooth passage between low and high field behaviors.

There is no doubt that metamodels have to be first assessed before being used as surrogates of any sophisticated system. In order to assess the prediction ability of the kriging metamodels on the subthreshold performance of nanoscale DG MOSFET device, the choice of the optimality criteria in addition to the number of training points should be carefully taken. Firstly, three optimality criteria namely A, D and V optimality criteria are tested and the optimal design is chosen according to the predictability resulted from the associated models. An initial size of the optimal design points is taken equal to 50 and is increased smoothly to verify the accuracy of the constructed krigin metamodels. It is found that the prediction tends to be constant when the number of design points is above 75. So, it is worthy recommended that the accuracy of the models would be satisfactory by constructing second-order polynomials kriging metamodels with 75 training points and by taking Gaussian correlation function. It should be mentioned that the prediction ability of models is estimated using the correlation (R) and Root Mean Square Error (RMSE) measures. For more precise metamodels, the value of the correlation coefficient should be close to 1 and the root mean square error between the estimated and the experimental values obtained by the ATLAS 2D simulator should be as small as possible (nearby zero). These statistical measures are defined as :

$$\mathrm{R} = \sqrt{1 - \frac{\sum_{i=1}^{\mathrm{N}_t} \left[y(x_i) - \hat{y}(x_i) \right]^2}{\sum_{i=1}^{\mathrm{N}_t} \left[y(x_i) - \bar{y}(x_i) \right]^2}} \tag{3.16}$$

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{\mathrm{N}_t} \left[y(x_i) - \hat{y}(x_i) \right]^2}{\mathrm{N}_t}} \tag{3.17}$$

where $y(x_i)$ and $\bar{y}(x_i)$ are the ATLAS 2D simulation values and the mean of these simulated values at validation points x_i. The corresponding predicted values obtained by the krigin metamodel are denoted by $\hat{y}(x_i)$ and N_t is the size of the validation set points. A summary of the effect of the selected optimal designs on the prediction measures of the kriging models is indicated in Table 3.2, in which the best quality is attributed to the D-optimal design.

The analytical expressions of kriging metamodels developed for the OFF-current, threshold voltage and swing factor using the D-optimal design are defined using the parameters $\hat{\beta}$, $\hat{\theta}$ and $\hat{\sigma}^2$, where the first parameter is related to the regression model and the latter two parameters are associated to the Gaussian correlation model. Table 3.3 recapitulates the corresponding estimated values of these parameters obtained for different subthreshold performance criteria.

TABLE 3.2 Comparison between the impacts of optimal design criteria considered in this work on the prediction measures.

Optimal design	**Value**	**Parameter**	**R**	**RMSE**	**p-val**
A-optimal	0.351	I_{OFF}	0.74	4.67×10^{-11}	0.0011
		V_{th}	0.99	0.01	4.07×10^{-12}
		S	1	5.68×10^{-4}	4.5×10^{-16}
V-optimal	0.059	I_{OFF}	0.76	4.22×10^{-11}	6.07×10^{-4}
		V_{th}	0.98	0.01	2.37×10^{-11}
		S	1	5.76×10^{-4}	2.21×10^{-15}
D-optimal	3.15	I_{OFF}	0.95	4.15×10^{-11}	1.36×10^{-8}
		V_{th}	0.98	0.01	7.08×10^{-12}
		S	1	4.69×10^{-4}	1.02×10^{-16}

TABLE 3.3 Estimated values of kriging metamodeling parameters associated to different subthreshold performance criteria.

Performance measure	**Parameter**	**Estimated value**
I_{OFF}	$\hat{\theta}$	$[7.07, 2.97]$
	$\hat{\beta}$	$[-0.79, -0.66, 0.44, 0.65, -0.60, 0.23]^{T}$
	$\hat{\sigma}^2$	2.92×10^{-20}
V_{th}	$\hat{\theta}$	$[16.82, 20]$
	$\hat{\beta}$	$[0.44, 0.36, -0.67, -0.23, 0.05, -0.24]^{T}$
	$\hat{\sigma}^2$	6.3×10^{-4}
S	$\hat{\theta}$	$[16.82, 20]$
	$\hat{\beta}$	$[-0.07, -0.92, 0.07, 0.23, -0.02, -0.17]^{T}$
	$\hat{\sigma}^2$	1.5×10^{-6}

Figure 3.4 depicts the predicted behavior of OFF-current using the associated kriging metamodel. As can be observed, the fitted model follows the pattern of the simulated sample points obtained using ATLAS 2D simulator rather closely, which indicates that the existing dependence among the OFF-current, channel length and channel thickness is well captured with the proposed metamodel. The OFF-current takes high values for short channel lengths and decreases steadily to stabilize at long channel lengths independently of the channel thickness values. It can be also noticed that the influence of the quantum confinement on the current becomes more pronounced at low values of the channel thickness and channel length, where the curvature of the graph is increased. The latter observation can be considered as a sign on the existence of a strong correlation between the quantum mechanical confinement and short channel effects, which may be of a big deal since many published works have stated contradictory results regarding the influence of quantum confinement on the short channel effect.

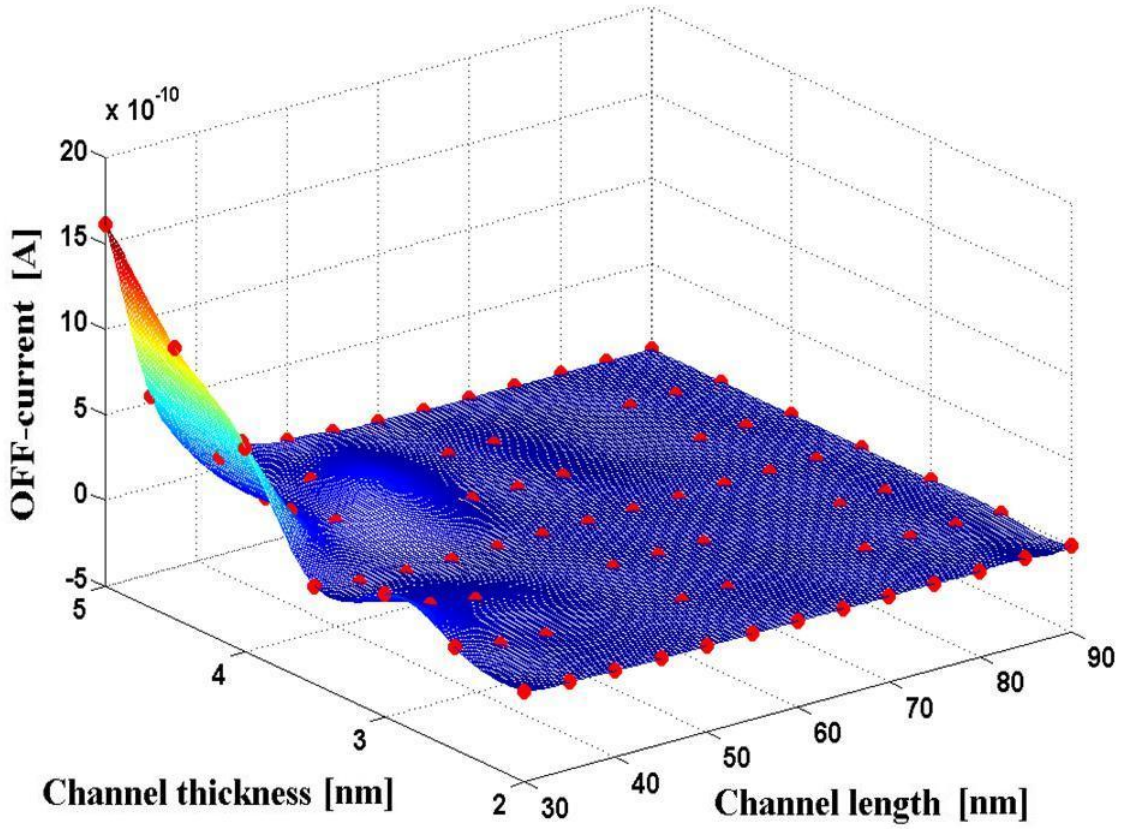

FIGURE 3.4 Graphical representation of the predicted OFF-current with respect to fitting (circle) and testing (square) sample points.

The predicted response of the kriging metamodel of the swing factor in addition to the simulated sample points are shown in Figure 3.5. It can be noticed that the tendency of the curve decreases gradually with the increase of the channel length to reach the near ideal value (60 mV/Dec). Furthermore, the swing factor is degraded when the channel length is located at the proximity of the middle at t_{si}=3.5 nm, especially for short channel lengths. This is can be interpreted by the influence of the interface traps jointly with the short channel effects on the inversion layer. On the other hand, the swing factor is slightly affected for the channel thickness values located near the bounds of the considered interval (2 and 5 nanometers) and is almost unchanged for a channel length superior or equals

to 65 nm.

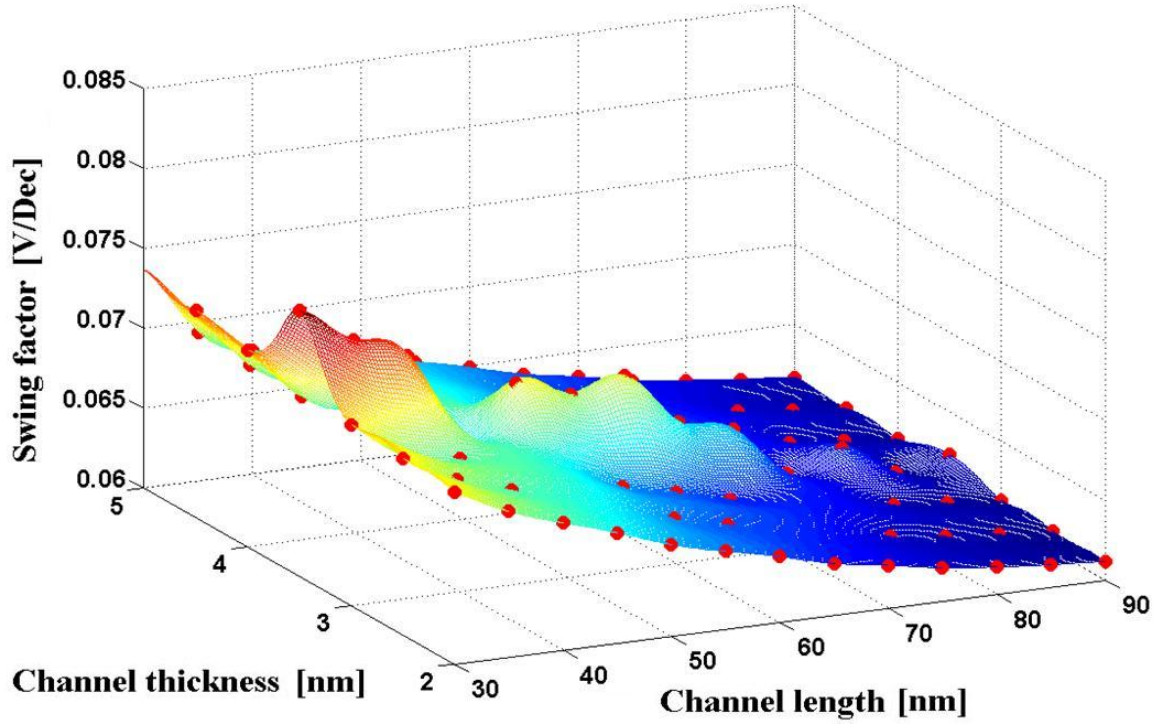

FIGURE 3.5 Graphical representation of the predicted Swing factor with respect to fitting (circle) and testing (square) sample points.

Our elaborated kriging metamodel also predicts a similar behavior for the degradation of the threshold voltage, which occurs depending on the channel thickness parameter. This is described in Figure 3.6, where the threshold voltage is plotted against the channel length and thickness. It can be clearly seen that as the channel thickness varies from 2 nm to 3 nm the threshold voltage decreases, it reaches a maximum value for a well defined position located between 3 nm and 4 nm and continues to decrease when the channel thickness value surpasses the middle of the channel and approaches 5 nm.

Based on the obtained prediction results, the sensitivity of the subthreshold performance criteria is tightly linked to the geometrical configuration of the device defined mainly by the channel length and channel thickness parameters. These results also delimit critical regions under which the device is subject to a little or a strong degradation in terms of subthreshold parameters. Consequently, innovative designs should be focused on regions having higher immunity against parasitic effects.

Since the developed kriging models have indicated sufficient prediction accuracy, they can be approved as robust objective functions for optimization tasks. The design of DG MOSFET devices can be seen as a multi-objective optimization problem in order to satisfy the requirements of different field applications. Hence, the problem of optimizing the subthreshold performance criteria can be described by the following multi-objective formulation :

$$\begin{cases} \min\limits_{(\mathrm{L},t_{Si})} (\mathrm{V}_{th}, \mathrm{S}, \mathrm{I}_{\mathrm{OFF}}) \\ Subject\ to \\ 30nm \leq \mathrm{L} \leq 90nm \\ 2nm \leq t_{\mathrm{S}i} \leq 5nm \end{cases} \tag{3.18}$$

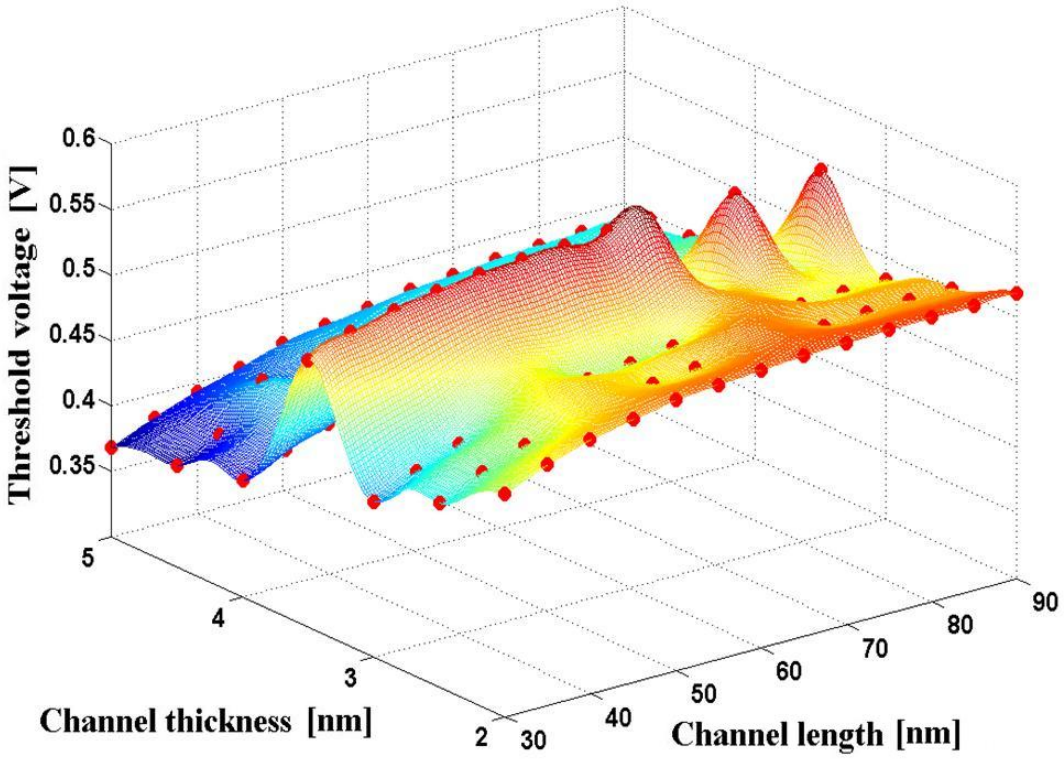

FIGURE 3.6 Graphical representation of the predicted threshold voltage with respect to fitting (circle) and testing (square) sample points.

where V_{th}, S and I_{OFF} represent the objective functions that have been expressed by kriging metamodels. This multi-objective optimization problem is to be solved using the well known non-dominated sorting genetic algorithm coded using the MATLAB software. The computer experiments are performed on a Desktop PC with Intel Pentium 4 CPU and 1.128 GB RAM.

In our NSGA II framework, we consider the following attributes. The tournament selection strategy is applied to a set of 5 solutions in order to get two parents using the crowded-comparison operator. Once both parents are selected, the two point crossover is applied to form an offspring. The principle consists in generating two random integers a and b between 1 and the number of genes in a solution. Then, genes at extremities are extracted from the first parent and genes of the middle are extracted from the second parent. All extracted genes are concatenated to create the offspring. By inverting the order of parents, it is possible to get a second offspring. The adaptive feasible mutation is adopted where randomly generated directions are adaptive with respect to the last successful or unsuccessful generation. A step length is chosen along each direction so that linear constraints and bounds are satisfied which allows preserving the feasibility of different solutions. A recapitulative table summarizing the main configuration values of the algorithm is introduced in Table 3.4.

The set of the best solutions presented by the Pareto front is visualized graphically using a three dimensional plot as indicated in Figure 3.7. For digital applications, the OFF-current and threshold voltage are required to be minimized in order to reduce the power dissipation. The swing factor is also minimized since it controls the commutation speed between "OFF" and "ON" states. In essence, it can be noticed that all these optimi-

TABLE 3.4 Configuration parameters of the adopted multi-objective genetic algorithm.

Parameter	**Value**
Population type	Double vector
Population size	100
Crossover rate	0.9
Pareto fraction	0.3
Number of generations	1000
Creation strategy	Uniform
Selection strategy	Tournment
Pool size	5
Crossover type	Two points
Mutation type	Adaptive feasible

zation problems are competing goals. This means that additional constraints imported from the application field to which the device is oriented need to be introduced by the designer for the selection of the most appropriate configuration.

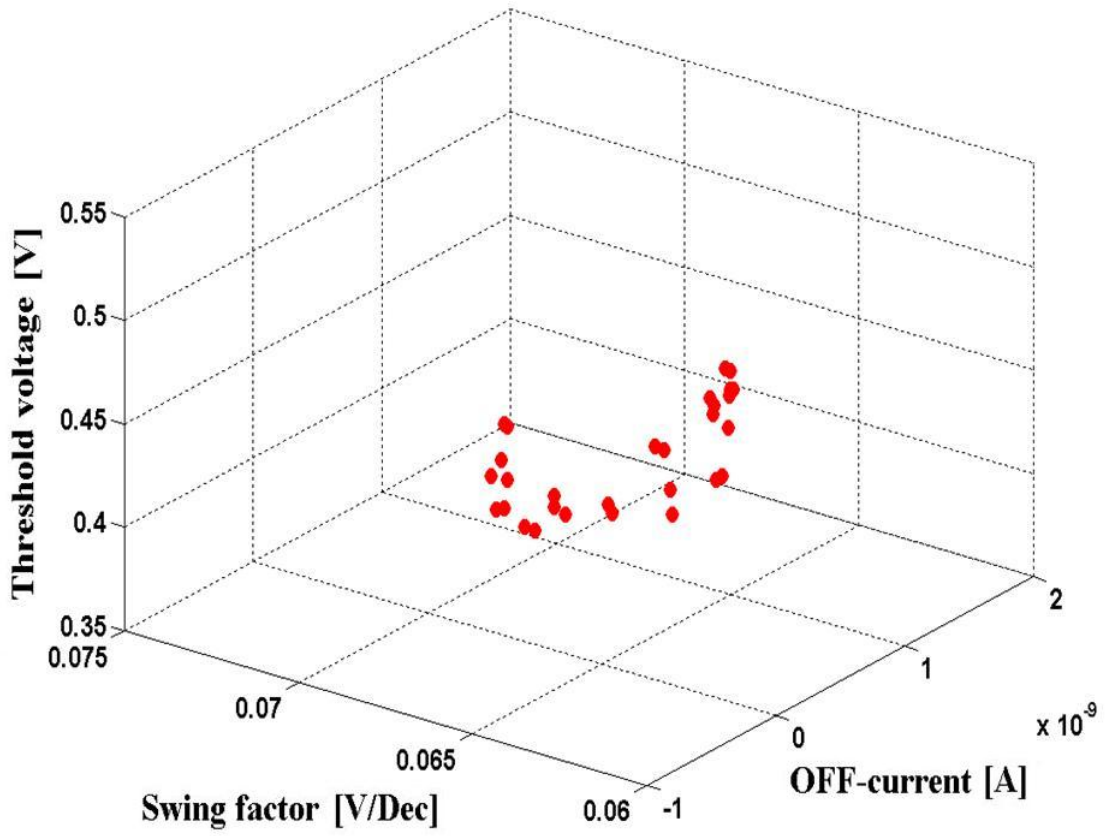

FIGURE 3.7 Pareto front obtained using NSGA II with the developed kriging metamodels as objective functions.

3.4.2 Modeling and optimization of the analog/RF performances

In the context of this work, the response function is used to express some analog/RF performance criteria namely the transconductance and cut-off frequency associated to nanoscale DG MOSFETs operating under critical conditions. The figure of merit known as the transconductance widely used in analog design reflects the change in the drain current with respect to the change in the gate/source voltage. The transconductance is usually calculated in saturation region to point out that the device transforms a voltage change to a current change [Razavi, 2017]. The cut-off frequency is a relevant RF performance measure since it represents the frequency for which the current gain of the device is unity [Sharma and Vishvakarma, 2015]. The transconductance and the cut-off frequency expressions are given by the following formula [Bentrcia et al., 2016] :

$$g_m = \left.\frac{\partial \mathrm{I}_{ds}}{\partial \mathrm{V}_{gs}}\right|_{\mathrm{V}_{ds}} \tag{3.19}$$

$$f_\mathrm{T} = \frac{g_m}{2\pi \mathrm{C}_{gs}\sqrt{1+2\frac{\mathrm{C}_{gd}}{\mathrm{C}_{gs}}}} \tag{3.20}$$

where I_{ds} is the drain current, V_{ds} and V_{gs} denote the applied gate and drain voltages, C_{gs} and C_{gd} refer to the gate-source and gate-drain capacitances, respectively.

