TCAD study of interface traps-related variability in ultra-scaled MOSFETs

Vikas Velayudhan

Universitat Autònoma de Barcelona
Escola d'Enginyeria
Electronic Engineering Department

TCAD study of interface traps-related variability in ultra-scaled MOSFETs

A dissertation submitted by
Vikas Velayudhan
in fulfillment of the requirements for the degree of
Doctor of Philosophy in Electrical Engineering

Supervised by Dra. Montserrat Nafría Maqueda and Dra. Rosana Rodríguez Martínez

Bellaterra, September 2016

Universitat Autònoma de Barcelona
Escola d'Enginyeria
Electronic Engineering Department

TCAD study of interface traps-related variability in ultra-scaled MOSFETs

A dissertation submitted by
Vikas Velayudhan
in fulfillment of the requirements for the degree of
Doctor of Philosophy in Electrical Engineering

Supervised by Dra. Montserrat Nafría Maqueda and Dra. Rosana Rodríguez Martínez

Vikas Velayudhan

Dra. Montserrat Nafría Maqueda

Dra. Rosana Rodríguez Martínez

Bellaterra, September 2016

Universitat Autònoma de Barcelona
Escola d'Enginyeria
Electronic Engineering Department

Dra. Montserrat Nafría, professor and Dra. Rosana Rodriguez, associate professor of the Electronic Engineering department of the Universidad Autònoma de Barcelona,

Certify

That the work "**TCAD study of the interface traps – related variability in ultra-scaled MOSFETs**" submitted by Vikas Velayudhan to the School of Engineering in fulfillment of the requirements of the degree of Doctor of Philosophy in Electronic Engineering has been performed under their supervision.

Dra. Montserrat Nafría

Bellaterra, September of 2016

Dra. Rosana Rodríguez

Bellaterra, September of 2016

Acknowledgments

First and foremost, I would like to express my sincere thanks to my mentors Prof. Montserrat Nafría and Prof. Rosana Rodríguez for their constant support throughout my PhD. I'm grateful for their insights, ideas and the constructive criticism towards the improvement and completion of this work. I greatly appreciate the patience they showed towards me during our numerous discussions not just in person but also over Skype. I would also like to mention my gratitude to the members of the Reliability of Electron Devices and Circuits group (REDEC) at the Universidad Autònoma de Barcelona for the shared time and discussions.

Besides my mentors, I would like to thank Prof. Francisco Gamiz from the Universidad Granada (Granada, Spain) who is truly an inspiration to me. This work would not have been complete without his collaboration. I'm also thankful for the fact that despite his busy schedule, he was accessible for having discussions. His enthusiasm towards the research was highly contagious and motivational, it was a joy to work with him.

My time at UAB was made memorable largely due to the friendship that I forged there. I fondly remember the discussions that ranged from politics to Game of thrones. I will always remember Vanessa's Spanish lessons that have helped me to speak more words than just "Hola". I'm really happy to have known you all, Albert, Ana, Paresh, …, to name a few.

Lastly, I would like to thank my family for their constant love and encouragement. For my parents who always encouraged me to pursue my dreams. My brother and sister-in-law who supported me in realizing my dreams. And most of all, I would like to thank my wife Varsha for the love, patience and support she had given me all this while during my PhD. Varsha, this work would not have been what it is without your support.

Contents

Publications related to this thesis

- V. Velayudhan, F. Gamiz, J. Martin-Martinez, R. Rodriguez, M. Nafria, and X. Aymerich, "Influence of the interface trap location on the performance and variability of ultra-scaled MOSFETs," in *Microelectronics Reliability*, 2013, vol. 53, no. 9–11, pp. 1243–1246.

- V. Velayudhan, J. Martin-Martinez, R. Rodriguez, M. Porti, M. Nafria, X. Aymerich, C. Medina, and F. Gamiz, "TCAD simulation of interface traps related variability in bulk decananometer mosfets," in *2014 5th European Workshop on CMOS Variability, VARI 2014*, 2014, pp. 1–6.

- V. Velayudhan, F. Gamiz, J. Martin-Martinez, R. Rodriguez, C. Marquez, M. Nafria, and X. Aymerich, "Threshold voltage and on-current Variability related to interface traps spatial distribution," in *45th European Solid State Device Research Conference (ESSDERC)*, 2015, pp. 230–233.

OTHER PUBLICATIONS

- A. Bayerl, M. Porti, J. Martin-Martinez, M. Lanza, R. Rodriguez, V. Velayudhan, E. Amat, M. Nafria, X. Aymerich, M. B. Gonzalez, and E. Simoen, "Channel hot-carriers degradation in MOSFETs: A conductive AFM study at the nanoscale," in *IEEE International Reliability Physics Symposium Proceedings*, 2013, p. 5D. 4.1–5D. 4.6.

- Q. Wu, A. Bayerl, M. Porti, J. Martin-Martinez, M. Lanza, R. Rodriguez, V. Velayudhan, M. Nafria, X. Aymerich, M. B. Gonzalez, and E. Simoen, "A conductive AFM nanoscale analysis of NBTI and channel hot-carrier degradation in MOSFETs," *IEEE Trans. Electron Devices*, vol. 61, no. 9, pp. 3118–3124, Sep. 2014.

- J. Martin-Martinez, M. Moras, N. Ayala, V. Velayudhan, R. Rodriguez, M. Nafria, and X. Aymerich, "Modeling of time-dependent variability caused by Bias Temperature Instability," in *Proceedings of the 2013 Spanish Conference on Electron Devices, CDE 2013*, 2013, pp. 241–244.

Preface

Over the past decades, an information world that envelops computers, the internet, wireless communication and global positioning system has emerged. Electronic devices have found a place in almost every aspect of our life and it is impossible to envisage a life without them. It is beyond our imagination as to how the semiconductor industry has evolved since its inception in the mid-twentieth century. This exhilarating growth was fueled by curious minds of scientist and engineers. To quote Shockley "accidents (inventions) favor the prepared mind", best describes the growth and the development of the electronics industry. In 1947, the team of John Bardeen, Walter Houser Brattain and William Bradford Shockley at Bell laboratories invented the first working point-contact transistor. This was a pivotal point in the history of electronics era, the introduction of transistor paved the way for many other devices and concepts, Metal Oxide Field Effect Semiconductor (MOSFET) being the more common and important one.

For many years, transistors were made as individual components and circuits based on individual transistors became too large. The integrated circuit, invented in 1960 made it possible to make and handle many devices in a single operation. These opened doors to cost reduction in electronics and increasingly larger systems. Since the 1960's the number of transistors per unit area has been doubling each year as a consequence of shrinkage in transistor sizes. Decade after decade, the transistor sizes have kept on reducing, leaving one wonder, how long can this continue. It appears now that this miniaturization has found its limits. For example, reduction of channel lengths in devices has given rise to a whole new set of problems known as short channel effects (SCE). New materials and novel architectures have been/are being developed to address these issues.

Variability of transistor characteristics has become a major concern associated with the scaling and integration of CMOS. Variability is the main factor restricting the scaling of supply voltages, which have for the last generations remained virtually constant. The statistical variability of transistor characteristics, that has been previously limited to analog design domain, has become a prominent concern in ultra-scaled devices. The statistical variability in scaled CMOS devices is introduced by the inevitability in the discreteness associated with charge and matter, atomic scale non-uniformity of the interfaces and the granularity of materials used in the fabrication of integrated circuits. Among different variability sources, Interface traps (IT), Random Dopant Distributions (RDD), Line Edge Roughness (LER) and Poly Gate Granularity (PGG) have been identified as the most prominent ones. As device dimensions are reduced, variations in the number and spatial distribution of dopants and IT can cause significant fluctuations in the device electrical characteristics. Apart from the increasing variability, the imminent difficulties related to statistical aspects of reliability will reduce the life span of integrated systems.

As a consequence of the pertaining issues and technical hitches, the numerical analysis and simulations have become popular. This has been supported considerably by the enormous progress in technology and performance of digital computers. Technology computer-aided design (TCAD) is an electronic design automation (EDA) tool that models the semiconductor device

operation based on fundamental physics. Simulation and optimization of the device before manufacture, though complex and time consuming, are nonetheless cost-effective. The cost of electronic product development, research and development, can be reduced by taking advantage of TCAD tools. TCAD tools and simulations have been increasingly used to study and analyze the impact of various sources of statistical variability in nano-scale devices. Simulations are being used to understand current variability mechanisms and also predicting the levels of variability in modern devices. With the aid of advanced computer simulations, researchers are able to identify the influence of individual and combined variability sources on the device performance. To this extent we can find many works addressing the variability in the device performance due to RDD, MGG and LER using TCAD simulations. However, there has not been many works that focus on understanding the role of interface traps (IT) in the device variability. Moreover, the tools that are used to perform the variability simulations have been developed in-house (some of them are now available commercially) and are computationally demanding. Thus, simulations that are not very expensive in resources and time could go a long way in helping to form a better understanding of the device variability. Following this line, in the work presented here, we use commercially available TCAD tools to perform the simulations to evaluate the variability in the transistor performance attending to the number and spatial distribution in the interface traps in nano-scale MOSFET's.

In chapter 1, the fundamental concepts of MOSFET, the basic component of the microelectronic industry is introduced. Among other things, the evolution of MOSFET over the years and the challenges/limitations in the path of its evolution are also presented. The concept of variability and different sources of variability are presented.

Chapter 2 is devoted to describe in detail the principles of TCAD simulations that are mainly used in the academia and industry. Here we also discuss how TCAD simulations are being used to study the variability in device performance in nano-scale devices. As this work focuses on the device variability perpetuated by interface traps, we also discuss how interface traps are simulated using TCAD.

Chapter 3 starts by outlining the methodology used in this work to perform TCAD simulations to address the variability in device performance perpetuated by spatial variation of interface traps. The results of 2D simulations focusing on the influence of the location (along the channel) of interface traps on device performance is presented. In this chapter, the variability of the threshold voltage due to the spatial distribution of discrete traps is analyzed.

In chapter 4, the 2D simulations are extended to 3D, where the spatial distribution of the traps is considered along the 2D interface. In this chapter, the variability in the threshold voltage perpetuated by the spatial distribution and number of interface traps along the length and the width of the channel is studied in detail.

Chapter 5 deals with the influence of the spatial distribution of interface traps on the variabilities of both the threshold voltage and on-current of the device.

Finally, the main conclusions are summarized in chapter 6.

CHAPTER I

Variability in CMOS technologies

The primary focus of the work presented in this thesis is to study the variability in the electrical characteristics of MOS structures due to interface traps. Therefore, this chapter introduces the concept and relevance of variability studies related to MOS devices. However, to have a better understanding of the concepts related to variability, in first half of this chapter, the fundamentals of MOS structure are discussed and then in the latter half of the chapter the challenges of variability in MOSFET devices are outlined.

1.1 MOSFET: Structure and electrical characteristics

The semiconductor industry has come a long way since the invention of the transistor in 1947. There has been an exponential growth of the industry which is now worth billions of dollars. This growth is fueled by the continuous advancement in the design of new devices and technological innovations. Early transistors were the so-called Bipolar Junction Transistor (BJT) which allowed faster speeds but were plagued by high power consumption. This limitation motivated researchers and scientists to look for transistors that had low power dissipation and this led to the development of Metal Oxide Semiconductor Field Effect Transistor, commonly referred by the acronym MOSFET.

The basic principle of MOSFET is the control (or modulation) of current conduction along the semiconductor surface by means of an applied electric field. Lilienfeld had described and patented the first field effect transistor, but probably could never make it work. The first functional

MOSFET based on the silicon substrate was demonstrated by Kahng and Atalla [1]. MOSFET is a unipolar device, that is, only one type of carrier is responsible for the current flow. Depending on the type of the carrier that contributes to the current flow, MOS transistors can be classified into two types: the n-MOSFET (electrons as carriers) and the p-MOSFET (holes as carriers). Figure 1-1 shows the schematic of a traditional MOSFET structure of n-type. A converse structure can be imagined for a MOSFET of p-type by substituting P for N and N for P.

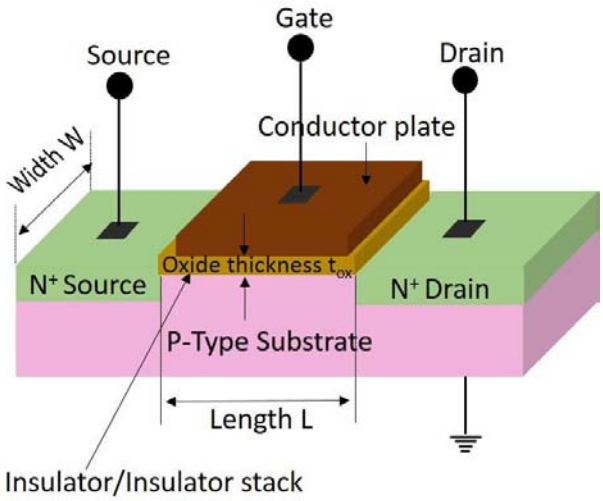


Figure 1-1: Schematic of an n-MOSFET.

In the figure, different regions of the MOS structure have been outlined. An nMOS transistor (as shown in the figure) is fabricated on a p-type semiconductor substrate which is typically silicon, however, devices with other materials (such as InGaAs, Ge) with faster carrier mobility have also been fabricated [2]. Two n-type regions are made in the substrate to enable current flow between them. The region with the lowest applied potential is called *Source* and that with the highest potential is called *drain*. A thin layer of the insulating material (or stack of different materials), known as the *gate oxide*, covers the region between the source and the drain. This layer is topped by a metal electrode (Conductor plate) called *gate*. Historically, the insulating material was predominantly SiO_2, however, due to performance and technological demands (will be outlined later) different insulator materials (or stack of materials) have replaced SiO_2.

In the most common operation mode of nMOSFET, the source (V_S) and substrate (V_B) terminals are grounded and the drain is biased to a positive voltage (V_D). When the gate is also biased at zero volts (V_G=0) then, there is no current flow from source to the substrate as they are at same potential (V_S=V_B=0). The back to back diodes that exist between the source and the drain prevent the carrier flow from source to drain. However, under a positive gate bias (V_G>0), free holes are repelled from the region under the gate creating a depletion region. The depletion region is then filled by the negatively charged electrons (which are in the majority in source and drain regions) that are attracted to the positive gate bias. For a large enough gate voltage (namely, the threshold voltage) a large number of electrons are accumulated near the surface underneath the gate, and an electron-rich layer called *channel* is created under the gate oxide. The channel is referred as an

inversion channel as the surface in this region has inverted from p-type to n-type. This channel forms a continuous electron bridge between the source and the drain enabling current flow (when a voltage is applied between source and drain) between these two electrodes. Channel length ('L' in Figure 1-1) is the length of this inverted channel under the gate and is an important transistor parameter. One of the central requirement for a field-effect-transistor (FET) is zero (ideal) or negligible gate leakage current. This is done by electrically isolating the gate from the substrate by placing a barrier for the electron (or hole) flow and in MOSFET, an insulator, with a thickness t_{ox}, provides this barrier.

The performance of a MOS device is evaluated by studying the so-called performance parameters (such as threshold voltage mentioned earlier) of the device from the electrical characteristics. So, in the following, an outline of the electrical characteristics (ideal) and extraction (or definition) of performance parameters of an n-MOSFET are discussed. The electrical characteristics of MOSFET under normal operation are influenced by the biases applied at its (four) terminals. In most applications, the source and bulk contacts are grounded (or biased at 0V) and depending on the application, gate and drain contacts are biased at different voltages.

In order to study the influence of the gate bias, let's have a look at the typical characteristics (I-V) of a nMOSFET. Figure 1-2 plots the transfer characteristics (drain current vs gate voltage) of a typical nMOSFET. It shows the plot of drain current (I_D) in both linear (red, circles) and logarithmic scale (black, square) as a function of gate voltage (V_G), for a positive drain bias (V_D). The characteristics (shown in Figure 1-2) were obtained for a device with a channel length of 250nm and an oxide (SiO_2) thickness of 2nm. In the example shown the drain is biased to 0.3V while the gate is ramped from 0 to 2V. We can see from the plot (with a linear scale, red circles) that there is no current below a particular gate voltage, called the *threshold voltage* (V_{th})[1]. However, when we use the logarithmic scale (black squares), it is clear that there is a non-zero current (known as 'off-current') which varies exponentially with the gate voltage below the so-called *threshold voltage*. This current flow between the drain and source (below *threshold voltage*) is known as *subthreshold current* [3] and is governed by the diffusion-dominated current [4] transport. The dependence of the subthreshold current on the gate voltage is characterized by a parameter called the *subthreshold slope* (SS) and is defined by the equation 1-1.

$$SS = \frac{dV_G}{d(\log_{10} I_D)} \qquad 1\text{-}1$$

The subthreshold slope (SS) is expressed in millivolts per decade of the current and is typically of 70-100mV/decade (in the example shown it is 80mV/decade). The subthreshold slope along with the threshold voltage is an important parameter in the transistor design. Above the *threshold voltage,* the drain current increases linearly with the gate voltage; i.e. the transistor operates in the linear region. The magnitude of drain current (I_D), depends on the density of electrons in the channel, which in turn depends on the gate voltage. Hence, an increase in the gate voltage would

[1] There are different ways of experimentally calculating the threshold voltage and they are discussed in the later sections.

result in an increase of the drain current by increasing the number of electrons being attracted to the surface.

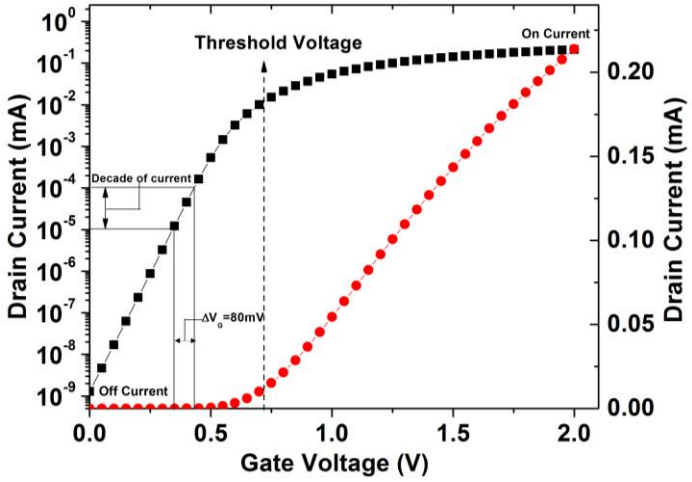


Figure 1-2: Drain current as a function of gate voltage. The two curves show identical data plotted in linear scale (blue circles, right y-axis) and logarithmic (black squares, left y-axis). The characteristics were obtained for a device with a channel length of 250nm and an oxide (SiO_2) thickness of 2nm.

In order to understand the influence of the drain bias, let's consider a situation where the channel is already formed ($V_G > V_{th}$). The drain contact is biased at 0V, which meant that there is no current flow between the source and the drain as both the contacts are at the same (0V) potential. Now, let's slowly increase the drain voltage. Due to the existing biasing conditions ($V_G>V_{th}$, $V_D>0$ and $V_S=V_B=0$), the channel is tapered as the voltage drop at the drain end is lower than the one at the source end. As the drain bias is increased further, the channel becomes more and more tapered and at a point where $V_G-V_D=V_{th}$, the channel depth at the drain end is almost zero. At this point, the channel is said to be pinched off, and increasing the drain bias any further has a negligible effect on the current (source to drain). The current is saturated and the device is said to be operating in the saturation region. The drain bias that causes this saturation is referred to as V_{Dsat} ($V_{Dsat}= V_G-V_{th}$) and the saturated value of the current is known as saturation current (I_{Dsat}). Figure 1-3 shows different regions of the transistor operation in the plot of output characteristics. The expressions of drain current in triode (or linear region) and saturation region are shown in equations 1-2 and 1-3 respectively.

$$I_D = \mu_n C_{ox} \frac{W}{L}\left[(V_G - V_{th})V_D - \frac{1}{2}V_D^2\right] \qquad 1\text{-}2$$

$$I_{D,max} = \frac{\mu_n C_{ox} W}{2L}(V_G - V_{th})^2 \qquad 1\text{-}3$$

Here, L is the effective length of the channel, W is the width of the gate, C_{ox} is the oxide capacitance and μ_n is the electron mobility.

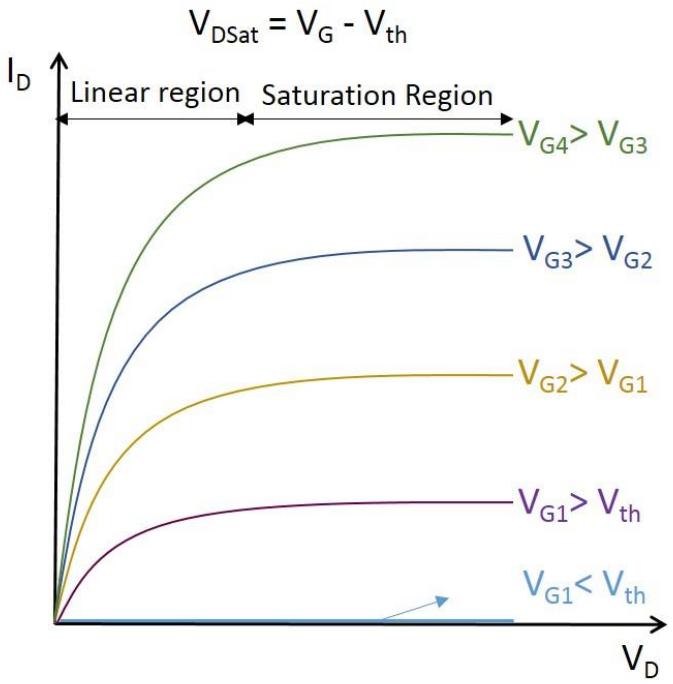


Figure 1-3: Output characteristics of an n-channel MOSFET, showing different operation regions of transistor.

Definition and extraction of the MOSFET performance parameters

As stated in the previous sections, the performance of a MOSFET is usually characterized by its *threshold voltage* (V_{th}) and often also by *on-current* (I_{on}). The measurement of *on-current* is straight forward as it is defined as the drain current for a particular gate voltage (usually at high V_D) [5]. However, V_{th} definition is a bit complex as there are several different definitions and measurement methods [6]–[8], each of them focusing on different aspects. Essentially, the threshold voltage can be understood as the gate voltage at which the transition between weak and strong inversion takes place in the MOSFET channel. The theoretical definition of V_{th} is based on the strong inversion condition at which the surface potential is twice the bulk Fermi potential. However, this definition is not practical as the surface potential is not a measurable parameter, hence, more practical methods are needed to measure (extract) V_{th}. The two most commonly used methods to extract V_{th} are the so-called "Extrapolation in linear regime (ELR)" and the "constant current (CC)" methods. In ELR method, a linear extrapolation of the I_DV_G curve at the maximum slope (also known as transconductance peak, g_m) is determined and the zero intercept ($I_D = 0$) of this line is defined as V_{th}. Figure 1-4(left) shows the extraction of V_{th} based on the ELR definition. For a better visualization, the inset in the figure shows the zoom of the curve delineated by the blue dotted box. In the CC method, a current value is selected and V_{th} is the gate voltage that corresponds to this selected (fixed) current. With the changes in the device dimensions, the value

of this fixed drain current is usually scaled by the ratio of W/L, where W and L are the channel width and the length respectively Figure 1-4 (right) shows the extraction of V_{th} based on the CC definition.

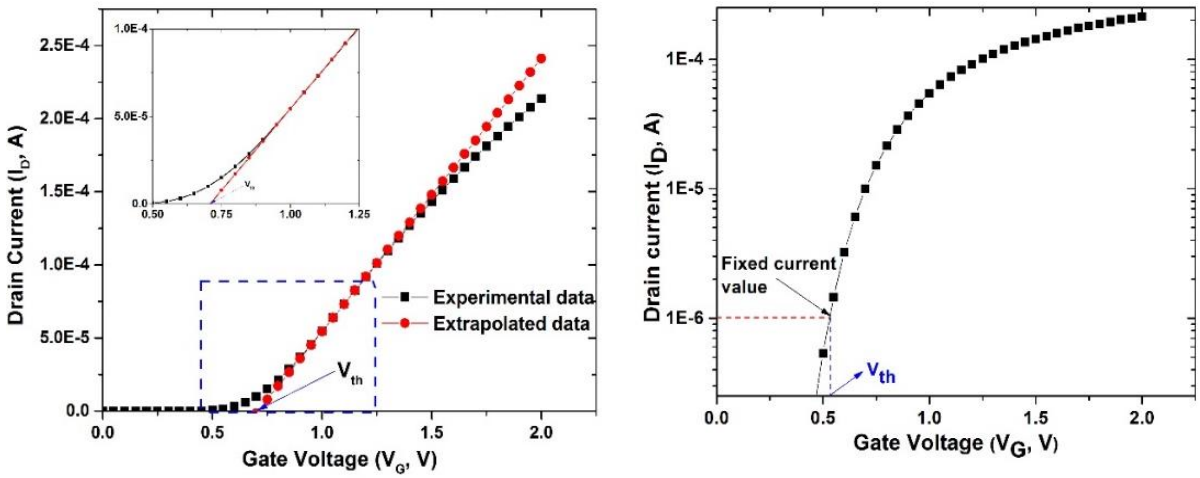


Figure 1-4: The figure on the left shows the extraction of Vth using ELR method. The intercept of the red line (with circles) on to the X-axis (Gate voltage) is the threshold voltage. The figure on the right shows the extraction of Vth using constant current method, here the current was fixed at 1µA.

Both the methods have their merits and limitations. The main drawback of ELR method is that it is only valid for the linear region of operation (low V_D) and the maximum slope point might be uncertain as the I_DV_G curves can deviate from the ideal straight line behavior even slightly above V_{th} and thus can significantly influence the calculation. Moreover, the ELR method is very sensitive to parasitic source/drain resistances [9]. On the other hand, using CC, V_{th} can be determined very quickly but is totally dependent on the arbitrarily chosen value of the drain current. Depending on the analysis required, in this work, V_{th} is defined using either 'ELR' or 'CC' method.

1.2 MOSFET Evolution: Past, Present and Future.

Since the invention of the first transistor almost half a century ago, the world has seen an unprecedented growth of the semiconductor industry. This growth has had an enormous impact on the lives of most of us. The sustained growth of the semiconductor industry is fueled by scaling which is the continuous shrinkage of transistor size and dimension. This race of transistor miniaturization started with the publication of a paper by Gordon Moore in which he made an important observation that the transistor density on a chip would double every 10 to 24 months [10]. This observation has since been dubbed "Moore's Law" and has been the central driving force of the semiconductor industry. Moore's law has also been viewed by many as a "self-fulfilling" prophecy because industries have set this standard as an objective rather than a consequence. Figure 1-5 shows how the number of transistors per chip has increased over the years.

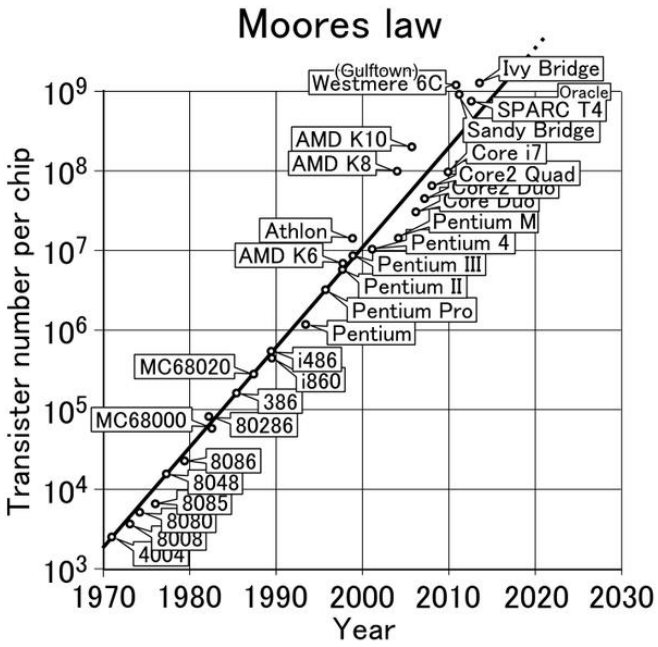


Figure 1-5: Transistor density per chip over the years [11]

This growth in the transistor density has helped the electronic industry to grow at an exponential rate and large part of this growth is due to the fact that transistors can be scaled to increasingly smaller dimensions. There are many motivations behind scaling, prominent among them are the increased packaging density and faster devices. The International Technology Roadmap for Semiconductors provides a taxonomy of scaling, which are Geometrical Scaling, Equivalent Scaling and Design Equivalent Scaling [2]. Geometrical scaling refers to the continued shrinking of the transistor feature sizes (horizontal and vertical) in order to improve density and performance. Equivalent scaling, which occurs in conjunction with geometrical scaling, refers to the improvements in the device structure (alternate device architectures) and other non-geometrical techniques (use of new and alternate materials). Design Equivalent scaling, which also occurs in conjunction with geometrical and equivalent scaling, refers to the improvement of design technologies that enable an increased performance, low power consumption, high reliability and low productivity costs. From a technology perspective, the continuous increase in the transistor density was made possible by dimensional (geometrical) scaling, the benefits of which were demonstrated in the works by Dennard [12]. The team led by Dennard described how the steady reduction of transistor feature dimensions (gate length, gate insulator thickness) could simultaneously improve switching speed, power consumption and transistor density. According to Dennard, scaling rules were governed by two scaling methods: constant field scaling and constant voltage scaling[12], [13]. Constant field scaling warrants a reduction in the power supply voltage as the minimum feature size is reduced. While the idea of constant voltage scaling was to keep the supply voltage at certain predetermined voltage nodes (5V, 3.3V…) as device dimensions are scaled down. In practice, constant voltage scaling is more useful than constant field scaling as the scaling of voltages require multiple power supply arrangements thus becoming impractical in many scenarios. The electronic industry has not followed the set of rules, known as Dennard's scaling law, as demand for faster devices drove down transistor gate length faster than Dennard's

Law [14]. And despite its merits, miniaturization of devices has also brought a host of issues. In the following, we discuss the challenges that arise as a result of device scaling and how these challenges are confronted with technological advancements.

Challenges to MOSFET evolution

Scaling of the devices resulted in a reduction of the device dimension in lateral and vertical directions. Lateral scaling meant a reduction in the channel lengths, which gave rise to the so-called short-channel effects (SCE). The evolution of MOSFET scaling is a continuous battle against the short-channel effects. For instance, for devices with very short channels, the distance between the source and drain is also very short and becomes comparable to the depletion layer widths of the source/drain junctions. Due to this short distance, the potential at the drain contact reduces the peak value of the energy barrier in the channel. This effect, known as drain-induced barrier lowering (DIBL), leads to a decrease in the threshold voltage with channel lengths. To mitigate these effects silicidation of the S/D junctions and lightly doped drain/source (LDD) structures with pocket halo implants and super-steep retrograde body doping were used [15][16]. The reduction in the distance between the source and the drain also weakens the gate control as current sneaks through the part of the channel that is farthest from the gate. Having shallow junctions reduces the influence of the drain on the channel and prevents the junction electric fields from penetrating deeper into the channel. However, it also increases the resistance of these regions and doping cannot be increased (to increase conductivity) beyond certain limits. For the realization of CMOS circuits, it was necessary to integrate n-MOS and p-MOS devices close to each other and the use of metal gates made it difficult to find materials with complimentary work functions. Polysilicon (poly-Si) gates allowed the modification of the work function by doping and hence were more suited for large-scale integration [17]. However, a depletion layer is formed at the interface to the dielectric when a voltage is applied to the gate and an undesired voltage drop occurs within this interface. This effect, known as 'polysilicon depletion' leads to a reduced electron concentration at the interface resulting in an increase of the effective dielectric thickness and increased threshold voltage [18].

Vertical scaling of devices reduced the thickness (t_{ox}) of the gate oxide, which resulted in an increased the gate capacitance and gave more control of channel to the gate [4]. The gate oxide thickness scales with almost the same factor as channel length scaling. However, it also increases the gate leakage (gate current) due to quantum mechanical tunneling. Quantum mechanical tunneling comes into effect when the energy barrier between the gate and the semiconductor becomes too small [19]. It is estimated that gate leakage current increases approximately 30 times every technology generation, as opposed to 3-5 times increase of channel leakage current [20]. For example by the 90nm node the thickness of the gate oxide reached the orders of Angstroms [21] and at these thicknesses, the gate leakages were too dominant to be neglected. One solution to this problem was to use dielectric materials which have a higher dielectric permittivity [22]. These materials allow the use of higher physical thickness (t_{ox}) but give lower electrical thickness (t_{eq}) than SiO_2. In devices with high-k dielectrics, the insulator is usually a stack consisting of a layer of high-k dielectric on top of a layer of SiO_2. In these devices, the oxide thickness is referred as equivalent oxide thickness (EOT), which is the equivalent thickness of SiO_2 layer that is needed to obtain the same capacitance as the one obtained by the use of high-k stack. The mathematical

expression for EOT is given by equation 1-4, where K and t are the dielectric constant and physical thickness, respectively, of the materials indicated in subscript (high-k and SiO_2).

$$EOT = t_{high-k} * \frac{K_{SiO_2}}{K_{high-k}} \quad 1\text{-}4$$

Several different binary metal oxides such as Ta_2O_5, TiO_2, ZrO_2, Al_2O_3, HfO_2 and their silicates have been extensively studied to be a suitable replacement for SiO_2 [23]. Among the many materials studied, HfO_2 is believed to be the most promising candidate [24], [25]. In 2007, Intel successfully introduced devices with HfO_2 in their 45nm technology node [26]. The combination of high-k and poly-Si gate is not suitable for high-performance logic applications due to high threshold voltage resulting from Fermi level pinning at the poly-Si/high-k interface [27]. A solution is to use metal gates instead of the poly-Si gates. Moreover, the use of metal gate also eliminates the poly depletion, resulting in an improved performance [26].

In 90nm node and beyond, as transistor channel lengths dropped to 45nm, strain engineering found a large acceptance as a promising technique to improve CMOS performance. The mobility of carriers in silicon is enhanced if a biaxial tensile strain is applied to it [28]. In early 2003, Intel was the first to fabricate chips with strain engineering in their 90nm node generation [29]. The silicon channel in these transistors had been physically squeezed to boost speed and reduce the power consumption. To improve device performance, non-silicon high-mobility compound semiconductors, such as Ge and III-V have also been studied [30].

As stated earlier, the introduction of high-k material helped to control the gate leakage current due to tunneling of carriers, however, the leakage from the source to drain (punch-through) was still a challenge in short channel devices. A direct method to address the punch-through was to use devices with fully or partially depleted substrates. The absence of free carriers (except in the channel) eliminated the punch-through [31]. Depleted-substrate devices are realized by using silicon-on-insulator (SOI) substrates. Figure 1-6 shows a cross-section of a SOI MOSFET. It consists of standard MOSFET with a substrate that is insulated from the wafer by a layer of insulator (typically SiO_2, 'BOX' in the figure). Based on the thickness of this insulator BOX, SOI transistors are classified into two types: Partially Depleted (PD) and Fully Depleted (FD). An SOI is considered as partially-depleted (PD) SOI if the silicon film (typically 100nm or more) on the box layer is thicker than the depletion region depth beneath the gate oxide and is considered fully-depleted (FD) SOI if the silicon film is thin enough (typically 50nm or less) or the doping concentration is low enough to be fully depleted. FD-SOI is also often referred as ultrathin body silicon-on-insulator, or simply UTB SOI. SOI technology provided the foundation for new device structures with multiple gates on a single device. For example, a solution to address the punch-through was to use the UTB SOI structure and turn the thin silicon channel on its side by 90 degrees, creating a 3-D device with a "fin" that juts out of the device plane. The channel of this new type of MOSFET resembled the 'fin' of a fish, hence the name FinFET [32]. The devices with multiple gates are known as multi-gate field effect transistors (MuGFETs) [33] and different multi-gate structures are shown in Figure 1-7.

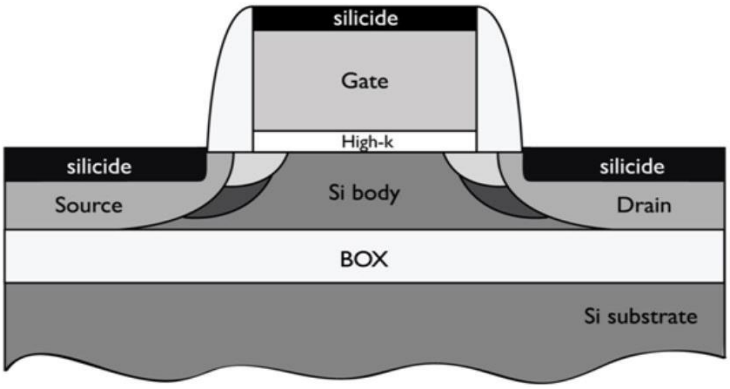


Figure 1-6: Cross-section of a SOI MOSFET [34]

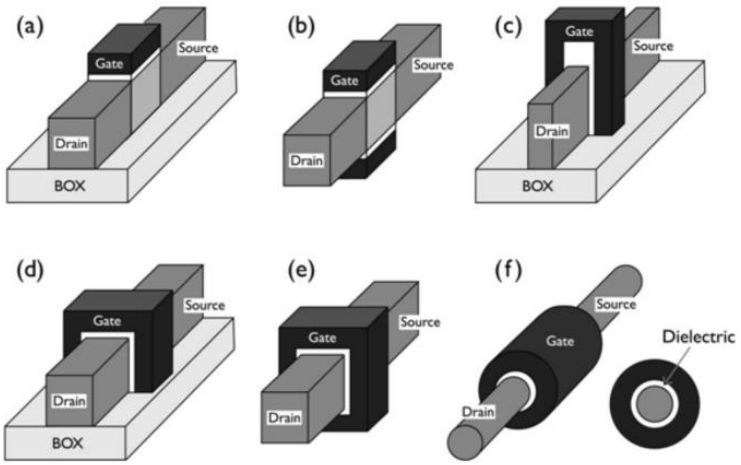


Figure 1-7: Different structures using SOI technology: (a) Single gate SOI, (b) double gate planar SOI, (c) double gate non-planar FinFET, (d) tri-gate FET (e) quadruple-gate (gate-all-around) FET and (f) nanowire (surrounding gate) FET[34] ***.***

Figure 1-7 (f) shows the structure of a nanowire FET with an all-around gate which provides superior control of the channel than conventional planar structures [35]. Recently, Intel has developed technologies to manufacture radically new 3D tri-gate transistors at 22nm and below to ensure the continuation of transistor growth for the foreseeable future [36], [37]. However, due to their small diameters, any variations in the dimensions due to fabrication imperfections can degrade the charge transport characteristics [38]. One of the fundamental limits of MOSFETs is the lack of scalability of the subthreshold slope (SS) below 60mV decade^{-1} at room temperature [39]. Researchers have suggested using impact ionization effects and quantum-tunneling effects to improve this limitation but these techniques have not proven to be very reliable [40], [41]. On the other hand, carbon nanotube FETs with surround-gate geometry have shown to improve subthreshold slope and devices with channel lengths scaled down to 9nm have shown the absence of SCE [42]. Graphene FETs have shown to have higher carrier mobility than carbon nanotube FETs and they are being extensively studied [43], [44].

Apart from the technological limitations and the challenges, evolution of MOSFET devices has been and still is confronted by reliability and variability. The channel doping required to control SCE in sub-32nm bulk MOSFETs is expected to be around 10^{18} cm^{-3} and above [45], [46]. These extreme levels of doping have a twofold consequence: they severely degrade the mobility due to impurity scattering and increase the variations in threshold voltage due to random dopant fluctuations [47], [48]. The increasing random dopant distributions (RDD), line-edge roughness (LER), non-rectangular gate and defects (or traps) along the silicon/insulator interface can cause significant statistical variations in V_{th} and, in-turn, affect the stability and performance of the device (or circuit). The issues surrounding variability are an important part of this thesis and are discussed in depth in the following sections.

1.3 Variability in MOSFETs

Variation is the deviation of the obtained value of a parameter from its design intent. While variation is certainly not new, the continued decrease of the feature sizes has made the management of this variation very challenging. Due to the continuous scaling of the transistor dimensions and the rapid introduction of sub-nanometer technology nodes (32nm and 22nm), the variability of transistor characteristics (V_{th}, I_{on}) have become a major concern associated with the further scaling and integration of CMOS devices [49], [50]. Figure 1-8 shows as an example, the experimental threshold voltage (V_{th}) distribution measured in 8000 devices in an array. Variations in the device parameter (or parameters) resulting from the stochastic nature of many physical processes involved during fabrication is referred as static variability (SV).

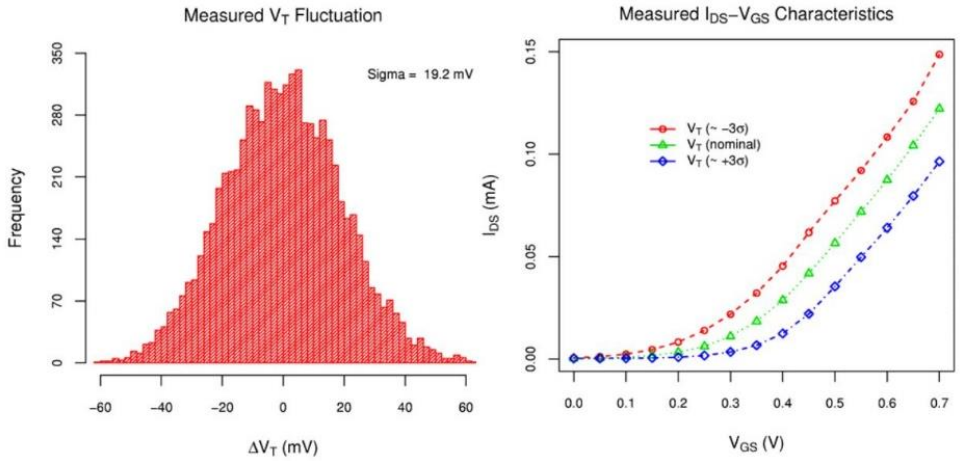


Figure 1-8: (left) Measured histogram of threshold voltage fluctuation and (right) I_D-V_G curves for three devices in the array with their threshold voltage at -3σ, nominal and 3σ [51].

On top of static variability, device parameters can change during circuit operation and hence might cause a fully functional circuit (of which the device is a part) to fail or be less functional. Such variability can be categorized as time-dependent variability or dynamic variability. There are diverse kinds of variability, which are different in their origins and behavior. Variability in

integrated circuits is usually classified by its correlation between multiple transistors placed at different positions as shown in Figure 1-9. The categories from (a) to (d) in the figure represent the systematic variability associated with the manufacturing process of the device. These variations, known as systematic variations, stem from the changes in the manufacturing apparatus (wafer-to-wafer, and wafer level), lithography equipment and lithography steps (die level, layout dependent). Systematic process variation has been of interest to semiconductor manufacturers as it is a strong driver of yield. The management of systematic variation is critical to maintaining competitive yield over multiple technological nodes [52]. The category (e) in Figure 1-9 represents Random Variability and is one of the most serious challenges to be confronted in order to continue scaling down of FETs. Variability is the main factor restricting the scaling of the supply voltage, which has remained virtually constant for the last many technological generations.

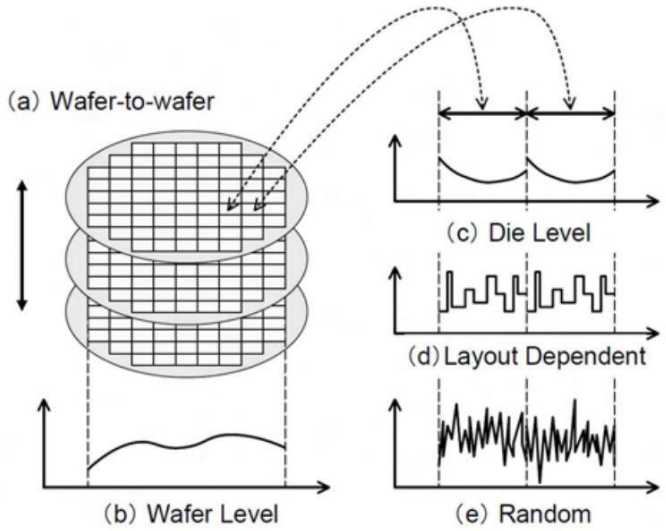


Figure 1-9: Variability categories and their impact [53].

Random Variability refers to a kind of variability that exhibits no correlation between neighboring devices. Because of random variability, the transistor characteristics are statistically dispersed and hence random variability is also referred as statistical variability [54]. The statistical variability of transistor characteristics has become a major roadblock associated with CMOS scaling and integration [55], [56].

1.4 Statistical Variability

As outlined earlier, statistical variability exhibits no correlation between neighboring devices and such an absence suggests that it is caused by some microscopic perturbations. These microscopic perturbations, which are exacerbated by scaling are introduced by discreteness of charge and matter, such as random placement of discrete charges, atomic scale irregularity of gate dielectric and so on [57]. Though statistical variability has been under investigation for a long period of time, it has only become industrially relevant as devices reached the deep sub-micron regime. As a matter of fact, it has been demonstrated that the magnitude of purely statistical variability in the

45nm technology has exceeded that of systematic variability [53], [58]. The dominant sources of statistical variability are random dopant fluctuations (RDF) [59]–[61], polysilicon/metal gate granularity (P/MGG) [55], [62]–[64]line edge roughness (LER), line width roughness (LWR) [65] and interface traps [66]–[70]. For 45nm technology, the random dopant fluctuations are known to be the principal physical sources of fluctuation among the most important ones (RDF, LER/LWR, MGG/PSG) [59]. Fully depleted thin –body silicon-on-insulator (FD-SOI) have shown to exhibit lower V_{th} variability associated with RDF and LER [71]. The 3D FinFET structure introduced by Intel in their 22nm technology was in response to reducing the statistical variability in nanoscale transistors. However, in these devices, there are new important sources of statistical fluctuations, such as, fin edge roughness (FER) and high-k/metal gate granularity which are associated with FinFET fabrication process [56].

Apart from being a serious concern for device scaling and integration, the statistical variability has also shown to adversely affect the yield and reliability of SRAM [49], [50], [72]. Owing to its true random nature, statistical variability generates performance differences in otherwise identical transistors. Statistical variability is impossible to reduce due to its atomistic nature and would be present even in a hypothetical "ideal" fabrication process where process and systematic variability are absent. This warrants the inclusion of statistical variability during the transistor and circuit design process. In the following sections, a brief discussion about different sources of statistical variability in modern CMOS devices is presented.

1.4.1 Random Dopant Fluctuations

Random dopant fluctuations (RDF) also referred as random discrete dopants distribution (RDD) results from the discreteness of dopant atoms in the channel of the transistor. Transistor channel is doped with dopant atoms to control its threshold voltage. With each technology generation, the transistor area is reduced, resulting in the exponential decrease of the number of dopant atoms in the channel. RDF is believed to be the dominant source of statistical variability in bulk transistors below 22nm technology node [49]. Figure 1-10 shows the cause of RDF variability in (left) ideal, (center) real and (right) scaled transistor. The problem was pointed out in the early seventies [47][73] and first treated analytically and numerically in [74]. The random fluctuations of the relatively small number of dopants and their discrete microscopic arrangement in the channel will lead to significant variations in the threshold voltage and drive current of nominally identical devices. The first order effect of random dopant fluctuation is a random shift in the threshold voltage of a device and studies have shown that the RDF induced V_{th} distribution is close to Gaussian, with deviation in the tails of the distribution (for a sufficiently large sample size) [75]. Although most works concerning RDF focus on the electrostatic effect on the threshold voltage [75]–[77], random dopants are also known to cause transport variability as a result of variations in ionized impurity scattering [78].

1.4.2 Line-edge and Line-width Roughness (LER and LWR)

Line-edge and Line-width roughness (LER/LWR) stem from the imperfections in the photoresist during the lithography and etching processes, affecting the shape of the edges of the transistor critical dimensions. LER refers to the distance between the edges of one side with respect to the nominal line, while LWR refers to the width variation from one side of the rectangular shape to

the other. Figure 1-11 shows the definition of LER and LWR as found in the literature. LER/LWR causes the channel length and width to vary in such a way that the transistor edges are no more rectilinear as drawn in the layout. LER/LWR are associated with increases in the sub-threshold current [80], [81] as well as degradation in threshold voltage characteristics [65], [82]. In larger devices LER and LWR caused little worry but for channel lengths below 85nm, they become critical factors for mismatch studies [81]. As channel lengths are shortened LER does not scale accordingly, becoming a larger fraction of the gate [65], [83]. A worrying effect of LER/LWR is the degradation of I_{on}/I_{off} ratio, due to enhanced short channel effects seriously impeding the device and circuit performance [84]. In [81] the combined effect of LER and RDF on current fluctuations was investigated and was demonstrated that these two sources of variability act in a statistically independent manner. Furthermore, it was shown that as devices are scaled to shorter dimensions (around 14nm) LER supplants RDD as the dominant variability source [65]. LER effects have shown to be the dominant source of variability in multi-gate devices such as FinFETs [56].

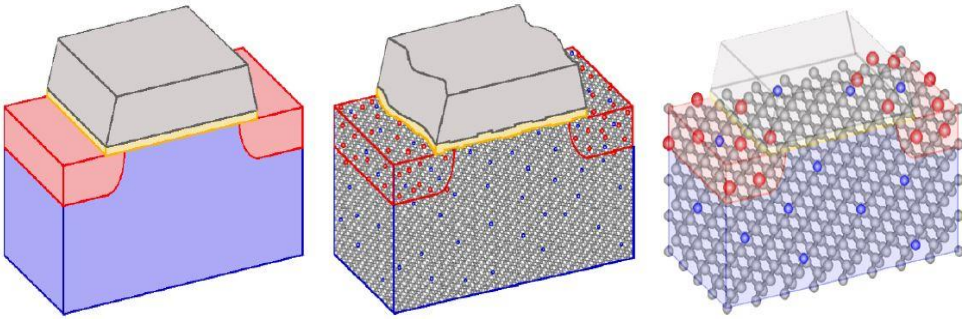


Figure 1-10: (left) An Ideal structure with continuous doping, (center) real transistor showing silicon lattice and dopants, (right) scaled transistor (channel length of 4.2nm) emphasizing a small number of random dopants [79]***.***

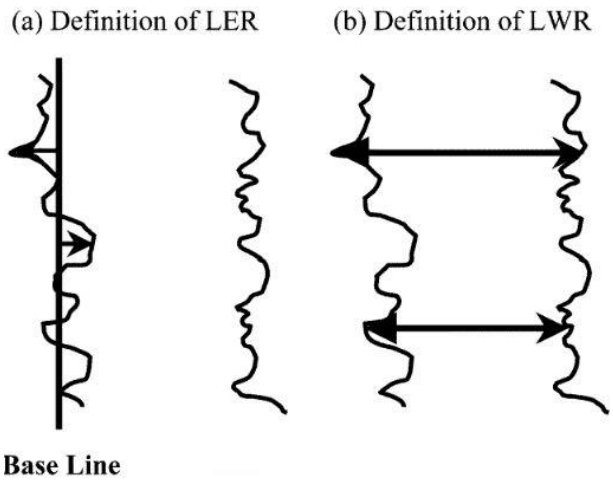


Figure 1-11: Definition of LER and LWR [81]

1.4.3 Poly-silicon/Metal Gate Granularity

In transistors with polysilicon gate, the polycrystalline granular structure of the polysilicon gate has been identified as an important source of variability [85]–[87]. Figure 1-12 is an AFM image of the polysilicon sample showing the grains; the average grain size (grain diameter) extracted using this image is 210nm.

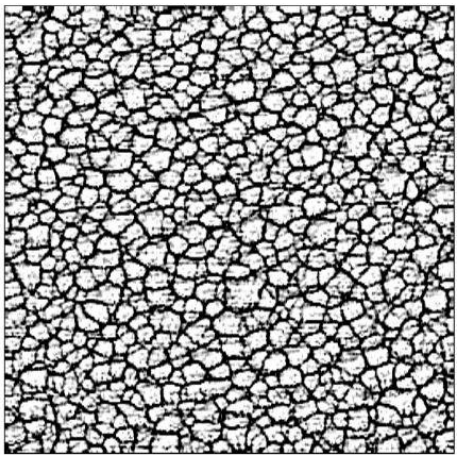


Figure 1-12: AFM image of the polysilicon sample showing grains [88].

There are many contributors to the variability within the polysilicon, such as penetration of dopants from the high doping regions in the gate into the channel through the gate oxide [89], extension and pocket definition variations due to implantation variation through polysilicon grain with different orientations [90], [91]. However, Fermi-level pinning due to the high density of defect states at the grain boundaries is likely to be the most significant source of fluctuations within the polysilicon [85], [92]. The Fermi-level pinning at the grain boundaries facing the polysilicon/gate-oxide interface induces fluctuations in the surface potential within the channel of the MOSFET. Depending on the location of the grain boundaries in the gate with respect to the channel, variations in the threshold voltage and current characteristics from device to device are generated.

In modern transistors high-k dielectric and metal gate are used to suppress the tunneling current. The use of metal is due to the incompatibility of polysilicon with high-k dielectrics [93] but it also increases the on-current by eliminating the poly-depletion [23]. The metal gate also eliminates the PGG (polysilicon gate granularity) and offers an improved screening of the mobile carriers (in the gate) from the RDF induced potential fluctuations in the channel, thus improving the RDF induced variability. However, the method of gate deposition determines the extent of statistical variability introduced [94]. In the "gate first" process technology, the gate metal is deposited before any high-temperature annealing process and during the high-temperature processing, the normally amorphous metal becomes polycrystalline [95]. This results in the formation of grains of differing metal work function associated with metal gate granularity. Both of these effects (PGG and MGG) have a serious impact on the device performance and the effect is highly dependent on the material grain size with respect to the overall gate size [94].

1.4.4 Area dependence

The variations in device parameters (mainly on V_{th}) are gaining importance with every technological node, as the stochastic spreading becomes increasingly pronounced with diminishing device dimensions. These undesired V_{th} variations introduce device mismatch and, in analog circuits, device mismatch between identically designed devices is the key limitation to the accuracy of the circuit [96]. For this end, the mismatch in MOS transistors has been extensively studied for more than three decades. Transistor mismatch was modeled using an approach in which the threshold voltage fluctuations in the whole device volume or surface stem from the random number of discrete entities controlling the V_{th}. These entities can be channel doping impurities, interface oxide charges and material granularities. The randomness of these numbers are governed by Poisson distribution law [47], [97].

It was demonstrated in [98] that the electrical parameter fluctuations in MOS transistors have a dimensional dependence. The rule, which is since then referred as Pelgrom's law, states that the variability of threshold voltage has an inverse relationship with the device active area. For example, the threshold voltage mismatch, ΔV_{th}, between identical transistors with the active area (Width * Length) is given by equation 1-5 and upon simplification yields equation 1-6 [97], [98]

$$\sigma^2(\Delta V_{th}) = \frac{A_{VT}^2}{WL} + S_{VT}^2 D^2 \qquad 1\text{-}5$$

$$\sigma V_{th} = \frac{A_{VT}}{\sqrt{L\,W}} \qquad 1\text{-}6$$

$$A_{VT} \propto t_{INV}\sqrt{N_{SUB} W_{DEP}}$$

Here σV_{th} (ΔV_{th}) is the standard deviation in the threshold voltage (threshold voltage shift), and A_{VT} is referred to as the Pelgrom's slope, an important merit parameter. A_{VT} is linearly dependent on the inversion layer thickness (t_{inv}) and square-root of depletion width (W_{DEP}) and channel doping (N_{SUB}). Pelgrom's law was further discussed by Mizunu, where they verified experimentally that the V_{th} mismatch is given by a Gaussian function and this distribution results from the random doping fluctuations in the depletion region [99]. In this work, it was shown that the σV_{th} (threshold voltage variation) depends on the fundamental process parameters (gate oxide thickness, channel dopant concentration) and is given by equation 1-7.

$$\sigma V_{th} = \left(\frac{\sqrt[4]{4q^3\epsilon_{Si}\phi_B}}{2}\right) * \frac{T_{ox}}{\epsilon_{ox}} * \frac{\sqrt[4]{N}}{\sqrt{W_{eff}L_{eff}}} \qquad 1\text{-}7$$

Here the key features are a linear dependence on the oxide thickness T_{ox}, an inverse square root dependence on the effective channel area (effective length (L_{eff}) * effective width (W_{eff})) and an inverse fourth-root dependence on the channel doping (N). It can be seen that matching improves with decreases in the channel doping and gate oxide thickness but worsens with the scaling of the device dimensions. The limitation of the equation is that it assumes that the only contribution to the random variation between two matched devices is random dopant fluctuation (RDF) while in practice it is known that additional effects also contribute to the measured variation [100].

