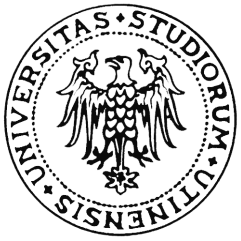

UNIVERSITÁ DEGLI STUDI DI UDINE

Dipartimento di Ingegneria Elettrica, Gestionale e Meccanica
Corso di Dottorato in Ingegneria Industriale e dell'Informazione

Institut de Microelectronique, Electromagnetisme et Photonique
Ecole Doctorale Electronique, Electrotechnique, Automatique,
Telecomunications, Signal (EEATS)

Tesi di Dottorato di Ricerca

Impact of the technology boosters on the MOSFET performance

Relatori:

Prof. Luca Selmi
Gerard Ghibaudo

Dottorando:

Paolo Toniutti

Commission esaminatrice:

Prof. Roberto Rinaldo
Prof. Wim Magnus
Prof. Andreas Schenk
Dr. Gerard Ghibaudo
Prof. Luca Selmi
Prof. Pierpaolo Palestri
Prof. Raphael Clerc

Anno Accademico 2011-12

Abstract

The understanding of the charge transport in nano-scale CMOS device is a very challenging issue that requires a physics-based modeling approach. In this respect, commercial T-CAD simulators have been shown evident limitations in the modeling of ultra-scaled devices, due to the strong off-equilibrium non-local transport that takes place in this kind of devices. On the other hand, full quantum approaches would in principle allow to correctly treat the problem, but their complexity and the difficult inclusion of many important scattering mechanisms limit their use to particular cases, decreasing their effectiveness in being a general tool for the comprehension of all the various aspects of the modeling problem. In this thesis we use a Multi Subband Monte Carlo simulation framework to assess the effects of some of the mostly used techniques to overcome the performances of the conventional ultra-scaled MOSFETs. This approach allows for a better understanding of the charge transport in short channel devices, using a software tool that remains relatively easy to use by the engineering community. The following technology boosters will be addressed: multi-gate structures, high-k dielectrics, strained silicon and germanium channels.

The modeling of screening in non-conventional structures such as double gate SOI and FinFETs is not trivial. In this respect, we analyze different formulations of the screening dielectric function approach in order to establish the correct methodology to properly model this physical mechanism in multi-gate MOSFETs.

Concerning the modeling of the high-k dielectric effects, we assess the effectiveness of various scattering mechanisms in degrading the electron and hole mobility. We also propose original models to explain the mobility degradation found experimentally. Finally, we analyze the influence of the high-k dielectrics on the ON-state current in short channel MOSFETs.

Concerning the alternative high-mobility channel materials, we perform Monte Carlo simulations of germanium and strained germanium channels to compare their performances with respect to silicon and strained silicon materials.

Finally, Monte Carlo simulations are used to analyze, validate and improve an existing technique used to experimentally extract the limiting velocity in short channel MOSFETs, that is useful to assess the real improvement provided by the alternative channel materials.

Contents

Chapter 1

Introduction

Abstract

In this chapter we introduce various examples of the scaling rules that, in the past decades, have lead the modern devices to dimensions of few nano-meters.

Then, we list some problems that raise in nowadays ultra short MOSFETs, and how they have been solved by means of the introduction of the technology boosters in the fabrication process.

Physical parameter	Constant Field	Generalized	Gen. Selective
Channel length	$1/\alpha$	$1/\alpha$	$1/\alpha_d$
Insulator thickness	$1/\alpha$	$1/\alpha$	$1/\alpha_d$
Channel width	$1/\alpha$	$1/\alpha$	$1/\alpha_w$
Wiring width	$1/\alpha$	$1/\alpha$	$1/\alpha_w$
Electric field in device	1	ϵ	ϵ
Voltage	$1/\alpha$	ϵ/α	ϵ/α_d
Doping	α	$\epsilon\alpha$	$\epsilon\alpha_d$
Area	$1/\alpha^2$	$1/\alpha^2$	$1/\alpha_w^2$
Capacitance	$1/\alpha$	$1/\alpha$	$1/\alpha_w$
Gate delay	$1/\alpha$	$1/\alpha$	$1/\alpha_d$
Power dissipation	$1/\alpha^2$	ϵ^2/α^2	$\epsilon^2/\alpha_d^2\alpha_w^2$
Power density	1	ϵ^2	$\epsilon^2\alpha_w/\alpha_d$

Table 1.1: Technology scaling rules: constant-electric field scaling, generalized scaling and generalized selective scaling.

1.1 Scaling of the MOSFET

1.1.1 The scaling scenario

In the past 40 years, most of the electronic fabrication activity has been focused to shrinking the dimensions of the MOSFETs (Metal Oxide Semiconductor Field Effect Transistors). The aim of this scaling is mainly to increase the number of transistors per chip which leads to lower cost per function and larger computing power.

The initial scaling rules proposed in [2] were constructed in order to obtain scaled devices with the same operative conditions as the original ones (i.e. the electric field in the device channel remains constant through the scaling process). This maintains a good reliability of the scaled devices. These rules change various device physical parameters by the same factor $\alpha > 1$, as shown in Table 1.1 (second column). In this methodology of scaling, called "constant field scaling", the threshold voltage shift should scale as the other parameters. However, in extremely scaled devices, it is difficult to scale the threshold voltage, since, otherwise, the leakage currents would become excessive.

This fact has lead to a more complex formulation of scaling rules, called "generalized scaling" [3, 4, 5, 6]. In these new methodologies, the supply voltage scales by a factor ϵ, with $1<\epsilon<\alpha$. The problem of this scaling approach is that the electric field in the devices increases by a factor ϵ, as we can see in the third column of Table 1.1. Thus, new reliability problems can arise.

Recently, a new scaling methodology, called "generalized selective scaling" has been proposed [7, 8, 9]. It consists of scaling the channel length and oxide thickness by a factor α_d, and the channel width and the wiring width by a factor

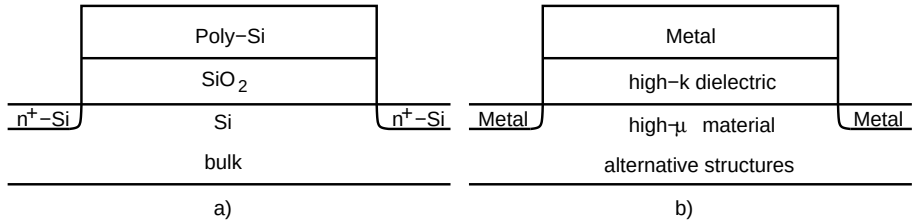

Figure 1.1: Comparison between a conventional MOSFET (a) and a device which features various technology boosters (b).

α_w. Choosing $\alpha_d > \alpha_w$, we maintain a good reliability, allowing the gate delay to scale faster than in the previous scaling technologies.

1.1.2 Problems related to the fabrication of ultra short MOSFETs

The scaling rules explained in Sec. 1.1.1 have been applied for entire decades, leading the modern MOSFET devices to dimensions of about 10 nano-meters. Due to the size of nowadays devices, some problems have become much more critical than in the past:

- static power dissipation : the leakage currents (of the gate contact and of the junctions) has augmented considerably;
- short channel effects : the electric potential in the channel of the device is not independent on the drain contact polarization, since the electric field of the channel-drain junction penetrates into the channel;
- limitation of the ON-current : the ON-current does not longer improve as the channel length is shrinked;
- dynamic power dissipation : the areal density of the operating power has increased sensibly.

In order to mitigate these negatives effects, some techniques called ”technology boosters“ have been introduced in the fabrication process of the modern MOS devices [10].

1.1.3 Overview of the technology boosters

The terminology ”technology boosters“ stands for all the techniques introduced in the fabrication process of the ultra scaled MOSFETs in order to continue the improvement of the performance of the devices beyond the classical scaling rules explained in Sec. 1.1.1. These techniques have modified all the regions of the MOSFET structure compared to the ”conventional“ device (see Fig. 1.1):

- channel material : the silicon (Si) has been substituted with the strained silicon (s-Si). Moreover, Germanium (Ge) and III-V materials (as for example the gallium arsenide, GaAs) are emerging as promising candidates to substitute the Si based materials in the device channel.

- gate oxide : the SiO_2 has been substituted with materials with a high electric permittivity k, in order to limit the gate leakage current and improve the reliability of the dielectric.
- gate contact : the poly-Si material has been substituted with a metal, for many reasons. The main reason is because of the better integrability with the high-k dielectrics. Moreover a metal gate reduces sensibly the poly-depletion region, that is one of the main problems affecting the poly-Si gates, allowing a better control of the channel region.
- source and drain contacts : the doped semiconducting regions will be substituted with metal regions, in order to limit the series resistances of the contacts.
- device architecture : much effort is devoted to substitute the conventional planar structure with Silicon On Insulator (SOI) or multi-gate structures. Mainly, this step will lead to improved capability of the gate contact to control the carriers in the channel, thus allowing to further scale the devices, while avoiding the short channel effects.

1.1.4 The state of the art

The guideline of the technology improvements in MOSFET fabrication is the International Technology Roadmap for Semiconductors (ITRS) [10]. According to the ITRS, devices which enter the mass production in 2012, would be a bulk transistor with a gate length of 22 nm for high-perfomance logic (the channel length is even shorter). The thickness of the SiO_2 dielectric should be around 1.06 nm. The mobility enhancement of the carriers in the inversion layer has to be around a factor 2 for both electrons and holes, with respect to a conventional device with unstrained Si channel. The supply voltage should be 0.9 V, while the threshold voltage is around 0.29 V.

Taking a look on most recent research activities, many publications have been focused on device structures alternative to the bulk ones. For instance, a 20 nm Single Gate Fully Depleted SOI structure has been presented in [11]. Much effort is also given to bulk FinFETS (triple gate structures) [12, 13], SOI FinFETs [14] and 3D vertical integration FinFETs [15]. Although almost all new devices feature the high-k dielectrics, much attention is still paid to the optimization of these type of dielectrics [16, 17, 18]. Concerning the introduction of the alternative channel materials, we can see that s-Si is nowadays employed in almost all Si-technologies. However, interesting publications can be found regarding the possible replacement of Si with Ge [19, 20] and III-V [21, 22, 23] materials.

1.2 High-k dielectrics

Observing Table 1.1, we can see that in all the mentioned scaling rules the gate oxide thickness must scale by the same factor as the channel length (see Fig. 1.2),

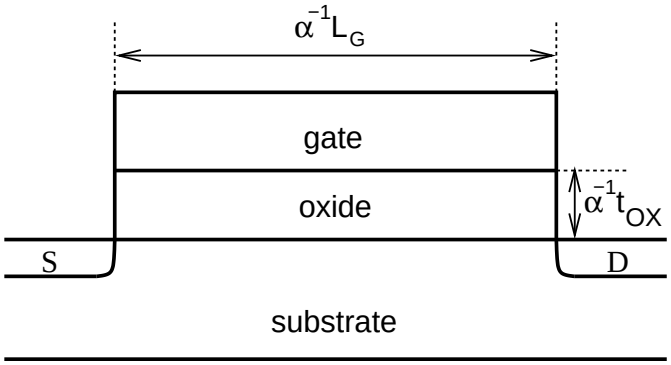

Figure 1.2: Sketch of the MOSFET device. The relation between the scaling of the channel length L_{G} and the dielectric thickness t_{OX} is shown.

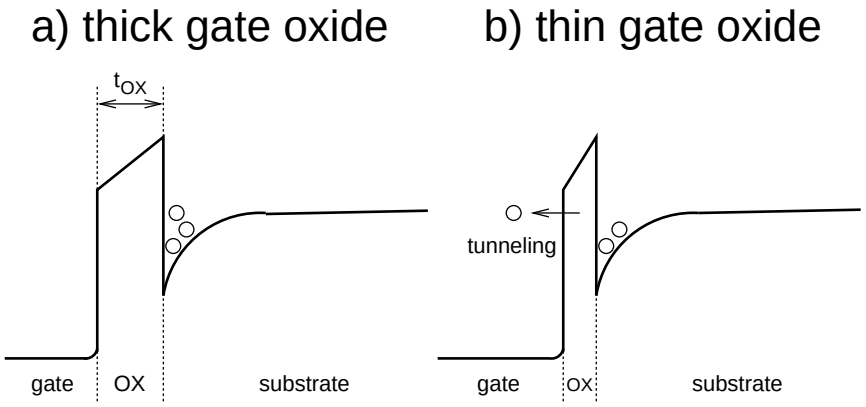

Figure 1.3: Band diagram of structures with a thick (a) and a thin (b) gate oxide. In the thin oxide structure, barrier tunneling is not negligible.

in order to avoid the so called short channel effects and thus allow the gate electrode to have an optimal control of the electrostatics in the channel. However, the oxide thickness shrinking leads to an increased leakage current.

1.2.1 The gate leakage current problem

Let's consider the potential barrier between channel and gate. If this barrier is thin in space or low in energy, the probability of barrier tunneling is high (see Fig. 1.3). For this reason, the probability that an electron passes through the dielectric depends on the insulator thickness (thickness of the barrier) and on the insulator material (barrier height, effective mass, ...).

The scaling rules have lead to silicon dioxide (SiO_2) thicknesses that result in significant tunneling of the carriers from the MOSFET channel to the gate electrode leading to an undesired gate leakage current. This is un undesired effect for mainly two reasons:

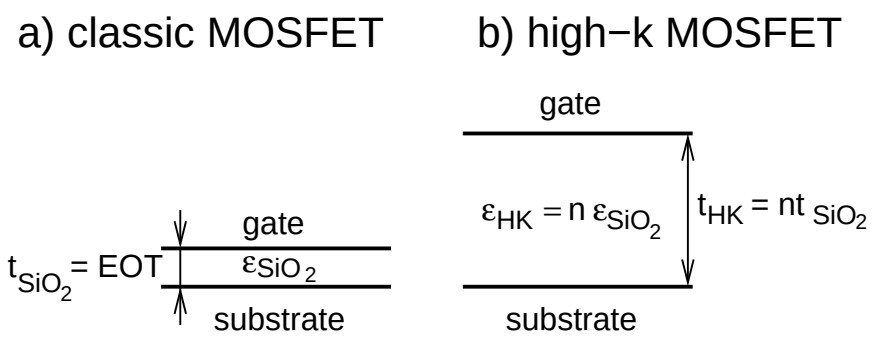

Figure 1.4: Comparison between structures with a SiO_2 (left) and a high-k (right) gate oxide. In the high-k structure, a thicker oxide layer can be used, reducing the gate leakage. n is the ratio between the high-k and SiO_2 electric permittivities.

- the gate current increases the static power dissipation.
- the gate current flows through a dielectric layer, decreasing the reliability of the device.

New solutions have been therefore devised to continue the nano-electronic development.

1.2.2 Benefit of using high-k dielectrics

One possible solution to reduce the gate leakage current is to change the insulator material. The most promising candidates to substitute the SiO_2 are the so called high-k dielectrics.

The introduction of high-k materials allows to increase the dielectric thickness (which should decrease the leakage current) without decreasing the equivalent oxide thickness (EOT). The EOT is the SiO_2 thickness that produces the same gate capacitance C_{OX} as the one of the high-k stack. The gate capacitance, per unit area, is:

$$C_{\mathrm{OX}} = \frac{\epsilon_{\mathrm{OX}}}{t_{\mathrm{OX}}} \tag{1.1}$$

where ϵ_{OX} is the electric permittivity of the gate dielectric and t_{OX} is the thickness.

Fig. 1.4 shows that if we use a material with a dielectric constant n times larger than the one of SiO_2, we can have the same C_{OX} with a thickness n times larger. This allows to decrease the gate leakage current, preserving the scalability of the device apart from two-dimensional effects [24].

However, while the use of the high-k dielectrics allows to increase the physical thickness of the gate oxide, they have a smaller height of the potential barrier. Thus, the carriers in the inversion layer of a MOSFET featuring high-k dielectrics feel a thicker but lower potential barrier than in classical devices. Anyway, despite of the lower potential barrier, high-k dielectric MOSFETs outperform SiO_2 devices [25, 26].

1.2.3 Mobility reduction in MOSFETs with high-k in the gate stack

The carrier mobility μ is relates the drift velocity of the carriers with the driving electric field, as:

$$\mu = \frac{v_{\mathrm{drift}}}{E_{\mathrm{lateral}}} \tag{1.2}$$

when the lateral electric field E_{lateral} is not too high (we will explain this aspect in Sec. 2.2.1). In the equation, v_{drift} is the drift velocity of the carriers in the inversion layer. Thus, the mobility, whose general definition can be found in [27], expresses the ability of the carriers to gain velocity when they are subjected to a driving electric field.

The use of high-k dielectrics leads to a lower value of the carrier mobility than the one obtained with SiO_2. For this reason two solutions that sensibly improve the mobility have been commonly introduced in the MOSFET fabrication:

- the usage of metal gate electrodes, that improves the quality of the interface between the high-k dielectric and the gate [28];
- a thin SiO_2 layer between the channel and the high-k material (the so called interfacial layer) has been introduced in order to place the high-k dielectric more distant from the free carriers of the inversion layer.

Nevertheless, these solutions do not completely solve the problem of the mobility reduction. Moreover, the interfacial layer introduction poses a trade off between the mobility reduction and the high-k effectiveness allowing for small EOT. However, since these two solutions are almost universally used, all data shown in this section refers to MOSFETs with metal gate and interfacial layer.

In Figs. 1.5 and 1.6 we show some experimental data that supports the fact that there is a mobility degradation due to the use of high-k insulators. In detail, Fig. 1.5 refers to devices with hafnium dioxide (HfO_2) and Fig. 1.6 refers to hafnium-silicon oxynitride (HfSiON). These are two of the most used high-k dielectrics. All the curves in the two graphs refer to measurements at 300 K. The HfO_2 data is published in [29, 30, 31, 32, 33, 34, 35, 28, 36, 37]. The HfSiON data is published in [38, 39, 40, 41, 42, 43, 44, 45, 46]

Considering the curves of Fig. 1.5, which refer to measurements at 300 K of HfO_2 devices with metal gate, Ref. [31] is quite useful since it contains a wide set of experimental data that can help us in the analysis of the mobility degradation, as for example the dependencies of the mobility on the interfacial or high-k layer thicknesses.

We collect some of these curves in Fig. 1.7. These curves refer to bulk devices with channel doping of 2×10^{17} cm^{-3}, with HfO_2 thickness of 3 nm and with various interfacial layer thicknesses ranging from 1.0 to 2.5 nm. In the figure we also show the mobility curve of the reference SiO_2 device with 2.5 nm of SiO_2 and metal gate. All the devices measured in [31] have titanium nitride (TiN) gate electrode. Moreover, Takagi reference curve is shown (which refers to a device with poly-Si gate electrode, but very thick SiO_2 dielectric). We can see that while the device with interfacial layer thickness of 2.5 nm shows no mobility degradation

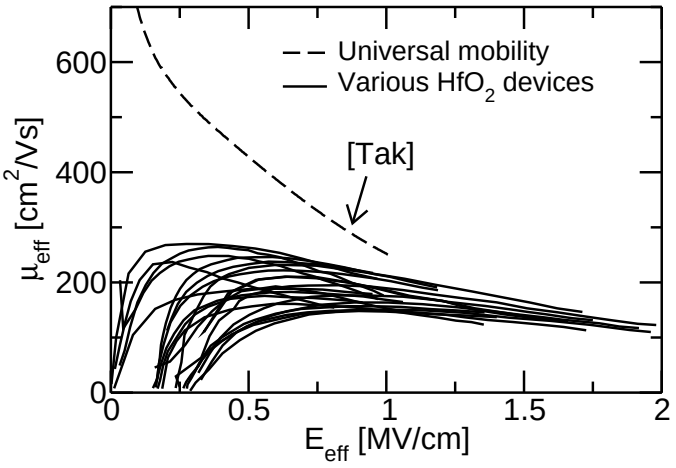

Figure 1.5: Comparison between the universal mobility curve [47] and various experimental data for HfO_2 devices with metal gate.

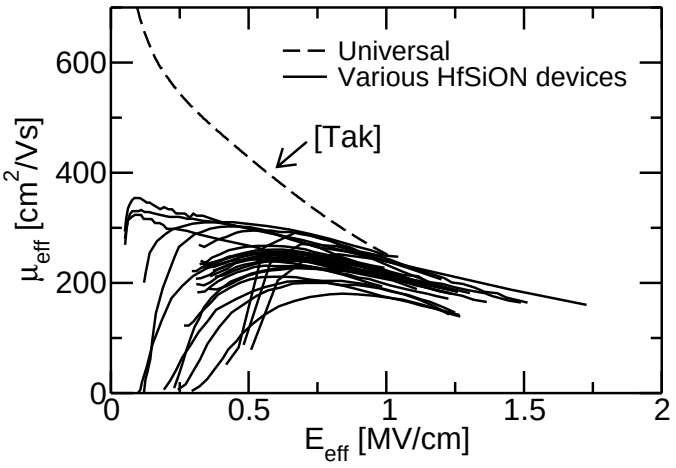

Figure 1.6: Same as in Fig. 1.5, but for HfSiON dielectric.

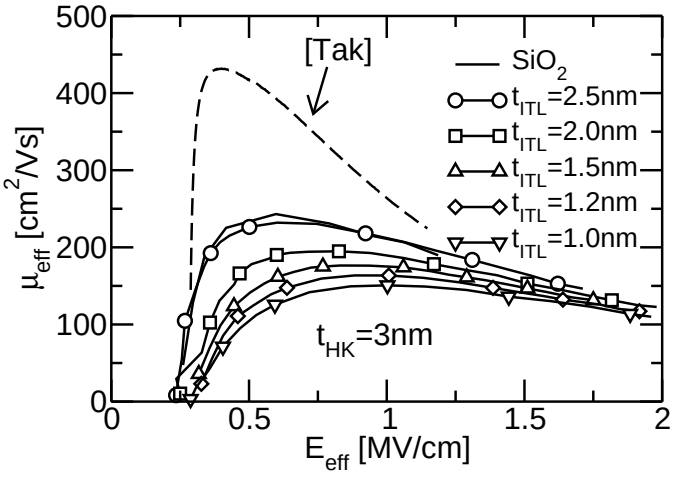

Figure 1.7: Experimental data measured in [31] at 300 K for bulk MOSFETs with doping 2×10^{17} cm^{-3}. The measurements refer to devices with 3 nm of HfO_2 and with various SiO_2 thicknesses from 1.0 nm to 2.5 nm. All devices measured have TiN metal gate. The SiO_2 reference device of [31] and the universal mobility curve of [47] are also shown.

with respect to the SiO_2 MOSFET, the device with interfacial layer thickness of 1.0 nm shows a peak mobility reduction of about 38% with respect to the SiO_2 device measured in [31], that is however much lower than the universal mobility curve [47].

1.2.4 Possible causes of mobility degradation in high-k stacks

In the previous section we have seen that the use of high-k dielectrics leads to a non-negligible mobility degradation with respect to conventional SiO_2 devices.

Fig. 1.7 shows that the mobility degradation seems to be strictly related to the interfacial layer thickness (and thus to the distance of the high-k insulator from the free carriers in the channel).

This mobility degradation has been ascribed to the vibration of the polar molecules of high-k dielectrics (the so-called soft optical phonons [48, 49, 50, 51]) or to fixed charges which are located in the gate stack. These charge can reside in the gate stack due to the non-ideality of the MOSFET fabrication process (in the form of single charges [31, 52]) or can be intrinsic and thus features of the interfaces in the gate stack (in the form of dipoles [53, 54, 55]).

In the first part of Chap. 5 we develop models for surface optical phonon scattering, single Coulomb centers and dipoles in gate stacks with interfacial layer, high-k material and metal gate. Moreover, we will compare a wide set of experimental data and Monte Carlo simulations in order to establish which is the main cause of the observed electron and hole mobility degradation.

Finally, we will compare simulated and experimental ON-current in short channel devices in order to assess the effect of the high-k dielectrics in modern MOSFETs.

1.3 Multi gate structures

Due to the short channel effects, in order to maintain a good control of the channel region by the gate contact, the channel length can not be scaled down below a given limit that depends on the depletion region thickness. The thinner is the depletion region, the shorter the channel can be. Indeed, for instance, this is one of the reason that has lead to highly doped channel regions in scaled bulk devices.

In order to avoid this technological limit, one possible solution is to change the structure of the MOSFET. A possible modification of the conventional bulk structure is to substitute the bulk region with a bottom oxide, obtaining the Single Gate SOI (SG-SOI) structure. Doing this, the minimum channel length of a technology is not related to the doping, but to the thickness of the channel region (with the assumption that the device channel is fully depleted) [56, 57].

Unfortunately, the substitution of the bulk doped region with an oxide is a very complex step and maybe not justified, even if the fully depleted SOI has potentially the possibility to overcome the bulk technology limits. However, once the bulk region has been eliminated, we can think to add a gate contact also at the bottom of the device obtaining the so-called Double Gate SOI (DG-SOI).

This further step in the process technology evolution, will expand the scenario of the ultra scaled MOSFETs to the multi-gate structures [58].

1.3.1 Benefits of using multi gate structures

The fabrication of the DG-SOI structure [5, 59] (see Fig. 1.8b) is even more difficult than the SG-SOI one, but it further improves the advantages of the SG-SOI structures over the bulk technologies [6, 60, 61].

It has been shown that the DG-SOI structure has better immunity to short channel effects, with the same Si thickness of the SG-SOI one. Indeed, the constraint of the maximum channel thickness as a function of the channel length is less limiting, leading the DG-SOI to be most promising to overcome the limits of the bulk technology. As an example, in the ITRS [10] a SG-SOI wich features a gate length of 17 nm (planned for 2015), requires a channel thickness of 5.5 nm. A DG-SOI device with the same channel length, needs a channel thickness of 8 nm.

Here we report a brief list of the main advantages of the DG-SOI structure with respect to the bulk technology:

– better immunity to short channel effects: for the reason previously explained, these kind of devices are easier to scale.

– larger gate capacitance: the larger is the number of gate contacts and the larger is the gate capacitance. With a large gate capacitance, the density of the carriers in the channel increases, leading to a larger ON-current (see Eq. 1.3).
– sub-threshold slope: the sub threshold slope (i.e. the slope of the logarithm of the I_{DS}-V_{GS} curve when $V_{\mathrm{GS}}<V_{\mathrm{TH}}$ [62]) is increased, getting closer to the theoretical limit of 60 mV/dec. This leads to have smaller OFF-current I_{OFF}, or equivalently, a lower threshold voltage V_{TH}, with the same ON-current.
– the channel region can be undoped leading to several advantages:
  • better mobility of the carriers
  • better immunity to doping variability
  • less parasitic capacitances

There is an additional possible improvement of the DG-SOI structure. We can add a third gate contact to the channel in order to further increase the gate capacitance. Such a 3D structure, called FinFET structure [63, 64, 65], if made on a bulk wafer (as in Fig. 1.8c) combines the advantages of the DG-SOI structures with the relative process simplicity of the bulk MOSFET fabrication. To be more precise, if on one hand the process fabrication is difficult since it is 3D for these structures, to the other hand the necessity to having a bottom gate is avoided.

Summarizing, if the total number of gate electrodes is 2, we obtain the DG-SOI. If the device has 3 gate electrodes, we have the triple gate structures (where the most famous architecture is the FinFET). Finally, if the gate electrodes are 4, we have the so called gate all around geometries [66] (see Fig. 1.8d).

1.3.2 Problems related to the modeling of multi gate structures

From the device modeling point of view, two main problems arise:

– the proper modeling of the quantization is more complex because the confinement of the carriers is no longer only due to the squeezing of the carrier at the channel/oxide interface; indeed the confinement is also due to the device geometry. Moreover, in FinFETs and in Gate all-around MOSFETs, the modeling of the quantization is a 2D problem, differently from bulk, SG-SOI and DG-SOI structures, where a 1D solution is required;
– the treatment of the surface roughness scattering mechanism requires a more complex formulation;
– more complex models for screening are required, due to the increased number of inversion layers in the device.

In Chap. 4 we explore the last aspect, by analyzing in detail the methodology to properly take into account the screening effect in multi gate structures.

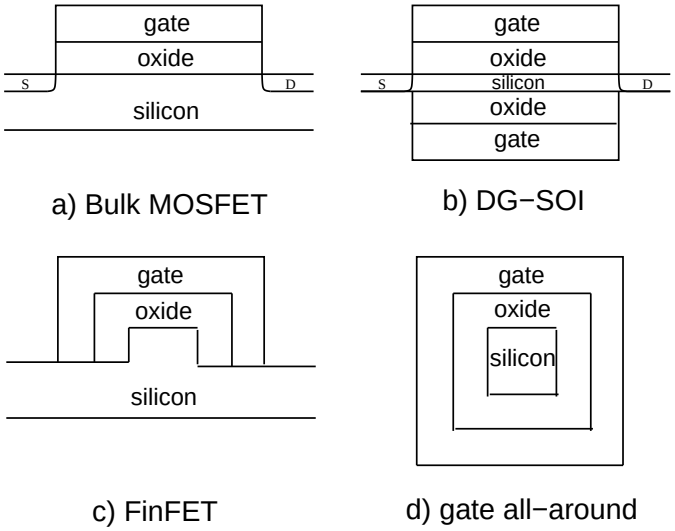

Figure 1.8: Comparison between various MOSFET device structures. a) classical bulk MOSFET. b) DG-SOI MOSFET. c) FinFET. d) gate-all-around MOSFET. In the a) and b) case, the device view is parallel to the transport direction. In the c) and b) case, the device view is orthogonal to the transport direction.

1.4 Alternative channel materials

In Sec. 1.1.2 we have seen that nowadays it is becoming extremely difficult to continue the MOSFET scaling. A possible solution is to change the channel material.

1.4.1 Benefits of using alternative materials

The low field mobility (Eq. 1.2) is a property of the channel material. If we choose more favorable materials, i.e. that have a higher low-field mobility than the conventional Si, we have carriers that can move faster through the channel. Thus, we can obtain higher ON-currents, congruently with the basic formula:

$$I_{\mathrm{ON}} \simeq Q_{\mathrm{eff}} \cdot v_{\mathrm{eff}} \tag{1.3}$$

In this regard, the easiest improvement with respect to the conventional Si MOSFET is to use the Strained-Si (s-Si). The s-Si technology have been demonstrated to yield mobility enhancements in CMOS devices and is currently used in the most advanced IC to boost performances beyond the improvements given by device scaling [67, 68, 69]. Fig. 1.9 shows the comparison between the Si and s-Si experimental data of [70, 71]. The device of [70] features a doping of $1{\times}10^{16}$ cm^{-3}, while the data of [71] refers to a device with channel doping of about $3{\times}10^{17}$ cm^{-3}. Both the experimental data refers to bi-axial tensile strain and to very thick SiO_2 oxide devices (13 and 6.7 nm). The mobility enhancement

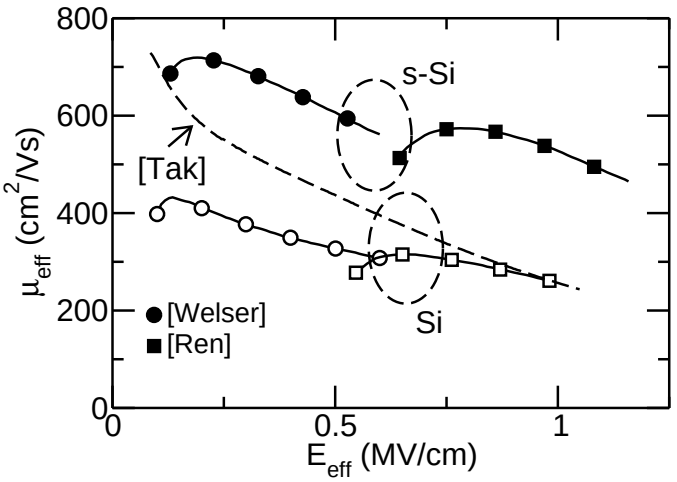

Figure 1.9: Comparison between the Si and s-Si electron mobility of [70, 71]. The universal mobility curve of [47] is also shown.

is evident and, in particular, it allows to overtake the universal mobility curve of [47].

Moreover, it is possible to make less conservative choises. New materials are indeed emerging as possible competitors of the Si and s-Si devices, for instance the Ge, strained germanium (s-Ge) and III-V materials.

The introduction of these materials is more difficult with respect to the s-Si, but interesting results have already been published for both the unstrained [72, 73] and strained Ge [74] for n-MOSFETs. In this regard, Fig. 1.10 shows the electron mobility for a device featuring a high-k dielectric (GeO_2) on the top of an unstrained Ge bulk. We can see that, despite of the use of a high-k dielectric, the mobility is comparable with the universal one for Si. Thus, Ge and s-Ge are promising candidates for overcome the Si limits.

Finally, much attention has been paid to III-V materials for high-performance devices. The research activity for the integration of these type of materials in these devices is not as advanced as the Ge based materials, but they appear very promising [75].

1.4.2 Strained silicon, germanium and strained germanium modeling

In the first part of Chap 6, we will briefly describe a methodology for the modeling of s-Si, Ge and s-Ge. Then we will compare the performances in ultra scaled n-MOSFETs, in order to assess the competitivity of Ge and s-Ge with respect to Si and s-Si.

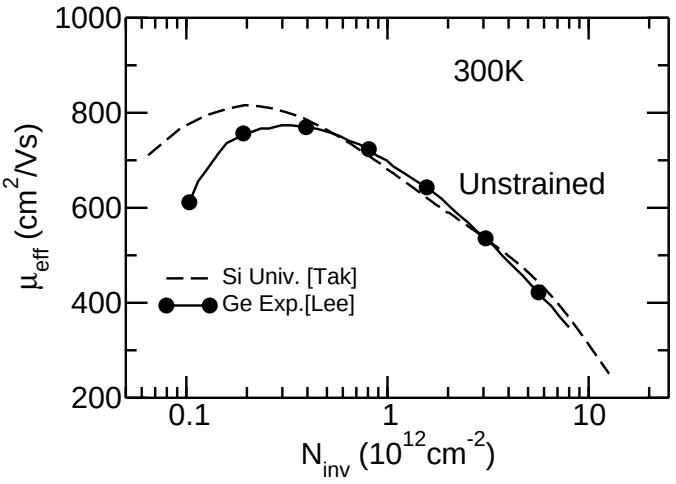

Figure 1.10: Comparison betwenn the Universal Si mobility of [47] and the Ge experimental data of [73].

1.5 Evaluation of the effect of technology boosters in real devices

High-mobility materials are introduced to boost the ON-current in modern MOSFETs.

We will see in Chap. 7 that the definition of mobility is valid only when the number of the scattering events of the carriers is very large. However, high mobility materials are used in order to reduce the number of scattering events. Moreover, if we consider short channel devices, the number is reduced even more, and the concept of mobility is hard to apply. Thus, the mobility is not a valid metric to describe the behaviour in short channel devices featuring high mobility materials.

Unfortunately, also the drain current, alone, can not give all the information that is needed to correctly characterize a modern MOSFET. Thus, we need other parameters as, for example, the limiting velocity v_{lim} of the carriers in the channel of the device. For instance, the v_{lim} can be a useful indicator of the regime where the transistor is operating and of the level of strain of the channel material.

In this regard, in Chap. 7 we use the Multi Subband Monte Carlo simulator in order to validate an existing extraction procedure of the limiting velocity in modern short channel devices [1]. After a detailed analysis of the method, we identify the sources of error of the method. Finally, we propose a new methodology for the extraction of the limiting velocity v_{lim} and we extensively validate it.

Chapter 2

The semi-classical approach and the Boltzmann transport equation

The exact solution of the motion of carriers in a nano structure as a modern MOSFET device is found solving a full quantum problem, where the carriers are described as wave-packets.

However, if the size of the wave-packets representing the carriers are much smaller than their mean free path (the length travelled by an electron between two successive collisions), we can treat the carriers as localized particles with a well defined position and momentum. This kind of approach is called *semi-classical*. In this chapter we present various modeling approaches for MOSFET devices in the semi-classical framework.

Firstly, we present the Boltzmann transport equation, which is the general equation governing the semi-classical transport.

Moreover, we derive the Drift Diffusion model as a simplified solution of the Boltzmann transport equation. We also see the analytic expressions for the drain current which can be derived from the Drift-Diffusion model and identify the main limitations of the Drift-Diffusion model.

Then, we describe the ballistic and quasi-ballistic transport regimes, and we show simplified analytic expressions for the drain currents in those regimes.

Finally, we briefly review the Monte Carlo method which allows to find an exact solution of the Boltzmann transport equation in a statistical way.

2.1 The Boltzmann transport equation

The Boltzmann Transport Equation (BTE) [76] is the basis for the semi-classical description of carrier transport in electron devices. The Boltzmann equation is a continuity equation in the phase space.

We start by defining some quantities that allow to completely describe each carrier in our system:

$$\begin{cases} \mathbf{R} = (x, y, z) \\ \mathbf{p} = (p_x, p_y, p_z) \\ t \end{cases} \tag{2.1}$$

where $\mathbf{R}$ is a vector which determines the position of the particle, $\mathbf{p}$ defines the momentum of the particle, and t is the instant of time considered.

We can describe the entire population of the carriers by considering the distribution function of the carriers $f(\mathbf{R}, \mathbf{p}, t)$, which is the probability to find carriers located at the position $\mathbf{R}$, with momentum $\mathbf{p}$ at the istant t [76]. The $f(\mathbf{R}, \mathbf{p}, t)$ can be obtained by solving the BTE [77, 78]:

$$\frac{\partial f}{\partial t} + \left(\frac{d\mathbf{R}}{dt}\right) \cdot \nabla_r f + \left(\frac{d\mathbf{p}}{dt}\right) \cdot \nabla_p f = \left(\frac{\partial f}{\partial t}\right)_{\mathrm{C}} + U(\mathbf{R}, \mathbf{p}, t) \tag{2.2}$$

In the left hand side of the equation, we can note that:

$$\left(\frac{d\mathbf{R}}{dt}\right) = \mathbf{v}_{\mathrm{g}} \tag{2.3}$$

where $\mathbf{v}_{\mathrm{g}}$ is the group velocity of the carriers, and that:

$$\left(\frac{d\mathbf{p}}{dt}\right) = \pm e\mathbf{E} \tag{2.4}$$

where $\mathbf{E}$ is the driving field. Eq. 2.4 is a version of the Newton's law. In the right hand side of the equation, $(\partial f/\partial t)_C$ is the change of the distribution function due to the scattering events of the carriers and $U(\mathbf{R}, \mathbf{p}, t)$ accounts for the changes of the distribution function due to the generation and recombination mechanisms.

Eq. 2.2 is simply the continuity equation of the carrier fluxes in each portion ($d\mathbf{R}$,$d\mathbf{p}$) in the ($\mathbf{R}$,$\mathbf{p}$) space, as we can see in Figure 2.1. Indeed, the BTE states that the number of carriers (expressed in terms of probability of occupation of a state) in position $\mathbf{R}$ with momentum $\mathbf{p}$ at the time t can change only if some carriers change their position from/to $\mathbf{R}$, or change their momentum from/to p (due to acceleration of the particle), or scatter (changing its momentum $\mathbf{p}$).

2.1.1 Validity of the Boltzmann transport equation

It is worth to ask ourselves when it is realistic to apply the BTE to the carrier transport in nano-devices. The limits of validity of the BTE are summarized below:

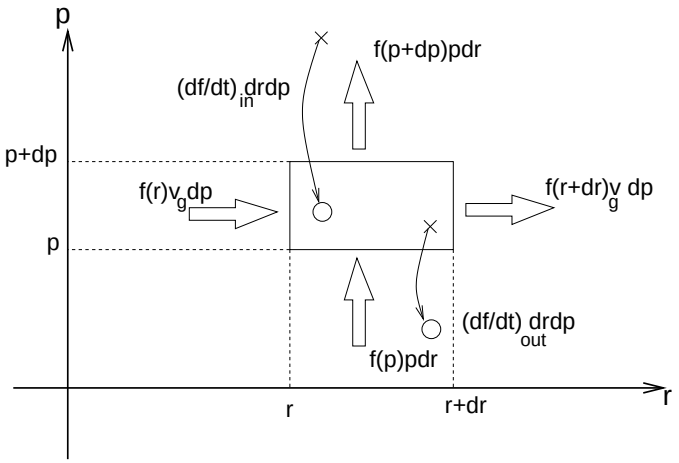

Figure 2.1: Cell in the phase space at time t, with all the possible carrier fluxes that may contribute to modify the carrier population in the cell.

- The BTE uses a classical description of the particle where both the position and the momentum of a particle can be determined. This approximation is reliable when the Heisenberg uncertainty principle of the quantum mechanics can be neglected.
- The semi-classical approach is valid only if the external electric field slowly varies over a distance comparable to the size of the wave packet which would correctly describe the carrier in a quantum approach. For the same reason, the applied potential variations have to be much smoother than the crystal potential variations. Otherwise, the Newton's law (Eq. 2.4) can not be applied to the carriers and a more complicated treatement based on wave equations is required.

2.2 The Drift - Diffusion model

We will see in this Chapter that the exact solution of the BTE (Eq. 2.2) requires computationally demanding methodologies. Thus, it is common to make some assumptions in order to simplify the Boltzmann transport equation.

Firstly, we assume the carriers as point particles with momentum:

$$\mathbf{p} = m^* \cdot \mathbf{v} \tag{2.5}$$

where m^* (assumed isotropic) is the effective mass of the electron in the semiconductor. We can express the electron density $n(\mathbf{R}, t)$ by means of the distribution function, as:

$$n(\mathbf{R}, t) = \frac{1}{\Omega} \sum_{\mathbf{p}} f(\mathbf{R}, \mathbf{p}, t) \tag{2.6}$$

where Ω is the normalization volume.

By computing the balance equation for the zero-th order moment of the BTE we can find the continuity equations [77, 78]:

$$\begin{cases} \frac{\partial n}{\partial t} - \frac{1}{e}\nabla_r \mathbf{J}_{\mathrm{n}} = U \\ \frac{\partial p}{\partial t} + \frac{1}{e}\nabla_r \mathbf{J}_{\mathrm{p}} = U \end{cases} \tag{2.7}$$

where $\mathbf{J}_{\mathrm{n}}$ and $\mathbf{J}_{\mathrm{p}}$ are the electron and hole current densities, and can be expressed as:

$$\begin{cases} \mathbf{J}_{\mathrm{n}} = -\frac{e}{\Omega}\sum_{\mathbf{p}}^{n\in CB} v(\mathbf{p}) f_n(\mathbf{R},\mathbf{p},t) \\ \mathbf{J}_{\mathrm{p}} = -\frac{e}{\Omega}\sum_{\mathbf{p}}^{n\in VB} v(\mathbf{p})(1 - f_n(\mathbf{R},\mathbf{p},t)) \end{cases} \tag{2.8}$$

where $\sum_{\mathbf{p}}^{\mathrm{CB}}$ and $\sum_{\mathbf{p}}^{\mathrm{VB}}$ are sums that are performed in the conduction and valence bands, respectively. Please note that while the probability to find an electron with energy E is expressed by $f(E)$, the probability to find a hole at that energy level is $1-f(E)$.

Assuming that the semiconductor is not degenerate, the distribution function becomes:

$$f(E) = e^{-\frac{E-E_{\mathrm{F}}}{KT}} \tag{2.9}$$

Eq. 2.9 describes a Maxwell-Boltzmann statistics, which is valid approximately when $E-E_{\mathrm{F}}>3KT$.

Assuming as closing condition of the system Eq. 2.9, from the balance equation for the first order moment of the BTE we can find the expression for the current density [77, 78]:

$$\begin{cases} \mathbf{J}_{\mathrm{n}} = -e\mu_{\mathrm{n}} n\nabla\phi + eD_n\nabla n \\ \mathbf{J}_{\mathrm{p}} = -e\mu_{\mathrm{p}} p\nabla\phi - eD_p\nabla p \end{cases} \tag{2.10}$$

where μ_{n} and μ_{p} are the mobilities for the electrons and holes (defined in Eq. 1.2), respectively. Moreover, ϕ is the electrostatic potential and D_n and D_p are the diffusion coefficients for electrons and holes [27] that can be expressed as a function of μ_{n} and μ_{p} by means of the Einstein relations for non-degenerate gas at thermodynamic equilibrium [77], which are:

$$\begin{cases} D_n = \frac{K_{\mathrm{B}}T}{e}\mu_{\mathrm{n}} \\ D_p = \frac{K_{\mathrm{B}}T}{e}\mu_{\mathrm{p}} \end{cases} \tag{2.11}$$

Finally, the electrostatic potential can be found by means of the Poisson equation:

$$\nabla\cdot\epsilon\nabla\phi = -e(p - n + N_{\mathrm{D}} - N_{\mathrm{A}}) \tag{2.12}$$

where N_{D} and N_{A} are the donor and acceptor concentrations of the semiconductor.

Eqs. 2.7, 2.10 and 2.12 form the so-called *drift-diffusion model*, which is used in the TCAD commercial tools. Please note that in this model the unknown values are n, p and ϕ, while the electron and hole mobilities μ_{n} and μ_{p} as well as U are model parameters.

Thus, models for the mobility evaluation are required in the Drift-Diffusion framework.

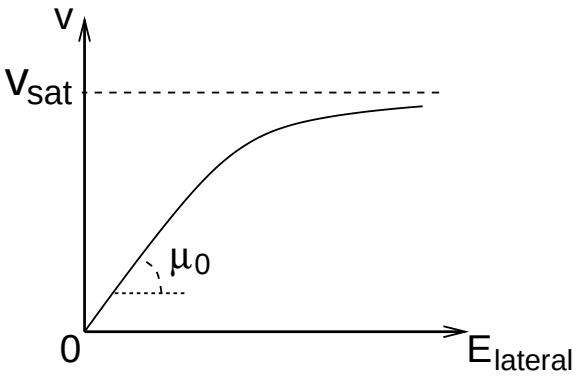

Figure 2.2: Graphical definition of the mobility and the saturation velocity.

2.2.1 The saturation velocity

Eq. 1.2 defines the mobility μ as the coefficient of proportionality between the velocity of the carriers and the driving electric field.

However, Eq. 1.2 is valid only when the driving electric field E_{lateral} is not too high. Otherwise, the increasing rate of interaction of the carriers with the crystal tends to limit the maximum attainable velocity.

This effect lead to the loss of the proportionality between the carrier velocity and the driving field, for high E_{lateral}. Indeed, experimental data [79] and Monte Carlo simulations [80] have shown that the relation between these quantities can be better modeled by [77]:

$$v = \frac{\mu E_{\mathrm{lateral}}}{\left[1 + \left(\frac{\mu E_{\mathrm{lateral}}}{v_{\mathrm{sat}}}\right)^{\beta}\right]^{1/\beta}} \tag{2.13}$$

where v_{sat} is the saturation velocity and β is a fitting parameter which for Si is 1.109 and 1.213 for electrons and holes, respectively. The saturation velocity v_{sat} has the physical meaning of the maximum velocity that carriers can have in a crystal under uniform transport conditions, as shown in Fig. 2.2. For the Si, the saturation velocity v_{sat} for electrons is $1.07{\times}10^{7}$ m/s while for holes has the value of $8.37{\times}10^{6}$ m/s.

Since the saturation velocity v_{sat} is due to a loss of energy of the carriers due to their interactions with the crystal when they are strongly accelerated, this values is strictly related to the inelastic scattering events.

2.2.2 Analytic expressions for the currents

Applying the Drift - Diffusion model to a MOSFET device structure, it is possible to derive some analytic expressions for the currents, in the gradual channel approximation [62], which assumes:

$$\frac{\partial E_{\mathrm{lateral}}(\mathbf{R})}{\partial t} \ll \frac{\partial E_{\mathrm{vertical}}(\mathbf{R})}{\partial t} \tag{2.14}$$

Eq. 2.14 is called Gradual Channel Approximation (GCA) since assumes that the channel potential ϕ slowly varies along the channel over a distance of the order of the oxide thickness. In this way the surface potential ϕ_{S} at the lateral coordinate x can be found solving a 1D vertical problem. Assuming Eq. 2.14 it is thus possible to find [62]:

$$I_{\mathrm{DD,lin}} = \frac{W}{L}\mu C_{\mathrm{G}}(V_{\mathrm{GS}} - V_{\mathrm{TH}})V_{\mathrm{DS}} \tag{2.15}$$

in the linear regime, i.e. $V_{\mathrm{DS}}<V_{\mathrm{GS}}-V_{\mathrm{TH}}$. V_{GS} and V_{DS} are the gate-to-source and the drain-to-source voltages, respectively. In Eq. 2.15 W and L are the MOSFET channel width and length, μ is the mobility of the carriers, C_{G} is the effective gate capacitance and V_{TH} is the threshold voltage [62].

Assuming infinite saturation velocity, the expression for the current in saturation regime is:

$$I_{\mathrm{DD,sat}} = \frac{W}{L}\mu C_{\mathrm{G}}\frac{(V_{\mathrm{GS}} - V_{\mathrm{TH}})^2}{2} \tag{2.16}$$

Accounting for v_{sat} with β=1 gives [81]:

$$I_{\mathrm{DD,sat}} = \frac{W}{L + \mu v_{\mathrm{sat}}^{-1} V_{\mathrm{DS,sat}}}\mu C_{\mathrm{G}}\left[(V_{\mathrm{GS}} - V_{\mathrm{TH}})V_{\mathrm{DS,sat}} - \frac{1}{2}V_{\mathrm{DS,sat}}^2\right] \tag{2.17}$$

where:

$$V_{\mathrm{DS,sat}} = \frac{-1 + \sqrt{1 + 2\mu(v_{\mathrm{sat}}L)^{-1}(V_{\mathrm{GS}} - V_{\mathrm{TH}})}}{\mu(v_{\mathrm{sat}}L)^{-1}} \tag{2.18}$$

Eq. 2.17 for $L\rightarrow 0$ reduces to:

$$I_{\mathrm{DD,sat}} = W\mu C_{\mathrm{G}}(V_{\mathrm{GS}} - V_{\mathrm{TH}})v_{\mathrm{sat}} \tag{2.19}$$

Eq. 2.19 states that for extremely short channel MOSFETs (usually called *ultimate CMOS devices*), the scaling rules explained in Sec. 1.1.1 can not be applied since there is no longer any current dependence on the channel length L. Moreover, it states that the saturation current in ultimate CMOS devices is limited by the saturation velocity v_{sat}. This means, that in this framework, the performances are expected to be strictly related to the inelastic scattering parameters and material/temperature dependence.

However, the validity of Eq. 2.19 is strictly related to the applicability of the Drift-Diffusion model to short channel devices, that is briefly discussed in the following.

2.2.3 Limits of validity of the Drift - Diffusion model

The Drift-Diffusion model has been very successful due to its relative simplicity with respect to more detailed solutions of the BTE. It's still the most used model in many TCAD commercial tools. Anyway, it is based on some assumptions that are questionable under certain conditions. Summarizing, the main assumptions at the basis of the Drift-Diffusion model are:

- the mobility can be defined only when the number of scattering events undergone by a carrier moving from Source to Drain is large;
- in the model presented in Sec. 2.2, the semiconductor is in a non-degenerate condition. Since this assumption is not valid in modern devices, more complicated Drift-Diffusion models have been generalized in the case of a degenerate semiconductor.

These assumptions are fulfilled in classical long-channel MOSFET devices, especially in linear regime.

However, in ultra-scaled MOSFETs, they become questionable. For example, in short channel device in saturation regime, the carriers are quite far from equilibrium conditions. Furthermore, the number of scattering events in the channel of a modern MOSFET is not so large, and the concept of mobility is no longer exploitable. Finally, there is no any evidence that the saturation velocity defined in Sec. 2.2.1 occurs in short channel MOSFET devices.

In the next sections we will see other methods for the solution of the BTE.

2.3 The ballistic transport regime

The DD model relies on the assumption that the device is in a scattering limited regime, namely in a regime where the scattering events thermalize the carriers at equilibrium. This is the condition which allows to define the mobility μ of the carriers (Eq. 1.2). In short channel devices it is no longer possible to define the mobility since the channel length of modern devices approaches the mean free path of the carriers (the average path the carriers travel without being scattered). Thus, we need to define new parameters to describe the performances of the nano MOSFET devices.

An approach that overcomes the problem of the impossibility to define the mobility in short channel MOSFETs is the one presented in [82, 83], which is directly derived from the flux theory. This approach is called the quasi-ballistic transport model, that will be descrbed in Sec. 2.4.

For the sake of clarity, we first introduce the ballistic model [84], that is equivalent to the QB model if we assume that the particles do not scatter during their flight from the source contact to the drain one. We will discuss later the validity of this assumption.

2.3.1 Basics on the ballistic model

In the ballistic model, carriers are injected from the source region (in which they are supposed to be at equilibrium) into the channel by crossing a potential barrier whose height is modulated by the gate voltage V_{GS}. Then, carriers move across the channel and are eventually collected by the drain contact. If scattering is null, all the carriers that are able to cross the potential barrier in the channel will eventually reach the drain contact [84].

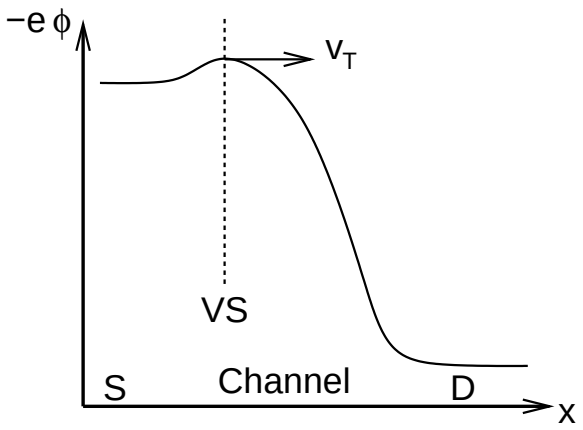

Figure 2.3: Graphical definition of the virtual source VS in the ballistic transport model.

Since we are interested in steady-state conditions, the current density in the channel can be considered solenoidal, i.e. constant in each section from the source to the drain. Thus, we can choose the most favorable section for the computation of the current I_{DS}. We decide to consider the section at the top of the potential barrier, which is called Virtual Source (VS), as shown in Fig. 2.3. The general expression for the current, if calculated at the VS, is [82]:

$$I_{\mathrm{DS}} = We(F^{+} - F^{-}) \tag{2.20}$$

where F^{+} and F^{-} are the charge fluxes at the VS directed to the drain and to the source, respectively, defined as:

$$\begin{cases} F^{+} = N^{+}_{\mathrm{inv,VS}} v^{+}_{\mathrm{VS}} \\ F^{-} = N^{-}_{\mathrm{inv,VS}} v^{-}_{\mathrm{VS}} \end{cases} \tag{2.21}$$

where N^{+}_{inv} and N^{-}_{inv} are the inversion densities of the carriers moving from source to drain and from drain to source, respectively. Similarly v^{+} and v^{-} are their average velocities.

Thus Eq. 2.20 becomes:

$$I_{\mathrm{DS,bal}} = We(N^{+}_{\mathrm{inv,VS}} v^{+}_{\mathrm{VS}} - N^{-}_{\mathrm{inv,VS}} v^{-}_{\mathrm{VS}}) \tag{2.22}$$

If we focus in the strong saturation regime, we can suppose that there are no carriers in the drain contact able to cross the barrier in the channel (which is larger than V_{DS}). Thus we can assume:

$$N_{\mathrm{inv,VS}} = N^{+}_{\mathrm{inv,VS}} + N^{-}_{\mathrm{inv,VS}} \tag{2.23}$$

$$\simeq N^{+}_{\mathrm{inv,VS}} \tag{2.24}$$

where $N_{\mathrm{inv,VS}}$ is the total inversion density of the carriers at the virtual source.

Thus, the total drain current is:

$$I_{\mathrm{bal,sat}} = WeN^{+}_{\mathrm{inv,VS}}v^{+}_{\mathrm{VS}} \simeq WeN_{\mathrm{inv,VS}}v_{\mathrm{VS}} \tag{2.25}$$

where v_{VS} is the average velocity of the carriers at the VS, since all the carriers are crossing the potential barrier in the source-drain direction.

In linear regime, the potential barrier seen by the carriers in the drain is much smaller compared to the one in the saturation regime. As a consequence, we have that, even in a case without scattering, we can not assume that all the carriers at the VS are moving from the source to the drain. Thus, the approximation in Eq. 2.24 is not valid in linear regime, whereas it is reliable to assume that the average velocity of the carriers moving in the source-drain direction is the same as the average velocity moving in the opposite direction [78]:

$$v^{+}_{\mathrm{VS}} = v^{-}_{\mathrm{VS}} \tag{2.26}$$

since V_{DS} is small.

Since the carriers are at equilibrium in the source and drain contacts with Fermi level different by eV_{DS}, assuming a Maxwell-Boltzmann distribution, we can express the inversion density of the carriers at the virtual source as [78]:

$$N^{-}_{\mathrm{inv,VS}} = N^{+}_{\mathrm{inv,VS}}e^{-\frac{eV_{\mathrm{DS}}}{K_{\mathrm{B}}T}} \tag{2.27}$$

the general expression for the drain current (Eq. 2.22) becomes:

$$I_{\mathrm{bal,lin}} = WeN_{\mathrm{inv,VS}}v^{+}_{\mathrm{VS}}\left(\frac{1-e^{-\frac{eV_{\mathrm{DS}}}{K_{\mathrm{B}}T}}}{1+e^{-\frac{eV_{\mathrm{DS}}}{K_{\mathrm{B}}T}}}\right) \tag{2.28}$$

When the device is in linear regime and V_{DS} is small, we can simplify Eq. 2.28 as:

$$I_{\mathrm{bal,lin}} = \frac{W}{2}eN_{\mathrm{inv,VS}}v^{+}_{\mathrm{VS}}\frac{eV_{\mathrm{DS}}}{K_{\mathrm{B}}T} \tag{2.29}$$

2.3.2 Analytic expressions for the current

From Fig. 2.3 we can see that the electric field driving the carriers at the VS is small (ideally null in the VS-section). Thus at first order we can assume a 1D electrostatics at the VS and write:

$$eN_{\mathrm{inv,VS}} = C_{\mathrm{G}}(V_{\mathrm{GS}} - V_{\mathrm{TH}}) \tag{2.30}$$

We can further note that if the semiconductor is not in degenerate conditions, since carriers at the VS are in equilibrium with the source fermi level, their mean velocity becomes the thermal velocity v_{T}, which takes the form [78]:

$$v_{\mathrm{T}} = \sqrt{\frac{2K_{\mathrm{B}}T}{\pi m^{*}}} \tag{2.31}$$

where m^* is the effective mass of the carrier as defined in Eq. 2.5. Using Eqs. 2.30 and 2.31, Eq. 2.25 can be re-written as:

$$I_{\mathrm{bal,sat}} \simeq W C_{\mathrm{G}} (V_{\mathrm{GS}} - V_{\mathrm{TH}}) v_{\mathrm{T}} \tag{2.32}$$

Concerning the linear regime, instead, Eq. 2.29 becomes:

$$I_{\mathrm{bal,lin}} \simeq \frac{W}{L} \mu_{\mathrm{bal}} C_{\mathrm{G}} (V_{\mathrm{GS}} - V_{\mathrm{TH}}) V_{\mathrm{DS}} \tag{2.33}$$

where we have used the so called *ballistic mobility* (usually called *apparent mobility*) which takes the form:

$$\mu_{\mathrm{bal}} = \frac{eLv_{\mathrm{T}}}{2K_{\mathrm{B}}T} \tag{2.34}$$

The dependence of the drain current in Eq. 2.33 on the channel length L is not a real effect, and it appears in the expression due to the definition of the apparent mobility (proportional to L).

2.3.3 Considerations about the ballistic transport model

We can see that Eq. 2.19 and Eq. 2.32 have similar expressions for the drain current in saturation regime, despite of the very different transport models used for their derivation.

Indeed, we saw that in the Drift - Diffusion model the current in ultimate MOSFETs is limited by the v_{sat} which is strictly related to the inelastic scattering events and assumes a kind of uniform field profile. Eq. 2.32, instead, has been deduced assuming a transport without scattering events. The thermal velocity v_{T} is related to the mass of the carrier in the crystal m^*. Thus, while v_{sat} is related to the scattering events, v_{T} derives from the properties of the conduction band only.

A further difference between v_{sat} and v_{T} is the following: while v_{sat} is present where the driving field is the highest (usually the end of the MOS channel), v_{T} can be defined only where the driving field is small (the VS), confirming that the Drift-Diffusion and ballistic models, although giving similar expressions for $L \to 0$, are conceptually very different.

2.4 The quasi ballistic transport regime

2.4.1 Basics of the quasi-ballistic model

In Sec. 2.2.3 we have seen that the Drift-Diffusion model (Sec. 2.2) can not be reliably applied to short channel devices, due to its several approximations. However, also the ballistic model (Sec. 2.3) can not be safely applied to nowadays MOSFETs due to the assumption of no-scattering in the channel region (Sec. 2.3) [85, 86].

What it is possible to do is to consider an extension of the ballistic transport model to the case with scattering events. This framework is the so called *quasi ballistic transport regime* and can describe a device in which the number of scattering events occured is not negligible, but it is not large enough to thermalize the carriers as in the Drift-Diffusion model. We will see that this framework is able to cover both the Drift-Diffusion and the ballistic transport regimes as limiting cases.

The validity of the quasi-ballistic model is still debated [83, 87]. However, it has been shown in [88] that this model can be considered valid, but the evaluation of the effects of the scattering events is critical. We will briefly discuss this point in Sec. 2.4.3.

For the sake of simplicity and to be congruent with Sec. 2.3.2 we will adopt the Maxwell-Boltzmann statistics to derive analytical expressions for the currents in Secs. 2.4.2 and 2.4.3. However, similar results can be derived in the case of Fermi-Dirac statistics [89].

2.4.2 Analytic expressions for the current

Firstly, we define the back-scattering coefficient r as [82]:

$$r = \frac{F^+}{F^-} \tag{2.35}$$

where F^+ and F^- have been defined in Eq. 2.21. Eq. 2.35 is valid at high V_{DS} where it has been assumed that there is no flux of the carriers from the drain contact to the source one. The proper modeling of the back-scattering coefficient r is the most critical point of the quasi ballistic model.

Starting from Eq. 2.20, and using Eqs. 2.35 and 2.21, we can write:

$$I_{\mathrm{DS,Q-bal}} = eWN^+_{\mathrm{inv,VS}} v^+_{\mathrm{VS}} (1 - r) \tag{2.36}$$

Now we assume that Eq. 2.26 is valid for both the linear and saturation regimes. While this assumption is fulfilled in linear regime, it may become inaccurate in the saturation regime [85]. However, the impact of this assumption is not critical in the cases of pratical interest where r is small.

Using Eqs. 2.26 and 2.23, and remembering the definition of r (Eq. 2.35), we obtain:

$$N^+_{\mathrm{inv,VS}} = \frac{N_{\mathrm{inv,VS}}}{1 + r} \tag{2.37}$$

Eq. 2.37 means that, regarding the saturation regime, since the carriers are back-scattered towards the VS-section, the portion N^+ of the inversion charge density at the VS effectively contributing to the I_{DS} is not equal to the N_{inv} (as in the ballistic case), but it is decreased by a factor r.

Thus, from Eq. 2.36 and 2.37 we obtain the expression valid in saturation regime [83]:

$$I_{\mathrm{Q-bal,sat}} \simeq WN_{\mathrm{inv,VS}} v^+_{\mathrm{VS}} \left(\frac{1 - r}{1 + r} \right) \tag{2.38}$$

Finally, using a Maxwell-Boltzmann statistics and using Eq. 2.30, we finally have:

$$I_{\mathrm{Q-bal,sat}} \simeq WC_{\mathrm{G}}(V_{\mathrm{GS}} - V_{\mathrm{TH}})v_{\mathrm{T}}\left(\frac{1-r}{1+r}\right) \tag{2.39}$$

Concerning the linear regime, instead, we can proceed as follows. It can be found that the generalization of Eq. 2.28 in the case with scattering events is [78]:

$$I_{\mathrm{Q-bal,lin}} = WeN_{\mathrm{inv,VS}}v^{+}_{\mathrm{VS}}\left(\frac{(1-r)-(1-r)e^{-\frac{eV_{\mathrm{DS}}}{K_{\mathrm{B}}T}}}{(1+r)+(1-r)e^{-\frac{eV_{\mathrm{DS}}}{K_{\mathrm{B}}T}}}\right) \tag{2.40}$$

Assuming $V_{\mathrm{DS}} \ll K_{\mathrm{B}}T/e$ and Eq. 2.34, Eq. 2.40 reduces to:

$$I_{\mathrm{Q-bal,lin}} \simeq \frac{W}{L}\mu_{\mathrm{bal}}C_{\mathrm{G}}(V_{\mathrm{GS}} - V_{\mathrm{TH}})V_{\mathrm{DS}}(1-r) \tag{2.41}$$

Comparing Eq. 2.33 with Eq. 2.41, we can see that, differently from the saturation case, the density of the carriers at the VS moving to the drain $N^{+}_{\mathrm{inv,VS}}$ is not modified by the introduction of the scattering events in the model.

2.4.3 Models for the back-scattering coefficient

As already said, the evaluation of the r parameter is the most critical point of the quasi-ballistic model. Indeed, it is very difficult to develop analytical solutions of the Boltzmann transport equation valid in the quasi-ballistic transport regime. Thus, to obtain a compact model for the drain current one needs to introduce many simplifications to the problem.

In the saturation regime, the coefficient r is assumed to be [82, 78]:

$$r = \frac{L_{\mathrm{KT}}}{L_{\mathrm{KT}} + \lambda} \tag{2.42}$$

where λ is the mean free path of the carriers and L_{KT} is called *KT-layer* and it is defined as the distance over which the potential barrier in the channel falls of $K_{\mathrm{B}}T/e$, as shown in Fig. 2.4. Since the average energy of the back-scattered carriers is $K_{\mathrm{B}}T/e$ [83], Eq. 2.42 means that only the back-scattered carriers that have to cross a barrier lower than their average energy can go back to the source contact. The validity of Eq. 2.42 has been analyzed in [88] where it has been found that the definitions of L_{KT} and λ are critical in short channel devices [88].

