POLITECNICO DI TORINO

Corso di Laurea Magistrale
in Nanotecnologie per le ICT

Tesi di Laurea Magistrale

Vertical Junctionless Nano Transistor TCAD modeling and performance evaluation

Relatori
Prof. Mariagrazia Graziano
Prof. Gianluca Piccinini
Dr. Fabrizio Mo
Dr. Yuri Ardesi
firma dei relatori

. . . . . .
. . . . . .

Candidato
Alessandro Bugliarelli

firma del candidato

. . . . . .

Anno Accademico 2022 - 2023

Alla mia famiglia
A Francesca

Summary

This work presents a TCAD (Technology Computer-Aided Design) calibration procedure specifically tailored for a vertical nanowire Gate-All-Around junctionless transistor developed at the LAAS-CNRS research facility in Toulouse. The junctionless transistor is an emerging device architecture that offers potential advantages over traditional junction-based transistors in terms of process simplicity, reduced fabrication cost, and improved performance characteristics. However, accurately modeling and simulating junctionless transistors using TCAD tools require a careful calibration process due to the unique device physics and material properties involved.

The proposed calibration procedure aims to optimize the TCAD model parameters to accurately represent the electrical behavior of junctionless transistors. It involves a systematic methodology that combines experimental data and simulation results to iteratively refine the model parameters. The calibration process takes into account various key factors, including the device geometry, material properties and interface effects, to ensure a comprehensive and accurate representation of the device behavior. To validate the effectiveness of the calibration procedure, a set of experimental data is used for comparison with the calibrated TCAD simulation results. The comparison involves various electrical characteristics, such as current-voltage (I-V) curves, subthreshold slope, threshold voltage, and other parameters specific to the inspected mechanisms. The calibrated TCAD models demonstrate reasonable agreement with the experimental data, indicating the accuracy and reliability of the proposed calibration procedure.

In the final part of this work, the calibrated TCAD model is validated through a full TCAD simulation of an inverter. Subsequently, the inverter logic cell is used to build and simulate the behavior of a ring oscillator circuit, consisting of multiple junctionless transistors interconnected in a feedback loop. This allows to investigate the frequency response of the junctionless transistor and to get a clear view of the overall performance of the studied device.

The calibrated TCAD model for junctionless transistors can serve as a valuable tool for device engineers and researchers to study and optimize the performance of novel JLNT-based architectures. It enables efficient design exploration, process optimization, and performance prediction, leading to improved device performance and enhanced manufacturing yield. Furthermore, the calibration procedure presented in this work could be extended to other emerging device architectures, facilitating their integration into mainstream semiconductor technologies.

Acknowledgements

I would like to thank all the people that accompanied me throughout this journey. A special thank you to my supervisors from LAAS-CNRS, Guilhem Larrieu and Anne Hemerick, whose guidance has been essential to me, and to Oskar Baumgartner and all the other GTS engineers I worked with for all the knowledge and the precious and inspiring insights shared with me. Another big thank you to my supervisors at PoliTo, Mariagrazia Graziano, Gianluca Piccinini, Fabrizio Mo and Yuri Ardesi, who helped me in further developing this work with great interest and passion.

Contents

List of Figures

Our technology, our machines,
is part of our humanity.
We created them to extend ourselves, and
that is what is unique about human beings.
[Ray Kurzweil]

Part I

Background

Chapter 1

Introduction

1.1 Outline

This work presents a TCAD model calibration process based on the electrical measurements of vertical junctionless nanowire nanotransistors fabricated at the LAAS-CNRS research center of Toulouse. In the first part of the work, some background knowledge about the technology is disclosed, then the methodology is illustrated and finally the electrical measurements of the studied devices are presented. In the second part, the calibration workflow is illustrated step by step and finally, the obtained calibrated TCAD model is validated through the simulation of an inverter and a ring oscillator.

1.2 Transistor scaling

Since the earliest years of integrated electronics, the scaling of transistors has been a fundamental driving force behind the rapid progress of the semiconductor industry. The concept of scaling basically involves reducing the size of transistors, thereby increasing the number of transistors that can be integrated into a chip, enabling designers to achieve higher performance and functionality. Specifically, transistor scaling leads to reduced channel widths and shorter gate lengths, thus enabling faster switching times, higher clock speeds and improved overall performance of electronics devices. Moreover, smaller devices require less power to switch on and off, resulting in reduced energy consumption. The principles of transistor scaling are summarized in the well known *Moore's Law*. This principle, named after Gordon E. Moore, co-founder of Intel Corporation, recites that the number of transistors per unit area in an integrated circuit doubles approximately every two years, leading to a rapid increase in computing power [3]. Thus, according to Moore's law, in order to obtain a device which is 50% smaller than the previous generation it is required to shrink the width and the length to 70% of the original size.

However, the scaled devices must meet some well defined requirements, such as gate length, oxide thickness, doping, threshold voltage or depletion region extension, to maintain the compatibility with the desired time delay and power consumption.
When scaling a device, it is necessary to take into account the bias voltage. In fact,

shrinking a device while keeping an unvaried bias would sensibly increase the electric fields, leading to several effects that designers usually want to avoid. Among said adverse effects we find:

- electromigration i.e., a momentum exchange between electrons and metal ions that takes place in the device's metal lines when they are traversed by a high current and leads to the migration of metallic ions in the direction of electron flow. It can lead to the degradation and eventual failure of the conductor, causing disruptions in the functionality and reliability of electronic devices;
- dielectric barrier breakdown due to gate oxide failure caused indeed by the higher electric fields;
- increased power density dissipation, which may cause thermal problems.

In order to avoid all these problems, it is necessary to scale the bias voltage accordingly. A widely used scaling principle is to keep the electric field constant, commonly known as Dennard's scaling [4]. Dennard's scaling is a principle that states that the as transistors get smaller, their power density stays constant. This implies that the device area and the power consumption scale in the same manner.
If, for example, one wants to scale a device by a scaling factor κ while keeping the electric field constant, will need to scale the relevant design parameters accordingly such as, for instance, oxide thickness $t_{ox} \rightarrow t_{ox}/\kappa$, channel length $L \rightarrow L/\kappa$ and bias voltage $V_{DD} \rightarrow V_{DD}/\kappa$. However, as transistor sizes approached the atomic scale and physical limitations emerged, Dennard's scaling began to face challenges. The leakage currents and power dissipation of transistors increased as their sizes decreased further. This led to heat management issues and limitations on the maximum clock frequencies achievable. Consequently, the traditional interpretation of Dennard's scaling began to break down, necessitating the adoption of new strategies to continue improving chip performance. Technological advancements, such as the introduction of FinFET (Fin Field-Effect Transistor) designs and other innovations, have helped address some of these challenges and allowed for further improvements in power efficiency and performance [5].

1.3 The Junctionless Nano Transistor (JLNT)

A junctionless nano transistor (JLNT) is a novel typology of field effect transistor that operates without traditional p-n junctions found in conventional transistors. It is a nanoscale device that can be used to control the flow of current in electronic circuits. In a junctionless nano transistor, the channel region is uniformly doped, unlike traditional transistors that have distinct regions of different doping types (p-type and n-type) to create junctions. The absence of junctions simplifies the fabrication process and allows for better control of the transistor's electrical properties.
The operation of a junctionless nano transistor relies on the modulation of the channel's conductivity through the application of a gate voltage. By applying a positive or negative voltage to the gate terminal, the transistor can be switched between an ON state (conducting current) and an OFF state (blocking current). The gate voltage controls

the electrostatic potential across the channel, thereby regulating the flow of carriers and determining the transistor's behavior. This working principle can be intuitively seen as a garden hose, representing the conduction channel which may be, for instance, a silicon nanowire, tightened by a hand, that represents the electrostatic control exerted by the gate and thus controls the flow of charge carriers inside the nanowire. When the device is turned on is in accumulation state, thus with $V_{th} > 0$ for the nMOS and $V_{th} < 0$ for the pMOS. The majority carrier concentration inside the nanowire grows and the electron/holes can flow. Conversely, when the device is off we find it in depletion mode: the voltage applied to the gate causes a depletion of majority carrier inside the nanowire, thus interrupting the conduction channel. Many differences can be seen between this device and a traditional planar MOSFET: the conduction happens thanks to the majority carriers and there is no need of forming an inversion channel below the gate, which means that if we want to obtain a pMOS we need to use p-doped Silicon for building the nanowire.

The junctionless design gives more versatility because it may eliminate the issues related to the need of abrupt doping profile at the source and drain regions [6], simplifying the fabrication process. Instead, the metallic contacts are obtained directly at the ends of the conduction channel. The junctionless design also helps to get rid of the parasitic junction capacitances that are formed in the planar device, between the highly-doped source-drain regions and the body. Among the advantages of the junctionless nano transistors we also find better electrostatic control, improved immunity to short-channel effects, and potentially higher current drive capabilities. These characteristics make them promising candidates for future nanoelectronic devices.

The first implementation of junctionless nanowire transistor is found in the work of Colinge et al. [7]: this article discusses the properties and design guidelines for junctionless nanowire transistors, including the effects of doping, channel length, and temperature. The JLNT exploits bulk conduction instead of surface channel conduction, thus the current drive is controlled by doping concentration and not by gate capacitance. Many of the fundamental aspects of Colinge's JLNT covered in the paper can also be found in the corresponding patent [8].

It's worth noting that while junctionless nano transistors have shown great potential in research and development, they are not yet widely used in commercial applications. However, ongoing advancements in nanotechnology and semiconductor manufacturing techniques may eventually lead to their practical implementation in various electronic devices.

1.4 The FVLLMONTI Project

FVLLMONTI is an acronym that stands for *Ferroelectric Vertical Low energy Low latency low volume Modules fOr Neural network Transformers In 3D* [9]. It is an European founded research project, part of the EU's Horizon 2020 research and innovation program. The goal of the project is to develop a lightweight in-ear device allowing instantaneous speech-to-speech translation, which takes place inside the device, thus completely offline. Choosing a *computing at the edge* approach allows to keep the computation and data storage of the system close to the source of data. In practice, this allows to have an implementation that is able to run all the logic locally, without the need to connect to a server to run computations. The FVLLMONTI project takes all the advantages of edge computing: the chosen design focuses on energy efficiency and compactness, since avoiding the transmission of data to a remote server allows to save a significant amount of energy.

One of the main challenges of the project is thus to build a suitable hardware platform for running the transformer neural network. Here a crucial role is played by the underlying technology: the project rests its foundations on an innovative silicon design, ditching the traditional 2D integration in favor of a stacked 3D architecture enabled by the **Vertical Nanowire Gate-All-Around FET**. The neuromorphic computing is an application that can fully take advantage of the 3D architecture, as the increased density and processing power of the proposed design significantly increase the performance per area figure of merit, making it particularly suited for embedded AI applications. Moreover, switching to a 3D architecture represents a further step in terms of neuromorphic computing, since the end is to have a design inspired to the human brain, which is an intrinsically 3D structure, with lots of interconnects between the neurons. One another crucial point in the technology is to have *logic in memory* capabilities, a necessary feature to improve the efficiency of AI application. To this end, the transistors will use a ferroelectric material as gate insulator.
Several European universities, research centers and companies are working jointly on the project. The overall organization presents the 5 main *work packages (WPs)*:

- WP1 Fabrication - LAAS-CNRS
- WP2 Technology parameter's extraction - Université de Bordeaux
- WP3 3D layout tool - Global TCAD Solutions GmbH
- WP4 N2C2 cell design - École Centrale de Lyon
- WP5 NN Transformer architecture - École Polytechnique Fédérale de Lausanne

The work done in LAAS is focused on the fabrication and the development of the JLNT, the integration of the ferroelectric material (HfO_2), done in collaboration with the company NaMLab gGmbH, and the modeling of the fabricated devices, done in collaboration with Global TCAD Solutions GmbH, which is what will be covered in this report. Some of the work packages require a tight cooperation between the different actors taking part in the project. Specifically, a continuous feedback between the fabrication team (LAAS-CNRS) and the layout team (Global TCAD Solutions GmbH) would be necessary in

order to enhance the design procedure thanks to the in-depth knowledge of the transistors physical behavior. This feedback loop is commonly known as Device Technology Co-Optimization (DTCO) loop, a modern workflow necessary for the development of the hardware platform that consists in a continuous synergy between the technology and design teams, which can give useful feedback to the other in order to enhance the device. The DTCO workflow will be presented in a more detailed way in the following section, as the work presented is part of the loop.

In this context, the final purpose of this work would be providing a reliable device model, that is able to faithfully predict the behavior of the fabricated devices, in order to empower the electronic design automation (EDA) of the more complex layouts based on such technology.

Chapter 2

Metodology

2.1 The Design-Technology Co-Optimization (DTCO) workflow

Design Technology Co-Optimization (DTCO) is a methodology that helps semiconductor fabs reduce costs and time-to-market in advanced process development. It consists in optimizing together design and process technology in order to improve performance, power efficiency, transistor density and cost. DTCO for a new technology node usually involves substantial architectural innovation instead of just delivering the exact same structure as the previous generation, only smaller [10]. In the last decade, DTCO has become more and more relevant for the development of new process nodes, as the Moore's law and the Dennard's scaling law [4] are coming close to an ending, since decreasing device dimensions of advanced technologies inherently lead to increased device variability.

In a typical DTCO workflow, first the structure of the device is created starting from the layout and the technological specifications. Once the cell is created, the full 3D TCAD model of the device can be used to run steady-state, transient, thermal and AC analyses, allowing for the extraction of several *key performance indicators* (KPI), such as characteristics, power/performance indicators, self-heating properties and parasitic resistances and capacitances [11].
To have a clearer view on the DTCO workflow, let us take a look at an example: a variability-aware DTCO flow, applied to the N3 FinFET and Nanosheet technologies for a 6-track (6T) SRAM [12]. In the presented work, the DTCO simulation flow uses accurate extraction of resistances, capacitances and device characteristics of 3D TCAD, enabling SPICE simulations incorporating variability and reliability for ring-oscillator (RO) cells, both at initial conditions and after 10 years of AC operation. One very interesting detail to notice is the adoption of SPICE models: by using the SPICE model for the simulation of the complete device, instead of using the full TCAD simulation, the workflow becomes significantly faster, since the full TCAD approach is very computationally expensive. After the 3D structures are generated based on a technology specification, the device characteristics of structures with single MOSFET's are simulated and a fast TCAD

PEX (parasitics extraction) is used to determine the netlist (including all resistances, capacitances and transistors) of both the single MOSFET structures and the full TCAD cell under investigation. The SPICE model cards for transistors in the netlist of the full cell are then evaluated. The variability is simulated by taking into account *random discrete dopants* (RDD), *metal gate granularity* (MGG) and discrete oxide defects, while for the reliability, the *bias-temperature instability* (BTI) is simulated. Based on the output of the simulations, one can determine which are the most significant factors that affect the performance of the device, and can use the gathered information for improving the layout and the technology specification. In particular, in the cited example the following results are obtained: concerning the FET variability, it is seen that the MGG is the main variability cause in the FinFET, while the geometrical variability is more relevant in the NS technology. Moreover, the NS device shows a better subthreshold slope than the FinFET. Concerning the SRAM, the variability issues seen on the FET's subsequently affect the performance of the SRAM: a large MGG causes an increase of the variability of all the key performance indicators (KPI) of the memory, both in the FinFET and the NS devices. The studied KPIs for the SRAM are *SNM (static noise margin) read*, *WTP (write-trip-points* and SRAM read current.

2.2 TCAD softwares

In this section the two TCAD softwares used for doing the simulation work will be briefly presented. It is useful to recall that the first part of this work, done during my internship period at LAAS-CNRS, has been completely done using Global TCAD Solutions GTS Framework, whereas the second part, done after the end of my internship period at PoliTo, has been done using Synopsys Sentaurus.

2.2.1 Global TCAD Solutions GTS Framework

GTS Framework [13] is a work environment for Technology CAD (TCAD) applications, such as semiconductor device simulation. Designed as a framework, its purpose is to embed various tools into a common workspace in order to facilitate ease of use, improve workflow, support structured data storage, and provide a consistent graphical user interface for all the embedded programs. The tools rely on basic functionalities provided by the framework (such as file i/o, network support, and a consistent user interface), plus a set of features for improved workflow, such as file-management and project storage. For tools running in Framework mode, GTS Framework provides simple features which ease common workflow scenarios, such as using multiple tools one after another, and keeping revisions and variants of simulation data. The main components of the concept are *ToolFolders* and the *NextTool* button. A ToolFolder is the working-space for each tool instance, it is like a private working-directory with extended features. The intention of ToolFolders is to record the steps of the workflow: for example, for creating a device and then simulate it, we need first to create a GTS Structure ToolFolder, and then a MinimosNT ToolFolder. In case we need to come back to the GTS Structure tool at a later time, to create a variant of your device, we can open the ToolFolder again: the Tool setup and data will be as we have left it before. Thus, a ToolFolder represents one step in the workflow and it holds the setup and data of one tool invocation. The NextTool button is the complement to the ToolFolders, the bridge that takes from one step to the next, i.e. from one ToolFolder to the next ToolFolder. Technically, the NextTool button creates a new ToolFolder, and copies the output file of the previous ToolFolder to the new ToolFolder, and uses it as input file. This speeds up the workflow and relieves the user of worrying about file names and paths.

Structure and Layout ToolFolder

GTS Structure is an intuitive tool for definition and modification of device structures. Devices can either be designed from scratch using the built-in 2D/3D CAD editor, can be imported from other TCAD formats, or can be created based on parameterized templates. Simulation quantities such as dopings or material definitions can be defined by analytical functions in the CAD editor, or can be interpolated from measurement data. Furthermore, GTS Structure provides an intuitive meshing engine for creation of structured as well as unstructured meshes in two and three dimensions. GTS Structure includes a 2D/3D device view, which instantly displays selected segments and quantities such as dopands. The Layout tool enables the layout-based structure generation, thus it is capable of creating a

structure starting from a GDSII mask layout file. The screenshots shown below in Figure 2.1 show the initial view of said ToolFolders.

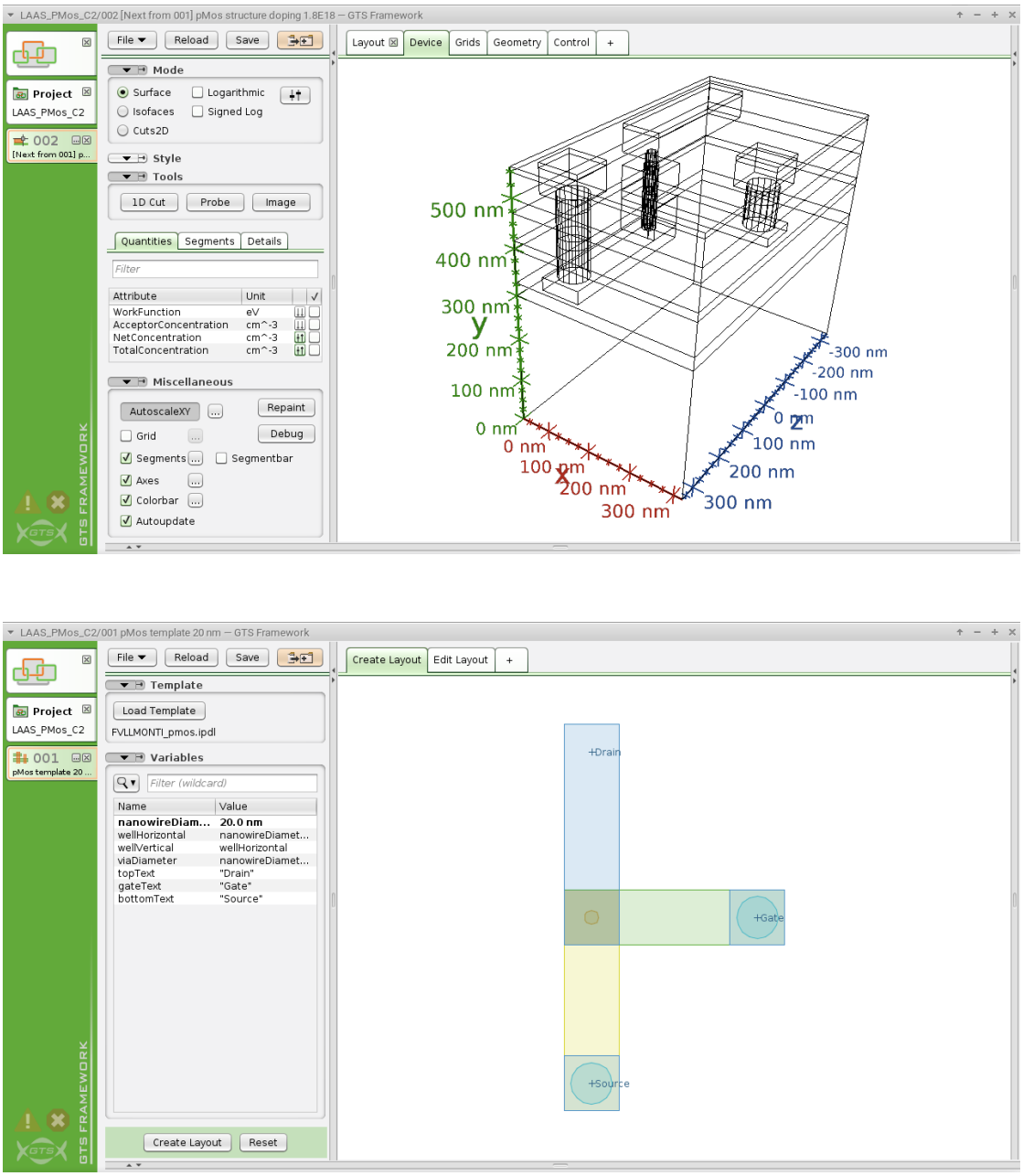


Figure 2.1: View of the GTS Structure (top) and Layout (bottom) tools

MinimosNT ToolFolder

MinimosNT is a general-purpose semiconductor device simulator providing steady-state, transient, and small-signal analysis of arbitrary two and three dimensional device geometries. In mixed-mode device and circuit simulation, numerically simulated devices can be embedded in circuits consisting of compact device models and passive elements. A comprehensive set of physical models allows for simulating various kinds of advanced device structures, such as present-day CMOS devices, silicon-on-insulator (SOI) devices, and hetero-structure devices. Aware of the atomistic nature of traps and dopants, MinimosNT provides reliability and variability modeling of highly scaled transistors such as bulk planar devices and silicon-on-insulator FinFETs having a channel length of 22nm or less. The setup can be modified either by the GUI or by editing the input-deck `.ipd` file. The simulator is developed in collaboration with Vienna University of Technology: its first release dates back to 1995 [14], as a successor of MINIMOS 6 [15].

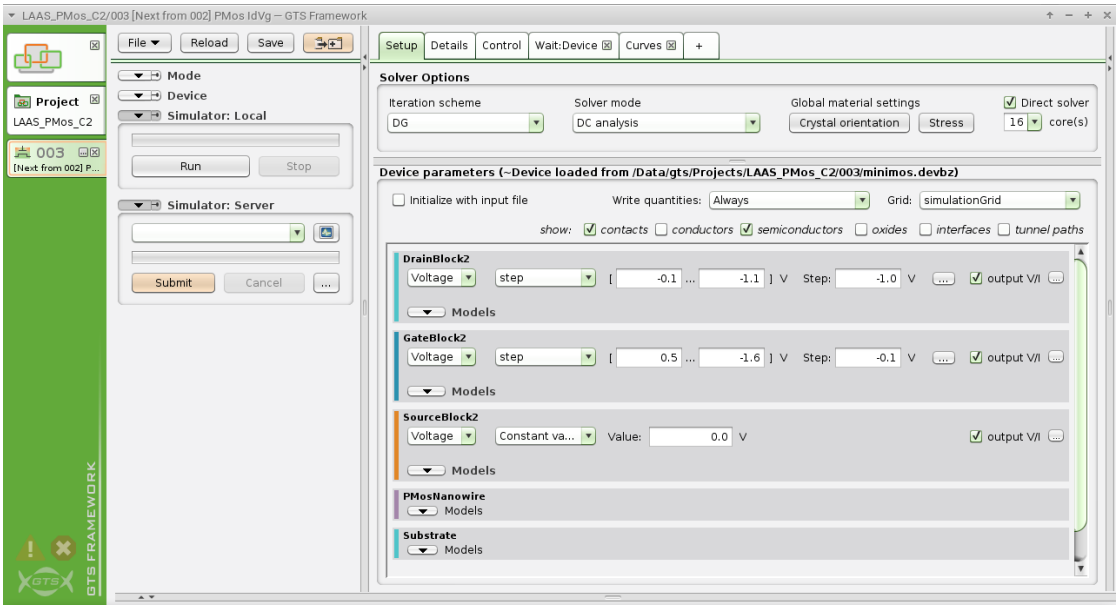


Figure 2.2: View of the MinimosNT configuration screen

Design Of Experiment (DOE)

The Design Of Experiment screen allows to run simulations with multiple values of one or more selected parameters. It is essential for the user because it allows to prepare experiments, through the graphical user interface, to study the effects that a certain parameter has on the output characteristics of a certain device. The DOE is not considered a ToolFolder, but a script, which in GTS Framework is described as a versatile tool for creating complex jobs which consist of multiple tool executions.

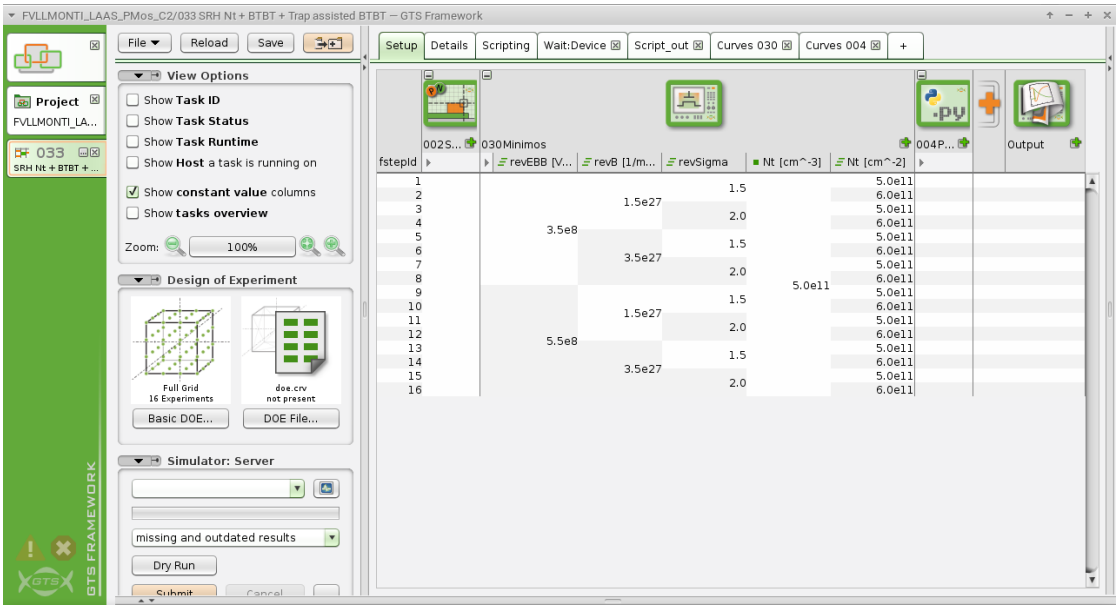


Figure 2.3: View of the GTS Design Of Experiment (DOE) screen

2.2.2 Synopsys Sentaurus Device

The softwares used for the second part of the work are the Sentaurus Device Editor (`sde`), which is used to build the geometrical model of the device, define the regions, materials and interfaces and generate the simulation mesh, and Sentaurus Device (`sdevice`), which include the solver and is used to run the simulation of the device model. The device structure can be modified in the Sentaurus Structure Editor (`sse`), shown in Figure 2.4, which provides a graphical user interface for `sde`.

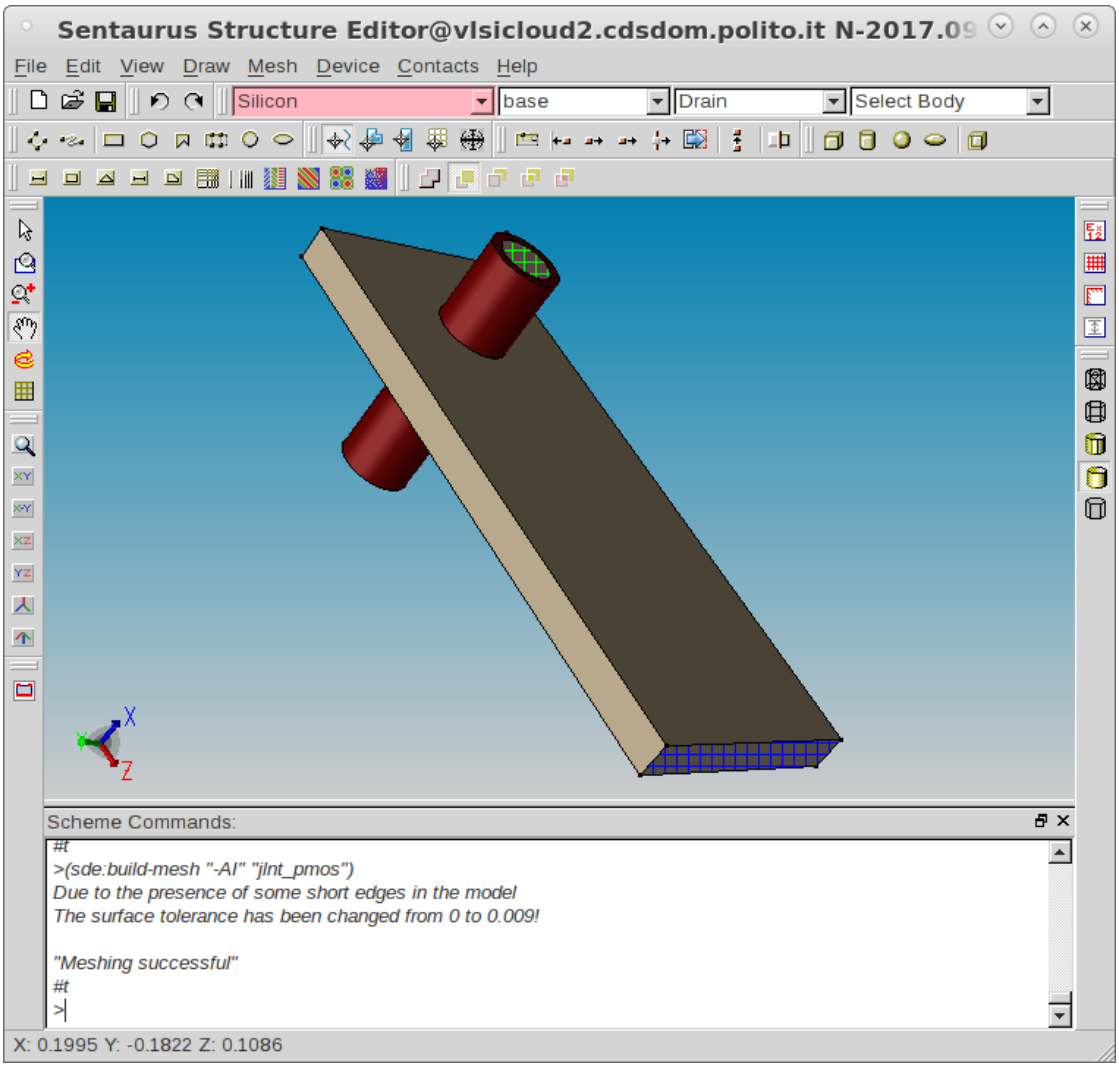


Figure 2.4: View of Synopsys Sentaurus Structure Editor

2.3 Definitions

Before introducing the data, it is useful to define some notations that will be used throughout the presentation and analysis of the results.

- The suffix *sat* denotes a parameter that has been extracted from a device in saturation region, hence with $V_D = -1.1$ V;
- The suffix *lin* denotes a parameter that has been extracted from a device in linear region, hence with $V_D = -0.1$ V;
- The threshold voltage V_{th} is computed as the gate voltage V_G at which the drain current I_D is equal respectively to $1 \cdot 10^{-8}$ A for the 20 nm and 27 nm devices, and to $1 \cdot 10^{-7}$ A for the 34 nm device;
- The drain induced barrier lowering (DIBL) is calculated as $(V_{th,sat} - V_{th,lin})/V_D$.

Chapter 3

Device under study

3.1 Introduction

The device under study at LAAS-CNRS was first demonstrated in 2013 by G. Larrieu and X. L. Han [16] and is a **vertical nanowire junctionless nanotransistor**, with surrounding gate all around.