1.5 Time-dependent variability

As it was discussed in the earlier sections MOSFETs are being scaled aggressively to reach their physical limits. However, the supply voltage at which these devices are operated has mostly remained constant over the generation. This non-scalability of the supply voltage has led the ever so shrinking (in dimensions) devices to be operated under extremely high electric fields. While the consequent increase in the internal electric field has increased the carrier velocity, thus increasing switching speeds, it has also presented major reliability problems to the long-term operation of the device. With the scaling of the devices, the reliability problems outweigh the benefits of higher electric fields.

The sources of variability that were discussed in the earlier section (statistical variability) were all static in nature, i.e. they occur during the fabrication of the device and do not depend on their operating conditions. However, transistor parameters can degrade as a function of time depending on the applied operating conditions (i.e. current/voltage for a period of time (short/long)) as well as operating temperature. This device degradation can partially or completely affect the functionality of circuit of which the device is a part. In other words, the degradation of a circuit's device (or devices) can cause the circuit parameters to vary partially or completely. These degradations are time and context variant and they aggravate the induced (static) statistical variability. As these variations depend on the operation (which is time-dependent) of the device, the associated variability is often referred as time-dependent variability. Therefore, to manage such sources of variability, it is imperative to have a better understanding of device degradation.

The degradation of device parameters over time is called aging and among different aging mechanisms in MOSFET, Bias Temperature Instability (BTI), Hot Carrier Injection (HCI) and Time-Dependent Dielectric Breakdown (TDDB) are noteworthy [101]–[105]. MOS structures also suffer from the random discrete fluctuations in drain current as a function of time, which resembles a Random telegraph noise (RTN) [106]. Performance degradation is usually monitored by estimating the shift in the I_D-V_G characteristics with respect to the fresh (i.e. before the application of stress) characteristics by the measurement of ΔV_{th} at a certain I_D as shown in Figure 1-13. It can be seen from the graph that a shift in the threshold voltage, ΔV_{th}, is measured as the difference in the voltages ($V_{t2} - V_{t1}$) corresponding to a drain current I_D. Fixed-charges and defect states in the oxide or at the insulator/substrate interface in the device are known to contribute (as will be explained in the following sections) towards the different aging mechanisms [107]–[110] In the following section, we discuss in brief about RTN, different prominent aging mechanisms (such as HCI, BTI) and the aging facilitator: interface traps

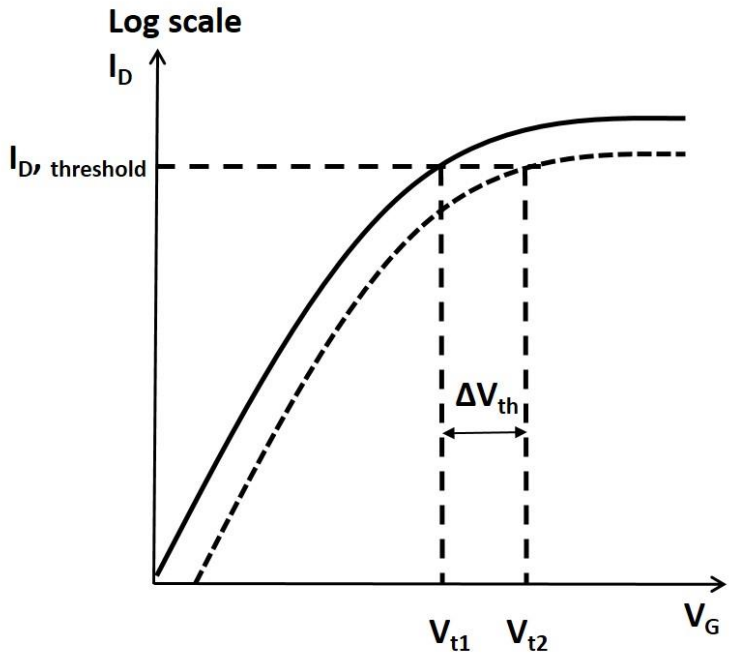


Figure 1-13: Performance degradation in any type of MOSFET is usually measured as a shift in the I_D-V_G characteristics giving rise to a change in V_{th} (ΔV_{th}).

1.5.1 Random Telegraph Noise (RTN)

In short channel devices, the MOSFET drain current (I_D) exhibits random discrete fluctuation or switching events as a function of time. These back and forth switch of channel current between two values resembles a telegraph signal with random timing hence the name Random telegraph noise (RTN) [106]. The variations in drain current due to RTN is schematically represented in Figure 1-14. RTN is due to the transient nature (capture and emission) of some lingering charge traps (in the oxide and at the interface) even when the devices have not been previously electrically stressed and can be modeled as transient changes in the threshold voltage V_{th} [111]. The effective V_{th} shift is shown to have an inverse relationship with channel area and to be significant in highly scaled devices and circuits such as SRAM and Flash memories [112]–[115]. When a trap emits an electron, becoming neutral, the effective threshold voltage (V_{th}) decreases and the channel becomes more conductive resulting in an increase in the drain current. And after a time (capture time, t_c), the trap captures an electron, resulting in an increase of the effective V_{th} and decrease in the drain current. It stays in this state for a time (emission time, t_e) until the electron is emitted and the cycle is repeated. The process is represented in Figure 1-15. The figure clearly shows the process where electrons are captured by empty traps (black circles) or released from a filled trap (blue circle).

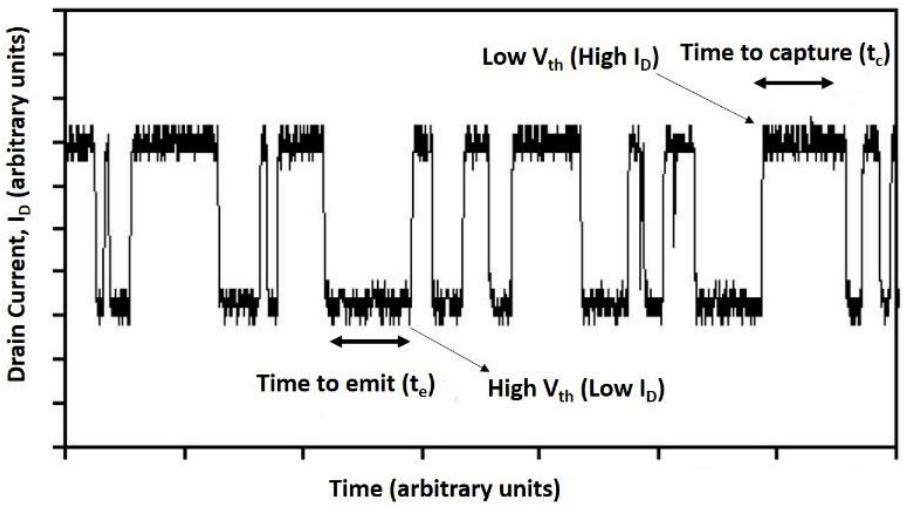


Figure 1-14: Time domain measurements of a stationary random telegraph noise (RTN). The low Vth state corresponds to the state where the trap is empty (electrically neutral) and high Vth corresponds to the state where the trap is occupied (electrically charged) [116]***.***

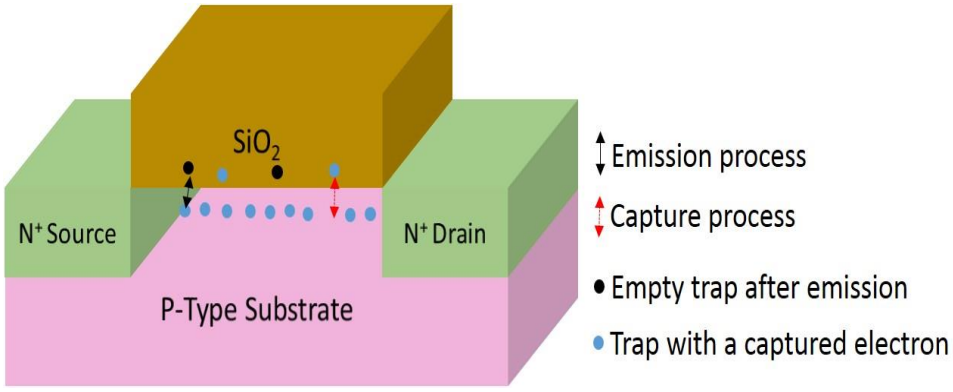


Figure 1-15: Trap-detrap of electrons at the Si-SiO_2 interface

1.5.2 Bias-Temperature Instability (BTI)

In ultra-scaled devices, BTI is considered as one of the most severe device degradation mechanisms. BTI happens when the gate of MOS structure is heavily biased while keeping the other contacts grounded and under these conditions the transistor parameters are known to degrade [117][118]. Based on the type of the MOSFET and the polarity of the voltage applied to the gate, BTI is classified into four categories: NBTI/pMOS, PBTI/pMOS, NBTI/nMOS and PBTI/nMOS. For p-channel MOSFETs, the degradation is called negative bias temperature instability (NBTI) and for n-channel MOSFETs, the degradation is called positive temperature instability (PBTI). Figure 1-16 shows the bias condition for the most discussed BTI combination: NBTI/pMOS (left) and PBTI/nMOS (right).

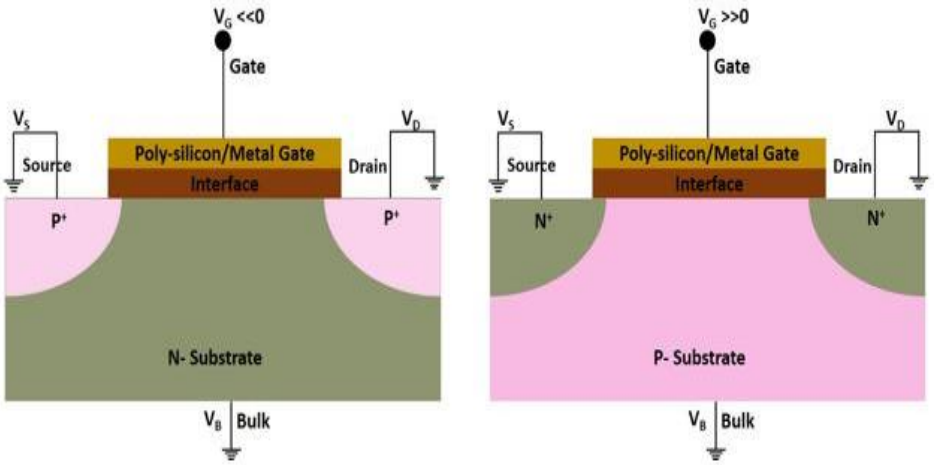


Figure 1-16: Bias condition for the two most important BTI configurations. In both cases the gate is biased to a voltage for strong inversion, (left) is for NBTI/pMOS and (right) is for PBTI/nMOS.

Bias Temperature Instability (BTI) is observed as an increase in the absolute values of threshold voltage (V_{th}) and off-current (I_{off}) and in the decrease of the absolute values of on-current (I_{on}) and transconductance (g_m). The mechanism is accelerated by temperature and voltage bias, regardless of the current flow [119]. It is widely believed that the physical phenomenon responsible for BTI is related to the generation and/or activation of states in the Si/SiO_2 interface and trapped charges in the oxide. BTI was first articulated in 1966 by Miura et al in which they linked the generation of charge due to the electrochemical reaction to the presence of strong electric field at the Si/SiO_2 interface [120]. However, since then BTI was largely forgotten until it became prominent as transistor scaling reached 60nm technology and below. In [117] the impact of BTI across nMOS and pMOS in several scenarios was studied and it was demonstrated that pMOS devices are more susceptible to BTI, regardless of the gate bias (positive or negative). As in digital circuits p-type MOSFETs are negatively biased, V_{th} degradation was only attributed to NBTI in pMOS and hence, BTI was often referred as only NBTI. However, with the advent of MOS structure using high-k/MG, PBTI/nMOS has started to gain prominence to a level similar to NBTI and have also accelerated the BTI degradation [121]. During the BTI stress, the oxide electric field (Eox) is nearly homogenous along the channel due to the presence of only the vertical electric field. Figure 1-17 shows the dependence of the threshold voltage on the oxide electric field (which is a function of applied voltage) for PMOS (left) and NMOS (right) devices for Intel's 45nm high-k/metal gate technology [122]. It can be seen from the figure that PBTI in NMOS is much worse than NBTI in PMOS with HKMG (high-k/metal gate) technology. Further evidence to this can be seen from Figure 1-18, where the BTI degradation is represented as the degradation of saturation-current as a function of stress time. We can see that the degradation of the saturation current follows a power law ($\Delta I_D \sim t^n$) and the value of n is lower in case of PBTI in NMOS compared to NBTI in PMOS.

One of the most interesting characteristics of BTI is its recovery behavior. The threshold voltage (V_{th}) increases when the device is under stress but when the stress is removed, the degradation relaxes [123], [124] and V_{th} partially recovers. An increasing opinion in the scientific community

is the presence of two components to BTI: a fast, universally recovering component and a slowly recovering or permanent component [117], [123], [125], [126] as shown in Figure 1-19.

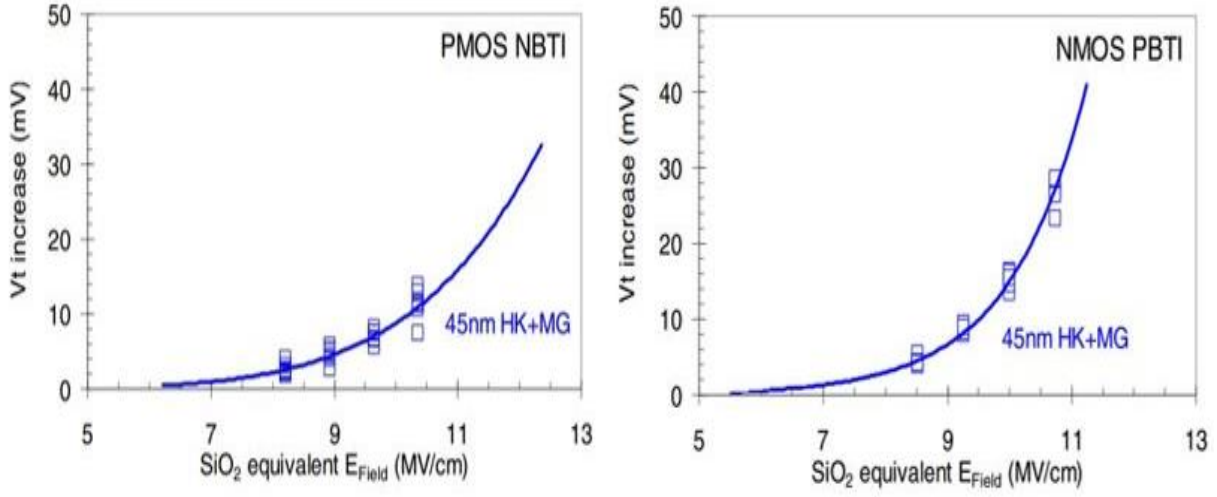


Figure 1-17: Dependence of Vth in response to the electric field due to (left) NBTI in PMOS and (right) PBTI in NMOS. The curves correspond to the 45nm HKMG technology [122]***.***

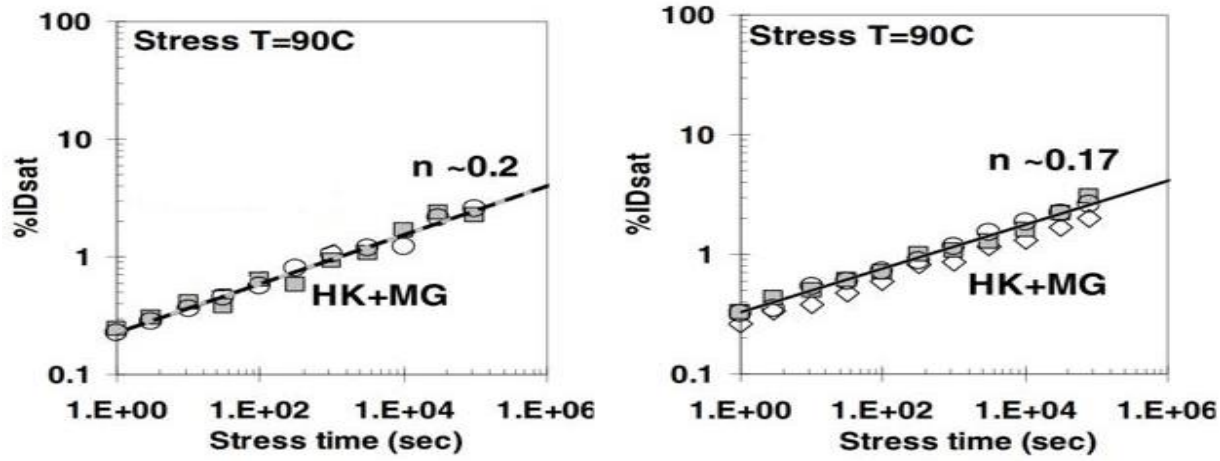


Figure 1-18: Time evolution of degradation in (left) PMOS NBTI and (right) NMOS PBTI [122]***.***

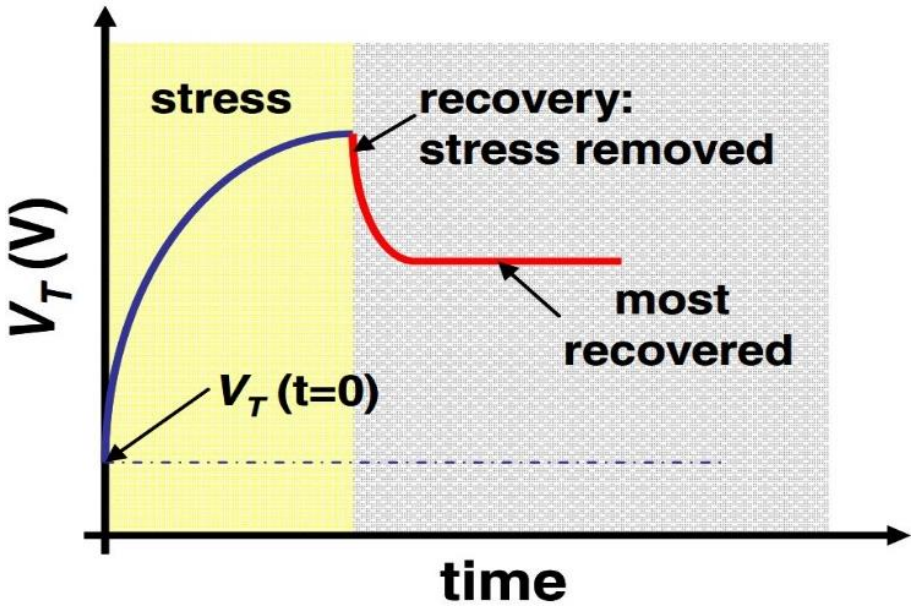


Figure 1-19: BTI as a function of time illustrating both degradation and recovery [122]**.**

We can see from the figure that the threshold voltage (V_T in the figure) degrades while the stress is being applied, however, once the stress is removed, a partial recovery is observed in the threshold voltage. The physical mechanisms to the origin of these two components (recovering and permanent) are still unclear. Some attribute interface states (P_b centers) to be responsible for the permanent component and oxide charges (E´centers) for the recoverable component [117]. While there are others of the opinion that interface traps are solely responsible for BTI [125], [127]. The partial recovery (after stress removal) of transistor characteristics is attributed to the de-trapping of charges from the trap centers. However, the fast NBTI relaxation makes the characterization of the failure mechanism extremely difficult warranting the need for special fast evaluation techniques [128]. In [129], a new ultra-fast technique was developed to accurately characterize the recovery of NBTI degradation. Numerous works have tried to explain the underlying physical phenomena for BTI indicating the contribution of defects, their number, spatial distribution and energies [104], [117], [121], [130]–[133]. However little or no consensus has been achieved in understanding the underlying physics and the research continues. Until recently, the reaction-diffusion (RD) theory was the most accepted theory to explain the experimental and theoretical aspects of NBTI. RD model is based on the assumption of the generation of traps at the Si/SiO_2 interface when the stress is applied and the subsequent annealing of these traps with the stress removal [134], [135]. However, the recovery process of NBTI cannot be explained by the diffusion-limited process suggested by RD model [134], [136], [137]. As suggested in [137], a two-stage model can be used to explain the entire process of NBTI (degradation and recovery) over a wide range of bias voltages and stress temperature. A detailed look of BTI degradation in small devices, shows that recovery proceeds in discrete steps whose properties are fully consistent with the charge trapping observed in the context of random telegraph

noise (RTN) [101] as shown in Figure 1-20. We can see from the figure that in small area devices, traces of RTN can be observed in the discrete recovery of BTI.

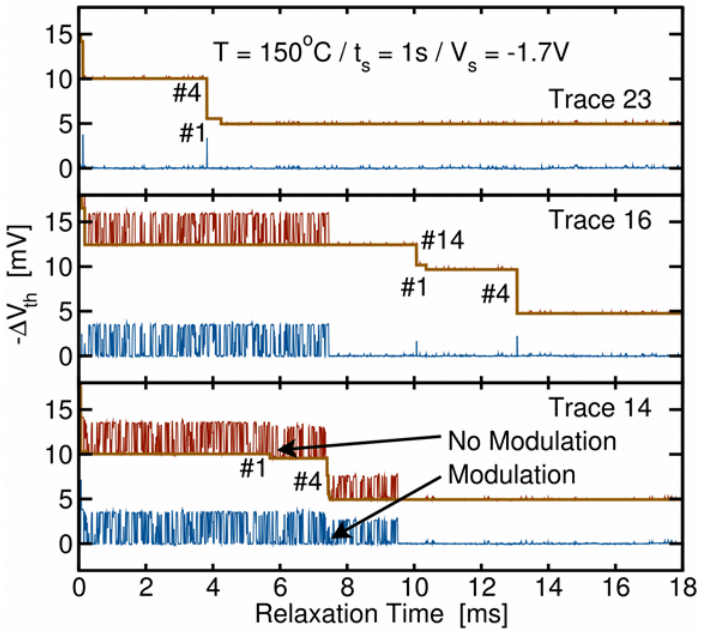


Figure 1-20: Traces of RTN can be found in the discrete recovery of BTI in small area devices. The traces can contain temporary RTN (tRTN), which disappears after a certain amount of time (see three selected traces above). The extracted step heights (brown) are subtracted from the raw data (red) to yield the noise trace (blue) [101].

Most effects of the bias stress (BTI) can be accounted for by the same sort of traps that cause RTN [101], [138]. The only difference is that bias stress measurements activate traps that require higher voltages and have longer time constants (t_c and t_e). BTI (specifically NBTI) can be considered as the non-equilibrium response of the defects, while RTN is a consequence of their quasi-stationary behavior [138].

1.5.3 Hot Carrier Injection (HCI)

MOSFET is essentially a voltage controlled resistor, made by the inversion layer induced by the gate voltage. When the gate to substrate voltage (V_G) is relatively high compared to the source to drain voltage (V_D), the resistivity along the channel is almost constant and the current (I_D) varies linearly between the source and the drain. If, however, V_G is comparable or lower than V_D, the inversion layer is much stronger at the source end than at the drain end (of course $V_D>V_S$). The voltage drop due to the channel current is concentrated on the drain side and carriers gain a significant amount of energy from this field. The electron (or hole) kinetic energy, which is written as a function of temperature, is much higher than the lattice temperature and hence, these energetic electrons (or holes) are referred as “hot-carriers (electrons or holes)”. The energetic carriers may lose their energy via impact ionization and contribute to substrate current. Some of these hot carriers can gain enough energy to break Si-H bonds at the interface; some hot-carriers can gain

enough energy to surmount the energy barrier at the Si/SiO_2 interface and can be injected into the oxide. Once injected, they can be trapped or create more defects in the bulk of the oxide. The trapped charges at the interface or (and) the insulator bulk cause time-dependent shifts in the measured device parameters, such as the threshold voltage (V_{th}), transconductance (g_m), linear ($I_{d,linear}$) and saturation ($I_{d,sat}$) drain current [139]. In time, substantial degradation can result in the partial or complete failure of the device. As holes are much "cooler" than electrons, hot-carrier effects in n-MOS transistors are more significant than in p-MOS transistors [140]. Figure 1-21 shows the basic process of hot carrier generation when appropriate voltages are applied to the gate and drain terminals.

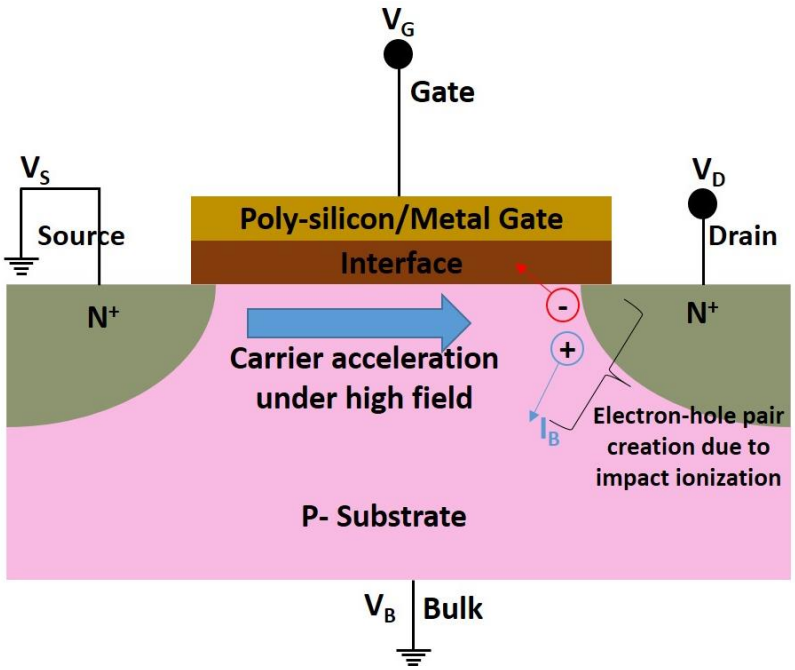


Figure 1-21: Basic process of hot-carrier generation in a MOSFET

Under HCI, a lateral electric field is also added by the application of a voltage between source and drain. To this extent, HCI can be related to BTI but more complex due to the additional lateral field. Hence, an in-depth knowledge of BTI is essential for understanding HCI. The worst HCI degradation for a MOSFET occurs at a condition that maximizes the bulk or substrate current (I_B). In long channel planar devices, the most degrading channel hot-carrier (CHC) condition is when the gate bias is 50% of the drain bias ($V_G = V_D/2$) [140], [141]. However, for short channel nMOS devices it was found that the most CHC degradation condition corresponds to the high vertical field ($V_G = V_D$) and this change is believed to be due to the loss of gate control over the channel [105]. In devices with high-k dielectrics, the shift may be attributed to the increased SiO_2 bulk defect contribution resulting from an increased equivalent oxide thickness (EOT) or due to the high-k bulk defect contribution in case of thin EOT [105], [142]. CHC stress can cause a significant degradation in the performance of strained MOSFETs [143]. In the case of Tri-gate/FinFET devices, hot-carrier degradation has varying trends in literature with some reporting an increased and some reporting decreased degradation for narrow fins [144]. This may be attributed to the

complex hot-carrier mechanisms for FinFETs/Tri-gate devices. However, the use of Tri-gate devices with proper design and process optimization provided significantly improved performance against HCI [145].

1.6 Interface traps and its role in device variability

Interface traps are a prominent variability source as they can be a static source of variability and a source of dynamic variability. Interface traps that exist after the fabrication process can be a static source of variability. Although the number of defects in deeply-downscaled devices is very few, their relative impact has become intolerable [101]. In atomistic devices, random dopant distribution (RDD) was identified as the main source of time zero variability in the form of RTN fluctuations in ultra-scaled devices [146]. Low doped channel device is used to reduce the variability, however, it was shown that in these devices, the adverse role of RDF is taken over by charged interface traps [68]. In addition to that, stress during the operation (BTI, HCI) induces interface state generation, causing time-dependent variability. Hence understanding the role of the interface traps as a source of variability could provide a deeper insight into the overall device variability.

Interface traps provide doping, enhance recombination and increase leakage through insulators. Hence, the presence of these defect centers, or traps, in semiconductor substrates may significantly influence the electrical characteristics of the device. The fundamental concept of interface states was put forward by Bardeen in 1947 [147]. Bardeen used the term ‘trap’ to indicate that the electronic states are electronic bound states whose bound electron wave functions are localized in all three space dimensions at these trapping centers [148]. Interface traps are also sometimes (actually in the initial years) referred as surface traps. However, the term interface trap is most appropriate as no practical surface is bare. To understand the origin of interface traps, let’s consider a silicon crystal along a crystallographic plane, as shown in Figure 1-22.

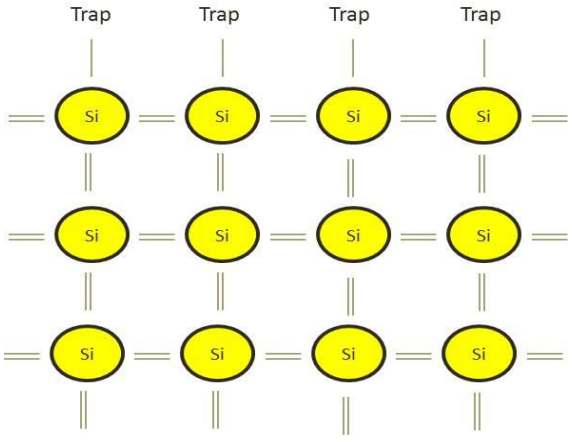
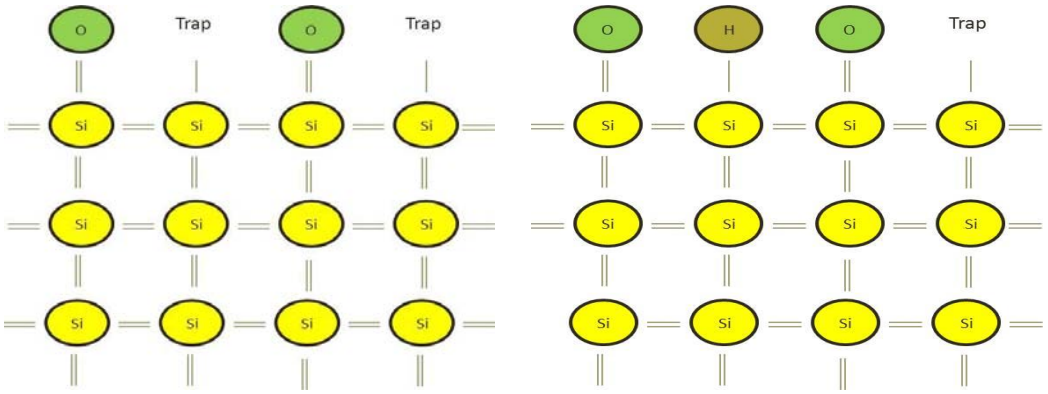


Figure 1-22: Silicon Surface.

We can see from the figure that the silicon atoms located at the surface are only bonded to the silicon atoms in the layer below and the forces experienced by the silicon atoms at the surface are

different to those experienced by the ones below. This perturbation of the surface layer of atoms produces potential wells at the semiconductor surface. The potential well contains quantum states, formerly in the valence band or conduction band, but now in the forbidden gap. These quantum states in the silicon band gap are called *interface states* or *interface traps.* When silicon surface is oxidized, all the valence electrons of the surface silicon atoms are satisfied by the silicon-oxygen bonds and the interface trap density drops to a low value as shown in Figure 1-23 (left). Each electrically active interface state leads to degradation of the important transistor parameters. Therefore, it is mandatory to increase the quality of the interface as much as possible and this is done by annealing the interface with hydrogen atoms. Therefore, the dangling silicon bonds are passivated by hydrogen forming Si-H bonds as shown in Figure 1-23 (right). With hydrogen passivation, the amount of electrically active interface states is reduced to an acceptable level. Ironically, these Si-H bonds are the cause for device degradation [104] and a source of variability.



Un-Passivated Si-SiO_2 surface

Passivated Si-SiO_2 with hydrogen atom forming Si-H bonds

Figure 1-23: Un-passivated (left) and passivated (right) Si-SiO_2 surface

There have been several works that have been carried out to study the various aspects of interface trap behavior. As early as in 1989, researchers have investigated in [149] the effect of interface traps and bulk traps in SiO_2 on hot-carrier-induced (HCI) degradation in MOSFET's. They found a strong correlation between device degradation and density of generated interface traps, indicating that the generated interface traps most significantly cause device degradation. In [150] an interface trap-assisted tunneling and thermionic emission model are used to study the increased drain leakage current in off-state n-MOSFET's after HCI stress. Traps at Si/high-k interface have degraded the performance of devices with the high-K dielectric material [151], [152]. Modern devices such as Tunnel FET's also suffer from performance degradation due to interface traps [153]. More recently, researchers have studied the impact of interface traps in tunneling FET's in comparison with modern MOSFET's [154]. They found that although V_{th} shifts and subthreshold swing (SS) degradation induced by interface traps have similar trends, the impact on I_{on} is different because of conduction mechanism.

Owing to the important role of interface traps on the performance of devices, appropriate characterization of interface states has always been of utmost importance. Charge pumping method is known to be one of the most powerful methods of qualitatively characterizing the interface state

properties of MOSFETs [155]–[157] Due to the considerable influence of interface traps on the performance and reliability of nanoscale devices, quantitative analysis of the interface states (trap density, spatial and energy distribution) is becoming more important. For example, the leakage current in DRAM cells is strongly dependent on the local electric field at the location of the traps [158]. This makes it extremely important to characterize the spatial distribution of interface traps and to simulate the localized electric field in these devices. In [159], a novel approach is presented to capture the impact of non-uniform energy distributions of interface traps in the calculations of surface potential (ψ_s). In the work presented in this thesis, we make use of TCAD simulations to study and analyze the contribution of the interface traps (attending to the number and spatial distribution of traps) towards the overall variability of the device performance. Since, we use TCAD simulations to carry out this study, next chapter is dedicated to discuss the fundamentals related to TCAD simulations and how they can be used to perform variability analysis.

CHAPTER II

TCAD Fundamentals and application in variability studies

As it was highlighted in the end of the last chapter, the primary focus of this thesis is studying the interface traps related variability in MOS structures by means of TCAD tools. Hence having a better understanding of the TCAD concepts is of paramount importance. The first part of this chapter discusses the basic concepts of TCAD simulations, its advantages and limitations. In the later part of this chapter, we discuss how TCAD tools are used to perform studies associated with variability in modern devices.

2.1 Semiconductor Device Modeling and Simulation

From the previous chapter, we have seen that there has been a continuous and an exponential growth of the semiconductor industry that is now worth billions of dollars. This steady growth has put an immense pressure on the industry to deliver products to the market in a short time. This has led to an intense development in the field of numerical simulation of the semiconductor process and device structures. In the absence of simulations, engineers are forced to rely on a large number of experimental data, impacting the development costs and time to market [2]. Simulation is the imitation of some real thing and the act of simulating something generally demands the representation of certain key characteristics or behavior of a selected system. Numerical analysis is based on fundamental, physical partial differential equations that represent a semiconductor device. The field of using simulations to study, model and analyze semiconductor process and

device operation is often referred as Technology Computer Aided Design (TCAD). TCAD tool is an electronic design automation tool that models the semiconductor device operation and fabrication based on fundamental physics through the aid of computer simulations. The key features of a simulator include the acquisition of valid source information about the device (or circuit being simulated), selection of key characteristics and behaviors, use of simplified approximations and assumptions within the simulation and finally validating the simulation outcome. Simulations provide the physical insight to optimize the process and maximize the yield. However, the accuracy of the simulation depends on the device models used in the simulators.

According to International Technology Roadmap for Semiconductors, development costs are reduced by about 40 percent by the efficient use of TCAD tools [160]. TCAD simulations have evolved from being traditionally used for research and development (to understand underlying physical concepts that are otherwise impossible to measure) to be increasingly used in manufacturing for advanced process control and yield improvement. Each technology generation has driven the device density higher, thus increasing the complexity of fabrication. This has fueled the development of simulation tools that can efficiently point in the right direction for technology during the development phase. TCAD tools are also being used more efficiently for exploring new device architectures and optimizing process flows. In the microelectronics industry, TCAD is widely referred to as *process CAD* and *device CAD*. Process CAD refers to the simulation of semiconductor device processing (fabrication) steps. Device CAD refers to the numerical simulation of the device operation to evaluate the electrical characteristics. These concepts will be addressed in the later sections of this chapter. In the next section, a historical overview of the different TCAD tools is presented.

Historical Overview of TCAD

During the last few decades, TCAD has evolved from being a one-dimensional (1D) simulator to full two-dimensional (2D) and three-dimensional (3D) simulation of today's complex circuits and devices. This evolution is the result of contributions from a large number of researchers over the years. Gummel reported the results of 1D simulations of BJT [161], where iterative methods to sequentially solve partial differential equations (PDE) in a drift-diffusion system were used. This work of Gummel was further developed and applied to p-n junctions by Mari [162] and to Read diode by Gummel and Scharfetter [163]. The first 2D numerical solution of Poisson's equation for MOSFET structure was reported simultaneously in the works of Loeb and Schroeder [164], [165]. The first finite element analysis of the semiconductor equations was reported in [166] and helped the development of more general purpose TCAD tools based on the reported method. In the early 1980's, as the device sizes became smaller and designs became more complex, 2D numerical simulations became increasingly inadequate leading to the developments of 3D numerical simulators [167]. Following that, the ability to perform 3D simulations were incorporated into almost every available 2D simulator programs [168], [169]. After the success of CAD tools to simulate the electrical characteristics, there was a growing interest in the research community for tools that could simulate the fabrication process. In the late 1970's researchers from Stanford University developed the tool SUPREM that was used as 1D process simulator [170]. The works of Professor Dutton and Professor Plummer led to the foundation of TMA which was the first supplier of commercial TCAD software. In 1998, TMA was acquired by Avant, which was later in 2001 acquired by Synopsys [171]. Currently, Silvaco [172] and Synopsys are the major

providers of commercial TCAD tools. In this thesis, we use both Synopsys TCAD (only in the initial period) and Silvaco TCAD.

2.2 Process and Device simulations

In the microelectronics industry, TCAD is widely referred to as *process simulator (process CAD)* and *device simulator (device CAD)*. In order to simulate the performance of a device structure, it is essential to represent the key characteristics (or behavior) of the device (and/or fabrication process) by means of physics based models. Since TCAD simulators use different physical models to represent device behavior (and/or fabrication process) they are often referred as physically-based simulators.

Process simulation

The first step in performing a simulation is the creation of the device structure and this is achieved by the use of a process simulation tool. Process simulation or process CAD refers to the front-end process modeling that includes the numerical simulation of the physical effects of IC processing steps used to fabricate transistors (or other devices) up to metallization. It is used to simulate the semiconductor processing steps like oxidation and diffusion, ion implantation, deposition, annealing and etching. Process CAD also generates the input file needed for device simulations and the scheme is shown in Figure 2-1.

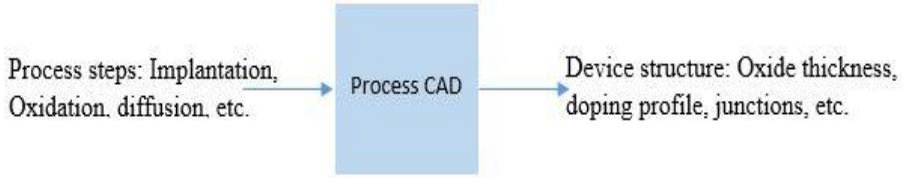


Figure 2-1: Scheme of the process simulation using process CAD to generate the input file (device structure) for device CAD.

Synopsys TCAD package includes Sentaurus process for 2D and 3D process simulations. Silvaco TCAD tools include ATHENA for 2D process simulation and Victory process for 2D/3D process simulation. Athena provides general capabilities for numerical, physically-based, 1D/2D simulation of process steps. Victory process is a general-purpose 1D/2D/3D process simulator with automatic switching between different modes (1D/2D/3D). In order to use process simulator, one needs to provide all the process steps involved. If the process steps are not available, then the device structure can be created through computer-aided design and process emulation steps. However, even for this procedure, it is essential to supply certain fundamental characteristics of the device (material properties, oxide thickness, doping etc.,). For example, Synopsys TCAD also includes Sentaurus Structure Editor (SSE), which is a 2D/3D device structure editor and a 3D process emulator. SSE can be used to create structures for which the fabrication process is unavailable. In process emulation mode, Sentaurus Structure Editor translates processing steps into geometrical operations (different device regions are created by drawing geometrical operations such as, drawing a rectangle or a polygon). Using Sentaurus Structure Editor, device structures can be created interactively using a graphical user interface (GUI) or in batch mode using Scheme scripting language. Silvaco TCAD provides the tool ATLAS to create the device structure in a

batch mode (command line input) and the tool DEVEDIT can be used to edit the device interactively. The created structure can be analyzed using the visualization tool that is included in the package.

Device simulation

Device simulation or device CAD refers to the numerical simulation of the device operation. Device CAD includes a set of physical device models that describes the carrier transport in materials. Physical device models account for the physics of the device operation and provide greater insight but usually require a lengthy analysis. The scheme of device CAD is shown in Figure 2-2.

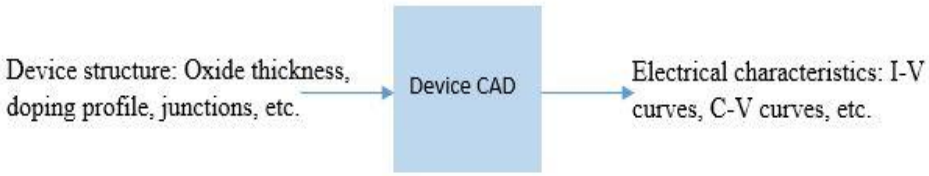


Figure 2-2: Scheme of device simulation using device CAD to generate electrical characteristics.

Synopsys TCAD tools include Sentaurus device for 2D and 3D device simulations. Sentaurus device is a numeric semiconductor device simulator that is capable of simulating the electrical, thermal and optical characteristics of various semiconductor devices. Silvaco TCAD tools include ATLAS for 2D/3D device simulations and Victory Device for 3D device simulations. Physically based device simulators predict the electrical characteristics associated with the specified structure and bias conditions. To accurately analyze an arbitrary semiconductor structure that is intended as a self-contained device under various operating conditions, a mathematical model has to be given. As we will see, these mathematical models are described onto 1D/2D/3D grid (mesh) of the structure.

2.3 Device Simulators: Physical Background

Since in this work devices were created using CAD procedures instead of process simulators, only the workings (physics) of device simulator are discussed in detail. As outlined earlier, physically-based device simulators predict the electrical characteristics that are associated with the specified structure and bias conditions. This is achieved by approximating the device operation described by a mathematical model onto a two or three-dimensional grid also referred as device mesh. The mesh consists of a number of mesh points known as nodes (N). The dimensions of the mesh of the simulated structure have a great influence on the simulated results (outputs) and these are discussed in section 2.4. Years of research have resulted in mathematical models represented by a set of differential equations. These mathematical models consist of a set of fundamental equations linking the electrostatic potential and the carrier densities within some simulation domain. These equations consisting of Poisson's equation, continuity equations and transport equations are discretized on the simulation grid (2D or 3D) to simulate the carrier transport through a structure.

Poisson's equation relates the variations in electrostatic potential to local charge densities. The continuity and transport equations describe the way the carrier densities (electrons and holes) evolve as a result of transport processes, generation and recombination processes. Due to the importance of these equations in any physically-based simulators, they are briefly discussed below.

Poisson's Equation

The electric field is defined as the negative gradient of the electrostatic potential, $\Psi(x)$, or in one dimension as the negative derivative of the electrostatic potential.

$$\frac{d\Psi(x)}{dx} = -E(x) \qquad 2\text{-}1$$

From Gauss's law, we can draw the relationship between the charge density and the electric field.

$$\frac{dE(x)}{dx} = \frac{\rho(x)}{\epsilon} \qquad 2\text{-}2$$

Substituting the equation 2-2 into 2-1, we get the relationship between the charge density and electrostatic potential and is given as

$$\frac{d^2\Psi(x)}{dx^2} = -\frac{\rho(x)}{\epsilon} \qquad 2\text{-}3$$

Equation 2-3 is referred as Poisson's equation, it relates the variations in electrostatic potential to local charge densities. Device physics has established the existence of three different mechanisms, which add to the space charge term in Poisson's equation, in addition to the ionized donor and acceptor impurities. These are interface fixed charge, interface trapped charge and bulk trapped charge. Interface fixed charge are always completely occupied. Interface traps (similarly bulk traps) can be categorized as donor traps and acceptor traps. A donor trap can be either positive or neutral like a donor dopant. A donor trap is positively charged (ionized) when empty and neutral when filled (with an electron). An empty donor trap which is positive, can capture an electron or emit a hole. A filled donor trap which is neutral, can emit an electron or capture a hole. An acceptor trap can be either negative or neutral like an acceptor dopant. It is neutral when empty and negatively charged (ionized) when filled (with an electron). A filled acceptor trap can emit an electron or capture a hole. An empty acceptor trap can capture an electron or emit a hole. Hence the total charge caused by the presence of traps is subtracted from the right-hand side of Poisson's equations (see eq.2-3) and is given as:

$$Q_T = q(N_{tD}^{+} - N_{tA}^{-}) \qquad 2\text{-}4$$

Where N_{tD}^{+} and N_{tA}^{-} are the densities of ionized (charged) donor-like and acceptor-like traps respectively and are described by eq. 2-5 and eq. 2-6.

$$N_{tD}^{+} = N_D * F_{tD} \quad (2\text{-}5)$$

$$N_{tA}^{+} = N_D * F_{tA} \quad (2\text{-}6)$$

The ionized trap density depends on the total trap density and probability of ionization. In eq. 2-5 and eq. 2-6, F_{tD}, F_{tA} are the ionization probability for donor-like and acceptor-like traps, N_D is the total density of traps. The probability of ionization of acceptor-like or donor-like traps is a function of different trap characteristics such as, capture cross-section and trap energy and is given by the eq. 2-7, where σ_n and σ_p are electron and hole trap cross-sections respectively, and E_t is the trap energy.

$$F_{tA}\ or\ F_{tD} = f(\sigma_n, \sigma_p, E_t) \quad (2\text{-}7)$$

The energy of the trap (E_t) is located in the forbidden gap of the material in which they (traps) exist. In the case of interface traps, one of the materials that form the interface can be considered as a reference from which the energy of the trap is measured. For donor type trap, the trap energy is measured from the valence band while for acceptor type trap, it is measured from the conduction band. The trap cross-section (σ_n/σ_p) is the effective physical size of the trap in the material.

Carrier Continuity Equations

The continuity equation relates the changes in the carrier density over time with the carrier fluxes and generation and recombination of carriers. The continuity equation is based on the conservation of mobile charges [173] and is:

$$\frac{\partial n}{\partial t} = \frac{1}{q}\frac{\partial J_n}{\partial x} - R_n + G_n \quad (2\text{-}8)$$

$$\frac{\partial p}{\partial t} = -\frac{1}{q}\frac{\partial J_n}{\partial x} - R_p + G_p \quad (2\text{-}9)$$

Here the equation 2-8 and equation 2-9 correspond to the case of electrons (n) and holes (p) respectively. Where G_n and G_p are the electron and hole generation rates, R_n and R_p are the electron

and hole recombination rates, J_n and J_p are the electron and hole current density, q is the magnitude of the charge on an electron. Recombination is a process by which an electron occupies an empty state associated with holes resulting in the loss of the carrier. At the end of a recombination process, energy is released and depending on how the energy is released we have different recombination processes. For radiative recombination, the energy is released in the form of a photon; in a non-radiative recombination, the energy is transferred to one or more phonons; the energy is transferred as a kinetic energy to another electron for Auger recombination [173]. In band-to-band recombination, an electron moves from conduction band to the valence band associated with a hole. In trap-assisted recombination, electrons first fall into trap energy in the band gap and then complete the recombination process by falling from the band gap to the valence band of a hole. This two-step recombination process assisted by a trap is known as Shockley-Read-Hall (SRH) recombination. The process of recombination, when reversed, causes generation process instead of recombination and a single expression can be used to describe both processes. However, carrier generation due to light absorption does not involve recombination and this process is known as ionization.

Carrier transport

Current inside a semiconductor (or any material) is defined as the rate of flow of charge carriers. The two main means of carrier flow inside a MOSFET structure are *drift current* and *diffusion current*. When an electric field is applied to a semiconductor, the electrostatic force causes the carriers to accelerate due to the field. The flow of charge carriers under an applied electric field is known as *carrier drift*, resulting in *drift current*. Due to collisions with impurities, the velocity of the carriers reaches a constant value, known as saturation velocity (*v*). The other means of current known as diffusion current is the result of movement of carriers due to thermal energy. This movement (or diffusion) of carriers happens due to a density gradient, created as a result of variation in doping density. The total current is the sum of the drift and the diffusion currents and equations 2-10 & 2-11 show the total current density (as the sum of drift and diffusion components) for electrons and holes.

$$J_n = qn\mu_n E + qD_n \frac{dn}{dx} \qquad 2\text{-}10$$

$$J_p = qp\mu_p E - qD_p \frac{dp}{dx} \qquad 2\text{-}11$$

In the above expressions, D_n and D_p are the diffusion constants, the terms μ_n and μ_p correspond to the electron and hole mobility respectively. Mobility is an important property of carriers and is deterministic in the performance of different devices, therefore, a better understanding of factors that influence carrier mobility is very crucial for device engineers. Carriers are accelerated by the applied electric field but they lose the momentum as a result of various scattering processes. These scattering processes include lattice vibrations, other carriers, surfaces and material imperfections. Since the effects of all these microscopic phenomena are lumped into the macroscopic mobilities introduced by the transport equations, they (mobilities) depend on the local electric field, lattice temperature etc.

Mobility modeling is consequently an important aspect of any device simulator and the device simulators must represent the mobility and its degradation as accurately as possible. In the simplest case, the mobility is a function of the lattice temperature, however, this is only valid for devices under a low electric field or with a low level of doping. Under such conditions, the mobility is dependent on the phonon scattering and impurity scattering. However, with high doping densities the mobility degradation due to impurity scattering is very high and models that describe mobility as a function of doping (apart from lattice temperature) should be considered. The carrier mobility is further degraded when operating at high electric fields. The mean drift velocity of carriers is saturated and the influence of impurity scattering on energetic carriers is reduced. Hence, when simulating devices with high electric fields, proper mobility models that take into account the velocity saturation should be considered. One of the most important parts of simulating a MOS structure involves the appropriate modeling of carrier mobility in the inversion layer. Mobility modeling in inversion layers introduces additional complications as the carriers in inversion layers are subjected to surface scattering, extreme carrier-carrier scattering, and quantum mechanical size quantization effects. So, when performing a simulation of MOSFET structure these effects should be taken into account to have more accurate results. Synopsys and Silvaco provide a wide range of mobility models and the choice of the appropriate model depends on the device being simulated and its bias (and temperature) conditions.

Quantum simulations

As device features are approaching nanoscale dimensions, the wave nature of electrons and holes can no longer be neglected. Hence, the physical models that describe their behavior inside device simulators must incorporate the necessary quantum effects. When the quantum nature of the electrons (holes) is neglected, carrier transport can be described by Boltzmann's transport equation. There are several approaches that can be used by modern physics-based simulators to incorporate quantum effects. One common method is coupling a Schrödinger-Poisson solver to the Boltzmann's transport equation. Here the carrier concentration is calculated quantum-mechanically and used in Poisson's equation to obtain the electrostatic potential, which is again used in the Schrödinger's equation until convergence is reached. The resulting quantum-mechanical carrier concentration is used to derive correction factors for the solution of Boltzmann's transport equation. The other method is to model the confinements of carriers associated with local potential using density gradients. The density-gradient model, based on the Wigner function allows a local representation of quantization effects. Thus is more suitable in simulators than the Schrödinger-Poisson solver (which depends on non-local quantities such as dielectric thickness). It has been reported that while the density-gradient method correctly models the carrier concentration in the inversion layer of a MOSFET, it fails to accurately reproduce the tunneling current [174]. Silvaco and Synopsys provide a suite of models with varying capabilities to incorporate quantum effects inside the device being simulated. The use of appropriate models depends on the purpose and conditions of the simulation. It has to be noted that a careful approach has to be employed in choosing the quantum models, as simulation time and convergence complexity is greatly influenced by the quantum models.

2.4 Numerical solution methods and Device mesh

TCAD simulators are based on Finite Element Method (FEM) which uses iterative method for solving. In an iterative method, we start with a guess for the solution and then successively renew this guess, getting closer to the solution at each stage. This iteration is usually performed until convergence is reached with the desired accuracy. The supremacy of most iterative methods lies in their ability to achieve this convergence efficiently. However, any particular iterative method is confronted by two entangled issues, speed and convergence. For example, to achieve convergence a large number of iterations are needed which adversely affects the speed of the solver. Conversely, a high-speed solver may not achieve a convergence. For semiconductor devices, different solution methods are used depending on the situation and it is also possible to use several different numerical methods (in conjunction) to obtain a solution. Hence, it is imperative for the numerical solvers included in TCAD simulators to find an appropriate balance between speed and convergence. There are three different numerical solution techniques commonly used for obtaining solutions for semiconductor devices and they are (i) GUMMEL (de-coupled) (ii) NEWTON (fully-coupled) and (iii) BLOCK. The GUMMEL method solves for each unknown in an equation while keeping the other variables constant and repeats until a stable solution is achieved. Hence, it is also referred as a de-coupled method. In NEWTON method, the total system of unknowns is solved together. In BLOCK method, the solutions are obtained by combining the fully-coupled and de-coupled techniques. Each method has its own merits and demerits and the choice of the method depends on the requirement of the simulations. For example, Gummel iteration converges relatively slowly but often tolerates poor initial guesses. However, NEWTON iteration will normally converge quickly so long as the initial guess is close to the final solution. Several strategies such as reducing the bias step in the event of non-convergence, loosen the requirement of a good initial guess. There are different parameters that are available for users by which one can modify the solution techniques. Much efforts have been put into developing reliable criteria for convergence and the default parameters work well for nearly all situations and most users will never need to change them.

As outlined earlier, TCAD simulations are based on finite element analysis of semiconductor equations that describe the device process and device operations. In physically based simulators, the real semiconductor device is represented in the simulator as ‘virtual’ device whose physical properties are approximated onto a two or three-dimensional non-uniform grid (called mesh), consisting of a number of grid points (N) called nodes. The layout of the mesh in a simulation is a very important aspect in determining the accuracy of the simulations. The problem of optimal mesh structure versus computational time and accuracy is always a dilemma in the world of TCAD simulations. For simulation accuracy, the number of grid points (nodes, N) allocated in the simulated structure must be maximized. However, for the computational efficiency, the number of nodes (N) must be minimized. Although optimal mesh generation is not exactly deterministic, one can always follow certain guidelines and heuristics for defining satisfactory meshes [175]. Denser meshes are allocated in regions where there is a rapid change in the values of potential, doping, etc., and the coarse mesh is applied in a region where these (doping, potential, etc.,) quantities do not change rapidly. Apart from the optimization techniques, certain consideration such as minimizing the number of obtuse triangles must be observed. It is often a good practice to ensure that the simulations results are not very sensitive to the device mesh being considered. In other words, the simulated structure must be robust so that the device performance (simulated) is independent of the grid density. To show this point as an example, Figure 2-3: top left shows the

simulated structure with a very fine grid spacing in the semiconductor under the gate insulator. It can be seen that the fine mesh also extends into the deeper regions of the substrate. In order to reduce the number of grids and thus reduce the simulation time, re-meshing of the structure is done. Figure 2-3: top right shows the same structure with a reduced number of grid points. Finer mesh is used in the gate insulator and the interface under the gate while a coarser mesh is considered for the other regions. The re-meshing reduced the total number of grid points by 40% and improved (faster) the simulation time by 40%. Figure 2-3: bottom compares the results (I_D-V_G) of the simulations before (Figure 2-3: top left) and after (Figure 2-3: top right) the re-meshing. The curve (a) corresponds to the mesh structure in Figure 2-3: top left and the curve (b) corresponds to the mesh structure in Figure 2-3: top right. It can be seen that the re-meshing has not changed the results but has improved the simulation time significantly. This ensures that the chosen mesh is robust to have device performance independent of the grid density.