For long channel devices, an approximated expression for L_{KT} can be derived [82, 90]:

$$L_{\mathrm{KT}} = L\frac{2K_{\mathrm{B}}T}{e(V_{\mathrm{GS}} - V_{\mathrm{TH}})} \tag{2.43}$$

Indeed, under the GCA (Eq. 2.14) [78, 77], we can write:

$$L_{\mathrm{KT}} = \frac{K_{\mathrm{B}}T/e}{E_{\mathrm{VS^{+}}}} \tag{2.44}$$

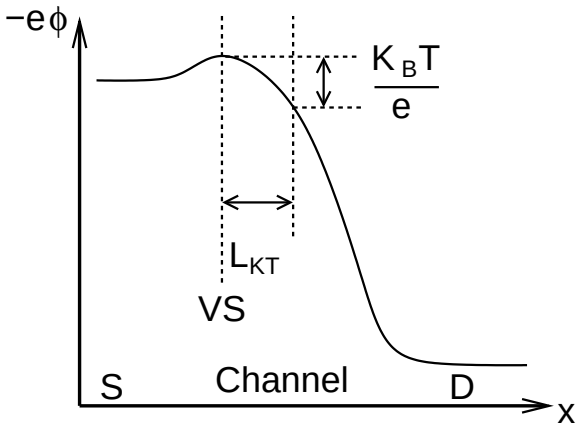

Figure 2.4: Graphical definition of the KT-layer in the ballistic transport model.

where $E_{\mathrm{VS^+}}$ is the electric field at the VS, at the side of the drain direction. The value of $E_{\mathrm{VS^+}}$ can be obtained from the expression of the surface potential profile in the channel. Under the same assumptions leading to the analytic expression for I_{DS} in the Drift-Diffusion model (Sec. 2.2.2), it is possible to express the surface potential as [81]:

$$\phi_{\mathrm{c}}(l) = (V_{\mathrm{GS}} - V_{\mathrm{TH}})\left(1 - \sqrt{1 - \frac{l}{L}}\right) \tag{2.45}$$

where l is:

$$l = x - x_{VS} \tag{2.46}$$

From Eq. 2.45 we can obtain:

$$E_{\mathrm{VS^+}} = -\frac{d\phi_{\mathrm{c}}(l)}{dl}\bigg|_{0^+} = \frac{V_{\mathrm{GS}} - V_{\mathrm{TH}}}{2L} \tag{2.47}$$

Eqs. 2.44 and 2.47 bring us to Eq. 2.43.

The mean free path λ can be approximated as [90, 91, 78, 77, 92]:

$$\lambda = \frac{2\mu K_{\mathrm{B}} T}{e v_{\mathrm{T}}} \tag{2.48}$$

where μ is the long channel mobility. The advantage of using Eqs. 2.42 and 2.48 is that the quasi ballistic model tends to the Drift - Diffusion when $L_{\mathrm{KT}} \gg \lambda$.

In the linear regime, we assume a zero-field condition, i.e. all the device channel contributes to the back-scattering, and we thus write [82, 83, 77, 78]:

$$r = \frac{L}{L + \lambda} \tag{2.49}$$

In [88] it has been shown that Eqs. 2.48 and 2.49 well reproduce Monte Carlo simulations at low field. Finally it is worth noting that Eq. 2.33, 2.34, 2.48 and 2.49, we can obtain:

$$I_{\mathrm{Q-bal,lin}} = \frac{W}{L+\lambda}\mu C_{\mathrm{G}}(V_{\mathrm{GS}} - V_{\mathrm{TH}})V_{\mathrm{DS}} \tag{2.50}$$

It is straightforward to see that for a long channel device, since we can assume $L \gg \lambda$, Eq. 2.50 reduces to Eq. 2.15. Thus, also in linear regime, the quasi-ballistic framework is consistent with the Drift-Diffusion model for long channel devices.

2.5 Exact solution of the BTE: the Monte Carlo method

Since the effectiveness of the mobility μ as performance indicator is hard to judge in short channel devices, we need other performance indicators. In this respect, v_{T} and r can be used as the main performance indicators of the modern short devices. However, as previously said, the quasi ballistic transport regime is a complex intermediate condition between the Drift-Diffusion and ballistic models where both the crystal structure and the scattering events contribute in determining the current. Thus, an exact solution of the BTE is required to correctly model this transport regime. We now explain the basics of the exact solution of the BTE by means of the Monte Carlo method.

2.5.1 Basics of the Monte Carlo method

The MonteCarlo (MC) is a direct (not approximate) statistical method to solve integral-differential equations, as the BTE [80]. In the case of transport in semiconductors, it consists of simulating the motion of one or many electrons inside the crystal, subject to the action of external forces due to applied electric field and of various scattering mechanisms.

In Sec. 2.4 we have seen that the quasi ballistic transport model can describe both short devices under off-equilibrium transport condition, but the modeling of the back scattering coefficient r (or the average velocity of the carriers v_{T}) is non trivial. The Monte Carlo method can be used to determine these values in modern devices.

Entering the details of the model, the motion of a particle is considered semi-classical. Indeed, it is a sequence of free flights obeying the Newton's law, which is:

$$\frac{d\mathbf{p}}{dt} = -\nabla_{\mathbf{R}} E \tag{2.51}$$

where E is the total energy. During the free flight, the motion of the particle is also governed by:

$$\frac{d\mathbf{R}}{dt} = \mathbf{v}_{\mathrm{g}} = \nabla_{\mathbf{p}} E \tag{2.52}$$

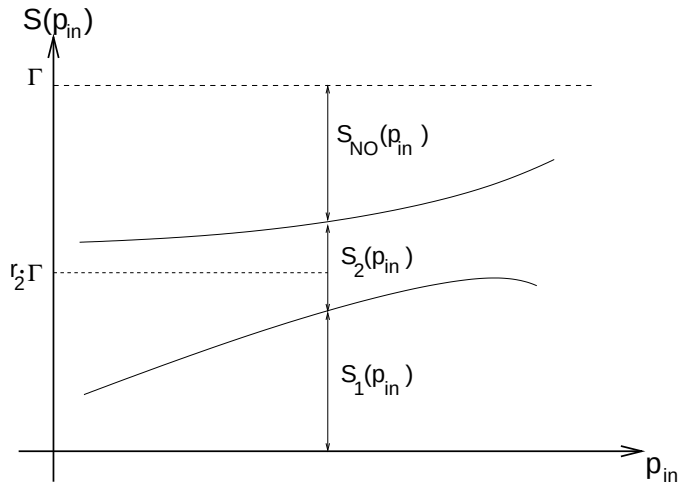

Figure 2.5: Graphic procedure for the determination of the type of the event suffered by a carrier whose state is $\mathbf{p}_{\rm in}$.

The free flights are interrupted by the scattering events. The scattering events are described by the scattering rates $S(\mathbf{R}, \mathbf{p}', \mathbf{p})$. $S(\mathbf{R}, \mathbf{p}', \mathbf{p})$ is defined as the probability per unit time that a carrier in the position $\mathbf{p}$ and with momentum $\mathbf{p}$ scatters to a state with momentum $\mathbf{p}'$.

In the self-scattering scheme [93, 78], the free flight duration is:

$$t = -\Gamma^{-1} \ln(r_1) \tag{2.53}$$

where r_1 is a random number whose probability distribution is uniform from 0 to 1. Γ is a number larger than the maximum total scattering rate (self-scattering technique), which is:

$$\Gamma > [S_{\rm TOT}(\mathbf{p})]_{\rm MAX} = \left[\sum_i \sum_{\mathbf{p}'} S_i(\mathbf{p}', \mathbf{p}) \right]_{\rm MAX} \tag{2.54}$$

where i indicates the scattering type. After the free flight, we decide either which scattering mechanism occurs or if the particles continues its motion (with the $\mathbf{p}$ that it had before the self scattering) by generating another random number r_2 which probability distribution is uniform from 0 to 1. Let us suppose, for example, that we have only two scattering mechanisms. The choice of the scattering mechanism is done as shown in Figure 2.5. If we have:

$$0 < r_2 < \frac{S_1(\mathbf{p})}{\Gamma} \tag{2.55}$$

the carrier will scatters with the mechanism "1". If we have:

$$\frac{S_1(\mathbf{p})}{\Gamma} < r_2 < \frac{S_2(\mathbf{p})}{\Gamma} \tag{2.56}$$

the carrier will scatters with the mechanism "2". Finally, if we have:

$$\frac{S_2(\mathbf{p})}{\Gamma} < r_2 < \Gamma \tag{2.57}$$

the carrier will no scatter, and will continue its previous free flight.

Finally, other random numbers are generated to establish the $\mathbf{p}'$ after the scattering, which will be used in an appropriate function depending on the scattering mechanism.

Finally, after having collected the information about the position and the energy of all the carriers [94, 95], we can determine the distribution function $f(\mathbf{R}, \mathbf{p}, t)$, which is the unknown value of the Boltzmann transport equation (Eq. 2.2). The gathering of the information about the position and momentum of the electrons can be done at the time t_n as:

$$f(\mathbf{R}_i, \mathbf{p}_j, t_n) = \frac{1}{N_\mathrm{e}} \sum_{1}^{N_\mathrm{e}} \mathrm{e}(\mathbf{R} = \mathbf{R}_i, \mathbf{p} = \mathbf{p}_j, t_n) \tag{2.58}$$

where $f(\mathbf{R}_i, \mathbf{p}_j, t_n)$ gives the probability for an electron to be at time t_n, at the position $\mathbf{R}_i$ and with momentum $\mathbf{p}_j$. N_e is the total number of electrons in the simulation. The statistics for the holes can be collected in a similar way.

In addition to this, appropriate boundary conditions have to be applied. For the interface between the channel and the gate stack, one usually imposes *reflecting* boundary conditions, i.e. a carrier wich reach the boundary of the domain is reflected in the channel. If the structure is bulk, the bottom boundary condition may be *absorbing*, i.e. a carrier reaching the bottom interface of the domain is destroyed. Considering instead the boundaries at the left and right limits of the domain, representing the source and drain reservoir of carriers, the contacts can be assumed *injecting*, i.e. the carriers are introduced into the simulating domain according to a distribution function, which typically is Fermi-Dirac distribution, since carriers are considered at equilibrium inside the source and drain regions [96]. However, the validity of this boundary condition is still debated, since it imposes that carriers enter the channel with thermal velocity (Eq. 2.31). For this reason, our group developed a different boundary conditions in which, in the source and drain contacts, the neutrality of the charge is imposed [97]. This boundary condition can take into account off-equilibrium conditions in source and drain regions.

In the next chapter we will describe how the Multi Subband Monte Carlo simulator exploits the Monte Carlo method to solve the Boltzmann transport equation in modern ultra scaled MOSFET devices.

Chapter 3

The Multi Subband Monte Carlo simulator

In this chapter we describe the basics of the Multi Subband Monte Carlo simulator employed in this thesis.

First, we see how the technique takes into account the quantization perpendicular to the tranport direction that occurs in modern MOSFET devices.

Then, we see how the simulator solves the Boltzmann transport equation in the plane perpendicular to the quantization direction. We also show how the BTE is modified by the fact that the problem is quantized in one direction.

Moreover, we review the scattering mechanisms occuring in conventional MOSFETs without the technology boosters listed in Chap. 1. In particular, we describe the perturbation potential due to phonon, ionized impurities and surface roughness scattering mechanisms.

Finally, we show the calibration of the simulator on the universal mobility curves for bulk Si devices featuring a SiO_2 dielectric.

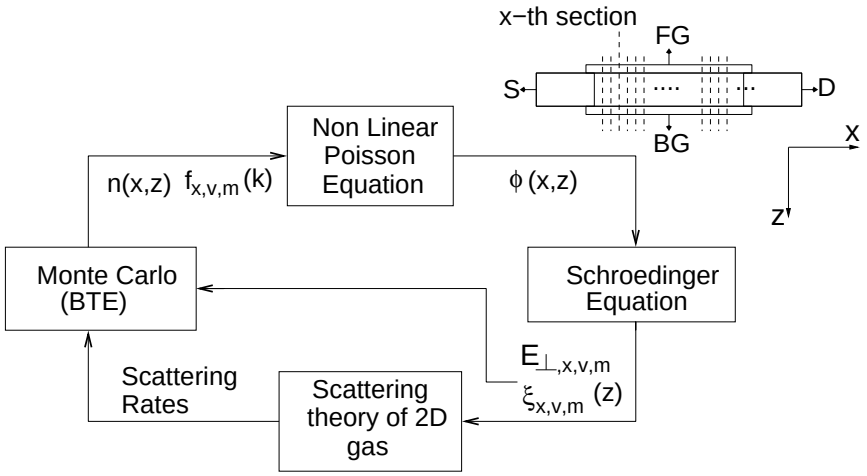

Figure 3.1: Flowchart of the Multi Subband Monte Carlo simulator. The symbols v and m indicate the valley and the subband in the x section. $\xi(z)$ is the envelope wave function, $E_\perp$ is the subband energy (that is supposed to be independent on the wave-vector k), $f(k)$ is the distribution function. n and ϕ are the electron concentration and the electrostatic potential in the x, z plane. The y-coordinate has been dropped since we consider indefinitely large devices that can be thus considered uniform in this direction.

3.1 The carrier transport framework

This simulator is called *multi subband* because it accounts for the fact that the allowed carrier energies in the direction normal to the Si/SiO_2 dielectric are discretized due to quantization [98, 99, 93, 100, 101]. This complicates the solution of the Boltzmann transport equation, since it is required to solve a BTE for each subband in the system. All these BTEs are then coupled by inter-subband scattering.

In this section we take a look to the distinct features of the technique used by our simulator to describe the carrier transport in a MOS device.

3.1.1 Flow chart

The solution of the tranport problem requires a self-consistent loop including the Schrödinger, the non-linear Poisson and the Monte Carlo solvers (see Fig. 3.1).

The first guess of the solution (potential and carrier concentration profile) is given by a drift-diffusion solution (e.g. with the Sentaurus TCAD). From the potential provided by the first guess, a one-dimensional Schrödinger equation (in the vertical direction z) is solved in each section x to find the subband energies.

At this point the scattering rates are calculated, and accordingly to them, the carriers are moved in the transport plane by the Monte Carlo transport core. The Monte Carlo solution is the carrier distribution in the $\mathbf{k}$ space, in each section x, in

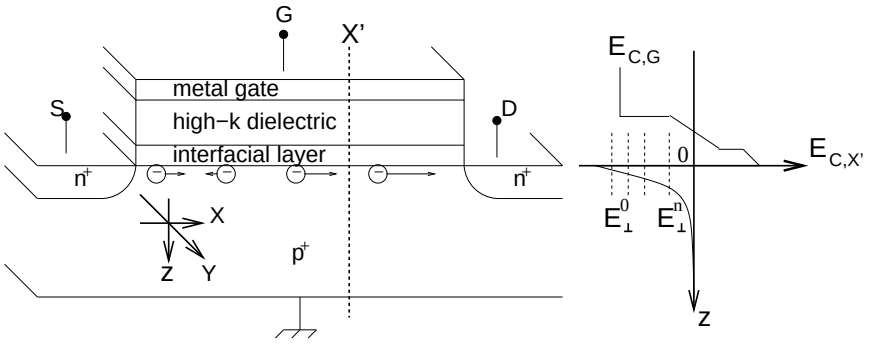

Figure 3.2: Sketch of the inversion sheet in a nano n-MOSFET. The potential well at the section with coordinate X' is also shown.

each valley v and in each subband s $f_{x,v,m}(\mathbf{k})$. From that we compute the electron concentration as:

$$n(x,z) = \frac{2}{A} \sum_{v,m,\mathbf{k}} f_{x,v,m}(\mathbf{k}) |\xi_{x,v,m}(z)|^2 \tag{3.1}$$

where $\xi_{x,v,m}(z)$ is the envelope wave function at the section x, of the wave-packet representing the carrier in the subband m belonging to the valley v, that has been found by solving the Schrödinger equation and whose meaning is explained in Eq. 3.2. A non-linear Poisson equation is solved to consider the changes of the electrostatic potential due to the movement of the particles, i.e. the new electron concentration $n(x,z)$.

In the new iteration, the electrostatic potential provided by the Poisson solver is directly used by the Schrödinger solver. Therefore, the simulation is obtained by iterating the Schrödinger solver, the Monte Carlo simulator and the Poisson solver.

3.1.2 The quantization effect

A carrier in a nano-MOSFET (Figure 3.2) is confined in one direction (the z direction in this thesis) and it can move as a free carrier (i.e. described by a plane wave) in the other two directions ($\mathbf{r}$) [93, 100].

Since the physical quantities slowly vary in the tranport direction, we assume that each section x_i is independent from the others. Therefore, the solution at each section x_i is found by considering a MOS transistor indefinitely long, with no-field in the x-direction. We can thus express the wave-function at each x-section as a wave-plane propagating in the x and y-directions, i.e:

$$\psi_{x,v,m}(\mathbf{r},z) = \xi_{x,v,m}(z) \frac{e^{i\mathbf{k}\cdot\mathbf{r}}}{\sqrt{A}} \tag{3.2}$$

where A is a normalization area. Moreover, the energy of a particle can be expressed as a sum of its energy in the vertical-direction (where it is quantized) and

the in-plane energy (where the particle is treated as classical). Thus the energy of a particle can be written as:

$$E_{\mathrm{TOT}} = E_{\perp} + E_{\parallel} \tag{3.3}$$

where $E_{\perp}$ is the potential energy (related to the quantization in the z-direction) and $E_{\parallel}$ is the kinetic energy in the transport plane. As previously said, the vertical quantization is accounted for by solving the steady-state Schrödinger equation in each device slice, which is:

$$[\hat{H} + E_{\mathrm{C}x}(z)]\xi_{x,v,m}(z) = E_{\perp x,v,m}\xi_{x,v,m}(z) \tag{3.4}$$

where H is the Hamiltonian that, under the effective mass approximation (Sec. 3.1.4), takes the form:

$$\hat{H}_{EMA} = -\frac{\hbar^2}{2m_z}\frac{\partial^2}{\partial z^2} \tag{3.5}$$

(m_z is the effective mass of the carriers in the z direction, that depend on the valley v) and $E_{\mathrm{C}x}$ is the conduction band edge in the x section, as:

$$E_{\mathrm{C}x}(z) = -e\phi(x,z) - \chi(x,z) \tag{3.6}$$

where $\phi(x,z)$ is the electrostatic potential and χ is the affinity of the material. x and s are indices which discriminate the x-coordinate and the subband in wich the electron is. Since the electrostatic potential profile changes from source to drain, the conduction band edge profile E_{C} depends also on the x coordinate. Thus, the wave-function of the carrier (which describes its vertical "position") depends on the section, the valley and the subband of the particle.

Using the Hamiltonian in Eq. 3.5, the energy of the carriers in the z-direction $E_{\perp}$ (which corresponds with the potential energy) does not depend on the wave-vector $\mathbf{k}$ [78, 102].

3.1.3 2-D multi-subband transport

Starting from the solution of the Schrödinger equation in each section x, the MC transport core simulates the motion of the particles in the in-plane direction, in a way similar to the one explained in Section 2.5 [93, 100].

A difference with the case explained in Sec. 2.5 is that now a BTE (Eq. 2.2) is solved for each subband. Moreover we choose to substitute the momentum of the carriers $\mathbf{p}$ with their wave-vectors $\mathbf{k}$ (The two representations are equivalent). Since in this case in the vertical direction the carriers can not move due to the quantization, the dimensionality of the $\mathbf{k}$-space is 2 (k_x,k_y), differently to $\mathbf{p}$ of Sec. 2.5, that has three components. Thus, here we represent the particles with the variables: $\mathbf{r}$, $\mathbf{k}$, v and s. Finally, we chose to neglect the generation-recombination term in Eq. 2.2. This means that we are neglecting the impact ionization, since we suppose that $E_{\parallel}$ does not increase too much in ultra scaled MOSFETs that typically

have V_{DS} of the order of 1 V. Moreover, we can neglect also the Shockley-Read-Hall mechanism since it is effective in long devices.

We have seen that the charge profile in the vertical direction is provided by the Schrödinger equation. Then, the Monte Carlo solver moves the carriers forcing them to remain in the same subband during the free flight. Thus, the only way a carrier has to change its subband is to have an inter-subband scattering event. The only way to change its valley is to have an inter-valley scattering event.

In detail, the Eq. 2.2 becomes in this case:

$$\frac{\partial f_{x,v,m}(\mathbf{k})}{\partial t} + \mathbf{v}_{\mathrm{g}} \cdot \nabla_r f_{x,v,m}(\mathbf{k}) + \left(\frac{\mathbf{F}_{v,m}}{\hbar}\right) \cdot \nabla_k f_{x,v,m}(\mathbf{k}) = \left(\frac{\partial f_{x,v,m}(\mathbf{k})}{\partial t}\right)_C \tag{3.7}$$

where $f_{x,v,m}(\mathbf{k})$ is function of the in-plane momentum of the carriers and $\mathbf{F}$ is the external force which is related to the eigen values of the Schrödinger equation by the relation:

$$\mathbf{F}_{v,m}(x) = -\frac{\partial E_{\perp v,m}(x)}{\partial x}\mathbf{x} \tag{3.8}$$

The carriers in different subbands are thus moved with different driving forces.

Finally, the term that indicates the change of the distribution function due to the scattering events can be expressed as:

$$\begin{aligned}\left(\frac{\partial f_{x,v,m}(\mathbf{r},\mathbf{k},t)}{\partial t}\right)_{\mathrm{C}} \;=\; & \frac{A}{4\pi^2}\sum_{v',m'}\int_{\mathbf{k}'}\big[S_{x,v',m',v,m}(\mathbf{r},\mathbf{k}',\mathbf{k}) \\ & f_{x,v',m'}(\mathbf{r},\mathbf{k}',t)\big(1-f_{x,v,m}(\mathbf{r},\mathbf{k},t)\big) - \\ & -S_{x,v,m,v',m'}(\mathbf{r},\mathbf{k},\mathbf{k}') \\ & f_{x,v,m}(r,\mathbf{k},t)\big(1-f_{x,v',m'}(\mathbf{r},\mathbf{k}',t)\big)\big]dk'_x dk'_y\end{aligned} \tag{3.9}$$

where $S_{x,v,m,v',m'}(\mathbf{r},\mathbf{k},\mathbf{k}')$ is the scattering rate. It is the probability per unit time for a carrier located in $\mathbf{r}$ to scatter from a state $v,m,\mathbf{k}$ to a state $v',m',\mathbf{k}'$. We remind that the solution at each section x_i is found by considering a MOS transistor indefinitely long, with no-field in the x-direction (a "dummy" device with uniform conditions in the transport plane). Thus the index x appears for the scattering rate $S_{x,v,m,v',m'}(\mathbf{r},\mathbf{k},\mathbf{k}')$. Eq. 3.9 considers the Pauli exclusion principle by means of the two terms $1-f$ which state that the final state has to be empty in order to have the scattering event.

In Sec. 3.2 we will see how the scattering rates of the scattering mechanisms occuring in conventional MOSFET devices can be found by means of the Fermi Golden Rule.

3.1.4 The effective mass approximation for the n-MOSFET

Here, we describe in more detail the hamiltonian in Eq. 3.5, that is valid under the Effective Mass Approximation (EMA) to solve the Schrödinger equation.

At low energies, the three-dimensional energy dispersion relationship $E(\mathbf{k})$ can be considered parabolic, i.e.:

$$E(\mathbf{k}) - E_{\mathrm{C}} = \frac{\hbar^2}{2}\left(\frac{k_x^2}{m_x^*} + \frac{k_y^2}{m_y^*} + \frac{k_z^2}{m_z^*}\right) \tag{3.10}$$

where $\hbar$ is the Planck constant and E_{C} is the bottom of the conduction band. m_x^*, m_y^* and m_z^* are the effective masses in the x, y and z directions, respectively. The expression in Eq. 3.10 takes into account the possible anisotropy of the valleys.

Using Eq. 3.10, and remembering the expression of the wave functions describing the carriers (Eq. 3.2), we obtain the Schrödinger equation in Eq. 3.4, where $\hat{H}$ is the one defined in Eq. 3.5. We obtain also:

$$E(\mathbf{k}) - E_{\perp x,v,m} = \frac{\hbar^2}{2}\left(\frac{k_x^2}{m_x^*} + \frac{k_y^2}{m_y^*}\right) \tag{3.11}$$

that is the two-dimensional energy dispersion relationship in the transport plane.

A carrier that can be described with Eq. 3.2 moves with group velocity [78]:

$$v_{\mathrm{g}}(\mathbf{k}) = \frac{1}{\hbar}\nabla_{\mathbf{k}} E(\mathbf{k}) \tag{3.12}$$

Thus, if the energy dispersion relationship is the one in Eq. 3.11, we can write:

$$v_{\mathrm{g}} = \frac{p_{x,y}}{m_{x,y}^*} = \frac{\hbar k_{x,y}}{m_{x,y}^*} \tag{3.13}$$

thus confirming that, with a parabolic dispersion relationship, m^* can be seen as the effective mass with which the carrier moves in a crystal.

Since in ultra scaled MOSFETs, the condition of not too strong off-equilibrium transport could be not verified, a modification of Eq. 3.10 should be required in order to take into account non-parabolicity effects, as:

$$(E(\mathbf{k}) - E_{\mathrm{C}})(1 + \alpha(E(\mathbf{k}) - E_{\mathrm{C}})) = \frac{\hbar^2}{2}\left(\frac{k_x^2}{m_x^*} + \frac{k_y^2}{m_y^*} + \frac{k_z^2}{m_z^*}\right) \tag{3.14}$$

However, to simplify the problem, we neglect the non-parabolicity of the valleys in the z-direction, i.e. we still use the hamiltonian in Eq. 3.5. Therefore, it is still possible to separate in-plane ($E_{\parallel}$) and quantization ($E_{\perp}$) energies, since $E_{\perp}$ does not depend on $\mathbf{k}$ if a parabolic dispersion relationship is assumed in the quantization direction [78, 102, 103]. Thus, we can express the non-parabolic two-dimensional Hamiltonian as:

$$(E(\mathbf{k}) - E_{\perp x,v,m})(1 + \alpha(E(\mathbf{k}) - E_{\perp x,v,m})) = \frac{\hbar^2}{2}\left(\frac{k_x^2}{m_x^*} + \frac{k_y^2}{m_y^*}\right) \tag{3.15}$$

that is a generalization of Eq. 3.11 and where α is the non-parabolicity factor. Eq. 3.15 better reproduces the two-dimensional energy dispersion relationship $E(\mathbf{k})$

when the energy of the carriers increases due to the heating of the horizontal electric field $F_{v,m}(x)$, and they can not be considered close to the bottom of the valley (in energy). For Si, α=0.5 eV^{-1} is typically used [99].

The approximation of neglecting the non-parabolicity is not too strong in Si, where the most populated valleys are the Δ_2. This approximation could be too strong in Ge, where the most populated valleys are the Λ_4. Indeed, when simulating Ge in Chap. 6, we will calibrate the quantization masses on the results obtained with the Linear Combination of Bulk Bands (LCBB) quantization model [104].

3.1.5 The semi-analytical model for the p-MOSFET

The expression of the energy dispersion relationship for the holes in a crystal is much more complicated than those expressed in Eq. 3.10 and 3.15. Thus, the effective mass approximation briefly explained in Sec. 3.1.4 can not be applied for the p-MOS case.

Thus, we have developed an analytic expression that is then calibrated on the $k{\cdot}p$ results [105, 78]. The analytic expression that our group developed to describe the energy dispersion relationship is:

$$\frac{\hbar^2 k_{v,d}^2}{2m_0 E_{\mathrm{V}}} = \left(\frac{1}{a_{v,d} + b_{v,d} E_{\mathrm{V}}} + c_{v,d} \right)^{-1} \tag{3.16}$$

where E_{V} is the energy referred to the valley minima, $k_{v,d}$ is the magnitude of the wave-vector in the direction d. $a_{v,d}$, $b_{v,d}$ and $c_{v,d}$ are the parameters that have to be fitted along the direction d, for each group of valleys v.

In order to derive a model valid for any $\mathbf{k}$ direction, our group has devised an appropriate function of the angle θ that is able to connect the values along the aforementioned directions d. For example, for the 001-oriented crystal, the equi-energy curves have periodicity $\pi/4$ [105].

Thus, we can fit the equi-energy curve between θ=0 and θ=$\pi/4$ with:

$$k_{\mathrm{V}}(E_{\mathrm{V}}, \theta) = A + B\cos(4\theta) + \cos(8\theta) \tag{3.17}$$

where θ is the angle between the direction $d(\theta)$ considered and direction $d(\theta{=}0)$ used to calibrate the analytical model of Eq. 3.16 on the $k{\cdot}p$ results.

Eq. 3.17 applies to (001) orientation of the crystal. However, the results can be extended to other crystal orientations and to strain materials [105].

The advantage of this model with respect to the $k{\cdot}p$ is that we can separate the in-plane ($E_{\parallel}$) and quantization ($E_{\perp}$) energies [78]. Indeed, the $E_{\perp x,v,m}$ is still found with Eqs. 3.4 and 3.5, once m_z has been calibrated on the $k{\cdot}p$ results obtained for $\mathbf{k}$=0 [103].

3.2 Scattering mechanisms for conventional MOSFET devices

In this Section we briefly review how the main scattering mechanisms included in the Multi Subband Monte Carlo simulator [78], before the starting of the activities of this thesis.

Eq. 3.9 shows that to model a scattering mechanism, we need its scattering rate $S(\mathbf{r}, \mathbf{k}', \mathbf{k})$. The calculation of the scattering rate is based on the Fermi's Golden Rule. The scattering rate from a state defined by $\mathbf{k}$ to a state defined by $\mathbf{k}'$ is:

$$\begin{aligned} S_{x,v,m,v',m'}(\mathbf{k},\mathbf{k}') &= \frac{2\pi\nu(E_\mathrm{a})}{\hbar}\left|M_{v,m,v',m'}(\mathbf{k}',\mathbf{k})\right|^2 \\ &\quad \delta(E_{x,v',m'}(\mathbf{k}') - E_{x,v,m}(\mathbf{k}) - E_\mathrm{a}) + \\ &\quad \frac{2\pi(1+\nu(E_\mathrm{e}))}{\hbar}\left|M_{v,m,v',m'}(\mathbf{k}',\mathbf{k})\right|^2 \\ &\quad \delta(E_{x,v',m'}(\mathbf{k}') - E_{x,v,m}(\mathbf{k}) + E_\mathrm{e}) \end{aligned} \tag{3.18}$$

where $\mathbf{k}$ is the electron wave vector before collision, $\mathbf{k}'$ is the wave vector after collision. $\mathbf{E}_{x,v,m}(\mathbf{k})$ is the energy of the electron (in the x-section, in the subband m of the valley v) before the collision, while $E_{x,v',m'}(\mathbf{k}')$ is the energy after the collision. Moreover, in Eq. 3.18, $E_{\mathrm{a/e}}$ is the absorbed/emitted energy during the scattering event, and $\nu(E_{\mathrm{a/e}})$ are the occupation numbers of the absorbed/emitted transitions. $M_{v,m,v',m'}(\mathbf{k}',\mathbf{k})$ are called *matrix elements*. Eq. 3.18 assumes that the transitions occur at thermal equilibrium. $S_{x,v,m,v',m'}(\mathbf{k},\mathbf{k}')$ has the units of an inverse of the time, so the scattering rate is a number that indicates how many collisions will occur on average per unit time.

In both the absorbing/emitting terms, the δ-function enforces the conservation of energy. If the scattering event is elastic, i.e. there is no exchange of energy during the transition, we have $E_\mathrm{a}{=}E_\mathrm{e}{=}0$.

So, it remains to define the matrix element, which, for elastic transitions, is:

$$M_{v,m,v',m'}(\mathbf{k}',\mathbf{k}) = \int e\psi^*_{\mathbf{k}'}(\mathbf{r},z)\phi_\mathrm{S}(\mathbf{r},z)\psi_{\mathbf{k}}(\mathbf{r},z)d^2r dz \tag{3.19}$$

where z is the vertical coordinate and $\mathbf{r}$ is the coordinate in the plane normal to the z direction. $\psi_{\mathbf{k}}(\mathbf{r},z)$ and $\psi_{\mathbf{k}}(\mathbf{r},z)$ are the wave functions before and after the collision, respectively. In the case of an inelastic transition, the integral in Eq. 3.19 has a further term as $e^{+i\omega t}$ for a transition which absorbes energy, or $e^{-i\omega t}$ if the energy is emitted.

By using Eq. 3.2, we can write:

$$M_{v,m,v',m'}(\mathbf{k}',\mathbf{k}) = \frac{e}{A}\int \xi_{\mathbf{k}'}(z)\xi_{\mathbf{k}}(z)\phi_\mathrm{S}(\mathbf{r},z)e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{r}}d^2r dz \tag{3.20}$$

in order to obtain:

$$M_{v,m,v',m'}(\mathbf{q}) = \frac{e}{A}\int \xi_{\mathbf{k}'}(z)\phi_\mathrm{S}(\mathbf{q},z)\xi_{\mathbf{k}}(z)dz \tag{3.21}$$

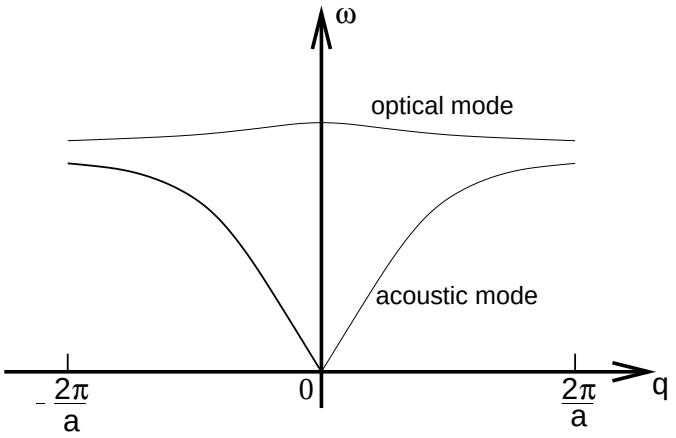

Figure 3.3: Comparison between the acoustic and the optical modes of the energy dispersion relationship of phonons.

where $\phi_{\mathrm{S}}(\mathbf{q}, z)$ is defined as the two-dimensional Fourier transform of the scattering potential $\phi_{\mathrm{S}}(\mathbf{q}, z)$ as:

$$\phi_{\mathrm{S}}(\mathbf{q}, z) = \int \phi_{\mathrm{S}}(\mathbf{r}, z) e^{i(\mathbf{k}-\mathbf{k}')\cdot\mathbf{r}} d^2 r \tag{3.22}$$

Thus, we can reconduce the evaluation of the scattering mechanisms to the Fourier transform of the scattering potential $\phi_{\mathrm{S}}(\mathbf{q}, z)$. The determination of the matrix elements for the phonons, surface roughness and ionized impurities scattering mechanisms are shown in in Sec. 3.2.1), Sec. 3.2.3 and Chap. 5, respectively.

Finally, it is worth noting that, actually, the scattering rates $\phi_{\mathrm{S}}(\mathbf{q}, z)$ are not calculated in the real device, but in a MOSFET featuring an infinitely long channel and without polarizing the drain contact (V_{DS}=0), i.e. we use Eq. 3.2 for describing the carriers, which assumes plane waves in the plane normal to the quantization direction.

3.2.1 Phonon scattering

The phonons of the bulk channel are the thermal vibrations of the crystal in the channel that interact with the free carriers [77, 78]. Indeed, when the atoms of the crystal lattice move from their nominal position, produce a perturbation potential that can result in carrier scattering events.

Since these particles are bosons, their statitisc obey to the Bose-Einstein statistics:

$$\nu(\omega_q) = \frac{1}{e^{\frac{\hbar\omega_q}{K_{\mathrm{B}}T_{\mathrm{L}}}} - 1} \tag{3.23}$$

where ω is the phonon energy, K_{B} is the Boltzmann constant and T_{L} is the lattice temperature.

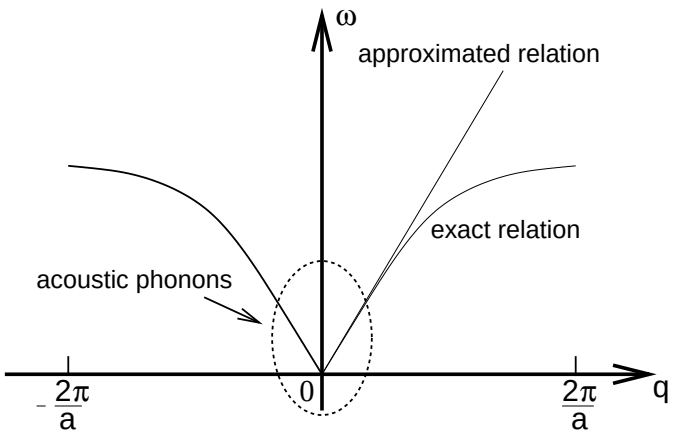

Figure 3.4: Comparison between the exact and the approximated (with the Debye model) acoustic mode of the phonon energy.

Fig. 3.3 shows a typical phonon dispersion relationship, and highlights the difference between the acoustic and optical branches. In the crystals in which we are interested in (Si and Ge), there are six branches: three of them are acoustic and the others are optical.

Acoustic phonons

We first consider acoustic phonons with small exchanged wave-vectors q. In this case we can adopt the Deby model and approximate the energy dispersion relationship of the acoustic phonons as:

$$\omega(q) \simeq v_{\mathrm{s}} q \tag{3.24}$$

where v_{s} is the sound velocity. Fig. 3.4 shows the acoustic branch close to the origin of the axis and its approximation with the Debye model.

The ω values are usually small when compared to $K_{\mathrm{B}}T$ and to the carrier energy, and this simplifies the calculation and allows to treat this scattering as elastic.

Under these assumptions, the squared matrix element of the intra-valley transitions due to acoustic phonons takes the form [98]:

$$|M_{v,m,m'}(\mathbf{q})|^2 = D_{\mathrm{ac}}^2 \frac{K_{\mathrm{B}} T_{\mathrm{L}}}{2 A v_{\mathrm{L}}^2 \rho} F_{v,m,m'} \tag{3.25}$$

where D_{ac} is the acoustic deformation potential, v_{L} is the sound velocity in the crystal, ρ is the semiconductor mass density and the term $F_{v,m,m'}$ is the so-called *form factor* and has the expression:

$$F_{v,m,m'} = \int_z |\xi_{v,m'}(z)\xi_{v,m}(z)|^2 dz \tag{3.26}$$

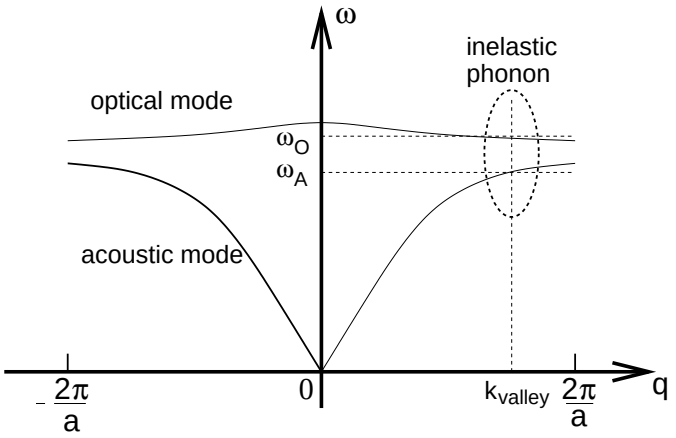

Figure 3.5: Inelastic inter-valley transitions originating from optical and acoustic phonons.

When considering acoustic phonons with large exchanged wave-vectors q, we adopt the Einstein model. This allows us to treat acoustic phonons with large q in the same way as we treat optical phonons. In the next subsection we briefly describe how to take into account the inter-valley transitions due to the optical phonons.

Optical phonons

The energy dispersion relationship for both the optical mode and, for large q values, for the acoustic mode can be approximated as:

$$\omega = \omega_0 \tag{3.27}$$

where we have assumed that the exchanged wave vector q is essentially given by the distance between the valleys in the k-space (k_{valley} in Fig. 3.5) and not by the position in the initial and final valleys. The exchanged energy $E_{\mathrm{a/e}}$ of Eq. 3.18 is:

$$E_{\mathrm{a/e}} = \hbar\omega_0 \tag{3.28}$$

This type of transitions are also called inter-valley since we suppose that they induce a change in the valley of the carrier. This approximation is not too strong for the materials we consider in this work, i.e Si and Ge.

In detail, in Si, the g-type phonons can scatter electrons from one valley to the other valley with the same orientation that is on the opposite side with respect to the origin (see Fig. 3.6). The f-type phonons scatter electrons from one valley to one of the other four valleys that have normal orientation (see Fig. 3.6).

The expression for the squared matrix element of the inter-valley transitions due to these phonons is [98]:

$$|M_{v,m,v',m'}(\mathbf{q})|^2 = D_{\mathrm{op}}^2 \frac{\hbar g^{v'}}{2\rho\omega A} F_{v,m,v',m'} \tag{3.29}$$

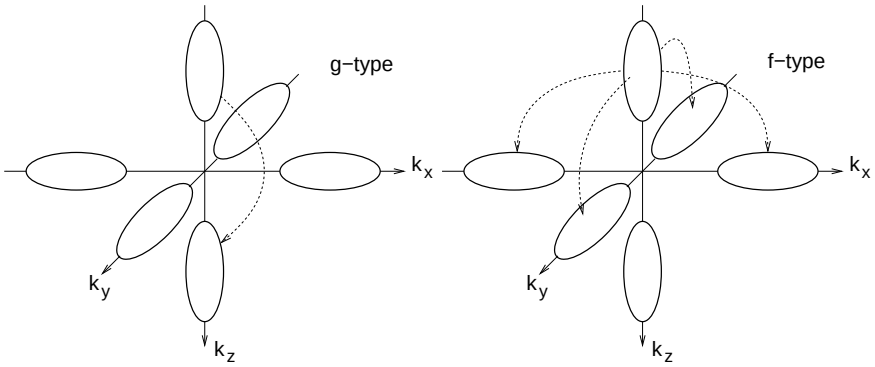

Figure 3.6: Sketch of the possible intervalley phonon transitions.

where D_{op} is the deformation potential and $g_{v'}$ is the number of available final valleys: $g^{v'}$ is 1 and 4 for g-type and for f-type transitions, respectively. The form factors $F_{v,m,v',m'}$ have been defined in Eq. 3.26.

3.2.2 Ionized impurities scattering

The ionized impurities scattering mechanism includes the scattering events due to the ionized dopants in the channel and to the charges at the channel/dielectric interface [106].

Here, we do not report any detail on the model for the ionized impurities, since it is equivalent to the one that will be presented in Chap. 5. In Sec. 5.3, we will derive the scattering potential produced by a point charge in the gate stack. The same derivation allows to find the scattering potential produced by a point charge in the channel or at the Si/ITL interface.

While the scattering mechanism does not need a calibration (the number of the Coulomb centers is the doping value), the value of the fixed charge density at the channel/dielectric interface needs to be calibrated on the experimental data.

3.2.3 Surface roughness

The treatment of the surface roughness scattering mechanism [107] is quite different with comparison to the previous models presented in Secs. 3.2.1 and 3.2.2. The operator representing the perturbation due this mechanism is:

$$-e\phi_{\mathrm{S}}(\mathbf{r}, z) = H_{\mathrm{PERT}}(\mathbf{r}, z) - H_{\mathrm{UNPERT}}(z) \tag{3.30}$$

where $H_{\mathrm{UNPERT}}(z)$ is the Hamiltonian of the unperturbed system, and the $H_{\mathrm{PERT}}(\mathbf{r}, z)$ is the Hamiltonian of the perturbed system. The unperturbed system is the one which has an ideal interface between the gate oxide and the channel, and thus its Hamiltonian is:

$$H_{\mathrm{UNPERT}}(z) = -\frac{\hbar^2}{2}\frac{\partial}{\partial z}\frac{1}{m_z}\frac{\partial}{\partial z} - e\phi + \phi_{\mathrm{B}}\theta(-z) \tag{3.31}$$

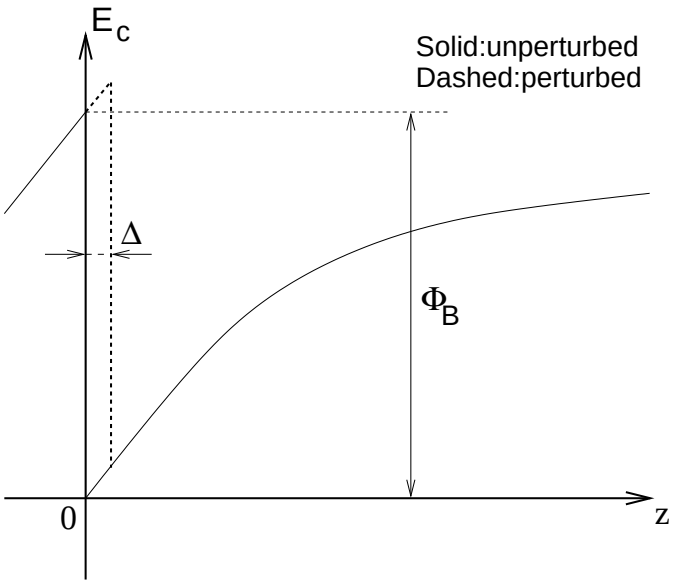

Figure 3.7: Potential well in the channel of an nMOSFET. The surface roughness is due to the non-ideality of the interface between the gate dielectric and the MOS channel. The roughness is managed by means of the displacement of the interface position from its mean value, called $\Delta(\mathbf{r})$.

where ϕ_{B} is the potential barrier height, $\theta(-z)$ is the shape of the potential barrier (step function for simplicity) and ϕ is the electrostatic potential. The perturbed system is the one which has a shifted interface between the gate oxide and the channel, and thus, under the EMA approximation (Sec. 3.1.4), its Hamiltonian is:

$$H_{\mathrm{PERT}}(\mathbf{r},z) = -\frac{\hbar^2}{2}\frac{\partial}{\partial z}\frac{1}{m_z(z-\Delta(\mathbf{r}))}\frac{\partial}{\partial z} - e\phi + \phi_{\mathrm{B}}\theta(-z+\Delta(\mathbf{r})) \tag{3.32}$$

where $\Delta(\mathbf{r})$ is the displacement of the interface at the position $\mathbf{r}$ from its average value in z-direction, as we can see in Figure 3.7. If we define:

$$z' = z - \Delta \tag{3.33}$$

and we assume an arbitrarily small $\Delta(\mathbf{r})$:

$$H_{\mathrm{PERT}}(\mathbf{r},z) \simeq H_{\mathrm{UNPERT}}(z') + \Delta(\mathbf{r})\frac{de\phi}{dz'} \tag{3.34}$$

at the same time, the wave-function of the unperturbed system can be written as:

$$\xi_{v,m}(z) \simeq \xi_{v,m}(z') + \Delta(\mathbf{r})\frac{d\xi_{v,m}(z')}{dz'} \tag{3.35}$$

It is possible to express the unscreened matrix element as:

$$M_{v,m,v',m'} = \frac{1}{A}\int_z \xi_{v',m'}\frac{de\phi}{dz}\xi_{v,m}dz + (E_{v',m'} - E_{v,m})\frac{1}{A}\int_z \xi_{v',m'}\frac{d\xi_{v,m}}{dz}dz \tag{3.36}$$

This formulation of the matrix element allows the wave function to be not null in the gate dielectric ($z<0$). This is the formulation that we will use in App. A when performing simulations accounting for the penetration of the wave functions describing the carriers.

If we impose that the wave functions can not penetrate into the gate dielectric, i.e.:

$$\begin{cases} \psi_{v,m}(0) = 0 \\ \psi_{v,m}(\infty) = 0 \end{cases} \tag{3.37}$$

we obtain the relation [78]:

$$\int_0^\infty \xi_{v,m} \frac{\partial e\phi}{\partial z} \xi_{v',m'} dz + (E_{v,m} - E_{v',m'}) \int_0^\infty \xi_{v,m} \frac{\partial \xi_{v',m'}}{\partial z} dz = \frac{\hbar^2}{2m_z} \left[\frac{\partial \xi_{v,m}}{\partial z} \frac{\partial \xi_{v',m'}}{\partial z} \right]_0 \tag{3.38}$$

This relation allows to easily compute Eq. 3.36 in order to find the final expression for the matrix element:

$$M_{v,m,v',m'} = \frac{\hbar^2}{2m_z} \left[\frac{\partial \xi_{v,m}}{\partial z} \frac{\partial \xi_{v',m'}}{\partial z} \right]_0 \int_{\mathbf{r}} \frac{1}{A} \Delta(\mathbf{r}) e^{i\mathbf{q}\cdot\mathbf{r}} d\mathbf{r} \tag{3.39}$$

Finally, when expressing the squared $M_{j,i}$, we obtain:

$$|M_{v,m,v',m'}|^2 = \frac{\hbar^4}{4m_z^2 A} \left[\frac{\partial \xi_{v,m}}{\partial z} \frac{\partial \xi_{v',m'}}{\partial z} \right]_0^2 S_{\mathrm{SR}}(q) \tag{3.40}$$

where $S_{\mathrm{SR}}(q)$ is the spectrum of the surface roughness, and is defined as:

$$S_{\mathrm{SR}}(q) = \frac{1}{A} \left| \int_{\mathbf{r}} \Delta(\mathbf{r}) e^{i\mathbf{q}\cdot\mathbf{r}} d\mathbf{r} \right|^2 \tag{3.41}$$

The form of the matrix elements given in Eq.3.40 is deliberately simple. In particular, we have neglected the terms in the scattering potential stemming from the change of the electron density produced by the variation of the interface position, although such terms have been shown to be quantitatively significant [108, 109]. Such a simplification is justified for the purpose of this work, because it does not alter the results found in this manuscript.

3.2.4 The screening effect

Actually, the perturbation potential felt by the carriers in the inversion layer is not directly the one shown in the previous sections. Indeed, when an external potential is applied to the channel, the inversion layer reacts in order to reduce the *effective* perturbation potential felt by the carriers. Indeed, this *effective* potential can be seen as composed of two different contributions: the perturbation potential of the scattering mechanism, and the potential produced by the inversion layer [108], that tends to decrease the effect of the perturbation potential of the scattering mechanism.

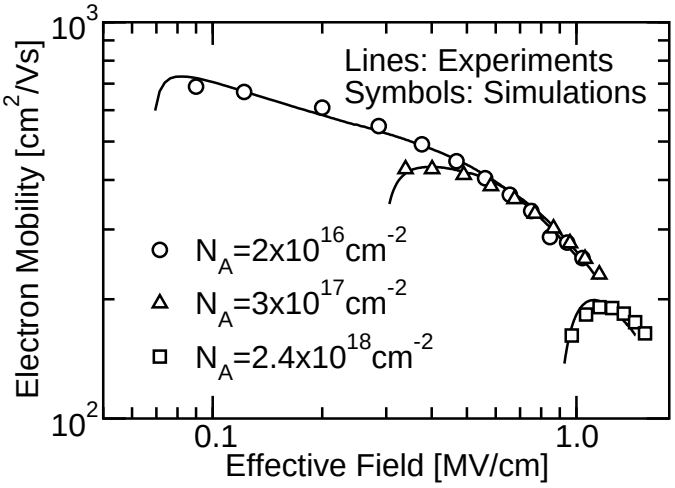

Figure 3.8: Calibration of the n-MOS MSMC simulator. Comparison between the universal mobility curves of [47] and the simulated ones [93] for different values of the channel doping, as a function of the effective field. Temperature is 300 K.

The screening effect models the re-distribution of the inversion layer in the presence of an external perturbation potential. While the ionized impurities and the surface roughness scattering mechanisms are strongly affected by the screening, we can assume that the phonons are not screened. Indeed, for inter-valley phonon assisted transitions in electron inversion layers, the very large q values make the screening very ineffective [98]. For intra-valley phonons, the situation is much more complicated. However, it is common believe that in an inversion layer also the screening for intra-valley transitions become ineffective [78].

The derivation for the various models for screening are reported in detail in Chap. 4. The model has no adjustable parameters.

3.3 Calibration of the simulator

The carrier mobility is the coefficient of proportionality between the drift velocity of the carriers v_{drift} and the driving field E_{lateral} in the x-direction (Eq. 1.2).

We choose to calibrate both the n- and p-MOS MSMC simulators with the so called universal mobility curves measured in [47]. Figs. 3.8 and 3.9 shows the comparison between the experimental curves of [47] and the simulated mobility for both the n-MOS [93] and p-MOS cases [100], as a function of the effective vertical electric field E_{eff}. The scattering mechanisms accounted for in Figs.3.8 and 3.9 are ionized impurities, acoustic and optical phonons and surface roughness.

We can see that our model correctly approaches the curves measured in [47] if we choose the simulation parameters shown in Tab. 3.1.

Concerning the phonons, the energies of the inter-valley transitions have been

Scattering mechanism	Parameter	n-MOS	p-MOS
Acoutic phonons	D_{ac} [eV]	13	5.2
Intervalley phonons	type	f	-
	ω [K]	220	710
	D_{op} [eV/m]	0.3×10^{10}	1.15×10^{11}
Intervalley phonons	type	f	
	ω [K]	550	-
	D_{op} [eV/m]	2.0×10^{10}	
Intervalley phonons	type	f	
	ω [K]	685	-
	D_{op} [eV/m]	2.0×10^{10}	
Intervalley phonons	type	g	
	ω [K]	140	-
	D_{op} [eV/m]	0.5×10^{10}	
Intervalley phonons	type	g	
	ω [K]	215	-
	D_{op} [eV/m]	0.8×10^{10}	
Intervalley phonons	type	g	
	ω [K]	720	-
	D_{op} [eV/m]	1.1×10^{10}	
Ionized Impurities	$N_{\mathrm{Si/SiO_2}}$ [cm^{-2}]	2×10^{10}	2×10^{10}
Surface Roughness	Spectrum	Gauss.	Exp.
	Δ [nm]	0.62	0.56
	Λ [nm]	1.00	2.6

Table 3.1: Parameters used to reproduce the universal mobility curves of [47] for both the n-MOS and p-MOS MSMC in the simulators.

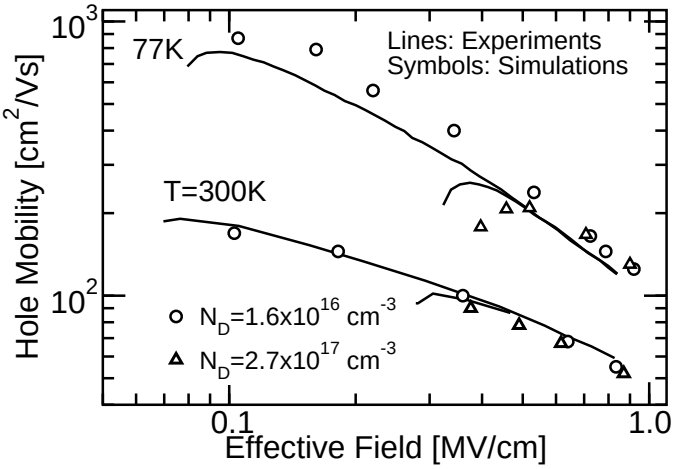

Figure 3.9: Calibration of the p-MOS MSMC simulator. Comparison between the universal mobility curves of [47] and the simulated ones [100] for different values of the channel doping, as a function of the effective field. The simulations for T=300 K and T=77 K are shown.

taken from [98] and from [110] for the n-MOS and for the p-MOS, respectively. In Tab. 3.1 the energy of the phonon are expressed in Kelvin. The relation between the energy of the phonon and its temperature is:

$$E_{\mathrm{PH}}[eV] = \frac{K_{\mathrm{B}}}{e} T_{\mathrm{PH}} \tag{3.42}$$

D_{op} has been kept as in [98, 110], whereas D_{ac} has been adjusted to reproduce the curves in [47].

Regarding the Ionized Impurities, as already said in Sec. 3.2.2, we can adjust the concentration of charges at the channel/gate oxide interface $N_{\mathrm{Si/SiO_2}}$, while the effect of the charged atoms in the channel is fixed by the doping concentration.

Finally, the calibration of the surface roughness reduces to the calibration of the parameters of the power spectral density of the roughness of the interface between the channel and the gate oxide (Eq. 3.41). For the n-MOS case, we chose a Gaussian spectrum:

$$S_{\mathrm{SR}}(q) = \pi \Lambda^2 \Delta^2 e^{\frac{\Lambda^2 q^2}{4}} \tag{3.43}$$

while for the p-MOS case we chose an exponential expression for the roughness spectrum:

$$S_{\mathrm{SR}}(q) = \frac{\pi \Lambda^2 \Delta^2}{(1 + q^2 \Lambda^2 / 2)^{3/2}} \tag{3.44}$$

The Λ and Δ in Eqs. 3.43 and 3.44 are the correlation length and the root mean squared values of the surface roughness.

As it can be seen in Tab. 3.1, the values of the spectral densities of the surface roughness of the Si/SiO_2 interface are different for n- and p-MOSFETs. This is still an unresolved point of the MOSFET modeling [111].

Chapter 4

Screening in multi-gate structures

Abstract

This chapter shows that modelling of the screening due to the inversion layer based on the widely employed scalar dielectric function fails in double-gate MOS transistors and in FinFETs. This leads to simulation results inconsistent with the experiments, especially at high channel inversion densities where the mobility is limited by surface roughness scattering. The use of the full tensorial dielectric function approach, instead, reconciles simulations with the mobility experiments.

In the first part of the chapter we compare the formulations of the scalar and the tensorial dielectric function approaches.

Then, we identify, using Multi Subband Monte Carlo simulations as well as analytical derivations for the screened matrix elements of the surface roughness scattering, what are the simplifying assumptions in the derivation of the scalar dielectric function that do not hold in a double-gate structure.

4.1 Introduction

A seen in Chap. 1, the thin body SOI MOSFETs and the FinFETs are very promising devices for nano-scale CMOS technologies because the double-gate (DG) operation of the transistors results in an excellent electrostatic integrity and in good potentials for high ON-currents [60, 61, 63, 65, 64]. Thus, not surprisingly, much attention has been recently devoted to low field mobility in DG transistors both on the experimental [112, 113, 114, 115, 116, 117, 118] and on the modelling side [119, 120, 121, 109].

A theoretically and computationally challenging issue in the physics-based modelling of mobility in MOS transistors is the screening produced by the carriers in the inversion layer [98, 99, 102, 106]. An accurate treatment of the screening is also very important, because the screening affects some features of the mobility curves versus inversion density $N_{\rm inv}$ that can be directly compared to experiments. In fact the screening of the Coulomb scattering centers is responsible for the roll-off of the mobility at small $N_{\rm inv}$ values [98, 102, 106]. Furthermore, at large inversion densities, where the mobility is essentially limited by the surface roughness scattering, the screening contributes to the slope by which the effective mobility decreases with $N_{\rm inv}$ [107].

The screening in a electron 2D gas can be described by resorting to the concept of the dielectric function, which is in general a six order tensor that governs the linear relation between the screened and the unscreened matrix elements [98, 99, 102, 106] (see also Sec.4.2 below). Such a general formulation based on the tensorial dielectric function (TDF) is computationally very demanding, so that a simplified procedure leading to a scalar dielectric function (SDF) has been derived [102], and frequently used in the literature for conventional bulk MOSFETs [122, 123, 124, 125, 126, 106], for single-gate (SG) and DG SOI transistors as well as for FinFETs [121, 127, 128, 129].

In this chapter we show that the SDF model should not be used in DG transistors and in FinFETs, because it leads to artifacts and to results that are clearly inconsistent with the experiments. This can be readily observed in Figs. 4.1 and 4.2. In fact Fig. 4.1 shows that the experimental mobility for thick SOI MOSFETs operated either in SG or in DG mode essentially coincides with the mobility of lightly doped bulk transistors [115] for large $N_{\rm inv}$ values. The simulations in Fig. 4.1 obtained with the TDF model reproduce nicely this behavior. If the SDF is used, instead, Fig.4.2 shows that the simulated mobility in DG mode is significantly larger than in SG or in bulk MOSFETs. We verified that this result is an artifact due to the failure of the SDF model. In fact Fig.4.2 demonstrates that the mobility curves for the different devices are newly in close agreement if we neglect the screening in the mobility calculation.

This chapter is focussed on the mobility of n-type MOSFETs and it is organized as follows. In Sec. 4.2 we introduce the concepts and basic equations behind the dielectric function model and clarify the differences between the tensorial and the scalar formulation. In Sec. 4.3 we discuss the application of the dielectric

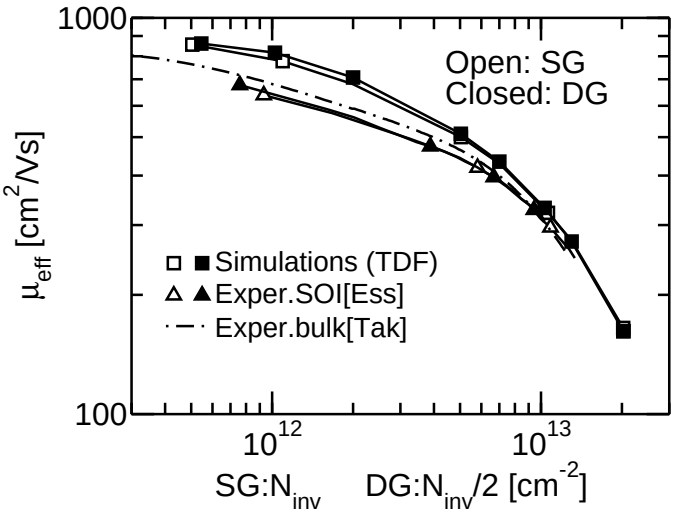

Figure 4.1: Comparison between experimental [115] and simulated mobility for 20 nm thick SG-SOI and DG-SOI MOSFETs obtained by using the TDF model. The universal mobility curve is also shown for reference (dot-dashed line) [47]. The simulations have been obtained accounting for optical and acoustic phonons and for surface roughness scattering with the Multi Subband Monte Carlo approach as in Tab. 3.1. Ionized impurity scattering mechanism has been neglected, leading to a slight disagreement between simulated and experimental mobility at low inversion density.

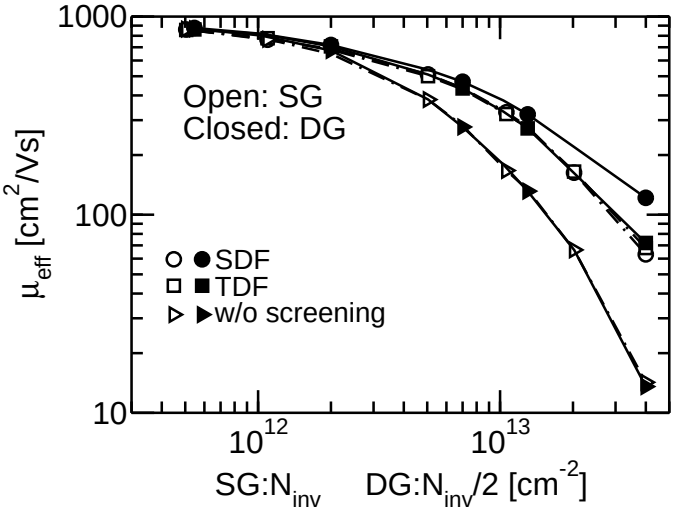

Figure 4.2: Mobility simulations for the same SOI MOSFET as in the Fig. 4.1 but obtained either with the SDF or by neglecting the screening (triangles). The electron mobility has been shown as a function of the $N_{\rm inv}$ or $N_{\rm inv}/2$ for the SG-SOI or DG-SOI, respectively.

function model to the surface roughness scattering and explain how some of the simplifications which lead to the SDF clearly fail in DG structure. In Sec. 4.4 we discuss analytically the artifacts in the DG mobility simulations stemming from the failure of the SDF model. Sec. 4.5 discusses the impact of the silicon thickness t_{Si} on the validity of the SDF model.

4.2 Dielectric function for the screening in MOSFETs

Let us consider an electron inversion layer described according to the effective mass approximation, where $\xi_{v,m}(z)$, as in Chap. 3, indicates the envelope wave-function of the subband m belonging to the valley v and $E_{v,m}(\mathbf{k})$ is the total energy in the subband for the in-plane wave-vector $\mathbf{k}$. For a static scattering potential, such as the Coulomb or the surface roughness scattering, we denote by $M_{v,m,m'}(\mathbf{q})$ the scattering matrix element between the subbands m and m' of the valley v, where $\mathbf{q}$ is the wave-vector variation produced by the scattering. As implied by the symbol $M_{v,m,m'}(\mathbf{q})$, the analysis will be restricted to intra-valley transitions, because for Coulomb and surface roughness scattering (see also Chap. 3) the much larger wave-vector variation necessary for inter-valley transitions reduces drastically the corresponding matrix elements [98, 102, 106]. The dielectric function is a general formalism suitable to account for the screening produced by the electrons in the inversion layer and to determine the screened matrix elements $M^{\mathrm{scr}}_{v,m,m'}(\mathbf{q})$ as a linear combination of the unscreened ones $M_{v,m,m'}(\mathbf{q})$ [102].

Before discussing the formulation of the SDF and TDF models, we notice that the dielectric function depends on the exchanged wave-vector $\mathbf{q}$ and not only on its magnitude q. For the elliptical electron energy relation, however, the dependence of the dielectric function on the direction of $\mathbf{q}$ is quite modest (also because the most populated unprimed 2-fold valleys in (100) silicon are circular), hence we will neglect it in the following of the Chapter. Furthermore, according to the model presented in Sec.4.3, the unscreened matrix elements for the surface roughness scattering depend only on q, the magnitude q of exchanged wave-vector $\mathbf{q}$, so that in the following both the scattering matrix elements and the dielectric function will be indicated as function of q only.