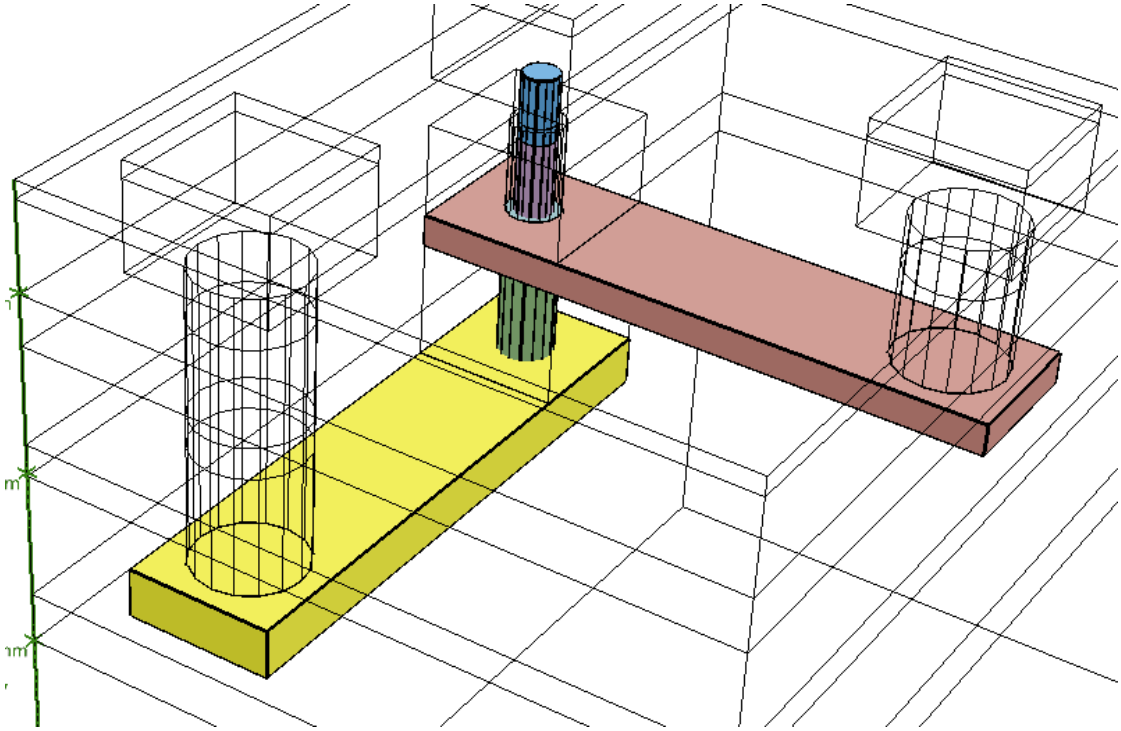


Figure 3.1: View of the device from the GTS Framework Structure tool.
Visible parts: Source silicide (yellow), Drain silicide (blue),
Gate metal (brown), Gate oxide (green), Silicon nanowire (purple)

The transistor has a monocrystalline Silicon nanowire channel, silicide source and drain contacts and metallic gate. The pMOS is based on a p-type Silicon nanowire and Platinum Silicide (PtSi) contacts, whereas the nMOS uses a n-type Si nanowire and Nickel Silicide (NiSi) contacts. In both devices, the metallic gate is made of Chromium, even though alternatives with Copper gate have been fabricated. A view of the device model obtained from GTS Framework is shown in Figure 3.1. From the picture one can see the main parts of the device: the silicon nanowire at the center, with the source and drain silicides at its ends and the metallic gate all around, with an oxide layer separating it from the nanowire. In this work, only the pMOS has been inspected and modeled on the TCAD softwares. The fabrication method and the most relevant experimental results will be disclosed in this chapter.

3.2 Fabrication

3.2.1 Fabrication of the Silicon nanowires

An essential requirement for building a high-performance GAA transistor is to start from high quality Silicon nanowires, with as few defects as possible. In this work a top-down approach has been chosen, because it is easier to integrate in a standard CMOS process flow and it is free from metallic contamination, a usual byproduct of the bottom-up methods based on catalytic growth. In particular, this method takes advantage of the strong anisotropy of the reactive ion etching process (RIE). The complete process flow resuming the fabrication steps is shown in Figure 3.2. In the following, bold numbers in parenthesis refer to the corresponding step shown in the flowchart.

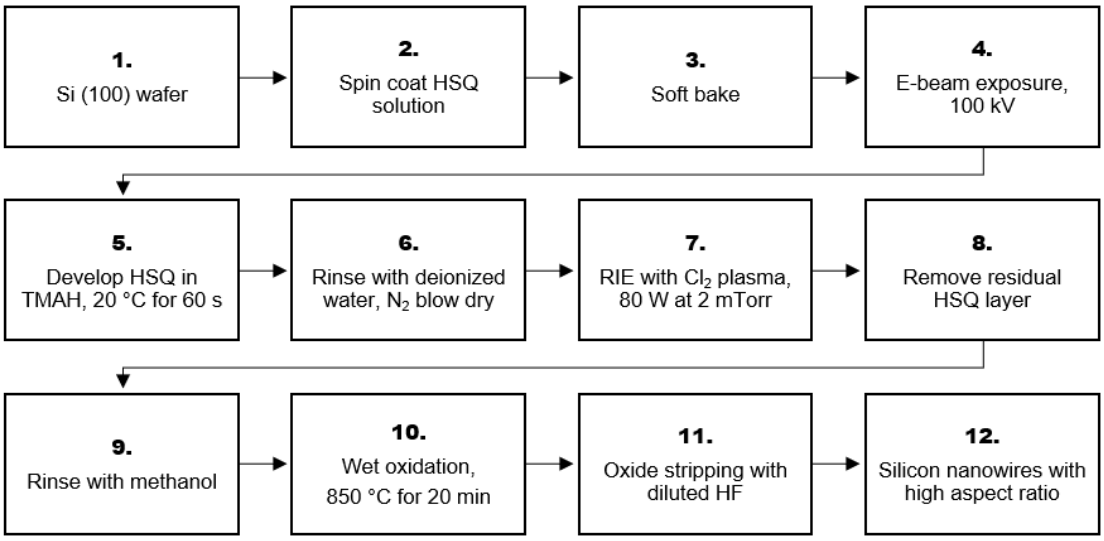


Figure 3.2: Scheme of the process steps for the Si nanowires [1]

The process starts with a <100> Silicon wafer. The first step is the spin coating of

the HSQ solution (**2**), comprising a negative tone resist Hydrogen Sylsesquioxane (HSQ) diluted in isobutyl ketone. The spin coated HSQ solution is then baked at 80 °C for 60 seconds (**3**), in order to evaporate the solvent. By varying the angular speed of the centrifuge it is possible to deposit resist thicknesses ranging from 70 to 330 nm.
After the soft baking the electron beam exposure (**4**) is carried out. The electron beam is accelerated at 20 kV, however higher acceleration values could be employed to ensure good results even with thicker HSQ coatings. In the first disclosure of this method, by Han, Larrieu and Dubois [1], an acceleration of 100 kV has been chosen. The almost perfectly circular shape is obtained thanks to a peculiar star-shaped design that forces the symmetry of the nanopillars and promotes an homogeneous distribution of Gaussian energy [17]. At the end of this step a HSQ nanocolumns network, with diameters ranging from 20 to 240 nm, is obtained. It is important to note that under electron beam exposure, HSQ has the remarkable property to evolve from a cage like monomer to a spaghetti-like polymer that resembles the structure of amorphous silicon dioxide [18]. Thanks to the said properties, the etching rates of HSQ and Silicon are respectively 45.9 nm/min and 77.8 nm/min, leading to a quite satisfactory selectivity. After the exposure, the HSQ resist is developed by manual immersion in 25% tetramethylammonium hydroxide (TMAH) at 20 °C for 60 seconds (**5**), then the wafer is rinsed with deionized water and blow dried with N_2 (**6**). At this point the silicon nanowires are ready to be etched from the substrate. A reactive ion etching step (**7**) is performed using a chlorine chemistry plasma with a bias power of 80 w and a pressure of 2 mTorr. The chlorine based chemistry, using pure Cl_2 plasma, is chosen because it allows to reach 94% anisotropy of the vertical nanowires, and leaves a clean substrate, free from needle-like silicon byproducts, typical of other reactants employed like, for instance, SF_6 combined to passivating gas CHF_3. After the exposure the residual HSQ layer is removed by a 10% hydrochloric acid diluted in methanol for 120 seconds (**8**) and then it is rinsed in methanol (**9**). After these steps, we have already an array of high aspect ratio silicon nanowires on the substrate. The obtained silicon sidewalls exhibit a bowl-like, concave profile, deeper for closely packed nanowires. This phenomenon is mainly due to ion sputtering scattered from the HSQ mask and because the top surface of the mask is rounded rather than being perfectly square shaped. As the distance between one nanowire and the other decreases, the undercut becomes more severe because this sputtered elements are concentrated on a smaller length, whereas by inspecting nanowires more distant from each other one can see that this effect is reduced. This effect was not observed in the case of isolated nanowires.
After rinsing the nanowires with methanol (**17**) there is the need to eliminate the defects due to plasma induced damage. Plasma induced damage is recognized as a source of device performance and reliability degradation, including UV radiation, electrostatic discharge and physical damage due to ionic bombardment iron induced appoint effects an surface or interface monster symmetric states resulting from plasma exposure hold a major responsibility in degradation of carrier mobility an sub threshold metal oxide semiconductor characteristic. In order to eliminate this problem, the vertical silicon nanowires are treated with wet oxidation (**10**) at 850 °C for 20 minutes. The silicon dioxide layer is then stripped in diluted hydrofluoric acid (**11**) resulting in a perfectly clean silicon surface. Consequently, the diameter of the silicon nanowires is trimmed. At the end of the procedure, high aspect ratio silicon nanowires are obtained.

3.2.2 Fabrication of the GAA JLNT

At this point, the Silicon nanowires are grown on the substrate (**2**) and we are ready to fabricate the devices around them. The flowchart representing the fabrication method is shown in Figure 3.3. In the following, bold numbers in parenthesis refer to the corresponding step shown in the flowchart. Firstly, the gate oxide is obtained by means of a dry oxidation step (**3**), carried out at 725 °C for over 30 minutes, resulting in an oxide thickness of 5 nm.

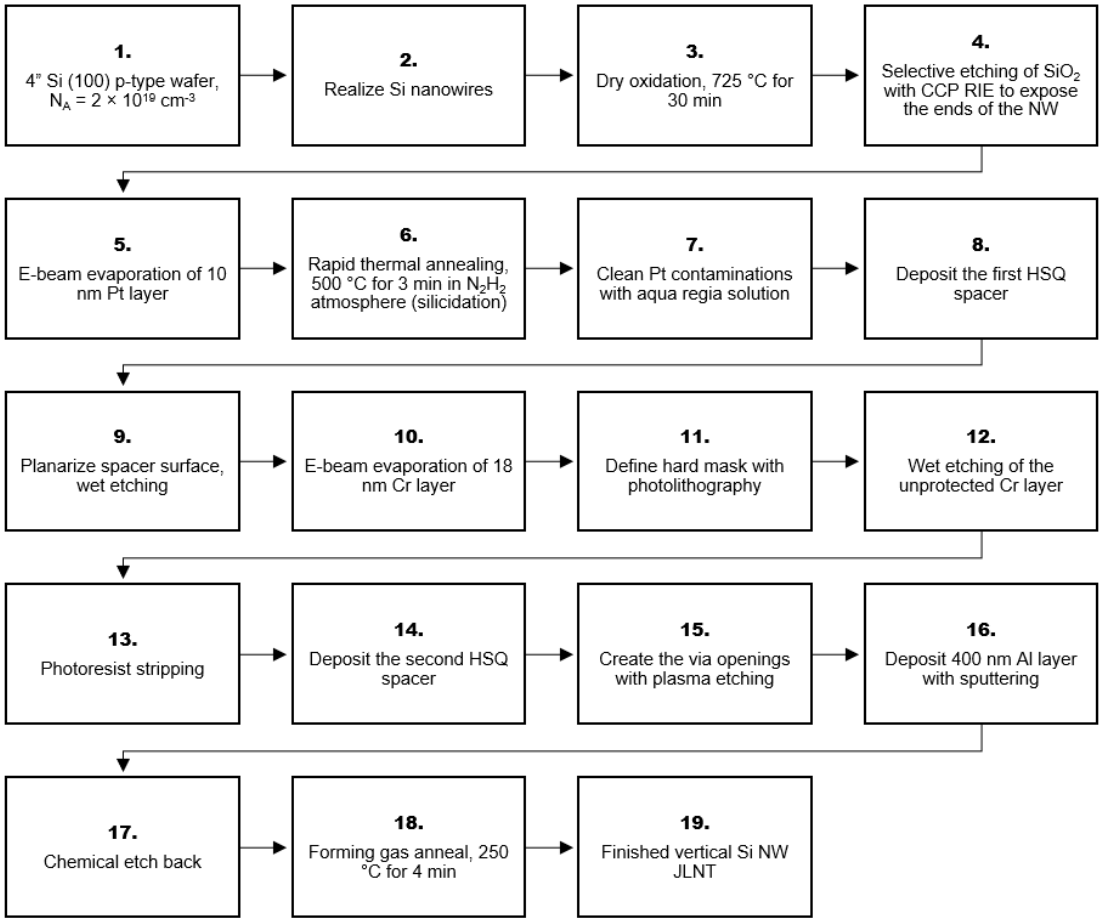


Figure 3.3: Scheme of the process steps for the VNWFET [2]

Successively, a selective etching of the Silicon dioxide (**4**) with charge coupled plasma (CCP) RIE is performed in order to expose both ends of the Silicon nanowire and thus allow the formation of the drain and source silicide contacts. The RIE step uses fluorine chemistry ($CHF_3 : CF_4 : Ar = 20 : 20 : 10$). The source contact is defined by a lift-off process. A 10 nm thick layer of Platinum is deposited on substrate by means of electron beam evaporation (**5**), then the PtSi contacts are formed with a rapid thermal annealing (RTA) (**6**) at 500 °C for three minutes in N_2H_2 atmosphere. The remaining Platinum

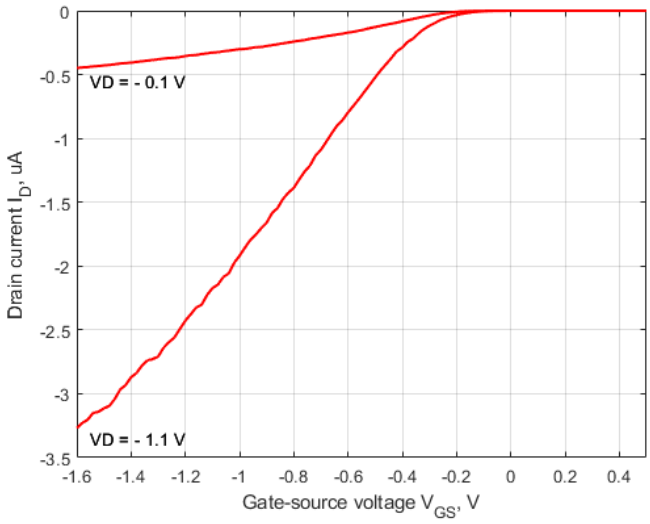
contamination is dissolved in an aqua regia solution (**7**).
At this point, the first HSQ spacer is deposited (**8**) and the upper surface of the spacer is planarised with wet etching (**9**). Specifically, this chemical etching step is done with a highly diluted hydrofluoric acid in deionized water (1:1000) solution, with the addition of a cationic surface agent, benzalkonium chloride. The latter is added in order to prevent bubble encroachment on the dielectric surface, issued from the releasing of Hydrogen gas during the etching of SiO_xH_y, leading to an increase of the surface roughness. By using said solution on the HSQ, an etching rate of 2.3 nm/s is achieved, resulting in a highly controllable process, and a very flat surface is obtained. With the aid of an atomic force microscope, it has been measured that the roughness of the planarized surface is below 2 nm [19]. The metallic gate is now ready to be formed. A Chromium gate is deposited by means of electron beam evaporation (**10**). This technique is advantageous because the gate length does not rely by any means on photolithographic steps but only on the deposition rate and time of the evaporation step. Gate contacts are defined with a photolithographic hard mask (**11**), then the unprotected Chromium layer is removed by wet etching (**12**). The remaining photoresist is stripped off (**13**) and the second HSQ spacer is deposited (**14**). At this point the via openings are created through the HSQ spacers with plasma etching (**15**), in order to gain access to the gate and the source, and the back end of line contacts are fabricated by depositing 400 nm of Aluminium with sputtering (**16**), followed by chemical etch back in a solution of $H_3PO_4 : HNO_3 : EDI = 5 : 40 : 7$ (**17**). Finally a forming gas anneal (FGA) (**18**) (N_2H_2) is performed at 250 °C for 4 minutes, in order to passivate defects at both $Si - SiO_2$ and PtSi interfaces.

3.3 Experimental results

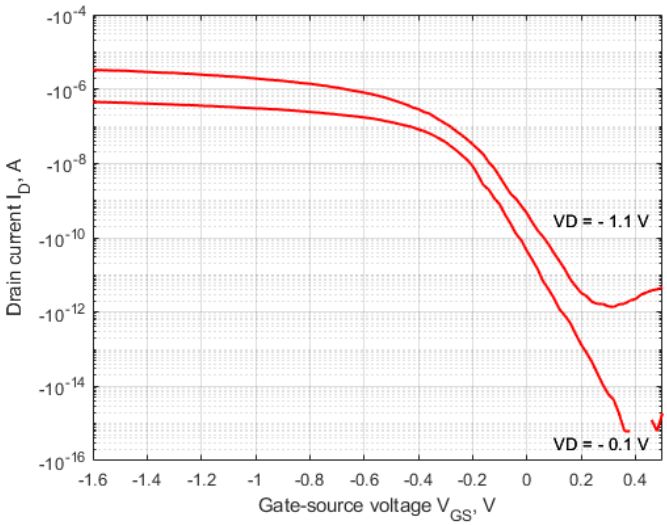
The junctionless nanotransistors studied at LAAS-CNRS during my internship semester are the result of several years of technological refinement and have been labeled as *devices of the third generation*. Of course, these devices inherit the majority of their features from the previous generations [16] [2], along with the fabrication method, which has been ameliorated from one generation to the other. The studied devices come in three different sizes, defined according to the nanowire diameter: we will investigate and model JLNTs with nanowire diameters of **20 nm**, **27 nm** and **34 nm**. In this section, the electrical measurements of the three different iterations of the studied devices are presented. It is immediate to notice how the electrical behavior of the three devices is different: the 20 nm device shows the best electrostatic control among the three, whereas the 34 nm device has the poorer. This is due to the fact that when the diameter of the nanowire is too large, the electric field exerted by the gate is not able to assert a sufficiently large depletion region across the whole silicon nanowire, and thus a current would still flow through the nanowire even when the device should be off. Conversely, thanks to the increased diameter, the 34 nm device shows the larger drive current. It is important to notice that these electrical measurements have been divided by a normalization factor of 16, in fact the devices have been fabricated starting from batches of 16 nanowires and thus the current read refers to 16 nano transistors in parallel.

3.3.1 20 nm device

$R_{ON,lin}$	$R_{ON,sat}$	SS	DIBL	$V_{th,lin}$	$V_{th,sat}$
223.06 kΩ	336.49 kΩ	73.784 mV/dec	0.052	-0.235 V	-0.182 V



(a) Linear plot

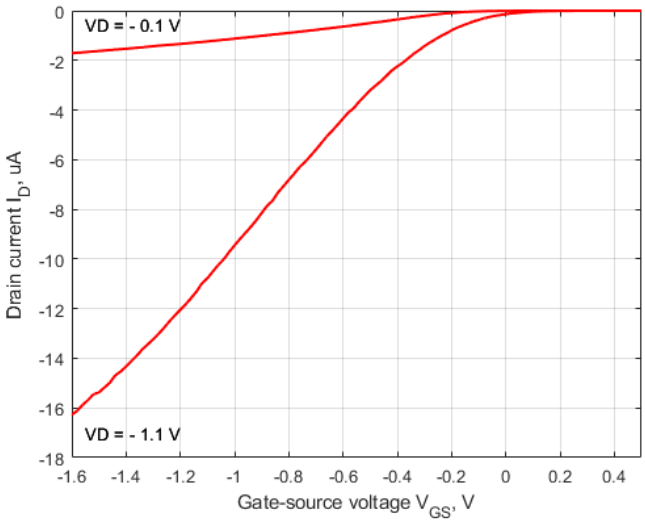


(b) Logarithmic plot

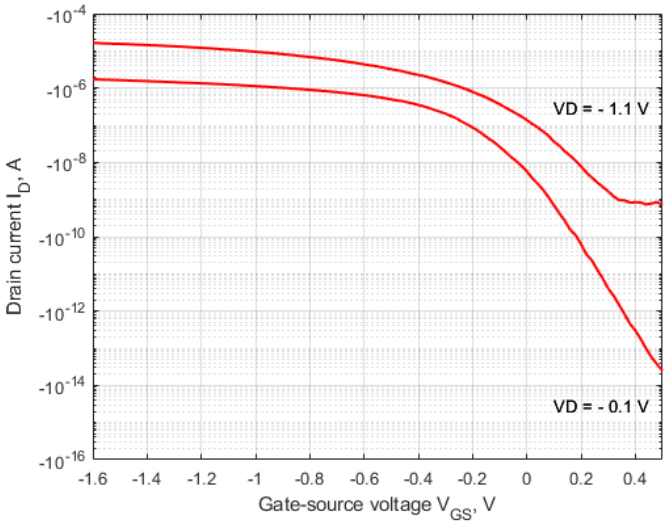
Figure 3.4: Transfer characteristics of the device with diameter = 20 nm

3.3.2 27 nm device

$R_{ON,lin}$	$R_{ON,sat}$	SS	DIBL	$V_{th,lin}$	$V_{th,sat}$
58.48 kΩ	67.60 kΩ	91.090 mV/dec	0.220	-0.020 V	0.200 V



(a) Linear plot

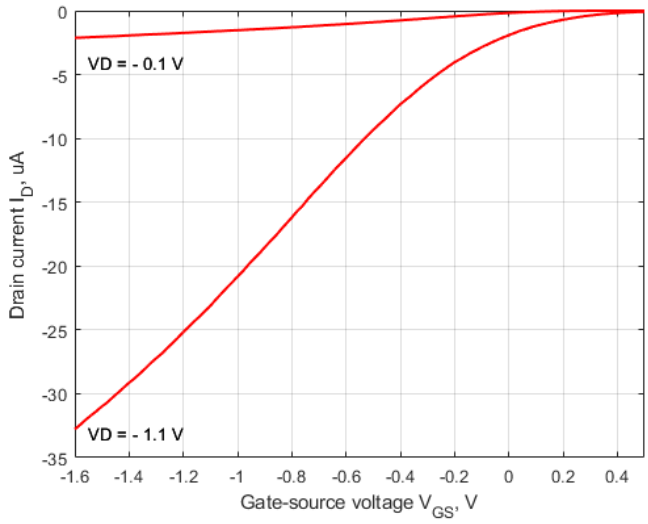


(b) Logarithmic plot

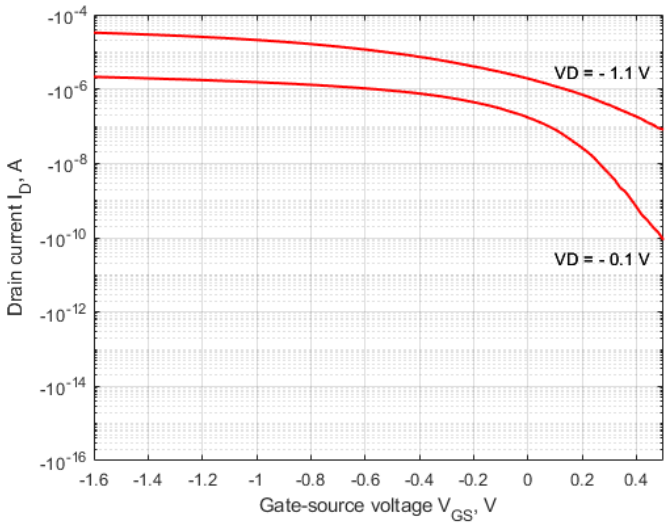
Figure 3.5: Transfer characteristics of the device with diameter = 27 nm

3.3.3 34 nm device

$R_{ON,lin}$	$R_{ON,sat}$	SS	DIBL	$V_{th,lin}$	$V_{th,sat}$
47.17 kΩ	33.56 kΩ	134.447 mV/dec	0.400	0.080 V	0.480 V



(a) Linear plot



(b) Logarithmic plot

Figure 3.6: Transfer characteristics of the device with diameter = 34 nm

Part II

Device TCAD modeling

In order to have a clearer picture of the parts inspected in this chapter, a 3D model of the device is presented in Figure 3.7. In the model we see the source silicide, marked in yellow, the drain silicide in blue, the silicon nanowire in purple, surrounded by the gate, in brown, and the oxide in green. In GTS Framework, each one of these blocks constitute a *segment*: a segment is used to define a specific portion of the device under study and is made of a specified material. Once the segment is defined, it inherits the properties from a *segment default*. For instance, the source and drain silicides are made in PtSi; when the structure is generated the TCAD loads all the material properties from the PtSi segment default. By modifying the properties of the segment default we can change the properties of a material through the whole device, while changing a property for a specified segment does not affect the other segments, even if they are made of the same material. A similar approach can be done in Sentaurus, hence we can modify physical parameters by acting on a single *region* or on the material properties defined in a `jlnt_pmos.par` parameters file. Many other parts visible in the 3D view are not colored and refer to other segments that have been left unmodified throughout the whole calibration process. These segments include the vias for the electrical contacts, the spacers and the filling layers.

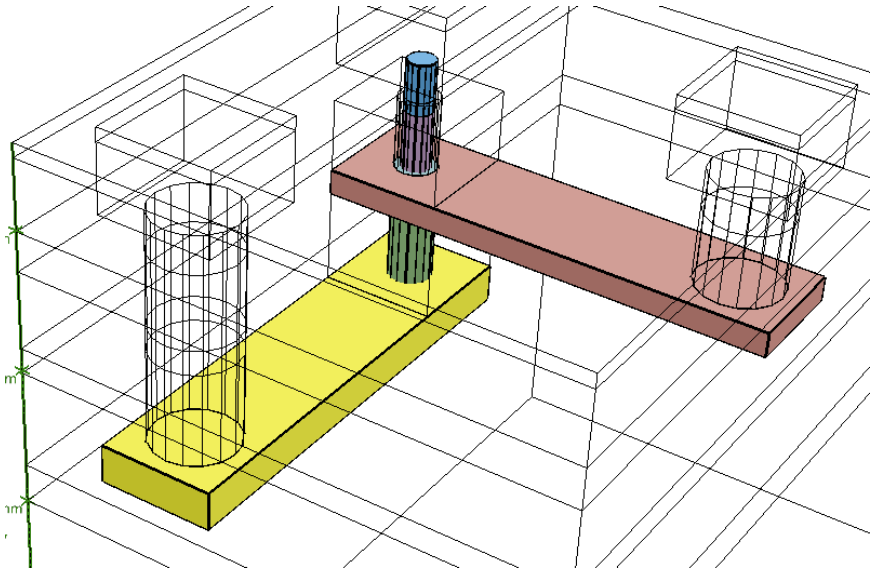


Figure 3.7: View of the device from the GTS Framework Structure tool.

The core part of the work done consists in the calibration of the TCAD model. Having a well calibrated model, which outputs results very close to the real world measurements, is very useful to have a good understanding of the physical behavior of the device and to make predictions for the device further developments, an essential ingredient for constructing an effective DTCO flow. In the calibration process several variables have been addressed separately, then merged in a single model to fit the experimental measurements. The variables inspected are shown in Figure 3.8. It is worth mentioning that this process has been done in an iterative way: in many cases, tuning a variable has an effect that counteracts other, thus many different trials were necessary to obtain a fitted model.

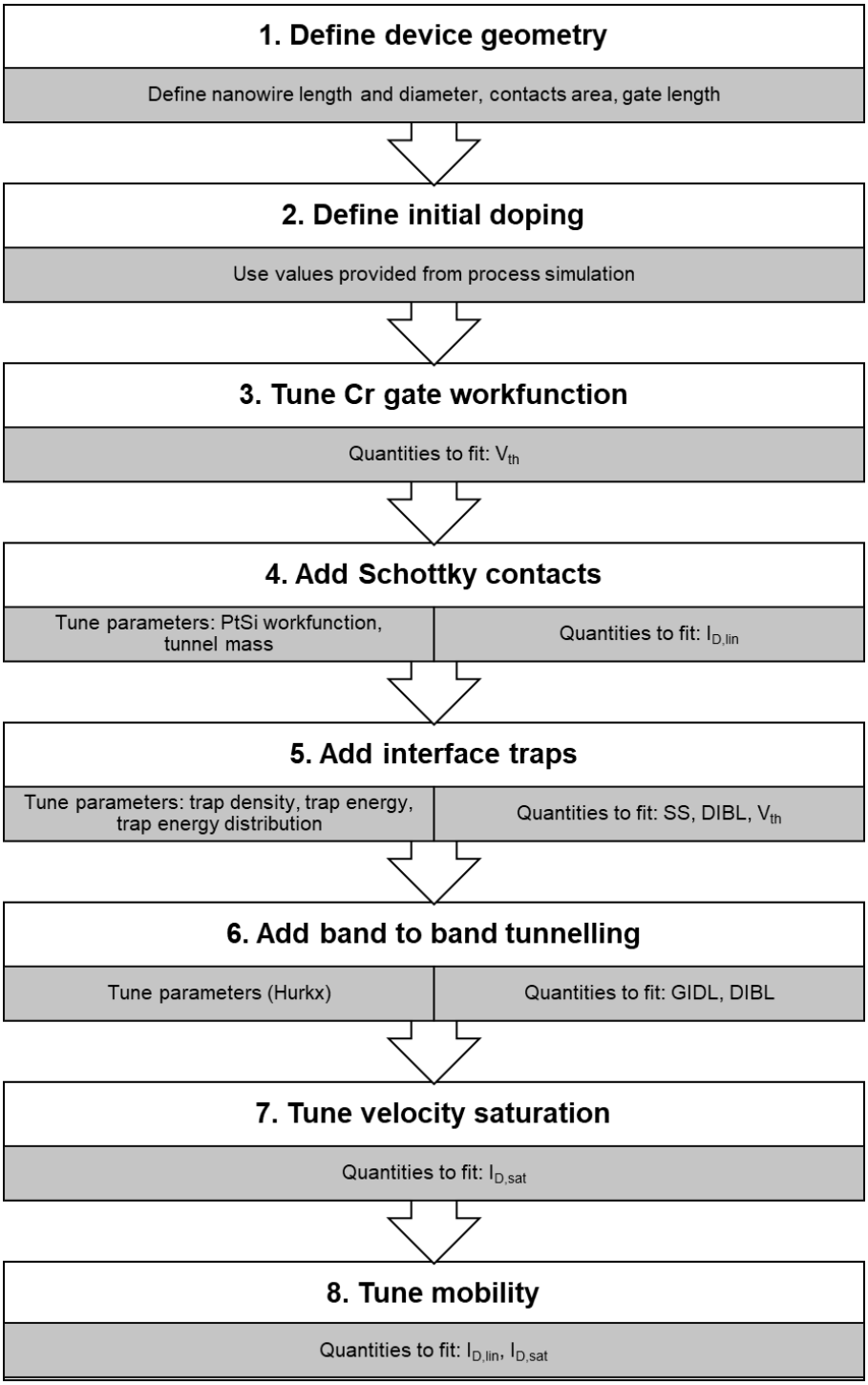


Figure 3.8: Summary of the parts inspected in the calibration of the model

Chapter 4

Silicide contacts modeling

Metal-semiconductor interfaces play a crucial role in electronic devices, as they are needed for the electrical access to the semiconductor. A desirable metal - semiconductor interface used for contacts provides low resistance, non rectifying behavior, and is commonly referred at as ohmic contact. In the studied device, the drain and source contacts are made in Platinum Silicide (PtSi). Silicides are a popular choice for making ohmic contacts, as they provide good conductivity and can be obtained without high processing temperatures. Depending on whether we have a p-doped or n-doped semiconductor, we need to chose a metal with a suitable work function accordingly.

