POLITECNICO DI TORINO

Corso di Laurea Magistrale
in Nanotechnologies for ICTs

Tesi di Laurea Magistrale

Gate-All-Around FET: analytical compact modeling and TCAD validation for system performance evaluation

Relatori
prof. Marco Vacca
M. Eng. Fabrizio Mo

Candidato
Silvio Gucciardo

Anno Accademico 2022-2023

Summary

Integrated electronics industry, since early 70s, realized how increasing the transistors integration density was the key to push towards the performances in CMOS technology. The notorious Moore's law chronicled, for almost fifty years, the evolution of chip industry. The main strategy to improve switching frequency, power dissipation and area was typically the constant field scaling approach traced out by Dennard et al. at IBM in 1974 which was based on keeping the electric field constant by scaling down device geometry and supply voltage by a constant factor to keep reliability unaltered and improving current capability. In more aggressively scaled nodes the short channel effects (SCE) compromised the feasibility of this approach leading to unbearable subthreshold leakage in planar MOSFETs. Multi-gate transistor technology soon became the most relevant solution to keep pushing the Moore's law and to mitigate SCEs.
Good quality models are then fundamental to exploit these kind of technologies pursuing the ability to predict the I-V and C-V characteristics of multi-gate devices. The objective of this work is to implement an accurate model in MATLAB for cylindrical section gate-all-around MOSFET, avoiding fitting parameters relying on experimental measurements. The focus is put onto the electrostatics inside the channel which give a meaningful insight of the device behavior. This is done starting from the BSIM-CMG model, the industrial compact model standard for multi-gate transistors developed by the *University of California, Berkeley.*
After a first chapter which serves as an introduction, focusing on the main issues plaguing scaled MOSFETs and the possible solutions, the main topics of the thesis are treated in the second and the third chapters, dedicated to the DG-MOSFET and the cylindrical Gate-all-around MOSFET models implementations and validation with TCAD extractions. The last chapter contains the conclusions of the work.
In the second chapter, the industry state-of-the-art double-gate MOSFET (DG-MOSFET) analytical model has been studied, and a MATLAB model developed, solving numerically the Poisson's equation inside the channel for intrinsic and doped substrate using the perturbation approach based on the BSIM model. Accurate electrostatic potential and charge density solution together with short channel effects corrections allowed to extract accurate I-V and C-V curves which allowed to develop a compact model which has been validated with physical-based device simulations (TCAD) in channel pinch-off and velocity saturation conditions at the end of the chapter. The simulations show a good matching even without process-specific fitting parameters up to $Lg = 13$ nm.
In the third chapter of the thesis, the DG-MOSFET model implementation has been used to develop the model of a possible candidate for high-density logic and memory applications: the cylindrical gate-All-Around FET (cy-GAAFET). An analytical model based on the electrostatic study of the channel for intrinsic channel, already present in literature has been implemented in MATLAB solving the Poisson's equation self-consistently with the gate bias equation. The limitation of this model is the assumption of intrinsic substrate.

In this work a proposed solution to overcome it is to employ the perturbation approach, similarly to what has been done for the finFET. The MATLAB implementation of this model is then developed, which is able to extract the potential profile inside the channel for a wide range of operational conditions. The model also takes into account SCEs which are modeled as potential corrections in the *Poisson's equation* so to be seaminglessly included into the model and predict accurately also the current degradation. I-V and C-V curves have been extracted and validated with TCAD physical simulations.
To validation of the MATLAB model is performed with *Sentaurus Device TCAD* for which a cy-GAAFET 3D structure has been developed from the ground-up and from it the extraction of output characteristics and transcharacteristics has been done. The validation shows good matching down to gate length $L_g = 13$ nm , in particular 18 nm shows relative errors of the ON-current lower than 5 % for a wide range of V_{DD}. The cylindrical GAAFET model shows better results compared to the DG-MOSFET's one at the same gate length because of the better electrostatic control the surrounding gate has on the channel, which limits the short channel effects that are more difficult to be fitted without empirical parameters and are the main contributors to the overall error of the models. In any case fitting parameters can be extracted from physical-based simulations to reduce the error. The compact models obtained by the analytical models, are well-suited for system performance evaluation simulators, such as TAMTAMS (previously developed at Politecnico di Torino), for this reason stand-alone PMOS and NMOS MATLAB codes have been developed to provide useful elements for this task.

Contents

Chapter 1

Introduction

1.1 From the origins to the state of the art of integrated electronics

The future of integrated electronics is the future of electronics itself (Moore [1998])

In 1965, Gordon E. Moore, head of R&D and co-founder of Fairchild semiconductors, taking advantage of a privileged point of view on the semiconductors industry, Moore [2006a] was able to deeply grasp the impact semiconductor integrated technology for digital and linear applications would have on the entire industry and society and, also, to extrapolate the trend over the years of the complexity of the integrated circuits corresponding to the minimum cost per component. This trend, that will be notoriously called Moore's law, showed an increment in complexity for minimum component costs of roughly a factor of two per year. It was expected to held up for a couple of years but instead the prediction applied much longer. In 1975 The Moore's law eventually was updated (Moore [2006b]), predicting a doubling of the complexity every two years. The Moore's law chronicled the evolution of integrated circuits since then and became the standard CMOS technology development benchmark, setting up, year by year, the next technological goal, for almost fifty years (Wong [2021]). But why increasing the transistor count inside a chip is so important?

Chip industry since its dawn has pursued the device scaling for achieving density, speed and power improvements at the lowest production cost (*power-performance-area-cost*, PPAC) (Taur and Ning [2021]). To improve PPAC, minimum size of the transistors has historically been reduced following the proposed *constant-field scaling* approach by Dennard et al. at IBM in 1974 (Dennard et al. [1974]). In this way all the device voltage and geometrical quantities would have scaled down by a constant scaling factor k in order to maintain the electric fields constant so that the reliability of the device remained unchanged. Also, keeping the electric field constant will reduce the mobility degradation and impact ionization effects.

In order to keep up the Moore's law pace, the scaling constant has been reduced by

approximately $\times 0.7$, that is $k \sim 1.4$, every technological node (Bohr and Young [2017]). Increasing integration density, the number of dies per wafer that can be fabricated will increase, reducing recurrent costs. Furthermore, all the capacitances (proportional to area and inversely proportional to thickness) are scaled by k implying circuit delay times which scale as $1/k$. Also, power dissipation per circuit will be aggressively scaled as $1/k^2$ because of the simultaneous scaling of drift current and voltage (Taur and Ning [2021]).
This kind of approach assumed the threshold voltage to scale down with k. This would improve active power and speed but unfortunately would also worsen subthreshold current, which was of minor concern in the early days but became more and more important when the size of the minimum transistor approached the ultimate scaling limit. Scaling of the threshold voltage became increasingly more difficult every generation due to subthreshold leakage (Bohr [2007]). This happens because of the factors which do not scale with k. The thermal voltage kT/q and the silicon bandgap E_g cannot scale with the dimension of the MOSFET. Nonscaling factors are the main contributors to the so-called *Short Channel effects* (SCE).
Depletion regions at the drain and source junctions become more and more comparable in size with the length of the channel when this is scaled down. The strong electric field inside of the space charge regions will then affect the potential barrier of the channel which will be then reduced and the channel conductivity ends to be controlled more and more by the drain voltage rather than by the gate voltage, the threshold voltage V_{th} is reduced by the effect of the drain-channel junction (Taur and Ning [2021]).

To limit the influence of the drain and source junction on the channel potential, it is convenient to reduce in size their depletion layers by increasing doping concentration in the body while the device is scaled. But this strategy incurs in the worsening of the free carrier mobility and also direct band-to-band tunneling in source and drain junction due to the increasing of the maximum electric field. (Bohr [2007])

Eventually the control of the channel by the gate contact will be lost unless the thickness of the dielectric is reduced. This has been feasible employing SiO_2 for over thirty years, scaling the gate oxide thickness linearly with the channel length; until gate-to-channel tunneling leakage started to impair the standby power consumption, one of the biggest prides of the CMOS technology. Despite technologically manufacturable, when the SiO_2 gate oxide thickness allowed a leakage gate current comparable to the subthreshold drain-to-source current, candidates to substitute silicon oxide but with higher dielectric constant have been proposed ending up to the introduction of HfO_2 in 2007 (Thompson [1998],Hu [2010]).
The key technology enabling the born of the first microprocessor, the 4004 by intel in 1972 was the self-aligned poly-silicon gate (Faggin [2021]) showing threshold voltage tunability by doping, high reliability and closer spacing between source and drain. Poly-silicon gate shows a depletion layer which will introduce a series capacitance with the oxide capacitance, leading to a lower inversion charge layer and eventually the reduction of the transconductance of the MOSFET (Taur and Ning [2021]).
Constant-field scaling applied very well up to the early 2000s with the 130 nm node (Bohr and Young [2017]), after that, SCE kicked in to such an extent that traditional scaling

techniques had to be replaced by new approaches based on innovation in material engineering and relying on design gimmicks which have characterized chip design at the device level in the 21th century.

The most notable innovations are

- Strained silicon to improve carrier mobility at 90 nm node (Thompson et al. [2002])
- high-k dielectric and metal gate at the 45 nm node, (Jan et al. [2008])
- 3D FinFET at the 22 nm node, (Jan et al. [2012])

In the 2 nm node, *lateral-gate-all-around-MOSFET, LGAA* are expected to reach mass production in 2025 (IRD [2022])

1.2 MOSFETs Scaling Limit

Conventional planar MOSFETs scaling stopped to be convenient when the physical constraints did not allow further improvements without worsening too much subthreshold slope, so stand-by power. In this section an intuitive explanation of the physics behind the short channel effects is given.

1.2.1 Drain Induced Barrier Lowering

The fundamental working principle of the MOSFET is to control the electrostatic potential of the channel with respect to the source by modulating the gate electrode, with a positive potential for the NMOS and negative for a PMOS. This kind of coupling happens by the polarization of a capacitance C_g which, although not constant (Taur [2021]) when V_g is above the threshold voltage V_{th} typically saturates to $C_{ox} = \frac{\varepsilon_{ox}}{t_{ox}}$.

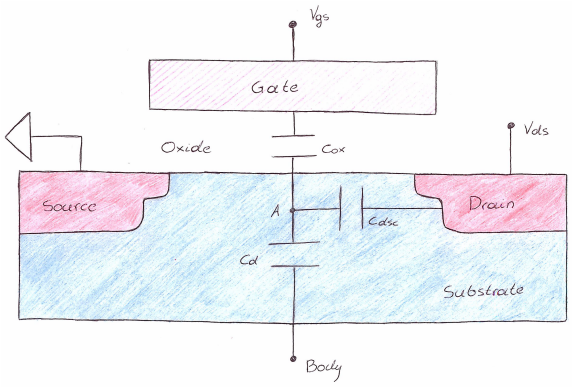


Figure 1.1: Effective capacitances inside the MOSFET affect its performances.

In the real device, channel potential is not affected only by the gate capacitance but also by the capacitance between the drain and the channel, C_{scd}.
As illustrated in figure 1.1, if we call A the mid-point of the channel, while the length of the gate L_g is reduced (and so the length of the channel), the capacitance between the the point A and the drain is going to increase.
Progressively C_{dsc} is going be comparable with the *depletion charge capacitance* C_d in the silicon and, for sufficiently short channels, the capacitive voltage divider causing the drain voltage to drop partially on C_{scd} and partially C_d, is now loading more significantly C_d, increasing the sensitivity of the channel potential to the V_d. This kind of effect is called *drain-induced barrier lowering (DIBL)* and manifests in a reduction of the threshold voltage as known as *threshold voltage roll-off*. By solving the Poisson's equation, can be shown that the voltage depends on L_g exponentially (Hu [2010]). In extreme cases, e.g. $L_g < 30\,\text{nm}$, the device does not switch off and the current is going to leak in any case, as long a drain voltage is applied.
DIBL set a lowerbound for the channel length. In order to keep this phenomenon under control, doping inside of the body has been increased with scaling, in this way the total surface formed by the space charge region is reduced, and also the absolute value of C_{dsc}.
A variation of the doping concentration N_b causes a variation of V_{th}, unless C_{ox} is increased; this is done by reducing the oxide thickness or employing a dielectric with a higher dielectric constant. Recalling the expression for V_{th} (Sze [2007])

$$V_{th} = V_{fb} + 2\psi_B + \frac{\sqrt{4\epsilon_{si} q N_b \psi_B}}{C_{ox}}, \tag{1.1}$$

we can have the evidence of this. The effort to get shorter devices, translates into the struggle to get thinner oxide at each new technological node (Hu [2010])

1.2.2 Process Variations

The strong sensitivity of V_{th} to L_g has detrimental effects on the circuit performances because of process variation. For long gates, the effect of the statistical distribution of the channel lengths is negligible, but for short devices a strong unwanted distribution in the threshold voltages across the circuit may be present. This may lead to impairment of the system performances, for example timing violations in clocked circuits (Alioto et al. [2015]).
But, maybe of higher concern in short MOSFET is the *random disctrete doping, RDD* becoming more severe in highly doped substrates (Saha [2021]).
It can be understood by thinking that a statistical variation of the number of impurities introduced in a substrate depends mostly on the process rather than the size of its volume, this means that the relative uncertainty of a smaller substrate will be higher than the one of a larger one. This means an higher dispersion of threshold voltages across the circuit.

1.2.3 Subthreshold Swing

Subthreshold current exponentially depends on V_{th}, and unfortunately it increases when the latter decreases.

Threshold voltage reduction has been a key strategy to increase the circuit speed but it decrease the subthreshold slope.
Subthreshold Swing S is used to understand how sharply the inversion region is formed with the control of the gate voltage V_g, and it is inversely proportional to the subthreshold slope.
This is strictly related to the depletion capacitance C_d and C_{ox} (Sze [2007])

$$S \equiv \left(\frac{d[log(I_d)]}{dV_g}\right)^{-1} = (\ln 10)\left(\frac{kT}{q}\right)\left(\frac{dV_g}{d\psi_s}\right) = (\ln 10)\left(\frac{kT}{q}\right)\left(\frac{C_{ox}+C_d}{C_{ox}}\right) \tag{1.2}$$

where ψ_s is the *surface potential*.

Also in this case, we got a problem with a capacitive voltage divider 1.1. To get a steeper transcharacteristic we have to minimize the subthreshold slope.
Two actions are available.

1. Increasing C_{ox} by minimizing the oxide thickness, leading to a certain increasing of gate-leakage, setting a lowerbound for the t_{ox}.

2. Decreasing C_d by reducing the doping concentration.

1.3 Multiple-Gate FET

Despite the efforts to reduce t_{ox}, even with a hypothetical zero-thickness oxide a leakage current will still be present. This happens because of the regions far from the silicon-oxide interface, where the influence of the gate electric field is weak or absent, which constitute a leakage path for electrons. So the potential barrier in these "leaky" regions can be easily controlled by the potential at the drain end (Chauhan [2015]).
To overcome the challenges SCEs had brought in, the complexity of the planar MOSFET increases without precedents. It is outside the scope of this work to discuss all the countermeasures industry has adopted through the years to keep pushing planar MOSFET to its limit. It is sufficient to say that at a certain point researchers understood that to relax SCEs constaints a better electrostatic control of the channel had to be triggered by increasing the number of gates leading to *multiple-gate MOSFET (MuGFET)* (Colinge).
Multiple-gate MOSFETs consist in a certain number of gates around a ultra-thin body which is typically fully depleted. This kind of device can be fabricated both in bulk or silicon-on-insulator (SOI) substrate and show superior performances with respect to the planar MOSFET.
In 2011, the first MuGFET to reach mass production became the triple-gate FinFET from *intel* (Jan et al. [2012]) and this kind of devices is nowadays the best solution for high speed, low power, VLSI ICs (Saha [2021])
Many variants of MuGFET have been proposed through the years (Colinge), whose differences of which are mainly on the number of gates control the potential inside the body. In the following, details about the *Double-Gate FinFET (DG-FinFET)* and the cylindrical *Gate-All-Around MOSFET (GAA-FET)*. The choice of these two structures among the

wide collection of multiple-gate MOSFETs is because they have a symmetrical geometry and, as will be shown, this permits a relatively easy calculation of the potential inside the body, at least in long devices, by solving the *Poisson's equation* with only minor approximations.

1.3.1 Double-Gate FinFET

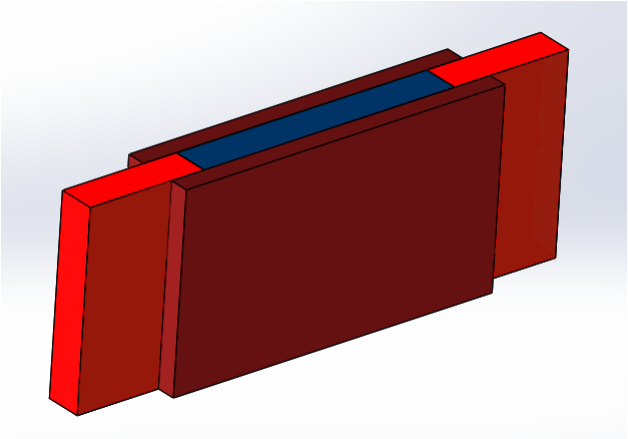


Figure 1.2: DG-FinFET

Double-Gate FinFET was initially proposed in 1996 by the University of California, Berkeley as solution to the SCEs (Chauhan [2015]). This structure comprise two metal gates, that at the same time control the potential inside a silicon thin-film body.
The body of the device is so thin that it is typically fully depleted and the width of the depletion region is controlled by the physical width of the body itself, rather than by the doping as in the planar device.
This is an fundamental aspect because the doping concentration of the channel can be reduced noticeably without worsening the SCEs. In this way, RDD is eliminated, mobility degradation reduced and band-to-band tunneling become less severe (Chauhan [2015]).
However, doping can be still used to finely adjust the threshold voltage.
The main advantage of the DG-FinFET is the ability to suppress the "leaky" part of the silicon body, that is the regions where the electrostatic control of the gate is weak.
This kind of technology can be implemented with the standard CMOS manufacturing processes and it fabrication costs are comparable with planar technology (Chauhan [2015]).
Another important advantage is the absence of *body effect* because of the absence of a body contact. Body effect causes a variation of the threshold voltage and eventually a reduction of the subthreshold slope. The absence of a body contact and the low means that the surface potential ψ_s in the inner region of the device is completely determined by the gate voltage V_g (Taur [2021]), that is

$$\frac{dV_g}{d\psi_s} = \left(\frac{C_{ox} + C_d}{C_{ox}}\right) \simeq 1 \tag{1.3}$$

C_d can get very small with respect to C_{ox} leading to an ideal subthreshold swing

$$S \equiv (\ln 10)\left(\frac{kT}{q}\right)\left(\frac{dV_g}{d\psi_s}\right) \simeq (\ln 10)\left(\frac{kT}{q}\right) = 60\,\text{mV/dec} \tag{1.4}$$

Furthermore, the DIBL is less severe because of the lower capacitance coupling the drain voltage and the body, since the smaller surface of the space charge region at the interface. The most important aspect of FinFET is that, with its 3D nature, introduces a new degree of freedom in controlling short channel effects and subthreshold swing, that is the thickness of the body t_{si}.
These wonderful characteristics and the compatibility with the standard CMOS technology allows this technology to push further the device scaling to improve performances.

