Geometrical Scaling Effects on Carrier Transport in Ultrathin-Body MOSFETs

Jan-Laurens P.J. van der Steen

Geometrical Scaling Effects on Carrier Transport in Ultrathin-Body MOSFETs

Jan-Laurens Pieter Jacobus van der Steen

The graduation committee consists of:

Chairman:	prof. dr. ir. A.J. Mouthaan	University of Twente
Secretary:	prof. dr. ir. A.J. Mouthaan	University of Twente
Supervisor:	prof. dr. J. Schmitz	University of Twente
Supervisor:	prof. dr. L. Selmi	University of Udine
Asst. Supervisor:	dr. ir. R.J.E. Hueting	University of Twente
Referee:	dr. D. Esseni	University of Udine
Referee:	dr. A.J. Scholten	NXP Semiconductors
Members:	prof. dr. ir. W.G. van der Wiel	University of Twente
	prof. dr. S. Cristoloveanu	IMEP-INP, MINATEC
	prof. dr. F. Crupi	University of Calabria

This research was supported by NXP Semiconductors, Eindhoven, The Netherlands. The project was carried out in the Semiconductor Components group, MESA+ Institute for Nanotechnology, University of Twente, The Netherlands and in DIEGM, University of Udine, Italy.

PhD. thesis – University of Twente, Enschede, The Netherlands

Geometrical Scaling Effects on Carrier Transport in Ultrathin-Body MOSFETs
Author: J.-L.P.J. van der Steen
ISBN: 978-90-365-3158-0
DOI: 10.3990/1.9789036531580

Cover design: Niek J. Bouman, Utrecht
Print: Gildeprint Drukkerijen, Enschede

© 2011, J.-L.P.J. van der Steen
All rights reserved.

The cover is based on the 2D energy dispersion in the silicon crystal with (100) quantization direction, obtained from full-band LCBB calculations (Ch. 2). The intersect depicts a typical energy subband profile along the transport direction of a MOSFET operating in the ON-condition (Ch. 4).

GEOMETRICAL SCALING EFFECTS ON CARRIER TRANSPORT IN ULTRATHIN-BODY MOSFETS

DISSERTATION

to obtain
the degree of doctor at the University of Twente,
on the authority of the rector magnificus,
prof. dr. H. Brinksma,
on account of the decision of the graduation committee
to be publicly defended
on Friday, April 1st, 2011 at 14:45

by

Jan-Laurens Pieter Jacobus van der Steen

born on December 26th, 1980
in Wageningen, The Netherlands

This dissertation is approved by:
prof. dr. J. Schmitz (supervisor)
prof. dr. L. Selmi (supervisor)
dr. ir. R.J.E. Hueting (asst. supervisor)

UNIVERSITÀ DEGLI STUDI DI UDINE

TESI DI DOTTORATO DI RICERCA

per ottenere il titolo di

Dottore di Ricerca dell'Università di Udine

e

Doctor at the University of Twente

Dipartimento di Ingegneria Elettrica, Gestionale e Meccanica

Faculty of Electrical Engineering, Mathematics and Computer Science

Jan-Laurens Pieter Jacobus VAN DER STEEN

Enschede (NL), April 1st, 2011

Geometrical Scaling Effects on Carrier Transport in Ultrathin-Body MOSFETs

Relatori: David ESSENI, Jurriaan SCHMITZ
Co-relatori: Luca SELMI, Ray HUETING

Commissione esaminatrice
prof. dr. ir. A.J. Mouthaan, *Presidente*
prof. dr. J. Schmitz, *Tutor*
prof. dr. D. Esseni, *Tutor*
dr. ir. R.J.E. Hueting, *Co-tutor*
prof. dr. S. Cristoloveanu, *Esaminatore esterno*
dr. A.J. Scholten, *Esaminatore esterno*
prof. dr. L. Selmi, *Invited*
prof. dr. F. Crupi, *Invited*
prof. dr. ir. W.G. van der Wiel, *Invited*

. . . in herinnering aan mijn Opa, Jan Mugge

Contents

1 Introduction **1**

1.1 The MOSFET . . . 1
1.2 Trends in the semiconductor device industry . . . 3
1.3 Key features of multiple-gate devices . . . 6
1.4 Challenges related to multiple-gate UTB MOSFETs . . . 8
1.5 Related topics . . . 12
1.6 Outline . . . 14

2 The Impact of Quantization **17**

2.1 Potential profile and charge distribution . . . 18
2.2 Bandstructure and carrier confinement . . . 23
2.3 Validity of the EMA . . . 36

3 Energy Band Offset Extraction **51**

3.1 Conventional methods . . . 52
3.2 Band offset extraction from $I_{DS}(T)$ . . . 53
3.3 Results . . . 57
3.4 Discussion . . . 69

4 Quasi-Ballistic Transport **71**

4.1 Carrier scattering . . . 73
4.2 What can we learn from Monte-Carlo simulations? . . . 75
4.3 Modelling quasi-ballistic transport . . . 80
4.4 The critical length for back-scattering . . . 86
4.5 A new model for the backscatter coefficient . . . 103

5 Conclusions **121**

5.1 The impact of quantization . . . 121
5.2 Energy band offset extraction . . . 122

5.3 Quasi-ballistic transport 123
5.4 Recommendations and closing remarks 124

Bibliography **127**

A The potential in subthreshold **143**

B The variational approach **145**

C UTB-SOI MOSFETs in subthreshold **147**

D Mobility extraction in subthreshold **149**

E The ballistic distribution function **151**

F The velocity of the ballistic flux **153**

G The critical length in the SSC model **155**

Summary **159**

Samenvatting **161**

Riassunto **163**

List of Publications **165**

Acknowledgements **167**

One

Introduction

This chapter presents a concise introduction to the terminology which is commonly used to describe the transistor, its behavior, and the trends in the semiconductor industry.

1.1 The MOSFET

The workhorse of integrated circuits (ICs), such as microprocessors and semiconductor memories, is the metal-oxide-semiconductor field-effect transistor (MOSFET). The main feature of the MOSFET is to control the output *current* by applying a *voltage* at the input.