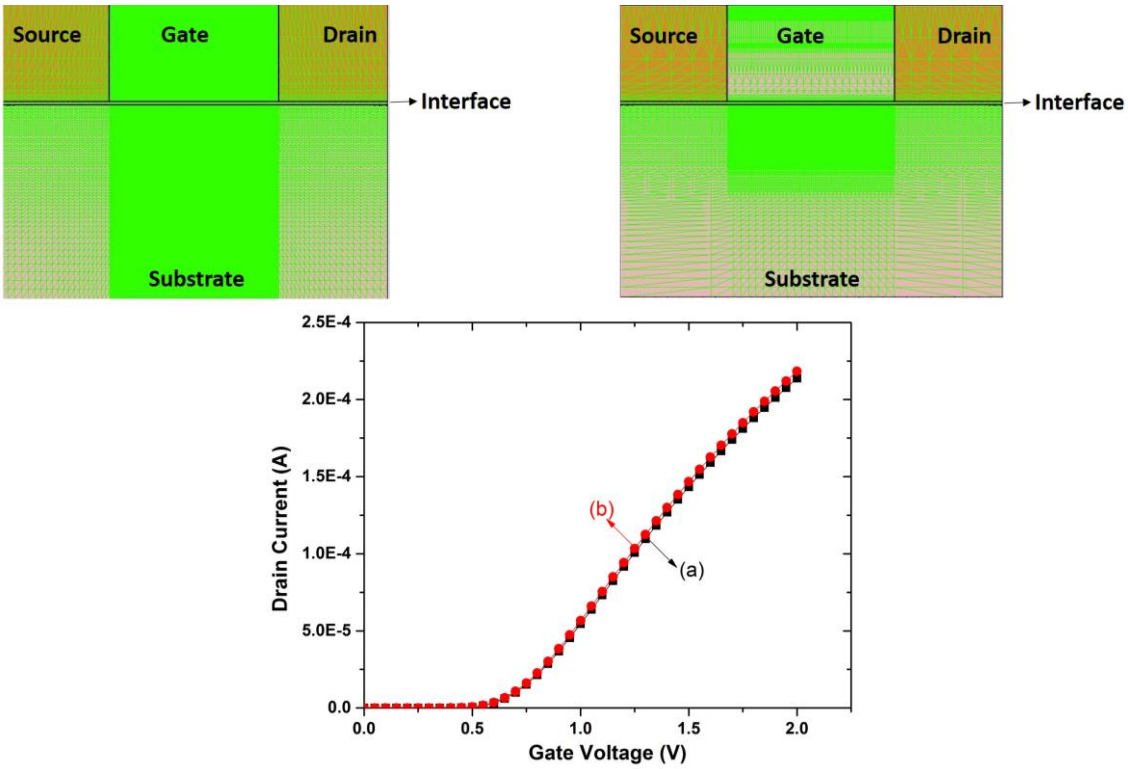


Figure 2-3: The simulated MOSFET Structure on the top left corresponds to a very fine mesh in the channel and also deep in the substrate. The structure on the top right corresponds to an optimized meshing of the structure with a fine mesh only in the channel, this resulted in reduced number of mesh points. In the plots (below), the curve (a) (squares, black) corresponds to the device with mesh shown at top left and curve (b) (circles, red) corresponds to the device with mesh shown at top right. The plots show that the re-meshing of the structure has improved the simulation efficiency without modifying the results.

Once the device mesh is defined, the different physical models that describe the mechanisms that happen inside the device need to be included in the simulator. The simulated structure can be a 2D or a 3D (real) structure based on the requirement of the simulation. 3D simulations take longer time than 2D but are useful in analyzing certain critical physical phenomena (corner effects, width

effects, etc.). In most cases the results from 2D simulations would suffice and do not warrant the need 3D simulations. It has to be noted that in general, a 2D simulation is, in fact, a 3D simulation of a device with a width of 1µm, with a very coarse mesh (in fact only two grid points at 0µm and 1µm) along the width. 3D simulations involve structure with a definite width and a definite number of grid points along the width. The typical simulation flow that is employed by device simulators is outlined in the following section.

2.5 Typical Simulation flow

In order to evaluate device characteristics, a simulator follows a typical flow which is shown in Figure 2-4. It shows different steps involved in the simulation of a device structure, from its creation to obtaining the electrical characteristics.

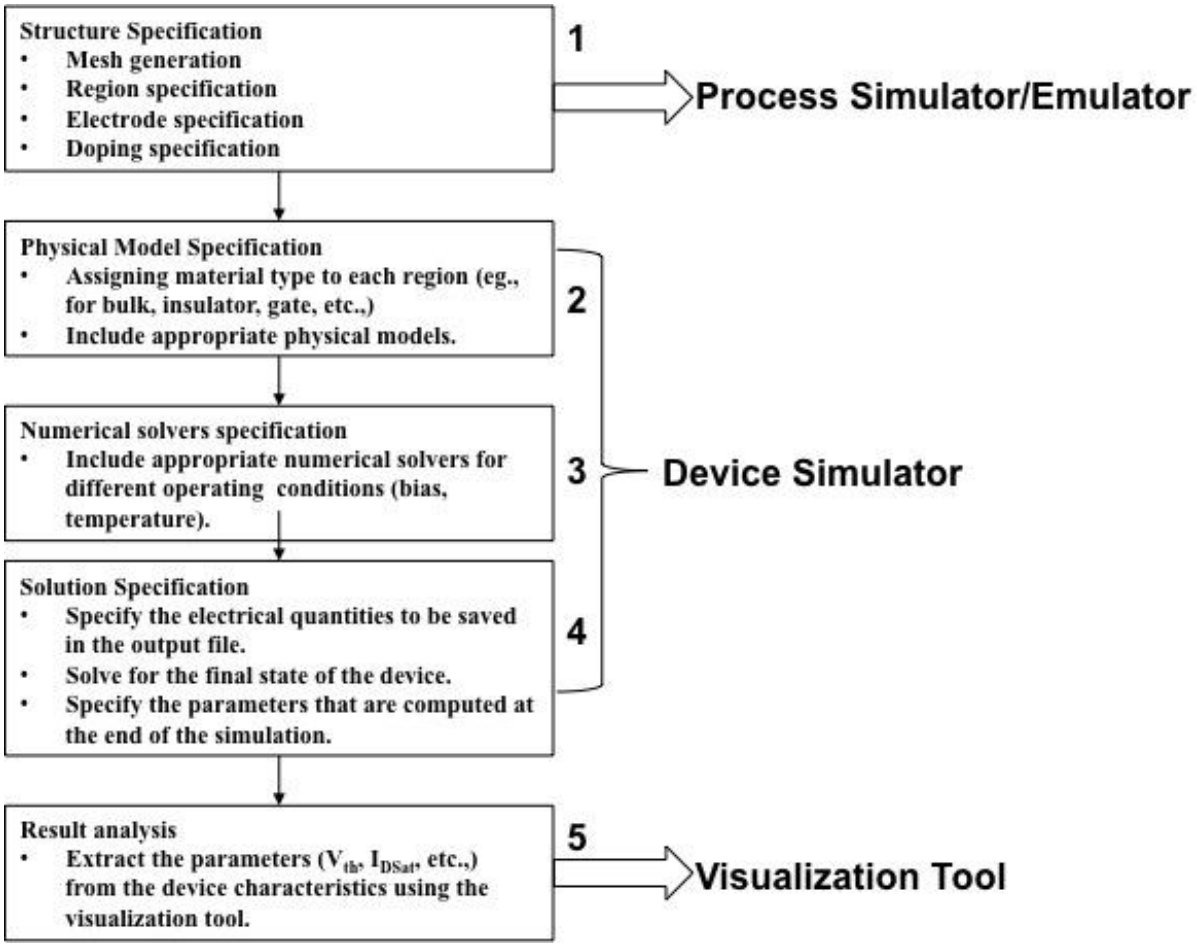


Figure 2-4: Simulation flow that is typically employed for a TCAD simulation.

As shown by the first block, the typical design flow starts with the creation of a device structure either by a process simulator tool or through computer-aided design (CAD) operation and process emulation steps. In order to create the structure, the geometrical information such as device dimensions should be provided. Along with the device geometry, device specific information such

as doping, electrode definition and mesh definition must also be provided. While specifying the device geometry, a defined width can be specified to create a 3D structure other than the 2D structure with default width. The constructed device is intelligently meshed and passed on to the device simulator to perform device simulations.

The entire process of performing device simulation is outlined by the second, third and fourth blocks in Figure 2-4. Inside device simulators, material type is assigned to different regions and different appropriate physical models are included. The inclusion of different physical models depends on the nature and intent of the simulations. They can range from simple carrier transport models such as drift-diffusion models to more complex quantum models. Device simulators numerically simulate the electrical behavior of a single semiconductor device or several devices combined in a circuit. Hence, appropriate numerical solvers are also specified within the device simulator (3rd block). Terminal currents, voltages and charges are computed based on the set of physical equations that are included in the different physical models and at the operating conditions (temperature and bias), these are represented by the 4th block in Figure 2-4. Finally, as shown in the 5th block, different graphing tools are used to plot and visualize the electrical characteristics of the simulated device (circuit). The visualization tools can also be used to visualize various other electrical quantities such as current density, the electric field across different regions etc. The magnitude and distribution of these electrical quantities are usually stored by default at the final biasing condition of the device. However, the simulator can also be forced to store them at intermediate or at every bias step. In the following sections, the simulation procedure used by the two different TCAD packages (Synopsys and Silvaco) is outlined.

Simulation using Synopsys TCAD

For the creation of the device structure through process simulation, Synopsys TCAD package offers *Sentaurus Process* or *Taurus TSUPREME-4* tool. To create device through computer aided design operations, *Sentaurus Structure Editor (SSE)* is used. Once the device structure is created, meshing is done before the device can be numerically solved to determine the electrical properties. In Sentaurus TCAD meshing can be done using three different meshing tools: *Sentaurus Mesh*, *Noffset3D*, and *Mesh*. All these tools generate high-quality spatial discretization using a variety of mesh generation algorithms. The methods and the choice of the tool to be used depends on the geometry of the device. *Noffset3D* is usually used to mesh 3D structures and *Sentaurus Mesh* is more suitable for 2D structures. *Sentaurus device* is used to simulate the electrical characteristics of the device. *Sentaurus device* offers both the 2D and 3D device simulations to be performed on the structure. *Tecplot SV* is used to visualize the output from the simulation in 2D and 3D, and *Inspect* tool is used to plot the electrical characteristics.

Simulations using Silvaco TCAD

Similar to Sentaurus TCAD, the first step in the simulation with Silvaco TCAD is the device creation. However, unlike Sentaurus TCAD, in Silvaco TCAD, the mesh is required to be defined before the structure is created. The mesh definition can be done within the tool used to create the device structure. Like Synopsys, the device structure can be created using the process simulator tool within the Silvaco TCAD package. Silvaco TCAD provides the tool *ATHENA* for process simulation. *ATHENA* is a framework program that integrates several smaller programs into a more

complex process simulation tool. Silvaco also provides the tool *ATLAS* that can be used to create the device using computer aided design (CAD) and process emulation. The tool *ATLAS* is also used for performing device simulations to obtain the electrical characteristics of the test (simulated) structure. *ATLAS* is very often used in conjunction with the process simulator *ATHENA* taking advantage of the automatic interface between them. *ATHENA* predicts the physical structure that results from the processing steps and the resulting physical structure is used as input by *ATLAS*, which then predicts the electrical characteristics for a specified bias and temperature. Therefore, it is possible to determine the effect of process parameters on device characteristics by the combination of *ATLAS* and *ATHENA*. However, it is much difficult to control the actual device performance in *ATHENA* process simulator environment in comparison to *ATLAS* device simulator environment. While a slight change in the individual parameters can affect the entire structure created using *ATHENA*, the device structure, material and doping concentration can be controlled precisely when creating a device through *ATLAS*. A device constructed through *ATHENA* is more difficult to characterize, however, the device is much closer to a real fabricated transistor. The tool *Tonyplot* is used to visualize the 2D/3D simulated structure and its electrical characteristics.

2.6 Limitations and challenges of TCAD simulations

As described in the previous sections, to perform any physically based simulation, we need to provide the following information

- Physical structure to be simulated.
- The physical models to be used.
- The bias conditions for which electrical characteristics are to be simulated.

Whether a device is created using a process simulator or with CAD operations, the creation procedures need to be calibrated to define a structure that is as close to the real device as possible. If CAD operations are used, then the doping profile of the virtual (simulated) device must match that of the actual device. If process simulators are used to create the structure, then in addition to the doping profiles, the process coefficients of the appropriate models that describe different process steps (oxidation, diffusion, etc.) must be well calibrated. It is a tedious task to tune a large number of undefined coefficients present in the empirical-based models in the process simulator. Moreover, this tuning procedure has to be repeated if any changes were made to the process recipe.

There is always skepticism about the ability of TCAD simulations to accurately predict the experimental results. This is largely due to the fact that device simulations are based on mathematical models derived from Maxwell's equations as outlined in the earlier sections. These equations are valid for a number of applications, especially for silicon devices. However, conditions do exist where their validity is not guaranteed (or at least in doubt). For example, when device dimensions are in nanoscale and quantum effects become dominant. The more sophisticated results in physics are too complex to give rigorous and sufficiently simple models for the purpose of device simulations. Aggressive research is being carried out to address this challenge and to develop newer and more reliable models [176]. The main limitation of these simulators is that all the relevant physics must be incorporated into the simulator and as we are venturing (by continuous scaling) into unfamiliar territories, the accuracy of available models is always contentious.

2.7 Variability studies using TCAD

We have seen from the section 1.3 that variability related issues are one of the roadblocks in the path of MOSFET evolution and a cause of concern in future technology devices involving conventional and novel CMOS devices such as FinFETs [177][178], nanowire MOSFETs [179] to name a few. While significant statistical variability effects like threshold voltage [49][180] and ON-current variations [181], leakage increase [182], random telegraph noise [183] are well understood and factored in transistor design, the atomic scale effects that contribute to this variability are still being carefully studied. The statistical variations in deca-nanometer devices have led to a shift in the paradigm of numerical simulations as it is no longer sufficient to simulate a single device to represent one macroscopic design. The aim of numerical simulation shifts from predicting the characteristics of a single device toward estimating the mean values and the variance of device parameters such as threshold voltage, drive current, etc., for a whole group of microscopically different devices in the system [184].

There has been an increasing interest in using TCAD simulations to analyze the source of statistical variations in nano-devices [57], [185]–[187]. Simulation of variability has become extremely important in terms of understanding the current variability mechanisms and sources. Simulations are also used for predicting the levels of variability in modern devices [188], [189]. TCAD based transistor design can be used at the early stage of technology development to introduce accurate statistical variability that can be of great help to designers. As outlined in earlier sections, most important sources of statistical variability (SV) are Random Discrete Dopants or Random Dopant fluctuations (RDD or RDF), Line Edge Roughness (LER)/Line Width Roughness (LWR) and Metal Gate Granularity (MGG). TCAD simulations incorporating these variability sources will help to gain reliable estimates for the magnitude of intrinsic parameter fluctuations. Among different groups working in developing variability aware simulations, the group of Asenov from the University of Glasgow have made significant strides. To show some examples, Figure 2-5 (left) identifies the impact of the various SV sources on the threshold voltage of fresh devices. Figure 2-5 (right) shows the impact of SV sources (combined and separate) on the on-current I_{on} from 1000 devices.

The RDD was generated from the continuous doping profile by placing dopant atoms on the silicon lattice sites in the S/D regions and in the channel. LER was introduced through 1D Fourier synthesis following the method described in [65] where random gate edges are generated from a power spectrum corresponding to a Gaussian autocorrelation function. The Poly-Gate Granularity (PGG) is simulated by randomly generating poly-grains for the whole gate region [55]. Since these simulations are ‘atomistic’, quantum confinement effects are needed to be considered by including density gradient (DG). From these pictures, we can see that RDF is inducing the largest dispersion in both V_{th} and I_{on} followed by MGG and LER. The plots shown in Figure 2-5 were obtained from the variability simulations performed using in-house TCAD simulator known as “GARAND”, a 3D ‘atomistic’ simulator, which is now commercially available from Gold Standard Simulations (GSS) Ltd [190]. As of May 2016, GSS has been acquired by Synopsys and now is a part of Synopsys TCAD package [171].

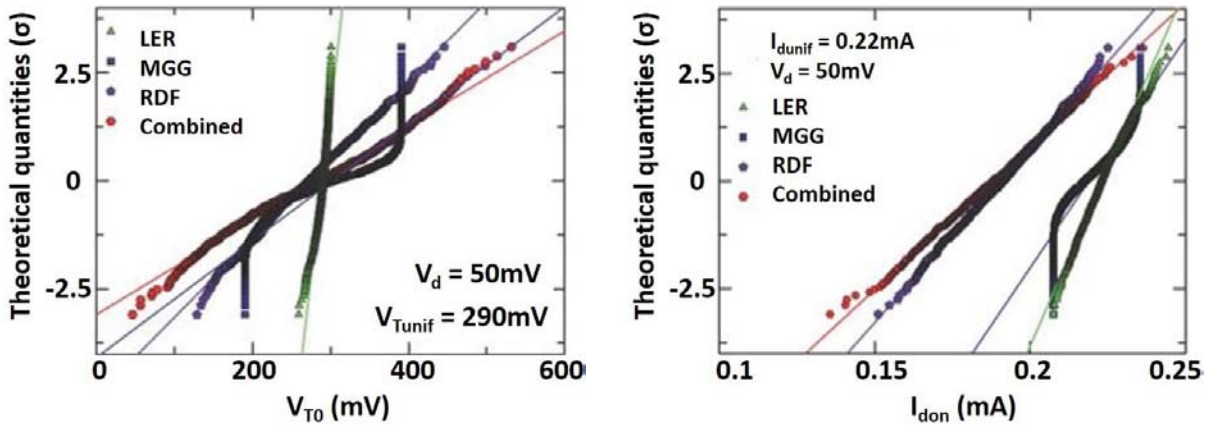


Figure 2-5: Variation in the threshold voltage (left) and on-current (right) from 1000 devices due the impact of combined and separate SV sources [191]***.***

Interface trap simulations using TCAD

On top of the SV sources that were discussed above, traps (in the oxide and/or interface) can capture carriers during the device lifetime and can be a variability source (discussed in chapter 1). These traps can be pre-existing traps [192] or generated over the course of device operation [127]. Due to the vital importance of interface traps in the performance degradation, the ability to simulate and analyze the impact of interface traps would be a great asset in predicting the device reliability.

Now, there are different approaches that are used to simulate the influence of interface traps. Trapped charge at the interface can be simulated by modifying the doping concentration at the nearest nodes which in turn introduces electron charge in the solution of Poisson equation [193]. In another approach, the interface traps are simulated by adding a sheet charge directly at the interface which adds space charge directly to the Poisson equation [194]. In this work the approach of adding a space charge to the Poisson's equation (described earlier) is used as it was much efficient (simulation time and accuracy of the results) than the former.

TCAD tools are being increasingly used to analyze the influence of interface traps with varying objectives in MOSFET and advanced structures. For example, studies using TCAD simulations have helped in understanding the influence of interface traps density on the device characteristics (I-V) of different III-V MOSFET devices [149]. TCAD tools have been extremely useful in understanding the role of interface traps in device degradation. Experimental studies accompanied by TCAD simulations were used to understand the generation and recovery of interface traps in MOSFETs subjected to different electrical stresses [133]. TCAD tools have also been useful in studying degradation in advanced structures such as Tunnel FET, bend gate structure and silicon nanowire MOSFETs [67], [195], [196].

As outlined in the earlier chapter, interface traps are an important variability source and with the scaling of devices, its prominence has only increased. In nano-scale structures interface traps can cause fluctuations in the threshold voltage due to variation in the number of interface traps and the random position of the traps [154]. Therefore, it is increasingly relevant to investigate the variability in the transistor performance due to the variations associated with interface traps. However, among different works [49], [57], [61], [72], [75], [180], [181], [183]–[185], [189]

studying the variability in devices by means of TCAD simulations, one can hardly find works that investigate the device variability associated with interface traps. Moreover, it has to be emphasized that the 'atomistic' simulation studies typically used (as those in GARAND) to analyze the impact of statistical variability is computationally very demanding and requires the use of cluster computing. In this work, we use commercially available TCAD tools to study the variability in device performance perpetuated by interface traps (attending to their number and spatial distribution) in MOSFETs. As the simulations done in this thesis were not atomistic in nature, employing cluster to perform the simulation were not necessary and the results from the simulations were helpful in understanding the role of interface traps in the device variability.

CHAPTER III

V_{th} variability perpetuated by interface trap distribution: 2D simulations.

As it was stated earlier, in small-area devices BTI induced V_{th} degradation and recovery in nanometer transistor proceed in discrete steps [132], [197], [198] due to augmentation of single-charge effects in scaled device. Therefore, we can no longer consider interface traps as being continuously distributed but discrete in space. Hence, we perform simulations by considering traps to be distributed discretely along the interface of the device. Our aim is to understand the variability in the transistor performance (threshold voltage) perpetuated by the number and the spatial distribution of interface traps. In this chapter we present the adopted methodology and the results related to the 2D simulations where traps are located along the 1D interface.

3.1 Device Structure and Simulation Methodology

As outlined in the earlier chapter, the prime objective of this thesis is to study the influence of the number and spatial distribution of interface traps on the variability of the device performance of a MOS structure. In order to make the analysis, we need to first create a structure that would be used for electrical simulations. In this section, we discuss creating the device structure to be simulated and the methodology of simulation.

Device Structure

For the analysis, nMOS structures were simulated as the convergence at low voltages needed fewer iterations owing to the larger mobility of electrons in comparison to holes (in pMOS). However, the study can be also applied to pMOS structures. Figure 3-1 shows the schematic of the MOSFET structure that is created and analyzed using ATLAS tool within the Silvaco TCAD package.

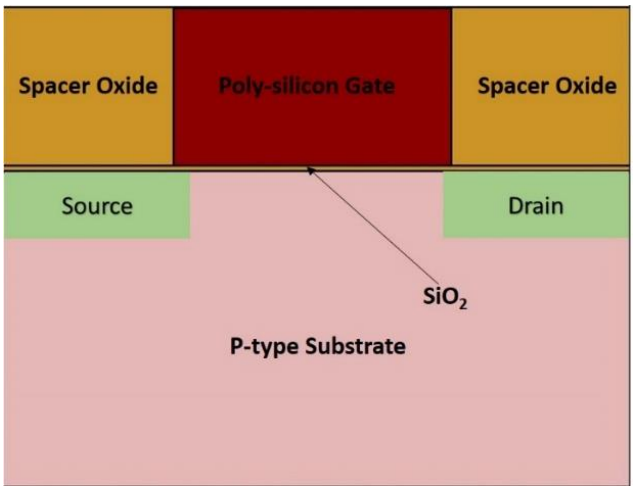


Figure 3-1: Schematic of the device structure to be simulated in 2D. Different regions of the device can also be identified.

Here, different regions (and the material used) of the device structure can be identified. Due to the absence of process recipe, the device was created using the process emulator within the ATLAS tool of Silvaco. We first analyze the case of an nMOS structure with a gate length of 250nm (long channel) and then extend the study to shorter channel lengths. The simulated long channel device has following attributes (material and doping), that were obtained from an example template within ATLAS.

- Oxide material was SiO_2.
- Oxide thickness was 2nm.
- Source and Drain doping values were 1e20 cm^{-3} with an n-type (Arsenic) impurity
- The doping profile for Source and Drain regions was a Gaussian profile with the peak value being 1e20 cm^{-3}.
- The doping profile used in the substrate was a constant profile of p-type (Boron) dopant with a doping concentration of 2e18 cm^{-3}.

Figure 3-2 shows the cross-section of the device with the 2D doping profile. The cut of the doping profile along the channel direction is also overlaid onto the device structure. From the figure we can see that the effective channel length (L_{eff}) is measured from this cut and for reference is marked in the figure. The effective channel length is the length of the channel where the doping is almost constant. For the simulated structure, L_{eff} is measured to be about 225nm. The origin of the x-axis in Figure 3-2 coincides with the left edge of the device. This device is electrically inactive and device (electrical) simulations are required to produce the electrical characteristics. The meshing used for the simulated structure is optimized (after several iterations) to obtain the desired accuracy

with the efficient use of the resources. The mesh density under the gate was chosen such that the distance between each node along the channel direction was 2.5nm.

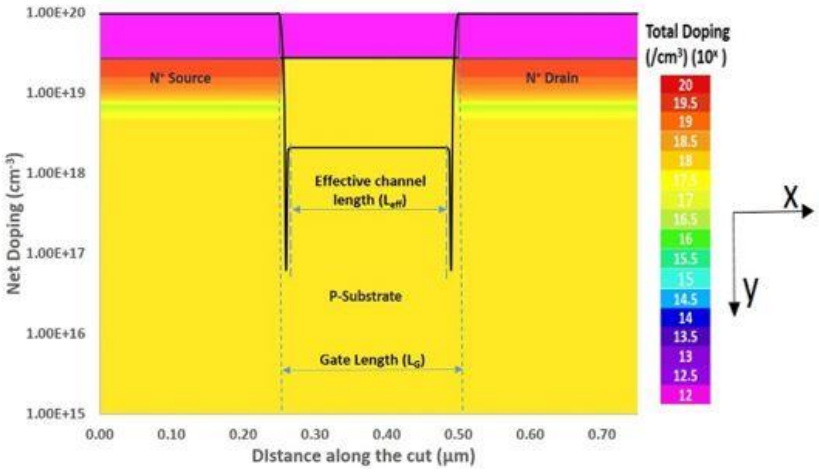


Figure 3-2: 2D cross-section of the simulated structure with doping profile. Cut of the doping profile along the channel direction is also overlaid onto the device structure.

Simulation Methodology

We intend to study the impact of the spatial distribution of interface traps on the device performance by monitoring the variations in device performance parameters (V_{th} and/or I_{on}). In order to achieve this objective, we needed to simulate a MOS structure with (charged) interface traps at discrete locations and analyze the impact of the trap location on device performance. To place the traps at discrete locations along the interface, the interface under the gate of the structure is divided into several discrete regions (also referred within this work as trap locations). This division is schematically shown in Figure 3-3. In the figure, we can see the schematic of a 2D nMOS structure (with a default width of 1µm) where the interface division is done only along the channel length. It can be seen from the figure that the division is only done along the effective channel length (L_{eff}) and not the actual gate length (L_G). The division was made such that each region encompasses only a single interface node (simulation node), this ensured the smallest possible division. As a first step, we perform device simulation without any interface traps and evaluate the performance parameter (V_{th} and/or I_{on}). This case (without traps or simply referred as 'pristine') is considered as a reference and interface trap effects are measured by monitoring the variations in the performance parameters from this reference case.

Since here, we are interested in studying the influence of trap location along the channel interface, we maintain the total interface charge (equivalently the number of interface traps) in the device always constant. For this, first we consider a predefined number of traps (charges) that are homogenously distributed along the entire interface and produce a shift in the performance parameters (as compared with the pristine case). Next the influence of trap location is then studied by equally distributing these traps at discrete trap locations (single or multiple) along the channel interface. So the number of traps (charges) in each interface trap location is scaled with the number of trap locations that are considered to be populated with traps. The influence of the spatial

distribution of trap location(s) on the variability of the device performance is analyzed by performing a set of simulations wherein each simulation the location of the trap region(s) is randomly chosen. The simulation schemes used for 2D and 3D simulations are explained in depth in the following sections. It has to be noted that unless stated otherwise, all simulations in the presented thesis were performed at T=25°C.

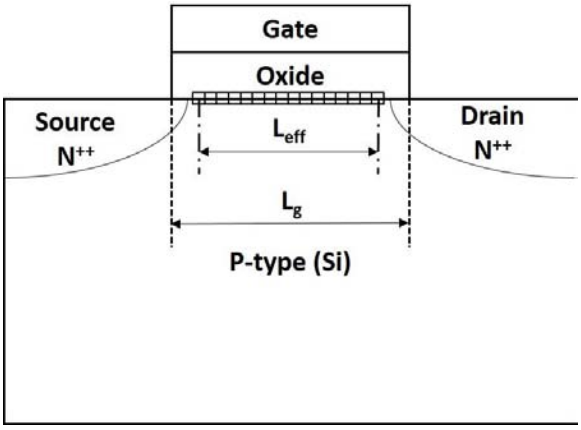


Figure 3-3: Schematic of the device structure to be simulated with the interface divisions.

Simulation scheme: 2D simulations

We begin the analysis by considering the simple case of 2D simulations. Here we analyze the influence of the spatial distribution of the trap locations on the variability of device performance using 2D simulations, where devices have a default width of 1µm. Figure 3-4 shows the schematic of the interface division (top view) employed.

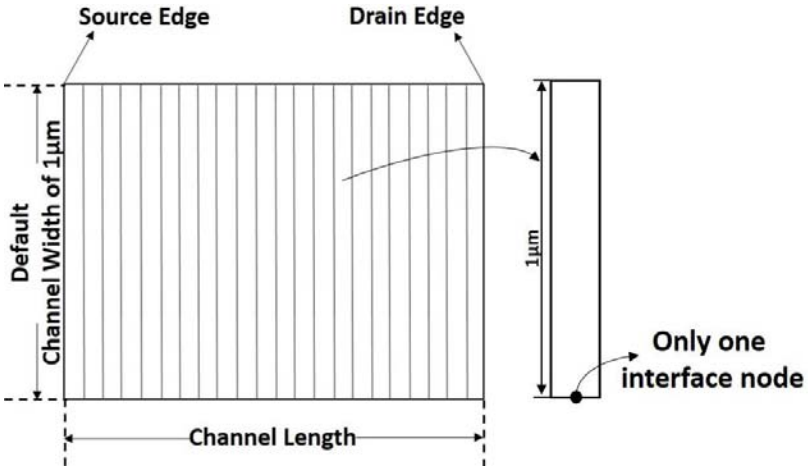


Figure 3-4: Cross-section of the interface of the simulated 2D structure. It can be seen that since a 2D structure has a default width of 1µm, each interface division along the x-axis (along the channel length) is also extended over the entire width of 1µm along the z-axis (along the channel width). However, each trap location encompassed only one interface node.

It can be seen that the interface division is done only along the channel length, which implies that each interface division (trap location) also has a width of 1µm. However, it is ensured that only a single interface node is encompassed by each trap location along the channel length. Hence, when traps are placed in each of this trap location (region) they are in fact spread across the entire default width. Figure 3-5 schematically shows the simulation scheme employed for 2D simulations.

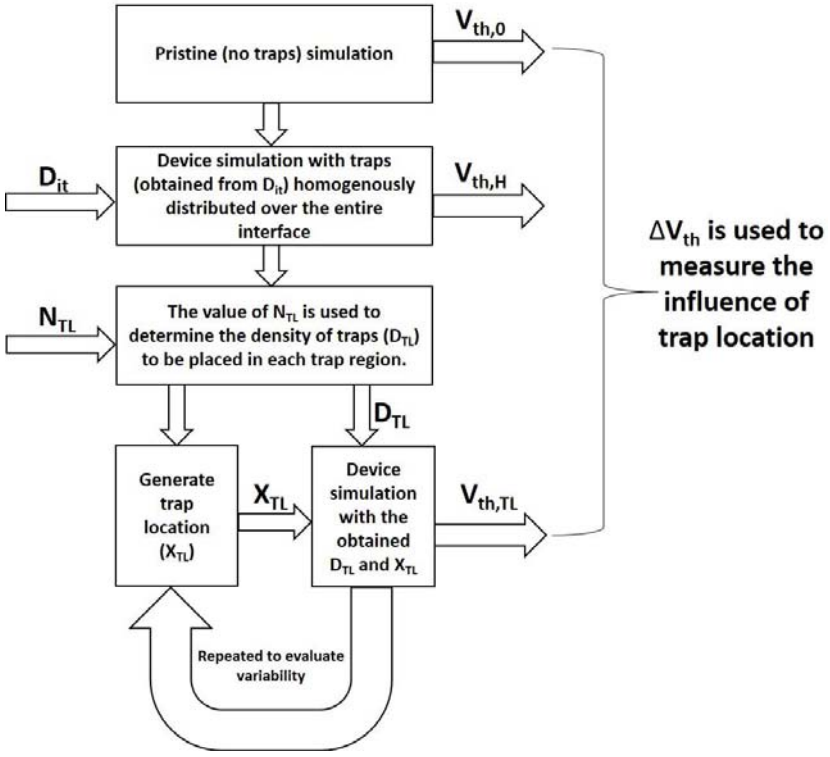


Figure 3-5: Simulation Scheme that is employed for 2D simulations.

We first simulate the device structure with no interface traps (pristine case) and then extract the threshold voltage ($V_{th,0}$) and this is represented by the first block in Figure 3-5. Then, the device is simulated with a homogenous distribution (i.e. all regions are populated) of interface traps of density 'D_{it}' over the entire interface (represented by the second block) and the threshold voltage ($V_{th,H}$) is extracted. To study the influence of the spatial distribution of the traps along the interface, traps are placed at discrete trap location(s). Note that the case of homogenous trap distributions can be treated as a case where all the trap locations are simultaneously populated with traps of equal density (D_{it}). After this, the homogenous distribution is discretized by placing traps at discrete trap locations. Since we are interested in the impact of the spatial distribution, it has to be emphasized that when placing traps at discrete location(s), we intend to keep the total number of interface traps (or equivalently, charge) in the device constant and same as that in the case of homogeneous distribution. Hence, depending on the number of trap location(s) (referred as 'N_{TL}') that is assumed to be populated with traps, the density of traps (referred as 'D_{TL}') at each of these location(s) is given by equation 3-1.

$$D_{TL} = \frac{D_{it} * N_{Tot}}{N_{TL}} \qquad 3\text{-}1$$

Here N_{Tot} is the total possible interface trap locations across the entire interface. Recall, that we divided the interface (along the effective channel) into different trap regions such that only one interface node is encompassed by each region. Hence, the total possible interface trap locations (N_{Tot}) is same as the total number of interface nodes in the effective channel length (L_{eff}). So in the case when traps are homogenously distributed along the entire interface, then the number of interface trap locations that are populated with traps (N_{TL}) is same as the total number of possible trap locations (N_{Tot}) and this would result in D_{TL} same as D_{it}. When a single trap location (simplest case) is considered the entire trap population (total number of traps) is concentrated in a single trap region along the channel interface and is reflected by the value of D_{TL}.

As done in the earlier cases (pristine and homogenous distribution) the threshold voltage ($V_{th,TL}$) is extracted at the end of the simulation. Since the interface trap spatial distribution can be varying across each device (leading to the observed variability in the device performance) we randomly generate the location(s) of the trap region(s) (fourth block left in Figure 3-5). This random trap location (X_{TL}) is a one-dimensional entity as in 2D simulations there are no variations along the device width. Depending on the values of N_{TL}, X_{TL} can get only one value (single trap location) or many (multiple trap location) values. The influence of trap location is studied by performing interface trap simulations with D_{TL} (fourth block, right) and extracting the threshold voltage ($V_{th,TL}$) at the end of each simulation and evaluating ΔV_{th} ($V_{th,TL} - V_{th,0}$). Device variability is then evaluated by repeating the simulations with different values (generated randomly or non-randomly) of X_{TL} i.e. different location of the traps. It has to be noted that in order to understand the obtained results better, some intermediate simulations were done with the specific (not random) location for the traps along the interface.

Simulation scheme: 3D simulations

To consider the spatial distribution of trap locations along the device width, it is essential to perform 3D simulations with a defined width (other than the default width of 1µm in 2D). Figure 3-6 shows the schematic of the interface division for a simulated 3D structure. In the similar manner as shown in Figure 3-3 the entire 2D interface (along the length and the width) was divided into different regions. However, for 3D simulations, the mesh is also specified along the width of the device. This meant that each interface division also has a defined width as opposed to the default 1µm used in 2D simulations (see Figure 3-4). It can be seen from the Figure 3-6 that as opposed to 2D simulations, the ability to also have a mesh along the channel width has facilitated the division of interface also along the width (z-axis). Using the similar approach as it was done earlier, the 2D interface is divided into many divisions and each interface trap location encompasses only a single interface node along the length and also along the width. The simulations scheme employed for 3D simulations is schematically shown in Figure 3-7 and follows the same procedure as 2D. It can be seen from the Figure 3-7 that in 3D simulations, the output of the generator is a 2D entity that corresponds to the x-coordinate (along the length) and the z-coordinate (along width) of the trap location. As stated earlier, in order to understand the obtained

results better, some intermediate simulations were also done with the specific (not random) location for the traps along the interface.

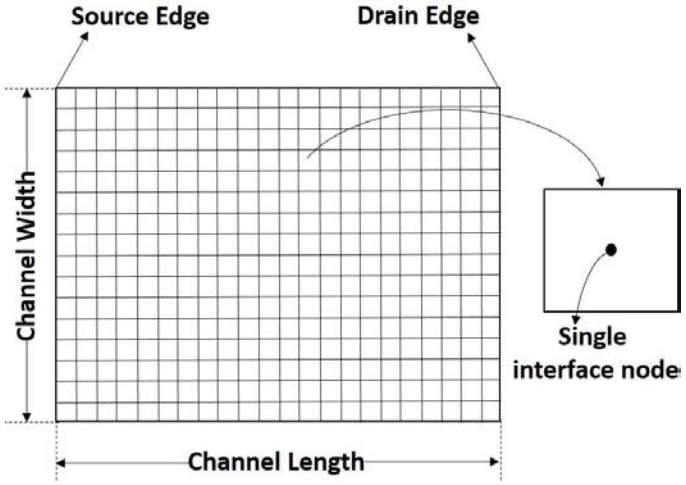


Figure 3-6: Cross-section of the interface of the simulated 3D structure. It can be seen each interface division encompasses only a single node along the length and along the width.

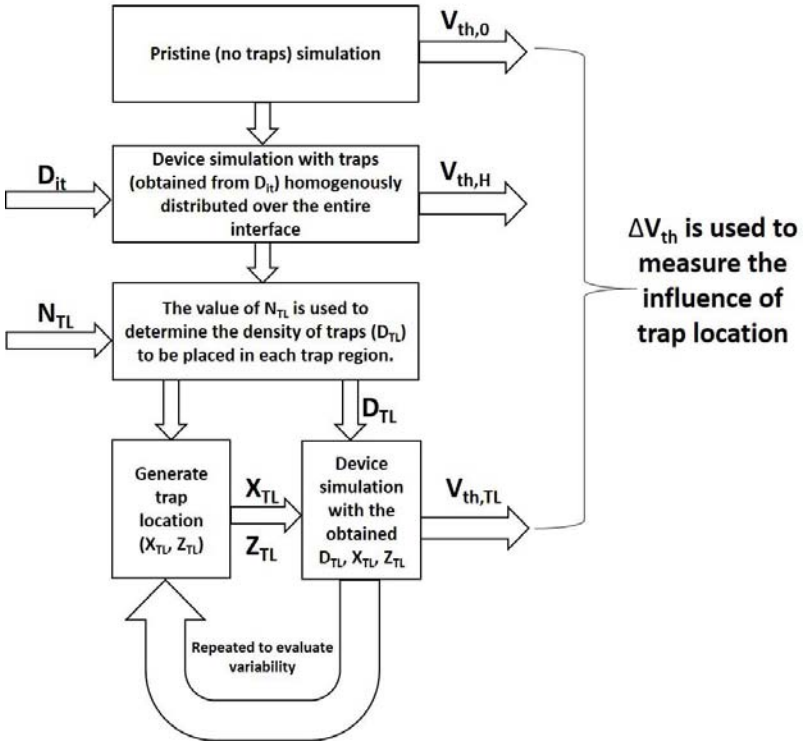


Figure 3-7: Simulation Scheme for 3D simulations.

In the beginning, Sentaurus TCAD was used to implement the above-stated strategy. The version of Sentaurus TCAD that was used was A-2008.09. However, it was not possible to implement this simulation strategy using Sentaurus TCAD. Sentaurus TCAD did not offer the possibility of dividing the interface into multiple regions using Sentaurus device editor. However, this limitation may be related to the tool or due to the lack of information in the user manual. Maybe these limitations are improved in the latest version of Sentaurus TCAD. However, the strategy was implemented using Silvaco with a great ease. The implementation was achieved as Silvaco allowed the placement of traps in the device by the use of spatial coordinates. Hence, it was decided to use Silvaco to evaluate the influence of the interface traps distribution along the channel on the variability of the device performance.

In order to evaluate the device characteristics and extract the performance parameters, it is essential to perform the electrical simulations of the structure that was created earlier (Figure 3-1). The ATLAS tool within the Silvaco package is used to perform device simulations and it is important to choose appropriate physical models to describe different physical phenomena. Hence, models that describe the carrier generation and recombination, carrier mobility, band-gap narrowing, etc., were included in device simulations. The drift-diffusion (DD) model is used to simulate the carrier transport inside the device. One of the most important decisions of TCAD simulations is to consider models that properly simulate the carrier mobility. Models should consider the carrier acceleration under the applied field and at the same time should consider the loss of momentum due to scattering processes and inside the inversion layer. In our simulations, we use the Lombardi CVT model, which takes into account the mobility degradation due to various scattering effects.

The variability in the device performance is analyzed by simulating several scenarios and they are explained in the following sections. In the rest of this chapter, 2D simulations are considered and the performance is analyzed by evaluating the V_{th}.

3.2 Electrical characteristics of devices without interface traps and with homogenous distribution of interface traps

As outlined in Figure 3-5, as a first step we perform device simulation of the MOS structure (see Figure 3-2) without any interface trap. As stated, in this chapter we perform the study by considering only 2D simulations, the case of 3D simulations is discussed in chapter 4. Using the physical models that were outlined in the above, we perform the device simulations to extract the pristine characteristics of the MOSFET. We first simulate the characteristics of the device without any interface traps and then compare them with characteristics obtained from the simulation of the device with a homogenous distribution of interface traps. Here we focus only on the variations in the threshold voltage by monitoring V_{th}.

Pristine device electrical characteristics

The plots in Figure 3-8 (left) show the I_DV_D characteristics of the pristine (i.e., without traps) device (L_{eff} = 225nm) without any interface traps at three different gate (V_G) biases of 1.0V (black squares), 2.0V (red circles) and 3.0V (blue triangles), all the while the drain was (V_D) ramped to 3.0V. The plots in Figure 3-8 (right) show the transfer characteristics (I_DV_G) corresponding to the pristine device. In the figure the transfer characteristics at three different drain biases can be

identified, namely, 0.1V (black, squares), 0.2V (red, circles) and 0.3V (blue, triangles). The performance parameter, the threshold voltage (V_{th}), is obtained from the I_DV_G plots using the ELR method (described in section 1.1) for a drain bias of 0.3V. The threshold voltage for the pristine case was found to be 0.6 V and is used as a reference to calculate the change in the device characteristics when devices with traps are considered.

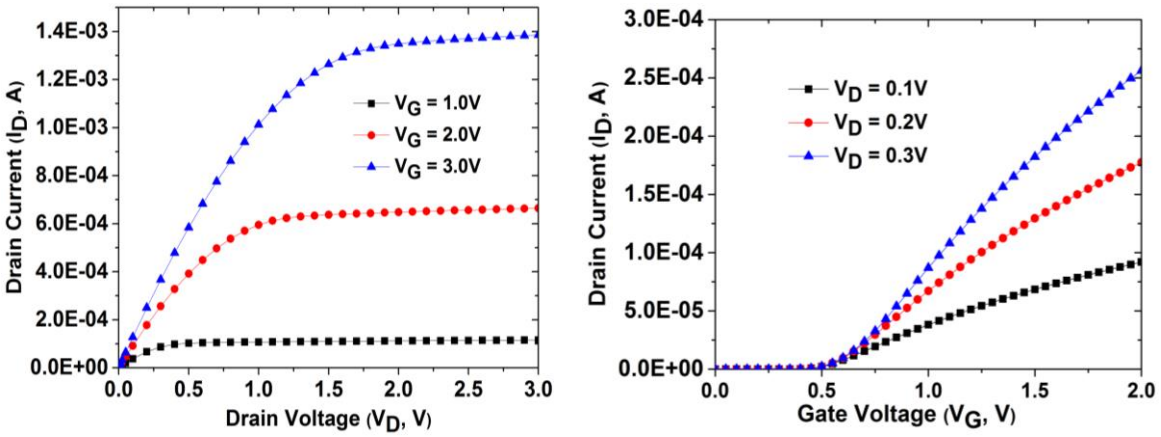


Figure 3-8: The plots on the left correspond to the I_DV_D characteristics of the pristine (fresh) device with an effective channel length (Leff) of 225nm. The plots on the right correspond to the I_DV_G characteristics of the pristine (fresh) device with an effective channel length (Leff) of 225nm

Homogenous Trap distribution

In line with our objective, we proceed to perform interface trap simulations to analyze the impact of their spatial distribution on the device characteristics. The plots in the pristine case serve as a reference to measure the change in the characteristics perpetuated by interface traps. The parameter 'threshold voltage shift' (ΔV_{th}) is used as a measure to monitor the change. ΔV_{th} is the difference in the threshold voltages of devices with and without traps (pristine device).

To simulate the effect of interface traps accurately, we need to assign reasonable values to the parameters that define the trap properties (such as density, cross-section, energy, etc.). There is a lot of ambiguity in the literature regarding the nature and number of interface traps that can exist at the interface between silicon and its oxide (SiO_2). This is largely due to the differences in processing and stress conditions that have been used for trapping studies. However, one can get a fair idea from different literature works. We consider the following attributes to interface traps with the respective values.

- Nature and density of the traps: Most of the trap states that are electrically active in SiO_2 are neutral traps. They are uncharged prior to capturing an electron and their density (D_{it}) in interfaces is about 10^{12} cm^{-2} [199]. In our simulations, we want to simulate the influence of very few discrete traps. So the chosen value of the trap density for the simulations done in this work use densities in the same range.
- Trap capture cross-section: The capture cross-section can be assumed as the area of influence of the trap and is in the order of 10^{-13} to 10^{-18} cm^{-2}, usually less than 10^{-14} cm^{-2}

[200], [201]. The atomic dimensions of the trap cross-section make the trap a highly localized potential well. To study the influence of this cross-section, several simulations with a different value for the cross-section (10, 100 and 1000 times) were performed. However, they did not have any effect on the device performance (by monitoring V_{th}). Hence, a value of 10^{-15} cm^{-2} was chosen to be consistent with those in the literature.

- Trap Energy: It has to be remembered that we are interested in traps that create unwanted energy bands in the forbidden gap. And as we are simulating traps that capture electrons (in other words acceptor traps), their energy (E_{trap}) is measured from the bottom of the conduction band. So having an energy E_{trap}=0 means the trap is located at the conduction band edge of silicon and higher values of E_{trap} mean the trap is located deeper in the energy band. In our simulations, after several trial simulations with different values for trap energy, the energy of the trap was chosen as 0.25 eV from the silicon conduction band. This value places the trap above the Fermi level so that they are empty and are available to trap an electron.

With the chosen physical models and trap parameters, we move towards performing the simulations of the device structure with interface traps. In order to understand the first-hand impact of the interface traps, we first consider the case of interface traps being homogenously distributed all along the interface. When traps (with a density D_{it}) are homogeneously distributed over the entire interface (in the effective channel), then they (with the same density D_{it}) are placed at every node along the interface. To evaluate the influence of traps on the device performance, simulations were performed with different values of trap densities (D_{it}). We know that interface traps (when charged) add space charge directly to the Poisson's equation and thus modify the electrostatic potential. This modification results in changes in the electrical characteristics of the device Figure 3-9 (left) shows the I_DV_G plots of the device structure without traps (pristine) and with a homogenous trap distribution (with different densities) along the entire interface.

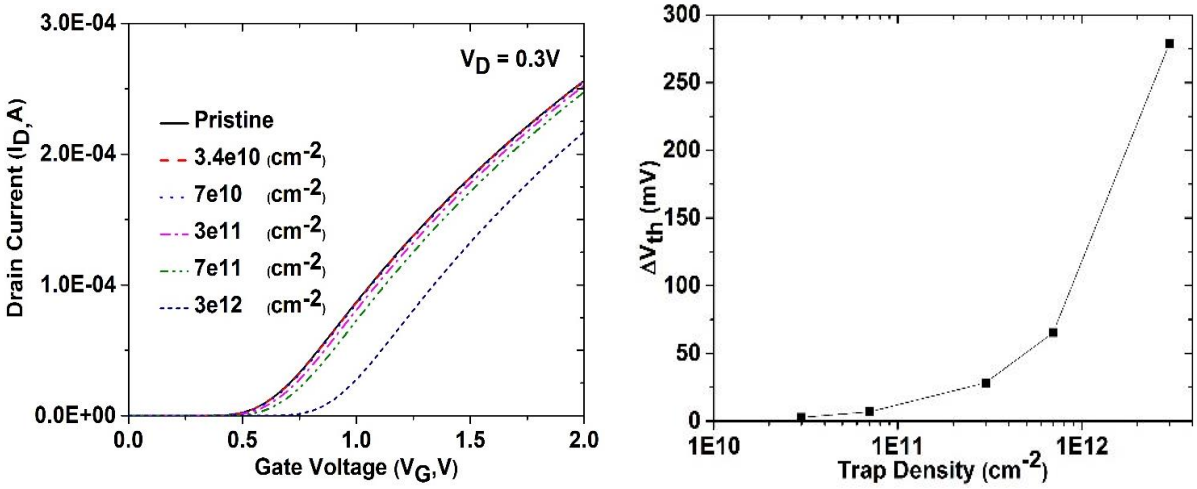


Figure 3-9: The plots on the left correspond to I_DV_G characteristics of the structure (L_{eff}) without traps and for devices with different densities of traps that are homogenously distributed along the channel and with a drain bias of V_D = 0.3V. The plots on the right shows the V_{th} shift as a function of trap density as obtained from the left figure. The fresh threshold voltage (0.6V) for the case of V_D = 0.3V is considered as reference.

We can see from the plots that the introduction of interface traps resulted in the shift (lower currents, higher V_{th}) of the device characteristics. This change in the device parameters is also reflected in device Figure 3-9 (right) where the threshold voltage shifts (ΔV_{th}) are plotted as a function of D_{it}. For these plots, the pristine threshold voltage (0.6V) at 0.3V of V_D is taken as a reference to calculate the ΔV_{th}. It can be seen from the plots that the introduction of traps along the interface has led to the change in the drain current, being more significant for higher values of $D_{it,}$ for example in case of D_{it} = 3e12 cm^{-2}, the drain current decreases by about 32µA from its pristine value. Note that interface traps can introduce a significant increase in the threshold voltage which can be as high as 50% for large enough D_{it} (3e12 cm^{-2}). Next the influence of trap distribution is analyzed by localizing the traps at discrete sites along the interface. In the following section, we discuss the influence of discrete trap distributions.

In line with our objective, we proceed to perform the simulations to analyze the influence of trap location on the device performance. For the sake of simplicity, we start with the case of single trap location (N_{TL} = 1), i.e. only one trap location is populated with traps at each simulation instant. We then move onto the case of multiple trap locations along the interface. At the end of each simulation, the threshold voltage (V_{th}) is extracted and the variability introduced due to the spatial distribution of trap location is analyzed by performing a simulation with a random location for the trap region (see Figure 3-5). It has to be recalled that traps were placed only at the interface along the effective channel length (L_{eff}, also referred in the work as just channel length) and not the entire gate length (L_G).

3.3 Single trap location

As stated earlier, the interface division is made such that only one node is encompassed by each region and with a node separation of 2.5nm, it would mean that the total number of possible trap regions (N_{Tot}) in a device with effective channel length of 225nm and width of 1µm is 90 (225/2.5), each with the same size (2.5nm X 1µm). The value of the homogenous trap density is used to determine the value of D_{TL} from the equation 3-1 which is also stated below.

$$D_{TL} = \frac{D_{it} * N_{Tot}}{N_{TL}}$$

With the value of D_{it} considered as 3.4e10 cm^{-2}, using the above equation (substituting N_{TL} = 1 and N_{tot} = 90) D_{TL} was found to be 3e12 cm^{-2}. Although a homogenous distribution of traps with this D_{it} (3.4e10 cm^{-2}) resulted in a ΔV_{th} of 3mV (device Figure 3-9), it was meaningful, as it has to be recalled that in 2D simulations the device has a default width of 1µm, which meant that in the device with an effective channel length of 0.225µm, the traps were distributed over a large area (0.225µm x 1µm). Moreover, using a higher value of D_{it} resulted in a value of D_{TL} that had an adverse effect on the simulation time and convergence. The pristine threshold voltage (see V_{th} definition) of 0.6V corresponding to V_D = 0.3V is used as a reference. The shift of the threshold voltage (ΔV_{th}) from this reference due to the presence of interface traps is used as an indicator to measure the trap influence.

From the previous chapters, we understand that an interface trap behaves as a localized potential well. It adds space charge directly to the Poisson's equation (see eq.2-4) and modifies the electrostatic potential locally and contributes to the threshold voltage shift. This can be visualized from the simulations. Figure 3-10 shows the (zoom of the interface region) potential profile (contour of the electrostatic potential) along the interface in the simulated structure without traps (Figure 3-10: left) and with traps at a specific location (marked by the rectangle in Figure 3-10: right). The contour maps were obtained when the gate is biased to a value that is close to the threshold voltage (in a pristine device) and drain is biased at 0.3V. Here we can observe that the placement of traps has modified the potential contour lines and this modification is more pronounced in the proximity of the traps.

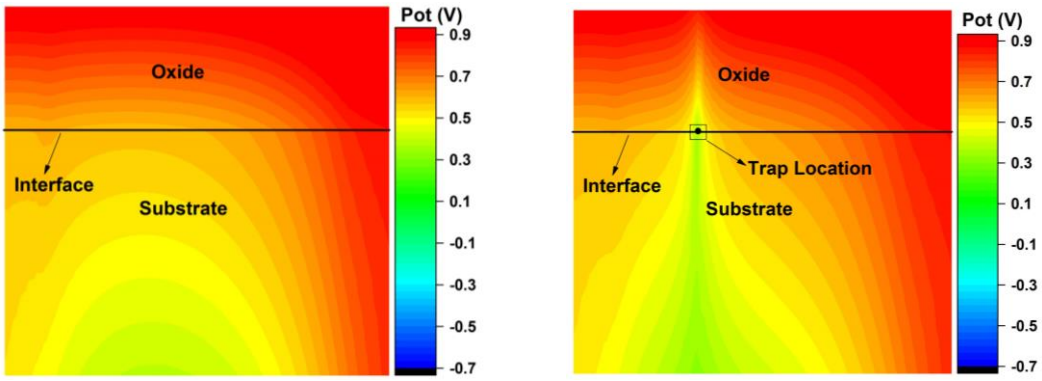


Figure 3-10: Zoom of the potential profiles along the interface in the simulated structure (left) without traps and (right) in the presence of traps.

To understand how each trap location modifies the device performance, simulations were done by varying the location of the interface trap along the channel interface. Figure 3-11 shows the plots of I_DV_G corresponding to three different scenarios: pristine (black, squares), homogenous (red, circles) and traps at a discrete location (blue, triangles) along the channel. The discrete trap location corresponds to a device with traps located close to the drain of the channel. It has to be noted that although homogenously distributed traps yielded a very low ΔV_{th} (a shift of 3mV), when the same number of traps are localized in a small region they assert a significant change in V_{th} (ΔV_{th} of 22mV for the case of Figure 3-11). The influence of the trap location on the device performance (ΔV_{th}) is analyzed by plotting ΔV_{th} as a function of the trap location that causes the shift and this is shown in Figure 3-12. In Figure 3-12 ΔV_{th} is plotted as a function of the distance of the trap location from the source edge.