1.3.2 Gate-All-Around MOSFET

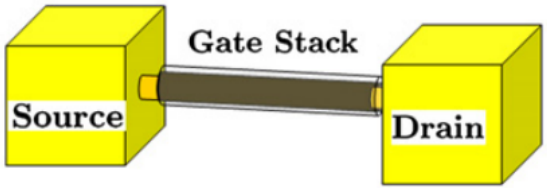


Figure 1.3: Horizontal Nanowire FET

When the gate contact wraps around the channel, the electrostatic gate control is enhanced and further SCEs mitigation allows further transistor miniaturization (Jimenez et al. [2004]). Square, Rectangular, Trapezoidal, Circular, Elliptical, and Triangular sections have been proposed (Kumari et al. [2021]) but only the circular one can be solved analytically for intrinsic channel, bringing to a detailed description of inner potential profile (Taur and Ning [2021]). Non-symmetrical GAAFETs, like the rectangular one need arbitrary assumptions on the potential profile inside the channel in order to simplify the mathematical treatment (Duarte et al. [2013]). This kind of structure is also referred to *nanowire*, and it gained attention because its small footprint and so its possible high density integration, especially useful in DRAM (Venugopalan et al. [2012]). Several type of nanowires exist. In this work the horizontal nanowire will be put under attention, whose structure is represented in Figure 1.3.

1.4 MOSFETs Modeling

In order to get working and predictable design at circuit level, an accurate device model has to be defined. Since the dawn of monolithic ICs invention in the '60s (Noyce [2007]) the importance of produce an accurate and, at the same time, computationally efficient model for MOSFET became crucial. In early days, when the number of components per chip was small, manual techniques were suitable. Today the highly non-linear equations describing the electrical quantities of MOSFETs leads to computer-aided circuit design mandatory.

This kind of approach allows dramatic reduction of the design cost and improve the understanding of the behavior of the circuit under design. In fact, the first version of the popular circuit simulator program SPICE (Simulation Program with Integrated Circuit Emphasis) which was implemented at the Berkely UC in 1973 was not the first circuit simulator ever created (e.g. BIAS, CANCER and TIME) but was the first thought for undergraduate classroom instruction, so that it was powerful enough to allow students to grasp the working principles of small circuits but, at the same time, sufficiently simple to be understood easily by the majority of them (Nagel and Pederson [1973]). Nowadays most of the commercially available circuit simulation programs are based on SPICE (Arora [2007]) that has been released under Public-domain licence, reason why it became so popular.

SPICE-based circuit simulators typically are able to simulate circuit of few hundreds of nodes and perform at least the following analyses (SPI)

- DC Analysis
- AC Small-Signal Analysis
- Transient Analysis
- Pole-Zero Analysis
- Small-Signal Distortion Analysis
- Sensitivity Analysis
- Noise Analysis

In general, these simulations, are performed solving numerically the nodal equations at each node. For this reason the accuracy of the models greatly influence the quality of the result. Involving non-linear and time-dependent quantities, like most of those associated with semiconductor based devices, often cumbersome system of non-linear differential equations have to be solved. To maintain the problem reasonably practical to solve and, at the same time, maintain the error as low as possible, some kind of approximations to the models have to be done, as is going to be showed in the following.

1.4.1 MOSFET Models

Many MOSFET models have been developed through the years. The progressive shrinking of the device caused the device models to become more and more complex, so that computationally efficiency became fundamental. For this reason three major kind of model can be distinguished and most of the circuit simulators allow the user to choose which one suits better the trade-off between accuracy and speed of choice.

Analytical Models

Expressions controlling I-V and C-V curves are directly derived from the device physics and by its geometry. This kind of models are the most accurate but also the most computationally expensive, they are seldom used without a certain amount of reasonable

approximations. Typically in MOSFETs the expressions are derived from electrostatic potential analysis. For example, for sufficiently long MOSFETs like the traditional planar MOSFETs and in some multi-gate MOSFETs as the DG-MOSFET and Cy-GAAFET, where the geometrical symmetry of the body can be exploited, the Poisson's equation, as will be shows, can be solved and the potential profile can be extracted quite accurately. Analytical models are based on continuous expressions along the entire domain of the device operation and allow deep understanding of the electrostatics of the system. Furthermore, they give accurate curves also when the device is scaled and the bias voltages modified. Analytical models can be used to simulate small circuits only, otherwise approximations on the expressions, to get simpler ones, can be done by neglecting secondary effects. In many situations, can be convenient the introduction of physical and empirical parameters, to include higher order effects without affecting complexity excessively. These parameters are often extracted by fitting of experimental curves or from device level simulations. Despite analytical models directly link the physics and geometry of the device with the electrical behavior, the model is typically hard to be developed and when the technology is changed and the device structure modified, the existing model often has to be adjusted and sometimes a whole new model have to be done from the ground up.
For circuit CAD, these kind of analytical models are referred to as *compact models*, they are based on a physics based analytical expressions which are fitted to technology and bias dependent experimental data in order to extract a set of *model parameters* that are fundamental to pursue accuracy in scaled technologies.
Compact models for circuit CAD are important for cost-effective design, performance optimization to get robust IC designs. Compact models also allow the integration of process variability of the different device parameters which can be associated to statistical distributions. Furthermore, knowing the manufacturing tolerances, worst case scenarios can be easily simulated (Saha [2016]).

Table Lookup Models

Data extracted from experimental devices or from a device level simulators are stored in a lookup table. It is a technology independent approach, simpler to develop than the analytical but it does not give any insight of the physical phenomena happening inside the device, hence the understanding of the system is scarce. Furthermore a large amount of memory may be required to get high accuracy.

Empirical Models

In this case, the model is based on expressions extracted by fitting of experimental curves or from device level simulators. The accuracy of these models is similar to table lookup models' but require lower time to be developed and lower memory. The main disadvantage is that this approach is technology dependent (Arora [2007]).

Chapter 2

Double-Gate FinFET

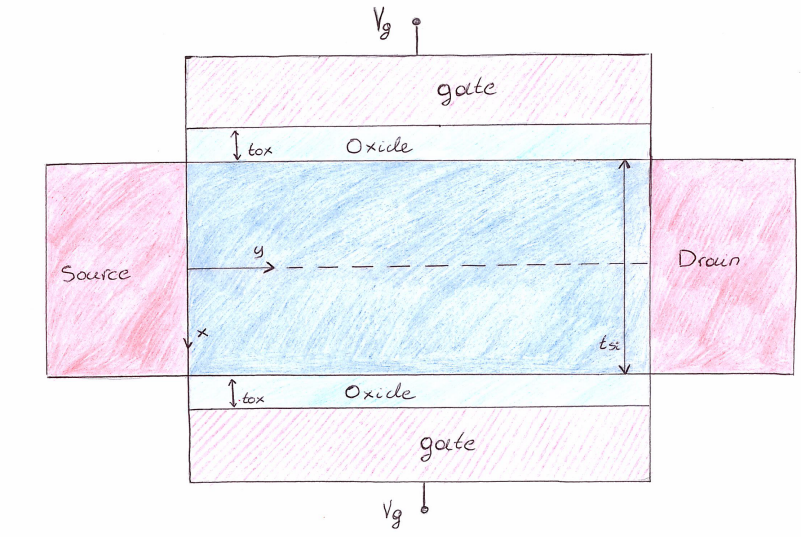


Figure 2.1: DG-FinFET section for $z = z_0$. The device geometry is invariant along the z axis.

The first step is to develop the *core model*, i.e. the first-order analytical description of the device behavior without using fitting parameters. It relies totally on physical assumptions and provides an accurate model of the electrostatics inside an *ideal* device. From the core model order-of-magnitude quantities can be extracted. The core model returns curves as accurate as the *real* device behaves "ideally", i.e., as far the assumptions the core model is based on are valid.
The main assumption is the *gradual channel approximation (GCA)* which strongly simplifies the solution of the *Poisson's equation*. This approximation tells that the device is long enough to safely neglect SCEs, gate-tunneling, charge quantization without losing accuracy (Chauhan [2015]).

Core models are great to compare different technologies at a first approximation but are insufficient to design real circuits, especially in the case of very scaled down transistors. Corrections based on empirical measurements are fundamental.
In the next section core model for FinFET is explained in great detail. It will be calculated numerically and the main electrical and physical quantities are extracted. Subsequently, a second-order effects are introduced to correct the core model.
This kind of approach, core model plus corrections, is shared with most of the silicon-based devices and, as well as for other technologies, the real devices suffers of the process variations that are stochastic by nature. This kind of variations affect strongly the electrical behavior of a specific device that will then inherit a stochastic nature. In a real model, aimed at manufacturing, also these variations have to be taken into account and they will be strongly dependent on the specific process used. Since these kind of variations may be difficult to be characterized analytically, fitting with empirical measurements became even more important.

2.1 Core Model

The device operation is controlled by the Poisson's equation

$$\frac{\partial^2\psi(x,y)}{\partial x^2} + \frac{\partial^2\psi(x,y)}{\partial y^2} = -\frac{\rho(x,y)}{\epsilon_{si}} \tag{2.1}$$

where ψ is the electrostatic potential, ϵ_{si} the dielectric constant of silicon and $\rho(x)$ is the total space-charge density

$$\rho(x,y) = q(p(x,y) - n(x,y) + N_d^+ - N_a^-) \ , \tag{2.2}$$

where N_d^+ and N_a^- are ionized donors and acceptors densities, whilst $p(x,y)$ and $n(x,y)$ are hole and electron densities, respectively. Now on the analysis will be focused on the NMOS which has a p-type channel.
Under the gradual channel approximation $\frac{\partial^2\psi(x,y)}{\partial x^2} \gg \frac{\partial^2\psi(x,y)}{\partial y^2}$. The meaning of the GCA is that the variation of the longitudinal electric field is much lower than the transversal.
Now, assuming Boltzmann statistics, a moderately doped substrate we have

$$\frac{\partial^2\psi(x,y)}{\partial x^2} = \frac{q}{\epsilon_{si}}\left(n_i e^{q\frac{(\psi(x,y)-\psi_B-V_{ch}(y))}{k_BT}} + N_a\right) \tag{2.3}$$

where $\psi_B = \|\phi_F - \phi_i\| = \frac{k_BT}{q}\log(N_a/n_i)$, i.e., the difference between the Fermi potential and the intrinsic potential; $V_{ch}(y)$ is the *quasi-Fermi* potential, in this context called *channel potential*.

Note that the quasi-Fermi potential along x is assumed to be constant. This is a reasonable assumption since no current is flowing in the transversal direction, because of the presence of oxide layers. In real devices a small gate leakage current is present but in the core model is neglected. Furthermore, no electric field along the z direction

has been considered; again this is a reasonable assumption for long-channel DG-FinFETs only. Equation (2.3) is a ordinary second-order non-linear differential equation. It could be solved analytically if the doping concentration N_a was small compared to the mobile charge $n(x, y)$ (Taur [2000]) but doping must be included because of its importance in fine tuning of the threshold voltage. In order to take it into account *perturbation approach* has been enforced (Dunga [2008]), meaning that the electrostatic potential is written as

$$\psi(x, y) \cong \psi_1(x, y) + \psi_2(x, y) \ , \tag{2.4}$$

and the Poisson's equation is split into two parts

$$\frac{\partial^2 \psi_1(x, y)}{\partial x^2} = \frac{q}{\epsilon_{si}} n_i e^{q\frac{(\psi(x,y)-\psi_B - V_{ch}(y))}{k_B T}} \tag{2.5}$$

$$\frac{\partial^2 \psi_2(x, y)}{\partial x^2} = q\frac{N_a}{\epsilon_{si}} \ . \tag{2.6}$$

Enforcing the symmetry of the system, expression (2.5) and (2.6) can be integrated twice with the boundary condition $\frac{\partial \psi_1(0,y)}{\partial x} = 0$

$$\psi_1(x, y) = V_{ch}(y) - \frac{2k_B T}{q} log\left(\frac{t_{si}}{2\beta}\sqrt{\frac{q^2 n_i^2}{2\epsilon_{si} k_B T N_a}} cos\left(\frac{2\beta x}{t_{si}}\right)\right) \ . \tag{2.7}$$

$$\psi_2(x, y) = \frac{qN_a x^2}{2\epsilon_{si}} \tag{2.8}$$

where t_{si} is the thickness of the fin and β is constant along x and is defined as

$$\beta = \sqrt{\frac{q^2}{2\epsilon_{si} k_B T}\frac{n_i^2}{N_a} e^{q\frac{\psi_0 - V_{ch}(y)}{k_B T}}}\frac{t_{si}}{2} \tag{2.9}$$

where ψ_0 is the electrostatic potential at the center of the channel. It is function of y and in order to get the its value it is important to solve the *gate bias equation*:

$$V_g - V_{fb} = V_{ox} + \psi_s(y) \ , \tag{2.10}$$

where $\psi_s(y) = \psi(x = \pm\frac{t_{si}}{2}, y)$, i.e. the surface potential, and $V_{ox} = \mathcal{E}_{ox}/t_{ox}$ the potential drop across the oxide, whilst $V_{fb} = (\phi_m - \phi_s)$ is the *flatband voltage*.
Considering no charge in the oxide, constant electric field across the oxide can be assumed. Boundary condition for the electric fields at the interface is $\epsilon_{ox}\mathcal{E}_{ox} = \epsilon_{si}\mathcal{E}_s$ and together with the *Gauss's law* allows to rewrite the gate bias equation as Y. Taur [2021]

$$V_g - V_{fb} = \frac{\epsilon_{si}}{C_{ox}}\frac{\partial \psi(x = \pm\frac{t_{si}}{2}, y)}{\partial x} + \psi_s(y) \ . \tag{2.11}$$

Equation (2.11), together with

$$\psi_s(y) \cong \psi_1(x = \pm\frac{t_{si}}{2}, y) + \psi_2(x = \pm\frac{t_{si}}{2}, y) \ , \tag{2.12}$$

has to be solved to get the value of β for each $V_{ch}(y)$.
This system can be written as Chauhan [2015]

$$log\beta - log(cos\beta) - q\frac{V_g - V_{fb} - V_{ch}(y)}{2k_BT} + log\left(\frac{2}{t_{si}}\sqrt{\frac{2\epsilon_{si}k_BTN_a}{q^2n_i^2}}\right)$$
$$+\frac{2\epsilon_{si}}{t_{si}C_{ox}}\sqrt{\beta^2\left(\frac{e^{q\frac{\psi_{pert}}{k_BT}}}{cos^2\beta} - 1\right) + q^2\frac{\psi_{pert}}{k_B^2T^2}[\psi_{pert} - 2\frac{k_BT}{q}log(cos\beta)]} = 0 \ ,$$

where $\psi_{pert} = \psi_2(x = \pm\frac{t_{si}}{2})$.
This is an implicit equation of β and can only be solved numerically.
Once β is calculated it can be used to get the potential distribution along x from (2.4), for different values of the channel potential.
The gate bias equation put in relationship the gate voltage with the surface potential and this gives the ability to study the behavior the electric field, the mobile charge density and the potential itself inside the channel as the gate potential is changed; but it also gives the ability to extract the gate capacitance, fundamental for transient analysis.
Under the assumption that that all the potentials are referred to the source and that fully depletion of the channel is considered, some results of this analysis are presented.

2.1.1 Electrostatic potential

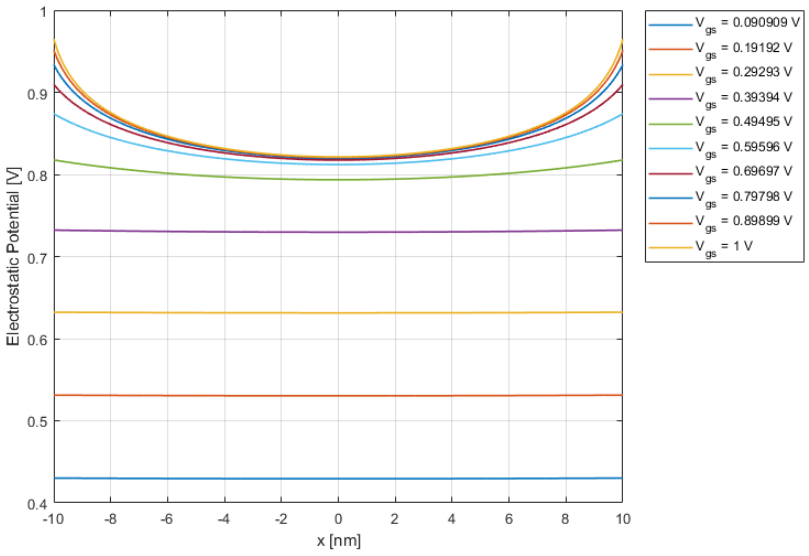


Figure 2.2: Electrostatic potential near the source, $\psi(x, y = 0)$, for different V_g. $N_a = 10^{16}$ cm^{-3}, $t_{si} = 20$ nm, $t_{ox} = 1$ nm.

The electrostatic potential in the channel is shown in Figure 2.2. In this case the gate bias equation has been solved for $V_{ch} = 0$, that is near the source end.
It is clear as the electrostatic potential $\psi(x)$ increases evenly as the gate potential is increased (in this specific case from 0V to 0.5V) meaning that the mobile charge density will increase exponentially equally in all the volume inside the channel, also in the central region where the mobility is higher because of the lower probability, for mobile carriers, to undergo scattering with the surface defects (Colinge).
Gradually, for higher values of V_g the potential start to saturate in the mid-point and to bend at the interfaces: mobile carriers are attracted by the gate potential.