A schematic representation of a MOSFET is shown in Fig. 1.1. The MOSFET contains a source and a drain terminal, which can be connected through a conductive channel at the semiconductor surface. The conductivity of the channel is controlled by the third terminal, the gate, which is separated from the channel by an insulator. Thus, the gate controls the current in the channel through a *capacitive* coupling (using the 'Field Effect').

We can distinguish two types of MOSFETs, based on the type of carriers that constitute the current: the n-MOSFET, where the channel is formed by electrons, and source and drain are highly doped n-type regions. The channel conductivity and the resulting current increases with increasing gate bias. Secondly, the p-MOSFET, where the channel consists of holes, and the source and drain are p-type. The p-MOSFET becomes more conductive with more negative gate bias. In integrated circuits (or 'chips') n-type and p-type MOSFETs are used together, mainly to reduce the static power consumption. Often the acronym CMOS is used, referring to Complementary-MOS [1].

In Fig. 1.1 we have indicated a few important MOSFET parameters: the *gate* length L_G, which can differ from the *channel* length L due to extension of the source and drain regions underneath the gate. Furthermore, the active device region is doped to level N_X, where 'X' denotes either acceptor (A) or

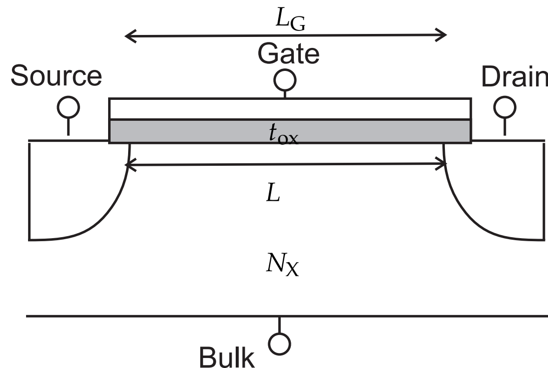


Figure 1.1: Schematic representation of a MOSFET, indicating the gate length L_G, the channel length L, the thickness of the gate dielectric t_{ox}, and the doping level in the channel N_X.

donor (D) type of dopant atoms, depending on whether the device is an *n*-channel or *p*-channel MOSFET, respectively. Since the MOSFET relies on the field effect to control the channel, the oxide (e.g. SiO_2) thickness t_{ox} is an important parameter. We note in passing that the device sketched in Fig. 1.1 is a bulk silicon MOSFET, which is a *planar* device. We will comment on other device architectures in the course of this chapter.

The conventional MOSFET has four terminals: a source, drain, gate and bulk. Usually, the bulk is connected to the source, which acts as the reference potential. In general terms, the gate controls the amount of charge in the channel, and the source/drain determine how fast the charge moves, i.e. the current.

In digital circuits (such as microprocessors) the MOSFET is used as a 'switch', which is either 'on' or 'off', depending on whether the gate voltage is above or below a certain *threshold voltage*, respectively. The MOSFET differs from an ideal switch in that the current in the off-state (I_{OFF}) is not exactly zero. It should, however, be as small as possible and much smaller than the on-state current (I_{ON}). Hence, an important transistor figure-of-merit is the ratio I_{ON}/I_{OFF}. Furthermore, another important characteristic is the 'steepness' of the switching characteristic or, more precisely, the gate voltage we have to apply in order to increase the output current by a factor of 10, which is called the *subthreshold swing* or its inverse, the *subthreshold slope*.

1.2 Trends in the semiconductor device industry

Although the principle of the MOSFET was invented by Lilienfeld already in the 1930s [2–4], the first actual MOSFET was fabricated in 1960 by Kahng and Atalla [5]. The biggest hurdle to overcome was to find and fabricate a good-quality gate dielectric. This issue was solved by Ligenza and Spitzer in 1960, who studied the mechanism of thermal oxidation of silicon and produced the first device-quality Si/SiO_2 stack [6].

Since the beginning of the IC era, the density of components in an IC has increased exponentially. The semiconductor industry is geared towards reducing the so-called 'minimum feature size', which essentially is the MOSFET's gate length. The transistor scaling and the improved circuit performance was mainly enabled by the improved control of the fabrication technology.

The trend in the semiconductor industry was first discussed by Gordon Moore in 1965 [7], starting from the notion that reducing costs is the major drive for the miniaturization of integrated electronics. The cost of a single component, such as a transistor, decreases as we put more and more components in a chip of a given silicon area. On the other hand, the increased complexity causes more circuits to fail, which counter-acts the cost advantage we obtained by increasing the number of components. Thus, there is an optimum number of components per unit silicon area that leads to minimum costs per component. Moore observed that, until then, the so-called 'complexity for minimum component costs' had increased by a factor of two per year. Because Moore predicted that the trend would continue, albeit in a slower pace from the early 80's onwards [8], his observation is commonly referred to as "Moore's Law". The implications of this 'Law' and the resulting requirements at the various levels are specified in the International Technology Roadmap for Semiconductors (ITRS) [9], which is the result of a joint effort of the entire semiconductor community, consisting of representatives from both industry and academia.

Extensive overviews of the scaling trends at several points in time are reported in [8,10,11] [12, Ch. 6]. In this chapter we will briefly outline the various scaling options at the transistor level. We will point out the 'design knobs' for future CMOS technology nodes, in order to explain the demand for a new type of MOSFET architecture, the so-called Multiple Gate MOSFET.

Since the beginning of the integrated circuit era, the minimum feature size (the gate length) has been reduced by more than two orders of magnitude [12]. Besides the resulting increased density of components, reduction of the MOS-

FET channel length has lead to an enormous performance gain, since the output current increases with decreasing channel length. While scaling the device dimensions, the challenge is to maintain the 'ideal' MOSFET characteristics, that is, the behavior of a 'long-channel' MOSFET in order to have a high I_{ON}/I_{OFF} ratio.

However, as mentioned before, the actual (electrical) channel length is often smaller than the gate length due to extension of the source and drain regions underneath the gate. Furthermore, as the gate length decreases, the depletion regions formed by the source-channel and drain-channel junctions become comparable to the channel length. One can imagine that if there is an unintentional conductive path between source and drain (*punch-through*) independent of the gate voltage, the transistor action is lost.