4.2.1 Tensorial dielectric function approach

The central equation of the TDF approach is [102]:

$$M_{v,m,m'}(q) = \sum_{w,n,n'} \epsilon^{w,n,n'}_{v,m,m'}(q) M^{\mathrm{scr}}_{w,n,n'}(q) \tag{4.1}$$

that is a linear algebraic system that must be solved to determine the screened matrix elements $M^{\mathrm{scr}}_{w,n,n'}(q)$ as a function of the unscreened ones $M_{v,m,m'}(q)$. The screening function is defined as [102]:

$$\epsilon^{w,n,n'}_{v,m,m'}(q) = \delta_{w,v}\delta_{n,m}\delta_{n',m'} + \alpha(q)\Pi_{w,n,n'}(q)F^{w,n,n'}_{v,m,m'}(q) \tag{4.2}$$

where $F^{w,n,n'}_{v,m,m'}$ and $\Pi_{w,n,n'}$ denote respectively the form factors and the polarization factors and $\alpha(q)$ is:

$$\alpha(q) = \frac{e^2}{q(\epsilon_{\rm Si} + \epsilon_{\rm OX})} \tag{4.3}$$

The expression of the polarization factor $\Pi_{w,n,n'}$ is [99, 102, 106]:

$$\Pi_{w,n,n'}(q) = \frac{1}{A}\sum_k \frac{f_0[E_{w,n'}(k+q)] - f_0[E_{w,n}(k)]}{E_{w,n'}(k+q) - E_{w,n}(k)} \tag{4.4}$$

where $f_0(E)$ is the Fermi-Dirac equilibrium occupation function. The term $\mathrm{F}^{w,n,n'}_{v,m,m'}(q)$ has been here defined as the dimensionless form factor:

$$F^{w,n,n'}_{v,m,m'}(q) = \int_0^{t_{\rm Si}} dz \int_0^{t_{\rm Si}} dz_0\, \xi_{v,m}(z)\, \xi^{\dagger}_{v,m'}(z) \xi^{\dagger}_{w,n}(z_0)\, \xi_{w,n'}(z_0) \phi_{\rm pcD}(q,z,z_0) \tag{4.5}$$

where $t_{\rm Si}$ is the thickness of the Si film and $\phi_{\rm pcD}(q,z,z_0)$ is the dimensionless potential produced by a point charge located at z_0:

$$\phi_{\rm pcD}(q,z,z_0) = \frac{\epsilon_{\rm Si} + \epsilon_{\rm OX}}{2\epsilon_{\rm Si}} \left[e^{-q|z-z_0|} + C_1 e^{qz} + C_2 e^{-qz} \right] \tag{4.6}$$

We remind that the expression in Eq. 4.6 is the two-dimensional Fourier transform that has been defined in 3.22. The two constants C_1 and C_2 depend on the structure of the device.

For the bulk and SOI cases we assume to have an infinitely thick gate dielectric. This assumption can be unreliable when simulating realistic devices with thin dielectrics. However, this is not the aim of this Chapter. Here, indeed, we compare different models to predict the effect on the mobility of the screening due to the carriers in the inversion layer in long channel devices.

In the bulk case, we also assume to have an infinitely deep substrate ($t_{\rm Si} \to \infty$) to obtain:

$$C_1 = 0 \qquad C_2 = \frac{\epsilon_{\rm Si} - \epsilon_{\rm OX}}{\epsilon_{\rm Si} + \epsilon_{\rm OX}} e^{-qz_0} \tag{4.7}$$

whereas in the SOI case we have:

$$\begin{cases} C_1 &= \dfrac{(\epsilon_{\rm Si} - \epsilon_{\rm OX})^2 e^{-qz_0} + (\epsilon^2_{\rm Si} - \epsilon^2_{\rm OX}) e^{qz_0}}{(\epsilon_{\rm Si} + \epsilon_{\rm OX})^2 e^{2qt_{\rm Si}} - (\epsilon_{\rm Si} - \epsilon_{\rm OX})^2} \\ C_2 &= \dfrac{\epsilon_{\rm Si} - \epsilon_{\rm OX}}{\epsilon_{\rm Si} + \epsilon_{\rm OX}} (e^{-qz_0} + C_1) \end{cases} \tag{4.8}$$

It is now worth noticing that the calculation of the tensorial dielectric function can be very CPU time expensive and that the computational burden is dominated by the calculation of the form factors defined in Eq.4.5. In fact, once the envelope wave-functions $\xi_{v,m}(z)$ have been determined by solving the self-consistent Schrödinger-Poisson problem in the inversion layer, then the double integral in Eq.4.5 must be evaluated numerically for a huge number of times. As an example,

for a (100) silicon inversion layer with 3 valleys, if we use 7 subbands per valley and we employ 50 discrete values for q, then the double integral must be calculated about 180000 times (even accounting for the fact that the tensor of the form factors has some elements which are equal).

One may argue that the calculation of the form factors can be completely parallelized if clusters with many CPU cores are available, however it is not surprising that a simplified and much less computationally expensive dielectric function model was developed [102], and used in many previous works [122, 123, 124, 125, 126, 106, 127, 128, 121, 129]. Such a scalar dielectric function approach is described in the next section.

4.2.2 Scalar dielectric function approach

The derivation of the SDF model is based on several approximations of the general formulation presented in the previous section. First of all this simplified formulation holds for very small q values, so that Eqs 4.6 to 4.8 allow us to write:

$$\begin{cases} \phi_{\mathrm{pcD}}(q \to 0) & \simeq 1 & \text{for bulk} \\ \phi_{\mathrm{pcD}}(q \to 0) & \simeq \dfrac{\epsilon_{\mathrm{Si}} + \epsilon_{\mathrm{OX}}}{2\epsilon_{\mathrm{OX}}} & \text{for SOI} \end{cases} \tag{4.9}$$

By substituting Eq. 4.9 in Eq. 4.5 and recalling the orthonormalization of the envelope wave-functions $\xi_{v,m}(z)$ belonging to the same valley v, we obtain that the inter-subband form factors tend to vanish for small q values:

$$F^{w,n,n'}_{v,m,m'}(q \to 0) \simeq 0 \qquad \text{when} \qquad (n \neq n') \ \text{ or } \ (m \neq m') \tag{4.10}$$

By inserting Eq. 4.10 in Eq. 4.2 we see that, according to the above simplifications, the inter-subband transitions are left unscreened, namely we have:

$$M^{\mathrm{scr}}_{v,m,m'}(q) = M_{v,m,m'}(q) \quad \text{for} \quad m' \neq m \tag{4.11}$$

Thus Eq. 4.1 can be restricted to intra-subband transitions as:

$$M_{v,m,m}(q) = \sum_{w,n} \epsilon^{w,n,n}_{v,m,m}(q)\, M^{\mathrm{scr}}_{w,n,n}(q) \tag{4.12}$$

We now recall Eqs. 4.9 and 4.5 and see that, for very small q values, we have $F^{w,n,n}_{v,m,m}(q{\to}0){\simeq}F^{w,n,n}_{w,n,n}(q{\to}0){\simeq}\phi_{\mathrm{pcD}}(q{\to}0)$. The condition $F^{w,n,n}_{v,m,m}{\simeq}F^{w,n,n}_{w,n,n}$ is necessary to derive a scalar dielectric function and we thus assume that its validity can be approximately extended also to non-zero q values. In other words we assume:

$$F^{w,n,n}_{v,m,m}(q) \simeq F^{w,n,n}_{w,n,n}(q) \qquad \forall\, w, v, m, n \tag{4.13}$$

which has been very widely used in literature [102, 123, 125, 126, 106, 121, 127, 128, 129]. By substituting Eq. 4.13 in the intra-subband dielectric function $\epsilon^{w,n,n}_{v,m,m}$

defined according to Eq. 4.2, one can realize that the $\epsilon^{w,n,n}_{v,m,m}$ takes the form of a matrix which can be inverted analytically, so that we obtain:

$$M^{\mathrm{scr}}_{v,m,m}(q) = \frac{M_{v,m,m}(q)}{\epsilon_{\mathrm{scal}}(q)} + \frac{\alpha(q)\sum_{(w,n)\neq(v,m)}\Pi_{w,n,n}(q)F^{w,n,n}_{w,n,n}(q)[M_{v,m,m}(q)-M_{w,n,n}(q)]}{\epsilon_{\mathrm{scal}}(q)} \tag{4.14}$$

where $\epsilon_{\mathrm{scal}}(q)$ is the scalar dielectric function:

$$\epsilon_{\mathrm{scal}}(q) = 1 + \alpha(q)\sum_{v,m}\Pi_{v,m,m}(q)F^{v,m,m}_{v,m,m}(q) \tag{4.15}$$

The SDF model further assumes that the unscreened intra-valley matrix elements for different subbands are quite similar, so that one can write:

$$|M_{v,m,m}(q) - M_{w,n,n}(q)| \ll |M_{v,m,m}(q)| \tag{4.16}$$

and Eq. 4.14 further simplifies to:

$$M^{\mathrm{scr}}_{v,m,m}(q) = \frac{M_{v,m,m}(q)}{\epsilon_{\mathrm{scal}}(q)} \tag{4.17}$$

In Eq. 4.17 each screened matrix element is obtained by simply dividing the unscreened one by the scalar dielectric function $\epsilon_{\mathrm{scal}}(q)$ defined in Eq. 4.15.

According to Eq.4.15 the form factors are arranged in a second rank tensor (namely a matrix) rather than in a six-order tensor as in Eq.4.5; hence the number of form factors to be calculated and the computational complexity is drastically reduced in the SDF with respect to the TDF model.

4.3 Validity of the SDF model for surface roughness scattering

In this section we discuss in detail the validity of the simplifying assumptions (i.e. Eqs. 4.10, 4.13 and 4.16) used in Sec. 4.2.2 to obtain the SDF model. The analysis is focussed on surface roughness scattering mechanism, which dominates the mobility at large inversion densities and is thus responsible for the features of the simulated mobility curves observed in Fig. 4.2.

To this purpose, we first discuss in Sec. 4.3.1 the expression for the unscreened matrix elements and the validity of Eq. 4.16. Then in Sec. 4.3.2 we discuss the application of the dielectric function approach to either SG-SOI or DG-SOI MOSFETs. Finally in Sec. 4.3.3 we address the behaviour of the form factors of the dielectric function to discuss the validity of Eqs. 4.10 and 4.13.

In order to obtain a fair comparison between bulk, SG-SOI and DG-SOI structures, we need a similar confining potential for all the examined structures. Thus

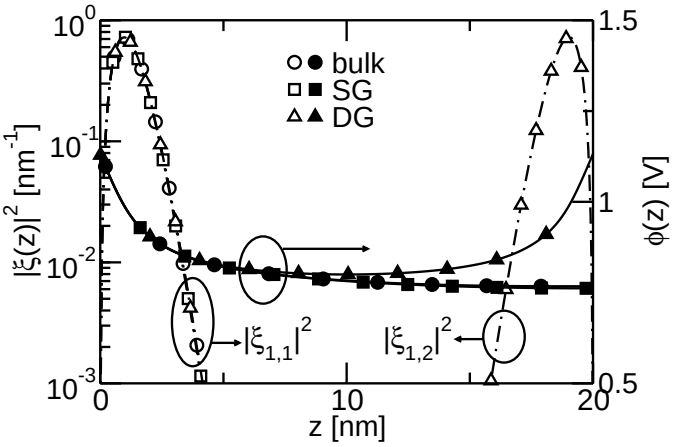

Figure 4.3: Comparison between the electrostatic potential $\phi(z)$ (closed symbols, right y-axis) and the wave-functions of the lowest subbands (open symbols, left y-axis) for bulk, SG-SOI and DG-SOI structures. The silicon film thickness is 20 nm for the SOI structures. The $N_{\rm inv}$ is 1×10^{13} cm^{-2} for bulk and SG-SOI and 2×10^{13} cm^{-2} for DG-SOI transistors.

we compare a bulk MOSFET with a fairly thick SOI structure ($t_{\rm Si}$=20 nm, as for the devices in Figs. 4.1 and 4.2). Fig. 4.3 shows the electrostatic potential $\phi(z)$ in bulk, SG-SOI and DG-SOI structures in strong inversion ($N_{\rm inv}$=1×10^{13} cm^{-2} for bulk and SG-SOI and 2×10^{13} cm^{-2} for DG-SOI). The same figure reports the squared magnitude $|\xi(z)|^2$ of the wave-function for the lowest subband of the unprimed valley and for the different device structures.

In the DG-SOI transistor the two lowest unprimed subbands, namely those labelled with (v,m)=$(1,1)$ and $(1,2)$, are pratically degenerate. Hence Fig. 4.3 reports both the corresponding wave-functions $|\xi_{1,1}(z)|^2$ and $|\xi_{1,2}(z)|^2$, where we have arbitrarily labelled with m=1 the subband whose $|\xi(z)|$ has the peak value close to the front interface (located at z=0). For the DG-SOI structure the potential profile is perfectly symmetric with respect to the center of the silicon film at $z_{\rm c}$=10 nm; thus the $|\xi_{1,2}(z)|^2$ is a mirror image of $|\xi_{1,1}(z)|^2$ with respect to $z_{\rm c}$.

We see that the potential profile and the envelope wave-functions of the SG-SOI are very close to their counterparts in the bulk MOSFET and those at the front interface of the DG-SOI MOSFET.

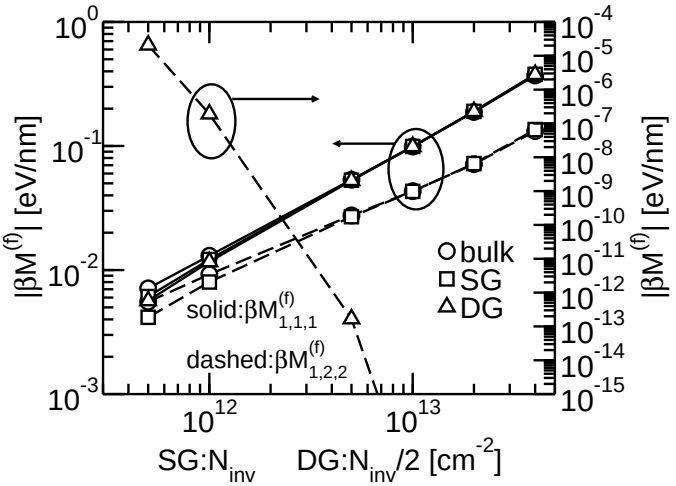

Figure 4.4: Normalized matrix elements $\beta M^{(\mathrm{f})}_{1,1,1}$ and $\beta M^{(\mathrm{f})}_{1,2,2}$ defined in Eq. 4.18 for the bulk, SG-SOI and DG-SOI structures versus the electron density. For the two SOI structures the silicon thickness is 20 nm. The normalization factor is $\beta{=}A/\Delta_{\mathrm{F}}(q)$.

4.3.1 Unscreened matrix elements

The unscreened matrix element $M^{(\mathrm{f})}_{v,m,m'}(q)$ and $M^{(\mathrm{b})}_{v,m,m'}(q)$, respectively at the front or at the back interface, can be expressed as [107, 122]:

$$M^{(\mathrm{f})}_{v,m,m'}(q) = \frac{\hbar^2}{2m_{z,v}A}\left[\frac{d\xi_{v,m}(y)}{dz}\frac{d\xi_{v,m'}(y)}{dz}\right]_{y=0}\Delta_{\mathrm{F}}(q) \qquad (4.18a)$$

$$M^{(\mathrm{b})}_{v,m,m'}(q) = \frac{\hbar^2}{2m_{z,v}A}\left[\frac{d\xi_{v,m}(y)}{dz}\frac{d\xi_{v,m'}(y)}{dz}\right]_{y=t_{\mathrm{Si}}}\Delta_{\mathrm{B}}(q) \qquad (4.18b)$$

which are a generalization of the expression seen in Chap. 3 and where m and m' are two subbands of the valley v and A is a normalization area. The terms $\Delta_{\mathrm{F}}(q)$ and $\Delta_{\mathrm{B}}(q)$ are the Fourier transform of the roughness at the front and back interfaces, respectively (Eq. 3.41). We assume that the random profiles of the interface roughness at the two interfaces are uncorrelated. We also suppose that the power spectrum of the roughness at the two interfaces is the same and denote it as $S_{\mathrm{SR}}(q)$. Hence the squared matrix elements of both interfaces are proportional to $S_{\mathrm{SR}}(q)$, as in Eq. 3.40.

Fig. 4.4 compares the unscreened intra-subband matrix elements $M^{(\mathrm{f})}_{1,1,1}$ and $M^{(\mathrm{f})}_{1,2,2}$ for the front interface roughness of the lowest and the second lowest subband of the unprimed valley; the calculations are illustrated for the different device structures of Fig. 4.3. The matrix elements in figure have been normalized to $\beta^{-1}{=}\Delta_{\mathrm{F}}(q)/A$ in order to obtain results independent on the exchanged wavevector q. It is easy to see that the Eq. 4.16 is not valid in the DG-SOI case because

the $M^{(\mathrm{f})}_{1,2,2}$ is orders of magnitude lower than the $M^{(\mathrm{f})}_{1,1,1}$, whereas it is fairly verified for the other device structures. This result for the DG-SOI is a direct consequence of the shapes of the corresponding $|\xi_{1,1}(z)|$ and $|\xi_{1,2}(z)|$ shown in Fig. 4.3.

4.3.2 Screening in bulk and SOI structures

In the case of bulk MOSFETs only one silicon-oxide interface exists and Eq. 4.18a can be used to calculate the unscreened matrix elements of such interface. Then the screened matrix elements are obtained from Eq. 4.1 or from Eq. 4.17 according to either the SDF or the TDF model, respectively.

In the SOI structures, instead, we have two interfaces. Since we assume that the surface roughness at the two interfaces are uncorrelated scattering mechanisms, we can apply the screening separately to the matrix elements of the two interfaces by using either Eq. 4.1 or Eq. 4.17 for the TDF or SDF model, respectively. Then the squared screened matrix elements $|M^{\mathrm{scr}}_{v,m,m'}|^2$ are obtained as:

$$|M^{\mathrm{scr}}_{v,m,m'}|^2 = |M^{(\mathrm{f,scr})}_{v,m,m'}|^2 + |M^{(\mathrm{b,scr})}_{v,m,m'}|^2 \tag{4.19}$$

where $|M^{(\mathrm{f,scr})}_{v,m,m'}|^2$ and $|M^{(\mathrm{b,scr})}_{v,m,m'}|^2$ are the screened matrix elements corresponding to the roughness at the two interfaces.

4.3.3 Form factors

Fig. 4.5 compares the form factors $\mathrm{F}^{1,1,1}_{1,1,1}$, $\mathrm{F}^{1,2,2}_{1,1,1}$, $\mathrm{F}^{1,1,2}_{1,1,2}$ and $\mathrm{F}^{1,2,2}_{1,2,2}$ in the bulk and SG-SOI structures. All the form factors are for subbands of the unprimed Δ_2 valley (i.e. first index equal to 1 in the form factor symbols). Fig. 4.6 shows a comparison similar to Fig. 4.5, but between bulk and DG-SOI structures.

The vanishing values of the inter-subband form factor $F^{1,1,2}_{1,1,2}$ for small q values is consistent with Eq. 4.10 for all the device structures. Furthermore, Fig. 4.5 shows that $F^{1,1,1}_{1,1,1}$ is essentially the same in the bulk and the SG device for q larger than approximately 0.2 nm^{-1} and the same holds for $F^{1,1,1}_{1,2,2}$ and $F^{1,2,2}_{1,2,2}$ for q larger than approximately 0.2 nm^{-1}. The same figure also shows that, for vanishing q values, the $F^{1,1,1}_{1,1,1}$, $F^{1,1,1}_{1,2,2}$ and $F^{1,2,2}_{1,2,2}$ for the bulk and SG device tend to the corresponding $\phi_{\mathrm{pcD}}(q{\to}0)$ expressed by Eq. 4.9. Fig. 4.6 shows that, instead, for the DG-SOI MOSFET the $F^{1,1,1}_{1,2,2}$ is much smaller than $F^{1,1,1}_{1,1,1}$ and $F^{1,2,2}_{1,2,2}$, except for very small q values below approximately $t_{\mathrm{Si}}^{-1}{=}0.05$ nm^{-1}.

The behaviour of $F^{1,1,1}_{1,2,2}$ in Fig. 4.6 for the DG-SOI device can be explained by recalling the definition of the form factor in Eq. 4.5. This expression is a double integral where the only term depending on the exchanged wave-vector q is the dimensionless potential $\phi_{\mathrm{pcD}}(q,z,z_0)$ (Eq 4.6). In the DG-SOI case, Fig. 4.3 shows that the $|\xi_{1,1}(z)|$ is mainly confined at the front interface and the $|\xi_{1,2}(z)|$ at the bottom interface. Thus, the only way to obtain a large $F^{1,1,1}_{1,2,2}$ form factor is to have significant $\phi_{\mathrm{pcD}}(q,z,z_0)$ values at both the interfaces, therefore over the whole silicon film; this requires $q{\lesssim}t_{\mathrm{Si}}^{-1}$, consistently with Fig. 4.6.

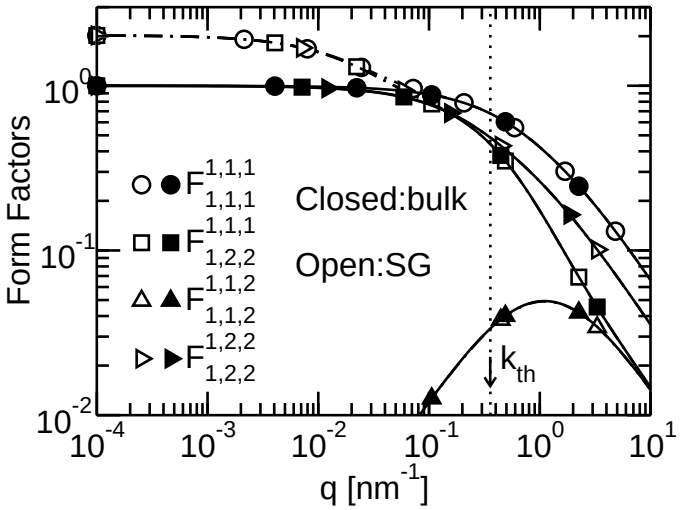

Figure 4.5: Comparison between the form factors $F^{1,1,1}_{1,1,1}$, $F^{1,2,2}_{1,1,1}$, $F^{1,1,2}_{1,1,2}$ and $F^{1,2,2}_{1,2,2}$ of the lowest and second lowest unprimed subbands in the bulk and SG-SOI structures. Same devices and inversion density as in Fig. 4.3.

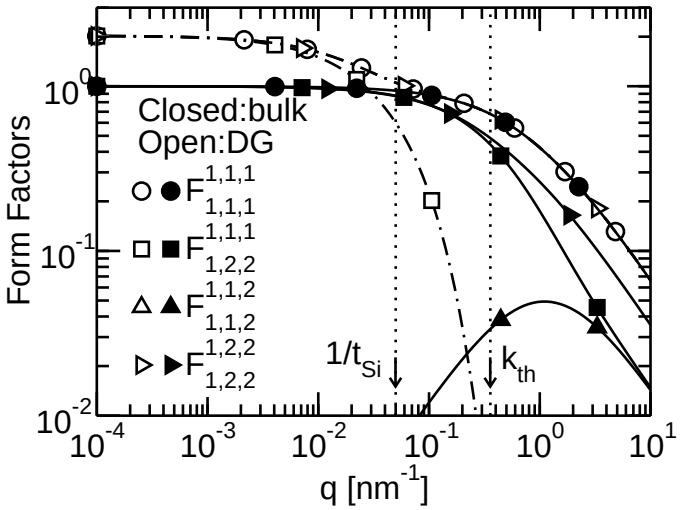

Figure 4.6: Same comparison as in Fig. 4.5 but between bulk and DG-SOI structures. Same devices and inversion density as in Fig. 4.3.

For values close to the thermal wave vector $k_{\rm th}{=}\sqrt{2\pi m^* K_{\rm B}T}/2\hbar$ (for Δ_2 subbands in (100) silicon we have $m^*{=}0.19m_0$ and $k_{\rm th}{\simeq}0.318$ nm^{-1}), the condition $q{\lesssim}t_{\rm Si}^{-1}$ giving $F_{1,2,2}^{1,1,1}{\simeq}F_{1,1,1}^{1,1,1}$ is valid for:

$$t_{\rm Si} \lesssim \frac{1}{k_{\rm th}} \simeq \frac{1}{0.318\ \text{nm}^{-1}} \backsimeq 3.14\ \text{nm} \tag{4.20}$$

Therefore we conclude that for all the $t_{\rm Si}$ values of technological interest for the MOS transistors, the assumption in the Eq. 4.13 is not verified for DG-SOI MOSFETs and FinFETs.

4.4 Explaining the artifacts in the mobility calculations

In this section we assume thick SOI layers and embrace the quantum limit approximation, where only the lowest subband is occupied. Since in thick DG-SOI devices the two lowest subbands are degenerate, in this case we will consider two subbands. This simplified picture allows us to compare analytically the screening models based either on the SDF or the TDF formulation.

Furthermore we assume a bias condition for the SG or DG-SOI device able to induce an inversion density in DG mode exactly twice as large as in SG mode. This results in the physical picture illustrated by Fig. 4.3 and corresponds to wave functions in DG mode which are essentially the same as the wave functions in SG mode, except for the fact that in DG mode the lowest subband is two times degenerate and we conventionally denote with $|\xi_1(z)|^2$ and $|\xi_2(z)^2|$ the wave functions confined respectively at the front and at the back interface (see again Fig. 4.3). The valley index has been dropped in this section consistently with the single valley approximation. In order to simplify the notation also the indication of the dependence on q will be dropped hereafter.

According to the features of the $|\xi_1(z)|^2$ and $|\xi_2(z)|^2$ illustrated in Fig. 4.3 and by recalling the formulation of the unscreened surface roughness matrix elements given in Eq. 4.18, one can realize that, in DG mode, the intra-subband matrix element $M_{1,1}^{(\rm f)}$ corresponding to the front-interface roughness and the wave-function $\xi_1(z)$ must be equal in magnitude to the $M_{2,2}^{(\rm b)}$ for the back interface roughness and $\xi_2(z)$, respectively. Namely we have:

$$|M_{1,1}^{(\rm f)}| \simeq |M_{2,2}^{(\rm b)}| \simeq |M_{1,1}^{(\rm bulk)}| \tag{4.21}$$

Furthermore, the features of the $\xi_1(z)$ and $\xi_2(z)$ make the remaining unscreened matrix elements calculated with Eq. 4.18 negligible with respect to $M_{1,1}^{(\rm f)}$ and $M_{2,2}^{(\rm b)}$, that is:

$$|M_{1,2}^{(\rm f)}|, |M_{1,2}^{(\rm b)}|, |M_{1,1}^{(\rm b)}|, |M_{2,2}^{(\rm f)}| \ll |M_{1,1}^{(\rm f)}| \simeq |M_{2,2}^{(\rm b)}| \tag{4.22}$$

4.4.1 Bulk and SG-SOI devices

For SG-SOI and bulk devices in the quantum limit there is only one relevant matrix element, thus the dielectric function is inherently scalar. In particular, Eq. 4.2 provides a dielectric function which is:

$$\epsilon_{\mathrm{SG}} = 1 + \alpha\,\Pi_{1,1}F_{1,1}^{1,1} \tag{4.23}$$

Therefore the squared magnitude of the screened matrix element is simply given by:

$$|M_{1,1}^{(\mathrm{SG,scr})}|^2 \simeq \frac{|M_{1,1}^{(\mathrm{f})}|^2}{(1+\alpha\,\Pi_{1,1}F_{1,1}^{1,1})^2} \tag{4.24}$$

4.4.2 DG-SOI devices

Since the DG-SOI devices have two degenerate subbands, the SDF and the TDF models are different.

Tensorial dielectric function in DG-SOI devices

In this formulation each element of the form factor matrix is in principle non-zero. However, Fig. 4.6 shows that, in DG-SOI, only $\mathrm{F}_{1,1}^{1,1}$ and $\mathrm{F}_{2,2}^{2,2}$ have a non negligible value in the range of q of pratical importance (approximately larger than 0.1 nm^{-1}). Besides, since the $|\xi_1|^2$ and the $|\xi_2|^2$ are symmetric with respect to the center of the silicon film (see Fig. 4.3), we have:

$$F_{1,1}^{1,1} \simeq F_{2,2}^{2,2} \tag{4.25}$$

Therefore, Eq. 4.1 can be approximately rewritten as:

$$\begin{bmatrix} M_{1,1} \\ M_{1,2} \\ M_{2,1} \\ M_{2,2} \end{bmatrix} = \begin{bmatrix} 1+\alpha\Pi_{1,1}F_{1,1}^{1,1} & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1+\alpha\Pi_{1,1}F_{1,1}^{1,1} \end{bmatrix} \begin{bmatrix} M_{1,1}^{\mathrm{scr}} \\ M_{1,2}^{\mathrm{scr}} \\ M_{2,1}^{\mathrm{scr}} \\ M_{2,2}^{\mathrm{scr}} \end{bmatrix} \tag{4.26}$$

Since the matrix is diagonal, it is easy to calculate the screened matrix elements. For the front-interface, using Eqs. 4.21 and 4.22, we obtain:

$$\begin{bmatrix} M_{1,1}^{(\mathrm{f,scr})} \\ M_{1,2}^{(\mathrm{f,scr})} \\ M_{2,1}^{(\mathrm{f,scr})} \\ M_{2,2}^{(\mathrm{f,scr})} \end{bmatrix} \simeq \begin{bmatrix} \left(1+\alpha\Pi_{1,1}F_{1,1}^{1,1}\right)^{-1} & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & \left(1+\alpha\Pi_{1,1}F_{1,1}^{1,1}\right)^{-1} \end{bmatrix} \begin{bmatrix} M_{1,1}^{(\mathrm{f})} \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{4.27}$$

Doing the same for the back-interface, we have:

$$\begin{bmatrix} M_{1,1}^{(\mathrm{b,scr})} \\ M_{1,2}^{(\mathrm{b,scr})} \\ M_{2,1}^{(\mathrm{b,scr})} \\ M_{2,2}^{(\mathrm{b,scr})} \end{bmatrix} \simeq \begin{bmatrix} \left(1+\alpha\Pi_{1,1}F_{1,1}^{1,1}\right)^{-1} & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & \left(1+\alpha\Pi_{1,1}F_{1,1}^{1,1}\right)^{-1} \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 0 \\ M_{2,2}^{(\mathrm{b})} \end{bmatrix} \tag{4.28}$$

Finally, Eq. 4.19 allows to write:

$$|M_{1,1}^{(\mathrm{DG,scr})}|^2 \simeq \frac{|M_{1,1}^{(\mathrm{f})}|^2}{(1+\alpha\Pi_{1,1}F_{11}^{11})^2} \tag{4.29a}$$

$$|M_{2,2}^{(\mathrm{DG,scr})}|^2 \simeq \frac{|M_{2,2}^{(\mathrm{b})}|^2}{(1+\alpha\Pi_{1,1}F_{11}^{11})^2} \tag{4.29b}$$

$$|M_{1,2}^{(\mathrm{DG,scr})}|^2 \simeq |M_{2,1}^{(\mathrm{DG,scr})}|^2 \simeq 0 \tag{4.29c}$$

As it can be seen, with the TDF model Eqs. 4.29a and Eqs. 4.29b are consistent with Eq.4.24 for the SG-SOI and the bulk case, and, moreover the Eq. 4.29c show that there is no inter-subband scattering. This explains the agreement between SG-SOI and DG-SOI results using the TDF model in Figs. 4.1 and 4.2.

Scalar dielectric function in DG-SOI devices

In this case, the dielectric function is:

$$\epsilon_{\mathrm{DG}} = 1 + 2\alpha\Pi_{1,1}F_{1,1}^{1,1} \tag{4.30}$$

since $\Pi_{1,1}$ and $\Pi_{2,2}$ as well as $F_{1,1}^{1,1}$ and $F_{2,2}^{2,2}$ and sum-up. Thus Eq. 4.17 yields:

$$M_{1,1}^{(\mathrm{f,scr})} = \frac{1}{1+2\alpha\Pi_{1,1}F_{1,1}^{1,1}} M_{1,1}^{(\mathrm{f})} \tag{4.31a}$$

$$M_{2,2}^{(\mathrm{f,scr})} = \frac{1}{1+2\alpha\Pi_{1,1}F_{1,1}^{1,1}} M_{2,2}^{(\mathrm{f})} \tag{4.31b}$$

$$M_{1,2}^{(\mathrm{f,scr})} = M_{1,2}^{(\mathrm{f})} \simeq 0\,,\; M_{2,1}^{(\mathrm{f,scr})} = M_{2,1}^{(\mathrm{f})} \simeq 0 \tag{4.31c}$$

and similar expressions hold for the back interface.

By recalling again Eqs. 4.19, 4.21 and 4.22, we obtain:

$$|M_{1,1}^{(\mathrm{DG,scr})}|^2 \simeq \frac{|M_{1,1}^{(\mathrm{f})}|^2}{(1+2\alpha\Pi_{1,1}F_{1,1}^{1,1})^2} \tag{4.32a}$$

$$|M_{2,2}^{(\mathrm{DG,scr})}|^2 \simeq \frac{|M_{2,2}^{(\mathrm{b})}|^2}{(1+2\alpha\Pi_{1,1}F_{1,1}^{1,1})^2} \tag{4.32b}$$

$$|M_{1,2}^{(\mathrm{DG,scr})}|^2 \simeq 0\,,\; |M_{2,1}^{(\mathrm{DG,scr})}|^2 \simeq 0 \tag{4.32c}$$

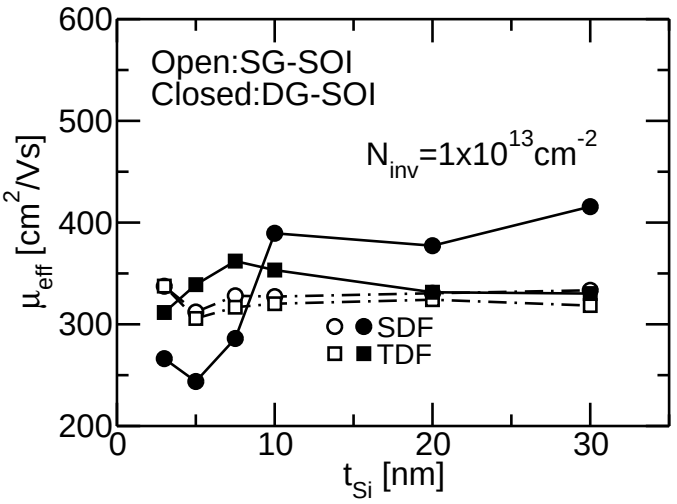

Figure 4.7: Comparison between the simulated mobilities obtained with the SDF and TDF model versus the silicon film thickness $t_{\rm Si}$ for both the SG and DG-SOI devices.

It easy to see that the Eqs. 4.32a and 4.32b are different from the Eq. 4.29a and 4.29b. Indeed, at large inversion densities the unit term is negligible in the denominator of Eqs. 4.29a, 4.29b, 4.32a and 4.32b and the screening of the matrix elements according to Eq. 4.32a and 4.32b can thus become up to four times larger than in the TDF case, hence four times larger than in the bulk and SG-SOI devices described by Eq. 4.24. This explains the larger mobilities at high $N_{\rm inv}$ obtained with the SDF model in DG with respect to SG mode observed in Fig. 4.2.

4.5 Mobility simulation results

In this section we compare the mobility simulation results obtained employing the SDF and the TDF models for the screening.

Due to small residual differences in bulk devices using either the SDF or the TDF models, in order to correctly reproduce the universal mobility curves [47] we set the r.m.s value of the surface roughness mechanism (Δ_{SR}) to 0.62 nm when the simulator employs the SDF model, as in Tab. 3.1, and to 0.66 nm for the TDF model. We assume to have infinitely thick gate oxides since, even if the impact of the oxide thickness on the simulated mobility would be not-negligible, the effect should be very similar between SDF and TDF model, not affecting the results of our analysis. Results are summarized in Fig. 4.7 reporting the simulated mobilities as a function of the silicon film thickness $t_{\rm Si}$ in the strong inversion regime ($N_{\rm inv}$=1×10^{13} cm^{-2} in SG-SOI, $N_{\rm inv}$=2×10^{13} cm^{-2} in DG-SOI). Fig. 4.8 compares the experimental mobility data versus $t_{\rm Si}$ for different values of the inversion density [115, 130].

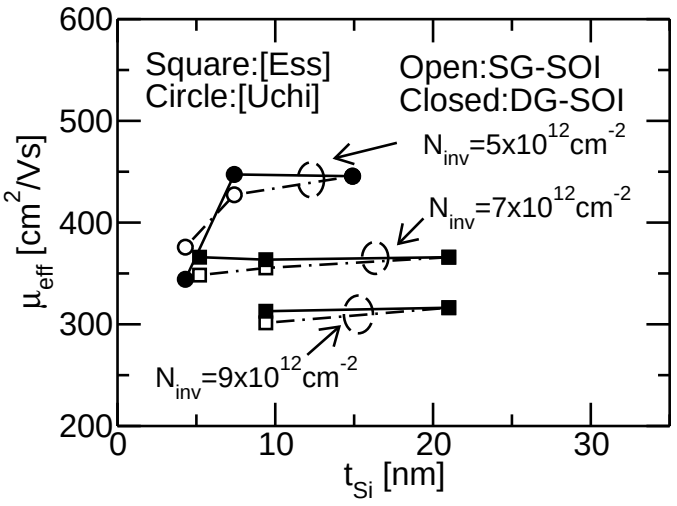

Figure 4.8: Comparison between experimental mobilities versus $t_{\rm Si}$ and for the SG-SOI and the DG-SOI structures; data from [115, 130].

4.5.1 SG-SOI vs. DG-SOI using the TDF approach

First of all we discuss the SG-SOI and DG-SOI simulated mobility of Fig. 4.7 when considering the TDF model. Consistently with Figs. 4.1 and 4.2, in strong inversion and for a large silicon thickness the mobility of the DG-SOI devices is comparable to the SG-SOI case (which is also consistent with experimental data of Fig. 4.1 and 4.8). When, instead, $t_{\rm Si}$ is smaller than 20 nm, the mobility in the DG-SOI case is somewhat larger than in the SG-SOI one. This quantitatively small effect (observed also in the experiments of Fig. 4.8) has been already studied in literature and it is commonly attributed to the volume inversion [131]. When the $t_{\rm Si}$ approaches a value of approximately 3 nm, the DG-SOI mobility is lower than the SG-SOI one, consistently with the results obtained in [132].

4.5.2 SDF vs. TDF in the SG-SOI structure

Let us then focus on the results for the SG-SOI obtained employing the SDF and the TDF models. We notice that simulations based on both the screening models provide very similar mobility values in the entire $t_{\rm Si}$ range. This is fully consistent with the analysis carried out in Sec. 4.4, thus confirming that the assumptions at the basis of the SDF model are verified in SG-SOI devices.

4.5.3 SDF vs. TDF in the DG-SOI structure

As for the DG-SOI case, we observe that the results obtained with the SDF and the TDF are quite different over the whole $t_{\rm Si}$ range. In particular, for $t_{\rm Si}$ larger than approximately 10 nm, the mobility with the SDF model is larger than with the TDF model, which is consistent with the discussion in Sec. 4.4. On the other

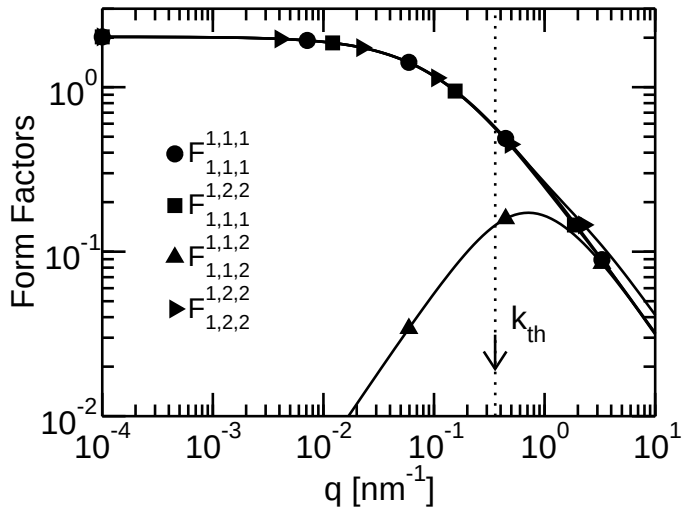

Figure 4.9: Comparison between the $F^{1,1,1}_{1,1,1}$, $F^{1,2,2}_{1,1,1}$, $F^{1,1,2}_{1,1,2}$ and $F^{1,2,2}_{1,2,2}$ in the 5 nm thick DG-SOI structure. In this figure the inversion layer density is 2×10^{13} cm^{-2}.

hand, for $t_{\rm Si}$ smaller than approximately 10 nm, the mobility with the SDF model is smaller than the TDF counterpart.

To explain this latter behaviour, Figs. 4.9 and 4.10 compare the form factors and the matrix elements of the surface roughness of a 5 nm thick DG-SOI structure. In Fig. 4.9 we can see that the assumptions in Eq. 4.10 and Eq. 4.13 can be considered verified, in contrast to what we observed in Fig. 4.6 for $t_{\rm Si}$=20 nm. Thus, Eq. 4.14 is valid in this case. In Fig. 4.10, however, we see that the $|M^{(\rm f)}_{1,1,1}|$ is approximately half as large as the $|M^{(\rm f)}_{1,2,2}|$ in the range of bias conditions of Fig. 4.7. Thus the condition in Eq. 4.16 is still not verified (as it was not for large $t_{\rm Si}$) and invalidates the Eq. 4.17.

We now go back to the quantum limit analysis of Sec. 4.4.2 and focus on the TDF model. Fig. 4.9 shows that the screening function with the TDF model (Eq. 4.2) cannot be written as in Eq. 4.26 due to the large value of the $F^{2,2}_{1,1}$. Thus in this case we assume:

$$F^{2,2}_{1,1} \simeq F^{1,1}_{1,1} \tag{4.33}$$

and Eq. 4.1 can be approximated as:

$$\begin{bmatrix} M_{1,1} \\ M_{1,2} \\ M_{2,1} \\ M_{2,2} \end{bmatrix} = \begin{bmatrix} 1+\alpha\Pi_{1,1}F^{1,1}_{1,1} & 0 & 0 & \alpha\Pi_{2,2}F^{2,2}_{2,2} \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \alpha\Pi_{1,1}F^{1,1}_{1,1} & 0 & 0 & 1+\alpha\Pi_{2,2}F^{2,2}_{2,2} \end{bmatrix} \begin{bmatrix} M^{\rm scr}_{1,1} \\ M^{\rm scr}_{1,2} \\ M^{\rm scr}_{2,1} \\ M^{\rm scr}_{2,2} \end{bmatrix} \tag{4.34}$$

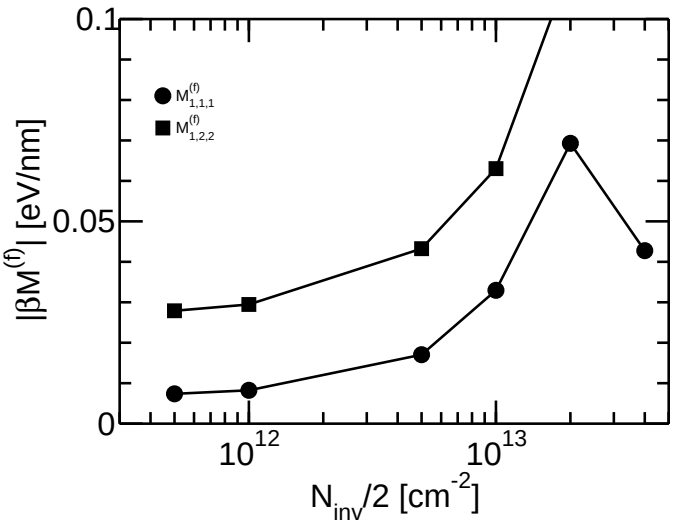

Figure 4.10: Comparison between the normalized unscreened matrix elements $\beta M^{(\mathrm{f})}_{1,1,1}$ and $\beta M^{(\mathrm{f})}_{1,2,2}$ as a function of the inversion density in the channel in DG-SOI structure. The β parameter is defined as the $A/\Delta_{SR}(q)$.

By inverting the Eq. 4.34, the $M^{\mathrm{scr}}_{1,1}$ can be calculated as:

$$M^{\mathrm{scr}}_{1,1}(q) = \frac{M_{1,1}(q)}{1+\alpha\Pi_{1,1}F^{1,1}_{1,1}+\alpha\Pi_{2,2}F^{2,2}_{2,2}} + \frac{\alpha(q)\Pi_{2,2}(q)F^{2,2}_{2,2}(q)[M_{1,1}(q)-M_{2,2}(q)]}{1+\alpha\Pi_{1,1}F^{1,1}_{1,1}+\alpha\Pi_{2,2}F^{2,2}_{2,2}} \quad (4.35)$$

which is an approximated model for the TDF approach. With respect to the results obtained by the SDF model in the quantum limit approximation and in DG mode (Eq. 4.31a), Eq. 4.35 contains a term proportional to $[M_{1,1,1}-M_{1,2,2}]$ which, by defining as positive all the unscreened intra-subband matrix elements (Fig. 4.10), is negative. As a result the screened matrix element in the TDF model is smaller than the one in SDF model and this explains the lower mobility provided by the SDF model in Fig. 4.7 for t_{Si} below 10 nm.

To complete the analysis, Tab. 4.1 shows that at small t_{Si} , there is a significant difference between the matrix elements of the surface roughness also in the SG-SOI device, however, this happens with a depopulation of the second lowest subband with respect to the lowest one that results in a polarization factor $\Pi_{2,2}$ much smaller than $\Pi_{1,1}$. Since the difference between the SDF and TDF models has been identified for the present analysis in the second term in Eq. 4.35, we see that for t_{Si}=5 nm, it is the reduction of the $\Pi_{2,2}$ that mainly contributes to maintain good agreement between the SDF and TDF models in SG mode. We verified also for different t_{Si}s that when a significant difference exists between $M_{1,1}$ and $M_{2,2}$ in SG mode, this happens contextually with a depopulation of the second

		$\alpha F_{i,i}^{i,i}$ [$\times 10^{-14}$ cm^{2}eV]	$\Pi_{i,i}$ [$\times 10^{14}$ cm^{-2}eV^{-1}]	$\lvert\beta M_{1,1}^{(\mathrm{f})}\rvert$ [$\times 10^{-2}$ eVnm^{-1}]	N_i [$\times 10^{12}$ cm^{-2}]
SG-SOI	i=1	2.688	1.3784	9.72	6.9186
	i=2	2.351	0.3114	4.71	0.85336
DG-SOI	i=1	2.232	1.4627	3.41	4.1362
	i=2	2.204	1.2472	6.41	2.8382

Table 4.1: Comparison between the intra subband form factor $\alpha F_{i,i}^{i,i}$, the intra subband polarization factor $\Pi_{i,i}$, the intra subband matrix element $|\beta M_{1,1}^{(\mathrm{f})}|$ and the inversion charge N_i of the lowest and second lowest subbands in 5 nm thick SG-SOI and DG-SOI devices.

lowest subband, which pushes the device toward the quantum limit and reduces the discrepancy between the SDF and TDF models.

Therefore, we can notice that when the assumption on the intra-subband form factors (Eq. 4.13) fails (t_{Si} thicker than 10 nm), the scalar model provides larger mobility values than the matrix one. When the aforementioned condition can be considered fairly verified (t_{Si} thinner than 10 nm), instead, the mobility of the SDF model is smaller than the TDFs one since Eq. 4.16 is not satisfied. Thus, the error associated to the use of the SDF model depends on the value t_{Si} and it is difficult to correct it in a simple way.

4.6 Conclusions

In this chapter we have examined the limits of validity of the SDF approach for the screening in bulk and SOI devices. We have found that the SDF model is fairly accurate in bulk and SG-SOI structures, whereas it becomes inaccurate in DG-SOI devices, in which case the SDF produces artifacts in the simulated mobility curves that are clearly inconsistent with the experiments.

We have presented an in detail analysis based both on numerical calculations and on analytical derivations that identified the assumptions behind the approximated SDF model that are clearly violated in DG-SOI MOSFETs, for reasons that are inherent to the electronic structure in DG mode.

It should be emphasized that the failure of the SDF model results in an overestimate or an underestimate of the mobility (with respect to the TDF model) depending on the thickness t_{Si} of the silicon film; this observation corroborates that the SDF model is not reliable in DG mode and, as such, it should not be used.

Chapter 5

MOSFETs with high-k dielectrics

Abstract

In this chapter we examine the mobility reduction associated to high-κ dielectrics in n- and p-MOSFETs already discussed in Sec. 1.2 by means of extensive comparison between accurate multi-subband Monte Carlo simulations and experimental data.

In the first part of the chapter we briefly recall the models we developed for the soft optical phonons and remote Coulomb scattering mechanism. Concerning the soft optical phonons, we first explain a model developed in a simplified structure and then we show the model for more realistic structures. as for the remote Coulomb scattering we describe the models for a single charge and for dipoles in a realistic gate stack structure.

In the second part of the chapter, we use Multi-Subband Monte Carlo simulations to understand which mechanism is mainly responsible for the mobility degradation observed in nMOSFETs featuring Hf-based high-κ dielectrics. Direct comparison with the experimental data points out that for realistic interfacial layer thicknesses the effect of surface optical phonons on the mobility is very modest in both the electron and hole inversion layers. Experimental data of devices featuring different HK materials from different process technologies can be reproduced only assuming consistently large concentrations of Coulomb scattering centers in the gate stack. However the corresponding remote charge or dipole density would result in a large threshold voltage shift not observed in the experiments.

Finally, assuming remote Coulomb centers in the form of a single charge sheet at the ITL/HK interface we simulate the drain current in real devices and we found that the current reduction in short channel devices is, instead, not as strong as the mobility reduction.

5.1 Introduction

Several mechanisms have been invoked to explain the mobility degradation related to the use of high-k dielectrics in MOSFETs, and a prominent role has been ascribed to soft optical phonons (SOph) [48, 49, 50, 51]. However, recent studies predict an influence of SOph on the *electron* mobility significantly weaker than previously thought [31, 133]. Coulomb centers in the gate stack (RemQ) have also been proposed as a possible cause of the mobility reduction [134, 135, 31, 52], but very large charge densities seem to be necessary to justify the experimental mobility degradation [31]. These densities seem to be not in agreement with the studies on the flat band voltage shift produced by the remote charges [136, 137, 138] or with direct measurements obtained with modified charge pumping methods [139, 140]. Furthermore recent experimental data and atomistic simulations suggest that the mobility degradation could be due to interface dipoles close to the HK/ITL interface (DipQ) [53, 54, 55]. The mobility reduction has been also ascribed to Nitrogen diffusion near the channel/dielectric interface [141, 142, 143]. From the device modeling perspective, while the models for the RemQ charges are quite well assessed [106, 144], different models for SOph scattering mechanism have been proposed, which result in quite different predictions [145, 146], thus making the overall scenario still unclear. A model for the DipQ scattering is proposed in this thesis [147].

In this Chapter we aim to provide the complete details of the theory and the results of an accurate comparison between multi-subband Monte Carlo transport simulations of the electron and hole low field mobility and a broad set of experimental data. We show that samples fabricated by different labs exhibit a remarkably consistent behavior, which can not be explained by soft optical phonon scattering. We additionally show that a very large amount of charges in the gate stack is necessary to reproduce the experimental data published by many authors. However, by modeling the Coulomb centers both as remote single charges (RemQ) or as remote dipoles (DipQ), the threshold voltage shifts are in contrast with the experimental evidences.

In Sec. 5.2, 5.3 and 5.4 we describe respectively the models for the SOph, RemQ and DipQ scattering employed in this work. In Sec. 5.5 we examine the effect of these scattering mechanisms on the n- and p-MOSFET mobility by means of comparison between experimental data and simulations. Sec. 5.6 examines the threshold voltage shifts associated to the RemQ and DipQ densities used to reproduce the experimental data. Finally, in Sec. 5.7 we investigate the effect of the high-k dielectrics on the current of short channel devices, by analyzing separately the effect of SOph and RemQ on the $I_D - V_G$ curves.

5.2 Modeling SO phonons

The surface soft optical phonons originate from the polar phonon modes present in high-k dielectrics. Indeed, the molecules of this type of insulators are strongly polarized, in accordance with the large electric permittivity. The vibration of this polar molecules (thermally activated) causes non-stationary electric fields which can penetrate into the Si channel, with exponential decay, as we will see in Secs. 5.2.2 and 5.2.3. Due to this type of behavior, they are called *surface optical phonons*. Moreover they are also called *soft* because the bond between the metal atom, for example Hf, and the atom of O is "soft", i.e. allows the molecule to vibrate strongly. Finally, they are also labeled as *optical* because their frequency is quite high and weakly dependent on their wave-vector (see Sec. 5.2.3).

5.2.1 Models for SO-phonons available in literature

One of the first models for soft optical phonon scattering have been proposed by Hess and Vogl [148]. They studied remote polar phonons of SiO_2 and they proposed these phonons as a possible cause of a "new" type of scattering mechanism. Indeed they observed that the vibrations of polar molecules of the gate oxide influence the mobility of channel carriers. In particular, they realized that phonon limited mobility of free carriers in inversion layers was somewhat lower than in bulk material for all the range of temperatures. They studied the effects of one polar phonon originated in SiO_2 and they observed that the potential scattering of this mechanism decays esponentially into the Si substrate. They obtained these results using a model implementing only one subband and without the screening effect.

Then, Moore and Ferry [149] calculated the effects of polar phonons of SiO_2 upon carrier mobility using a more complex model. They noticed that the mobility degradation due to the polar phonons of SiO_2 was negligible. They considered the lowest and second lowest phonons that originate in the gate dielectric. They used a three-energy-level model for transport in the quasi-two-dimensional inversion layer assuming a drifted-Maxwellian distribution.

In 1978 Kim, Das and Senturia [150] proposed a generalized formulation of the electron scattering interaction with coupled plasmon-phonon modes in degenerate polar semiconductors. They derived the scattering lifetime in terms of a dielectric response formalism.

In 1993 Fischetti and Laux [98] studied what are the main limiting scattering mechanisms affecting the mobility of Si inversion layers using a multi-subband Monte Carlo simulator. They included the surface optical phonon scattering, developing two models. The first model considers two semi-infinite regions: a semiconductor and an insulator which permittivity depends on frequency. The other model considers a finite thickness of the SiO_2, with on top the metal gate. These models accounts for the lowest and second lowest optical phonons that originates in the SiO_2.

In 2001 Fischetti *et al.* [48] studied the effects on the carrier mobility of the

polar phonons in high-k gate stacks. To study this type of scattering mechanism, they used a triangular approximation of the potential well in the channel. Firstly, they also assumed an infinite thickness of the high-k dielectric, showing that, while the polar phonons of the SiO_2 have negligible impact on the mobility, polar phonon of popular high-k dielectrics such as HfO_2 degrade carrier mobility in the channel. The paper also proposes a model to describe the plasmons that originate in the poly-Si gate and another model which accounts for the presence of the interfacial layer.

In 2003 Ren, Fischetti and Gusev [151] used Kubo's formula and a self-consistent solution of Poisson and Schrödinger equation, to study mobility degradation due to high-k dielectrics. They showed that there is a mobility degradation due to the presence of these polar phonons, and they proposed the use of high-k silicates to try to improve the mobility of high-k stacks.

In 2004 Kotlyar et al. [152] used Fischetti's SO-phonons model [48] to show that the use of the metal gate is effective to improve carrier mobility in the channel. Indeed they observed that, considering the metal gate as an ideal conductor, all gate plasmons are suppressed.

In 2007 Shah et al. [153] developed a model for the plasmon-phonon coupling modes that originate when the metal gate is a non-ideal conductor. They used the Kubo-Greenwood integral to study the SO-phonon limited mobility changes due to these modes. Accounting for five primed and unprimed subbands as well as surface roughness and bulk phonon scattering, their model does not consider Coulomb scattering and the presence of the interfacial layer.

5.2.2 SOph in structures without interfacial layer

Modeling details

The perturbation potential for SOph scattering in a structure with an infinite high-k layer on top of a Si channel (see Fig. 5.1) has been computed following the approach in [48].

The dielectric constant of the channel material is set to its low frequency value. The lowest (TO1) and the second lowest (TO2) modes in the HK layer have been taken into account according to [48], so that the permittivity of the high-k layer is:

$$\epsilon_{\mathrm{HK}}(\omega) = \epsilon_\infty + \frac{\epsilon_0 - \epsilon_\mathrm{i}}{1 - (\frac{\omega}{\omega_{\mathrm{TO},1}})^2} + \frac{\epsilon_\mathrm{i} - \epsilon_\infty}{1 - (\frac{\omega}{\omega_{\mathrm{TO},2}})^2} \tag{5.1}$$

where ϵ_0, ϵ_i, ϵ_∞, $\omega_{\mathrm{TO},1}$ and $\omega_{\mathrm{TO},2}$ are the electric permittivity of the HK material at low, intermediate and infinite frequency and the energies of the lowest and second lowest phonon modes, respectively. Tab. 5.1 reports the values of ϵ_0, ϵ_i, ϵ_∞ $\omega_{\mathrm{TO},1}$ and $\omega_{\mathrm{TO},2}$ for some common high-k dielectrics.

Neglecting the coupling between the SOph modes and the plasmons originating from the inversion layer (this approximation will be discussed in App. B), the

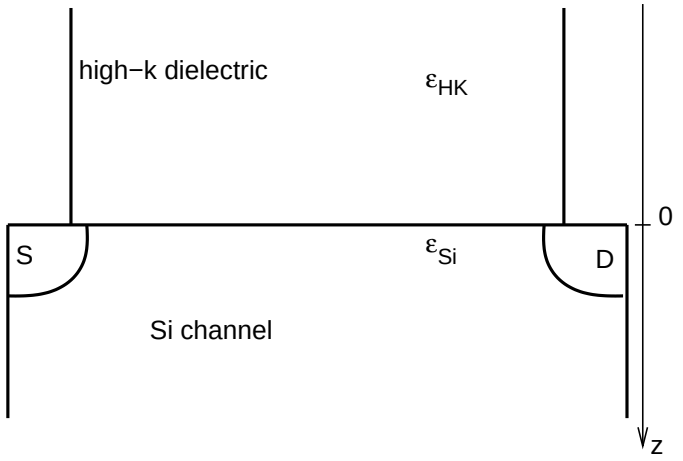

Figure 5.1: Sketch of the gate stack structure considered for the SOph model without ITL and definition of the symbols used. The dielectric constant of the various layers is also indicated.

Quantity	SiO_2	Al_2O_3	AlN	ZrO_2	HfO_2	$ZrSiO_4$
ϵ_0	3.90	12.53	9.14	24.0	22.00	11.75
ϵ_i	3.05	7.27	7.35	7.75	6.58	9.73
ϵ_∞	2.50	3.20	4.80	4.00	5.03	4.20
$\hbar\omega_{\mathrm{TO},1}$ [meV]	55.60	48.18	81.40	16.67	12.40	38.62
$\hbar\omega_{\mathrm{TO},2}$ [meV]	138.10	71.41	88.55	57.70	48.35	116.00

Table 5.1: Parameters of some gate dielectrics from [48]. ϵ_0 is the relative permittivity at zero frequency, ϵ_i at intermediate frequency and ϵ_∞ at infinite frequency for each material. $\omega_{\mathrm{TO},1}$ and $\omega_{\mathrm{TO},2}$ are the energies of the lowest and second lowest phonons modes. The static permittivity ϵ_0 of the Si is 11.9.

condition allowing the modes to propagate is [154, 78]:

$$\epsilon_{\mathrm{HK}} = -\epsilon_{\mathrm{Si}} \tag{5.2}$$

Combining Eqs. 5.1 and 5.2, we can find the angular frequency for the lowest SOph mode, that is:

$$\omega_{\mathrm{SO},1}(q) = \omega_{\mathrm{TO},1}\sqrt{\frac{\epsilon_0 + \epsilon_{\mathrm{Si}}}{\epsilon_{\mathrm{i}} + \epsilon_{\mathrm{Si}}}} \tag{5.3}$$

and for the second lowest SOph mode:

$$\omega_{\mathrm{SO},2}(q) = \omega_{\mathrm{TO},2}\sqrt{\frac{\epsilon_{\mathrm{i}} + \epsilon_{\mathrm{Si}}}{\epsilon_{\infty} + \epsilon_{\mathrm{Si}}}} \tag{5.4}$$

The expression of the perturbation potential in the Si channel for a SOph mode is [154, 48, 78]:

$$\phi_{\mathrm{SO}}(q, z) = A_0 e^{-qz} \tag{5.5}$$

The amplitude of the perturbation potential at the Si/HK interface A_0 is found by equating the classical and the quantum mechanical phonon energy (as in [48]) [78]. For the lowest SOph mode we thus have:

$$\frac{(n_{\mathrm{SO}} + \frac{1}{2} \pm \frac{1}{2})\hbar\omega_{\mathrm{SO},1}}{A} = W_{\mathrm{EM}} \tag{5.6}$$

where A is the normalization area and n_{SO} is the number of phonons, given by the Bose-Einstein statistics (Eq. 3.23). The classical energy W for the lowest mode can be expressed as [155]:

$$W_{\mathrm{EM}} = 2q|A_0|^2 \Re\left\{\omega_{\mathrm{SO},1}\frac{\partial\epsilon_{\mathrm{AVE}}(\omega)}{\partial\omega}\bigg|_{\omega_{\mathrm{SO},1}}\right\} \tag{5.7}$$

where ϵ_{AVE} is an average electric permittivity of the system and has the expression:

$$\epsilon_{\mathrm{AVE}} = \frac{\epsilon_{\mathrm{Si}} + \epsilon_{\mathrm{HK}}}{2} \tag{5.8}$$

Using Eqs. 5.6 and 5.7 we can find the square of the amplitude of the scattering potential for the lowest mode at the interface between the semiconductor and the channel as:

$$|A_0|^2 = \frac{(n_{\mathrm{SO}} + \frac{1}{2} \pm \frac{1}{2})\hbar\omega_{\mathrm{SO},1}}{2qA\Re\left\{\omega_{\mathrm{SO},1}\frac{\partial\epsilon_{\mathrm{AVE}}(\omega)}{\partial\omega}\Big|_{\omega_{\mathrm{SO},1}}\right\}} \tag{5.9}$$

From Eqs. 5.1 (where we have to neglect the third right term, since we are analyzing only the lowest mode) and 5.8 we obtain:

$$\frac{\partial\epsilon_{\mathrm{AVE}}(\omega)}{\partial\omega} = \frac{\epsilon_0 - \epsilon_{\mathrm{i}}}{\left[1 - \left(\frac{\omega}{\omega_{\mathrm{TO},1}}\right)^2\right]^2}\left(\frac{\omega}{\omega_{\mathrm{TO},1}^2}\right) \tag{5.10}$$

Remembering Eqs. 5.5, 5.9 and 5.10, and multiplying for a term A to have the correct units, we can finally obtain:

$$\phi_{\mathrm{SO},1}(q,z) = \sqrt{\frac{(n_{\mathrm{SO}} + \frac{1}{2} \pm \frac{1}{2})\hbar\omega_{\mathrm{SO},1}A}{2q}\left(\frac{1}{\epsilon_{\mathrm{i}} + \epsilon_{\mathrm{Si}}} - \frac{1}{\epsilon_0 + \epsilon_{\mathrm{Si}}}\right)}e^{-qz} \quad (5.11)$$

for the lowest SOph mode. Similarly, it is possible to find also the amplitude of the scattering potential of the second lowest SOph mode, that is:

$$\phi_{\mathrm{SO},2}(q,z) = \sqrt{\frac{(n_{\mathrm{SO}} + \frac{1}{2} \pm \frac{1}{2})\hbar\omega_{\mathrm{SO},2}A}{2q}\left(\frac{1}{\epsilon_{\infty} + \epsilon_{\mathrm{Si}}} - \frac{1}{\epsilon_{\mathrm{i}} + \epsilon_{\mathrm{Si}}}\right)}e^{-qz} \quad (5.12)$$

$\phi_{\mathrm{SO},1}(q,z)$ and $\phi_{\mathrm{SO},2}(q,z)$ are the two-dimensional Fourier transform of the scattering potential that has been defined in Eq. 3.22. This method is fully consistent with the results obtained in [154, 48, 78].

Comparison with other authors

Before discussing the details of mobility modeling for a realistic gate stack featuring an ITL, the HK dielectric and a metal gate, we show in Fig.5.2 the simulated effective mobility (μ_{eff}) for an infinite HfO_2 layer on top of a Si channel without scattering from remote Coulomb centers and neglecting the coupling of the SOph with electron gas in the inversion layer. The matrix elements for the SOph scattering in such a simple situation (Eqs. 5.3–5.12) have a well established formulation [148, 149]. However, the figure shows that calculations from different authors (open symbols) disagree even in this simple situation, indicating that there is still the need for a modeling based analysis of the importance of SOph scattering in HK stacks. Good agreement is found between our model (filled squares) and the curve obtained combining by means of the Matthiessen's rule the experimental mobility curve for a SiO_2 dielectric (N_{A}=2×10^{16} cm^{-3} [47]) with the SOph limited mobility computed in [133] (open circles).