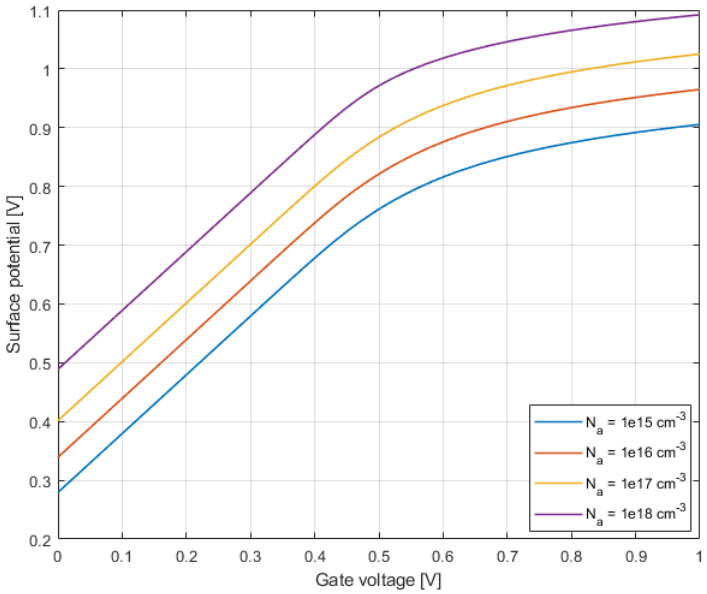


Figure 2.3: Surface potential near the source, $\psi_s(x = \pm\frac{t_{si}}{2}, y = 0)$. $t_{si} = 20$ nm, $t_{ox} = 1$ nm.

In Figure 2.3, can be seen how the surface potential , similarly to the planar MOSFET, saturates because of the presence the inversion layer which screens the silicon from the penetration of the electric field which will be confined inside the oxide layer (Y. Taur [2021]). The saturation of the surface potential typically coincides with the gate potential equal to what is called threshold voltage V_{th} in standard MOSFETs. In fact, this kind of definition may be unpractical from a design point of view and different definitions of V_{th} have been proposed by Sánchez et al. [2006].

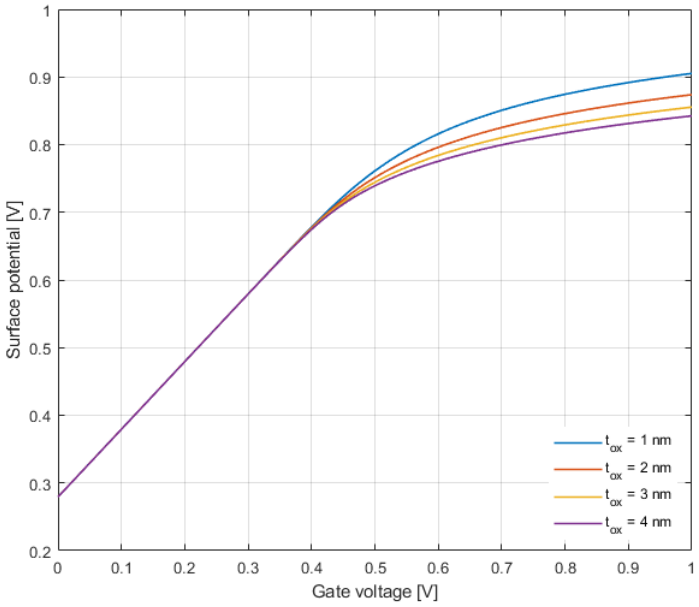
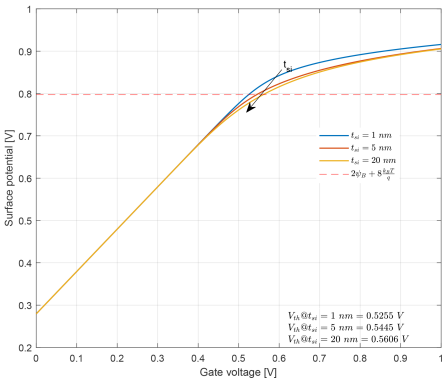


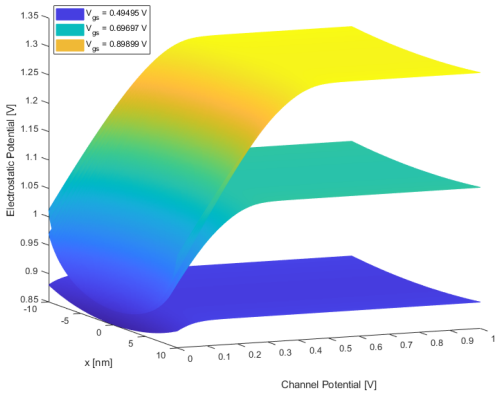
Figure 2.4: Surface potential near the source, $\psi_s(x = \pm\frac{t_{si}}{2}, y = 0)$.
$N_a = 1e15$ cm^{-3}, $t_{si} = 20$ nm.

In Figure 2.4, is shown as the surface potential changes with the thickness of the oxide. In Figure 2.5a, instead, as it changes when the thickness of the silicon is varied. In this case the threshold voltage is calculated as the gate voltage at which the surface potential is equal to $2\psi_B + 6k_BT/q$, conventionally set as the point at which saturation happens (Arora [1993]).
It is clear as the threshold voltage increases as the thickness of t_{si} increases, as well as the t_{ox} increasing.

In Figure 2.5b, visually, the dependence of the electrostatic potential to the channel potential is shown. For high values of V_{ch} the potential curvature is reduced, this due to the *pinch-off* condition, that is when the transversal electric field goes to zero because of the effect of an high drain voltage. This kind of condition happens typically near the drain end (Arora [2007]). Here the exact relationship between a V_{ch} and y is not visible. To get it the continuity equation for the current has to be solved in a self-consistent manner together with the 2D Poisson's equation. In general, the solution has to be carried out numerically (Hong and Taur [2021]).



(a) Surface potential at $V_{ch} = 0$, $\psi_s(x = \pm\frac{t_{si}}{2}, y = 0)$. $N_a = 1e15$ cm^{-3}, $t_{ox} = 1$ nm.



(b) Electrostatic potential as function of V_{ch}. $N_a = 1e17$ cm^{-3}, $t_{ox} = 1$ nm, $t_{si} = 20$ nm.

Figure 2.5

2.1.2 Charge density

Exploiting the Gauss's law (Hu [2010])

$$\left(\frac{\partial\psi(x,y)}{\partial x}\right)_{x=\frac{t_{si}}{2}} - \left(\frac{\partial\psi(x,y)}{\partial x}\right)_{x=0} = -\frac{Q_s(y)}{\epsilon_{si}} \tag{2.13}$$

$Q_s(y)$, i.e., the total charge density per unit area inside the channel can be extracted. The LHS of (2.13) can be taken analytically by the derivation of (2.1)

$$Q_s(y) = -\epsilon_{si}\left(\frac{\partial\psi(x,y)}{\partial x}\right)_{x=\frac{t_{si}}{2}} = -\epsilon_{si}\left(4\frac{k_BT}{q}\frac{sin\,(\beta)}{cos\,(\beta)}\frac{\beta}{t_{si}} + \frac{qN_a}{\epsilon_{si}}\frac{t_{si}}{2}\right) \tag{2.14}$$

This quantity depends on the channel potential V_{ch} as well as to V_g and in figure 2.6a and 2.6b the total charge density inside the fin is shown for different doping concentrations.

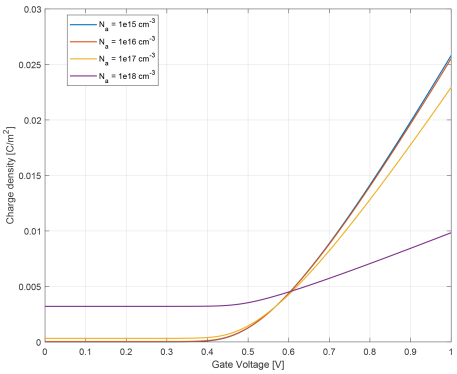
Especially in figure 2.6b, the depletion fixed charge is clearly visible in subthreshold region.

Assuming complete depletion inside the channel, inversion charge density Q_i can be simply obtained by subtraction

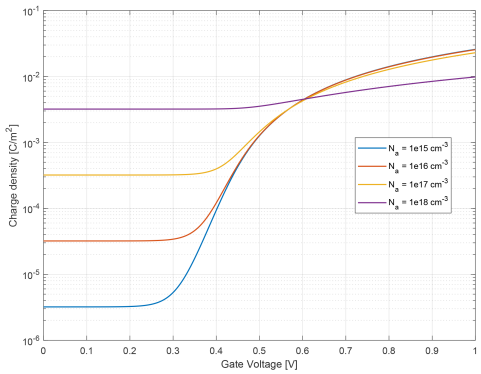
$$Q_i = Q_s - qN_at_{si}\ . \tag{2.15}$$

Q_i is the mobile charge component of the total charge density; it contributes to the current density inside the device.

.

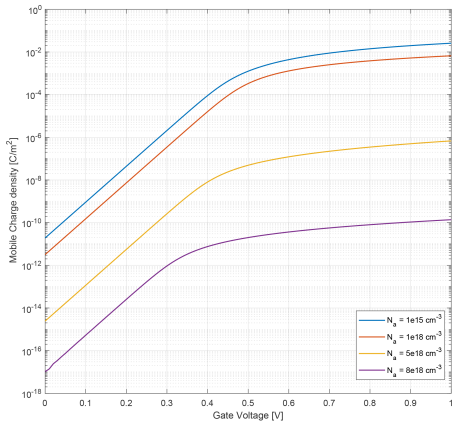


(a) Q_s in linear scale for different N_a.

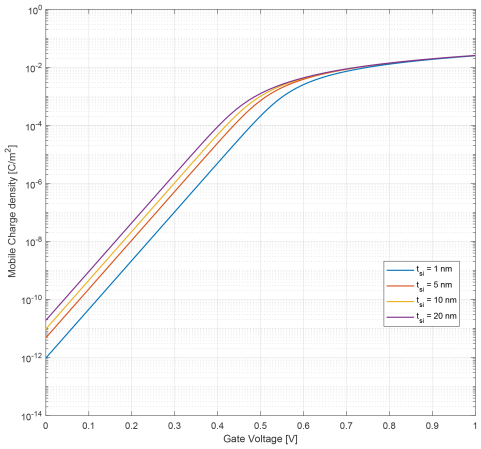


(b) Q_s in logarithmic scale for different N_a.

Figure 2.6: Q_s at $V_{ch} = 0$. $t_{si} = 20$ nm, $t_{ox} = 1$ nm.



(a) Q_i in logarithmic scale for different N_a. $t_{si} = 20$ nm.



(b) Q_i in logarithmic scale for different t_{si} $N_a = 1e15$ cm^{-3}.

Figure 2.7: Q_i at $V_{ch} = 0$. $t_{ox} = 1$ nm.

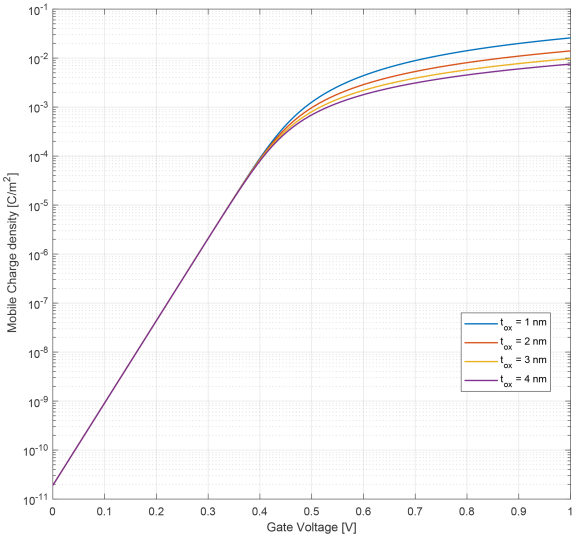


Figure 2.8: Q_i in logarithmic scale at $V_{ch} = 0$. $N_a = 1e15$ cm^{-3}, $t_{si} = 20$ nm.

2.1.3 Drain current

Drain current can be obtained by the application of the current continuity to the current density definition (Taur et al. [2004])

$$J_n(x,y) = (x,y) = -q\mu_n n(x,y)\frac{dV_{ch}}{dy} \ , \tag{2.16}$$

where μ_n is the effective mobility of mobile carriers, $n(x,y)$ their volume density and $J_n(x,y)$ is the electron carrier density.
This kind of current is both diffusion and drift in nature, and the current continuity state that at any point of y in the channel, the total current I_n, is constant. With this assumption integrating 2.16 in x

$$I_D(y) = qW\mu_n\frac{dV_{ch}}{dy}\int_0^{\frac{t_{si}}{2}} n(x,y)dx \ , \tag{2.17}$$

where W is equal to two times the height of the fin of the DG-FinFET and $I_D(y)$ is the drain current per unit length.
Knowing that inversion charge density per unit are is

$$Q_i(y) = -q\int_0^{\frac{t_{si}}{2}} n(x,y)dx \tag{2.18}$$

The drain current per unit length can be written as

$$I_D(y) = -Wv_eQ_i(V_{ch}) \ , \tag{2.19}$$

where $v_e = \mu_n \frac{dV_{ch}}{dy}$ is the electron velocity in the channel and so

$$I_D(y) = -W\mu_n \frac{dV_{ch}}{dy} Q_i(y) = -W\mu_n \frac{dV_{ch}}{dy} Q_i(V_{ch}) \ , \tag{2.20}$$

where y can be substituted with V_{ch} because the latter depends on the first only (Taur and Ning [2021]).

The last step is to multiply dy both sides of 2.20 and integrate from 0 to L_g

$$\int_0^L I_D(y)dy = -W\mu_n \int_0^{V_{ds}} Q_i(V_{ch})dV_{ch} \ . \tag{2.21}$$

Finally, since $I_D(y) = const.$, because of the continuity condition,

$$I_D = -\mu_n \frac{W}{L} \int_0^{V_{ds}} Q_i(V_{ch})dV_{ch} \ . \tag{2.22}$$

Combining (2.22) and (2.15) the total current can be calculated by integration. The first approach employed in this work is numerical. This approach's accuracy strongly depends on the resolution of the integration domain, i.e., how small dV_{ch} is, and, of course, higher is the accuracy lower will be the program to complete the computation.

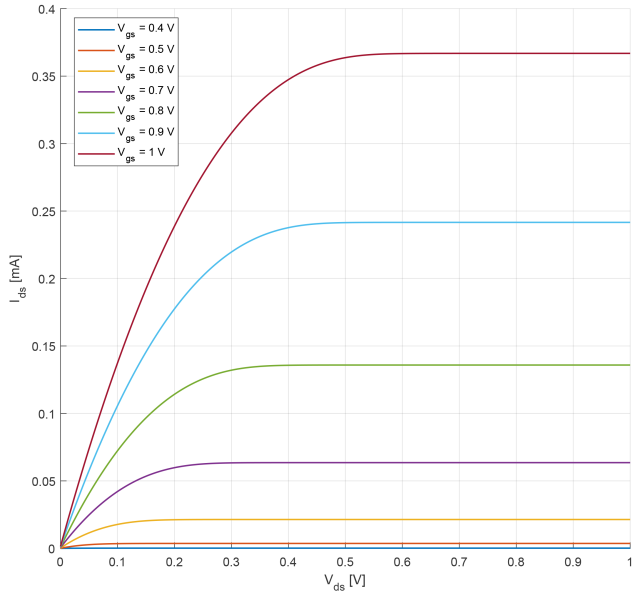


Figure 2.9: Output characteristics of the DG-FinFET. $N_a = 1e17$ cm^{-3}, $t_{si} = 20$ nm, $t_{ox} = 1$ nm, $L_g = 1$ μm, $\mu_n = 136$ cm^{-2}/(V $-$ s), $W = 1$ μm

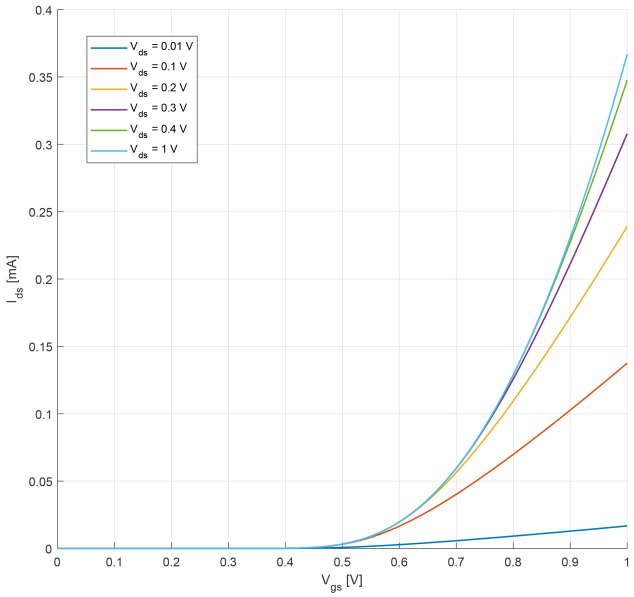


Figure 2.10: Transcharacteristics of the DG-FinFET. $N_a = 1e17$ cm^{-3}, $t_{si} = 20$ nm, $t_{ox} = 1$ nm, $L_g = 1$ μm, $\mu_n = 136$ cm^{-2}/(V − s), $W = 1$ μm

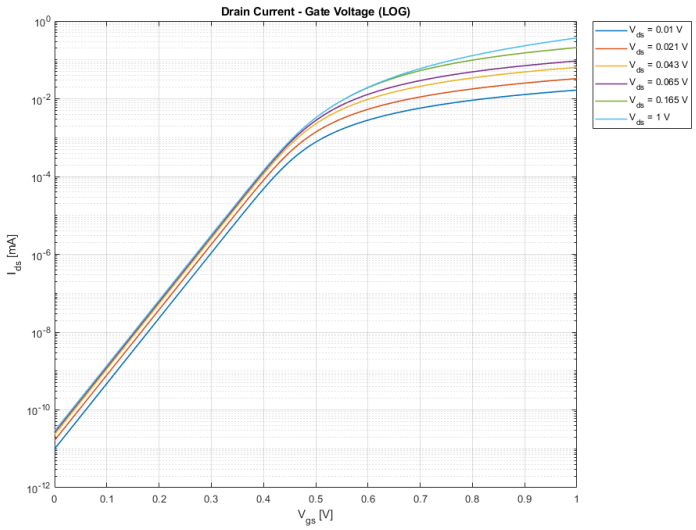


Figure 2.11: Transcharacteristics of the DG-FinFET in logarithmic scale. $N_a = 1e17$ cm^{-3}, $t_{si} = 20$ nm, $t_{ox} = 1$ nm, $L_g = 1$ μm, $\mu_n = 136$ cm^{-2}/(V − s), $W = 1$ μm

Device models for circuit simulations must be as fast as possible, in order to deal with increasing complexity, and for this reason, expression involving integration as in (2.22) should be avoided, considering the numerical perspective, also because the dependency onto the quantity dV_{ch} controls the accuracy of the output of the integration. However this kind of integration showed to be the most accurate when dV_{ch} is chosen to be very small, to complexity and resource costs.
However, in Colinge, researchers managed to carry out equation (2.22), using the gate bias equation 2.10 discussed in 2.1, expressing the silicon charge density as

$$Q_s(y) = C_{ox}(V_{gs} - V_{fb} - \psi_s(y)). \tag{2.23}$$

The drain current can be written as

$$I_D = \mu_n \frac{W}{L} \left(f(\psi_s) - f(\psi_d) \right), \tag{2.24}$$

where

$$f(\psi_s) = \frac{Q_i^2}{2C_{ox}} + 2\frac{kT}{q}Q_i - \frac{kT}{q}\left(5\frac{\epsilon_{si}kT}{qt_{si}} + Q_{bulk}\right) \cdot \left(5\frac{\epsilon_{si}kT}{qt_{si}} + Q_{bulk} + Q_i\right), \tag{2.25}$$

where $Q_{bulk} = qN_at_{si}/2$ The accuracy of the current expressed in (2.24) is slightly lower with respect of what is obtained from the numerical integration but the advantage in terms of efficiency justifies its application.