One way to reduce the extension of the source/drain depletion regions into the channel is to increase the channel doping. One of the drawbacks, however, is an increased threshold voltage V_{TH}. In order to maintain V_{TH} at a reasonable level, the oxide thickness t_{ox} must be reduced, which results in an increased 'field effect' from the gate. Hence, this typical example indicates that the various device parameters are coupled. Different scaling rules have been employed to increase the device performance while, for example, maintaining either a constant lateral electric field, or a constant supply voltage [12, p. 329].

Whichever scaling rule is employed, device scaling is far from trivial, even more because some factors such as the band gap and the built-in voltages do not scale. The end of the 'happy scaling' era is explored in many reports, an overview of which is presented in [13, and references therein]. Today the semiconductor industry is facing limits that are –besides the technological challenges– of a more fundamental nature, i.e. related to the silicon properties and device architecture.

Recently, alternative solutions are being explored for their possible application in future CMOS technologies. These options include the choice for other channel materials, such as germanium and various III-V compounds which generally feature a high mobility [14–16], or strained Si devices, see e.g. [17]. The main merit is the increase of the maximum drive current I_{ON}.

A second strategy concerns the pursuit of new device architectures, which aim at enhancing the 'field effect': if we are somehow able to increase the gate control over the channel, i.e. the electrostatic integrity, we can further decrease the channel length while keeping the impact of short-channel effects under control. One way of enhancing the gate control is to switch to a *multiple gate* architecture. In addition to the multiple gates, the device performance can be

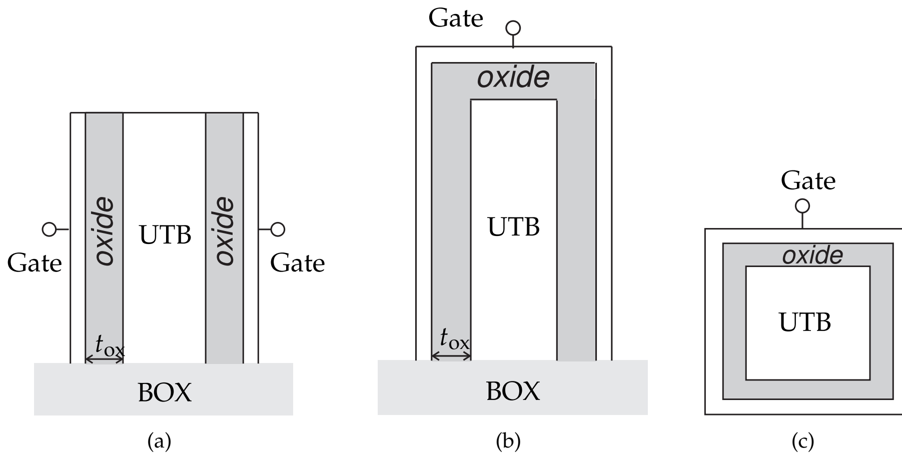


Figure 1.2: Schematic representation of various types of multiple-gate structures on a Buried Oxide (BOX) layer: a Double-Gate MOSFET (a), FinFET (b) and a Gate-All-Around or Nanowire FET (c). 'UTB' denotes the ultrathin body. The current flow is in the direction normal to the plane.

further optimized by reducing the thickness of the active device region, the 'body', which is then labeled 'ultrathin body' (UTB).

Various types of novel architectures are considered, some of which are exemplified in Fig. 1.2. Fig. 1.2(c) depicts a Gate-All-Around or, similarly, a Nanowire FET, where the body is entirely surrounded by the gate. In contrast to the conventional bulk and Single Gate (SG) SOI MOSFET, the multiple gate devices in Fig. 1.2 and similar are *non-planar* devices, where at least part of the conductive channel is not aligned with the wafer surface. The idea of a vertical ultrathin silicon-on-insulator (UTB-SOI) MOSFET was proposed by Hisamoto in 1989 [18], and it was named 'DELTA'. Later, this new type of device gained interest as a candidate to replace the bulk MOSFET in future CMOS technologies, and it is nowadays referred to as the FinFET [19–21]. An extensive overview of the various multiple-gate devices, along with the broad variety of names they have in literature, is presented in [22]. In summary, these novel device architectures aim at improving the electrostatic device behavior, which enables a further reduction of the channel length and thus an increase in I_{ON}/I_{OFF}.

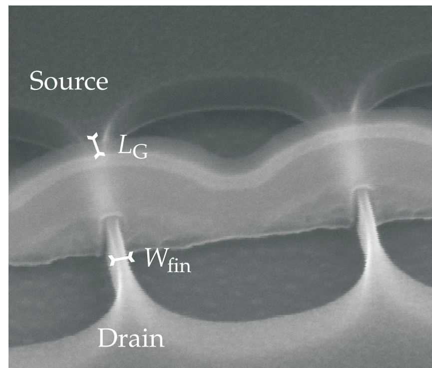


Figure 1.3: SEM image of two FinFETs in parallel, with shorted drain and source regions and a common gate. Image courtesy of M. van Dal, NXP-TSMC Research Center, Leuven, Belgium.

1.3 Key features of multiple-gate devices

This chapter proceeds by pointing out the typical features of FinFETs and similar multiple gate UTB devices, along with the currently open issues and challenges related to those new device architectures. An example of a FinFET is given in Fig. 1.3, showing a SEM image of two FinFETs in parallel, fabricated in NXP-TSMC Leuven, Belgium. The source and drain regions of the respective fins are shorted, and the common gate runs over all fins. The fin width is around 10 nm, and the height is 60 nm.

Improved electrostatics

One of the main incentives to consider multiple gate architectures is the improved electrostatic behavior. In fact, the combination of multiple gates and an ultrathin body enhances the control over the charge in the channel, which translates into a steeper subthreshold slope compared to a single-gate bulk MOSFET of equal channel length [23–25]. Ideally, the subthreshold slope is independent of the applied drain voltage: the subthreshold current is determined by the source-channel barrier which, on its turn, is controlled by the gate. However, in a short channel, the barrier near the source can become sensitive to the applied drain voltage due the close proximity of the source and drain. This effect is called drain-induced barrier lowering (DIBL), and it is a typical short-channel effect (SCE). The enhanced gate-induced field effect in

multiple gate UTB devices increases the device's immunity to short channel effects [26, 27].