The validity of the obtained kriging model should be checked before being used as a surrogate of any analog/RF DG MOSFET criterion. Arbitrary subset from the geometrical configuration space is used for the test. Three statistical measures namely the Sum of Relative Errors (SRE), the Mean Percentage of Absolute Errors (MAPE) besides the Correlation coefficient (R) are adopted. Such statistical measures allow assessing the goodness of fit associated to our surrogate models [Abbott, 2014]. The sum of relative errors, the mean percentage of absolute errors and the correlation coefficient are calculated using the following equations :

$$\mathrm{SRE} = \sum_{i=1}^{\mathrm{N}} \left| \frac{\mathrm{Y}(x_i) - \hat{\mathrm{Y}}(x_i)}{\mathrm{Y}(x_i)} \right| \tag{3.21}$$

$$\mathrm{MAPE} = \left(\frac{1}{\mathrm{N}} \sum_{i=1}^{\mathrm{N}} \left| \frac{\mathrm{Y}(x_i) - \hat{\mathrm{Y}}(x_i)}{\mathrm{Y}(x_i)} \right| \right) \times 100 \tag{3.22}$$

$$\mathrm{R}^2 = 1 - \left[\frac{\sum_{i=1}^{\mathrm{N}} \left(\mathrm{Y}(x_i) - \hat{\mathrm{Y}}(x_i)\right)^2}{\sum_{i=1}^{\mathrm{N}} \left(\mathrm{Y}(x_i) - \bar{\mathrm{Y}}(x_i)\right)^2} \right] \tag{3.23}$$

with N the size of the test sample, Y the response delivered by ATLAS 2D simulator and $\hat{\mathrm{Y}}$ the response predicted by the surrogate model. If the computation of these statistical criteria leads to near ideal values (0 for the SRE, 0% for the MAPE and 1 for the R^2) then

the prediction ability of the kriging models is satisfactory. The generation of near ideal values for the statistical criteria enables the use of metamodels for prediction or optimization goals. Otherwise, obtained metamodels should be revised with the selection of appropriate regression and correlation functions.

In the context of this study, the design of a DG MOSFET device dedicated to analog applications involves two objectives to be optimized namely the transconductance and the cut-off frequency. However, the difficulty that can arise is that different solutions will generate trade-off between the objective functions. As a matter of fact, the optimization procedure can be formulated as a multi-objective optimization procedure rather than a single objective optimization problem as illustrated below :

$$\begin{cases} \underset{(t_{si},\mathrm{L})}{Maximize}\left(g_m, f_\mathrm{T}\right) \\ Subject to \\ t_{si} \in \left[t_{si}^{\min}, t_{si}^{\max}\right] \\ \mathrm{L} \in \left[\mathrm{L}^{\min}, \mathrm{L}^{\max}\right] \end{cases} \tag{3.24}$$

The proposed optimization framework described above is applied to the modeling and design of a nanoscale DG MOSFET including traps with the device oriented towards analog/RF applications. The first stage is to build an equivalent discrete system of the device using ATLAS 2D simulator. The bird eye's view of the generated design is reported in Figure 3.8.

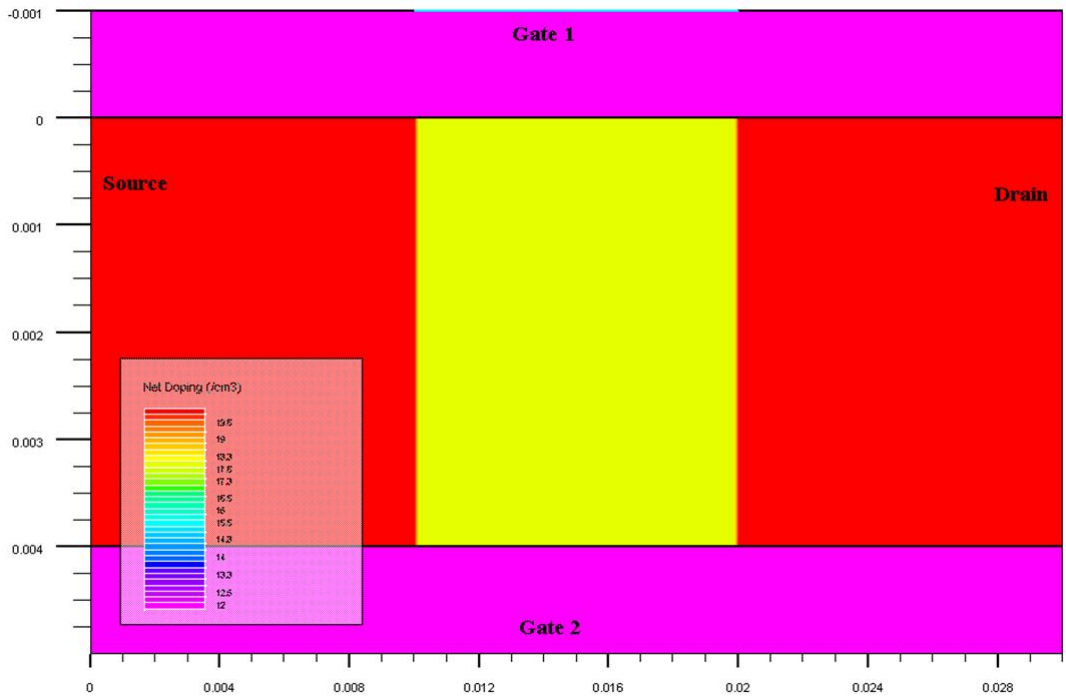

FIGURE 3.8 Two dimensional view of the investigated device.

The sample data needed for the elaboration of the kriging models are collected by parsing the channel length and the thickness values in the ranges $[10nm-30nm]\times[2nm-4nm]$ respectively. Both intervals are divided into 4 subregions which results in 25 nodes. It should be highlighted that the sample data is partitioned into two sets, the training set containing 21 nodes and the validation set containing 4 nodes. We have focused on these

bounds for the channel length and channel thickness to meet the tremendous demand in terms of miniaturization and performance. The localization of various sample data nodes is depicted in Figure 3.9.

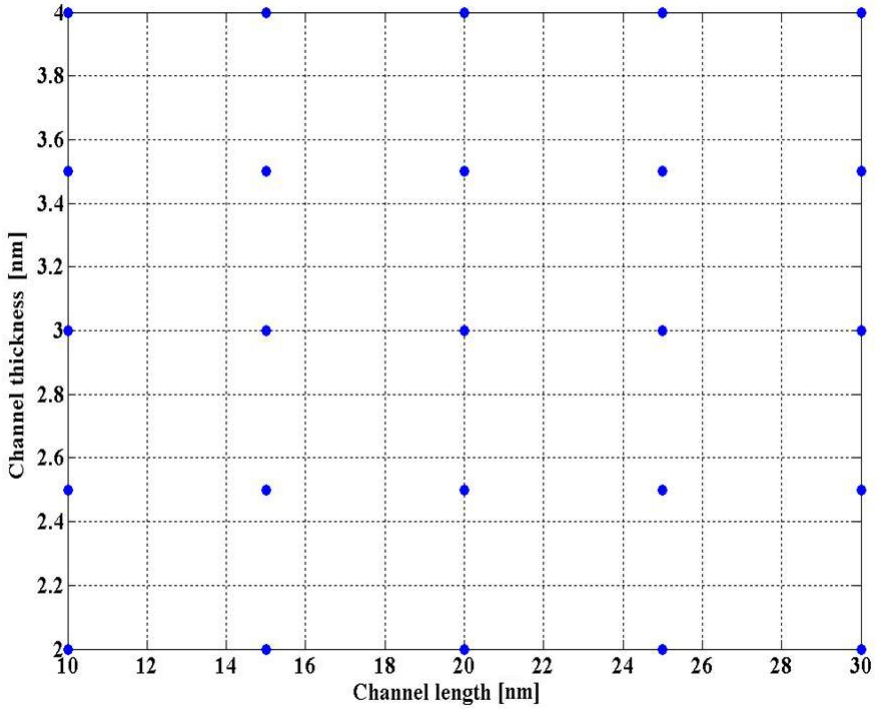

FIGURE 3.9 Various sample points used in building kriging models.

In Table 3.5, we summarize the statistical criteria of the accuracy evaluation of the best kriging models associated with the transconductance and the cut-off frequency in predicting the test set nodes. As can be observed from this table, the correlation factor values for both responses is very close to 1 which indicates a strong dependence between predicted and numerical results. Furthermore, the sum of the relative errors and the mean absolute percentage error are both small indicating a high precision for the obtained models. The hyperparameters values assigned to the best kriging models are introduced in Table 3.6. Regarding regression function type, we associate a first polynomial order (second polynomial order) to the transconductance (cut-off frequency) response, respectively. For the correlation function, we select the Gaussian and linear functions for metamodeling of the transconductance and the cut-off frequency, respectively.

TABLE 3.5 Statistical assessment of analog/RF surrogate models.

Performance measure	**Notation**	R	SRE	MAPE
Transconductance	g_m	0.88	0.16	4%
Cut-off frequency	f_T	0.99	0.20	5%

Figures 3.10 and 3.11 illustrate the predicted behavior of the transconductance and cut-off frequency as deduced by the obtained kriging models. It should be mentioned that the validation configurations are not highlighted on these figures since they were already evaluated using more elaborated statistical criteria as provided in Table 3.5. The fitted

TABLE 3.6 Hyperparameters values associated with the best surrogate models.

Performance measure	**Parameter**	**Estimated value**
Transconductance	$\hat{\theta}$	$[47.57, 50]$
	$\hat{\beta}$	$[-7.45 \times 10^{-12}, -0.04, 0.58]^{\mathrm{T}}$
	$\hat{\sigma}^2$	6.22×10^{-8}
Cut-off frequency	$\hat{\theta}$	$[10, 10]$
	$\hat{\beta}$	$[-0.25, -1, 0.04, 0.3, -0.08, -0.03]^{\mathrm{T}}$
	$\hat{\sigma}^2$	1.72×10^{21}

models follow accurately the patterns of the numerical sample points generated by ATLAS 2D simulator indicating that the dependence of analog/RF responses to geometrical dimensions is well captured. The transconductance increases steadily with the channel thickness with the existence of some local optima, which expresses the difficulty of developing a simplified closed form. Furthermore, the accuracy of the predicted transconductance can be estimated to be high for configurations having a sufficient number of nearby simulation points compared to configurations located at boundaries. This can be justified by the fact that the kriging formalism is based on the prediction of a response value at a given configuration by computing the weighted average of the known response values at the surrounding configuration points. Due to the consideration of quantum confinement, interface traps and short channel effects in the device numerical model, the behavior of the device operation basically above threshold voltage becomes very complex and leads to the presence of some local minima for which it is very difficult to provide a physical interpretation. The cut-off frequency decreases as a function of the channel length, where the high values of this measure covers the existence of any curvature in the graph. It can be concluded that the cut-off frequency response has higher immunity against geometrical variation in comparison to the transconductance.

Despite that the effect of the parasitic capacitances is included within the developed empirical model of the cut-off frequency through the adjusted hyperparameters, it is also possible to exploit the influence of the capacitance using the developed empirical models jointly with the relation $\frac{g_m}{2\pi f_{\mathrm{T}}} = \mathrm{C}_{gs}\sqrt{1+2\frac{\mathrm{C}_{gd}}{\mathrm{C}_{gs}}} \approx \mathrm{C}_{gs} + \mathrm{C}_{gd} \approx \mathrm{C}_{gg}$. However, this will necessitate supplementary tedious computations. For this reason, the capacitances are considered as intermediate parameters and are not treated separately in an explicit manner especially with the main goal of this work orientated towards confirming the adaptability of surrogate metamodeling for analog/RF criteria at the deep nanoscale level. In Figure 3.12, we elucidate Pareto-optimal points. It is clear that maximum transconductance results in minimum cut-off frequency and vice versa. Moreover, it is worth noting that the Pareto front configurations of the device exhibit better behavior in comparison to the initial training configurations. Since under such situation we have two objective functions, a compromise is required in order to select a configuration from the Pareto front as the final design. The Pareto optimal set permits to gain more flexibility in the design procedure before implementation. The selection of configurations located at the middle of Pareto front achieves a compromise between the transconductance and the cut-off frequency and results in nanoscale devices adequate to both analog and RF applications. It is worthy to mention that our hybrid framework based on the integration of metamodeling techniques and multi objective optimization algorithms is a generic approach. Consequently, it has been applied successfully to other important parameters namely Gain (A_v), Trans-

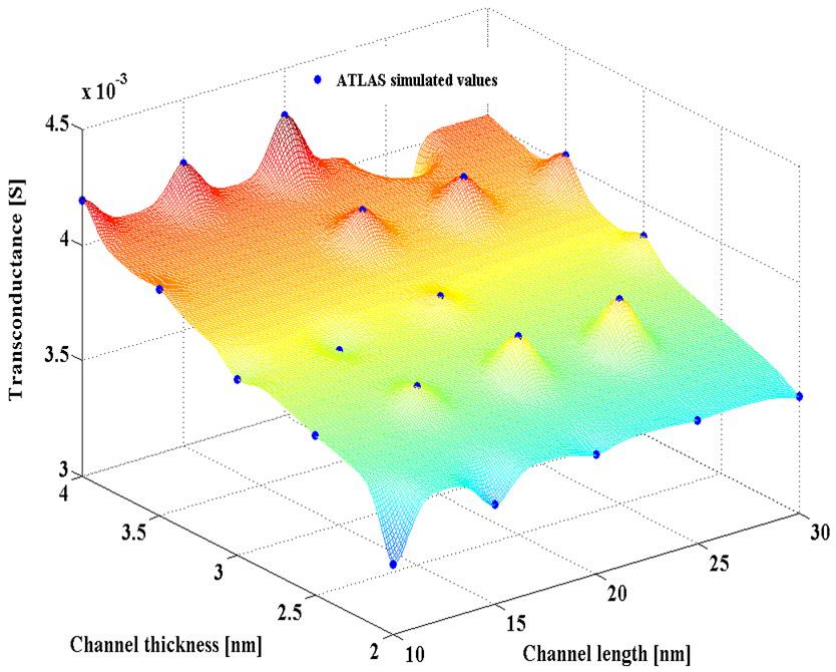

FIGURE 3.10 Predicted transconductance values as a function of the channel thickness and length.

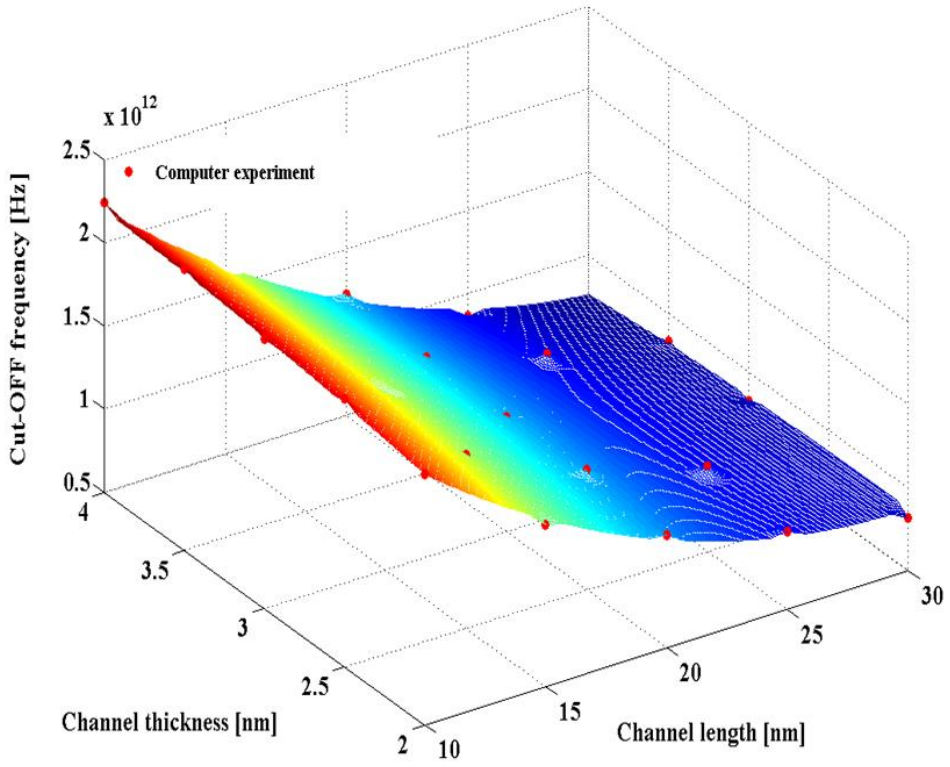

FIGURE 3.11 Predicted cut-off frequency values as a function of the channel thickness and length.

conductance Generation Factor (TGF), Threshold voltage (V_{th}), Swing factor (S) and OFF-

current (I_{OFF}) ([Bentrcia et al., 2017a] and [Bentrcia et al., 2017b]). Our approach can be also applied to digital performance by including phenomena and mechanisms dominant at subthreshold regime of the device during the simulation phase and performing again the hyperparameters adjustment. However, supplementary extensions should be in this case conducted on the original ATLAS 2D numerical model in order to enrich the training data set and account for the additional performance criteria.

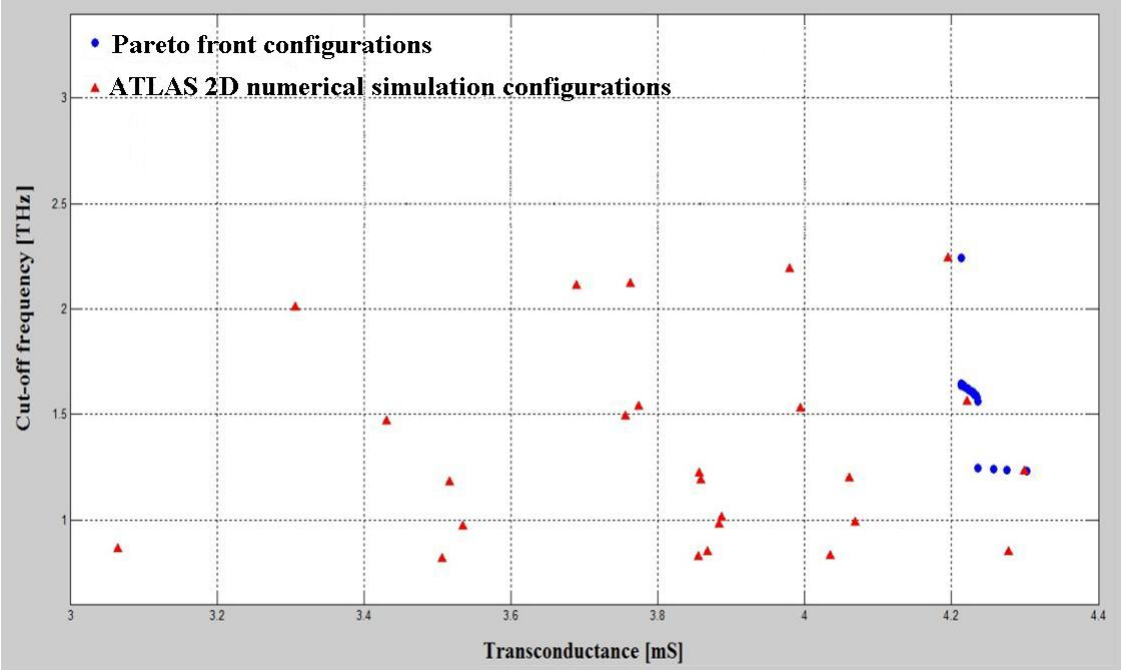

FIGURE 3.12 Pareto front obtained using best surrogate models.

3.5 Conclusion

In this chapter, we have proposed a hybrid approach, based on the integration of surrogate modeling and multi-objective approaches, for the design of efficient DG MOSFET devices at nanoscale level. We have attempted to make our metamodels as close as possible to practical conditions by including parasitic effects namely the short channel, the quantum mechanical confinement and the hot carrier injection effects. In order to reduce the computation time, a limited number of computer experiments have been designed using three design optimality criteria. ATLAS 2D simulator was adopted to perform numerical simulations and to extract necessary information for fitting the kriging metamodels associated with the OFF current, threshold voltage, swing factor, transconductance and cut-off frequency. The computation of statistical measures of prediction ability has confirmed the validity of the proposed kriging metamodels as surrogate of the device under consideration. Therefore, they have been used in the context of a multi-objective optimization problem to enhance both the subthreshold and the analog/RF performances, in which the design variables are related to the geometry shape namely the channel length and channel thickness. The results obtained by NSGA II approach have shown that such framework constitutes a reliable tool for the accurate design of nanoscale DG MOSFETs, especially when the trade-off between goals is inevitable. The resulted Pareto front may

provide a wide range of configuration possibilities according to the conditions imposed by the environment.