2.2 Real Device Effects

To have an excellent prediction of the I-V and C-V curves, together with the core model, *real device effects* must be included, they are typically modeled as correction of core model quantities like the threshold voltage or a geometrical parameter. These kind of corrections, to give high accuracy cannot preclude the use fitting parameters which are extracted from experimental measurements. In this work all the corrections included in the models are only based on physical and geometrical assumptions with no fitting parameters involved; this is mainly because of the practical impossibility to have access to experimental measurements and, even trying to match physical (TCAD) simulations becomes pointless over a certain level of accuracy without technology related information, since the fitting with real devices is highly process dependent.
Another reason to rely only on physical and geometrical assumptions is to keep the model as simple as possible in order to give the ability to the user to have an insight of what is going on inside the device.

2.2.1 Channel Length Modulation (CLM)

Similarly to the drain current in the planar MOSFET, I_D in real devices shows a non-zero conductance in the saturation region. One of the contributions to this effect is the channel length modulation. When the drain current exceed the saturation voltage $V_{D_{sat}} = V_g - V_{th}$,

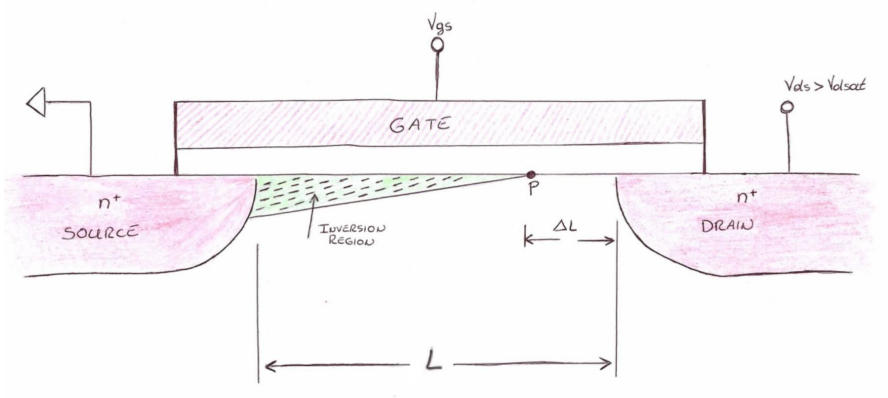


Figure 2.12: The depletion region width between P and the drain end is modulated by the drain overdrive.

the channel undergoes to pinch-off condition, where the transversal electric field goes to zero. Any further potential applied to the drain exceeding $V_{D_{sat}}$ will drop onto the depleted region between the drain and the channel at a certain point P, in Figure 2.12. In fact, the GCA cannot be applied in the "pinched-off" region of the channel and full 2D Poisson's equation must be solved to get the exact shape of the potential near the drain. This region can be treated as a reversely biased PN junction whose space charge region width is modulated by the drain potential overdrive ($V_D - V_{D_{sat}}$). The result is the reduction of the effective length of the channel of a quantity ΔL and the consequent increasing of the current in the saturation region.
In order to take into account this effect without invoking the full Poisson's equation solution, many semi-empirical approaches have been proposed (Arora [1993]). The same models used for planar MOSFET can be applied also here.
In order to avoid the use of *fitting parameters* as much as possible to give priority to the understanding of the device working principles rather than ultimate accuracy, the CLM model used here (Reddi and Sah [1965]) treats the pinch-off region, as a simple inversely biased PN junction between the drain and the channel with a potential drop across its space charge region equal to $V_D - V_{D_{sat}}$.
From basic semiconductor theory, space charge region width of an N^+P junction can be expressed as (Hu [2010])

$$W_{dep} = \sqrt{\frac{2\epsilon_{si}(V_R + V_{bi})}{qN_b}} \ , \tag{2.26}$$

where $V_{bi} = \frac{k_B T}{q} log\left(\frac{N_a N_d}{n_i^2}\right)$ is the *built-in potential* between the drain and the channel and V_R is the reverse bias applied: in this case $V_R = V_D - V_{D_{sat}}$.
The *saturation voltage* is defined as $V_{D_{sat}} = V_g - V_{th}$; it is the voltage at which $\frac{dI_D}{dV_D} = 0$.

Now, from 2.22 L is substituted with $L_{eff} = L - W_{dep} = L - \Delta L$.
The output characteristics, in Figure 2.13, happens to have a non-zero conductance in

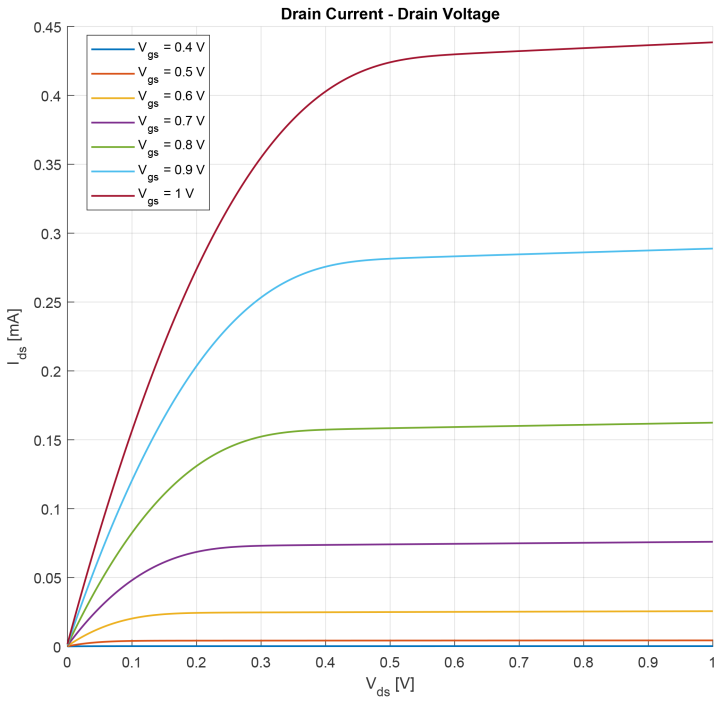


Figure 2.13: Channel Length Modulation causes a non-zero conductance in saturation region. $N_a = 1e17$ cm^{-3}, $N_d = 1e21$ cm^{-3} $t_{si} = 20$ nm, $t_{ox} = 1$ nm, $L_g = 1$ μm, $\mu_n = 136$ cm^{-2}/(V $-$ s), $W = 1$ μm.

the saturation region. This has an impact on the gain of amplifiers which is lowered by the CLM and on the transition slope in CMOS technology which is also reduced. CLM is mostly a problem in analog applications rather than in digital (Weste and Harris [2011]).

2.2.2 Drain induced barrier lowering (DIBL)

When $V_g < V_{th}$, electrons in the source-end encounter an high potential barrier preventing them to flow towards the drain. In long devices, this barrier, is mainly flat, except for the regions near the source and the drain, where the electric field due to the junctions affects the barrier, lowering it.
When the channel length is shortened the potential barrier shrinks because of the effect of source and drain junctions and below a certain L_g, its maximum becomes a point near the source end (Taur and Ning [2021]). In this condition, it is easier for electrons to overcome the potential barrier and to reach the drain end.
Modulating the drain voltage, the maximum of the barrier will be at the same time lowered and pushed against the source. The impact on the output characteristics is first of all, an

increasing of the subthreshold current, but also an increasing of the output conductance. Because of 2D nature of the electric field near the drain and the source end, in order to include this kind of phenomena in the DG-FinFET model (in general, in any kind of MOSFET model), the full 2D Poisson's equation must be solved (Chauhan [2015]). In order to do so without the cumbersome task to find the solution in a self-consistent manner with the current continuity equation, the 2D Poisson's equation has been solved in subthreshold regime, that is neglecting the mobile charge density.

$$\frac{\partial^2\psi(x,y)}{\partial x^2} + \frac{\partial^2\psi(x,y)}{\partial y^2} = \frac{qN_a}{\epsilon_{si}} \ . \tag{2.27}$$

In this way the problem to be solved is linear and do not involve the need to find the relationship between V_{ch} an y.
This kind of calculation is quite long and it is based on the assumption of a parabolic potential distribution along x direction. It is carried out and well explained by Taur and Ning [2021] and in the book of Colinge.
Here, only the effects of this analysis are integrated in the core model.
It has been proven that the DIBL affects the threshold voltage by lowering it by a certain ΔV_{th} depending on the geometrical parameters of the FinFET, mainly the channel length, and also on the drain voltage and gate voltage. In particular, the *scale length* λ is defined as the measure of how much the drain electric field penetrates into the silicon body and then how severe the SCEs will be (Chauhan [2015]).

$$\Delta V_{th} = \frac{2((V_{bi} - V_{SL}) + V_{ch}(L_g))}{2cosh(\frac{L_g}{2\lambda}) - 2} \ , \tag{2.28}$$

where V_{SL} is the center potential for the long-channel transistors

$$V_g - V_{fb} - q\frac{N_a}{\epsilon_{si}\lambda^2} \tag{2.29}$$

$$\lambda = \sqrt{\frac{\epsilon_{si}}{2\epsilon_{ox}}\left(1 + \frac{\epsilon_{ox}t_{si}}{4\epsilon_{si}t_{ox}}\right)t_{si}t_{ox}} \ . \tag{2.30}$$

To include 2.28 into the core model, it has added V_g in the implicit equation of β 2.1, that becomes

$$log\beta - log(cos\beta) - q\frac{V_g + \Delta V_{th} - V_{fb} - V_{ch}(y)}{2k_BT} + log\left(\frac{2}{t_{si}}\sqrt{\frac{2\epsilon_{si}k_BTN_a}{q^2n_i^2}}\right)$$
$$+\frac{2\epsilon_{si}}{t_{si}C_{ox}}\sqrt{\beta^2\left(\frac{e^{q\frac{\psi_{pert}}{k_BT}}}{cos^2\beta} - 1\right) + q^2\frac{\psi_{pert}}{k_B^2T^2}[\psi_{pert} - 2\frac{k_BT}{q}log(cos\beta)]} = 0 \ ,$$

which has to be solved in the same way.
From 2.28, shows to be exponentially dependent to L_g, lower the channel length, higher the threshold voltage roll-off, but also higher the its sensitivity the gate length itself.
This kind of analysis is performed below threshold, to get a comprehensive prediction of this effect, including above-threshold characterization inversion carrier density must be included in 2.27. The accuracy of this model worsen as the inversion charge increases its magnitude with V_g.

2.2.3 Velocity saturation

When the device get shorter, *velocity saturation* becomes so important that it limits the advantages to scale L_g. In effect, the drain current won't scale as $1/L$ but becomes hand by hand less sensitive to the channel length (Taur et al. [1993]).
Also, the saturation of the drain voltage arrives much before than in the long-channel case.
Velocity saturation happens when an intense longitudinal electric field leads the carrier velocity to be non-linearly dependent to the electric field and in particular saturating with an fields higher than a critical value (Sze [2007]).
In order to include this effect to the finFET model, the empirical velocity-field relationship by Caughey and Thomas [1967] has been employed

$$v_e = \frac{\mu_e \left(\frac{dV_{ch}}{dy}\right)}{1 + \left(\frac{\mu_e}{v_{sat}}\right)\left(\frac{dV_{ch}}{dy}\right)} \ , \tag{2.31}$$

where $v_{sat} = \mu_e E_c$ is the velocity saturation, a constant value independent on the electric field and E_c is the *critical field* at which saturation happens.
In 2.31 the term $\frac{dV_{ch}}{dy}$ contains the information about both drift and diffusion currents, however in saturation region the current is mainly drift (Taur and Ning [2021]).
if 2.31 is substituted to v_e in 2.19

$$I_D(y) = -WQ_i(V_{ch})\frac{\mu_e \left(\frac{dV_{ch}}{dy}\right)}{1 + \left(\frac{\mu_e}{v_{sat}}\right)\left(\frac{dV_{ch}}{dy}\right)} = -\left(\mu_n WQ_i(V_{ch}) + \frac{\mu_n I_D}{v_{sat}}\right)\frac{dV_{ch}}{dy} \ , \tag{2.32}$$

as already done for 2.19, both sides are multiplied to dy and integration is done from 0 to V_D. The expression for the drain current becomes

$$I_D = \frac{-\mu_n (W/L)\int_0^{V_D} Q_i(V_{ch})dV_{ch}}{1 + \left(\frac{\mu_n V_D}{v_{sat} L}\right)} \ . \tag{2.33}$$

v_{sat} for electrons is a temperature dependent parameter. It has been extracted by Taur et al. [1993] which showed to be more or less equal to $7 \cdot 10^6$ cm/s for both electrons and holes in the inversion region at 300 K. Lower temperatures gives higher saturation velocities.

2.2.4 Quantum mechanical effects (QME)

In ultra-scaled, transistors, in the nanometer range, electrons in the inversion layer must be treated as two-dimensional electron gas (2DEG) (Colinge), i.e. taking into account the quantum confinement in the normal direction, while electrons are free to move in the other two directions. The main effects of this confinement are:

- The inversion layer carriers are now occupying discrete sub-bands.

- The peak of the inversion layer carriers is few nanometers away from the $Si - SiO_2$ interface.

Effectively, the first effect is the cause of the increasing of the effective E_g so the reduction of the inversion carrier concentration because (Sze [2007])

$$n_i \propto e^{-\frac{E_g}{k_B T}} \tag{2.34}$$

$$n(x) = n_i e^{\frac{E_F - E_i}{k_B T}} . \tag{2.35}$$

The shift of the inversion layer , instead, causes the increase of the effective oxide thickness t_{ox} reducing the electrostatic control of the gate to the channel and the consequent increasing of the threshold voltage. In DG-FinFET, the inversion populations induced by the two gates merge together at the center of the fin causing *volume inversion*. These two effects cause the degradation of the I-V and C-V curves. All together, in order to predict this effects quantitatively the Poisson's equation, self-consistently with the *Schrödinger's equation* have to be solved (Saha [2021]).
Another consequence of having the concentration peak further from the surface is the reduction of interface scattering and the improvement of carrier mobility with respect to the classical theory. In any case however, in very thin fins, the interface scattering increases because of the physical proximity of the surfaces with the center of the channel, where the carrier concentration has its peak. Both the reduction of the inversion concentration due to bands quantization and the increase of the effective oxide thickness can be modelled as a variation of the threshold voltage. When the quantization of the conduction band happens, the ground level determines the energy bandgap which start to depend on the fin thickness: lower is the thickness, higher is the energy bandgap (Chauhan [2015]). The quantum correction to be applied in this model are based on quantum consideration which cannot be predicted without fitting parameters that depend on experimental measurements. However, approximating the fin section as an infinite potential well and the potential to be flat, that is something fairly true in the subthreshold region but not above it, where in 2.1.1 it as been shown to rather be parabolic, the *Schrödinger's equation* has a simple solution (Saha [2021]) for the first conduction sub-band which has to be added to conduction band edge without quantum confinement. The flat potential assumption gives more accurate results if the doping level inside the channel is low. The result is a variation of the threshold voltage equal to

$$V_{th_{QM}} = \frac{\pi^2 \hbar^2}{2 q m^* t_{si}^2} , \tag{2.36}$$

affecting the threshold voltage as

$$V_{th} = V_{th_0} + V_{th_{QM}} . \tag{2.37}$$

2.3 Capacitances

The ability to calculate accurate C-V characteristics enables the possibility to perform transient analysis in circuit simulators.