An attractive feature of an UTB is that it is sufficiently thin to be entirely depleted, which enables the occurrence of *volume inversion*. This means that, depending on the bias and geometry, the inversion layer can extend throughout the entire UTB. The principle of volume inversion, originally reported by Balestra [28], is central to the operation of UTB-SOI devices and it will be discussed in more detail in Chapter 2. Here we highlight the main implications.

Because of the fully depleted body, UTB devices exhibit an inverse subthreshold slope close to the theoretically ideal value of around 60 mV/dec. at 300 K [29]. Furthermore, there is no need for channel doping to set the threshold voltage. In fact, the threshold voltage is determined by the workfunction of the gate material rather than by the channel doping, as pointed out in [30]. With the reduction of the channel length, random dopant fluctuation has become one of the major sources of mismatch in conventional bulk MOSFETs [31–34]. Hence, the fact that no doping is required makes UTB devices promising CMOS candidates from the viewpoint of stochastic variability (random device-to-device variation).

To exemplify the FinFET IV characteristics, Fig. 1.4 plots the typical measured drain current I_D in a long (L_G = 1 μm) and short channel (L_G = 35 nm) device for low and high drain bias. The processing and device details are reported in [35].

The long channel device exhibits an ideal subthreshold slope of 60 mV/decade. The subthreshold slope in the short channel devices is less steep (88 mV/dec), but still better than a bulk MOSFET with the same gate length. Furthermore, the shift in the short-channel high-V_{DS} subthreshold regime is due to DIBL, a typical short channel effect [12]. The level of DIBL actually poses a limit on the minimum gate length, because it reduces the I_{ON}/I_{OFF} ratio.

Besides the improved electrostatic properties, there might be some additional advantages linked to the UTB. The distribution of the inversion charge depends both on the bias and the UTB thickness. Volume inversion is believed to be beneficial in terms of mobility [28], since a significant fraction of the carriers reside around the center of the channel, away from the semiconductor-oxide interface. As we will see in Ch. 2, the carrier distribution is tightly related to the strength of quantum mechanical confinement, which tends to repel the carriers from the interface. Another effect related to the mobility is that the low channel doping will reduce the impact of impurity (Coulomb)

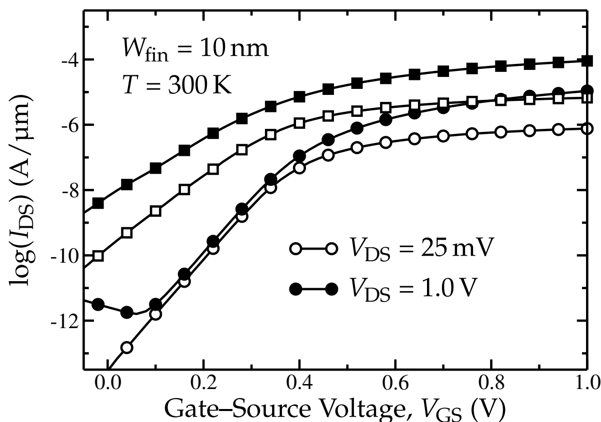


Figure 1.4: Measured drain current I_{DS} in a long (L_G = 1 µm, circles) and a short channel (L_G = 35 nm, squares) FinFET for low and high drain bias. The devices were fabricated by NXP-TSMC Research Center, Leuven, Belgium.

scattering in the channel. However, the proximity of the gates may adversely affect the mobility, since the relative importance of carrier scattering against the semiconductor-oxide interface increases. In short, the interaction of the various mechanisms which determine the mobility in UTB-SOI devices has been, and still is, an active area of research [36–41].

From the technology point of view, the FinFET is an interesting option for future CMOS technology nodes since most of its processing steps are very similar to the 'conventional' CMOS steps. In addition, the non-planar technology allows for high-density stacking, which makes the FinFET a promising candidate for future logic applications.

1.4 Challenges related to multiple-gate UTB MOSFETs

Despite the fact that FinFET-like structures are widely recognized as promising candidates for future CMOS technologies, there are some challenges to face. A compact overview will be given in the following.

Access resistance and parasitic capacitance

Although beneficial in terms of electrostatic properties, reducing the UTB thickness increases the series resistance and, furthermore, makes it progressively

more difficult to maintain the source and drain doping concentration at the desired level. Hence, without any special precautions, the source/drain access resistance will significantly degrade the device performance. Increasing the implantation dose in order to achieve a sufficiently high source/drain doping concentration is limited, because high implantation dose leads to amorphization: a high implantation dose induces significant lattice damage, yielding a poor crystallinity [42]. So, there is a limit to the reduction in access resistance through ion implantation. Instead, a viable strategy to lower the source/drain resistance is to enlarge the source/drain access regions to the channel using selective epitaxial growth (SEG). This was shown to be a successful way of lowering the source/drain resistance [35, 43].

We have so far focussed on the static (DC) device behavior. For RF applications, besides the characteristics mentioned above, properties like the cut-off frequency determine the dynamic switching behavior. Hence, an additional aspect to consider is the impact of parasitic capacitive effects, due to the unintentional capacitive coupling of the gate to the various regions of the device. These so-called *fringing* capacitances could have a detrimental impact on the high-frequency behavior [44, 45].

Variability

As mentioned before, in FinFETs there is no need for channel doping to set the threshold voltage, thanks to the UTB. Hence, the impact of random doping fluctuation (RDF) due to channel doping will be reduced. This comes, however, at the expense of a more profound effect of other sources of variability. More precisely, UTB thickness variations cause direct device-to-device variation in the resulting drive currents [46]. Furthermore, even within a single device, UTB thickness fluctuations could produce both directly and indirectly local variations in mobility, band gap and, hence, the current density (e.g. [37]).