3.6 References

Abbott, D. (2014). *Applied Predictive Analytics : Principles and Techniques for the Professional Data Analyst*. Wiley Press. 93

Abdelhamid, H. M. (2007). *Compact Modeling of Multiple Gate MOS Devices*. PhD thesis, University of Rovira i Virgili. 72

Amat, E., Kauerauf, T., Rodriguez, R., Nafria, M., Aymerich, X., Degraeve, R., and Groeseneken, G. (2013). A comprehensive study of channel hot-carrier degradation in short channel MOSFETs with high-k dielectrics. *Microelectronic Engineering*, 103 :144–149. 72

Arora, N. (2007). *MOSFET Modeling for VLSI Simulation Theory and Practice*. World Scientific Press. 72, 75

Asenov, A., Brown, A. R., Roy, G., Cheng, B., Alexander, C., Riddet, C., Kovac, U., Martinez, A., Seoane, N., and Roy, S. (2009). Simulation of statistical variability in nano-CMOS transistors using drift-diffusion, monte carlo and non-equilibrium green's function techniques. *Journal of Computational Electronics*, 08(03) :349–373. 71

Bentrcia, T. and Djeffal, F. (2011). *Compact modeling of multi-gate MOSFET including hot-carrier effects*, chapter 04. in M. J. Kwon (eds.), CMOS Technology, Nova Science Press. 71, 72

Bentrcia, T. and Djeffal, F. (2014). *An ANFIS based approach for prediction of threshold voltage degradation in nanoscale DG MOSFET device*, chapter 25. in G. C. Yang and S. I. Ao and L. Gelman (eds.), IAENG Transactions on Engineering Technologies, Springer Press. 71

Bentrcia, T., Djeffal, F., Arar, D., and Dibi, Z. (2014). Gate-engineering-based approach to improve the nanoscale DG MOSFET behavior against interfacial trap effects. *Physica Status Solidi (C)*, 11(01) :77–80. 71

Bentrcia, T., Djeffal, F., and Benhaya, A. (2012a). Continuous analytic i-v model for GS DG MOSFETs including hot-carrier degradation effects. *Journal of Semiconductors*, 33(01) :1–6. 72

Bentrcia, T., Djeffal, F., and Chebaki, E. (2012b). ANFIS-based approach to studying subthreshold behavior including the traps effect for nanoscale thin-film DG MOSFETs. *Journal of Semiconductors*, 34(08) :1–9. 72

Bentrcia, T., Djeffal, F., and Chebaki, E. (2017a). *Multi-objective design of nanoscale double gate MOSFET devices using surrogate modeling and global optimization*, chapter 04. in A. Tiwari, and Y. K. Mishra, and H. Kobayashi, and A. P. F. Turner (eds.), Intelligent Nanomaterials, Wiley-Scrivener Press. 98

Bentrcia, T., Djeffal, F., Chebaki, E., and Arar, D. (2016). Impact of the drain and source extensions on nanoscale double-gate junctionless MOSFET analog and RF performances. *Materials Science in Semiconductor Processing*, 42(02) :264–267. 93

Bentrcia, T., Djeffal, F., Chebaki, E., and Arar, D. (2017b). A kriging framework for the efficient exploitation of the nanoscale junctioless DG MOSFETs including source/drain extensions and hot carrier effect. *Materials Today : Proceedings*, 04(07) :6804–6813. 98

Bindu, B., Cheng, B., Roy, G., Wang, X., Roy, S., and Asenov, A. (2010). Parameter set and data sampling strategy for accurate yet efficient statistical MOSFET compact model extraction. *Solid-State Electronics*, 54(03) :307–315. 71

Camiola, V. D., Mascali, G., and Romano, V. (2012). Numerical simulation of a double-gate MOSFET with a subband model for semiconductors based on the maximum entropy principle. *Continuum Mechanics and Thermodynamics*, 24(04) :417–436. 73

Caughey, D. M. and Thomas, R. E. (1967). Carrier mobilities in silicon empirically related to doping and field. *Proceedings of the IEEE*, 55(12) :2192–2193. 87

Chau, Q. (2014). Investigating and modeling impact ionization current in MOSFETs. *Solid-State Electronics*, 94 :66–71. 78

Chaudhry, A. (2013). *Fundamentals of Nanoscaled Field Effect Transistors*. Springer Press. 72, 77

Chaudhry, A. and Roy, J. N. (2010a). An analytical model of inversion layer quantization and gate oxide quantum mechanical tunneling in nano-p-MOSFETs. *Electronics*, 14(02) :86–89. 76

Chaudhry, A. and Roy, J. N. (2010b). Mathematical modeling of nanoscale MOS capacitance in the presence of depletion and energy quantization in a poly-silicon gate. *Journal of Semiconductors*, 31(11) :1–4. 77

Chen, V. C. P., Tsui, K. L., Barton, R. R., and Allen, J. K. (2003). A review of design and modeling in computer experiments. *Handbook of Statistics*, 22 :231–261. 80

Darbandy, G., Lime, F., Cerdeira, A., Estrada, M., Garduño, S. I., and Iñiguez, B. (2012). Gate leakage current partitioning in nanoscale double gate MOSFETs, using compact analytical model. *Solid-State Electronics*, 75 :22–27. 71

De Aguiar, P. F., Bourguignon, B., Khots, M. S., Massart, D. L., and Phan-Than-Luu, R. (1995). D-optimal designs. *Chemometrics and Intelligent Laboratory Systems*, 30(02) :199–210. 80

Deb, K. (2001). *Multi-Objective Optimization Using Evolutionary Algorithms*. Wiley Press. 83

Djeffal, F., Abdi, M. A., Dibi, Z., Chahdi, M., and Benhaya, A. (2008). A neural approach to study the scaling capability of the undoped double-gate and cylindrical gate all around MOSFETs. *Materials Science and Engineering : B*, 147(2-3) :239–244. 71

Djeffal, F., Dibi, Z., Hafiane, M. L., and Arar, D. (2007). Design and simulation of a nanoelectronic DG MOSFET current source using artificial neural networks. *Materials Science and Engineering : C*, 25(7-8) :1111–1116. 72

Durillo, J. J., Nebro, A. J., Coello Coello, C. A., García-Nieto, J., Luna, F., and Alba, E. (2010). A study of multiobjective metaheuristics when solving parameter scalable problems. *IEEE Transactions on Evolutionary Computation*, 14(04) :618–635. 83

Forrester, A. I. J., Sóbester, A., and Keane, A. J. (2008). *Engineering Design via Surrogate Modeling : A Practical Guide*. Wiley Press. 81

Galup-Montoro, C. and Schneider, M. C. (2007). *MOSFET Modeling for Circuit Analysis and Design*. World Scientific Press. 76

Gary Wang, G. and Shan, S. (2006). Review of metamodeling techniques in support of engineering design optimization. *Journal of Mechanical Design*, 129(04) :370–380. 74

Gaspar, B., Teixeira, A. P., and Soares, C. G. (2014). Assessment of the efficiency of kriging surrogate models for structural reliability analysis. *Probabilistic Engineering Mechanics*, 37 :24–34. 82

Geng, H. (2005). *Semiconductor Manufacturing Handbook*. McGraw-Hill Press. 71

Gildenblat, G. (2010). *Compact Modeling Principles, Techniques and Applications*. Springer Press. 73

Goudon, T., Miljanović, V., and Schmeiser, C. (2007). On the Shockley-Read-Hall model : Generation-recombination in semiconductors. *SIAM Journal on Applied Mathematics*, 67(04) :1183–1201. 87

Grabinski, W., Nauwelaers, B., and Schreurs, D. (2006). *Transistor Level Modeling for Analog/RF IC Design*. Springer Press. 74

Hametner, C., Stadlbauer, M., Deregnaucourt, M., Jakubek, S., and Winsel, T. (2013). Optimal experiment design based on local model networks and multilayer perceptron networks. *Engineering Applications of Artificial Intelligence*, 26(01) :251–261. 80

http ://itrs.net (2013). International technology roadmap for semiconductors (ITRS). 71

Iannaccone, G., Curatola, G., and Fiori, G. (2004). *Effective Bohm quantum potential for device simulation based on drift-diffusion and energy transport*, chapter 64. in G. Wachutka and G. Schrag (eds.), Simulation of Semiconductor Processes and Devices, Springer Press. 86

Kleijnen, J. P. C. (2015). *Design and Analysis of Simulation Experiments*. Springer Press. 79, 81

Li, M., Lin, D., and Wang, S. (2010). Solving a type of biobjective bilevel programming problem using NSGA-II. *Computers and Mathematics with Applications*, 59(02) :706–715. 83

Lucifredi, A., Mazzieri, C., and Rossi, M. (2000). Application of multi-regressive linear models, dynamic kriging models and neural network models to predictive maintenance of hydroelectric power systems. *Mechanical Systems and Signal Processing*, 14(03) :471–494. 82

Maricau, E., De Wit, P., and Gielen, G. (2008). An analytical model for hot carrier degradation in nanoscale CMOS suitable for the simulation of degradation in analog IC applications. *Microelectronics Reliability*, 48(8-9) :1576–1580. 78

Mason, R. L., Gunst, R. F., and Hess, J. L. (2003). *Statistical Design and Analysis of Experiments with Applications to Engineering and Science*. Wiley Press. 79

Matta, A., Pezzoni, M., and Semeraro, Q. (2012). A kriging-based algorithm to optimize production systems approximated by analytical models. *Journal of Intelligent Manufacturing*, 23(03) :587–597. 82

Miyano, T., Fujito, M., Kato, M., and Tsuge, H. (1992). A mechanism for impact ionization of Si n-channel MOSFETs. *Solid-State Electronics*, 35(01) :89–94. 78

Mohammadi, S. and Afzali-Kusha, A. (2010). Modeling of drain current, capacitance and transconductance in thin film undoped symmetric DG MOSFETs including quantum effects. *Microelectronics Reliability*, 50(03) :338–345. 72

Montero, J. M., Fernàndez-Avilés, G., and J. Mateu, J. (2015). *Spatial and Spatio-Temporal Geostatistical Modeling and Kriging*. Wiley Press. 82

Pagey, M. P. (2002). Characterization and modeling of hot-carrier degradation in submicron nmosfets. Master's thesis, Vanderbilt University. 78

Pradhan, K. P., Mohapatra, S. K., Sahu, P. K., and Behera, D. K. (2014). Impact of high-κ gate dielectric on analog and RF performance of nanoscale DG-MOSFET. *Microelectronics Journal*, 45(02) :144–151. 74

Razavi, B. (2017). *Design of Analog CMOS Integrated Circuits*. McGraw-Hill Press. 93

Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P. (1989). Design and analysis of computer experiments. *Statistical Science*, 04(04) :409–423. 74

Schonlau, M. (1997). *Computer Experiments and Global Optimization*. PhD thesis, University of Waterloo. 73

Schroder, D. K. (2006). *Semiconductor Material and Device Characterization*. Wiley-IEEE Press. 78

Sharma, D. and Vishvakarma, S. K. (2015). Analyses of DC and analog/RF performances for short channel quadruple-gate gate-all-around MOSFET. *Microelectronics Journal*, 46(08) :731–739. 93

Silvaco (2012). *Atlas User Manual : Device Simulation Software*. 85

Song, Y., Zhou, H., Xu, Q., Luo, J., Yin, H., Yan, J., and Zhong, H. (2011). Mobility enhancement technology for scaling of CMOS devices : Overview and status. *Journal of Electronic Materials*, 40(07) :1584–1610. 76

Tachi, K., Vulliet, N., Barraud, S., Kakushima, K., Iwai, H., Cristoloveanu, S., and Ernst, T. (2011). Influence of source/drain formation process on resistance and effective mobility for scaled multi-channel MOSFET. *Solid-State Electronics*, 65-66 :16–21. 76

Taur, Y. (2000). An analytical solution to a double-gate MOSFET with undoped body. *IEEE Electron Device Letters*, 21(05) :245–247. 72

Telen, D., Logist, F., Van Derlinden, E., Tack, I., and Van Impe, J. (2012). Optimal experiment design for dynamic bioprocesses : A multi-objective approach. *Chemical Engineering Science*, 78 :82–97. 80

Wagner, M., Karner, M., Cervenka, J., Vasicek, M., Kosina, H., Holzer, S., and Grasser, T. (2006). Quantum correction for DG MOSFETs. *Journal of Computational Electronics*, 5(4) :397–400. 71

Wang, W., Lu, H., Song, J., Lo, S. H., and Taur, Y. (2010). Compact modeling of quantum effects in symmetric double-gate MOSFETs. *Microelectronics Journal*, 41(10) :688–692. 72

Weber, O., Casse, M., Thevenod, L., Ducroquet, F., Ernst, T., and Deleonibus, S. (2006). On the mobility in high-κ /metal gate MOSFETs : Evaluation of the high-κ phonon scattering impact. *Solid-State Electronics*, 50(04) :626–631. 76

Zhang, Y., Hu, S., Wu, J., Zhang, Y., and Chen, L. (2014). Multi-objective optimization of double suction centrifugal pump using kriging metamodels. *Advances in Engineering Software*, 24 :16–26. 74

Chapter 4

A TCAD based approach for the performance improvement of nanoscale DG MOSFETs including interface traps

« Design is so simple. That's why it's so complicated. »

Paul Rand

Abstract- *The current revolution manifesting in semiconductor industry has made many innovative designs feasible especially with the availability of powerful TCAD tools. As a usual practice in design procedures, some structural parameters are varied within a specific range and the obtained figures of merit are compared. The idea behind this chapter follows this vision, where we propose various modifications in order to improve the efficiency of scaled DG MOSFET devices using ATLAS 2D simulator. So, we scrutinize the influence of the gate work function, the molar fraction of silicon in a SiGe alloy based channel, the source/drain extensions and the channel doping shape. All these amendments are carried out in the context of digital, analog/RF applications. The conducted numerical experiments have revealed the vital role played by TCAD tools in providing optimized circuits and opening up the door towards reliable technologic solutions under low costs.*

4.1 Introduction

The actual advances achieved in semiconductor industry have been rendered feasible with the availability of a set of powerful software packages known as Technology Computer Aided Design (TCAD) tools. Because of the possibility of conducting both process and device simulations using a TCAD environment, such paradigm is of a great deal not only for practical purposes but also for integrated circuits designers. Moreover, the adoption of these numerical artifacts allows gaining deeper knowledge regarding the fabrication process flow in addition to the implemented device without the need to be present in a real cleanroom [Ismail et al., 2012].

The operation of a TCAD tool is based on the imitation of the fundamental properties of the considered system. This is achieved through the investigation of physical partial differential equations governing the semiconductor process and the device operation jointly with computer experiments. The cornerstone of a simulator includes the selection of key models, exploitation of introduced approximations and assessment of the simulation findings. Despite that the simulation procedure delivers some insights about the optimization of performances, but it is strongly dependent on the choice of the device models [Velayudhan, 2016].

It is a usual procedure for device designers to modify some design specifications such as geometric or electronic parameters, which consequently affects the fabrication process besides the figures of merit of the elaborated device. The incentive behind this chapter is issued from this vision, where the ATLAS 2D simulator is qualified as the workhorse for the performed numerical simulations. In this chapter, we propose several structural modifications to improve the efficiency of the DG MOSFET device. Hence, we consider the impact of various critical aspects including the gate work function, the molar fraction of silicon in a SiGe based channel, the source/drain extensions and the channel doping shape. All these factors are accounted for in the evaluation of the subthreshold, the analog/RF and linearity performance. Several numerical codes implemented using ATLAS 2D simulator are investigated for this objective.

4.2 Technology Computer-Aided Design simulation for nanoscale design

During the last four decades, Technology Computer-Aided Design (TCAD) has passed from a one dimensional simulation in the 1970s to a full two dimensional and three dimensional simulations of nowadays fabrication processes and devices. Such mature stage of development couldn't be reached without the accelerated advance of both computer technology and mathematical modeling techniques. Therefore, the accurate forecasting of costs and performance can be achieved. The International Technology Roadmap for Semiconductors (ITRS) states that the adoption of technology computer-aided design may lead to the lowering of technology development costs up to 40% by reducing the number of experiment tests and the development cycle time. Hence, it is essential to acquire a deep comprehension of the correlation between the process configuration variables and the device parameters, where TCAD tools may offer pertinent insights regarding various important physical mechanisms. It is worthy to use TCAD-based transistor design at the early stage of development to account for statistical variability and reliability information in the process design kit. TCAD consists mainly of two parts, process design and device simulation as depicted in Figure 4.1. The process design refers to the modeling of various

fabrication stages. The objective of the device design resides in obtaining components with reduced power consumption, fabrication cost and elevated reliability. This is achieved through computer based experiments rather than expensive practical tests in order to predict accurately the electrical behavior of a new device [Maiti, 2017].

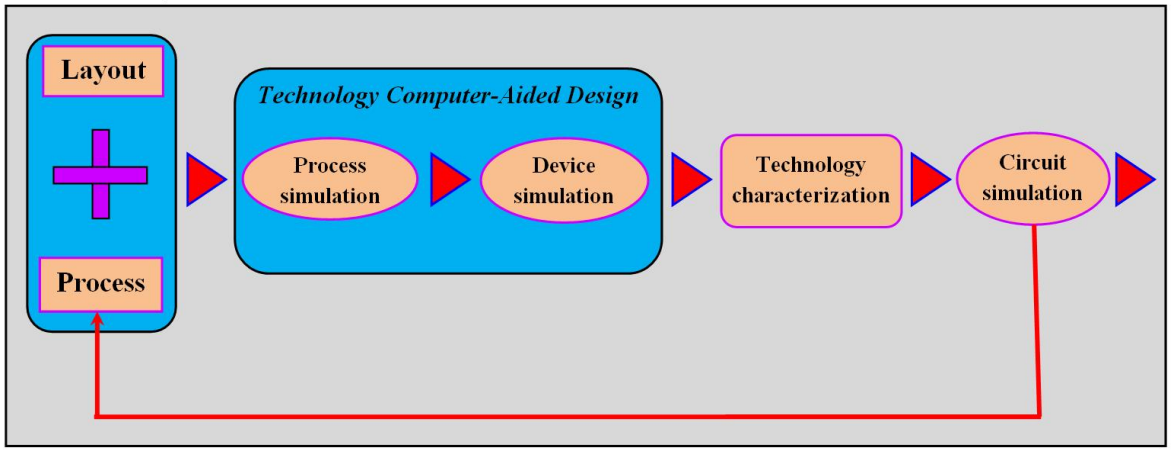

FIGURE 4.1 Conventional role of technology computer-aided design in integrated circuits processing.

For the description of the transport properties in the device, the drift-diffusion model remains the milestone of the industry due to its validity for explaining many important characteristics. The most relevant elements of the drift-diffusion theory are the Poisson and current continuity equations. An optional equation is the hydrodynamic model governing the carrier energy/temperature distribution. The aforementioned equations are solved self-consistently in order to determine the electrostatic potential and the carriers' concentrations energies. Since the obtained set of equations is coupled nonlinearly, an exact method does not exist and a nonlinear iteration schema starting from an initial guess is needed. Various kinds of initial guess solutions are used in the TCAD simulator. The simple charge neutral assumption is commonly used to obtain the first bias point. The convergence is strongly related to the choice of the initial guess solution because an initial guess close to the solution leads to less nonlinear iteration. The TCAD tool implements an adaptive method to control the bias step after the successful convergence of a previous solution [Fu et al., 2014]. A diagram of the biasing procedure is provided in Figure 4.2.

It should be mentioned that the three top TCAD vendors in terms of software product revenue market size are Synopsys-Avanti-TMA, Integrated Systems Engineering (ISE) and Silvaco. The TCAD software market in 2001 was estimated to 30 million dollars, with prospects to grow at a compound annual growth rate of 10%. The TCAD landscape is elucidated in Table 4.1, [Wu, 2003].

TABLE 4.1 Principal technology computer-aided design software vendors.

Company	Process simulation	Device simulation	Platform
Synopsys	Taurus-TSuprem4	Taurus-Medici	Unix
ISE	DIOS	DESSIS	Unix/Windows
Silvaco	ATHENA	ATLAS	Unix/Windows

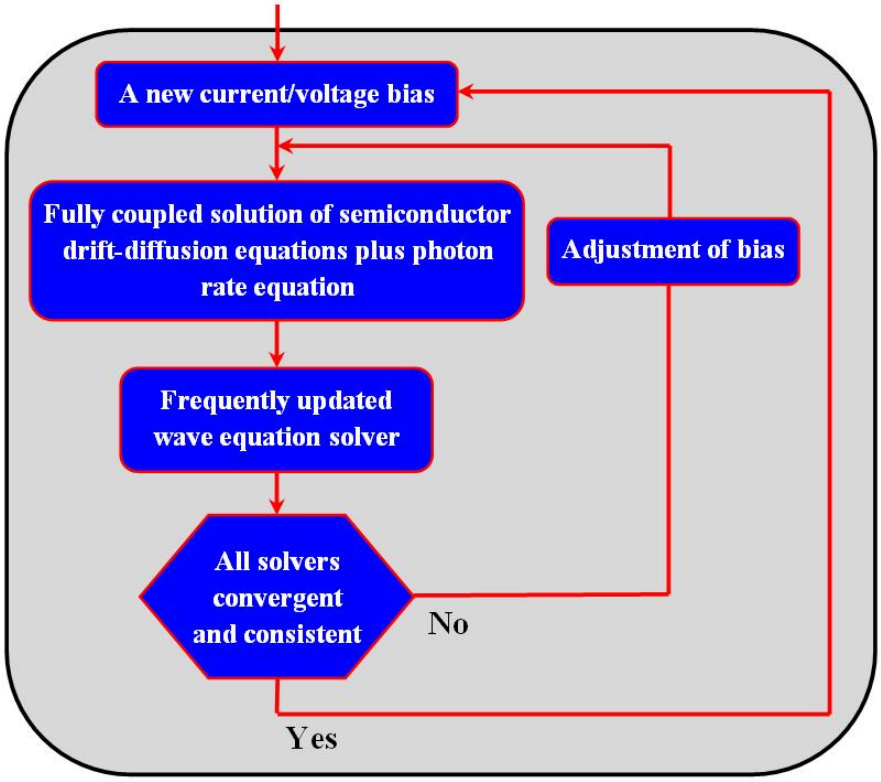

FIGURE 4.2 Flow diagram of a standard device simulation running.

4.3 An improved gate engineering based design for nanoscale DG JL MOSFETs including interface traps

Due to the work function difference between the metal and the semiconductor in MOSFET devices, a band bending occurs. Additional band bending under the gate are generated by the interface charge density, which in turn introduces changes in flat band voltage in the damaged region [Pagey, 2003]. Mainly, two device properties are significantly degraded, the threshold voltage increases in presence of interface traps making the device less sensitive to be turned on. The second property is related to the local channel electron mobility, which is reduced as a result of the scattering phenomena. Both effects can be handled analytically by the introduction of additional terms in the compact models associated to a parameter in the fresh case [Bentrcia et al., 2012].

In this section, we aim at highlighting the immunity of a new design called Dual Material Double Gate Junctionless MOSFET (DM DG JL MOSFET) against the hot carrier degradation effect at nanoscale level. The effect of interface traps on the threshold voltage (V_{th}) has been carried out by extensive simulation using ATLAS 2D simulator. In order to illustrate the superior performance of the proposed structure, the single material junctionless and conventional DG MOSFETs are used as benchmarks, where the effect of the interface trap density and damaged region length are also taken into account in the proposed investigation.

4.3.1 Numerical simulations

The structures of the proposed DM DG JL MOSFET in addition to the Single Material (SM) DG JL MOSFET designs are depicted in Figure 4.3. The subthreshold characteristics

of these devices are handled using the ATLAS 2D simulator [Silvaco, 2012]. The channel region is lightly doped with similar polarity as the source and the drain extensions, which allows the elimination of source and drain/channel junctions, so that less variability in doping profile is provided [Chebaki et al., 2012]. In the case of acceptor type interface traps, they can be expressed by fixed negative charges if the trap level is located beneath the Fermi level [Gautam et al., 2011]. As an additional feature in the proposed design, we consider a more complicated shape for the gate, where it is formed by the adjunction of materials with different work function values. Hence, the channel is divided into two parts, where L_1 and L denote the second gate length and the channel length, respectively. The gradual step change at the interface of two-gate materials results in the suppression of short channel and hot carrier effects by substantially reducing the electric field peak at the drain side [Djeffal et al., 2011].