Our calculations predict a less than 40% reduction of μ_{eff} with respect to the universal curve when no ITL is present. As for the differences between our results and the simulations reported in [48, 156, 157, 158], possible reasons are:

1. with respect to [156] we have a different treatment of the surface roughness scattering mechanism; In fact, Fig. 5.3a shows that our results are close to [156] at low N_{inv}, when plotted vs. the inversion density N_{inv}. Moreover, Fig. 5.3b shows that our SOph limited mobility is close to the one reported in [156] for a wide range of temperatures.
2. the model in [48] assumes a triangular well approximation whereas the actual shape of the potential well may be critical because the scattering potential has an exponential decay from the Si/dielectric interface [78].
3. the momentum-relaxation-time approximation for the calculation of μ_{eff} (as in [48]), whose formulation and implementation is critical when considering inelastic and anisotropic scattering mechanisms such as SOph.

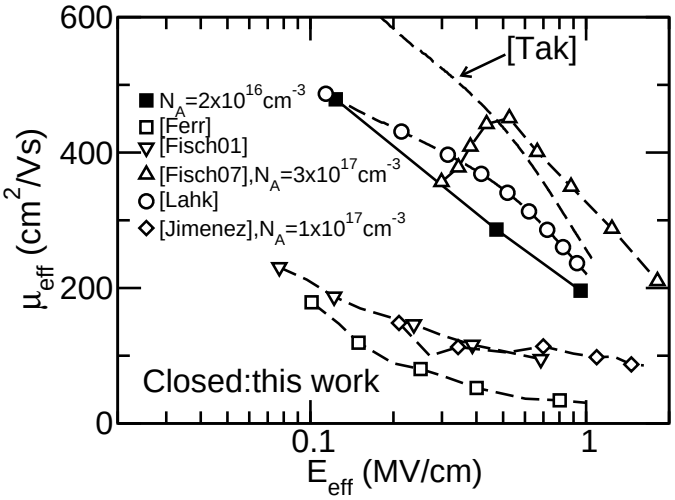

Figure 5.2: Simulated effective mobility vs. effective field for a gate stack consisting of a thick layer of HfO_2 without ITL as obtained from simulations in this work (filled symbols, Eqs. 5.3–5.12) and in the literature [48, 157, 158, 156, 133]. The considered scattering mechanisms are: acoustic and optical phonons in the Si substrate, surface roughness and soft optical phonons. Parameters for phonons and surface roughness have been calibrated to reproduce the universal mobility curves (dashed line in figure [47]) in SiO_2/Si stacks [93, 78].

4. the free-electron gas approximation in [157], which overestimates SOph scattering, since carriers are much closer to the interface than they are in a model accounting for quantization. Indeed, in Fig. 5.4, the curve obtained with our model without accounting for quantization in the calculation of the SOph scattering rate (filled diamonds) is in good agreement with the results of [157]. These results have been obtained by substituting the values of ξ_{k_2} and ξ_{k_1} in Eq. 3.21 with the classical solution of the Poisson Equation, i.e. using:

$$\xi(z) = \sqrt{\frac{2eE_{\text{eff}}}{K_{\text{B}}T}}e^{-\left(\frac{eE_{\text{eff}}z}{K_{\text{B}}T}\right)} \tag{5.13}$$

that somehow turns off the quantization in the computation of the matrix elements of the SOph scattering.

Fig. 5.5 shows the simulated electron and hole mobility in bulk devices versus the effective field E_{eff} taking into account SOph scattering for various HK dielectrics *without* ITL. The parameters of the SOph model are taken from [48] and are shown in Tab. 5.1. In the n-MOS case (closed symbols) SOph scattering causes a large mobility reduction (up to about 40%), whereas in the p-MOS case (open symbols) the effect is negligible, for HfO_2, which instead produces the largest mobility reduction in the n-MOS case.

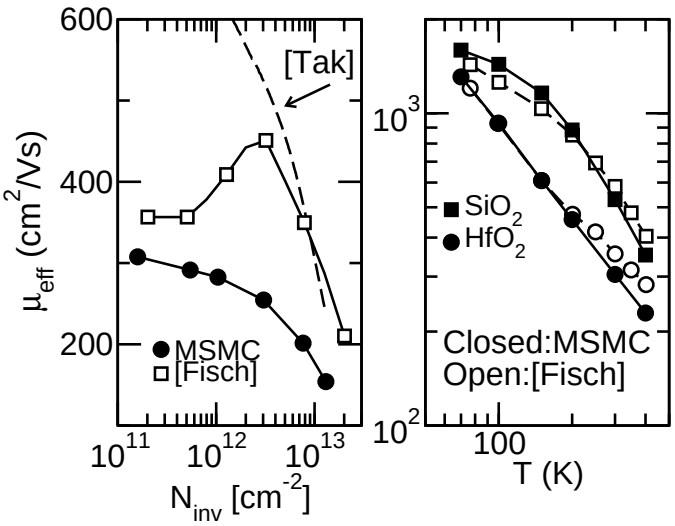

Figure 5.3: Simulated effective mobility for a gate stack consisting of a thick layer of HfO_2 without ITL. The considered scattering mechanisms are: acoustic and optical phonons in the Si substrate, surface roughness and soft optical phonons. Our simulations are directly compared with those of [156]. a) Effective mobility at 300 K as a function of the inversion charge density $N_{\rm inv}$. b) Comparison between the SOph limited mobility for the SOph of SiO_2 and HfO_2 , as a function of the temperature ($N_{\rm inv}$=2×10^{11} cm^{-2}).

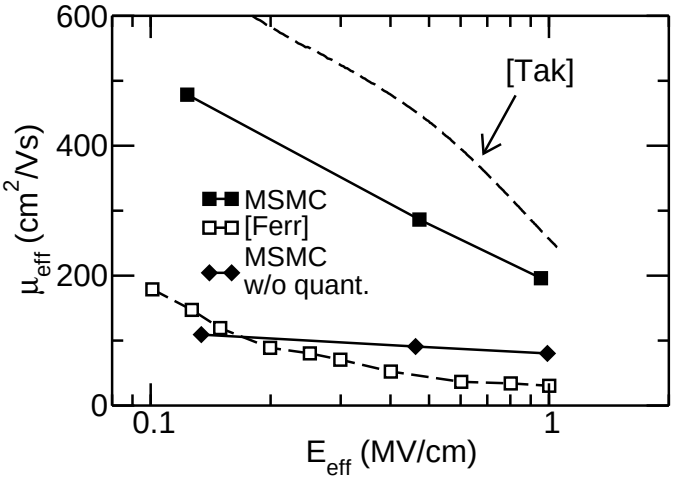

Figure 5.4: Simulated effective mobility for a gate stack consisting of a thick layer of HfO_2 without ITL. The considered scattering mechanisms are: acoustic and optical phonons in the Si substrate, surface roughness and soft optical phonons. Results of [157] are compared with our results when the quantization is taken into account (squares) or neglected (diamonds) in the matrix element computation.

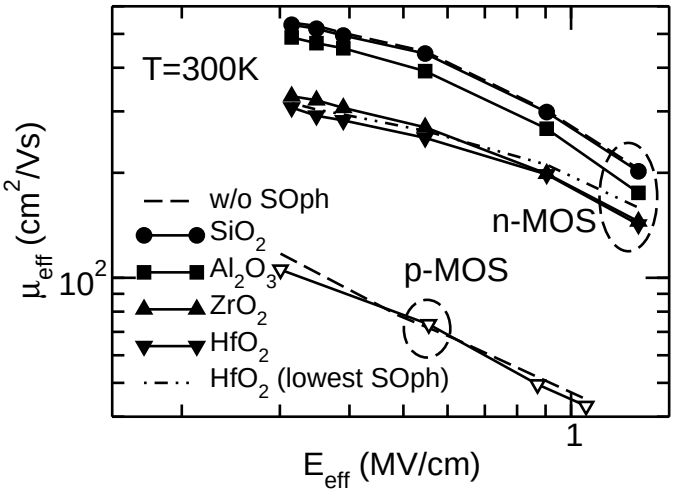

Figure 5.5: Simulated effective mobility vs. effective field taking into account the SOph scattering mechanism for various HK dielectrics without ITL. Results for electron (filled symbols) and hole (open symbols) inversion layers are shown, with doping $N_{\rm A}$=3×10^{17} cm^{-3} and $N_{\rm D}$=2.7× 10^{17} cm^{-3}, respectively. Scattering mechanisms are the same as in Fig. 5.2. The dot-dashed line shows the mobility degradation obtained accounting only for the lowest SOph of the HfO_2.

In the pMOS case the effect of the SOph on the mobility is much smaller for several reasons. Firstly because the mobility without SOph is much lower than in the nMOS case, so that the additional scattering mechanism plays a minor role. Secondly, the SOph limited mobility is larger in pMOS than in nMOS. In fact, the matrix element contains terms in the form $\exp(-qz)$ where q is the exchanged wave-vector (see Eqs. 5.11 and 5.12). Trying to simplify the picture, we can assume that the transitions involve carriers only at the Fermi electron surface (or curve, since the Gas is 2D). The curvature of the $E-k$ relationship for holes is much smaller than for electrons, meaning that the q of the transitions is larger than in the nMOS case. The transitions with larger q are those affecting the mobility the most, since they strongly change the wave vector k. However, the term $\exp(-qz)$ in the scattering matrix element of the SOph decreases with the increasing of q, leading to a reduced influence of the SOph scattering mechanism on the overall mobility. In the same spirit, this explains why SOph becomes very relevant in high mobility (and low transport mass) materials such as III-V [159] and graphene [160].

Moreover, the dot-dashed line in Fig. 5.5 is obtained for the HfO_2 n-MOS case, in which only the effect of the first lowest SOph is accounted for. We can see that the effect of the second lowest SOph (Eqs. 5.4 and 5.12) is modest for HfO_2 and can be safely neglected.

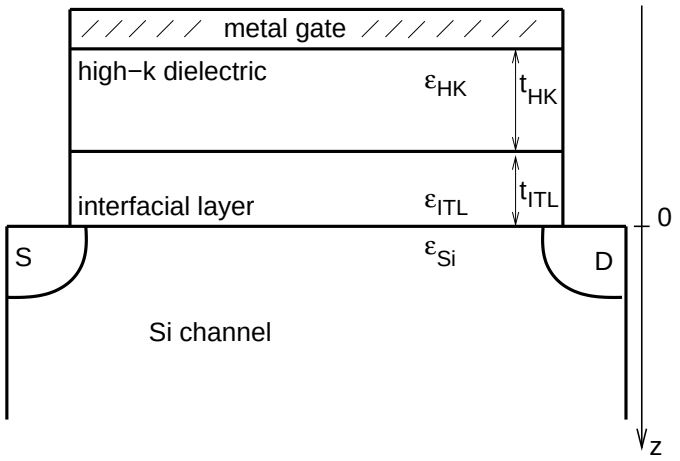

Figure 5.6: Sketch of a realistic gate stack structure and definition of the symbols used in the text. The dielectric constant of the various layers is also indicated.

5.2.3 SOph scattering in MG/HK/ITL structures

The perturbation potential for SOph scattering in a gate stack featuring a SiO_2 ITL, the HK dielectric and an ideal metal gate (see Fig. 5.6) has been computed following the approach in [48].

In principle, all remote phonon modes that originate in both the HK and the ITL material should be considered. However, an analytic expression for the phonon energy dispersion is possible only if we consider a single phonon. It is expected that the HK modes are predominant due to the larger dielectric permittivity compared to the SiO_2 ITL. Moreover, Fig. 5.5 shows that the effect of the second lowest SOph can be neglected even in that simplified structure. Indeed, the low energy modes are those that affect the mobility the most, because their energy is closer to the thermal carrier energy. We thus conclude that the effect of the second lowest SOph can be safely neglected in a structure as the one in Fig. 5.6. Therefore, the dielectric constants of the ITL and of the Si channel are set to their low frequency values and only the lowest mode (TO1) in the HK layer has been taken into account simplifying Eq. 5.1 as:

$$\epsilon_{\mathrm{HK}}(\omega) = \epsilon_{\mathrm{i}} + \frac{\epsilon_0 - \epsilon_{\mathrm{i}}}{1 - (\frac{\omega}{\omega_{\mathrm{TO},1}})^2} \tag{5.14}$$

where ϵ_0, ϵ_{i} and $\omega_{\mathrm{TO},1}$ are the electric permittivity of the HK material at low and intermediate frequency and the energy of the lowest phonon mode, respectively.

The expression of the perturbation potential $\phi_{\mathrm{SO},i}(q,z)$ in the generic i-th layer (namely Si, ITL or HK layer, see Fig.5.6) is [48, 154, 78]:

$$\phi_{\mathrm{SO},i}(q,z) = A_{i,1}e^{-qz} + A_{i,2}e^{qz} \tag{5.15}$$

where $A_{i,1}$ and $A_{i,2}$ are two constants to be determined imposing the continuity of the potential $\phi_{\mathrm{SO}}(q,z)$ and of the electric displacement field $\epsilon\partial\phi(q,z)/\partial z$ at

the ITL/HK and Si/ITL interfaces. At the HK/MG interface we have imposed $\phi_{\rm SO}(q, z{=}{-}t_{\rm ITL}{-}t_{\rm HK}){=}0$, thus considering the metal gate as an ideal conductor. Imposing that the determinant of the corresponding linear system of equations, that is:

$$\begin{cases} A_{\rm HK,1}e^{-q(t_{\rm ITL}+t_{\rm HK})} + A_{\rm HK,2}e^{q(t_{\rm ITL}+t_{\rm HK})} = 0 \\ A_{\rm HK,1}e^{-qt_{\rm ITL}} + A_{\rm HK,2}e^{qt_{\rm ITL}} = A_{\rm ITL,1}e^{-qt_{\rm ITL}} + A_{\rm ITL,2}e^{qt_{\rm ITL}} \\ A_{\rm ITL,1} + A_{\rm ITL,2} = A_{\rm Si,1} \\ \epsilon_{\rm HK}(A_{\rm HK,1}e^{-qt_{\rm ITL}} - A_{\rm HK,2}e^{qt_{\rm ITL}}) = \epsilon_{\rm ITL}(A_{\rm ITL,1}e^{-qt_{\rm ITL}} - A_{\rm ITL,2}e^{qt_{\rm ITL}}) \\ \epsilon_{\rm ITL}(A_{\rm ITL,1} - A_{\rm ITL,2}) = -\epsilon_{\rm Si}A_{\rm Si,1} \end{cases} \tag{5.16}$$

is zero, we determine the angular frequency of the resulting SOph mode as a function of the magnitude q of the exchanged wave-vector, that is [154, 78]:

$$\omega_{\rm SO}(q) = \omega_{\rm TO,1}\sqrt{\frac{\gamma - \frac{\epsilon_0}{\epsilon_{\rm ITL}}}{\gamma - \frac{\epsilon_{\rm i}}{\epsilon_{\rm ITL}}}} \tag{5.17}$$

where:

$$\gamma = \left[\frac{1 - e^{-2qt_{\rm HK}}}{1 + e^{-2qt_{\rm HK}}}\right]\left[\frac{1 - \frac{\epsilon_{\rm ITL}+\epsilon_{\rm Si}}{\epsilon_{\rm ITL}-\epsilon_{\rm Si}}e^{2qt_{\rm ITL}}}{1 + \frac{\epsilon_{\rm ITL}+\epsilon_{\rm Si}}{\epsilon_{\rm ITL}-\epsilon_{\rm Si}}e^{2qt_{\rm ITL}}}\right] \tag{5.18}$$

while the other symbols are defined in Fig. 5.6. Fig. 5.7 shows the modes originating from a complete numerical solution (explained in App. C) obtained accounting for the TO1 and TO2 modes in both the HK and the ITL layers, and with $t_{\rm ITL}$=1 nm and $t_{\rm HK}$=5 nm. The two TO modes in the ITL produce four branches, whereas the two TO modes in the HK produce two additional branches [145]. However, Fig. 5.7 shows that the other remote phonon modes have a large energy, meaning that their effect on the mobility is less pronounced. The approximated dispersion relationship in Eq. 5.17 is compared in Fig. 5.8 with the lowest mode of the complete numerical solution, for various values of the $t_{\rm ITL}$. Fig. 5.9 shows the same comparison as in Fig. 5.8, but for various values of the $t_{\rm HK}$. We see that Eq. 5.17 reproduces well the features of the lowest mode in the full numerical solution.

Proceeding in a similar way as in Sec. 5.2.2, we equate the classical and the quantum mechanical phonon energy (as in [48]), and we express the classical electromagnetic wave as in Eq. 5.7. Thus, the amplitude of the perturbation potential in the Si channel is found to be [154, 48, 78]:

$$\phi_{\rm SO}(q, z) = \sqrt{\frac{\hbar\omega_{\rm SO}A}{2q}\left(\frac{1}{\alpha\epsilon_{\rm i} + \beta} - \frac{1}{\alpha\epsilon_0 + \beta}\right)}e^{-qz} \tag{5.19}$$

where:

$$\begin{aligned} \alpha \;=\; & \left[\left(\frac{\epsilon_{\rm ITL} - \epsilon_{\rm Si}}{2\epsilon_{\rm ITL}}\right)^2 e^{-2qt_{\rm ITL}} + \right. \\ & \left. + \left(\frac{\epsilon_{\rm ITL} + \epsilon_{\rm Si}}{2\epsilon_{\rm ITL}}\right)^2 e^{2qt_{\rm ITL}} + 2\frac{\epsilon_{\rm ITL}^2 - \epsilon_{\rm Si}^2}{(2\epsilon_{\rm ITL})^2}\right]\frac{1 + e^{-2qt_{\rm HK}}}{1 - e^{-2qt_{\rm HK}}} \end{aligned} \tag{5.20}$$

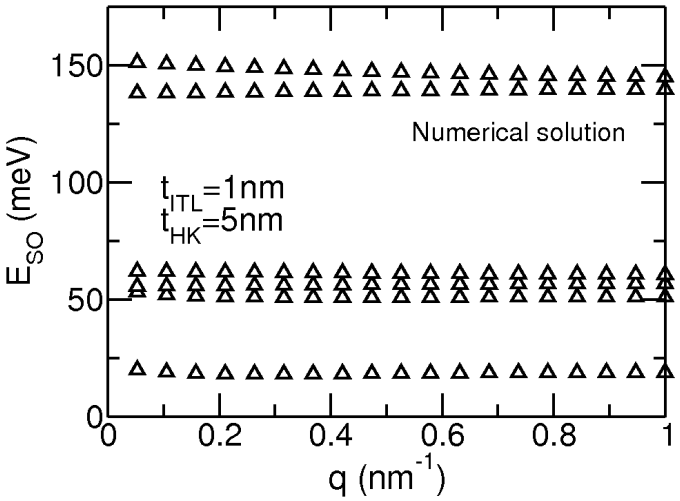

Figure 5.7: Modes of the full dispersion relationship of the SOph modes in stacks featuring an HfO_2 high-k layer with $t_{\rm ITL}$=1 nm (SiO_2) and $t_{\rm HK}$=5 nm. The numerical model accounts for the two TO modes in the ITL as well as the two TO modes in the high-k layer. .

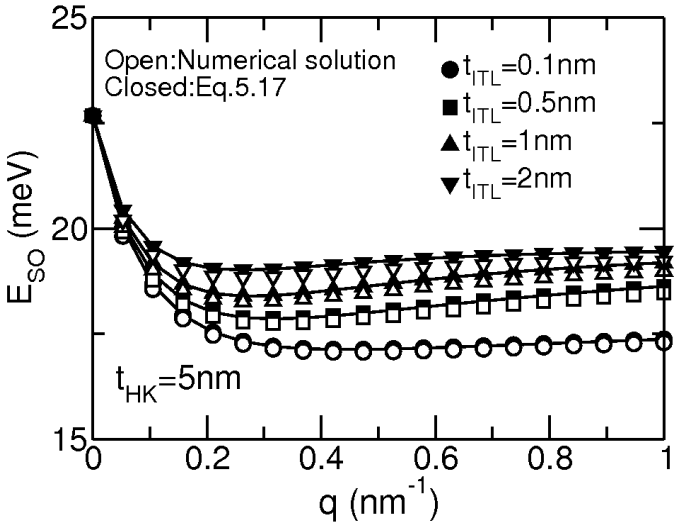

Figure 5.8: Open symbols: lowest mode of the full dispersion relationship of the SOph modes in stacks featuring an HfO_2 high-k layer with $t_{\rm HK}$=5 nm and various values of $t_{\rm ITL}$. The numerical model accounts for the two TO modes in the ITL as well as the two TO modes in the high-k layer. Closed symbols: lowest mode (originating from the TO1 mode in the HfO_2) as obtained from Eq.5.17. Similar results have been found over a wide range of $t_{\rm HK}$ values

.

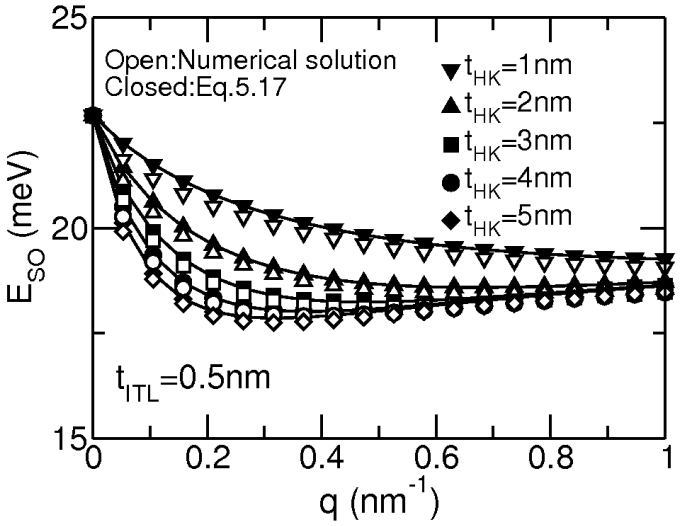

Figure 5.9: Same as in Fig. 5.8, but for different $t_{\rm HK}$ and fixed $t_{\rm ITL}$.

and:

$$\beta = \frac{(\epsilon_{\rm ITL}+\epsilon_{\rm Si})^2}{4\epsilon_{\rm ITL}} e^{2qt_{\rm ITL}} - \frac{(\epsilon_{\rm ITL}-\epsilon_{\rm Si})^2}{4\epsilon_{\rm ITL}} e^{-2qt_{\rm ITL}} + \epsilon_{\rm Si} \tag{5.21}$$

The q dependence of $\omega_{\rm SO}$ (and thus of the phonon energy) significantly complicates the computation of the state-after-scattering in the MSMC. In the following we will use an effective average phonon energy $\hbar\omega_{\rm SO,ave}$ (averaged over the 0.01 nm^{-1}<q<12 nm^{-1} interval) in the computation of the state-after-scattering. Fig. 5.10 reports simulations performed using the full $\omega_{\rm SO}(q)$ relationship in the computation of the state-after-scattering, instead of $\omega_{\rm SO,ave}$, showing that the approximation of a q independent phonon energy has a negligible influence on $\mu_{\rm eff}$, whereas the simulation time can be 10 times shorter.

Fig.5.10 also shows the effect of the ITL thickness on $\mu_{\rm eff}$: when $t_{\rm ITL}$ becomes larger than about 2 nm, the mobility reduction due to SOph vanishes. While the effect of $t_{\rm ITL}$ on $\mu_{\rm eff}$ is significant, Fig. 5.11 shows that the impact of $t_{\rm HK}$ is very weak when $t_{\rm HK}$>2 nm. Otherwise the screening induced by the metal gate reduces the perturbation potential of the SOph, increasing the mobility. Notice that in the simulations of Figs. 5.10 and 5.11 the effect of SOph is magnified by the fact of neglecting Coulomb scattering due to ionized dopants in the channel (the dominant mechanism at low inversion charge).

5.3 Modeling RemQ scattering

The perturbation potential of the RemQ scattering strongly depends on the thickness and on the dielectric constant of the materials in the gate stack as well as on the screening produced by the metal gate and by the inversion layer [106].

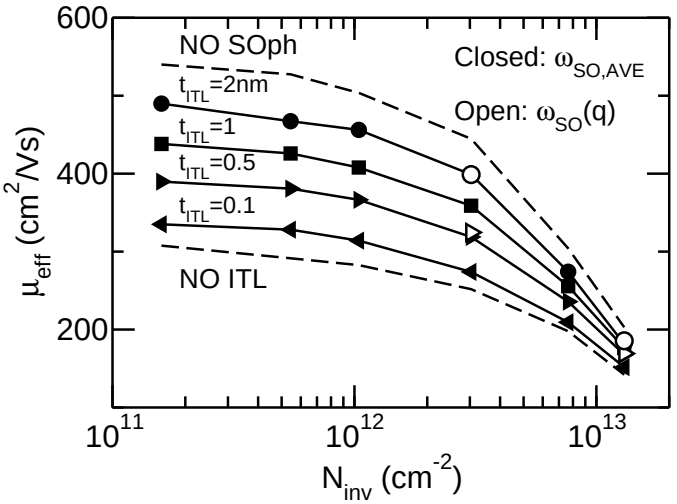

Figure 5.10: Simulated effect of $t_{\rm ITL}$ on the effective mobility when SOph scattering is considered in addition to Si phonons and surface roughness. Bulk device with doping $N_{\rm A}$=3×10^{17} cm^{-3} and $t_{\rm HK}$=5 nm (HfO_2). Closed symbols: $\omega_{\rm SO,ave}$ used in the determination of the state-after-scattering averaged in the range of q from 0.01 nm^{-1} to 12 nm^{-1}. Open symbols: $\omega_{\rm SO}(q)$ (Eq.5.17) is used.

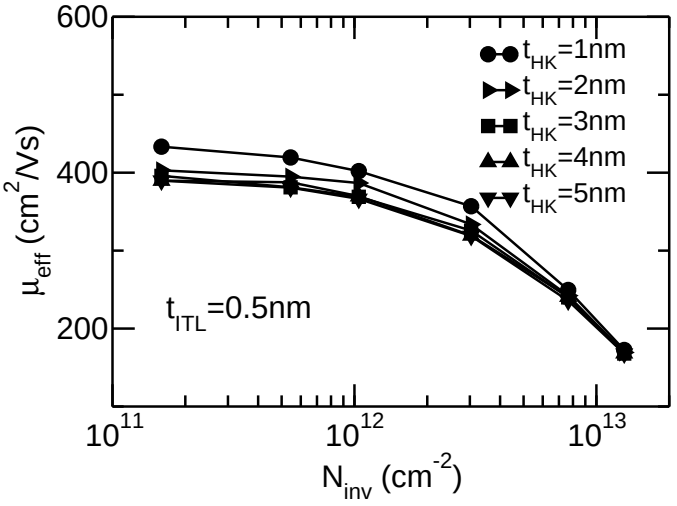

Figure 5.11: Simulated effect of $t_{\rm HK}$ on the effective mobility when SOph scattering is considered in addition to Si phonons and surface roughness. Bulk device with doping $N_{\rm A}$=3×10^{17} cm^{-3} and $t_{\rm HK}$=5 nm (HfO_2).

In this work we have first calculated the scattering potential produced by a point charge accounting only for the metal gate screening.

The potential produced by a point charge placed at the position (0,0,z_0) can be found by solving the two-dimensional transformed Poisson equation, that is:

$$\left[\frac{\partial^2}{\partial z^2} - q^2\right]\Phi(q,z) = -\frac{e}{\epsilon}\delta(z-z_0) \tag{5.22}$$

Eq. 5.22 is a particular version of the Poisson equation, in which we have applied the spatial Fourier transform in the transport plane normal to the vertical direction z. We consider a one-dimensional problem since, as already said, each x section is represented as a MOS infinitely long and large (with no electric field applied in the transport plane). Thus the problem is uniform in the x and y directions.

The general solution of the Eq. 5.22, which is the unscreened scattering potential in the i-th layer of the stack (namely the Si, ITL or HK layer, see Fig.5.6) is given by [106, 161, 78]:

$$\phi_{\mathrm{RemQ},i}(q,z,z_0) = \frac{e}{2\epsilon_i q}e^{-q|z-z_0|} + B_{i,1}e^{-q\,z} + B_{i,2}e^{q\,z} \tag{5.23}$$

where $B_{i,1}$ and $B_{i,2}$ are two constants to be determined. In the Si substrate we have $B_{i,2}$=0, whereas in the high-κ layer a relation between the two constants is obtained by setting $\phi(q,z,z_0)$=0 at the HK to metal gate interface. Due to this boundary condition, the mobility depends on the HK layer thickness. Comparing the mobility obtained assuming an infinitely thick HK material with a case with a negligible thickness of the HK material, we verified that this boundary condition does not affect too much the mobility in realistic devices. In this respect, Fig. 5.12 shows the electron mobility for the device in [31] with t_{ITL}=1 nm, for different values of the high-k dielectric thickness t_{HK} between 2 nm and 10 nm. We can see that the ideal screening of the gate can enhance the mobility up to 10% in real devices, thus confirming the validity our models. The four remaining constants in the HK, ITL and Si region are determined by imposing the continuity of the potential $\phi(q,z,z_0)$ and of the electric displacement field $\epsilon\partial\phi(q,z)/\partial z$ at the ITL/HK and SI/ITL interfaces [106]. The matrix element $M_{i,j}(q,z_0)$ (for a fixed charge at the position (0,0,z_0)) is then obtained using Eq. 3.21, as:

$$M_{i,j}(q,z_0) = \frac{e}{A}\int_0^\infty \xi_j(z)\phi_{\mathrm{RemQ}}(q,z,z_0)\xi_i(z)dz \tag{5.24}$$

Since the positions of the charges are assumed to be uncorrelated, the overall matrix element accountong for all the Coulomb centers in the device can be expressed as:

$$|M_{i,j}(q)|^2 \simeq A\int_{-t_{\mathrm{ITL}}-t_{\mathrm{HK}}}^\infty N_{z_0}|M_{i,j}(q,z_0)|^2 dz_0 \tag{5.25}$$

where N_{z_0} is the concentration of fixed charges at the coordinate z_0. In detail, we

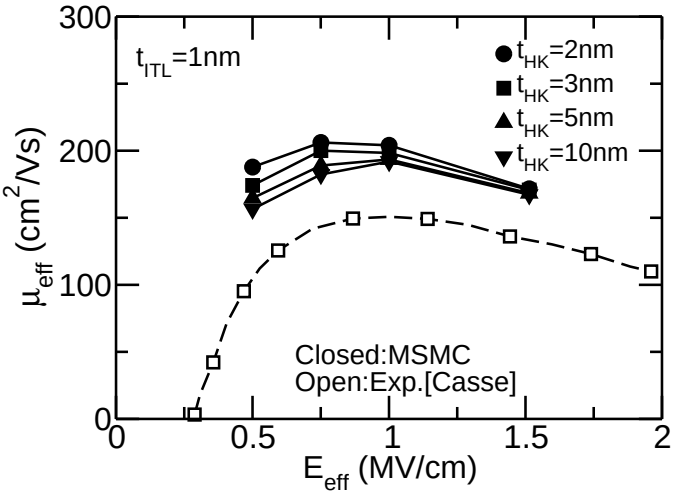

Figure 5.12: Electron mobility vs. effective field for the bulk device in [31] with $t_{\rm ITL}$=1 nm and $t_{\rm HK}$=3 nm at T=300 K. The value of the high-k dielectric thickness $t_{\rm HK}$ is assumed to vary between 2 nm and 10 nm in the simulations. The phonons are calibrated as in Fig. 3.8. The $N_{\rm Si/SiO_2}$ is 2×10^{10} cm^{-2}. The Δ for the surface roughness is 0.5 nm, while Λ is the same as in Tab. 3.1.

obtain:

$$\begin{aligned}
|M_{i,j}(q)|^2 \;\simeq\;& A\int_{-t_{\rm HK}}^{-t_{\rm ITL}} N_{\rm HK}|M_{i,j}(q,z_0)|^2dz_0 + \\
+\;& A\cdot N_{\rm ITL/HK}|M_{i,j}(q,z_0=-t_{\rm ITL})|^2 + \\
+\;& A\int_{-t_{\rm ITL}}^{0} N_{\rm ITL}|M_{i,j}(q,z_0)|^2dz_0 + \\
+\;& A\cdot N_{\rm Si/ITL}|M_{i,j}(q,z_0=0)|^2 + \\
+\;& A\int_{0}^{\infty} N_{\rm A}|M_{i,j}(q,z_0)|^2dz_0
\end{aligned} \tag{5.26}$$

It is woth noting that $N_{\rm HK}$, $N_{\rm ITL}$ and $N_{\rm A}$ are volumetric concentrations, while $N_{\rm ITL/HK}$ and $N_{\rm Si/ITL}$ are areal densities.

Finally, we have used the scalar dielectric function approach to include the screening of the intra-subband transitions produced by the inversion charge [78]:

$$|M^{\rm scr}_{i,i}(q)|^2 = \frac{|M_{i,i}(q)|^2}{\epsilon(q)^2} \tag{5.27}$$

where $M^{\rm scr}_{i,i}(q)$ is the screened matrix element in the i-th subband and $\epsilon(q)$ is the dielectric function [78, 106]. We have shown in Chap. 4 that this simplified model for the screening can be safely used in bulk structures [162].

As a sample result, Fig. 5.13 shows the effect on the electron mobility of the ITL thickness, considering a layer of fixed charge at the HfO_2/SiO_2 interface with

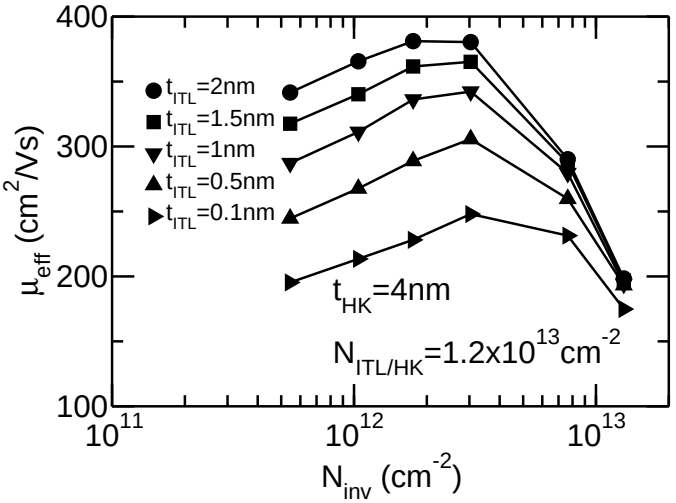

Figure 5.13: Simulated effect of $t_{\rm ITL}$ on the electron mobility when remote Coulomb scattering is activated (in addition to Si phonons and surface roughness calibrated as in Fig. 3.8). The charge (1.2×10^{13} cm^{-2}) is placed at the SiO_2/HfO_2 interface. Bulk device with $N_{\rm A}$=3×10^{17} cm^{-3} and $t_{\rm HK}$=4 nm.

density equal to 1.2×10^{13} cm^{-2} (see Fig. 5.15). The mobility reduction induced by the fixed charge increases for decreasing $t_{\rm ITL}$ because the charge gets closer to the channel.

Fig. 5.14 show the effect on the electron mobility of the HK layer thickness, considering again a layer of fixed charge at the HfO_2/SiO_2 interface with density equal to 1.2×10^{13}cm^{-2}. Differently from the SOph case, the mobility is also sensitive to the HK layer thickness $t_{\rm HK}$, since a thinner HK layer leads to a better screening produced by the metal gate.

However, the mobility dependence on $t_{\rm ITL}$ and $t_{\rm HK}$ of Figs. 5.10 and 5.11 is magnified by the fact that no charges at the Si/ITL have been assumed. In fact, Fig. 5.12, that has been obtained with a realistic value of $N_{\rm Si/ITL}$ of a modern MOSFET, shows that in fact, the presence of the metal gate deas not affect significantly the mobility.

5.4 Modeling DipQ scattering

The two-dimensional transformed Poisson equation describing the perturbation potential of one of the vertical dipoles in Fig. 5.15 is:

$$\left(\frac{\partial^2}{\partial z^2}-q^2\right)\Phi(q,z)=-\frac{e}{\epsilon}[\delta(z-z_{q1})-\delta(z-(z_{q1}+d_{\rm dipQ}))] \qquad (5.28)$$

Due to the linearity of the laplacian operator, the solution of Eq. 5.28 can be found by applying the superposition principle. Thus, the solution of Eq. 5.28 is the sum

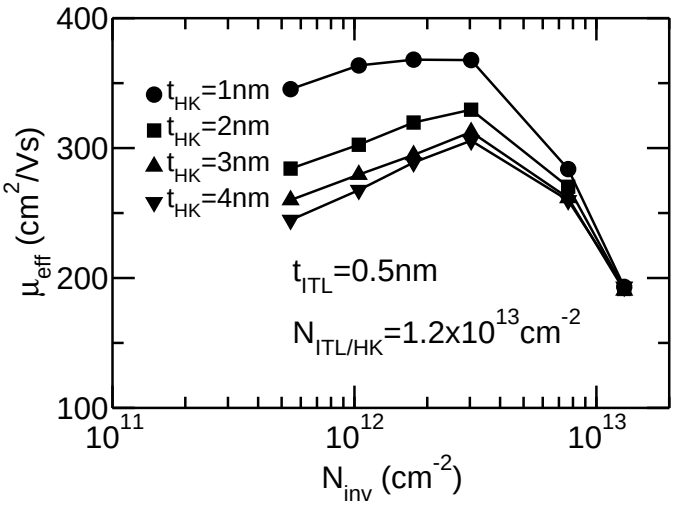

Figure 5.14: Same as in Fig. 5.13, but for varying t_{HK} with a fixed t_{ITL}.

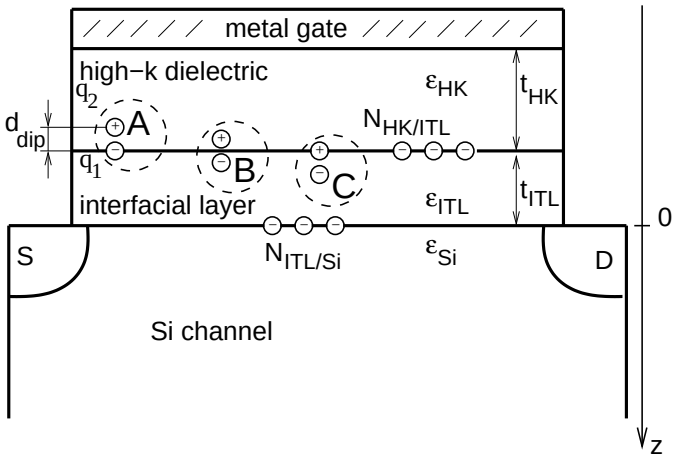

Figure 5.15: The considered positions (A, B, C) for the dipole, for the RemQ and Si/ITL charges at the interfaces are shown. The $-t_{\mathrm{ITL}}$ is equal to the positions z_{q1} and z_{q2} in the configuration A and C, respectively. In the configuration B, instead, the charges q_1 and q_2 are placed at $-t_{\mathrm{ITL}} \pm \frac{d_{\mathrm{dipQ}}}{2}$.

of the transformed scattering potentials of the single charges (Eq. 5.23) constituting the dipole, as:

$$\phi_{DipQ}(q,z) = \frac{e}{2q\epsilon_{\mathrm{Si}}} e^{-q|z-z_{q1}|}(1-e^{-qd_{\mathrm{dipQ}}}) + B_q e^{-qz} \tag{5.29}$$

where d_{dipQ} is the distance between the charges $q1$ (placed at z_{q1}) and $q2$ of the dipole (see Fig. 5.15); B_q is a new constant stemming from the sum of the scattering potentials of the charges q_1 and q_2 of the dipole that can be expressed in terms of the $B_{q_i,1}$ coefficients in Eq. 5.23 as:

$$B_q = B_{q_1,1} - B_{q_2,1} \tag{5.30}$$

After using Eq. 3.19, which gives the matrix element assuming a single dipole $M_{i,j}^{\mathrm{single}}(q)$, the overall matrix element for the DipQ scattering can be obtained as:

$$|M_{i,j}(q)|^2 \simeq AN_{\mathrm{dipQ}}|M_{i,j}^{\mathrm{single}}(q)|^2 \tag{5.31}$$

As in the RemQ case, the screening produced by the inversion layer has been included in the evaluation of the scattering rate by using Eq. 5.27.

As a sample result, Fig. 5.16 shows the effect on the electron mobility of the distance between the charges of the dipole d_{dipQ}, considering a layer of dipoles at the HfO_2/SiO_2 interface with density equal to 1×10^{15} cm^{-2}. The dipole in figure has one charge at the ITL/HK interface and the other one inside the HK layer. (A-position in Fig. 5.6). The mobility reduction induced by the fixed charge increases for decreasing t_{ITL} because the charge gets closer to the channel. The results are in between the case with d_{dipQ}=0 (the effect of one charge is cancelled by the other one) and d_{dipQ}=t_{HK} (corresponding to the mobility degradation produced by the RemQ mechanism since q_2 is completely screened by the metal gate).

5.5 Comparison with experimental mobility data

In this section we compare a large number of experimental mobility data for HfO_2 gate stacks (see Tab. 5.2) with MSMC simulations in order to assess the effect of the SOph, RemQ and DipQ scattering when simulating realistic n- and p-MOSFET devices fabricated by different companies and labs.

5.5.1 Calibration of the models

In Sec. 3.3, the simulation models had been calibrated on the universal mobility curves [47], as described in [93, 100].

However, the reference devices with SiO_2 dielectric corresponding to the Bulk-A, Bulk-C and Bulk-D devices (see [31, 164, 30]) exhibit lower mobility than the universal curves. Thus, before simulating the corresponding transistors featuring HfO_2, we calibrated the parameters $N_{\mathrm{Si/SiO_2}}$, Δ_{SR} and Λ_{SR} on the specific SiO_2 reference devices.

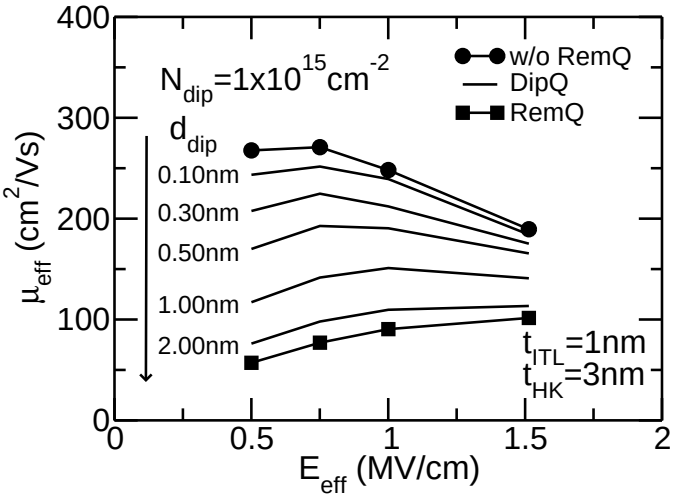

Figure 5.16: Simulated effect of d_{dipQ} on the electron mobility when DipQ is activated (in addition to Si phonons, surface roughness and SOph scattering mechanisms). The DipQ (1×10^{15} cm^{-2}) is placed in the A-position (see Fig. 5.15), close to the SiO_2/HfO_2 interface. Bulk device with N_{A}=2×10^{17} cm^{-3}, t_{ITL}=1 nm and t_{HK}=3 nm.

	from	t_{ITL} (nm)	t_{HK} (nm)	channel doping (cm^{-3})	type
nBulk-A pBulk-A	[31]	1 - 2	3	2×10^{17}	n-bulk p-bulk
nBulk-B	[163]	1	1.6 - 3	3×10^{17}	n-bulk
nBulk-C	[164]	0.8	3	1×10^{17}	n-bulk
nBulk-D	[30]	1	2.54	2×10^{17}	n-bulk
nSG-E pSG-E	this work	0.9	3	1×10^{15}	n-SG-SOI p-SG-SOI

Table 5.2: Summary of the devices simulated in this work. The experimental data for the bulk devices are taken from literature, whereas the SG-SOI devices have been measured in this work.

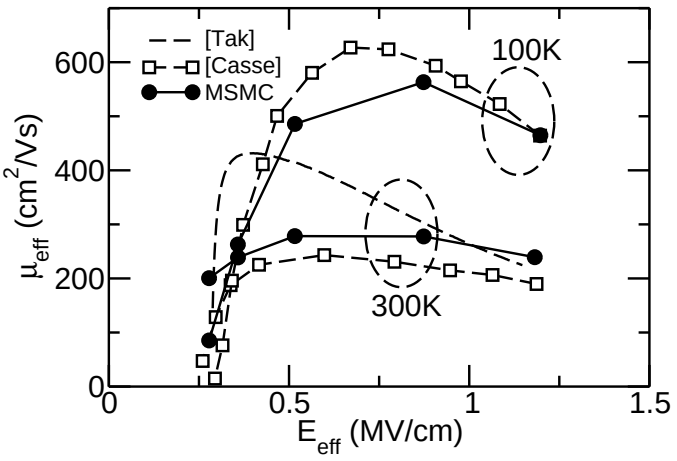

Figure 5.17: Calibration of the MSMC model (filled symbols) on the reference bulk SiO_2 device (t_{SiO_2}=2.5 nm, N_A=2×10^{17} cm^{-3}) with TiN gate of the nBulk-A devices of [31] (open symbols). The charge at the Si/SiO_2 interface is 2.5×10^{12} cm^{-2}. The r.m.s. value of the surface roughness spectrum (gaussian) has been reduced to 0.5 nm from the 0.62 nm used to reproduce the universal curve [93].

Fig. 5.17 shows the calibration on the SiO_2 control of the nBulk-A devices of [31]. Starting from the set of parameters that reproduces the universal mobility curves of [47] with SiO_2 (see [93]), and following an empirical approach, we have slightly decreased the r.m.s. value of the surface roughness, based on the observation that the slope of mobility of the metal gate devices in [31] is lower than the universal curve (that refers to Poly-Si gate) for large $E_{\rm eff}$. Then, we have adjusted the concentration of interface states at the Si/SiO_2 interface, obtaining the good agreement with the experiments at room and low temperature reported in Fig. 5.17. We have proceded in a similar way for the SiO_2 control of the pBulk-A devices (Fig. 5.18).

The calibration for the nBulk-C and nBulk-D cases is shown in Fig. 5.19 and 5.20, respectively (open circles vs. dashed line).

For the Bulk-B devices, instead, the SiO_2 reference device was not available. However, the nBulk-A and the nBulk-B devices have the same channel doping concentration and the same gate stack. So, for the purpose of calibration in this case, we compared the experimental mobility data of the HK devices with $t_{\rm ITL}$=1 nm and $t_{\rm HK}$=3 nm in [31] (nBulk-A) with the device in [163] (nBulk-B) and we attributed the discrepancy in terms of mobility to the quality of the Si/ITL interface. The two specified devices have a similar mobility at high effective field. Thus, for the simulations of the nBulk-B devices we use a value of the $\Delta_{\rm SR}$ parameter ($\Delta_{\rm SR}$=0.55 nm) similar to the one used in the simulation of the nBulk-A devices ($\Delta_{\rm SR}$=0.50 nm), that is lower than the $\Delta_{\rm SR}$=0.62 nm reproducing universal curves (Tab. 3.1). Moreover, the nBulk-B device has a higher peak mobility which sug-

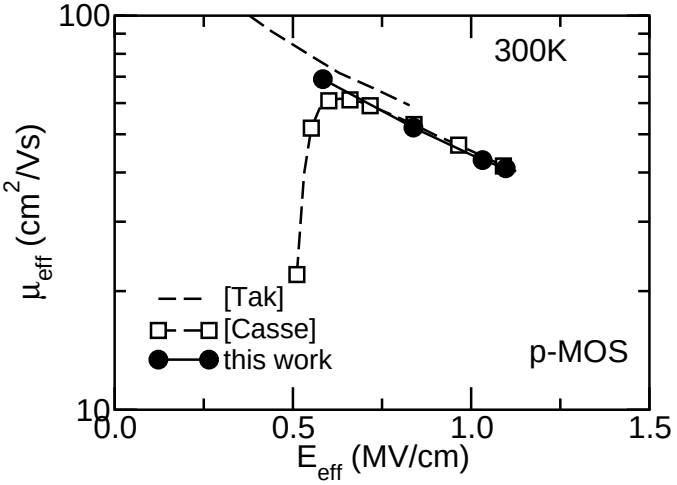

Figure 5.18: Calibration of the MSMC model (filled symbols) on the reference bulk SiO_2 device (t_{SiO_2}=2.5 nm, N_A=2×10^{17} cm^{-3}) with TiN gate of the pBulk-A devices of [31] (open symbols). The charge at the Si/SiO_2 interface is 5×10^{11} cm^{-2}. The values of the surface roughness spectrum (exponential) are Δ_{SR}=0.52 nm and Λ_{SR}=2.0 nm.

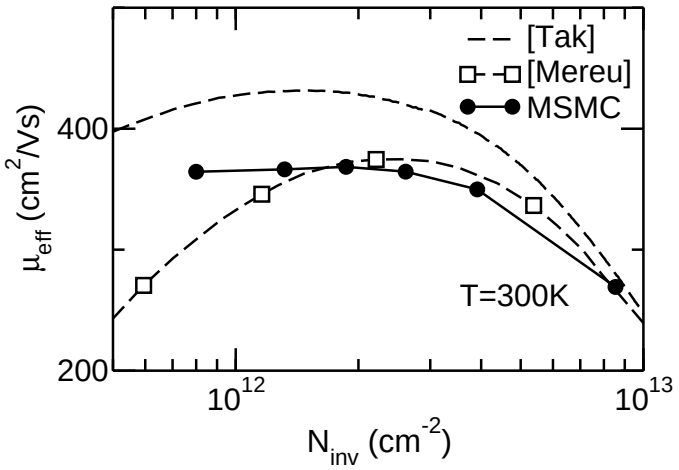

Figure 5.19: Calibration of the MSMC model (filled symbols) on the reference bulk SiO_2 device (N_A=3×10^{17} cm^{-3}) with Poly-Si gate of the nBulk-C devices of [164] (open symbols). The charge at the Si/SiO_2 interface is 5.2×10^{11} cm^{-2}. The r.m.s. value of the surface roughness spectrum has been augmented to 0.7 nm from the 0.62 nm used to reproduce the universal curve. The curve from [47] refers to a channel doping of 3×10^{17} cm^{-3}.

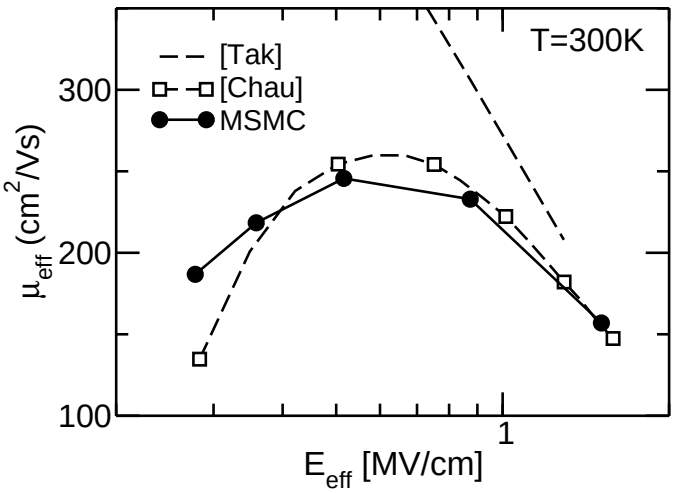

Figure 5.20: Calibration of the MSMC model (filled symbols) on the reference bulk SiO_2 device (N_A=3×10^{17} cm^{-3}) with TiN gate of the nBulk-D devices of [164] (open symbols). The charge at the Si/SiO_2 interface is 1.5×10^{12} cm^{-2}. The r.m.s. value of the surface roughness spectrum is 0.62 nm, as the one used to reproduce the universal curve. The curve from [47] refers to a channel doping of 3×10^{17} cm^{-3}.

gests a lower density of charged states at the Si/ITL interface. For these reasons, the simulations of the nBulk-B devices were performed using a very small concentration N_{Si/SiO_2}=2×10^{10} cm^{-2} (the same as in Tab. 3.1 for the devices in [47]).

Finally, for the SG-E devices, we calibrated the simulators for n- and p-MOSFETs on the universal mobility curves of [47] (using the parameter values in Tab. 3.1) since SiO_2 control devices were not available, but the same fabrication process produces bulk devices with mobility close to the universal curves [165].

For the sake of clarity, Tab. 5.3 summarizes the calibration parameters used for each set of devices.

5.5.2 Effect of the SOph and RemQ scattering

LETI data

The models described in Secs. 5.2.3 and 5.3 have been firstly used to simulate the nBulk-A and pBulk-A devices measured in [31]. After having calibrated the simulator on the correspondent SiO_2 reference device (previous section), we have considered the devices featuring a HfO_2 dielectric. Fig.5.21 shows that SOph scattering has a very weak influence on $\mu_{\rm eff}$. In fact the $\mu_{\rm eff}$ reduction observed in Fig.5.10, for devices without ITL and with no Si/SiO_2 charges, is here masked by the lower mobility of the SiO_2 control, due to Coulomb centers in the doped Si substrate and at the Si/SiO_2 interface.

In order to reproduce the mobility reduction induced by the high-κ, we have to

	$N_{\mathrm{Si/SiO_2}}$ (cm^{-2})	Δ_{SR} (nm)	Λ_{SR} (nm)
nBulk-A	2.5×10^{12}	0.50	1.0
pBulk-A	5×10^{11}	0.52	2.0
nBulk-B	2×10^{10}	0.55	1.0
nBulk-C	5.2×10^{11}	0.70	1.0
nBulk-D	2.5×10^{12}	0.62	1.0
nSG-E	2×10^{10}	0.62	1.0
pSG-E	2×10^{10}	0.56	2.6

Table 5.3: Summary of the parameters used to simulate the devices in Tab. 5.2 which have been found calibrating the simulations on the respective SiO_2 control devices (Figs. 5.17, 5.18, 5.19 and 5.20).

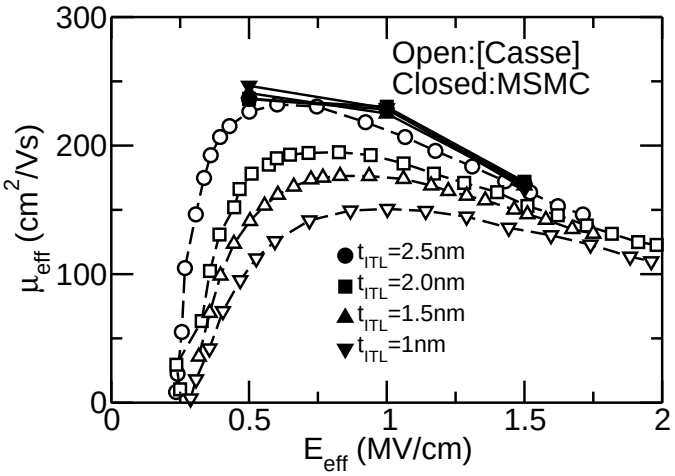

Figure 5.21: Mobility from MSMC simulations with SOph scattering (filled symbols) vs. experimental data of the nBulk-A devices [31] for HfO_2 stacks with different t_{ITL}. Temperature is 300 K.

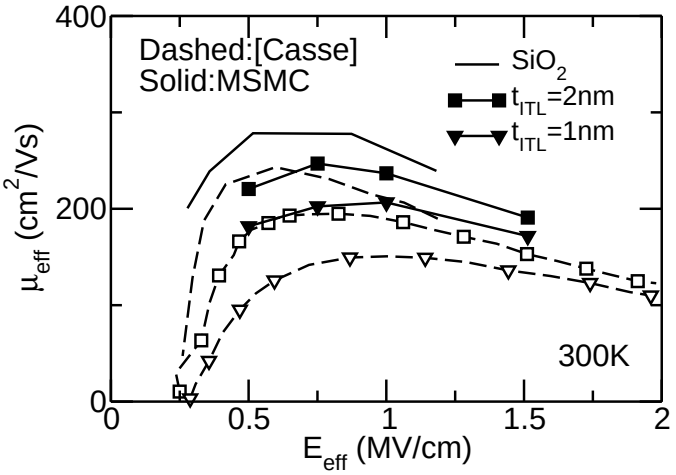

Figure 5.22: Mobility from MSMC simulations with SOph and RemQ ($N_{\mathrm{ITL/HK}}$=9×10^{13} cm^{-2}) scattering (filled symbols) vs. experimental data of the nBulk-A devices [31] for HfO_2 stacks with different t_{ITL}. Temperature is 300 K.

activate the RemQ scattering mechanism, assuming a large concentration of charge at the ITL/HfO_2 interface which is equal to $N_{\mathrm{ITL/HK}}$=9×10^{13} cm^{-2}, as shown in Figs.5.22 and 5.23. These findings confirm the results obtained in [31] for the n-MOS, where the $N_{\mathrm{ITL/HK}}$ was estimated to be 7×10^{13} cm^{-2} using a 2 subband momentum relaxation time model for RemQ.

We have proceded in a similar way for the p-MOS case. Starting with the simulator calibrated on the mobility curve of the SiO_2 control p-MOS device of [31] (Fig. 5.18 dashed line in Fig.5.24), the SOph of HfO_2 have been activated but Fig.5.24 shows that the mobility reduction is much smaller than in the experiments for HfO_2 devices. The hole mobility reduction induced by the HK measured on the pMOS-A devices of [31] can be reproduced with the same large charge concentration at the ITL/HK interface (see in Fig.5.24) necessary in the n-MOS devices (Fig. 5.22). Thus we can conclude that the experimental mobility reduction induced by the high-k dielectric can be reproduced by using the same concentration of charges for the nMOS and pMOS transistors ($N_{\mathrm{ITL/HK}}$=9×10^{13} cm^{-2}), whereas SOph alone have a negligible impact on the mobility.

Here we extend the analysis to data from different sources in order to assess the effect of SOph and RemQ and the general validity of the previous findings in many realistic devices.

Tyndall data

Fig. 5.25 shows simulations of the nBulk-B devices [163]. SOph scattering has a very weak influence on μ_{eff} for the HfO_2 samples, as for the devices in [31] as seen in the previous paragraph. The figure also reports the curve obtained accounting for the RemQ mechanism, that approaches the experimental mobility

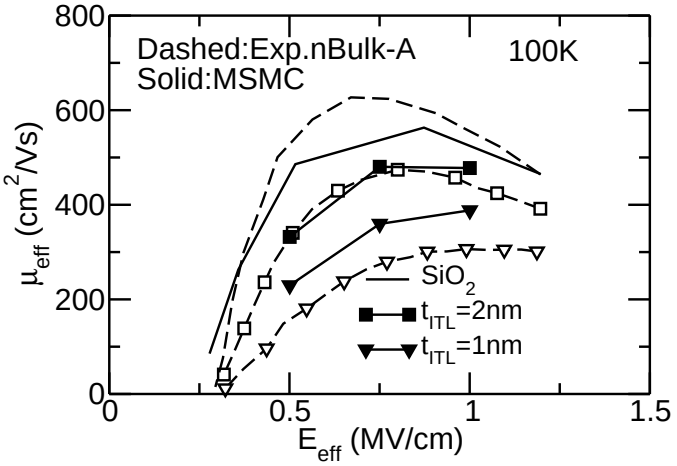

Figure 5.23: Same as in Fig. 5.22, but for a temperature of 100 K.

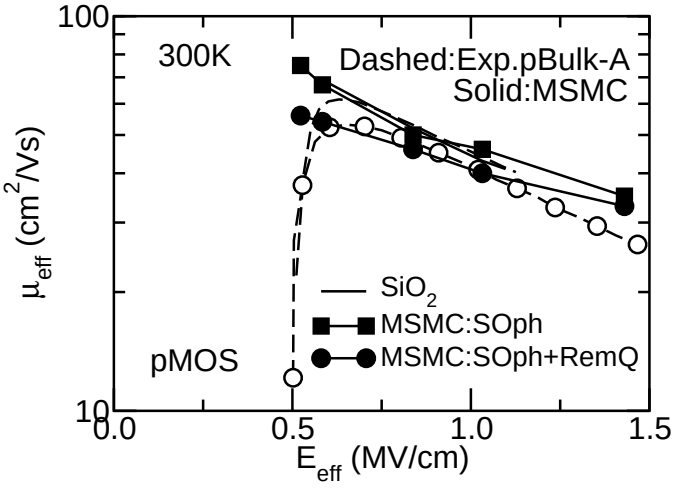

Figure 5.24: Comparison between simulated hole mobility accounting for SOph mechanism (filled squares), for the SOph plus RemQ ($N_{\mathrm{ITL/HK}}$=9×10^{13} cm^{-2}, filled circles) and experimental data (open circles) for the pBulk-A device with t_{ITL}=1 nm and t_{HK}=3 nm in [31]. The doping of the devices is N_{D}=2×10^{17} cm^{-3}.

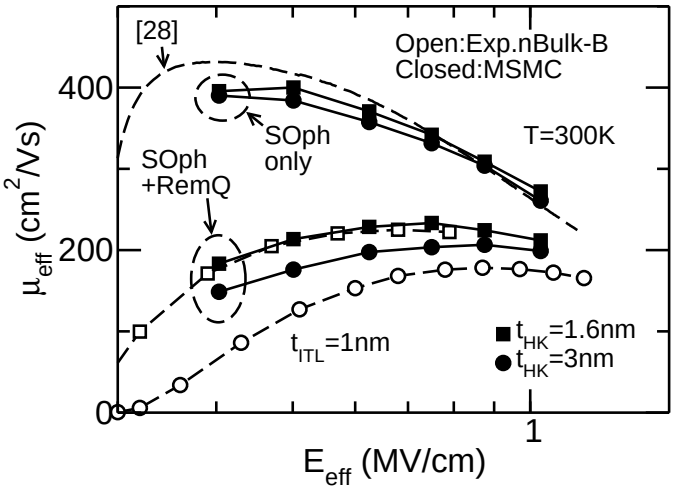

Figure 5.25: Comparison between simulated electron mobility accounting for the SOph and RemQ ($N_{\mathrm{ITL/HK}}$=1×10^{14} cm^{-2}) mechanisms (filled symbols) and experimental data (open symbols) for the HfO_2 devices with t_{HK}=1.6 nm and t_{HK}=3 nm in [163] (t_{ITL}=1 nm). The scattering mechanisms explained in Sec. 3.2 are active in all the MSMC curves. The curve from [47] refers to a channel doping of 3×10^{17} cm^{-3}.

by assuming a concentration of RemQ centers $N_{\mathrm{ITL/HK}}$=10^{14} cm^{-2}, which is very close to the one assumed in the nBulk-A and pBulk-A transistors.

IBM data

Concerning the nBulk-C devices, Fig. 5.26 shows that, when activating the HfO_2 SOph scattering mechanism only (filled squares), results are far from the experimental data. Once again, in order to reproduce the experimental mobility reduction, we have to activate the RemQ mechanism, assuming a charge density of at the ITL/HK interface $N_{\mathrm{ITL/HK}}$=4×10^{13} cm^{-2} which is fairly close to the value used for the nBulk-A, pBulk-A and nBulk-B devices.

Intel data

The simulation of the nBulk-D device (ref. [30]) has been slightly more complicated. In fact the t_{ITL} value was not available, whereas the EOT was given instead. We have assumed a realistic value for the t_{ITL}, namely 1 nm, and we thus obtained the t_{HK} (2.54 nm) from the EOT (assuming ϵ_{SiO_2}=3.9 and ϵ_{HfO_2}=22). Fig.5.27 shows that, once again, SOph scattering has a very weak influence on μ_{eff} while a large concentration of fixed charges at the ITL/HK interface helps approach the experimental mobility curve. Also for this device, the concentration of RemQ scattering centers ($N_{\mathrm{ITL/HK}}$=5×10^{13} cm^{-2}) is close to the value used to repro-

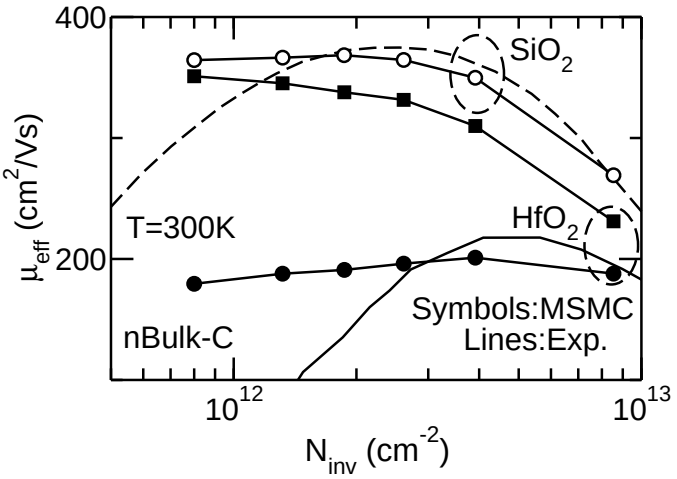

Figure 5.26: Comparison between simulated electron mobility accounting for the SOph mechanism (filled squares), for the SOph and the RemQ ($N_{\mathrm{ITL/HK}}$=4×10^{13} cm^{-2}) mechanisms (filled circles) and experimental data of [164] (solid line). HfO_2 devices with t_{ITL}=0.8 nm and t_{HK}=3 nm. The scattering mechanisms explained in Sec. 3.2 are active in all the MSMC curves.

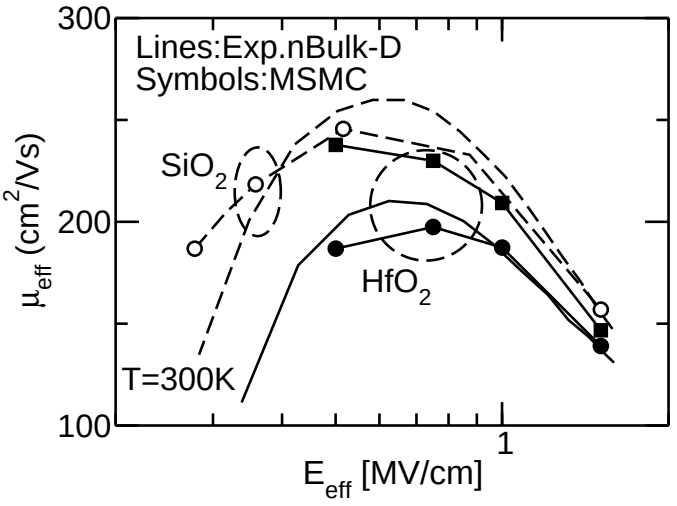

Figure 5.27: Comparison between simulated electron mobility accounting for the SOph mechanism (filled squares), for the SOph and the RemQ ($N_{\mathrm{ITL/HK}}$=5×10^{13} cm^{-2}) mechanisms (filled circles) and experimental data of [30] (solid line). The scattering mechanisms explained in Sec. 3.2 are active in all the MSMC curves.