The transport mechanism that allows the flow of carriers through the metal - semiconductor interface changes according to the doping level of the semiconductor, resulting in three different regimes: **thermionic emission (TE)**, **thermionic - field emission (TFE)** and **field emission (FE)**.

The three cases can be roughly distinguished by comparing k_BT to the value of E_{00}, defined as

$$E_{00} = \frac{q\hbar}{2}\sqrt{\frac{N}{m^*\varepsilon_s}} \tag{4.1}$$

As disclosed by the work of Ng and Liu [20], we will appreciate:

- TE regime for $kT \gg E_{00}$ and doping levels smaller than 10^{18} cm^{-3}
- TFE for $kT \approx E_{00}$ and doping levels comprised between 10^{18} and $5 \cdot 10^{19}$ cm^{-3}
- FE for $kT \ll E_{00}$ and doping levels over $5 \cdot 10^{19}$ cm^{-3}

The specific contact resistances R_c are respectively in the order of

- $10^0 - 10^{-2}\ \Omega \cdot \text{cm}^2$ for TE
- $10^{-1} - 10^{-4}\ \Omega \cdot \text{cm}^2$ for TFE
- $10^{-2} - 10^{-7}\ \Omega \cdot \text{cm}^2$ for FE

The specific contact resistance can vary in these ranges according to the Schottky barrier height (SBH), defined for holes as:

$$SBH = E_g - q(\Phi_m - \chi) \tag{4.2}$$

with E_g being the Silicon energy gap, Φ_m the metal work function and χ the Silicon electron affinity. We have $E_g = 1.12$eV and $\chi = 4.05$eV.

The work of Ng and Liu [20] also shows the dependence of the contact resistivity on the Schottky barrier height: lower barrier heights will yield lower contact resistances, whereas higher barrier heights will yield higher contact resistances. In thermionic emission (TE), the contact resistance does not vary with the doping, but only with the barrier height, whereas as we move towards the thermionic - field emission (TFE) and field emission (FE) cases, the doping dependence becomes noticeable. Thus, it is clear that in order to obtain a good ohmic contact, the doping must be sufficiently high.

4.1 Ohmic contacts simulated with Sentaurus

A preliminary version of the device model is studied on Synopsys Sentaurus with ideal Ohmic drain and source contacts. This in order to verify that the model works correctly and is ready to accommodate the next calibration steps. In order to define Ohmic contacts in Sentaurus, it is sufficient to declare the contacts in the *Electrode* section of the sdevice `jlnt_pmos_des.cmd` file, without the need of adding additional parameters, it is the default choice in the TCAD.

```
Electrode {
    { name="Source"    Voltage=0.0 }
    { name="Drain"     Voltage=0.0 }
    { name="Gate"      Voltage=0.0 }
}
```

Charge neutrality and equilibrium conditions are assumed at Ohmic contacts [21]:

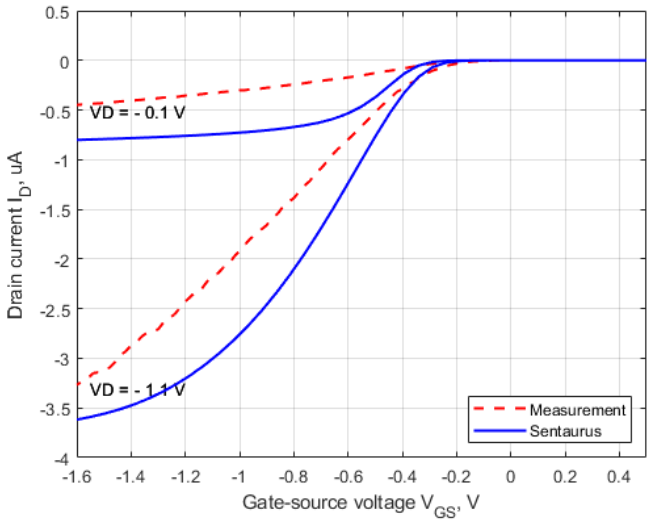
$$n_0 - p_0 = N_D - N_A \qquad\qquad n_0 p_0 = n_{i,eff}^2 \tag{4.3}$$

In the following, the devices with nanowire diameters of 20, 27 and 34 nm are simulated with Ohmic contacts in Synopsys Sentaurus.

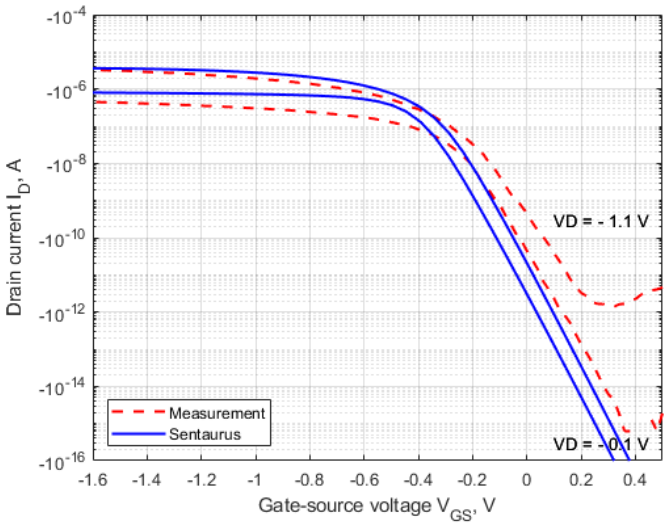
In Figure 4.1 the transfer characteristics of the 20 nm device is shown. It can be seen that the on-current is higher than the measured one, both for the saturation ($V_{DS} = -1.1$ V) and linear ($V_{DS} = -0.1$ V) regimes, a foreseeable result given the fact that the contact resistance, null in this case, should dominate the overall R_{ON} of the device. Conversely, the transfer characteristics of the 27 and 34 nm devices, shown in Figures 4.2 and 4.3, are showing an on-current lower than the measured one, but this will be addressed in the next chapter. For the moment the results shown can still be considered as a best-case scenario concerning the contact resistance.

NW diameter = 20 nm

Doping	Cr workfunction	SS	DIBL	$V_{th,lin}$	$V_{th,sat}$
$N_A = 1.2e18\ cm^{-3}$	4.50 eV	75.325 mV/dec	0.052	-0.235 V	-0.182 V



(a) Linear plot

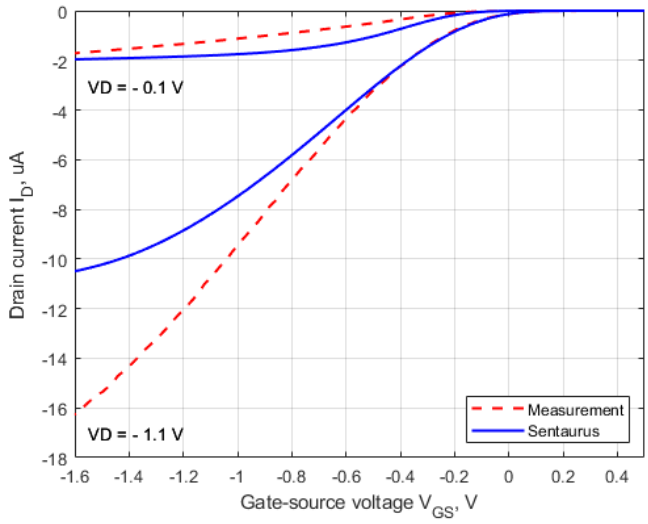


(b) Logarithmic plot

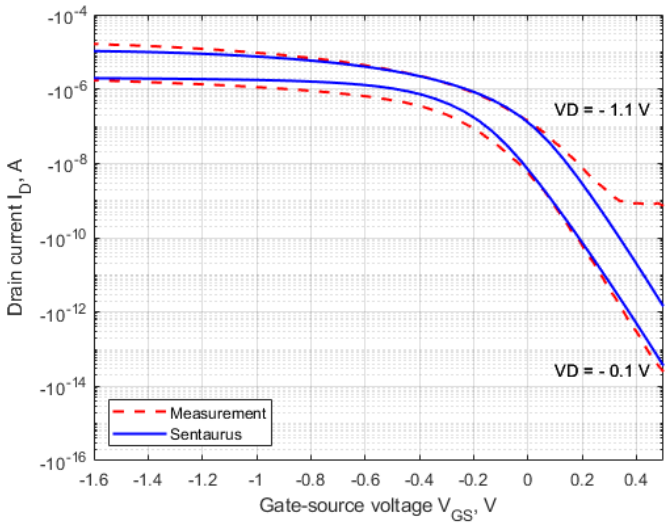
Figure 4.1: Transfer characteristics of 20 nm device with Ohmic source/drain contacts

NW diameter = 27 nm

Doping	Cr workfunction	SS	DIBL	$V_{th,lin}$	$V_{th,sat}$
$N_A = 2.2e18\ cm^{-3}$	4.36 eV	93.624 mV/dec	0.158	0.028 V	0.185 V



(a) Linear plot

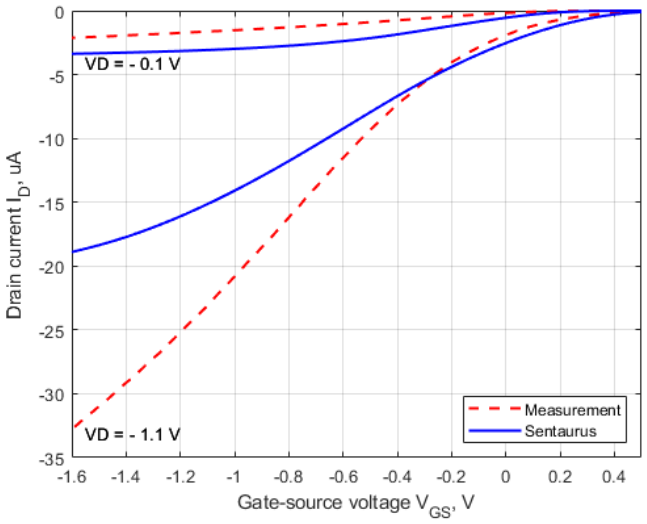


(b) Logarithmic plot

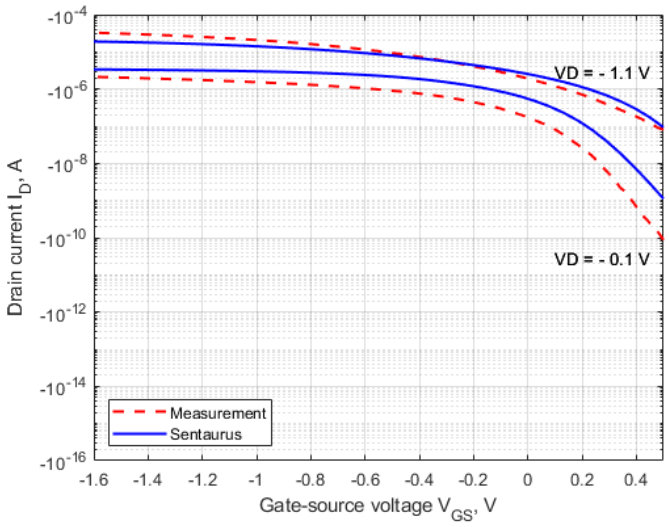
Figure 4.2: Transfer characteristics of 27 nm device with Ohmic source/drain contacts

NW diameter = 34 nm

Doping	Cr workfunction	SS	DIBL	$V_{th,lin}$	$V_{th,sat}$
$N_A = 2.7e18\ cm^{-3}$	4.25 eV	141.560 mV/dec	0.263	0.237 V	0.500 V



(a) Linear plot



(b) Logarithmic plot

Figure 4.3: Transfer characteristics of 34 nm device with Ohmic source/drain contacts

4.2 Schottky contacts - Varahramyan model

The model presented by Varahramyan and Verret [22] includes the influence of field emission and thermionic emission in a unified manner, requiring only one simple relation for the determination of the specific contact resistance. Since the device studied has a doping in the order of 10^{18} cm^{-3}, the case we are interested in is the TFE (thermionic-field emission) regime. In this case the contact resistance can be expressed as

$$R_c = \frac{k}{qA^*T} c_{TFE} \exp\left(\frac{q\phi_b}{E_0}\right) \tag{4.4}$$

where

$$c_{TFE} = \frac{kT}{\sqrt{\pi(q\phi_b + u_F)E_{00}}} \cosh\left(\frac{E_{00}}{kT}\right) \sqrt{\coth\left(\frac{E_{00}}{kT}\right)} \times \exp\left(\frac{u_F}{E_0} - \frac{u_F}{kT}\right)$$

and

$$E_{00} = \frac{qh}{4\pi}\sqrt{\frac{N}{\epsilon_s m_t^*}} \qquad E_0 = E_{00} \coth\left(\frac{E_{00}}{kT}\right)$$

Considering that E_0 approaches E_{00} when $kT/E_{00} \ll 1$, the Equation 4.4 can be used also for the FE case, by substituting the constant c_{TFE} with an average value $\bar{c} = (m\ c_{TFE} + n\ c_{FE})/(m+n)$ where m and n are the number of points in the doping range over which the coefficients are calculated. Selecting the Varahramyan model has some noticeable advantages in terms of computing costs: with this model there is no need to enable the placement of tunneling paths at the metal - semiconductor junctions since, as discussed previously, the model is capable to present the TE, TFE and FE regimes in a unified way. With this model selected, the time required for extracting an IV curve is reduced with respect to the Schottky model however, it also reduces the flexibility as it only provides three fitting parameters, thus limiting the control that the user has on the thermionic and field emission mechanisms separately.

4.2.1 Varahramyan model in GTS Framework

We recall the expression of the contact resistance, Equation 4.4, now extracted from the MinimosNT user manual [23]:

$$R_c = \frac{k_B}{qA^*T} \bar{c} \exp\left(\frac{q\phi_b}{E_0}\right) \tag{4.5}$$

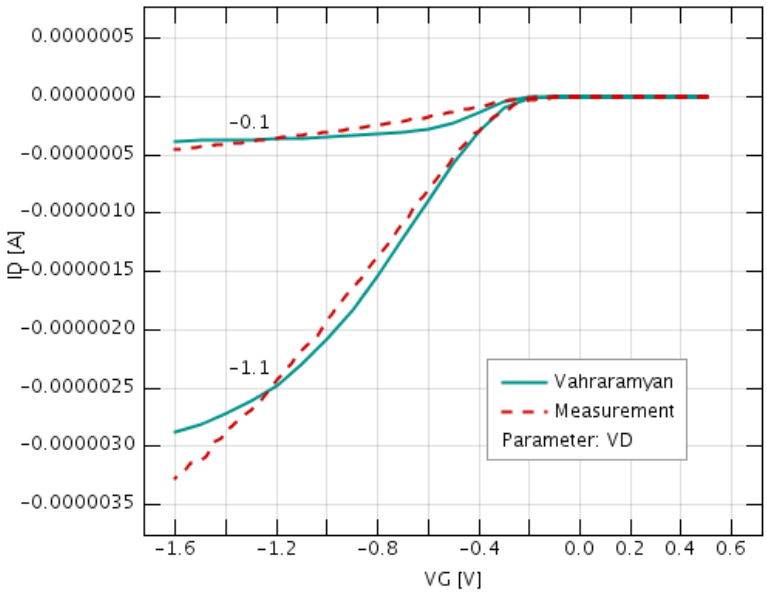
with

$$E_{00} = \frac{qh}{4\pi}\sqrt{\frac{N}{\epsilon_s m_t^*}} \qquad E_0 = E_{00} \coth\left(\frac{E_{00}}{kT}\right)$$

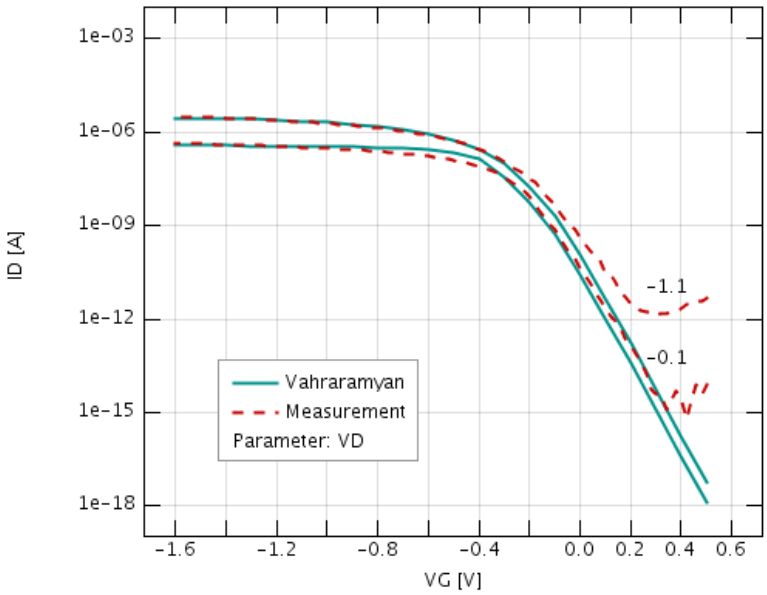
We recall that with the equation shown above we can effectively model the transport in TE, TFE and FE cases, by simply tuning the value of c using equation 4.4, for the TFE case. In case it is necessary to model a TE scenario, the c coefficient should be equal to 1.

The values used for the parameters are

$$m_t^* = 0.16\ m_0 \quad c = 0.03 \quad \phi_B = 0.25\ \text{V}$$



(a) Linear plot



(b) Logarithmic plot

Figure 4.4: Transfer characteristics obtained with the Varahramyan model

4.2.2 Varahramyan model in Sentaurus

The metal/semiconductor interfaces can modeled in an effective way in Sentaurus by selecting the Varahramyan model. This model, as we have already seen in the previous section, offers a good trade-off between accuracy and computational complexity.
The Varahramyan model can be activated in Sentaurus by declaring the drain in the *Electrode* section as follows:

```
Electrode {
    { name="Source"    Voltage=0.0 }
    { name="Drain"     Voltage=0.0 DistResist = SchottkyResist }
    { name="Gate"      Voltage=0.0 }
}
```

After defining the electrodes at which we want to apply the Varahramyan model, it is necessary to define the parameters of the model. These can be defined in the `jlnt_pmos.par` parameters file, inside the section corresponding to the Silicon nanowire region.

```
SchottkyResistance {
    Electrode = "Drain" {
        Rinf = 2.4000e-09 , 5.2000e-09 # [Ohm*cm^2]
        PhiB = 0.6 , 0.12  # [eV]
        mt = 0.19 , 0.16   # [*m0]
    }
}
```

Once the parameters have been defined, the TCAD is able to compute the contact resistance, which is given by the equations [21]:

$$R_d = R_\infty \frac{300\text{K}}{T_0} \exp\left(\frac{q\phi_B}{E_0}\right) \tag{4.6}$$

where

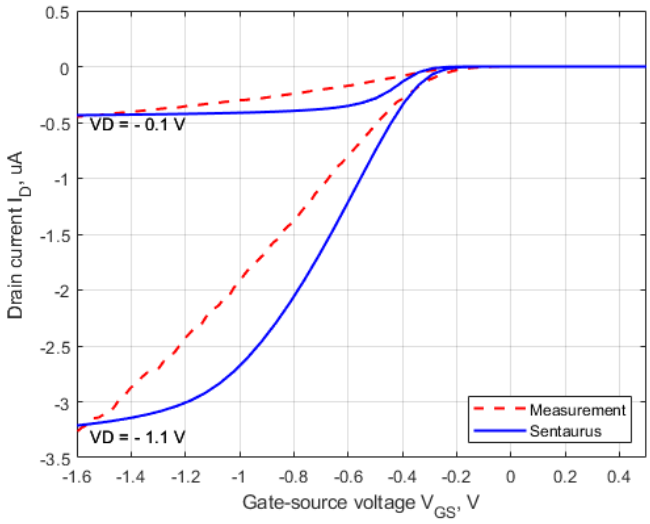
$$E_0 = E_{00} \coth\left(\frac{E_{00}}{kT_0}\right) \qquad E_{00} = \frac{qh}{4\pi}\sqrt{\frac{|N_{D,0} - N_{A,0}|}{\epsilon_s m_t}}$$

The parameter ϕ_B is the Schottky barrier height, corresponding to the difference between the valence band energy of the semiconductor and the metal workfunction for holes, to the difference between the metal workfunction and the electron affinity of the semiconductor for electrons. R_∞ is the Schottky resistance for an infinite doping concentration i.e., a null Schottky barrier, and its value is taken from the default parameters of Silicon.

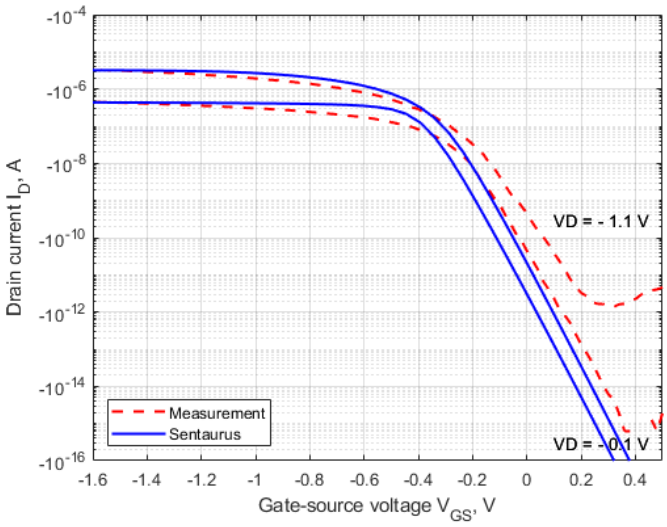
In the following, the 20, 27 and 34 nm devices are simulated in Sentaurus with the Varahramyan model implementing the Schottky contacts. In all the three cases, it can be appreciated that the on-current decreases with respect to the plots shown in the previous section, due to the presence of the calculated Schottky resistance. This allows us to have a better fit on the linear on-current, which now matches in a better way the measurements. Furthermore, it is interesting to notice a curvature of the transfer characteristics at $V_{DS} = -1.1$ V and V_{GS} approaching -1.6 V, particularly visible in the 20 nm device by comparing Figure 4.5 with Figure 4.1, wherein Ohmic contacts are used.

NW diameter = 20 nm

Doping	Cr workfunction	SS	DIBL	$V_{th,lin}$	$V_{th,sat}$
$N_A = 1.2e18\ cm^{-3}$	4.50 eV	73.784 mV/dec	0.052	-0.235 V	-0.182 V



(a) Linear plot

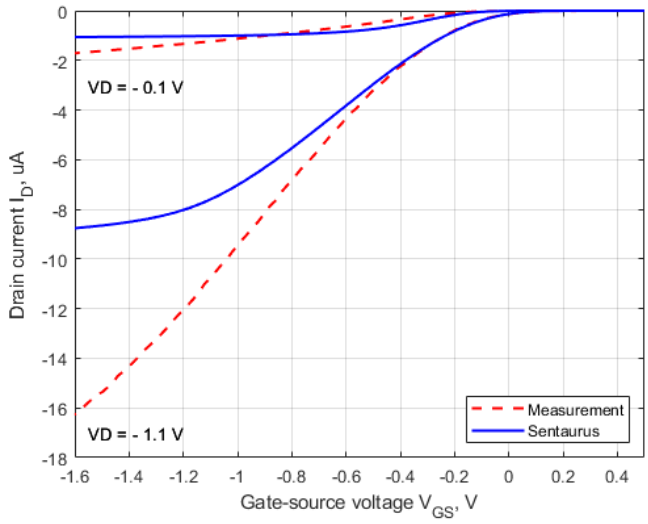


(b) Logarithmic plot

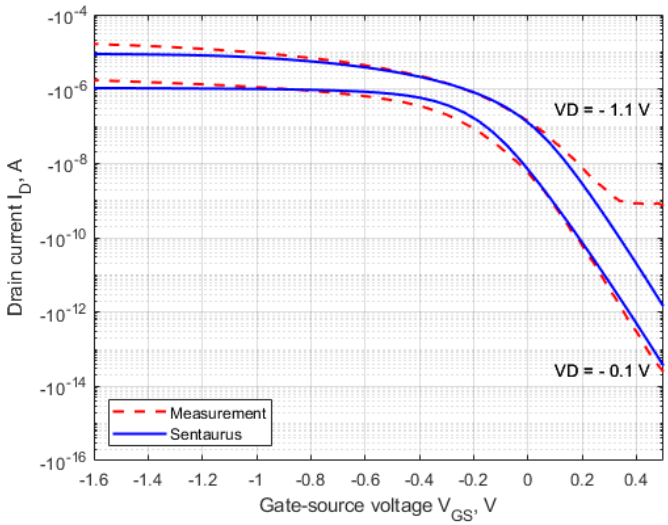
Figure 4.5: Transfer characteristics of 20 nm device with Varahramyan model

NW diameter = 27 nm

Doping	Cr workfunction	SS	DIBL	$V_{th,lin}$	$V_{th,sat}$
$N_A = 2.2e18\ cm^{-3}$	4.36 eV	93.630 mV/dec	0.158	0.028 V	0.185 V



(a) Linear plot

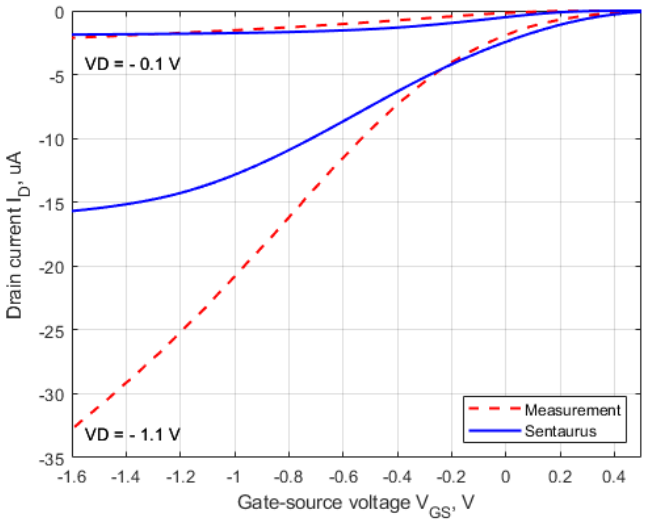


(b) Logarithmic plot

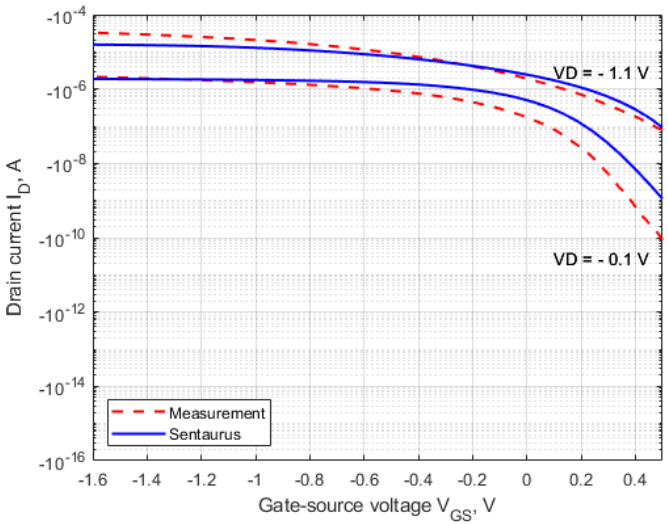
Figure 4.6: Transfer characteristics of 27 nm device with Varahramyan model

NW diameter = 34 nm

Doping	Cr workfunction	SS	DIBL	$V_{th,lin}$	$V_{th,sat}$
$N_A = 2.7e18\ cm^{-3}$	4.25 eV	141.965 mV/dec	0.263	0.237 V	0.500 V



(a) Linear plot



(b) Logarithmic plot

Figure 4.7: Transfer characteristics of 34 nm device with Varahramyan model

4.3 Schottky contacts - thermionic and field emission

The most accurate simulation of the metal - semiconductor interface can be achieved with the Schottky model, wherein the contributions due to thermionic emission and field emission are computed separately. The contribution due to thermionic emission is due to carriers having enough energy to overcome the Schottky barrier, while in the case of field emission carriers can tunnel through a potential barrier, narrowed by the applied V_{SD}.

In the case of Ohmic contacts, as seen previously, the contribution due to field emission is the prevalent one: the Schottky barrier has a low resistance and no rectifying behavior. This approach allows the user to select a wide number of parameters to tune for having an effective calibration of the model such as, for instance, the carriers effective mass for tunneling or the desired model to use for computing the transmission probability across the potential barrier (for FE case).

The main drawback is that it is a computationally expensive model, more than the Varahramyan model presented previously, and the use of non-refined settings can easily make the solver to diverge.

In this section the results obtained with GTS Framework will be exploited to present the model in depth, covering all the main parameters that have been tuned in order to calibrate the model and inspecting the effect that changing said parameters has. Successively, the same model implemented in Sentaurus will be used to demonstrate the 20, 27 and 34 devices.

Interestingly, the ways in which this model is implemented in GTS Framework and Sentaurus are slightly different. In fact, Sentaurus builds a so-called **nonlocal tunneling mesh** at the selected interfaces, which can have a refinement level independent from the device mesh. On the other hand, in GTS Framework the mesh is built at the usual meshing step and thus the tunneling mesh depends on the overall mesh refinement.

It is important to notice that this is the model that will be kept for the further calibration steps, as it yields the most accurate results.

4.3.1 Calibration of Schottky model in GTS Framework

The Schottky model is activated in GTS Framework by declaring it in the conductor / semiconductor interface section of the input-deck file `minimos.ipdm`. This enables the thermionic emission at the metal / semiconductor contacts, for which it is possible to tune the electron and hole masses.

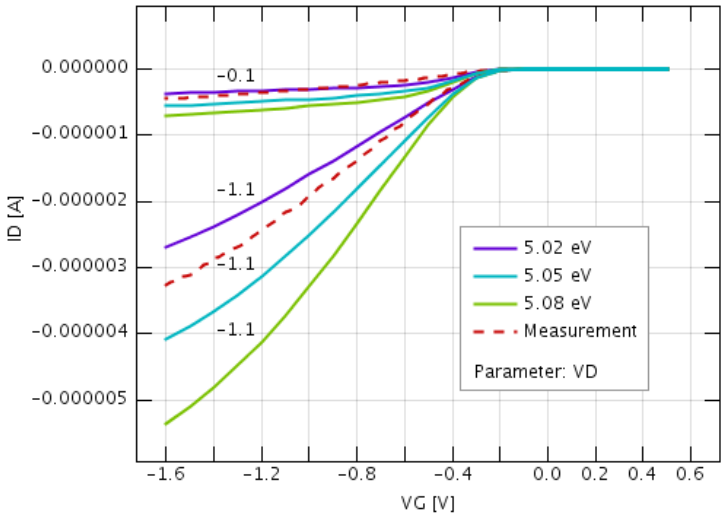
```
FvllmontiConductor_Semiconductor {
    CondSemiInterface {
        Schottky {
            me = 0.3;
            mh = 0.2;
}   }   }
```

Likewise, in order to activate the field emission it is necessary to change the value of parameter `tunnelPermeation` from `"None"` to `"TsuEsaki"`, thus selecting the Tsu - Esaki model for tunneling. The parameter `holePermeation` must be set to `"*"` in order to compute the tunneling transmission probability in all the regions of the model.

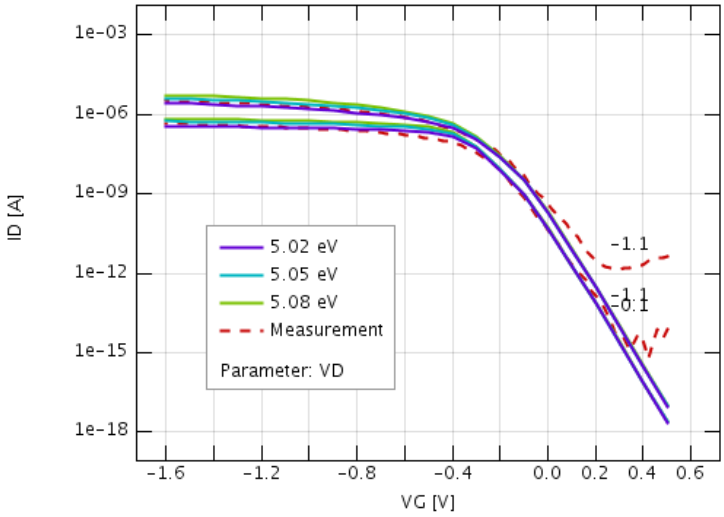
```
+PMosNanowire {
    Hole {
        TunnelMass {
            Const {
                relMass = 0.16;
        }   }
        tunnelPermeation = "TsuEsaki";
        TunnelPermeation {
            TsuEsaki {
                fromSegment = "*Silicide";
                acceleration = 2.0;
                pathLengthLimit = 1e-08 "m";
    }   }   }
    electronPermeation = "";
    holePermeation = "*";
}
```

Platinum Silicide work function

As seen in Equation 4.2, the Schottky barrier height depends on the work function of the metal constituting the contact: selecting a higher work function will result in a decrease of the barrier height, whereas using a lower work function will increase the barrier height, with a subsequent increase of the contact resistance. Starting from the values provided by Drummond [24], where the work function for bulk PtSi is estimated to be $\Phi_m = 5.05\text{eV}$, we tried different values for the work function in order to get the best fit on the output characteristics. As it can be noticed from Figure 4.8, the effect of changing the work function has a very noticeable impact on the output characteristics of the device.