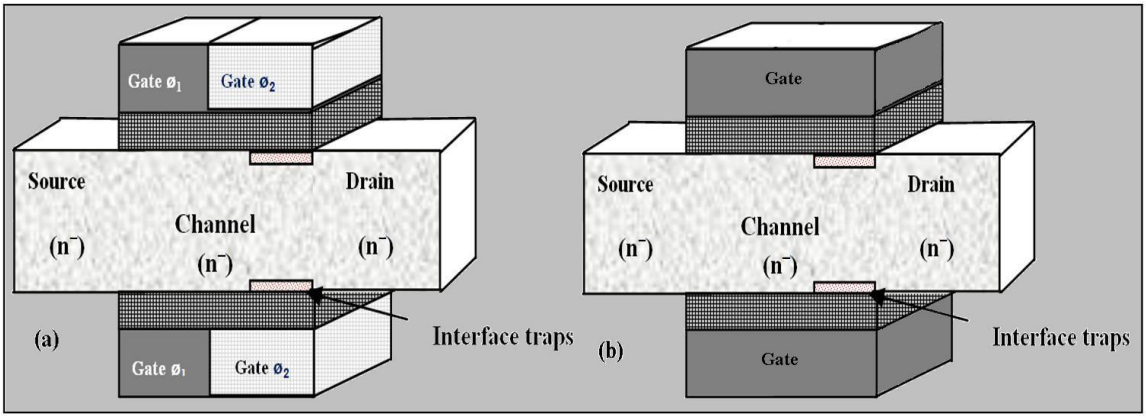

FIGURE 4.3 Longitudinal cross sectional views of (a) DM DG JL MOSFET structure (b) SM DG JL MOSFET structure.

During the simulation, we focus on analyzing the effect of the gate material work function (ϕ_2), the interface trap density (N_F) and the damaged region length (L_d) variations on the threshold voltage behavior. The drain voltage (V_{ds}) is fixed at 0.2 V, whereas the gate voltage (V_{gs}) is varied from -1V to 2 V respectively. The interface trap density has two extreme values (0 cm^{-2} for the fresh device and 5×10^{-12} cm^{-2} for the damaged one). The values of parameters which remain unchanged for various simulations are summarized in Table 4.2. It should be noted that more effects have been activated in the simulation model section to specify additional constraints such as the field dependent mobility, which is needed to model the velocity saturation effect. Shockley-Read-Hall recombination/generation is also included in the simulation to account for leakage currents. All simulations have been carried out at ordinary room temperature.

4.3.2 Results and discussion

In Figure 4.4, we compare electron concentrations at the Si/SiO_2 interface for different devices at a gate voltage equals to 2 V. It is shown that the electron concentration in the middle of the channel is reduced for the junctionless devices as a consequence of the uniformity of dopants in the whole channel and remains unaffected by the interface traps near the drain side in contrast to the conventional structure. Therefore, the threshold voltage shifts of hundreds of mV between various devices can be obtained. The electron concentrations of devices at the source and the drain are constants, which indicate the achieved convergence of our simulation.

TABLE 4.2 Main simulation parameters.

Parameter	SM DG MOSFET	SM DG JL MOSFET	DM DG JL MOSFET
L (nm)	50	50	50
L_1 (nm)	/	/	30
t_{si} (nm)	10	10	10
t_{ox} (nm)	2	2	2
$N_{D/S}$ (cm^{-3})	10^{20}	/	/
N_A (cm^{-3})	10^{16}	10^{16}	10^{16}
ϕ_1 (eV)	4	4	4

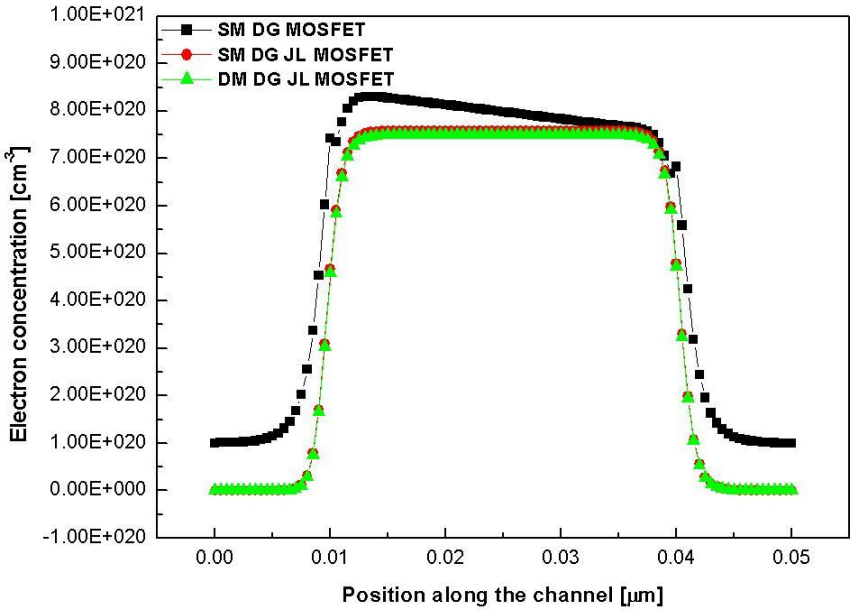

FIGURE 4.4 Electronic concentration as a function of the position along the channel for the three considered structures with the damaged region length L_d equals to 10 nm.

Figure 4.5 depicts the surface potential versus the position along the channel for a given gate voltage (V_{gs}= 2 V). We observe that the potential drops across the source and the drain regions is significant. This is can be explained by the electron accumulation at the surface, which makes the channel on-resistance smaller than the source/drain resistance. We notice also that the surface potential in the conventional case is altered by the existence of the interface charge density near the drain side, which expresses less immunity compared to both junctionless devices.

Figure 4.6 illustrates the relative degradation of the threshold voltage as a function of the gate material work function (ϕ_2). The influence of the gate material is evaluated by considering work function values comprised between 3 eV and 4 eV with a step of 0.2 eV. In this interval, we can find many special elements that are widely used for gate engineering purposes such as high-κ magnesium [Wilk et al., 2001]. From the presented curve, it is clearly indicated that the proposed design has higher immunity against the hot carrier effect in comparison to single material junctionless design. The resulted degradation is negligible for values inferior to 3.8 eV and increases rapidly beyond this value to reach a maximum relative degradation of 26.17% for the single material junctionless DG MOSFET.

The simulation results of the threshold voltage relative degradation versus the inter-

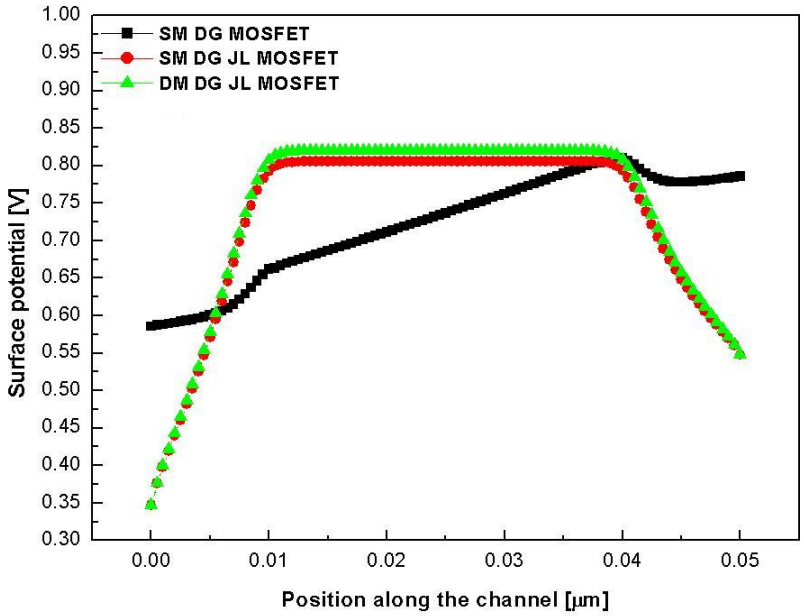

FIGURE 4.5 Surface Potential as a function of the position along the channel for the three considered structures with L_d equals to 10 nm.

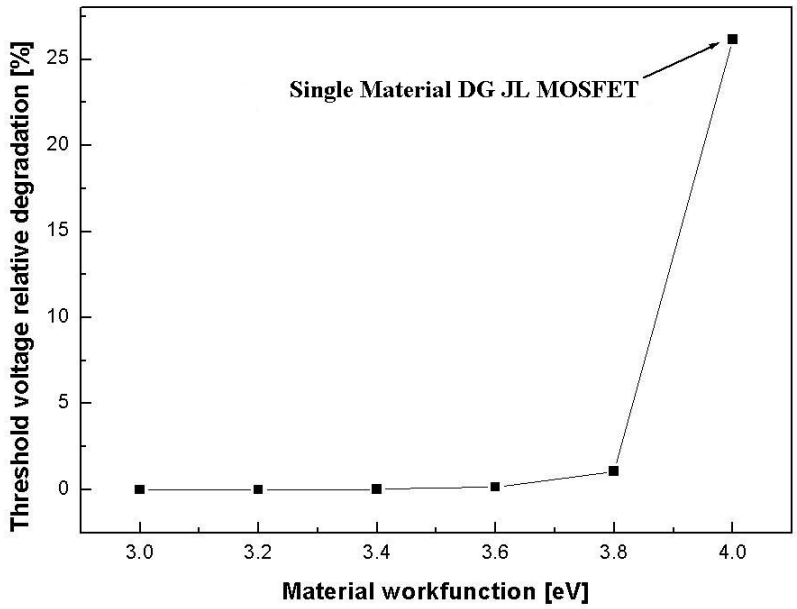

FIGURE 4.6 Relative degradation of the threshold voltage versus the material work function with L_d equals 20 nm.

face trap density for different values of the material work function are plotted in Figure 4.7. It is shown that the relative degradation remains insignificant for all trap density values in the case of both dual material junctionless DG MOSFET structures. While for the single material case and by approaching high values of trap density ($N_F = 5 \times 10^{11}$ cm^{-2}), the threshold voltage degrades drastically by about the quarter of its initial value. The ratio between the relative degradation in the best and worst cases corresponding to the dual material and single material based designs, respectively, is around 10^{-5}, which proves the efficiency and the impact of the proposed gate engineering based approach on the device scaling capability including the hot carrier effect at nanoscale level.

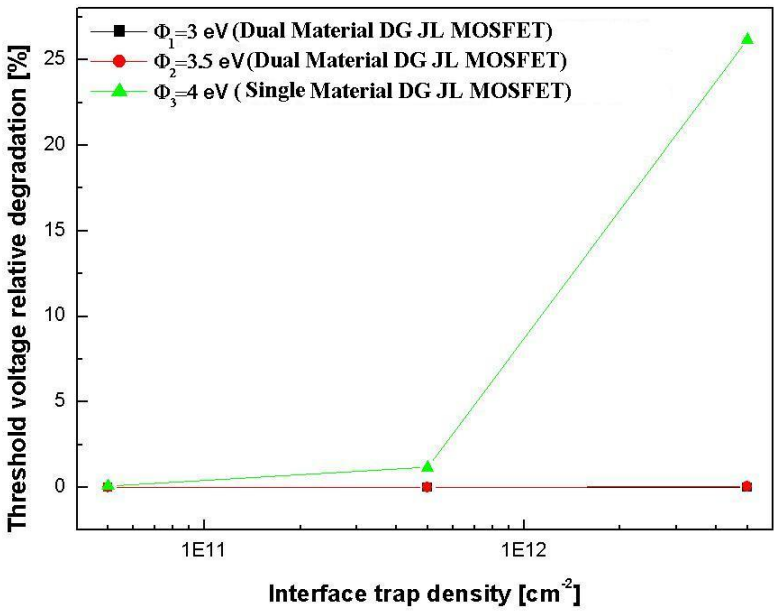

FIGURE 4.7 Relative degradation of the threshold voltage as a function of the interface trap density for different material work function values with L_d equals to 20 nm.

Figure 4.8 represents the threshold voltage relative degradation versus the damaged region length L_d for various material work function values. In general, the relative degradation shifts upward with the increase of the work function values. Moreover, it is only slightly increased for the dual material design, but it is increased considerably with the damaged region length in the case of single material design. This Figure concludes that as the interface charge region length increases, the effect of the interface charge density on the threshold voltage relative degradation becomes more apparent in the case of single material structures. Clearly, the DM DG JL MOSFET provides better threshold voltage immunity against hot carrier effect with respect to SM DG JL MOSFET in nanoscale domain.

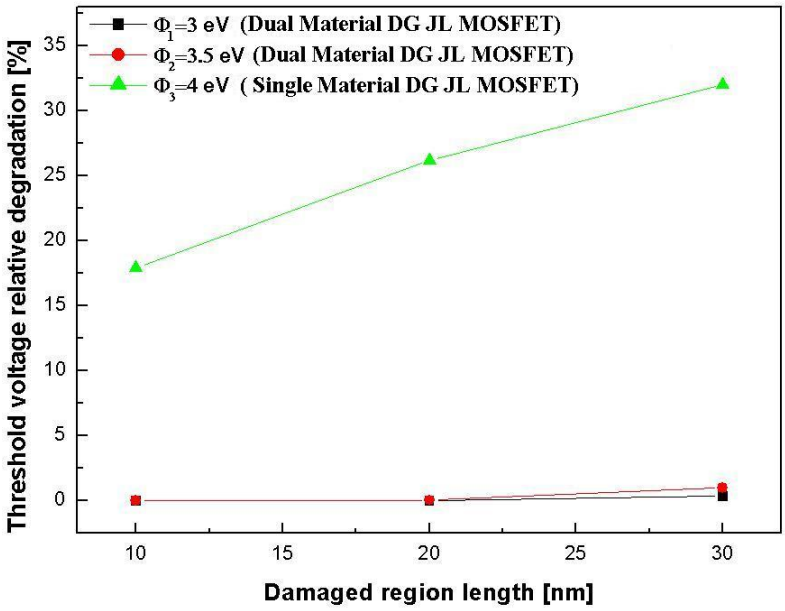

FIGURE 4.8 Relative degradation of the threshold voltage versus the damaged region length for different material work function values.

Based on the obtained simulation results, some general comments can be made regarding the reliability behavior of our novel design at nanoscale level. Low values of threshold voltage relative degradation have been recorded in the case of the DM DG JL MOSFET, which means better electrical and scaling performances in comparison to the conventional and single material junctionless designs. The following comparative table (Table 4.3) provides more accurate numerical results for the three investigated structures. In the case of the dual material device, we pick the best configuration obtained so far.

TABLE 4.3 Summary of the numerical simulation results.

Parameter	**SM DG MOSFET**	**SM DG JL MOSFET**	**DM DG JL MOSFET**
V_{th0} (mV)	-167.54	-476.32	-500.55
V_{th} (mV)	-68.05	-351.66	-500.54
$(\triangle V_{th}/V_{th0})$ (%)	59.39	26.17	10^{-5}

It can be deduced from this table that the DM DG JL MOSFET offers the most competitive design among the three included structures. The sensitivity of the later is highly improved, where it is possible to switch the device to the "ON" state by applying very low gate voltages. Furthermore, the immunity of such device against the hot carrier effect is incomparable to both other alternatives. All these advantages make our device a good choice for low power consumption and highly reliable circuits.

4.4 An improved SiGe based design for nanoscale DG MOSFETs including interface traps

Silicon has played a fundamental role in semiconductor industry, where it was employed as a basic material for the device channel material. The use of such material is strongly encouraged due to two main causes, the widespread existence of silicon within the earth's surface and the stability of its associated dioxide [Geng, 2005]. However, it is to mention that the first integrated circuit was based on germanium and since then other materials have been proposed and exploited in various microelectronics fields. The adoption of SiGe alloy as a channel material has attracted the attention of many research groups because of pertinent performance being gained in a few tens of nanoscale devices [Dollfus et al., 1997]. The main objective of introducing SiGe is the enhancement of the carrier mobility in the channel region based on the lower effective masses than in bulk silicon. By adopting this vision, it is possible to increase the drain current without suffering from mobility reduction obtained if high doping concentrations are considered. Another concern that should be also taken into account is the channel resistance, which needs to be suppressed or at least reduced without significantly affecting other performance parameters of the device. The use of SiGe channel has been reported among the techniques of reducing the channel resistance and shows to give up to 10% improvement in the on-resistance [Wang et al., 2011]. It is also demonstrated that advanced SiGe MOS technologies have the potential to make a significant impact on millimeter wave consumer products in terms of cost and ability to integrate multiple functions on the same die.

Our objective in this section is to explore physical insights into the device modeling of small signal parameters for SiGe DG MOSFET, while including the interface defects. As we know, ATLAS 2D numerical simulator is a good choice for studying the behavior of electronic devices including complex phenomena, which may be difficult or even impossible

to handle using available measurement setups. Hence, in order to take into account the influence of interface defects on the small signal parameters at nanoscale level, numerical simulation results instead of real measurement data are used for parameter extraction in this work. The current contribution may be considered as an attempt aiming at highlighting the response of more complicated SiGe multigate MOSFET devices under miniaturization constraints, in addition to their appropriateness to be included within different digital and analog applications.

4.4.1 Numerical simulations

In Figure 4.9, we depict the cross sectional view of SiGe DG MOSFET used in this work. It is clear from the figure that the device structure consists of highly n-doped source and drain regions, an intrinsic SiGe channel with two sided gates, and two dioxide (SiO_2) layers. Both the channel and the source/drain regions have uniform doping concentrations. In order to take into consideration the interface defects influence on the small signal parameters, we assume the existence of a fixed charge density near the drain side along a distance L_d at the interface between the SiGe channel and SiO_2 oxide regions. The creation mechanism can be explained by the injection of hot carriers. All simulations are carried out using two carrier types and the simulations are realized at room temperature (300 K). The generation/recombination effects are supported through the introduction of Shockley-Read-Hall (SRH) and Auger models. These models allow taking into account many phenomena such as carrier velocity saturation, carrier-carrier scattering in the high doping concentration, in addition to the dependence of mobility on vertical electric field. Although, we use the drift-diffusion model representing a first order approximation to the Boltzmann transport equation, where Poisson's equation is solved self consistently with the carrier continuity equation. The carrier distribution follows a Maxwell-Boltzmann distribution, with a carrier temperature equals to the lattice temperature [Silvaco, 2012]. There are some benefits in using classical simulation over quantum simulation based approaches for some classes of device regimes. Drift-diffusion simulations are considerably faster than quantum ballistic simulations and are also well matched to experimental results ([Fjeldly et al., 2006] and [https ://nanohub.org, 2014]).

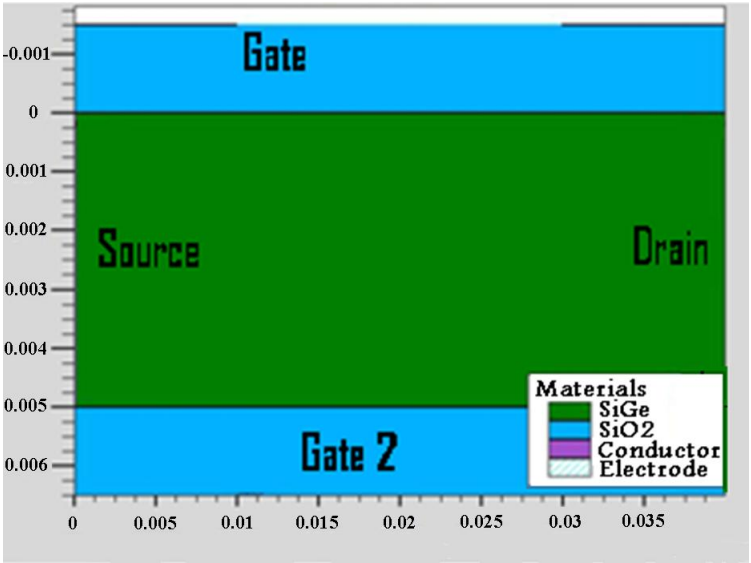

FIGURE 4.9 Cross sectional view of the DG MOSGET structure.

The values of the parameters in all simulations selected for the device design are given in Table 4.4. However, some input parameters (Ge mole fraction, interface defects, channel length and channel thickness) take values in a specified interval with a fixed step. The Ge mole fraction x denotes the ratio of Ge material in the $Si_{1-x}Ge_x$ alloy and is varied from 0 to 100% with a step of 20%. The channel length (L) is varied from 20 nm to 100 nm with a step of 10 nm, the channel thickness (t_{si}) is varied from 3 nm to 8 nm with a step of 1 nm, the gate voltage (V_{gs}) is varied from -1 V to 4 V with a step of 0.1 V and finally, the drain voltage (V_{ds}) which is varied from 0 V to 3 V with a step of 0.1V. During the computation of the values of the transconductance and output conductance, the drain and gate voltages are maintained to fixed values 0.3 V and 0.7 V, respectively.

TABLE 4.4 Simulation parameter values of the SiGe DG MOSFET structure.

Parameter	**Notation**	**Value**
Oxide thickness	t_{ox}	1.5 nm
Drain/source doping	$N_{D/S}$	$10^{20}cm^{-3}$
Channel doping	N_{Ch}	10^{15} cm^{-3}
Work function	ϕ_{MS}	4.55 eV
Interface trap distance	L_d	L/3

4.4.2 Results and discussion

Since the numerical investigation is focused on the quantification of the impact of interface defects on the device small signal parameters at nanoscale dimensions, we conduct a sensitivity analysis based on some input variables in order to obtain a complete scene of the behavior of SiGe DG MOSFET. In what follows, two basic small signal parameters are studied (the transconductance and the output conductance factors). The transconductance is required to be the highest possible. Whereas, the output conductance should be as small as possible so that better cut-off frequency and gain can be obtained [Arora, 2007].

The maximum transconductance of the DG MOSFET device is determined for various Ge mole fractions for both fresh and damaged cases as indicated in Figure 4.10. It is shown that the fresh DG MOSFET device has higher transconductance when compared to the damaged device. When there is an increase in the Ge mole fraction, the transconductance has two opposite tendencies. It increases slightly in the fresh case and decreases significantly in the device including interface traps. Hence, it can be concluded that Si based DG MOSFET device has higher immunity against the interface traps and such immunity decreases with the increase of introduced Ge mole fraction. In addition, a larger Ge mole fraction may adversely affect the interface between the gate oxide (SiO_2) and SiGe channel leading to a modification of the effect of traps generated at these interfaces. This is why a careful attention should be dedicated to the Ge amount used in analog circuit applications where higher transconductance is needed. The calculation of the relative degradation of the transconductance caused by the interface traps is smaller in the Si channel (1.4%) compared to Ge channel (2.3%), which means that the introduction of any amount of germanium leads to a decrease in the immunity of the DG MOSFET device.