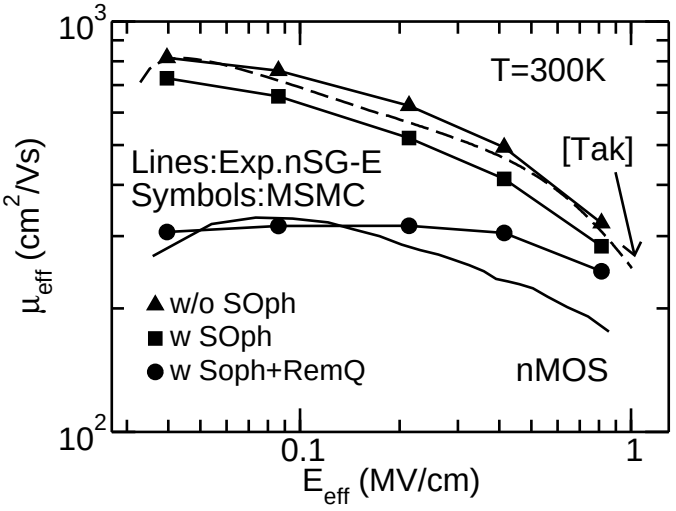

Figure 5.28: Comparison between the simulated electron mobility accounting for the SOph mechanism (squares), for the SOph and the RemQ mechanisms (circles) and experimental data (solid line) for undoped 12 nm thick SG-SOI n-device (nSG-E) with t_{ITL}=0.9 nm (SiO_2), t_{HK}=3 nm (HfO_2) and metal gate. The curves obtained excluding SOph and RemQ scattering mechanisms (triangles) are also shown. When the RemQ is actived, the concentration of charges is $N_{\mathrm{ITL/HK}}$=3×10^{13} cm^{-2}.

duce the measurements of the n- and pBulk-A, nBulk-B and nBulk-C devices in the previous paragraphs.

ST data

We have also analyzed SG-SOI n- and p-MOS devices to assess if a different electrostatic configuration, as the one in an SOI substrate, can change the effect of SOph and RemQ scattering mechanisms on the mobility. To this purpose, we have measured SG-SOI MOSFETs fabricated by ST Microelectronics Crolles with an undoped 12 nm thick Si channel (nSG-E and pSG-E). Figs. 5.28 and 5.29 compare measurements and simulation results accounting for SOph scattering whose effect is almost negligible, giving a mobility close to the universal curve for both the n- and p-MOSFETs, whereas the experimental mobility reduction with respect to the universal curves is significant. The same figure shows that, in order to approach the experimental mobility of the SG-SOI devices, RemQ densities of 3×10^{13} cm^{-2} and 6×10^{13} cm^{-2} are needed in the n- and p-MOS case, respectively. These values are once again consistent with those used for the simulations of the samples in the previous cases.

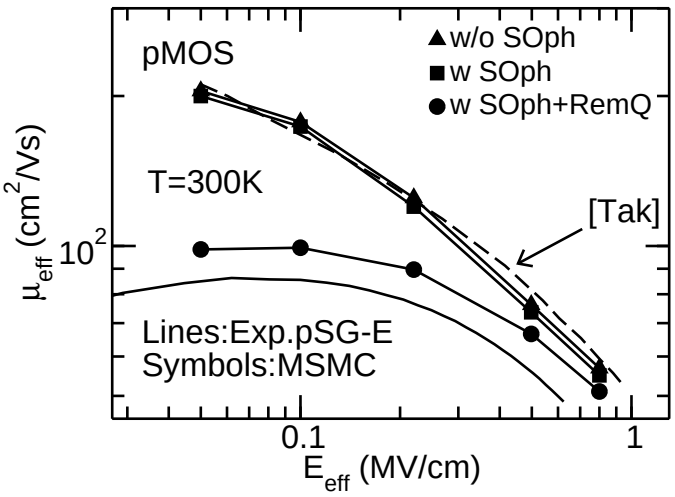

Figure 5.29: Same as in Fig. 5.28, but for the p-MOSFET (pSG-E). When the RemQ is actived, the concentration of charges is $N_{\mathrm{ITL/HK}}$=6×10^{13} cm^{-2}.

Other dielectrics

To show that the previous results are not peculiar to HfO_2 gate stack, we consider here other materials. Since the mobility reduction associated to SOph scattering mechanism is almost negligible, we can safely decide to neglect this scattering mechanism in the following simulations.

Fig. 5.30 shows the simulation of an undoped 10 nm thick SG-SOI transistor featuring a 2.5 nm thick HfSiON gate dielectric with t_{ITL}=1 nm [165]. The same figure shows also the simulation of a bulk device featuring $HfZrO_2$ (t_{HK} is 2.5 nm). The channel doping of the bulk device is 4×10^{17} cm^{-3} and the t_{ITL} value is 0.7 nm. We have calibrated our simulator as done for the nSG-E device, since all these devices have a similar fabrication process. We observe that the experimental data is reproduced assuming a concentration of RemQ charges $N_{\mathrm{ITL/HK}}$=2×10^{13} cm^{-2} and $N_{\mathrm{ITL/HK}}$=3×10^{13} cm^{-2} for the HfSiON and $HfZrO_2$ devices, respectively.

The concentration of charges that has to be assumed in order to reproduce experimental data for HfSiON and $HfZrO_2$ devices are very similar to those assumed for the HfO_2 devices, thus confirming that the results of this analysis are of general relevance for HK devices.

5.5.3 Effect of the correlation between the charges

We have seen that very large concentrations of Coulomb centers in the gate stack are needed in order to reproduce the experimental data.

With such concentrations of charges, the scattering potentials of the single charge (Eq. 5.23) can not be considered uncorrelated from each other due to the

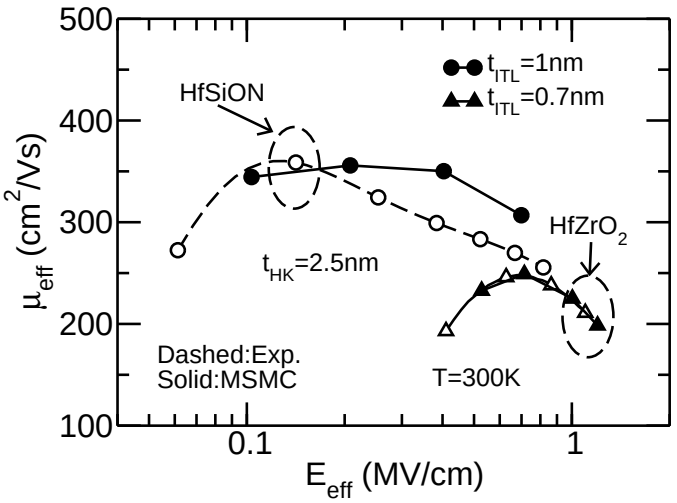

Figure 5.30: Comparison between simulated electron mobility accounting for the RemQ mechanism and experimental data for undoped 10 nm thick SG-SOI featuring HfSiON (open circles) and for a bulk device featuring $HfZrO_2$. The channel doping of the $HfZrO_2$ bulk devices is 4×10^{17} cm^{-3}.

proximity of the charges. In this condition, Eq. 5.25 may become unreliable and, according to the model presented in [166], it should be more appropriate to use:

$$|M_{i,j}(q)|^2 \simeq \frac{1}{A}|M_{i,j}(q,z_0)|^2 N_{z_0}\left[1-\frac{2CFJ_1(qR_c)}{qR_c}\right] \tag{5.32}$$

where J_1 is the first order Bessel function, CF is the correlation of the scattering potential between two adiacent Coulomb charges and R_c is a critical radius defined as:

$$R_c = \sqrt{\frac{CF}{\pi N_{z_0}}} \tag{5.33}$$

Fig. 5.31 shows the mobility dependence on the Correlation Factor (CF) between the charges at the ITL/HK interface. The simulated curves are obtained assuming no correlation (CF=0.0), partial correlation between the charges (CF=0.5) and total correlation between the charges (CF=1.0). We can see that the larger is the CF value, the lower is the effect of the RemQ charges on the mobility.

Finally, Fig. 5.32 shows the RemQ density that inserted in the model reproduce the experimental data of the devices in [31] at T=300 K as a function of the Correlation Factor CF. We can see that if we take into account the correlation, the RemQ density needed to reproduce the experimental data increases significantly for values of the CF larger than 0.5, and eventually diverge for values of the CF close to 1.

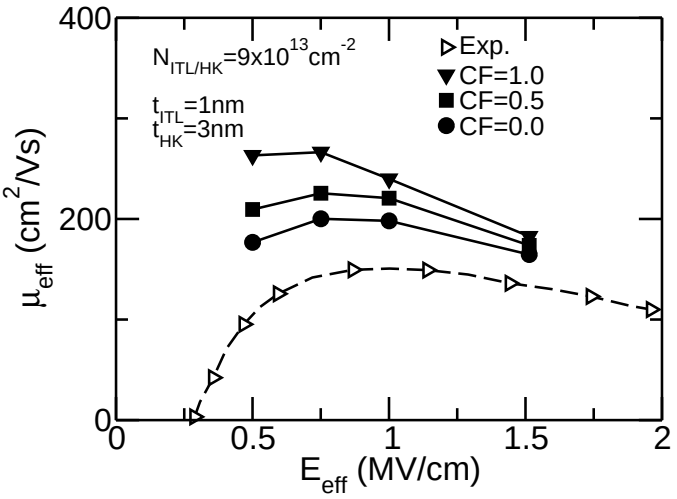

Figure 5.31: Electron mobility vs. effective field for the nBulk-A device with $t_{\rm ITL}$=1 nm and $t_{\rm HK}$=3 nm at T=300 K, for different values of the correlation factor (CF) between the RemQ charges. The RemQ density is the one allowing to reproduce experimental data ($N_{\rm ITL/HK}$=9×10^{13} cm^{-2}) with CF=0 as in Fig. 5.22.

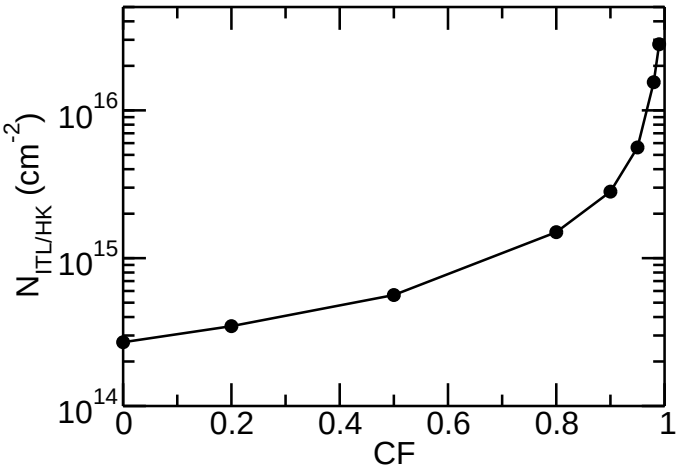

Figure 5.32: The concentration of RemQ charges needed to reproduce experimental data as a function of the Correlation Factor. The experimental data is taken from the nBulk-A device with $t_{\rm ITL}$=1 nm and $t_{\rm HK}$=3 nm at T=300 K.

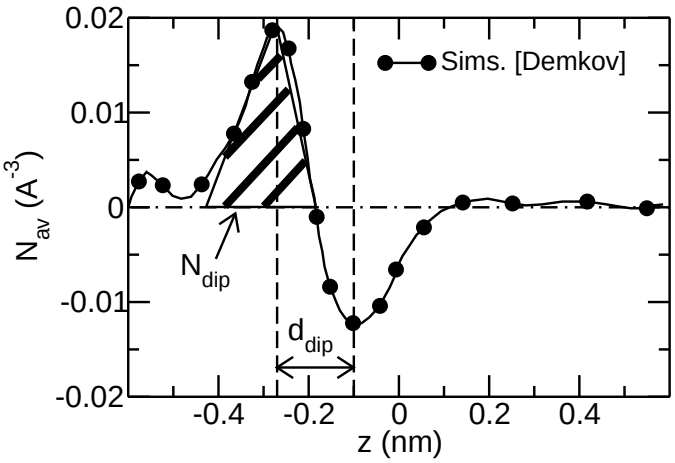

Figure 5.33: Average number of the electron N_{av} as a function of the vertical coordinate z. Simulated data from [55].

5.5.4 Effect of the DipQ scattering

Assessement from *ab initio* simulations

Recent *ab initio* studies concerning the HfO_2/SiO_2 interface have tried to assess the features of the dipole originating at that interface. From [55], we can extract the concentration and the distance between the dipole charges close to 2.8×10^{14} cm^{-2} and 0.175 nm, respectively. Fig. 5.33 shows the *ab initio* simulation results from [55] and the way we performed the extraction of the $N_{\rm dipQ}$ and $d_{\rm dipQ}$ from the simulated data of [55].

However, Fig. 5.34 clearly shows that the experimental mobility of the nBulk-A devices cannot be reproduced by simulation if we take into account the presence of the dipoles with the parameters suggested in [55], in addition to the SOph scattering mechanism. The situation is better evidenced in Fig. 5.35, where the simulation of the nBulk-A devices accounting for SOph and the DipQ (same parameters as in Fig. 5.34) is performed at T=100 K. The same holds for the other devices considered in this study (not shown).

Much larger DipQ concentrations and dipole distances than the one in Figs. 5.34 and 5.35 are needed to reproduce the experimental data.

Assessement from the density of HfO_2 atoms

The density of the $d_{\rm HfO_2}$ of the HfO_2 is 9.68 g/cm^3. The molar mass ($M_{\rm mol}$) of this dielectric is 210.49 g/mol. From the molar mass we can find the molecular mass of the HfO_2 $m_{\rm mol}$, which is:

$$m_{\rm mol} = \frac{M_{\rm mol}}{N_{\rm A}} = \frac{210.49\ {\rm g/mol}}{6.022\times10^{-23}\ {\rm mol}^{-1}} = 9.495\times10^{-22}\ {\rm g} \qquad (5.34)$$

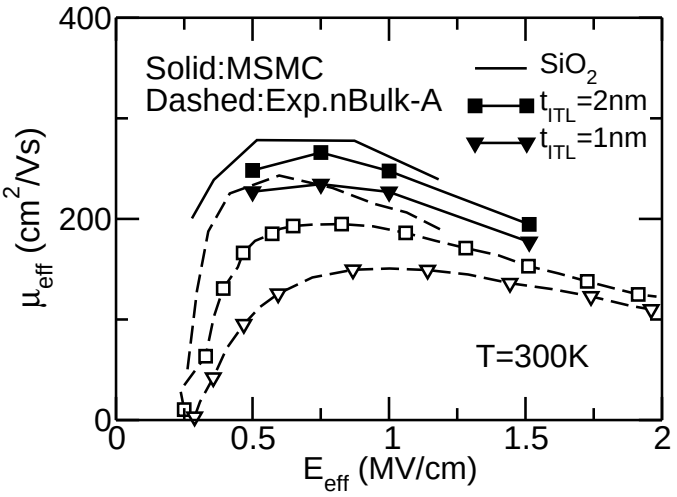

Figure 5.34: Comparison between simulated and experimental [31] electron mobility versus the effective field for 3 nm thick HfO_2 devices with $t_{\rm ITL}$=1 nm and $t_{\rm ITL}$=2 nm at T=300 K. Simulations have been obtained accounting for SOph and DipQ scattering mechanisms, as well as the mechanisms explained in Sec. 3.2. The concentration of the DipQ centres is $N_{\rm dipQ}$=2.8×10^{14} cm^{-2}–$d_{\rm dipQ}$=0.175 nm [55]. The experimental and simulated SiO_2 reference of [31] are also shown (solid and dashed lines, respectively).

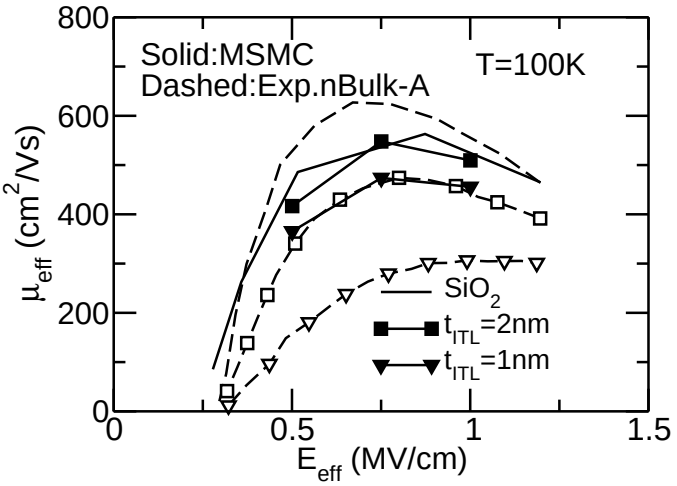

Figure 5.35: Same as Fig. 5.34, but for a temperature of 100 K.

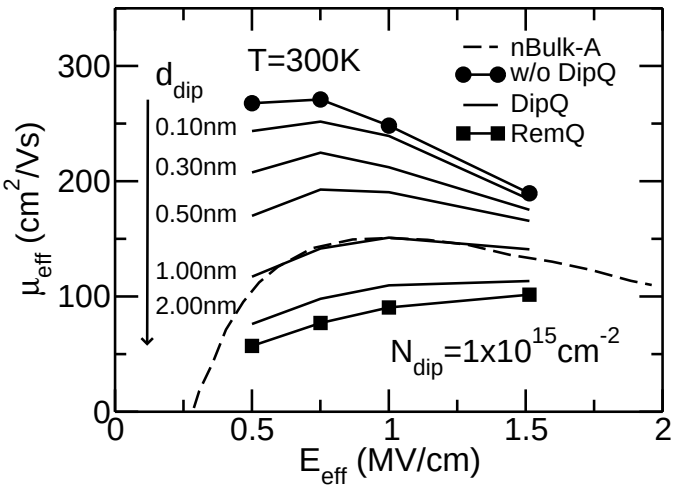

Figure 5.36: Simulated electron mobility versus the effective field for the nBulk-A device with t_{ITL}=1 nm and t_{HK}=3 nm of [31]. The curves obtained with or without the RemQ scattering and the experimental data of [31] are also shown. The concentration of the DipQ or the RemQ centres is 1×10^{15} cm^{-2}.

where N_{A} is the Avogadro constant. Then, we can obtain the molecolar concentration of the HfO_2 $N_{3\mathrm{D}}$ by means of:

$$N_{3\mathrm{D}} = \frac{d_{\mathrm{HfO_2}}}{m_{\mathrm{mol}}} = \frac{9.68\ \mathrm{gcm^{-3}}}{9.495\times10^{-22}\ \mathrm{g}} = 2.769\times10^{22}\ \mathrm{cm^{-3}} \tag{5.35}$$

From C_{ND} we can easily obtain the average distance between two molecules l_{mol}:

$$l_{\mathrm{mol}} = \frac{1}{\sqrt[3]{N_{3\mathrm{D}}}} = \frac{1}{\sqrt[3]{2.769\times10^{22}\ \mathrm{cm^{-3}}}} = 3.3\times10^{-8}\ \mathrm{cm} = 0.33\ \mathrm{nm} \tag{5.36}$$

From Eq. 5.36 and from the fact that the the HfO_2 lattice is cubic we finally find the areal concentration of the HfO_2 molecules:

$$N_{2\mathrm{D}} = \frac{1}{l_{\mathrm{mol}}^2} = \frac{1}{(3.3\times10^{-8}\ \mathrm{cm})^2} = 9.153\times10^{14}\ \mathrm{cm^{-2}} \tag{5.37}$$

The other value we need is the distance between two atoms of the HfO_2 molecule. From the literature, this value is about 0.2 nm. This is a reasonable value, since it has to be slightly smaller than the distance between two atoms of the HfO_2 l_{mol}, that has been found in Eq. 5.36.

Fig. 5.36 analyzes the effect of dipole charge distance d_{dipQ} on the simulated electron mobility for dipoles having one charge at the ITL/HK interface and the other one inside the HK layer. (A-position in Fig. 5.6). The results are in between the case with d_{dipQ}=0 (the effect of one charge is cancelled by the other one) and d_{dipQ}=t_{HK} (corresponding to the mobility degradation produced by the

RemQ mechanism since q_2 is completely screened because it has reached the metal gate). For N_{dipQ}=10^{15} cm^{-2}, the value d_{dipQ}=1 nm allows us to reproduce the experimental data of the nBulk-A device of [31].

Fig. 5.36 clearly shows that a layer of dipoles with concentration as N_{2D} in Eq. 5.37, and distance d_{dipQ} as l_{mol} in Eq. 5.36 has a negligible effect on the electron mobility.

5.6 Threshold voltage shift

In the previous section we have seen that large amounts of charges (sheets or dipoles) in the gate stack have to be assumed in order to reproduce the experimental data. Obviously, such densities of charges are expected to produce large threshold voltage shifts. In this subsection we examine how strongly these charges affect the threshold voltages of the n- and p-MOSFETs and assess the consistency with the mobility data.

5.6.1 Threshold voltage associated to charges at the ITL/HK interface

The threshold voltage shift ΔV_{TH} produced by a sheet of remote charges at the ITL/HK interface assumed to be of the same type (i.e. all positive or all negative) is:

$$\Delta V_{\mathrm{TH}}^{N_{\mathrm{ITL/HK}}} = \frac{e \cdot N_{\mathrm{ITL/HK}} \cdot t_{\mathrm{HK}}}{\epsilon_{\mathrm{HK}}} \tag{5.38}$$

By means of Eq. 5.38 the numerical values of ΔV_{TH} corresponding to the $N_{\mathrm{ITL/HK}}$ necessary to reproduce the experimental data of the devices in Tab. 5.2 can be calculated. The results are reported in Tab. 5.4: these values are unrealistic and, to our knowledge, never observed in any experiment. Moreover, both in the simulations and in Eq. 5.38 we have assumed an ideal HfO_2 dielectric with the relative electric permittivity equal to 22, thus close to its theoretical value [48]. Assuming lower values of the electric permittivity of HfO_2 allows us to reproduce the experimental data with a slightly lower concentration of remote charges, but the ΔV_{TH} due to the RemQ remains unacceptable. For instance, if we assume ϵ_{HK}=12 (instead of 22), we can reproduce the nBulk-A devices of [31] assuming $N_{\mathrm{ITL/HK}}$=5×10^{13} cm^{-2} (instead of 9×10^{13} cm^{-2}). However, according to Eq. 5.38, the ΔV_{TH} would be 2.26 V, thus slightly increased with respect to the value reported in Tab. 5.4 for the nBulk-A devices.

5.6.2 Coulomb centers in various positions

So far we have assumed remote charge at the ITL/HK interface. We see now if different charge configurations can reproduce the experimental data and with which parameters.

device	$N_{\mathrm{ITL/HK}}$ (cm^{-2})	ΔV_{TH} [V] n-MOS	ΔV_{TH} [V] p-MOS
Bulk-A	9×10^{13}	2.22	2.22
Bulk-B	1×10^{14}	2.46 (t_{HK}=3nm)	-
	1×10^{14}	1.31 (t_{HK}=1.6nm)	-
Bulk-C	4×10^{13}	0.99	-
Bulk-D	5×10^{13}	1.23	-
nSG-E	3×10^{13}	0.74	
pSG-E	6×10^{13}		1.48

Table 5.4: ΔV_{TH} due to the RemQ concentration of charges needed to reproduce the experimental data. All the charges are assumed to have the same sign and to be at the ITL/HK interface.

Bulk charges in the ITL

Fig. 5.37 compares the experimental data for the nBulk-A devices with t_{ITL}=1 nm and 2 nm (t_{HK} is 3 nm) at T=300 K and the MSMC simulations in which it is assumed to have the Coulomb centers located in the ITL in addition to the Si/ITL charges calibrated in Sec. 5.5.1 on the SiO_2 control devices.

The threshold voltage shift produced by a constant distribution of charges in the interfacial layer:

$$\Delta V_{\mathrm{TH}} = \int_0^{t_{\mathrm{ITL}}} \frac{qN_{\mathrm{ITL}}}{\epsilon_{\mathrm{HK}}\epsilon_{\mathrm{ITL}}}[\epsilon_{\mathrm{HK}}x + \epsilon_{\mathrm{ITL}}t_{\mathrm{HK}}]dx \tag{5.39}$$

Solving the integral for the variable x we obtain:

$$\Delta V_{\mathrm{TH}} = \frac{qN_{\mathrm{ITL}}}{t_{\mathrm{ITL}}^2}2\epsilon_{\mathrm{ITL}} + \frac{qN_{\mathrm{ITL}}t_{\mathrm{HK}}t_{\mathrm{ITL}}}{\epsilon_{\mathrm{HK}}} \tag{5.40}$$

Tab. 5.5 summarizes the densities of charges that have to be assumed to reproduce the experimental data and the threshold voltages produced by that charges. The densities of N_{ITL} assumed and the ΔV_{TH} obtained are not so different from that obtained assuming the charges to be at the ITL/HK interface. However, in this case we need to recalibrate the RemQ density when the t_{ITL} changes.

Charge sheet at the Si/ITL interface

Fig. 5.38 compares the experimental data for the nBulk-A devices with t_{ITL}=1 nm and 2 nm (t_{HK} is 3 nm) at T=300 K and the MSMC simulations in which only charge at the Si/ITL interface are considered and used as fitting parameters to reproduce experimental data of the high-k devices. Here we are assuming that all the imperfections due to the HK/metal-gate process are mimicked by a layer of charges

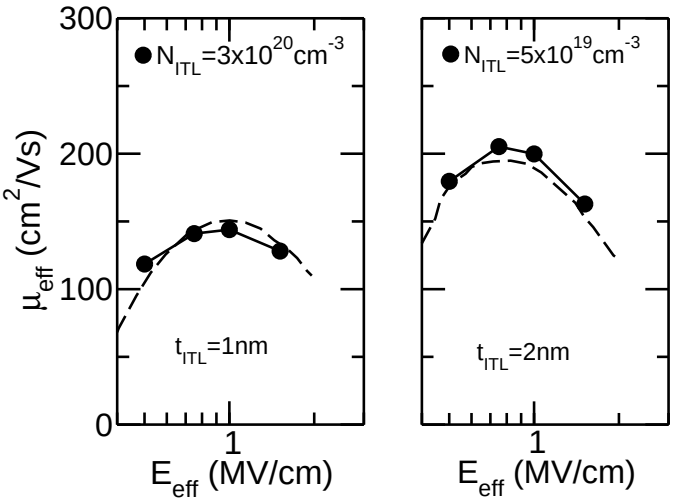

Figure 5.37: Electron mobility vs. effective field for the nBulk-A devices [31] with $t_{\rm ITL}$=1 nm and 2 nm at T=300 K. $t_{\rm HK}$=3 nm for both the devices. The experimental data is reproduced assuming the presence of bulk charges in the interfacial layer plus the charge at the Si/ITL interface calibrated on the SiO_2 control (Tab. 5.3).

	$N_{\rm ITL}$ (cm^{-3})	$\Delta V_{\rm TH}$ (V)
$t_{\rm ITL}$=1 nm	3×10^{20}	1.43
$t_{\rm ITL}$=2 nm	5×10^{19}	0.711

Table 5.5: Density of charges in the bulk of the ITL ($N_{\rm ITL}$) that has to be inserted in the simulations to reproduce the nBulk-A devices with $t_{\rm ITL}$=1 or 2 nm and $t_{\rm HK}$=3 nm. The $\Delta V_{\rm TH}$ produced by such density of charges is also reported.

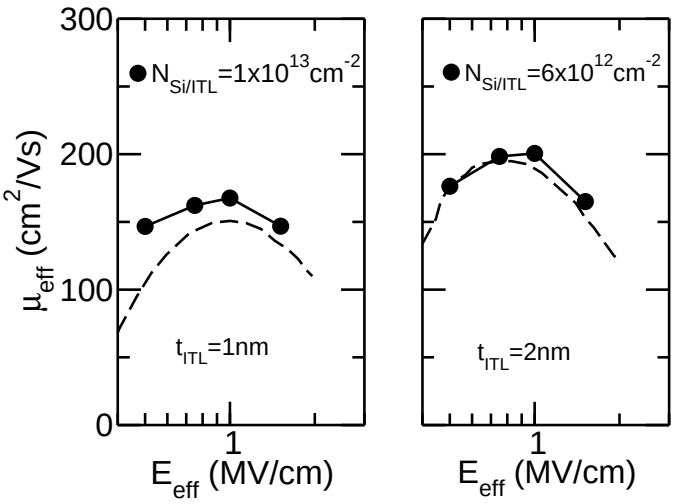

Figure 5.38: Electron mobility vs. effective field for the nBulk-A devices with $t_{\rm ITL}$=1 nm and 2 nm at T=300 K. The $t_{\rm HK}$ is 3 nm for both the devices. The experimental data is reproduced by adjusting the charges at the Si/ITL interface while no other remote charges are assumed.

at the Si/ITL. We are thus assuming that the imperfections due to the non-ideal gate stack fabrication approach the channel region.

The threshold voltage shift produced by a layer of charges at the Si/ITL interface can be found using:

$$\Delta V_{\rm TH} = \frac{qN_{\rm Si/ITL}(\epsilon_{\rm HK}t_{\rm ITL} + \epsilon_{\rm ITL}t_{\rm HK})}{\epsilon_{\rm HK}\epsilon_{\rm ITL}} \tag{5.41}$$

Tab. 5.6 summarizes the densities of charges at the Si/ITL interface that have to be inserted in the simulations to reproduce the experimental data and the threshold voltages produced by that charges. As in the case with bulk charges in the ITL, the $N_{\rm Si/ITL}$ that allows to reproduce the experimental data changes when the $t_{\rm ITL}$ value is changed. However, differently from the previous case, the $\Delta V_{\rm TH}$ seems to be independent of the $t_{\rm ITL}$ value, but again very large and never observed in experiments.

Arbitrary position in the gate stack

To conclude this analysis, we want to see if there is any optimal position for a layer of charges in the gate stack, namely if there exists a position for the charges that maximizes the mobility reduction having a low $\Delta V_{\rm TH}$ associated to it. Fig. 5.39 shows the matrix element squared (that can be considered proportional to the induced mobility reduction) and the corresponding threshold voltage shift for a sheet of charges with density N_{it}=1$\times$10^{13} cm^{-2} as a function of its position in

	$N_{\mathrm{Si/ITL}}$ (cm^{-2})	ΔV_{TH} (V)
t_{ITL}=1 nm	1×10^{13}	0.711
t_{ITL}=2 nm	6×10^{12}	0.706

Table 5.6: Density of charges at the Si/ITL interface ($N_{\mathrm{Si/ITL}}$) that has to be assumed to reproduce the nBulk-A devices with t_{ITL}=1 or 2 nm and t_{HK}=3 nm. The ΔV_{TH} produced by such density of charges is also reported.

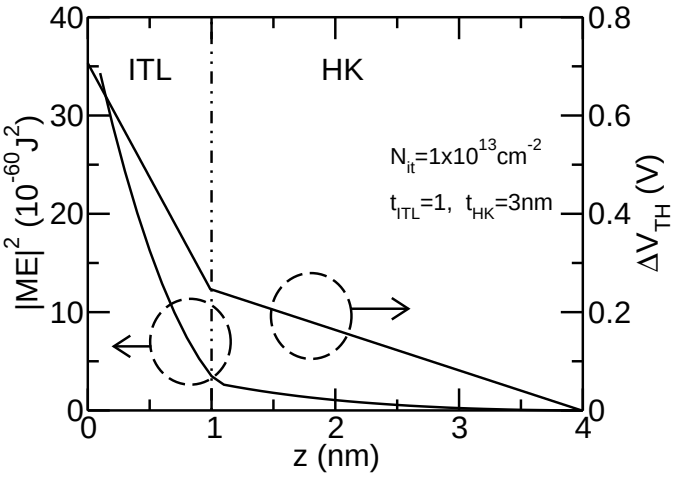

Figure 5.39: Comparison between the matrix element squared and the threshold voltage shift produced by a sheet of charges with density of 1×10^{13} cm^{-2} as a function of the position of the gate stack from the Si/ITL interface to the HK/metal gate interface. The t_{ITL} is 1 nm and the t_{HK} is 3 nm.

the gate stack. The device considered for this analysis is the nBulk-A device [31] with t_{ITL}=1 nm and t_{HK}=3 nm.

We can see that, when the layer approaches the channel, the squared matrix element is enhanced, but the ΔV_{TH} produced increases in the same way. Thus, the results obtained so far are qualitatively valid also for Coulomb centers in other positions in the gate stack.

Alternatively, models for the remote charge distributions should assume a random placement of positive and negative charges or dipoles parallel to the ITL/HK interface statistically yielding a zero ΔV_{TH}. However it is difficult to believe that a density of 10^{13} or 10^{14} cm^{-2} random placed charges (which would be comparable or larger than the inverson layer density) could exist at any interface without interacting with each other (the distance between the charges would be of few Angstroms). The possible superimposition of the scattering potentials stemming from charges with different sign thus implies a reduction of the scattering potential per Coulomb scattering center with respect to the form of the scattering potential

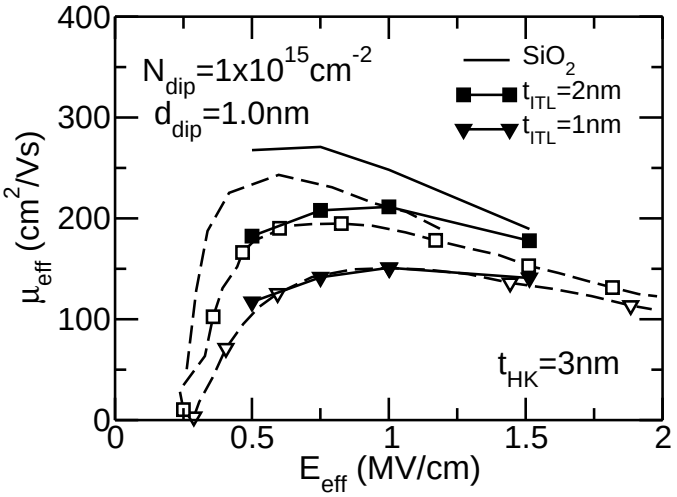

Figure 5.40: Simulated effect of d_{dipQ} on the electron mobility when DipQ is activated (in addition to Si phonons, surface roughness and SOph scattering mechanisms). The DipQ (1×10^{15} cm^{-2}) is placed in the A-position (see Fig. 5.15), close to the SiO_2/HfO_2 interface. Bulk device with N_{A}=2×10^{17} cm^{-3}, t_{ITL}=1 nm and t_{HK}=3 nm.

assumed so far. This, in turn, reduces the effect of the Coulomb centers and does not contribute to explain the experiments, but instead would force us to further increase the density of the Coulomb centers in the simulations in order to reproduce the experimental data.

5.6.3 Threshold voltage associated to DipQ

One may expect dipoles to provide a smaller ΔV_{TH} with respect to the RemQ model for a given mobility degradation. However, we will see in the following that this is not the case.

Firstly, Fig. 5.40 shows that the same (N_{dipQ},d_{dipQ}) pair reproducing the mobility of the t_{ITL}=1 nm device reproduce also that of the HfO_2 device of [31] with t_{ITL}=2 nm and t_{HK}=3 nm. Thus, from now on, we will limit our analysis to the nBulk-A device featuring t_{ITL}=1 nm.

Fig. 5.41 shows the pairs of values (N_{dipQ},d_{dipQ}) allowing to reproduce the experimental mobility results of [31] proceeding as in Fig. 5.36 at 300 K and 100 K, for dipoles in the A-position. As it can be seen the same sets of (N_{dipQ},d_{dipQ}) can approximately reproduce the experimental mobility behaviour at different temperatures. The small spread between the (N_{dipQ},d_{dipQ}) needed to reproduce results at 300 K and 100 K is not directly related to the modeling of RemQ and DipQ scattering: even when considering the SiO_2 control devices (see Fig. 5.17), slightly different model parameters would be needed to reproduce the data at 100 K and 300 K.

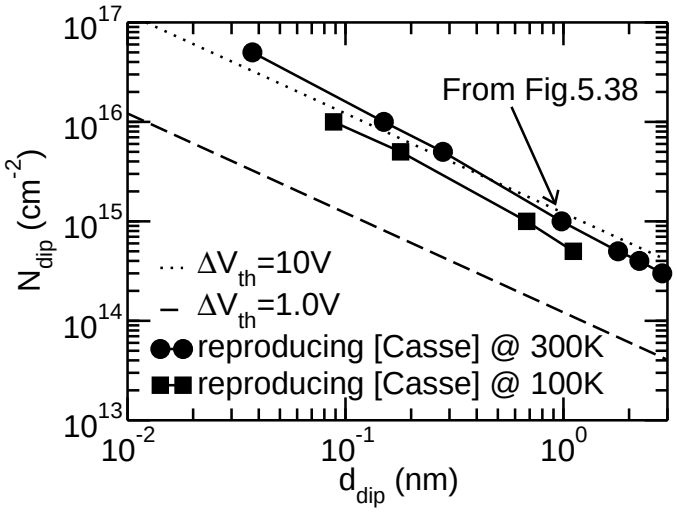

Figure 5.41: Configurations in the ($N_{\rm dipQ}$,$d_{\rm dipQ}$) plane of the dipole in the position A reproducing the experimental mobility of [31] as in Fig. 5.36. The configurations giving $\Delta V_{\rm TH}$=10 V or $\Delta V_{\rm TH}$=1.0 V are also indicated.

By using:

$$\Delta V_{\rm TH}^{\rm A-DipQ} = \frac{e \cdot N_{\rm dipQ} \cdot d_{\rm dipQ}}{\epsilon_{\rm HK}} \tag{5.42}$$

which gives the $\Delta V_{\rm TH}$ produced by a sheet of dipoles in the A-position (Fig. 5.6), the configurations giving a $\Delta V_{\rm TH}$ of 1.0 V and 10 V can be identified (lines in Fig. 5.41). We see that the lowest $\Delta V_{\rm TH}$ corresponds to $d_{\rm dipQ}$=$t_{\rm HK}$, which is a condition equivalent to the RemQ, namely a single layer of charges at the ITL/HK interface, as considered in Sec. 5.3.

The situation is not significantly improved if we consider dipoles in position B (Fig. 5.42) or C (Fig. 5.43). In fact, noting that the corresponding $\Delta V_{\rm TH}$ is:

$$\Delta V_{\rm TH}^{\rm B-DipQ} = \frac{e \cdot N_{\rm dipQ} \cdot d_{\rm dipQ} \cdot (\epsilon_{\rm HK} + \epsilon_{\rm ITL})}{2 \cdot \epsilon_{\rm HK} \cdot \epsilon_{\rm ITL}} \tag{5.43}$$

and:

$$\Delta V_{\rm TH}^{\rm C-DipQ} = \frac{e \cdot N_{\rm dipQ} \cdot d_{\rm dipQ}}{\epsilon_{\rm ITL}} \tag{5.44}$$

for a layer of dipoles in the B and C position, respectively, we could find few ($N_{\rm dipQ}$,$d_{\rm dipQ}$) pairs reproducing the experimental data while giving a $\Delta V_{\rm TH}$ slightly smaller than 1 V. However, the configurations with the lowest $\Delta V_{\rm TH}$ are those for $d_{\rm dipQ}$$\simeq$$2t_{\rm ITL}$ for case B and $d_{\rm dipQ}$$\simeq$$t_{\rm ITL}$ for case C. Hence they are essentially equivalent to the RemQ case with $N_{\rm Si/ITL}$=$N_{\rm dipQ}$; since they are dominated by the sheet of charge at the Si/ITL interface. Therefore, for the dipole configurations A, B and C in Fig. 5.6, the densities that reproduce the experimental mobility give a $\Delta V_{\rm TH}$ shift comparable to or even larger than the RemQ case.

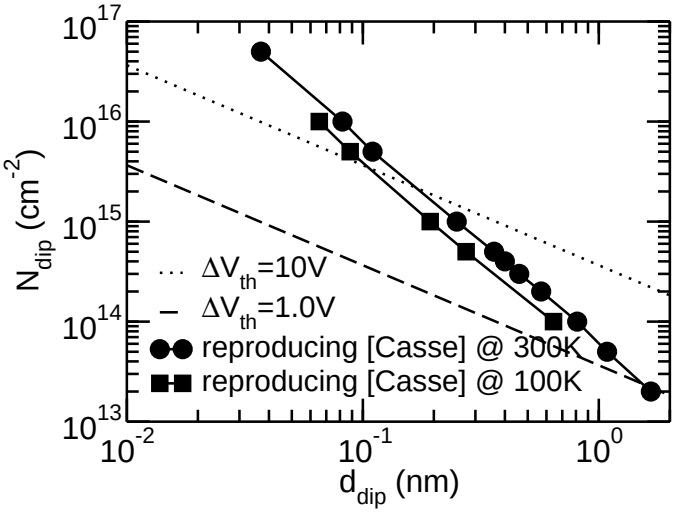

Figure 5.42: Same as in Fig. 5.41 but for dipoles in position B.

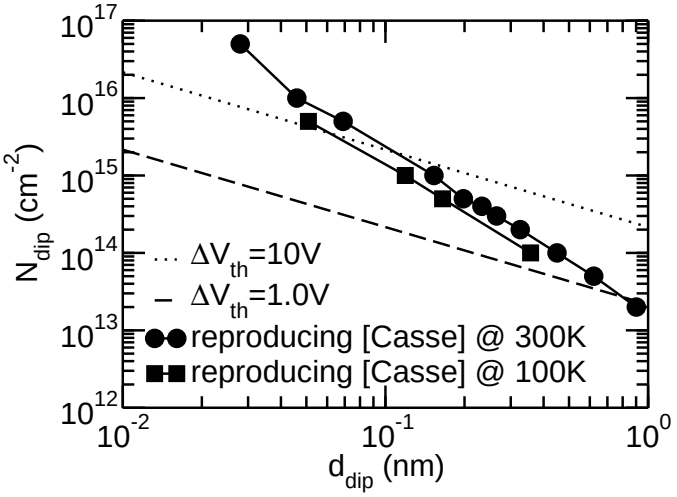

Figure 5.43: Same as in Fig. 5.41 but for dipoles in position C.

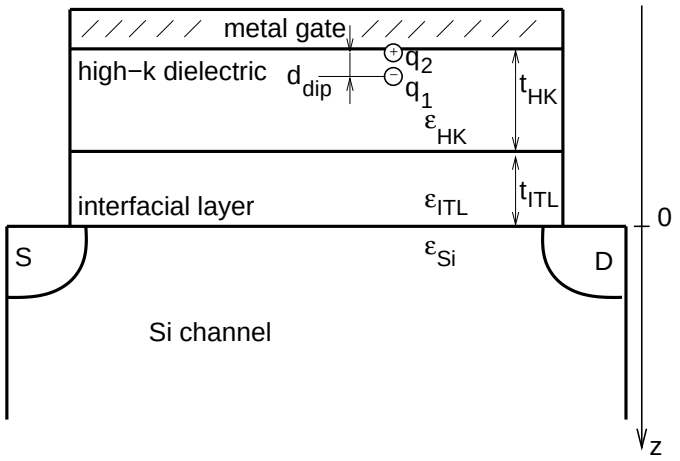

Figure 5.44: Sketch of the dipole at the metal gate/HK interface (HK/MG-DipQ). z_{q2} is placed at $-t_{\rm ITL}-t_{\rm HK}$ and z_{q1} is placed in the HK dielectric.

We finally note that such densities and dipole distances are much larger than the ones provided by *ab initio* calculations on the SiO_2/HfO_2 interface [55] as discussed in Sec. 5.5.4.

5.6.4 Dipoles at the MG/HK interface

A possible explanation of the apparent inconsistency between the RemQ or DipQ densities needed to reproduce the mobility and the $\Delta V_{\rm TH}$ may be the presence of a sheet of dipoles at the metal gate/HK interface (see Fig. 5.44) as reported for instance in [167, 168, 169, 170, 171, 172]. These dipoles could compensate with an opposite $\Delta V_{\rm TH}$ the ones produced by RemQ or DipQ, thus masking their effect on the overall $\Delta V_{\rm TH}$, while still having marginal impact on the inversion layer mobility due to their distance from the channel. To test this hypothesis, the concentration of the dipoles at the metal gate/high-k interface (HK/MG-DipQ) has been estimated using Eq. 5.45, consistently with [167]. In detail, we have chosen the commonly accepted value for $d_{\rm dipQ}\approx$0.2 nm [171] and the largest estimated dipole in literature ($\Delta V_{\rm TH}\approx$0.7 V in [168]) which corresponds to a $N_{\rm HK/MG-DipQ}$ equal to:

$$N_{\rm dipQ} = \frac{\epsilon_{\rm HK}\cdot\Delta V_{\rm TH}}{e\cdot d_{\rm dipQ}} \simeq 4.25\times10^{14}\ {\rm cm}^{-2} \tag{5.45}$$

similarly to what found in [169]. Fig. 5.45 demonstrates that the estimated HK/MG-DipQ layer has a negligible effect on the electron mobility in the case of the nBulk-A devices of [31].

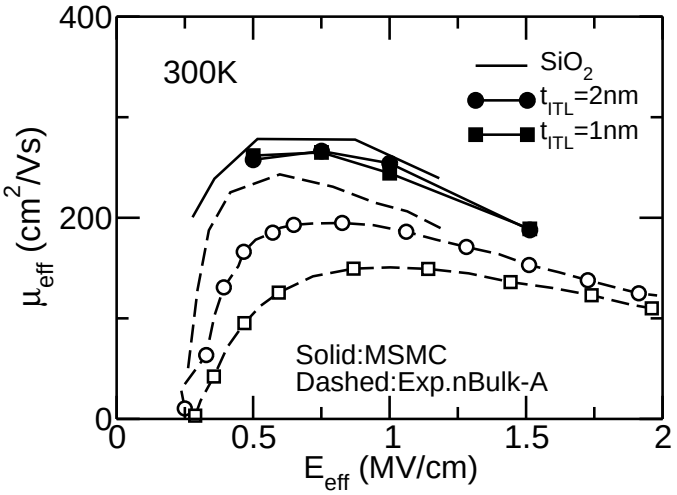

Figure 5.45: Comparison between simulated and experimental [31] electron mobility versus the effective field for 3 nm thick HfO_2 devices with $t_{\rm ITL}$=1 nm and $t_{\rm ITL}$=2 nm. Simulations have been performed accounting for SOph and HK/MG-DipQ scattering mechanisms. The concentration of the DipQ centres is 4.25×10^{14} cm^{-2} and $d_{\rm dipQ}$ is 0.2 nm [168].The experimental SiO_2 reference data of [31] and its simulation are also shown (dashed and solid lines).

5.7 Impact of the high-k dielectrics on the ON-current

So far we have addressed the influence of the SOph, RemQ and DipQ on the n and p-MOS mobility. Now we want to analyze the effect of the high-k dielectrics on the drain current in short devices. Indeed, an important and not yet fully addressed point concerns the possible on-current degradation produced by the use of a high-κ dielectric in very short MOSFETs [156].

Since we have previously shown that DipQ model does not lead any advantage on the trade off between the mobility reduction and the threshold voltage shift with respect to the RemQ one, from now on we simulate the high-k dielectrics accounting for only the SOph and RemQ mechanisms.

5.7.1 SOph vs RemQ in HfO_2 in a template structure

We have simulated first the drain current of a 32 nm SG-SOI device in the presence of SOph and RemQ scattering. The simulated device has been designed according to the LSTP device of the 32 nm technology node of [10]. Main features of this template device are reported in Tab. 5.7.

Fig.5.46 shows the simulated I_D-V_G curves in saturation regime of the device described in Tab. 5.7, assuming that the RemQ density at the ITL/HK interface is the same as for the Bulk-A devices in Tab 5.2, namely 9×10^{13} cm^{-2} (see Tab. 5.4). We see that the impact of SOph scattering alone in this device is essentially negli-

	32-LSTP
Structure	SG-SOI
Stress [GPa]	-
Gate Stack	SiO_2 + HfO_2
t_{ITL} [nm]	0.8
t_{HK} [nm]	3.2
V_{DD} [V]	1.0
N_{A} [cm^{-2}]	Undoped
t_{Si} [nm]	7
Type	LSTP

Table 5.7: Main parameters of the template 32 nm device representation of the technology node indicated in [10].

gible. On the other hand, the activation of RemQ scattering reduces I_{DS}, although the I_{ON} (I_{DS} at V_{GS}=V_{DS}=V_{DD}) is not as degraded as much as the mobility. As expected, the relative I_{DS} reduction is stronger at low gate voltages, where Coulomb scattering is stronger because the screening is weak. In order to emphasise the impact of RemQ on the transport alone, the charge introduced in the stack is considered for the RemQ scattering but it is not included in the Poisson equation; thus, the threshold voltage is the same with and without RemQ.

Analysis of the internal quantities relevant in quasi-ballistic transport (Sec. 2.4) revealed that: 1) the inversion charge N_{inv} and the average velocity of the injected carriers v_{inj} at the virtual source (VS) are almost the same with and without RemQ scattering; 2) the activation of RemQ scattering mostly affects the back-scattering coefficient r, that increases from 0.24 to 0.31 at V_{GS}=1 V, and from 0.23 to 0.39 at V_{GS}=0.7 V, consistently with the larger mobility reduction induced by RemQ at low V_{GS}.

5.7.2 ON-current of realistic devices

Now we want to simulate realistic devices, which main features have been reported in Tab. 5.8 [173]. Firstly we calibrate our MSMC simulator on the long channel mobility curves [173], and then we analyze the impact of the high-k dielectrics on the short channel n- and p-MOS devices.

It is worth noting that the high-k material is $HfZrO_2$. We could not find the SOph parameters for this material. However, in this chapter, we have shown that the SOph scattering in real devices is negligible even for the HfO_2 which has very strong polar phonons (see Tab. 5.1). So, we can safely neglect the SOph scattering and account for the RemQ scattering only.

Fig. 5.47 shows the calibration of our simulators on the electron and hole long channel mobilities. The experimental mobility curves can be reproduced assuming

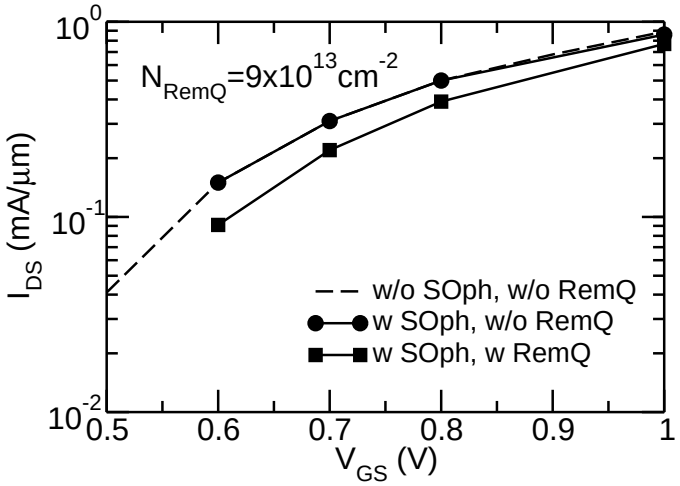

Figure 5.46: Simulated I_D-V_G curves in saturation regime ($V_{\rm DS}$=$V_{\rm DD}$) for a SG-SOI device with 32 nm gate length. The drain current reduction due to the effect of the SOph and RemQ mechanisms are shown.

	32-LSTP
Structure	SG-SOI
Stress [GPa]	-
Oxide	SiO_2 + $HfZrO_2$
$t_{\rm ITL}$ [nm]	1.3
$t_{\rm HK}$ [nm]	1.9
$V_{\rm DD}$ [V]	1.0
$N_{\rm A}$ [cm^{-2}]	Undoped
$t_{\rm Si}$ [nm]	7
Type	LSTP

Table 5.8: Main parameters of the 32 nm n- and p-MOS devices measured in [173].

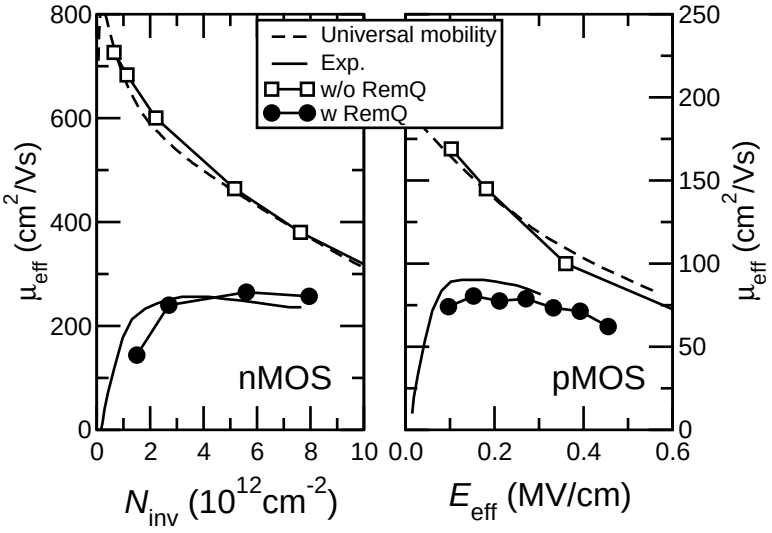

Figure 5.47: Comparison between the experimental mobility [173] for high-k Si n-MOS (left, vs. inversion density $N_{\rm inv}$) and p-MOS (right, vs. the effective field $E_{\rm eff}$) and the corresponding MSMC simulations. The measured [47] and simulated universal mobility curves are also shown for reference. The RemQ density is 2×10^{14} cm^{-2} and 1×10^{14} cm^{-2} for the n-MOS and p-MOS case, respectively.

RemQ densities that are very close to each other. Indeed the experimental mobility has been reproduced assuming 2×10^{14} cm^{-2} and 1×10^{14} cm^{-2} for the n-MOS and the p-MOS devices, respectively. Moreover, these densities are close also to the values reported in Tab. 5.4 for a wide range of experimental devices.

Fig. 5.48 shows the I_D-V_G curves for both n- and p-MOS case, in both saturation and linear regime. The RemQ density employed in this simulations is the one that has been calibrated on the mobility curves in Fig. 5.47. Good agreement is found between the simulated and the experimental [146] curves.

We can see that the effect of the high-k dielectrics on the drain current is weaker than on the mobility (as already explained in Sec. 5.7.1). However, it can not be neglected in the short channel n- and p-MOSFETs.

5.8 Conclusions

Multi Subband Monte Carlo simulations compared to experimental data for HK MOSFETs fabricated in different $R\&D$ lines provide a consistent indication that the experimentally observed mobility reduction cannot be attributed to soft optical phonons neither in n- nor in p-MOS devices.

The bias and temperature dependence of the reduction appears to be consistent with significant remote Coulomb scattering with charge trapped in the gate stack for all the HK materials considered. Assuming the charges to have the same sign,

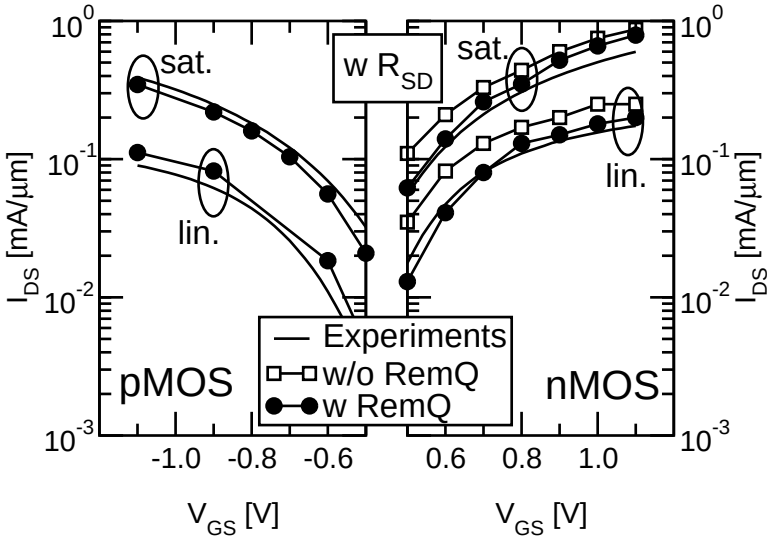

Figure 5.48: Comparison between the experimental [146] and the simulated drain current I_D for a 32 nm n-MOS and p-MOS as a function of the gate voltage $V_{\rm GS}$ in both linear and saturation regimes. The RemQ density employed for both n-MOS and p-MOS has been determined in Fig. 5.47. Simulations account for the series resistance $R_{\rm SD}$=200 $\Omega\mu$m (source plus drain resistance) estimated in [146].

they should cause very large threshold voltage shifts inconsistent with the experiments. The situation does not improve if we assume the charge to be in the form of dipoles with a dipole moment normal to the ITL/HK interface. Possible models for the remote charge should then assume a random distribution of positive and negative charge or dipoles placed parallel to the ITL/HK interface.

Further efforts appear to be necessary to understand the causes of the mobility reduction.

In this respect, a recent work [142] ascribes the mobility reduction to neutral defects located at the interface between the channel and the gate stack which are induced by nitrogen diffusion during the fabrication process. Such a scattering mechanism would obviously solve the inconsistency with the $V_{\rm TH}$ shift. The inclusion of this model in physics-based Multi Subband Monte Carlo transport models is however made difficult by the lack of a generally accepted expression for the scattering potential.

Chapter 6

Modeling of alternative channel materials

Abstract

In this chapter we use the Multi-Subband Monte Carlo approach to analyze the ON-current in Si, sSi, Ge and sGe n-MOSFETs by accounting for all the relevant scattering mechanisms (including the remote surface-optical phonons, SOph, and remote Coulomb scattering, RemQ, related to the presence of high-κ dielectrics), in which strain is implicitly introduced by a modification of the band structure.

In the first part of the chapter we explain the main differences in simulating Si and Ge as channel materials. Thus, firstly we show the different groups of valleys that are accounted for in the simulation of Ge inversion layers. Then, we show the modifications that have been applied to the scattering mechanisms, namely the bulk phonons and the calibration of the surface roughness.

Finally, we simulate high-performance DG-SOI MOSFETs with Si, s-Si, Ge or s-Ge as channel material in order to analyze the beneficts in introducing the Ge technology in short channel devices.

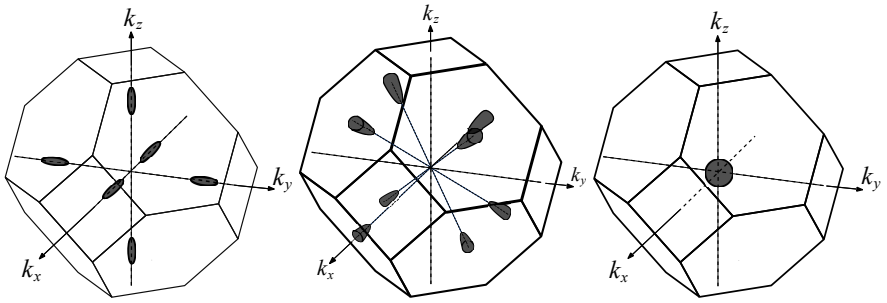

Figure 6.1: Sketch of the three main groups of valleys in the Brillouin zone of a zincblende lattice: Δ (left), Λ (center) and Γ (right) valleys are shown.

6.1 Introduction

The competitive edge of Ge compared to Si MOSFETs is a hot topic already investigated for n-MOSFETs by using quantum ballistic simulations [174, 175] and for p-MOSFETs with a semi-classical approach [176]. Recently, promising experimental data for electron and hole mobility in Ge MOSFETs was reported [177, 72, 73], as well as the first data for the electron mobility enhancement in strained Ge (sGe) transistors [74]. Thus a more complete and quantitative study of the on-current $I_{\rm ON}$ in nanoscale Ge MOSFETs is now possible.

In Sec. 6.2 we first validate our models comparing our simulations to recent mobility experiments and then study the $I_{\rm ON}$ in n-MOSFETs devices designed for a HP application. Then, in Sec. 6.3, we compare the ON-current in short channel in DG-SOI n-MOSFETs feauring Si, sSi, Ge or sGe as channel material to establish the theorical advantages in introducing the Ge material in the CMOS technology.

6.2 Transport modeling description

The mobility $\mu_{\rm eff}$ and the drain current $I_{\rm DS}$ are simulated with the multi-subband Monte Carlo (MSMC) approach described in Chap. 3. In the simulations of this chapter we will account for the presence of the high-k dielectrics using the models for soft optical phonons (SOph) and remote charges (RemQ) described in Chap. 5.

6.2.1 The simulation of germanium inversion layers

Modeling of the crystal properties

The simulations of Ge n-MOS transistors account for the Λ, Δ and Γ valleys (shown in Fig. 6.1) within the effective mass approximation EMA (explained in Sec. 3.1.4) for the different crystal orientations with non parabolic corrections in the transport plane [99].

Valley	m_l/m_0	m_t/m_0	α [eV^{-1}]	ΔE_v [eV]	ν
Λ	1.588	0.081	0.3	0.0	4
Δ	1.353	0.29	0.0	0.18	6
Γ	0.037	0.037	0.0	0.14	1

Table 6.1: Parameters of the conduction band minima in Ge, taken from [178]. m_l and m_t are the longitudinal and transverse masses, respectively. α is the non-parabolicity factor defined in Eq. 3.15. ΔE_v is the shift of the valley minima with respect of the minimum of the conduction band. ν is the multiplicity of the valley.

	Ge (100)		**Ge (111)**		**Ge (110)**	
	ν	m_q	ν	m_q	ν	m_q
Λ	4	0.1539 (0.1185)	1	2.0756 (1.588)	2	0.2726 (0.2205)
			3	0.1187 (0.0905)	2	0.1039 (0.081)
Δ	2	0.8355 (1.353)	6	0.3094 (0.3929)	2	0.2341 (0.29)
	4	0.1943 (0.29)			4	0.3136 (0.4776)

Table 6.2: Quantization mass m_q extracted by fitting with an EMA model the subband minima calculated with the LCBB method for triangular wells with different confining electric fields. The values in parenthesis are those inferred from transverse and the longitudinal masses of bulk Ge, i.e. m_t and m_l in Tab. 6.1 [178], using the orientation matrices reported in [78].

Tab. 6.1 shows the values of the masses for the three groups of valleys, taken from [178]. However, in order to improve the accuracy in the calculation of the different valleys subband minima, whose relative position critically affects the subband population and hence the transport, we used the Linear Combination of Bulk Bands (LCBB) quantization model [104] to extract modified quantization masses m_q to be used in the simplified EMA model. The m_q values used in this work are reported in Tab. 6.2 and Fig. 6.2 shows the good agreement between the subband minima obtained with either the EMA or the LCBB model (using the same confining potential).

Modeling of the scattering mechanisms

With the phonon set in [179] our simulations reproduce the velocity vs. field curves in bulk Ge [180]. For Ge MOSFET simulations, instead, the deformation potential for acoustic phonons is empirically increased with respect to [179] by the same enhancement factor $13/9$ used also in Si inversion layers [99, 121]. Tab. 6.3 shows the parameters we used to account for intra-valley and intervalley phonon scattering. Moreover, in Ge devices phonon transitions can also move a carrier from a valley group to another. Tab. 6.4 shows the values we used to take into

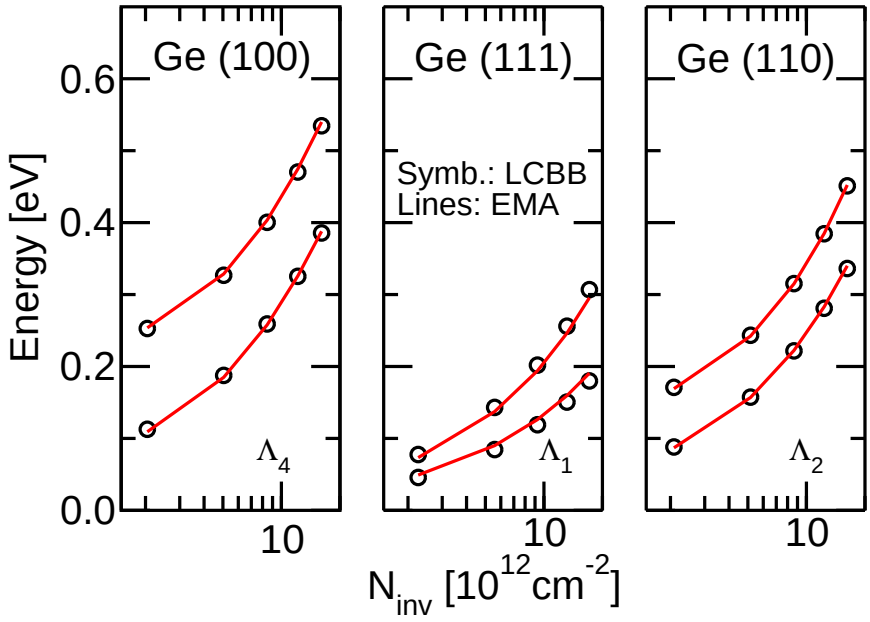

Figure 6.2: Subband minima calculated with the LCBB (symbols) and the EMA model (lines, m_q from Tab. 6.2) for the same confining potential obtained with an EMA based self-consistent Schrödinger-Poisson solver. The two lowest subbands for the (100), the (111) and the (110) crystal orientations are shown.