For a clearer view of the device edges, doping profile is overlaid onto the same plot. Here '0' along the X-axis corresponds to the source edge of the channel and '225' corresponds to the end (drain edge) of the channel. By plotting the ΔV_{th} as a function of the location of the traps in the channel, we can clearly observe that ΔV_{th} is almost always (except when the traps are located at the channel edges) always larger than the same number of traps were homogenously distributed. The smaller values were obtained when the traps were located at the channel extremes. We can also observe that when traps are located close to the drain edge, they tend to provoke a higher shift in the threshold voltage. In fact, the maximum shift in V_{th} happens when traps are located at a region close to the drain edge of the channel, which in this case is at 175nm from the source edge. To

understand the result better, we investigate the device band profile which is plotted in Figure 3-13. Figure 3-13 (a) shows the profile of the conduction band of the device without traps and when traps are located at two different locations (25nm and 100nm). The conduction band (CB) profiles were obtained for a gate voltage of 0.5V (close to the threshold voltage in a pristine device) and drain voltage of 0.3V.

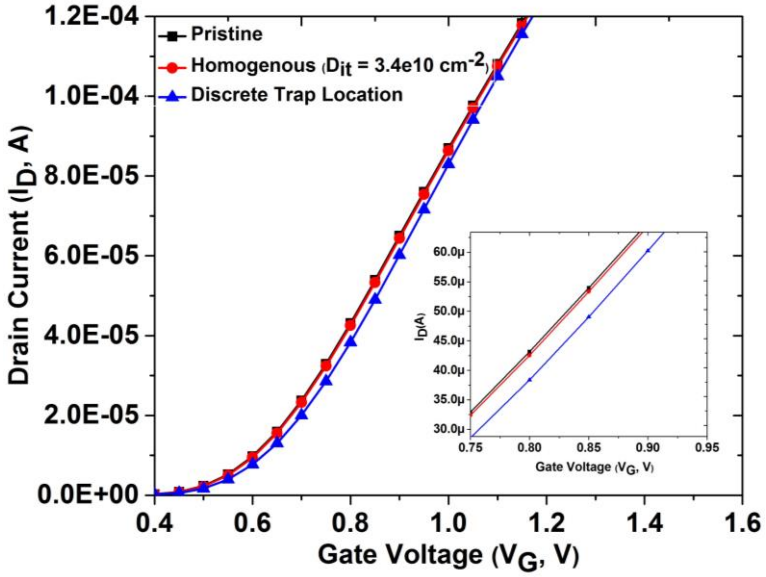


Figure 3-11: I_DV_G plots corresponding to the Pristine (black, squares), Homogenous (Red, circles) and traps at a discrete location (blue, triangles) for the device with L_{eff} =225nm. Localized traps evoke larger change in the device characteristics.

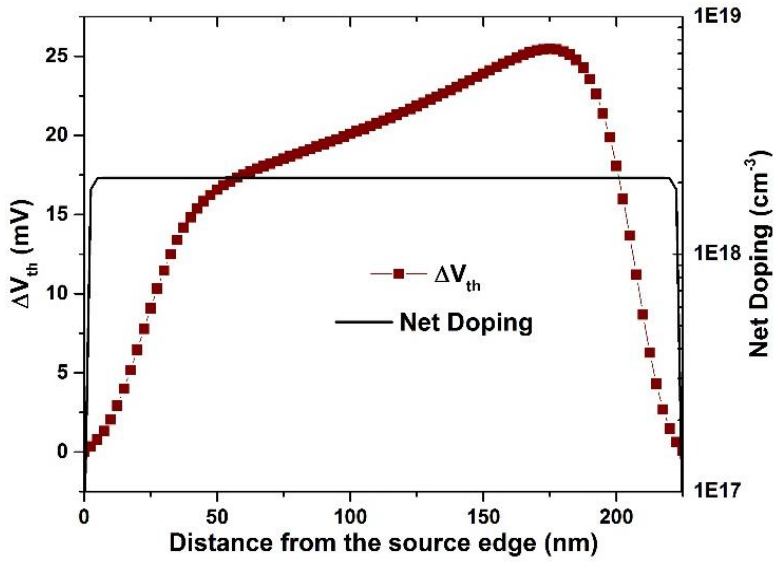


Figure 3-12: ΔV_{th} as a function of distance of the trap location from source edge for a device with 225nm effective channel length. ΔV_{th} is overlaid onto the 1D cut of the doping profile. It can be seen that the maximum ΔV_{th} is obtained when traps are located at the drain edge of the channel.

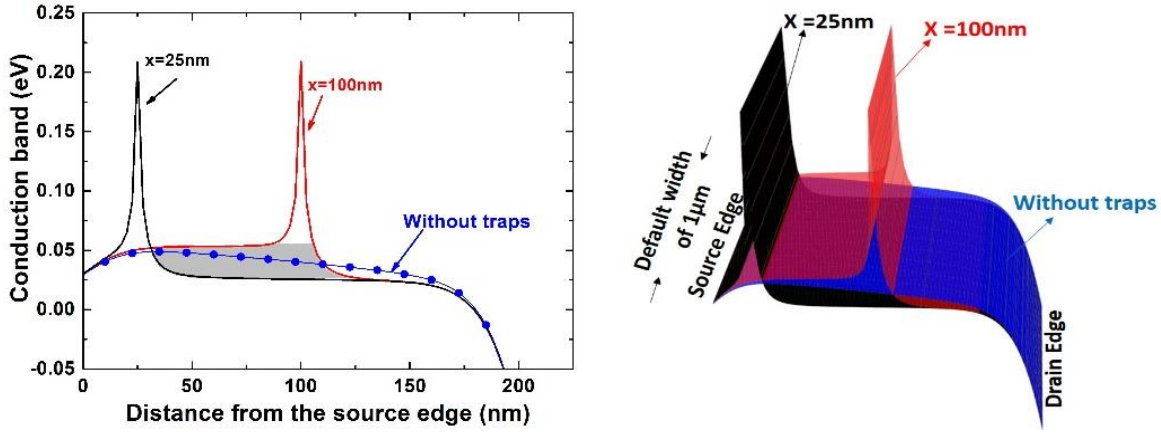


Figure 3-13: The plots on the left correspond to the CB edge for two different trap location (red and black) and also without traps (blue dotted). The plot on the left correspond to the same plot as on the left but shown in 3D with the default width of 1µm.

Here the x-axis represents the distance from the source edge of the channel in nm, with x=0 being the source end of the channel and x=225nm being the drain end of the channel. The black curve represents the case when traps were located at x= 25nm and the red one corresponds to the case when traps were located at x = 100nm. The blue curve (with circles) corresponds to the profile of the conduction band when there are no traps (pristine device). Recall that a simulated 2D structure has a default width of 1µm. which means that the influence of the trap is extended over the entire device width and this is shown in Figure 3-13 (b). We can see that placement of traps has created an increased height of the barrier to carrier flow from source to drain. It can also be seen from the plot that the trap at x=100nm (0.1µm) creates an extra offset (shaded zone) of the barrier in comparison with the potential barrier obtained when the trap is located at x=25nm. Then the effect of the trap which is to create an additional opposition to the carrier flow has a dependence on the trap location in the channel, being larger when closer to the drain. However, it should be noted that the maximum height of the barrier has no dependence on the trap location.

The importance of the trap location can be demonstrated further in Figure 3-14, where traps at different locations are compared. Also shown is the CB profile when there are no traps in the device. In the figure, the cut of the band profile is made for the entire device (not just in the channel) and the portion that forms the effective channel is clearly marked. It can be seen that when traps are located close to the drain end (in this case at 175nm) of the channel, then the resulting increased barrier is extended over the entire length. This results in a larger value of the threshold voltage. From these results, we can conclude that the location of interface traps along the channel has a significant influence on the threshold voltage (or its shift). In order to study the influence of the spatial distribution of traps on the variability of the device performance, different devices (100) were simulated with different spatial distribution of traps i.e., changing the location of the traps randomly. Figure 3-15 shows the variability in the threshold voltage shift across the 100 simulated devices.

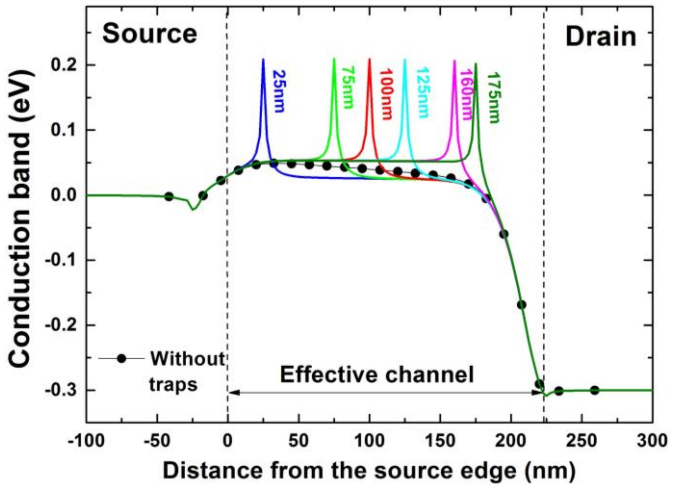


Figure 3-14: CB edge for six different trap locations along the channel, also included in the plot is the band profile corresponding to the device without traps (pristine).

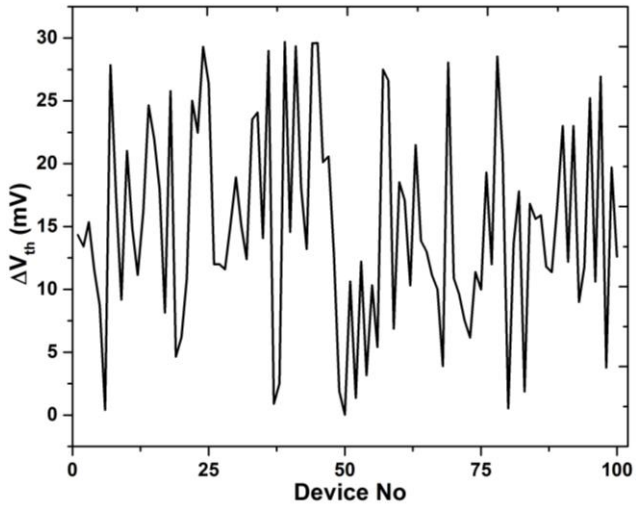


Figure 3-15: Variability in the ΔV_{th} across devices with varying distribution of traps.

The x-axis in Figure 3-15 corresponds to devices with the different spatial position of traps and y-axis plots the shift in the threshold voltage (ΔV_{th}) from the pristine value. It can be seen from the figure that there is large variability in the ΔV_{th} resulting from variations in the location of the traps.

Influence of drain bias

Up until now, the influence of the trap location was analyzed when the drain was biased to 0.3V. However, simulations were performed to evaluate if the V_{th} change abetted by interface traps has any dependence on the drain bias. So simulations at different drain biases were considered. Three different values of drain biases (apart from 0.3V which was already simulated) were chosen

(V_D=0.2V, 0.4V and 0.5V) and the simulations described earlier were repeated. The plots in Figure 3-16 show ΔV_{th} as a function of the trap location for different values of drain biases. It has to be noted that, the drain voltage seems to influence the V_{th} when the traps are located in the central region (between 25 and 200nm) of the channel while being ineffective when the traps are at either (source/drain) edges. However, it can be seen that irrespective of the applied drain bias, the maximum shift in the threshold voltage always happens at the same location, i.e. close to the drain edge.

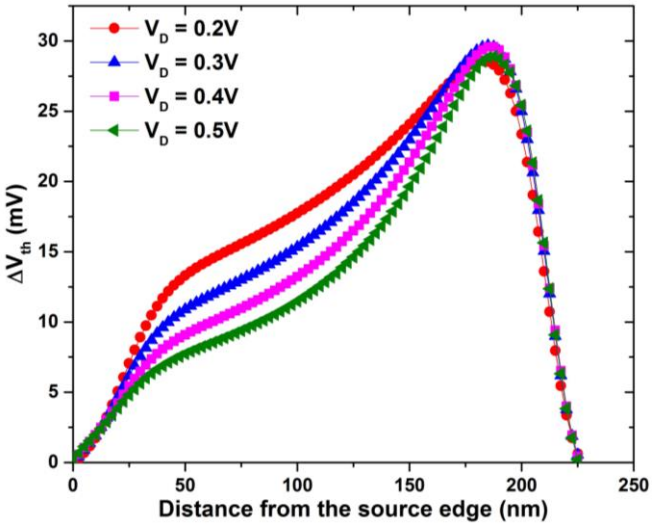


Figure 3-16: V_{th} shift as a function of the trap location, for different drain voltages. It can be seen that the influence of drain bias on V_{th} is significant when traps are located in the central region (between 25 and 200nm) of the channel. However, the location of the traps that leads to a maximum ΔV_{th} is always the same (close to the drain edge).

3.3.1 Channel length dependence

The simulation performed until now considered a device with a gate length (L_G) of 250nm (L_{eff} = 225nm), however, as discussed in chapter 1, channel lengths of modern devices can be much smaller. Hence to have a realistic understanding, TCAD simulations are used to extend the analysis to smaller channel lengths. In the following sections, we discuss the simulations that correspond to shorter channel lengths.

Device Structures and Simulation Setup

Device structures with channel lengths ranging from 90nm (apart from the case of 225nm which was analyzed earlier) down to 30nm were simulated and the case of 50nm is studied in depth. The device structure that was simulated before (225nm) used a doping profile that was calibrated to the standard fabrication process of long channel devices. However, as simulation ventured into smaller channel lengths it was decided to use doping profiles that were well calibrated to a standard fabrication process of short channel devices. Hence, the devices with shorter channel lengths (90nm and below) were created using the well calibrated MIT template [202] with the same physical oxide thickness of 2nm that was considered for the long channel device. Figure 3-17

shows an example of defining L_{eff} from the cut of the doping profile. The x-axis in the figure corresponds to the coordinates of the simulated structure, '0' being the channel center.

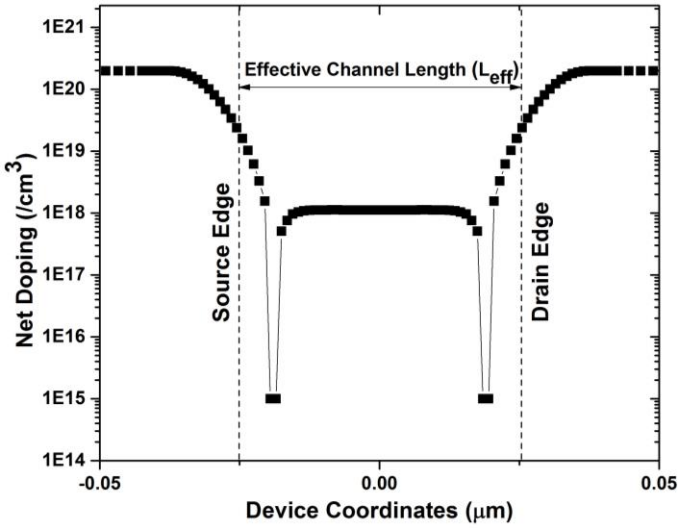


Figure 3-17: Cut of the doping profile of the simulated structure. The effective channel length (Leff) is defined between where the source/drain dopings fall to 2e19 cm-3. The channel edges are delimited by source and drain edge.

The effective channel length defined between the points where the source/drain dopings fall to 2e19 cm^{-3} (also marked in Figure 3-17). Also marked are the channel extremes referred here as source edge (at the source side) and drain edge (at the drain side). Devices with different effective channel lengths (defined between where the source/drain dopings fall to 2e19 cm^{-3}) were created by varying the doping profile defined by the template. We use the same strategy that was employed in the case of long channel devices to perform simulations of devices with different gate lengths. The physical models used in the simulation are the same ones that were considered for the case of long channel device. However, the effect of band gap narrowing was also included to describe the carrier statistics. This was done as in shorter gate lengths the band gap separation decreases with the increase in the doping. Although this decrease is also observed in large devices, the effect is more pronounced in nanoscale devices [203].

In order to evaluate the influence of the trap location on the V_{th}, we divide the interface into many smaller regions according to the strategy outlined in section 3. With respect to the mesh employed, the separation between nodes at the interface is 1nm, then the total number of interface nodes (N_{Tot}) is same as the channel length (for example, for a 50nm long channel device N_{Tot} =50). As with the long channel device, the trap density that was used to populate the trap region is derived from a chosen uniform trap density (D_{it}). In this case, a D_{it} of 5e11 cm^{-2} over the entire channel is considered as a reference. This value is higher than the one that was chosen earlier (for L_G = 250nm) and this was done to have meaningful comparisons when moving to 3D, where the device width is reduced and lower D_{it} would have reduced the number of traps to be of any significance. Also, this value of the homogenous density (D_{it} = 5e11 cm^{-2}) is compatible with that in the current technology nodes and implies a measurable change in the V_{th}. Then to keep the total charge in the device constant, the trap density (D_{TL}) at each trap location is calculated according to the equation 3-1 which is also listed below for reference.

$$D_{TL} = \frac{D_{it} * N_{Tot}}{N_{TL}}$$

As it was done in the case of a long channel, we first start our analysis by simulating the device structure without traps (pristine or reference) and with homogenous distribution of interface traps. The results of the simulations with traps are then compared with the pristine case to measure the shift in V_{th}.

Pristine characteristics and homogenous trap distribution

Although the devices with various channel lengths will be analyzed, unless otherwise stated, the case of L_g=50nm will be described in detail as an example. The threshold voltage is extracted using a constant current method as opposed to the maximum slope method that was used in the case of long channel device. The definition of V_{th} using a constant current (CC) method is more appropriate for short channel devices, as ELR method is sensitive to mobility degradation (see section 1.1) and greatly influences the V_{th} measurement. Figure 3-18 shows the I_DV_G curves obtained for devices without traps (squares) and with a uniform interface trap distribution (circles) with a density (D_{it}) of 5e11cm^{-2}. The threshold voltage is defined as the gate voltage that corresponds to the drain current of 1µA.The characteristics shown in the figure (Figure 3-18) were obtained when the gate was ramped to 1.5V (V_G) at a fixed drain bias of 0.3V (V_D). Using the CC method, threshold voltages of 0.407V and 0.457V were obtained for the pristine (without traps) and uniform trap distributions, respectively. This meant that the introduction of uniformly distributed interface traps leads to a threshold voltage shift (ΔV_{th}) of 50mV.

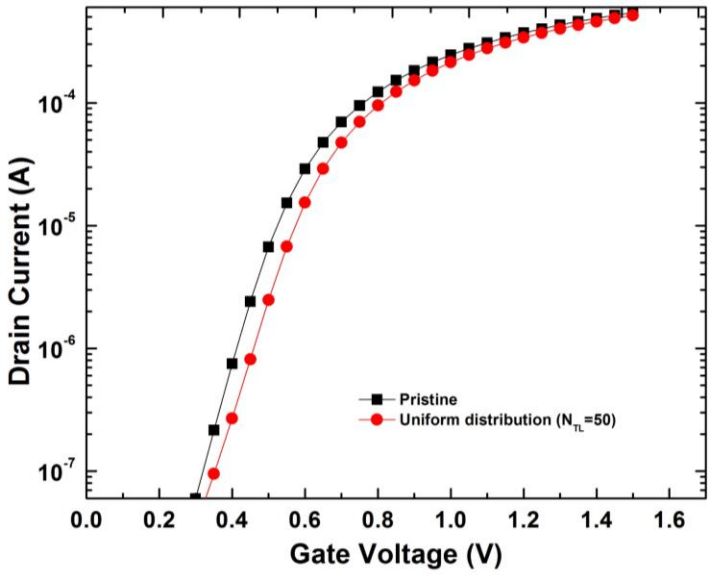


Figure 3-18: Simulated transfer characteristics in case of pristine device and when traps are uniformly distributed throughout the channel in a device with L_{eff} =50nm and D_{it} = 5e11 cm^{-2}.

Discrete trap distribution: Influence of trap location

The influence of the discrete trap locations in devices with shorter channel length is analyzed by starting with the simple case of single trap location (N_{TL} = 1). The density at each trap location D_{TL} is calculated using the equation 3-1 and substituting N_{TL}, $D_{it,}$ as 1 and 5e11 cm^{-2} respectively, where $N_{Tot} = L_{eff}/1nm$. Table 1 lists the D_{TL} that is used for different channel lengths. Now, as it was done in the case of long channel device, the influence of trap location on ΔV_{th} can be analyzed by plotting ΔV_{th} as a function of the location of the traps along the interface in the device.

Channel Length, L_{eff} (nm)	**Density at each location (D_{TL}, cm^{-2})**
30	1.50E+13
35	1.75E+13
40	2.00E+13
45	2.25E+13
50	2.50E+13
55	2.75E+13
60	3.00E+13
65	3.25E+13
90	4.50E+13

Table 1: Interface trap density at each interface trap location for different channel lengths.

The dependence of the V_{th} shift (ΔV_{th}) on the trap location (measured from the source edge) for L_{eff} =50nm is shown in Figure 3-19. The plot of ΔV_{th} vs trap location is overlaid onto the channel doping as a guide to distinguish between regions (source, drain and channel). It can be seen from the plot that traps localized at a single location has produced a significant shift in the threshold voltage (max of about 300mV), six times larger than that observed in the case of homogenous distribution (50mV) (Figure 3-18). It is interesting to highlight that the larger ΔV_{th} is obtained for devices where traps (along the interface) are located close to the center of the channel (at 22nm from the source edge), as for longer devices, larger drifts in the V_{th} (ΔV_{th}) occur in devices where the traps (along the interface) are located close to the drain edge of the channel.

As done before (for long channel) this behavior can be well explained by investigating the conduction band (CB) of the device as shown in Figure 3-20. In this figure (Figure 3-20) we compare the CB edge of the devices with four different location of the traps along the interface. The band profile was extracted at a gate bias of 0.35V (close to the V_{th} of the pristine sample) and a drain bias of 0.3V, for a 50nm device. In Figure 3-20, the modification of the barrier corresponding to different trap locations (38, 31, 22 and 10nm from the source edge) can be clearly seen. The trap located at 22nm (close to the center) produces a barrier that is higher compared to the rest and this is reflected in Figure 3-19 where the maximum ΔV_{th} happens when traps are located in a region close to the center. Then, the maximum ΔV_{th} takes place at those channel positions where the barrier to be surmounted by the carrier is larger.

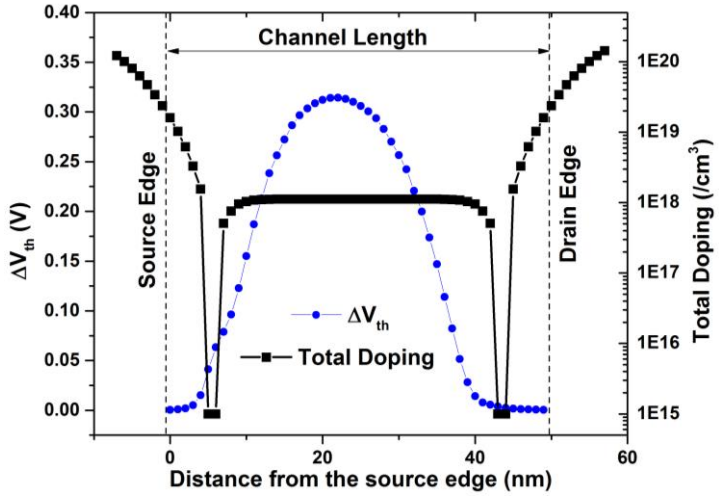


Figure 3-19: ΔV_{th} (circles) as a function of the distance of the trap location from the source edge overlaid onto the channel doping profile (squares). The plot corresponds to the device with channel length of 50nm with drain biased to 0.3V.

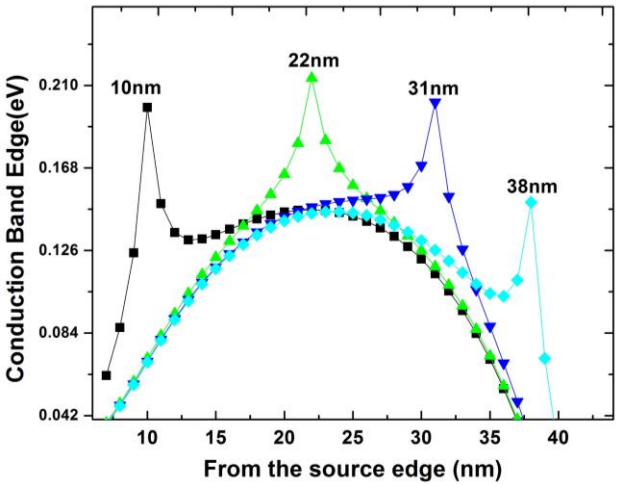


Figure 3-20: Conduction band edge corresponding to different interface trap locations along the channel. The bands were obtained at V_G=0.35V and V_D=0.3V.

The simulations have shown that for a short channel device (L_{eff} = 50nm), traps located close to the center (in contrast to the case of the long channel) of the channel have resulted in a maximum shift in the device threshold voltage. So in order to see if this trend is related to the variation in channel lengths, simulations were repeated for devices with different channel lengths and with D_{TL} obtained from Table 1. The threshold voltage for different channel lengths was extracted using the CC method. However, the considered value of the fixed current was scaled with the gate length. Figure 3-21 shows the location along the channel from the source edge (normalized to L_{eff}) that corresponds to the maximum V_{th}, versus the channel lengths. We can see that for shorter channel lengths (less than 100nm) the location of the traps that correspond to the maximum V_{th} is from the source edge at a distance that is between 45%-60% of the channel length, which means that they

are located in a region that is closer to the center of the channel. For larger channel lengths this maximum V_{th} happens at a location that is closer to the drain edge (distance is greater 80% of the channel length). Then from these results, it can be concluded that, as channel lengths are reduced, the location (measured from the source edge) that leads to the maximum V_{th} (or ΔV_{th}) moves closer to the source edge of the channel.

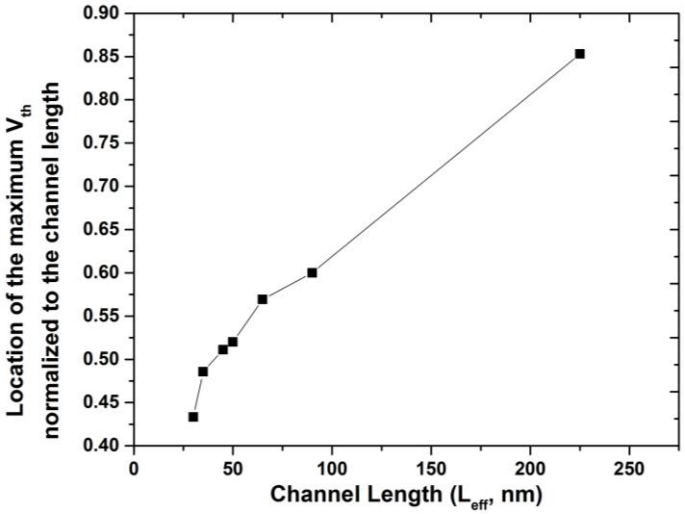


Figure 3-21: Relative distance from the source edge of the trap location that leads to max V_{th}, as a function of the transistor channel lengths. It can be seen that the location of the maximum V_{th} moves from a region closer to the drain end of the channel to the source end of the channel as channel lengths are reduced.

V_{th} Variability

As it was done for the long channel device, we perform simulations of a set of 100 devices (with effective channel length of 50nm) that are nominally identical but each with a different location of trap region in the device. The 100 simulated devices are identical to each other with the only difference of having a different location of trap region. Figure 3-22 (left) plots the variability in V_{th} and (right) plots the variability in ΔV_{th} across the 100 simulated devices. In the plot of V_{th} (left) the value of average V_{th} (0.579V) is also shown (red horizontal line). The standard deviation of V_{th} (σV_{th}) across 100 devices was found to be 0.139V. In all the simulations the gate was biased to 1.5V for a fixed drain bias of 0.3V. We can see that changes in the location of the traps along the interface introduce a large variability in V_{th} (ΔV_{th}) of the set of devices.

Influence of drain bias

It was outlined earlier (in chapter 1), that as channel lengths are reduced, the influence of drain potential on the channel region is substantial. The effect known as Drain-Induced Barrier lowering (DIBL) is similar to the punch-through effect, which is often referred as “subsurface DIBL”. In the weak inversion regime, there is a potential barrier between the source and the channel region and the height of this barrier is a result of a balance between the drift and diffusion current between these two regions. The application of higher drain bias can reduce the height of this barrier leading to an increased drain current. Thus, the drain current is not only controlled by the gate bias but

also by the drain bias. Figure 3-23 plots the cut of the conduction profile along the interface of the device with a channel length of 50nm for different drain biases. It has to be noted that these profiles are obtained for the device with no interface traps. We can clearly see from the figure that the increase of the drain bias results in a modification of the barrier (consequently reducing it). For device modeling, this parasitic effect of barrier height reduction can be accounted for by a threshold voltage reduction depending on the drain voltage [204]. Figure 3-24 plots the V_{th} (without interface traps) as a function of the drain voltage. Here we can see that as drain voltages are increased, there is a reduction in the measured V_{th}. Hence, we can state that the consequence of the barrier reduction observed in Figure 3-23 results in the decrease of the V_{th} that is observed in Figure 3-24. We have seen that the drain bias has a great influence on the device performance. Then, the drain bias can also influence the way localized traps modify the V_{th}.

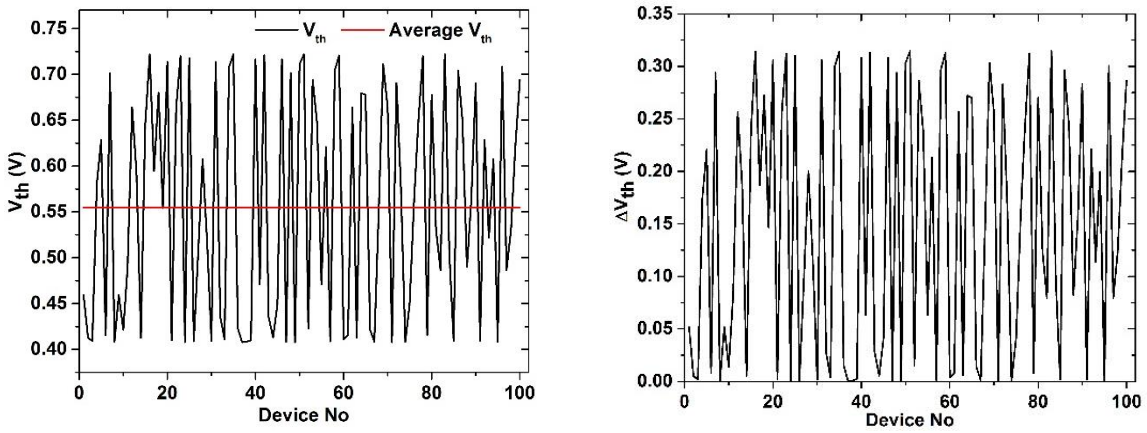


Figure 3-22: Variability in V_{th} (left) and ΔV_{th} (right) across 100 simulated devices with varying location of the single trap region. In the plot of V_{th} (left) average value of V_{th} is also plotted as a red line.

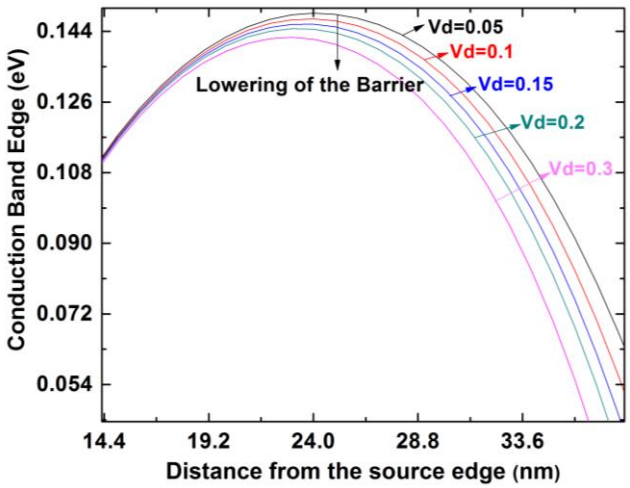


Figure 3-23: Conduction profile edge along the interface corresponding to different biases of the drain.

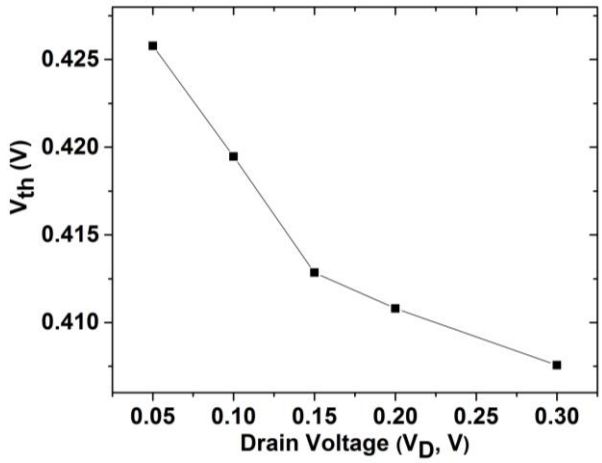


Figure 3-24: Threshold voltage in a device without traps as a function of the drain bias.

In order to understand the influence of the drain bias on the threshold voltage in devices with discrete trap distributions, we repeat the device simulations with interface traps and with N_{TL} = 1 for 50nm channel length device, considering different drain voltages. Figure 3-25 plots the shift in the threshold voltage (ΔV_{th}) as a function of distance (from the source edge) of the trap location, for the different drain biases. From the plots in Figure 3-25, it can be seen that the V_{th} is not influenced by the drain bias when the traps are located close to the source edge. However, when traps are located in the center of the channel or closer to the drain edge, then we can see that the V_{th} is influenced by the drain bias. It can be seen that the ΔV_{th} induced by these traps (located in the center or closer to the drain edge) decreases with the increasing drain bias. In any case, the location of the maximum ΔV_{th} does not depend on the drain bias and always happened when traps are located along the channel center.

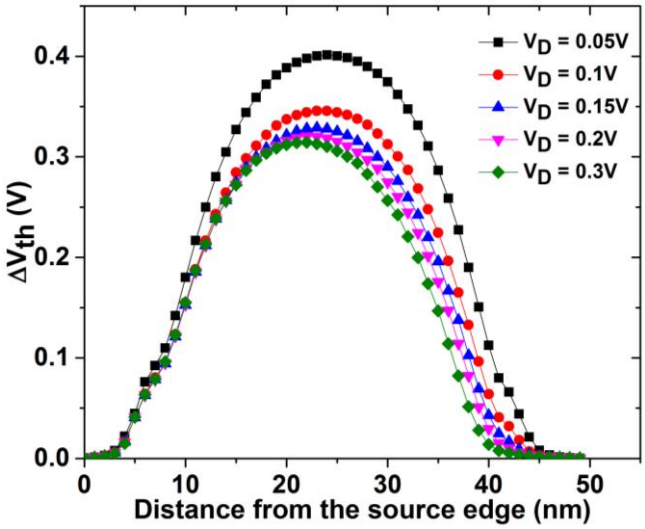


Figure 3-25: The variation of the ΔV_{th} with trap location and at different drain biases.

3.4 Two trap locations

Until now, we simulated device structures with traps located only at a single location along the interface. However, in reality, traps can be spread over multiple locations along the interface. In the following sections, we discuss the impact of traps simultaneously spread at multiple locations along the interface on V_{th}. We start with the simplest case of 2 trap locations and then extend our study to include more than 2 trap locations. For this analysis, we consider the device with an effective channel length (L_{eff}) of 50nm that was used earlier in the above sections. Then the density at each trap location D_{TL} is calculated using the equation 3-1 and substituting N_{TL}, $D_{it,}$ and N_{Tot} as 2, 5e11 cm^{-2} and 50 respectively. In this case, D_{TL} was found to be 1.25e13 cm^{-2}. Like it was done earlier, to evaluate the variability in the device performance, we perform simulations of a set of 100 devices with the different spatial distribution of traps (different positions of the 2 trap locations). Figure 3-26 (left) shows the plot of the threshold voltage (V_{th}) variability and on the right is the plot of the variability in the threshold voltage shift (ΔV_{th}) across different simulated devices. Also in the plot of V_{th}, the average value (0.531V) is shown as a red horizontal line. It can be seen that there is a significant variability in the threshold voltage (σV_{th} = 0.067) perpetuated by the spatial distribution of traps, although the total charge in the device remains constant. However, it can be observed that when traps are placed at two different locations, then there is a decrease of the average V_{th} (0.531V) and the σV_{th} compared to the case of only one trap location (average V_{th} = 0.579V and σV_{th} = 0.139V, see Figure 3-22). In order to understand further the origin of the observed variability and the relative values of V_{th}, let's plot the V_{th} as a function of the relative distance between the two trap locations (Figure 3-27).

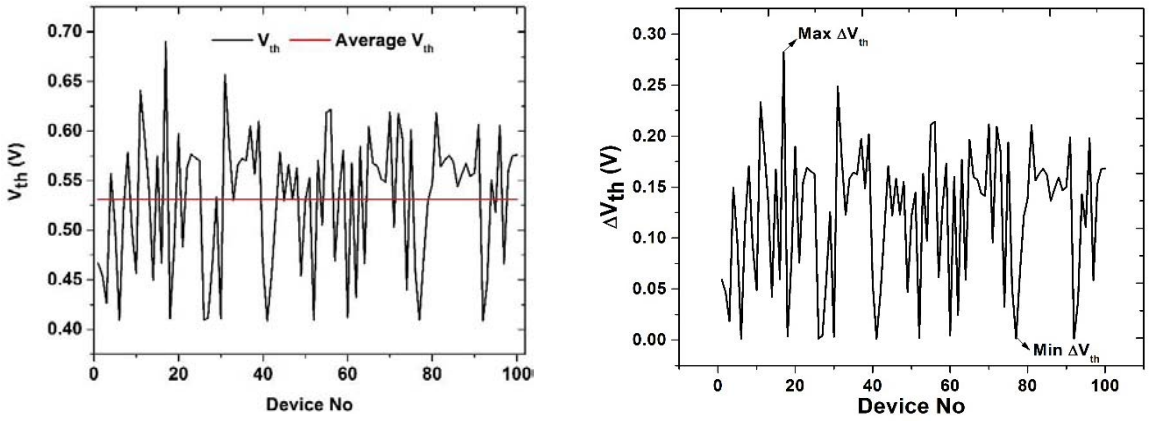


Figure 3-26: Variability in V_{th} (left) and ΔV_{th} (right) due to the spatial variation of the two trap locations across 100 simulated devices. In the plot of V_{th} on the left the value of average V_{th} is shown as a red horizontal line.

In Figure 3-27, four regions of special interest can be distinguished, they are marked by small rectangles and assigned different quadrant names (Q1, Q2, Q3 and Q4). Q1 and Q3 correspond to devices where the trap regions are close to each other and resulting in larger (Q1) or smaller (Q3) V_{th}. However, the actual location of the traps along the interface is not clear from the picture. Q2 corresponds to the devices where the separation between the traps is almost 50% of the channel length and resulting in intermediate values of V_{th}. This implies that one of the trap regions is located

close to the center and the other is located close to the channel edge (source/drain). Q4 corresponds to the devices where the separation between the trap regions is very large (almost the 80% of the channel length) and resulting in smaller values of V_{th}. This implies that for devices in the Q4, both the trap regions are located far from each other and close to the channel edges.

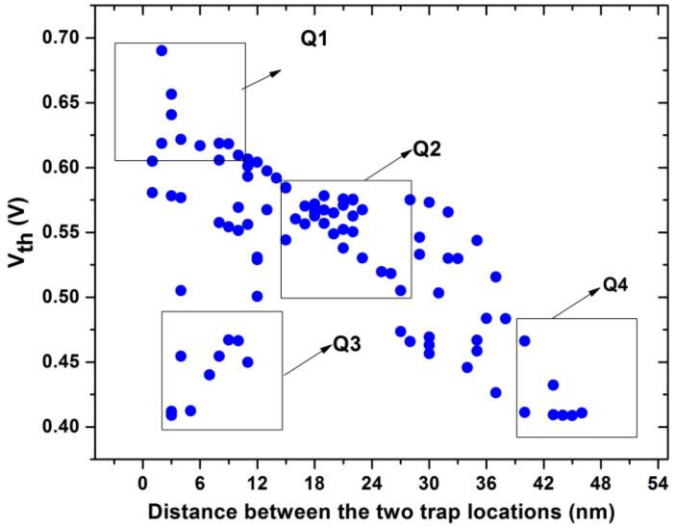


Figure 3-27: V_{th} as a function of the separation between the trap locations (N_{TL}=2) along the channel length. Four regions of interest (marked with squares) can be distinguished.

In order to understand how the spatial distribution of traps lead to a different magnitude of V_{th}'s (seen in Figure 3-27), it is imperative to understand the role of each trap location towards the total V_{th}. Each V_{th} in Figure 3-27 corresponds to a device with traps at two locations, let's call these locations 'T1' and 'T2'. A device is chosen from Q1 (shown in Figure 3-27) and simulations are done with (a) traps only at T1 and (b) traps only at T2. The V_{th} from each simulation (a, b) is compared with the V_{th} of the device when traps were located simultaneously at T1 and T2. It has to be noted that in all the three scenarios (with traps at T1 and T2, only at T1 and only at T2) the number of traps in each location was always the same. For this, we choose a device from the quadrant Q1 with a V_{th} of 0.690V and with trap locations as shown in Figure 3-28 (left). In the figure, the map of the locations of the traps is overlaid onto the zoom of the interface (with potential distribution) along the channel region. For comparison the potential distribution in a device without traps is shown in the (right) Figure 3-28. This aids in visualizing the modification of the potential along the interface due to the presence of interface traps at the respective locations (T1 and T2). For the considered device (in Figure 3-28), traps were located at a distance of 24nm (T1) and 22nm (T2) from the source edge. Simulation of a device with traps located only at T1 (24nm) lead to a V_{th} of 0.573V. While simulation of a device with traps located only at T2 (22nm) led to a V_{th} of 0.576V. We can see from the values of V_{th}, that the traps at these locations (T1=24nm and T=22nm) have almost the same influence on the threshold voltage (evident from the similar values of V_{th}). Now, we repeat simulations with a device from Q4. Here the trap locations are very far (at the channel edges) from each other (thus the larger distance) with T1=2nm and T2=46nm. Figure 3-29 shows the trap locations of the device from Q4 with V_{th} of 0.409V. Simulation of a device with traps located only at T1 (2nm) lead to a V_{th} of 0.408V. Simulation of a device with traps located

only at T2 (46nm) led to a V_{th} of 0.408V. We can see that the influence of these traps on the threshold voltage is similar and very small due to their position along the channel.

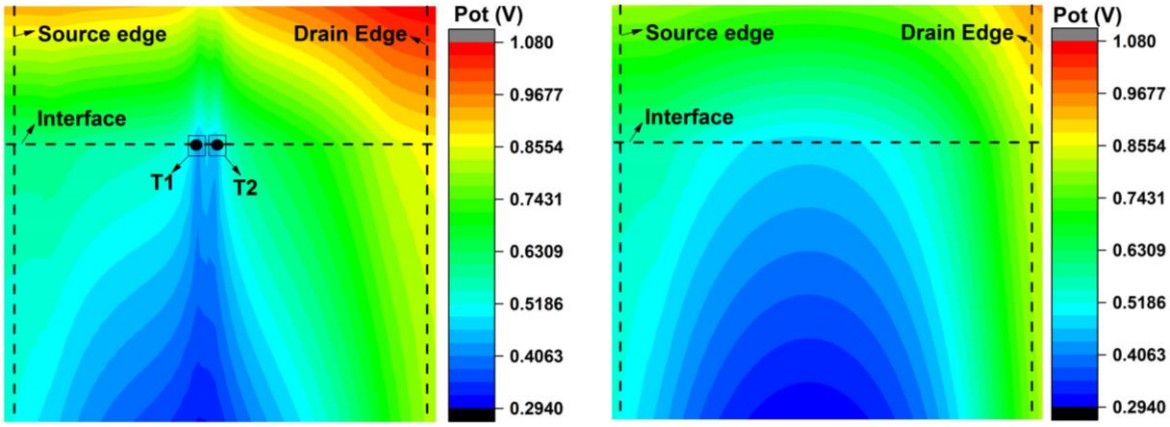


Figure 3-28 : (left) Potential distribution corresponding to the device with a V_{th} of 0.641V from Q1. Traps at the corresponding locations are also shown, for the example shown, traps are located at sites that are close to each other and located along the channel center. (right) Potential distribution of the device without traps.

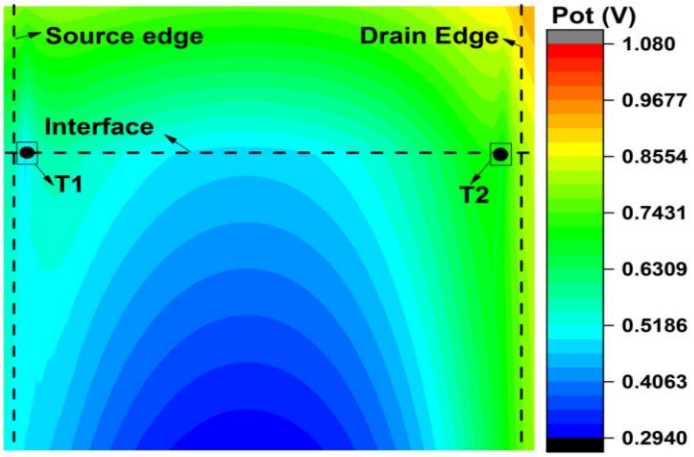


Figure 3-29: Potential distribution corresponding to the device with a V_{th} of 0.409V from Q4. In this case T1 is located at a site closer to the source edge and T2 is located at a site closer to the drain edge of the channel.

Now, Q3 in Figure 3-27 corresponds to devices with a small separation of trap regions but with V_{th}'s similar to that of devices in Q4. Same approach (as used before) can be employed to analyze the behavior of the devices in this quadrant. Figure 3-30 shows the trap locations of a device from Q3 with V_{th} of 0.412V where T1=41nm and T2=46nm. Simulation of a device with traps located only at T1 (41nm) lead to a V_{th} of 0.411V. While simulation of a device with traps located only at T2 (46nm) led to a V_{th} of 0.408V. We can see from the values of V_{th}'s that among the traps at T1 and at T2, dominant contribution towards the total V_{th} comes from the trap at T1. This is an interesting result as it suggests that there can be situations where traps located in a certain location assert a dominant influence on the device performance as compared to traps in other locations. This

behavior can be verified by considering a device in the quadrant Q2 where one of the two traps locations is close to the center while the other is close to the channel edge (source/drain). Figure 3-31 shows the trap locations of a device from Q3 with V_{th} of 0.565V where T1=7nm and T2=27nm. Simulation of a device with traps located only at T1 (7nm) lead to a V_{th} of 0.410V while traps located only at T2 (27nm) led to a V_{th} of 0.560V. These values of V_{th}'s also show that traps located close to the center of the channel (in this case T2) asserts a dominant influence on the total V_{th} to an extent of making the traps located close to the edge (T1, which in this case is close to the source edge) almost insignificant.

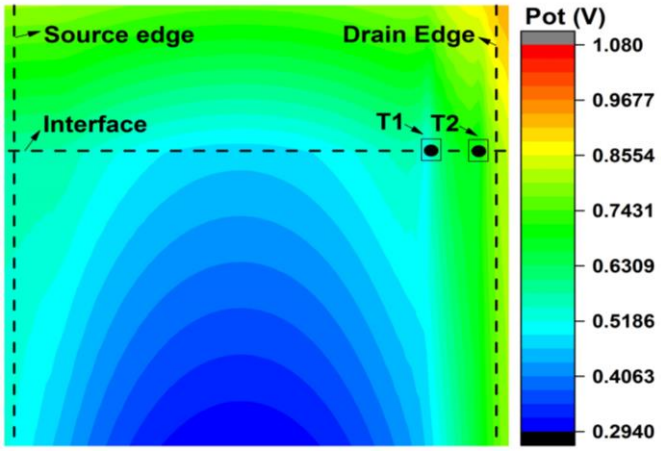


Figure 3-30: Potential distribution corresponding to the device with a V_{th} of 0.412V from Q3. In this case the traps are located at sites closer to the source edge of the channel.

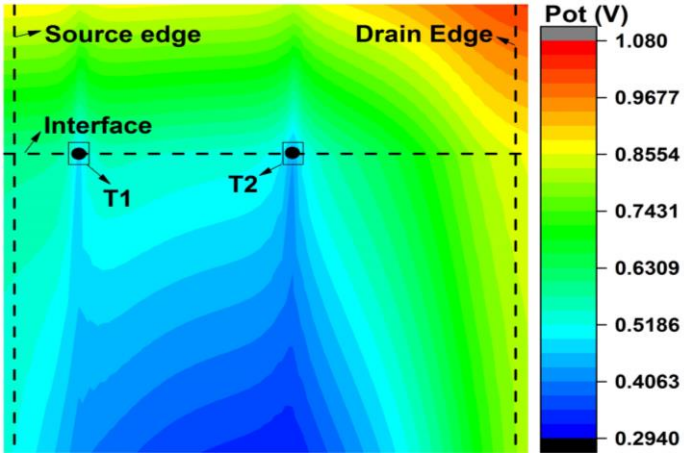


Figure 3-31: Potential distribution corresponding to the device with a V_{th} of 0.565V from Q2. In this case T1 is located close to the source edge and T2 is located close to the channel center.

The relationship ΔV_{th} (V_{th}) as a function of the relative location of the traps can be visualized from Figure 3-32. The figure plots the contour map of the threshold voltage across the 100 simulated

devices as a function of the distance of T1 and T2 from the source edge. Here we can see that larger V_{th}'s happen when both the traps are located closer to the center of the channel. Moreover, the traps located at the device edges do not significantly contribute towards the total V_{th}. With this understanding, we move to study the case of multiple trap locations (N_{TL} >2) and they are discussed below.

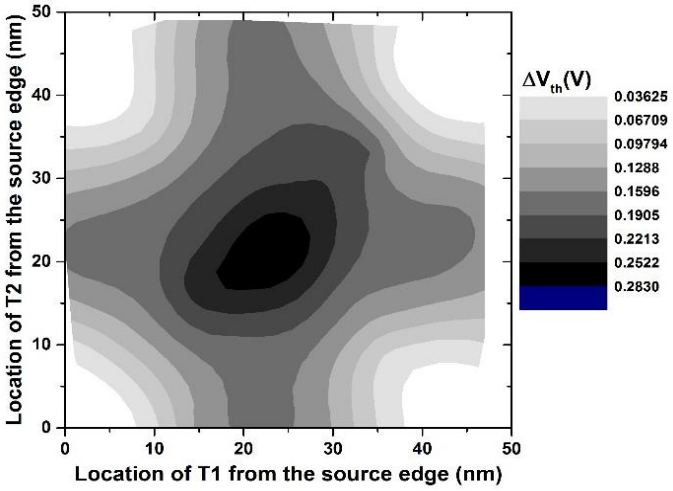


Figure 3-32: Contour map of the threshold voltage shift for different locations of T1 and T2 measured from the source edge of the channel. It can be seen from the plot that larger ΔV_{th} is obtained when both the traps are closer to the channel center.

3.5 Multiple trap locations

In the same lines discussed earlier, the case of multiple trap locations (N_{TL} >2) is considered. Recall that, according to equation 3-1, with the change in the number of trap locations populated with charge (N_{TL}), the density of traps at each location (D_{TL}) is scaled. As usual 100 devices (with L_{eff} = 50nm) were simulated with N_{TL} varying from 1 (single trap location) to 50 (uniform distribution) and D_{TL} correspondingly varying from 2.5e13 cm^{-2} (single trap location) to 5e11 (uniform distribution). Note that when N_{TL} is 50 then D_{TL} is same as the homogenous density D_{it}. Across the 100 simulated devices, the spatial distribution of the trap locations is varied randomly in each device to evaluate the variability of the device performance. For each value of N_{TL}, the average V_{th}, and its standard deviation is evaluated from the 100 simulations, and plotted in Figure 3-33 are the average V_{th} (black circle) and its standard deviation (red squares) as a function of N_{TL}. The extreme values of N_{TL}=1 (all the traps are concentrated at one location) and N_{TL}=50 (uniform trap distribution) are included in the graph. It can be observed from the plot that, as N_{TL} increases, both the average value of the V_{th} distribution and its standard deviation (i.e., the variability) decrease. These results can be explained taking into account that the total number of traps in the device is kept constant, regardless of how they are distributed in the device area. In other words, when the number of trap locations (N_{TL}) increases, to keep the total number of traps constant, D_{TL} at each trap location (N_{TL}) must be decreased. Then, the charge at each trap location contributes to the total V_{th} by a smaller amount. In addition, only those traps that are close to the center of the channel will significantly contribute to V_{th} (as shown in the case of N_{TL} = 2). So that, an overall

reduction in V_{th} is expected and is shown in the figure. As for the variability, larger N_{TL} would imply a smaller number of position combinations for the traps to be located at influential positions (close to the center), so that the randomness decreases. In the extreme case of a uniform distribution (N_{TL}=50), the randomness disappears, so that σ=0, as Figure 3-33 shows.

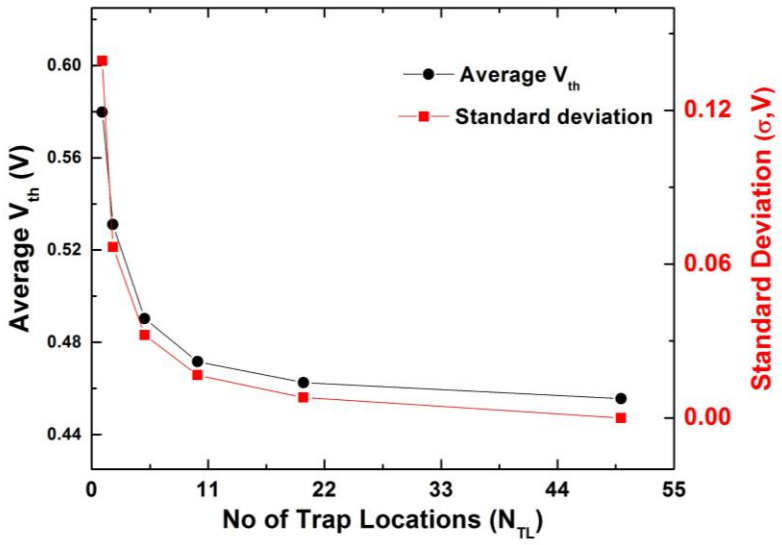


Figure 3-33: Average value (circles) of the threshold voltage and the standard deviation (squares) as a function of the number of trap locations. 100 devices where considered for each case. The extreme cases of N_{TL}=1 (single) and N_{TL}= 50 (homogenous) are also included.

In the next chapter, we will extend the 2D simulations into 3D to include the additional variability along the width of the device.

CHAPTER IV

V_{th} variability perpetuated by interface trap distribution: 3D simulations.

Until now the simulations done were in 2D which meant that the simulated structure had a default width of 1µm. However, to get more realistic results, simulations needed to be done in 3D, considering a device with a defined width. Hence, we perform TCAD simulations using the simulator tool Silvaco to study the influence of the number and spatial distribution of the traps on the V_{th} variability of 3D structures. In this chapter, we present and discuss the results of 3D simulations, and also study the dependence of the variability in V_{th} on the device geometry.

4.1 Device Structure, Simulation Setup and Methodology

As was done in the case of 2D simulations, the first step involves the definition of the device structure. It has to be recalled that when 2D simulations are performed, the device structure is, in fact, a 3D structure with a width of 1µm, however, in 3D simulations the width of the device is defined at a desired value. This implies that a mesh must be also defined in the direction along the width of the channel. As a result, the number of simulation nodes and consequently the simulation time increases. A 3D device structure is created by extending the 2D structure in the third (z-direction) dimension. We use the 2D device structure with 50nm channel length that was used earlier (Chapter 3) and extend it into the third dimension. Figure 4-1 shows the simulated 3D

structure with the map of the 2D doping profile. The simulated structure shown in Figure 4-1 has a width of 20nm and a channel length of 50nm. The chosen dimension of the channel length ensured that the considered device is a short channel device, and the chosen thickness (width) ensured that the simulations can be completed in a reasonable time. The device attributes (such as doping, oxide thickness, material, etc.,) were the same as the 2D structure simulated before.