2.3.1 Intrinsic Capacitance

In 2.1.2 the charge density has been extracted thanks to the Gauss' law and the accurate prediction of the potential inside the silicon fin. Now, a step further can be done calculating the total gate charge as the integral of the charge density along the channel. That is:

$$Q_g = W \int_0^L Q_s(y)dy = WC_{ox} \int_0^L (V_{gs} - V_{fb} - \psi_s(y))dy, \tag{2.38}$$

which requires to link the charge density to the variable y. This can be done be solving 2.41 consistently with the current continuity equation

$$I_D(L) = I_D(y). \tag{2.39}$$

Using (2.24) leads to a very convenient analytical expression for the total gate charge (Colinge), that is

$$Q_g = WLC_{ox} \left(V_{gs} - V_{fb} - \frac{(\psi_s(0) + \psi_s(L))}{2} + \frac{(\psi_s(0) - \psi_s(L))^2}{6(B - \psi_s(0) - \psi_s(L))} \right), \tag{2.40}$$

where $B = 2\left(V_{gs} - V_{fb} - \frac{Q_{bulk}}{C_{ox}} + 2\frac{k_B T}{q}\right)$. (2.40) contains all the charge inside the channel, however, depending on the bias point, its density distribution with respect the length, from the source to the drain changes. In particular, above the threshold, but below the saturation, the channel is symmetric allowing the *charge partitioning* of the charge in two equal contributions Q_S and Q_D, which are equal to $Q_g/2$, to respect the charge conservation. Above the saturation condition, the channel becomes asymmetric and the inversion charge near the drain shrinks, for this reason the charge must be divided in a certain ratio between the source and the drain. One effective approach is the one proposed by Ward and Dutton [1978] where the drain charge can be calculated as

$$Q_D = W \int_0^L \frac{y}{L}(Q_s(y) - Q_{bulk})dy = WC_{ox} \int_0^L \frac{y}{L} \left(V_{gs} - V_{fb} - \psi_s(y) - \frac{Q_{bulk}}{C_{ox}} \right) dy. \tag{2.41}$$

similarly to what done for the total gate charge,

$$\begin{aligned} Q_D = WLC_{ox}(& \frac{V_{gs} - V_{fb} - \frac{Q_{bulk}}{C_{ox}}}{2} - \frac{(\psi_s(0) + \psi_s(L))}{4} + \frac{(\psi_s(L) - \psi_s(0))^2}{60(B - \psi_s(L) - \psi_s(0))} \\ & + \frac{(5B - 4\psi_s(L) - 6\psi_s(0))(B - 2\psi_s(L))(\psi_s(0) - \psi_s(L))}{60(B - \psi_s(L) - \psi_s(0))^2}). \end{aligned}$$

To ensure charge conservation the the charge associated with the source is

$$Q_S = Qg - Q_D - Q_{bulk}. \tag{2.42}$$

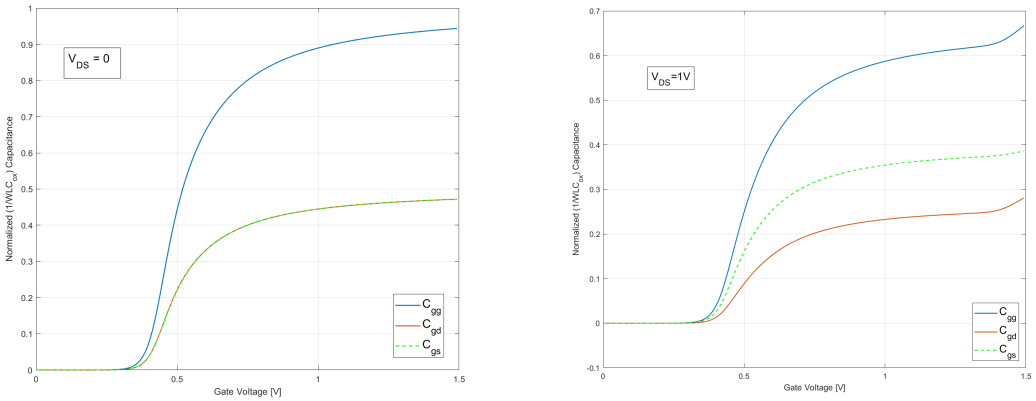
The total gate capacitance is

$$C_{gg} = \frac{\partial Q_g}{\partial V_{gs}} \tag{2.43}$$

which is, because of the charge partitioning, above threshold, composed by two terms

$$C_{gg} = C_{gs} + C_{gd} \tag{2.44}$$

where, at low V_{ds}, i.e. in linear region $C_{gs} = C_{gd} = \frac{1}{2}C_{gg}$ (figure 2.14a) whilst in saturation region, the two components are divided asymmetrically in a ratio about 60% for the C_{gs} and 40% for the C_{gd} as clearly visible in 2.14b



(a) Linear region.

(b) Saturation region.

Figure 2.14: Intrinsic Capacitance. $N_a = 1e16$ cm^{-3}, $t_{si} = 15$ nm, $t_{ox} = 2$ nm

With this approach the total capacitance is conserved.

On the other hands, below threshold the channel is not formed so C_{gs} and C_{gd} are equal to zero.

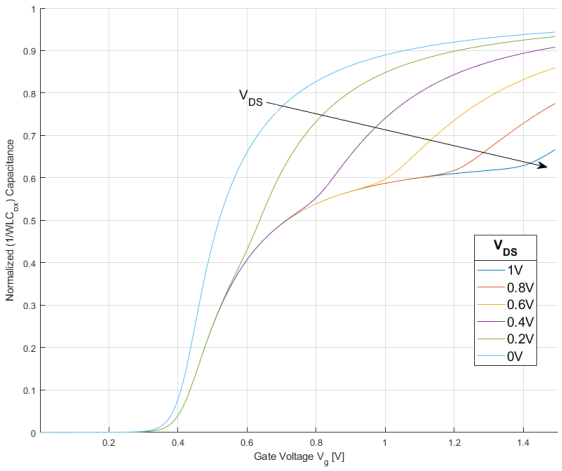


Figure 2.15: C_{gg} parametrized for V_{DS}. $N_a = 1e16$ cm^{-3}, $t_{si} = 15$ nm, $t_{ox} = 2$ nm

2.4 Double-Gate FinFET model MATLAB implementation

In the present work I have implemented a MATLAB code predicting the behavior of the presented double-gate transistor. The main structure is based by the *core model* which it is based on the calculation of the potential inside the channel. This is done by employing the calculation of the β and the gate bias equation presented in 2.1 using the MATLAB function *fzero* which is a useful way to find the roots of non-linear implicit equations when a proper interval is provided. In this case the interval used has been $(0, \frac{\pi}{2})$.
Many iterations of the gate bias equations has been performed for each V_G, V_{ch} couples contained into an array.
The potential obtained has been then derived analytically in order to keep computational time as small as possible to get the charge density useful to extract a specific current at a certain V_D and V_G.
Real device effects are included as additional potential terms in the gate bias equation. Each of them is calculated into a separate MATLAB function which can be excluded in the case a specific effect has to be excluded in the calculation by setting a variable in the main of the code.
More specifically the following variables have been included into the code which allows more flexibility

```
SAT = 1; % {0,1}  0: No Vel Sat   1: Yes Vel Sat
CLM = 0; % {*,1}  1: Channel Length Modulation Enabled
DIBL = 0; % {*,1}  1: Drain Induced Barrier Lowering
QME = 0; % {*,1}  1: Quantum Mechanical Effects Enabled
```

A specific function dedicated to capacitance calculation extracts a set of curves C_{gg}, C_{gd} and C_{gs} as function of the drain and gate voltages.
I also developed a stand-alone monolithic version for both NMOS and PMOS which can be useful for circuit simulators. In this regards few more toggles to reduce complexity has been included

```
MODE = 1;  % {1,2} 1: slow/accurate  2: fast
CAPMODE = 2; % {1,2} 1: accurate/unstable  2: approx./stable
```

The first variable gives the ability to choose between two different current calculation methods: a slow one which calculate 2.22 by numerical integration. This solution is the most accurate but the slowest. It gives best results when the substrate doping is $N_b > 1e16cm^{-3}$. When `MODE=2`, the analytical expression 2.24 is calculated.
These stand-alone scripts also provide the proper set of three capacitances C_{gg}, C_{gd} and C_{gs} for the specific bias provided by the user.

2.5 Double-Gate FinFET Validation and error evaluation

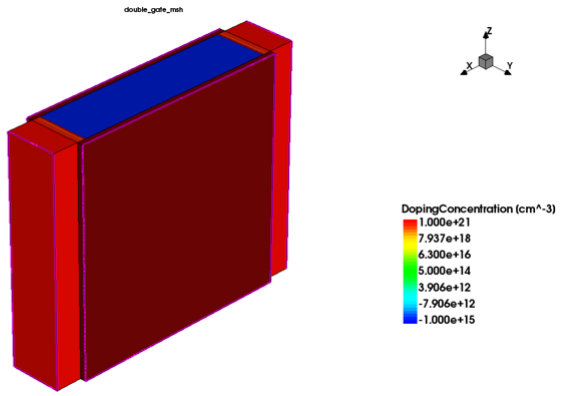
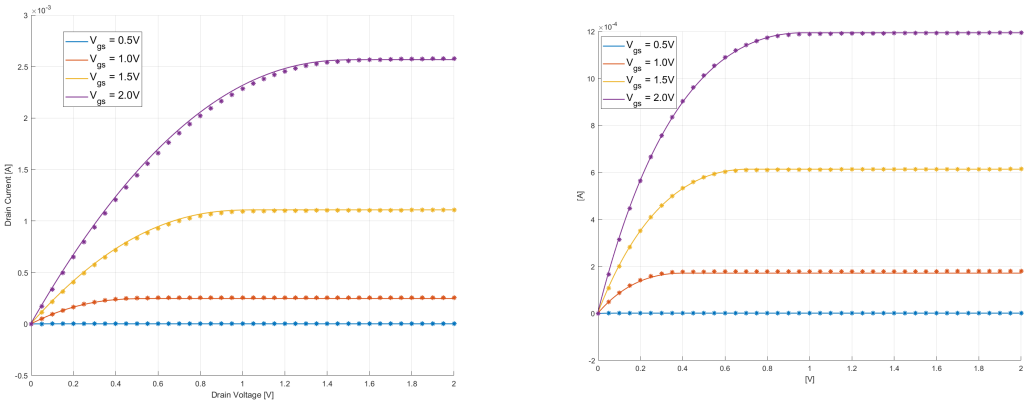


Figure 2.16: Short DG-FinFET generated in Sentaurus TCAD simulation tool

By employing the TCAD physical simulator I developed a double-gate FinFET structure from the ground up which can be used to validate the analytical model just developed. The drain current for long and short FinFET have been extracted as function of the drain and gate voltages. Short channel effects and velocity saturation has been considered.
No fitting factor have been used except for the mobility and saturation velocity which has been extracted from the TCAD model at the equilibrium condition.

.



(a) without vel. Saturation

(b) with vel. Saturation

Figure 2.17: FinFET Drain current, stars are TCAD extracted points. $L_g = 1$ μm, $W = 1$ μm, $N_a = 1e16$ cm^{-3}, $t_{si} = 5$ nm, $t_{ox} = 1.5$ nm

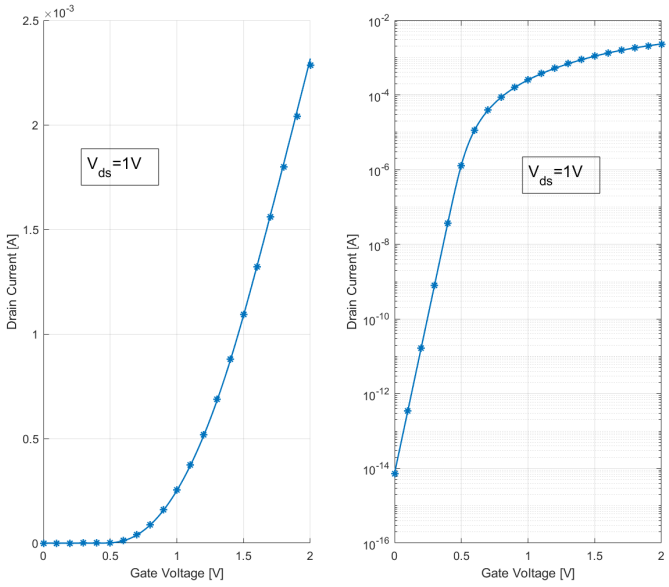


Figure 2.18: FinFET (w/o vel. Saturation) Transchar. Stars are TCAD extracted points. $L_g = 1$ μm, $W = 1$ μm, $N_a = 1e16$ cm^{-3}, $t_{si} = 5$ nm, $t_{ox} = 1.5$ nm

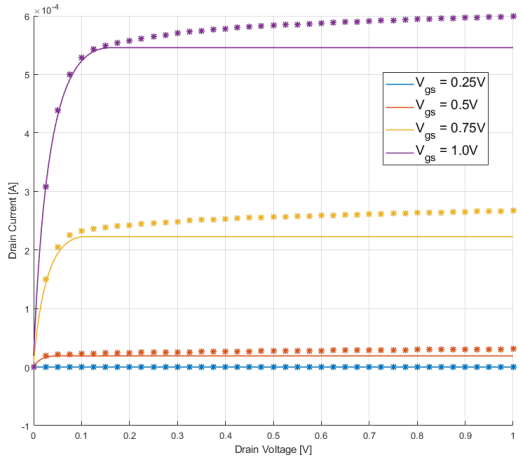


Figure 2.19: Velocity saturated FinFET Drain current, stars are TCAD extracted points. $L_g = 50$ nm, $W = 1$ μm, $N_a = 1e15$ cm^{-3}, $t_{si} = 10$ nm, $t_{ox} = 2$ nm

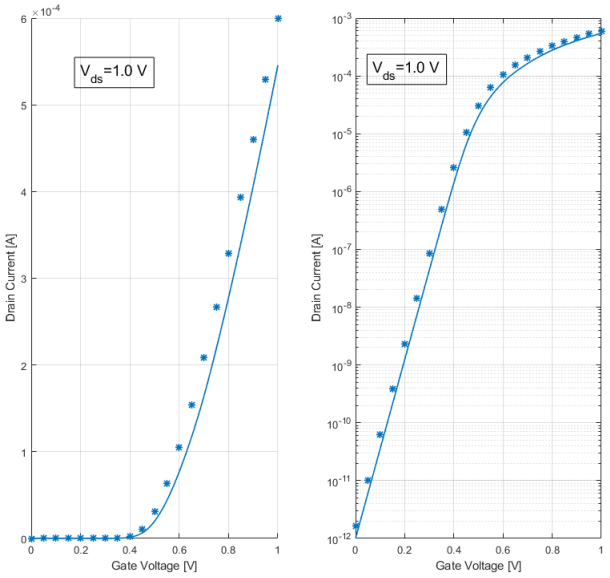


Figure 2.20: Velocity saturated FinFET Transchar. Stars are TCAD extracted points. $L_g = 50$ nm, $W = 1$ μm, $N_a = 1e15$ cm^{-3}, $t_{si} = 10$ nm, $t_{ox} = 2$ nm

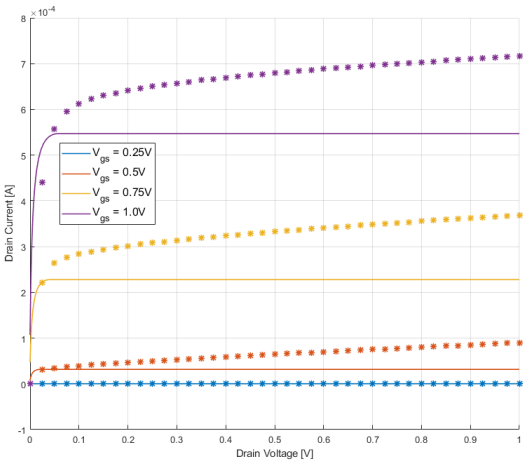


Figure 2.21: Velocity saturated FinFET Drain current, stars are TCAD extracted points. $L_g = 18\,\text{nm}$, $W = 1\ \mu m$, $N_a = 1e15\ cm^{-3}$, $t_{si} = 7\,\text{nm}$, $t_{ox} = 0.9\,\text{nm}$

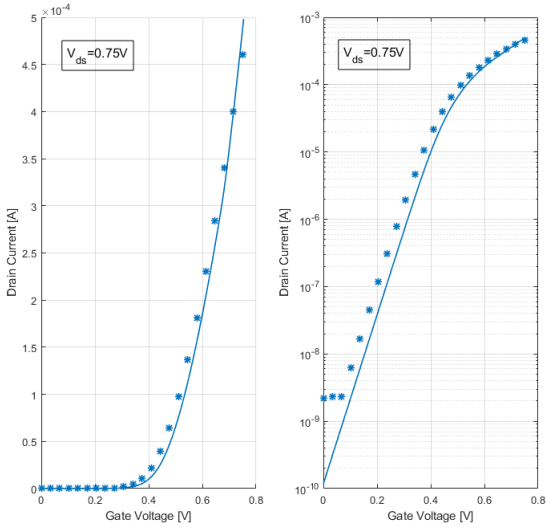


Figure 2.22: Velocity saturated FinFET Transchar. Stars are TCAD extracted points. $L_g = 18\,\text{nm}$, $W = 1\ \mu\text{m}$, $N_a = 1e15\ \text{cm}^{-3}$, $t_{si} = 7\,\text{nm}$, $t_{ox} = 0.9\,\text{nm}$

The model shows to be extremely accurate for the long channel case where SCEs are negligible. Scaling down the gate length the error between the TCAD and analytical models increases, especially in saturation region.
The quality of the model is still remarkable considering no fitting factor has been used which surely can contain the error.

In the following tables, the drain currents extracted from the TCAD simulation and from the model has been compared. In both the cases drain and gate voltage are forced to V_{DD} so the transistor is always in saturation region, that is the most interesting in digital applications.
The geometrical parameters are summarized here below and grouped in "nodes" from 50 nm to 13 nm. In all the situations the height of the fin has been taken equal to 0.5 μm so that $W = 1\,\mu$m and each current extracted can be considered as quantities per unit length with unit measure of $\frac{A}{\mu m}$. Each scaled node has been obtained using the constant field scaling approach with a scaling factor of $k = 1.4$.

Node's name	50 nm	35 nm	24 nm	18 nm	13 nm
L_g	50 nm	35 nm	24 nm	18 nm	13 nm
T_{ox}	2 nm	1.4 nm	1 nm	0.7 nm	0.5 nm
T_{si}	10 nm	7 nm	5 nm	3.5 nm	2.45 nm

Table 2.1: Geometrical parameters for the nodes under test.

Node's name	50 nm	35 nm	24 nm	18 nm	13 nm
$V_{DD} = 0.25V$	6.9e-9	4.4e-9	3.8e-9	2.5e-9	1.8e-9
$V_{DD} = 0.5V$	25.7e-6	27.5e-6	28.1e-6	23.7e-6	19.5e-6
$V_{DD} = 0.75$	274.3e-6	395.7e-6	520.1e-6	469.8e-6	418.8e-6

Table 2.2: FinFET TCAD I_D current per unit length $\left(\frac{A}{\mu m}\right)$ with $V_G = V_D = V_{DD}$.