Self-Heating

In presence of high current densities, not only the carrier temperature (actually, the carrier energy) may be significantly higher than the lattice temperature, also the average temperature of the lattice itself may well exceed the ambient temperature. This effect, called self-heating (SH), is often observed in power devices and SOI based devices [47–49]. Recent simulation studies [50–54] have indicated that also FinFETs may exhibit self-heating ef-

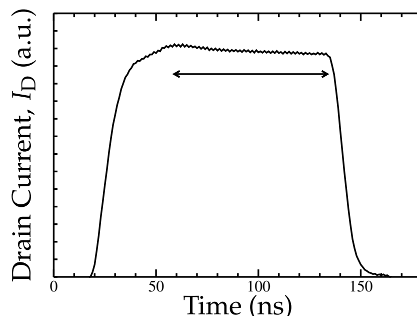


Figure 1.5: Typical trace of drain current vs. time, when applying a voltage pulse on the gate. The arrow indicates the measurement window.

fects: the UTB is entirely encapsulated in an oxide with a low thermal conductivity. In addition, the connection with regions of the device that do feature a good thermal conductivity (such as the source and drain), and thus could act as 'heat sink', is limited. In this respect, FinFETs on a bulk silicon substrate may be more favorable, since –at least part of– the lattice heat is dissipated through the bulk.

If the device dimensions are comparable to the bulk phonon mean free path (around 100 nm in silicon), the thermal conductivity decreases rapidly with decreasing geometry [55]. Furthermore we note that, in case of FinFETs and similar devices, the dimensions in all directions are comparable, hence estimating the thermal resistance R_{th} based on a simple 1D analysis may become questionable.

In order to complement the simulation data presented in literature with experimental results, we performed pulsed IV measurements on SOI FinFETs for various geometries. In the pulsed IV measurement, following [56], the device characteristics can be measured without the effect of self-heating. To this end, instead of a DC bias, a short pulse is applied to the gate. Meanwhile, V_{DS} is kept constant, and the drain current I_D is measured after a certain settling time. A typical graph of $I_D(t)$ is shown in Fig. 1.5.

In the cases shown in this work, the pulse width was set to 100 ns and averaging was performed over 100 successive pulses. To reduce the impact of unwanted parasitic effects (e.g. from the bias-tee), the drain current variations are recorded only in the last 60% of the pulse, as indicated in Fig. 1.5.

Fig. 1.6 plots the resulting pulsed-IV data along with the DC IV curves, for

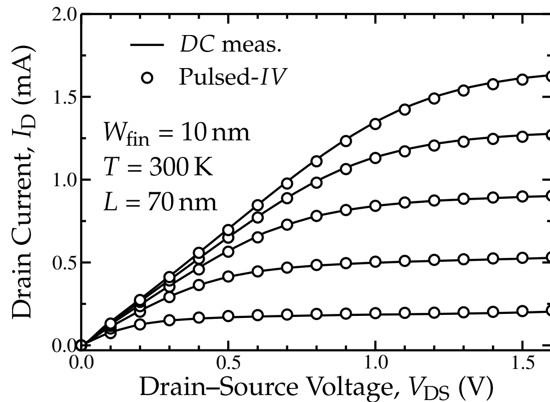


Figure 1.6: Comparison of pulsed and DC *IV* measurements for an SOI FinFET with 30 parallel fins, and 70 nm gate length.

30 parallel fins with 70 nm gate length, 60 nm fin height and 200 nm spacing between the neighboring fins. Further device details are documented in [57]. No significant difference between the DC and the pulsed-*IV* measurements can be observed. This statement was verified for various FinFET geometries, with gate length ranging from 1 μm down to 70 nm, and devices consisting of 1 up to 168 fins in parallel [58].

The above results indicate that either the impact of Self-Heating on the drain current is small, or that the thermal time constant is smaller than the pulse width in our measurement, i.e. below 100 ns. Hence, in the latter case we cannot give any conclusive answer whether Self-Heating significantly affects the device current.

Recently, it was demonstrated that *s*-parameter measurements are instrumental in assessing the impact of self-heating on the device characteristics [58]. Central to this method is the fact that the crystal lattice is able to 'adjust' its temperature to slowly varying signals whereas it fails to track the fast AC signals. This concept was theoretically investigated for bipolar transistors in [59]. The response of the thermal effects to the applied signals is determined by the thermal resistance and capacitance. As for MOSFETs, particularly the drain capacitances seem to exhibit a high sensitivity to Self-Heating [58].

Using *s*-parameter measurements the dominant thermal capacitances and resistances can be extracted. In fact, simulation of the aforementioned pulsed-*IV* experiment using a 4th order thermal network connected to a FinFET com-

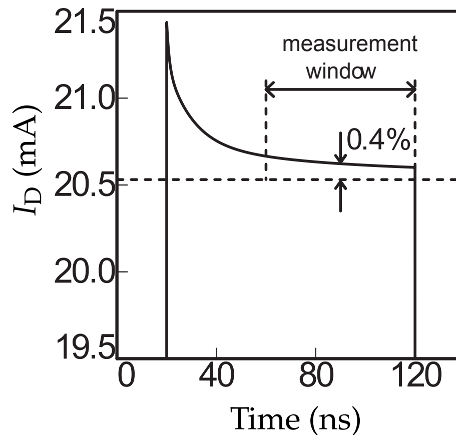


Figure 1.7: Simulation of the pulsed *IV* experiment. In the measurement window, the pulsed I_D (solid line) has nearly saturated to the DC I_D, which explains the virtual absence of Self-Heating effects in the pulsed-*IV* measurements (e.g. Fig. 1.6). Figure reproduced from [58].

pact model [60] reveals that the difference between the 'experimental' pulsed and DC values of I_D in the measurement window has dropped to below 1%, as indicated in Fig. 1.7. This small difference explains why no self-heating effect was observed in the pulsed-*IV* measurements. Furthermore, the dominant thermal time constant is found to be around 100 ns, which is comparable to the pulse width used in the measurements.

Fig. 1.8(a) reports the simulated DC *IV* curves with and without self-heating, demonstrating that the resulting change in drain current due to self-heating is modest, at most 4% for the highest bias and the shortest channel. Yet, Fig. 1.8(b) shows that for the same condition the corresponding average channel temperature rise, extracted from the thermal network, is expected to be around 80 K; the peak value in the channel may well exceed this value. Hence, self-heating might become a serious issue for even further scaled devices, which feature a higher I_{ON} and thus an increased power dissipation.