(a) Linear plot



(b) Logarithmic plot

Figure 4.8: Transfer characteristics with different PtSi work functions

As expected, lower work function values yield higher Schottky barrier heights and thus lower currents due to the increased contact resistance, while higher work function values have the opposite effect. Tuning the PtSi work function can be helpful to achieve a better calibration however, the range in which we can pick the values is relatively small, due to the fact that the work function is determined with a good accuracy and there are only a few factors that could introduce a deviation from the reference value, such as, for instance, the formation of small amounts of Pt_2Si, which has a slightly higher work function.

Hole mass in thermionic emission

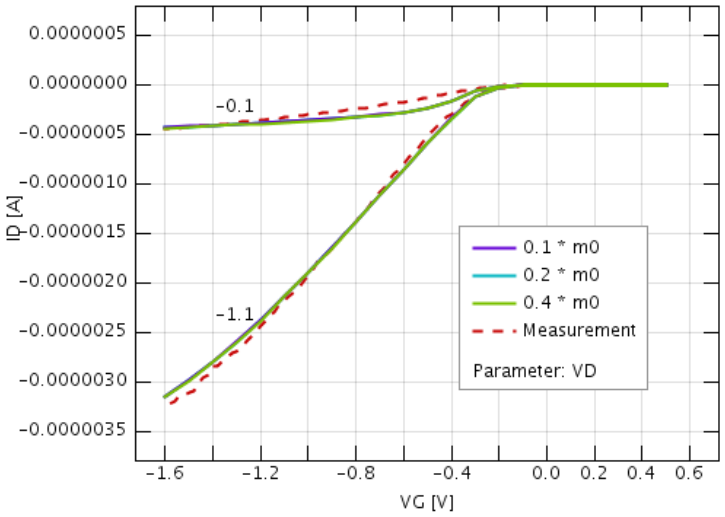
The current density contribution due to thermionic emission is described by the equation:

$$J_p = \left[A^* T^2 \exp\left(-\frac{q\Phi_{Bp}}{kT}\right)\right]\left[\exp\left(\frac{qV}{kT}\right) - 1\right] \tag{4.7}$$

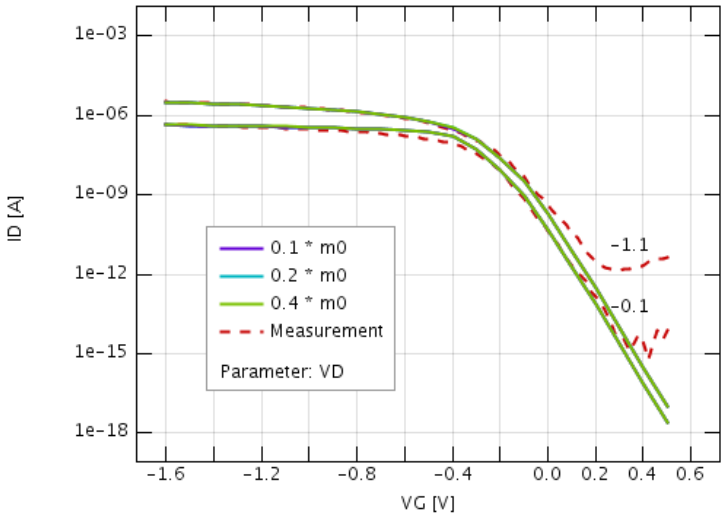
Where A^* is the effective Richardson constant, calculated as:

$$A^* = \frac{4\pi q m^* k^2}{h^3} \tag{4.8}$$

The Richardson constant for the free electron is defined as $A = 120$ A/cm^{2}K^2.
The plots in Figure 4.9 show that the variation of hole mass does not cause visible changes in the IV curves. The reason is that the contribution of thermionic emission is quite small with respect to the current due to field emission. This can be confirmed either by looking at a longitudinal cut in the nanowire, where the current density is displayed, or by disabling the *Tunnel Permeation* option in GTS Framework, thus leaving only thermionic emission as a way to overcome the Schottky barrier.



(a) Linear plot



(b) Logarithmic plot

Figure 4.9: Transfer characteristics with different hole masses for the thermionic emission

Hole mass in field emission

The current density is proportional to the hole effective mass. The tunneling current can be obtained from the Tsu-Esaki formula [25] [26]:

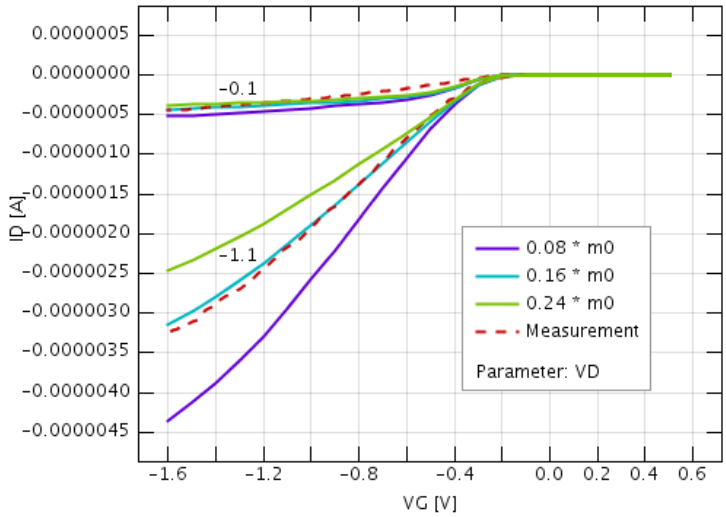
$$J = \frac{4\pi m_h q}{h^3} \int_{\varepsilon_{min}}^{\varepsilon_{max}} TC(\mathcal{E}) N(\mathcal{E}) \mathrm{d}\mathcal{E} \tag{4.9}$$

Where TC is the transmission coefficient of the potential barrier, N is the supply function and $\mathcal{E}$ is the energy. The transmission coefficient is computed using the Wentzel - Kramers - Brillouin (WKB) approximation:

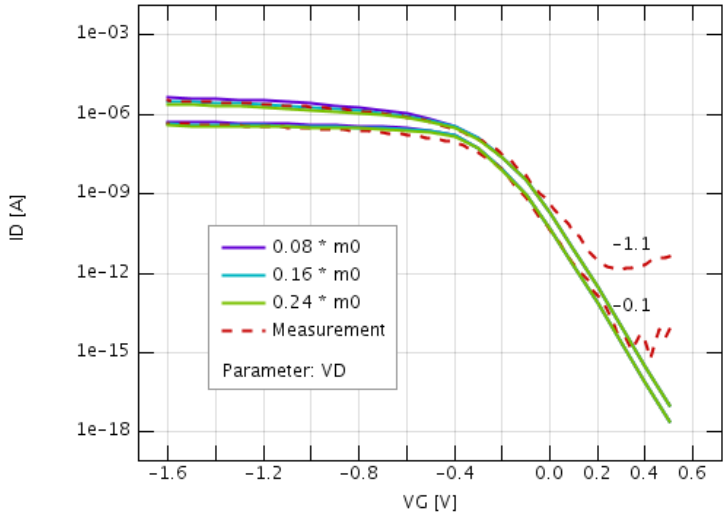
$$TC = \exp\left(-\frac{2}{\hbar} \int_{x_1}^{x_2} \sqrt{2m(W(x) - \mathcal{E})} \mathrm{d}x \right) \tag{4.10}$$

In the case of a linear energy barrier the integral can be solved analytically:

$$\begin{aligned}
TC(\mathcal{E}) &= 1 && \text{for } \mathcal{E} > \Phi \\
TC(\mathcal{E}) &= \exp\left(-\frac{4\sqrt{2 \cdot m}}{3 \cdot q \cdot |\mathbf{E}| \cdot \hbar} \cdot (\Phi - \mathcal{E})^{3/2} \right) && \text{for } \Phi > \mathcal{E} > \Phi_0 \\
TC(\mathcal{E}) &= \exp\left(-\frac{4\sqrt{2 \cdot m}}{3 \cdot q \cdot |\mathbf{E}| \cdot \hbar} \cdot \left[(\Phi - \mathcal{E})^{3/2} - (\Phi_0 - \mathcal{E})^{3/2} \right] \right) && \text{for } \mathcal{E} < \Phi_0
\end{aligned} \tag{4.11}$$



(a) Linear plot



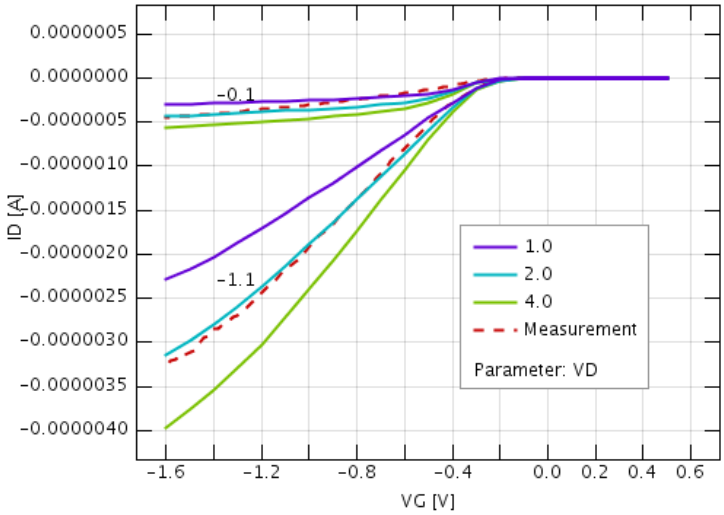
(b) Logarithmic plot

Figure 4.10: Transfer characteristics with different hole masses for the field emission

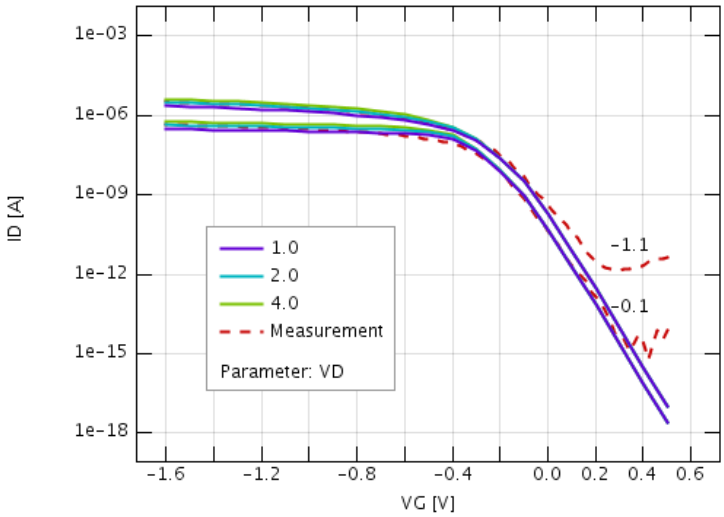
The plots shown in Figure 4.10 highlight the great influence that the hole effective mass has on the ON current. The value that yield the best fit, $m_h^* = 0.16\ m_0$, is coherent with the one used in the work of Ng and Liu [20]. Minor adjustments can be made, however a m_h value outside the range $0.10 - 0.20\ m_0$ may indicate that something in the simulation is not working properly.

Acceleration in field emission

The acceleration is used as a fitting parameter: the tunneling current is directly proportional to the acceleration. It can be noted that the acceleration behaves similarly to the tunnel mass of the hole: doubling the acceleration has roughly the same effect as halving the hole mass.



(a) Linear plot

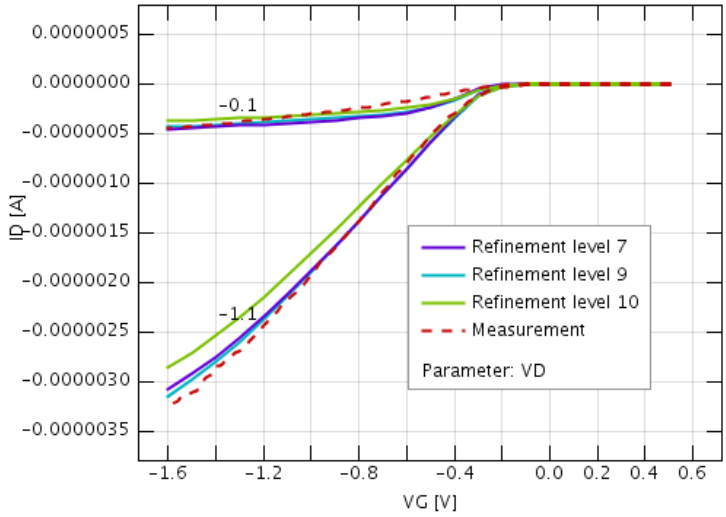


(b) Logarithmic plot

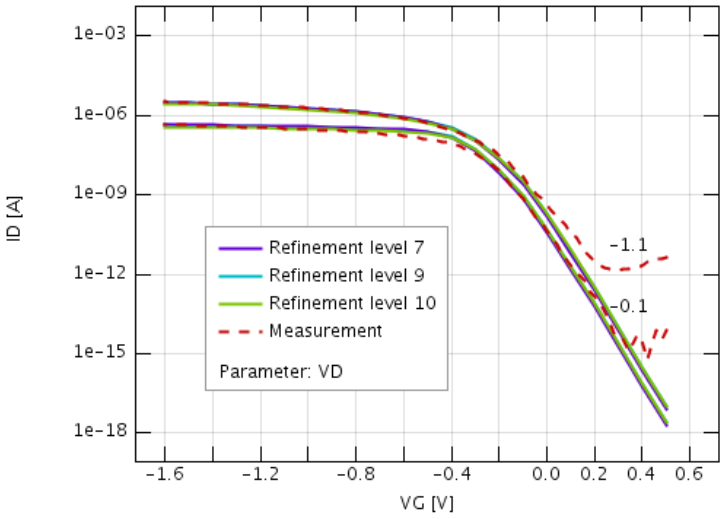
Figure 4.11: Transfer characteristics with different accelerations

Mesh refinement of the nanowire

The refinement level of the mesh plays an important role in the simulation process, as it defines the number of tunnel paths i.e., the segments at the metal-semiconductor interface across which the tunneling current is calculated. Using a non sufficiently refined mesh can lead to an underestimation of the tunneling current, due to the fact that the number of generated tunnel paths is not sufficient to provide a realistic description of the tunneling current.



(a) Linear plot



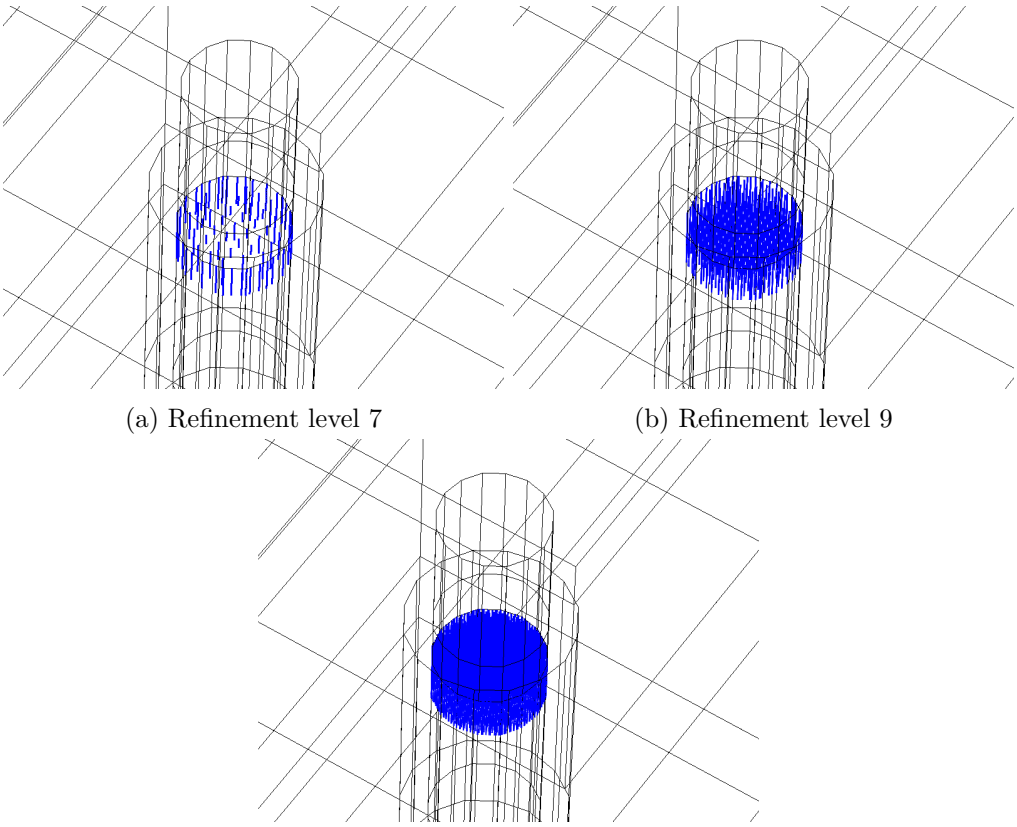
(b) Logarithmic plot

Figure 4.12: Transfer characteristics with different levels of refinement of the nanowire

On the other hand, using a very refined mesh lead to a very computationally expensive simulation. For this reason, it is important to find a good compromise between the mesh refinement level and the obtained output current. The results of three different mesh refinement levels are reported, along with the computation time required for extracting the IV curve.

Refinement level	7	9	10
Total run time	17 min	46 min	15 h, 37 min

To emphasize the impact that the mesh refinement has on the simulation, in Figure 4.13 are shown the tunnel paths of the drain contact in the three different cases, and in Figure 4.13 are shown the respective tunneling meshes generated.



(a) Refinement level 7

(b) Refinement level 9

(c) Refinement level 10

Figure 4.13: Placement of the tunnel paths with three different levels of mesh refinement

4.3.2 Calibration of Schottky model in Sentaurus

In order to activate this model in Sentaurus it is necessary to declare the source and drain contacts as *Schottky* in the *Electrodes* section. In the parenthesis, the **pinning** is activated, with the model of Sze [27] selected, and the workfunction of the metal is defined. The Schottky contacts are implemented in Sentaurus as *boundary conditions*, thus applied to the areas where the contacts are defined, which in this case are places at both ends of the silicon nanowire. In this way there is no need to include the PtSi source and drain contacts in the geometrical description. Conversely, the model used in GTS Framework has the source and drain silicides included in the geometrical description and ideal electrodes placed at one end of the aluminum vias connected to the source and drain silicides.

```
Electrode {
    { name="Source"    Voltage=0.0 Schottky(Pinning(Model="Sze"))
        Workfunction = 5.05 }
    { name="Drain"     Voltage=0.0 Schottky(Pinning(Model="Sze"))
        Workfunction = 5.05 }
    { name="Gate"      Voltage=0.0 }
}
```

Then, enable the `hBarrierTunneling` through the nonlocal mesh "NLM" in the physics section and activate the **barrier lowering**. It is possible to keep the `eBarrierTunneling` deactivated (here commented), since the minority carriers contribution is very little with respect to the conduction current of the device studied.

```
Physics {
    Fermi
    Mobility( DopingDep HighFieldSaturation Enormal )
    EffectiveIntrinsicDensity( Bennett )
    Recombination(SRH( DopingDep ))
    eQuantumPotential
    hQuantumPotential
#   eBarrierTunneling "NLM" (BarrierLowering)
    hBarrierTunneling "NLM" (BarrierLowering)
}
```

Finally, initialize a nonlocal mesh named "NLM" in the *Math* section of `jlnt_pmos_des.cmd`.

```
NonLocal "NLM" (
    Electrode = "Drain"
    Electrode = "Source"
    Length = 1.0e-6
    Digits = 5
    EnergyResolution = 0.001
)
```

The parameter `Length` defines the maximum length for the generated tunneling paths. The parameter `Digits` determines the number of decimal digits to which the integrals reported in equation 4.12 is computed. The parameter `EnergyResolution` sets the lower limit for the energy step that Sentaurus uses to perform the integration, and its purpose is to limit the run time for computing the tunneling currents if the value of `Digits` is too large [21].

The **hole nonlocal tunneling current** for holes that tunnel from the valence band at points above l to the valence band at point l is computed with the integral:

$$\frac{dj_{VV}}{dl}(l) = -q \sum_{\nu} \int_{l}^{\infty} \int_{-\infty}^{\infty} [R_{VV,\nu}(u,l,\varepsilon) - G_{VV,\nu}(u,l,\varepsilon)] d\varepsilon du \tag{4.12}$$

wherein the net recombination rate $R_{VV,\nu}(u,l,\varepsilon) - G_{VV,\nu}(u,l,\varepsilon)$ is expressed as:

$$\begin{aligned} R_{VV,\nu} - G_{VV,\nu} = \frac{A_{VV}}{qk} \vartheta \left[E_{V,\nu}(u) - \varepsilon \, , -\frac{dE_{V,\nu}}{du}(u) \right] \vartheta \left[E_{V,\nu}(l) - \varepsilon \, , -\frac{dE_{V,\nu}}{dl}(l) \right] \times \\ \left[T_n(u) \ln \left(1 + \exp \left[\frac{\varepsilon - E_{F,p}(u)}{kT_n(u)} \right] \right) - T_n(l) \ln \left(1 + \exp \left[\frac{\varepsilon - E_{F,p}(l)}{kT_n(l)} \right] \right) \right] \end{aligned} \tag{4.13}$$

where $\vartheta(x,y) = \delta(x)|y|\Theta(y)$, $A_{VV} = g_V A_0$ wherein A_0 is the Richardson constant $A_0 = 4\pi m_0 kq/h^3$ and g_V is a fit parameters. $\Gamma_{VV,\nu}$ is the tunneling probability, computed according to the Wentzel - Kramers - Brillouin (WKB) approximation.

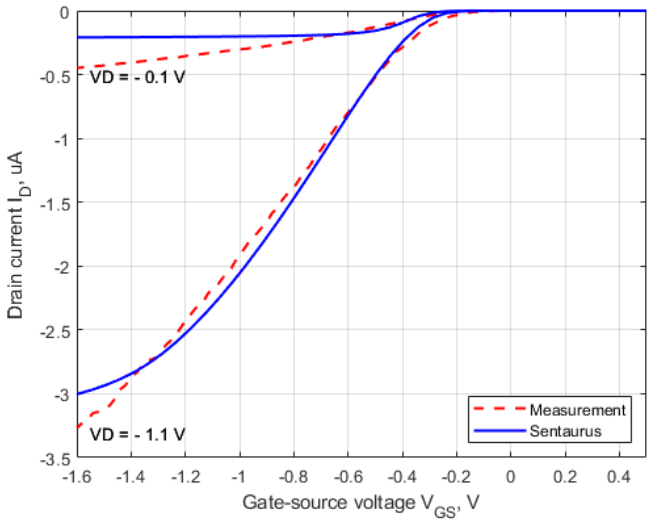
In the following are shown the transfer characteristics extracted from the TCAD model configured with the settings shown above. It is important to note that the results extracted from Sentaurus don't have a fitting factor such as the *Acceleration* parameter used in GTS Framework applied. Conversely, in the 20 nm device model fitted on GTS Framework it has been necessary to use the *Acceleration* parameter equal to 2 in order to obtain a good fit of the on-current, as shown in Figure 4.11.

The transfer characteristics of the 20 nm device, shown in Figure 4.14 shows a sensibly better fit with respect to the one shown for the Varahramyan model in Figure 4.5. This is because the Schottky barrier height selected in this case is slightly higher, $\phi_B = 0.15$ V, and thus a slightly lower on-current at each value considered for V_{DS}.

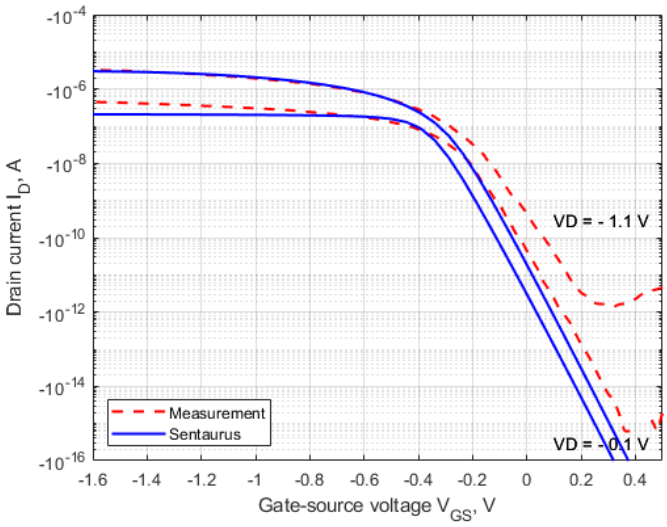
In the case of the 27 nm device, the Cr gate workfunction has been slightly decreased in order to fix some convergence issues of the solver. The fit of the on-current in linear regime is now acceptable, the values of the model are comparable with the measurements of the real device. However, the on-current in saturation still needs work. Similar considerations hold for the 34 nm device. Interestingly, if we compare the plots with the ones obtained with the Varahramyan model, we can see that the *saturation* effect that takes place at $V_{DS} = -1.1$ V is less pronounced and that the saturation on-current is around 2 μA larger.

NW diameter = 20 nm

Doping	Cr workfunction	SS	DIBL	$V_{th,lin}$	$V_{th,sat}$
$N_A = 1.2e18\ cm^{-3}$	4.50 eV	71.584 mV/dec	0.052	-0.235 V	-0.182 V



(a) Linear plot

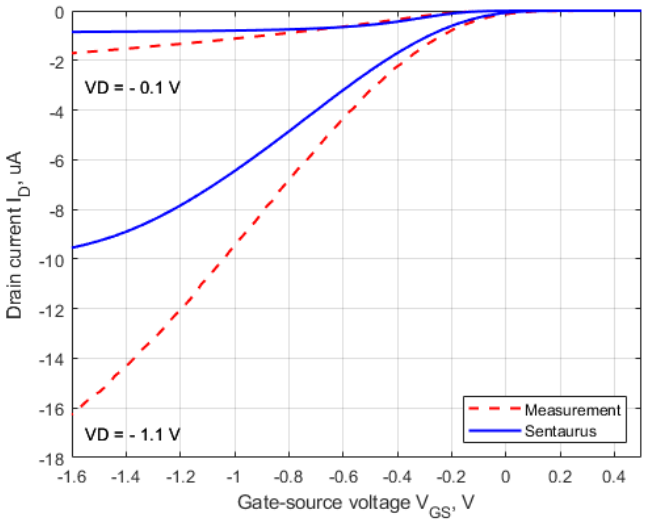


(b) Logarithmic plot

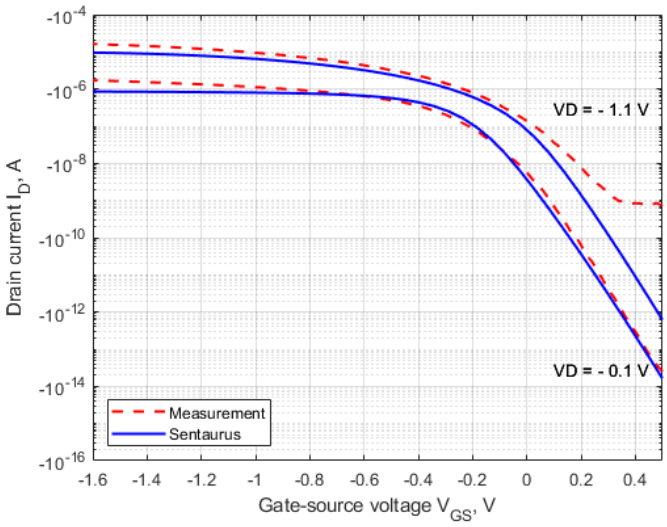
Figure 4.14: Transfer characteristics of 20 nm device with Schottky contacts and nonlocal tunneling at source and drain

NW diameter = 27 nm

Doping	Cr workfunction	SS	DIBL	$V_{th,lin}$	$V_{th,sat}$
$N_A = 2.2e18\ cm^{-3}$	4.33 eV	93.639 mV/dec	0.157	-0.025 V	0.133 V



(a) Linear plot

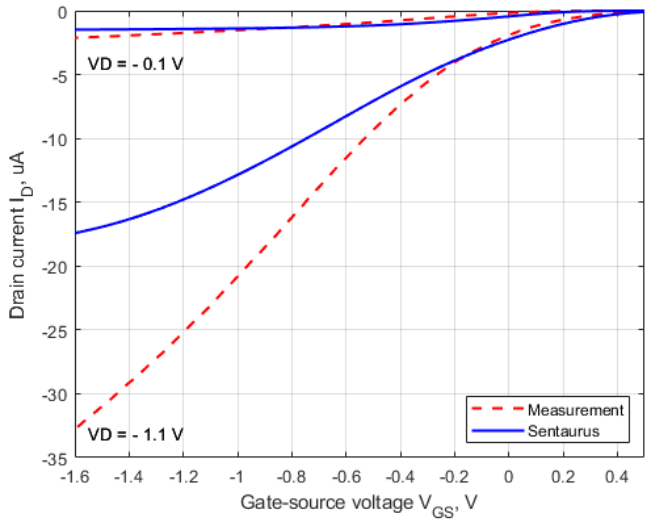


(b) Logarithmic plot

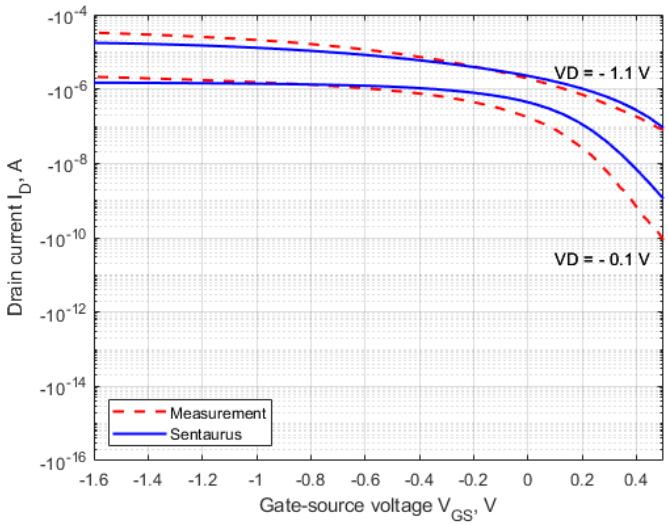
Figure 4.15: Transfer characteristics of 27 nm device with Schottky contacts and nonlocal tunneling at source and drain

NW diameter = 34 nm

Doping	Cr workfunction	SS	DIBL	$V_{th,lin}$	$V_{th,sat}$
$N_A = 2.7e18\ cm^{-3}$	4.25 eV	142.933 mV/dec	0.263	0.237 V	0.500 V



(a) Linear plot



(b) Logarithmic plot

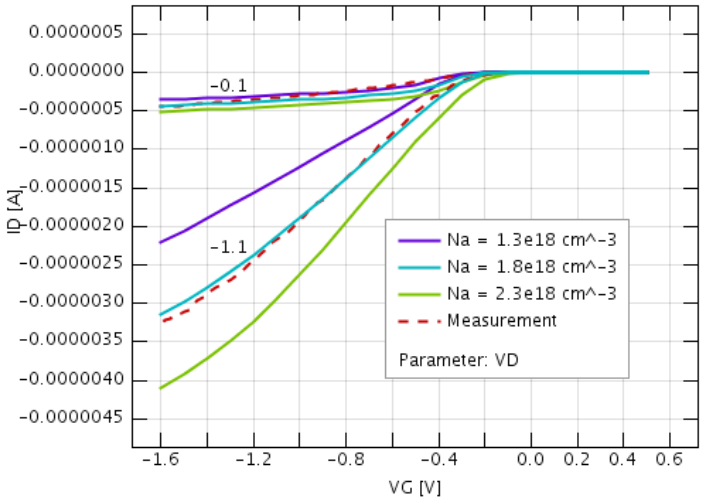
Figure 4.16: Transfer characteristics of 34 nm device with Schottky contacts and nonlocal tunneling at source and drain

Chapter 5

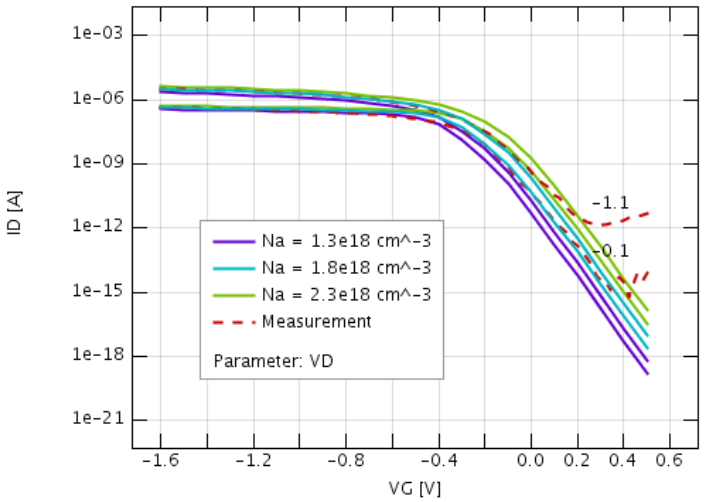
Silicon nanowire

5.1 Doping

Estimating the correct doping concentration is crucial to obtain an effective simulation of the device. The doping controls the total number of carriers available for transport, and heavily influences the threshold voltage and a wide number of transport properties such as the carrier mobility and, more generally, the conductivity of the Silicon nanowire. As a starting point, the doping concentration has been estimated by other partners using a process simulation TCAD (Synopsys Sentaurus Process). Starting from those values, several values of doping density has been tried in order to narrow down the doping range to a more precise value.
As it can be noticed from Figure 5.1, the doping affects the IV curve both in ON and OFF regime. It is worth noticing that during the calibration of a TCAD model, modifying the doping level would affect a broad number of physical phenomena occurring in the device, hence impacting a broad number of electrical characteristics such as threshold voltage, mobility, contacts resistance, gate electrostatic control and short channel effects. Thus, it would be advisable to determine the doping level at the beginning of the calibration process, and then modify it in the middle of the calibration only if strictly necessary.