Figure 4.11 shows the dependency of the maximum transconductance upon the device channel thickness for three values of the density of the interface defects. It can be easily noticed that the transconductance has an increasing behavior with respect to the

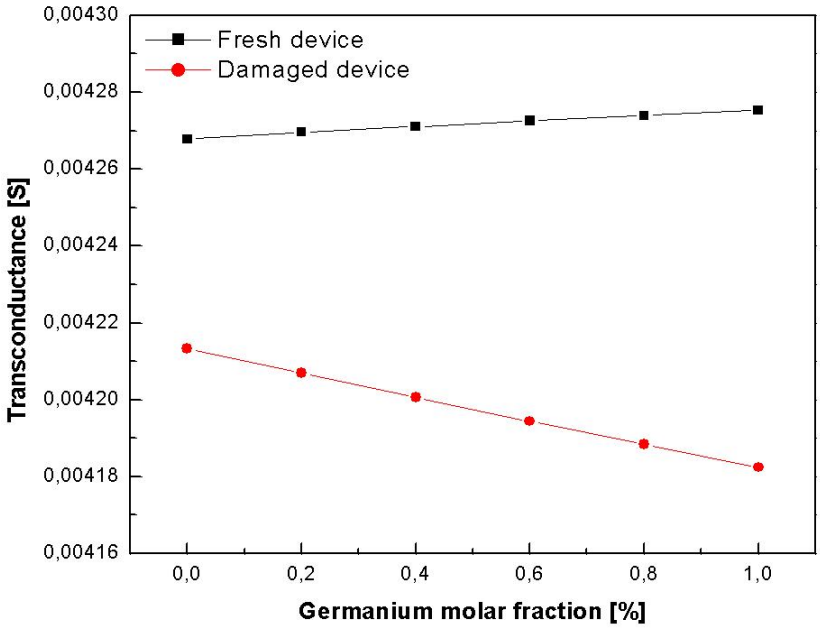

FIGURE 4.10 Variation of the maximum transconductance versus the Ge mole fraction (L=40 nm, t_{si}=5 nm, $N_F = 5 \times 10^{12}$ cm^{-2}).

channel thickness, which can be explained by the fact that the scattering effect in thin channels is more pronounced compared to the case of thick channels. Besides, the performance of the device is strongly affected by the density of the interface defects because of the inversion charge being trapped, which implies that for a damaged device, the transconductance is reduced with the increase of traps' density comparatively with a fresh device.

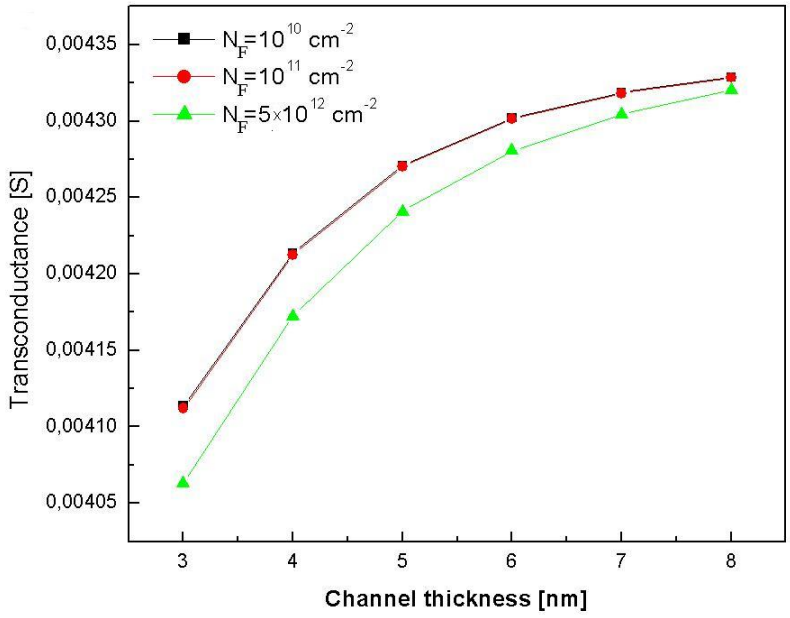

FIGURE 4.11 Variation of the maximum transconductance as a function of the channel thickness (L=40 nm, x=40%).

The maximum transconductance is depicted in Figure 4.12 for various channel lengths, where it is observed that as the length of the channel is increased, the device

has lower transconductance. For small channel lengths (L < 40 nm), the effect of the interface defects ($N_F = 5 \times 10^{12}$ cm^{-2}) is noticeable compared to other two cases, while for lengths superior to 40 nm, the transconductance values remain approximately the same for the three considered cases regardless of the defects density values. In other words, it may be inferred that the clear discrepancy for small lengths over device transconductance characteristics may be due to the short channel effect.

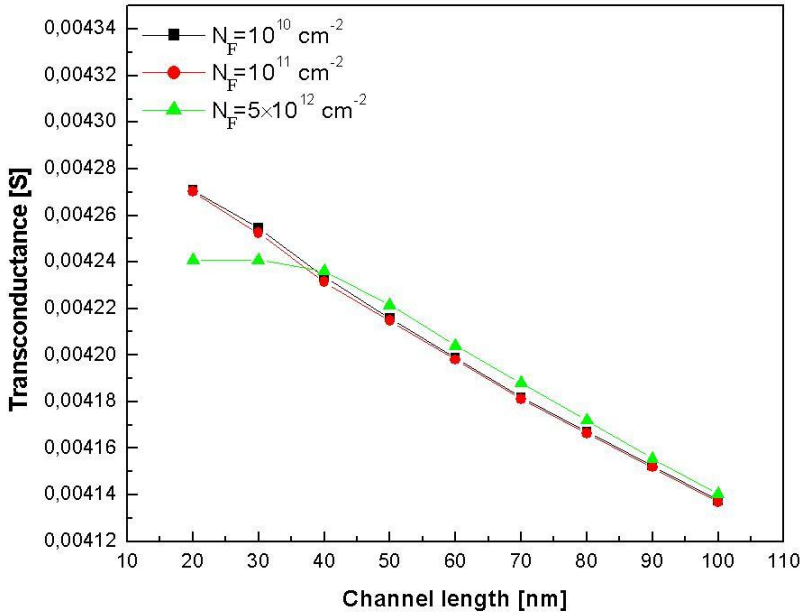

FIGURE 4.12 Variation of the maximum transconductance as a function of the channel length (t_{si}=5 nm, x=40%).

The minimum output conductance versus Ge mole fraction curves of DG MOSFET with and without traps is represented in Figure 4.13. The damaged device shows lower output conductance than the fresh one. Note that the discrepancy between both curves decreases with the increase of the Ge mole fraction to get the same value for a pure Ge channel device. It can be deduced from this figure that using a Ge based channel gives higher immunity against the interface defects regarding the output conductance parameter.

Figure 4.14 shows the minimum output conductance as a function of the channel thickness. The obtained values are relatively independent of the interface defect densities, indicating that the selected Ge mole fraction (x=40%) leads to a good immunity against the interface defects even for high values.

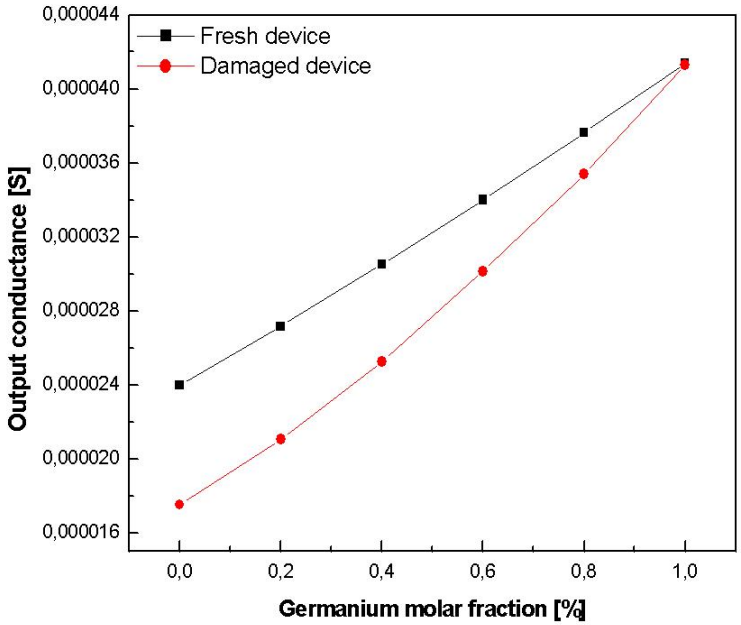

FIGURE 4.13 Variation of the minimum output conductance as a function of the Ge mole fraction (L=40 nm, t_{si}=5 nm).

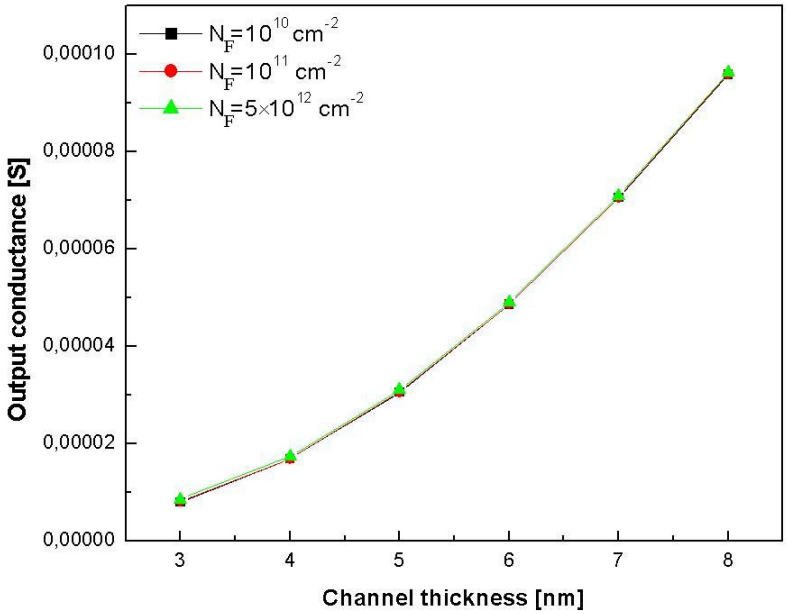

FIGURE 4.14 Variation of the minimum output conductance as a function of the channel thickness (L=20 nm x=40%).

4.5 An improved source/drain extensions based design for nanoscale DG JL MOSFETs including interface traps

Junctionless multigate MOSFETs have emerged in recent years as a paramount candidate to remedy the many parasitic effects arising from scaling down to sub 100 nanometer channel lengths. In the junctionless elaboration of transistors, a uniform doping with the same type (n- or p-type) is affected to the device channel making the manufacturing process easier due to the elimination of both lateral abrupt doping and high thermal budget

([Colinge et al., 2011] and [Holtij et al., 2015]). Reasons behind the wide accreditation of such paradigm are due to the reduced short channel effects, excellent subthreshold slope behavior and lower drain induced barrier lowering ([Rios et al., 2011] and [Jeon et al., 2013]). However, the continuous shrinking of the DG JL MOSFETs necessitates the reduction of the doping concentration along the channel in order to get appropriate threshold voltage and subthreshold slope values. Under such condition, source and drain resistance values become larger leading to a degradation of the ON-current. The impact of parasitic resistance is even more crucial in nanoscale DG device because it holds two channels for current flow ([Orouji et al., 2012] and [Charmi et al., 2013]). Hence, additional efforts should be dedicated to the analysis of the source/drain extension regions in addition to their influence on the device analog and digital behavior.

Our aim in this section is to investigate numerically the analog/RF performance of deeply scaled double gate junctionless MOSFETs when the device is subject to both interface defects and quantum confinement effect. In order to elucidate the impact of source/drain extensions doping, a comparative analysis is carried out between the conventional DG JL MOSFET and the DG JL MOSFET with highly doped source/drain extensions. Several numerical simulations are conducted in terms of some performance criteria such as the linearity parameters.

4.5.1 Numerical simulations

Since almost closed form models are based on approximations or simplifying hypothesis, which reduce their efficiency in dealing with deep nanoscale devices, numerical frameworks are finding increasing deployment by designers' community [Mohammadi and Afzali-Kusha, 2010]. Hence, we use ATLAS 2D simulator to develop a numerical model that is more accurate and allows catching difficult aspects of analog/RF device performance. The main features of devices under consideration (Conventional and including highly doped source/drain extensions DG JL MOSFET) are outlined in Figure 4.15, where both structures are assumed subject to critical channel length (L=10 nm) and channel thickness. In this case, the quantum and ballistic transport effects should be taken into account in order to get accurate numerical models [Fiori and Iannaccone, 2003].

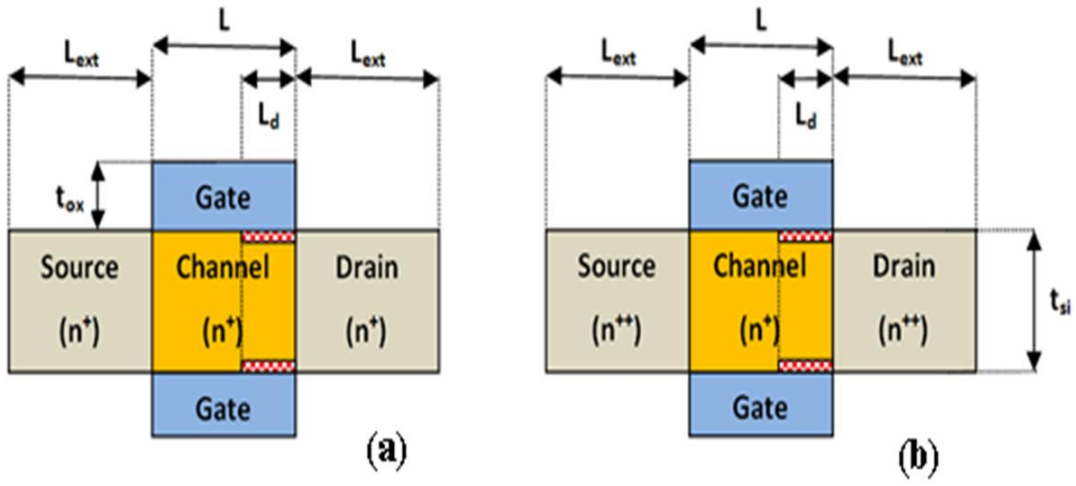

FIGURE 4.15 Cross sectional view of (a) DG JL MOSFET device including interface defects (b) DG JL MOSFET with high doped source/drain extensions and including interface defects.

It should be stated that both devices have identical geometrical and electrical characteristics except for the source/drain extensions which are highly doped in our proposed design compared to the conventional DG JL MOSFET. The full list of adopted parameters values is summarized in Table 4.5.

TABLE 4.5 Configuration of the conventional and the proposed DG JL MOSFET structures.

Parameter	Notation	Conventional design	Proposed design
Channel length	L	10 nm	10 nm
Extensions length	L_{ext}	10 nm	10 nm
Interface traps localisation	L_d	L/3	L/3
Interface traps density	N_F	5×10^{12} cm^{-2}	5×10^{12} cm^{-2}
Channel thickness	t_{si}	4 nm	4 nm
Oxide thickness	t_{ox}	2 nm	2 nm
Channel doping	N_d	10^{18} cm^{-3}	10^{18} cm^{-3}
Extensions doping	N_{ext}	10^{18} cm^{-3}	10^{20} cm^{-3}

During simulation, the following models have been invoked in order to approach the real behavior of the device [Silvaco, 2012]. For carriers' transport, Schrödinger and Poisson equations are self-consistently solved in order to provide a quantized description of states density when quantum confinement effects are present. The carrier recombination is expressed using the well known Shockley-Read-Hall model, where a two stage transition between bands is achieved through a localized state within the band gap. An additional recombination mechanism is Auger recombination in which a third carrier obtain the emitted energy. The hot carriers' injection is proposed to describe the passage of carriers from the channel to the oxide, which depends on several factors such as the gained energy and the phonon scattering at the level of the interface. Once the device structure is completely identified, a set of partial differential equations are used to deduce an equivalent discreet system, which needs to be solved on a mesh of points within the device. Therefore, we use iterative procedures that refine successively estimates of the solution. Since there is no guarantee that a particular method will converge, the following iterative methods are applied in this order : Newton, Gummel and Block respectively for which the maximum allowed number of trials started with the same guess is fixed to 10.

4.5.2 Results and discussion

In this section, the analog/RF performance of the aforementioned devices is assessed. To this end, several criteria are extracted and compared to study the impact of source/drain extensions doping and interface defects on the overall device operation. The surface potential along the channel with V_{ds}= 0.4 V and V_{gs}= 5 V is shown in Figure 4.16. From this figure, it can be observed that the introduction of highly doped extensions leads to a reduction of the surface potential beneath the gate, which allows better control of the threshold voltage. In addition, the incorporation of the interface defects introduces a shift in the surface potential profile near the drain side of the channel. Such alteration has been proved to cause a degradation of DG MOSFET devices especially their subthreshold behavior parameters.

Figure 4.17 presents the drain current responses at V_{gs} = 0.8 V. One can notice that the magnitude of the drain current in the case of the device with highly doped extensions with or without interface defects is significantly larger than its conventional counterpart.

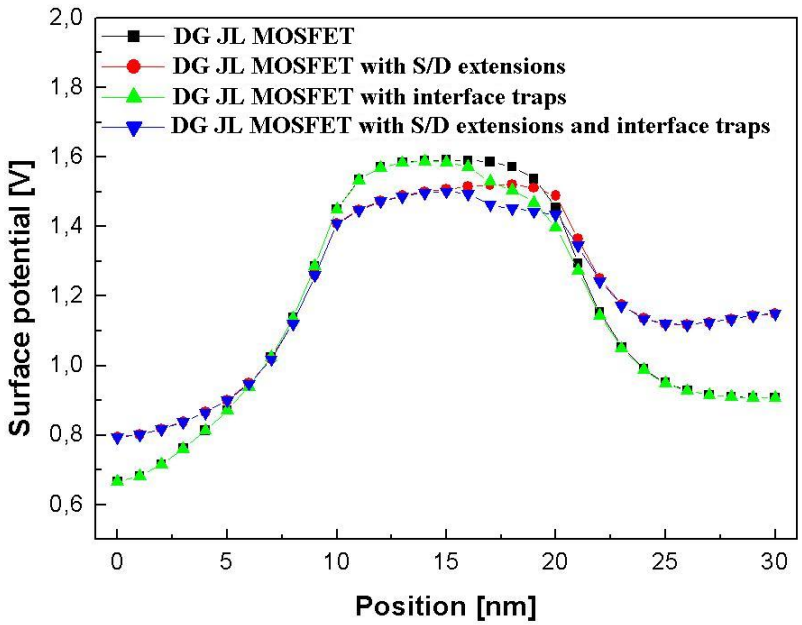

FIGURE 4.16 Surface potential Variation as a function of the channel position.

Some calculations allow us to deduce that the ratio of improvement is about 73 times for the fresh design and for the design with interface defects.

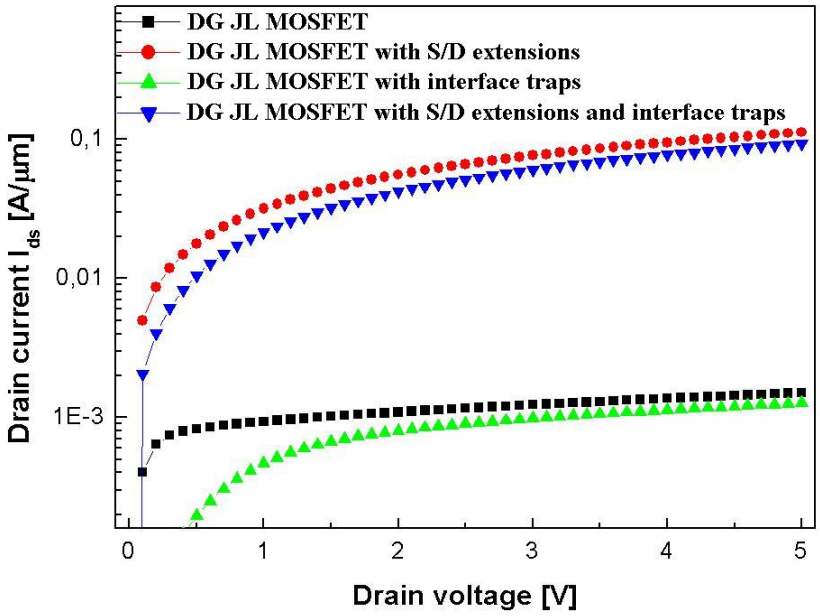

FIGURE 4.17 Variation of the drain current I_{ds} in log scale as a function of the drain voltage V_{ds}.

The variation of the transconductance versus the gate voltage for V_{ds}= 0.4 V is given in Figure 4.18. It is clearly shown that the transconductance values are higher under all regimes of operation (subthreshold, linear and saturation) for the junctionless design with highly doped extensions compared to the conventional one. Besides, the presence of interface defects reduces the transconductance maximal value and shifts the curve to the right which can be interpreted by the increase of the threshold voltage due to trap filling.

The Transconductance Generation Factor (TGF) reflects the performance of em-

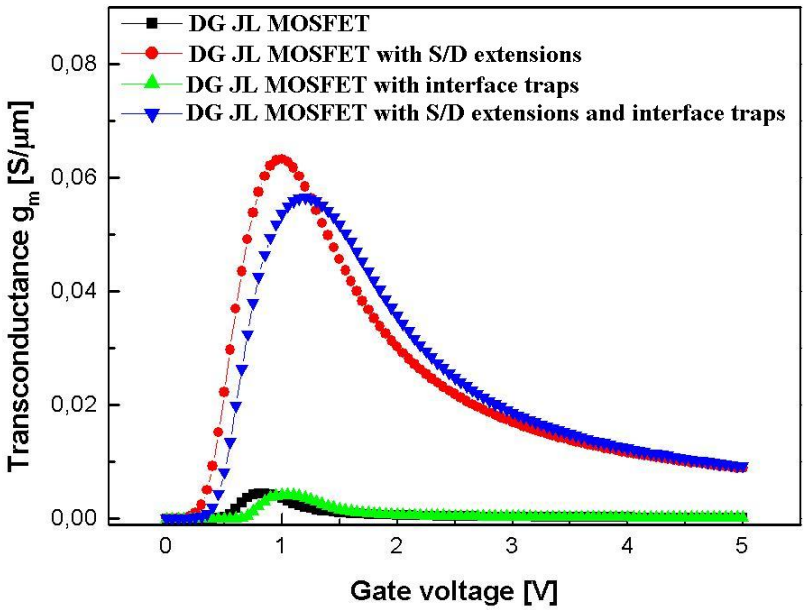

FIGURE 4.18 Evolution of the transconductance g_m versus the gate voltage V_{gs}.

ploying the drain current to reach a well defined transconductance value. Therefore, high values for such parameter is a desirable requirement for the analog and RF circuits development [Djeffal et al., 2016]. The transconductance generation factor can be calculated using the formula :

$$\mathrm{TGF} = \frac{g_m}{I_{ds}} \tag{4.1}$$

where g_m is the transconductance and I_{ds} denotes the drain current. As per the Figure 4.19, the DG JL MOSFET device indicates small values of TGF as compared with that for conventional devices for an applied drain voltage V_{ds}= 0.4 V. Therefore, such source/drain region modification can improve this criterion because the TGF can be termed as the available gain per unit value of the power dissipation. As predicted and due basically to the degradation of the drain current, designs with interface defects have much higher TGF.