1.5 Related topics

Besides moving to a new type of device architecture, some other trends are currently observed. They do, however, not strictly apply to multiple gate de-

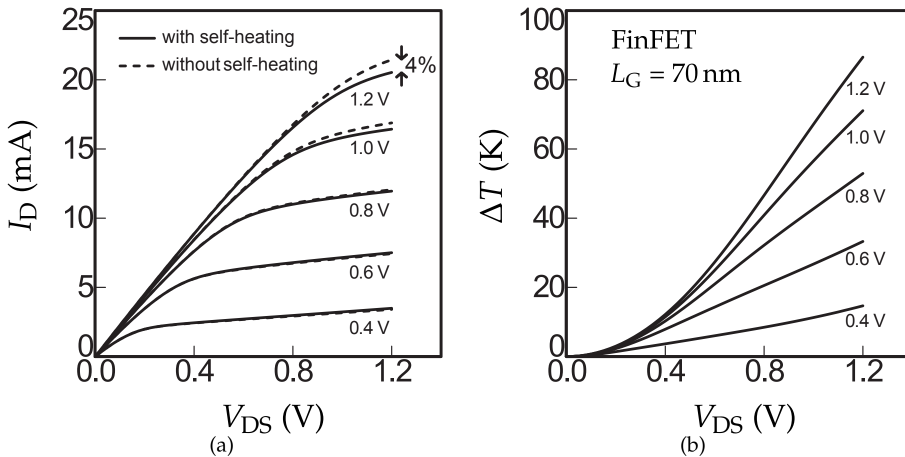


Figure 1.8: (a) Simulated DC drain current with (solid lines) and without self-heating, obtained from a FinFET compact model [60] with a 4th order thermal network. The values of the thermal elements have been obtained from *s*-parameter measurements. (b) the corresponding increase in average temperature ΔT in the FinFET. Figures reproduced from [58].

vices only, and are therefore briefly discussed in this section along with some references for further reading.

High-κ dielectrics

Increasing the gate capacitance by reducing the oxide thickness t_{ox} has reached its limits, due to the gate leakage current which increases with decreasing t_{ox} [12]. In order to enhance the gate-to-channel coupling while keeping the gate leakage current at an acceptable level, alternative gate dielectric materials are employed, which generally feature a high dielectric constant (κ). By employing high-κ materials, such as HfO_2 (κ = 25) we can obtain a given gate capacitance using a dielectric layer which is thicker compared to SiO_2 (κ = 3.9), thereby suppressing the gate (tunnelling) current. However, the drawback is that high-κ dielectrics tend to have a higher density of interface states and fixed charge which can lead to mobility reduction [61] and matching issues [62, 63].

Replacing the gate material

As for the gate itself, conventionally the MOSFET gate is polysilicon. If sufficiently high doped, the polysilicon behaves electrically metal-like. With decreasing device dimensions and increasing normal field and inversion densities, depletion effects in the polysilicon gate can occur. To circumvent this problem, and to have a better interface with high-κ dielectrics, the polysilicon gate is replaced by a metal gate electrode [17].

Schottky Source/Drain

Replacing the silicon based source and drain with a metal is considered to be a viable strategy to reduce the access resistance to the channel [64], because it allows for abrupt source/drain junctions. The channel entrance and exit then become *Schottky* junctions, because of the metal/semiconductor junction. This poses some interesting questions on modelling of the transport across these barriers, which is an active area of research [65–69].

Strain

By applying strain to the active device region, the band structure is modified. It directly alters the alignment of the valleys or, more precisely, the corresponding conduction effective masses and thus depends on the crystallographic orientation. The aim of 'strain engineering' is to enhance the mobility. The deformation of the band structure can be obtained through gate stressors or by incorporating germanium in the source and drain (SiGe). In effect, the lattice constant in the channel is modified, inducing changes in the bandstructure. Reports on strained-silicon MOSFETs can be found in [17,70–74].

1.6 Outline

The above strategies concerning alternative materials and new device architectures are not mutually exclusive and certainly deserve a careful and detailed study. In this work, we will mainly focus the impact of reduction of the UTB and the channel length on the DG/FinFET device characteristics.

When moving to UTB devices, some *material* properties become *device* properties, since they become geometry dependent and, in particular, sensitive to the thickness of the UTB. These device properties include the mobility, band gap, but also the thermal device characteristics. Hence, one should be very

careful with extrapolating device characteristics towards, say, the 'limit', since the predominant behavior of one effect over the other involves careful optimization of the device geometry.

Directly linked to the above, with the miniaturization of the device dimensions we enter several new regimes. From a bandstructure and electrostatic viewpoint, we have to account for the effect of carrier confinement. This topic will be discussed in Chapters 2 and 3. As for reducing the channel length, the distance from source to drain becomes comparable to the carrier mean-free-path, which implies that carriers encounter only a limited number of scattering events. Thus, carrier transport enters the so-called quasi-ballistic regime, which is the subject of Chapter 4. The main findings and key results of this work will be summarized in chapter 5.

Two

The Impact of Quantization

In the introductory Chapter 1, we have briefly discussed the key features of FinFETs and similar multiple-gate devices. In summary, these devices are widely recognized as promising candidates for future CMOS technology nodes, mainly because the combination of multiple gates and an ultrathin body (UTB) offers an enhanced control over the charge in the channel [19–21] ("improved electrostatics").

In terms of input–output characteristics, this translates, amongst others, into a steeper subthreshold slope and an improved immunity against short-channel effects (SCE) compared to a conventional (single gate) bulk MOSFET with the same channel length. Alternatively, for a given impact of SCE in terms of subthreshold slope and DIBL, the multiple-gate architecture allows for further reduction of the channel length compared to the single-gate bulk MOSFET.

Along with the introduction of the UTB comes a new crucial device 'parameter', namely the body thickness (t_{Si}) or, for FinFETs, the fin width (W_{fin}). In fact, as pointed out in [22], it is the ratio of channel length and body thickness that determines the so-called *electrostatic integrity*, which is a measure of the device's immunity to SCE and the (unwanted) impact of the drain potential on the charge distribution in the channel. As a result, we find that the smaller t_{Si}, the shorter we can make the channel without running into severe SCE.