Scattering mechanism	Parameter	Δ	Λ	Γ
Acoustic phonons	D_{ac} [eV]	13.0	15.8	7.22
Intervalley phonons	type	g	g	-
	ω [K]	100	430	-
	D_{op} [eV/m]	7.89×10^{9}	5.50×10^{10}	-
Intervalley phonons	type	g	f	-
	ω [K]	430	120	-
	D_{op} [eV/m]	9.46×10^{10}	2.0×10^{9}	-
Intervalley phonons	type	f	f	-
	ω [K]	430	320	-
	D_{op} [eV/m]	3.15×10^{10}	3.0×10^{10}	-

Table 6.3: Parameters used in all Ge simulations for acoustic and inter-valley phonons. Intra-valley and inter-valley electron phonons are from [179] but the deformation potential of intra-valley acoustic phonons has been empirically increased by (13/9) (with respect to [179]), namely by the same factor used also in Si to reproduce phonon limited mobility in inversion layers [99, 121].

Scattering mechanism	Parameter	$\Delta\leftrightarrow\Lambda$	$\Lambda\leftrightarrow\Gamma$	$\Gamma\leftrightarrow\Delta$
Intergroup phonons	ω [K]	320	320	320
	D_{op} [eV/m]	4.06×10^{10}	2.0×10^{10}	1.0×10^{11}

Table 6.4: Parameters used in all Ge simulations for inter-group phonons. Parameters from [179].

Quantity	GeO_2
ϵ_0	13.00
ϵ_{i}	6.58
ϵ_∞	5.03
$\hbar\omega_{\mathrm{TO},1}/e$ [meV]	37.1
$\hbar\omega_{\mathrm{TO},2}/e$ [meV]	45.8

Table 6.5: Parameters for the SOph scattering mechanism of the GeO_2. ϵ_0 is the relative permittivity at zero frequency, ϵ_{i} at intermediate frequency and ϵ_∞ at infinite frequency for each material. $\omega_{\mathrm{TO},1}$ and $\omega_{\mathrm{TO},2}$ are the energies of the two phonons.

account inter-group phonon scattering mechanism [179].

Concerning the high-k dielectrics, the RemQ density at the GeO_2/Al_2O_3 interface has been taken as $N_{\mathrm{GeO_2/Al_2O_3}}=6\times10^{12}$ cm^{-2}, that has been experimentally estimated in [72]. The SOph for the GeO_2 dielectric were included according to the GeO_2 parameters in Tab. 6.5, that have been taken from [181]; their impact on mobility is modest (about 3% at high effective field E_{eff}). The density of interface states at the Ge/GeO_2 interface has been assumed to be $N_{\mathrm{Ge/GeO_2}}=4\times10^{11}$ cm^{-2}, as estimated in [72].

Figs. 6.3 and 6.4 show the calibration of the surface roughness scattering mechanism for the (100) and (111) orientations on the experimental mobility data of [72, 73], respectively. For the (110) orientation, instead, the roughness parameters of the (100) have been maintained. The effective mobility is thus somewhat better for the (110) compared to the (100) orientation (see Fig. 6.3). Tab. 6.6 summarizes the parameters used for the surface roughness scattering mechanism in the different orientations.

Modeling of strain

The modeling of the uniaxial stress in n-type Si MOSFETs was described in detail in [182, 183]. Fig. 6.6 shows the good agreement between simulated and measured mobility enhancements in Si inversion layers. The uniaxial strain in (100) Ge n-MOSFETs was modeled introducing in the simulations the splitting between the $\Lambda_{||}$ and the $\Lambda_{\perp}$ valleys reported in [74] and obtained with DFT cal-

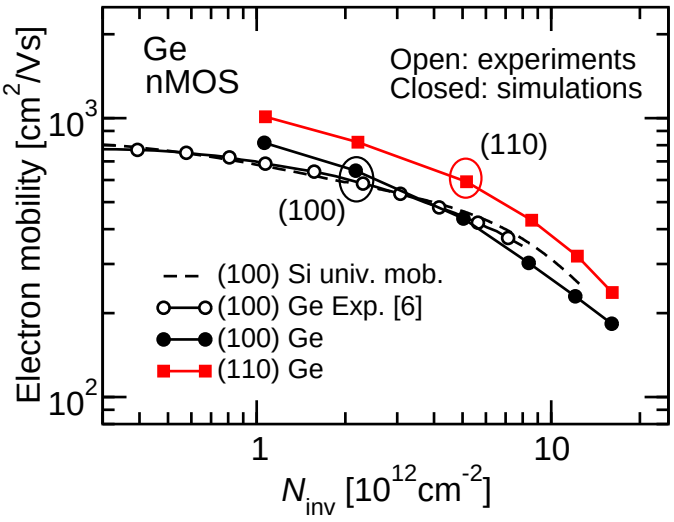

Figure 6.3: Experimental [73] and simulated (100) and (110) Ge electron mobility versus electron inversion density for different crystal orientations. A GeO_2/Al_2O_3 gate stack was assumed with t_{GeO_2}=0.5 nm. (100) Si universal mobility curve [47] is reported for comparison.

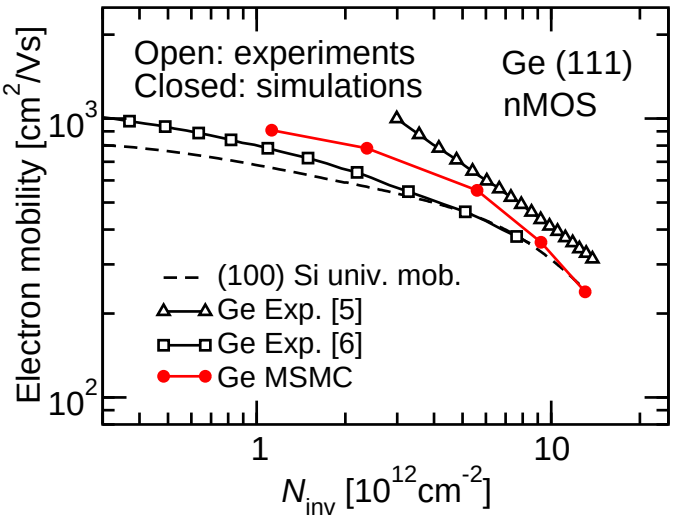

Figure 6.4: Experimental [72, 73] and simulated (111) Ge electron mobility versus electron inversion density for different crystal orientations. Same gate stack and trap densities as in Fig. 6.3. (100) Si universal mobility curve [47] is reported for comparison.

Scattering mechanism	Parameter	(100)	(110)	(111)
Surface Roughness	Spectrum	Gauss.	Gauss.	Gauss.
	Δ [nm]	1.2	1.2	1.0
	Λ [nm]	1.3	1.3	1.5

Table 6.6: Parameters used in all Ge simulations for the surface roughness scattering mechanism.

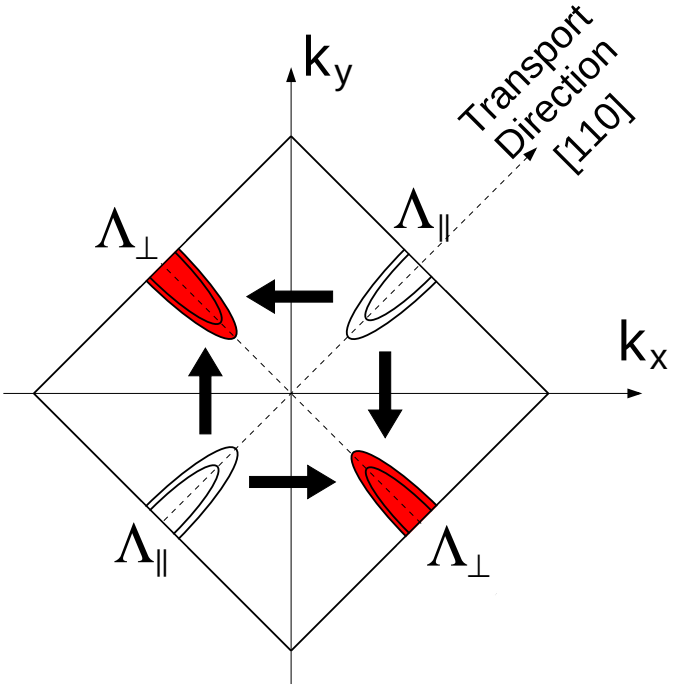

Figure 6.5: Sketch of the Λ valleys for a (100) Ge n-MOSFET with [110] transport direction. The $\Lambda_{||}$ valleys are aligned with transport and have transport mass 1.086 m_0; the $\Lambda_{\perp}$ valleys have a remarkably lower transport mass 0.082 m_0 [178]. The arrows indicate the stress induced repopulation of the $\Lambda_{\perp}$ valleys (see Fig. 6.8).

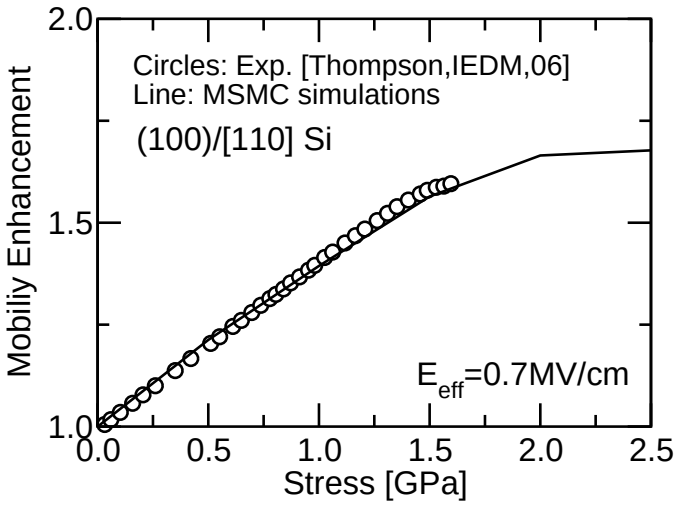

Figure 6.6: (100)/[110] Si. Measured (symbols [184]) and simulated (lines) electron mobility enhancement vs. stress at $E_{\rm eff}$=0.7 MV/cm. The stress is uniaxial along the [110] direction.

culations (see Fig. 6.5 for (100)-oriented Ge); the corresponding strain induced modulation of the transport masses was considered to be negligible [74]. Fig. 6.7 illustrates the good agreement between the simulated and experimental $\mu_{\rm eff}$ enhancements employing the same scattering parameters as in Figs.6.3 and 6.4 (see Tabs. 6.3, 6.4 and 6.6 for bulk phonons and surface roughness as well as Tab. 6.5 for the SOph scattering mechanisms and the values of $N_{\rm GeO_2/Al_2O_3}$ and $N_{\rm Ge/GeO_2}$ previously reported). Fig. 6.8 shows that the physical mechanism responsible for the $\mu_{\rm eff}$ improvement is the stress induced repopulation of the $\Lambda_\perp$ valleys that have a smaller transport mass than the $\Lambda_{||}$ valleys (see Fig. 6.5). The population of the Δ_2, Δ_4 and Γ valleys is always negligible.

6.3 $I_{\rm ON}$ in Si and Ge MOSFETs

The transport model described in Sec. 6.2 has been used to simulate high performance DG-SOI devices with $L_{\rm G}$=25 nm and 16 nm. The devices were designed as high performance MOSFETs and ideal work function flexibility is assumed to achieve the same inversion density $N_{\rm inv}$ in ON-state conditions. Series resistances $R_{\rm SD}$ were accounted for as external lumped elements; unless otherwise specified we used $R_{\rm SD}$=140 $\Omega\mu$m and $R_{\rm SD}$=160 $\Omega\mu$m for $L_{\rm G}$=25 nm and $L_{\rm G}$=16 nm, respectively, consistently with the ITRS roadmap [10]. $I_{\rm ON}$ is defined as $I_{\rm DS}$ for extrinsic voltages $V_{\rm GS}$=$V_{\rm DS}$=1 V. The $I_{\rm DS}$ is always reported per unit width, per gate. For Si and sSi devices we assumed an SiO_2/HfO_2 gate stack and for Ge devices a GeO_2 gate dielectric which give EOT=1 nm and 0.85 nm for $L_{\rm G}$=25 nm and $L_{\rm G}$=16 nm, respectively [73]. The appropriate SOph and RemQ scattering

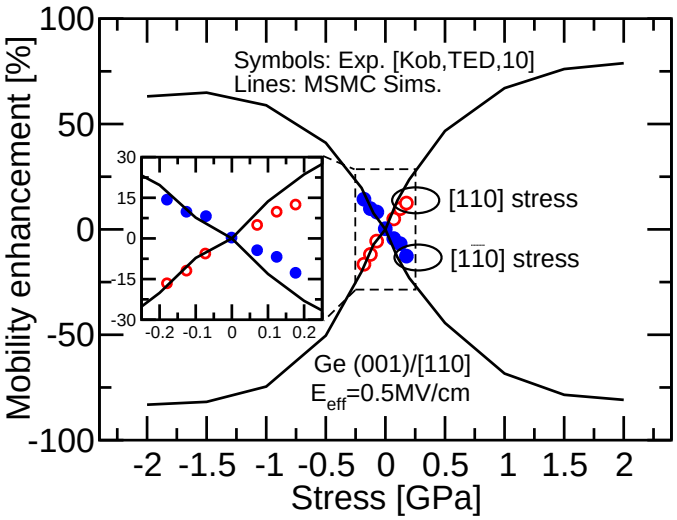

Figure 6.7: (100)/[110] Ge. Measured (symbols, [74]) and simulated (lines) electron mobility vs. uniaxial tensile stress along the [110] and [1$\bar{1}$0] directions and at E_{eff}=0.5 MV/cm. The inset shows a zoom for low stress values. The simulated mobility enhancement saturates at about 1.5 GPa.

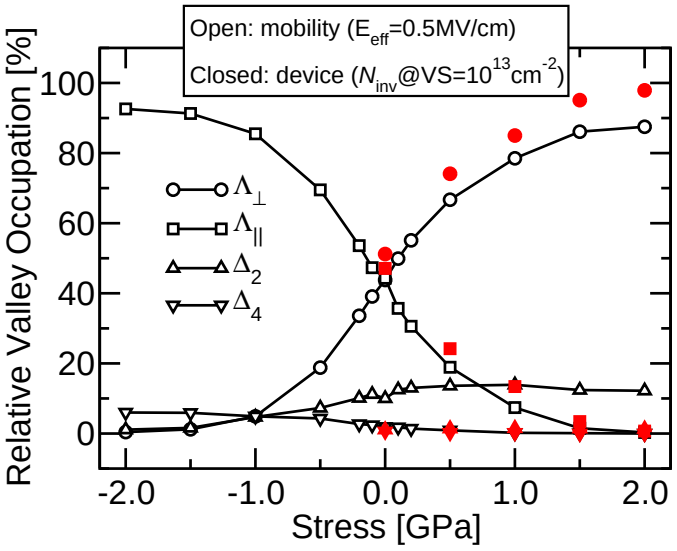

Figure 6.8: (100) Ge. Simulated occupation of the $\Lambda_{||}$, $\Lambda_\perp$ (see Fig. 6.5), and Δ_2, Δ_4 valleys versus stress along the [110] direction. Open symbols are results obtained with an equilibrium self-consistent Schrödinger-Poisson solver used for mobility calculation and for E_{eff}=0.5 MV/cm (as in Fig. 6.7). Tensile stress results in a strong repopulation of the $\Lambda_\perp$ valley. Closed symbols indicate the relative population at the virtual source of a 25 nm Ge n-MOS obtained with self-consistent MSMC simulations and for an inversion density at the VS N_{inv}=10^{13} cm^{-2}.

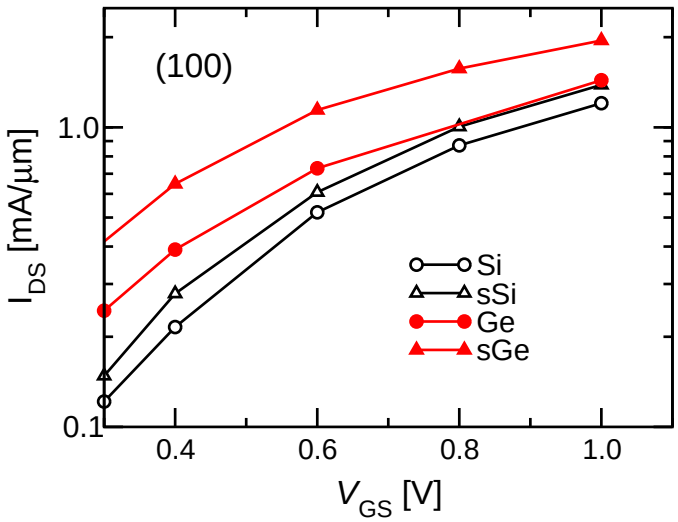

Figure 6.9: Simulated drain current versus gate voltage for the 25 nm DG device and for unstrained and strained Si and Ge (EOT=1 nm, t_{Si}=11 nm). The 1.5 GPa tensile stress is along the [110] transport direction.

mechanisms were accounted for using the parameters shown in Tabs. 5.1 and 6.5 for HfO_2 and GeO_2, respectively. However, the SOph and RemQ mechanisms have a modest impact on the I_{ON}, as seen in Sec. 5.7 for Si devices.

Fig. 6.9 shows that unstrained Ge n-MOSFETs are competitive with but do not outperform sSi devices (for a 1.5 GPa uniaxial stress). At the same time, Fig. 6.11(a) demonstrates that the injection velocities v_{inj} at the virtual source (VS) are comparable for sSi and Ge n-MOSFETs.

Fig. 6.10(a) reports the I_{ON} versus the stress for both Si and Ge n-MOSFETs; different crystal orientations for unstrained Ge are also shown. The I_{ON} for (110) and (111) Ge is larger than for sSi, essentially because of the larger v_{inj} shown in Fig. 6.11(a). Furthermore, the strained (100) Ge can outperform sSi significantly for a given R_{SD} value. Fig. 6.8 suggests that the I_{ON} enhancement is due to the repopulation of the $\Lambda_\perp$ valleys at the VS [82], which is the same physical mechanism responsible for the mobility enhancement. Fig. 6.11 consistently shows the increase of the v_{inj} at the VS of n-MOSFETs with the tensile stress; the reflection coefficient r, instead, is not much improved by the stress. The advantages of the sGe are confirmed also for the MOSFETs scaled to 16 nm (see Fig. 6.10(b)). Fig. 6.10(a) finally shows that, if the R_{SD} of the Ge n-MOSFETs is increased by 50% with respect to the Si devices, then the I_{ON} improvement of the sGe is completely lost.

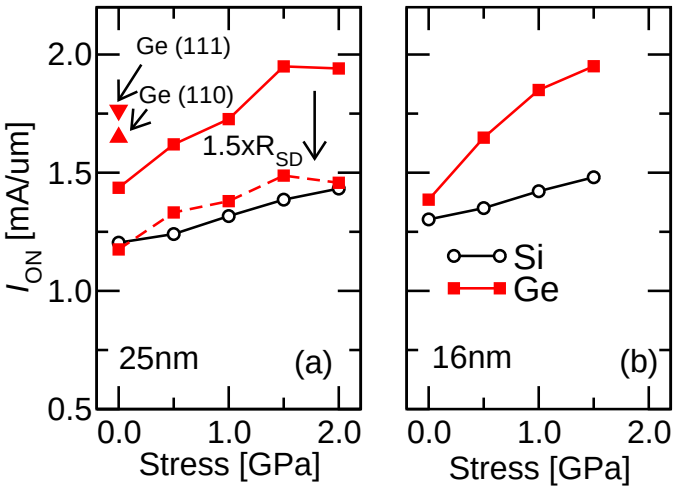

Figure 6.10: (a) I_{ON} vs. stress for Si and Ge n-MOSFETs; different crystal orientations for unstrained Ge are also shown. (b) Same as (a) for a 16 nm n-MOSFET (EOT=0.85 nm, t_{Si}=7 nm).

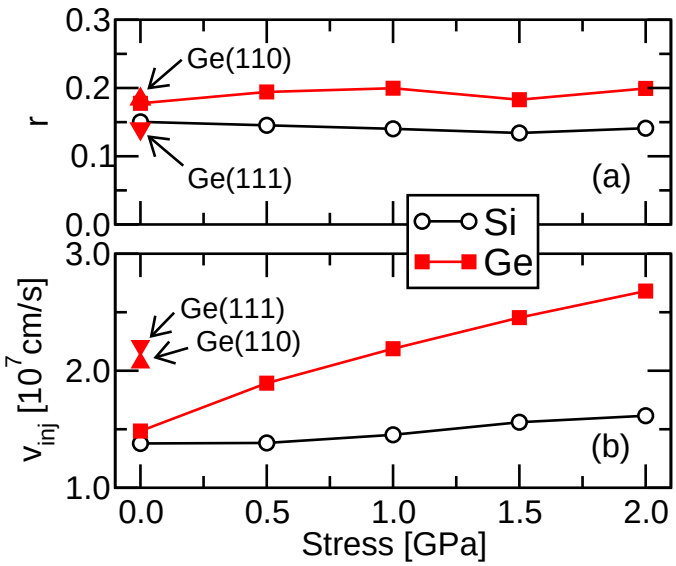

Figure 6.11: Injection velocity at the VS (a) and reflection coefficients (b) for some of the Si and Ge n-MOSFETs of Fig. 6.10(a).

6.4 Conclusions

In this chapter, we have first validated our Multi-Subband Monte Carlo model by means of comparison with with experimental mobility data. Then, we have simulated short channel MOSFETs showing that Ge MOSFETs are competitive with but do not outperform s-Si devices in terms of ON-current.

The s-Ge, instead, has great potentials for n-MOSFETs, however the engineering of the R_{SD} is a crucial issue to exploit the potential advantages of Ge transistors.

Chapter 7

Extracting and Understanding Carrier Velocity in nano-MOSFETs

Abstract

In this chapter, we use the accurate and calibrated transport model based on the Multi Subband Monte Carlo technique described in the previous chapters, to validate the extraction technique in [1] for the limiting velocity $v_{\rm lim}$ in nano-MOSFETs. We apply the technique to simulated $I_{\rm DS}-V_{\rm GS}$ curves and compare the extracted $v_{\rm lim}$ to the injection velocity $v_{\rm inj}$ given by the Monte Carlo. We critically discuss the validity of the experimental technique and we identify the main sources of error. Then, we propose a new $v_{\rm lim}$ extraction method and extensively validate it. Our results reconcile the values and trends of the $v_{\rm lim}$ with the expectations stemming from Quasi Ballistic transport theory.

7.1 Introduction

As discussed in Chap. 2, recent studies [83, 185, 186, 187, 188, 189] pointed out the importance of the injection velocity $v_{\rm inj}$ on the performance of nano-MOSFETs.

Since $v_{\rm inj}$ is significantly enhanced by strain [187, 188] these studies explain the effectiveness of strain technologies in improving the MOSFET on-current [185, 186]. Many techniques have been proposed to determine the $v_{\rm inj}$ in order to compare different technologies [190, 191, 189, 192, 193, 1]. In this context, Ref. [1] presented a promising technique to extract the $v_{\rm lim}$ which is here defined as the velocity that yields the current $I_{\rm DS,lim}$, according to:

$$I_{\rm DS,lim} = C_{\rm G} V_{\rm GT} v_{\rm lim} \tag{7.1}$$

where $I_{\rm DS,lim}$ is the upper bound of the $I_{\rm DS}$ obtained when $L_{\rm G}$ tends to zero, $V_{\rm GT}$ is $V_{\rm GS}-V_{\rm TH}$ and $C_{\rm G}$ is the effective gate capacitance. Both the Quasi-Ballistic (QB) and the drift diffusion (DD) models lead to Eq. 7.1 for $L_{\rm G}\rightarrow 0$ (Eqs. 2.32 and 2.19), but with $v_{\rm lim}$ being $v_{\rm inj}$ in the QB and $v_{\rm sat}$ in the DD picture. The $v_{\rm lim}$ extracted in [1] has been found higher than $v_{\rm sat}$, but with almost the same temperature dependence and an unexplained $L_{\rm G}$ dependence [1], thus raising doubts about the actual transport regime in nano-MOSFETs.

7.2 Review of existing extraction techniques

In [190] the backscattering model proposed in [82] has been extended to include the temperature dependency. Moreover, approximated temperature dependencies have been supposed for the $v_{\rm T}$, the $L_{\rm KT}$, and μ. Moreover, it has been assumed that the $N_{\rm inv,VS}$ in short channel devices in saturation is equal to the $N_{\rm inv}$ of long channel devices with $V_{\rm DS}=0$, i.e. Eq. 2.30. Then, $I_{\rm DS,lin}$ and $I_{\rm DS,sat}$ have been measured ($I_{\rm DS,sat}$ has been corrected to take into account the $R_{\rm SD}$ effect) on bulk devices down to 75 nm gate length. Finally, $V_{\rm TH,lin}$ has been extracted as the maximum of the $g_{\rm m}$ and $V_{\rm TH,sat}$ has been found by measuring the Drain Induced Barrier Lowering (DIBL) in sub-threshold regime. By means of linear extrapolation, it has been found that the value of the back-scattering ratio r (Eq. 2.35) should be close to 0.21 for 10nm long MOSFETs.

In [191] the ballisticity ratio has been found by comparison the between the measured $I_{\rm DS,sat}$ and the ballistic $I_{\rm bal,sat}$ predicted by a Schrödinger-Poisson solver [194]. For both the measures and the simulations, the $N_{\rm inv,VS}$ has been assumed to take its long channel expression (Eq. 2.30). It has been found that for 50 nm long MOSFETs, the measured $I_{\rm DS,sat}$ was up to 40% of the simulated $I_{\rm bal,sat}$, namely the backscattering coefficient r is about 0.43.

In [189] an improved version of the method proposed in [191] has been shown. In this paper, the measured $I_{\rm DS,sat}$ has been fitted with a 9-parameter model [195]. Of these parameters, 5 are measured: $C_{\rm OX}$, $I_{\rm OFF}$, DIBL, Sub-threshold swing and

$L_{\rm G}$. The adjustable parameters are the $R_{\rm SD}$, the μ, $v_{\rm inj}$ and the effective channel length $L_{\rm eff}$. The $I_{\rm bal,sat}$ is still calculated by means of a Schrödinger-Poisson solver. Using this method, it has been found that the ballisticity ratio of a 60 nm long MOSFET which operates in saturation regime is 0.72, namely r=0.16.

In [192] approximated expression for $v_{\rm inj}$ in linear and saturation regime are carried out assuming the quantum limit regime and strong degeneracy. After, these simplified expressions are corrected using f-functions that have been found in the ballistic regime. The $N_{\rm inv,VS}$ has been assumed to take its long channel expression (Eq. 2.30). With this approach, the backscattering r for 30 nm long MOSFETs has been found to be approximately equal to 0.33.

Finally, in [193] the quasi-ballistic model proposed in [82] has been extended to exactly take into account the two-dimensional nature of planar bulk structures. The $N_{\rm inv,VS}$ has been assumed to take its long channel expression (Eq. 2.30), but a direct evaluation of the strong inversion DIBL has been proposed. Using this method, the backscattering coefficient r for 70 nm long MOSFETs has been found to be approximately 0.60.

It is important to note the wide spread between the different backscattering coefficients r extracted with the different methods. Moreover, it is also worth noting that all these methods make the assumption of assuming that the $N_{\rm inv,VS}$ in short channel devices is equal to the $N_{\rm inv}$ in long the corresponding long channel device with $V_{\rm DS}$=0, but corrected to take into account the DIBL and the $V_{\rm TH,lin}$ roll-off in short channel devices. However, these corrections could be not sufficient, since the VS is usually placed close to the source region where the $N_{\rm inv}$ profile is steep in the x-direction [85].

7.3 The $v_{\rm lim}$ extraction procedure of [1]

The method starts by determining the series resistances [196], the gate capacitance $C_{\rm G}$, and the linear and saturation threshold voltages ($V_{\rm TH,lin}$ and $V_{\rm TH,sat}$) as in [197]. Then, normalized "apparent" mobilities are extracted using:

$$\left.\frac{\mu_{\rm app}}{L_{\rm G}}\right|_{\rm lin} = \frac{I_{\rm DS,lin}}{C_G V_{\rm GT,lin} V_{\rm DS}} \tag{7.2}$$

in linear regime, and:

$$\left.\frac{\mu_{\rm app}}{L_{\rm G}}\right|_{\rm sat} = \frac{2I_{\rm DS,sat}}{C_G V_{\rm GT,sat}^2} \tag{7.3}$$

in the saturation one. Eqs. 7.2 and 7.3 have been obtained by inverting the long channel DD model expressions (Eqs. 2.15 and 2.16), using $\mu_{\rm app}$ as a parameter to reproduce the experimental $I_{\rm DS}$−$V_{\rm GS}$ curves.

The velocity $v_{\rm lim}$ is finally deduced from the apparent mobilities according to:

$$v_{\rm lim} = \frac{e(V_{\rm GS} - V_{\rm TH,sat}) - 4K_{\rm B}T}{2e\left[\left.\frac{L_{\rm G}}{\mu_{\rm app}}\right|_{\rm sat} - \left.\frac{L_{\rm G}}{\mu_{\rm app}}\right|_{\rm lin}\right]} \tag{7.4}$$

Eq. 7.4 can be derived casting the expression for I_{DS} as:

$$I_{\mathrm{DS}}^{-1} = I_{\mathrm{DS,lim}}^{-1} + I_{\mathrm{DS,DD}}^{-1} \tag{7.5}$$

which is a combination of the long channel drift-diffusion current $I_{\mathrm{DS,DD}}$, namely:

$$\begin{cases} I_{\mathrm{DD,lin}} = \mu C_{\mathrm{G}} \dfrac{V_{\mathrm{GT,lin}} V_{\mathrm{DS,lin}}}{L_{\mathrm{G}}} \\ I_{\mathrm{DD,sat}} = \mu C_{\mathrm{G}} \dfrac{V_{\mathrm{GT,sat}}^2}{2L_{\mathrm{G}}} \end{cases} \tag{7.6}$$

and the $I_{\mathrm{DS,lim}}$ of Eq. 7.1.

Eq. 7.5 can be derived from the QB ballistic theory in the linear regime [83], using Eq. 2.32 as $I_{\mathrm{DS,lim}}$ and Eq. 2.39 as the effective I_{DS}, if we assume the back-scattering ratio r as in Eq. 2.49 and λ as in Eq. 2.48.

In the saturation regime, instead, Eq. 7.5 is valid using Eqs. 2.33 as $I_{\mathrm{DS,lim}}$ and 2.41 as I_{DS}, if we assume r as in Eq. 2.49, λ as in Eq. 2.48 and if we use the long channel expression for the KT-layer length as in Eq. 2.43, namely:

$$L_{\mathrm{KT}} = \frac{2L_{\mathrm{G}} K_{\mathrm{B}} T}{e(V_{\mathrm{GS}} - V_{\mathrm{TH,sat}})} \tag{7.7}$$

Eq. 7.5 can also be derived from the DD model with velocity saturation, provided the Caughey-Thomas model with β=1 [198] is used for the velocity-field relation and the electric field at the source end of the channel takes the form:

$$F_{\mathrm{S}} = \frac{V_{\mathrm{GS}} - V_{\mathrm{TH,sat}}}{2L_{\mathrm{G}}} \tag{7.8}$$

7.4 Comparison and calibration of the MSMC and T-CAD simulators

In this section we compare the results of the Monte Carlo simulator and of the T-CAD simulator.

We chose to simulate firstly the 15 nm technology node of the ITRS [10], which should enter in production in the year 2018.

After having created the 15 nm Double Gate (DG) MOSFET, we have obtained the other devices by changing only the L_{G} parameter, while the other values in Tab. 7.1 have been left the same for all the devices.

7.4.1 Series resistances

The T-CAD simulator intrinsically accounts for the series resistances of the Source and Drain regions due to Ionized Impurities scattering (Sec. 3.2.2). In the MSMC simulator, instead, the scattering mechanim is turned off in the Source and Drain regions, and the series resistances are taken into account in the simulations as

Structure	L_{G} [nm]	t_{Si} [nm]	EOT [nm]	V_{DD} [V]	R_{SD} [$\Omega\mu$m]	Type
DG-SOI	15	7	0.85	1.0	140	HP

Table 7.1: Parameters used for the calibration of the DG-SOI device on the 15 nm technology node [10].

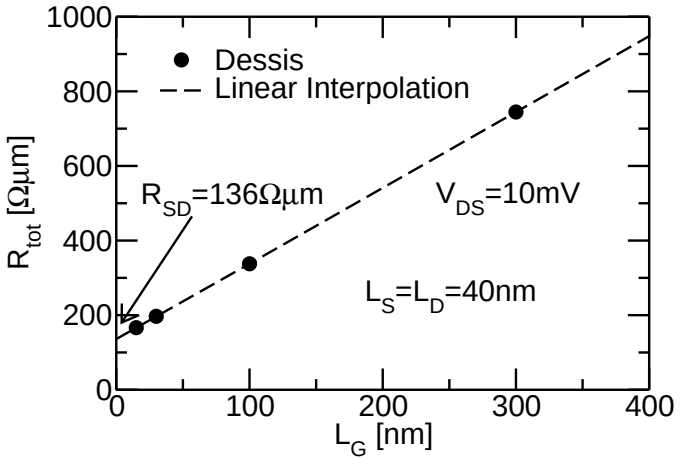

Figure 7.1: Extraction of the value of the serie resistances of the Source and Drain regions R_{SD}, in the T-CAD simulations. The exctracted value is in agreement with Tab. 7.1 [10].

lumped elements. This simplified approach has been validated by comparison with other MC simulators including proper modeling of impurity scattering in [146]. This simplification is further validated by the fact that in the Source and Drain regions the plasmon scattering is very effective in thermalizing the carriers [85]. Finally, we also note that the models for the ionized impurities scattering mechanism are developed for low doping concentrations, i.e. the channel doping rather than the Source and Drain ones in a MOSFET.

In order to make a fair comparison between the two simulators, we have however decided to de-embedd the series resistances R_{SD} from the T-CAD simulations. Firstly, we have exctracted the R_{SD} values from Fig. 7.1, and obtained 136 $\Omega\mu$m (the value is congruent with Tab. 7.1 [10]).

Then, once the R_{SD} is known, we have de-embedded this value from the $I_{\mathrm{DS}}-V_{\mathrm{GS}}$ of the T-CAD simulator by using:

$$\begin{cases} V^*_{\mathrm{GS}} = V_{\mathrm{GS}} - \dfrac{R_{\mathrm{SD}}}{2} I_{\mathrm{DS}} \\ V^*_{\mathrm{DS}} = V_{\mathrm{DS}} - R_{\mathrm{SD}} I_{\mathrm{DS}} \end{cases} \tag{7.9}$$

where the symbols are defined in Fig. 7.2.

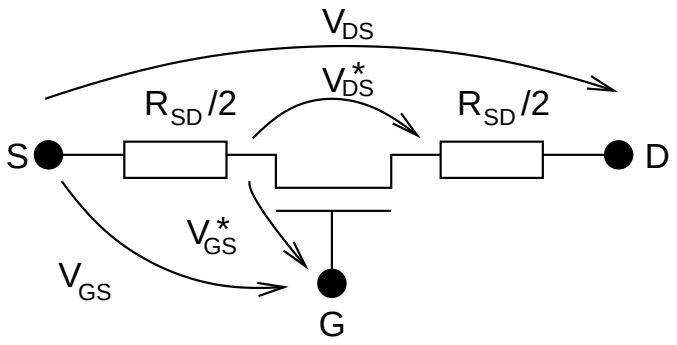

Figure 7.2: Sketch of a MOSFET when accouning for the series resistances of Source and Drain as lumped elements.

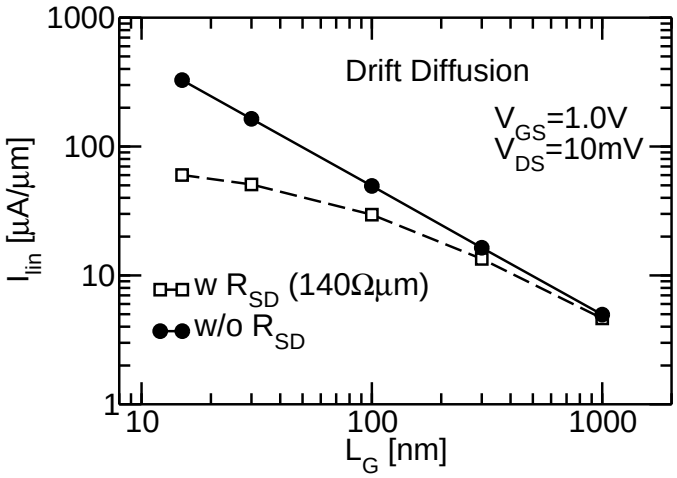

Figure 7.3: De-embedding of the serie resistances $R_{\rm SD}$ in the T-CAD simulations, in the linear regime.

The result is shown in Fig. 7.3, where the slope of the $I_{\rm DS,DD}$, without the $R_{\rm SD}$ effect, has the correct $1/L_{\rm G}$ behaviour. Fig. 7.3 also shows that the $R_{\rm SD}$ effect is important in even in the linear regime (i.e. low currents) in short devices.

7.4.2 Comparison of the quantization models

The MSMC uses in each section vertical section a self-consistent 1D Schrödinger-Poisson solver (see Sec. 3.1.2). The T-CAD simulator, instead, uses the Density Gradient Approximation (DGA) to take into account the quantization effects [199, 200]. Fig. 7.4 compares the $N_{\rm inv}$ as a function of the $V_{\rm GS}$ obtained with the MSMC and T-CAD simulations on long channel devices. The DGA gives reliable results in terms of $N_{\rm inv}$ and $V_{\rm TH}$.

However, Fig. 7.5 shows that, despite of the fair results in terms of $N_{\rm inv}$ ob-

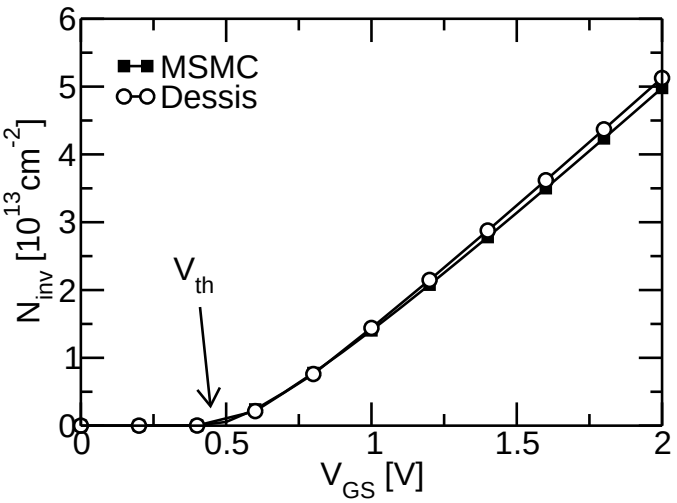

Figure 7.4: Comparison between the $N_{\rm inv}$ in linear regime ($V_{\rm DS}$=10 mV), in a long channel device ($L_{\rm G}$=1000 nm) obtained with the MSMC and with the T-CAD simulators. The value of the threshold voltages $V_{\rm TH}$ in the simulations is also shown.

tained with the GDA in all the $V_{\rm GS}$ range, the charge profile as a function of the vertical direction $n(z)$ is different between the MSMC and T-CAD simulations. The incongruence of the $n(z)$ obtained with the GDA with respect to the solution obtained with the MSMC (i.e. the 1D Schrödinger-Poisson solver), is due to the fact that the GDA has been calibrated on bulk devices, while the devices we are analyzing are DG-SOI.

7.4.3 Mobility models

The MSMC simulated mobility values are obtained computing the scattering rates of each perturbation mechanism (Sec. 3.2). The T-CAD simulator, instead, uses analytical models for the mobility computation, which are mainly function of the density and position of the carriers.

Fig. 7.6 compares the simulated mobility obtained with the T-CAD and MSMC simulators. We can see that the results provided by the two simulators are quite different. This result could be explained by the different charge profile between the two simulators shown in Fig. 7.5. Indeed, the carriers in the T-CAD simulation are closer to the Si/SiON interface, and this could result in a lower mobility due to the surface roughness scattering mechanism, that in the T-CAD simulator is computed as:

$$\mu_{\rm SR}(x,z) = B(T)\left(\frac{N_{\rm A}(x,z)+N_{\rm D}(x,z)+N_3}{N_4}\right)^b \frac{1}{E^{\gamma}_{\rm vertical}(x,z)} \qquad (7.10)$$

if using the *University of Bologna surface mobility model* [201]. In Eq. 7.10, $B(T)$,

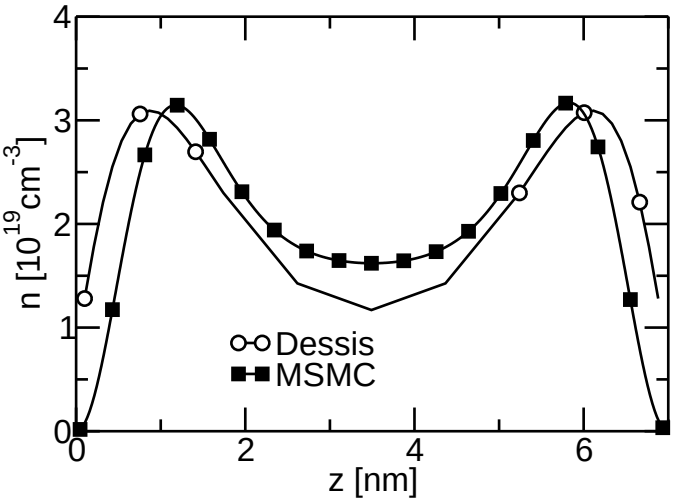

Figure 7.5: Comparison between the charge profile as a function of the vertical coordinate in the T-CAD and MSMC simulators, in a long channel device ($L_{\rm G}$=1000 nm) in linear regime ($V_{\rm DS}$=10 mV).

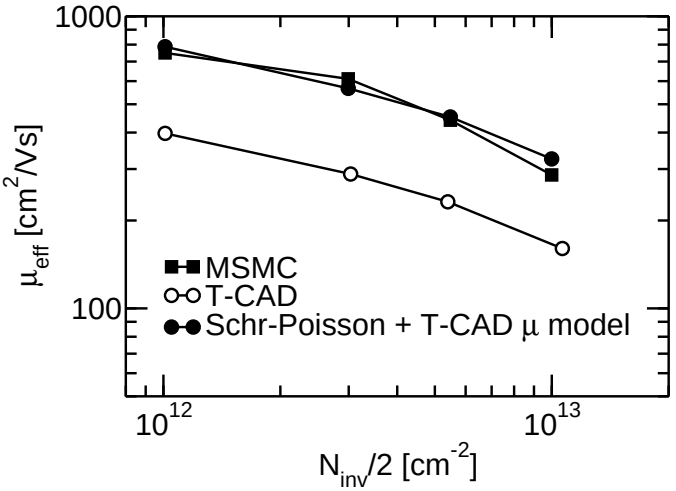

Figure 7.6: Comparison between the inversion layer mobility obtained with the MSMC and T-CAD simulators. The mobility obtained applying the T-CAD mobility models to the Schrödinger-Poisson solution of the MSMC is also shown.

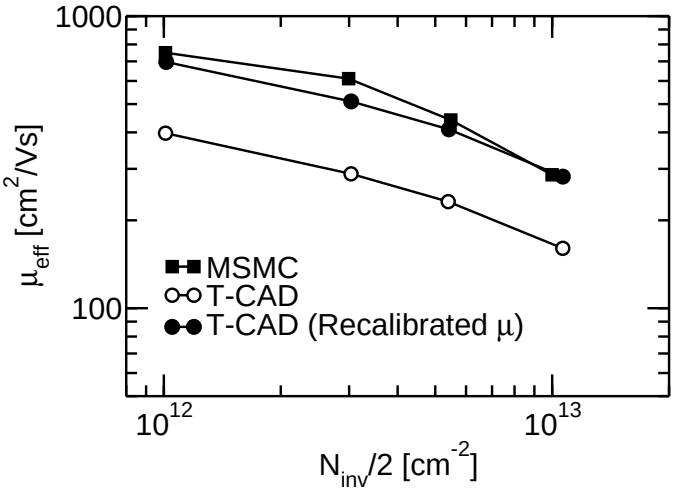

Figure 7.7: Comparison between the inversion layer mobility obtained with the MSMC and T-CAD simulators. The result obtained recalibrating the T-CAD mobility models on the MSMC results is also shown.

N_3, N_4, γ and b have been fitted on bulk structures [201].

To demonstrate this, we have applied the analytical mobility models used by the T-CAD for the phonon and surface roughness computation to the charge profile given by the Schrödinger-Poisson solver of the MSMC simulator. Using the Bologna mobility model, the overall mobility in the T-CAD simulator is computed as [201]:

$$\frac{1}{\mu(x,z)} = \frac{1}{\mu_{\rm PH}(x,z)} + \frac{1}{\mu_{\rm SR}(x,z)} \tag{7.11}$$

where $\mu_{\rm SR}(x,z)$ is defined in Eq. 7.10 and $\mu_{\rm PH}(x,z)$ is the phonon limited mobility, computed as:

$$\mu_{\rm PH}(x,z) = C(T)\left(\frac{N_{\rm A}(x,z)+N_{\rm D}(x,z)}{N_2}\right)^a \frac{1}{E^{\lambda}_{vertical}(x,z)} \tag{7.12}$$

where $C(T)$, N_2, λ and c have been fitted on bulk structures [201].

The mobility thus obtained is congruent with the one obtained with the MSMC (see Fig. 7.6). This validates our supposition that the incongruency between the MSMC and the T-CAD simulators in terms of mobility is due to the GDA, and not to the correctness of the analytical mobility models of the Drift-Diffusion simulator.

However, since the recalibration of the GDA method in our structure would be a time demanding activity, we chose to simply recalibrate the analytical mobility models of the T-CAD simulator, in order to have similar values of the mobility on long channel devices in the TCAD and MSMC simulations. Fig. 7.7 shows the recalibration of the T-CAD mobility models, obtained by doubling the values of $B(T)$ and $C(T)$ in Eqs. 7.10 and 7.12, respectively.

Structure	L_G [nm]	EOT [nm]	N_A [cm^{-3}]	V_{DD} [V]	R_{SD} [$\Omega\mu$m]	Type
Bulk	35	1.15	8×10^{17}	1.0	180	LSTP

Table 7.2: Parameters used for the calibration of the bulk device on the 35 nm technology node [10].

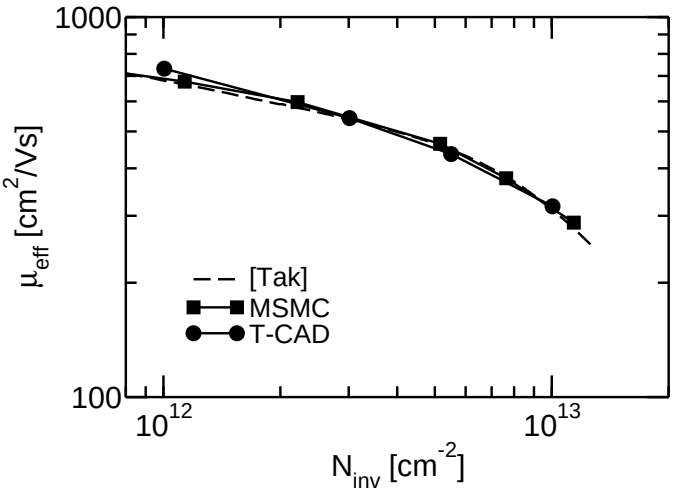

Figure 7.8: Comparison between the inversion layer mobility obtained with the MSMC and T-CAD simulators in an undoped bulk device. The universal mobility curve of [47] is also shown.

7.4.4 Bulk devices

Then, we chose to simulate also a LSTP bulk device calibrated on the 35 nm technology node of the ITRS [10]. The main parameters of this device are shown in Tab. 7.2.

The calibration of the T-CAD and the MSMC has been performed as for the DG-SOI. Thus, firstly we de-embedded in T-CAD simulator the R_{SD} resistences, that in this case are about 180 $\Omega\mu$m.

Then we verified that the N_{inv} provided by the T-CAD and MSMC simulators are similar. Since the DGA approximation of the T-CAD simulator is calibrated on bulk devices [201], we verified that in the analyzed structure the $n(x,z)$ profile in the z direction in long channel devices (L_G=1000 nm) in linear regime (V_{DS}=10 mV) is close to that the Schrödinger-Poisson solver of the MSMC.

Thus, the mobility provided by the T-CAD and MSMC simulators are close in the bulk structures (see Fig. 7.8), without the need of any fitting procedure.

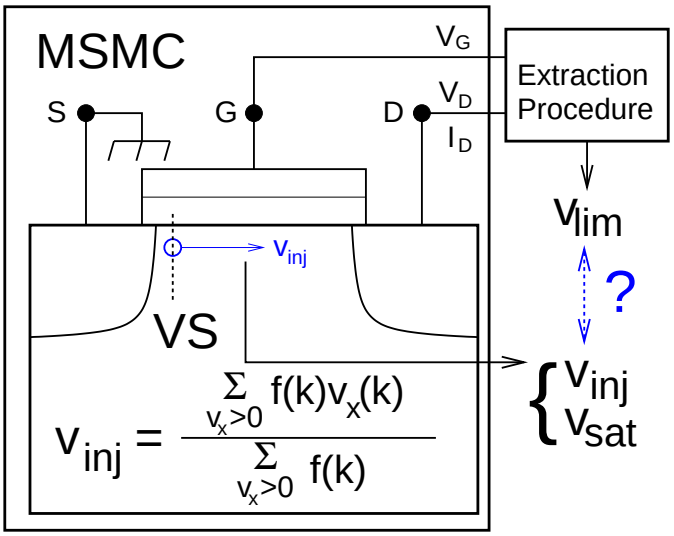

Figure 7.9: The extraction procedure of [1] is applied to the $I_{\rm DS}-V_{\rm GS}$ obtained from accurate MSMC simulations. The extracted $v_{\rm lim}$ is compared with the $v_{\rm inj}$ calculated from the carrier velocity distributions and with the $v_{\rm sat}$ used in TCAD simulations. VS denotes the Virtual Source.

7.5 Methodology

The extraction procedure of [1] has been applied to the $I_{\rm DS}-V_{\rm GS}$ calculated with accurate Multi Subband Monte Carlo (MSMC) simulations [93] of scaled DG (parameters in Tab. 7.1) and Bulk transistors (parameters in Tab. 7.2).

After having simulated the devices, we apply the method of Sec. 7.3 to extract the $v_{\rm lim}$. Then, we compare the extracted $v_{\rm lim}$ to the $v_{\rm inj}$ calculated from the MSMC simulator by means of:

$$v_{\rm inj} = \left.\frac{\sum_{v_x>0} f(\mathbf{k})v_x(\mathbf{k})}{\sum_{v_x>0} f(\mathbf{k})}\right|_{\rm VS} \tag{7.13}$$

where $f(k)$ is the distribution function in the k-space, and v_x is the component of the group velocity of the carriers in the Source to Drain direction. Eq. 7.13 states that the injection velocity is the average velocity of the carrier that are crossing the virtual source from the Source to Drain direction. $v_{\rm inj}$ is the velocity of Eq. 2.39, when the ditribution of the carriers at the VS does not follow a Maxwell-Boltzmann distribution. A sketch of the methodology of this work is shown in Fig. 7.9)

7.6 Analysis and results

We now simulate the devices in Tabs. 7.1 and 7.2 and we apply the extraction procedure of Sec. 7.3. Extracted $v_{\rm lim}$ values are directly compared with simulated

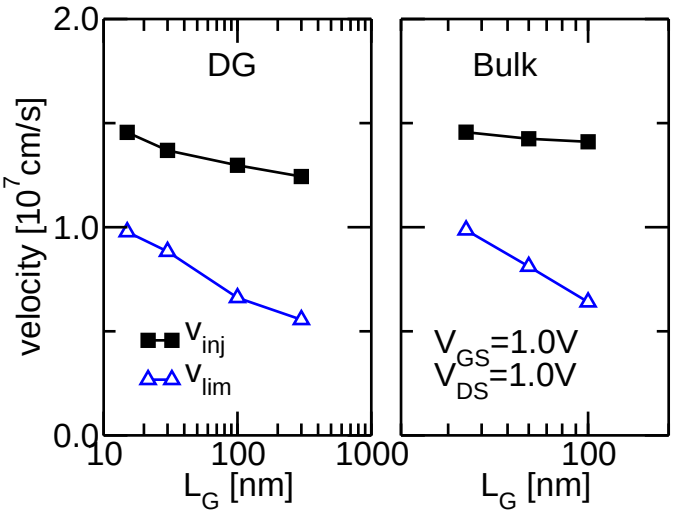

Figure 7.10: Comparison between $v_{\rm inj}$ calculated with the MSMC transport model and $v_{\rm lim}$ extracted as in [1] (Eq. 7.4) from the simulated MSMC $I_{\rm DS}-V_{\rm GS}$ curves. The inaccuracy of the extraction technique is evident.

$v_{\rm inj}$ in Fig. 7.10, showing a significant discrepancy. In particular, for $L_{\rm G}$>300 nm the extraction procedure is not reliable due to the extreme sensitivity of $v_{\rm lim}$ to the inaccuracy on the threshold voltage $V_{\rm TH,sat}$ (not shown). For shorter devices the inaccuracy of the $v_{\rm inj}$ extraction is due to the failure of Eq. 7.5. In fact, as shown in Figs. 7.11, 7.12 and Fig. 7.13, the current deduced from Eq. 7.5 is in acceptable agreement with MSMC simulations in the linear region, but not in the saturation region. Interestingly, the extracted $v_{\rm lim}$ shows the same unexplained gate length (Fig. 7.10) and temperature (Fig. 7.14) dependencies as the experiments in [1]. Fig. 7.15 shows that the $v_{\rm lim}$ increases with strain as expected for $v_{\rm inj}$ [185, 186], but the velocity values are significantly underestimated.

The method in [1] has been also applied to drift-diffusion simulations accounting for velocity saturation and calibrated as in Sec. 7.4. Fig. 7.16 clearly shows that the procedure fails to extract $v_{\rm sat}$ even when the limiting velocity has been forced to be $v_{\rm sat}$, due to the above mentioned unrealistic assumptions on β and $F_{\rm S}$ at the basis of the validity of Eqs. 7.4 and 7.5 in the DD framework.

Moreover, Fig. 7.17 shows that the velocity profile in MSMC simulations does not show any saturation, confirming that the concept of $v_{\rm sat}$ is not applicable to short devices.

We thus conclude that:

- the results in [1] do not prove that nano-MOSFET transport is limited by $v_{\rm sat}$;
- the $v_{\rm lim}$ extracted according to the procedure in [1] is at the best an inaccurate estimate of $v_{\rm inj}$ because Eq. 7.5 is inaccurate in the saturation region.

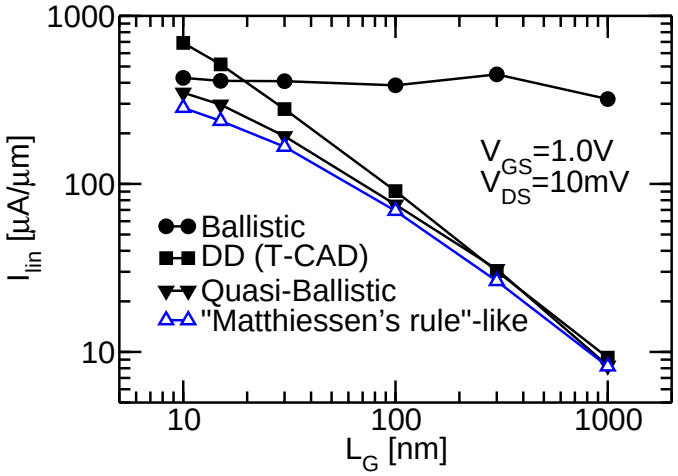

Figure 7.11: Comparison between Eq. 7.5 (open triangles) and the results of the MSMC simulations in linear region (closed triangles). $I_{\mathrm{DS,lim}}$ (filled circles) is taken from ballistic MSMC simulations (no scattering events); $I_{\mathrm{DS,DD}}$ is taken from TCAD simulations without velocity saturation (filled squares).

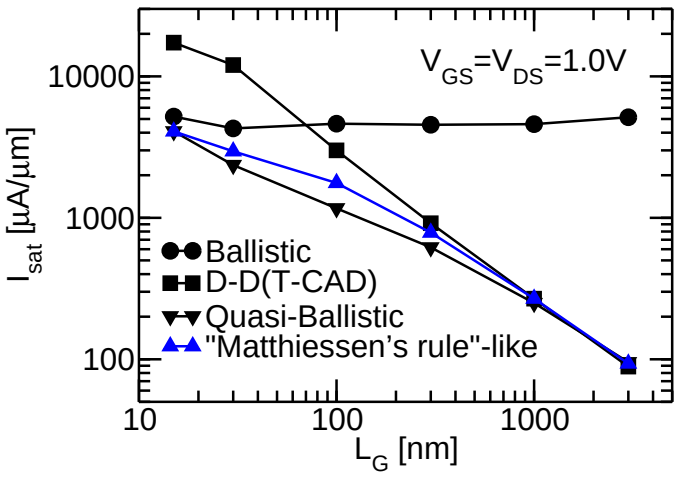

Figure 7.12: Same as in Fig. 7.11, but in saturation regime.

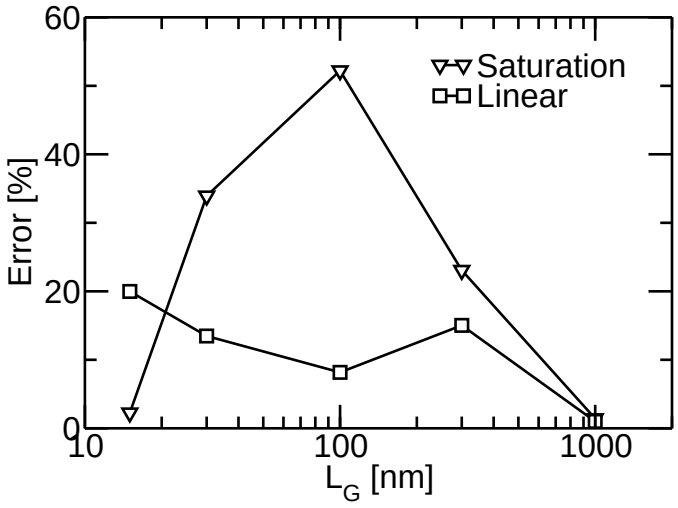

Figure 7.13: Percentage error of Eq. 7.5 with respect to the $I_{\rm DS}$ of the MSMC model for the linear and saturation regimes, as extracted from Figs. 7.11 and 7.12. The calibrated mobility models employed in the T-CAD (Sec. 7.4) yield almost zero error in long channel devices.

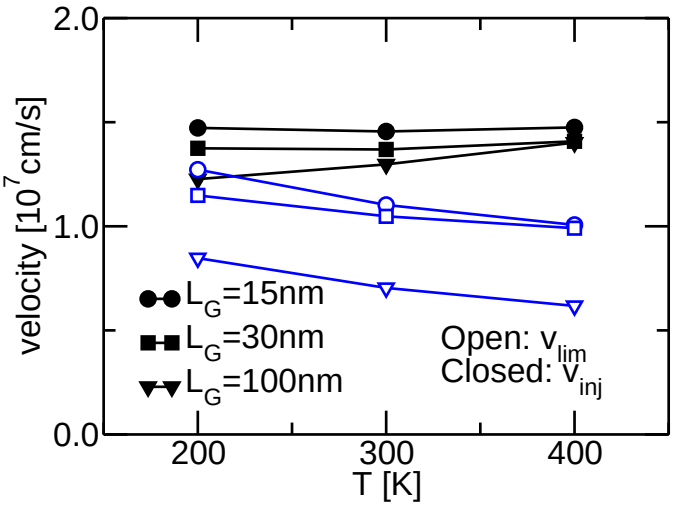

Figure 7.14: Same as in Fig. 7.10, but as a function of the temperature, for different gate length in the DG device. The decrement of $v_{\rm lim}$ with T is an artifact of the extraction procedure.

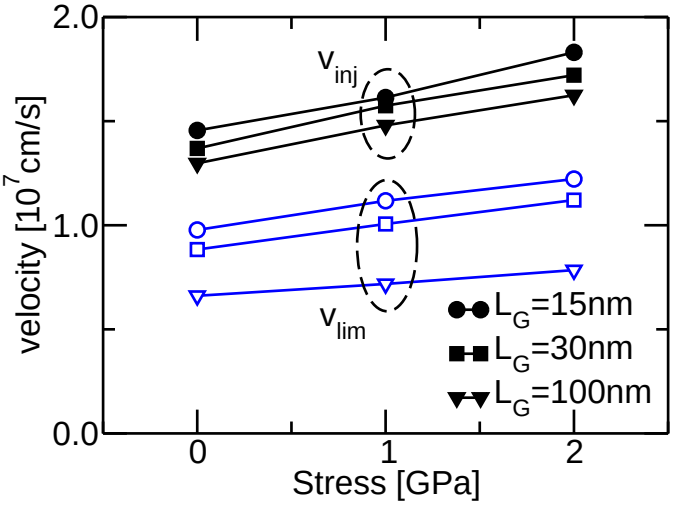

Figure 7.15: Same as in Fig. 7.14, but as a function of the uniaxial stress in the transport direction, for different gate length in the DG device. $v_{\rm lim}$ has the same strain dependence as $v_{\rm inj}$.

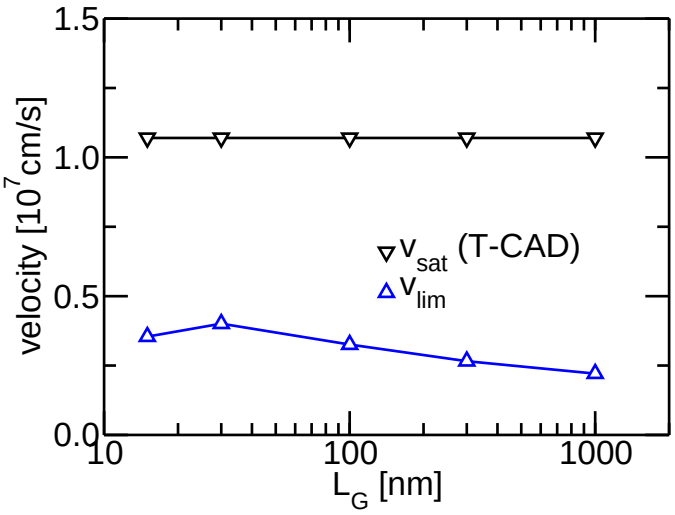

Figure 7.16: Comparison between the saturation velocity used in T-CAD simulations (down triangles) and the one obtaind applying the method in [1] (Eq. 7.4) to the same simulations. Although in this simulation $v_{\rm sat}$ is limiting the $I_{\rm DS}$ of the shorter devices (see open circles in Fig. 7.17), the extracted $v_{\rm lim}$ differs from $v_{\rm sat}$.

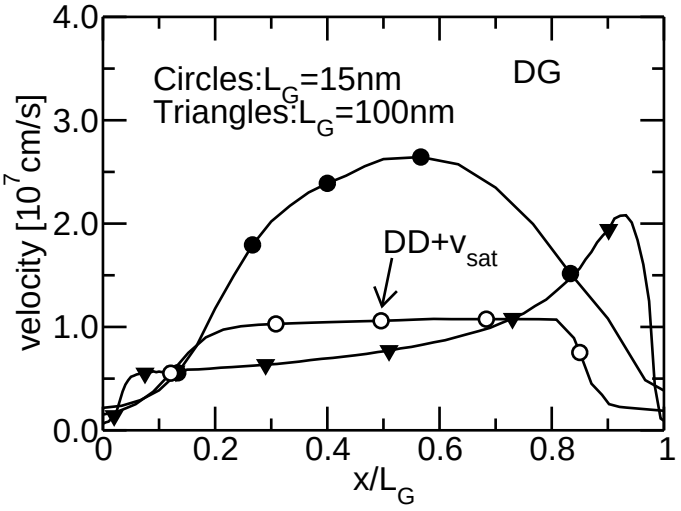

Figure 7.17: Drift velocity versus channel position normalized to $L_{\rm G}$ according to MSMC and DD simulations. x=0 is the source end of the channel. The VS is not at x=0 due to the presence of the gate overlap. Note the absence of any velocity saturated region in the MSMC results.

7.7 The new extraction procedure

Fig. 7.13 shows that Eq. 7.5 is in good agreement with MSMC simulations in saturation only when the $I_{\rm DS}$ is close to either $I_{\rm DS,lim}$ ($L_{\rm G}\leq$15 nm) or $I_{\rm DS,DD}$ ($L_{\rm G}\geq$300 nm). To better represent the intermediate QB regime we replace Eq. 7.5 with an empirical generalization, which is:

$$I_{\rm DS}^{-\alpha} = I_{\rm DS,bal}^{-\alpha} + I_{\rm DS,DD}^{-\alpha} \tag{7.14}$$

where α is a parameter to be determined.