The cut-off frequency f_T is one of the most important parameters for evaluating analog/RF performance of devices. Generally, f_T is the frequency when the current gain is unity. The cut-off frequency is defined by the equation [Sze and Ng, 2007] :

$$f_T = \frac{g_m}{2\pi C_{gs}} \tag{4.2}$$

with C_{gs} is the gate to source capacitance obtained numerically to be equal to $6.4 \times 10^{-12}(F/\mu m)$ for all four types of structures. However, it should be mentioned that the behavior of C_{gs} may be more complicated for other case studies.

The variation of the cut-off frequency is depicted in Figure 4.20 for different considered structures. It can be seen that junctionless devices with source/drain extensions provide higher cut-off frequency, which is mainly due to the superior values of the transconductance since the parasitic gate to source capacitance is approximately constant and is not affected by the source/drain regions neither by the presence of the interface defects near the drain side.

A summary of analog obtained performance is provided in Table 4.6. For different

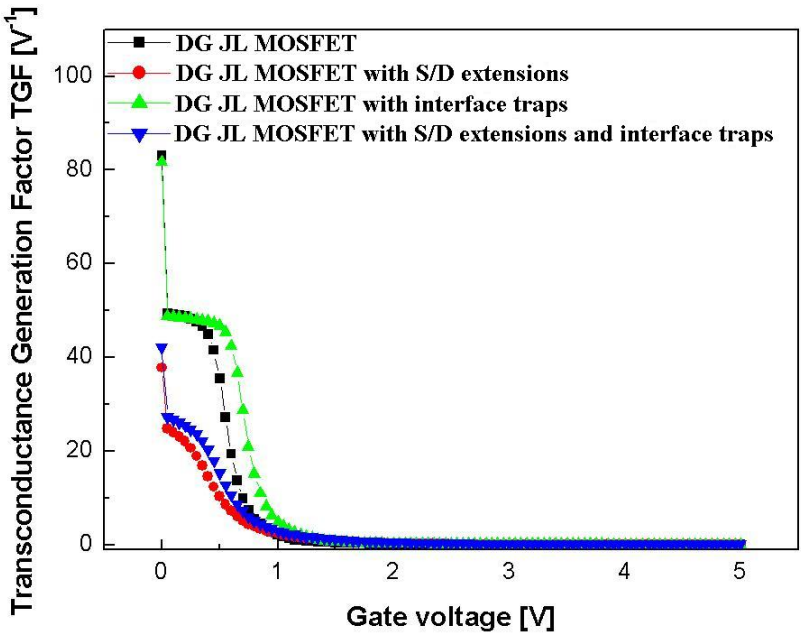

FIGURE 4.19 Variation of the transconductance generation factor TGF as a function of the gate voltage V_{gs}.

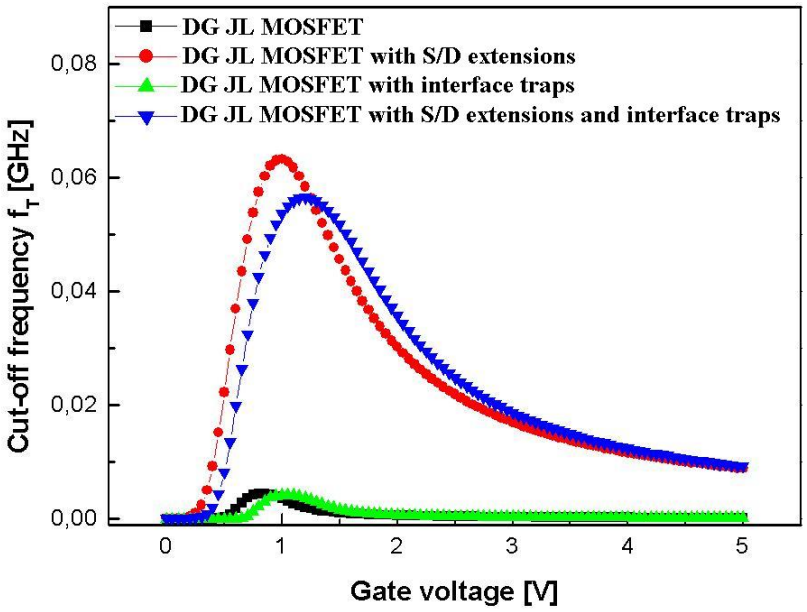

FIGURE 4.20 Variation of the cut-off frequency f_T versus the applied gate voltage V_{gs}.

considered devices, the maximal value of performance parameters with respect to the applied gate voltage is selected, where the ON-current is given in mA/μm, the transconductance in mS/μm and the cut-off frequency in GHz. The intrinsic gain is calculated from the transconductance using the formula :

$$|A_v| = g_m r_0 \tag{4.3}$$

where r_0 is the load resistance fixed to a value of 1 kΩ. On this table, it is shown that the device with source/drain extensions expresses higher performance in terms of analog

criteria and such superior performance remains valid even with the consideration of an interface trap density near the drain side.

TABLE 4.6 Comparison of analog performance parameters of the conventional and the proposed DG JL MOSFET devices.

Design	I_{ON}	g_m	A_v	f_T
DG JL MOSFET	1.5	4.5	4.5	0.12
DG JL MOSFET with S/D extensions	112.3	63.2	63.5	1.55
DG JL MOSFET with interface traps	1.3	4.2	4.2	0.1
DG JL MOSFET with S/D extensions and interface traps	93.6	56.5	56.5	1.4

Despite that the conventional MOSFET devices show higher linear property in comparison to advanced bipolar devices because of the absence of exponential terms in I-V characteristics. However, this statement is subject to doubt in deep nanoscale MOSFETs having innovative gate geometries. For a reliable evaluation of RF linearity performance of the four types of DG JL MOSFET, criteria deduced from higher derivatives of the transconductance are more appropriate since they reflect well the losses of useful output power [Gautam et al., 2012]. The Second and the Third Order Voltage Intercept Points (denoted by VIP2 and VIP3, respectively) express the extrapolated amplitudes of the gate voltage at which the second and the third harmonic become equal to the principal tone in the drain current of the device under study [Biswas et al., 2015]. In this work, these two linearity metrics are given by :

$$\text{VIP2} = 4\frac{g_{m1}}{g_{m2}} \tag{4.4}$$

and

$$\text{VIP3} = \sqrt{24\frac{g_{m1}}{g_{m3}}} \tag{4.5}$$

where g_{mi} denotes the i^{th} derivative of the current with respect to the gate voltage.

The obtained results related to linearity metrics are given in Table 4.7. Regarding the parameter VIP2, we notice that despite the occurrence of the first peaks associated to the fresh and damaged DG JL MOSFET devices with highly doped source/drain extensions at gate voltages higher than those associated to the conventional DG JL MOSFET devices but their magnitude outperforms significantly the later with a ratio of approximately 3 :1 for both cases. For the second parameter VIP3, fresh and damaged junctionless designs having highly doped source/drain extensions reach higher peak values at earlier applied gate voltages in comparison to the conventional junctionless designs.

TABLE 4.7 Comparison of analog performance parameters of the conventional and the proposed DG JL MOSFET devices.

Design	VIP2	V_{gs}	VIP3	V_{gs}
DG JL MOSFET	6.89	0.8	0.99	0.6
DG JL MOSFET with S/D extensions	21.42	0.95	1.74	0.5
DG JL MOSFET with interface traps	12.31	1	0.64	0.75
DG JL MOSFET with S/D extensions and interface traps	34.79	1.15	2.15	0.6

4.6 Impact of the source/drain extensions on the nanoscale DG JL MOSFETs including interface traps in terms of analog/RF criteria

Despite that several experimental and numerical studies published recently have shown the superior fabrication process properties provided by the junctionless device in comparison with the conventional one ([Cerdeira et al., 2013], [Cerdeira et al., 2014] and [Paz et al., 2015]), the junctionless structure is more vulnerable to the impact of the source/drain parasitic resistance compared to an inversion mode structure. The high series resistance associated to the source and drain regions can arise as a serious problem when dealing with uniformly doped channel, which leads to the degradation of the device performance. Moreover, a dramatic increase in the series resistance is obtained when lowering the temperature which makes it a difficult problem to tackle even at room temperature ([Doria et al., 2012] and [Holtij et al., 2014]). Therefore, in order to obtain a global view of DG JL MOSFET performance under critical conditions, new designs and models of nanoscale DG JL MOSFET including analog and RF performances are important for the comprehension of the fundamentals of such device characteristics. In this context, several models and design methodologies have been proposed to improve the nanoscale DG JL MOSFET performances. However, till now, no design approach based on the source and drain extensions is proposed to improve the electrical performance for RF and analog applications. Therefore, it is required to develop accurate numerical models, which are able to describe the device behavior in nanoscale regime. In addition, it is valuable to analyze how the analog and RF performances in the investigated nanoscale DG JL MOSFET is affected by the source and drain extensions under different electrical conditions.

The investigation of the nanoscale DG JL MOSFET behavior through device fabrication or measurement through experimental testing is found very expensive and time consuming. Because of all these limitations, the analysis of the device performance is determined by using device simulation. In this context, our main objective in this section consists in the numerical investigation using ATLAS 2D simulator to study the impact of introducing high doped source/drain extensions basically on the analog performance of the DG JL MOSFET device. The obtained numerical results are compared with respect to the conventional case in order to highlight any observed improvement.

4.6.1 Numerical simulations

In what follows, we present the DG JL MOSFET under study and describe the simulation framework used in the numerical modeling of the analog/RF behavior properties. Both schemas of DG JL MOSFET devices in the cases with and without high doped

source/drain extensions are elucidated in Figure 4.21. For the conventional DG JL MOSFET, the channel body and the source/drain extensions are uniform lightly doped regions, which can be noted by $n^+/n^+/n^+$. Whereas for the proposed structure, the source/drain extensions are both highly doped regions compared to that of the channel body. So, the doping concentration distribution is given by $n^{++}/n^+/n^{++}$. It should be noted that the channel length (L) is varied during simulations from 20 nm to 80 nm in order to analyze the scaling capability of the studied performance criteria.

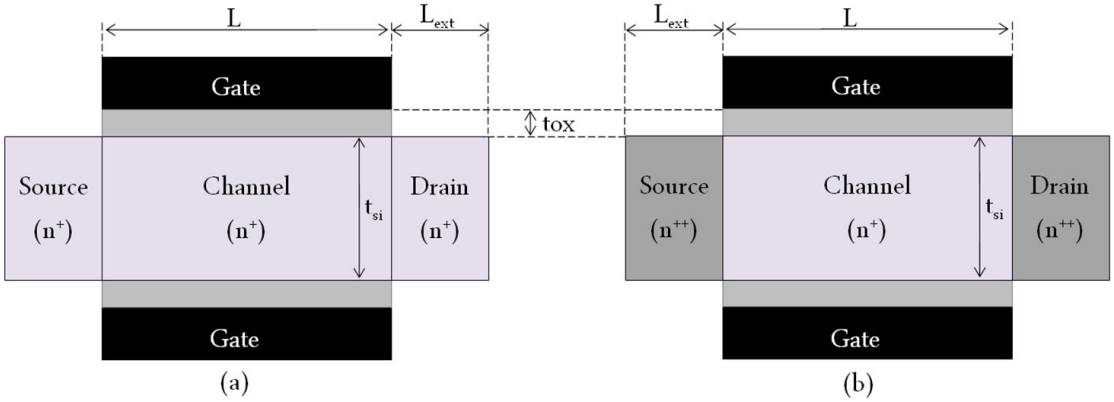

FIGURE 4.21 Cross sectional view of (a) Conventional DG JL MOSFET device (b) DG JL MOSFET with high doped source/drain extensions (t_{si} =5 nm, t_{ox} =5 nm, L_{ext}= 10 nm, $N_{ch} = 10^{18}$ cm^{-3} and $N_{ext} = 10^{20}$ cm^{-3}).

The electrical behavior of these devices is handled using the ATLAS 2D simulator with appropriate tuning, where all simulations are carried out at room temperature (300 K) [Silvaco, 2012]. To enrich the numerical model in order to be very close to realistic behavior, additional effects are included within the models section. The generation/recombination effects are supported through Shockely-Read-Hall (SRH), which permits to take into account different phenomena such as the leakage currents due to thermal generation. The Lombardi CVT model is employed to express the transverse field, doping dependent and temperature dependent components of the mobility using Matthiessen's rule. Since the quantum confinement of carriers at nanoscale level cannot be neglected, the expressions for electron/holes need to be modified according to the density gradient model given for the electrons by :

$$J_n = qD_n\nabla n - qn\mu_n\nabla(\Psi-\xi) - \mu_n n(kT_L\nabla(\ln(n_{ie}))) \tag{4.6}$$

and for the holes :

$$J_p = -qD_p\nabla p - qp\mu_p\nabla(\Psi-\xi) + \mu_p n(kT_L\nabla(\ln(n_{ie}))) \tag{4.7}$$

where n_{ie} is the effective intrinsic concentration and T_L is the lattice temperature. The parameter ξ denotes the quantum correction potential and is defined as :

$$\xi = -\frac{\eta h^2}{12m}\left[\nabla^2\log(c) + \frac{1}{2}\left(\nabla\log(c)\right)^2\right] \tag{4.8}$$

with η a fitting factor, m the carrier effective mass and c the electron (n) or hole (p)

concentration as appropriate. A set of partial differential equations for the modeled device is solved self-consistently on the discrete mesh in an iterative fashion. The currents, voltages and charges for each electrode are calculated after each step of bias through quasi-stationary ramp. All the structure junctions, in the conventional design, are assumed as abrupt, the generation of smooth mesh is done in the simulation.

It should be noted that some parameters are varied in order to get a global view about the behavior when devices are scaled down to nanoscale level. For instance, the channel length (L) is varied from 20 nm to 80 nm with a 10 nm step. Regarding the applied voltages, the gate voltage (V_{gs}) is varied from 0 V to 5 V with a step equals to 0.05 V and the drain voltage (V_{ds}) is varied from 0 V to 2 V with a step of 0.1 V. During the computation of the values of the transconductance based parameters, the drain and gate voltages are maintained to fixed values 0.8 V and 1 V, respectively.

4.6.2 Results and discussion

The quantification of the impact of the source/drain extensions on the analog performance of the device is achieved through the sensitivity analysis of some basic parameters. In what follows, we investigate the alteration of the surface potential, drain current characteristics and the transconductance as a function of some electrical and geometrical parameters. Obtained results are compared with those associated to the conventional case. In Figure 4.22, we depict the variation of the surface potential with the position along the channel for both structures. It can be easily observed that the DG JL MOSFET including highly doped extensions is more stable against the variation of the drain voltage, where the surface potential values inside the extension regions remain approximately constant. It is also observed that the shape of the curve changes above the threshold voltage causing the disappearance of the minimum surface potential at the middle of the channel.

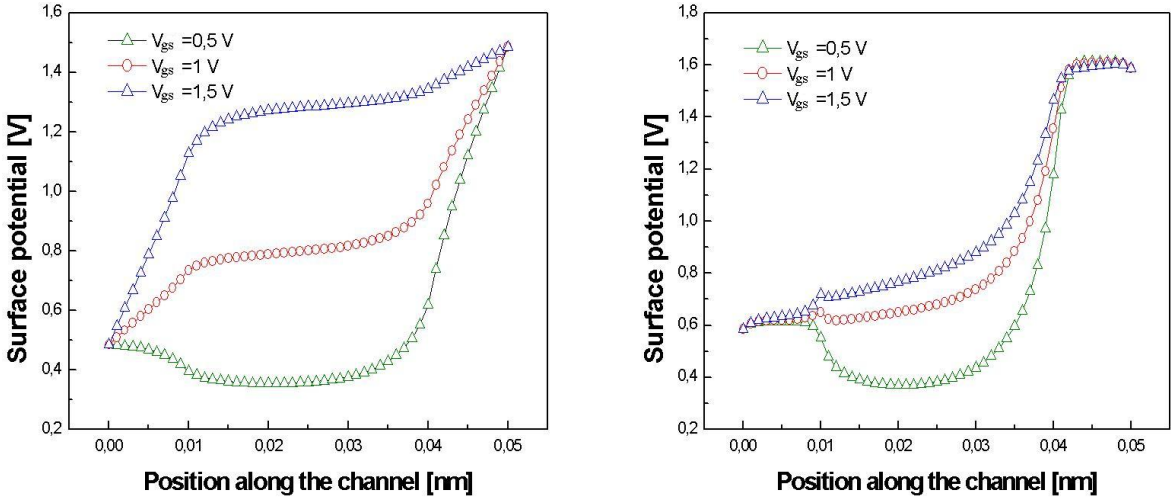

FIGURE 4.22 Surface potential for a drain voltage V_{ds}= 1 V in the case of (a) DG JL MOSFET device (b) DG JL MOSFET with high doped extensions (t_{si} =5 nm, t_{ox} =5 nm, L_{ext}= 10 nm, $N_{ch} = 10^{18}$ cm^{-3} and $N_{ext} = 10^{20}$ cm^{-3}).

In order to assess the scaling capability of the investigated devices, the ON-current obtained at V_{gs}=0.8 V for different channel lengths is represented in Figure 4.23. The current is significantly higher when doped extensions are included, where the drain current has a rapid monotone decreasing tendency for DG JL MOSFET with high doped extensions compared to the conventional device having a slight degradation. The increment in the

ON-current is extremely important in order to obtain more efficiency with the same transistor dimensions at nanoscale regime. For DG JL MOSFET with high doped extensions, the increase in ON-current is due to the highly doped regions and consequently the series resistance can be neglected in comparison to that of the conventional structure. The drain current I_{ds} as a function of the applied drain voltage V_{ds} is also plotted in Figure 4.23, where the high doped extensions lead to an increase in the current magnitude estimated by a ratio of about 230% in the saturation regime. However, the conventional device reaches the saturation regime earlier than the case with high doped extensions, which is mainly caused by the surface potential profile shift in the conventional design as indicated in Figure 4.22. The improvement in the drain current can be explained by the decreasing in the series resistance caused by the highly doped extension regions. The transfer characteristics $I_{ds} - V_{gs}$ for different devices can be extracted using the output curves $I_{ds} - V_{ds}$ as it is shown in Figure 4.23.

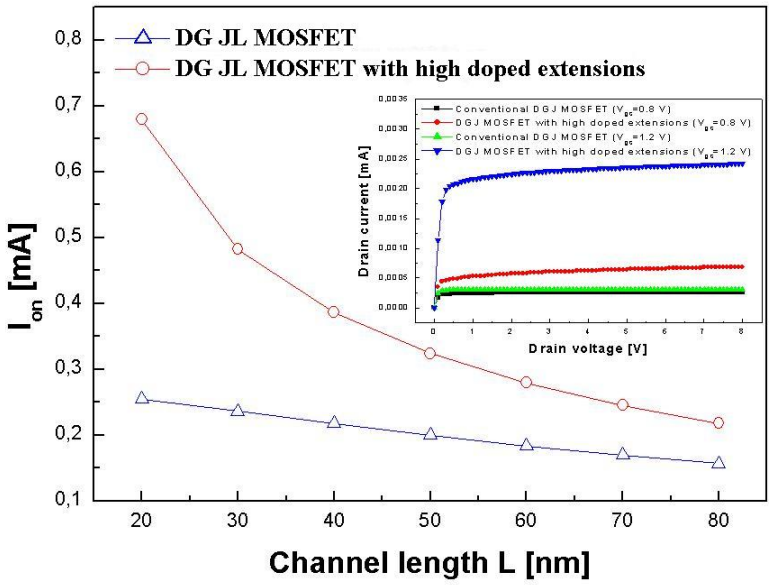

FIGURE 4.23 Variation of the ON-current as a function of the channel length.

Since the transconductance is a critical parameter in building efficient RF and analog-based applications, we study the scaling process of this criterion as a function of the channel length as provided in Figure 4.24. From this figure, it is clear that the design with high doped extensions provides larger values in comparison with the conventional one. This is mainly due to the improved surface potential profile recorded in the case of our proposed design. This improvement can lead to the boost of the transport mechanism in the channel.

The cut-off frequency can be interpreted as the frequency for which the current gain equals to unity [Sarkar et al., 2012]. Such parameter can be computed based on transconductance, gate to source capacitance and gate to drain capacitance using the following expression :

$$f_T = \frac{g_m}{2\pi C_{gs}\sqrt{1+2\frac{C_{gd}}{C_{gs}}}} \tag{4.9}$$

In Figure 4.24, we depict the variation of the cut-off frequency for both structures as a function of the channel length. This criterion has a decreasing tendency with respect

to the channel length, where the DG JL MOSFET with high doping extensions exhibits the highest performance. Moreover, the proposed design provides very high cut-off frequency values when the channel length decreased to the nanoscale domain. This result makes the proposed design a very attractive solution for high frequency analog applications.