(a) Linear plot



(b) Logarithmic plot

Figure 5.1: Output characteristics with different doping levels

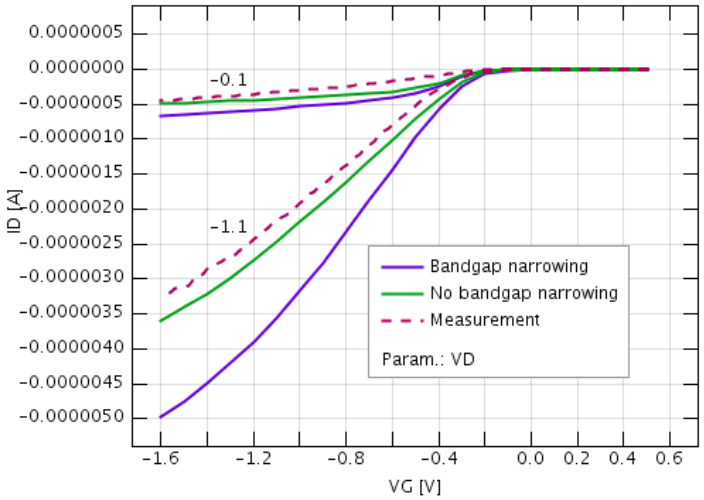
5.2 Bandgap narrowing

In intrinsic semiconductors, the bandgap is determined by the electronic structure of the material and its composition. However, introducing impurities or defects into the semiconductor crystal lattice can lead to a shrinkage of the bandgap. This effect is called **bandgap narrowing** and is substantially ascribed to the emerging of the impurity band formed by the overlapping impurity states [28]: when the dopant concentration is high, the swollen orbitals associated with the impurity atoms begin to overlap and as a result the discrete energy level associated with the impurities broadens to form a band of finite width. Potential fluctuations due to the random distribution of the impurities also lead to a broadening of the impurity band. The ionization level of the impurities is consequently reduced. Eventually, the impurity band overlaps the conduction band or valence band effectively narrowing the bandgap. It's important to note that bandgap narrowing is a complex phenomenon influenced by several factors, including material composition, impurity concentration, strain conditions, and quantum effects. The degree of bandgap narrowing can vary widely depending on these factors and the specific semiconductor material involved. Specifically, also applying mechanical strain to a semiconductor material can modify its band structure and induce bandgap narrowing. Strain alters the interatomic distances and affects the electron energy levels, leading to changes in the bandgap. Compressive strain tends to reduce the bandgap, while tensile strain can increase it. In nanostructured semiconductors, such as quantum dots or nanowires, the confinement of charge carriers in a small volume can lead to quantization effects. The discrete energy levels resulting from quantum confinement can cause the bandgap to become narrower than that of the bulk material. In both TCAD softwares, the model chosen for implementing the bandgap narrowing in the simulation is the Bennett - Wilson [29] model, which was initially developed from absorption and luminescence data of heavily doped n-type materials. In Synopsys Sentaurus the bandgap narrowing would be computed as: $E_{g,eff}(T) = E_g(T) - E_{bgn}$ [21]. In the Bennett - Wilson model, the bandgap narrowing is expressed as follows:

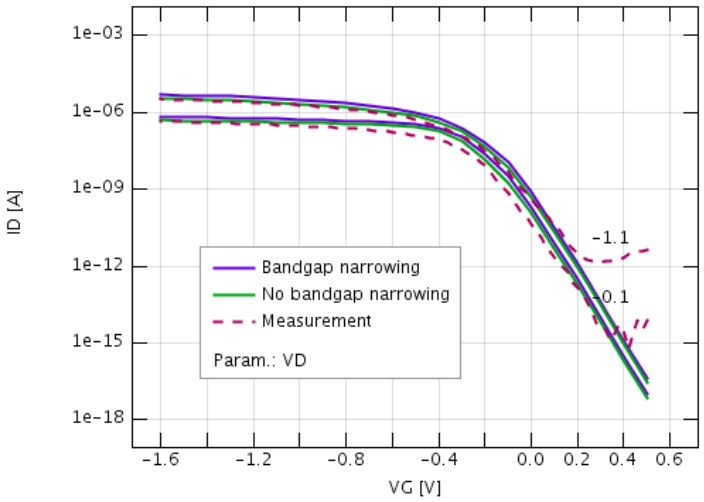
$$E_{bgn} = E_{ref}\left[\ln\left(\frac{N_{tot}}{N_{ref}}\right)\right]^2 \tag{5.1}$$

The parameters N_{ref} and E_{ref} are defined in the parameters file for the Bennett - Wilson model. It can be noted that in this model the bandgap narrowing is defined only if $N_{tot} \geq N_{ref}$. An analogous expression is implemented in the GTS Framework TCAD.

NW diameter = 20 nm



(a) Linear plot



(b) Logarithmic plot

Figure 5.2: Output characteristics of the 20 nm device with bandgap narrowing enabled and disabled

5.3 Impurities at Si/Oxide interface

The presence of impurities in the semiconductor may introduce energy level in the forbidden bandgap. Since the transition probability from one band to another depends exponentially on the energy difference, having trap levels inside the bandgap sensibly increases it. The Shockley-Read-Hall (SRH) recombination-generation rate is expressed as

$$R^{SRH} = \frac{n \cdot p - n_i^2}{\tau_p \cdot (n + n_1) + \tau_n \cdot (p + p_1)} \tag{5.2}$$

with the auxiliary variables n_1 and p_1 defined as

$$n_1 = N_c \cdot \exp\left(\frac{-E_C + E_T}{k_B \cdot T}\right) \qquad p_1 = N_v \cdot \exp\left(\frac{-E_T + E_V}{k_B \cdot T}\right) \tag{5.3}$$

The variables N_C and N_V are the carrier effective density of states. For a trap energy level E_T located in the middle of the bandgap, $n_1 = p_1 = n_i$ and the recombination rate is maximum. The thermal carrier velocity v_ν and the recombination lifetime τ_ν are expressed as:

$$v_\nu = \sqrt{\frac{3 \cdot k_B \cdot T}{m_\nu}} \qquad \tau_\nu = \frac{1}{\sigma_{T,\nu} \cdot N_T \cdot v_\nu + S_\nu / y} \tag{5.4}$$

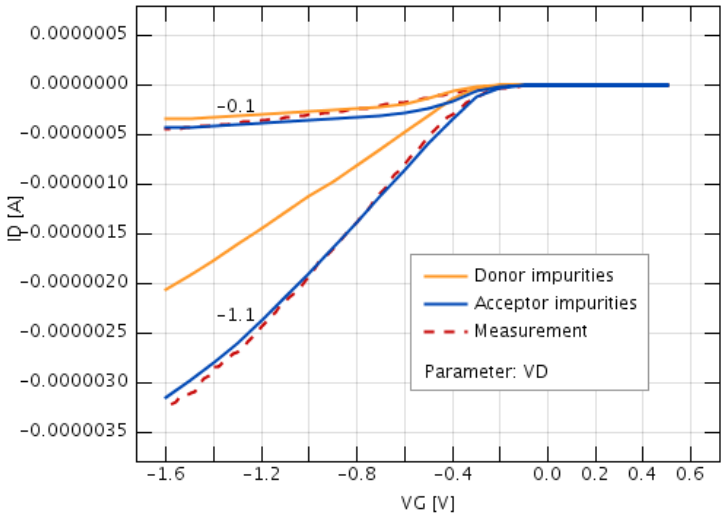
where N_T is the trap density, $\sigma_{T,\nu}$ the trap capture cross section, S_ν the surface recombination velocity and y the distance to the surface. The SRH recombination is enabled and impurities are added at the Silicon - oxide interface. The impurities can be either acceptors or donors: ionized electron donors will release an electron in the conduction band, whereas ionized electron acceptors will release a free hole in the valence band. The ionization of the impurity atoms will affect the electrostatics of the device due to the fact that they are fixed charges located under the gate.

In order to model the presence of impurities at the silicon/oxide interface we can tune different parameters. Initially we adjust the trap concentration N_T and the trap energy E_T, assuming a constant and uniform energy level for the traps. Then, we will try to increase the variability of the energy levels by selecting, for instance, a Gaussian distribution for the trap energy levels. In both TCAD softwares there are several distributions that can be selected, such as uniform, exponential and Gaussian, and it is also possible to insert user-defined tables containing the trap energy levels.

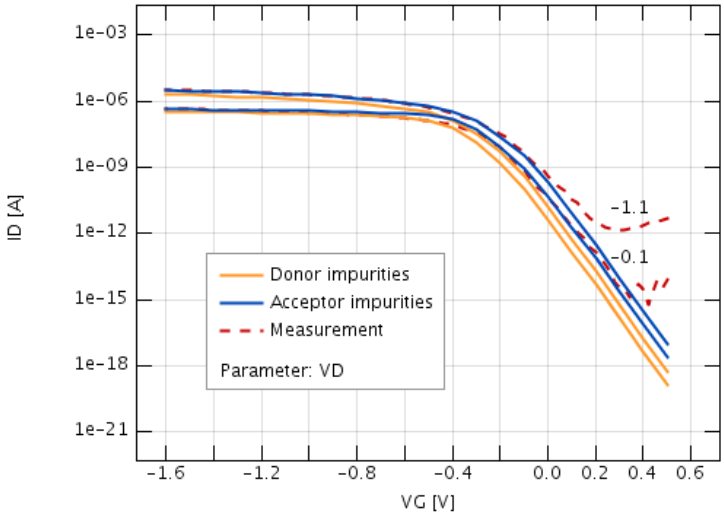
As it will be seen in the following sections, a correct modeling of the impurities at the interface is essential for obtaining a good representation of the subthreshold behavior. In particular, subthreshold slope (SS) and drain induced barrier lowering (DIBL) are sensibly affected by the impurities at the silicon/oxide interface, hence a good check for asserting the validity of the tuning is to compare these two figures of merit with the ones extracted from the measurements.

5.3.1 Adding interface traps in GTS Framework

The selected trap density is $N_T = 3 \cdot 10^{11}\text{cm}^{-3}$ and the energy is $E_T = -0.3\text{eV}$.



(a) Linear plot



(b) Logarithmic plot

Figure 5.3: Transfer characteristics with SRH recombination enabled

5.3.2 Adding interface traps in Sentaurus

The interface trap levels can be added in Sentaurus by specifying them in the *Physics* section of the Silicon/SiO_2 interface. The following code snippet shows the the settings for acceptor impurities, uniformly placed at -0.3 eV from the midgap and with a concentration $N_T = 4.5 \cdot 10^{11}$ cm^{-3}. The electron and hole capture cross sections are set to default values for Si, $1 \cdot 10^{-14}$ cm^2.

```
Physics(MaterialInterface = "Silicon/SiO2"){
    Traps(
        Acceptor Level EnergyMid=-0.3 fromMidBandGap
        Conc=4.5e11 eXsection=1e-14 hXsection=1e-14
    )
}
```

Similarly, to define impurities with Gaussian-distributed energy levels around a mean value `EnergyMid` = -0.3, with standard deviation `EnergySig` = 0.2 one must write:

```
Physics(MaterialInterface = "Silicon/SiO2"){
    Traps(
        Acceptor Gaussian fromMidBandGap Conc=4.5e11
        EnergyMid=-0.3 EnergySig=0.2
        eXsection=1e-14 hXsection=1e-14
    )
}
```

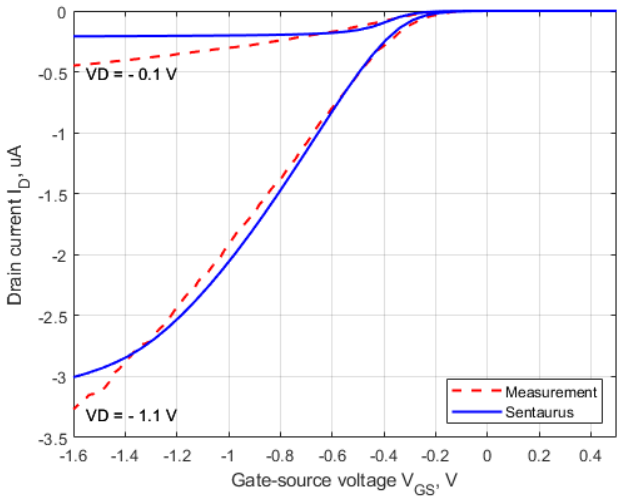
In the following, we will simulate first the three different devices with a single trap energy level, then we'll move to Gaussian-distributed energy levels.

Firstly, it is interesting to notice that in order to obtain a good fit on the DIBL and subthreshold slope, we have to specify trap densities that increase with the nanowire diameter. A value $N_T = 4.0 \cdot 10^{11}$ cm^{-3} has been used for the 20 nm device, up to $N_T = 5.0 \cdot 10^{11}$ cm^{-3} for the 34 nm device. The difference is directly connected to the fabrication process of the devices, in particular to the oxidation steps that are necessary to form the gate oxide.

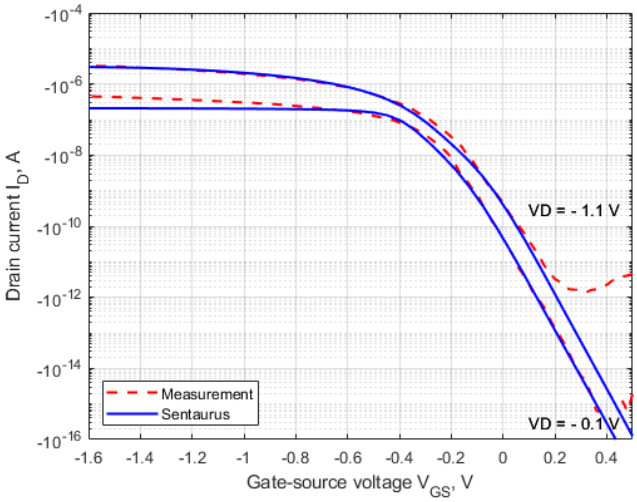
As it can be noted, introducing a Gaussian energy distribution causes a shift of the threshold voltage towards negative values and a slight decrease of the subthreshold slope. This is due to the fact that, since the energy of the trapping centers has changed, also the occupation probability of the trapping centers has changed. Recalling that an acceptor trap has neutral charge when unoccupied and carries the charge of an electron when occupied, we can understand that the average occupation of traps using the Gaussian energy distribution is higher, due to the shift towards negative values. This is reasonable if we think that using such distribution, a certain number of trapping centers will be *shallower* and thus closer to the valence band edge and more likely to be occupied.

NW diameter = 20 nm

Doping	Cr workfunction	SS	DIBL	$V_{th,lin}$	$V_{th,sat}$
N_A = 1.2e18 cm^{-3}	4.50 eV	80.840 mV/dec	0.105	-0.235 V	-0.130 V
Trap type	E_T distribution	N_T	E_T	σ_T	
Acceptor	Single level	4.0e11 cm^{-3}	-0.3 eV	–	



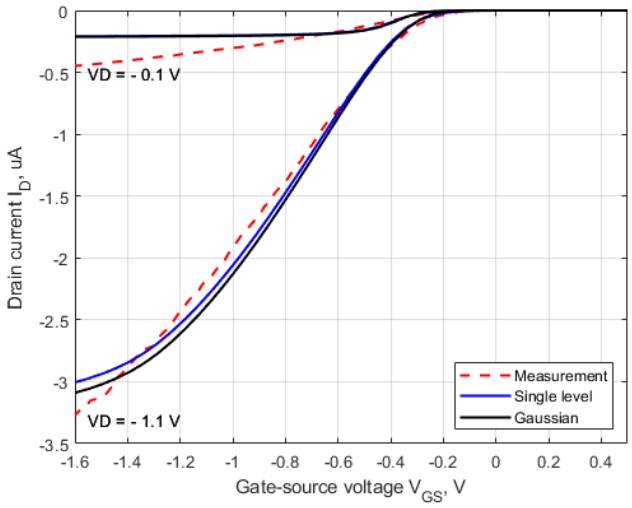
(a) Linear plot



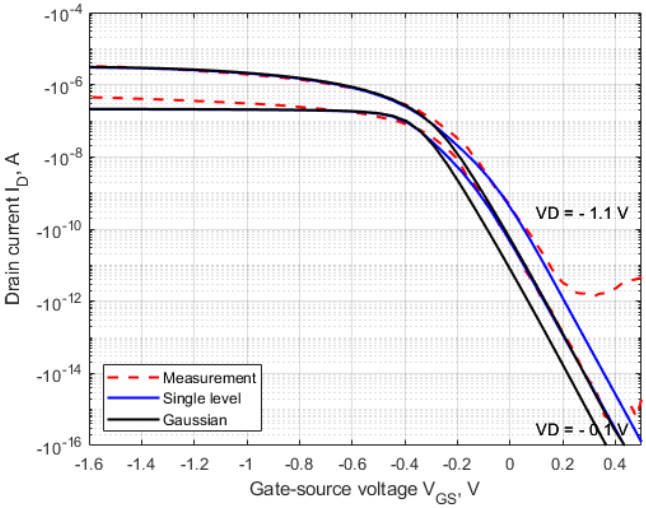
(b) Logarithmic plot

Figure 5.4: Transfer characteristics of 20 nm device with interface traps, single trap energy

Doping	Cr workfunction	SS	DIBL	$V_{th,lin}$	$V_{th,sat}$
$N_A = 1.2e18$ cm^{-3}	4.50 eV	76.815 mV/dec	0.052	-0.235 V	-0.182 V
Trap type	E_T distribution	N_T	E_T	σ_T	
Acceptor	Gaussian	$3.5e11$ cm^{-3}	-0.3 eV	0.2 eV	



(a) Linear plot

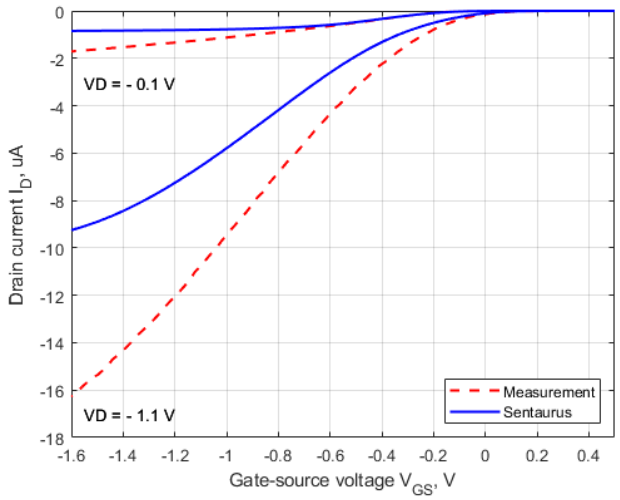


(b) Logarithmic plot

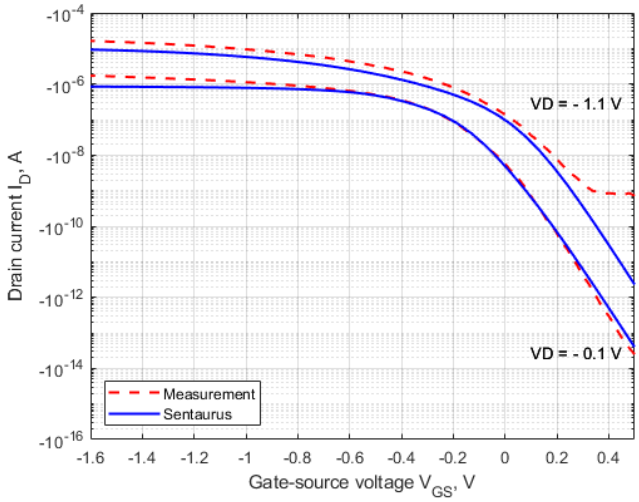
Figure 5.5: Transfer characteristics of 20 nm device with interface traps, traps energy levels Gaussian distributed. A slightly lower traps concentration has been used for stability.

NW diameter = 27 nm

Doping	Cr workfunction	SS	DIBL	$V_{th,lin}$	$V_{th,sat}$
N_A = 2.2e18 cm^{-3}	4.25 eV	98.266 mV/dec	0.210	-0.025 V	0.185 V
Trap type	E_T distribution	N_T	E_T	σ_T	
Acceptor	Single level	4.5e11 cm^{-3}	-0.3 eV	–	



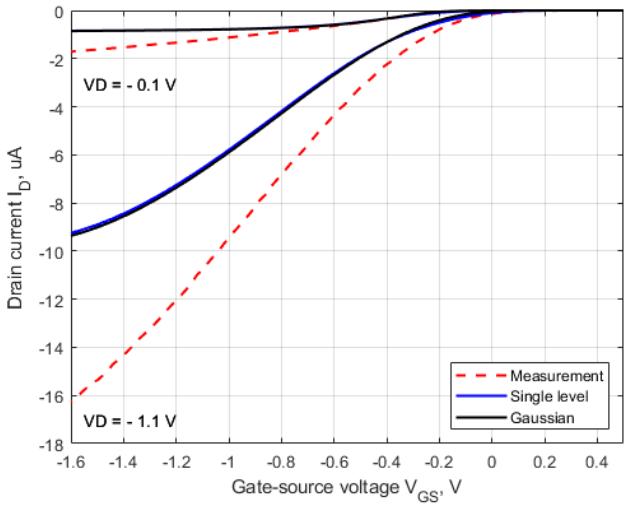
(a) Linear plot



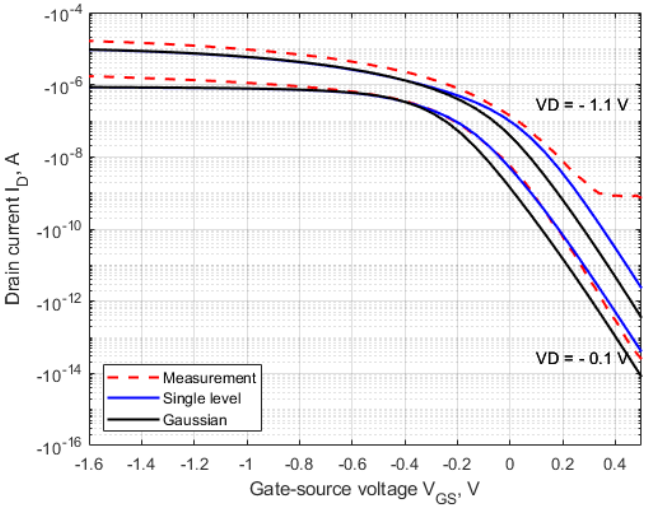
(b) Logarithmic plot

Figure 5.6: Transfer characteristics of 27 nm device with interface traps, single trap energy

Doping	Cr workfunction	SS	DIBL	$V_{th,lin}$	$V_{th,sat}$
$N_A = 2.2e18$ cm^{-3}	4.25 eV	96.948 mV/dec	0.157	-0.077 V	0.080 V
Trap type	E_T distribution	N_T	E_T	σ_T	
Acceptor	Gaussian	$4.5e11$ cm^{-3}	-0.3 eV	0.2 eV	



(a) Linear plot

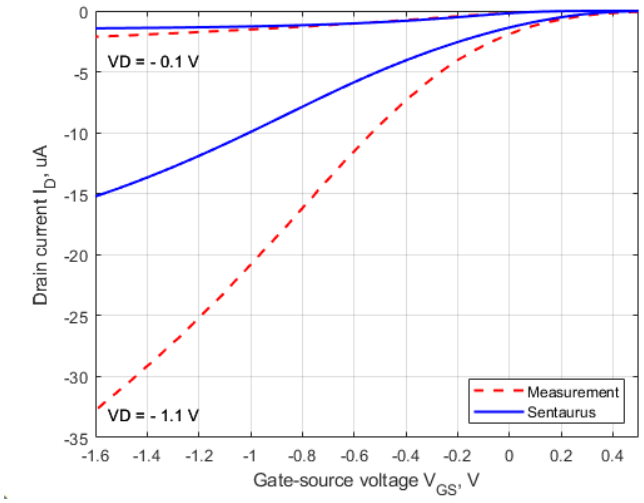


(b) Logarithmic plot

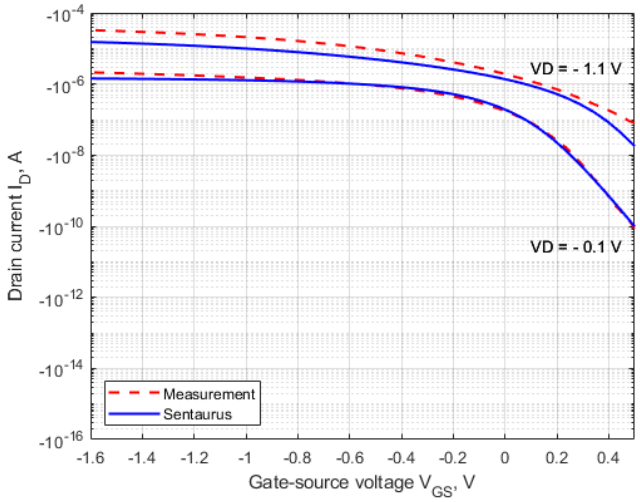
Figure 5.7: Transfer characteristics of 27 nm device with interface traps, traps energy levels Gaussian distributed

NW diameter = 34 nm

Doping	Cr workfunction	SS	DIBL	$V_{th,lin}$	$V_{th,sat}$
N_A = 2.7e18 cm^{-3}	4.00 eV	138.437 mV/dec	0.315	0.080 V	0.395 V
Trap type	E_T distribution	N_T	E_T	σ_T	
Acceptor	Single level	5.0e11 cm^{-3}	-0.3 eV	–	



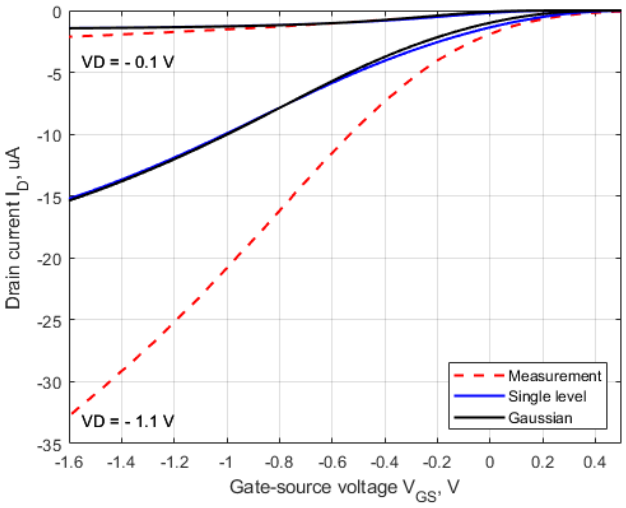
(a) Linear plot



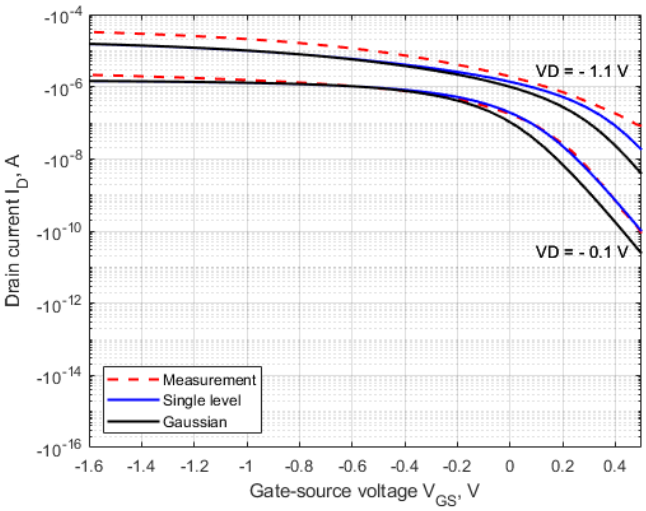
(b) Logarithmic plot

Figure 5.8: Transfer characteristics of 34 nm device with interface traps, single trap energy

Doping	Cr workfunction	SS	DIBL	$V_{th,lin}$	$V_{th,sat}$
$N_A = 2.7e18$ cm^{-3}	4.00 eV	132.610 mV/dec	0.315	0.028 V	0.343 V
Trap type	E_T distribution	N_T	E_T	σ_T	
Acceptor	Gaussian	$5.0e11$ cm^{-3}	-0.3 eV	0.2 eV	



(a) Linear plot



(b) Logarithmic plot

Figure 5.9: Transfer characteristics of 34 nm device with interface traps, traps energy levels Gaussian distributed

5.4 Band to band tunneling

The band to band tunneling (BTBT) is necessary to describe the GIDL (Gate Induced Drain Leakage). The direct band to band tunneling describes the carrier generation in the high field region without any influence of local traps i.e., the field emission of valence electrons leaving back holes. A commonly used model for describing the phenomenon is the Hurkx model [30]. This empirical model treats separately the rates of recombination and generation. Generation occurs in reverse bias i.e. when $n_i^2 - n \cdot p \gg 0$, whereas recombination occurs in forward bias, when $n_i^2 - n \cdot p \ll 0$. The rate is calculated as

$$R^{BB} = B \cdot \left(\frac{|\mathbf{E}|}{1 \frac{V}{cm}} \right)^{\sigma} \cdot D \cdot \exp\left(- \frac{E_{BB}}{|\mathbf{E}|} \right) \tag{5.5}$$

To have a better understanding of the process, it is interesting to see *where* the band to band tunneling causes an increase of carrier concentration in the device. To this end, we can look at a cut of the nanowire taken at $V_G = 0.5$ V and $V_D = -1.1$ V that shows the SRH net recombination rate, in the cases of no band to band tunneling and interface traps (Figure 5.10), only with interface traps (Figure 5.11), and with interface traps and band to band tunneling (Figure 5.12). It can be noticed how enabling the band to band tunneling changes the net recombination ratio: under the gate there are no more areas in which the carriers recombine and the generation ratio sensibly increases. Conversely, near the drain contact the generation ratio is slightly lower than in the cases without BTBT.