Starting from Eq. 7.14, that is valid in saturation regime, we can write:

$$\frac{1}{I_{\rm DS,sat}} = \frac{1}{I_{\rm DD,sat}}\left(1+\left(\frac{I_{\rm DD,sat}}{I_{\rm bal,sat}}\right)^{\alpha}\right)^{\frac{1}{\alpha}} \tag{7.15}$$

Then, using Eqs. 2.32 and 2.16, we obtain:

$$\frac{1}{I_{\rm DS,sat}} = \frac{1}{I_{\rm DD,sat}}\left(1+\left(\frac{\mu V_{\rm GT}}{2L_{\rm G}v_{\rm T}}\right)^{\alpha}\right)^{\frac{1}{\alpha}} \tag{7.16}$$

Rembering Eq. 7.3 and noting that:

$$\left.\frac{L_{\rm G}}{\mu_{\rm app}}\right|_{\rm lin} = \frac{L}{\mu} + \frac{2K_{\rm B}T}{ev_{\rm T}} \tag{7.17}$$

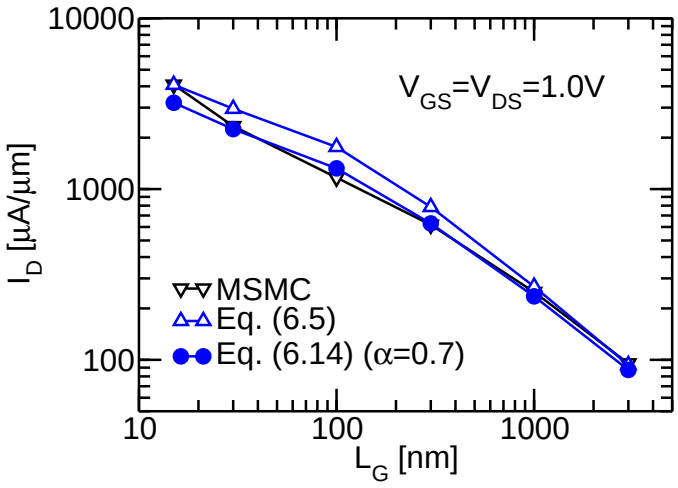

Figure 7.18: Comparison between MSMC simulations of the drain current in saturation and Eq. 7.14, allowing to calibrate $\alpha \simeq 0.7$.

we can obtain:

$$\left(\frac{L_{\rm G}}{\mu_{\rm app}}\bigg|_{\rm sat}\right)^{\alpha} = \left(\frac{L_{\rm G}}{\mu_{\rm app}}\bigg|_{\rm lin} - \frac{2K_{\rm B}T}{ev_{\rm T}}\right)^{\alpha} + \left(\frac{V_{\rm GT}}{2v_{\rm T}} \cdot \frac{L_{\rm G}}{\mu_{\rm app}}\bigg|_{\rm lin}\right)^{\alpha} \tag{7.18}$$

Finally, a new expression for the $v_{\rm lim}$ can be thus worked out:

$$\left(\frac{L/\mu_{\rm app}|_{\rm sat}}{L/\mu_{\rm app}|_{\rm lin}}\right)^{\alpha} = (1-\gamma)^{\alpha} + \left(\frac{e(V_{\rm GS}-V_{\rm TH,sat})}{4K_{\rm B}T}\gamma\right)^{\alpha} \tag{7.19}$$

where:

$$\gamma = \frac{2K_{\rm B}T}{ev_{\rm lim}}\frac{\mu_{\rm app}}{L}\bigg|_{\rm lin} \tag{7.20}$$

Figs. 7.18 and 7.19 show that $\alpha \simeq 0.7$ allows us to reproduce the whole $I_{\rm DS}$ vs. $L_{\rm G}$. Fig. 7.20 shows that the same α value yields good agreement between the extracted $v_{\rm lim}$ and the expected $v_{\rm inj}$ for the DG structure of Tab. 7.1. A slightly modified $\alpha \simeq 0.8$ yields satisfactory results also for bulk transistors of Tab. 7.2. Figs. 7.21 and 7.22 show that the $v_{\rm lim}$ extracted by the new method proposed here has the expected dependence upon $L_{\rm G}$, T and strain and it is in good quantitative agreement with the $v_{\rm inj}$ at the VS of the MSMC simulations.

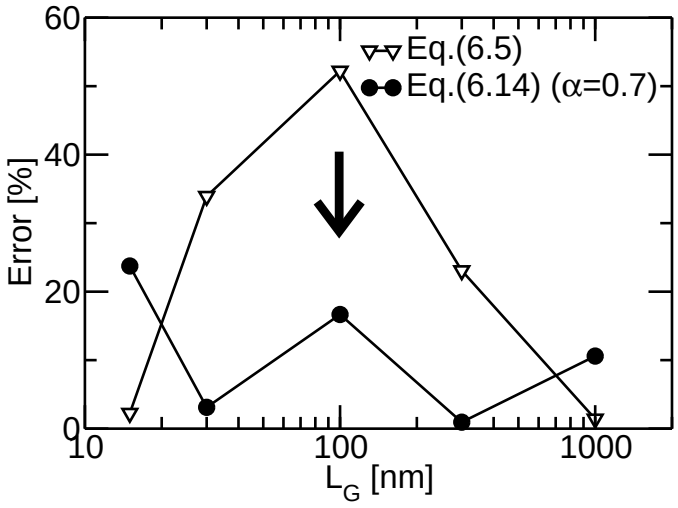

Figure 7.19: Percentage error of Eq. 7.5 (α=1 as in [1]) and Eq. 7.14 with α=0.7 in saturation regime. Eq. 7.14 reduces the errors to values comparable to those of the linear case in Fig. 7.13.

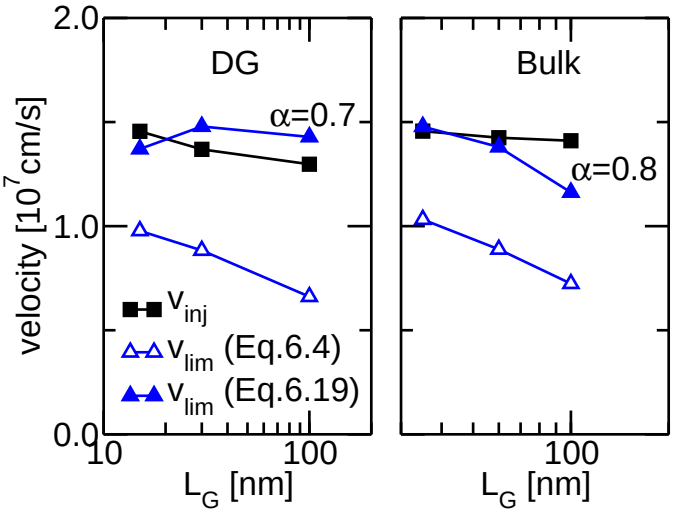

Figure 7.20: Comparison between the $v_{\rm inj}$ (v^+ at the VS of the MSMC), the $v_{\rm lim}$ obtained applying the method of [1] (Eq. 7.14) and the new procedure (Eq. 7.19). $\alpha\simeq$0.7 and 0.8 fits the DG and Bulk device data, respectively.

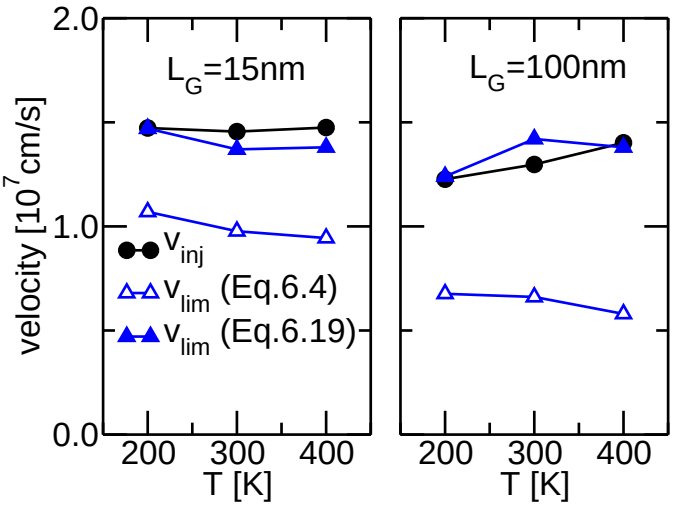

Figure 7.21: Comparison between the $v_{\rm inj}$ (MSMC) and the $v_{\rm lim}$ by the new method (with α=0.7) as a function of the temperature, for different gate lenght in the DG structure. The weak temperature variation of $v_{\rm inj}$ is better reproduced using the new method.

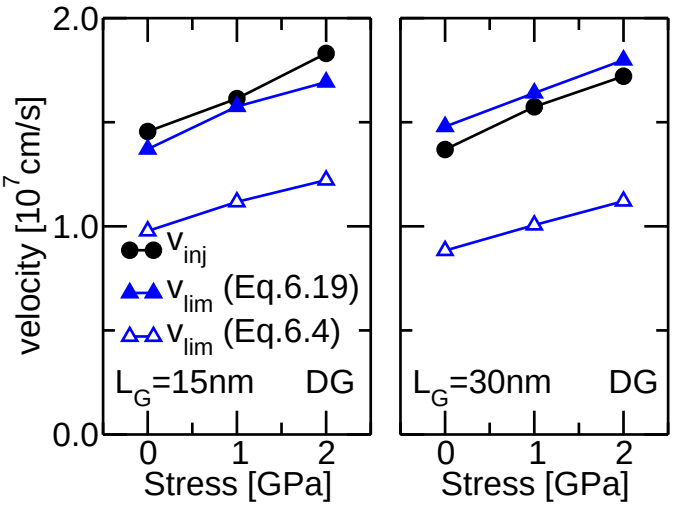

Figure 7.22: Comparison between the $v_{\rm inj}$ (MSMC) and the $v_{\rm lim}$ by the new method as a function of the uniaxial stress level in the transport direction, for different gate lenght in the DG structure. The strain variation of $v_{\rm inj}$ is better reproduced using the new method.

	28-LSTP
Structure	Bulk
Stress [GPa]	1.5GPa (Uni)
Oxide	SiON + HfSiON
EOT [nm]	1.11
$V_{\rm DD}$ [v]	1.0
$N_{\rm A}$ [cm^{-2}]	4×10^{17}
Type	LSTP

Table 7.3: Main parameters of the measured devices

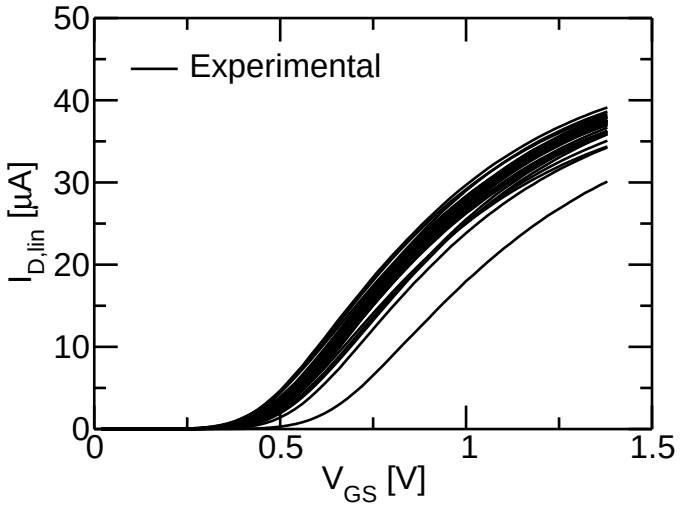

Figure 7.23: Measured currents in linear regime for the devices with $L_{\rm G}$=45 nm.

7.8 Improved method applied to experimental data

We have applied the new technique to bulk devices fabricated with a 28 nm LSTP technology (parameters in Tab. 7.3) [202] fabricated and measured in ST Microelectronics, Crolles. The devices are strained, n-MOS bulk devices. The gate lengths range from 40 nm to 1 µm.

7.8.1 Extraction in linear regime

The extraction starts with collecting the $I_{\rm DS}-V_{\rm GS}$ curves (see Fig. 7.23 for a 45 nm long device).

The second step is to calculate the transconductance $g_{\rm m}$, wich is defined as:

$$g_{\rm m} = \frac{\partial I_{\rm DS}}{\partial V_{\rm GS}} \tag{7.21}$$

Fig. 7.24 shows the $g_{\rm m}$ obtained from the curves in Fig. 7.23.

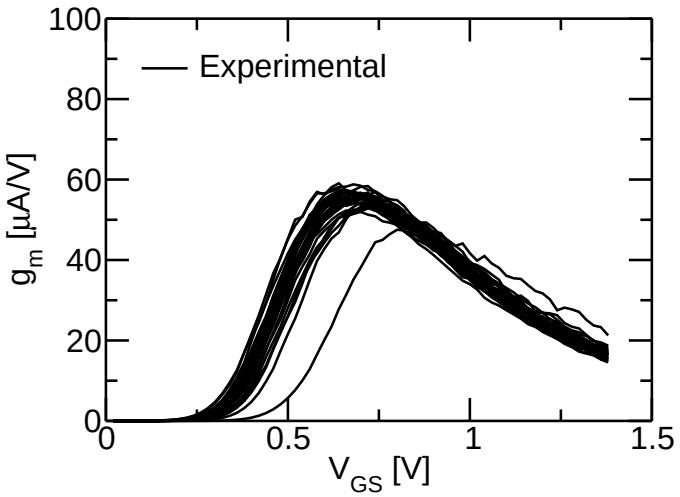

Figure 7.24: Measured transconductance $g_{\rm m}$ in linear regime for the devices with $L_{\rm G}$=45 nm.

The Y function

The second quantity we need is the so called Y function [197], which is defined as:

$$Y = \frac{I_{\rm DS}}{\sqrt{g_{\rm m}}} \tag{7.22}$$

from the experimental data in Figs. 7.23 and 7.24 we obtain the Y function for the 45 nm device reported in Fig. 7.25. Then, by means of:

$$Y = a_Y V_{\rm GS} + b_Y \tag{7.23}$$

we extract the slope a and the intercept b of the linearized Y function curves in Fig. 7.25. These two quantities are used to obtain the threshold voltage in linear regime $V_{\rm TH,lin}$, as:

$$V_{\rm TH,lin} = -\frac{b_Y}{a_Y} \tag{7.24}$$

Fig. 7.26 shows the results obtained applying Eq. 7.24 to the dashed curves in Fig. 7.25. Moreover, the coefficient a can be also used to extract the parameter β, using:

$$\beta = \frac{a_Y^2}{V_{\rm DS,lin}} \tag{7.25}$$

where β is the current factor of the transistor (see Eq. 7.31), namely:

$$\beta = \frac{W \mu_0 C_{\rm G}}{L} \tag{7.26}$$

where μ_0 is the carrier mobility without any dependence of the mobility on $E_{\rm eff}$ [196]. The extraction of β for the 45 nm long device is shown in Fig. 7.27.

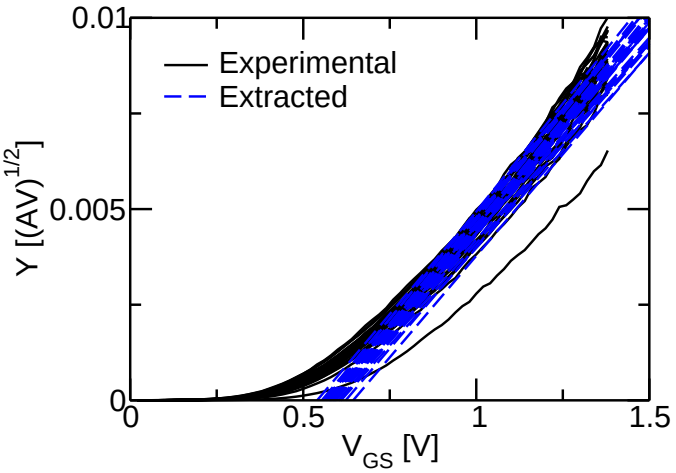

Figure 7.25: Solid lines: calculated Y function obtained from the curves in Figs. 7.23 and 7.24. Dashed lines: linearized Y function calculated in the linear region of the solid lines.

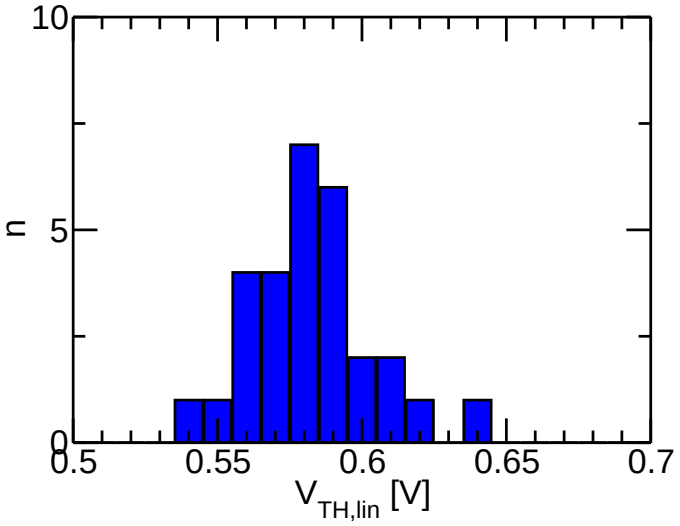

Figure 7.26: Threshold voltage in linear regime $V_{\mathrm{TH,lin}}$ for the curves in Fig. 7.23 corresponding to the devices with L_{G}=45 nm.

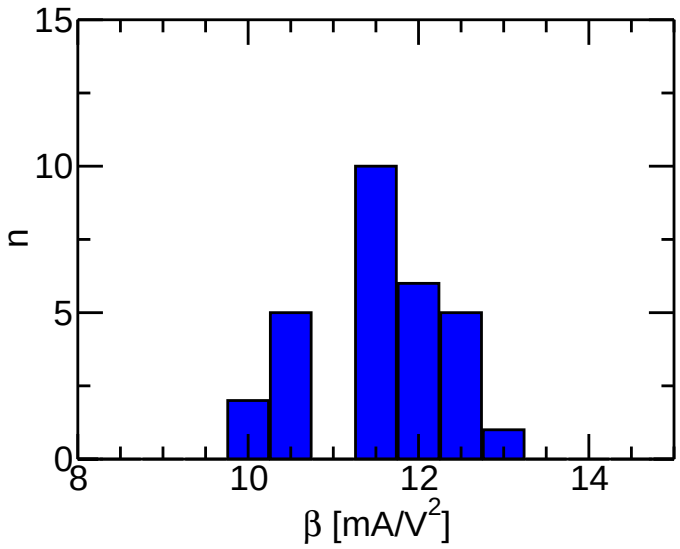

Figure 7.27: Extracted β for the curves in Fig. 7.23 corresponding to the devices with L_{G}=45 nm.

Once the value of β is known, we can also extract the R_{SD} of the devices, using the technique explained in [196]. Fig. 7.28 shows the extraction on the 45 nm long devices. If V_{GS}=V_{GT} (the gate polarization at which the v_{lim} extraction will be performed) and V_{DS}=10 mV, the R_{SD} value has been estimated to be 128.9 Ωμm.

The X function

Then, from the g_{m} curves, it is possible to extract the X function [197] as:

$$X = (\sqrt{g_{\mathrm{m}}})^{-1} = \left(\sqrt{\frac{\partial I_{\mathrm{DS}}}{\partial V_{\mathrm{GS}}}}\right)^{-1} \tag{7.27}$$

shown in Fig. 7.29 for the 45 nm long device.

Moreover, we can express the linearized X function defined in Eq. 7.27 (dashed lines in Fig. 7.29) as:

$$X = a_X V_{\mathrm{GS}} + b_X \tag{7.28}$$

The parameters a and b of Eq. 7.28 are the slope and the intercept of the curves in Fig. 7.29. Their determination allow to find the parameter θ_1 with:

$$\theta_1 = a_Y \cdot a_X \tag{7.29}$$

that will help us to fit the curves in Fig. 7.23 using Eq. 7.31.

Using θ_1 (from Eq. 7.29), β (from Eq. 7.25) and the value of the series resistances of the regions of source and drain R_{SD} (previously evaluated in Fig. 7.28), we can find $\theta_{1,0}$:

$$\theta_{1,0} = \theta_1 - \beta R_{\mathrm{SD}} \tag{7.30}$$

θ_1 and $\theta_{1,0}$ are reported in Figs. 7.30 and 7.31.

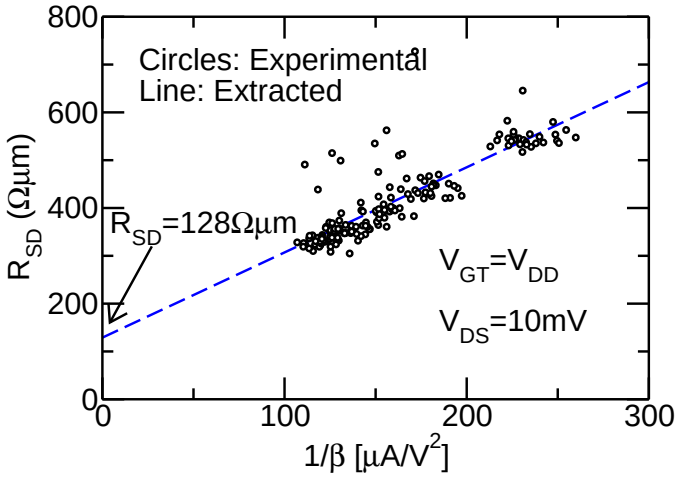

Figure 7.28: Extracted R_{SD} for the 45 nm long devices in Tab. 7.3.

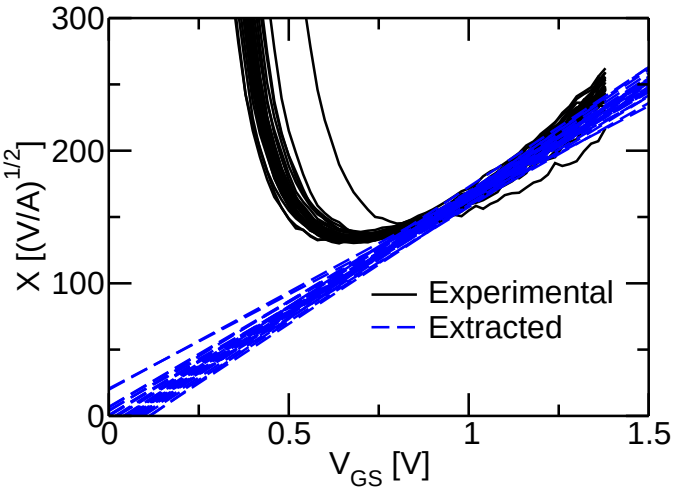

Figure 7.29: Assessment of the X function for the curves in Fig. 7.23.

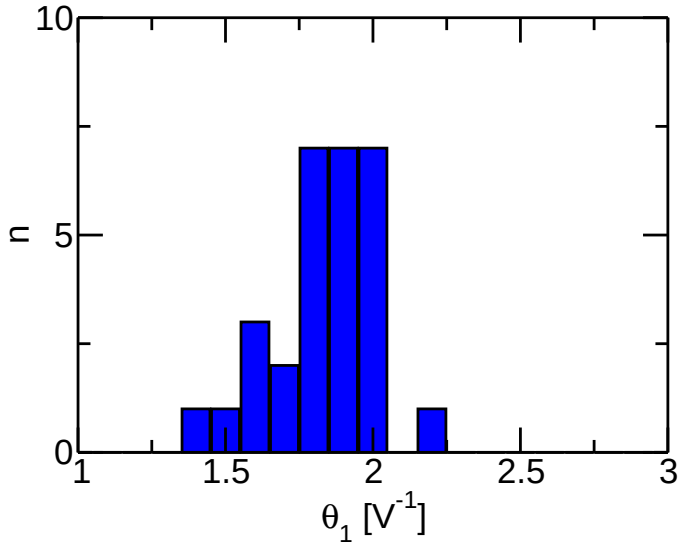

Figure 7.30: Extracted θ_1 for the curves in Fig. 7.23 corresponding to the devices with L_{G}=45 nm.

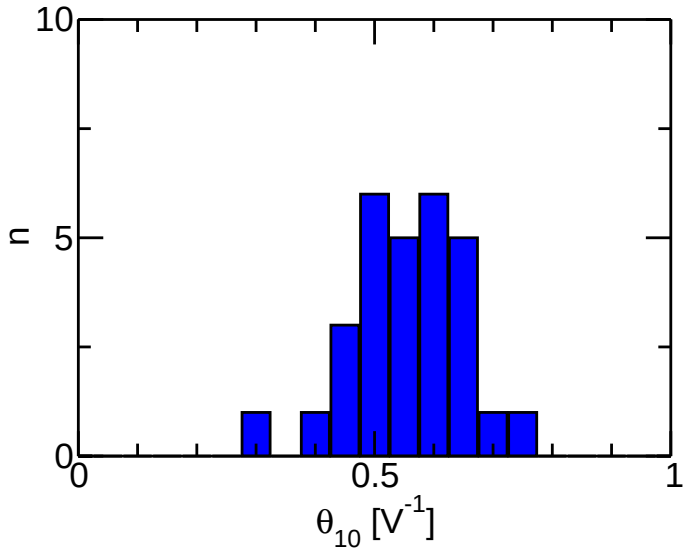

Figure 7.31: Extracted $\theta_{1,0}$ for the curves in Fig. 7.23 corresponding to the devices with L_{G}=45 nm.

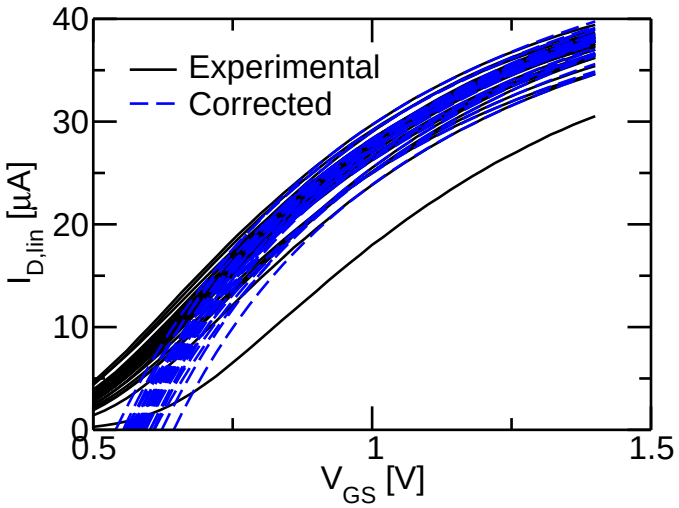

Figure 7.32: Comparison between the experimental $I_{\mathrm{DS,lin}}$ of Fig. 7.23 and the fitting obtained by using Eq. 7.31.

7.8.2 Corrected current in linear regime

By means of the quantities extracted in the first part of this Section, we can fit the linear current $I_{\mathrm{DS,lin}}$ with [197]:

$$I_{\mathrm{DS,lin}} = \frac{\beta V_{\mathrm{DS,lin}}(V_{\mathrm{GT,lin}})}{1 + \theta_1 V_{\mathrm{GT,lin}} + \theta_2 V_{\mathrm{GT,lin}}^2} \tag{7.31}$$

where θ_1 is a mobility attenuator factor which accounts for the dependence of μ on the E_{eff} and the presence of the R_{SD} [203]. θ_2 models the second order dependence of the mobility on the E_{eff} which is mainly observable when the gate dielectric is thin and the device is in strong inversion, i.e. high V_{GS} [197].

Fig. 7.32 shows the fitting of the curves in Fig. 7.23 by means of Eq. 7.31. We can see that the fit is satisfactory in the high V_{GS} regime.

To obtain the $I_{\mathrm{DS,lin}}$ corrected by the R_{SD} term, we can use:

$$I_{\mathrm{DS,lin}} = \frac{\beta V_{\mathrm{DS,lin}}(V_{\mathrm{GT,lin}})}{1 + \theta_{1,0} V_{\mathrm{GT,lin}} + \theta_2 V_{\mathrm{GT,lin}}^2} \tag{7.32}$$

that is similar to Eq. 7.31 but $\theta_{1,0}$ has been substituted by its corrected counterpart θ_1.

Thus, using:

$$I_{\mathrm{DS,lin}}^{\mathrm{corr}} = I_{\mathrm{DS,lin}} \left(\frac{1 + \theta_1 V_{\mathrm{GT,lin}} + \theta_2 V_{\mathrm{GT,lin}}^2}{1 + \theta_{1,0} V_{\mathrm{GT,lin}} + \theta_2 V_{\mathrm{GT,lin}}^2} \right) \tag{7.33}$$

it is possible to find the $I_{\mathrm{DS,lin}}^{\mathrm{corr}}$ which is the $I_{\mathrm{DS,lin}}$ of Fig. 7.23 after having been corrected from the mobility variations (the term in brackets in the right side of

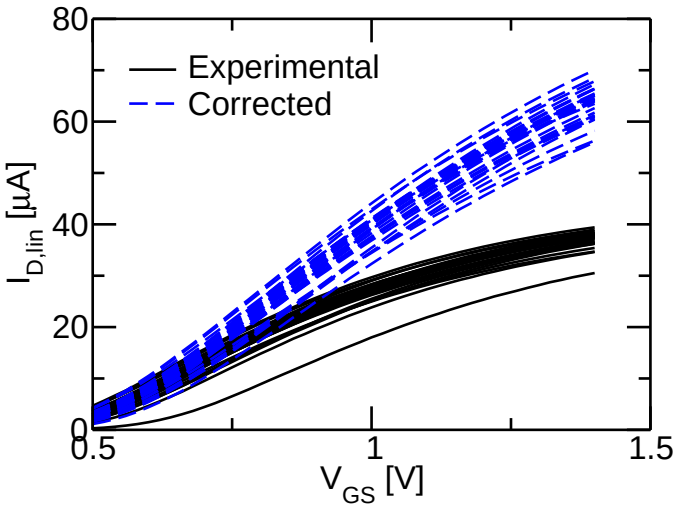

Figure 7.33: Comparison between the experimental $I_{\mathrm{DS,lin}}$ of Fig. 7.23 and its corrected version from the mobility variations and from the R_{SD} presence, obtained from Eq. 7.33.

Eq. 7.33) and from the de-embedding of the source-drain series resistances R_{SD} (the substitution of $\theta_{1,0}$ with θ_1).

7.8.3 Extraction in saturation regime

Threshold voltage shift

We chose to extract the threshold voltage in the saturation regime $V_{\mathrm{TH,sat}}$ as:

$$V_{\mathrm{TH,sat}} = V_{\mathrm{TH,lin}} - \Delta V_{\mathrm{TH}} \tag{7.34}$$

where ΔV_{TH} is the voltage shift that allow the $I_{\mathrm{DS,lin}}$ and the $I_{\mathrm{DS,sat}}$ in logarithmic scales to overlap in the sub-threshold regime. It has also been graphically defined in Fig. 7.34. Results are reported in Fig. 7.35.

Corrected current in saturation regime

In saturation regime, we only correct for the R_{SD} by defining a corrected gate voltage $V_{\mathrm{GS}}^{\mathrm{corr}}$, as:

$$V_{\mathrm{GS}}^{\mathrm{corr}} = V_{\mathrm{GS}} - \frac{R_{\mathrm{SD}}}{2} I_{\mathrm{DS,sat}} \tag{7.35}$$

Fig. 7.36 shows the comparison between the measured $I_{\mathrm{DS,sat}} - V_{\mathrm{GS}}$ characteristics and the characteristics obtained by applying the correction in Eq. 7.35.

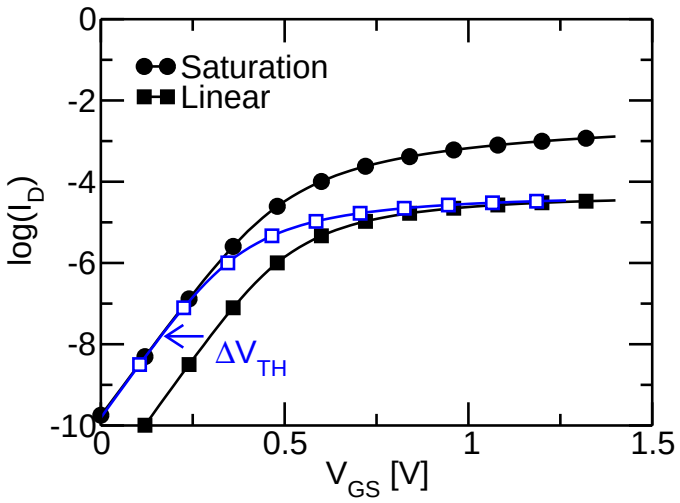

Figure 7.34: Comparison between the linear and saturation currents of one of the 45 nm devices. A graphic definition of the ΔV_{TH} is also shown.

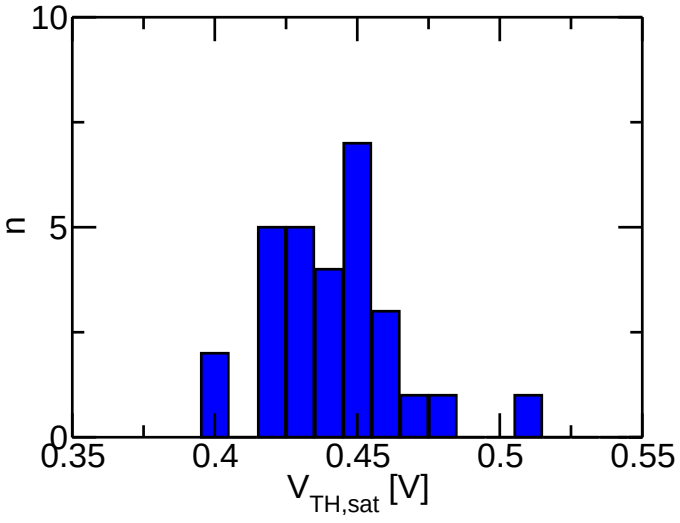

Figure 7.35: Threshold voltage in saturation regime $V_{\mathrm{TH,sat}}$ corresponding to the devices with L_{G}=45 nm, obtained by means of Eq. 7.34.

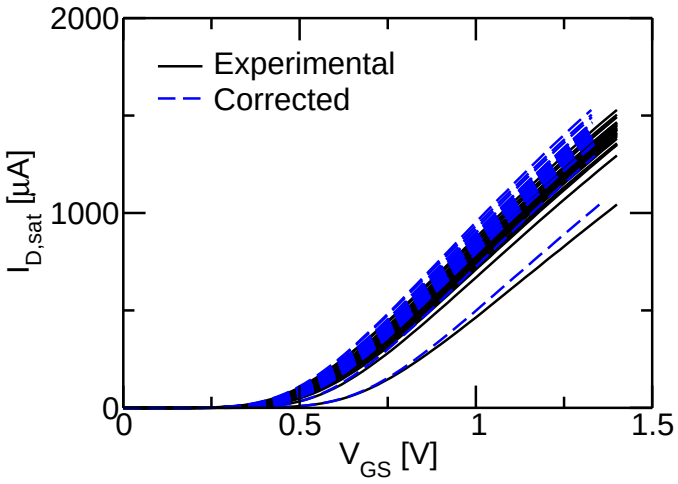

Figure 7.36: Comparison between the experimental $I_{\mathrm{DS,sat}}$ and its corrected version from the R_{SD} presence, obtained from Eq. 7.35.

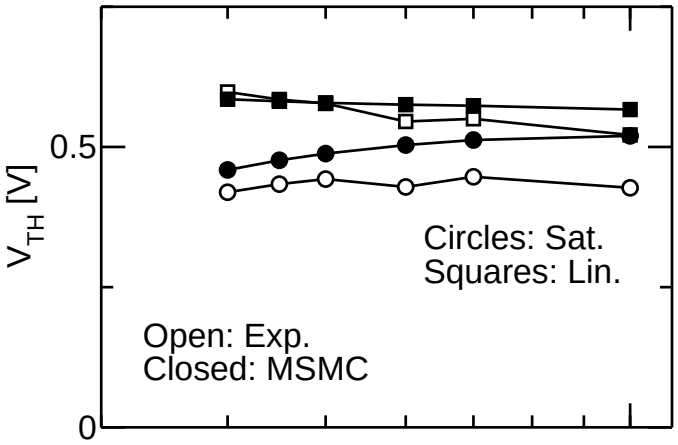

Figure 7.37: Comparison between experimental V_{TH} for the device described in Tab. 7.3 and MSMC simulations.

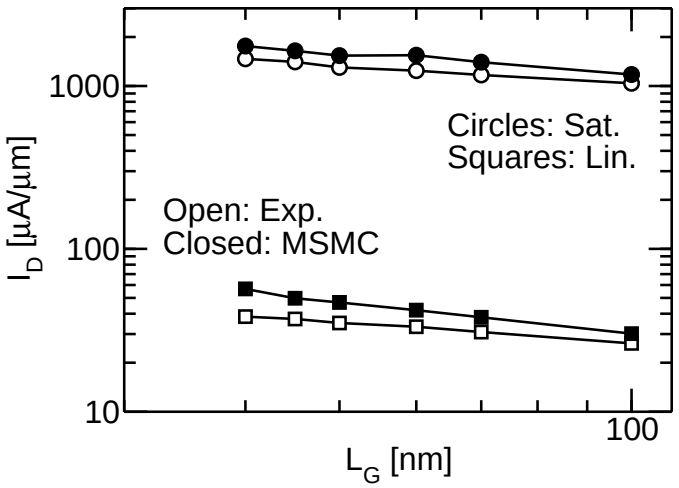

Figure 7.38: Comparison between experimental $I_{\rm DS}$ for the device described in Tab. 7.3 and MSMC simulations. Uniaxial strain (1.5 GPa) has been included in the MSMC as described in [185].

7.8.4 The extraction of the limiting velocity

We applied the procedure explained in Secs. 7.8.1 and 7.8.3 to the experimental data ranging from $L_{\rm G}$=40 nm to $L_{\rm G}$=100 nm. Figs. 7.37 and 7.38 shows the median values of the experimental $V_{\rm TH}$ and $I_{\rm DS}$ in both the linear and saturation regimes (open symbols).

We have exploited the experimental data in Figs. 7.37 and 7.38 as well as the data in Tab. 7.3 to calibrate the MSMC simulator. The default model parameters that reproduce the universal mobility curves of long devices [47] have been maintained (Tab. 3.1). Results of the MSMC calibration are shown in Figs. 7.37 and 7.38 (closed symbols).

The experimental data in Figs. 7.37 and 7.38 has been also used to extract the $v_{\rm lim}$ in the devices from ST using the method in [1] (Eq. 7.4) and the new method (Eq. 7.19), shown as up-triangles in Fig. 7.39. The filled circles, instead, report the calculated $v_{\rm inj}$ with the MSMC method. As can be seen, in spite of the measured $I_{\rm DS}$ being slightly smaller than the simulated one, the calculated $v_{\rm inj}$ is in excellent agreement with the extracted $v_{\rm lim}$ providing reassuring indications on the validity of the new extraction procedure This is an expected result because the accuracy of the calculated $v_{\rm inj}$ which is mainly set by the band structure and by the electrostatics at the VS and not by the absolute value of the drain current, that depends also on the rate of the scattering events in the channel.

Moreover, Fig. 7.40 compares the extraction of the $v_{\rm lim}$ using Eqs. 7.4 and 7.19 and the $v_{\rm inj}$ provided by the MSMC simulations, as a function of the temperature. Figs. 7.39 and 7.40 show that the new procedure yields significantly higher velocity, free of anomaly in the $L_{\rm G}$ or T dependencies.

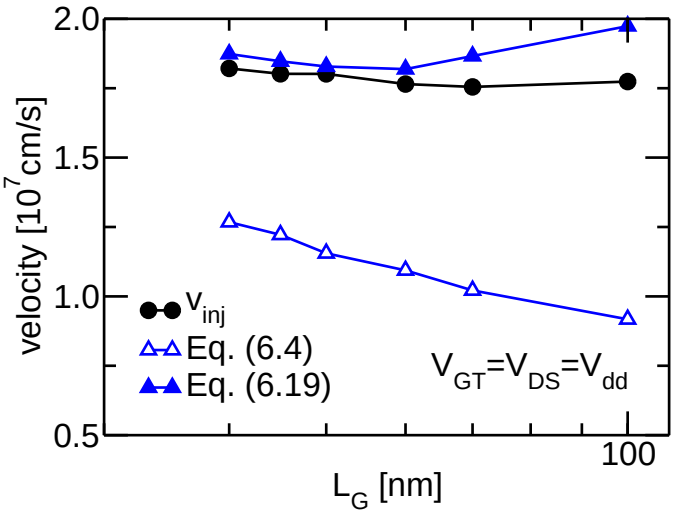

Figure 7.39: v_{lim} extracted using the old (Eq. 7.4) and the new (Eq 7.19) method from the experimental data on the devices of Tab. 7.3. The v_{inj} calculated with MSMC simulations is in good agreement with the v_{lim} extracted by the new method.

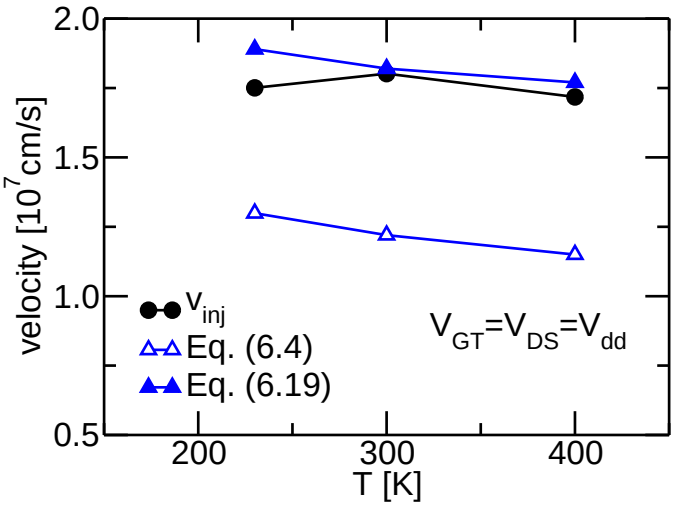

Figure 7.40: v_{lim} extracted using the old (Eq. 7.4) and the new (Eq. 7.19) method from the experimental data on the devices of Tab. 7.3. The v_{inj} calculated with MSMC simulations is in good agreement with the v_{lim} extracted by the new method.

7.9 Conclusions

A detailed analysis of the results in [1] based on an accurate transport model for nanoscale transistors led us to identify the limitations of the existing method and to propose a new extraction technique. The new method has been extensively validated by simulations and when applied to 28 nm technology devices provides realistic results.

Chapter 8

Conclusions

In this thesis we have used a Multi Subband Monte Carlo simulation framework to assess the effects of some of the Technology Boosters, (i.e. the techniques introduced in the fabrication process of the ultra scaled MOSFETs in order to continue the performance improvement beyond the classical *happy scaling* era) in terms of carrier mobility and ON-current. We have focused our analysis on multi-gate structures, high-k dielectrics, germanium channel devices and on the techniques to assess the advantages of the alternative channel materials from measured $I-V$ in short channel devices.

In more detail, in Chap. 4 we have examined the limits of validity of the scalar dielectric function approach for the screening in bulk, silicon-on-insulator and FinFET devices. After a detailed comparison between the Monte Carlo simulations with experimental data, we have found that the scalar dielectric function approach is inaccurate in multi-gate structures. Thus, in Double-Gate SOI and FinFET structures, the full tensorial dielectric function approach is needed.

The modeling of the high-k dielectric effects has been presented in Chap. 5. A physics-based modeling of the soft optical phonons has been derived. Then, a detailed analysis of the mobility reduction predicted by these models has been carried out. The same has been done for the remote Coulomb scattering mechanism. By comparing Multi Subband Monte Carlo simulations with experimental mobility from high-k MOSFETs fabricated in different $R\&D$ lines, we have shown that the effect of the soft optical phonons is negligible in silicon n- and p-MOSFETs. A large amount of Coulomb centers is needed to correctly reproduce the experimental data, which however would produce very large threshold voltage shifts inconsistent with the experiments, independently of the position of the charges in the gate stack. After having proposed and validated an original model for the dipole Coulomb scattering mechanism, we have demonstrated that this scattering mechanism can reproduce the experimental data, but however producing larger threshold voltage shifts than the single-charge Coulomb scattering mechanism. Next, we have shown that the effect of the high-k dielectric in ultra-scaled MOSFETs is lower than it is on long channel devices, but it cannot be neglected to correctly predict the ON-state

current.

In Chap. 6, Monte Carlo simulations show that germanium MOSFETs are competitive with but do not outperform strained-silicon devices in terms of ON-current. The strained germanium has great potentials for n-MOSFETs, however the engineering of the source-drain series resistances is a crucial issue to exploit the potential advantages of germanium transistors.

In Chap. 7, Monte Carlo simulations have been used to identify the sources of error in an existing experimental method to extract the limiting velocity in short channel MOSFETs. Then, we have proposed a new extraction method whose results, applied on simulations, are compatible with the Quasi-Ballistic transport theory. Moreover, when the new method has been applied to real short channel devices it has provided the same results as predicted by the Monte Carlo simulator for devices with strained-silicon channels.

To conclude, in this thesis we have seen the effectiveness of the semi-classical approach to assess the effects of the Technology Boosters in modern MOSFETs. We have shown that the Multi Subband Monte Carlo approach is able to quantitatively describe the physics that govern the charge transport in nano MOSFET devices, despite of its relative simplicity compared to much more complicated full quantum approaches.

Appendix A

The effects of wave function penetration into the high-k dielectric on the surface roughness limited mobility

In this Appendix we consider the surface roughness limited mobility variations when the wave function penetration is taken into accounting in MOSFETs featuring high-k dielectrics. Indeed, while the wave function penetration could be negligible when simulating SiO_2 MOSFETs with sufficiently thick t_{SiO_2}, this could be not correct for HfO_2 MOSFETs.

When simulating SiO_2 MOSFETs, it is commonly assumed that the potential barrier between the channel and the gate oxide is infinitely high. However, this approximation could be too strong for the HfO_2, because its potential barrier is lower than the one of SiO_2. Indeed, the potential barrier height between Si and SiO_2 is 3.1 eV [78] while the potential barrier height of the Si/HfO_2 interface is about 2.2 eV [204].

A.1 Evaluation of mobility when accounting for wave function penetration in high-k stacks

Fig. A.1 compares our results with those of [156] for a gate stack consisting of a thick layer of HfO_2 without ITL (comparison from Fig. 5.3a). In detail, the simulations in Fig. A.1 refer to a bulk MOS with doping 3×10^{17} cm^{-3} and with an infinitely thick HfO_2 layer as gate insulator (thus, without the ITL and the MG).

In this Appendix we focus on the additional points shown in Fig A.1. Precisely, the down triangles have been obtained allowing the wave-function to penetrate into the gate dielectric and using for the surface scattering mechanism the formulation in Eq. 3.36, instead of the one in Eq. 3.40. The observed mobility degradation is

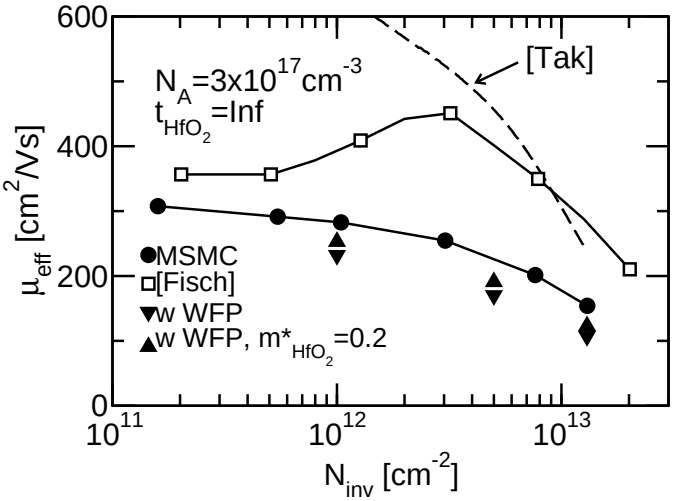

Figure A.1: Comparison between the electron mobility vs. $N_{\rm inv}$ obtained when the gate oxide is an infinitely thick layer of HfO_2 and the results in [156]. The down-triangles are obtained accounting for the wave-function penetration. The up-triangles are obtained obtained accounting for the wave-function penetration and considering that the relative electron mass in HfO_2 is 0.2 [205] (instead of 0.5 for the SiO_2 [78]).

about 25%.

However this mobility degradation is partially canceled if we consider that the electron mass in HfO_2 is lower than the one SiO_2, since the relative electron mass in the SiO_2 is 0.5 and the relative electron mass in the HfO_2 is about 0.2 [205]. Indeed, this feature lead to a lower mobility degradation of about 13% with respect to the one obtained without accounting for the wave-function penetration. This low percentage error suggests that the wave-function penetration can not sensibly degrade the mobility in real MOSFETs with the ITL.

However, Fig. A.1 shows also that the the gap between our results and the ones in [156] for values of $N_{\rm inv}$ around 4×10^{12} cm^{-2}, can not be attributed to the neglecting of the wave-function penetration. Indeed, the wave-function penetration leads to a decreasing of the mobility for high values of $N_{\rm inv}$. Instead, the model in [156] shows an electron mobility increase, which can be due to the image charges position variations when considering that the interface between the channel and the gate oxide is not ideal. However, a direct comparison is not possible since this effect is not modeled by our simulator.

Appendix B

The phonon-plasmon coupling

As it has been thoroughly discussed in [48], the polar phonon modes of the high-κ dielectrics can couple with the collective excitations of the carriers in the inversion layer and thus produce coupled phonon-plasmon modes. This is a known issue also in III-V polar semiconductors, where polar phonons are native in the semiconductor, rather than originating in an adjacent dielectric [206]. When such a phonon-plasmon coupling is important several relevant consequences come along [48]:

- the $\hbar\omega$ versus q relation of the coupled phonon-plasmon modes can be different with respect to the expressions obtained by neglecting the coupling (such as Eq. 5.17);
- the amplitude of the coupled modes can be enlarged by the anti-screening effect, namely by the fact that the dielectric function for silicon can be much smaller than $\epsilon_{Si,\infty} \simeq 11.7\epsilon_0$ and even negative;
- the phonon and plasmon content of the coupled modes must be distinguished in order to calculate the scattering rates.

The quantitative assessment of the possible role played by the phonon-plasmon coupling is a delicate issue. Even if the energy dispersion of the modes is typically obtained by considering only the real part of the electronic dielectric response, it is well known that, if the corresponding imaginary part is non negligible, then the supposedly coupled modes tend to vanish because of the Landau damping [48, 206].

We investigated the possible remote phonon-plasmon coupling in an inversion layer by using a numerical determination of both the real and the imaginary electronic response. To this purpose we assumed a simplified system consisting of the interface between a semiconductor and a dielectric, both with an infinite extension. In such conditions, the $\hbar\omega$ versus q relation for the coupled phonon-plasmon modes can be obtained by solving for ω the equation [48, 206]:

$$\epsilon_{(\infty)} + \frac{\epsilon_{(0)} - \epsilon_i}{1 - \left(\frac{\omega}{\omega_{TO1}}\right)^2} + \frac{\epsilon_i - \epsilon_{(\infty)}}{1 - \left(\frac{\omega}{\omega_{TO2}}\right)^2} + \Re\{\epsilon_{Si}(q,\omega)\} = 0 \qquad \text{(B.1)}$$

The silicon dielectric function $\epsilon_{si}(q,\omega)$ consists of both a real and an imaginary part and it was calculated by considering only intra-subbands polarizability and assuming that the intra subband form factors are approximately one [162]. The resulting, simplified expression for $\epsilon_{si}(q,\omega)$ reads [78]:

$$\epsilon_{Si} = \epsilon_{Si,\infty}\left[1 - \sum_{v,m} \frac{e^2}{2q\epsilon_{Si,\infty}} \Pi_{v,m,m}(q,\omega)\right] \tag{B.2}$$

where $\epsilon_{Si,\infty}$ is the optic dielectric constant of the silicon, v and s are respectively the valley and subband index, and the polarization factors $\Pi_{v,m,m}$ are defined as:

$$\Pi_{v,m,m}(q,\omega) = \frac{1}{A}\sum_{\mathbf{k}} \frac{f_0[E_{v,m}(\mathbf{k}+\mathbf{q})] - f_0[E_{v,m}(\mathbf{k})]}{E_{v,m}(\mathbf{k}+\mathbf{q}) - E_{v,m}(\mathbf{k}) - \hbar\omega - i\alpha\hbar} \tag{B.3}$$

where $E_{v,m}(\mathbf{k})$ is the energy at the wave-vector k in the subband (v,m), $f_0(E)$ is the Fermi-Dirac equilibrium occupation function and α is a positive real number that can be taken as vanishingly small in practical calculations [207, 206].

The polarization factors $\Pi_{v,m,m}(q,\omega)$ defined in Eq. B.3 were evaluated by converting the sum over **k** to an appropriate integral according to the standard prescriptions, and by employing a circular and parabolic energy relation, that makes the $\Pi_{v,m,m}(q,\omega)$ depend only on the magnitude q of the wave-vector **q** [78]. The real and the imaginary part of the $\Pi_{v,m,m}(q,\omega)$ are obtained by using the properties of the integral over **k** for a vanishingly value of α [207].

After the calculation of $\Pi_{v,m,m}(q,\omega)$ and thus $\epsilon_{Si}(q,\omega)$, Eq. B.1 was solved numerically to find the $\hbar\omega$ versus q relation for the coupled phonon-plasmon modes; the imaginary part of the silicon dielectric function $\epsilon_{Si}(q,\omega)$ was also monitored.

Fig. B.1 shows the energy dispersion of the modes, for $N_{inv}\simeq 3.2\times10^{12}$ cm^{-2} and for $N_{inv}\simeq 6.45\times10^{12}$ cm^{-2}, calculated by solving numerically Eq. B.1, compared to the q independent $\hbar\omega$ values obtained by neglecting the possible phonon-plasmon coupling (dot-dashed horizontal lines); the expressions for these latter $\hbar\omega$ values are given by Eqs. 5.3 and 5.4. Fig. B.1 also shows the boundaries of the Landau damping region (solid lines), defined as the region where the magnitude of the imaginary part of $\epsilon_{Si}(q,\omega)$ is larger than ϵ_0. We can see that the damped region is wider than in the case with T=0 (dot-dashed lines), that has been calculated as:

$$\hbar\omega = \frac{\hbar^2}{2m}(q^2 \pm q\sqrt{\pi N_{inv}}) \tag{B.4}$$

Fig. B.1 conveys at least two important messages:

- at the two considered N_{inv} the energy dispersion of the modes is very well approximated by the simple expressions obtained by neglecting the phonon-plasmon coupling;
- the energy dispersion belongs almost entirely to a region of the $\hbar\omega-q$ plane where the Landau damping is large. The second point reinforces the first one and indicates that the phonon-plasmon coupling is expected to have a modest role in the conditions at study.

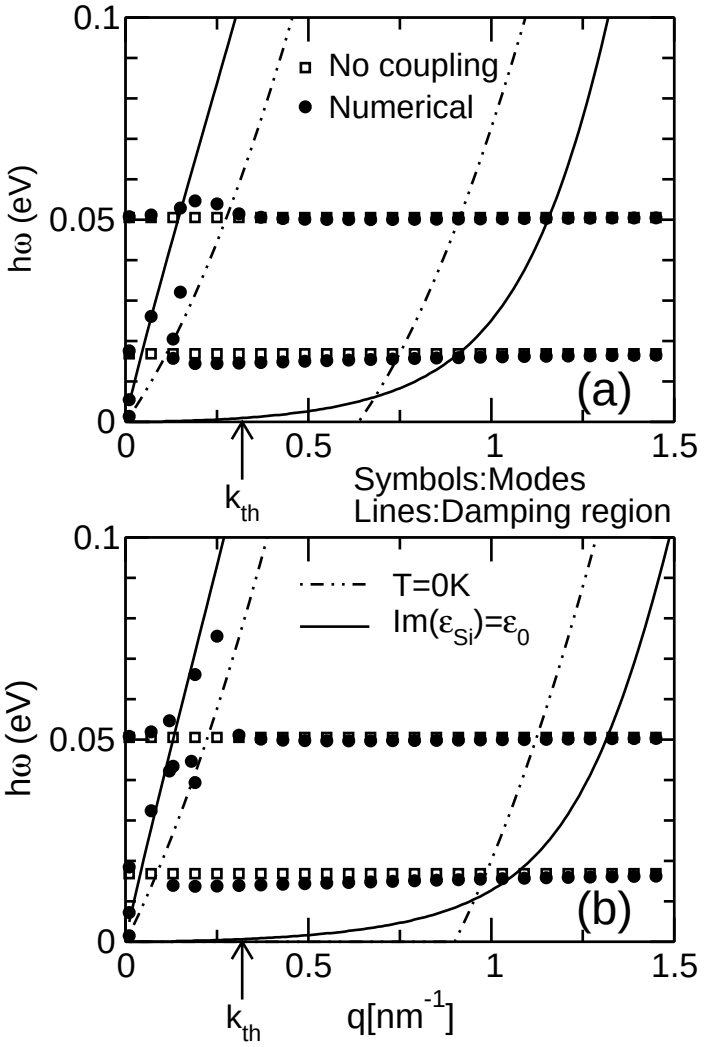

Figure B.1: Energy dispersion, $\hbar\omega$ versus q, for the coupled phonon-plasmon modes calculated by solving numerically Eq. B.1 (filled circles) in a structure with an infinitely thick HfO_2 layer on top of the bulk Si with N_A=2×10^{17} cm^{-3}. The open squares indicate the two $\hbar\omega$ values obtained by neglecting the phonon-plasmon coupling (Eqs. 5.3 and 5.4). The solid lines identify the boundaries of the Landau damping region assuming T=0 (Eq. B.4). The dot-dashed lines identify the boundaries of the Landau damping region, obtained as the region where the magnitude of the imaginary part of $\epsilon_{Si}(q,\omega)$ is larger than ϵ_0. The values of the thermal wave vector k at T=300K is also indicated ($k_{th}\simeq$0.317 nm^{-1} [162]). a) Inversion density $N_{inv}\simeq3.22\times10^{12}$ cm^{-2}. b) Inversion density $N_{inv}\simeq6.45\times10^{12}$ cm^{-2}.

An important consequence of this reasoning is that, because of the damping, the anti-screening of the electronic response is also expected to be small for the $\hbar\omega{-}q$ values of most practical relevance, so that the expressions for the phonon amplitudes reported in Sec. 5.2.3 are quite defendable approximations.

The analysis reported in Fig. B.1 was repeated also for different inversion densities. As a general trend, the phonon-plasmon coupling effects appear to have a more visible effect on the energy dispersion with the increase of N_{inv}, however, even at large inversion densities close to 10^{13} cm^{-2}, almost the entire $\hbar\omega{-}q$ relation belongs to a region where the imaginary part of the $\epsilon_{Si}(q,\omega)$ is significant, and consequently the Landau damping is expected to be large.

The results and discussion in this appendix support the simplifying assumptions for the modeling of remote phonons embraced in Sec. 5.2 and, in particular, we believe that the neglect of the phonon-plasmon coupling used in our simulations does not play a critical role in the reaching of the main conclusions of the paper.

Appendix C

Numerical algorithm for the determination of the phonon modes in generic gate structures

In this appendix we describe a numerical algorithm which allows to find the dispersion relation of the phonon modes which can originate in a gate structure composed of an arbitrary number of dielectric layers.

This algorithm has been used to generate the Figs. 5.7, 5.8 and 5.9 in Chap. 5.

C.1 General structure approach

We consider a generic interface i of an arbitrary gate structure, as shown in Fig. C.1. The two dimensional Fourier transform of the scattering potential (as defined in Eq. 3.22) in the layers above and below the interface can be expressed, according to Eq. 5.15, as:

$$\begin{cases} \phi_{\mathrm{SO},i}(q,z) = A_{i,1}e^{-qz} + A_{i,2}e^{qz} \\ \phi_{\mathrm{SO},i+1}(q,z) = A_{i+1,1}e^{-qz} + A_{i+1,2}e^{qz} \end{cases} \tag{C.1}$$

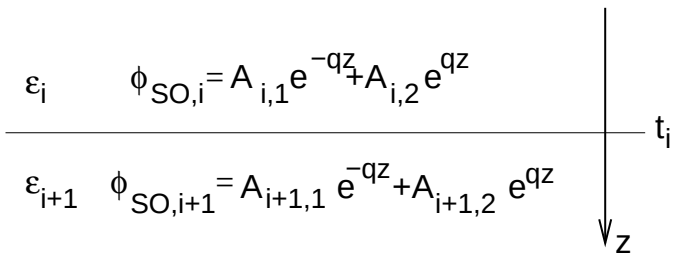

Figure C.1: Generic interface between two layers of an arbitrary gate structure. The electric permittivities ϵ_i and the scattering potentials $\phi_{\mathrm{SO}}(q,z)$ of the two materials are also shown.

At the section t_i, we can impose the continuity of the potential and the continuity of the displacement field, obtaining:

$$\begin{cases} A_{i,1}e^{-qt_i} + A_{i,2}e^{qt_i} = A_{i+1,1}e^{-qt_i} + A_{i+1,2}e^{qt_i} \\ \epsilon_i(A_{i,2}e^{qt_i} - A_{i,1}e^{-qt_i}) = \epsilon_{i+1}(A_{i+1,2}e^{qt_i} - A_{i+1,1}e^{-qt_i}) \end{cases} \tag{C.2}$$

Afer some manipulations, from (C.2) we have:

$$\begin{pmatrix} A_{i+1,1} \\ A_{i+1,2} \end{pmatrix} = \begin{pmatrix} \frac{\epsilon_{i+1}+\epsilon_i}{2\epsilon_{i+1}} & \frac{\epsilon_{i+1}-\epsilon_i}{2\epsilon_{i+1}}e^{2qt_i} \\ \frac{\epsilon_{i+1}-\epsilon_i}{2\epsilon_{i+1}}e^{-2qt_i} & \frac{\epsilon_{i+1}+\epsilon_i}{2\epsilon_{i+1}} \end{pmatrix} \begin{pmatrix} A_i, 1 \\ A_i, 2 \end{pmatrix} \tag{C.3}$$

This matrix allows to find the amplitude of the potential in the $(i+1)$-th layer once we know the amplitude of the potential in the i-th layer. Thus, if the gate structure is composed by N layers of dielectric, we can relate the amplitude of the potential in the $(N+1)$-th layer with the amplitudes in the section 0 with the equation:

$$\begin{pmatrix} A_{N+1,1} \\ A_{N+1,2} \end{pmatrix} = M_{\mathrm{TOT}}(\omega) \begin{pmatrix} A_{0,1} \\ A_{0,2} \end{pmatrix} \tag{C.4}$$

where M_{TOT} is simply the multiplication of all the matrices M_i (shown in Eq. C.3) of each i-th section.

Since the relative permittivities ϵ_i of the materials depend on the pulsation ω, also M_{TOT} is function of the pulsation.

C.2 Solution in the case of infinitely thick dielectric

Now we look for the modes which can originate in a gate structure composed by N different material layers, with on top an infinitely thick dielectric and to the bottom an infinitely thick bulk MOSFET channel. In this structure we have $N+1$ discontinuity interfaces, as shown in the first image of Fig. C.2. We can assume that the amplitudes $A_{0,2}$ and $A_{N+1,1}$ of the scattering potential are null, in order to avoid infinite power fields.

Thus, from Eq. C.4, we obtain in this case:

$$M_{\mathrm{TOT},11}(\omega) = 0 \tag{C.5}$$

Thus, to find the dispersion relation in this structure, for each value of the wave-vector q, we have to determine the values of the pulsation ω which verify Eq. C.5. Once established the singularities of the function $M_{\mathrm{TOT},11}(\omega)$ we can apply the Newton's method to find the values $\omega_{\mathrm{SO}}(q)$ which obey Eq. C.5.

C.3 Solution in the case of metal gate electrode

Now we look for the modes which can originate in a gate structure composed by $N+1$ different material layers, with on top the metal gate and to the bottom an

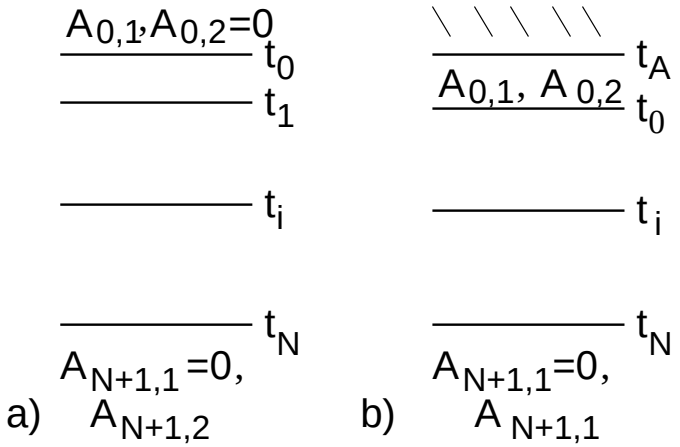

Figure C.2: a) Sketch of a gate structure with $N+1$ discontinuity interfaces, with on top an infinitely thick dielectric. b) Sketch of a gate structure with $N+2$ discontinuity interfaces, with on top the metal gate. The amplitudes of the scattering potential ϕ_{SO} in the dielectric in the highest position and in the MOSFET channel are also shown.

infinitely thick bulk MOSFET channel. In this structure we have N+2 discontinuity interfaces, as we can see in the second image of Fig. C.2. At the interface t_A (the interface between the metal gate and the layer at the bottom) we have that the potential has to be null, i.e.:

$$\phi(-t_A) = A_{0,1}e^{qt_A} + A_{0,2}e^{-qt_A} = 0 \tag{C.6}$$

and thus:

$$A_{0,2} = -A_{0,1}e^{2qt_A} \tag{C.7}$$

Since $A_{N+1,1}$ has to be null and remembering Eq. C.4, we can impose:

$$M_{TOT,11}(\omega) - M_{TOT,12}(\omega)e^{2qt_A} = 0 \tag{C.8}$$

Thus, to find the dispersion relation in this structure, for each value of the wavevector q, we have to determine the values of the pulsation ω which verify Eq. C.8. Once established the singularities of the function $M_{TOT,11}(\omega) - M_{TOT,12}(\omega)e^{2qt_A}$ we can apply the Newton's method to find the values $\omega_{SO}(q)$ which obey Eq. C.8.

C. Numerical algorithm for the determination of the phonon modes in generic gate structures

Bibliography

[1] D. Fleury, G. Bidal, A. Cros, F. Boeuf, T. Skotnicki, and G. Ghibaudo, "New experimental insight into ballisticity of transport in strained bulk mosfets," in *VLSI Technology, 2009 Symposium on*, 2009, pp. 16 –17.