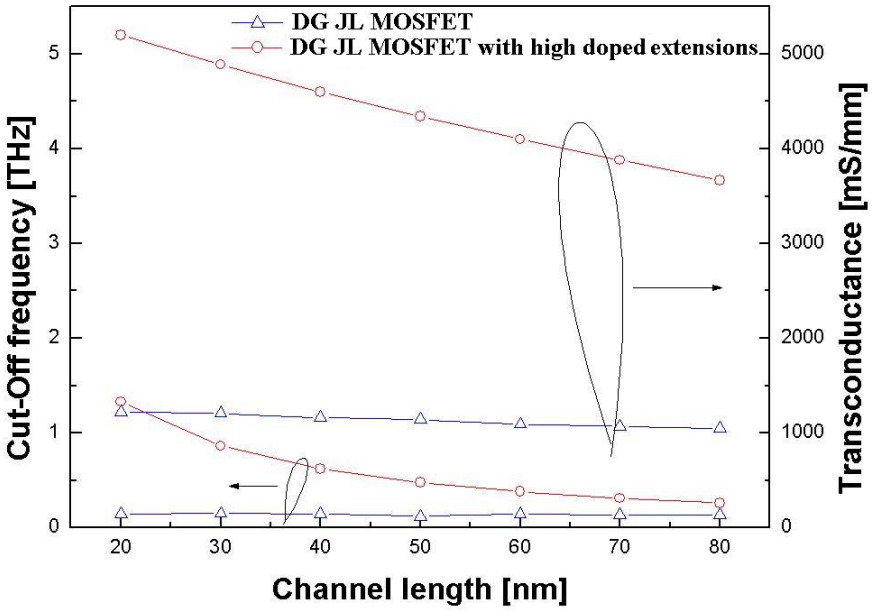

FIGURE 4.24 Variation of the cut-off frequency and transconductance as a function of the channel length.

Most of the useful output signal is wasted if the transistor shows non-linear behavior. The theoretical foundation of linearity analysis is based on Taylor series development of the drain current as a function of the applied voltages. If we consider only the gate voltage, the drain current can be written as :

$$I_{ds} = I_0 + \sum_{i=1}^{n} g_{mi} V_{gs}^{i} \tag{4.10}$$

where I_0 is the current at DC operation point and $g_{mi} = \frac{\partial^i I_{ds}}{\partial V_{gs}^i}$. In practice, losses of useful output power can be generated in the presence of non-linearity in the behavior of MOSFET devices. In addition, higher transconductance derivatives can lead to intermodulation for RF applications. To evaluate the immunity of both devices under study, we extract the values of g_{m1}, g_{m2} and g_{m3} at a fixed value of I_{ds} =0.3mA (R_s can be computed using the surface potential profile for both structures), then we compute the following metrics [Gautam et al., 2012] :

$$VIP2 = 4\frac{g_{m1}}{g_{m2}} \tag{4.11}$$

$$VIP3 = \sqrt{24\frac{g_{m1}}{g_{m3}}} \tag{4.12}$$

and

$$IIP3 = \frac{2}{3}\frac{g_{m1}}{g_{m2}R_s} \tag{4.13}$$

These Figures of merit are used to estimate the distortion characteristics from DC analysis. VIP2 represents the input voltage at which the first and second harmonic voltages are equal. Similarly, VIP3 can be interpreted as the input voltage at which the first and the third order harmonic voltages are the same. IIP3 serves as the input power at which first and third order harmonic powers are equal. VIP2, VIP3 and IIP3 should be as high as possible for better RF and analog performances [Mohapatra et al., 2014]. A summary of the obtained results is given in Table 4.8.

TABLE 4.8 Comparison of linearity metric parameters of the conventional and the proposed DG JL MOSFET devices.

Design	VIP2	VIP3	IIP3
Conventional DG JL MOSFET	0.8624	0.8899	0.0029
DG JL MOSFET with high doped extensions	1.5080	1.3612	0.0500

From the obtained results, we deduce that the consideration of source/drain extensions with high doping concentrations in the context of junctionless devices leads to higher analog performance and increases the scalability of devices to nanoscale level. Moreover, the proposed design predicts significant high linearity metric values, which further enhances the signal distortion.

4.7 An improved graded channel doping based design for nanoscale DG JL MOSFETs including interface traps

The junctionless device is characterized by a different transport mechanism in the sense that the current flow is along the entire volume of the channel in contrast to an inversion mode DG MOSFET for which the current is flowing at the Si/SiO_2 interface. A significant amount of the published literature has elucidated the superior performance of the junctionless designs with respect to their conventional counterparts in terms of many criteria like high I_{ON}/I_{OFF} ratio, low drain induced barrier lowering and high temperature behavior ([Gnani et al., 2011] and [Koukab et al., 2013]). Unfortunately, some hindrances are still persisting while scaling down the device. For example, both subthreshold slope and drain induced barrier lowering may be altered due to the doping density. Furthermore, the random dopant fluctuation can trigger subthreshold behavior degradation especially for high doping levels [Leung and Chui, 2012].

In this section, we showcase the impact of channel engineering aspect on the analog/RF performance of the DG JL MOSFET. The channel engineering technique is expressed through the introduction of a highly doped subregion near the drain side. We conduct numerical simulation experiments to examine the dynamical responses in comparison with the conventional DG JL MOSFET. We adopt ATLAS 2D simulator as a numerical modeling tool to account for various parasitic phenomena related mainly to short channel, quantum mechanical confinement and hot carrier effects.

4.7.1 Numerical simulations

In Figure 4.25(a), we depict the cross sectional view of the reference DG JL MOSFET. It can be easily seen the absence of p-n junctions along the source/channel/drain path, where both source and drain extensions have a highly doped concentration (10^{20} cm^{-3}).

Besides, the channel has a uniform doping concentration (5×10^{19} cm^{-3}). Hence, the three consecutive subregions forming the whole channel region and source/drain extensions can be expressed as $n^{++}/n^{+}/n^{++}$. The channel length and channel thickness are represented by L and t_{si}, respectively. The gate oxide thickness is given by t_{ox} and the work function assigned to both gate electrodes is denoted by ϕ_M. Regarding the proposed nanoscale DG JL MOSFET with channel engineering aspect illustrated in Figure 4.25(b), we consider that the channel is divided into two subregions having distinct doping levels. We assume that the subregion near the drain side is relatively high doped with respect to the subregion near the source side (5×10^{19} cm^{-3} > 5×10^{18} cm^{-3}) and such assumption will be proved numerically later to be the best choice. Such vision permits to obtain a graded doping profile in the lateral direction (along x axis).

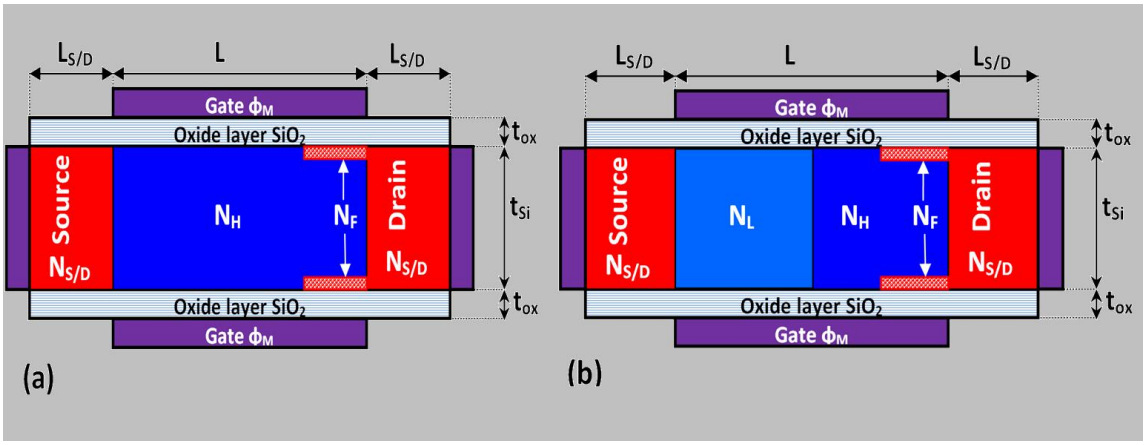

FIGURE 4.25 Cross sectional view of (a) the conventional DG JL MOSFET and (b) the GC DG JL MOSFET.

The set of values affected to different configuration parameters of both devices is summarized in Table 4.9.

TABLE 4.9 Values of the design parameters used in our simulation.

Parameter	**Notation**	**Value**
Channel thickness	t_{si}	10 nm
Oxide thickness	t_{ox}	1 nm
Source side doping	N_L	5×10^{18} cm^{-3}
Drain side doping	N_H	5×10^{19} cm^{-3}
Source/drain extension doping	$N_{S/D}$	10^{20} cm^{-3}
Source/drain extension length	$L_{S/D}$	10 nm
Gate work function	ϕ_M	4.9 eV

For the implementation of the numerical model related to the conventional and Graded Channel Double Gate Junctionless(GC DG JL) MOSFETs, we use ATLAS 2D simulator emerging in the last few years as a powerful computation tool for the modeling of a wide range of electronic components [Silvaco, 2012]. In order to account for various phenomena governing the operation of nanoscale devices, we include some mechanisms in the models section. For the introduction of the mobility degradation of carriers in silicon inversion layer, we employ the semi-empirical CVT model. It is developed on the basis of

Mathiessen rule, where impurity concentration, normal/parallel electric fields and temperature effects are all considered [Lombardi et al., 1988]. For the description of the generation/recombination statistics of carriers among the valence band and the conduction band, we adopt the Shockely-Read-Hall model. In this context, the presence of intermediate trap level within the silicon band gap due to the crystal impurities facilitates the carriers' emission and capture [Goudon et al., 2007]. Since the inclusion of quantum mechanical confinement is indispensable to accurately express the electrical behavior of nanoscale devices, we exploit the advantages of the density gradient model elaborated using the moments of the Wigner function equations of motion. The confinement effects are supported by the quantum correction of the carrier temperatures in the current and energy flux equations [Wettstein et al., 2001]. In particular, such model provides a fitting parameter allowing the calibration of the model with respect to various situations in addition to the stability of its convergence properties. The presence of a trap density at the Si/SiO_2 interface is one of the degradation mechanisms originated mainly from the hot carrier effect [Acovic et al., 1996]. Therefore, it is introduced in this work through the consideration of fixed charges near the drain side along the quarter of the total channel length (L/4). The operation of proposed devices can be modeled through a set of coupled non linear partial differential equations. A discretized version is generated using a mesh of points within the device structure, where an iterative method permits to obtain approximated solutions. In our numerical models, we adopt the Newton method for solving the linearized schema of the whole non linear algebraic system. The Newton method is recommended under several circumstances such as strong coupling between equations, direct current analysis including transient calculations and when frequency based computations of small signal responses are conducted. Moreover, continuity equations of both types of carriers (holes and electrons) are considered in our numerical model [Sarkar, 2013].

4.7.2 Results and discussion

In order to investigate the influence of the graded channel approach on the performance of the junctionless device, sensitivity analysis tests based on some fundamental parameters are realized. Hence, we study in what follows the alteration of the device behavior in terms of the surface potential, drain current, transconductance and cut-off frequency versus the variation of the principal electrical and geometrical parameters. It should be also mentioned that in almost simulation experiments, the results obtained for the graded channel device are assessed against their conventional counterpart not only in the fresh case but also when subject to the interface trap degradation.

We introduce in Figure 4.26 the variation of the surface potential as a function of the position along the channel for the graded channel and conventional DG JL MOSFET devices with and without interface traps. Both gate and drain voltages are fixed to a value of 0.5 V. From the generated curves, it can be observed that the introduction of a highly doped region leads to a step up profile in the surface potential. Since the surface potential slope near the drain side for the GC DG JL MOSFET is smaller than the slope of the conventional junctionless design, this implies the reduction of the electric field leading to improved RF/analog performance [Sarkar et al., 2012]. In addition, it is worthy to mention that the presence of the interface traps increases the slope of the surface potential near the drain side in comparison to the fresh case. Therefore, we can deduce that the graded channel structure exhibits higher reliability performance against the interface trap degradation effect.

Figure 4.27 provides the plot of the drain current for different applied gate voltages

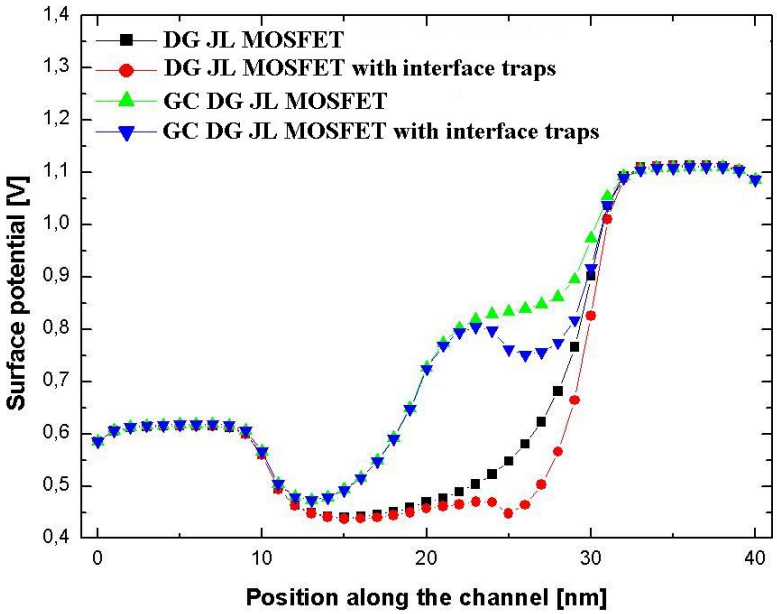

FIGURE 4.26 Variation of the surface potential profile versus the position along the channel for both considered devices with and without interface traps (V_{gs}=0.5 V, V_{ds}=0.5V, $N_F = 5 \times 10^{12}$ cm^{-2} and L= 20 nm).

with the drain voltage equals to 0.5 V. The variation of the drain current for both junctionless devices shows the following features. Firstly, the drain current magnitude in the case of the junctionless device including the channel engineering aspect is higher in comparison to the conventional junctionless device even with the consideration of the interface traps. Secondly, the GC DG JL MOSFET demonstrates superior immunity against the interface trap degradation effect, which is the reverse for the conventional device for which a significant alteration in the current is obtained.

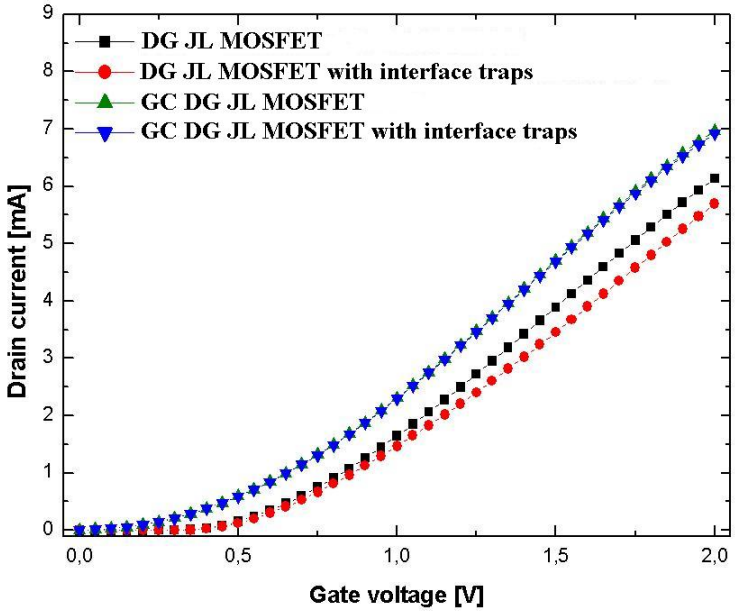

FIGURE 4.27 Variation of the drain current as a function of the gate voltage for both considered devices with and without interface traps (V_{ds}=0.5V, $N_F = 5 \times 10^{12}$ cm^{-2} and L= 20 nm).

In order to justify our choice regarding the position of the highly doped subregion in the channel, we fix the doping of the subregion near the drain side to a value of 5×10^{19} cm^{-3} and we vary the doping near the source side from 5×10^{16} cm^{-3} to 5×10^{20} cm^{-3}. As can be noticed in Figure 4.28, the degradation of the current for doping levels N_L less than or equal to 5×10^{18} cm^{-3} is insignificant and may be neglected. Whereas for doping levels of N_L strictly higher than 5×10^{18} cm^{-3}, the decrease in the drain current is pronounced. Such sensitivity analysis may provide a solid argument in favor to the inclusion of a highly doped subregion near the drain side in the channel in order to boost the performance of the conventional junctionless device at nanoscale level.

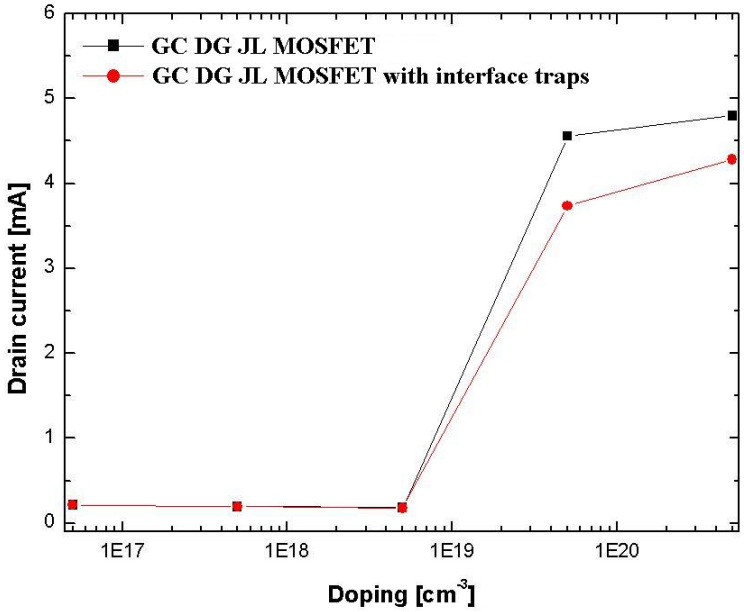

FIGURE 4.28 Variation of the drain current as a function of the doping concentration near the source side for the proposed device with and without interface traps (V_{gs}=2 V, V_{ds}=0.5V, $N_F = 5 \times 10^{12}$ cm^{-2} and L= 20 nm).

The immunity of the GC DG JL MOSFET against the interface trap effect is evaluated in terms of the transconductance and cut-off frequency with respect to the conventional junctionless device. The expressions of the transconductance and cut-off frequency parameters are given below [Pradhan et al., 2014] :

$$g_m = \frac{\partial I_{ds}(V_{ds}, V_{gs})}{V_{gs}} \tag{4.14}$$

$$f_T = \frac{g_m}{2\pi C_{gs}\sqrt{1 + 2\left(C_{gd}/C_{gs}\right)}} \tag{4.15}$$

The maximum transconductance, obtained in the range of the gate voltage [0 V : 2 V], of both devices is plotted as a function of the interface trap density in Figure 4.29 with a drain voltage of 0.5 V. The use of the graded channel paradigm enhances the transconductance at various levels of interface traps. Despite that the conventional junctionless device degrades drastically for an interface trap density more than 10^{12} cm^{-2}, it is interesting to note that the maximum transconductance of the proposed design remains roughly unaffected by the increase of the trap density. This behavior may be attributed to the step

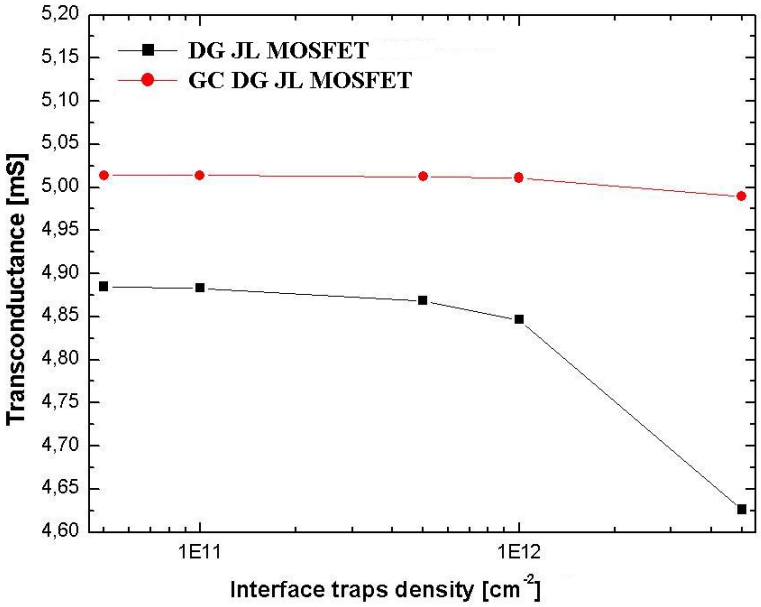

FIGURE 4.29 Variation of the maximum transconductance as a function of the interface trap density for both considered devices (V_{ds}=0.5V and L= 20 nm).

change in the surface potential which guarantees the efficient screening of the channel region from the interface trap influence.

A similar behavior can be recorded for the cut-off frequency parameter deduced from the maximum transconductance (Figure 4.30), where the latter is higher for the graded channel device expressing superior gate controllability in comparison to the conventional junctionless device. This may be due to the high transconductance and low parasitic gate capacitances obtained by the graded channel. Besides, the immunity of GC DG JL MOSFET is relatively high with respect to the conventional device for which the degradation of the cut-off frequency is more important.

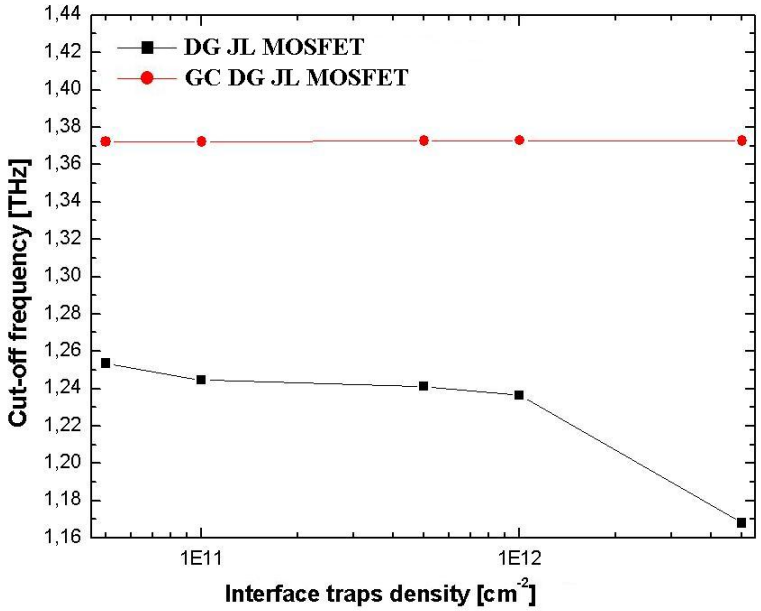

FIGURE 4.30 Variation of the cut-off frequency as a function of the interface trap density for both considered devices (V_{ds}=0.5V and L= 20 nm).