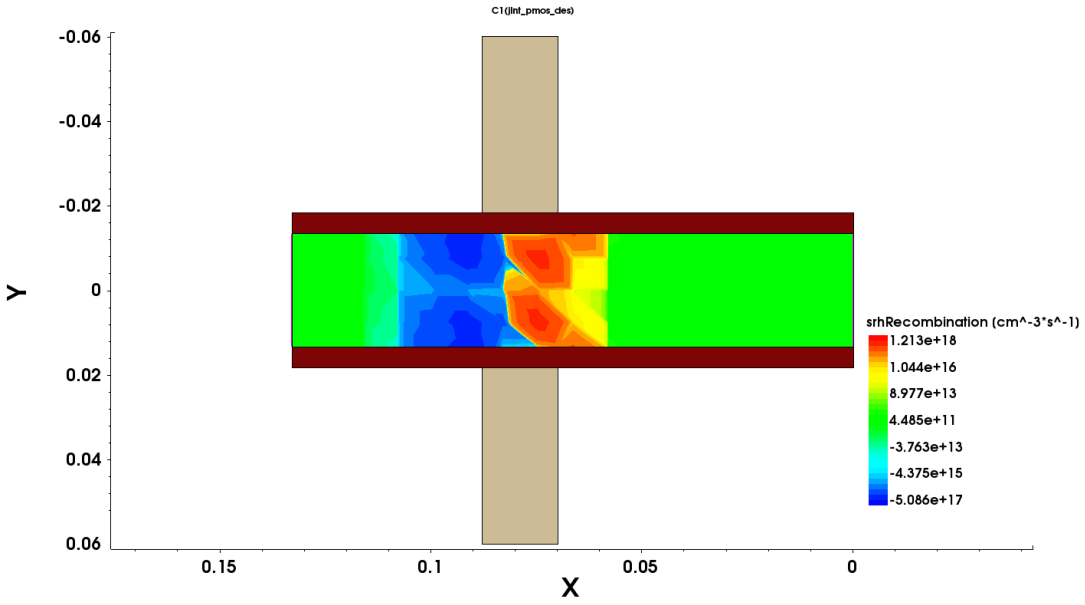


Figure 5.10: Z-cut of the 27 nm device at $V_G = 0.5$ V and $V_D = -1.1$ V showing the SRH net recombination ratio. No band to band tunneling and interface traps.

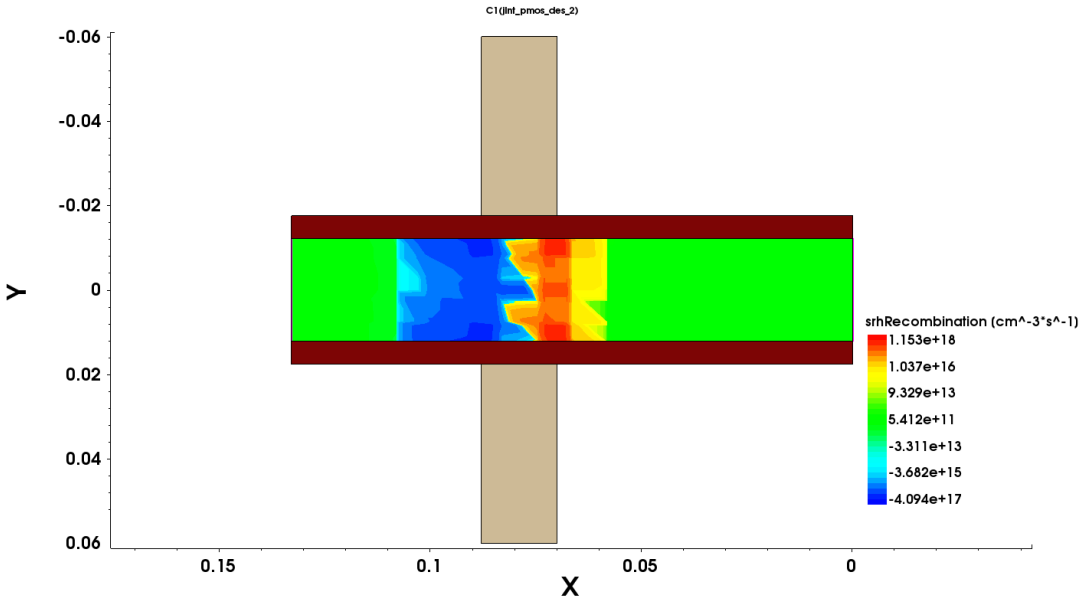


Figure 5.11: Z-cut of the 27 nm device at $V_G = 0.5$ V and $V_D = -1.1$ V showing the SRH net recombination ratio. Interface traps present, band to band tunneling disabled.

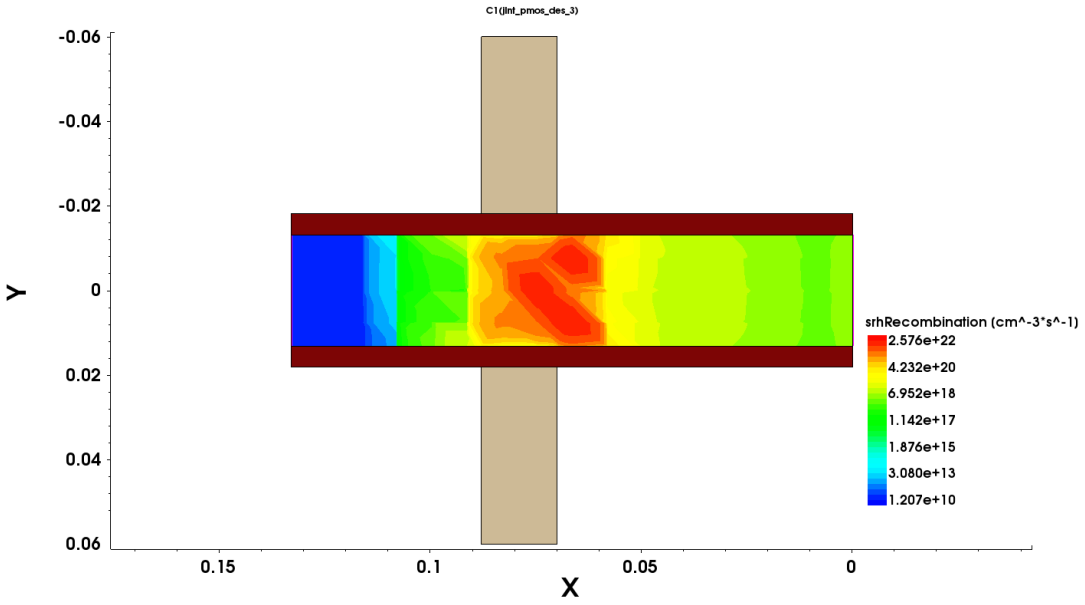
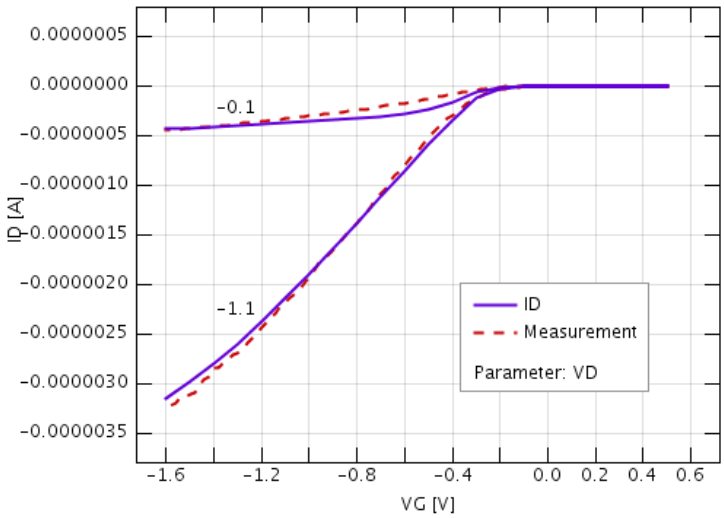


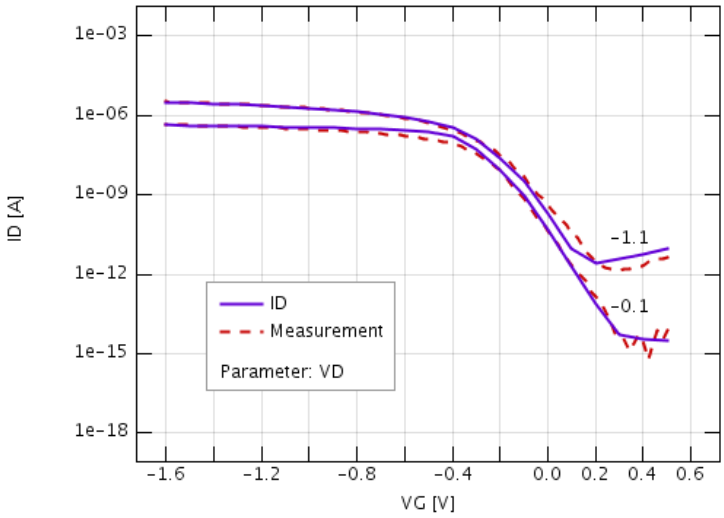
Figure 5.12: Z-cut of the 27 nm device at $V_G = 0.5$ V and $V_D = -1.1$ V showing the SRH net recombination ratio. Interface traps present, band to band tunneling enabled.

5.4.1 Hurkx model in GTS Framework

NW diameter = 20 nm



(a) Linear plot



(b) Logarithmic plot

Figure 5.13: Transfer characteristics with band to band tunneling enabled

The results shown in Figure 5.13 are obtained using the following values:

E_{BB}	$5.5 \cdot 10^8$ V/m
B	$3.5 \cdot 10^{27}$ 1/m$^3 \cdot$ s
σ	2.0

5.4.2 Hurkx model in Sentaurus

Band to band tunneling is enabled by specifying under *Recombination* of the *Physics* section of file `jlnt_pmos.des` the following. It is possible to select different models, and according to the model selected the parameters are specified in the parameter file.

```
Recombination(
    SRH( DopingDep )
    Band2Band (
        Model = Hurkx
        DensityCorrection = Local
        ParameterSetName = ("myBTBT") )
)
```

According to the Sentaurus Device user manual [21], the net recombination rate given by the Hurkx model [30] is expressed as:

$$R_{net}^{bb} = A \cdot D \cdot \left(\frac{F}{1\ \mathrm{V/cm}} \right)^{P} exp\left(\frac{B E_g T^{3/2}}{E_g (300\mathrm{K})^{3/2} F} \right) \tag{5.6}$$

with F being the electric field, A a recombination rate, B an electric field threshold

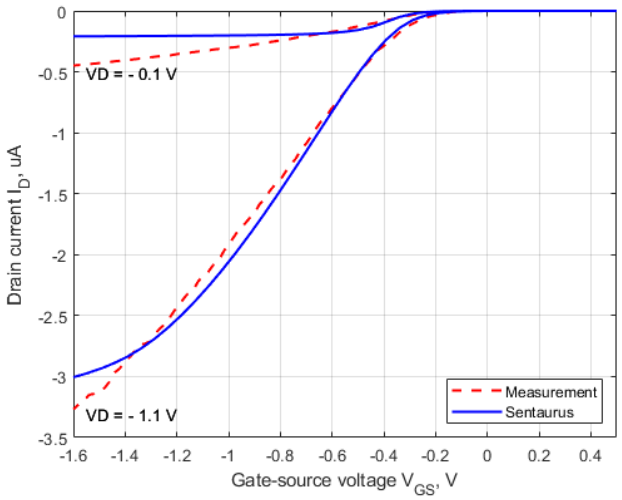
$$D = \frac{np - n_{i,eff}^2}{(n + n_{i,eff})(p + n_{i,eff})} (1 - |\alpha|) + \alpha \tag{5.7}$$

The code snipped below shows the parameters for the Hurkx model in Sentaurus. In this case, these parameters belong to `Region = "core"`, in such a way that only the silicon nanowire will use said values. The parameters are declared separately for the cases of generation and recombination, and the parameter `alpha` allows to select, if desired, exclusively the case of generation or recombination.

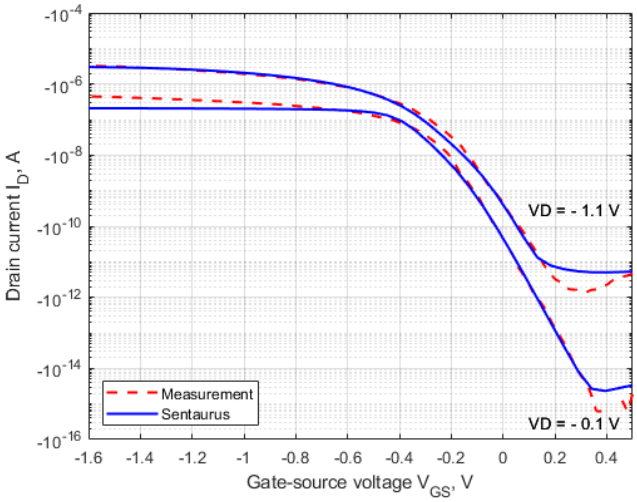
```
Band2BandTunneling "myBTBT" {
    Agen = 3.5e15   # [1/(cm 3 s)]
    Bgen = 3.5e6    # [V/cm]
    Pgen = 2.0      # [1]
    Arec = 3.5e15   # [1/(cm 3 s)]
    Brec = 3.5e6    # [V/cm]
    Prec = 2.0      # [1]
    alpha = 0       # [1] 1 = R  0 = R/G  -1 = G
}
```

NW diameter = 20 nm

Doping	Cr workfunction	SS	DIBL	$V_{th,lin}$	$V_{th,sat}$
$N_A = 1.2e18$ cm^{-3}	4.50 eV	100.900 mV/dec	0.105	-0.393 V	-0.287 V
A_{gen}	B_{gen}	P_{gen}	α		
$5.5e17$ 1/cm$^3 \cdot$ s	$5.5e6$ V/cm	2.0	-1		



(a) Linear plot

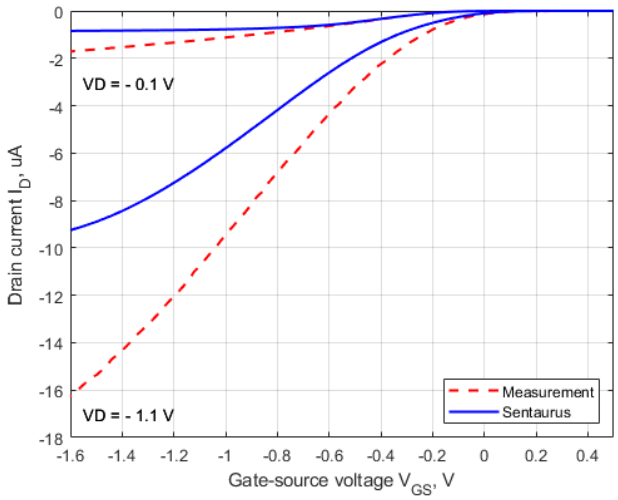


(b) Logarithmic plot

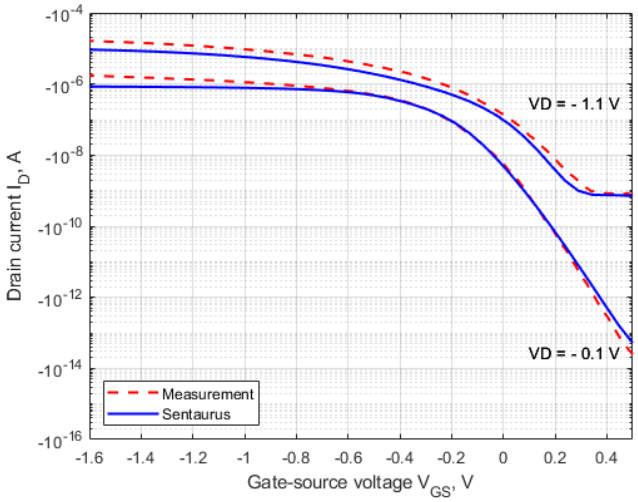
Figure 5.14: Transfer characteristics of 20 nm device with band to band tunneling enabled

NW diameter = 27 nm

Doping	Cr workfunction	SS	DIBL	$V_{th,lin}$	$V_{th,sat}$
$N_A = 2.2e18$ cm^{-3}	4.25 eV	99.426 mV/dec	0.210	-0.025 V	0.185 V
A_{gen}	B_{gen}	P_{gen}	α		
$3.5e15$ 1/cm$^3 \cdot$ s	$3.5e6$ V/cm	2.0	0		



(a) Linear plot

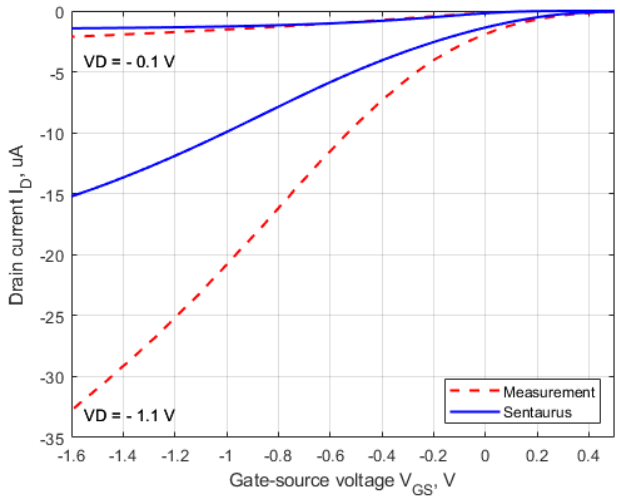


(b) Logarithmic plot

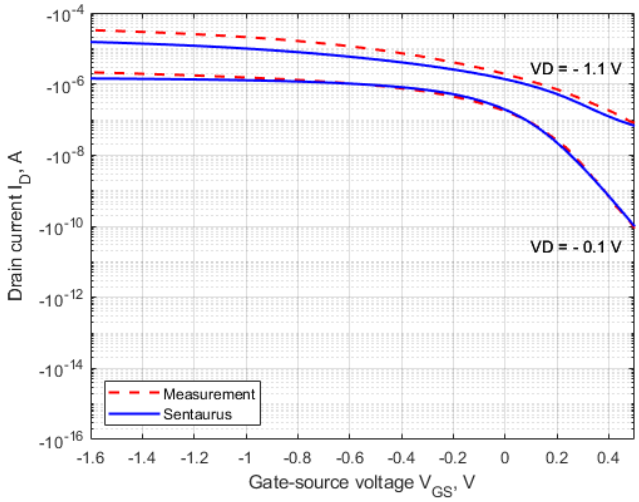
Figure 5.15: Transfer characteristics of 27 nm device with band to band tunneling enabled

NW diameter = 34 nm

Doping	Cr workfunction	SS	DIBL	$V_{th,lin}$	$V_{th,sat}$
$N_A = 2.7e18$ cm^{-3}	4.00 eV	138.441 mV/dec	0.367	0.080 V	0.448 V
A_{gen}	B_{gen}	P_{gen}	α		
$3.5e14$ 1/cm$^3 \cdot$ s	$3.5e6$ V/cm	2.0	0		



(a) Linear plot



(b) Logarithmic plot

Figure 5.16: Transfer characteristics of 34 nm device with band to band tunneling enabled

5.5 Velocity saturation

Velocity saturation refers to a phenomenon that occurs when the speed of charge carriers reaches a maximum value and does not increase further even with an increase in the electric field. This saturation effect is primarily observed in high electric fields or high carrier concentrations. In an ideal semiconductor, the electron drift velocity increases linearly with the applied electric field. However, in reality, due to various scattering mechanisms, such as phonon scattering, impurity scattering, and carrier-carrier scattering, the electron velocity does not continue to increase indefinitely with the electric field strength. As the electric field increases, the scattering mechanisms become more pronounced, leading to a reduction in the average time between scattering events. At high electric fields, the scattering events become so frequent that the carriers do not gain additional momentum between the scattering events. As a result, the carrier velocity reaches a maximum, beyond which it does not increase further. This maximum velocity is referred to as the velocity saturation. In silicon, velocity saturation is more pronounced for electrons than for holes. This is because electrons have a higher mobility and, therefore, experience fewer scattering events compared to holes. As a result, electrons in silicon exhibit a higher velocity saturation compared to holes. Velocity saturation has significant implications for the performance of electronic devices, particularly in high-frequency and high-power applications. It limits the maximum achievable electron velocity and, consequently, the saturation velocity determines the maximum frequency at which a device can operate efficiently.
In Sentaurus, the default model for implementing velocity saturation is the Canali [31] model, which originates from the Caughey-Thomas [32] formula further including temperature dependent parameters:

$$v_{sat} = v_{sat,0}\left(\frac{300\ \mathrm{K}}{T}\right)^{v_{sat,exp}} \tag{5.8}$$

The Canali model yields also the high field mobility whose effect will be addressed more in depth in the following chapter. The velocity saturation can be enabled by specifying the `HighFieldSaturation` keyword in the *mobility* parameters of the physics section. The velocity saturation value can then be tuned in the parameters file:

```
HighFieldDependence "myHighField" {
* Caughey-Thomas model:
* mu_highfield = ( (alpha+1)*mu_lowfield ) /
* ( alpha + ( 1 + ( (alpha+1)*mu_lowfield*E/vsat)^beta )^(1/beta) )
* beta = beta0 (T/T0)^betaexp.
beta0 = 1.109 ,1.213 # [1]
betaexp = 0.66 ,0.17 # [1]
alpha = 0.0000e+00 ,0.0000e+00 # [1]

* For vsat either Formula1 or Formula2 can be used.
Vsat_Formula = 1 ,1 # [1]
* Formula1 for saturation velocity:
```

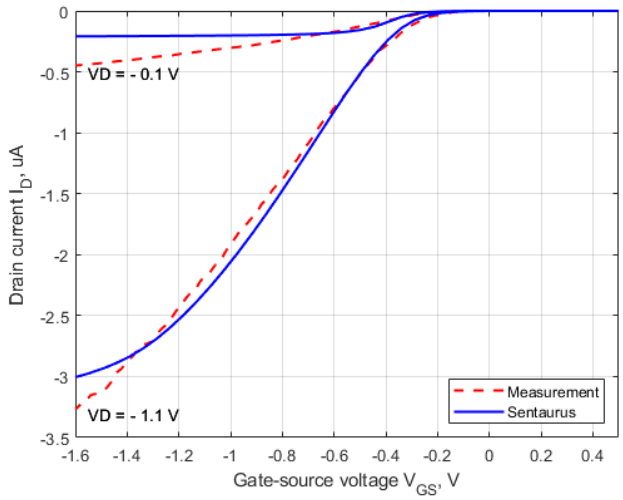
```
* vsat = vsat0 (T/T0)^(-Vsatexp)
* (Parameter Vsat_Formula has to be not equal to 2)
vsat0 = 1.0700e+07 ,8.3700e+06 # [1]
vsatexp = 0.87 ,0.52 # [1]

* Mobility scaling parameters:
* mu_lowfield -> ku * mu_lowfield
* vsat        -> kv * vsat
ku = 1 ,1.0 # [1]
kv = 1 ,2.0 # [1]
}
```

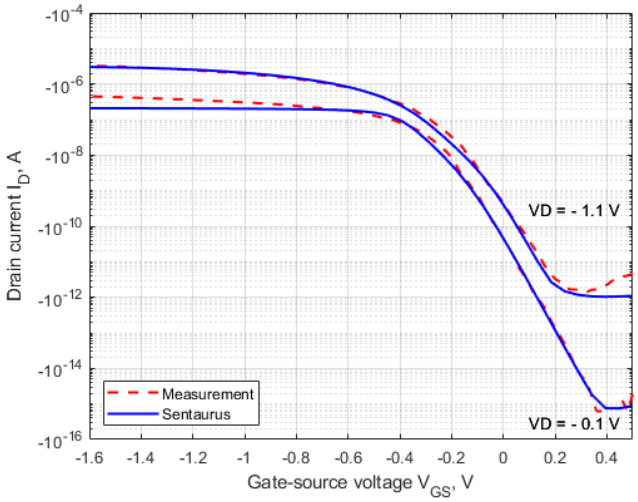
Specifically, in the parameters file one can define the parameters of the Canali formula and then act on the scaling parameters. In the following we will demonstrate the effects of modifying the velocity saturation scaling parameter `kv`. It is a simple approach to account for several non-trivial phenomena that may cause a change of the velocity saturation. An example may be, for instance, the change of velocity saturation due to compressive strain, hence causing compression in the lattice, which can decrease the effective mass of holes, leading to increased mobility. This enhanced mobility can potentially result in higher velocity saturation, as carriers experience fewer scattering events and can achieve higher velocities before saturation occurs.

NW diameter = 20 nm

Doping	Cr workfunction	SS	DIBL	$V_{th,lin}$	$V_{th,sat}$
N_A = 1.2e18 cm^{-3}	4.50 eV	91.325 mV/dec	0.105	-0.235 V	-0.130 V
`kv`	$v_{sat,0}$	$v_{sat,exp}$	v_{sat}		
1.0	8.37e6 cm/s	0.52	8.37e6 cm/s		



(a) Linear plot

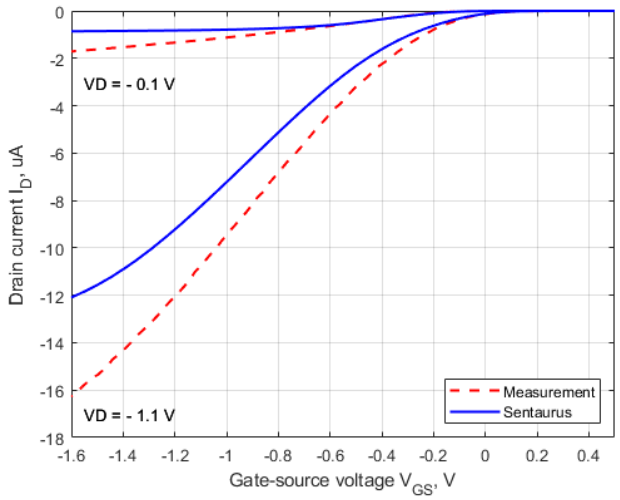


(b) Logarithmic plot

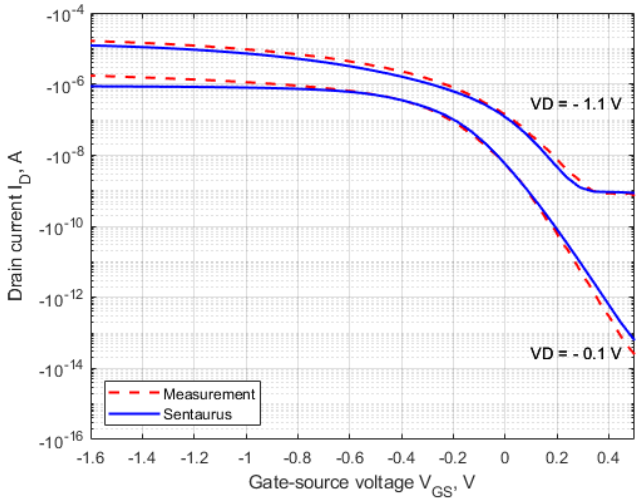
Figure 5.17: Transfer characteristics of 20 nm device with velocity saturation scaling *kv*

NW diameter = 27 nm

Doping	Cr workfunction	SS	DIBL	$V_{th,lin}$	$V_{th,sat}$
N_A = 2.2e18 cm^{-3}	4.25 eV	99.751 mV/dec	0.210	-0.025 V	0.185 V
kv	$v_{sat,0}$	$v_{sat,exp}$	v_{sat}		
2.0	8.37e6 cm/s	0.52	1.67e7 cm/s		



(a) Linear plot

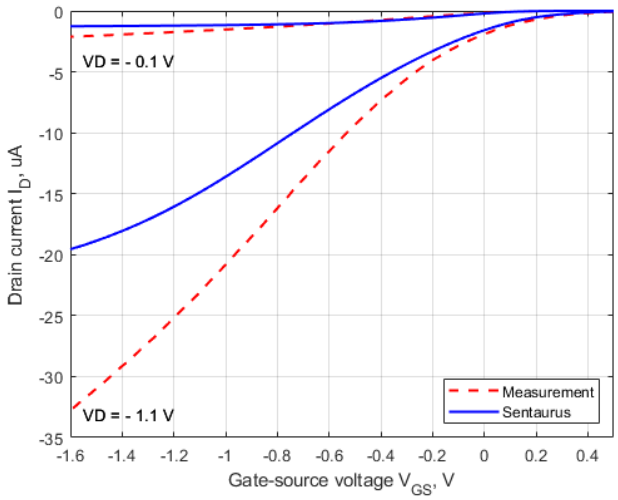


(b) Logarithmic plot

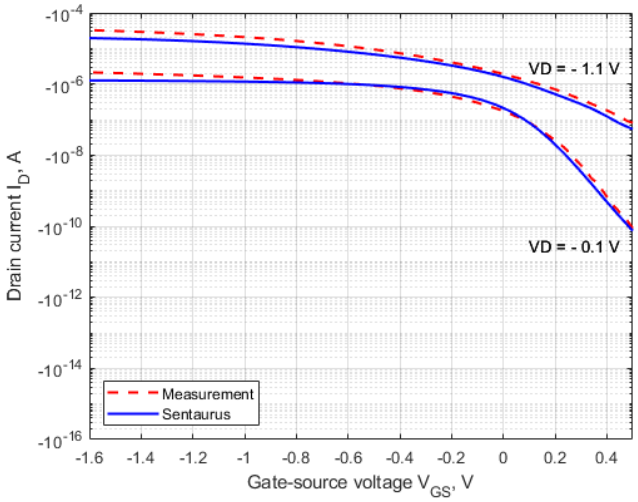
Figure 5.18: Transfer characteristics of 27 nm device with velocity saturation scaling *kv*

NW diameter = 34 nm

Doping	Cr workfunction	SS	DIBL	$V_{th,lin}$	$V_{th,sat}$
$N_A = 2.7e18\ cm^{-3}$	4.25 eV	126.102 mV/dec	0.315	0.133 V	0.448 V
kv	$v_{sat,0}$	$v_{sat,exp}$	v_{sat}		
3.5	8.37e6 cm/s	0.52	2.93e7 cm/s		



(a) Linear plot



(b) Logarithmic plot

Figure 5.19: Transfer characteristics of 34 nm device with velocity saturation scaling *kv*

5.6 Mobility

Carrier mobility is a fundamental property that plays a crucial role in the performance of semiconductor devices. It refers to the ability of charge carriers to move through the semiconductor in response to an electric field. The mobility of carriers in silicon is influenced by various factors, including lattice structure, temperature, impurity scattering, and electric field strength. The dopants affect carrier mobility by introducing impurity scattering, which can hinder carrier movement. The extent of impurity scattering thus depends on the type and concentration of dopants. Also, carrier mobility in silicon decreases with increasing temperature due to lattice vibrations and increased scattering events. The decrease in mobility with temperature is more significant for holes than for electrons [33]. Generally, the mobility of carriers decreases by approximately 2-3 % for every 1°C rise in temperature. Scattering mechanisms can be categorized into several types: phonon scattering, impurity scattering, surface scattering, and carrier-carrier scattering [34]. Each scattering mechanism has a different impact on carrier mobility, with impurity scattering being one of the primary factors affecting mobility in doped silicon. The overall effect on mobility is accounted with the well-known Matthiessen rule, always under the assumption that the scattering mechanisms are independent from each other [35]. Finally, it is fundamental to recall that the mobility of carriers in silicon is not constant but depends on the strength of the applied electric field. This phenomenon is known as field-dependent mobility or velocity saturation. At low electric fields, carrier mobility remains relatively constant (low-field mobility). However, at high electric fields, carrier mobility decreases due to velocity saturation, treated in the previous section, limiting the overall current flow in a semiconductor device.
The work of Gunawan [36] has been very useful to get a better understanding of the carriers mobility in nanowires with similar diameters to the ones studied in this work. Gunawan demonstrates a very interesting dependence of both electron and hole mobility on the diameter of the nanowire, and also discloses a modulation of the conductance by applying stress with a piezoactuator.
In Sentaurus, the mobility models are selected in the *Physics* section:

```
Physics{
Fermi
Mobility( DopingDep HighFieldSaturation
( ParameterSetName = "myHighField" ) Enormal )
EffectiveIntrinsicDensity( Bennett )
Recombination(
SRH( DopingDep ) [...]
```

By default, the **constant mobility model** [37] is enabled. It accounts only for phonon scattering and, therefore, it is dependent only on lattice temperature:

$$\mu_{const} = \mu_L \left(\frac{T}{300\ \mathrm{K}} \right)^{-\zeta} \tag{5.9}$$

where $\mu_L = 470.5\ \mathrm{cm}^2/\mathrm{Vs}$ and $\zeta = 2.2$ for holes.

The `DopingDep` keyword enables the computation of **doping dependent mobility**, which for silicon is accounted by the Masetti model [38]:

$$\mu_{dop} = \mu_{min1} \exp\left(-\frac{P_c}{N_{A,0}+N_{D,0}}\right) + \frac{\mu_{const}-\mu_{min2}}{1+((N_{A,0}+N_{D,0})/C_r)^{\alpha}} - \frac{\mu_1}{1+(C_s/(N_{A,0}+N_{D,0}))^{\beta}}$$

To account the mobility degradation due to interaction of carriers with the **semiconductor-oxide interface**, forced by the transverse electric field, the keyword `Enormal` is used, which enables by default the aforementioned Lombardi model [37] in an enhanced form that calculates a contribution attributed by surface roughness scattering.
Finally, the keyword `HighFieldSaturation`, already presented in the previous section, enables the **velocity saturation** caused by the high electric field.