This chapter is focussed on the impact of reducing the thickness of the UTB ('vertical scaling'). In particular, if the thickness of the UTB is reduced to values in the order of the De Broglie wavelength [12][1], we have to account for quantum-mechanical confinement of the carriers in the UTB, which is referred to as *quantization*. When presenting the results, the various quantities

[1] approximately 17 nm for an electron with energy 26 meV and mass $0.19m_0$, with m_0 the electron rest mass

are shown as a function of the body thickness t_{Si} in a double-gate MOSFET. Unless stated otherwise, instead of t_{Si} one could equivalently read W_{fin}, i.e., the width of the FinFET body.

The first part of this chapter deals with the impact of quantization on the carrier distribution in the channel. Then, in the second part, we will focus on the band structure with a detailed comparison of two quantization models. To be more precise, we examine the validity of the almost universally used *effective mass approximation* (EMA) by means of a systematic comparison with results obtained from rigorous Full-Band calculations.

The definition of the axis labels is depicted in Fig. 2.1, showing a cross-section along the longitudinal direction (a) and along the height of the fin (b). We label the longitudinal (transport) direction x, the direction normal to the gates z; the height of the fin is directed along y. Fig. 2.1(b) also represents a double-gate (DG) device, cut along the width direction.

We identify two 'special cases' of the FinFET; first, if the top oxide ($t_{\mathrm{ox,t}}$) is thick, the influence of the top gate is small, which effectively makes the device a DG device. Second, if the aspect ratio is high, i.e. $H_{\mathrm{fin}} \gg W_{\mathrm{fin}}$, the device analysis can often be simplified to the 2D cross-section shown in Fig. 2.1(b).

2.1 Potential profile and charge distribution

In this section, we will have a closer look at the potential profile and charge distribution in the channel of a long-channel Fully Depleted (FD) symmetric DG MOSFET, or equivalently, a long-channel FinFET with $H_{\mathrm{fin}} \gg W_{\mathrm{fin}}$, as e.g. shown in 2.1(b). At this stage, we restrict our analysis to the *classical* approach, that is, we do not account for quantization. In section 2.2, instead, the impact of quantum confinement on the carrier distribution will be discussed.

Modeling of the potential and charge in multiple-gate devices is an active area of research. A comprehensive overview can be found in [75], which covers and compares a variety of models reported in literature. In this work, we adhere to the derivation and definitions as presented by Taur in [76], and reprint the resulting key equations.

The carrier distribution and potential profile are related through Poisson's equation, given by

$$\frac{\partial^2 \psi}{\partial x^2} + \frac{\partial^2 \psi}{\partial z^2} = \frac{\rho_{\mathrm{sc}}}{\varepsilon_{\mathrm{Si}}}(x, z) \tag{2.1}$$

with $\varepsilon_{\mathrm{Si}}$ the permittivity of silicon, ρ_{sc} the space charge density and $\psi(x, z)$ the potential as a function of the position along the channel (x) and the direction

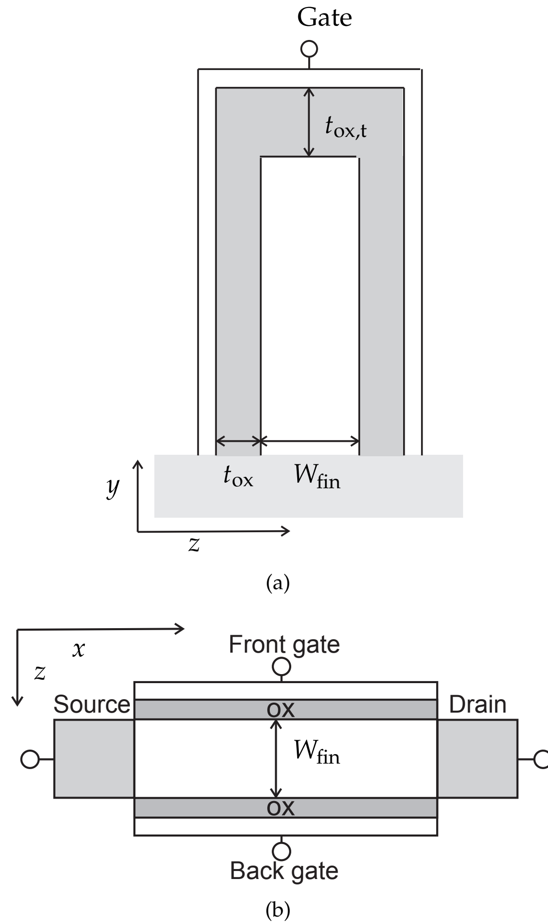


Figure 2.1: (a) Schematic cross-section of the FinFET along the channel direction (x). (b) a cross-section along the height of the FinFET (y), or along the width of a Double-Gate MOSFET. The direction perpendicular to the gates is labeled z.

normal to the gates (z). In the case of a long-channel device, we assume that the lateral variation of the longitudinal electric field, $E_x = -\partial\psi/\partial x$, is much less than the variation of the normal electric field $E_z = -\partial\psi/\partial z$, i.e.

$$\frac{\partial^2\psi}{\partial x^2} \ll \frac{\partial^2\psi}{\partial z^2} \tag{2.2}$$

This approximation is commonly referred to as the Gradual Channel Approximation (GCA) [77] and it is often used because it reduces the 2D poisson equation (2.1) to one dimension [78]. By embracing the GCA, the Poisson equation reads

$$\frac{d^2\psi}{dz^2} = \frac{\rho_{sc}}{\varepsilon_{Si}} = -\frac{q}{\varepsilon_{Si}}\left[p(z) - n(z) + N_D^+(z) - N_A^-(z)\right] \tag{2.3}$$

with n and p the electron and hole concentration, respectively; q is the elementary charge, N_D^+ and N_A^- are the ionized donor and acceptor concentrations, respectively. We now assume a lightly or undoped UTB, which means that N_D^+(z) and N_A^-(z) in (2.3) are small compared to the mobile carrier concentration (p and n). Furthermore, in case of an n-type source and drain, the electron density in the channel will exceed the hole density, so we include only the electron concentration in (2.3). The resulting Poisson equation reads

$$\frac{d^2\psi}{dz^2} = \frac{q}{\varepsilon_{Si}} n_i \exp\left(\frac{q\psi(z)}{k_B T}\right) \tag{2.4}$$

where we have used

$$n(z) = n_i \exp\left(\frac{q\psi(z)}{k_B T}\right) \tag{2.5}$$

with

$$n_i = \sqrt{N_{VB}N_{CB}} \exp\left(-\frac{E_C - E_V}{2k_B T}\right) \tag{2.6}$$

assuming Boltzmann's approximation; n_i is the intrinsic carrier concentration, k_B is the Boltzmann constant and T the absolute temperature; N_{CB} and N_{VB} are the effective *Density of States* (DOS) in the bulk conduction and valence band respectively.