In 2.4 is shown how the model provides good accuracy overall. In particular, can be noted how the relative error for the 50nm column is similar along from interdiction to saturation region maintaining the same sign, this suggests that a better fitting of the physical simulation can be obtained multiplying the output current to a constant. Below the 18nm node the error is more severe in saturation region where the electric field in the longitudinal direction intensifies the DIBL effect. This stresses out the importance of fitting parameters coming from experimental measurement.

Node's name	50 nm	35 nm	24 nm	18 nm	13 nm
$V_{DD} = 0.25V$	5.1e-9	4.3e-9	4.2e-9	2.7e-9	1.9e-9
$V_{DD} = 0.5V$	20.1e-6	25.9e-6	24.9e-6	22.3e-6	18.1e-6
$V_{DD} = 0.75$	228.7e-6	326.4e-6	454.2e-6	620.3e-6	733.9e-6

Table 2.3: FinFET compact model I_D current per unit length $\left(\frac{A}{\mu m}\right)$ with $V_G = V_D = V_{DD}$.

Node's name	50 nm	35 nm	24 nm	18 nm	13 nm
$V_{DD} = 0.25V$	-26.22%	-2.53%	9.8%	8.29%	3.67%
$V_{DD} = 0.5V$	-21.57%	-5.89%	-11.46%	-5.84%	-7.29%
$V_{DD} = 0.75$	-16.62%	-17.51%	-12.67%	32.04%	75.23%

Table 2.4: FinFET compact model I_D current relative errors with $V_G = V_D = V_{DD}$.

Chapter 3

Cylindrical Gate-all-around FET

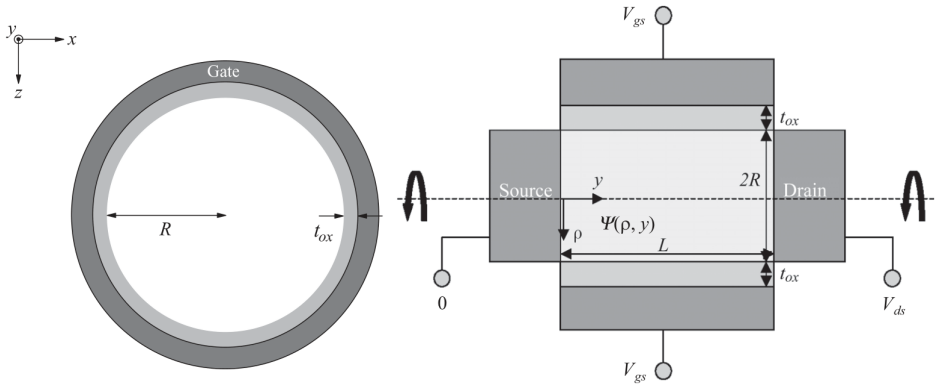


Figure 3.1: Cy-GAAFET. The device geometry is invariant for rotation around the $y-axis$.

In this case cylindrical section gate-all-around FET, also referred to as nanowire MOSFET, model is going to be derived. As already mentioned, this kind of structure allows an analytical solution, although not explicit, of the Poisson's equation, as well as for rectangular section DG-FinFET. The only limitation of the model present in literature Jimenez et al. [2004] is the necessity to consider an intrinsic substrate. To overcome it, in this work, a perturbation approach, similarly to what done in the DG-FinFET, is proposed which is based on the calculation of two potential contributions, considering only mobile carriers and one considering only doping charges; the two contribution will be summed up. The analysis starts with the derivation of the core model, based on Gradual Channel Approximation (GCA), and *Boltzmann* statistics for the carriers. Then real device effects will be added, trying to predict as well as possible the TCAD I-V characteristics which are going to be validated in the last section. The theory will treated taking NMOS as

reference.

In this work I developed a MATLAB code extracting I-V and C-V curves based on the nanowire MOSFET models presented here. The model is mainly based on the electrostatic studies of this structure already present in literature. From the numerical calculation of the potential with MATLAB I evaluate the charge inside the channel enforcing Gauss' law and consequently the current by integration along the length of the channel.
After that, I build up a TCAD 3D structure for the cylindrical gate-all-around FET and the I used to extract the output characteristics and the transcharacteristics. A comparison of these quantities is done in the last part of the chapter showing excellent results.

3.1 Core Model

Taking advantage of the cylindrical geometry, a Poisson's equation in cylindrical coordinates can be conveniently written. Assuming the *Boltzmann* distribution of mobile carriers and under GCA assumption

$$\frac{d^2\psi}{d^2\rho} = \frac{q}{\epsilon_{si}} n(\rho) + \frac{q}{\epsilon_{si}} N_a = \frac{q}{\epsilon_{si}} n_i e^{q(\psi - V(y))/k_B T} + \frac{q}{\epsilon_{si}} N_a \tag{3.1}$$

where ρ is the radius of the circular section. In order to calculate the potential inside the channel the perturbation approach can be employed as already done in the DG-FinFET. In this way the problem can be easily solved dividing the potential in two contributions, one for the mobile carriers, ψ_1, and one for the doping concentration N_a which is seen as a small perturbation of the potential profile inside the channel, ψ_2.

$$\psi(\rho) = \psi_1(\rho) + \psi_2(\rho) \tag{3.2}$$

$$\frac{d^2\psi_1}{d^2\rho} = \frac{q}{\epsilon_{si}} n_i e^{q(\psi - V(y))/k_B T} \tag{3.3}$$

$$\frac{d^2\psi_2}{d\rho^2} = \frac{q}{\epsilon_{si}} N_a \tag{3.4}$$

by integrating once the first term on the RHS of 3.1, under the boundary condition $\frac{d\psi_1}{d\rho} = 0$ which takes into account the symmetry on the channel section.

$$\frac{d\psi_1(\rho)}{d\rho} = 4\frac{k_B T}{q} \frac{(1-\beta)\left(\frac{\rho}{R}\right)}{(1-\beta)\left(\frac{\rho}{R}\right)^2 - 1} \tag{3.5}$$

Where β is a constant to be calculated from the *gate bias condition* (Jimenez et al. [2004])

$$V_g - V_{fb} - \Delta V_{th} = \frac{\epsilon_{si}}{C_{ox}} \frac{\partial\psi(\rho = R, y)}{\partial\rho} + \psi(\rho = R, y) \ , \tag{3.6}$$

where ΔV_{th} is the correction due to the short channel effects. From (3.5) under the boundary condition of $\psi_1(\rho = 0) = V_{ch}$

$$\psi_1(\rho) = V_{ch} + \frac{k_B T}{q} log\left(\frac{8N_a\epsilon_{si}k_B T(1-\beta)}{q^2 n_i^2 R^2[1-(1-\beta)(\rho/R)^2)]^2}\right). \tag{3.7}$$

Similarly, for the perturbation

$$\frac{d\psi_2(\rho)}{d\rho} = \frac{qN_a}{\epsilon_{si}}\rho \tag{3.8}$$

$$\psi_2 = \frac{qN_a}{2\epsilon_{si}}\rho^2. \tag{3.9}$$

The equation (3.6) can be solved including the perturbation term inside the derivation term. Being an implicit equation, numerically, using MATLAB and its function *fzero*, the *beta* term can be calculated.

Summing up (3.7) and (3.9) the total electrostatic potential profile as function of the radius and gate voltage can be traced

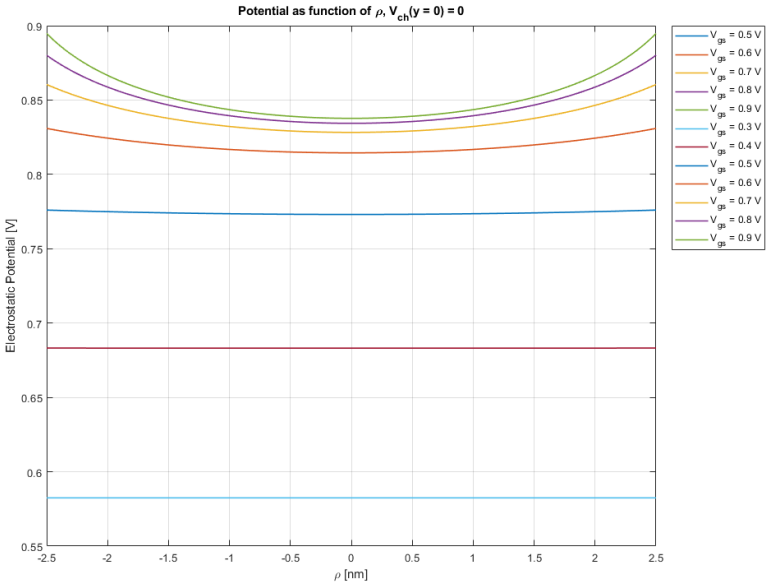


Figure 3.2: Electrostatic potential as function of V_{gs} and ρ. $N_a = 1e15$ cm^{-3}, $t_{ox} = 1.5$ nm ,$R = 2.5$ nm.

The electrostatic potential in 3.2 and 3.3 are calculated at the source end.
Fixing $\rho = R$ the surface potential can be calculated from the electrostatic potential and its dependencies from the geometrical and doping parameters.

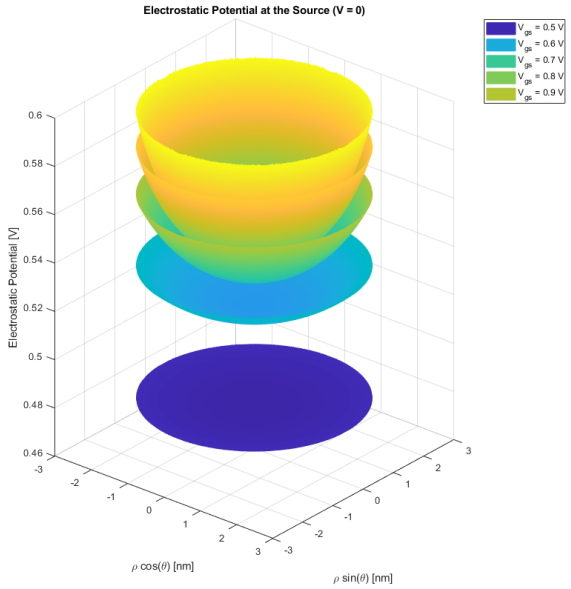


Figure 3.3: Electrostatic potential as function of V_{gs} and ρ in 3D. $N_a = 1e15\text{cm}^{-3}$, $t_{ox} = 1.5$ nm , $R = 2.5$ nm.

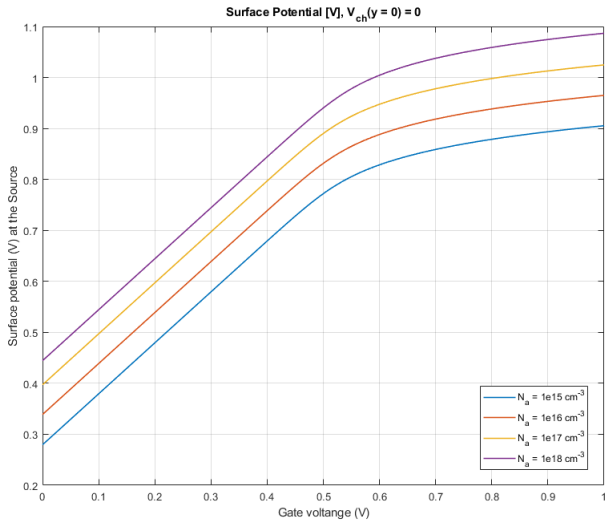


Figure 3.4: Surface potential near the source, $\psi_s(\rho = R\ V_{ch} = 0)$. $R = 2.5$ nm, $t_{ox} = 1.5$ nm.

The extraction of the potential inside the channel is fundamental to build a compact model with satisfactory accuracy.

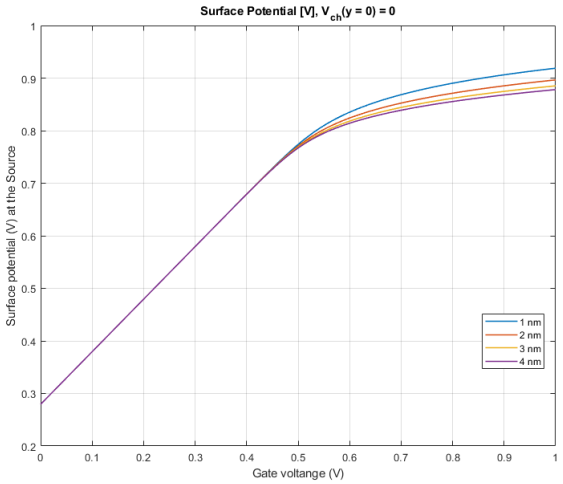


Figure 3.5: Surface potential near the source, $\psi_s(\rho = R\ V_{ch} = 0)$. $R = 2.5$ nm, $N_a = 1e^{15}$ cm^{-3}.

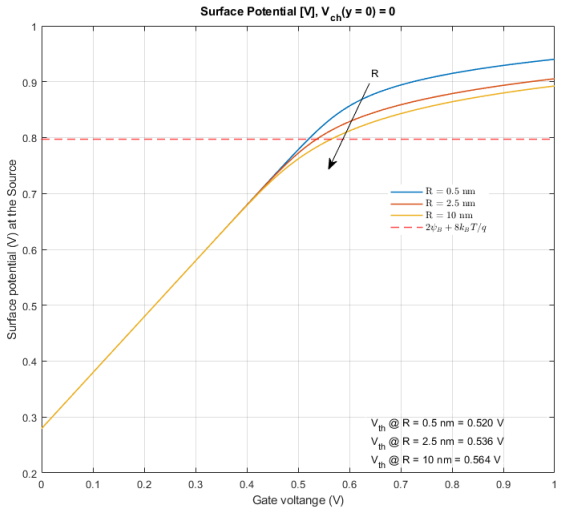


Figure 3.6: Surface potential near the source, $\psi_s(\rho = R\ V_{ch} = 0)$. $t_{ox} = 1.5$ nm, $N_a = 1e^{15}$ cm^{-3}.

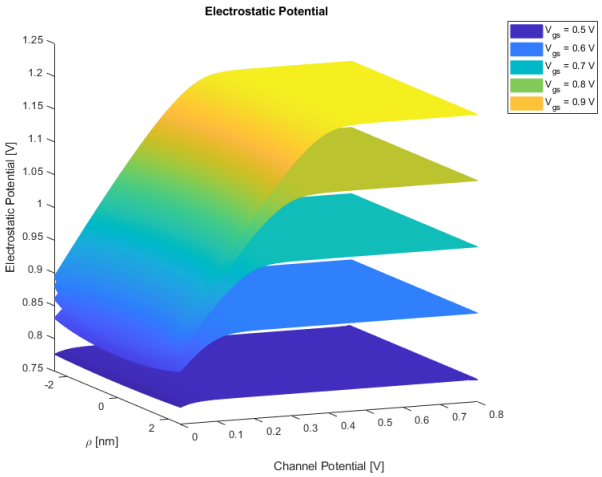


Figure 3.7: Electrostatic potential as function of V_{ch} and ρ. $R = 2.5$ nm, $t_{ox} = 1.5$ nm , $N_a = 1e^{15}$ cm^{-3}, $L_g = 1$ μm.

3.1.1 Charge density

In order to extract the charge density *Gauss's law* is applied

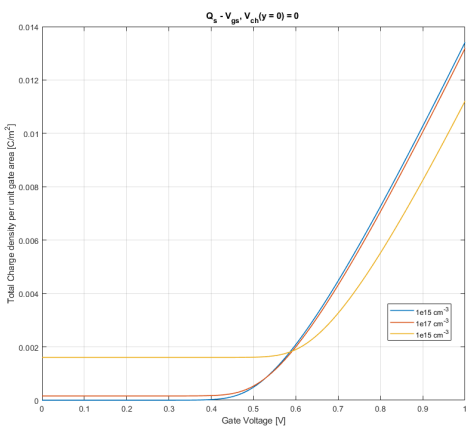
$$\left(\frac{\partial \psi}{\partial \rho}\right)_{\rho=R} - \left(\frac{\partial \psi}{\partial \rho}\right)_{\rho=0} = -\frac{Q_s}{\epsilon_{si}}, \tag{3.10}$$

$$Q_s(V_g, V_{ch}) = 4\epsilon_{si}\frac{k_B T}{q}\frac{(1-\beta)}{(1-\beta)-1} + qN_aR \tag{3.11}$$

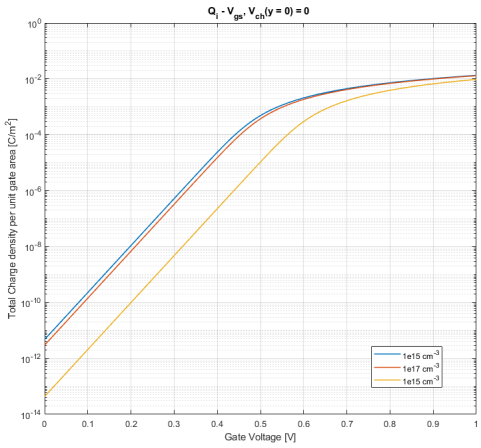
where the second term in (3.12) is related to the fixed charge due to ionized dopants $Q_{bulk} = qN_aR$. The first term instead is related to the mobile charges which are those contributing to the current

$$Q_i(V_g, V_{ch}) = Q_s - Q_{bulk} = 4\epsilon_{si}\frac{k_B T}{q}\frac{(1-\beta)}{(1-\beta)-1}. \tag{3.12}$$

.

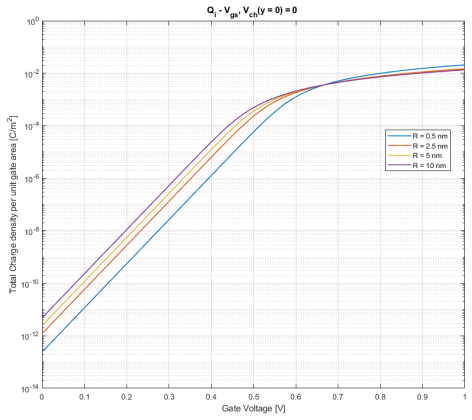


(a) Q_s in linear scale at $V_{ch} = 0$ in the Cy-GAAFET. $R = 10$ nm, $t_{ox} = 1$ nm.