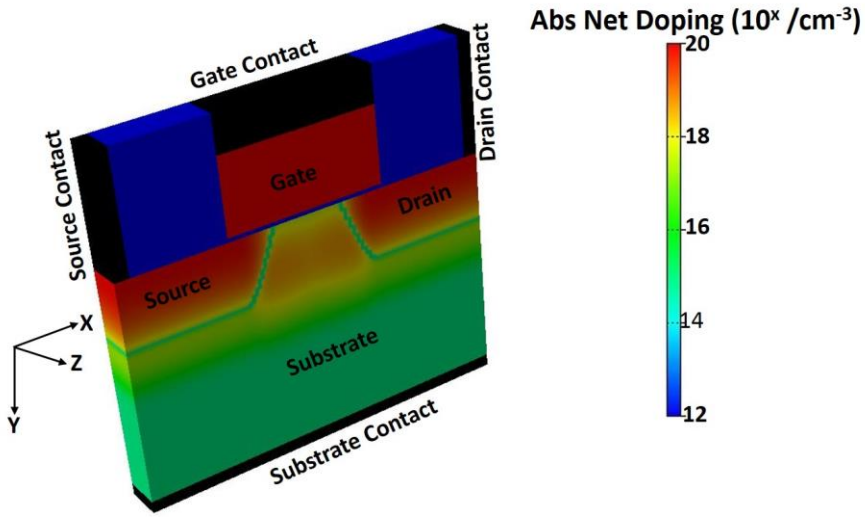


Figure 4-1: 3D simulated structure of 50nm channel length and 20nm channel width. The doping profile is also overlaid on the device structure.

As stated, 3D simulations require meshing the device structure also along the channel width. A fine mesh with a node separation of 1nm (same as in 2D) is used along interface parallel to the channel length (x-axis) and a node separation of 2nm is used along the interface perpendicular (width, z-axis) to the channel length. The adopted mesh structure meant that in the 2D interface there are 550 nodes (50 nodes along the length, across 11 nodes along the width). In order to analyze the influence of the location of the traps on the device performance (V_{th}), we proceed by dividing the interface into several smaller trap regions (as done in the case of 2D). However, unlike in 2D where the area of each trap region (location) was 1nm x 1µm (see Figure 3-4), in 3D the area of each trap region is 1nm (node separation along the length) x 2nm (node separation along the width, see Figure 3-6). We applied the same constraint that each trap region encompasses only one interface node, thus, the total number of trap locations is same as the total number of nodes (550). Following the methodology that was outlined in the Figure 3-7, we start our analysis by considering a reference device (Figure 4-1) and perform the electrical simulation without interface traps to evaluate the pristine characteristics.

4.2 Device characteristics: pristine device and with homogenous interface trap distribution.

As before, the cases of pristine (without traps) and a device with homogenous distribution where all the available interface nodes are populated with traps are used as reference. For the considered structure (W x L=20nm x 50nm) a uniform trap density (D_{it}) of 5e11 cm^{-2} (same as that considered

in 2D simulations) was chosen to study the influence of interface traps on device performance. Figure 4-2 shows the I_DV_G plots for the pristine device and that if a device with uniform distribution of traps, under a drain bias of 0.3V.

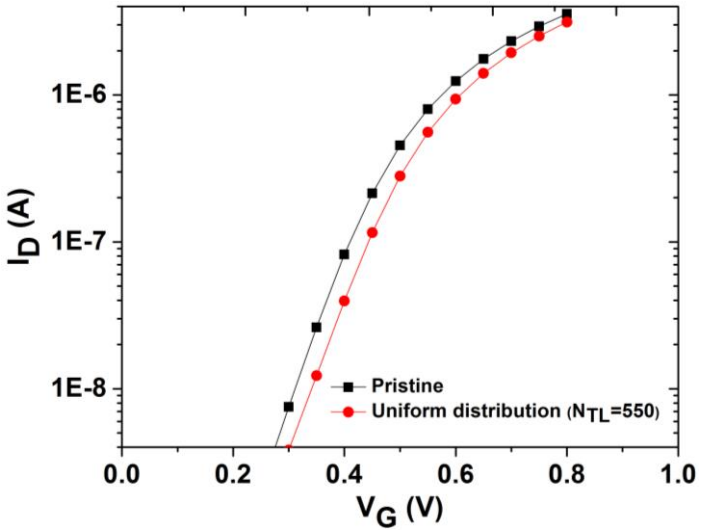


Figure 4-2: Simulated transfer characteristics in case of pristine device and when traps are uniformly (D_{it}=5e11 cm^{-2}) distributed throughout the channel.

The threshold voltages were obtained using the CC method with a fixed current of 1µA. From these plots, $V_{th,0}$ = 0.572V and $V_{th,H}$ = 0.606V were obtained for the pristine and uniform trap distribution cases, respectively. It can be seen that the introduction of interface traps (D_{it}) led to a ΔV_{th} of 34mV. Note that this shift is smaller than that observed in the case of 2D simulations (50mV) when same D_{it} (5e11 cm^{-2}) was used. This is because the device width in 2D was 1µm, which meant a device with a larger area than the device considered (here) for 3D simulations (width=20nm). Larger area meant more charge (traps) in the device (with the same D_{it}) hence larger V_{th}. The results corresponding to that of the discrete trap locations are discussed in the following sections.

4.3 Single trap location

In 2D simulations that were discussed earlier, the locations of the trap were only varied along the length. However, when performing 3D simulations, the trap locations can be varied along the length and along the width. As it was done earlier (in 2D simulations), we start out analysis with the simplest case of single trap location (N_{TL} = 1) and then move on the case of multiple trap locations. Since we are only considering a single trap location along the channel area, the influence of the trap location along the length of the device shouldn't be any different from the ones observed in the earlier chapter (see Figure 3-19). In order to verify this, simulations were done by varying the location of the trap only along the channel length and fixing the location along the channel width. Recall that the device has the dimensions of 50 X 20 (L X W) and as it was done earlier (in the case of 2D simulations) the value of interface trap density (D_{TL}) that is used to populate the selected number of trap location (N_{TL}, which in this case is 1) is obtained from the equation 3-1.

By substituting for N_{TL} (=1), D_{it} (=5E11) and N_{Tot} (=550), D_{TL} was obtained as 2.75e14 cm^{-2}. As an example, Figure 4-3 shows the plots of V_{th} (left) and ΔV_{th} (right) as a function of the trap location varied only along the length (from the source edge to the drain edge), with a fixed location along the width (which here was fixed at 6nm). As explained in section 3.2, ΔV_{th} is the difference in the V_{th} of a device with traps and without traps (pristine device). It can be seen from the plot that similar to the one observed in Figure 3-19, the maximum V_{th} / ΔV_{th} (0.591V / 0.0199V) happens when traps are located close to the channel center (in the shown case at 22nm).

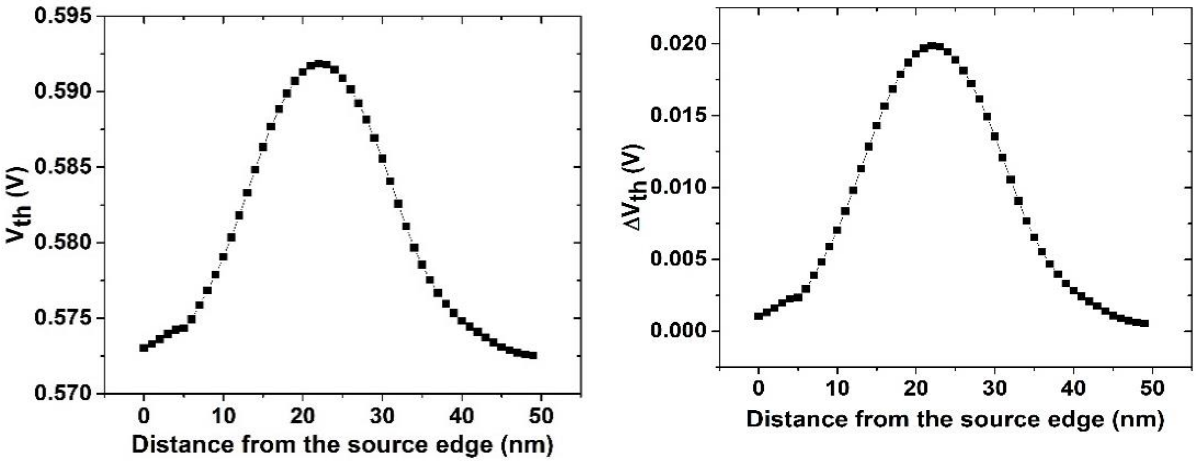


Figure 4-3: (left) V_{th} and (right) ΔV_{th} as a function of the location of traps along the channel length, for a fixed trap location at 6nm (measured from the device edge) along the width of a 20x50nm (W x L) device. Using equation 3-1 D_{TL} was found to be 2.75e14 cm^{-2}.

Now, to analyze the influence of the trap location along the width, we perform simulations by varying the location of the trap along the width, for a fixed position along the length. Again, as an example, Figure 4-4 shows V_{th} (left) and ΔV_{th} (right) as a function of the trap location along the channel width for a fixed position along the channel length (which was fixed here at 25nm from the source edge). The initial increase and the final decrease in the values of V_{th} (ΔV_{th}) as seen in Figure 4-4 are related to the border effect associated with the simulation mesh. Apart from the situations corresponding to the traps at the borders (where the increase/decrease is observed), the values of $V_{th}/\Delta V_{th}$ (0.591V/0.0187V) at any location along the width is identical. This implies that the location of trap along the width does not induce a remarkable change in the V_{th} or its shift (ΔV_{th}).

To explain this observation, the band profile of the device is investigated and is plotted in Figure 4-5. The figure shows the plot of the conduction band profile when traps are located at a single trap location (in this case at x = 22nm and z = 16nm) along the channel area.The CB (conduction band) profile was obtained at a gate voltage close to the threshold voltage in pristine case (0.572V) and drain bias of 0.3V. The plot in Figure 4-5 provides a visual aid into understanding the behavior that is observed in Figure 4-4. It has to be noted that unlike Figure 3-13 (b), where the barrier generated by the trap extends over the entire width, here the trap influence does not extend over the entire device width. Hence, contrarily to the 2D case, the generated barrier will not affect all carriers on their path to drain from source. Moreover, even those affected carriers can go around

the barrier along their path. The influence of the arbitrary trap locations along the 2D interface on ΔV_{th} can be clearly seen from the Figure 4-6 where the contour map of ΔV_{th} as a response to different positions of the trap location along the 2D channel area is represented. We can see from the plot that traps are very influential when they are located (along the channel length) close to the channel center, with smaller impacts when moving towards the source and drain edges. Their spatial variation along the channel width has no effect on the V_{th} (ΔV_{th}). From these results, we can conclude that when considering N_{TL} =1, V_{th} (ΔV_{th}) is only influenced by the trap location along the channel length and moreover those that are located close to the channel center are the most influential.

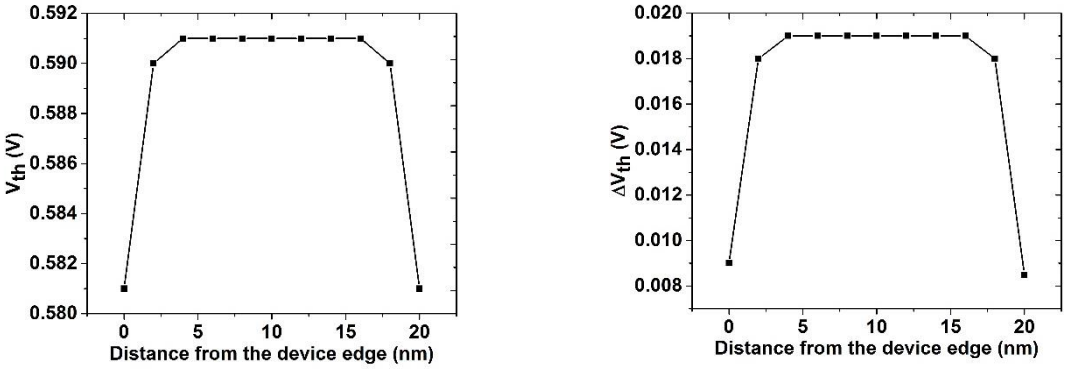


Figure 4-4: V_{th} (left) and ΔV_{th} (right) as a function of the location of traps along the channel width, for a fixed position along the channel length (25nm from the source edge) for a 20 x 50nm (W x L) device. It can be seen that the location of the trap along the width has no influence on V_{th} or its shift (ΔV_{th}).

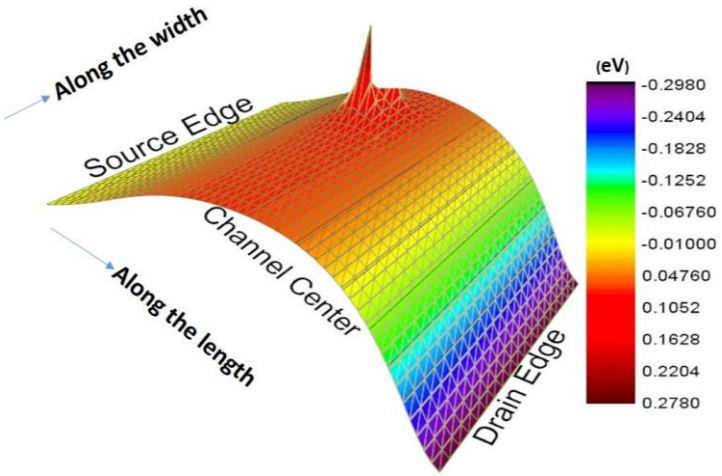


Figure 4-5: Conduction band profile with traps at single location (x = 22nm and z= 16nm) along the 2D interface.

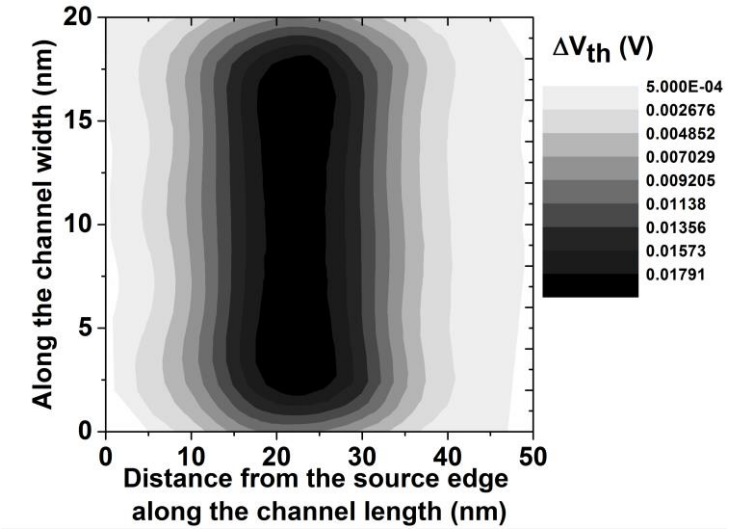


Figure 4-6: Contour map showing ΔVth as a function of the trap location along the 2D interface. It can be seen that traps are most effective when located along the channel center.

Now, to study the variability in the device performance (V_{th}) due to the variations in the positions of trap location along the 2D interface, device simulations of a set of 100 devices with the different (random) location for the trap site (region populated with traps) are performed. At the end of each simulation, V_{th} is extracted and ΔV_{th} is evaluated by calculating the difference from the pristine V_{th} (V_{th}^0 = 0.572V). Figure 4-7 plots the variability in V_{th} (left) and ΔV_{th} (right) for the 100 simulated devices.

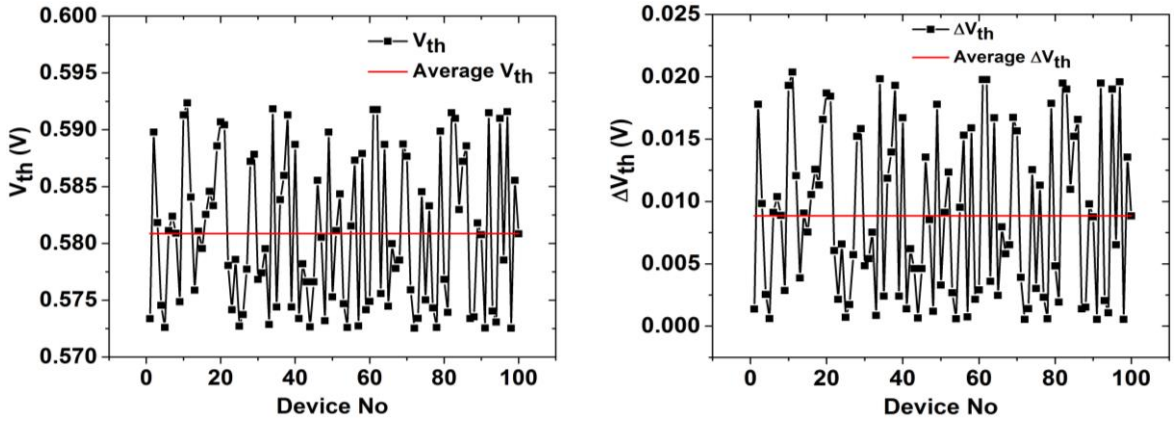


Figure 4-7: Variability in V_{th} (left) and in ΔV_{th} (right) due to the spatial variation of the trap location along the 2D interface across 100 simulated devices. Average V_{th} (0.581V) and average ΔV_{th} (8.87mV) are represented by red lines in the plots.

The plots in the figure (Figure 4-7) show a significant variability in V_{th} (left) and in ΔV_{th} (right) due to the location of traps (σV_{th} = 6. 5mV). From the results seen earlier (Figure 4-3, Figure 4-4 and Figure 4-6) we can state that this variability in ΔV_{th} is due to the variation of the trap location

along the channel length. Figure 4-7 also plots the average (horizontal red line) value of V_{th} and ΔV_{th} across the 100 simulated devices and it corresponds to 0.581V and 8.87mV respectively. This average V_{th} is about 25mV lower than the V_{th} corresponding to the device when the same number of traps were homogenously distributed along the entire interface. This result is interesting as in 2D simulations, we found that average V_{th} when traps were confined (to a single location) was higher than when the same traps were homogenously distributed (over the entire interface) (see Figure 3-22). This behavior is related to the difference in the way the simulator handles the influence of traps in 2D and 3D. As it was outlined earlier, in 2D simulations, the trap site is spread over the entire device width of 1µm. So, even though the number of traps in the trap site is same for 3D and 2D (this is maintained by the scaling of D_{TL}, in 2D, D_{TL} was 2.5e13 cm^{-2} and in 3D, D_{TL} was 2.75e14 cm^{-2}), these traps affect all carriers along the width in 2D, whereas in 3D they only affect the carriers along a confined region (both along the length and along the width).

In order to understand this better, we perform simulations to reproduce the plots in Figure 4-3 but with different values of D_{TL}. The resulting plots from these simulations are shown in Figure 4-8.

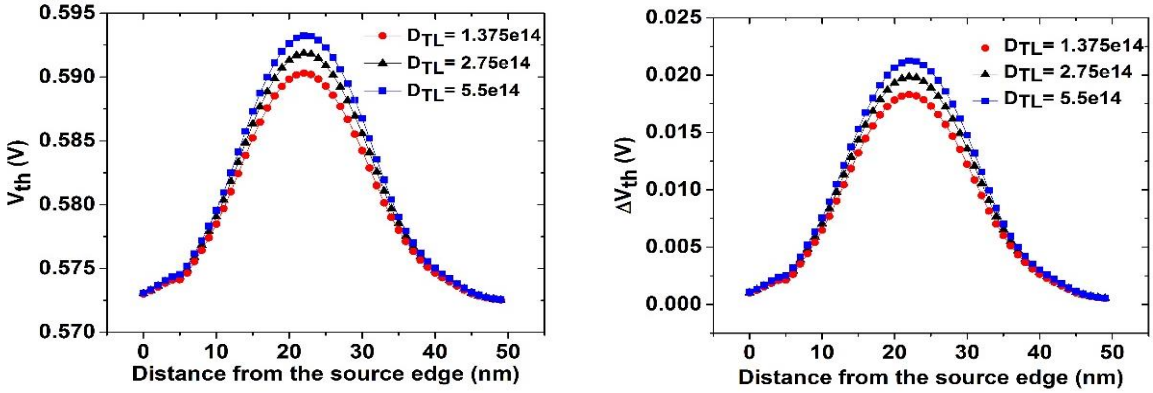


Figure 4-8: V_{th} (left) and ΔV_{th} (right) as a function of the location of traps along the channel length, for a fixed trap location of 6nm along the width and for different values of D_{TL}.

From these plots (in Figure 4-8) we can see that chosen range of D_{TL}'s does not modify the ΔV_{th} significantly. It can be seen that although D_{TL} is varied linearly (1.375e14, 2.75e14 and 5.5e14) the corresponding variation in V_{th} (ΔV_{th}) is non-linear. For example, an increase in value of D_{TL} by 4 times (from 1.375e14 cm^{-2} to 5.5e14 cm^{-2}) has only resulted in an increase of V_{th} by a factor of 1.005 (from 0.590V to 0.593V). Also, the modification is more pronounced when traps are located close to the channel center. These results help us in understanding how V_{th} (or ΔV_{th}) is modified with varying values of D_{TL}. This understanding would be crucial in analyzing the results pertaining to $N_{TL} > 1$, as D_{TL} is scaled with increasing N_{TL}. Next we move to the case of multiple trap locations ($N_{TL} > 1$) starting with the simplest among them, the case of $N_{TL} = 2$.

4.4 Two trap locations

In this case, the entire charge in the device is split and is placed at two different locations along the 2D interface. The density of traps at each trap location (D_{TL}) was obtained as 1.375e14 cm^{-2} from equation 3-1 (also shown below) by substituting for D_{it} (5e11 cm^{-2}), N_{Tot} (550) and N_{TL} (2).

$$D_{TL} = \frac{D_{it} * N_{Tot}}{N_{TL}}$$

In order to understand the effect of splitting the charge on the V_{th} (or ΔV_{th}) we start with special cases of non-random simulations and then move to random simulations. The non-random simulations are performed by considering traps at two different sites, say T1 and T2 along the 2D interface. Two different scenarios are considered, the location of T1 is fixed while the location of T2 is (a) varied along the channel length and (b) varied along the channel width. In the following sections, the results of these simulations are presented and discussed.

T1 is fixed and T2 is varied along the length

We start our analysis by considering a situation where T1 and T2 have the same fixed position along the width. However, the position of T2 is varied along the length while the position of T1 along the length is kept fixed. Simulations were repeated with three different fixed length positions for T1 (i) close to the source edge (ii) close to the channel center and (iii) close to the drain end of the channel. It has to be noted that for all the simulated cases, the position of both T1 and T2 along the channel width was fixed at 6nm (from the device edge). Figure 4-9 (left) plots the V_{th} and (right) plots the ΔV_{th} as a function of the location of the mobile trap site (T2) for three different locations of T1 (5nm, 25nm and 36nm). These locations are close to the source end, close to the channel center and close to the drain end respectively. Several different observations can be made from the plots shown in Figure 4-9 we will outline and discuss the most notable ones.

V_{th}'s (or ΔV_{th}'s) corresponding to the situation when T1 is fixed (along the length) at a position closer to the source/drain edge are much smaller than the V_{th}'s (or ΔV_{th}'s) compared to the situation when T1 is fixed (along the length) at a position closer to the channel center. This observation can be understood by taking into context the results seen earlier, where it was shown that location of the trap (along the channel length) has an influence on the V_{th} of the device (see Figure 3-20 and Figure 4-3). We have also seen (Figure 3-19) that this dependence of the V_{th} on the trap location (along the length) is due to the height of the additional barrier created by the traps at the respective location, being more significant when traps were located at sites close to the channel center and less significant when traps were located at sites close to the channel edges. Based on these, it is reasonable that in Figure 4-9, we obtain a lower overall V_{th} (due to traps at T1 and T2) when T1 is fixed at locations close to the channel edges as compared to the situation where T1 is fixed at sites located close to the channel center. We can also observe from Figure 4-9 that the maximum V_{th} (ΔV_{th}) happens when the position of either of the two locations are close to the channel center (so long as they are not too close to each other). This behavior is also understandable as traps when located close to the channel center, assert a significant influence on the threshold voltage.

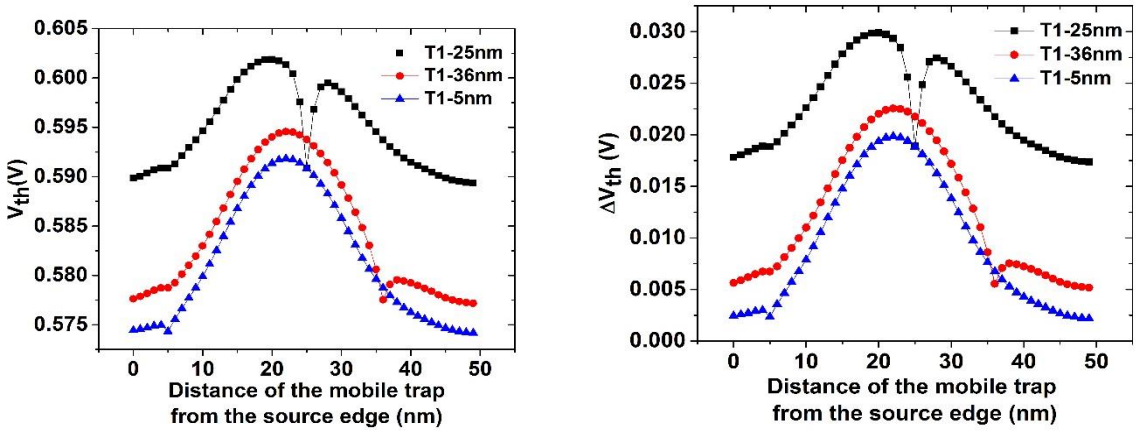


Figure 4-9: The plots correspond to (left) V_{th} and (right) ΔV_{th} as a function of the position of T2 for three different positions of T1. T1 was fixed at a location close to the source end (5nm), close to the channel center (25nm) and close to the drain end of the channel (36nm). It is interesting to see that when traps are close enough to each other there is a reduction in V_{th} (ΔV_{th}).

The other interesting behavior that can be observed (from Figure 4-9) is that the V_{th} (ΔV_{th}) tends to decrease when T1 and T2 are located close enough to each other and, this behavior is observed for all the fixed positions of T1 (5nm, 25nm and 36nm). In order to understand this phenomenon, we investigate the band profiles of the device corresponding to particular positions of T1 and T2. It should be noted that at both trap sites (T1 and T2), the value of D_{TL} is kept the same (1.375e14 cm^{-2}). For our analysis, we consider three different positions of T2 while the position of T1 is fixed at 25nm along the length. The corresponding plots of the CB profile are shown Figure 4-10 In the figure, the plot corresponding to the CB edge for the device with no traps (fresh/pristine) is also overlaid. It has to be noted that these profiles were obtained at the gate voltage of 0.75V and a drain bias of 0.3V. In the figure, the top (left) plot corresponds to the situation when T2 is located at 3nm (from the source edge), top (right) corresponds to the situation when T2 is located at 15nm and bottom corresponds to the situation when T2 is located at 26nm. The shaded portion in the plots represents the additional barrier area created by the traps. The total effective barrier area for the case of plots at the top left was calculated to be ≈ 0.0173eV-nm, for the plots at the top right it was calculated to be ≈ 0.0187eV-nm and for the plots at the bottom it was calculated to be ≈ 0.0177eV-nm. From these values, we can see that the effective barrier area is lower when (a) the traps sites are very far apart (either or both of them being located at the channel edges) or (b) the trap sites are very close to each other. The decrease of the V_{th} (ΔV_{th}) as observed in Figure 4-9 can be correlated to the decrease of this effective barrier area. Based on this, it can be stated that the overall V_{th} due to both trap locations (T1 and T2) depends upon the effective barrier area created by traps (at these two locations), which in-turn depends upon the location traps along the channel length and the separation between them.

It was stated earlier that, when moving from N_{TL} = 1 to N_{TL} = 2, we scaled the value of D_{TL} by 0.5. This essentially implies that charge at a single location (when N_{TL} = 1) is split and spread across two different locations (when N_{TL} = 2). So, it would be interesting to observe how the splitting of the charge has influenced the V_{th}. In order to do this, we overlay the plots in Figure 4-9 (left, traps

at two different locations) with the plots in Figure 4-3 (left, traps only at single location) and this overlap is shown in Figure 4-11. In other words, we are overlapping the plots of V_{th} vs trap location corresponding to the cases of N_{TL} = 1 (with D_{TL} = 2.75e14 cm^{-2}) and N_{TL} = 2 (with D_{TL} = 1.375e14 cm^{-2} at each location). Note that the x-axis in the plot corresponds to the position of the location that is varied, so in case of N_{TL} = 2, this would correspond to the position of T2. This plot can help us visualize the effect on V_{th} as a result of splitting the charge and spreading it across different locations (in this case 2). In the figure, the plot corresponding to 'ST' represents V_{th} as function of trap location when N_{TL} = 1.

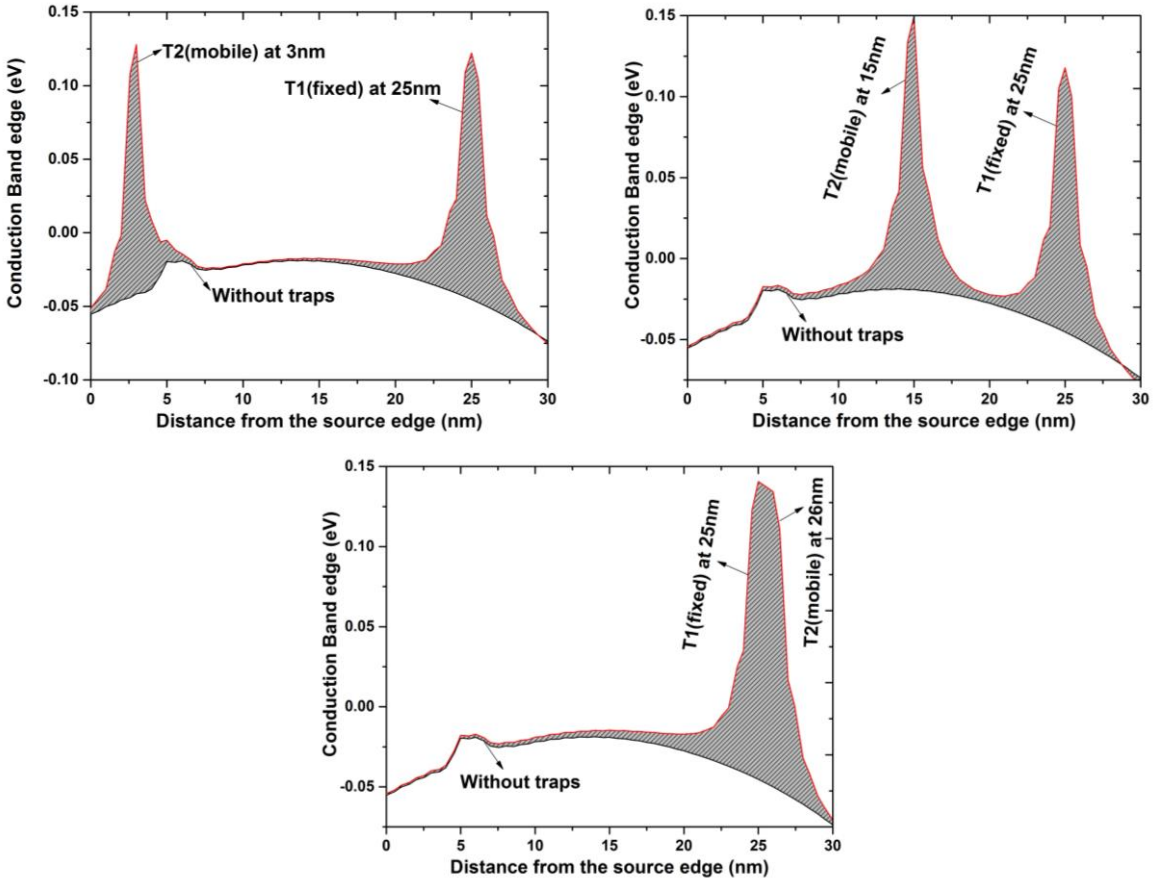


Figure 4-10: Profiles of the CB edges along the interface corresponding to cases when the position of T2 at 3nm (top left), 15nm (top right) and 26nm (bottom), all the while the position of T1 is fixed at 25nm. The CB edge corresponding to the device with no traps is also overlaid to have a clear picture of the barrier. As T1 and T2 get closer to each other, there is a decrease in the barrier area (shaded region) thus resulting in a decrease of V_{th} (ΔV_{th}). For all the cases, D_{TL} = 1.375e14 cm^2 is kept at both T1 and T2.

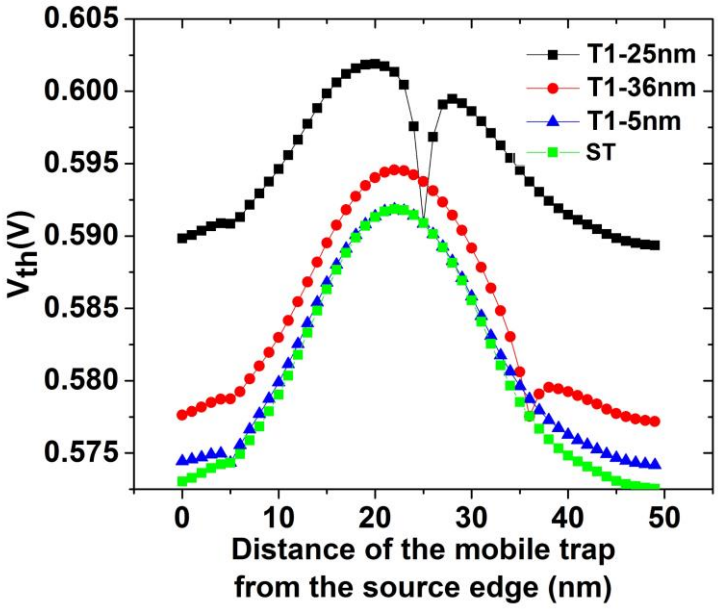


Figure 4-11: Comparison of V_{th} due to traps at two locations (Figure 4-9, left) with the V_{th} when the same number of traps were located at a single location (Figure 4-3, left). It can be seen that the V_{th} corresponding to the situation when the charge is split (D_{TL} = 1.375e14 cm^{-2} at each location) is always greater than or equal to the V_{th} corresponding to situation when the entire charge was confined at one location (D_{TL} = 2.75e14 cm^{-2} at each location).

From Figure 4-11, we can observe that the V_{th} when charge is split and spread at two separate locations is always greater than or equal to the V_{th} when the entire charge is confined at one trap site. Note that when the positions of T1 and T2 are same (both along the length and along the width) then the resulting V_{th} is same as that in the case of single trap location. In order to understand this behavior, Figure 4-12 (bottom) shows the CB profile with the position of T1 (with D_{TL} = 1.375e14 cm^{-2}) fixed at 25nm (channel center) and the position of T2 (also with D_{TL} = 1.375e14 cm^{-2}) located at 3nm (closer to the source edge), and compares this plot with the CB profile when traps (with D_{TL} = 2.75e14 cm^{-2}) were individually located only at 3nm (top left) and only at 25nm (top right). For comparison, the plot corresponding to the pristine device (without traps) is also included. The shaded portion in the plots corresponds to the additional barrier area created due to the traps. In the case of the plot at the top left (traps only at 3nm) the barrier area was calculated to be ≈ 0.012eV-nm, for the plots at the top right (traps only at 25nm), the total barrier area was calculated to be ≈ 0.0135eV-nm and for the plots at the bottom (traps are split across 3nm and 25nm), the total barrier area was calculated to be ≈ 0.0173eV-nm. From these results, we can see that when traps were located simultaneously at T1 and T2 then the total barrier area is more than when they were located individually (at either only T1 or only T2). This increased barrier area results in an increased V_{th} (ΔV_{th}).

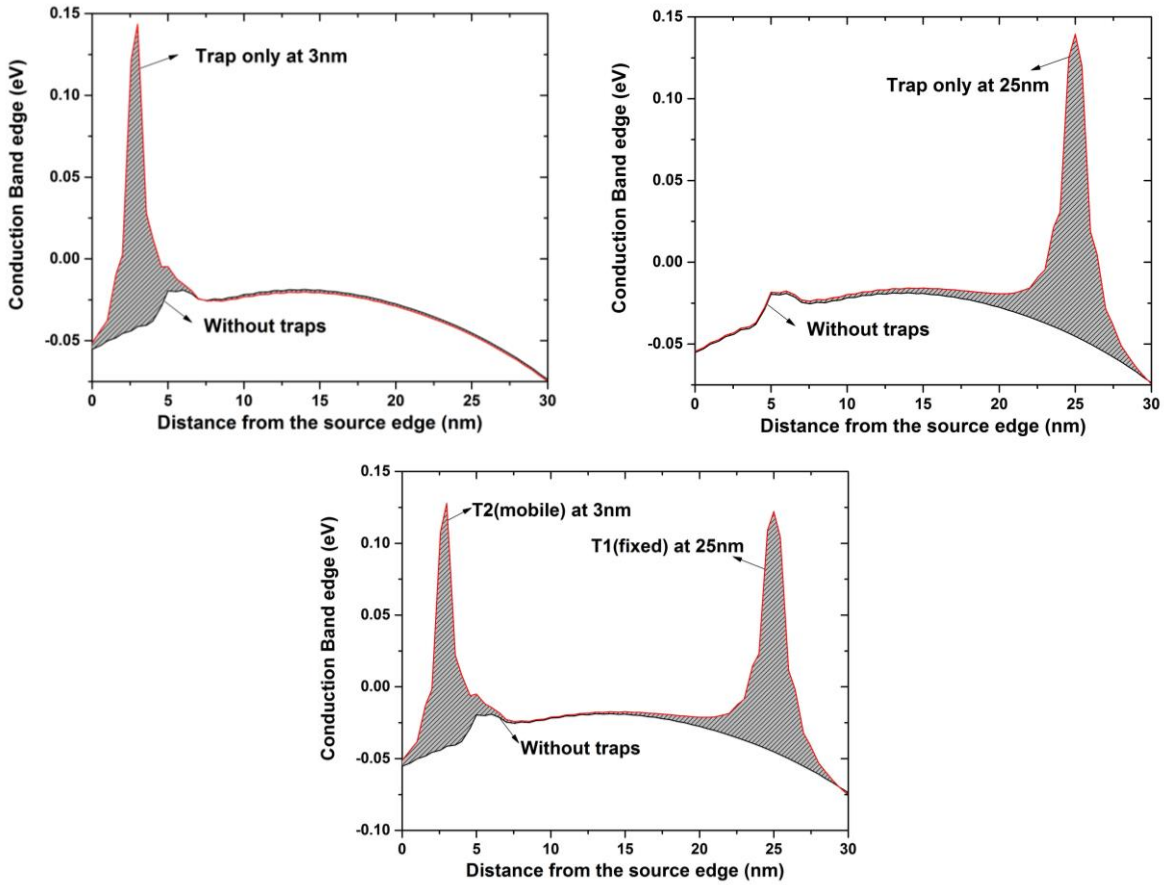


Figure 4-12: Top (left) corresponds to the CB edge when trap is only located at 3nm (closer to source edge) and top (right) corresponds to the CB edge when trap is only located at 25nm (channel center) from the source edge. The plot in the bottom corresponds to the CB edge when traps were simultaneously placed at two different trap locations, T1 (at 25nm) and T2 (at 3nm). The shaded area in the plots represents the additional barrier area created by the included traps, also shown is the CB edge corresponding to the device without traps (pristine). When traps were located simultaneously at T1 and T2 then the total barrier area is more than when they were located individually resulting in an increased V_{th} (ΔV_{th}).

On the other hand, we would also like to see if having traps simultaneously present at two different locations would result in a ΔV_{th} that is superposition of ΔV_{th} when traps were considered at these locations individually. To make this comparison, same density (D_{TL} = 1.375e14 cm^{-2} at each location) is used if traps were placed at two locations or at a single location. Figure 4-13 compares the two cases, in the first case ((ΔV_{th} (T1, T2), green circles) traps are placed at two different locations (T1 and T2, both with D_{TL} = 1.375e14 cm^{-2}) with the position of T1 is fixed at 25nm along the length and the position of T2 is mobile along the entire channel length. In the other case (ΔV_{th} (T1), red squares), we consider traps only at a single location (T1 with D_{TL} = 1.375e14 cm^{2}) that is mobile along the entire channel length. In both the cases, the position of the trap locations along the channel width is fixed at 6nm. In order to check for the superposition, a data point from the green (circles) curve is selected and compared with the corresponding data points from the red (squares) curve. For example, from the green curve a ΔV_{th} (0.019V) corresponding to the position of T2 at 5nm is considered. So as T1 is fixed at 25nm, ΔV_{th} (25,5) = 0.019V, now this ΔV_{th} is

compared to the sum of the ΔV_{th} corresponding to locations 5nm (0.002V) and 25nm (0.017V) from the red curve. From this we can observe two things (a) ΔV_{th} (T1, T2) follows the superposition relation and (b) the trap closer to channel center (25nm) provides the significant contribution towards the total ΔV_{th} (i.e., **ΔV_{th} (T1, T2) = ΔV_{th} (T1) + ΔV_{th} (T2)**). Now, to confirm this trend, we consider a different set of data points from the red and green curves, and this time we consider a ΔV_{th} (0.028V) corresponding to the position of T2 at 15nm from the green curve. This value is compared to the ΔV_{th} corresponding to locations 15nm (0.013V) and 25nm (0.017) on the red curve. We can observe from this that (a) ΔV_{th} (T1, T2) is less than the superposition (**ΔV_{th} (T1, T2) < ΔV_{th} (T1) + ΔV_{th} (T2)**) and (b) the contributions of traps at both the locations are equally significant toward the total ΔV_{th}. Based on these observations we can state that ΔV_{th} due to the simultaneous presence of traps at T1 and T2, follows

$$\Delta V_{th}\,(T1, T2) \leq \Delta V_{th}\,(T1) + \Delta V_{th}\,(T2)$$

This observation can be interpreted from the maps of the effective barrier area seen earlier (Figure 4-10 and Figure 4-12), it can be stated that when T1 and T2 are located close enough to each other then it results in lower effective barrier area, which leads to a total ΔV_{th} that is lower than the superposition of ΔV_{th} when traps were individually located at these locations. However, the farther they are from each other, the total ΔV_{th} is almost equal to the value of superposition of individual ΔV_{th}'s.

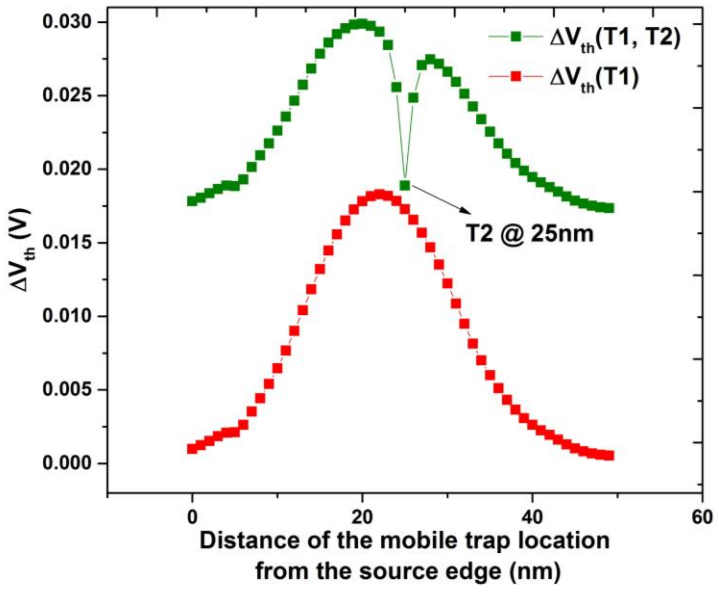


Figure 4-13: (green circles, top) shows the ΔV_{th} with traps simultaneously present at two locations (with identical D_{TL} of 1.375e14 at each location) with T1 fixed at 25nm and T2 mobile along the channel length. (red squares) shows the ΔV_{th} with traps located only at single location (that is mobile along the channel length) but with a D_{TL} = 1.375e14. In both the cases the positon of the trap locations along the channel width was fixed at 6nm.

T1 is fixed and T2 is varied along the width

In order to investigate the behavior of having traps spread at two different locations along the channel width, we consider the second scenario where the position of T1 is fixed along the length and along the width, and the position of T2 is fixed along the length but varied along the channel width. Note that, as it was done earlier, at each location (T1 and T2) identical value of D_{TL} (1.375e14 cm^{-2}) is used. Figure 4-14 plots the V_{th} (left) and ΔV_{th} (right) corresponding to the situation where the position of T1 was fixed at z=6nm (along the width and measured from the device edge) and x =25nm (along the length, measured from the source edge) while the position of T2 is varied along the channel width with a fixed position along the channel length at x=25nm. For comparison, the plots also show the V_{th} (identified by the label 'V_{th}-ST') when there is only one trap (with double D_{TL} = 2.7e14 cm^{-2}) located at different z's. The plots on the right present the same relationship in terms of ΔV_{th}.

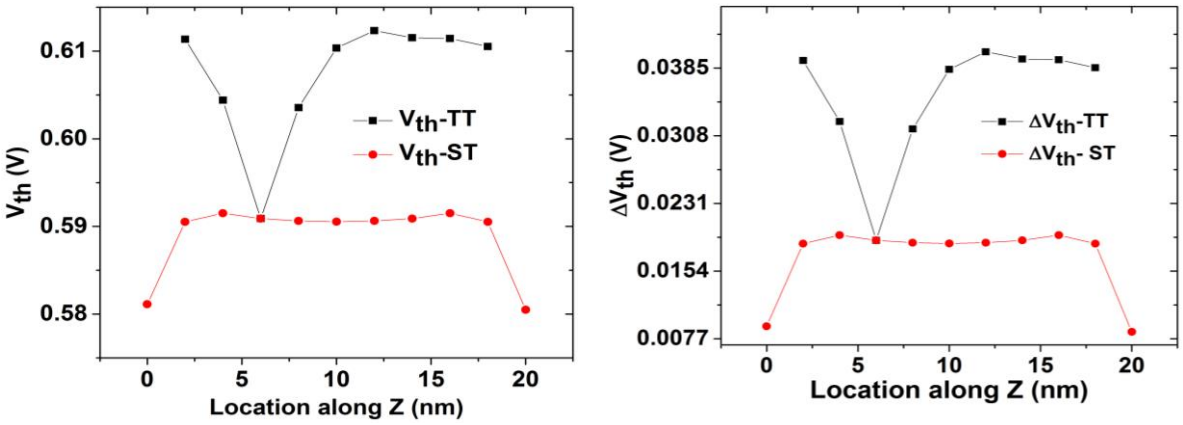


Figure 4-14: (left) V_{th} and (right) ΔV_{th} as a function of the position of mobile trap location (T2 along the channel width in case of 2 trap locations (TT)). In case of two trap locations (TT), the position of T2 was varied only along the width while the position of both T1 and T2 along the length was fixed at 25nm, also the position of T1 along the z was fixed at 6nm. It can be seen from the plots that similar to that seen in Figure 4-9 the splitting of charge has resulted in an increment of V_{th} (ΔV_{th}). It can also be seen that when T1 and T2 are close enough to each other then, it results in lower values of V_{th} and ΔV_{th}.

From the plots, we can observe that similar to one seen Figure 4-11, splitting the charge has resulted in an increase of V_{th} (ΔV_{th}) when N_{TL} is increased from 1 to 2. We also observe that V_{th} (ΔV_{th}) is smaller when T1 and T2 are close enough as compared to when they are further apart. This observation is also similar to the one when trap location (see Figure 4-11) was varied only along the channel length. This increase/decrease of ΔV_{th} can be understood on the same lines with the increase/decrease of the effective barrier area from the plots of the CB cuts profile as shown in Figure 4-15. In the figure, the plot at the top shows the band profiles corresponding to the situation where z-location of T1 is 6nm and z-location of T2 is 4nm (one node to the left of T1). The plot at the bottom shows the CB profile along the 2D interface corresponding to the situation where the z-location of T1 is 6nm and the z-location of T2 is 12nm (six nodes to the right of T2). In both the cases, the x-location of the both T1 and T2 is fixed at 25nm. The band profiles were obtained for the gate bias of 0.75V and drain bias of 0.3V. The shaded graph on the right of each 3D plot corresponds to the effective barrier area as a result of T1 and T2.

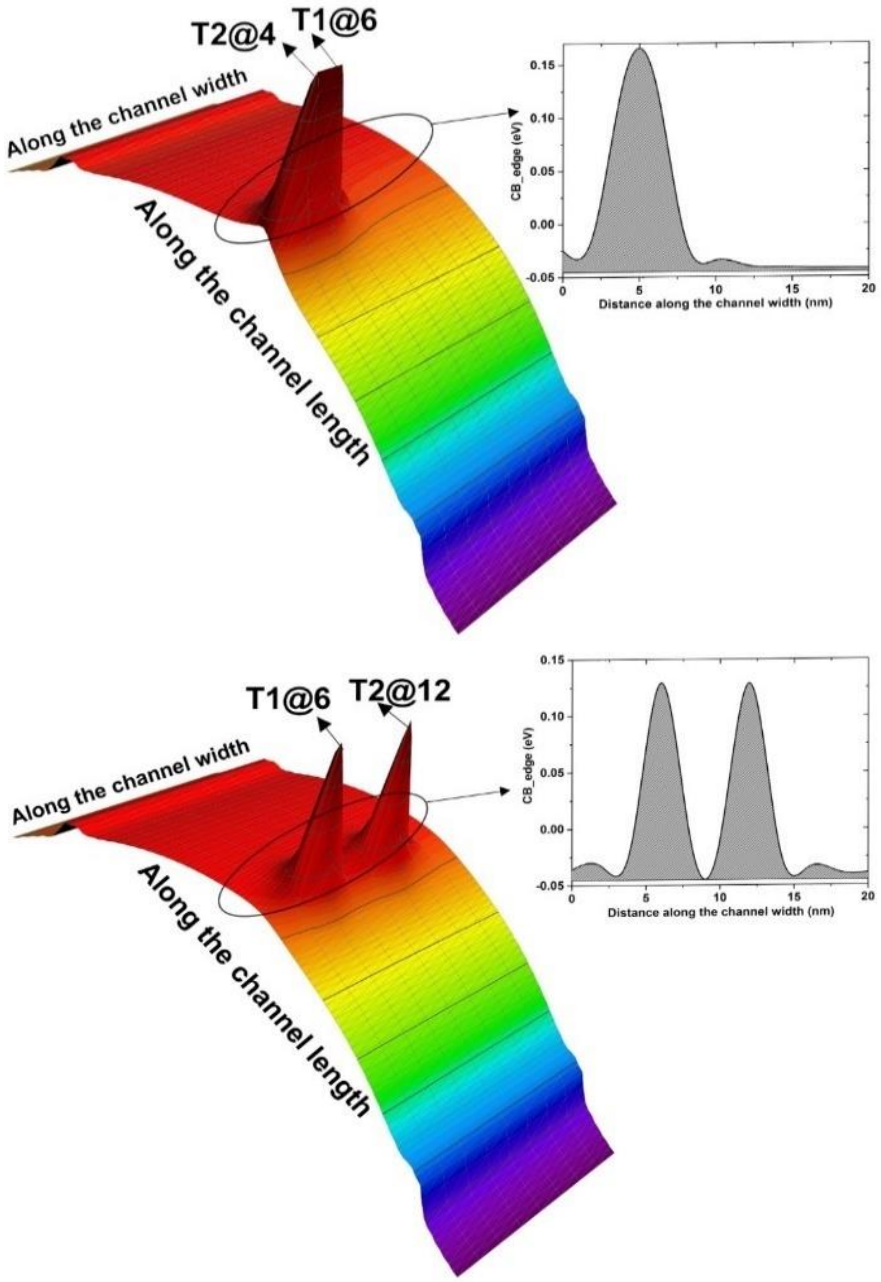


Figure 4-15: The CB profile along the 2D interface shows (top) the barriers created by traps at T1=6nm and T2=4nm and (bottom) the barriers created by traps at T1=6nm and T2=12nm. The shaded regions on the right of each plot corresponds to the effective barrier area created due to T1 and T2. When the locations are sufficiently separated, they result in an increased barrier area and in-turn results in an increased V_{th} (ΔV_{th}) as seen from Figure 4-14.

In the considered example, the effective barrier area for the case corresponding to the plot at the top (Figure 4-15) was calculated to be ≈ 0.921eV-nm and for the case corresponding to the plot at the bottom (Figure 4-15) was calculated to be ≈ 0.963eV-nm. It can be seen that the increased

barrier area resulted in an increased threshold voltage (V_{th}) or its shift (ΔV_{th}), when the locations (T1 and T2) are sufficiently separated further apart.

Once again, we would like to see if having traps simultaneously present at two different locations would result in a ΔV_{th} that is superposition of ΔV_{th} when traps were considered at these locations individually. To make this comparison, same density (D_{TL} = 1.375e14 cm^{-2} at each location) is used if traps were placed at two locations or at a single location. Figure 4-16 compares the two cases, in the first case ((ΔV_{th} (T1, T2), green circles) traps are placed at two different locations (T1 and T2, both with D_{TL} = 1.375e14 cm^{-2}) with the position of T1 is fixed at 6nm along the width and the position of T2 is mobile along the entire channel width. In the other case ((ΔV_{th} (T1), blue squares), we consider traps only at a single location (T1 with D_{TL} = 1.375e14 cm^{-2}) that is mobile along the entire channel width. In both the cases, the position of the trap locations along the channel length is fixed at 25nm.

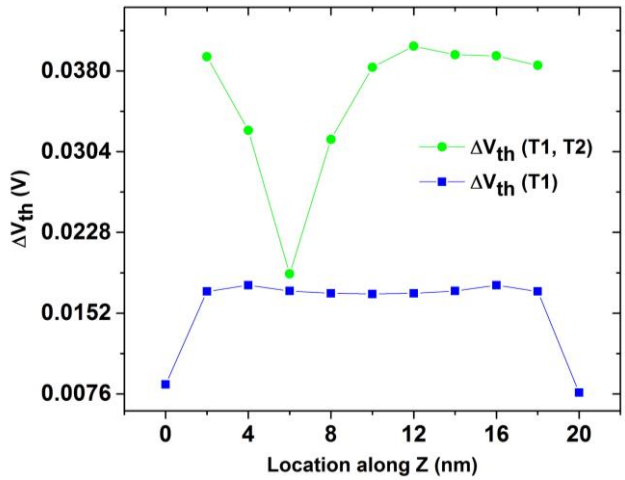


Figure 4-16: (green circles) shows the ΔV_{th} with traps simultaneously present at two locations (with identical D_{TL} of 1.375e14 at each location) with T1 fixed at 6nm and T2 mobile along the channel width. (blue squares) shows the ΔV_{th} with traps located only at single location (that is mobile along the channel width) but with a D_{TL} = 1.375e14. In both the cases the positon of the trap locations along the channel length was fixed at 25nm.

Upon investigating the values in Figure 4-16, we can observe that in cases when T1 and T2 are far apart, the overall ΔV_{th} due to traps at T1 and T2 follows ΔV_{th} (T1, T2) > ΔV_{th} (T1) + ΔV_{th} (T2), while in some other cases (when the locations are close enough to each other) it follows ΔV_{th} (T1, T2) < ΔV_{th} (T1) + ΔV_{th} (T2). This can be understood by looking at the CB profile seen in Figure 4-5 where we can see that in the absence of the traps the band profile is uniform along the channel width. This is understandable as the drain bias along the channel width is always the same. Also from the curve corresponding to ΔV_{th} (T1) we can see that the influence of the traps on the ΔV_{th} is always the same no matter where they are located along the channel width. Hence, when considering traps at 2 locations along the channel width, the overall V_{th} (ΔV_{th}) depends on the additional effective barrier area, which (as seen from Figure 4-10, Figure 4-12 and Figure 4-15) depends on the distance between the trap locations.