[2] R. Dennard, F. Gaensslen, L. Kuhn, and H. Yu, "Design of micron mos switching devices," in *IEEE IEDM Technical Digest*, vol. 18, 1972, pp. 168–170.

[3] G. Baccarani, M. Wordeman, and R. Dennard, "Generalized scaling theory and its application to a $1/4$; micrometer MOSFET design," *Electron Devices, IEEE Transactions on*, vol. 31, no. 4, pp. 452–462, 1984.

[4] B. Davari, R. Dennard, and G. Shahidi, "CMOS technology for low voltage/low power applications," in *IEEE Proceedings*, 1994, pp. 3–10.

[5] C. Fiegna, H. Iwai, T. Wada, T. Saito, E. Sangiorgi, and B. Ricco, "A New Scaling Methodology For The $0.1-0.025$ μm MOSFET," in *VLSI Technology, Symposium on*, 1993, pp. 33–34.

[6] C. Fiegna, H. Iwai, T. Wada, M. Saito, E. Sangiorgi, and B. Ricco, "Scaling the MOS transistor below 0.1 μm: methodology, device structures, and technology requirements," *Electron Devices, IEEE Transactions on*, vol. 41, no. 6, pp. 941–951, 1994.

[7] J.-J. Maa and C.-Y. Wu, "A new constant-field scaling theory for MOSFET's," *Electron Devices, IEEE Transactions on*, vol. 42, no. 7, pp. 1262–1268, 1995.

[8] D. Frank, Y. Taur, and H.-S. Wong, "Generalized scale length for two-dimensional effects in MOSFETs," *Electron Device Letters, IEEE*, vol. 19, no. 10, pp. 385–387, 1998.

[9] D. Frank, R. Dennard, E. Nowak, P. Solomon, Y. Taur, and H.-S. P. Wong, "Device scaling limits of Si MOSFETs and their application dependencies," *Proceedings of the IEEE*, vol. 89, no. 3, pp. 259–288, 2001.

[10] International Technology Roadmap for Semiconductors (ITRS), *Process Integration, Devices, and Structures, http://www.itrs.net/*, 2010.

[11] O. Faynot, F. Andrieu, O. Weber, C. Fenouillet-Béranger, P. Perreau, J. Mazurier, T. Benoist, O. Rozeau, T. Poiroux, M. Vinet, L. Grenouillet,

J.-P. Noel, N. Posseme, S. Barnola, F. Martin, C. Lapeyre, M. Casse and, X. Garros, M.-A. Jaud, O. Thomas, G. Cibrario, L. Tosti, L. Brevard, C. Tabone, P. Gaud, S. Barraud, T. Ernst, and S. Deleonibus, "Planar Fully depleted SOI technology: A powerful architecture for the 20nm node and beyond," in *IEEE IEDM Technical Digest*, 2010, pp. 3.2.1–3.2.4.

[12] C. Wu, D. Lin, A. Keshavarzi, C. Huang, C. Chan, C. Tseng, C. Chen, C. Hsieh, K. Wong, M. Cheng, T. Li, Y. Lin, L. Yang, C. Lin, C. Hou, H. Lin, J. Yang, K. Yu, M. Chen, T. Hsieh, Y. Peng, C. Chou, C. Lee, C. Huang, C. Lu, F. Yang, H. Chen, L. Weng, P. Yen, S. Wang, S. Chang, S. Chuang, T. Gan, T. Wu, T. Lee, W. Huang, Y. Huang, Y. Tseng, C. Wu, E. Ou-Yang, K. Hsu, L. Lin, S. Wang, T. Kwok, C. Su, C. Tsai, M. Huang, H. Lin, A. Chang, S. Liao, L. Chen, J. Chen, P. Lim, X. Yu, S. Ku, Y. Lee, P. Hsieh, P. Wang, Y. Chiu, S. Lin, H. Tao, M. Cao, and Y. Mii, "High performance 22/20nm FinFET CMOS devices with advanced high-K/metal gate scheme," in *IEEE IEDM Technical Digest*, 2010, pp. 27.1.1–27.1.4.

[13] C.-C. Yeh, C.-S. Chang, H.-N. Lin, W.-H. Tseng, L.-S. Lai, T.-H. Perng, T.-L. Lee, C.-Y. Chang, L.-G. Yao, C.-C. Chen, T.-M. Kuan, J. Xu, C.-C. Ho, T.-C. Chen, S.-S. Lin, H.-J. Tao, M. Cao, C.-H. Chang, T.-C. Ko, N.-K. Chen, S.-C. Chen, C.-P. Lin, H.-C. Lin, C.-Y. Chan, H.-T. Lin, S.-T. Yang, J.-C. Sheu, C.-Y. Fu, S.-T. Hung, F. Yuan, M.-F. Shieh, C.-F. Hu, and C. Wann, "A low operating power FinFET transistor module featuring scaled gate stack and strain engineering for 32/28nm SoC technology," in *IEEE IEDM Technical Digest*, 2010, pp. 34.1.1–34.1.4.

[14] I. Ok, K. Akarvardar, S. Lin, M. Baykan, C. Young, P. Hung, M. Rodgers, S. Bennett, H. Stamper, D. Franca, J. Yum, J. Nadeau, C. Hobbs, P. Kirsch, P. Majhi, and R. Jammy, "Strained SiGe and Si FinFETs for high performance logic with SiGe/Si stack on SOI," in *IEEE IEDM Technical Digest*, 2010, pp. 34.2.1–34.2.4.

[15] J. Lin, W. Chiou, K. Yang, H. Chang, Y. Lin, E. Liao, J. Hung, Y. Lin, P. Tsai, Y. Shih, T. Wu, W. Wu, F. Tsai, Y. Huang, T. Wang, C. Yu, C. Chang, M. Chen, S. Hou, C. Tung, S. Jeng, and D. Yu, "High density 3D integration using CMOS foundry technologies for 28 nm node and beyond," in *IEEE IEDM Technical Digest*, 2010, pp. 2.1.1–2.1.4.

[16] L. Wu, K. Yew, D. Ang, W. Liu, T. Le, T. Duan, C. Hou, X. Yu, D. Lee, K. Hsu, J. Xu, H. Tao, M. Cao, and H. Yu, "A novel multi deposition multi room-temperature annealing technique via Ultraviolet-Ozone to improve high-K/metal (HfZrO/TiN) gate stack integrity for a gate-last process," in *IEEE IEDM Technical Digest*, 2010, pp. 11.6.1–11.6.4.

[17] X. Garros, L. Brunet, M. Rafik, J. Coignus, G. Reimbold, E. Vincent, A. Bravaix, and F. Boulanger, "PBTI mechanisms in La containing Hf-based oxides assessed by very Fast IV measurements," in *IEEE IEDM Technical Digest*, 2010, pp. 4.6.1–4.6.4.

[18] G. Du, D. Ang, and C. Gu, "Abnormal slow recovery characteristic of La-doped HfSiOx n-MOSFET bias-temperature instability," in *IEEE IEDM Technical Digest*, 2010, pp. 4.7.1–4.7.4.

[19] G. Thareja, J. Liang, S. Chopra, B. Adams, N. Patil, S.-L. Cheng, A. Nainani, E. Tasyurek, Y. Kim, S. Moffatt, R. Brennan, J. McVittie, T. Kamins, K. Saraswat, and Y. Nishi, "High performance germanium n-MOSFET with antimony dopant activation beyond 1×10^{20} cm^{-3}," in *IEEE IEDM Technical Digest*, 2010, pp. 10.5.1–10.5.4.

[20] C. Lee, T. Nishimura, T. Tabata, S. Wang, K. Nagashio, K. Kita, and A. Toriumi, "Ge MOSFETs performance: Impact of Ge interface passivation," in *IEEE IEDM Technical Digest*, 2010, pp. 18.1.1–18.1.4.

[21] R. Hill, C. Park, J. Barnett, J. Price, J. Huang, N. Goel, W. Loh, J. Oh, C. Smith, P. Kirsch, P. Majhi, and R. Jammy, "Self-aligned III-V MOSFETs heterointegrated on a 200 mm Si substrate using an industry standard process flow," in *IEEE IEDM Technical Digest*, 2010, pp. 6.2.1–6.2.4.

[22] Y. Urabe, N. Miyata, H. Ishii, T. Itatani, T. Maeda, T. Yasuda, H. Yamada, N. Fukuhara, M. Hata, M. Yokoyama, N. Taoka, M. Takenaka, and S. Takagi, "Correlation between channel mobility improvements and negative Vth shifts in III-V MISFETs: Dipole fluctuation as new scattering mechanism," in *IEEE IEDM Technical Digest*, 2010, pp. 6.5.1–6.5.4.

[23] M. Yokoyama, R. Iida, S. Kim, N. Taoka, Y. Urabe, T. Yasuda, H. Takagi, H. Yamada, N. Fukuhara, M. Hata, M. Sugiyama, Y. Nakano, M. Takenaka, and S. Takagi, "Extremely-thin-body InGaAs-on-insulator MOSFETs on Si fabricated by direct wafer bonding," in *IEEE IEDM Technical Digest*, 2010, pp. 3.1.1–3.1.4.

[24] N. Barin, M. Braccioli, C. Fiegna, and E. Sangiorgi, "Scaling the high-performance double-gate SOI MOSFET down to the 32 nm technology node with SiO_2-based gate stacks," in *IEEE IEDM Technical Digest*, 2005, pp. 609–612.

[25] T. Kauerauf, R. Degraeve, E. Cartier, B. Govoreanu, P. Blomme, B. Kaczer, L. Pantisano, A. Kerber, and G. Groeseneken, "Towards understanding degradation and breakdown of SiO_2/high-k stacks," in *IEEE IEDM Technical Digest*, 2002, pp. 521–524.

[26] B. Govoreanu, P. Blomme, K. Henson, J. Van Houdt, and K. De Meyer, "An investigation of the electron tunneling leakage current through ultrathin oxides/high-k gate stacks at inversion conditions," in *SISPAD*, 2003, pp. 287–290.

[27] S. Brugger, A. Schenk, and W. Fichtner, "Moments of the Inverse Scattering Operator of the Boltzmann Equation: Theory and Applications," *SIAM Journal of Applied Mathematics*, vol. 66, pp. 1209–1226, 2006.

[28] V. Narayanan, V. K. Paruchuri, E. Cartier, B. P. Linder, N. Bojarczuk, S. Guha, S. L. Brown, Y. Wang, M. Copel, and T. C. Chen, "Recent advances and current challenges in the search for high mobility band-edge high-k/metal gate stacks," *Microelectronic Engineering*, vol. 84, no. 9-10, pp. 1853–1856, 2007.

[29] B. Guillaumot, X. Garros, F. Lime, K. Oshima, B. Tavel, J. Chroboczek, P. Masson, R. Truche, A. Papon, F. Martin, J. Damlencourt, S. Maitrejean, M. Rivoire, C. Leroux, S. Cristoloveanu, G. Ghibaudo, J. Autran, T. Skotnicki, and S. Deleonibus, "75 nm damascene metal gate and high-k integration for advanced CMOS devices," *Proc. IEEE IEDM*, pp. 355–358, 2002.

[30] R. Chau, J. Brask, S. Datta, G. Dewey, M. Doczy, B. Doyle, J. Kavalieros, B. Jin, M. Metz, A. Majumdar, and M. Radosavljevic, "Application of high-k gate dielectrics and metal gate electrodes to enable silicon and non-silicon logic nanotechnology," *Microelectronic Engineering*, vol. 80, pp. 1–6, 2005.

[31] M. Casse, L. Thevenod, B. Guillaumot, L. Tosti, F. Martin, J. Mitard, O. Weber, F. Andrieu, T. Ernst, G. Reimbold, T. Billon, M. Mouis, and F. Boulanger, "Carrier transport in HfO_2/metal gate MOSFETs: physical insight into critical parameters," *IEEE Transactions on Electron Devices*, vol. 53, no. 4, pp. 759–768, 2006.

[32] F. Andrieu, O. Faynot, X. Garros, D. Lafond, C. Buj-Dufournet, L. Tosti, S. Minoret, V. Vidal, J. Barbe, F. Allain, E. Rouchouze, L. Vandroux, V. Cosnier, M. Casse, V. Delaye, C. Carabasse, M. Burdin, G. Rolland, B. Guillaumot, J. Colonna, P. Besson, L. Brevard, D. Mariolle, P. Holliger, A. Vandooren, C. Fenouillet-Beranger, F. Martin, and S. Deleonibus, "Comparative Scalability of PVD and CVD TiN on HfO_2 as a Metal Gate Stack for FDSOI cMOSFETs down to 25nm Gate Length and Width," *Proc. IEEE IEDM*, pp. 1–4, 2006.

[33] S. Samavedam, L. La, J. Smith, S. Dakshina-Murthy, E. Luckowski, J. Schaeffer, M. Zavala, R. Martin, V. Dhandapani, D. Triyoso, H. Tseng, P. Tobin, D. Gilmer, C. Hobbs, W. Taylor, J. Grant, R. Hegde, J. Mogab, C. Thomas, P. Abramowitz, M. Moosa, J. Conner, J. Jiang, V. Arunachalarn, M. Sadd, B.-Y. Nguyen, and B. White, "Dual-metal gate CMOS with HfO_2 gate dielectric," *Proc. IEEE IEDM*, pp. 433–436, 2002.

[34] O. Weber, F. Andrieu, M. Casse, T. Ernst, J. Mitard, F. Ducroquet, J.-F. Damlencourt, J.-M. Hartmann, D. Lafond, A.-M. Papon, L. Militaru, L. Thevenod, K. Romanjek, C. Leroux, F. Martin, B. Guillaumot, G. Ghibaudo, and S. Deleonibus, "Experimental determination of mobility scattering mechanisms in Si/HfO_2/TiN and SiGe:C/HfO_2/TiN surface channel n- and p-MOSFETs," *Proc. IEEE IEDM*, pp. 867–870, 2004.

[35] S. Datta, G. Dewey, M. Doczy, B. Doyle, B. Jin, J. Kavalieros, R. Kotlyar, M. Metz, N. Zelick, and R. Chau, "High mobility Si/SiGe strained channel

MOS transistors with HfO_2/TiN gate stack," *Proc. IEEE IEDM*, pp. 2811–2814, 2003.

[36] O. Weber, Y. Bogumilowicz, T. Ernst, J.-M. Hartmann, F. Ducroquet, F. Andrieu, C. Dupre, L. Clavelier, C. L. Royer, N. Cherkashin, M. Hytch, D. Rouchon, H. Dansas, A.-M. Papon, V. Carron, C. Tabone, and S. Deleonibus, "Strained Si and Ge MOSFETs with high-k/metal gate stack for high mobility dual channel CMOS," *Proc. IEEE IEDM*, pp. 137–140, 2005.

[37] Y. Kim, J. Cabral, C., E. Gusev, R. Carruthers, L. Gignac, M. Gribelyuk, E. Cartier, S. Zafar, M. Copel, V. Narayanan, J. Newbury, B. Price, J. Acevedo, P. Jamison, B. Linder, W. Natzle, J. Cai, R. Jammy, and M. Ieong, "Systematic study of work function engineering and scavenging effect using NiSi alloy FUSI metal gates with advanced gate stacks," *Proc. IEEE IEDM*, pp. 645–649, 2005.

[38] S. Inumiya, Y. Akasaka, T. Matsuki, F. Ootsuka, K. Torii, and Y. Nara, "A thermally-stable sub-0.9nm EOT TaSix/HfSiON gate stack with high electron mobility, suitable for gate-first fabrication of hp45 LOP devices," *IEEE IEDM Technical Digest*, pp. 23–26, 2005.

[39] P. Kirsch, M. Quevedo-Lopez, S. Krishnan, C. Krug, H. AlShareef, C. Park, R. Harris, N. Moumen, A. Neugroschel, G. Bersuker, B. Lee, J. Wang, G. Pant, B. Gnade, M. Kim, R. Wallace, J. Jur, D. Lichtenwalner, A. Kingon, and R. Jammy, "Band Edge n-MOSFETs with High-k/Metal Gate Stacks Scaled to EOT=0.9nm with Excellent Carrier Mobility and High Temperature Stability," *IEEE IEDM Technical Digest*, pp. 1–4, 2006.

[40] M. Quevedo-Lopez, S. Krishnan, D. Kirsch, C. Li, J. Sim, C. Huffman, J. Peterson, B. Lee, G. Pant, B. Gnade, M. Kim, R. Wallace, D. Guo, H. Bu, and T. Ma, "High performance gate first HfSiON dielectric satisfying 45nm node requirements," *IEEE IEDM Technical Digest*, pp. pp.4–8, 2005.

[41] M. Sato, Y. Nakasaki, K. Watanabe, T. Aoyama, E. Hasegawa, M. Koyama, K. Sekine, K. Eguchi, M. Saito, and Y. Tsunashima, "Impact of Very Low Hf Concentration (Hf=6%) Cap Layer on Performance and Reliability Improvement of HfSiON - CMOSFET with EOT Scalable to 1nm," *IEEE IEDM Technical Digest*, pp. 1–4, 2006.

[42] T. Hoffmann, A. Veloso, A. Lauwers, H. Yu, H. Tigelaar, M. Van Dal, T. Chiarella, C. Kerner, T. Kauerauf, A. Shickova, R. Mitsuhashi, I. Satoru, M. Niwa, A. Rothschild, B. Froment, J. Ramos, A. Nackaerts, M. Rosmeulen, S. Brus, C. Vrancken, P. Absil, M. Jurczak, S. Biesemans, and J. Kittl, "Ni-based FUSI gates: CMOS Integration for 45nm node and beyond," *IEEE IEDM Technical Digest*, pp. 1–4, 2006.

[43] K. Terashima, K. Manabe, K. Takahashi, K. Watanabe, T. Ogura, M. Saitoh, M. Oshida, N. Ikarashi, T. Tatsumi, and H. Watanabe, "Practical Vth Control

Methods for Ni-FUSI/HfSiON MOSFETs on SOI Substrates," *IEEE IEDM Technical Digest*, pp. 1–4, 2006.

[44] T. Aoyama, T. Maeda, K. Torii, K. Yamashita, Y. Kobayashi, S. Kamiyama, T. Miura, H. Kitajima, and T. Arikado, "Proposal of new HfSiON CMOS fabrication process (HAMDAMA) for low standby power device," *IEEE IEDM Technical Digest*, pp. 95–98, 2004.

[45] H. Harris, H. Alshareef, H. Wen, S. Krishnan, K. Choi, H. Luan, D. Heh, C. Park, H. Park, M. Hussain, B. Ju, P. Kirsch, S. Song, P. Majhi, B. Lee, and R. Jammy, "Simplified manufacturable band edge metal gate solution for NMOS without a capping layer," *IEEE IEDM Technical Digest*, pp. 1–4, 2006.

[46] H.-J. Cho, H. Lee, S. Park, H. Park, T. Jeon, B. Jin, S. Kang, S. Lee, Y. Kim, I. Jung, J. Lee, Y. Shin, U.-I. Chung, J. Moon, J. Choi, and Y. Jeong, "The effects of TaN thickness and strained substrate on the performance and PBTI characteristics of poly-Si/TaN/HfSiON MOSFETs," *IEEE IEDM Technical Digest*, pp. 503–506, 2004.

[47] S. Takagi, A. Toriumi, M. Iwase, and H. Tango, "On the universality of inversion layer mobility in Si MOSFET's: Part I-effects of substrate impurity concentration," *IEEE Transactions on Electron Devices*, vol. 41, no. 12, pp. 2357–2362, 1994.

[48] M. V. Fischetti, D. A. Neumayer, and E. A. Cartier, "Effective electron mobility in Si inversion layers in metal-oxide-semiconductor systems with a high-k insulator: The role of remote phonon scattering," *Journal of Applied Physics*, vol. 90, no. 9, pp. 4587–4608, 2001.

[49] T. Ando, M. Frank, K. Choi, C. Choi, J. Bruley, M. Hopstaken, M. Copel, E. Cartier, A. Kerber, A. Callegari, D. Lacey, S. Brown, Q. Yang, and V. Narayanan, "Understanding mobility mechanisms in extremely scaled HfO2 (EOT 0.42 nm) using remote interfacial layer scavenging technique and Vt-tuning dipoles with gate-first process," *Proc. IEEE IEDM*, pp. 423–426, 2009.

[50] G. Hyvert, T. Nguyen, L. Militaru, A. Poncet, and C. Plossu, "A study on mobility degradation in nMOSFETs with HfO_2 based gate oxide," *Materials Science and Engineering: B*, vol. 165, pp. 129–131, 2009.

[51] K. Choi, H. Jagannathan, C. Choi, L. Edge, T. Ando, M. Frank, P. Jamison, M. Wang, E. Cartier, S. Zafar, J. Bruley, A. Kerber, B. Linder, A. Callegari, Q. Yang, S. Brown, J. Stathis, J. Iacoponi, V. Paruchuri, and V. Narayanan, "Extremely scaled gate-first high-k/metal gate stack with EOT of 0.55 nm using novel interfacial layer scavenging techniques for 22nm technology node and beyond," *VLSI Technology, 2009 Symposium on*, pp. 138–139, 2009.

[52] S. Barraud, O. Bonno, and M. Casse, "The influence of Coulomb centers located in HfO_2/SiO_2 gate stacks on the effective electron mobility," *Journal of Applied Physics*, vol. 104, no. 7, p. 073725, 2008.

[53] K. Kita and A. Toriumi, "Intrinsic origin of electric dipoles formed at high-k/SiO_2 interface," in *Proc. IEEE IEDM*, 2008, pp. 29–32.

[54] K. Tatsumura, T. Ishiara, I. S., K. Nakajima, A. Kaneko, M. Goto, S. Kawanaka, and A. Kinoshita, "Intrinsic correlation between mobility reduction and Vt shift due to interface dipole modulation in HfSiON/SiO_2 stack by La or Al addition," *Proc. IEEE IEDM*, pp. 25–28, 2008.

[55] O. Sharia, A. A. Demkov, G. Bersuker, and B. H. Lee, "Theoretical study of the insulator/insulator interface: Band alignment at the SiO_2/HfO_2 junction," *Physical Review B*, vol. 75, no. 3, p. 035306, 2007.

[56] H.-S. Wong, D. Frank, P. Solomon, C. Wann, and J. Welser, "Nanoscale CMOS," *Proceedings of the IEEE*, vol. 87, no. 4, pp. 537–570, apr 1999.

[57] J.-P. Colinge, "Thin-film SOI technology: the solution to many submicron CMOS problems," in *IEEE IEDM Technical Digest*, 1989, pp. 817–820.

[58] Intel 22 nm technology, *http://www.intel.com/content/www/us/en/silicon-innovations/intel-22nm-technology.html*, 2011.

[59] T. Sekigawa and Y. Hayashi, "Calculated threshold-voltage characteristics of an XMOS transistor having an additional bottom gate," *Solid-State Electronics*, vol. 27, no. 8-9, pp. 827–828, 1984.

[60] Y. Taur, D.A. Buchanan, W. Chen, D.J. Frank, K.E. Ismail, S.-H. Lo, G.A. Sai-Halasz, R.G.Viswanathan, H.C. Wann, S.J.Wind and H.-S. Wong, "CMOS scaling into the nanometer regime," *IEEE Proceedings*, vol. 85, no. 4, pp. 486–503, 1997.

[61] P. C-T. Chuang and C.J.Anderson, "SOI for digital VLSI: Design consideration and advances," *IEEE Proceedings*, vol. 86, no. 4, pp. 689–720, 1998.

[62] Y. Tsividis, *Operation and modeling of the MOS transistor*. Oxford University Press, 2003.

[63] M.J.H. van Dal, N. Collaert, G. Doornbos, G. Vellianitis, G. Curatola, B.J. Pawlak, R. Duffy, C. Jonville, B. Degroote, E. Altamirano, E. Kunnen, M. Demand, S. Beckx, T. Vandeweyer, C. Delvaux, F. Leys, A. Hikavyy, R. Rooyackers, M. Kaiser, R.G.R. Weemaes, S. Biesemans, M. Jurczak, K. Anil, L. Witters and R.J.P. Lander, "Highly manufacturable FinFETs with sub-10nm fin width and high aspect ratio fabricated with immersion lithography," in *VLSI Technology, Symposium on*, 2007, pp. 110–111.

[64] S. Suthram, M.M. Hussain, H.R. Harris, C. Smith, H.-H. Tseng, R. Jammy and S.E. Thompson, "Comparison of Uniaxial Wafer Bending and Contact-Etch-Stop-Liner Stress Induced Performance Enhancement on Double-Gate FinFETs," *Electron Device Letters, IEEE*, vol. 29, no. 5, pp. 480–482, May 2008.

[65] G. Vellianitis, M.J.H van Dal, L. Witters, G. Curatola, G. Doornbos, N. Collaert, C. Jonville, C. Torregiani, L.-S. Lai, J. Petty, B.J. Pawlak, R. Duffy, M. Demand, S. Beckx, S. Mertens, A. Delabie, T. Vandeweyer, C. Delvaux, F. Leys, A. Hikavyy, R. Rooyackers, M. Kaiser, R.G. Weemaes, F. Voogt, H. Roberts, D. Donnet, S. Biesemans, M. Jurczak and R.J.R. Lander, "Gate-stacks for scalable high-performance FinFETs," in *IEEE IEDM Technical Digest*, Dec. 2007, pp. 681–684.

[66] N. Singh, A. Agarwal, L. Bera, T. Liow, R. Yang, S. Rustagi, C. Tung, R. Kumar, G. Lo, N. Balasubramanian, and D.-L. Kwong, "High-performance fully depleted silicon nanowire (diameter < 5 nm) gate-all-around CMOS devices," *IEEE Electron Device Letters*, vol. 27, no. 5, pp. 383–386, 2006.

[67] T. Sanuki, A. Oishi, Y. Morimasa, S. Aota, T. Kinoshita, R. Hasumi, Y. Takegawa, K. Isobe, H. Yoshimura, M. Iwai, K. Sunouchi, and T. Noguchi, "Scalability of strained silicon CMOSFET and high drive current enhancement in the 40 nm gate length technology," in *IEEE IEDM Technical Digest*, 2003, pp. 3.5.1–3.5.4.

[68] K. Uchida, R. Zednik, C.-H. Lu, H. Jagannathan, J. McVittie, P. McIntyre, and Y. Nishi, "Experimental study of biaxial and uniaxial strain effects on carrier mobility in bulk and ultrathin-body SOI MOSFETs," in *IEEE IEDM Technical Digest*, 2004, pp. 229–232.

[69] S. Thompson, G. Sun, Y. S. Choi, and T. Nishida, "Uniaxial-process-induced strained-Si: extending the CMOS roadmap," *Electron Devices, IEEE Transactions on*, vol. 53, no. 5, pp. 1010–1020, 2006.

[70] J. Welser, J. Hoyt, S. Takagi, and J. Gibbons, "Strain dependence of the performance enhancement in strained-Si n-MOSFETs," in *IEEE IEDM Technical Digest*, 1994, pp. 373–376.

[71] K. Rim, J. Hoyt, and J. Gibbons, "Fabrication and analysis of deep submicron strained-Si n-MOSFET's," *IEEE Transactions on Electron Devices*, vol. 47, no. 7, pp. 1406–1415, 2000.

[72] D. Kuzum, T. Krishnamohan, A. Nainani, Y. Sun, P. Pianetta, H.-P. Wong, and K. Saraswat, "Experimental demonstration of high mobility Ge NMOS," in *IEEE IEDM Technical Digest*, dec. 2009, pp. 1 –4.

[73] C. Lee, T. Nishimura, N. Saido, K. Nagashio, K. Kita, and A. Toriumi, "Record-high electron mobility in Ge n-MOSFETs exceeding Si universality," in *IEEE IEDM Technical Digest*, 2009, pp. 1–4.

[74] M. Kobayashi, T. Irisawa, B. Magyari-Kope, K. Saraswat, H.-S. Wong, and Y. Nishi, "Uniaxial Stress Engineering for High-Performance Ge NMOSFETs," *IEEE Transactions on Electron Devices*, vol. 57, no. 5, pp. 1037–1046, 2010.

[75] G. Doornbos and M. Passlack, "Benchmarking of III-V n-MOSFET Maturity and Feasibility for Future CMOS," *IEEE Electron Device Letters*, vol. 31, no. 10, pp. 1110–1112, 2010.

[76] K. Huang, *Statistical Mechanics*. Wiley & Sons, 1987.

[77] M. Lundstrom, *Fundamentals of carrier transport*. Cambridge University Press, 2000.

[78] D. Esseni, P. Palestri, and L. Selmi, *Nanoscale MOS Transistors: Semi-Classical Transport and Applications*. Cambridge University Press, 2011.

[79] J. A. Cooper and D. F. Nelson, "High-field drift velocity of electrons at the Si-SiO2 interface as determined by a time-of-flight technique," *Journal of Applied Physics*, vol. 54, no. 3, pp. 1445–1456, 1983.

[80] C. Jacoboni and L. Reggiani, "The Monte Carlo method for the solution of charge transport in semiconductors with applications to covalent materials," *Reviews of Modern Physics*, vol. 55, pp. 645–705, 1983.

[81] Y. Taur and T. H. Ning, *Fundamentals of Modern VLSI Devices*. Cambridge University Press, 1988.

[82] M. Lundstrom, "Elementary scattering theory of the Si MOSFET," *IEEE Electron Device Letters*, vol. 18, no. 7, pp. 361–363, 1997.

[83] M. Lundstrom and Z. Ren, "Essential physics of carrier transport in nanoscale MOSFETs," *IEEE Transactions on Electron Devices*, vol. 49, no. 1, pp. 133–141, 2002.

[84] K. Natori, "Ballistic metal-oxide-semiconductor field effect transistor," *Journal of Applied Physics*, vol. 76, no. 8, pp. 4879–4890, 1994.

[85] P. Palestri, D. Esseni, S. Eminente, C. Fiegna, E. Sangiorgi, and L. Selmi, "Understanding quasi-ballistic transport in nano-MOSFETs: part I-scattering in the channel and in the drain," *IEEE Transactions on Electron Devices*, vol. 52, no. 12, pp. 2727–2735, 2005.

[86] S. Eminente, D. Esseni, P. Palestri, C. Fiegna, L. Selmi, and E. Sangiorgi, "Understanding quasi-ballistic transport in nano-MOSFETs: part II-Technology scaling along the ITRS," *IEEE Transactions on Electron Devices*, vol. 52, no. 12, pp. 2736–2743, 2005.

[87] M. Fischetti, S. Jin, T.-w. Tang, P. Asbeck, Y. Taur, S. Laux, and N. Sano, "Scaling MOSFETs to 10 nm: Coulomb Effects, Source Starvation, and Virtual Source," in *International Workshop on Computational Electronics*, 2009, pp. 1–4.

[88] P. Palestri, R. Clerc, D. Esseni, L. Lucci, and L. Selmi, "Multi-Subband-Monte-Carlo investigation of the mean free path and of the kT layer in degenerated quasi ballistic nanoMOSFETs," in *IEEE IEDM Technical Digest*, 2006, pp. 1–4.

[89] A. Rahman and M. Lundstrom, "A compact scattering model for the nanoscale double-gate MOSFET," *IEEE Transactions on Electron Devices*, vol. 49, no. 3, pp. 481–489, 2002.

[90] R. Clerc, P. Palestri, and L. Selmi, "On the physical understanding of the kT-layer concept in quasi-ballistic regime of transport in nanoscale devices," *IEEE Transactions on Electron Devices*, vol. 53, no. 7, pp. 1634–1640, 2006.

[91] R. Clerc, P. Palestri, L. Selmi, and G. Ghibaudo, "Impact of carrier heating on backscattering in inversion layers," *Journal of Applied Physics*, vol. 110, no. 10, pp. 104 502–104 502–10, 2011.

[92] J. P. McKelvey, R. L. Longini, and T. P. Brody, "Alternative Approach to the Solution of Added Carrier Transport Problems in Semiconductors," *Physical Review*, vol. 123, pp. 51–57, 1961.

[93] L. Lucci, P. Palestri, D. Esseni, L. Bergagnini, and L. Selmi, "Multisubband Monte Carlo Study of Transport, Quantization, and Electron-Gas Degeneration in Ultrathin SOI n-MOSFETs," *IEEE Transactions on Electron Devices*, vol. 54, no. 5, pp. 1156–1164, 2007.

[94] C. Jungemann, S. Yamaguchi, and H. Goto, "Convergence estimation for stationary ensemble Monte Carlo simulations," *SISPAD*, p. 209, 1997.

[95] A. Pacelli and U. Ravaioli, "Analysis of variance-reduction schemes for ensemble Monte Carlo simulation of semiconductor devices," *Solid-State Electronics*, vol. 41, no. 4, pp. 599–605, 1997.

[96] T. Gonzalez and D. Pardo, "Physical models of ohmic contact for Monte Carlo device simulation," *Solid-State Electronics*, vol. 39, no. 4, pp. 555–562, 1996.

[97] I. Riolino, M. Braccioli, L. Lucci, P. Palestri, D. Esseni, C. Fiegna, and L. Selmi, "Monte-Carlo simulation of decananometric nMOSFETs: Multi-subband vs. 3D-electron gas with quantum corrections," *Solid-State Electronics*, vol. 51, no. 11-12, pp. 1558–1564, 2007.

[98] M. V. Fischetti and S. E. Laux, "Monte carlo study of electron transport in silicon inversion layers," *Physical Rewiev B*, vol. 48, no. 4, pp. 2244–2274, Jul 1993.

[99] C. Jungemann, A. Emunds and W.L. Engl, "Simulation of linear and nonlinear electron transport in homogeneous silicon inversion layers," *Solid-State Electronics*, vol. 36, no. 11, pp. 1529–1540, 1993.

[100] M. De Michielis, D. Esseni, P. Palestri, and L. Selmi, "Semiclassical Modeling of Quasi-Ballistic Hole Transport in Nanoscale pMOSFETs Based on a Multi-Subband Monte Carlo Approach," *IEEE Transactions on Electron Devices*, vol. 56, no. 9, pp. 2081–2091, 2009.

[101] J. Saint-Martin, D. Querlioz, A. Bournel, and P. Dollfus, "Efficient multi sub-band monte carlo simulation of nano-scaled double gate mosfets," in

International Conference on Simulation of Semiconductor Processes and Devices, 2006, pp. 216–219.

[102] D.K. Ferry and S.M. Goodnick, *Transport in Nanostructures (Cambridge Studies in Semiconductor Physics and Microelectronic Engineering)*. Cambridge University Press, 1997.

[103] E. O. Kane, "Energy Band Structure in P-Type Germanium and Silicon," *Journal of Physics and Chemistry of Solids*, vol. 1, pp. 82–99, 1956.

[104] D. Esseni and P. Palestri, "Linear combination of bulk bands method for investigating the low-dimensional electron gas in nanostructured devices," *Physical Review B*, vol. 72, pp. 165 342–165 342.16, 2005.

[105] M. De Michielis, D. Esseni, Y. Tsang, P. Palestri, L. Selmi, A. O'Neill, and S. Chattopadhyay, "A Semianalytical Description of the Hole Band Structure in Inversion Layers for the Physically Based Modeling of pMOS Transistors," *IEEE Transactions on Electron Devices*, vol. 54, no. 9, pp. 2164–2173, 2007.

[106] D. Esseni and A. Abramo, "Modeling of electron mobility degradation by remote Coulomb scattering in ultrathin oxide MOSFETs," *IEEE Transactions on Electron Devices*, vol. 50, no. 7, pp. 1665–1674, 2003.

[107] D. Esseni, "On the modeling of surface roughness limited mobility in SOI MOSFETs and its correlation to the transistor effective field," *Electron Devices, IEEE Transactions on*, vol. 51, no. 3, pp. 394–401, March 2004.

[108] A. Tsuneya, "Screening Effect and Quantum Transport in a Silicon Inversion Layer in Strong Magnetic Field," *Journal of the Physical Society of Japan*, vol. 43, no. 5, pp. 1616–1626, 1977.

[109] J. Seonghoon, M.V. Fischetti, T. Ting-Wei, "Modeling of Surface-Roughness Scattering in Ultrathin-Body SOI MOSFETs," *Electron Devices, IEEE Transactions on*, vol. 54, no. 9, pp. 2191–2203, Sept. 2007.

[110] M. V. Fischetti, Z. Ren, P. M. Solomon, M. Yang, and K. Rim, "Six-band $k{\cdot}p$ calculation of the hole mobility in silicon inversion layers: Dependence on surface orientation, strain, and silicon thickness," *Journal of Applied Physics*, vol. 94, no. 2, pp. 1079–1095, 2003.

[111] M. De Michielis, F. Conzatti, D. Esseni, and L. Selmi, "On the Surface-Roughness Scattering in Biaxially Strained n- and p-MOS Transistors," *IEEE Transactions on Electron Devices*, vol. 58, no. 9, pp. 3219–3223, 2011.

[112] J.-H. Choi, Y.-J. Park and H.-S. Min, "Electron Mobility Behavior in Extremely Thin SOI MOSFET's," *Electron Device Letters, IEEE*, vol. 16, no. 11, pp. 527–529, 1995.

[113] A. Toriumi, J. Koga, H. Satake and A. Ohata, "Performance and Reliability Concerns of Ultra-Thin SOI and Ultra-Thin Gate Oxide MOSFETs," in *IEEE IEDM Technical Digest*, 1995, pp. 847–850.

[114] D. Esseni, M. Mastrapasqua, G.K. Celler, C. Fiegna, L. Selmi and E. Sangiorgi, "Low Field Electron and Hole Mobility of SOI Transistors Fabricated on Ultra-Thin Silicon Films for Deep Sub-Micron Technology Application," *Electron Devices, IEEE Transactions on*, vol. 48, no. 12, pp. 2842–2850, 2001.

[115] ——, "An experimental study of mobility enhancement in ultrathin SOI transistors operated in double-gate mode," *Electron Devices, IEEE Transactions on*, vol. 50, no. 3, pp. 802–808, 2003.

[116] J. Koga, S. Takagi and A. Toriumi, "Influencies of Buried-Oxide Interface on Inversion-Layer Mobility in Ultra-Thin SOI MOSFETs," *Electron Devices, IEEE Transactions on*, vol. 49, no. 6, pp. 1042–1048, 2002.

[117] K. Uchida, J. Koga, R. Ohba, T. Numata and S. Takagi, "Experimental Evidences of Quantum-Mechanical Effects on Low-field Mobility, Gate-channel Capacitance and Threshold Voltage of Utrathin Body SOI MOSFETs," in *IEEE IEDM Technical Digest*, 2001, pp. 633–636.

[118] K. Uchida, H. Watanabe, A. Kinoshita, J. Koga, T. Numata and S. Takagi, "Experimental Study on Carrier Transport Mechanisms in Ultrathin-body SOI n- and p-MOSFETs with SOI Thickness less than 5nm," in *IEEE IEDM Technical Digest*, 2002, pp. 47–50.

[119] F. Gámiz, J.A. Lopez-Villanueva, J.B. Roldan, J.E. Carceller and P. Cartujo, "Monte Carlo Simulation of Electron Transoprt Properties in Extremely Thin SOI MOSFET's," *Electron Devices, IEEE Transactions on*, vol. 45, no. 5, pp. 1122–1126, 1998.

[120] M. Shoji and S. Horiguchi, "Electronic structures and phonon-limited electron mobility of double-gate silicon-on-insulator Si inversion layers," *Journal of Applied Physics*, vol. 85, no. 5, pp. 2722–2731, 1999.

[121] D. Esseni, A. Abramo, L. Selmi and E. Sangiorgi, "Physically Based modeling of low field electron mobility in Ultra-Thin SIngle- and Double-Gate SOI n-MOSFETs ," *Electron Devices, IEEE Transactions on*, vol. 50, no. 12, pp. 2445–2455, 2003.

[122] T. Ando, A.B. Fowler and F. Stern, "Electronic properties of two-dimensional systems," *Reviews of Modern Physics*, vol. 54, no. 2, pp. 437–672, Apr 1982.

[123] A. Pirovano, A.L. Lacaita, G. Zandler and R. Oberhuber, "Explaining the dependences of the hole and electron mobilities in Si inversion layers," *Electron Devices, IEEE Transactions on*, vol. 47, no. 4, pp. 718–724, Apr 2000.

[124] G. Mazzoni, A.L. Lacaita, L.M. Perron and A. Pirovano, "On surface roughness-limited mobility in highly doped n-MOSFET's ," *Electron Devices, IEEE Transactions on*, vol. 46, no. 7, pp. 1423–1428, Jul 1999.

[125] S. Yamakawa, H. Ueno, K. Taniguchi, C. Hamaguchi, K. Miyatsuji, K. Masaki and U. Ravaioli, "Study of interface roughness dependence of elec-

tron mobility in Si inversion layers using the Monte Carlo method," *Journal of Applied Physics*, vol. 79, no. 2, pp. 911–916, 1996.

[126] F. Monsef, P. Dollfus, S. Galdin-Retailleau, H.-J. Herzog and T. Hackbarth, "Electron transport in Si/SiGe modulation-doped heterostructures using Monte Carlo simulation," *Journal of Applied Physics*, vol. 95, no. 7, pp. 3587–3593, 2004.

[127] F. Gámiz, J. B. Roldán, P. Cartujo-Cassinello, J. A. López-Villanueva and P. Cartujo, "Role of surface-roughness scattering in double gate silicon-on-insulator inversion layers," *Journal of Applied Physics*, vol. 89, no. 3, pp. 1764–1770, 2001.

[128] V. Sverdlov, S.E. Ungersboeck and H. Kosina, "Theoretical Electron Mobility Analysis in Thin-Body FETs: Dependence on Substrate Orientation and Biaxial Strain," *Nanotechnology, IEEE Transactions on*, vol. 6, no. 3, pp. 334–340, May 2007.

[129] D. Querlioz, J. Saint-Martin, V.-N. Do, A. Bournel and P. Dollfus , "A Study of Quantum Transport in End-of-Roadmap DG-MOSFETs Using a Fully Self-Consistent Wigner Monte Carlo Approach," *Nanotechnology, IEEE Transactions on*, vol. 5, no. 6, pp. 737–744, Nov. 2006.

[130] K. Uchida, M. Saitoh, and S. Kobayashi, "Carrier transport and stress engineering in advanced nanoscale transistors from (100) and (110) transistors to carbon nanotube FETs and beyond," in *IEEE IEDM Technical Digest*, 2008, pp. 569–572.

[131] F. Gámiz and M.V. Fischetti, "Monte Carlo simulation of double-gate silicon-on-insulator inversion layers: The role of volume inversion," *Journal of Applied Physics*, vol. 89, no. 10, pp. 5478–5487, May 2001.

[132] K. Shimizu, T. Saraya, and T. Hiramoto, "Suppression of electron mobility degradation in (100)-oriented double-gate ultrathin body nmosfets," *Electron Device Letters, IEEE*, vol. 31, no. 4, pp. 284–286, april 2010.

[133] B. Laikhtman and P. M. Solomon, "Remote phonon scattering in field-effect transistors with a high-k insulating layer," *Journal of Applied Physics*, vol. 103, no. 1, p. 014501, 2008.

[134] G. Lujan, S. Kubicek, S. De Gendt, M. Heyns, W. Magnus, and K. De Meyer, "Mobility degradation in high-k transistors: the role of the charge scattering," in *European Solid-State Device Research Conference*, 2003, pp. 399–402.

[135] G. S. Lujan, W. Magnus, L.-. Ragnarsson, S. Kubicek, S. D. Gendt, M. M. Heyns, and K. D. Meyer, "Modelling mobility degradation due to remote Coulomb scattering from dielectric charges and its impact on MOS device performance," *Microelectronics Reliability*, pp. 794–797, 2005.

[136] S. Zafar, V. Narayanan, A. Callegari, F. McFeely, P. Jamison, E. Gusev, J. Cabral, C., and R. Jammy, "HfO_2/metal stacks: determination of energy

level diagram, work functions their dependence on metal deposition," *VLSI Technology, 2005 Symposium on*, pp. 44–45, june 2005.

[137] A. Chin, M. Chang, S. Lin, W. Chen, P. Lee, F. Yeh, C. Liao, M.-F. Li, N. Su, and S. Wang, "Flat band voltage control on low Vt metal-gate/high- CMOSFETs with small EOT," *Microelectronic Engineering*, vol. 86, pp. 1728–1732, 2009.

[138] C.-L. Lin, M.-Y. Chou, T.-K. Kang, and S.-C. Wu, "Electrical characteristics and TDDB breakdown mechanism of N_2-RTA-treated Hf-based high- gate dielectrics," *Microelectronic Engineering*, vol. 88, pp. 950–958, 2011.

[139] B. P. Linder, V. Narayanan, and E. A. Cartier, "Interfacial layer optimization of high-k/metal gate stacks for low temperature processing," *Microelectronic Engineering*, vol. 86, pp. 1632–1634, 2009.

[140] S. Sahhaf, R. Degraeve, M. Cho, K. D. Brabanter, P. Roussel, M. Zahid, and G. Groeseneken, "Detailed analysis of charge pumping and I_dV_g hysteresis for profiling traps in SiO_2/HfSiO(N)," *Microelectronic Engineering*, vol. 87, pp. 2614–2619, 2010.

[141] C. Choi, C. Kang, C. Kang, R. Choi, H. Cho, Y. Kim, S. Rhee, M. Akbar, and J. Lee, "The effects of nitrogen and silicon profile on high-k MOSFET performance and Bias Temperature Instability," *VLSI Technology, 2004 Symposium on*, pp. 214–215, 2004.

[142] X. Garros, M. Casse, G. Reimbold, M. Rafik, F. Martin, F. Andrieu, V. Cosnier, and F. Boulanger, "Performance and reliability of advanced High-K/Metal gate stacks," *Microelectronic Engineering*, vol. 86, no. 7-9, pp. 1609–1614, 2009.

[143] C. Gaumer, E. Martinez, S. Lhostis, M.-J. Guittet, M. Gros-Jean, J.-P. Barnes, C. Licitra, N. Rochat, N. Barrett, F. Bertin, and A. Chabli, "Impact of the TiN electrode deposition on the HfO_2 band gap for advanced MOSFET gate stacks," *Microelectronic Engineering*, vol. 88, pp. 72–75, 2011.

[144] T. Ishihara, D. Matsushita, K. Tatsumura, Y. Nakabayashi, J. Koga, and K. Kato, "Novel Carrier-Mobility Modeling with Interface States for MOSFETs with Highly Scaled Gate Oxide Based on First-Principles Calculations," *Proc. IEEE IEDM*, pp. 101–104, 2007.

[145] P. Toniutti, P. Palestri, D. Esseni, and L. Selmi, "Revised analysis of the mobility and I_{ON} degradation in high-k gate stacks: Surface optical phonons vs. remote Coulomb scattering," in *European Solid-State Device Research Conference*, 2008, pp. 246–249.

[146] P. Palestri, C. Alexander, A. Asenov, V. Aubry-Fortuna, G. Baccarani, A. Bournel, M. Braccioli, B. Cheng, P. Dollfus, A. Esposito, D. Esseni, C. Fenouillet-Beranger, C. Fiegna, G. Fiori, A. Ghetti, G. Iannaccone,

A. Martinez, B. Majkusiak, S. Monfray, V. Peikert, S. Reggiani, C. Riddet, J. Saint-Martin, E. Sangiorgi, A. Schenk, L. Selmi, L. Silvestri, P. Toniutti, and J. Walczak, "A comparison of advanced transport models for the computation of the drain current in nanoscale nMOSFETs," *Solid-State Electronics*, vol. 53, no. 12, pp. 1293–1302, 2009.

[147] P. Toniutti, M. De Michielis, P. Palestri, F. Driussi, D. Esseni, and L. Selmi, "Understanding the mobility reduction in MOSFETs featuring High-k dielectrics," in *ULIS*, 2010, pp. 65–68.

[148] K. Hess and P. Vogl, "Remote polar phonon scattering in silicon inversion layers," *Solid State Communications*, vol. 30, no. 12, pp. 797–799, 1979.

[149] B. T. Moore and D. K. Ferry, "Remote polar phonon scattering in Si inversion layers," *Journal of Applied Physics*, vol. 51, no. 5, pp. 2603–2605, 1980.

[150] M. E. Kim, A. Das, and S. D. Senturia, "Electron scattering interaction with coupled plasmon-polar-phonon modes in degenerate semiconductors," *Physical Rewiev B*, vol. 18, no. 12, pp. 6890–6899, 1978.

[151] Z.Ren, M. Fischetti, E. Gusev, E. Cartier, and M.Chudzik, "Inversion channel mobility in high-k high performance MOSFETs," *IEEE IEDM Technical Digest*, pp. 33–37, 2003.

[152] R. Kotlyar, M. Giles, P. Matagne, B. Obradovic, L. Shifren, M. Stettler, and E. Wang, "Inversion mobility and gate leakage in high-k/metal gate MOSFETs," *IEEE IEDM Technical Digest*, pp. 391–394, 2004.

[153] R. Shah and M. De Souza, "Impact of a Nonideal Metal Gate on Surface Optical Phonon-Limited Mobility in High- k Gated MOSFETs," *IEEE Transactions on Electron Devices*, vol. 54, no. 11, pp. 2991–2997, 2007.

[154] P. Toniutti, *Impact of high-k dielectrics on the transport properties of nanoscale nMOSFET*. Master Degree Thesis, University of Udine, 2008.

[155] F. Rana, J. H. Strait, H. Wang, and C. Manolatou, "Ultrafast carrier recombination and generation rates for plasmon emission and absorption in graphene," *Phys. Rev. B*, vol. 84, p. 045437, 2011.

[156] M. Fischetti, T. O'Regan, S. Narayanan, C. Sachs, S. Jin, J. Kim, and Y. Zhang, "Theoretical Study of Some Physical Aspects of Electronic Transport in nMOSFETs at the 10-nm Gate-Length," *IEEE Transactions on Electron Devices*, vol. 54, no. 9, pp. 2116–2136, 2007.

[157] G. Ferrari, J. Watling, S. Roy, J. Barker, P. Zeitzoff, G. Bersuker, and A. Asenov, "Monte carlo study of mobility in Si devices with HfO_2-based oxides," *Materials Science in Semiconductor Processing*, vol. 9, no. 6, pp. 995–999, 2006.

[158] F. Jimenez-Molinos, F. Gamiz, A. Godoy, and J. B. Roldan, "Combined influence of Coulomb and Remote phonon scattering mechanisms on the

electron mobility in SOI-MOSFETs with high-k dielectrics," in *ULIS*, 2006, pp. 111–114.

[159] T. P. O'Regan, M. V. Fischetti, B. Soree, S. Jin, W. Magnus, and M. Meuris, "Calculation of the electron mobility in III-V inversion layers with high-kappa dielectrics," *Journal of Applied Physics*, vol. 108, no. 10, p. 103705, 2010.

[160] M. Bresciani, A. Paussa, P. Palestri, D. Esseni, and L. Selmi, "Low-field mobility and high-field drift velocity in graphene nanoribbons and graphene bilayers," in *IEEE IEDM Technical Digest*, 2010, pp. 32.1.1–32.1.4.

[161] F. Gamiz, J. Lopez-Villanueva, J. Banqueri, J. Carceller, and P. Cartujo, "Universality of electron mobility curves in MOSFETs: a Monte Carlo study," *IEEE Transactions on Electron Devices*, vol. 42, no. 2, pp. 258–265, 1995.

[162] P. Toniutti, D. Esseni, and P. Palestri, "Failure of the Scalar Dielectric Function Approach for the Screening Modeling in Double-Gate SOI MOSFETs and in FinFETs," *Electron Devices, IEEE Transactions on*, vol. 57, no. 11, pp. 3074–3083, 2010.

[163] M. A. Negara, K. Cherkaoui, P. K. Hurley, C. D. Young, P. Majhi, W. Tsai, D. Bauza, and G. Ghibaudo, "Analysis of electron mobility in HfO_2/TiN gate metal-oxide-semiconductor field effect transistors: The influence of HfO_2 thickness, temperature, and oxide charge," *Journal of Applied Physics*, vol. 105, no. 2, p. 024510, 2009.

[164] B. Mereu, C. Rossel, E. P. Gusev, and M. Yang, "The role of Si orientation and temperature on the carrier mobility in metal oxide semiconductor field-effect transistors with ultrathin HfO_2 gate dielectrics," *Journal of Applied Physics*, vol. 100, no. 1, p. 014504, 2006.

[165] C. Fenouillet-Beranger, S. Denorme, B. Icard, F. Boeuf, J. Coignus, O. Faynot, L. Brevard, C. Buj, C. Soonekindt, J. Todeschini, J. Le-Denmat, N. Loubet, C. Gallon, P. Perreau, S. Manakli, B. Mmghetti, L. Pain, V. Arnal, A. Vandooren, D. Aime, L. Tosti, C. Savardi, F. Martin, T. Salvetat, S. Lhostis, C. Laviron, N. Auriac, T. Kormann, G. Chabanne, S. Gaillard, O. Belmont, E. Laffosse, D. Barge, A. Zauner, A. Tarnowka, K. Romanjec, H. Brut, A. Lagha, S. Bonnetier, F. Joly, N. Mayet, A. Cathignol, D. Galpin, D. Pop, R. Delsol, R. Pantel, F. Pionnier, G. Thomas, D. Bensahel, S. Deleombus, T. Skotnicki, and H. Mmgam, "Fully-depleted SOI technology using high-k and single-metal gate for 32 nm node LSTP applications featuring 0.179 μm^2 6T-SRAM bitcell," *Proc. IEEE IEDM*, pp. 267–270, 2007.

[166] T. H. Ning and C. T. Sah, "Theory of Scattering of Electrons in a Nondegenerate-Semiconductor-Surface Inversion Layer by Surface-Oxide Charges," vol. 6, pp. 4605–4613, 1972.

[167] J. Liu and D. L. Kwong, "Investigation of work function adjustments by electric dipole formation at the gate/oxide interface in preimplanted NiSi fully silicided metal gates," *Applied Physics Letters*, vol. 88, pp. 192 111–192 111–3, 2006.

[168] E. Cartier, F. McFeely, V. Narayanan, P. Jamison, B. Linder, M. Copel, V. Paruchuri, V. Basker, R. Haight, D. Lim, R. Carruthers, T. Shaw, M. Steen, J. Sleight, J. Rubino, H. Deligianni, S. Guha, R. Jammy, and G. Shahidi, "Role of oxygen vacancies in V_{FB}/V_t stability of pFET metals on HfO_2," *VLSI Technology, Symposium on*, pp. 230–231, 2005.

[169] M. Kadoshima, A. Ogawa, H. Ota, M. Ikeda, M. Takahashi, H. Satake, T. Nabatame, and A. Toriumi, "Two Different Mechanisms for Determining Effective Work Function ($f_{m,eff}$) on High-k - Physical Understanding and Wider Tunability of $f_{m,eff}$," *VLSI Technology, Symposium on*, pp. 180–181, 2006.

[170] K. Iwamoto, H. Ito, Y. Kamimuta, Y. Watanabe, W. Mizubayashi, S. Migita, Y. Morita, M. Takahashi, H. Ota, T. Nabatame, and A. Toriumi, "Re-examination of Flat-Band Voltage Shift for High-k MOS Devices," *VLSI Technology, Symposium on*, pp. 70–71, 2007.

[171] C. Hinkle, R. Galatage, R. Chapman, E. Vogel, H. Alshareef, C. Freeman, E. Wimmer, H. Niimi, A. Li-Fatou, J. Shaw, and J. Chambers, "Dipole controlled metal gate with hybrid low resistivity cladding for gate-last CMOS with low Vt," *VLSI Technology, Symposium on*, pp. 183–184, 2010.

[172] Changhwan and Choi, "Thickness and material dependence of capping layers on flatband voltage (V_{FB}) and equivalent oxide thickness (EOT) with high-k gate dielectric/metal gate stack for gate-first process applications," *Microelectronic Engineering*, vol. In press, 2011.

[173] C. Fenouillet-Beranger, S. Denorme, P. Perreau, C. Buj, O. Faynot, F. Andrieu, L. Tosti, S. Barnola, T. Salvetat, X. Garros, M. Casse, F. Allain, N. Loubet, L. Pham-NGuyen, E. Deloffre, M. Gros-Jean, R. Beneyton, C. Laviron, M. Marin, C. Leyris, S. Haendler, F. Leverd, P. Gouraud, P. Scheiblin, L. Clement, R. Pantel, S. Deleonibus, and T. Skotnicki, "Fd-soi devices with thin box and ground plane integration for 32nm node and below," pp. 206–209, 2008.

[174] T. Low, Y. Hou, M. Li, C. Zhu, A. Chin, G. Samudra, L. Chan, and D.-L. Kwong, "Investigation of performance limits of germanium double-gated MOSFETs," in *IEEE IEDM Technical Digest*, 2003, pp. 29.4.1–29.4.4.

[175] J. Wang, A. Rahman, G. Klimeck, and M. Lundstrom, "Bandstructure and orientation effects in ballistic Si and Ge nanowire FETs," in *IEEE IEDM Technical Digest*, 2005, pp. 533–537.

[176] T. Krishnamohan, D. Kim, T. Dinh, A.-t. Pham, B. Meinerzhagen, C. Jungemann, and K. Saraswat, "Comparison of (001), (110) and (111) uniaxial-

and biaxial- strained-Ge and strained-Si PMOS DGFETs for all channel orientations: Mobility enhancement, drive current, delay and off-state leakage," in *IEEE IEDM Technical Digest*, 2008, pp. 1–4.

[177] P. Zimmerman, G. Nicholas, B. De Jaeger, B. Kaczer, A. Stesmans, L.-A. Ragnarsson, D. P. Brunco, F. E. Leys, M. Caymax, G. Winderickx, K. Opsomer, M. Meuris, and M. M. Heyns, "High performance Ge pMOS devices using a Si-compatible process flow," in *IEEE IEDM Technical Digest*, 2006, pp. 1–4.

[178] F. Stern and W. E. Howard, "Properties of Semiconductor Surface Inversion Layers in the Electric Quantum Limit," *Physical Review*, vol. 163, pp. 816–835, 1967.

[179] C. Jacoboni, F. Nava, C. Canali, and G. Ottaviani, "Electron drift velocity and diffusivity in germanium," *Physical Review B*, vol. 24, pp. 1014–1026, 1981.

[180] Q. Rafhay, P. Palestri, D. Esseni, R. Clerc, and L. Selmi, "Mobility and Backscattering in Germanium n-type Inversion Layers," *International Conference on Solid State Devices and Materials*, pp. 46–47, 2007.

[181] D. Roessler and W. A. Jr., "Infrared reflectance of single crystal tetragonal GeO2," *Journal of Physics and Chemistry of Solids*, vol. 33, no. 2, pp. 293–296, 1972.

[182] N. Serra, F. Conzatti, D. Esseni, M. De Michielis, P. Palestri, L. Selmi, S. Thomas, T. Whall, E. Parker, D. Leadley, L. Witters, A. Hikavyy, M. Hytch, F. Houdellier, E. Snoeck, T. Wang, W. Lee, G. Vellianitis, M. van Dal, B. Duriez, G. Doornbos, and R. Lander, "Experimental and physics-based modeling assessment of strain induced mobility enhancement in FinFETs," in *IEEE International Electron Devices Meeting*, 2009, pp. 71–73.

[183] D. Esseni, F. Conzatti, M. De Michielis, N. Serra, P. Palestri, and L. Selmi, "Semi-classical transport modelling of CMOS transistors witharbitrary crystal orientations and strain engineering," *Journal of Computational Electronics*, vol. 8, pp. 209–224, 2009.

[184] S. E. Thompson, S. Suthram, Y. Sun, G. Sun, S. Parthasarathy, M. Chu, and T. Nishida, "Future of Strained Si/Semiconductors in Nanoscale MOSFETs," in *IEEE IEDM Technical Digest*, 2006, pp. 681–684.

[185] D. Ponton, L. Lucci, P. Palestri, D. Esseni, and L. Selmi, "Assessment of the Impact of Biaxial Strain on the Drain Current of Decanometric n-MOSFET," in *European Solid-State Device Research Conference*, 2006, pp. 166–169.

[186] F. Conzatti, M. De Michielis, D. Esseni, and P. Palestri, "Drain current improvements in uniaxially strained p-MOSFETs: A Multi-Subband Monte Carlo study," in *European Solid-State Device Research Conference*, 2008, pp. 250–253.

[187] S. Takagi, T. Mizuno, T. Tezuka, N. Sugiyama, S. Nakaharai, T. Numata, J. Koga, and K. Uchida, "Sub-band structure engineering for advanced cmos channels," *Solid-State Electronics*, vol. 49, no. 5, pp. 684 – 694, 2005.

[188] M. Ferrier, R. Clerc, L. Lucci, Q. Rafhay, G. Pananakakis, G. Ghibaudo, F. Boeuf, and T. Skotnicki, "Conventional Technological Boosters for Injection Velocity in Ultrathin-Body MOSFETs," *Nanotechnology, IEEE Transactions on*, vol. 6, no. 6, pp. 613 –621, 2007.

[189] C. Jeong, D. Antoniadis, and M. Lundstrom, "On Backscattering and Mobility in Nanoscale Silicon MOSFETs," *IEEE Transactions on Electron Devices*, vol. 56, no. 11, pp. 2762–2769, 2009.

[190] M.-J. Chen, H.-T. Huang, K.-C. Huang, P.-N. Chen, C.-S. Chang, and C. Diaz, "Temperature dependent channel backscattering coefficients in nanoscale mosfets," in *Electron Devices Meeting, 2002. IEDM '02. Digest. International*, 2002, pp. 39 – 42.

[191] A. Lochtefeld and D. Antoniadis, "On experimental determination of carrier velocity in deeply scaled NMOS: how close to the thermal limit?" *IEEE Electron Device Letters*, vol. 22, no. 2, pp. 95–97, 2001.

[192] V. Barral, T. Poiroux, J. Saint-Martin, D. Munteanu, J.-L. Autran, and S. Deleonibus, "Experimental investigation on the quasi-ballistic transport: Part i - determination of a new backscattering coefficient extraction methodology," *Electron Devices, IEEE Transactions on*, vol. 56, no. 3, pp. 408 –419, 2009.

[193] G. Giusi, G. Iannaccone, D. Maji, and F. Crupi, "Barrier lowering and backscattering extraction in short-channel mosfets," *Electron Devices, IEEE Transactions on*, vol. 57, no. 9, pp. 2132 –2137, 2010.

[194] F. Assad, Z. Ren, S. Datta, M. Lundstrom, and P. Bendix, "Performance limits of silicon mosfet's," in *Electron Devices Meeting, 1999. IEDM Technical Digest. International*, 1999, pp. 547–550.

[195] A. Khakifirooz, O. Nayfeh, and D. Antoniadis, "A Simple Semiempirical Short-Channel MOSFET Current-Voltage Model Continuous Across All Regions of Operation and Employing Only Physical Parameters," *IEEE Transactions on Electron Devices*, vol. 56, no. 8, pp. 1674–1680, 2009.

[196] D. Fleury, A. Cros, G. Bidal, J. Rosa, and G. Ghibaudo, "A new technique to extract the source/drain series resistance of mosfets," *Electron Device Letters, IEEE*, vol. 30, no. 9, pp. 975 –977, 2009.

[197] D. Fleury, A. Cros, H. Brut, and G. Ghibaudo, "New y-function-based methodology for accurate extraction of electrical parameters on nano-scaled mosfets," in *Microelectronic Test Structures, 2008. ICMTS 2008. IEEE International Conference on*, 2008, pp. 160 –165.

[198] D. Caughey and R. Thomas, "Carrier mobilities in silicon empirically related to doping and field," *Proceedings of the IEEE*, vol. 55, no. 12, pp. 2192 – 2193, 1967.

[199] M. G. Ancona and H. F. Tiersten, "Macroscopic physics of the silicon inversion layer," *Physical Review B*, vol. 35, pp. 7959–7965, 1987.

[200] M. G. Ancona and G. J. Iafrate, "Quantum correction to the equation of state of an electron gas in a semiconductor," *Physical Review B*, vol. 39, pp. 9536–9540, 1989.

[201] S. Reggiani, M. Valdinoci, L. Colalongo, and G. Baccarani, "A Unified Analytical Model for Bulk and Surface Mobility in Si n- and p-Channel MOSFETs," in *European Solid-State Device Research Conference*, 1999, pp. 240–243.

[202] F. Arnaud, A. Thean, M. Eller, M. Lipinski, Y. Teh, M. Ostermayr, K. Kang, N. Kim, K. Ohuchi, J.-P. Han, D. Nair, J. Lian, S. Uchimura, S. Kohler, S. Miyaki, P. Ferreira, J.-H. Park, M. Hamaguchi, K. Miyashita, R. Augur, Q. Zhang, K. Strahrenberg, S. ElGhouli, J. Bonnouvrier, F. Matsuoka, R. Lindsay, J. Sudijono, F. Johnson, J. Ku, M. Sekine, A. Steegen, and R. Sampson, "Competitive and cost effective high-k based 28nm cmos technology for low power applications," in *Electron Devices Meeting (IEDM), 2009 IEEE International*, 2009, pp. 1 –4.

[203] G. Ghibaudo, "New method for the extraction of MOSFET parameters," *Electronics Letters*, vol. 24, no. 9, pp. 543–545, 1988.

[204] V. V. Afanas'ev, A. Stesmans, F. Chen, X. Shi, and S. A. Campbell, "Internal photoemission of electrons and holes from (100)Si into HfO_2," *Applied Physics Letters*, vol. 81, no. 6, pp. 1053–1055, 2002.

[205] S. Spiga, G. Congedo, U. Russo, A. Lamperti, O. Salicio, F. Driussi, and E. Vianello, "Experimental and simulation study of the program efficiency of HfO_2 based charge trapping memories," *European Solid-State Device Research Conference*, pp. 408–411, 2010.

[206] B. K. Ridley, *Quantum Processes in Semicondutors*. Oxford Science Publication, 1993.

[207] O. Madelung, *Introduction to Solid-State Theory*. Springer, 1996.