The evolution of the cut-off frequency parameter, associated to the maximum transconductance, as a function of the channel length is highlighted in Figure 4.31. It is clearly shown that the cut-off frequency decreases smoothly with the channel length values. Curves associated with the GC DG JL MOSFET are overlaid which expresses the good immunity of the device against the interface trap degradation. Whereas for the conventional junctionless design, a significant decrease of the cut-off frequency is detected and such discrepancy tends to decrease with the channel length values. The latter remark can be explained by the fact that the short channel effects are more pronounced at small values of the channel length.

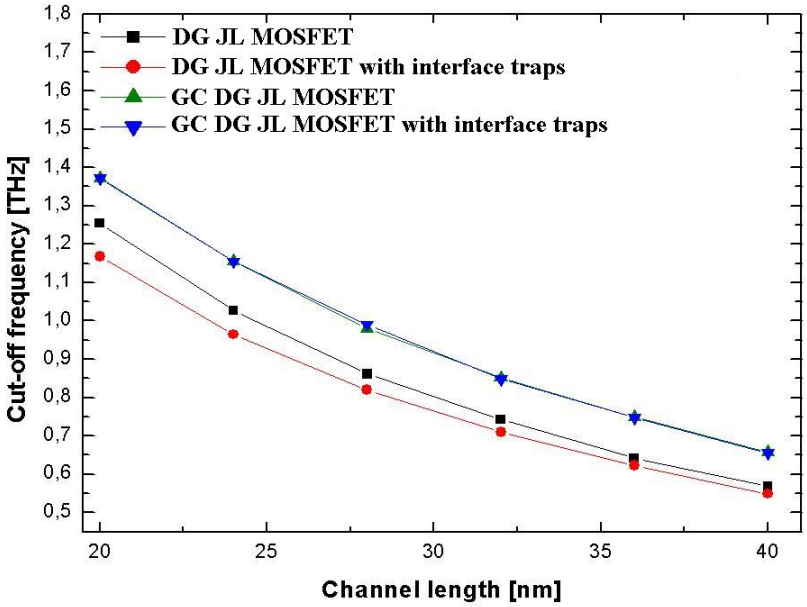

FIGURE 4.31 Variation of the cut-off frequency as a function of the channel length for both considered devices with and without traps (V_{ds}=0.5V and $N_F = 5 \times 10^{12}$ cm^{-2}).

4.8 Conclusion

In conclusion, we have centered our efforts in this chapter on the thorough investigation of some engineering approaches with the aim of bringing additional benefits at the level of both performance measures and operating lifetime. In this extent, we have included the channel, the gate and the source/drain engineering techniques to significantly boost the response of the scaled DG MOSFET device, where the extensive numerical experiments have demonstrated that it exhibits superior performance in comparison to the conventional structures. The performance assessment has been conducted in terms of distinct aspects including the adaptability of the design for digital and analog/RF applications jointly with its immunity against the impact of interface traps. The realized simulations have revealed the paramount importance of TCAD tools not only for offering optimized components but also for opening up the door towards well adapted solutions with reduced costs.

4.9 References

Acovic, A., La Rosa, G., and Sun, Y. C. (1996). A review of hot-carrier degradation mechanisms in MOSFETs. *Microelectronics Reliability*, 36(7-8) :845–869. 133

Arora, N. (2007). *MOSFET Modeling for VLSI Simulation Theory and Practice*. World Scientific Press. 116

Bentrcia, T., Djeffal, F., and Benhaya, A. (2012). Continuous analytic I-V model for GS DG MOSFETs including hot-carrier degradation effects. *Journal of Semiconductors*, 33(01) :014001(1–6). 109

Biswas, K., Sarkar, A., and Sarkar, C. K. (2015). Impact of barrier thickness on analog, RF and linearity performance of nanoscale DG heterostructure MOSFET. *Superlattices and Microstructures*, 86 :95–104. 125

Cerdeira, A., Ávila, F., Iñíguez, B., de Souza, M., Pavanello, M. A., and Estrada, M. (2014). Compact core model for symmetric double-gate junctionless transistors. *Solid-State Electronics*, 94 :91–97. 126

Cerdeira, A., Estrada, M., Iñíguez, B., Trevisoli, R. D., Doria, R. T., de Souza, M., and Pavanello, M. A. (2013). Charge-based continuous model for long-channel symmetric double-gate junctionless transistors. *Solid-State Electronics*, 85 :59–63. 126

Charmi, M., Orouji, A. A., and Mashayekhi, H. R. (2013). Design considerations of underlapped source/drain regions with the gaussian doping profile in nano-double-gate MOSFETs : A quantum simulation. *Materials Science in Semiconductor Processing*, 16(02) :311–317. 120

Chebaki, E., Djeffal, F., and Bentrcia, T. (2012). Two-dimensional numerical analysis of nanoscale junctionless and conventional double gate MOSFETs including the effect of interfacial traps. *Physica Status Solidi (c) - Current Topics in Solid State Physics*, 09(10-11) :2041–2044. 110

Colinge, J. P., Kranti, A., Yan, R., Lee, C. W., Ferain, I., Yu, R., Dehdashti Akhavan, N., and Razavi, P. (2011). Junctionless nanowire transistor (JNT) : Properties and design guidelines. *Solid-State Electronics*, 65-66 :33–37. 120

Djeffal, F., Ferhati, H., and Bentrcia, T. (2016). Improved analog and RF performances of gate-all-around junctionless MOSFET with drain and source extensions. *Superlattices and Microstructures*, 90 :132–140. 123

Djeffal, F., Lakhdar, N., and Yousfi, A. (2011). An optimized design of 10-nm-scale dual-material surrounded gate MOSFETs for digital circuit applications. *Physica E : Low-dimensional Systems and Nanostructures*, 44(01) :339–344. 110

Dollfus, P., Galdin, S., Arbey, M. E., and Hesto, P. (1997). Study of uniform and graded sige channel heterojunction p-MOSFETs using monte carlo simulation. *Thin Solid Films*, 294(1-2) :259–262. 114

Doria, R. T., Trevisoli, R. D., de Souza, M., and Pavanello, M. A. (2012). Impact of the series resistance in the I-V characteristics of junctionless nanowire transistors and its dependence on the temperature. *Journal Integrated Circuits and Systems*, 07(01) :121–129. 126

Fiori, G. and Iannaccone, G. (2003). Atomistic, quantum and ballistic effects in nanoscale MOSFETs. *Journal of Computational Electronics*, 02(02-04) :123–126. 120

Fjeldly, T. A., Kolberg, S., and Iñíguez, B. (2006). Precise 2D compact modeling of nanoscale DG MOSFETs based on conformal mapping techniques. In *The* 9^{th} *annual NSTI Nanotech, Nanotech'2006*, Boston, USA. 115

Fu, Y., Li, Z., Ng, W. T., and Sin, J. K. O. (2014). *Integrated Power Devices and TCAD Simulation (Devices, Circuits, and Systems)*. CRC Press. 108

Gautam, R., Saxena, M., Gupta, R. S., and Gupta, M. (2011). *Channel Material Engineered Nanoscale Cylindrical Surrounding Gate MOSFET with Interface Fixed Charges*. in D. C. Wyld and M. Wozniak and N. Chaki and N. Meghanathan and D. Nagamalai (eds.), Trends in Network and Communications. Communications in Computer and Information Science, Springer Press. 110

Gautam, R., Saxena, M., Gupta, R. S., and Gupta, M. (2012). Effect of localised charges on nanoscale cylindrical surrounding gate MOSFET : Analog performance and linearity analysis. *Microelectronics Reliability*, 52(06) :989–994. 125, 130

Geng, H. (2005). *Semiconductor Manufacturing Handbook*. McGraw-Hill Press. 114

Gnani, E., Gnudi, A., Reggiani, S., and Baccarani, G. (2011). Theory of the junctionless nanowire fet. *IEEE Transactions On Electron Devices*, 58(09) :2903–2910. 131

Goudon, T., Miljanović, V., and Schmeiser, C. (2007). On the shockley-read-hall model : Generation-recombination in semiconductors. *SIAM Journal on Applied Mathematics*, 67(04) :1183–1201. 133

Holtij, T., Graef, M., Kloes, A., and Iñíguez, B. (2014). Modeling and performance study of nanoscale double gate junctionless and inversion mode MOSFETs including carrier quantization effects. *Microelectronics Journal*, 45(09) :1220–1225. 126

Holtij, T., Kloesa, A., and Iñíguez, B. (2015). 3-d compact model for nanoscale junctionless triple-gate nanowire MOSFETs, including simple treatment of quantization effects. *Solid-State Electronics*, 112. 120

https ://nanohub.org (2014). 115

Ismail, M. A., Bakar, M. H. A., Zaini, K. M. M., and Nasir, I. M. (2012). *Refinement of Diffusion Profiles in TCAD for Calibrated and Predictive MOSFET Simulations*. in T. H. Kim, and A. Stoica, and W. C. Fang, and T. Vasilakos, and J. Garcia Villalba, and K. P. Arnett, and M. Khurram Khan, and B. H. Kang (eds.), Computer Applications for Security, Control and System Engineering. Springer Press. 107

Jeon, D. Y., Park, S. J., Mouis, M., Barraud, S., Kim, G. T., and Ghibaudo, G. (2013). Low-temperature electrical characterization of junctionless transistors. *Solid-State Electronics*, 80 :135–141. 120

Koukab, A., Jazaeri, F., and Sallese, J. M. (2013). On performance scaling and speed of junctionless transistors. *Solid-State Electronics*, 79 :18–21. 131

Leung, G. and Chui, C. O. (2012). Variability impact of random dopant fluctuation on nanoscale junctionless finfets. *IEEE Electron Device Letters*, 33(06) :767–769. 131

Lombardi, C., Manzini, S., Saporito, A., and Vanzi, M. (1988). A physically based mobility model for numerical simulation of nonplanar devices. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 07(11) :1164–1171. 133

Maiti, C. K. (2017). *Introducing Technology Computer-Aided Design (TCAD) : Fundamentals, Simulations, and Applications*. Pan Stanford Press. 108

Mohammadi, S. and Afzali-Kusha, A. (2010). Modeling of drain current, capacitance and transconductance in thin film undoped symmetric dg mosfets including quantum effects. *Microelectronics Reliability*, 50(03) :338–345. 120

Mohapatra, S. K., Pradhan, K. P., and Sahu, P. K. (2014). Linearity and analog performance analysis in GSDG-MOSFET with gate and channel engineering. In *Annual IEEE India Conference, INDICON*, Pune, India. 131

Orouji, A. A., Mashayekhi, H. R., and Charmi, M. (2012). Design considerations of source and drain regions in nano double gate MOSFETs. *Materials Science in Semiconductor Processing*, 15(05) :572–577. 120

Pagey, M. P. (2003). *Hot-carrier reliability simulation in aggressively scaled MOS transistors*. PhD thesis, Vanderbilt University. 109

Paz, B. C., Ávila-Herrera, F., Cerdeira, A., and Pavanello, M. A. (2015). Double-gate junctionless transistor model including short-channel effects. *Semiconductor Science and Technology*, 30(05) :055011(1–11). 126

Pradhan, K. P., Mohapatra, S. K., Sahu, P. K., and Behera, D. K. (2014). Impact of high-k gate dielectric on analog and RF performance of nanoscale DG-MOSFET. *Microelectronics Journal*, 45(02) :144–151. 135

Rios, R., Cappellani, A., Armstrong, M., Budrevich, A., Gomez, H., Pai, R., Rahal-Orabi, N., and Kuhn, K. (2011). Comparison of junctionless and conventional trigate transistors with l_g down to 26 nm. *IEEE Electron Device Letters*, 32(09) :1170–1172. 120

Sarkar, A., Das, A. K., De, S., and Sarkar, C. K. (2012). Effect of gate engineering in double-gate MOSFETs for analog/RF applications. *Microelectronics Journal*, 43(11) :873–882. 129, 133

Sarkar, C. K. (2013). *Technology Computer Aided Design : Simulation for VLSI MOSFET*. CRC Press. 133

Silvaco (2012). *Atlas User Manual : Device Simulation Software*. 110, 115, 121, 127, 132

Sze, S. M. and Ng, K. K. (2007). *Physics of Semiconductor Devices*. Wiley Press. 123

Velayudhan, V. (2016). *TCAD study of interface traps-related variability in ultra-scaled MOSFETs*. PhD thesis, Universitat Autònoma de Barcelona. 107

Wang, Y., Hu, H. F., and Cheng, C. (2011). Improved performance of trench power MOSFET with sigec-based channel. *Microelectronics Reliability*, 51(02) :376–380. 114

Wettstein, A., Schenk, A., and Fichtner, W. (2001). Quantum device-simulation with the density-gradient model on unstructured grids. *IEEE Transactions on Electron Devices*, 48(02) :279–284. 133

Wilk, G. D., Wallace, R. M., and Anthony, J. M. (2001). High-κ gate dielectrics : Current status and materials properties considerations. *Journal of Applied Physics*, 89(10) :5243–5275. 111

Wu, N. (2003). TCAD bridging electronic design and manufacturing. *Gartner Dataquest*, pages 1–4. 108

General conclusion

« *Without faith, nothing is possible. With it, nothing is impossible.* »

Mary McLeod Bethune

Summary

Along this dissertation, the conducted bibliographical review relative to the modeling and design at the nanoscale level of the double gate MOSFETs including interface traps has permitted to disclose the significant gap existing between the theoretical development and practical recommendations mainly in terms of realistic operating and structural constraints. In fact, the conventional MOSFETs have been accurately described by easy to use models that are founded under some simplifying and approximation assumptions. Such modeling frameworks have preserved their validity for a long period and have been extensively exploited in simulation tools. However, with the continuous shrinking of electronics components, the elaborated models have been proved to be inefficient in expressing accurately many performance features and the demand for pertinent artifacts becomes obviously an inevitable reality. Subsequently, the context of our dissertation has been focalized on the proposition of powerful yet flexible artifacts that are able to handle sever constraints governing the geometrical and operation features of nowadays double gate MOSFET devices.

The research work performed in this dissertation has allowed from a physical viewpoint not only the assimilation of various mechanisms and phenomena occurring during the operation of the DG MOSFET device, but also the extraction of the main parameters having an impact on the evolution of the device behavior in terms of performance measures. It should be pointed out that our contributions belong to the intersection of two thematic fields. The first one is relative to the elaboration of modeling frameworks devoted to the nanoscale double gate MOSFET subject to both quantum mechanical confinement and presence of interface traps. While the second field concerns implementing optimization strategies and evaluating their performances. In the first field, we have proposed two modeling approaches (neuro-fuzzy and surrogate techniques) and we have considered the channel length and the channel thickness as the independent variables since they possess a strong impact on the short channel, hot carrier and quantum confinement degradation effects. The developed models have been used to express a wide range of performance criteria namely subthreshold, analog and radio frequency measures. In the second field of this research, we have covered the DG MOSFET design aspects in addition to their assessment with respect to the conventional structures. According to such vision, we have suggested two fundamental design schemas. The first schema was based on the adoption of multi-objective evolutionary methods such as NSGA II, which provides a set of configurations to the designer. The second schema was established using an accredited simulation environment (ATLAS simulator) with its rich packages. While guided by the principles of the device operation, the trial and error procedure has been carried out in this case to enhance the device performance. Moreover, we have also consolidated the efficiency of the proposed designs through statistical or quantitative evaluation of the responses associated to these designs.

The investigation of the proposed models and designs has permitted us to gain the following general remarks :

- With the tremendous sophistication of downscaled devices, the balance is oriented towards empirical models due to their accuracy and low elaboration costs in comparison to the resolution of Schrödinger/Poisson equations self-consistently;
- The use of soft computing or surrogate based models jointly with dedicated simulation tools may be of paramount importance for designers, where the included fitting parameters can be efficiently adjusted using numerical data issued by simulators. In addition, such models can be integrated within IC simulation packages;

- Despite that the trial and error procedure using specialized simulation environments is seemingly an easy task, it requires a satisfactory comprehension of MOSFET properties in order to introduce amendments at the correct location and with the correct magnitude;
- Even with the very competitive outcomes of the established modeling and design techniques, the feasibility validation of the resulted geometries via practical experiments cannot be neglected especially with the numerous conditions imposed at the level of technological fabrication processes.

We would like to mention that various ideas implemented in the context of this dissertation can be easily extended to support other types of multi-gate devices (such as the gate all around structure). Furthermore, fabrication constraints can be accounted for during the optimization procedure running with the aim of discarding non feasible configurations.

Suggestions for future work

The current trends in nanoscale devices to push further the miniaturization process has opened up a plethora of intensive research in terms of modeling and design approaches. The modest research work realized in this study can be considered as an initial step towards a more thorough analysis of the following axes :

- Since the channel to gate tunneling current can impose serious drawbacks on the performance of circuits besides the device downscaling, taking into account such parameter will probably be essential especially with oxide thicknesses less than 1.5 nm. The use of novel high-κ dielectrics can bring additional benefits in reducing the magnitude of such leakage current;
- The selection of high-κ material with low interface trap density has been demonstrated to be very challenging. The presence of discrete charges at the interface leads to the generation of threshold voltage fluctuations because of the non-uniform flowing of the current inside the channel. Moreover, some studies have put into evidence the correlation between the interface traps and short channel effects. Hence, dealing with the interface trap density as a homogeneous interface layer near the drain side may be in some cases obsolete and needs a revision in the context of quantum mechanics;
- A significant lack of work related to the lifetime prediction of nanoscale multi-gate MOSFETs can be clearly observed in the available literature. Although we have not understood if such situation is originated due to commercial reasons or the difficulty of combining multiple degradation mechanisms in the same framework, we think that it becomes indispensable to propose reliability enhancements to make these nanoscale devices meet prolonged periods of operation.

Scientific production

Journal papers

- T. Bentrcia, F. Djeffal, E. Chebaki, Approach for designing and modelling of nanoscale DG MOSFET devices using Kriging metamodelling technique, IET Circuits, Devices and Systems, Vol. 11, No. 06, pp. 618-623, 2017.
- T. Bentrcia, F. Djeffal, E. Chebaki, and D. Arar, Impact of the drain and source extensions on nanoscale double-gate junctionless MOSFET analog and RF performances, Materials Science in Semiconductor Processing, Vol. 42, No. 02, pp. 264-267, 2016.
- T. Bentrcia, F. Djeffal and E. Chebaaki, ANFIS-based approach to studying subthreshold behavior including the traps effect for nanoscale thin-film DG MOSFETs, Journal of Semiconductors, Vol. 34, No. 08, pp. 084001(01-09), 2013.

Book chapters

- T. Bentrcia, F. Djeffal, and E. Chebaki, Multi-objective design of nanoscale double gate MOSFET devices using surrogate modeling and global optimization, chapter 04. in A. Tiwari, Y.K. Mishra, H. Kobayashi, and A.P.F. Turner (eds.), Intelligent Nanomaterials, Wiley-Scrivener Press, 2017.
- T. Bentrcia, and F. Djeffal, An ANFIS based approach for prediction of threshold voltage degradation in nanoscale DG MOSFET device, Chapter 25. in G.C. Yang and S.I. Ao and L. Gelman (eds.), IAENG Transactions on Engineering Technologies, Springer Press, 2014.

International conferences

- T. Bentrcia, F. Djeffal, E. Chebaki and D. Arar, A kriging framework for the efficient exploitation of the nanoscale junctionless DG MOSFETs including source/drain extensions and hot carrier effect, The 13^{th} Conference on Nanosciences and Nanotechnologies NN'2016, Thessaloniki, Greece, July 05-08, 2016.
- T. Bentrcia, F. Djeffal and M. Meguellati, Analytical investigation of swing factor for nanoscale Double Gate MOSFET including gate work function effect, World Congress on Engineering WCE'2015, Vol. I, pp. 426-429, London, United Kingdom, July 01-03, 2015.
- T. Bentrcia, F. Djeffal, M. Meguellati and D. Arar, Numerical investigation of nanoscale SiGe DG MOSFET scalability including interfacial trap effects, International Parallel Conferences on Researches in Industrial and Applied Sciences IARIAS'2014, pp. 63-69, Dubai, United Arab Emirates, April 25-26, 2014.

Abstract

During the few past decades, we have seen an accelerated evolution of the semi-conductor industry. Without doubt, the MOSFET structure constitutes the workhorse of the prevailing technological era. Despite the many advantages gained by the continuous scaling of MOSFETs, it has lead to the amplification of parasitic phenomena. Our objective in this dissertation is to propose efficient modeling/design approaches for the nanoscale double gate MOSFET structure with the consideration of the short channel, quantum confinement and hot carrier effects. The appropriateness of introduced designs with respect to some application fields has been investigated. We hope that the findings of this dissertation will contribute to further developments.

Keywords : Double gate MOSFET, Parasitic effect, Numerical simulation, Modeling and Design.

Résumé

Au cours des dernières décennies, nous avons assisté à une évolution accélérée de l'industrie des semi-conducteurs. Sans aucun doute, la structure MOSFET constitue le fil conducteur de l'ère technologique dominante. Malgré les nombreux avantages obtenus par la miniaturisation continue des MOSFETs, cela a conduit à l'amplification des phénomènes parasites. Notre objectif dans cette thèse est de proposer des approches efficaces de modélisation/conception pour la structure nanométrique MOSFET à double grille en tenant compte des effets canal court, confinement quantique et porteurs chauds. La pertinence des conceptions introduites relativement à certains champs d'application a été étudiée. Nous espérons que les résultats de cette thèse contribueront à des futurs développements.

Mots clefs : MOSFET à double grille, Effet parasite, Simulation numérique, Modélisation et Conception.

ملخص

لقد شهدنا خلال العقود القليلة الماضية تطورا متسارعا لصناعة أشباه الموصلات. بدون ريب ان البنية MOSFET تشكل العمود الفقري للعصر التكنولوجي السائد. بالرغم من المزايا العديدة المكتسبة من التصغير المستمر للبنيات MOSFETs، فقد أدى ذلك بالمقابل إلى تضخيم الظواهر الطفيلية .هدفنا في هذه الاطروحة هو اقتراح طرق نمذجة/تصميم فعالة للبنية النانومترية MOSFET ثنائية البوابات مع الاخذ بعين الاعتبار لتأثيرات القناة القصيرة، التقييد الكمي و الشحنات الساخنة. و تم التحقق من مدى ملاءمة التصاميم المقدمة فيما يتعلق ببعض مجالات التطبيق. نأمل أن تسهم نتائج هذه الاطروحة في التطورات المستقبلية.

الكلمات المفتاحية: بنية MOSFET ثنائية البوابات، تأثيرات طفيلية، محاكاة رقمية، نمذجة وتصميم.