We can see (from Figure 4-16) that when the trap locations are closer to each other, the overall ΔV_{th} is less than the super position of the individual (**ΔV_{th} (T1, T2) < ΔV_{th} (T1) + ΔV_{th} (T2)**), however, as the distance between the traps locations increases, the overall ΔV_{th} starts to increase eventually becoming greater than superposition (**ΔV_{th} (T1, T2) > ΔV_{th} (T1) + ΔV_{th} (T2)**). The value of the overall ΔV_{th} reaches a maximum value at a critical separation between the trap locations (which in this case is 6nm) and then starts to decrease (but still greater than the superposition). For example, from the Figure 4-16 we can see that when considering traps simultaneously placed at 2nm (T2) and 6nm (T1), the overall 'ΔV_{th} (2,6)' is 0.039V which is greater than the sum of the ΔV_{th} (0.172V at 2 and 0.172V at 6, the sum being 0.0344V) when traps were located individually located at respective locations. And when considering the traps simultaneously placed at 12nm (T1) and 6nm (T2), the overall 'ΔV_{th} (12,6)' is 0.040V, which is the maximum value of ΔV_{th}. However, as the locations are separated further apart, the values of overall ΔV_{th} tends to slightly decrease. We have seen from Figure 4-15 that this behavior is related to the effective barrier area created by the traps and this area becomes almost constant after a particular separation between the trap locations. Thus we observe the value of overall ΔV_{th} to be almost saturated to the superposition (**ΔV_{th} (T1, T2) ≈ ΔV_{th} (T1) + ΔV_{th} (T2)**) for larger separations.

Up until now we only discussed the situations where traps were either aligned along the length or along the width. However, in reality, they can be randomly located at different angles to each other. Figure 4-17 shows a contour plot of V_{th} corresponding to the different locations (along the length and along the width) of T2 with respect to T1, which is fixed at 25nm (from the source edge) along the channel length and 2nm along the channel width.

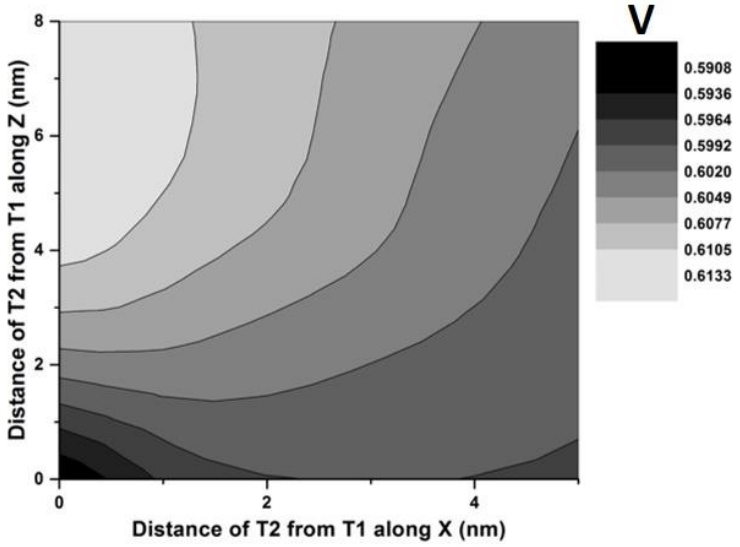


Figure 4-17: The figure shows the contour plot of the V_{th} corresponding to different positions of T2 (both along the channel length and channel width) with the position of T1 being fixed at 25nm (from the source edge) along the channel length and 2nm along the channel width.

Due to the symmetrical nature of the structure, the results in the figure only correspond to the locations along the channel center (area of significant influence), where T2 is moved along the

channel length from x = 25nm (channel center) to x = 30nm and along the channel width from z = 0nm (device edge) to z = 10nm (channel center along the width). We can see from the figure that the overall V_{th} depends on both the relative separation of the traps and its position along the 2D interface. From the results presented so far, we have seen that the location of the trap along the length determines the additional barrier height which in combination with the separation between the trap locations determines the effective barrier area that in-turn influences the V_{th}. However, the height of the additional barrier created by traps along the width is always the same irrespective of their location. But the separation between the trap locations has an influence on the effective barrier area that in-turn has an effect on the overall V_{th}. Thus, it can be stated that the overall effect of traps along the interface on V_{th} (or ΔV_{th}) is the combination of their individual locations (only along the length) and the effective barrier area (which depends on the relative distance between them) created by them.

V_{th} Variability

The variability in V_{th} was evaluated by performing simulations of a set of 100 devices with different combinations for the locations T1 and T2. Figure 4-18 shows the plot of V_{th} (left) and ΔV_{th} (right) across 100 different simulated devices.

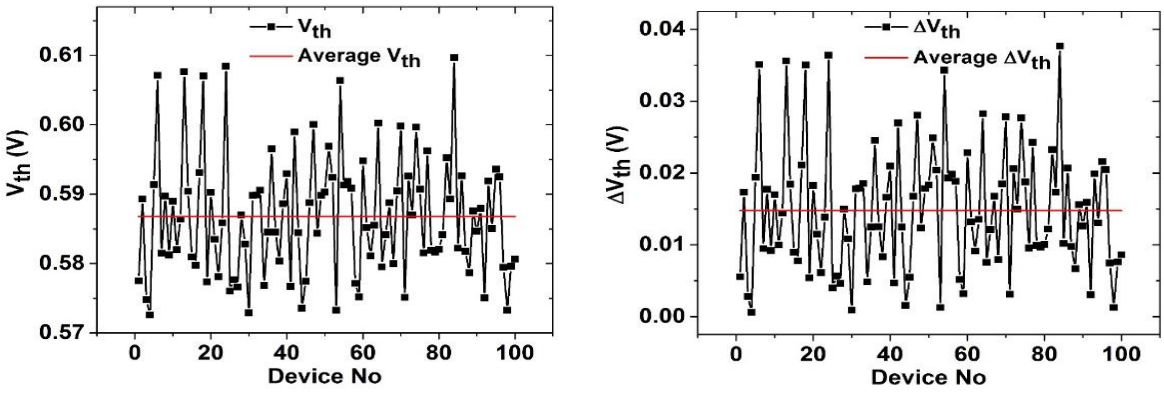


Figure 4-18: Variability in V_{th} (left) and in ΔV_{th} (right) due to different positions of the trap locations (T1 and T2) along the 2D interface across 100 simulated devices. The average value is marked by the red line.

The figure shows the average (horizontal red line) value of V_{th} and ΔV_{th} across the 100 simulated devices and they correspond to 0.587V and 14.75mV respectively. It can be seen that there is an increase in the average values of V_{th} and ΔV_{th} for the case N_{TL}=2 as compared to the case of N_{TL} = 1. These are coherent with the results shown earlier (Figure 4-11 and Figure 4-16) wherein V_{th} increased when N_{TL} is increased from 1 to 2. Figure 4-18 also shows the variability (σV_{th} = 8.72mV) in the V_{th} (left) and variability in the ΔV_{th} (right) due to the spatial location of traps. The increase (compared to 6.5mV in the case of N_{TL}=1) in the value of sigma (σ) can be attributed to the probability of having traps at additional sites along the length and width. With this understanding, we now proceed to study the variability in V_{th} (ΔV_{th}) due to multiple values of N_{TL} (N_{TL} > 2).

4.5 Multiple trap locations

The influence of multiple trap locations along the entire channel area is studied for several different values of N_{TL}. D_{TL} is obtained from the equation 3-1 by substituting the values of N_{TL} and D_{it}. Variability is analyzed by performing 100 simulations for each value of N_{TL} with the random spatial distribution of the trap locations over the entire device area. As an example, Figure 4-19 shows the variability in the V_{th} (left) and ΔV_{th} (right) corresponding to N_{TL} = 5. Also shown are the corresponding values of average V_{th} (0.602V, red line) and average ΔV_{th} (29.76mV, red line). Also marked (by arrows) in the figures are the maximum and minimum values (of V_{th} and ΔV_{th}) corresponding to N_{TL}=5.

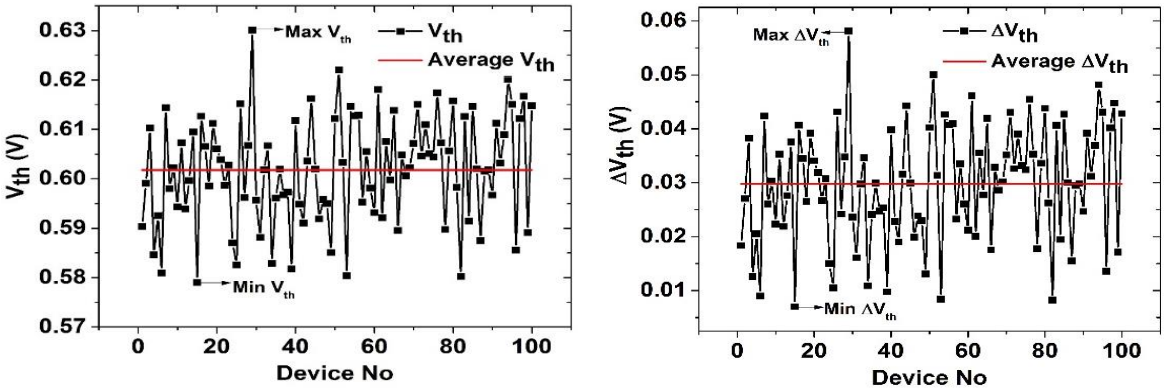


Figure 4-19: Variability in V_{th} (left) and in ΔV_{th} (right) due to the spatial variation of the trap locations along the 2D interface across 100 simulated devices for N_{TL} = 5. Also marked in the figure are the corresponding maximum and minimum values among the 100 simulated devices.

From Figure 4-19 we obtain the values of minimum V_{th} as 0.579V (min ΔV_{th} =7mV) and maximum V_{th} as 0.630V (max ΔV_{th} =58.1mV) corresponding to N_{TL}=5. As seen in the earlier cases (N_{TL}=1, 2), the average V_{th} (ΔV_{th}) has increased with increasing N_{TL}. We have already seen from earlier results for N_{TL}=1 and N_{TL}=2 that when traps are located in the center of the channel then they result in larger values of V_{th} (ΔV_{th}). However, when traps are located along the channel edges then they are not very influential and hence result in smaller changes in the V_{th}. Moreover, we have also seen that when traps are very close to each other then, the resulting V_{th} is lower. So in order to investigate what distributions of traps lead to the marked (in Figure 4-19) maximum and minimum values of V_{th} (or ΔV_{th}), we look at the corresponding spatial distribution of traps. Figure 4-20 plots the spatial distribution corresponding to these minimum and maximum values. In Figure 4-20 the plot on the left represents the spatial distribution of the traps that result in the minimum V_{th} and the plot on the right corresponds to the spatial distribution of the traps that lead to maximum V_{th}. From the maps, it is clear that max V_{th} is obtained when most of the traps are located close to the channel center and minimum V_{th} happens when most of the traps are located close to the channel edges. Similarly, the variability in ΔV_{th} across the 100 simulated devices for N_{TL}=10 is also evaluated. The corresponding values for average V_{th} and average ΔV_{th} were found to be 0.613V and 43.18mV respectively. Again we observe an increase in the values of average V_{th} and average ΔV_{th} as compared to those corresponding to lower values of N_{TL}.

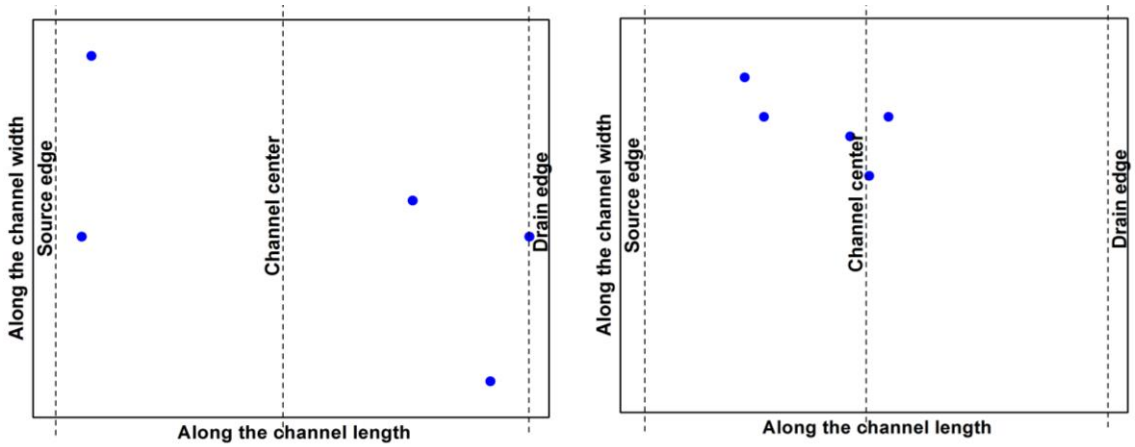


Figure 4-20: The map of the spatial distribution of the traps that lead to min V_{th} (left) and max V_{th} (right) for the corresponding N_{TL} of 5. It can be seen that the max V_{th} happens when more traps are located close to the channel center while minimum V_{th} happens when more traps are located close to the device edges.

In order to study the dependence of N_{TL} on the variability of V_{th}, simulations were done with various values for N_{TL} and, for each value of N_{TL}, the average value of V_{th} and its standard deviation (σV_{th}) from the simulations of 100 devices with varying spatial distribution of trap locations were evaluated (D_{TL} was scaled accordingly). Figure 4-21 shows the average V_{th} and standard deviation (σ) in V_{th} as a function of the number of trap locations (N_{TL}). Once again the extreme cases of single trap location (N_{TL}=1) and that of uniform distribution (N_{TL}=550) are also included. Two different behaviors for both σV_{th} and average V_{th} can be observed in the graph and these behaviors depend on the value of N_{TL} in the device. The standard deviation in V_{th} initially increases but after a certain N_{TL} (>11) it starts to decrease, eventually reaching zero for homogenous distribution (N_{TL} = 550). The behavior of the decrease of σV_{th} with increasing N_{TL} is also seen in 2D simulations (see Figure 3-33) however, the increase for small enough values of N_{TL} is unique to 3D simulations. Similar behavior can be observed in the average V_{th}, however unlike in 2D, after a certain N_{TL} ($\approx$>50) the average V_{th} only decreases slightly (is almost saturated).

So as to understand these behaviors, the two parameters (average V_{th} and σV_{th}) are analyzed separately starting with the σV_{th}. From the point of variability, in the 3D simulations considered, two sources of variability can be identified: (a) trap locations along the length and (b) trap locations along the width. This is the fundamental difference compared to 2D simulations, where the only source of variability was location of the traps along the channel length. Now, we have to recall that the simulated structure has a width of 20nm and based on the mesh used, we have 11 possible unique locations along the width. So for a large enough sample size (100 simulations), beyond N_{TL}=11, we would not expect any additional variability due to the width. Thus, beyond this value (N_{TL} > 11) the only source of variability is the location of traps along the channel length. And the shape of the curve of σV_{th} in Figure 4-21 would be similar to the one observed in 2D simulations (Figure 3-33). For all values of N_{TL} < 10, with each increase of N_{TL} (from 1 to 10) we are adding an additional variability in the form of trap locations along the width. Thus, we would expect the curve of σV_{th} to increase up until N_{TL} < 11 and this is reflected in Figure 4-21.

Now, let's examine the other magnitude in Figure 4-21 which is the average V_{th}. We have seen from our earlier results that as we increase the value of N_{TL}, irrespective of their spatial distribution (along length and width) there would be an increase in the threshold voltage due to increased effective barrier area (Figure 4-10, Figure 4-12 and Figure 4-15). However, we also observe a decrease in the V_{th} when traps are located close enough to each other (Figure 4-11 to Figure 4-15) and the reduction in the effective barrier area was identified as the reason for this decrease. Moreover, this reduction in V_{th} due to the proximity of traps with each other is observed both along the channel length and along the channel width. So with increasing N_{TL}, the probability of traps being located closer to each other also increases, thus resulting in a decrease of the average V_{th}. In addition, in our study, we have always maintained the total charge in the device constant by scaling the value of D_{TL} with increasing N_{TL}. This would mean that with increasing N_{TL} (resulting in the scaling of D_{TL}), the influence of individual trap location should also decrease (see Figure 4-8). Eventually, this reduction in the influence would reflect in the overall decrease of the average V_{th}. Thus, the decrease in the average V_{th} with increasing N_{TL} can be attributed to the combined impact of decreasing D_{TL} and increasing the probability of traps being closely placed. It has to be remarked that, although the average V_{th} decreases with increasing N_{TL}, the average V_{th} (with a minimum charge at each location) for a homogenous distribution (N_{TL} = 550) reaches a value that is larger than when N_{TL} =1.

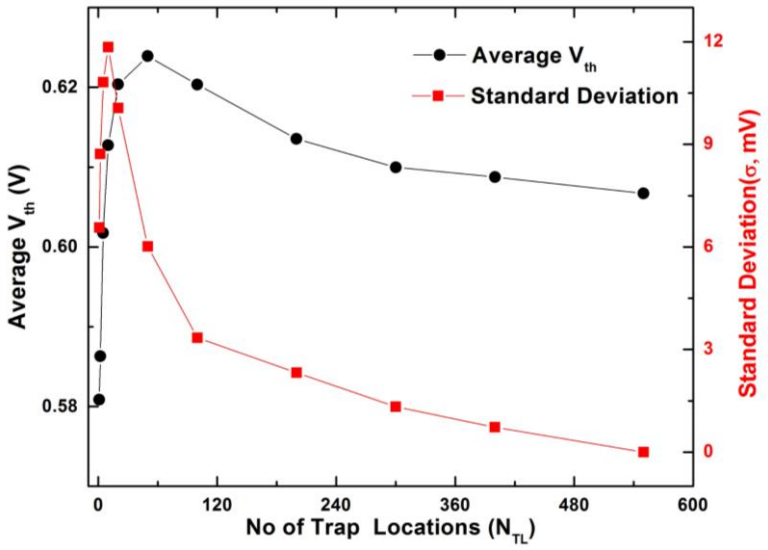


Figure 4-21: Average V_{th} and standard deviation as a function of the number of trap locations in the device (N_{TL}). 3D simulations of 100 devices were performed for each case.

4.6 Dependence of V_{th} variability on device dimensions

As stated in chapter 1, there has been a tremendous increase in efforts to analyze the threshold voltage variability dependence on the device geometry due to several statistical sources in modern MOSFET devices. However, most of these works focus on the variability sources such as RDD and LER [180], [205] and not much has been done with the focus on interface traps. In this section

of the thesis, we present our analysis of the dependency of V_{th} variability on the device dimensions by considering the interface traps as the variation source.

In order to perform the analysis, we consider the same device (doping, material and oxide thickness) that was used earlier but with varying dimensions of lengths and widths. Two sets of simulations namely 'Set1' and 'Set 2' are considered for each device dimension. The variability in V_{th} is evaluated by considering the following approach:

- The device with a channel length of 50nm and a channel width of 20nm is chosen as a reference device in both 'Set 1' and 'Set 2'.
- For this reference device (L X W = 50nm X 20nm) the value of N_{TL} is fixed at 100 and is referred as '$N_{TL, Ref}$'.
- The value of N_{TL} for all other device dimensions is scaled according to the respective device area from the reference value ($N_{TL, Ref}$). For example, for a device with the dimension 35nm X 10nm, the value of N_{TL} was scaled to 35. The value of D_{TL} was obtained from equation 3-1, using a D_{it} = 5e11 cm^{-2}.
- The threshold voltage (V_{th}) is defined using the constant current (CC) method, where the V_{th} is defined as the gate voltage for a fixed value of drain current. However, to make sure that we calculate the threshold voltage at the same electrostatic condition for all device dimensions, we scale the value of the fixed current according to the device area based on the following expression

$$I_{D,F} = I_{D,R} * \frac{W_D}{L_D} * \frac{L_R}{W_R} \qquad 4\text{-}1$$

- In the above expression, $I_{D,F}$ and $I_{D,R}$ corresponds to the fixed value of drain current (for the considered device) and the reference value of drain current (the fixed value of I_D corresponding to the reference device). W_D and L_D are the channel width and channel length for the considered device while W_R and L_R correspond to the channel width and the channel length for the reference device.
- In our study the value of $I_{D,R}$ for the reference device (L_R X W_R = 50nm X 20nm) was fixed at 1µA. So for a device with dimensions 35nm X 10nm (L_D X W_D), the fixed value of drain current to evaluate the V_{th} was found to be 0.714µA using the eq.4-1. It has to be noted that for the entire simulations (Set 1 and Set 2) the drain was always biased to 0.1V.
- In 'Set 1' 50 simulations are performed with the value of N_{TL} fixed across the entire set. Thus, the only source of variability in this set is in the form of the spatial distribution of N_{TL}.
- As the number of charged defects per device is Poisson distributed [104], [206] we perform a new set of simulations say 'Set 2'. In 'Set 2' 50 simulations are performed with N_{TL} varying as per a Poisson distribution with a mean that is equal to the corresponding value of N_{TL} in 'Set 1'. Thus apart from the variability in spatial distribution, an additional source of variability is introduced in the form of varying N_{TL}.

The simulations were repeated for various device areas with channel lengths varying from 60nm down to 35nm and channel widths varying from 20nm down to 10nm. The dimensions of the devices (in nm) and the corresponding values of N_{TL} are tabulated in below Table 2 for reference.

Dim. (LXW)	Area (nm^2)	N_{TL}	Dim. (LXW)	Area (nm^2)	N_{TL}
35X10	350	35	50X14	700	70
35X14	490	49	50X20	1000	100
35X20	700	70	50X30	1500	150
35X30	1050	105	55X10	550	55
40X10	400	40	55X14	770	77
40X14	560	56	55X20	1100	110
45X10	450	45	55X30	1650	165
45X14	630	63	60X10	600	60
45X20	900	90	60X14	840	84
45X30	1350	135	60X20	1200	120
50X10	500	50	60X30	1800	180

Table 2: Dimensions of devices that are used in the study and the corresponding values of N_{TL}. The device with the dimension 50nmX20nm and with N_{TL} of 100 is used as a reference. The value of N_{TL} corresponding to all other devices scaled according to the area from this value.

Thus, based on our considerations, in 'Set 1', we have only one source of variability in the form of the spatial distribution of N_{TL}, however, in 'Set 2' we have two sources of variabilities in the form of the value of N_{TL} and its spatial distribution. Now, at the end of each simulation, V_{th} is extracted and the values of average V_{th} and σV_{th} are evaluated for each set. In the following, the results of the simulations are presented and analyzed. To be clear, the plots and results corresponding to the case of 'Set 1' will be referred by 'SD' ("spatial distribution") as the only source of variability, in this case, is in the form of spatial distribution. And the plots and results corresponding to the case of 'Set 2' will be referred by 'NSD' ("number and spatial distribution") as in this case, apart from the existing variability in the form of spatial distribution, an additional variability is added in the form N_{TL}. Figure 4-22 plots the V_{th} as a function of channel lengths (left) and channel widths (right) for a pristine device (without traps). As expected, we observe that the V_{th} decreases with channel lengths, an effect related to the 'Short Channel Effects (SCE)'. We see that in the considered range the threshold voltage is independent of the channel widths, this could be due to the fact that all the devices being simulated have very narrow channel.

Now, we perform simulations by introducing traps along the interface. Figure 4-23 plots the average V_{th} as a function of channel lengths on the left and as a function of channel widths on the right. For comparison, the threshold voltage corresponding to the pristine device is also plotted in the same graph (on left). As one would expect, we can see that the introduction of the traps has caused a shift (almost parallel) in the threshold voltage in the form of an increase in the average V_{th}. We can also see from the plots that the randomness in the value of N_{TL} does not seem to have any effect on the average V_{th} as the plots corresponding to SD and NSD overlap. According to [98] the dependence of σV_{th} on device area follows the Pelgrom's rule, which states that σV_{th} across several device areas is inversely proportional to the square root of the device area (equation 1-7). Then when the graph of σV_{th} vs $1/\sqrt{Area}$, known as 'Pelgrom's plot' is plotted, a straight line

is obtained. So, in order to verify if the variability (σV_{th}) due to interface traps follows the Pelgrom's rule, in Figure 4-24 we make a graph of σV_{th} vs $1/\sqrt{Area}$. In the figure, the plots corresponding to the SD case are represented by red circles and the case corresponding to NSD are represented by black squares. The linear fittings for the respective data are also shown in the figure. The linear fit corresponding to SD has a slope of 0.144V-nm and the linear fit corresponding to the NSD has a slope of 0.175V-nm. We can observe that larger variability is obtained in the case of NSD. This is reasonable as in this case, an additional variability in the form of N_{TL} is introduced.

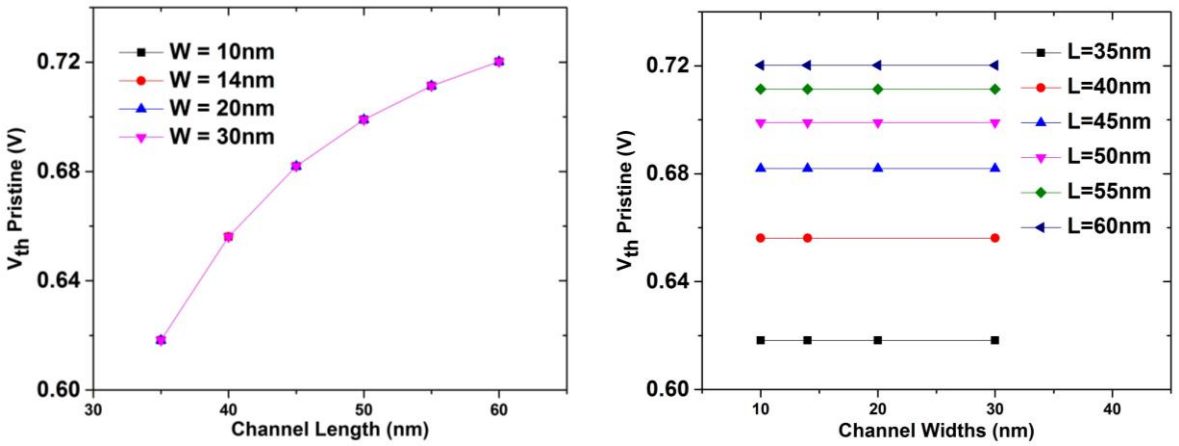


Figure 4-22: (left) V_{th} in a pristine device (without traps) as a function of the channel length for different channel widths. (right) V_{th} as a function of channel width for different channel lengths. The extracted threshold voltage shows the expected dependency on the length and is independent of the channel width.

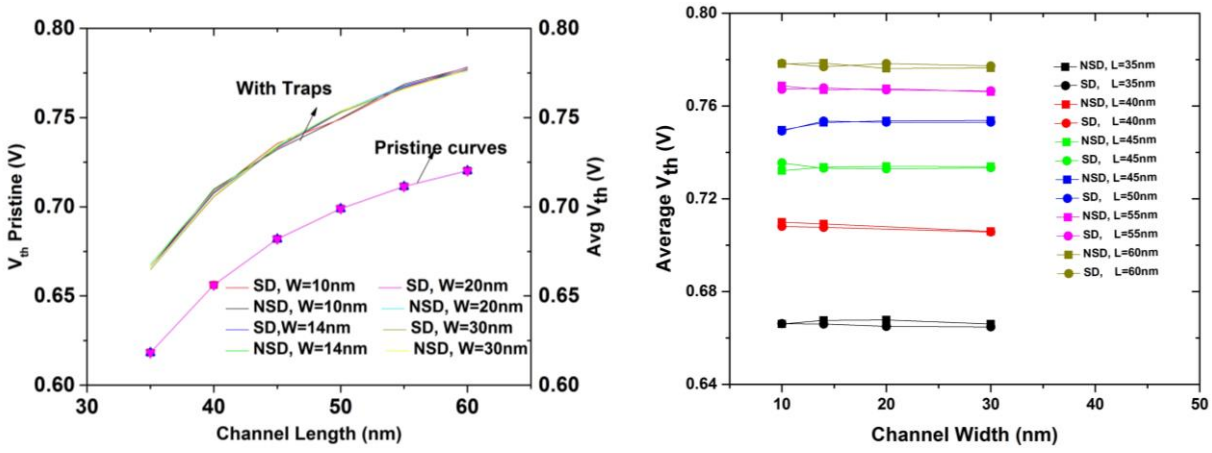


Figure 4-23: (left) Average V_{th} (across 50 simulations) as a function of channel length for different channel widths. (right) Average V_{th} (across 50 simulations) as a function channel width for different channel lengths. 'SD' the simulations where only the spatial distribution is varied while 'NSD' indicates the simulations where both N_{TL} and the spatial distribution are varied. For comparison the plots belonging to the pristine case is overlaid onto the average plot on the left.

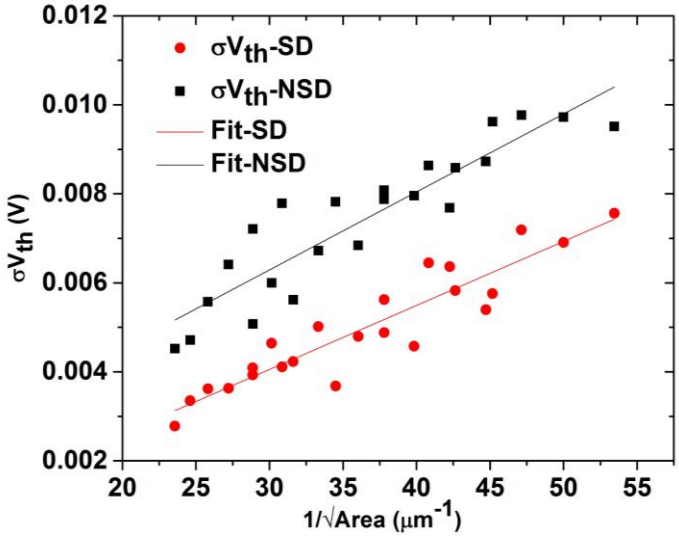


Figure 4-24: The plots correspond to σV_{th} as a function of device area for the SD case (red circles) and NSD case (black squares). Also shown are the linear fits of the data with the slopes of 0.144 V-nm and 0.175 V-nm for SD and NSD respectively.

Now, as area (A) can be expressed in terms of the device dimensions (A = L * W), to better understand the dimensional dependence of σV_{th}, we examine the relationship of σV_{th} on the device width and length separately. Figure 4-25 shows σV_{th} as a function of channel length for different channel widths. The plots on the left (in Figure 4-25) correspond to the SD case and the plots on the right correspond to that of NSD, also shown are the corresponding linear fits (lines). The fit lines have a lower slope values with an average (across all the widths) of about -54.85μVnm^{-1} for the case of SD and 66.02μVnm^{-1} for the case of NSD.

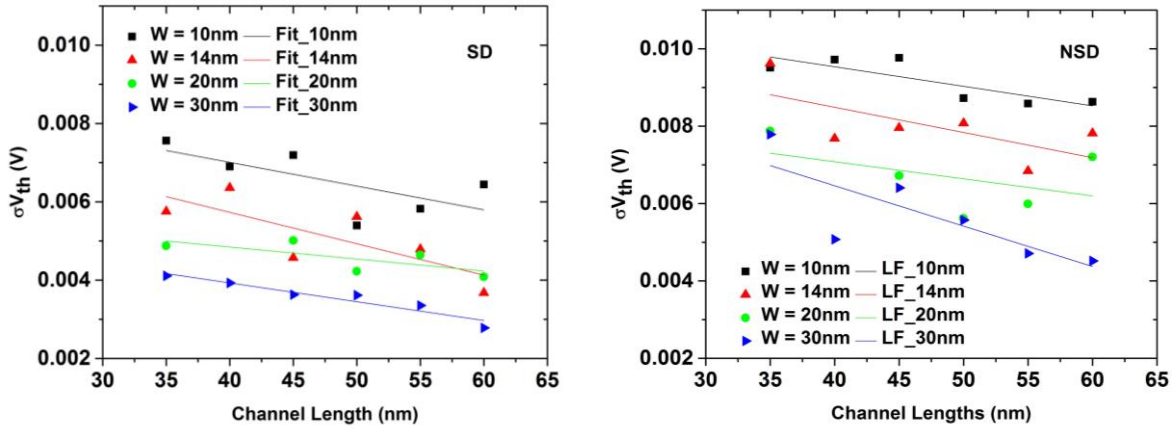


Figure 4-25: (left) σV_{th} in case of SD as a function of the channel length for different channel widths and (right) σV_{th} in the case of NSD as a function of channel length for different channel widths. The variability in the threshold voltage (V_{th}) has a very small dependence on the channel lengths. However, the introduction of additional variability (in NSD) has resulted in an increase of σV_{th} (compared to SD).

From these low slope values, it is evident that σV_{th} has a very low dependence on the channel lengths. However, we can see that the values of σV_{th} are larger for the case of NSD (N_{TL} is varied according to Poisson distribution across the 50 simulations for each device area) as compared to the case of SD (where only the spatial distribution is varied).

Similarly, we can examine the dependence of σV_{th} on the channel width. For this, we plot σV_{th} as a function of channel widths for different channel lengths and this is shown in Figure 4-26.In the figure the plots on the left correspond to the case of SD and the plots on the right correspond to that of NSD. The dependencies can be fitted by a power function with the value of average exponent (across all the lengths) as -0.53 for the case of SD and -0.46 for the case of NSD. Based on this, we can see that (from Figure 4-25) when considering interface traps as the only source of variability, the dependency of σV_{th} on channel lengths is very small. This could mean that when considering only interface traps as the source of variability, the dependence of σV_{th} on the device area is primarily from the channel widths.

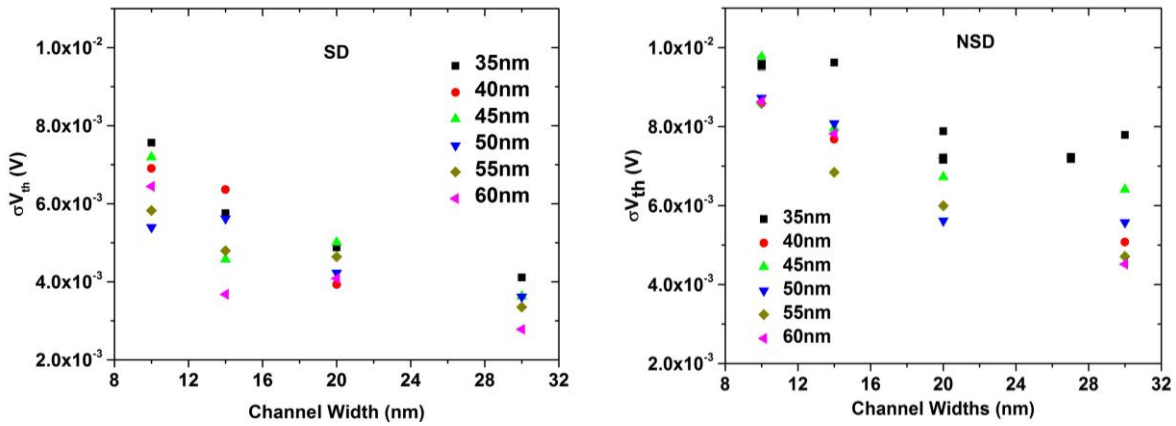


Figure 4-26: (left) σV_{th} in case of SD as a function of the channel width for different channel lengths and (right) σV_{th} in the case of NSD as a function of channel width for different channel lengths. The dependencies can be fitted by power function with an average exponent (across all lengths) of -0.53 for the case of SD and -0.46 for the case of NSD.

As outlined earlier, Pelgrom's rule states that σV_{th} is inversely proportional to the square root of the device area (A=L*W), which is written as follows

$$\sigma V_{th} \propto \frac{1}{\sqrt{L * W}}$$

The influence of channel width (W) can be isolated by multiplying the above expression with $\sqrt{L}$, thus the above expression can be rewritten as below

$$\sigma V_{th} * \sqrt{L} \propto \frac{1}{\sqrt{W}}$$

In order to verify this, we make a graph similar to the Pelgrom's plot but instead of σV_{th} vs $1/\sqrt{Area}$, we plot σV_{th}* $\sqrt{L}$ vs $1/\sqrt{W}$ as shown in Figure 4-27.

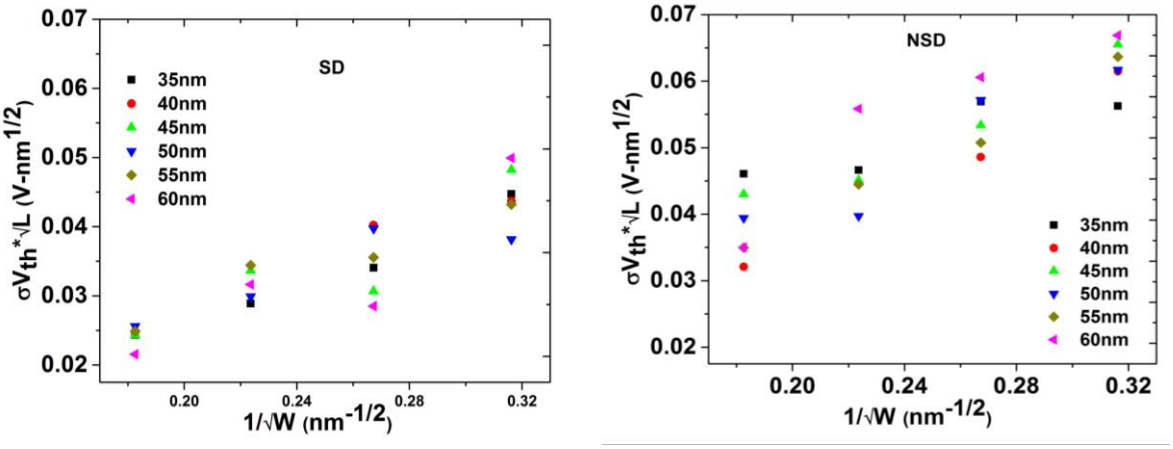


Figure 4-27: (left) Corresponds to the plots for the case of SD and (right) for the case of NSD. We can see an inverse linear dependency of σV_{th} on the channel widths.

This plot helps us to observe if σV_{th} is inversely proportional to square root of the channel widths. Again the plots on the left correspond to the case of Set2 (SD) and the plots on the right correspond to the case of Set2 (NSD). For a better understanding let's make a plot of the average σV_{th} (across several different lengths) for each case (Set1 and Set2) as shown in Figure 4-28

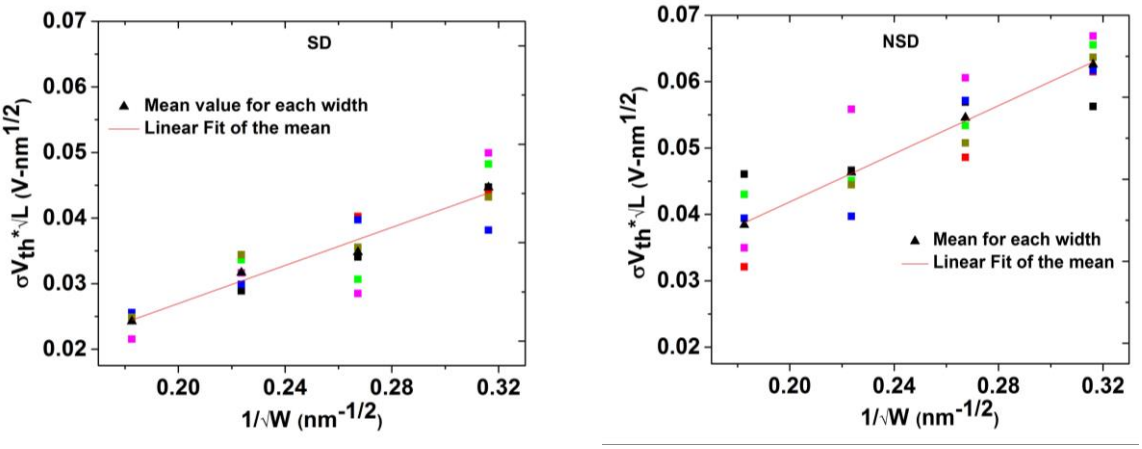


Figure 4-28: The figure shows the same plots as shown in Figure 4-27 but also included is the plot of average σV_{th} (across different lengths) and the linear fit curve for the average σV_{th}. The plots on the left correspond to the case of SD and on the right corresponds to the case of NSD. The slope of the fit curves was found to be 0.145 V-nm and 0.1813 V-nm corresponding to the case of SD and NSD respectively.

Also included in the figure is the linear fit of the average σV_{th}. The plots on the left correspond to the case of SD and the plots on the right correspond to the case of NSD. The slope of the linear fit of the mean was found to be 0.145 V-nm and 0.1813 V-nm for SD and NSD respectively. We can

see that these slopes are almost coincidental to the slopes of the Pelgrom's plot shown in Figure 4-24. Thus, it can be stated that the geometry dependence of σV_{th} due to interface traps is mostly determined by the channel widths. The reason for this behavior can be understood by taking into context the results seen earlier sections, where it was shown that only those traps that are located along the maximum of the barrier (which happens close to the channel center) along the channel length contribute significantly to the variation in the σV_{th}. Therefore, several different combination of traps along the edges of the channel length could lead to similar V_{th} (thus the smaller dependence on 'L'). However, along the width of the channel, the contribution of the trap is independent of its (trap's) location along the width. Therefore, different combinations (spatial distribution) of traps along the channel width would lead to different V_{th}. Thus based on this, we can conclude that when considering interface traps as the variability source, the geometrical dependence of σV_{th} is mostly determined by the channel widths.

Up until now, we only studied the influence of the trap location and its spatial distribution on the variability of the device performance by observing the threshold voltage. In the next chapter, by means of 3D simulations, we investigate the influence of trap location and its spatial distribution on the variability of both threshold voltage (V_{th}) and on-current (I_{on}).

CHAPTER V

On-current and threshold voltage variability due to interface traps spatial distribution

Shrinking of transistor sizes meant that the number of interface traps in the channel area of MOSFETs becomes smaller to the level of few traps. In [207] the existence of a single trap across various short channel devices was experimentally demonstrated using charge pumping method. Recently, in [208] the impact of single-charged trap induced random telegraph noise (RTN) on FinFET devices were studied. In small-area devices BTI induced V_{th} degradation and recovery in nanometer transistor proceed in discrete steps [132], [197], [198]due to augmentation of single-charge effects in scaled device. Charges trapped/de-trapped in/from the traps located at the semiconductor/insulator interface lead to fluctuation in the threshold voltage and (or) the on-current (saturation drive current) [117][137]. Therefore, while analyzing the impact of spatial variability of interface traps on device performance, it is important to also include the changes in the on-current (apart from V_{th}). In this chapter we study how the spatial distribution of single and multiple traps effects the variabilities in on-current (I_{on}) and threshold voltage (V_{th}) of the device.

5.1 Device structure and Simulation Methodology

To analyze the influence of discrete traps on the variability of device performance (by analyzing V_{th} and I_{on}), we simulate a 3D structure with a width of 50nm and channel length of 40nm. The

schematic of the simulated 3D structure is shown in Figure 5-1. The simulated structures had silicon oxide with 1.5nm thickness as an interfacial layer. The device is based on the super-halo doping from the well-calibrated template that was used in the earlier simulations [202].

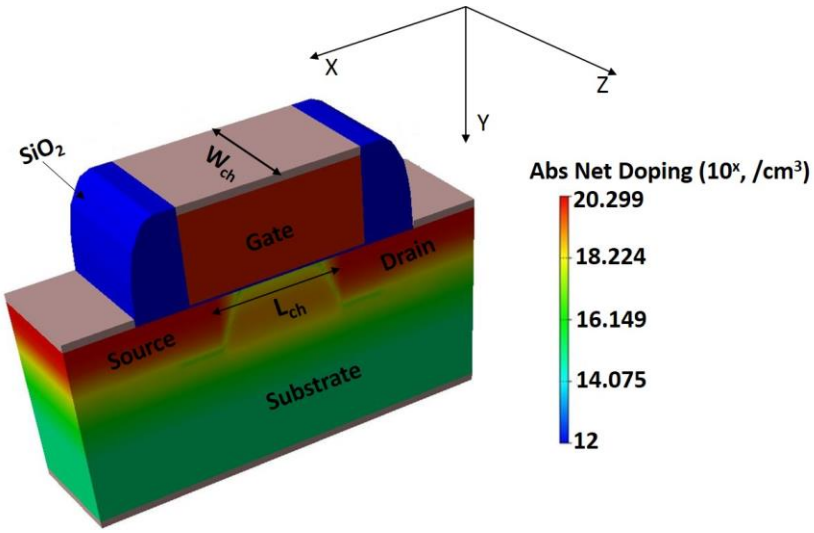


Figure 5-1: Simulated 3D structure with the doping profile overlaid

We use the same simulation approaches that were discussed earlier to analyze the impact of interface trap(s) related variability. However, here we use a much finer mesh along the interface with 1nm along the length and 1nm along the width. This allowed us to simulate a higher number of traps using the earlier strategy. However, we only analyze the case of 3D device structures with a 2D interface. This was because simulating the effect of a single trap required the device to have a defined width other than the default width of 1µm.

To evaluate the electrical characteristics of devices, 3D device simulations were performed on the device structure shown in Figure 5-1. The characteristics were obtained when the gate voltage was ramped to 1.5V and drain was biased at 1.1V while all other contacts (source and bulk) were biased at 0V. The drain is biased at a larger voltage as usually I_{on} is defined at higher drain bias (see section 1.1, parameter extraction). Figure 5-2 shows the transfer characteristics (I_D-V_G) of a device without interface traps and this case will be considered as the reference device (to measure the shifts in the performance parameter due to traps). The figure also shows the definition of the performance parameters used in the presented work. The threshold voltage (V_{th}) is defined as the gate voltage corresponding to the drain current of 0.1µA. On-current (I_{on}) is defined (see section 1.1, parameter extraction) as the maximum drain current obtained from the transfer characteristics (I_D-V_G) at a high drain bias (V_D) which in the analysis was set to 1.1V. The threshold voltage (V_{th}) and on-current (I_{on}) of the reference device was extracted to values of 0.301V and 5.83E-5 A, respectively.

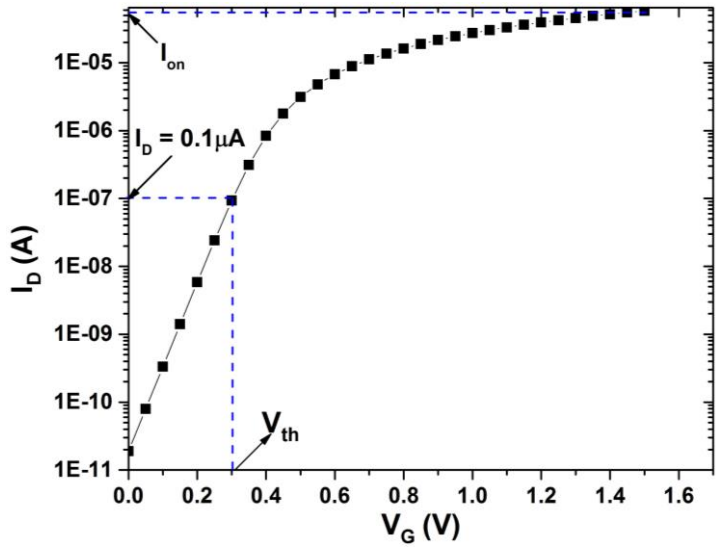


Figure 5-2: I_DV_G curve of the device without traps. Also shown are the definition of I_{on} and V_{th}. The gate was ramped to 1.5V while the drain was biased at 1.1V. V_{th} is defined as the gate voltage at a fixed drain current of 0.1μA and I_{on} is defined as the maximum drain current.

5.2 Single Trap

We start the analysis by considering the case of a single trap, which may not be realistic but is the simplest case. Figure 5-3 (left) shows the 3D simulated structure with a single interface trap along the 2D interface and Figure 5-3 (right) shows the cross-section (in XZ plane) of the channel region with the trap. Figure 5-4 shows the transfer characteristics (I_D-V_G) of a device without traps (reference) and with a trap at a particular location along the 2D interface. In this case, the trap was located in a region close to the channel center and the lower part (along the width) of the channel (corresponding to Figure 5-3) (in this case at x= 18nm and z=8nm). The transfer characteristics shown in Figure 5-4 is obtained by biasing the gate to 1.5V and drain to 1.1V. The threshold voltage and the on-current of the device with a trap at that location were found to be 0.307V and 55.5μA respectively. This meant that the trap has caused the threshold voltage to shift by 6mV (ΔV_{th}) and the on-current to (negative) shift by 2.8μA (ΔI_{on}) from the reference device. As two different performance parameters are analyzed, for the sake of simplicity, threshold voltage and on-current for devices will be given as (V_{th}, I_{on}) vectors. As an example, in the case of the structure (shown in Figure 5-2) without traps, the device is characterized by (0.301V, 58.3μA) and the device with a single trap (as shown in Figure 5-4) is characterized by (0.307V, 55.5μA).

The statistical variability in a set of devices is studied by performing a simulation of 100 different devices with the location randomly varied along the 2D interface between the semiconductor and the insulator of the MOSFET. At the end of each simulation, performance parameters (threshold voltage and on-current) were extracted. Figure 5-5 plots the variability in threshold voltage (V_{th}) and on-current (I_{on}) for the 100 simulated devices. In Figure 5-5, X-axis represents the device number, left (black) Y-axis represents the threshold voltage (V_{th}) corresponding to each device and right (red) Y-axis represents the on-current (I_{on}) corresponding to each device. Device-to-

device variability in both parameters is clearly observed and, in general, larger values of threshold voltage lead to smaller values of the on-current, as one would expect. The relationship of the on-current and the threshold voltage can be clearly visualized by plotting I_{on} as a function of V_{th}, as shown in Figure 5-6. From this figure, it can be seen that smaller values in V_{th} are usually followed by larger values in I_{on}

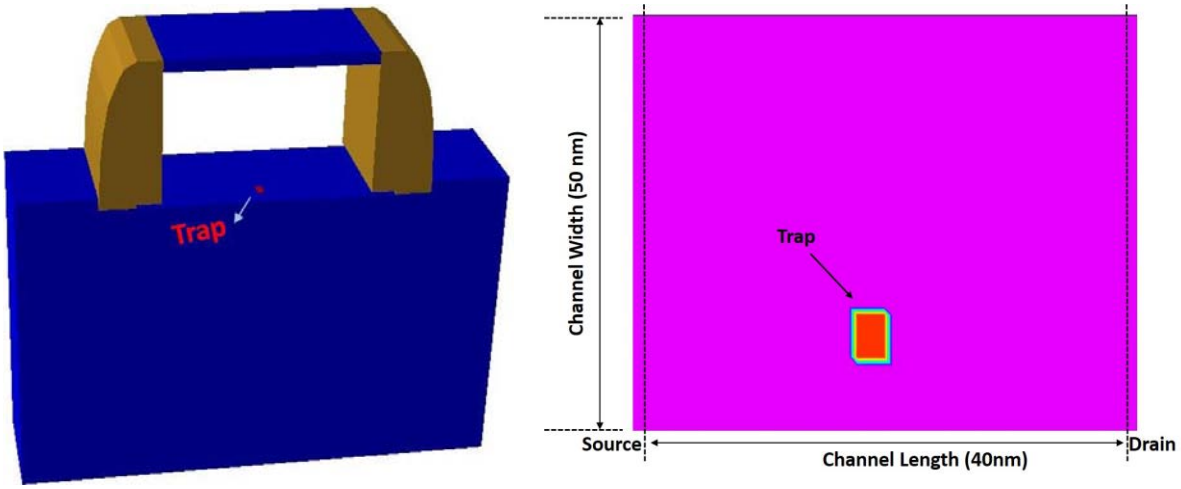


Figure 5-3: (left) 3D simulated structure with a single trap along the 2D interface and (right) the cross-section of the interface with the trap. The device has a channel length of 40nm and channel width of 50nm.

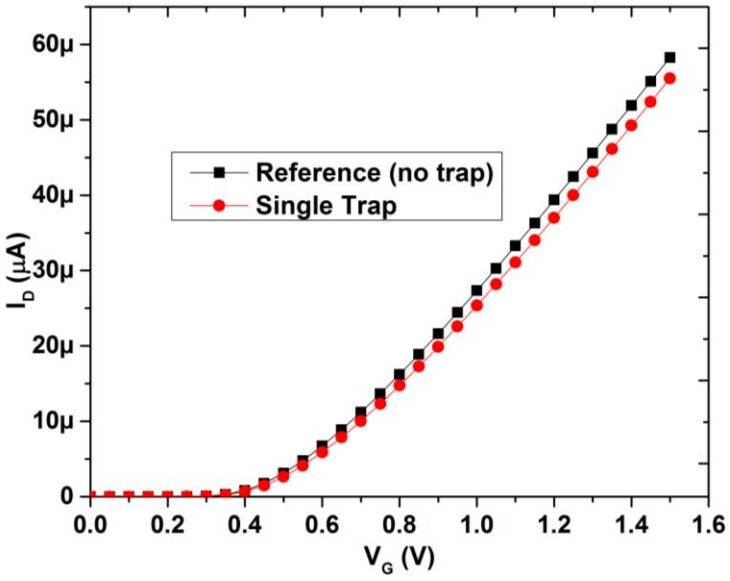


Figure 5-4: I_DV_G curve of the reference device without traps (squares) and with a trap (circles) corresponding to the Figure 5-3.

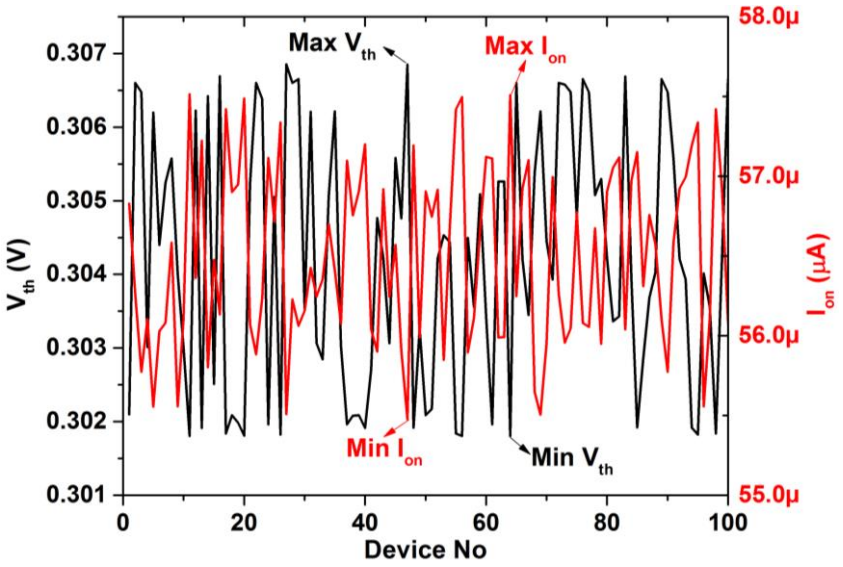


Figure 5-5: Threshold voltage (left, black) and on-current (right, red) of the devices for 100 simulated devices with a single interface trap. The location of this trap is varied randomly from one device to the other. In general, large values of V_{th} lead to smaller values of I_{on}.

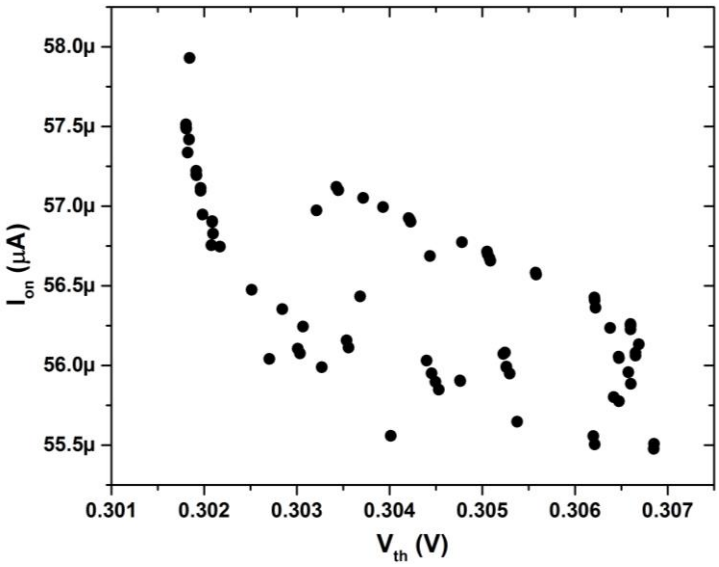


Figure 5-6: I_{on} as a function of the extracted V_{th}. It can be seen that in general, devices with smaller V_{th} have larger I_{on}.