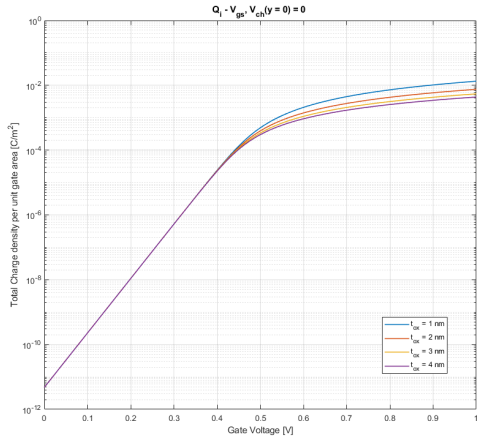


(b) Q_i in logarithmic scale at $V_{ch} = 0$ in the Cy-GAAFET. $R = 10$ nm, $t_{ox} = 1$ nm.

Figure 3.8: Q_i at $V_{ch} = 0$ in the Cy-GAAFET. $R = 10\ nm$, $t_{ox} = 1$ nm.



(a) Q_i in logarithmic scale at $V_{ch} = 0$ in the Cy-GAAFET. $N_a = 1e15$ cm^{-3}, $t_{ox} = 1$ nm.



(b) Q_i in logarithmic scale at $V_{ch} = 0$ in the Cy-GAAFET. $N_a = 1e15$ cm^{-3}, $R = 10$ nm.

Figure 3.9: Q_i in logarithmic scale at $V_{ch} = 0$ in the Cy-GAAFET. $R = 10$ nm, $t_{ox} = 1$ nm.

3.1.2 Drain current

.

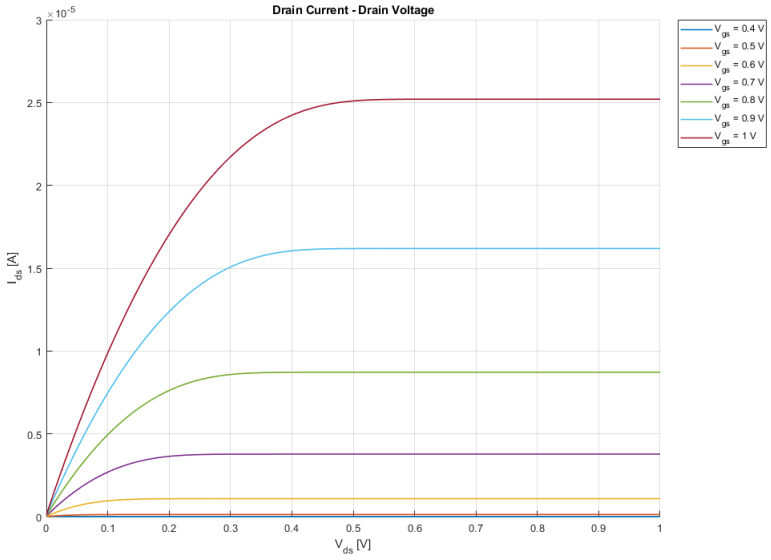


Figure 3.10: Output characteristics of the Cy-GAAFET. $N_a = 1e17$ cm^{-3}, $R = 10$ nm, $t_{ox} = 1$ nm, $L_g = 1$ μm, $\mu_n = 136$ cm^{-2}/(V $-$ s)

3.2 Real Device Effects

The approach I have used to include real device effects into the model is the same used in the FinFET which can be considered as a template for the GAAFET model. In fact, the main difference between the two models is in the set of equations leading to the electrostatic potential function in the channel.
The short channel effects are modeled as a corrective term ΔV_{th} to be included into

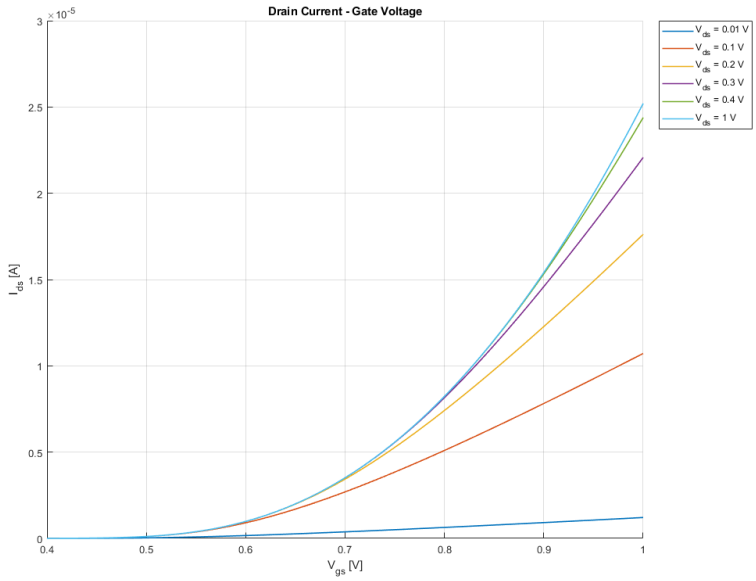


Figure 3.11: Transcharacteristics of the Cy-GAAFET. $N_a = 1e17$ cm^{-3}, $R = 10$ nm, $t_{ox} = 1$ nm, $L_g = 1$ μm, $\mu_n = 136$ cm^{-2}/(V − s)

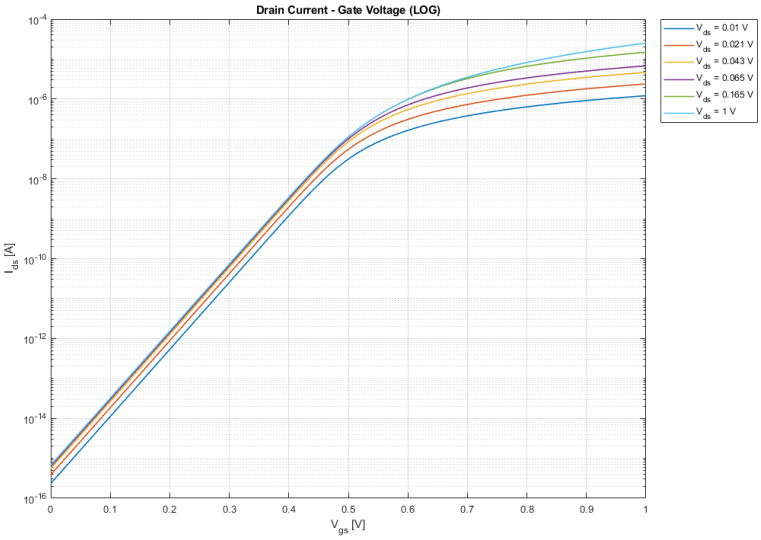


Figure 3.12: Transcharacteristics of the Cy-GAAFET in logarithmic scale. $N_a = 1e17$ cm^{-3}, $R = 10$ nm, $t_{ox} = 1$ nm, $L_g = 1$ μm, $\mu_n = 136$ cm^{-2}/(V − s)

the gate bias equation 3.6 which will modify the extracted electrostatic potential profile consequently, here the *scale length* λ term is different and from Y. Taur [2021]

$$\lambda = 1.3(R + t_{ox}) \tag{3.13}$$

Furthermore, the quantum contribution, following (2.36) is now

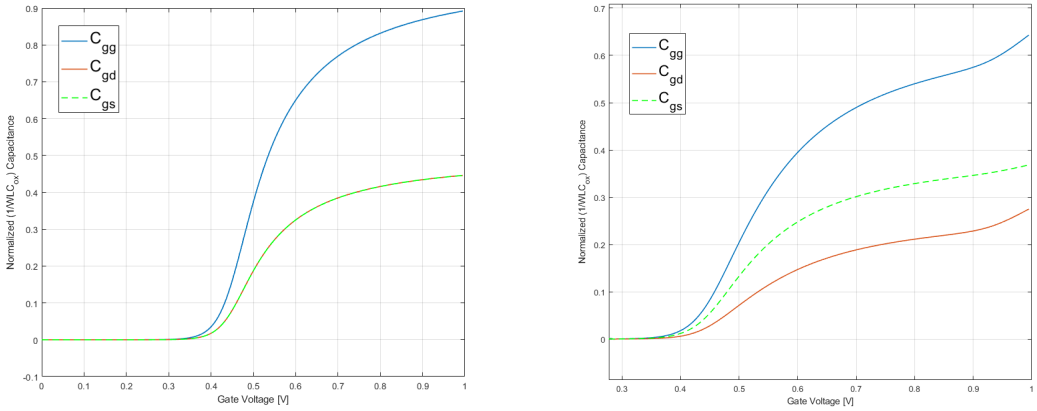
$$V_{thQM} = \frac{\hbar^2 \pi^2}{2qm^*(2R)^2} \tag{3.14}$$

Channel length modulation is taken into account as a subtracting term to the L_g, in the same way as in tradition MOSFETs. Velocity saturation, finally is modeled by modifying the mobility term as discussed in 2.2.3.

3.3 Capacitances

.

.



(a) Linear region.

(b) Saturation region.

Figure 3.13: Intrinsic Capacitance. $N_a = 1e15$ cm^{-3}, $R = 7.5$ nm, $t_{ox} = 2$ nm

3.4 cy-GAAFET model MATLAB implementation

The FinFET model implementation discussed in the previous chapter is here used as a template for the GAAFET model implementation. As well as the model discussed in 2.4, for the cylindrical GAAFET, I have implemented a NMOS and PMOS stand-alone

MATLAB codes. Many similarities in the MATLAB code structure can be found. Also in this case, a number of toggles to include or exclude specific real device effects can be found in the first lines of the codes.

```
MODE = 1;  % {1,2} 1: slow/accurate  2: fast
SAT = 1; % {0,1}  0: No Vel Sat   1: Yes Vel Sat
CLM = 0; % {*,1}  1: Channel Length Modulation Enabled
DIBL = 1; % {*,1}  1: Drain Induced Barrier Lowering
QME = 0; % {*,1}  1: Quantum Mechanical Effects Enabled
CAPMODE = 2; % {1,2} 1: accurate/unstable  2: approx./stable
```

Also in this case the user is free to choose the desired gate, drain and source external voltage and the algorithm is able to extract the drain current at that specific bias point and a set of three capacitances C_{gg}, C_{gd} and C_{gs}.

3.5 GAAFET Validation and error evaluation

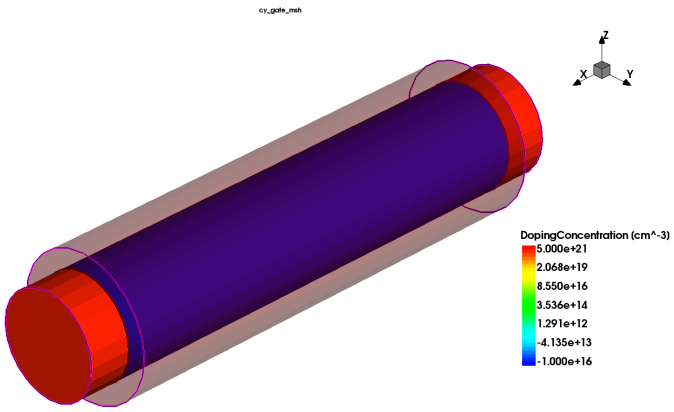
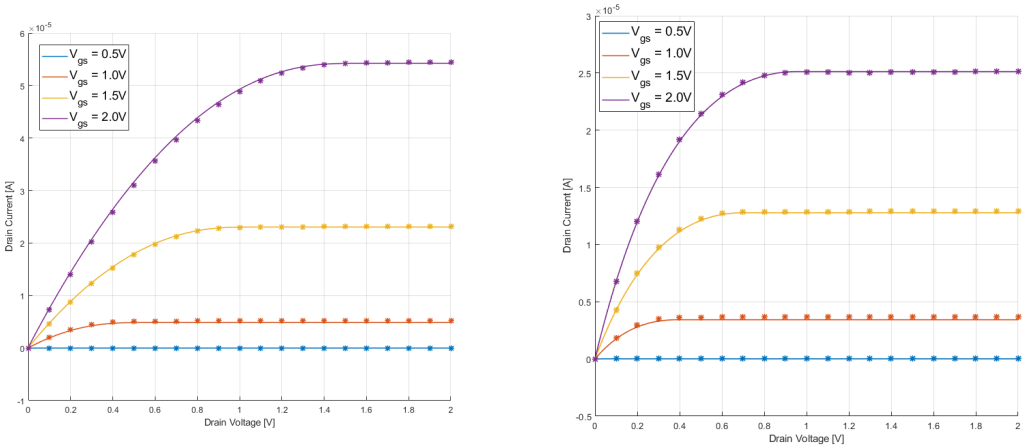


Figure 3.14: Short GAAFET generated in Sentaurus TCAD simulation tool

To validate the model a set of *Sentaurus Device TCAD* testbenches the cylindrical GAAFET are developed. A set of examples of comparison between TCAD physical simulations and model extraction are shown. The accuracy for very scaled nodes is very high, even for 18 nm node. This is because the cylindrical GAAFET, by its nature, is less prone to suffer short channel effects that are the main contributors to the overall error.



(a) without vel. Saturation

(b) with vel. Saturation

Figure 3.15: GAAFET Drain current, stars are TCAD extracted points. $L_g = 1$ μm, $R = 2.5$ nm, $N_a = 1e16$ cm^{-3}, $t_{ox} = 1.5$ nm

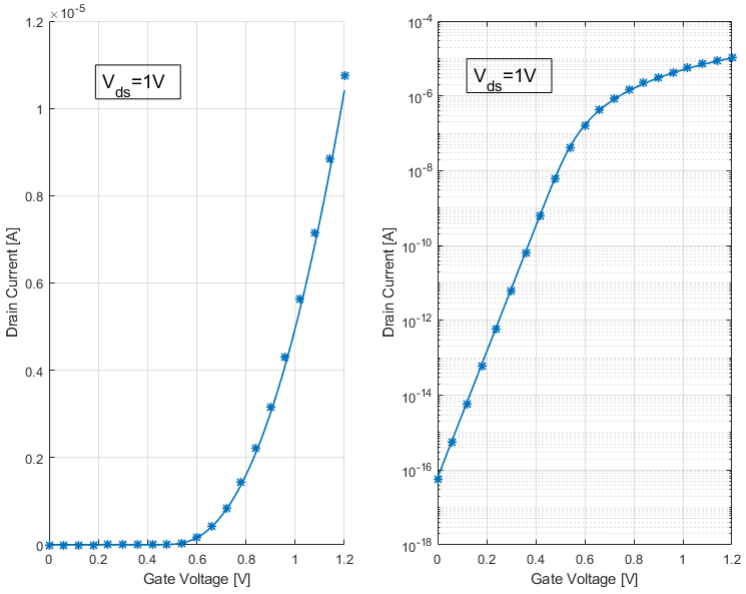


Figure 3.16: GAAFET Transchar. Stars are TCAD extracted points. $L_g = 1$ μm, $R = 2.5$ nm, $N_a = 1e16$ cm^{-3}, $t_{ox} = 1.5$ nm

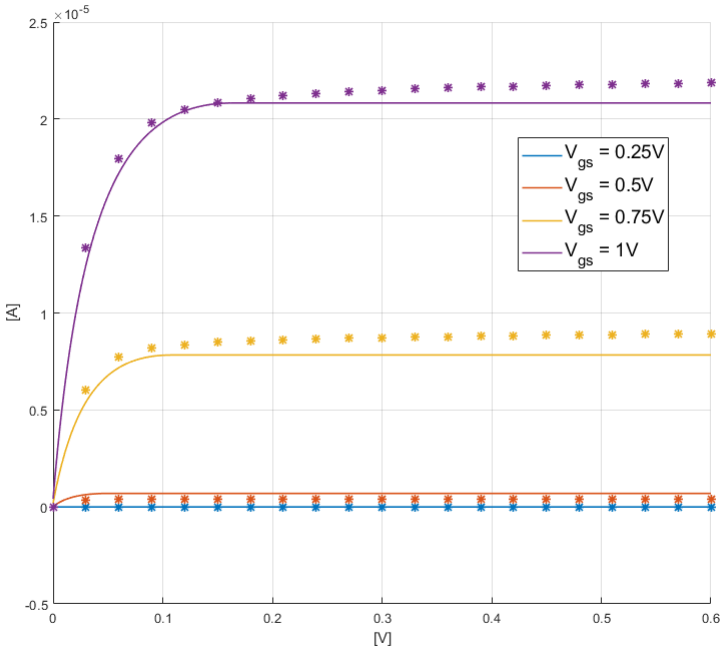


Figure 3.17: GAAFET Drain current with vel. Saturation, stars are TCAD extracted points. $L_g = 50$ nm, $R = 5$ nm, $N_a = 1e15$ cm^{-3}, $t_{ox} = 2$ nm

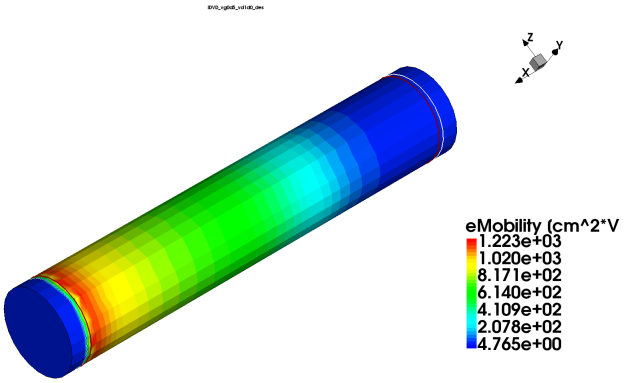


Figure 3.18: GAAFET electron mobility under velocity saturation. $L_g = 50$ nm, $R = 5$ nm, $N_a = 1e15$ cm^{-3}, $t_{ox} = 2$ nm, $V_{gs} = 0.5\,\mathrm{V}$

.