The TCAD combines all the mentioned mobility models by using the Matthiessen's rule by default, hence:

$$\frac{1}{\mu} = \frac{1}{\mu_{Doping}} + \frac{1}{\mu_{HighField}} + \frac{1}{\mu_{Enormal}} \tag{5.10}$$

With the mobility calculated in such way, we can run a simulation of the device and see how the mobility changes across the nanowire. A Z-cut of the nanowire is shown in Figure 5.20. We notice a degradation of the hole mobility under the gate, probably due to the strong transverse electric field, and near the source and drain contacts, probably due to the high carrier concentration thanks to the presence of the tunneling meshes that inject carriers in the channel.

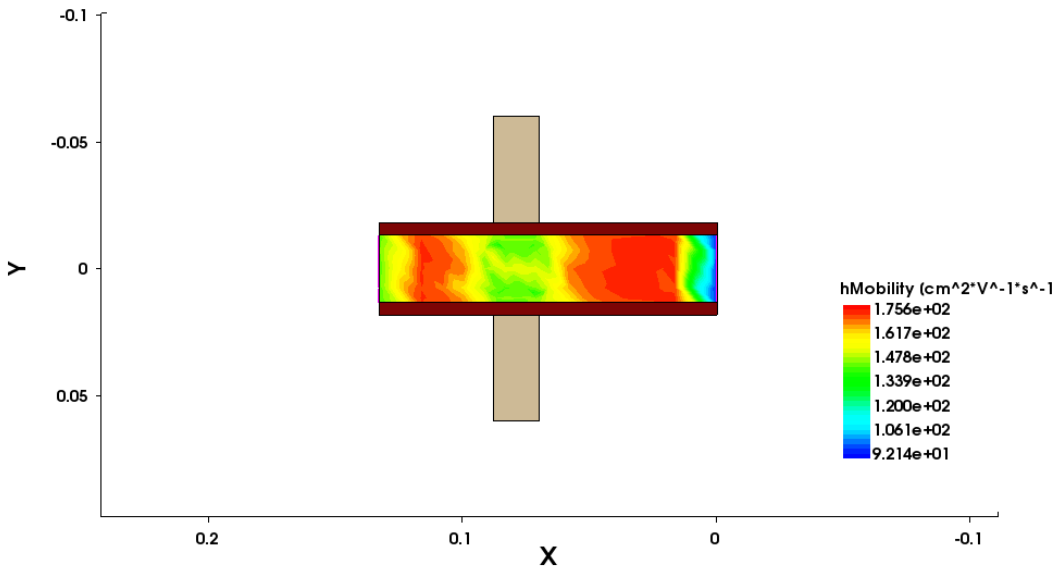
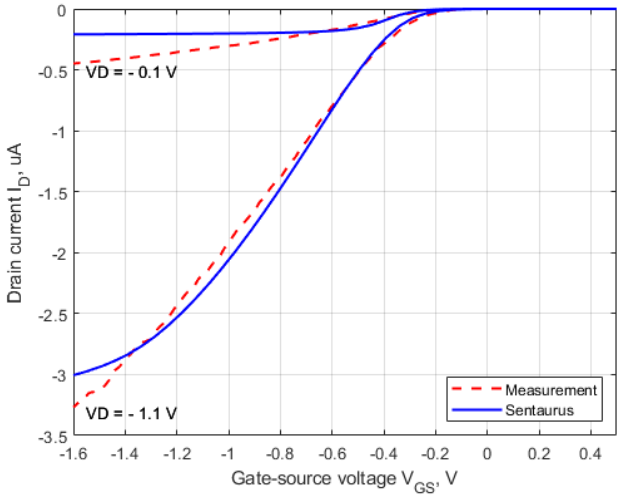


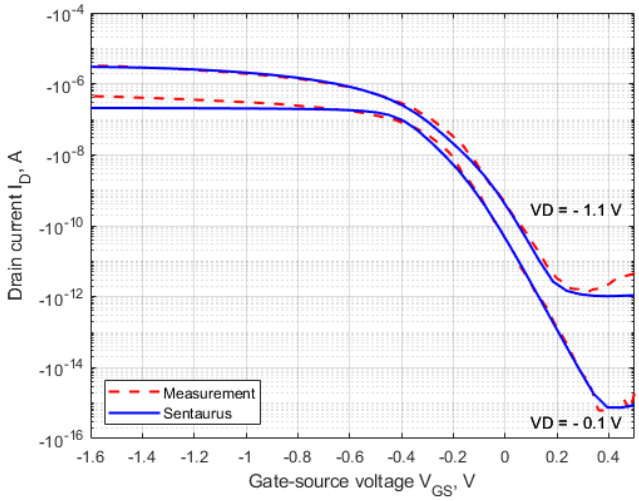
Figure 5.20: Z-cut of the 27 nm device at $V_G = -1.6$ V and $V_D = -0.1$ V showing the hole mobility across the nanowire.

NW diameter = 20 nm

Doping	Φ_M Cr	SS	DIBL	$V_{th,lin}$	$V_{th,sat}$	ku
$N_A = 1.2e18$ cm^{-3}	4.50 eV	91.325 mV/dec	0.105	-0.235 V	-0.130 V	1.0



(a) Linear plot

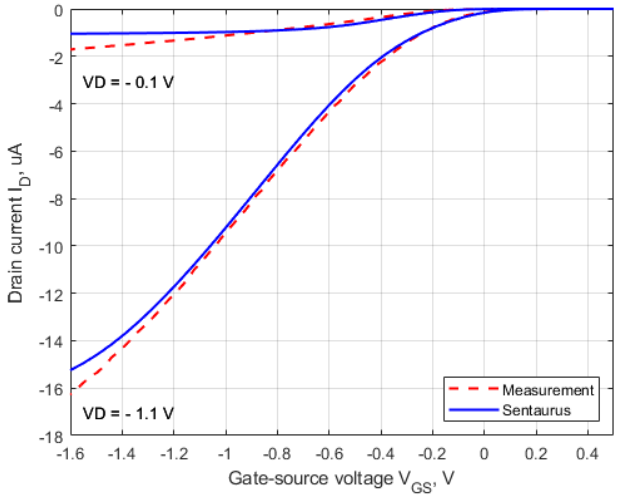


(b) Logarithmic plot

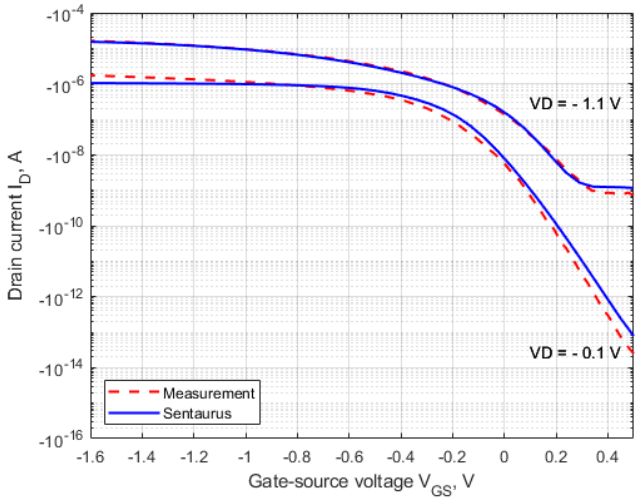
Figure 5.21: Transfer characteristics of 20 nm device with low-field mobility scaling *ku*

NW diameter = 27 nm

Doping	Φ_M Cr	SS	DIBL	$V_{th,lin}$	$V_{th,sat}$	ku
N_A = 2.2e18 cm^{-3}	4.25 eV	97.529 mV/dec	0.158	0.028 V	0.185 V	1.5



(a) Linear plot

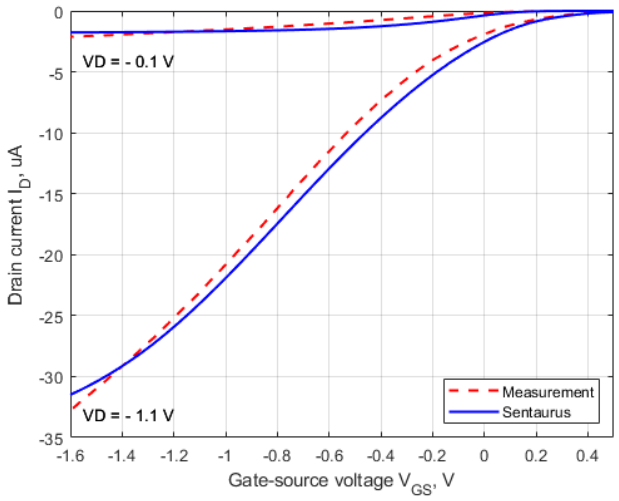


(b) Logarithmic plot

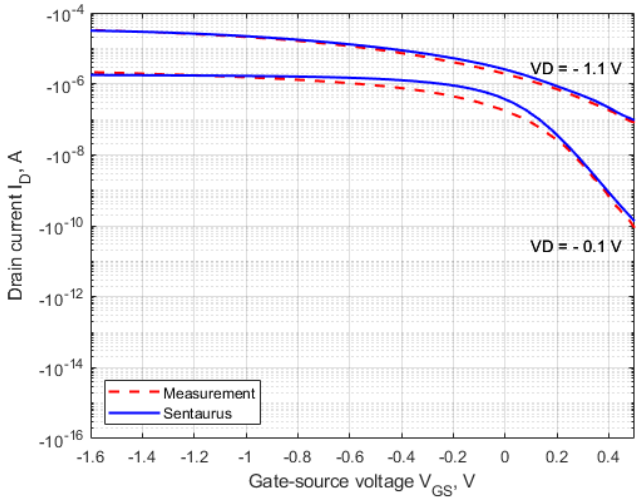
Figure 5.22: Transfer characteristics of 27 nm device with low-field mobility scaling *ku*

NW diameter = 34 nm

Doping	Φ_M Cr	SS	DIBL	$V_{th,lin}$	$V_{th,sat}$	ku
$N_A = 2.7e18\ cm^{-3}$	4.25 eV	125.932 mV/dec	0.367	0.133 V	0.500 V	2.0



(a) Linear plot

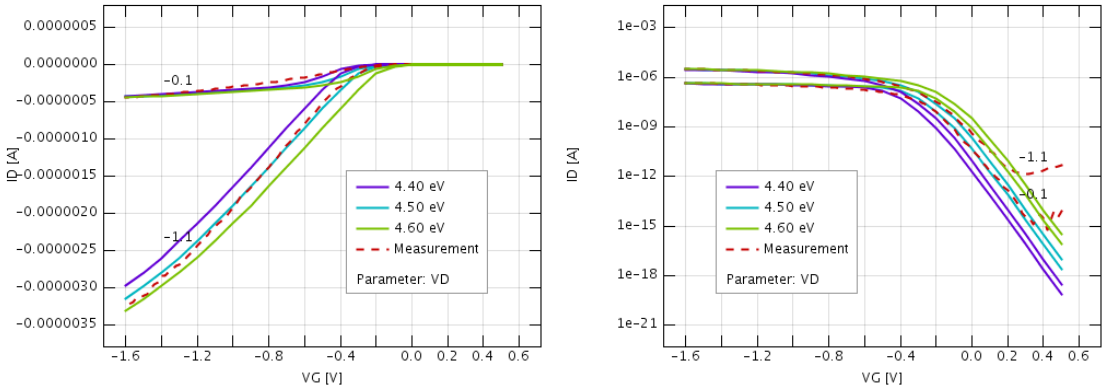


(b) Logarithmic plot

Figure 5.23: Transfer characteristics of 34 nm device with low-field mobility scaling *ku*

5.7 Metallic gate

The work function of the Chromium gate can be adjusted to obtain a good fit on the threshold voltage. Starting from the reference value of 4.5 eV, it is possible to adjust this work function within a relatively large range. The reason is not necessarily related to the work function of the Chromium itself, but it is for taking into account other fabrication defects that may shift the threshold voltage such as, for instance, the presence of fixed charges at the metal-oxide interface and the metal gate granularity (MGG). Tuning the gate work function has an important role in the calibration of the TCAD model: it allows to compensate the variation of threshold voltage if, for instance, there is the necessity to increase the doping level.



(a) Linear plot

(b) Logarithmic plot

Figure 5.24: Transfer characteristics of 20 nm device with different Cr work functions

It must be noted that the gate metal work function will be modified throughout the whole calibration process to adjust the threshold voltage. For this reason, the Cr work-function will be listed in many of the tables of this part. In some cases, the workfunction will vary by some tenths of eV in order to match the desired threshold voltage. This large variation could be considered as a sign of presence of fixed charges under the gate and should not worry an user calibrating a model.

Chapter 6

Simulation of a pMOS inverter

In this chapter, the calibrated TCAD models will be validated through the full TCAD simulation of a pMOS inverter and a 5 stages ring oscillator. As a preliminary study, the transistor parasitic capacitances will be extracted then, successively, the pMOS inverter and the ring oscillator will be demonstrated. Specifically, the static output characteristics of the inverter will be firstly computed, then the transient output characteristics will be extracted. Similarly, for studying the ring oscillator we will extract firstly the static output characteristics to verify that the circuit model is working properly, then we will proceed with the extraction of the transient output characteristics. The specific device model employed in this chapter is the one with a nanowire diameter of 20 nm: this choice has been made due to the fact that the 20 nm device seems to be the one among the others that shows a better subthreshold behavior and it also has a negative threshold voltage, sufficiently far from 0 V, around -0.2 V. The choice of this threshold voltage is convenient because it allows to build logic gates that have consistent logic levels at both input and output, whereas, for instance with the 34 nm device, a positive threshold voltage would complicate the design of a gate using pMOS logic. Always for this reason, the 27 nm device has been discarded due to the fact that its threshold voltage is too much close to 0 V and thus could cause some complications in the realization and simulation of the logic gates. However, by using the 27 nm and 34 nm devices, one could expect better performances due to the higher drive currents.

6.1 pMOS capacitance extraction

The first step for having a rough estimate of the device performance is the extraction of the parasitic capacitances, in fact, by looking at the capacitances extracted from the model one can have an idea of the order of magnitude of the time delay introduced by the device. Extracting the parasitic capacitances requires a different setup of Synopsys Sentaurus, specifically tailored to carry out a small signal AC analysis. To this end, one must define a *system* section in the `jlnt_pmos_des.cmd` file, comprising a voltage source for each terminal of the device i.e., source, gate and drain. Consequently, the *system* section for the studied device would look like:

```
System {
    Vsource_pset vd ( d 0 ) { dc = 0.0 }
    Vsource_pset vg ( g 0 ) { dc = 0.0 }
    Vsource_pset vs ( s 0 ) { dc = 0.0 }
    PMOS pmos ( "Drain" = d "Gate" = g "Source" = s )
}
```

where the `PMOS pmos` is an instance of the TCAD model of the 20 nm device and it is declared in the same `jlnt_pmos_des.cmd` file. Once the system has been defined, it is necessary to define a new simulation setup in the *solve* section:

```
Solve {
    NewCurrentPrefix="init"
    Coupled(Iterations=100){ Poisson }
    Coupled{ Poisson Electron Hole }

    Quasistationary(
    InitialStep=0.05 MaxStep=0.05 MinStep=0.0001
    Goal{ Parameter=vg.dc Voltage= 0.5 }
    ){ Coupled{ Poisson Electron Hole } }

    NewCurrentPrefix=""
    Quasistationary (
    InitialStep=0.01 Increment=1.3 MaxStep=0.05 Minstep=1.e-5
    Goal { Parameter=vg.dc Voltage=-1.6 }
    ){ ACCoupled (
      StartFrequency=1e6 EndFrequency=1e6 NumberOfPoints=1 Decade
      Node(s d g) Exclude(vs vd vg)
      ACCompute (Time = (Range = (0 1)  Intervals = 40))
    ){ Poisson Electron Hole }
    }
}
```

The first `Quasistationary` command biases the device to the desired starting gate voltage for extracting the C-V curves. In our case, the voltage sweeps from 0.5 V to

-1.6 V, as the simulations that have been done in the previous chapters. The following `Quasistationary` command comprises the `ACCoupled` directive, which activates the small signal analysis and has parameters that define the starting and ending frequency and that allow to select the desired terminals to test. The capacitances extraction has been made at a single frequency point, namely 1 MHz. The `Exclude(vs vd vg)` directive is to exclude the voltage sources from the small signal analysis, that would be accounted as short circuits otherwise. The solver computes the Y matrix that maps the current - voltage relation $\delta\mathbf{i} = \mathbf{Y} \cdot \delta\mathbf{v} = (\mathbf{A} + j\omega \cdot \mathbf{C}) \cdot \delta\mathbf{v}$. Thus, once the simulation has run, the results will be stored in the conductance matrix $\mathbf{A}$ and the capacitance matrix $\mathbf{C}$ [21]:

$$\begin{bmatrix} i(s) \\ i(g) \\ i(d) \end{bmatrix} = \left(\begin{bmatrix} a(s,s) & a(s,g) & a(s,d) \\ a(g,s) & a(g,g) & a(g,d) \\ a(d,s) & a(d,g) & a(d,d) \end{bmatrix} + j \cdot \omega \begin{bmatrix} c(s,s) & c(s,g) & c(s,d) \\ c(g,s) & c(g,g) & c(g,d) \\ c(d,s) & c(d,g) & c(d,d) \end{bmatrix} \right) \cdot \begin{bmatrix} v(s) \\ v(g) \\ v(d) \end{bmatrix} \tag{6.1}$$

At this point, one can extract the relevant capacitances, hence gate to drain C_{gd}, gate to source C_{gs} and drain to source C_{ds}. The extracted C-V curve is shown in Figure 6.1.

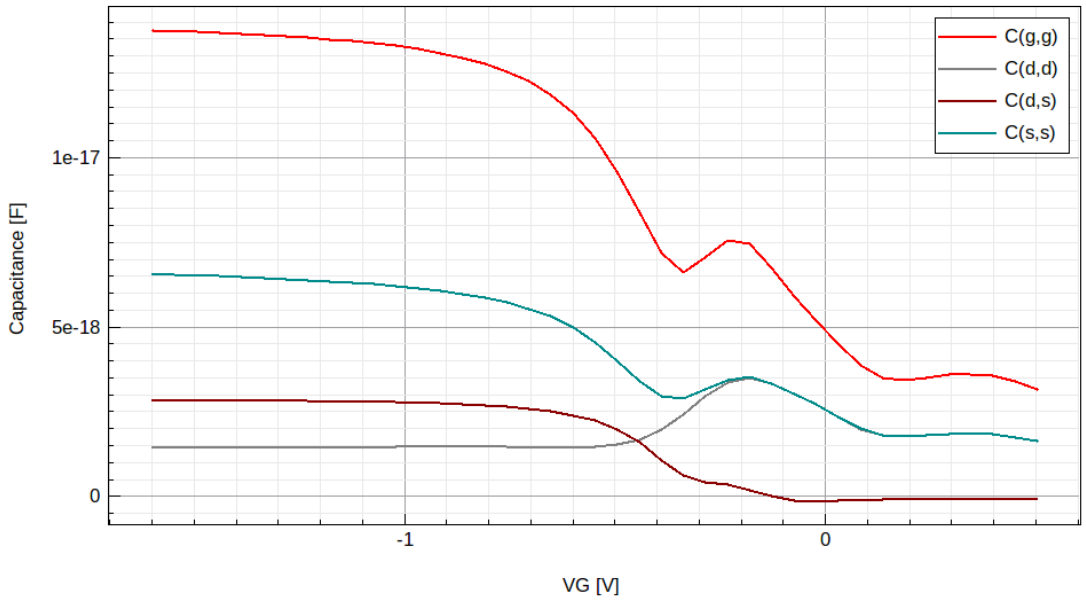


Figure 6.1: Parasitic capacitances extracted at f = 1 MHz

6.2 Inverter static output characteristics

In order to build an inverter using only a pMOS transistor it is necessary to use the device as a pull-up switch. In first place the logic levels are defined: the logic 0 will be set at -1.1 V, whereas the logic 1 will be set at 0 V. The complete circuit, shown in Figure 6.2, includes a load resistor of value 100 MΩ and positioned upstream with respect to the switching transistor. The load resistance is calculated in such a way to obtain a negligible voltage drop on the switch resistance R_S i.e., the drain to source voltage drop of the pMOS, when the transistor is in conduction, and a negligible voltage drop on the load resistance R_L when the transistor is interdicted. Thus, the load resistance R_L will have to be small with respect to the drain to source resistance when the pMOS in OFF, and large with respect to the drain to source resistance when the pMOS is ON. By extracting the resistance values from the measured $I_D - V_G$ curves at $V_D = -1.1$ V of Figure 3.4, it is found that $R_{DS}(V_G = -1.1 \text{ V}) = 506.52$ **kΩ** and $R_{DS}(V_G = 0 \text{ V}) = 2.41$ **GΩ**. Thus a value $R_L = 100$ MΩ would yield satisfactory results. It is interesting to note that an equivalent circuit could be implemented by substituting the load resistance R_L with an active load i.e., a pMOS properly biased in order to obtain the desired drain to source resistance. In this way one can implement the pMOS circuit by using only pMOS devices. Unfortunately, this option has been discarded in favor of the solution with the passive load to decrease the simulation complexity and thus prevent convergence issues of the solver.

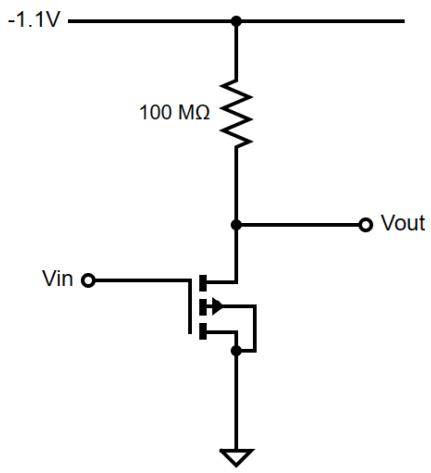


Figure 6.2: Circuit schematics of the pMOS inverter with passive load

By inspecting the circuit shown in Figure 6.2, one can notice that when the pMOS will be ON, hence the gate voltage will be $V_G = -1.1$ V, the output voltage V_{out} should approach 0 V; conversely when the pMOS will be OFF, hence with a gate voltage $V_G =$ V, the output voltage V_{out} should be close to $V_{DD} = -1.1$ V.

Now that the main circuit characteristics have been defined, it is time to configure the simulation in Sentaurus. The *System* section will look as follows:

```
System {
    Vsource_pset vdd (dd 0) { dc = 0.0 }
    Vsource_pset vin (in 0) { dc = 0.0 }
    PMOS pulldown ( "Drain" = out "Gate" = in "Source" = 0 )
    Resistor_pset rload ( dd out ){ resistance = 100e6 }
    Plot "n@node@_sys_des.plt" (time() v(in) v(out) i(pulldown,out) )
}
```

while the *Solve* section will include sweeping the bias voltage up to V_{DD} = - 1.1 V and then sweeping the gate voltage V_{in} from -1.1 to 0.5 V. The corresponding V_{out} will be saved in the output characteristics, shown in Figure 6.3. From the plot below, we can verify that the circuit has an inverting behavior and that the output characteristics is symmetrical with respect to the pMOS threshold voltage V_{th} = - 0.130 V. However, it must be noted that the 0 - 1 transition is not as sharp as the output characteristic of a CMOS inverter, as the one demonstrated in the work of Mukherjee et al. [6], but the absence of a complementary nMOS device model limits us in this perspective. Another major drawback of the chosen circuit topology is that when the transistor is ON, a significant current flows from the supply line V_{DD} to ground, causing a much higher power consumption compared to the CMOS counterpart.

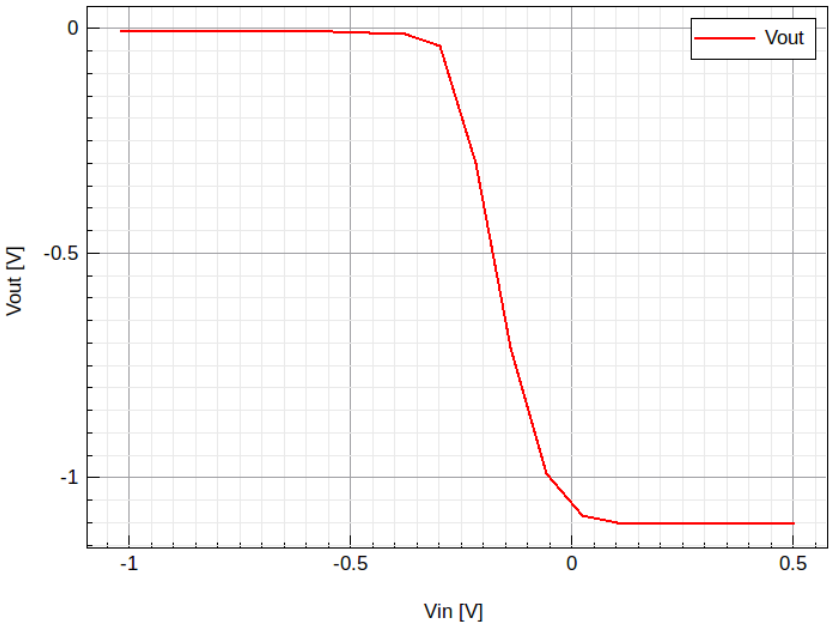


Figure 6.3: Static output characteristics of the pMOS inverter

6.3 Inverter transient output characteristics

After simulating the static output characteristics, it is interesting to investigate the transient behavior of the inverter. To this end, we can keep the simulation setup used for extracting the static output characteristics and change the probing voltage source and the solver directives. To include a square wave with defined amplitude, period, duty cycle, rise time and fall time one must define the source `vin` as follows:

```
System {
Vsource_pset vdd (dd 0) { dc = 0.0 }
Vsource_pset vin (in 0) { pulse = (
0.0       # dc
-1.1      # amplitude
10e-8     # td
10e-8     # tr
10e-8     # tf
40e-8     # ton
100e-8)   # period }
PMOS pulldown ( "Drain" = out "Gate" = in "Source" = 0 )
Resistor_pset rload ( dd out ){ resistance = 100e6 }
Plot "n@node@_sys_des.plt" (time() v(in) v(out) i(pulldown,out) )}
```

In the following, the extracted transient responses of the inverter at frequencies sweeping from 1 MHz to 1 GHz will be shown.

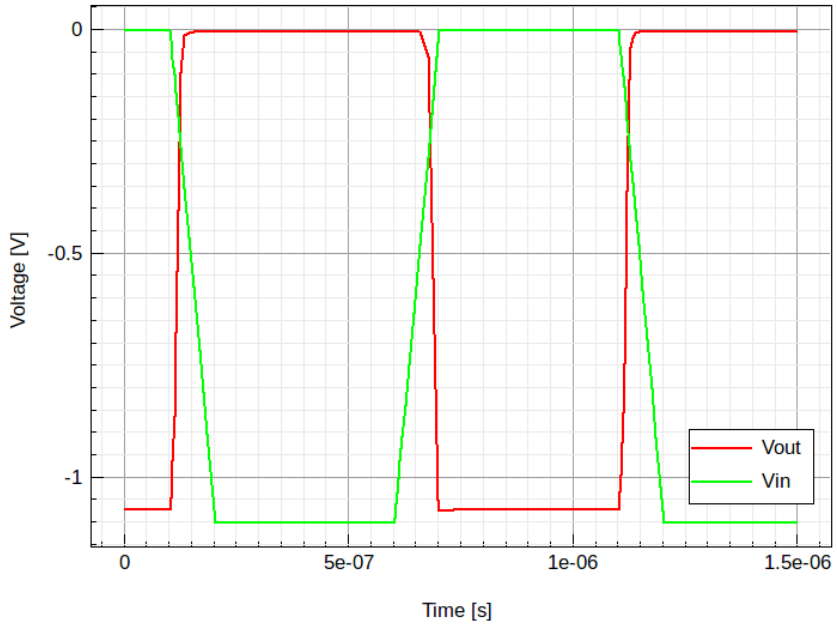


Figure 6.4: Transient response of the pMOS inverter, $f = 1\text{MHz}$

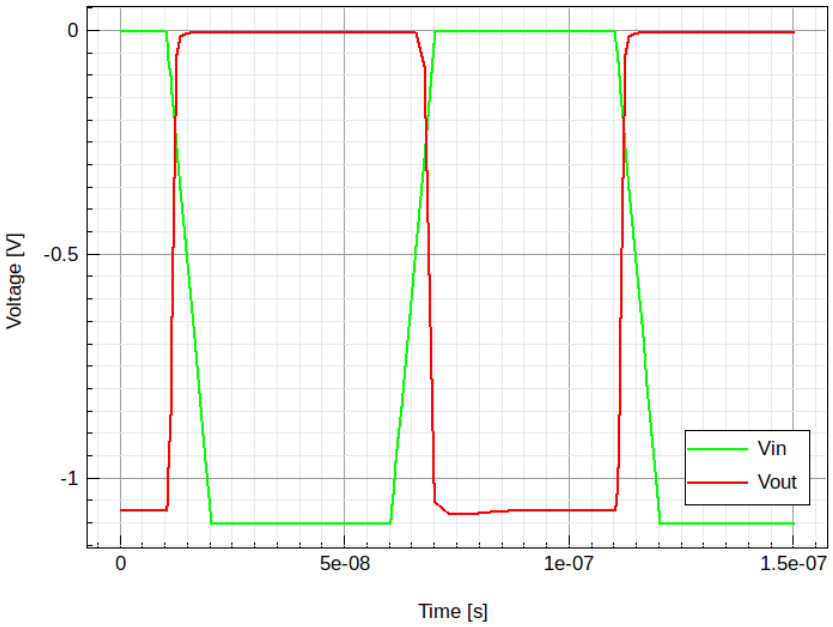


Figure 6.5: Transient response of the pMOS inverter, $f = 10\text{MHz}$

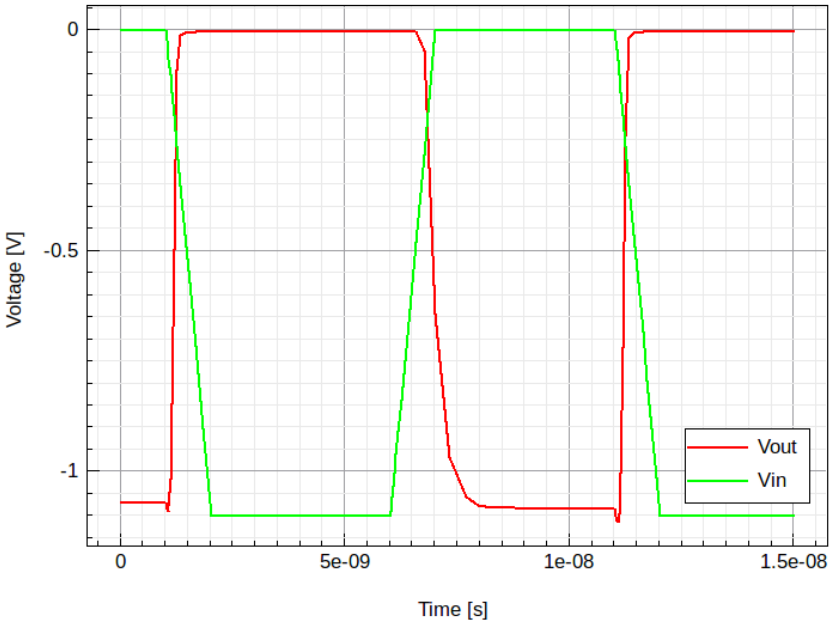


Figure 6.6: Transient response of the pMOS inverter, $f = 100\text{MHz}$

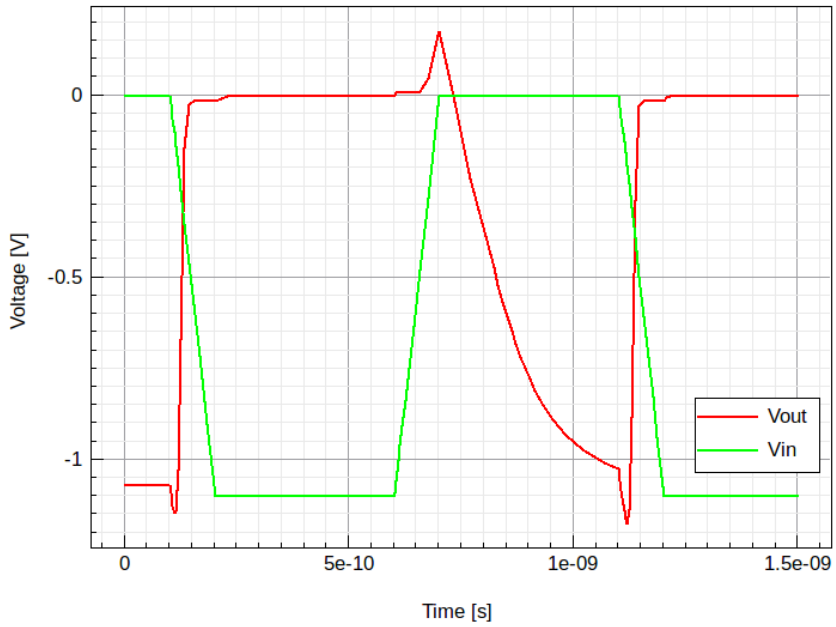


Figure 6.7: Transient response of the pMOS inverter, $f = 1\text{GHz}$

In order to extract the transient characteristics, the *Solve* section must include the `Transient` directive, which enables the solver to simulate the time evolution of the circuit:

```
Transient (
InitialTime=0 FinalTime=150e-8
InitialStep=1e-8 Increment=1.3
MaxStep=5e-7 MinStep=1e-11
TurningPoints (
( Condition ( Time ( 1e-7 ; 6e-7 ; 1.1e-6 ) ) Value= 1.0e-9 )
( Condition ( Time ( Range= ( 1e-7 1.5e-7 ) ; Range= ( 8e-7 9e-7 ) ;
Range= ( 1.1e-6 1.2e-6 ) ) ) Value= 1.0e-8 ) )
){ Coupled{ pulldown.poisson pulldown.electron
pulldown.hole pulldown.contact circuit } }
```

As it can be seen, the transient response of the inverter at 1 MHz of Figure 6.4 shows almost no distortion, whereas at 100 MHz - Figure 6.6 - the distortion of the output signal becomes quite noticeable. At 1 GHz frequency, Figure 6.7, the output characteristics shows severe degradation with the presence of noticeable voltage spikes.