From Figure 5-6, we can also observe that there are some devices which have similar V_{th}'s (I_{on}) but slightly different I_{on}'s (V_{th}). To understand the influence of trap location on individual parameters (V_{th} and I_{on}) separately, Figure 5-7 plots the contour maps of V_{th} and Figure 5-8 plots the contour map of I_{on} as a function of trap locations along the 2D interface. The X-axis of the map is the coordinate along the channel length (with -20nm being the source edge and 20nm being the drain edge) and Y-axis corresponds to the channel width (Z of the device).

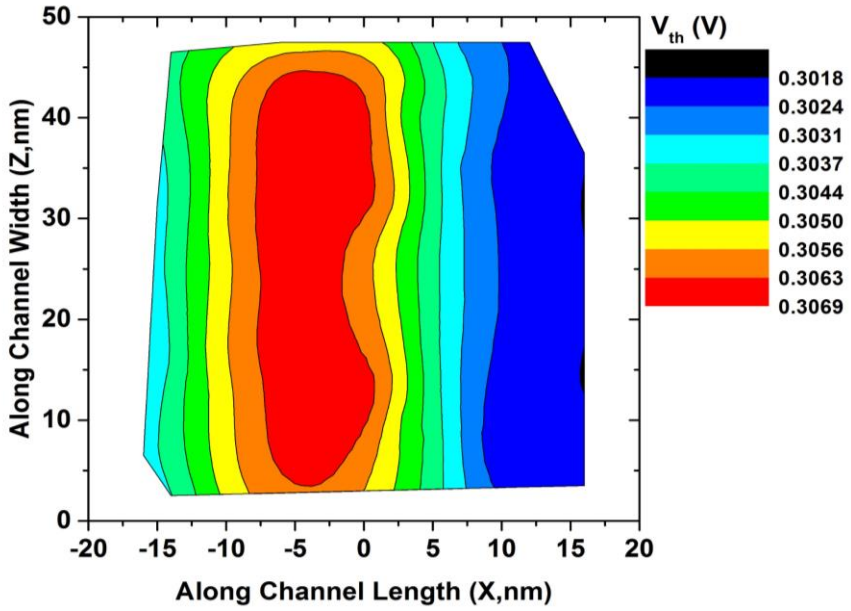


Figure 5-7: Contour map of V_{th} corresponding to different trap locations along the 2D interface. We can see that larger V_{th} is obtained when the trap is located along the channel center and lower V_{th} is obtained when the trap is located along the device edges.

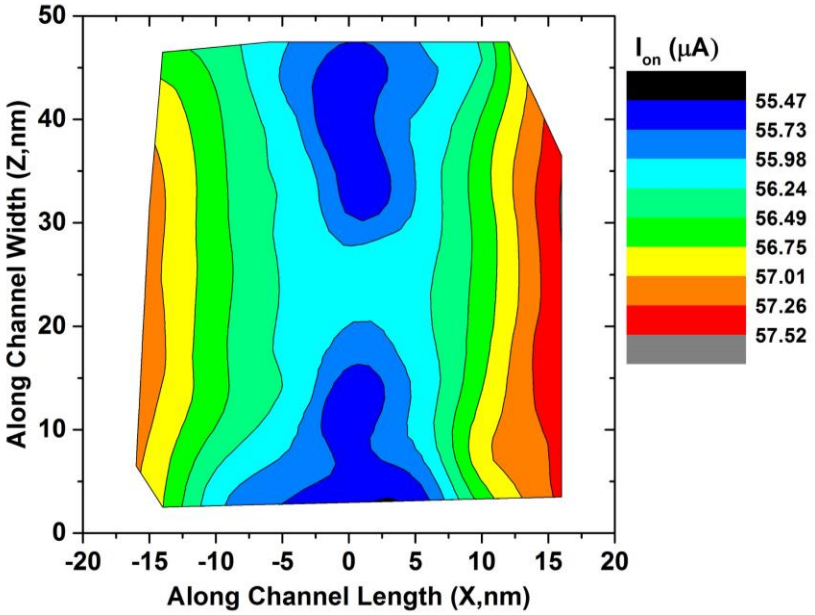


Figure 5-8: Contour map of I_{on} corresponding to different trap locations along the 2D interface. We can see that in general, larger I_{on} is obtained when the trap is located along the channel edges smaller I_{on} is obtained when the trap is located along the channel center.

From the contour maps of V_{th} in Figure 5-7, we can clearly observe that larger V_{th}'s are obtained when the trap is located close to the channel center (along the channel length). We can also observe that V_{th} is not influenced by the location of the trap along the channel width. These results are coherent with what we have seen in the earlier chapters, where it was shown that traps assert a significant influence when they are located along the channel center. The contour map of I_{on} in Figure 5-8 provides a clear picture of influence of the location of trap on the on-current. We can see from the figure that in general, lower values of I_{on} are obtained when the trap is located close to the channel center and larger values are obtained when the trap is located along the channel edges. Now, we can observe (Figure 5-8) that there is a slight variation in values of I_{on} corresponding to when the trap is located at the channel edges along the width (along z-direction) and channel center along the length (along x-direction). However, from the magnitudes in the contour plot (Figure 5-8), we can see that this variation I_{on} is in the range of $\approx 0.5\mu A$, suggesting that these variations may be related to the resolution (I_{on} is extracted at max I_D) of simulations and not related to any physical mechanism. Now, to analyze the variability related to more than one trap, simulations were repeated with an increased number of traps along the interface and the analysis of the results are presented in the following section.

5.3 Multiple Traps

Until now we have studied the influence of a single discrete trap on the device performance (V_{th}, I_{on}). The analysis was focused on the spatial location of this trap along the 2D interface of the structure. However, having more than one trap can contribute to the variability in the form of the spatial distribution of these traps. Hence, in this section we study the impact of the spatial distribution of multiple interface traps on the device electrical characteristics. To do this we increase the total number of traps along the interface to 10, as this number ensured an appreciable shift in the electrical characteristics. As an example, Figure 5-9 compares the I_D-V_G curves of the device with a particular spatial distribution of 10 traps (red, circles) to that of the pristine device (black, squares). The figure also shows the modifications in the V_{th} and I_{on} due to the introduction of traps. The device with traps is characterized by the vector (V_{th}', I_{on}') (0.413V, 29.2µA). It can be seen that the increased number of traps has significantly increased the V_{th} ($\Delta V_{th} = 0.112V$) and decreased I_{on} ($\Delta I_{on} = 29.1\mu A$). To be unambiguous, the threshold voltage and on-current for the device with traps is also referred by the notation V_{th} and I_{on} unless they are represented together with the pristine device, in that case, they will be referred by the notation V_{th}' and I_{on}'.

To analyze the impact of the spatial distribution, 100 devices with randomly generated trap distributions were simulated. Figure 5-10 plots the variability in threshold voltage (V_{th}) and on-current (I_{on}) for the 100 simulated devices, obtained from curves as those in Figure 5-9. Maximum and minimum values are indicated by the arrows in the figure. In Figure 5-11, the same data shown in Figure 5-10 is plotted in a different way, by plotting the on-current across different devices as a function of the corresponding threshold voltage. Large variability in both parameters is observed with (V_{th}, I_{on}) vector varying from (0.32V, 44.67µA) to (0.413V, 29.1µA). In general, large V_{th}'s lead to small I_{on}'s and vice versa, however, unlike in the case of single trap (Figure 5-6), here the dispersion is significant ($\approx$15µA) and we can see (from Figure 5-10 and Figure 5-11) that there are some distributions that lead to the same threshold voltage (on-currents) but have different on-currents (threshold voltages).

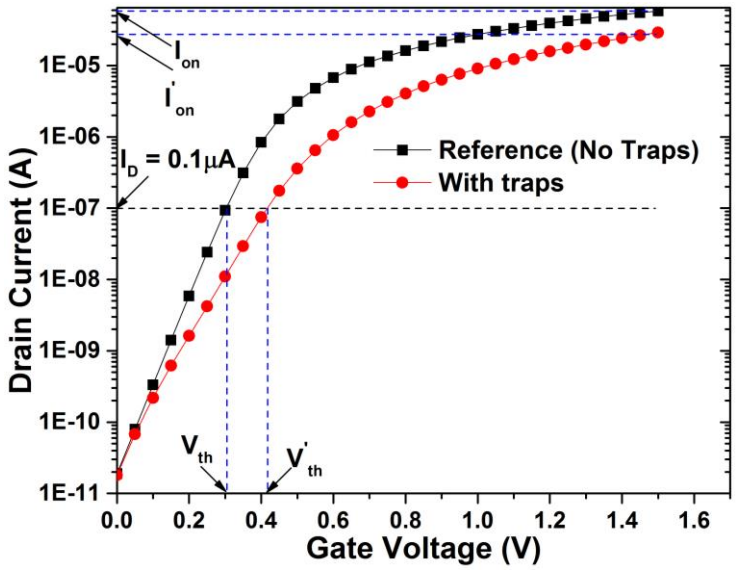


Figure 5-9. I_DV_G curve of the reference device without traps (squares, black)) compared to that of a particular case of 10 trap distribution (circles, red). Here V_{th} and I_{on} are the threshold voltage and on-current of the pristine device while V_{th}' and I_{on}' are the threshold voltage and the on-current of the device with traps (of a particular spatial distribution).

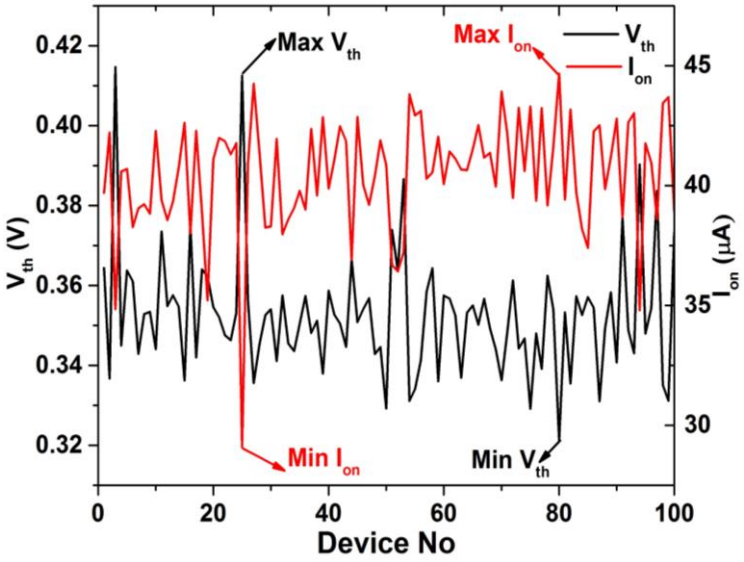


Figure 5-10: Threshold voltage and on-current of the devices for the 100 considered trap distributions. Large variability is observed both in I_{on} and V_{th}.

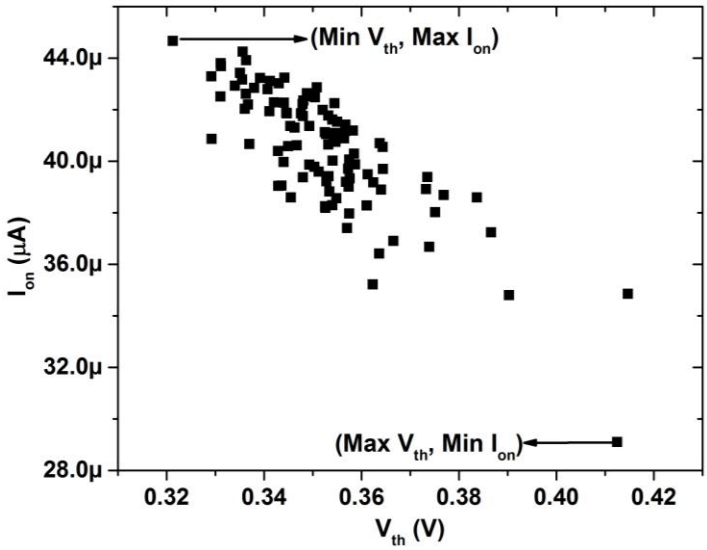


Figure 5-11: On-current and threshold voltages of the 100 considered devices with randomly varying spatial distribution of the 10 traps. Devices with extreme V_{th}-I_{on} vectors are also indicated in the figure.

In order to understand the general trend and the anomalous behavior, we will investigate these two behaviors individually. We start the analysis by having a look at the trap maps corresponding to the min and max vectors marked in Figure 5-11. Figure 5-12 shows the I_D-V_G curves corresponding to these devices. The insets in the figure show the maps of the corresponding trap distributions.

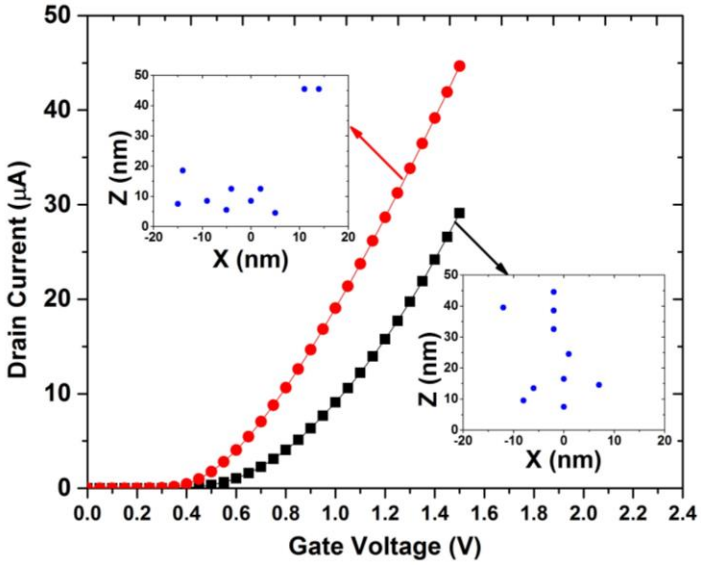


Figure 5-12: I_DV_G curves of the devices with (max V_{th}, min I_{on}) (squares) and (min V_{th}, max I_{on}) (circles) vectors. The insets show the 2D maps of the corresponding trap distributions.

The X-axis of the plots in the inset correspond to the coordinate along the channel length and Y-axis corresponds to the coordinate along the width of the channel. So, -20nm along the X-axis

represents the channel at the source edge, 0 represents the channel center and 20nm represents the channel at the drain edge. From the maps of the trap distributions, we can see that in the case of the I_D-V_G that leads to max V_{th} (black squares) most traps are located along the channel center and aligned along the width as compared to the traps corresponding to the case I_D-V_G that leads to min V_{th} (red circles) where the traps are aligned along the length. We can see that when traps are aligned along the channel width and close to the center of the channel, then these distributions led to the vector of ($V_{th,\,max}$, $I_{on,\,min}$) as compared to the traps that lead to the vector ($V_{th,\,min}$, $I_{on,\,max}$). Based on the results seen in the previous chapters, we can attribute this increase of V_{th} (conversely decrease of I_{on}) to the increase of the effective barrier area along the channel width. The anomalous behavior (different V_{th}/I_{on} but same I_{on}/V_{th}) seen in Figure 5-11 suggests that in some cases the trap distribution only impacts one of the electrical parameters. In order to understand how the distribution along the 2D interface (x (channel length) and z (channel width) coordinates) affect the device V_{th} and $I_{on,}$ pairs of such devices have been considered. As an example, Figure 5-13 shows the I_D-V_G curves (with insets of trap maps) for two distributions that imply the same V_{th} but different $I_{on,}$ i.e., (V_{th}, I_{on}) vectors of (0.33V, 40.86µA) (black squares) and (0.33V, 43.3µA) (red circles). To understand the influence of trap distribution on the device performance, the trap maps have been overlaid onto the conduction band profile in Figure 5-14. It has to be noted that the conduction band was obtained from the device (reference/pristine) with no interface traps. From the trap maps, we can see that most of the traps are located 'far' from the channel center (along the channel), where the peak of the conduction band is observed. This could explain the low value of the threshold voltages in both cases. We can also observe from the figure that there is a trap alignment along the width in both the devices, and this alignment is delineated by a dashed rectangle in Figure 5-14. This alignment of the traps (inside the dashed box) along the width happens at a location with same barrier height.

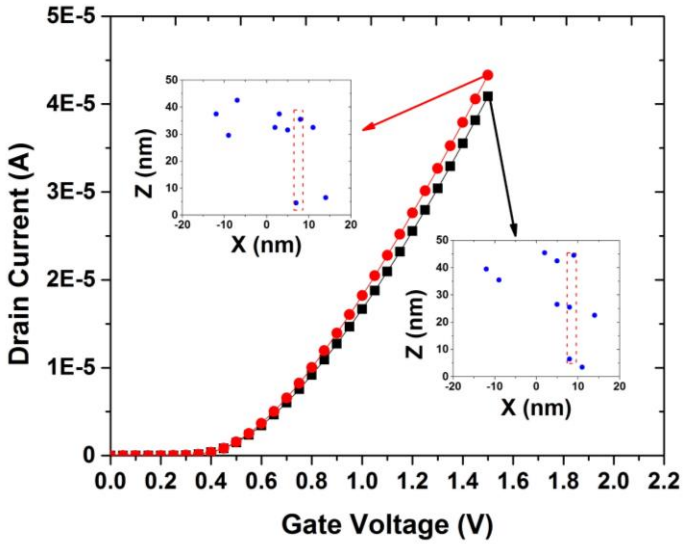


Figure 5-13: I_DV_G curves for the devices with similar V_{th} but different I_{on}. The insets show the corresponding trap distribution maps. The distribution belonging to the black curve (squares) has the vector (0.33V, 40.86 µA) and the red curve (circles) has the vector (0.33V, 43.3µA).

From the results discussed in the earlier chapters, we know that the trap location along the channel length (X-direction) modifies the barrier height and hence modifies the V_{th}. By taking them into account it is reasonable that the V_{th} in both of the trap distribution (presented above) is same. However, in the case of Figure 5-14 (right) the alignment along the width is more pronounced (3 traps) than in the case of Figure 5-14 (left). This greater alignment (at the same x) along the width could be the reason for a smaller value of I_{on} while being ineffective on V_{th}. However, before making any conclusions or hypothesis, other vector (V_{th}, I_{on}) pairs should also be analyzed. Figure 5-15 shows the I_D-V_G curves (with insets of trap maps) for two distributions that have similar (almost same) I_{on} but different $V_{th,}$ i.e., (V_{th}, I_{on}) vectors of (0.36V, 35.22µA) (black squares) and (0.41V, 34.85µA). The device with the distribution as shown in the right inset has a higher V_{th} than the device with the distribution shown in the left inset.

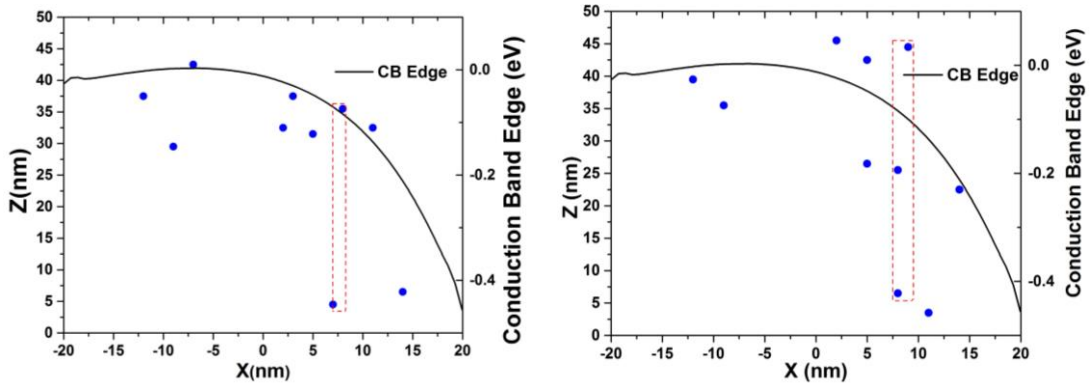


Figure 5-14: Trap maps of the devices corresponding to similar V_{th} and different I_{on} (IV 's in Figure 5-13) overlaid onto the CB, I_{on} is larger for the case of the distribution shown in the left.

Similar to the previously analyzed case (Figure 5-13 and Figure 5-14), the trap distributions have been overlaid onto the conduction band edge and is shown in Figure 5-16. The dashed rectangles correspond to the traps that are aligned along the width of the device. From the above trap maps, we can see a clear alignment along the width (Z-axis) but at different locations along the device channel length (X-axis). Taking the results from the earlier chapters into consideration, the same/different threshold voltages for the distributions in Figure 5-14 and Figure 5-16 can be attributed to the same/different positions of the traps along the channel direction (X-axis). The differences/similarities in I_{on} could be attributed to the variations in the extent of alignment along the channel width (Z-axis). This could be easily explained by considering a classical picture: the aligned traps along the device width would represent a "barrier" to the current flow from source to drain, reducing the current collected. However, a large decrease in the V_{th} is also reflected as an increase in the value of I_{on} as less energy (voltage) is needed by the carriers to overcome the barrier. We can correlate this reason to the increase of V_{th} (or the decrease of I_{on}) for the case of I_D-V_G corresponding to the distribution in Figure 5-16 as compared to I_D-V_G corresponding to the distributions in Figure 5-14.

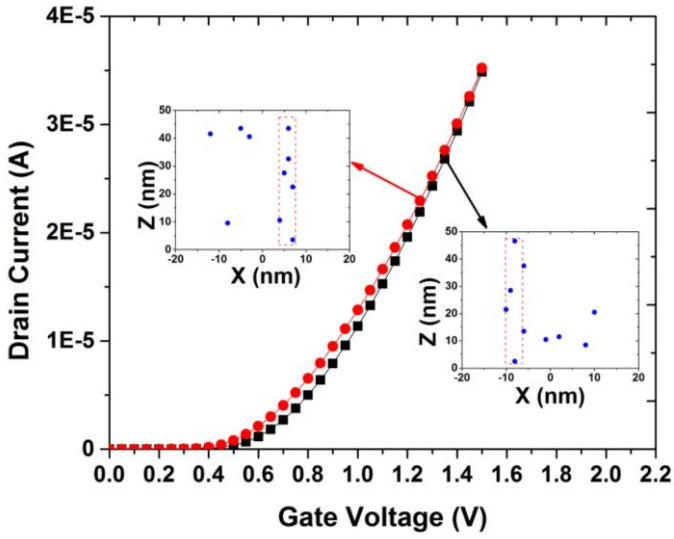


Figure 5-15: I_DV_G curves with the inset of the trap maps for the distributions corresponding to the vectors (0.36V, 35.22µA) and (0.41V, 34.85µA) that yield similar I_{on} but different V_{th}.

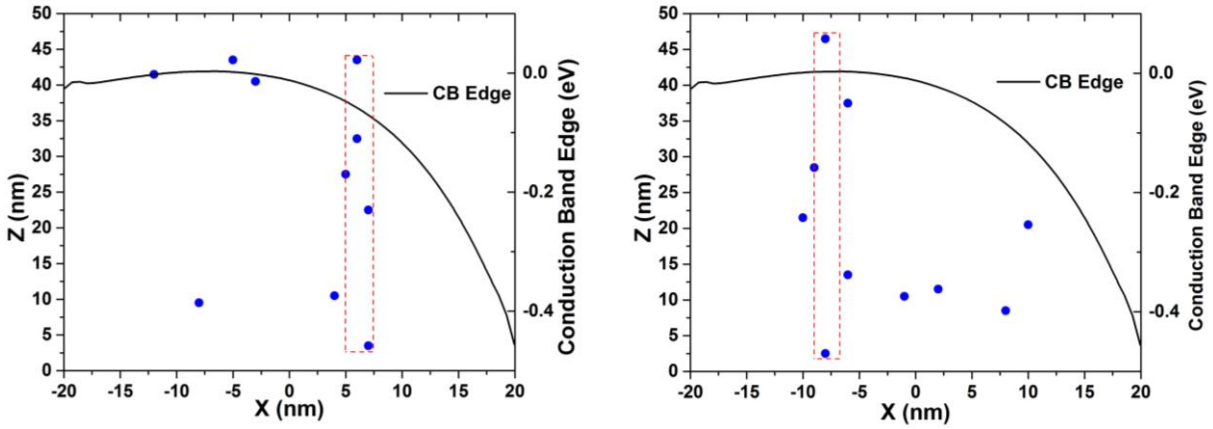


Figure 5-16: Trap maps of the devices corresponding to similar I_{on} and different V_{th}. The device with the distribution shown on the right has a larger V_{th}.

To corroborate the hypothesis and decouple the effect of trap distribution (along X and along Z axis) on the V_{th} and $I_{on,}$ we consider certain special distributions, such as (i) the 10 traps are completely aligned along the channel length (horizontal distribution) as shown in Figure 5-17 (a) and (ii) the 10 traps are completely aligned along the channel width (vertical direction) as shown in Figure 5-17 (b). In Figure 5-17 (a) all the traps are aligned along the channel length at a particular location along the channel width. In Figure 5-17 (b) the traps are aligned along the channel width at a particular location along the channel length.

Several simulations were done for these cases, sweeping the position of the x-aligned traps along the z coordinate and the z-aligned traps along the x coordinate. Figure 5-18 shows the I_D-V_G curves

of two distributions corresponding to X-aligned traps at two different locations along the channel width. The results obtained for the horizontal distribution show that, independently of its location along the z coordinate (width of the device), the I_D-V_G curves overlap (as shown in Figure 5-18), so that no remarkable changes neither on the threshold voltage nor on the on-current are observed. Similarly, Figure 5-19 shows the I_D-V_G curves corresponding to Z-aligned traps at two different locations along the channel length. In this case, important changes are observed in the threshold voltage depending on the trap position along the channel: V_{th} decreases when the x-aligned traps move towards the drain edge.

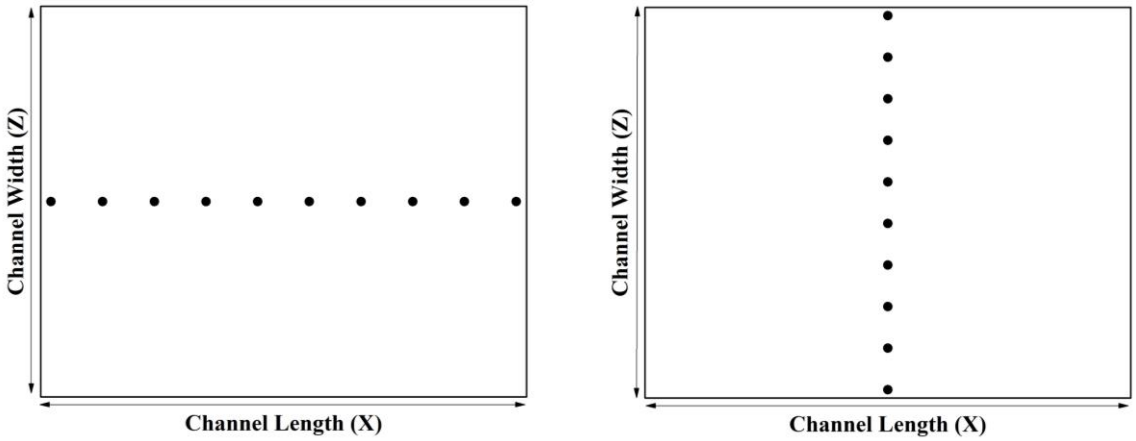


Figure 5-17: Schematic of the special trap distributions (a) X-aligned traps (b) Z-aligned traps.

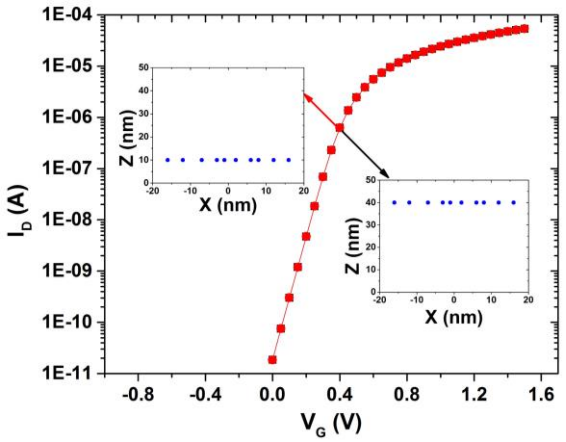


Figure 5-18: I_DV_G curves of two devices with x-aligned traps. The curves show no changes in the device characteristics.

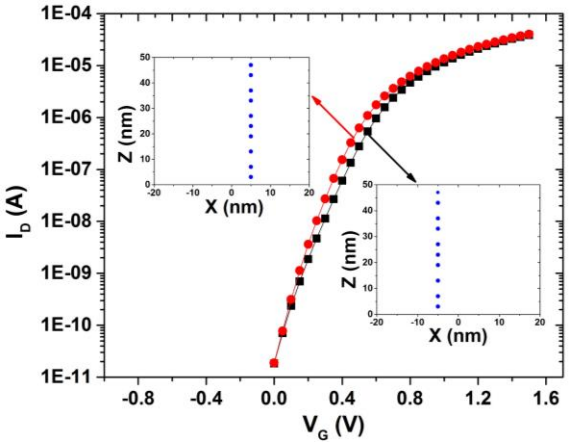


Figure 5-19: I_DV_G curves of two devices with x-aligned traps, the curves show negligible changes in the characteristics.

However, to emphasize the importance of the alignment of traps along the width, we consider a distribution where the traps are aligned along the width but at a location along the length that is closer to the drain edge. Figure 5-20 compares the I_D-V_G curves of this special distribution to a distribution with no z-alignment (x- aligned traps). The red (circles) curve corresponds to the

device with x-aligned traps and with a (V_{th}, I_{on}) vector of (0.313V, 53.61µA). The black (squares) curve corresponds to the device with z-aligned traps and with a (V_{th}, I_{on}) vector of (0.316V, 40.22µA). From these values, we can see that in this case, the alignment of the traps along the width (black, squares) has resulted in a change of 3mV in the V_{th} and a change of about 13 µA in the I_{on} from the device where the traps were aligned along the length (red, circles). It is clear from this that this alignment has greatly influenced the I_{on} while not be so effective on V_{th}.

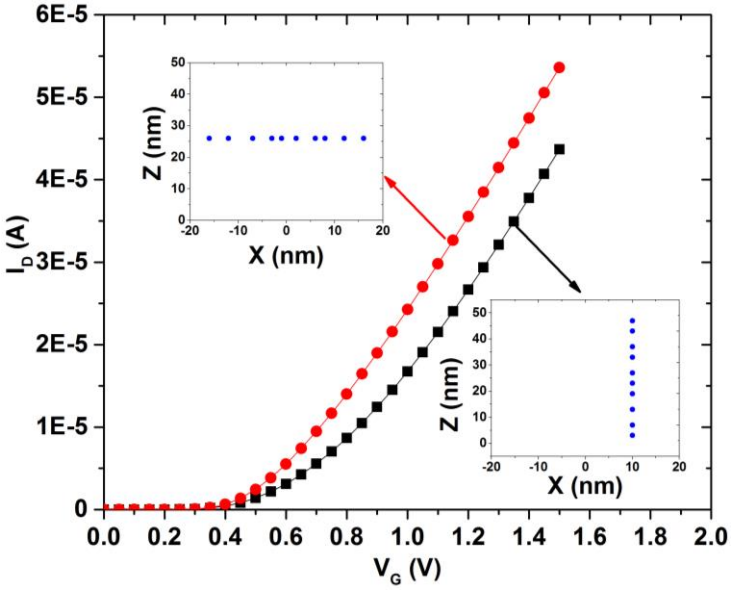


Figure 5-20: I_DV_G curves of two devices with x-aligned traps (circles) and z-aligned traps (squares), as a function of V_G. The devices have same V_{th} but different I_{on}. The insets show the 2D maps of the corresponding trap distributions.

The influence of the alignment on the current is further evidenced in the map of the conduction current as shown in Figure 5-21. The map of the conduction current on the left in Figure 5-21 corresponds to the device with red (circles) curve in Figure 5-20. It can be seen from the current map that the barrier created by the traps aligned along the length is not seen by all the carriers and those that are affected can go around the barrier to reach the drain. Hence, we can see that this distribution does not result in a large reduction of the drain current. Similarly, the picture on the right in Figure 5-21 shows the total current density calculated in case of the device with black (squares) curve in Figure 5-20. It can be seen that when traps are aligned along the channel width then the barrier created by them is seen by all the carriers resulting in a significant reduction of the current. Then these results suggest that the alignment of traps along the z coordinate has an important effect in the reduction of the device I_{on} current while their position along the channel length (x coordinate) influences the threshold voltage.

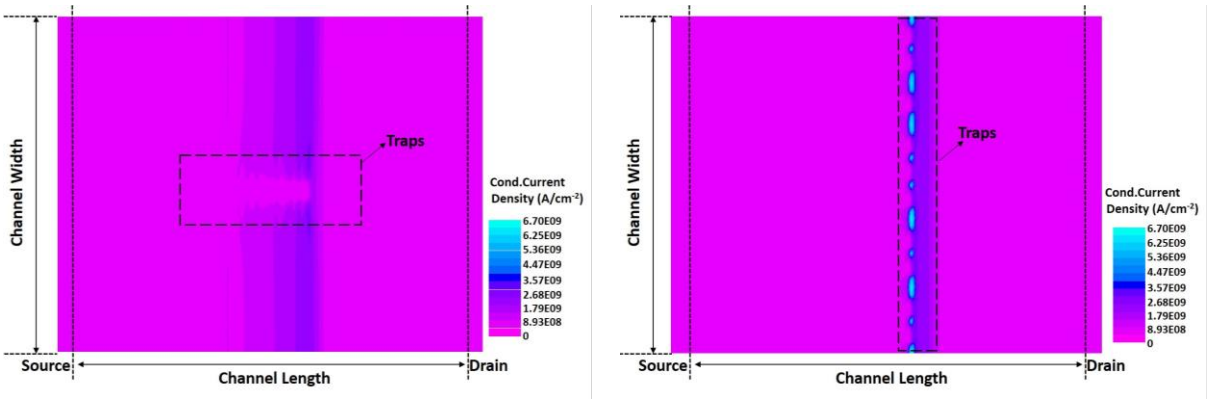


Figure 5-21: (left) The total current density calculated in the case of the device in Figure 5-20 with traps aligned along the length of the channel. There is an increase in the current around the traps, as carriers can go around them on their way to reach the drain. (right) The total current density calculated in the case of the device in Figure 5-20 with traps aligned along the width of the channel. Due to the alignment of the traps along the width, the barrier affects the carriers all along the channel and thus results in a significant decrease of current.

5.4 Influence of trap spatial distribution on the device electrical symmetry

The observed dependence of the electrical properties of the device on the spatial distribution of interface traps could break the intrinsic electrical symmetry of a MOSFET. To show this point, as an example, a device with trap distribution in Figure 5-22 is considered, and simulations are performed with the default source and drain biases (source is biased to 0V and drain is biased to 1.1V). To study the influence of the spatial distribution on the electrical symmetry, simulations are performed by swapping the source/drain biases (source is biased to 1.1V and drain is biased to 0V). The advantage of choosing the distribution as shown in Figure 5-22 is that for a default bias, the alignment of traps is closer to the source edge and when the biases are swapped, the alignment is closer to the drain edge. In other words, when the biases are swapped, it can be treated as the same device (with default bias) but with the trap distribution mirrored along the x coordinate, as shown in Figure 5-23.

The 'I-V' curves of the two devices (same trap distribution, default/swapped biasing) are shown in Figure 5-24 (left). It can be seen from the plots that mirroring the distribution along the channel length has modified the characteristics. In this particular case, initially (see Figure 5-22) most of the traps are close to the source, where they have the larger influence, leading to a (V_{th}-I_{on}) vector of (0.41V, 34.85µA). Then, swapping the biases meant that most traps are now located closer to the drain (see Figure 5-23) and results in a (V_{th}, I_{on}) vector of (0.36V, 36.5µA), i.e., V_{th} decreases because now traps are closer to the drain side. Consequently, the distribution of interface traps along the channel length introduces an electrical asymmetry in physically symmetrical devices, by changing the threshold voltage (from 0.41 V to 0.36 V, in the shown example). However, once the change in the V_{th} has been accounted for, changing the roles played by source and drain should have a small effect on I_{on} since trap alignment along the z-coordinate remains the same. This is

corroborated in Figure 5-24 (right), where the two 'I-V' curves overlap when differences in V_{th} are accounted for by plotting the drain current as a function of V_G-V_{th}.

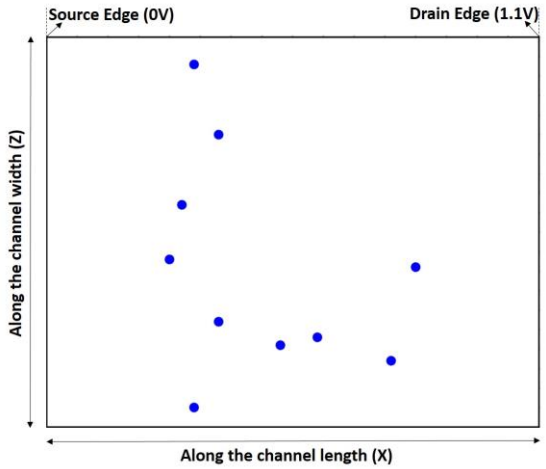


Figure 5-22: Distribution of the traps along the 2D area with default bias conditions. The source is biased at 0V and the drain is biased at 1.1V.

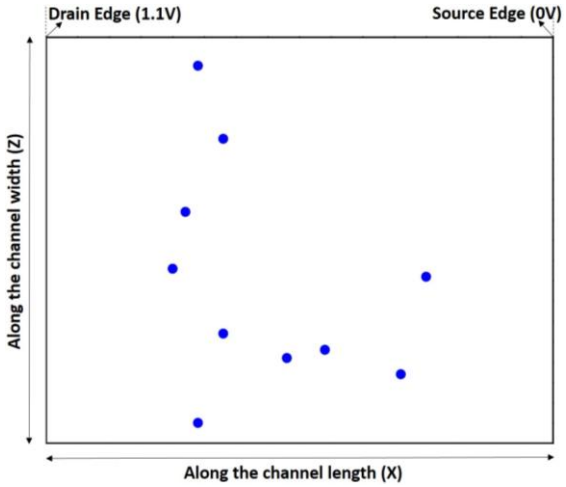


Figure 5-23: Mirroring the distribution of traps shown in Figure 5-22 along the length to reflect the swapping of the biases.

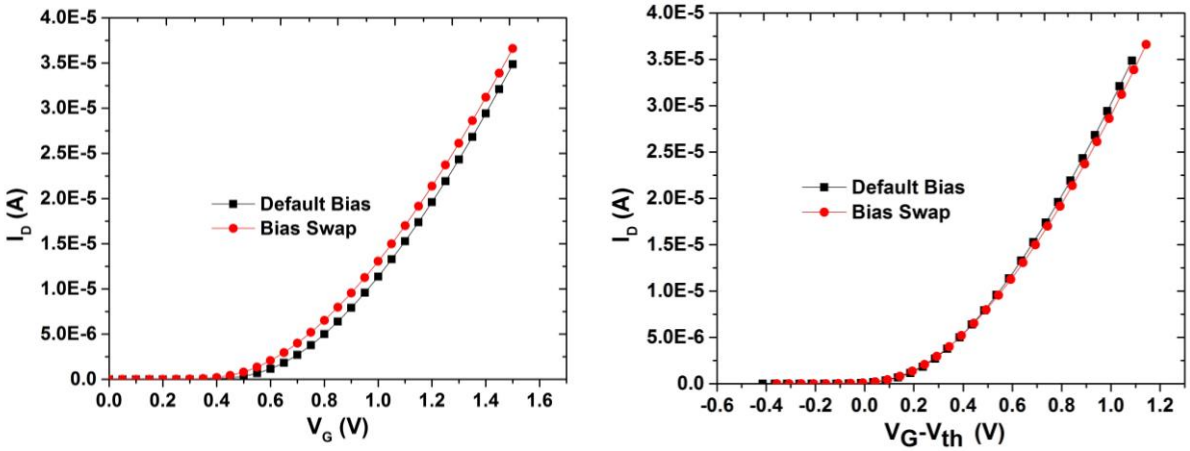


Figure 5-24: I-V curves of the device with trap distribution as shown in Figure 5-22 (black, squares) and with trap distribution as shown in Figure 5-23 (red, circles) is plotted as a function of V_G (left) and V_G-V_{th} (right).

This asymmetry in identical devices has been experimentally verified in a set of several devices. As an example, Figure 5-25 depicts statistical data measured in 50 different SOI MOSFETs. Black squares in the figure correspond to the threshold voltage of the device when drain (source grounded) was biased to 50mV while the red circles correspond to when the source (drain grounded) was biased to 50mV. It can be seen from the figure that for some devices, the V_{th} measured when bias is swapped overlap while for the others they don't. Two such cases have been marked in the figure, 'A' being the case where the bias swapping resulted in different V_{th} and 'B' being the case where the swapping of biases yielded no change in the V_{th}. The figure clearly shows the asymmetry (when the points do not overlap) in the measured electrical characteristics.

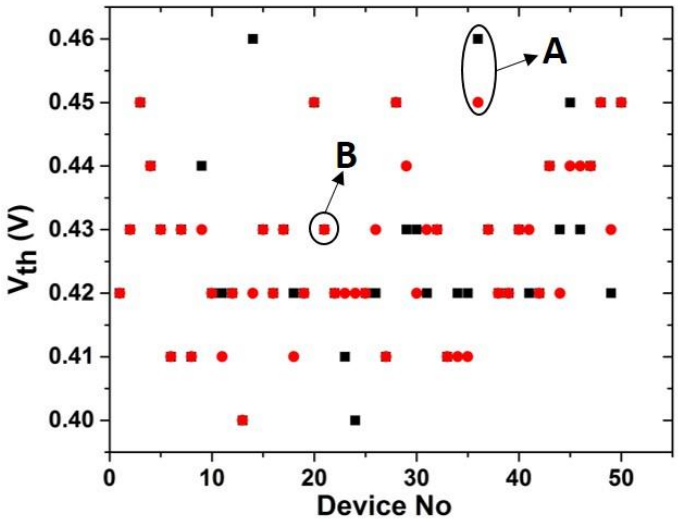


Figure 5-25: V_{th} measured in 50 MOSFETs, when source and drain are swapped. Black squares correspond to V_{th} when source is grounded and drain is biased to 50mV. Red circles correspond to V_{th} when drain is grounded and source is biased to 50mV.

Similar behavior is also observed in devices based on bulk MOSFET technology. Figure 5-26: left plots the I_DV_G curve ((black squares) of a device with drain biased at 350mV and source biased at 0V and the red curve corresponds to the I_DV_G curve (red circles) of the same device with drain biased to 0V and source biased to 350mV. The same figure (on the left) shows the zoom of the portion of the curve where the curves don't coincide.

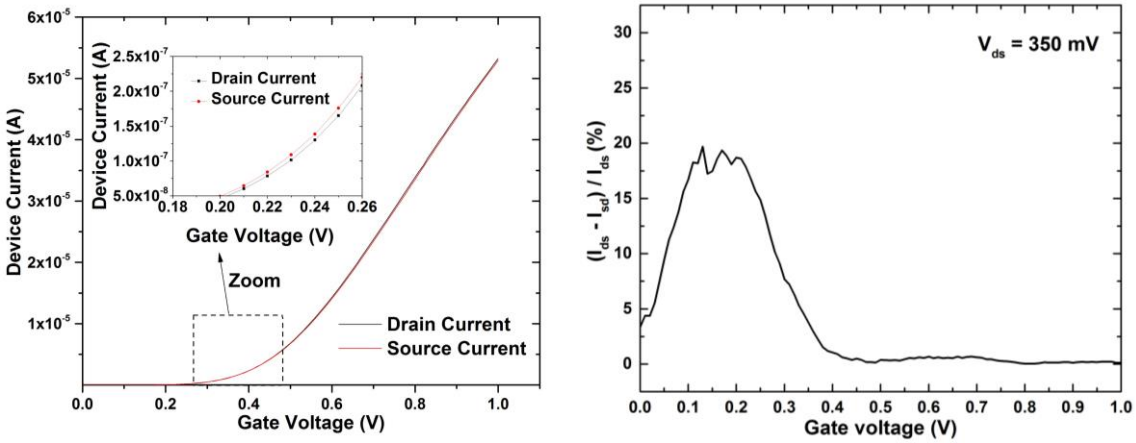


Figure 5-26: (left) The black curve corresponds to the I-V curve of the device with drain biased to 350mV and source biased to 0V. The red curve corresponds to the I-V curve of the same device with source biased to 350mV and drain biased to 0V. The zoom (inset in the left figure) swapping the biases has led to a change in the curves. On the right the ration of the change in the current is plotted as function of the gate voltage.

We can see that the swapping of the biases has led to changes in the current curves at lower gate voltages. The plot on the right (Figure 5-26) shows the relative percentage change in the current

(when the biases are swapped) as a function of the gate voltage. 'I_{ds}' is the current when the drain is biased to 350mV while the source is grounded and 'I_{sd}' is the current when the source is biased to 350mV while the drain is grounded. It can be seen from the figure that the difference is higher at the lower gate voltages (where V_{th} is extracted). From Figure 5-25 and Figure 5-26, we can see that the observed asymmetry seems to be independent of the technology, as the effect is seen in both SOI (Figure 5-25) and bulk (Figure 5-26) technologies. The observed asymmetry could be partially (maybe there are other effects that are beyond the scope of this work) related to random changes in the interface trap distribution from device to device, as expected for a random variability source.

Thus in this work, we have investigated and discussed the influence of the number and spatial distribution of interface traps as a variability source. In the next chapter we summarize the main ideas presented in this thesis and also outline the important conclusions of this work.

CHAPTER VI

Summary and Conclusions

Over the last decades, driven by the continuous demand for high performance, MOSFET devices have undergone a shrinkage in the size at an exponential rate. Although this downscaling has enhanced the device performance, it has brought a host of issues that have become extremely relevant with every generation of the technological node. Apart from the short channel effects (SCE), variability in transistor characteristics is one of the primary roadblocks to device scaling. Despite considerable efforts to control extrinsic process variations, there exist intrinsic sources of dispersion in devices, which arise due to the discrete nature of matter and charge. Among others, interface states have been identified as a prominent source of device variability. As devices are scaled to nano dimensions, these inherent dispersions cause significant fluctuations in the device characteristics among nominally identical transistors. This transistor mismatch might have a significant impact not only on analog circuits but also on digital circuits. Studies using computer aided simulations provide an excellent way in understanding and evaluating the impact of these variation sources on the device performance. In this thesis, commercially available TCAD tools from Silvaco have been used to study the impact of the number and spatial distribution of charged interface traps on the variability of transistor characteristics (threshold voltage (V_{th}), on-current (I_{on})). We start our analysis by performing 2D simulations and then extend the simulation into 3D. In following lines, we summarize the work done in the thesis.

Simulation Approach (Valid for both 2D and 3D)

- We first perform simulations of the device without traps and the characteristics (referred to as the pristine device characteristics) obtained is used as a reference. The impact of the interface traps is measured by monitoring the variation in the characteristics of the device (with traps) from the pristine device characteristics.
- To make a comparative analysis, we first consider a homogenous distribution of traps along the entire interface with a density that produces an appreciable change in the device characteristics.
- The impact of the spatial distribution of traps are studied by confining the traps at different sites along the interface. We start the analysis of the impact of the spatial distribution by considering the traps to be placed at a single location along the 1D (in 2D simulations) or 2D (in 3D simulations) interface and then move onto the situation where the traps are spread across multiple locations along the interface.
- As we are only interested in analyzing the impact of spatial distribution of traps, the total interface charge in the device is always maintained constant. Hence the density of traps (D_{TL}) at each trap site is scaled with number of locations (where traps are placed) considered.

Main results from 2D simulations

- We start the study by considering traps to be placed at a single location along the 2D interface. Simulations are performed by moving the position of this trap site from the source end to the drain end of the channel.
- The location of the traps along the channel significantly impacted the device performance by increasing the threshold voltage. The maximum shift (from the case of pristine V_{th}) in the threshold voltage of a device with a channel length of 250nm is obtained when traps were placed close to the drain edge of the channel. The results were interpreted by considering the increase in the height of the barrier (created by the charged interface traps) to the carrier flow, which is larger when traps were located at the drain edge of the channel, thus resulting larger ΔV_{th}.
- The influence of the drain bias was also investigated, concluding that larger change in V_{th} happens when traps were located close to the channel center. However, irrespective of the applied drain bias, the maximum threshold voltage always happens when traps were placed close to the drain end of the channel.
- As channel lengths were reduced from 250nm to 50nm, the location of the trap (along the interface) that produced maximum threshold voltage (or its shift) moved from the drain end of the channel to the center of the channel.

Next, we analyzed the situation where traps were spread across multiple locations along the interface, starting our analysis from the simplest case of two trap sites and then moving to the case of traps placed at multiple sites along the 2D channel interface. As modern devices have shorter

channel lengths, this analysis (of traps spread across multiple locations) was done by considering a device with a channel length of 50nm.

- Results from the simulations have shown that, only those traps that were located close to the channel center contributed significantly towards the total V_{th}.
- To analyze the influence of the trap locations on the variability of V_{th}, several devices with random positions for the two trap sites were simulated. It was found that the variability in the V_{th} (σV_{th}) and the value of average V_{th} decreased in comparison to the situation when the same number of traps were located at a single location.
- Simulations were further performed by considering devices with traps randomly located at multiple locations (>2) along the interface. The results showed that as the number of trap sites were increased, the variability in the V_{th} (σV_{th}) and the average value of V_{th} (across several devices) decreased. The decrease in the average V_{th} with increasing number of trap locations was understood by taking into account both the location of the traps along the channel and the scaled charge per location. The scaled charge at each trap site contributed to the total V_{th} by a smaller amount, and moreover, only those traps that were placed close to the channel center contributed significantly towards the overall V_{th}. As for the variability, larger number of trap locations would imply a smaller number of position combinations for the traps to be located at influential positions (close to the center), so that the randomness decreases.

To have a more realistic understanding of the variability of V_{th} due to the spatial distribution of interface traps, 3D simulations were performed by considering the traps along a 2D interface (in a 3D structure) to study the influence of trap distribution not just along the length but also along the channel width. We considered a reference device with a short channel length (50nm) and a width (20nm), dimensions that did not adversely influence the simulations time. As it was done in the case of 2D, we start with the simplest case of a single location along the 2D interface and move onto the case of multiple locations.

The results of the 3D simulations are summarized below:

- The location of the traps was varied both along the length (x-direction) and along the width (z-direction). The results showed the threshold voltage (V_{th}) is only influenced by the location of the trap along the channel length. Moreover, as observed in 2D, those traps that were located closer to the channel center asserted larger influence on V_{th}.
- To help understand more complex situations, traps were first placed at two predefined sites along the interface. The location of one trap site was fixed (both along the length and along the width) while the location of the other trap site was varied along the length or along the width. When trap sites were aligned along the length then, the resulting overall V_{th} is mostly contributed by traps that were located close to the channel center. Moreover, when the trap sites were positioned close enough to each other, the overall influence of traps (from both trap sites) on the V_{th} remarkably decreases (compared to when they are apart).
- When the trap sites were aligned along the width, the overall influence of traps (from both trap sites) on the ΔV_{th} was less than the superposition of the individual ΔV_{th}'s (when traps were only located at each trap site individually). And as the separation between the trap

sites increased, it resulted in an increase in the value of ΔV_{th} that is greater than the superposition. It was found that after ΔV_{th} reaches a maximum value, it (ΔV_{th}) tends to slowly decrease with further increase of the separation (between the trap sites) eventually saturating (to a value that is ≈ to the superposition) for large values of trap separation.

- These behaviors were interpreted by considering the effective barrier area created by the traps, which depends on the separation between them and their locations along the channel length and width. Larger barrier area resulted in an overall increase of the V_{th}.
- When the traps were located at two different random sites, both the average V_{th} and σV_{th} increased in comparison to the single location case. The increase of the average V_{th} was shown to be due to higher probability of having traps at sites that were located further from each other (along the length or along the width) while the increase in σV_{th} was shown to be related to having traps at an additional site compared to the case of having them at a single trap location.

Next moving to a more realistic case, to evaluate the influence of traps spread across multiple trap locations (>2) on the variability in V_{th}, devices with different number of trap locations and random distribution of the trap sites were simulated.

- The results showed a 'turn-around effect': i.e. for small enough number of locations, an increased value of average V_{th} and σV_{th} is observed. But, as we increased the number of trap locations, a decrease is observed. The initial increase in the value of average V_{th} was attributed to the increase in the effective barrier area (with trap sites not very close to each other) associated with multiple trap locations. The decrease in the average V_{th} for large enough number of trap locations was explained by the increasing probability (with increasing number of trap locations) of having trap sites very close to each other, resulting in a decrease of the barrier area (which resulted in a decrease of the overall V_{th}) accompanied by the scaling of the charge (with increasing number of trap locations) at each trap site. The initial increase of σV_{th} is attributed to the additional variability introduced by the channel width. However, given a certain width, this additional variability is saturated beyond a certain number of trap locations, hence, recovering the previously observed 2D behavior.
- The dependence of the V_{th} variability on the device dimensions due to interface traps was investigated. This dependence of σV_{th} on the device geometry (length, width) followed the generally accepted Pelgrom's rule (σV_{th} α $1/\sqrt{Area}$), however, it was found that this dependency is largely determined by the width of the transistors.

Finally, we studied the simultaneous influence of the spatial distribution of single trap/multiple traps on the threshold voltage (electrostatics) and on the on-current (conduction) of the device. Since I_{on} (on-current) is affected by the V_{th} (threshold voltage), the main results were interpreted taking into account the previous observations.

- When a single trap along the 2D interface is considered, as expected, an increase in V_{th} is followed by a decrease in the I_{on}. When the trap is located close to the channel center, then it asserted larger influence on the device characteristics (V_{th} and I_{on}).
- However, when multiple traps were considered, although in general an increase in V_{th} is followed by a decrease in the I_{on}, there were certain distributions of traps that led to a non-correlated V_{th} and I_{on}. The analysis of the trap distributions showed that the distribution of

traps in the channel direction (along the length) mainly determined the threshold voltage, leading to larger V_{th} when traps are located close to the channel center. Additionally, the trap distribution along the channel width mostly effects the I_{on}. Lower on-currents are obtained for distributions with greater alignment of traps along the channel width. This influence of spatial distribution of the traps can partially explain the experimentally observed asymmetry in the I_D-V_G characteristics of devices.

Therefore, to conclude, the work done in this thesis has shown that the number and spatial distribution of interface traps is a significant contributor to the variability of device performance in ultra-scaled MOSFETs. This work could be continued to further investigate the role of interface traps on the device variability in dynamic situations under RTN and BTI.

References

[1] W. F. Brinkman, D. E. Haggan, and W. W. Troutman, "A history of the invention of the transistor and where it will lead us," *IEEE J. Solid-State Circuits*, vol. 32, no. 12, pp. 1858–1864, 1997.

[2] "International Technology Roadmap for Semiconductors," 2009.

[3] N. H. E. Weste and D. F. Harris, *Book: CMOS VLSI Design: A Circuits and Systems Perspective*, 4th ed. Addison Wesley, 2011.

[4] Y. Taur and T. H. Ning, *Book: Fundamentals of Modern VLSI Devices*, vol. 25, no. 11. Cambridge University Press, 2002.

[5] R. Chau, S. Datta, M. Doczy, B. Doyle, B. Jin, J. Kavalieros, A. Majumdar, M. Metz, and M. Radosavljevic, "Benchmarking nanotechnology for high-performance and low-power logic transistor applications," *IEEE Trans. Nanotechnol.*, vol. 4, no. 2, pp. 153–158, Mar. 2005.

[6] J. J. Liou, A. Ortiz-Condez, and F. G. Sanchez, "Extraction of the threshold voltage of MOSFETs: an overview," in *IEEE Proceedings Electron Devices Meeting*, 1997, pp. 31–38.

[7] M. Tsuno, M. Suga, M. Tanaka, K. Shibahara, M. Miura-Mattausch, and M. Hirose, "Physically-based threshold voltage determination for MOSFET's of all gate lengths," *IEEE Trans. Electron Devices*, vol. 46, no. 7, pp. 1429–1434, Jul. 1999.