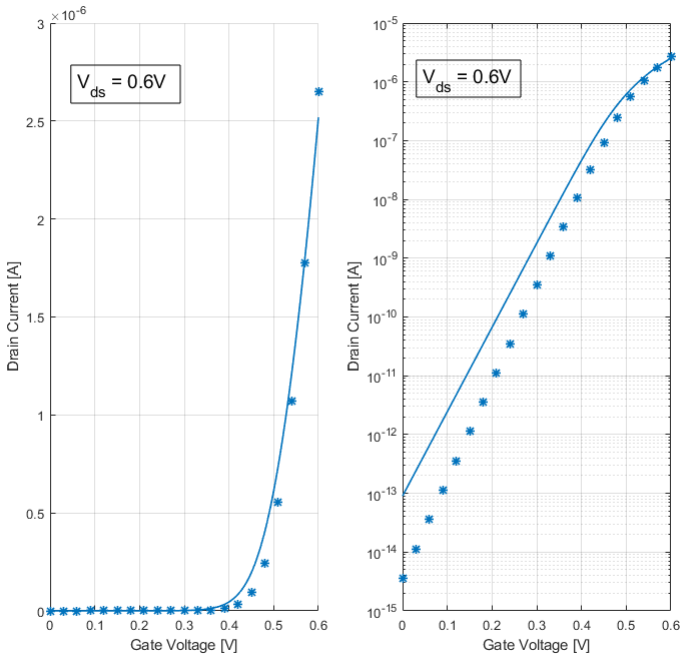
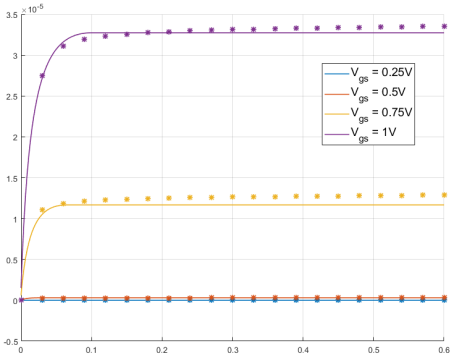
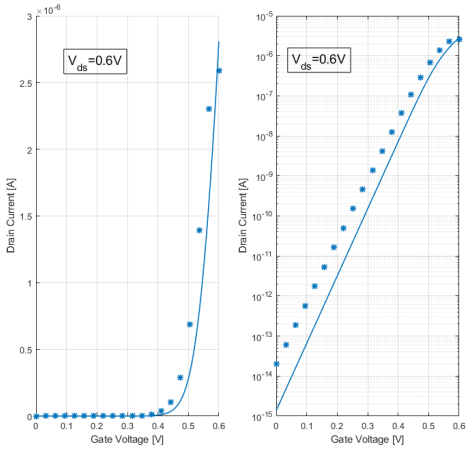


Figure 3.19: GAAFET Transchar. Stars are TCAD extracted points. $L_g = 50$ nm, $R = 5$ nm, $N_a = 1e15$ cm^{-3}, $t_{ox} = 2$ nm



(a) GAAFET Drain current



(b) GAAFET Transchar.

Figure 3.20: GAAFET current. Stars are TCAD extracted points. $L_g = 18$ nm, $R = 3.5$ nm, $N_a = 1e15$ cm^{-3}, $t_{ox} = 0.9$ nm

Node's name	50 nm	35 nm	24 nm	18 nm	13 nm
L_g	50 nm	35 nm	24 nm	18 nm	13 nm
R	5 nm	3.5 nm	2.5 nm	1.75 nm	1.225 nm
T_{ox}	2 nm	1.4 nm	1 nm	0.7 nm	0.5 nm

Table 3.1: Geometrical parameters for the nodes under test.

Node's name	50 nm	35 nm	24 nm	18 nm	13 nm
$V_{DD} = 0.25V$	1.7e-9	1.4e-9	1.1e-9	833.6e-12	628.6e-12
$V_{DD} = 0.5V$	14.7e-6	14.1e-6	13.1e-6	10.7e-6	8.6e-6
$V_{DD} = 0.75$	291.2e-6	272.7e-6	542.7e-6	700.0e-6	649.4e-6

Table 3.2: GAAFET TCAD I_D current per unit length $\left(\frac{A}{\mu m}\right)$ with $V_G = V_D = V_{DD}$.

In 3.2 and 3.3 tables, all the current have been normalized to the circumference of the specific device under test, i.e. $\phi = 2\pi R$, in order to have a quantity in unit measure of $(A/(\mu m))$ similarly to what done in the finFET case. Starting from node 50 nm the constant field scaling approach has been used with a scaling factor, again, of $k = 1.4$. Focusing on 3.4, model shows satisfactory results and good level of accuracy, especially considering that no fitting parameters have been used, with the exceptions of mobility at zero-bias condition and saturation velocity. The same considerations regarding the finFET surely do apply: fitting may be improved with a simple re-scaling of the currents by putting a multiplicative factor in front of the drain current related array in the MATLAB script, this can be seen as a scaling of the mobility term which can be fine tuned to reduce error, especially in longer nodes where the error looks to be more or less the same for different gate voltages.

Node's name	50 nm	35 nm	24 nm	18 nm	13 nm
$V_{DD} = 0.25V$	1.2e-9	1.1e-9	903.9e-9	717.7e-12	573.1e-12
$V_{DD} = 0.5V$	12.2e-6	12.5e-6	12.2e-6	10.6e-6	8.9e-6
$V_{DD} = 0.75$	258.1e-6	367.8e-6	511.1e-6	693.1e-6	923.6e-6

Table 3.3: GAAFET compact model I_D current per unit length $\left(\frac{A}{\mu m}\right)$ with $V_G = V_D = V_{DD}$.

Node's name	50 nm	35 nm	24 nm	18 nm	13 nm
$V_{DD} = 0.25V$	-27%	-22%	-20%	-14%	-9%
$V_{DD} = 0.5V$	-17%	-11%	-7%	-1%	4%
$V_{DD} = 0.75$	-11%	35%	-6%	-1%	42%

Table 3.4: GAAFET compact model I_D current relative errors with $V_G = V_D = V_{DD}$.

Chapter 4

Conclusions and future works

In this work FinFET and cy-GAAFET core models have been developed with the most relevant real device effects included as corrective parameters. The extracted quantities showed a good level of accuracy which can certainly be improved by including other secondary effects like gate-tunneling leakage, Band-to-Band tunneling at the drain level, nonuniform doping effects, strain effects and noise. Even a better modeling of the included ones would, for sure, reduce the relative error in saturation region, where the simulations here presented showed the worst accuracy overall, especially for high supply voltages. Also drain and source parassitic capacitances are missing since they strongly depend on the contact geometry and technology. Speed of the code is another critical point to be considered, especially in large circuits and can be certainly improved by code optimization.
In any case these MATLAB implementations might, I hope, be useful to students and researchers as a support for their work. Devices modeling is about constantly improving while the technology evolve and when the gate length enters in the single-digit nanometric domain a set of quantum mechanical effects start to dominate offering new challenges to researchers. The power of this models is indeed its modularity; around the backbone of the core model, based on simple electrostatics, many different corrections are included which characterizes, with different level of accuracy, many other effects. This gives a big margin of improvement.
The devices studied are highly symmetric and this symmetry has been exploited to get an analytical expression for the electrostatic potential. In asymmetric devices the extraction of the core model is way more challenging and in many cases cannot be possible at all. For this reason every advanced compact model for multi-gate transistors involve the use of a big set of fitting parameters which have to be extracted by experimental measurements.
The models implemented here can be used for system performance estimation based on simple RC models of CMOS gates which may give indications about the advantages and drawbacks of these technologies. In fact, they perfectly suite the TAMTAMS tool developed at the Politecnico di Torino where the finFET and GAAFET technologies could be compared with others from a system level point of view.

Bibliography

Spice overview. URL http://bwrcs.eecs.berkeley.edu/Classes/IcBook/SPICE/UserGuide/overview_fr.html.

International roadmap for devices and systems™ 2022 update more moore. 2022.

Massimo Alioto, Elio Consoli, and Gaetano Palumbo. Variations in nanometer cmos flip-flops: Part i—impact of process variations on timing. *IEEE Transactions on Circuits and Systems I: Regular Papers*, 62(8):2035–2043, 2015. doi: 10.1109/TCSI.2014.2366811.

Narain Arora. *MOSFET DC Model*, pages 230–324. Springer Vienna, Vienna, 1993. ISBN 978-3-7091-9247-4. doi: 10.1007/978-3-7091-9247-4_6. URL https://doi.org/10.1007/978-3-7091-9247-4_6.

Narain Arora. *MOSFET MODELING FOR VLSl SIMULATION Theory and Practice*. World Scientific, 2007.

Mark Bohr. A 30 year retrospective on dennard's mosfet scaling paper. *IEEE Solid-State Circuits Society Newsletter*, 12(1):11–13, 2007. doi: 10.1109/N-SSC.2007.4785534.

Mark T. Bohr and Ian A. Young. Cmos scaling trends and beyond. *IEEE Micro*, 37(6):20–29, 2017. doi: 10.1109/MM.2017.4241347.

D.M. Caughey and R.E. Thomas. Carrier mobilities in silicon empirically related to doping and field. *Proceedings of the IEEE*, 55(12):2192–2193, 1967. doi: 10.1109/PROC.1967.6123.

Y.S. Chauhan. *FinFET Modeling for IC Simulation and Design*. Elsevier, 2015.

J.P. Colinge. *FinFETs and Other Multi-Gate Transistors*. Springer.

R.H. Dennard, F.H. Gaensslen, Hwa-Nien Yu, V.L. Rideout, E. Bassous, and A.R. LeBlanc. Design of ion-implanted mosfet's with very small physical dimensions. *IEEE Journal of Solid-State Circuits*, 9(5):256–268, 1974. doi: 10.1109/JSSC.1974.1050511.

Juan Pablo Duarte, Sung-Jin Choi, Dong-Il Moon, Jae-Hyuk Ahn, Jee-Yeon Kim, Sungho Kim, and Yang-Kyu Choi. A universal core model for multiple-gate field-effect transistors. part i: Charge model. *IEEE Transactions on Electron Devices*, 60(2):840–847, 2013. doi: 10.1109/TED.2012.2233478.

Mohan Vamsi Dunga. *Nanoscale CMOS Modeling*. PhD thesis, EECS Department, University of California, Berkeley, Mar 2008. URL http://www2.eecs.berkeley.edu/Pubs/TechRpts/2008/EECS-2008-20.html.

Federico Faggin. The birth of the microprocessor. *IEEE Micro*, 41(6):16–19, 2021. doi: 10.1109/MM.2021.3112302.

David Chuyang Hong and Yuan Taur. An above threshold model for short-channel dg mosfets. *IEEE Transactions on Electron Devices*, 68(8):3734–3739, 2021. doi: 10.1109/TED.2021.3092310.

C.C. Hu. *Modern Semiconductor Devices for Integrated Circuits*. Prentice Hall, 2010.

C.-H. Jan, P. Bai, S. Biswas, M. Buehler, Z.-P. Chen, G. Curello, S. Gannavaram, W. Hafez, J. He, J. Hicks, U. Jalan, N. Lazo, J. Lin, N. Lindert, C. Litteken, M. Jones, M. Kang, K. Komeyli, A. Mezhiba, S. Naskar, S. Olson, J. Park, R. Parker, L. Pei, I. Post, N. Pradhan, C. Prasad, M. Prince, J. Rizk, G. Sacks, H. Tashiro, D. Towner, C. Tsai, Y. Wang, L. Yang, J.-Y. Yeh, J. Yip, and K. Mistry. A 45nm low power system-on-chip technology with dual gate (logic and i/o) high-k/metal gate strained silicon transistors. In *2008 IEEE International Electron Devices Meeting*, pages 1–4, 2008. doi: 10.1109/IEDM.2008.4796772.

C.-H. Jan, U. Bhattacharya, R. Brain, S.-J. Choi, G. Curello, G. Gupta, W. Hafez, M. Jang, M. Kang, K. Komeyli, T. Leo, N. Nidhi, L. Pan, J. Park, K. Phoa, A. Rahman, C. Staus, H. Tashiro, C. Tsai, P. Vandervoorn, L. Yang, J.-Y. Yeh, and P. Bai. A 22nm soc platform technology featuring 3-d tri-gate and high-k/metal gate, optimized for ultra low power, high performance and high density soc applications. In *2012 International Electron Devices Meeting*, pages 3.1.1–3.1.4, 2012. doi: 10.1109/IEDM.2012.6478969.

D. Jimenez, B. Iniguez, J. Sune, L.F. Marsal, J. Pallares, J. Roig, and D. Flores. Continuous analytic i-v model for surrounding-gate mosfets. *IEEE Electron Device Letters*, 25(8):571–573, 2004. doi: 10.1109/LED.2004.831902.

Tripty Kumari, Jawar Singh, and Pramod Kumar Tiwari. Impact of non-rectangular cross-section on electrical performances of gaa fets. In *2021 IEEE 18th India Council International Conference (INDICON)*, pages 1–6, 2021. doi: 10.1109/INDICON52576.2021.9691574.

G.E. Moore. Cramming more components onto integrated circuits. *Proceedings of the IEEE*, 86(1):82–85, 1998. doi: 10.1109/JPROC.1998.658762.

Gordon E. Moore. Cramming more components onto integrated circuits, reprinted from electronics, volume 38, number 8, april 19, 1965, pp.114 ff. *IEEE Solid-State Circuits Society Newsletter*, 11(3):33–35, 2006a. doi: 10.1109/N-SSC.2006.4785860.

Gordon E. Moore. Progress in digital integrated electronics [technical literaiture, copyright 1975 ieee. reprinted, with permission. technical digest. international electron devices meeting, ieee, 1975, pp. 11-13.]. *IEEE Solid-State Circuits Society Newsletter*, 11(3):36–37, 2006b. doi: 10.1109/N-SSC.2006.4804410.

Laurence W. Nagel and D.O. Pederson. Spice (simulation program with integrated circuit emphasis). Technical Report UCB/ERL M382, EECS Department, University of California, Berkeley, Apr 1973. URL http://www2.eecs.berkeley.edu/Pubs/TechRpts/1973/22871.html.

Robert N. Noyce. Semiconductor device-and-lead structure, reprint of u.s. patent 2,981,877 (issued april 25, 1961. filed july 30, 1959). *IEEE Solid-State Circuits Society Newsletter*, 12(2):34–40, 2007. doi: 10.1109/N-SSC.2007.4785577.

V.G.K. Reddi and C.T. Sah. Source to drain resistance beyond pinch-off in metal-oxide-semiconductor transistors (most). *IEEE Transactions on Electron Devices*, 12(3):139–141, 1965. doi: 10.1109/T-ED.1965.15469.

Samar Saha. *FinFET Devices for VLSI Circuits and Systems*. CRC Press, 2021.

Samar K. Saha. Compact models for integrated circuit design. 2016.

S.M. Sze. *Physics of Semiconductor Devices*. Wiley, 2007.

F.J. Garcıa Sánchez, A. Ortiz-Conde, and J. Muci. Understanding threshold voltage in undoped-body mosfets: An appraisal of various criteria. *Microelectronics Reliability*, 46(5):731–742, 2006. ISSN 0026-2714. doi: https://doi.org/10.1016/j.microrel.2005.07.116. URL https://www.sciencedirect.com/science/article/pii/S0026271405003033.

Y. Taur, C.H. Hsu, B. Wu, R. Kiehl, B. Davari, and G. Shahidi. Saturation transconductance of deep-submicron-channel mosfets. *Solid-State Electronics*, 36(8):1085–1087, 1993. ISSN 0038-1101. doi: https://doi.org/10.1016/0038-1101(93)90185-S. URL https://www.sciencedirect.com/science/article/pii/003811019390185S.

Yuan Taur. An analytical solution to a double-gate mosfet with undoped body. *IEEE Electron Device Letters*, 21(5):245–247, 2000. doi: 10.1109/55.841310.

Yuan Taur. *Modern VLSI devices*. Cambridge University Press, 2021.

Yuan Taur and Tak H. Ning. *Fundamentals of Modern VLSI Devices*. Cambridge University Press, 3 edition, 2021. doi: 10.1017/9781108847087.

Yuan Taur, Xiaoping Liang, Wei Wang, and Huaxin Lu. A continuous, analytic drain-current model for dg mosfets. *IEEE Electron Device Letters*, 25(2):107–109, 2004. doi: 10.1109/LED.2003.822661.

S. Thompson, N. Anand, M. Armstrong, C. Auth, B. Arcot, M. Alavi, P. Bai, J. Bielefeld, R. Bigwood, J. Brandenburg, M. Buehler, S. Cea, V. Chikarmane, C. Choi, R. Frankovic, T. Ghani, G. Glass, W. Han, T. Hoffmann, M. Hussein, P. Jacob, A. Jain, C. Jan, S. Joshi, C. Kenyon, J. Klaus, S. Klopcic, J. Luce, Z. Ma, B. Mcintyre, K. Mistry, A. Murthy, P. Nguyen, H. Pearson, T. Sandford, R. Schweinfurth, R. Shaheed, S. Sivakumar, M. Taylor, B. Tufts, C. Wallace, P. Wang, C. Weber, and M. Bohr. A 90 nm logic technology featuring 50 nm strained silicon channel transistors, 7 layers of

cu interconnects, low k ild, and 1 /spl mu/m/sup 2/ sram cell. In *Digest. International Electron Devices Meeting,*, pages 61–64, 2002. doi: 10.1109/IEDM.2002.1175779.

Scott E. Thompson. Mos scaling: Transistor challenges for the 21st century. 1998.

S. Venugopalan, Darsen D. Lu, Yukiya Kawakami, Peter M. Lee, Ali M. Niknejad, and Chenming Hu. Bsim-cg: A compact model of cylindrical/surround gate mosfet for circuit simulations. *Solid-State Electronics*, 67(1):79–89, 2012. ISSN 0038-1101. doi: https://doi.org/10.1016/j.sse.2011.09.001. URL https://www.sciencedirect.com/science/article/pii/S0038110111003194.

D.E. Ward and R.W. Dutton. A charge-oriented model for mos transistor capacitances. *IEEE Journal of Solid-State Circuits*, 13(5):703–708, 1978. doi: 10.1109/JSSC.1978.1051123.

N. Weste and D. Harris. *CMOS VLSI Design: A Circuits and Systems Perspective*. Pearson Education, 2011. ISBN 9780133001471. URL https://books.google.it/books?id=4SYrAAAAQBAJ.

H. Wong. On the cmos device downsizing, more moore, more than moore, and more-than-moore for more moore. In *2021 IEEE 32nd International Conference on Microelectronics (MIEL)*, pages 9–15, 2021. doi: 10.1109/MIEL52794.2021.9569101.

Tak H. Ning Y. Taur. *Fundamentals of Modern VLSI devices*. Cambridge University Press, 2021.