6.4 Ring oscillator

As a final validation test for the device, a 5 stages ring oscillator, shown in Figure 6.8 is simulated. A ring oscillator is a device composed of an odd number of inverters arranged in a ring. The output of the last inverter is fed back into the first, causing the output to oscillate between two voltage levels, representing 0 (-1.1 V) and 1 (0 V) logic levels. Because a single inverter computes the logical NOT of its input, it can be shown that the last output of a chain of an odd number of inverters is the logical NOT of the first input. The final output is asserted a finite amount of time after the first input is asserted and the feedback of the last output to the input causes oscillation. The frequency of a ring oscillator is determined by the number of stages in the ring and the delay of each stage. The oscillator period is equal to twice the sum of the individual delays of all stages. To increase the frequency of oscillation, one could make the ring from a smaller number of inverters results in a higher frequency of oscillation or, alternatively, may increase the supply voltage, thus reducing the propagation delay through the chain of stages, increasing both the frequency of the oscillation and the current consumed. The delay time for each gate in a ring oscillator can be calculated by measuring the ring oscillator's fundamental frequency. The total delay time is equal to the product of single gate-delay time multiplied by the number of stages. Thus, by measuring a ring oscillator's fundamental frequency, we can calculate the average gate-delay time $t = \frac{1}{2fn}$ [39].

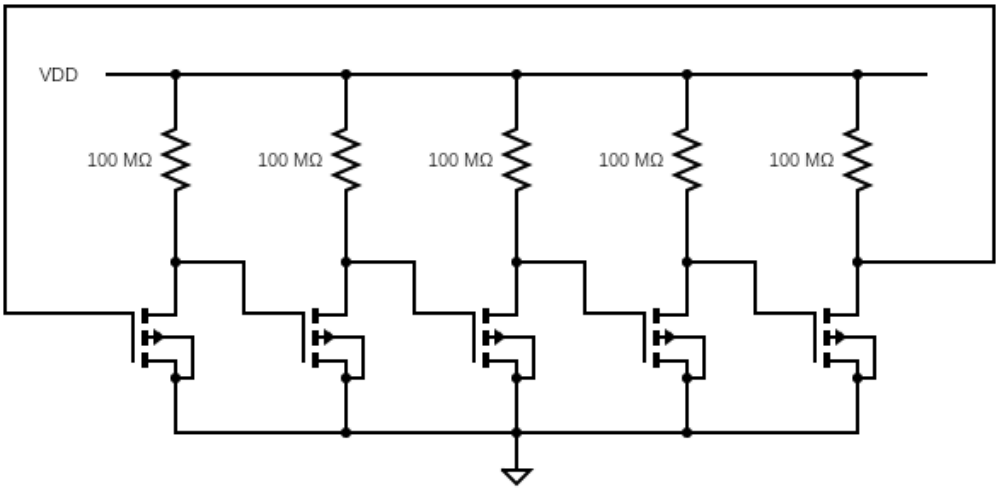


Figure 6.8: Circuit schematics of a 5 stages ring oscillator

For simulating the ring oscillator we have chosen a full TCAD approach: this way assures very accurate results but it has a noticeable shortcoming in terms of versatility, in fact the simulation of such circuit happened to be significantly time consuming. Employing a SPICE model for simulating the ring oscillator would sensibly reduce the simulation time and add improved flexibility to the design flow.

6.4.1 Static characteristics

As done previously with the inverter, we firstly extract the static characteristics of the ring oscillator. The static characteristics can be extracted by opening the feedback loop, as shown in the circuit of Figure 6.9, and connect a voltage source to the input of the first inverter, then measuring the voltage at the output of each inverter. The resulting characteristics is shown in Figure 6.10. It is interesting to notice how the 0 - 1 transitions become sharper as the output is taken from a further stage.

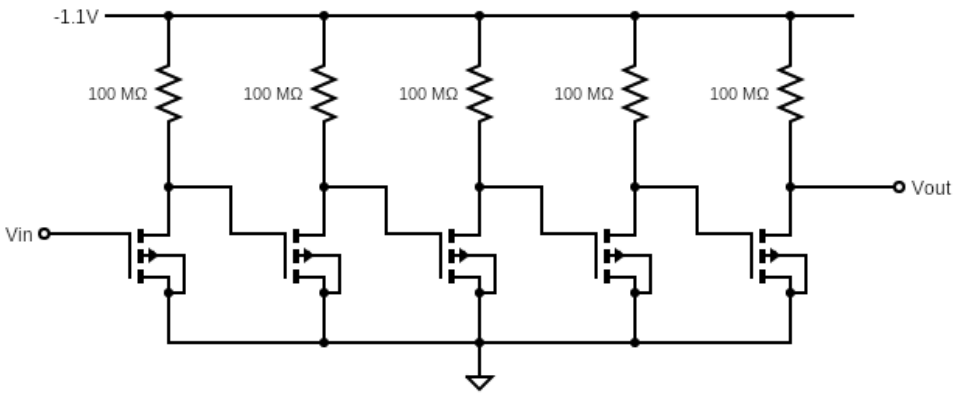


Figure 6.9: Circuit schematics of the ring oscillator with open loop

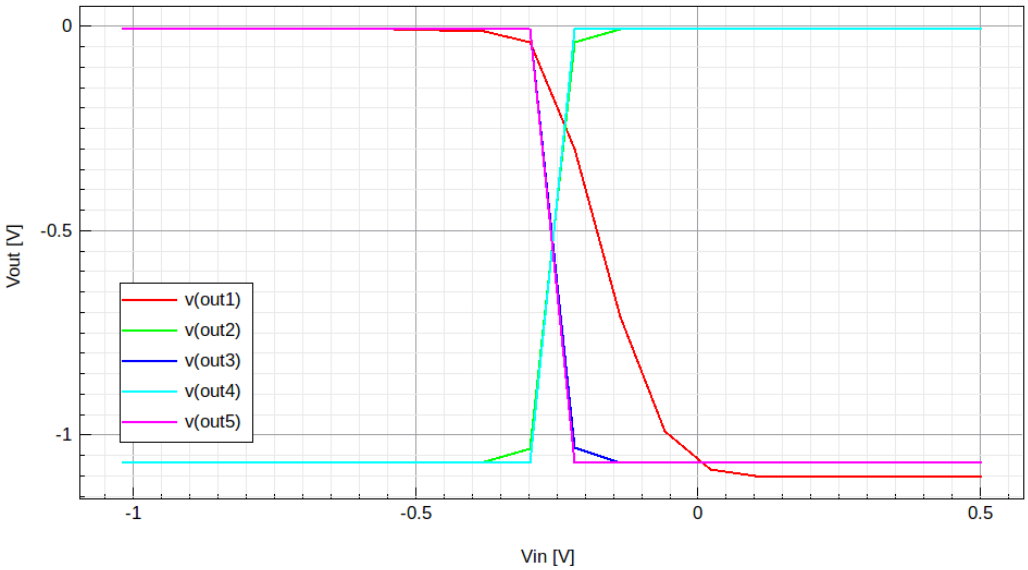


Figure 6.10: Static output characteristics of the 5 stages ring oscillator.

6.4.2 Transient behavior

Finally, the transient behavior of the ring oscillator is simulated. In order to let the circuit oscillate quickly, the output nodes of each stage are initialized with a voltage $V_0 = -0.25$ V. The circuit is simulated at different values of V_{DD}, -0.6 V, -1.1 V and -1.6 V, for a sufficient amount of time that allows to save few oscillation periods. As seen in the work of Marin and Marin [39], the oscillation period will depend on the parasitic capacitance of the single transistor, which has been extracted in the previous section. After measuring the frequency of oscillation of each transient characteristics, we can derive the time delay of a single inverter according to the equation:

$$t = \frac{1}{2fn} \tag{6.2}$$

where f is the oscillation frequency and n is the number of stages, hence 5 in our case. Firstly, we let a simulation run at V_{DD} = -1.1 V and we observe the voltage at each output node. The output curve of Figure 6.11 shows that the signals taken at each output node are shifted by a time difference, which should correspond to the time delay of the single device. After that, a single signal taken at the first stage is plotted at said values of V_{DD} and is shown in Figures 6.12, 6.13 and 6.14. The period T has been calculated by the plots in order to derive the frequency.

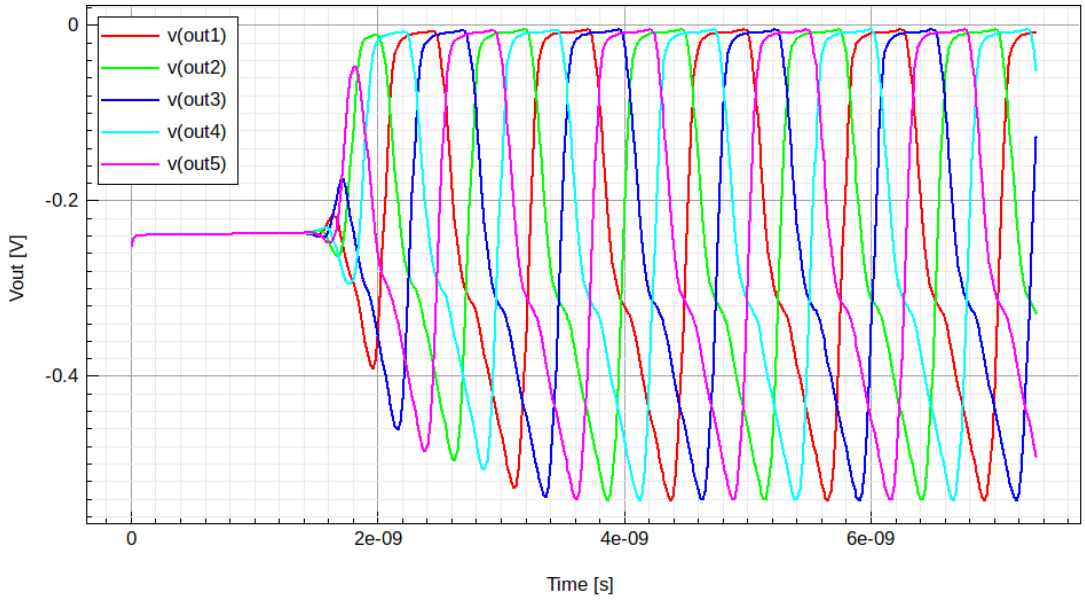


Figure 6.11: Time evolution of the 5 stages ring oscillator, the output of each stage is shown.

V_{DD}	f	T	A	t
-0.6 V	461.0 MHz	2.169 ns	0.407 V	0.217 ns
-1.1 V	783.1 MHz	1.277 ns	0.536 V	0.128 ns
-1.6 V	1.091 GHz	0.917 ns	0.601 V	0.092 ns

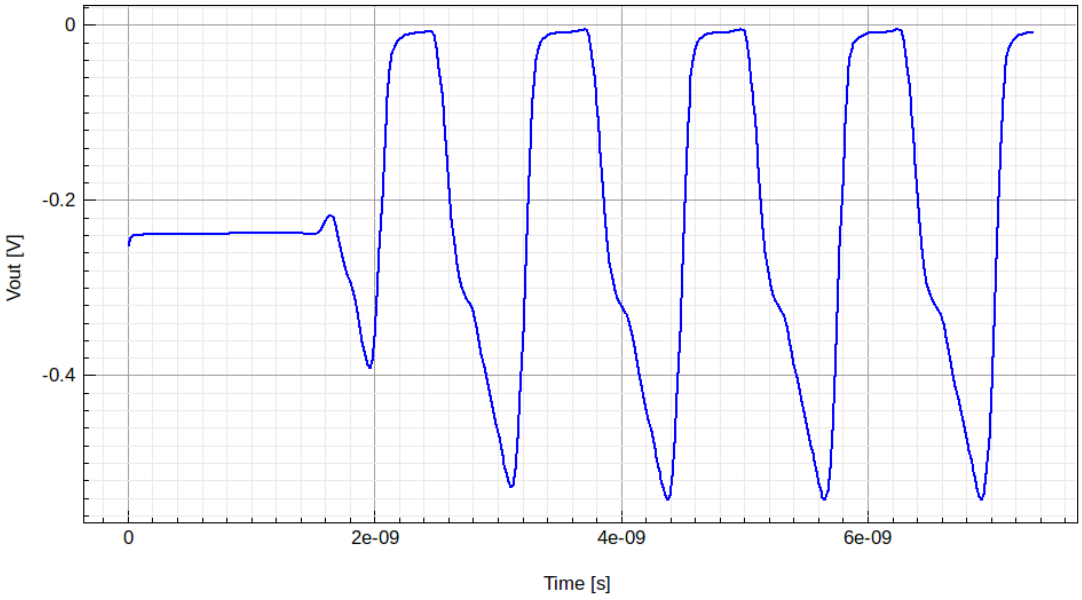


Figure 6.12: Time evolution of the 5 stages ring oscillator, VDD = -1.1 V.

The results show that the oscillation frequency, as well as the amplitude, depends on the applied voltage thus, as expected, the ring oscillator can be considered as a voltage controlled oscillator. Even though it is not clearly visible, the oscillation frequency can vary from period to period and this variation, commonly referred at as *jitter* can be measured. An interesting application of this is that this frequency fluctuations are intrinsically random, and thus the ring oscillator can be employed for building a true random number generator (TRNG) [40].
Curiously, the obtained oscillation frequency is quite small, if compared with other ring oscillators demonstrated. For instance, the work of Deepak et al. [41] discloses a 5 stages ring oscillator built on CMOS inverters realized with 32 nm FinFET devices and is able to oscillate at a frequency of 39 GHz, while retaining a satisfactory power efficiency. Probably, the oscillation frequency of the circuit presented here is smaller due to the fact that we are employing pMOS inverters with passive load, instead of CMOS inverters.

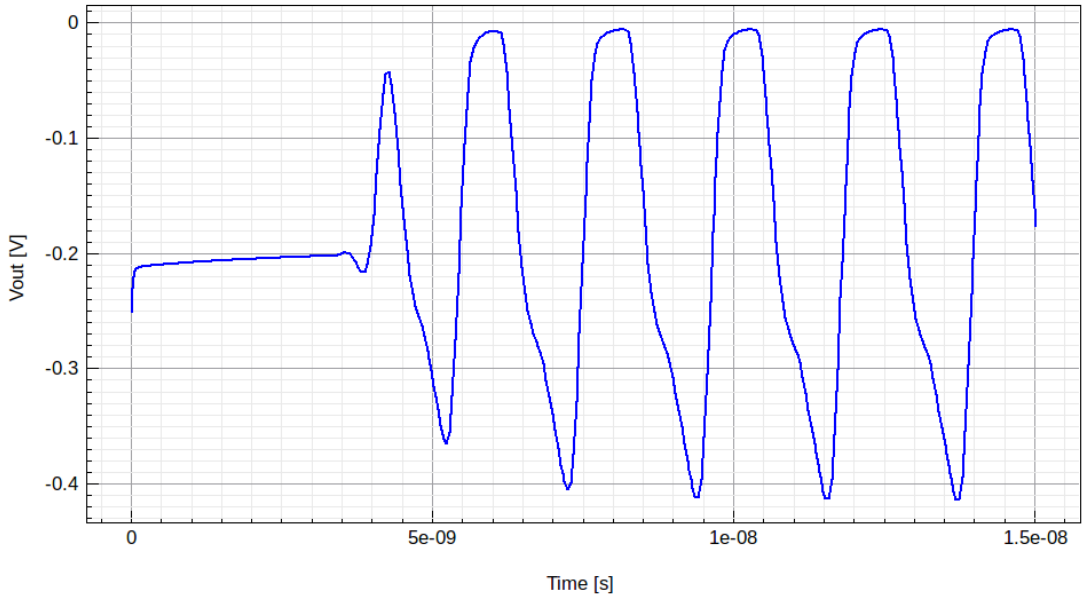


Figure 6.13: Time evolution of the 5 stages ring oscillator, VDD = -0.6 V.

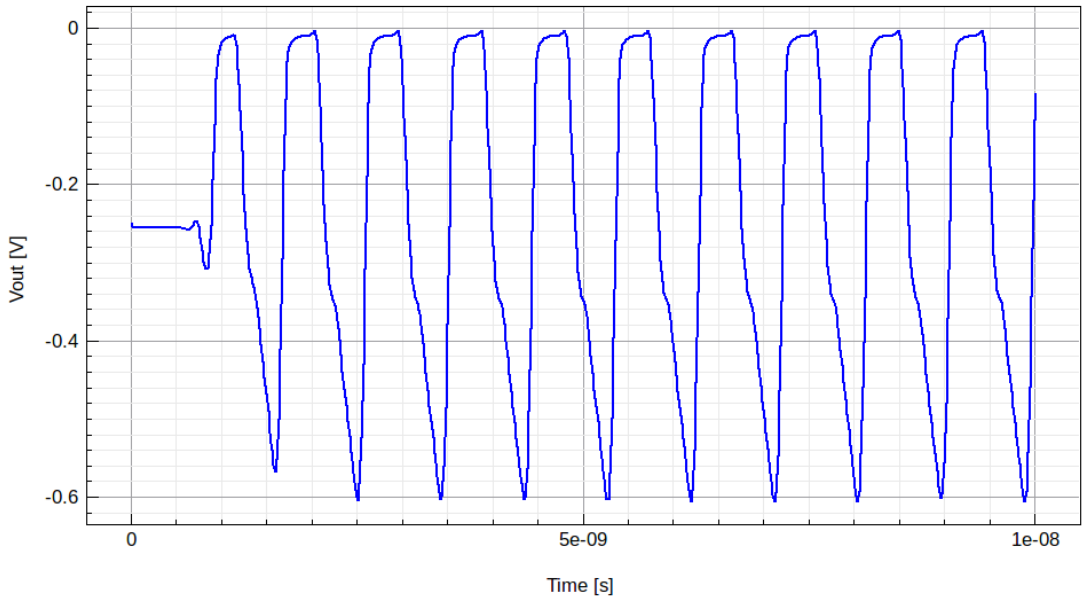


Figure 6.14: Time evolution of the 5 stages ring oscillator, VDD = -1.6 V.

Chapter 7

Conclusion

In this work the Vertical Junctionless Nanotransistor has been studied in depth, yielding three separate TCAD models corresponding to three different device geometries. I consider building and calibrating a TCAD model a very constructive experience, that allowed me to review and put in practice much of the theory on semiconductor devices that I have learned in these years at PoliTo, and, of course, to discover and understand more complicated, yet fascinating, topics concerning this specific technology. Along with all the technical aspects, there also has been a methodology improvement, in fact I had to organize the work, keep track of all the changes that I was applying in the middle of the calibration process, make choices based on the outcomes, schedule the more time consuming simulations and, sometimes repeat some parts of the work trying to avoid mistakes. On top of that, working at LAAS, along with the Global TCAD Solutions engineers and within the framework of the FVLLMONTI project has been an exceptional opportunity for me to get in contact with a real device development workflow, and work and be mentored by many experts. However, there is still some work that could be done, so, in conclusion, the following section will contain some interesting follow-ups.

7.1 Future works

This work allowed me to get a quite comprehensive understanding of the operation of the vertical junctionless nanotransistor. However, there is still plenty of work that could be done to ameliorate the TCAD model, that could lead to a deeper understanding of the device and the underlying physical phenomena. Some interesting studies that could be addressed are:

- The near-threshold and subthreshold behavior could be refined by simulating and inspecting **trap assisted band to band tunneling**, a phenomenon which TCAD environments usually treat in a separate manner with respect to band to band tunneling. Inspecting the trap assisted band to band tunneling could lead to a more accurate fitting of the model, as it is a mechanism that becomes noticeable at lower electric fields than band to band tunneling [42].

- The high values of mobility needed to fit the ON current, considered along with the fabrication process, strongly suggest the presence of **compressive strain** in the silicon nanowire. A further study then could be related to the implementation of specific models relating the stress to the carrier mobility in order to quantify the compressive strain occurring in the device [43] [44].

- At LAAS-CNRS there is also the complementary nMOS device under development. Using this work as a reference could enable a fast and effective TCAD modeling and calibration of such device, even though some electrical characteristics will be different. Having a working nMOS model could allow to simulate CMOS gates and subsequently more complex logic designs or logic cells.

- The studied device has been chosen for being fitted with a ferroelectric material, such as HfO_2 as insulator. Simulating a device with a ferroelectric gate material would certainly give some useful insights on the behavior of the device and would enhance the development of device aimed at logic-in-memory applications.

- Once we have a calibrated model, it of great importance to extract a compact model that has to be used in the subsequent workflows for developing more complex logic gates and subsequently integrated circuits. A suitable and interesting follow-up would be then the extraction of a BSIM-CMG [45] SPICE model. On this regard, a preliminary work has been done starting from the compact model calibrated on a similar, non-junctionless GAAFET of Gucciardo [46], but, apparently, there is still work to be done to adapt such model to the junctionless technology.

Bibliography

[1] X.-L. Han, G. Larrieu, and E. Dubois, "Realization of vertical silicon nanowire networks with an ultra high density using a top–down approach," *Journal of Nanoscience and Nanotechnology*, vol. 10, 11 2010.

[2] Y. Guerfi and G. Larrieu, "Vertical silicon nanowire field effect transistors with nanoscale gate-all-around," *Nanoscale Res. Lett.*, vol. 11, p. 210, Dec. 2016.

[3] C. A. Mack, "Fifty years of moore's law," *IEEE Transactions on Semiconductor Manufacturing*, vol. 24, no. 2, pp. 202–207, 2011.

[4] R. Dennard, F. Gaensslen, H.-N. Yu, V. Rideout, E. Bassous, and A. LeBlanc, "Design of ion-implanted mosfet's with very small physical dimensions," *IEEE Journal of Solid-State Circuits*, vol. 9, no. 5, pp. 256–268, 1974.

[5] IEEE, 2021. The International Roadmap for Devices and Systems.

[6] C. Mukherjee, A. Poittevin, I. O'Connor, G. Larrieu, and C. Maneux, "Compact modeling of 3d vertical junctionless gate-all-around silicon nanowire transistors towards 3d logic design," *Solid-State Electronics*, vol. 183, p. 108125, 2021.

[7] J. Colinge, A. Kranti, R. Yan, C. Lee, I. Ferain, R. Yu, N. Dehdashti Akhavan, and P. Razavi, "Junctionless nanowire transistor (jnt): Properties and design guidelines," *Solid-State Electronics*, vol. 65-66, pp. 33–37, 2011. Selected Papers from the ESSDERC 2010 Conference.

[8] J. P. Colinge. United States Patent No. US 8,178,862 B2, May 15, 2012.

[9] http://fvllmonti.eu/. FVLLMONTI Consortium homepage.

[10] L. Yuan, "What is dtco?: An introduction to design-technology co-optimization." https://www.tsmc.com/english/news-events/blog-article-20220615, June 2022.

[11] V. Moroz, X.-W. Lin, P. Asenov, D. Sherlekar, M. Choi, L. Sponton, L. S. Melvin, J. Lee, B. Cheng, A. Nannipieri, J. Huang, and S. Jones, "Dtco launches moore's law over the feature scaling wall," in *2020 IEEE International Electron Devices Meeting (IEDM)*, pp. 41.1.1–41.1.4, 2020.

[12] M. Karner, G. Rzepa, O. Baumgartner, G. Strof, F. Schanovsky, F. Mitterbauer, C. Kernstock, H. Karner, and Z. Stanojevic, "Variability-aware dtco flow: Projections to n3 finfet and nanosheet 6t sram," in *2021 International Conference on Simulation of Semiconductor Processes and Devices (SISPAD)*, pp. 15–18, 2021.

[13] https://www.globaltcad.com/. Global TCAD Solutions GmbH.

[14] T. Simlinger, H. Kosina, M. Rottinger, and S. Selberherr, "Minimos-nt: A generic simulator for complex semiconductor devices," in *ESSDERC '95: Proceedings of the*

25th European Solid State Device Research Conference, pp. 83–86, 1995.
[15] S. Selberherr, A. Schutz, and H. W. Potzl, "Minimos—a two-dimensional mos transistor analyzer," *IEEE Transactions on Electron Devices*, vol. 27, pp. 1540–1550, 1980.
[16] G. Larrieu and X.-L. Han, "Vertical nanowire array-based field effect transistors for ultimate scaling," *Nanoscale*, vol. 5, pp. 2437–2441, 2013.
[17] Y. Guerfi, F. Carcenac, and G. Larrieu, "High resolution hsq nanopillar arrays with low energy electron beam lithography," *Microelectronic Engineering*, vol. 110, pp. 173–176, 2013.
[18] S. Trellenkamp, J. Moers, A. van der Hart, P. Kordoš, and H. Lüth, "Patterning of 25-nm-wide silicon webs with an aspect ratio of 13," *Microelectronic Engineering*, vol. 67-68, pp. 376–380, 2003. Proceedings of the 28th International Conference on Micro- and Nano-Engineering.
[19] Y. Guerfi, J. Doucet, and G. Larrieu, "Thin-dielectric-layer engineering for 3d nanostructure integration using an innovative planarization approach," *Nanotechnology*, vol. 26, p. 425302, 10 2015.
[20] K. Ng and R. Liu, "On the calculation of specific contact resistivity on ⟨100⟩ si," *IEEE Transactions on Electron Devices*, vol. 37, no. 6, pp. 1535–1537, 1990.
[21] Synopsys, Inc., June 2021. Sentaurus Device User Guide.
[22] K. Varahramyan and E. Verret, "A model for specific contact resistance applicable for titanium silicide-silicon contacts," *Solid-State Electronics*, vol. 39, no. 11, pp. 1601–1607, 1996.
[23] Global TCAD Solutions GmbH and Institute for Microelectronics TU Vienna, 2022. Minimos-NT – User Manual.
[24] T. J. Drummond, "Work functions of the transition metals and metal silicides," *Journal of Applied Physics*, 2 1999.
[25] R. Tsu and L. Esaki, "Tunneling in a finite superlattice," *Appl. Phys. Lett.*, vol. 22, pp. 562–564, June 1973.
[26] C. Duke, *Tunneling in Solids*. Solid State physics, Academic Press, 1969.
[27] A. M. Cowley and S. M. Sze, "Surface States and Barrier Height of Metal-Semiconductor Systems," *Journal of Applied Physics*, vol. 36, pp. 3212–3220, Oct. 1965.
[28] T. Ayalew, "Sic semiconductor devices technology, modeling, and simulation," 2004.
[29] H. S. Bennett and C. L. Wilson, "Statistical comparisons of data on band-gap narrowing in heavily doped silicon: Electrical and optical measurements," *Journal of Applied Physics*, vol. 55, pp. 3582–3587, 1984.
[30] G. Hurkx, D. Klaassen, and M. Knuvers, "A new recombination model for device simulation including tunneling," *IEEE Transactions on Electron Devices*, vol. 39, no. 2, pp. 331–338, 1992.
[31] C. Canali, G. Majni, R. Minder, and G. Ottaviani, "Electron and hole drift velocity measurements in silicon and their empirical relation to electric field and temperature," *IEEE Transactions on Electron Devices*, vol. 22, no. 11, pp. 1045–1047, 1975.
[32] D. Caughey and R. Thomas, "Carrier mobilities in silicon empirically related to doping and field," *Proceedings of the IEEE*, vol. 55, no. 12, pp. 2192–2193, 1967.
[33] J. Van Zeghbroeck, *Principles of Semiconductor Devices*. Bart Van Zeghbroeck, 2011.

[34] M. G. Francesco Bertazzi, *Monte Carlo modeling of carrier transport in crystalline materials*. Politecnico di Torino, 2021.

[35] Y. Takeda and T. P. Pearsall, "Failure of matthiessen's rule in the calculation of carrier mobility and alloy scattering effects in ga0.47in0.53as," *Electronics Letters*, vol. 17, pp. 573–574, 1981.

[36] O. Gunawan, L. Sekaric, A. Majumdar, M. Rooks, J. Appenzeller, J. W. Sleight, S. Guha, and W. Haensch, "Measurement of silicon nanowires carrier mobility and its size dependence," in *2008 Device Research Conference*, pp. 189–190, 2008.

[37] C. Lombardi, S. Manzini, A. Saporito, and M. Vanzi, "A physically based mobility model for numerical simulation of nonplanar devices," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 7, no. 11, pp. 1164–1171, 1988.

[38] G. Masetti, M. Severi, and S. Solmi, "Modeling of carrier mobility against carrier concentration in arsenic-, phosphorus-, and boron-doped silicon," *IEEE Transactions on Electron Devices*, vol. 30, no. 7, pp. 764–769, 1983.

[39] R. C. Marin and M. E. Marin, "Simulation of a ring oscillator with cmos inverters," *University Politehnica Bucharest*, 2011.

[40] M.-J. O. Saarinen, "On entropy and bit patterns of ring oscillator jitter," in *2021 Asian Hardware Oriented Security and Trust Symposium (AsianHOST)*, pp. 1–6, 2021.

[41] A. Lourts Deepak, L. Dhulipalla, S. Chaitra, and C. Basha Shaik, "Designing of finfet based 5-stage and 3-stage ring oscillator high frequency generation in 32nm," in *IEEE-International Conference On Advances In Engineering, Science And Management (ICAESM -2012)*, pp. 222–227, 2012.

[42] R. N. Sajjad, W. Chern, J. L. Hoyt, and D. A. Antoniadis, "Trap assisted tunneling and its effect on subthreshold swing of tunnel fets," *IEEE Transactions on Electron Devices*, vol. 63, no. 11, pp. 4380–4387, 2016.

[43] H. Minari, T. Kitayama, M. Yamamoto, and N. Mori, "Strain effects on hole current in silicon nanowire fets," in *2010 International Conference on Simulation of Semiconductor Processes and Devices*, pp. 207–210, 2010.

[44] J. Qin, J. Zhang, G. Du, X. Zhang, and X. Liu, "Strain impacts on electron mobility in silicon nanowires," in *2012 IEEE 11th International Conference on Solid-State and Integrated Circuit Technology*, pp. 1–3, 2012.

[45] J. P. Duarte, S. Khandelwal, A. Medury, C. Hu, P. Kushwaha, H. Agarwal, A. Dasgupta, and Y. S. Chauhan, "Bsim-cmg: Standard finfet compact model for advanced circuit design," in *ESSCIRC Conference 2015 - 41st European Solid-State Circuits Conference (ESSCIRC)*, pp. 196–201, 2015.

[46] S. Gucciardo, "Gate-all-around fet: analytical compact modeling and tcad validation for system performance evaluation," *Politecnico di Torino*, 2022.