[8] A. Ortiz-Conde, F. J. García Sánchez, J. J. Liou, A. Cerdeira, M. Estrada, and Y. Yue, "A review of recent MOSFET threshold voltage extraction methods," *Microelectron. Reliab.*, vol. 42, no. 4–5, pp. 583–596, 2002.

[9] J. J. Liou, A. Ortiz-Conde, and F. Garcia-Sanchez, *Book: Analysis and Design of MOSFETs*, 1st ed. Springer, 2012.

[10] G. E. Moore, "Cramming more components onto integrated circuits," *Proc. IEEE*, vol. 86, no. 1, pp. 82–85, Jan. 1998.

[11] Cringely.com, "Breaking Moores-Law," 2013. [Online]. Available: http://www.cringely.com/2013/10/15/breaking-moores-law/.

[12] R. H. Dennard, F. H. Gaensslen, H. N. Yu, V. L. Rideout, E. Bassous, and A. R. Leblanc, "Design of Ion-Implanted MOSFET's With Very Small Physical Dimensions," *IEEE J. Solid-State Circuits*, vol. 9, no. 5, pp. 256–268, Apr. 1974.

[13] D. L. Critchlow, "Mosfet Scaling-the Driver of VLSI Technology," *Proc. IEEE*, vol. 87, no. 4, pp. 659–667, Apr. 1999.

[14] M. Bohr, “A 30 Year Retrospective on Dennard’s MOSFET Scaling Paper,” *IEEE Solid-State Circuits Newsl.*, vol. 12, no. 1, pp. 11–13, 2007.

[15] P. J. Tsang, W. W. Walker, J. F. Shepard, D. L. Critchlow, and S. Ogura, “Fabrication of High-Performance Lddfet’S With Oxide Sidewall-Spacer Technology.,” *IEEE Trans. Electron Devices*, vol. 29, no. 4, pp. 590–596, Apr. 1982.

[16] S. Wolf, *Book: Silicon processing for the VLSI era - vol 2: Process integration*. Lattice Press, 1990.

[17] P. J. French, “Polysilicon: A versatile material for microsystems,” *Sensors Actuators, A Phys.*, vol. 99, no. 1–2, pp. 3–12, 2002.

[18] F. Faggin and T. Klein, “Silicon gate technology,” *Solid State Electron.*, vol. 13, no. 8, pp. 1125–1144, 1970.

[19] S. H. Lo, D. A. Buchanan, Y. Taur, and W. Wang, “Quantum-mechanical modeling of electron tunneling current from the inversion layer of ultra-thin-oxide nMOSFET’s,” *IEEE Electron Device Lett.*, vol. 18, no. 5, pp. 209–211, May 1997.

[20] C. Te Chuang, K. Bernstein, R. V. Joshi, R. Puri, K. Kim, E. J. Nowak, T. Ludwig, and I. Aller, “Scaling planar silicon devices,” *IEEE Circuits Devices Mag.*, vol. 20, no. 1, pp. 6–19, Jan. 2004.

[21] K. A. Bowman, L. Wang, X. Tang, and J. D. Meindl, “A circuit-level perspective of the optimum gate oxide thickness,” *IEEE Trans. Electron Devices*, vol. 48, no. 8, pp. 1800–1810, 2001.

[22] H. Wong and H. Iwai, “On the scaling issues and high-k replacement of ultrathin gate dielectrics for nanoscale MOS transistors,” *Microelectron. Eng.*, vol. 83, no. 10, pp. 1867–1904, 2006.

[23] E. P. Gusev, D. A. Buchanan, E. Cartier, A. Kumar, D. DiMaria, S. Guha, A. Callegari, S. Zafar, P. C. Jamison, D. A. Neumayer, M. Copel, M. A. Gribelyuk, H. Okorn-Schmidt, C. D’Emic, P. Kozlowski, K. Chan, N. Bojarczuk, L.-A. Ragnarsson, P. Ronsheim, K. Rim, R. J. Fleming, A. Mocuta, and A. Ajmera, “Ultrathin high-K gate stacks for advanced CMOS devices,” in *International Electron Devices Meeting*, 2001, pp. 20.1.1–20.1.4.

[24] R. M. Wallace and G. D. Wilk, “High-κ Dielectric Materials for Microelectronics,” *Crit. Rev. Solid State Mater. Sci.*, vol. 28, no. 4, pp. 231–285, 2003.

[25] M. Wu, Y. I. Alivov, and H. Morkoç, “High-κ dielectrics and advanced channel concepts for Si MOSFET,” *J. Mater. Sci. Mater. Electron.*, vol. 19, no. 10, pp. 915–951, 2008.

[26] C. R., “A 45 nm Logic Technology with High-K+ Metal Gate Transistors, Strained Silicon, 9 Cu Interconnect Layers, 193 nm Dry Patterning, and 100% Pb-Free Packaging,” in *Technol. Dig. - Int. Electron Devices Meet. IEDM*, 2007, p. 247.

[27] C. Hobbs, L. Fonseca, V. Dhandapani, S. Samavedam, B. Taylor, J. Grant, L. Dip, D. Triyoso, R. Hegde, D. Gilmer, R. Garcia, D. Roan, L. Lovejoy, R. Rai, L. Hebert, H. Tseng, B. White,

and P. Tobin, “Fermi level pinning at the polySi/metal oxide interface,” in *Symposium on VLSI Technology. Digest of Technical Papers*, 2003, pp. 3–4.

[28] H. S. P. Wong, “Beyond the conventional transistor,” *Solid. State. Electron.*, vol. 49, no. 5, pp. 755–762, 2005.

[29] S. E. Thompson, M. Armstrong, C. Auth, M. Alavi, M. Buehler, R. Chau, S. Cea, T. Ghani, G. Glass, T. Hoffman, C.-H. Jan, C. Kenyon, J. Klaus, K. Kuhn, Z. Ma, B. Mcintyre, K. Mistry, A. Murthy, B. Obradovic, R. Nagisetty, P. Nguyen, S. Sivakumar, R. Shaheed, L. Shifren, B. Tufts, S. Tyagi, M. Bohr, and Y. El-Mansy, “90-Nm Logic Technology Featuring Strained-Silicon,” *IEEE Trans. Electron Devices*, vol. 51, no. 11, pp. 1790–1797, 2004.

[30] C. H. Lee, C. Lu, T. Nishimura, K. Nagashio, and A. Toriumi, “Thermally robust CMOS-aware Ge MOSFETs with high mobility at high-carrier densities on a single orientation Ge substrate,” in *Digest of Technical Papers - Symposium on VLSI Technology*, 2014, pp. 1–2.

[31] J.-P. Colinge, *Book: Silicon-on-Insulator Technology: Materials to VLSI*. 1997.

[32] D. Hisamoto, W. C. Lee, J. Kedzierski, H. Takeuchi, K. Asano, C. Kuo, E. Anderson, T. J. King, F. Jeffrey Bokor, and C. Hu, “FinFET-A self-aligned double-gate MOSFET scalable to 20 nm,” *IEEE Trans. Electron Devices*, vol. 47, no. 12, pp. 2320–2325, 2000.

[33] I. Ferain, C. a. Colinge, and J.-P. Colinge, “Multigate transistors as the future of classical metal–oxide–semiconductor field-effect transistors,” *Nature*, vol. 479, no. 7373, pp. 310–316, 2011.

[34] Y.-B. Kim, “Challenges for Nanoscale MOSFETs and Emerging Nanoelectronics,” *Trans. Electr. Electron. Mater.*, vol. 11, no. 3, pp. 93–105, 2010.

[35] C. M. Lieber and Z. L. Wang, “Functional Nanowires,” *MRS Bull.*, vol. 32, no. 02, pp. 99–108, 2007.

[36] C.-H. Jan, U. Bhattacharya, R. Brain, S.-J. Choi, G. Curello, G. Gupta, W. Hafez, M. Jang, M. Kang, K. Komeyli, T. Leo, N. Nidhi, L. Pan, J. Park, K. Phoa, A. Rahman, C. Staus, H. Tashiro, C. Tsai, P. Vandervoorn, L. Yang, J.-Y. Yeh, and P. Bai, “A 22nm SoC platform technology featuring 3-D tri-gate and high-k/metal gate, optimized for ultra low power, high performance and high density SoC applications,” in *International Electron Devices Meeting*, 2012, vol. 48, no. 8, pp. 3.1.1–3.1.4.

[37] C. H. Jan, F. Al-Amoody, H. Y. Chang, T. Chang, Y. W. Chen, N. Dias, W. Hafez, D. Ingerly, M. Jang, E. Karl, S. K. Y. Shi, K. Komeyli, H. Kilambi, A. Kumar, K. Byon, C. G. Lee, J. Lee, T. Leo, P. C. Liu, N. Nidhi, R. Olac-Vaw, C. Petersburg, K. Phoa, C. Prasad, C. Quincy, R. Ramaswamy, T. Rana, L. Rockford, A. Subramaniam, C. Tsai, P. Vandervoorn, L. Yang, A. Zainuddin, and P. Bai, “A 14 nm SoC platform technology featuring 2nd generation Tri-Gate transistors, 70 nm gate pitch, 52 nm metal pitch, and 0.0499 um2 SRAM cells, optimized for low power, high performance and high density SoC products,” in *Symposium on VLSI Technology (VLSI Technology)*, 2015, pp. T12–T13.

[38] S. Ramey, A. Ashutosh, C. Auth, J. Clifford, M. Hattendorf, J. Hicks, R. James, A. Rahman, V.

Sharma, A. St Amour, and C. Wiegand, “Intrinsic transistor reliability improvements from 22nm tri-gate technology,” in *IEEE International Reliability Physics Symposium Proceedings*, 2013, pp. 4C.5.1–4C.5.5.

[39] B. G. Streetman and S. K. Banerjee, *Book: Solid State Electronic Devices*. 2009.

[40] Q. Zhang, W. Zhao, and A. Seabaugh, “Low-subthreshold-swing tunnel transistors,” *IEEE Electron Device Lett.*, vol. 27, no. 4, pp. 297–300, Apr. 2006.

[41] A. Afzalian, J. P. Colinge, and D. Flandre, “Physics of Gate Modulated Resonant Tunneling (RT)-FETs: Multi-barrier MOSFET for steep slope and high on-current,” *Solid. State. Electron.*, vol. 59, no. 1, pp. 50–61, 2011.

[42] K. V. Prakash, A. Kumar, P. Jain, D. Sgdcoe, and M. Jalgaon, “Circumventing Short Channel Effects in FETs: Review,” *Int. J. Comput. Appl.*, vol. 117, no. 17, pp. 975–8887, 2015.

[43] Y. Awano, “Graphene for VLSI: FET and interconnect applications,” in *Technical Digest - International Electron Devices Meeting, IEDM*, 2009, pp. 1–4.

[44] F. Schwierz, “Graphene transistors.,” *Nat. Nanotechnol.*, vol. 5, no. 7, pp. 487–496, Jul. 2010.

[45] Y. Taur, C. H. Wann, and D. J. Frank, “25 nm CMOS design considerations,” in *International Electron Devices Meeting. Technical Digest*, 1998, pp. 789–792.

[46] T. Ghani, K. Mistry, P. Packan, S. Thompson, M. Stealer, S. Tyagi, M. Bohr, and M. Stettler, “Scaling challenges and device design requirements for high performance sub-50 nm gate length planar CMOS transistors,” in *Symposium on VLSI Technology. Digest of Technical Papers*, 2000, pp. 174–175.

[47] R. W. Keyes, “The effect of randomness in the distribution of impurity atoms on FET thresholds,” *Appl. Phys.*, vol. 8, no. 3, pp. 251–259, 1975.

[48] H. Philip wong, Y. Taur, and D. Frank, “Discrete random dopant distribution effects in nanometer-scale MOSFETs,” *Microelectron. Reliab.*, vol. 38, no. 9, pp. 1447–1456, 1998.

[49] G. Roy, A. R. Brown, F. Adamu-Lema, S. Roy, and A. Asenov, “Simulation study of individual and combined sources of intrinsic parameter fluctuations in conventional nano-MOSFETs,” *IEEE Trans. Electron Devices*, vol. 53, no. 12, pp. 3063–3069, 2006.

[50] K. Kuhn, C. Kenyon, A. Kornfeld, M. Liu, A. Maheshwari, W. Shih, S. Sivakumar, G. Taylor, P. Vandervoorn, and K. Zawadzki, “Minimizing Process Variation in Intel’s 45nm CMOS Technology,” *Intel Technol. J.*, vol. 12, no. 02, pp. 93–110, 2008.

[51] K. Agarwal and S. Nassif, “Characterizing process variation in nanometer CMOS,” in *44th annual Design Automation Conference*, 2007, pp. 396–399.

[52] K. J. Kuhn, M. D. Giles, D. Becher, P. Kolar, A. Kornfeld, R. Kotlyar, S. T. Ma, A. Maheshwari, and S. Mudanai, “Process technology variation,” *IEEE Trans. Electron Devices*, vol. 58, no. 8, pp. 2197–2208, Aug. 2011.

[53] K. Takeuchi, A. Nishida, and T. Hiramoto, “Random fluctuations in scaled MOS devices,” in *International Conference on Simulation of Semiconductor Processes and Devices, SISPAD*, 2009, pp. 1–7.

[54] V. Bonfiglio and G. Iannaccone, “An Approach Based on Sensitivity Analysis for the Evaluation of Process Variability in Nanoscale MOSFETs,” *Electron Devices, IEEE Trans.*, vol. 58, no. 8, pp. 2266–2273, 2011.

[55] A. R. Brown, G. Roy, and A. Asenov, “Poly-Si-gate-related variability in decananometer MOSFETs with conventional architecture,” *IEEE Trans. Electron Devices*, vol. 54, no. 11, pp. 3056–3063, 2007.

[56] X. Wang, A. R. Brown, B. Cheng, and A. Asenov, “Statistical variability and reliability in nanoscale FinFETs,” in *Technical Digest - International Electron Devices Meeting, IEDM*, 2011, pp. 5.4.1–5.4.4.

[57] A. Asenov, S. Roy, R. A. Brown, G. Roy, C. Alexander, C. Riddet, C. Millar, B. Cheng, A. Martinez, N. Seoane, D. Reid, M. F. Bukhori, X. Wang, and U. Kovac, “Advanced simulation of statistical variability and reliability in nano CMOS transistors,” in *Technical Digest - International Electron Devices Meeting, IEDM*, 2008, p. 1.

[58] K. Nagase, S. I. Ohkawa, M. Aoki, and H. Masuda, “Variation status in 100nm CMOS process and below,” in *The International Conference on Microelectronic Test Structures (ICMTS)*, 2004, pp. 257 – 261.

[59] A. Cathignol, B. Cheng, D. Chanemougame, A. R. Brown, K. Rochereau, G. Ghibaudo, and A. Asenov, “Quantitative evaluation of statistical variability sources in a 45-nm technological node LP N-MOSFET,” *IEEE Electron Device Lett.*, vol. 29, no. 6, pp. 609–611, 2008.

[60] T. Chen and S. Naffziger, “Comparison of Adaptive Body Bias (ABB) and Adaptive Supply Voltage (ASV) for Improving Delay and Leakage Under the Presence of Process Variation,” *IEEE Trans. Very Large Scale Integr. Syst.*, vol. 11, no. 5, pp. 888–899, 2003.

[61] A. Asenov, “Random dopant induced threshold voltage lowering and fluctuations in sub-0.1 µm MOSFET’s: A 3-D ‘atomistic’ simulation study,” *IEEE Trans. Electron Devices*, vol. 45, no. 12, pp. 2505–2513, 1998.

[62] H. Dadgour, K. Endo, D. Vivek, and K. Banerjee, “Modeling and analysis of grain-orientation effects in emerging metal-gate devices and implications for sram reliability,” in *Technical Digest - International Electron Devices Meeting, IEDM*, 2008, pp. 1–4.

[63] H. F. Dadgour, K. Endo, V. K. De, and K. Banerjee, “Grain-orientation induced work function variation in nanoscale metal-gate transistors—Part I: Modeling, analysis, and experimental validation,” *IEEE Trans. Electron Devices*, vol. 57, no. 10, pp. 2504–2514, 2010.

[64] H. F. Dadgour, K. Endo, V. K. De, and K. Banerjee, “Grain-orientation induced work function variation in nanoscale metal-gate transistors—Part II: Implications for process, device, and circuit design,” *IEEE Trans. Electron Devices*, vol. 57, no. 10, pp. 2515–2525, 2010.

[65] A. Asenov, S. Kaya, and A. R. Brown, “Intrinsic parameter fluctuations in decananometer MOSFETs introduced by gate line edge roughness,” *IEEE Trans. Electron Devices*, vol. 50, no. 5, pp. 1254–1260, May 2003.

[66] M. G. Pala, D. Esseni, and F. Conzatti, “Impact of interface traps on the IV curves of InAs Tunnel-FETs and MOSFETs: A full quantum study,” in *Technical Digest - International Electron Devices Meeting, IEDM*, 2012, pp. 6.6.1–6.6.4.

[67] D. Esseni and M. G. Pala, “Interface traps in InAs nanowire tunnel FETs and MOSFETs - Part II: Comparative analysis and trap-induced variability,” *IEEE Trans. Electron Devices*, vol. 60, no. 9, pp. 2802–2807, Sep. 2013.

[68] M. Toledano-Luque and B. Kaczer, “Degradation of time dependent variability due to interface state generation,” in *Symposium on VLSI Technology (VLSIT)*, 2013, no. 261868, pp. 190–191.

[69] D. M. Fleetwood, P. S. Winokur, R. A. Reber, T. L. Meisenheimer, J. R. Schwank, M. R. Shaneyfelt, and L. C. Riewe, “Effects of oxide traps, interface traps, and ‘border traps’ on metal-oxide-semiconductor devices,” *J. Appl. Phys.*, vol. 73, no. 10, pp. 5058–5074, 1993.

[70] Y. Qiu, R. Wang, Q. Huang, and R. Huang, “A comparative study on the impacts of interface traps on tunneling FET and MOSFET,” *IEEE Trans. Electron Devices*, vol. 61, no. 5, pp. 1284–1291, May 2014.

[71] S. Markov, B. Cheng, and A. Asenov, “Statistical variability in fully depleted SOI MOSFETs due to random dopant fluctuations in the source and drain extensions,” *IEEE Electron Device Lett.*, vol. 33, no. 3, pp. 315–317, Mar. 2012.

[72] B. Cheng, S. Roy, G. Roy, F. Adamu-Lema, and A. Asenov, “Impact of intrinsic parameter fluctuations in decanano MOSFETs on yield and functionality of SRAM cells,” in *Solid-State Electronics*, 2005, vol. 49, no. 5, pp. 740–746.

[73] B. Hoeneisen and C. A. Mead, “Fundamental limitations in microelectronics-I. MOS technology,” *Solid State Electron.*, vol. 15, no. 7, pp. 819–829, Jul. 1972.

[74] T. Hagiwara, K. Yamaguchi, and S. Asai, “Threshold Voltage Deviation in Very Small MOS Transistors Due to Local Impurity Fluctuations,” in *VLSI Technology, 1982. Digest of Technical Papers. Symposium on*, 1982, pp. 46–47.

[75] D. Reid, C. Millar, G. Roy, S. Roy, and A. Asenov, “Analysis of Threshold Voltage Distribution Due to Random Dopants: A 100 000-Sample 3-D Simulation Study,” *IEEE Trans. Electron Devices*, vol. 56, no. 10, pp. 2255–2263, Oct. 2009.

[76] Y. Li and C. H. Hwang, “High-frequency characteristic fluctuations of nano-MOSFET circuit induced by random dopants,” *IEEE Trans. Microw. Theory Tech.*, vol. 56, no. 12, pp. 2726–2733, 2008.

[77] N. Sano, K. Matsuzawa, M. Mukai, and N. Nakayama, “On discrete random dopant modeling in drift-diffusion simulations: Physical meaning of ‘atomistic’ dopants,”

Microelectron. Reliab., vol. 42, no. 2, pp. 189–199, 2002.

[78] C. Alexander, G. Roy, A. Asenov, and S. Member, "Random-Dopant-Induced Drain Current Variation Self-Consistent Monte Carlo Simulation Study Using ' Ab Initio ' Ionized Impurity Scattering," *Electron Devices, IEEE Trans.*, vol. 55, no. 11, pp. 3251–3258, 2008.

[79] G. Roy, "Thesis: Simulation of Intrinsic Parameter Fluctuations in Nano-{CMOS} Devices," University of Glasgow, 2005.

[80] C. H. Díaz, H. J. Tao, Y. C. Ku, A. Yen, and K. Young, "An experimentally validated analytical model for gate line-edge roughness (LER) effects on technology scaling," *IEEE Electron Device Lett.*, vol. 22, no. 6, pp. 287–289, 2001.

[81] H. W. Kim, J. Y. Lee, J. Shin, S. G. Woo, H. K. Cho, and J. T. Moon, "Experimental investigation of the impact of LWR on sub-100-nm device performance," *IEEE Trans. Electron Devices*, vol. 51, no. 12, pp. 1984–1988, 2004.

[82] H. Fukutome, Y. Momiyama, T. Kubo, Y. Tagawa, T. Aoyama, and H. Arimoto, "Direct evaluation of gate line edge roughness impact on extension profiles in sub-50-nm n-MOSFETs," *IEEE Trans. Electron Devices*, vol. 53, no. 11, pp. 2755–2763, 2006.

[83] D. Reid, C. Millar, S. Roy, A. Asenov, and S. Member, "Variability — Part I : Three-Dimensional Simulation of Large Statistical Samples," *Electron Devices, IEEE*, vol. 57, no. 11, pp. 2801–2807, 2010.

[84] S. Xiong and J. Bokor, "A simulation study of gate line edge roughness effects on doping profiles of short-channel MOSFET devices," *IEEE Trans. Electron Devices*, vol. 51, no. 2, pp. 228–232, 2004.

[85] A. Asenov and S. Saini, "Polysilicon gate enhancement of the random dopant induced threshold voltage fluctuations in sub-100 nm MOSFET's with ultrathin gate oxide," *IEEE Trans. Electron Devices*, vol. 47, no. 4, pp. 805–812, 2000.

[86] R. Difrenza, J. C. Vildeuil, P. Llinares, and G. Ghibaudo, "Impact of grain number fluctuations in the MOS transistor gate on matching performance," in *International Conference on Microelectronic Test Structures*, 2003, pp. 244–249.

[87] H. Watanabe, "Statistics of grain boundaries in polysilicon," *IEEE Trans. Electron Devices*, vol. 54, no. 1, pp. 38–44, 2007.

[88] S. Uma and A. McConnell, "Temperature Dependent Thermal Conductivity of Undoped Polycrystalline Silicon Layers," *Int. J. Thermophys.*, vol. 22, no. 2, pp. 605–616, 2001.

[89] H. Tuinhout, "Impact of Parametric Fluctuations on Performance and Yield of Deep-Submicron Technologies," in *32nd European Solid-State Device Research Conference*, 2002, pp. 95–102.

[90] M. Hane, T. Ikezawa, and T. Ezaki, "Coupled atomistic 3D process/device simulation considering both line-edge roughness and random-discrete-dopant effects," in *International Conference on Simulation of Semiconductor Processes and Devices,. SISPAD*,

2003, pp. 99–102.

[91] H. Fukutome, Y. Momiyama, T. Kubo, E. Yoshida, H. Morioka, M. Tajima, and T. Aoyama, “Suppression of poly-gate-induced fluctuations in carrier profiles of sub-50nm MOSFETs,” in *Technical Digest - International Electron Devices Meeting, IEDM*, 2006, pp. 1–4.

[92] A. Cathignol, K. Rochereau, and G. Ghibaudo, “Impact of a single grain boundary in the polycrystalline silicon gate on sub 100nm bulk MOSFET characteristics-Implication on matching properties,” in *7th Eur. Workshop ULIS*, 2006, pp. 145–148.

[93] E. P. Gusev, V. Narayanan, and M. M. Frank, “Advanced high-k dielectric stacks with polySi and metal gates: Recent progress and current challenges,” *IBM J. Res. Dev.*, vol. 50, no. 4.5, pp. 387–410, 2006.

[94] A. R. Brown, N. M. Idris, J. R. Watling, and A. Asenov, “Impact of metal gate granularity on threshold voltage variability: A full-scale three-dimensional statistical simulation study,” *IEEE Electron Device Lett.*, vol. 31, no. 11, pp. 1199–1201, 2010.

[95] H. Zhao, J. Huang, Y.-T. Chen, J. H. Yum, Y. Wang, F. Zhou, F. Xue, and J. C. Lee, “Effects of gate-first and gate-last process on interface quality of In[sub 0.53]Ga[sub 0.47]As metal-oxide-semiconductor capacitors using atomic-layer-deposited Al[sub 2]O[sub 3] and HfO[sub 2] oxides,” *Appl. Phys. Lett.*, vol. 95, no. 25, p. 253501, 2009.

[96] J. Martin-Martinez, R. Rodriguez, M. Nafria, and X. Aymerich, “Time-dependent variability related to BTI effects in MOSFETs: Impact on CMOS differential amplifiers,” *IEEE Trans. Device Mater. Reliab.*, vol. 9, no. 2, pp. 305–310, Jun. 2009.

[97] K. R. Lakshmikumar, R. A. Hadaway, and M. A. Copeland, “Characterization and Modeling of Mismatch in MOS Transistors for Precision Analog Design,” *IEEE J. Solid-State Circuits*, vol. 21, no. 6, pp. 1057–1066, Dec. 1986.

[98] M. J. M. Pelgrom, A. C. J. Duinmaijer, and A. P. G. Welbers, “Matching Properties of MOS Transistors,” *IEEE J. Solid-State Circuits*, vol. 24, no. 5, pp. 1433–1439, Oct. 1989.

[99] T. Mizuno, J. Okumtura, and A. Toriumi, “Experimental study of threshold voltage fluctuation due to statistical variation of channel dopant number in MOSFET’s,” *IEEE Trans. Electron Devices*, vol. 41, no. 11, pp. 2216–2221, 1994.

[100] K. Takeuchi, T. Fukai, T. Tsunomura, A. T. Putra, A. Nishida, S. Kamohara, and T. Hiramoto, “Understanding random threshold voltage fluctuation by comparing multiple fabs and technologies,” in *Technical Digest - International Electron Devices Meeting, IEDM*, 2007, pp. 467–470.

[101] T. Grasser, B. Kaczer, W. Goes, H. Reisinger, T. Aichinger, P. Hehenberger, P. J. Wagner, F. Schanovsky, J. Franco, P. Roussel, and M. Nelhiebel, “Recent advances in understanding the bias temperature instability,” in *Technical Digest - International Electron Devices Meeting, IEDM*, 2010, pp. 4.4.1–4.4.4.

[102] E. Amat, T. Kauerauf, R. Rodriguez, M. Nafria, X. Aymerich, R. Degraeve, and G.

Groeseneken, “A comprehensive study of channel hot-carrier degradation in short channel MOSFETs with high-k dielectrics,” *Microelectron. Eng.*, vol. 103, pp. 144–149, 2013.

[103] J. Martin-Martinez, N. Ayala, R. Rodriguez, M. Nafria, and X. Aymerich, “Bias Temperature Instability: Characterization, modeling and circuit aging evaluation,” in *IEEE 11th International Conference on Solid-State and Integrated Circuit Technology*, 2012, pp. 1–4.

[104] B. Kaczer, T. Grasser, P. J. Roussel, J. Franco, R. Degraeve, L. A. Ragnarsson, E. Simoen, G. Groeseneken, and H. Reisinger, “Origin of NBTI variability in deeply scaled pFETs,” in *IEEE International Reliability Physics Symposium Proceedings*, 2010, pp. 26–32.

[105] E. Amat, T. Kauerauf, R. Degraeve, R. Rodriguez, M. Nafria, X. Aymerich, and G. Groeseneken, “Channel Hot-Carrier degradation in short channel devices with high-k/metal gate stacks,” *Proc. 2009 Spanish Conf. Electron Devices, CDE’09*, vol. 9, no. 3, pp. 238–241, 2009.

[106] S. Machlup, “Noise in semiconductors: Spectrum of a two-parameter random signal,” *J. Appl. Phys.*, vol. 25, no. 3, pp. 341–343, 1954.

[107] G. Lucovsky, “Intrinsic limitations on the performance and reliability of high-k gate dielectrics for advanced silicon devices,” in *IEEE International Integrated Reliability Workshop*, 2005, pp. 145–157.

[108] G. Bersuker, J. H. Sim, C. D. Young, R. Choi, B. H. Lee, P. Lysaght, G. A. Brown, P. Zeitzoff, M. Gardner, R. W. Murto, and H. R. Huff, “Effects of Structural Properties of Hf-Based Gate Stack on Transistor Performance,” in *Spring Meeting of the Material Research Society*, 2004, vol. 811, pp. 31–35.

[109] J. L. Gavartin, A. L. Shluger, A. S. Foster, and G. I. Bersuker, “The role of nitrogen-related defects in high- k dielectric oxides: Density-functional studies,” *J. Appl. Phys.*, vol. 97, no. 5, p. 053704, 2005.

[110] H. Wen, S. S. Member, H. R. Harris, C. D. Young, S. S. Member, H. Luan, H. N. Alshareef, K. Choi, S. S. Member, D. Kwong, S. S. Member, P. Majhi, G. Bersuker, S. S. Member, B. H. Lee, and S. S. Member, “On Oxygen Deficiency and Fast Transient Charge-Trapping Effects in High- k Dielectrics,” *IEEE Electron Device Lett.*, vol. 27, no. 12, pp. 984–987, Dec. 2006.

[111] S. M. Amoroso, L. Gerrer, S. Markov, F. Adamu-Lema, and A. Asenov, “RTN and BTI in nanoscale MOSFETs: A comprehensive statistical simulation study,” *Solid. State. Electron.*, vol. 84, pp. 120–126, 2013.

[112] K. Takeuchi, “Impact of discrete-charge-induced variability on scaled MOS devices,” *IEICE Trans. Electron.*, vol. 95, no. 4, pp. 414–420, 2012.

[113] N. Tega, H. Miki, F. Pagette, D. J. Frank, A. Ray, M. J. Rooks, W. Haensch, and K. Torii, “Increasing threshold voltage variation due to random telegraph noise in FETs as gate lengths scale to 20 nm,” in *VLSI Technology, 2009 Symposium on*, 2009, pp. 50–51.

[114] N. Tega, H. Miki, M. Yamaoka, H. Kume, T. Mine, T. Ishida, Y. Mori, R. Yamada, and K. Torii,

"Impact of threshold voltage fluctuation due to random telegraph noise on scaled-down SRAM," in *IEEE International Reliability Physics Symposium Proceedings*, 2008, pp. 541–546.

[115] A. S. Spinelli, C. Monzio Compagnoni, R. Gusmeroli, M. Ghidotti, and A. Visconti, "Investigation of the random telegraph noise instability in scaled flash memory arrays," *Jpn. J. Appl. Phys.*, vol. 47, no. 4 PART 2, pp. 2598–2601, Apr. 2008.

[116] R. Da Silva, G. I. Wirth, and R. Brederlow, "Novel analytical and numerical approach to modeling low-frequency noise in semiconductor devices," *Phys. A Stat. Mech. its Appl.*, vol. 362, no. 2, pp. 277–288, 2006.

[117] V. Huard, M. Denais, and C. Parthasarathy, "NBTI degradation: From physical mechanisms to modelling," *Microelectron. Reliab.*, vol. 46, no. 1, pp. 1–23, 2006.

[118] D. K. Schroder and J. A. Babcock, "Negative bias temperature instability: Road to cross in deep submicron silicon semiconductor manufacturing," *J. Appl. Phys.*, vol. 94, no. 1, pp. 1–18, 2003.

[119] H. Reisinger, T. Grasser, W. Gustin, and C. Schluender, "The statistical analysis of individual defects constituting NBTI and its implications for modeling DC- and AC-stress," in *IEEE International Reliability Physics Symposium Proceedings*, 2010, pp. 7–15.

[120] Y. Miura and Y. Matukura, "Investigation of Silicon-Silicon Dioxide Interface Using MOS Structure," *Jpn. J. Appl. Phys.*, vol. 5, no. 2, pp. 180–180, Feb. 1966.

[121] S. Zafar, Y. H. Kim, V. Narayanan, C. Cabral, V. Paruchuri, B. Doris, J. Stathis, A. Callegari, and M. Chudzik, "A Comparative Study of NBTI and PBTI (Charge Trapping) in SiO2/HfO2 Stacks with FUSI, TiN, Re Gates," in *Symposium on VLSI Technology,. Digest of Technical Papers.*, 2006, pp. 23–25.

[122] J. Hicks, "Intel's 45nm CMOS Technology," *Intel Technol. J.*, vol. 12, no. 2, pp. 77–156, 2008.

[123] T. Grasser, B. Kaczer, P. Hehenberger, W. Gös, R. O'Connor, H. Reisinger, W. Gustin, and C. Schlünder, "Simultaneous extraction of recoverable and permanent components contributing to bias-temperature instability," in *Technical Digest - International Electron Devices Meeting, IEDM*, 2007, pp. 801–804.

[124] S. Rangan, N. Mielke, and E. C. C. Yeh, "Universal recovery behavior of negative bias temperature instability [PMOSFETs]," in *IEEE International Electron Devices Meeting 2003*, 2003, pp. 14.3.1–14.3.4.

[125] S. Mahapatra, K. Ahmed, D. Varghese, A. E. Islam, G. Gupta, L. Madhav, D. Saha, and M. A. Alam, "On the physical mechanism of NBTI in silicon oxynitride p-MOSFETs: Can differences in insulator processing conditions resolve the interface trap generation versus hole trapping controversy?," in *Annual Proceedings - Reliability Physics (Symposium)*, 2007, pp. 1–9.

[126] V. Huard, C. Parthasarathy, C. Guerin, T. Valentin, E. Pion, M. Mammasse, N. Planes, and

L. Camus, “NBTI degradation: From transistor to SRAM arrays,” in *IEEE International Reliability Physics Symposium Proceedings*, 2008, pp. 289–300.

[127] M. A. Alam, H. Kufluoglu, D. Varghese, and S. Mahapatra, “A comprehensive model for PMOS NBTI degradation: Recent progress,” *Microelectron. Reliab.*, vol. 47, no. 6, pp. 853–862, 2007.

[128] H. Reisinger, U. Brunner, W. Heinrigs, W. Gustin, and C. Schlünder, “A comparison of fast methods for measuring NBTI degradation,” *IEEE Trans. Device Mater. Reliab.*, vol. 7, no. 4, pp. 531–539, Dec. 2007.

[129] M. Moras, J. Martin-Martinez, R. Rodriguez, M. Nafria, X. Aymerich, and E. Simoen, “Negative Bias Temperature Instabilities induced in devices with millisecond anneal for ultra-shallow junctions,” *Solid. State. Electron.*, vol. 101, pp. 131–136, 2014.

[130] G. Chen, K. Y. Chuah, M. F. Li, D. S. H. Chan, C. H. Ang, J. Z. Zheng, Y. Jin, and D. L. Kwong, “Dynamic NBTI of PMOS transistors and its impact on device lifetime,” in *IEEE International Reliability Physics Symposium Proceedings*, 2003, pp. 196–202.

[131] M. Agostinelli, S. Lau, S. Pae, P. Marzolf, H. Muthali, and S. Jacobs, “PMOS NBTI-induced circuit mismatch in advanced technologies,” in *IEEE International Integrated Reliability Workshop Final Report*, 2004, pp. 171–175.

[132] C. T. Chan, H. C. Ma, C. J. Tang, and T. Wang, “Investigation of post-NBTI stress recovery in pMOSFETs by direct measurement of single oxide charge De-trapping,” in *Digest of Technical Papers - Symposium on VLSI Technology*, 2005, pp. 90–91.

[133] S. Mahapatra, D. Saha, D. Varghese, and P. B. Kumar, “On the generation and recovery of interface traps in MOSFETs subjected to NBTI, FN, and HCI stress,” *IEEE Trans. Electron Devices*, vol. 53, no. 7, pp. 1583–1592, Jul. 2006.

[134] H. Reisinger, O. Blank, W. Heinrigs, A. Mühlhoff, W. Gustin, and C. Schlünder, “Analysis of NBTI degradation- and recovery-behavior based on ultra fast VT -measurements,” in *IEEE International Reliability Physics Symposium Proceedings*, 2006, pp. 448–453.

[135] S. Chakravarthi, A. Krishnan, V. Reddy, C. F. Machala, and S. Krishnan, “A comprehensive framework for predictive modeling of negative bias temperature instability,” in *2004 IEEE International Reliability Physics Symposium. Proceedings*, 2004, pp. 273–282.

[136] T. Grasser, B. Kaczer, and W. Goes, “An energy-level perspective of bias temperature instability,” in *IEEE International Reliability Physics Symposium Proceedings*, 2008, pp. 28–38.

[137] T. Grasser, B. Kaczer, W. Goes, T. Aichinger, P. Hehenberger, and M. Nelhiebel, “A two-stage model for negative bias temperature instability,” in *International Reliability Physics Symposium (IRPS),* 2009, pp. 33–44.

[138] T. Grasser, H. Reisinger, W. Goes, T. Aichinger, P. Hehenberger, P.-J. Wagner, M. Nelhiebel, J. Franco, and B. Kaczer, “Switching oxide traps as the missing link between negative bias

temperature instability and random telegraph noise," in *Electron Devices Meeting (IEDM), 2009 IEEE International*, 2009, pp. 729–732.

[139] A. Bravaix and V. Huard, "Hot-Carrier Degradation Issues in Advanced CMOS Nodes," in *Proc. European Symposium on Reliability of Electron Devices Failure Physics and Analysis (ESREF), tutorial*, 2010, pp. 1267–1272.

[140] C. Hu, S. C. Tam, F. C. Hsu, P. K. Ko, T. Y. Chan, and K. W. Terrill, "Hot-Electron-Induced MOSFET Degradation-Model, Monitor, and Improvement," *IEEE Trans. Electron Devices*, vol. 32, no. 2, pp. 375–385, Feb. 1985.

[141] G. Groeseneken, R. Bellens, G. Van Den Bosch, and H. E. Maes, "Hot-carrier degradation in submicrometre MOSFETs: from uniform injection towards the real operating conditions," *Semicond. Sci. Technol.*, vol. 10, no. 9, pp. 1208–1220, Sep. 1995.

[142] A. Bravaix, Y. Mamy Randriamihaja, V. Huard, D. Angot, X. Federspiel, W. Arfaoui, P. Mora, F. Cacho, M. Saliva, C. Besset, S. Renard, D. Roy, and E. Vincent, "Impact of the gate-stack change from 40nm node SiON to 28nm High-K Metal Gate on the Hot-Carrier and Bias Temperature damage," in *IEEE International Reliability Physics Symposium Proceedings*, 2013, p. 2D. 6.1–2D. 6.9.

[143] Q. Wu, A. Bayerl, M. Porti, J. Martin-Martinez, M. Lanza, R. Rodriguez, V. Velayudhan, M. Nafria, X. Aymerich, M. B. Gonzalez, and E. Simoen, "A conductive AFM nanoscale analysis of NBTI and channel hot-carrier degradation in MOSFETs," *IEEE Trans. Electron Devices*, vol. 61, no. 9, pp. 3118–3124, Sep. 2014.

[144] C. Prasad, L. Jiang, D. Singh, M. Agostinelli, C. Auth, P. Bai, T. Eiles, J. Hicks, C. H. Jan, K. Mistry, S. Natarajan, B. Niu, P. Packan, D. Pantuso, I. Post, S. Ramey, A. Schmitz, B. Sell, S. Suthram, J. Thomas, C. Tsai, and P. Vandervoorn, "Self-heat reliability considerations on Intel's 22nm Tri-Gate technology," in *IEEE International Reliability Physics Symposium Proceedings*, 2013, pp. 5D.1.1–5D.1.5.

[145] C. Prasad, "Advanced CMOS reliability challenges," in *Proceedings of Technical Program - International Symposium on VLSI Technology, Systems and Application, VLSI-TSA*, 2014, pp. 1–2.

[146] A. Asenov, R. Balasubramaniam, A. R. Brown, and J. H. Davies, "RTS amplitudes in decananometer MOSFETs: 3-D simulation study," *IEEE Trans. Electron Devices*, vol. 50, no. 3, pp. 839–845, 2003.

[147] J. Bardeen, "Surface states and rectification at a metal semi-conductor contact," *Phys. Rev.*, vol. 71, no. 10, pp. 717–727, May 1947.

[148] C.-T. Sah, *Book: Fundamentals of Solid State Electronics*, vol. 1. World Scientific Publishing Company, 1991.

[149] B. Benbakhti, J. S. Ayubi-Moak, K. Kalna, D. Lin, G. Hellings, G. Brammertz, K. De Meyer, I. Thayne, and A. Asenov, "Impact of interface state trap density on the performance characteristics of different III-V MOSFET architectures," *Microelectron. Reliab.*, vol. 50, no.

3, pp. 360–364, 2010.

[150] T. W. T. Wang, T.-E. C. T.-E. Chang, and C. H. C. Huang, “Interface trap induced thermionic and field emission current in\noff-state MOSFET’s,” in *Proceedings of 1994 IEEE International Electron Devices Meeting*, 1994, pp. 161–164.

[151] Y. Li, H. W. Su, Y. Y. Chen, S. C. Hsu, and W. T. Huang, “The intrinsic parameter fluctuation on high-k/metal gate bulk FinFET devices,” *Microelectron. Eng.*, vol. 109, pp. 302–305, Sep. 2013.

[152] L. Brunet, X. Garros, M. Cassé, O. Weber, F. Andrieu, C. Fenouillet-Béranger, P. Perreau, F. Martin, M. Charbonnier, D. Lafond, C. Gaumer, S. Lhostis, V. Vidal, L. Brévard, L. Tosti, S. Denorme, S. Barnola, J. F. Damlencourt, V. Loup, G. Reimbold, F. Boulanger, O. Faynot, and A. Bravaix, “New insight on VT stability of HK/MG stacks with scaling in 30nm FDSOI technology,” in *Digest of Technical Papers - Symposium on VLSI Technology*, 2010, pp. 29–30.

[153] D. Esseni, M. Pala, E. Gnani, and E. Sangiorgi, “Influence of interface traps on the performance of Tunnel FETs,” in *2015 Fourth Berkeley Symposium on Energy Efficient Electronic Systems (E3S)*, 2015, pp. 1–1.

[154] T. Tsuchiya, “Characterization of Individual Si/SiO2 Interface Traps: Direct Observation of Single Pb0 Centers by the Charge Pumping (CP) Method and Correction of the Conventional CP Theory.,” *ECS Trans.*, vol. 69, no. 10, pp. 145–154, Oct. 2015.

[155] S. Okhonin, T. Hessler, and M. Dutoit, “Comparison of gate-induced drain leakage and charge pumping measurements for determining lateral interface trap profiles in electrically stressed MOSFET’s,” *IEEE Trans. Electron Devices*, vol. 43, no. 4, pp. 605–612, Apr. 1996.

[156] S. Mahapatra, C. D. Parikh, J. Vasi, V. Ramgopal Rao, and C. R. Viswanathan, “Direct charge pumping technique for spatial profiling of hot-carrier induced interface and oxide traps in MOSFETs,” *Solid. State. Electron.*, vol. 43, no. 5, pp. 915–922, 1999.

[157] Y. L. Chu, D. W. Lin, and C. Y. Wu, “A new charge-pumping technique for profiling the interface-states and oxide-trapped charges in MOSFET’s,” *IEEE Trans. Electron Devices*, vol. 47, no. 2, pp. 348–353, 2000.

[158] G. Vincent, A. Chantre, and D. Bois, “Electric field effect on the thermal emission of traps in semiconductor junctions,” *J. Appl. Phys.*, vol. 50, no. 8, pp. 5484–5487, 1979.

[159] I. S. Esqueda and H. J. Barnaby, “Modeling the non-uniform distribution of radiation-induced interface traps,” in *IEEE Transactions on Nuclear Science*, 2012, vol. 59, no. 4 PART 1, pp. 723–727.

[160] “International Technology Roadmap for Semiconductors,” in *Semiconductor Industry Association*, 2005.

[161] H. K. Gummel, “A Self-Consistent Iterative Scheme for One-Dimensional Steady State

Transistor Calculations," *IEEE Trans. Electron Devices*, vol. 11, no. 10, pp. 455–465, Oct. 1964.

[162] A. De Mari, "An accurate numerical steady-state one-dimensional solution of the P-N junction," *Solid. State. Electron.*, vol. 11, no. 1, pp. 33–58, 1968.

[163] D. L. Scharfetter and H. K. Gummel, "Large-Signal Analysis of a Silicon Read Diode Oscillator," *IEEE Trans. Electron Devices*, vol. 16, no. 1, pp. 64–77, Jan. 1969.

[164] H. W. Loeb, R. Andrew, and W. Love, "Application of 2-Dimensional Solutions of the Shockley-Poisson Equation to Inversion-Layer M.O.S.T. Devices," *Electron. Lett.*, vol. 4, no. 17, pp. 352–354, 1968.

[165] J. E. Schroeder and R. S. Muller, "IGFET analysis through numerical solution of Poisson's equation," *IEEE Trans. Electron Devices*, vol. 15, no. 12, pp. 954–961, Dec. 1968.

[166] J. J. Barnes and R. J. Lomax, "Finite-element methods in semiconductor device simulation," *IEEE Trans. Electron Devices*, vol. 24, no. 8, pp. 1082–1089, Aug. 1977.

[167] E. Buturla, P. Cottrell, B. Grossman, K. Salsburg, M. Lawlor, and C. McMullen, "Three-dimensional finite element simulation of semiconductor devices," in *IEEE International Solid-State Circuits Conference. Digest of Technical Papers.*, 1980, vol. 23, pp. 76–77.

[168] P. Ciampolini, A. Pierantoni, A. Forghieri, and G. Baccarani, "Simulation of three-dimensional effects in VLSI devices," *Microelectronics J.*, vol. 21, no. 6, pp. 5–12, Jan. 1990.

[169] G. Heiser, C. Pommerell, J. Weis, and W. Fichtner, "Three-dimensional numerical semiconductor device simulation: algorithms, architectures, results," *IEEE Trans. Comput. Des. Integr. Circuits Syst.*, vol. 10, no. 10, pp. 1218–1230, 1991.

[170] C. P. Ho, J. D. Plummer, S. E. Hansen, and R. W. Dutton, "VLSI Process Modeling: SUPREM III," *IEEE Trans. Electron Devices*, vol. 30, no. 11, pp. 1438–1453, Nov. 1983.

[171] Synopsys, "Synopsys TCAD," 2016. .

[172] C. Silvaco International, "Silvaco," 2016. [Online]. Available: www.silvaco.com.

[173] D. K. Schroder, *Book: Semiconductor Material and Device Characterization*. IEEE Press, 2006.

[174] H. Timm, A. Schenk, and A. Wettstein, "On density-gradient modeling of tunneling through insulators," *IEICE Trans.*, 2003.

[175] C. Soo-Won and K. J. Bathe, "On automatic mesh construction and mesh refinement in finite element analysis," *Comput. Struct.*, vol. 32, no. 3–4, pp. 911–936, 1989.

[176] A. Kawamoto, J. Jameson, K. Cho, R. W. Dutton, E. Gnani, S. Reggiani, R. Colle, M. Rudan, A. Nakano, and M. E. Bachlechner, *Book: Special Issue on Computational Electronics: New Challenges and Directions*, vol. 47, no. 10. IEEE, 2000.

[177] A. Islam, A. Imran, and M. Hasan, "Variability analysis and FinFET-based design of XOR and

XNOR circuit," in *International Conference on Computer and Communication Technology, ICCCT-2011*, 2011, pp. 239–245.

[178] S. Mittal, S. Kurude, S. Dutta, P. Debashis, S. Ganguly, S. Lodha, A. Laha, and U. Ganguly, "Epitaxial rare earth oxide (EOx) FinFET: A variability-resistant Ge FinFET architecture with multi VT," in *Device Research Conference - Conference Digest, DRC*, 2014, pp. 97–98.

[179] A. Martinez, N. Seoane, A. R. Brown, J. R. Barker, and A. Asenov, "Variability in Si nanowire MOSFETs due to the combined effect of interface roughness and random dopants: A fully three-dimensional NEGF simulation study," *IEEE Trans. Electron Devices*, vol. 57, no. 7, pp. 1626–1635, 2010.

[180] X. Wang, A. R. Brown, N. Idris, S. Markov, G. Roy, and A. Asenov, "Statistical threshold-voltage variability in scaled decananometer bulk HKMG MOSFETs: A full-scale 3-D simulation scaling study," *IEEE Trans. Electron Devices*, vol. 58, no. 8, pp. 2293–2301, 2011.

[181] C. Alexander, G. Roy, and A. Asenov, "Random-Dopant-Induced Drain Current Variation in Nano-MOSFETs: A Three-Dimensional Self-Consistent Monte Carlo Simulation Study Using 'Ab Initio' Ionized Impurity Scattering," *Electron Devices, IEEE Trans.*, vol. 55, no. 11, pp. 3251–3258, 2008.

[182] S. Markov, S. Roy, and A. Asenov, "Direct tunnelling gate leakage variability in Nano-CMOS transistors," *IEEE Trans. Electron Devices*, vol. 57, no. 11, pp. 3106–3114, 2010.

[183] M. F. Bukhori, S. Roy, and A. Asenov, "Simulation of statistical aspects of charge trapping and related degradation in bulk mosfets in the presence of random discrete dopants," *IEEE Trans. Electron Devices*, vol. 57, no. 4, pp. 795–803, 2010.

[184] A. Asenov, A. R. Brown, J. H. Davies, S. Kaya, and G. Slavcheva, "Simulation of intrinsic parameter fluctuations in decananometer and nanometer-scale MOSFETs," *IEEE Trans. Electron Devices*, vol. 50, no. 9, pp. 1837–1852, Sep. 2003.

[185] A. Asenov, B. Cheng, X. Wang, A. R. Brown, D. Reid, C. Millar, and C. Alexander, "Simulation based transistor-SRAM co-design in the presence of statistical variability and reliability," in *Technical Digest - International Electron Devices Meeting, IEDM*, 2013, pp. 33.1.1–33.1.4.

[186] V. Bonfiglio and G. Iannaccone, "Evaluation of threshold voltage dispersion in 45 nm CMOS technology with TCAD-based sensitivity analysis," in *14th International Workshop on Computational Electronics, IWCE*, 2010, pp. 101–104.

[187] H. G. Mohammadi, P. E. Gaillardon, M. Yazdani, and G. De Micheli, "A fast TCAD-based methodology for Variation analysis of emerging nano-devices," in *Proceedings - IEEE International Symposium on Defect and Fault Tolerance in VLSI Systems*, 2013, pp. 83–88.

[188] A. R. Brown, N. Daval, K. K. Bourdelle, B. Y. Nguyen, and A. Asenov, "Comparative simulation analysis of process-induced variability in nanoscale soi and bulk trigate finfets," *IEEE Trans. Electron Devices*, vol. 60, no. 11, pp. 3611–3617, 2013.

[189] L. Wang, A. R. Brown, C. Millar, A. Burenkov, X. Wang, A. Asenov, and J. Lorenz, “Simulation for statistical variability in realistic 20nm MOSFET,” in *15th International Conference on Ultimate Integration on Silicon, (ULIS)*, 2014, pp. 5–8.

[190] “Gold Standard Simulations,” 2015. [Online]. Available: http://www.goldstandardsimulations.com.

[191] L. Gerrer, J. Ding, S. M. Amoroso, F. Adamu-Lema, R. Hussin, D. Reid, C. Millar, and A. Asenov, “Modelling RTN and BTI in nanoscale MOSFETs from device to circuit: A review,” *Microelectron. Reliab.*, vol. 54, no. 4, pp. 682–697, 2014.

[192] M. Toledano-Luque, B. Kaczer, J. Franco, P. J. Roussel, T. Grasser, T. Y. Hoffmann, and G. Groeseneken, “From mean values to distributions of BTI lifetime of deeply scaled FETs through atomistic understanding of the degradation,” in *2011 Symposium on VLSI Technology - Digest of Technical Papers*, 2011, pp. 152–153.

[193] A. Asenov, R. Balasubramaniam, A. R. Brown, and J. H. Davies, “Effect of single-electron interface trapping in decanano MOSFETs: A 3D atomistic simulation study,” *Superlattices Microstruct.*, vol. 27, no. 5–6, pp. 411–416, May 2000.

[194] S. Aleksic, B. Pesic, and D. Pantic, “Simulation of semiconductor bulk trap influence on the electrical characteristics of the n-channel power VDMOS transistor,” *Electron. Components Mater.*, vol. 43, no. 2, pp. 124–130, 2013.

[195] Z. Chen, X. Zhou, G. Zhu, and S. Lin, “Interface-trap modeling for silicon-nanowire MOSFETs,” in *IEEE International Reliability Physics Symposium Proceedings*, 2010, pp. 977–980.

[196] B. A. Prabowo, S. K. Amethystna, J. R. Tsai, S. M. Yang, and G. Sheu, “Interface trap distribution for HCI reliability assessment on bend gate structure by 3D TCAD simulation,” in *Proceedings of the International Symposium on the Physical and Failure Analysis of Integrated Circuits, IPFA*, 2012, pp. 1–4.

[197] T. Grasser, H. Reisinger, P. Wagner, and B. Kaczer, “The Time Dependent Defect Spectroscopy (TDDS) Technique for the Bias Temperature Instability,” in *International Reliability Physics Symposium (IRPS),* 2010, vol. 1835, pp. 9–10.

[198] T. Wang, C. T. Chan, C. J. Tang, C. W. Tsai, H. C. H. Wang, M. H. Chi, and D. D. Tang, “A Novel transient characterization technique to investigate trap properties in HfSiON gate dielectric MOSFETs - From single electron emission to PBTI recovery transient,” *IEEE Trans. Electron Devices*, vol. 53, no. 5, pp. 1073–1079, 2006.

[199] A. H. Edwards, “Interaction of H and H 2 with the silicon dangling orbital at the 〈111〉 Si/SiO2 interface,” *Phys. Rev. B*, vol. 44, no. 4, pp. 1832–1838, Jul. 1991.

[200] M. Lax, “Cascade capture of electrons in solids,” *Phys. Rev.*, vol. 119, no. 5, pp. 1502–1523, Sep. 1960.

[201] B. Keramati and J. N. Zemel, *Book: The Physics of SiO2 and its Interfaces*. Pergamon Press,

1978.

[202] “M. I. T. Microsystems Technology Laboratory.” [Online]. Available: http://www-mtl.mit.edu/researchgroups/Well/.

[203] A. Schenk, “Finite-temperature full random-phase approximation model of band gap narrowing for silicon device simulation,” *J. Appl. Phys.*, vol. 84, no. 7, pp. 3684–3695, 1998.

[204] Y. Cheng, M. Jeng, Z. Liu, and J. Huang, “A Physical and Scalable I-V Model in BSIM3v3 for Analog / Digital Circuit Simulation,” *IEEE Trans. Electron Devices*, vol. 44, no. 2, pp. 277–287, 1997.

[205] X. Wang, F. Adamu-Lema, B. Cheng, and A. Asenov, “Geometry, temperature, and body bias dependence of statistical variability in 20-nm bulk CMOS technology: A comprehensive simulation analysis,” *IEEE Trans. Electron Devices*, vol. 60, no. 5, pp. 1547–1554, May 2013.

[206] J. Franco, B. Kaczer, P. J. Roussel, M. Toledano-Luque, P. Weckx, and T. Grasser, “Relevance of non-exponential single-defect-induced threshold voltage shifts for NBTI variability,” in *IEEE International Integrated Reliability Workshop Final Report*, 2013, pp. 69–72.

[207] G. V. Groeseneken, I. De Wolf, R. Bellens, and H. E. Maes, “Observation of single interface traps in submicron MOSFET’s by charge pumping,” *IEEE Trans. Electron Devices*, vol. 43, no. 6, pp. 940–945, 1996.

[208] M. L. Fan, V. P. H. Hu, Y. N. Chen, P. Su, and C. Te Chuang, “Impacts of random telegraph noise on FinFET devices, 6T SRAM cell, and logic circuits,” in *IEEE International Reliability Physics Symposium Proceedings*, 2012, vol. 59, no. 8, pp. 2227–2234.