Unless stated otherwise, in the following we assume a symmetric DG n-MOSFET: the front and back-gate are at the same potential, and have the same work function and oxide thickness. Then, at $z = 0$, i.e. the center of the channel, we find that $d\psi/dz = 0$. Integrating (2.4) twice, $\psi(z)$ can be shown to be [76]

$$\psi(z) = \psi_0 - 2u_{th} \ln\left(\cos\left[\sqrt{\frac{qn_i}{2\varepsilon_{Si}u_{th}}} \exp\left(\frac{\psi_0}{2u_{th}}\right) z\right]\right) \tag{2.7}$$

with ψ_0 the potential at $z = 0$, and the thermal voltage $u_{\mathrm{th}} = k_{\mathrm{B}}T/q$. In the following we define $\psi_{\mathrm{s}} \equiv \psi(\pm\frac{t_{\mathrm{Si}}}{2})$, which we call the *surface potential*. The surface potential can be related to the applied gate bias V_{GS} through the boundary condition at the Si/SiO$_2$ interface ($z = \pm t_{\mathrm{Si}}/2$), resulting in

$$\frac{\varepsilon_{\mathrm{ox}}}{t_{\mathrm{ox}}}\left(V_{\mathrm{GS}} - \Delta\phi_{\mathrm{ms}} - \psi_{\mathrm{s}}\right) = \sqrt{2\varepsilon_{\mathrm{Si}} k_{\mathrm{B}} T n_{\mathrm{i}} \left[\exp\left(\frac{q\psi_{\mathrm{s}}}{k_{\mathrm{B}}T}\right) - \exp\left(\frac{q\psi_0}{k_{\mathrm{B}}T}\right)\right]} \tag{2.8}$$

with $\Delta\phi_{\mathrm{ms}}$ the workfunction difference between gate and intrinsic silicon and $\varepsilon_{\mathrm{ox}}$ the dielectric constant of the gate oxide.

Eq. (2.7) defines an implicit link between the surface potential ψ_{s} and the potential at the center of the channel ψ_0. One cannot find a closed-form expression for $\psi(z)$ for the entire range of V_{GS}, without making any simplifying assumptions on either the potential profile, or the charge distribution. However, we can identify two limiting cases: $\psi_{\mathrm{s}} \approx \psi_0$, corresponding to the subthreshold region, and $\psi_{\mathrm{s}} \gg \psi_0$, which occurs at V_{GS} well above the threshold voltage (strong inversion). We will illustrate these cases with some numerical examples.

The potential perpendicular to the gates is shown in Fig. 2.2(a), obtained by numerical evaluation of (2.7) (solid lines). The corresponding electron density, calculated with (2.5), is shown in Fig. 2.2(b). In weak inversion, i.e. the subthreshold regime, the potential across the UTB essentially tracks ψ_{s}, thus $\psi_{\mathrm{s}} \approx \psi_0$. As shown in Appendix A on p. 143, assuming $\psi_{\mathrm{s}} \approx \psi_0$ and using Taylor's expansion around $\psi_{\mathrm{s}} - \psi_0$, we find the following approximate expression for $\psi(z)$

$$\psi(z) \approx \frac{q n_{\mathrm{i}}}{2\varepsilon_{\mathrm{Si}}} \exp\left(\frac{q\psi_0}{k_{\mathrm{B}}T}\right) z^2 + \psi_0 \tag{2.9}$$

In Fig. 2.2(a) the above parabolic approximation is indicated with the dashed lines .

If we further increase V_{GS}, the inversion charge density is no longer uniformly distributed throughout the UTB and, instead, shows a maximum at the Si/SiO$_2$ interfaces. The inversion charge gradually screens the potential at the center of the UTB from the gate, resulting in a significant voltage drop across either half of the UTB and a reduced coupling of ψ_{s} to ψ_0. Eventually ψ_0 saturates to $\psi_{0,\max}$, which is given by [76]

$$\psi_{0,\max} = \frac{k_{\mathrm{B}}T}{q} \ln\left(\frac{2\pi^2 \varepsilon_{\mathrm{Si}} k_{\mathrm{B}} T}{q^2 n_{\mathrm{i}} t_{\mathrm{Si}}^2}\right) \tag{2.10}$$

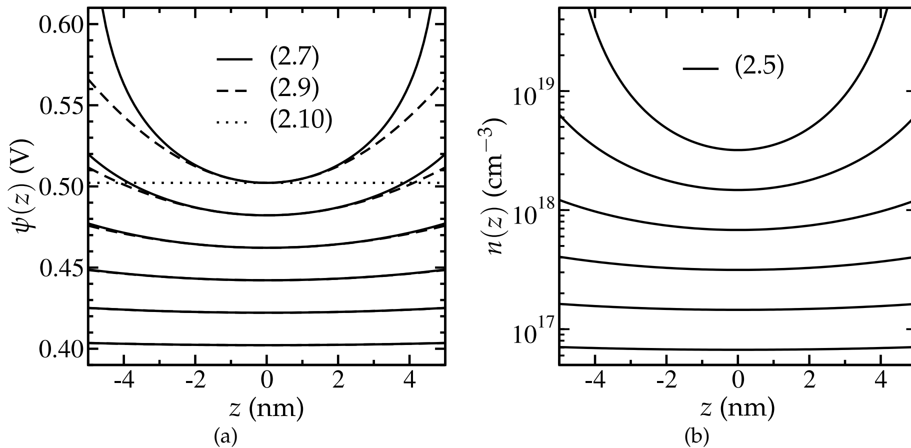


Figure 2.2: (a) $\psi(z)$, obtained by numerical evaluation of (2.7) (solid lines) and its parabolic approximation (dashed lines) calculated with (2.9). The maximum value of the center potential ψ_0 is shown as well (dotted line). (b) the electron density n obtained from (2.5). The considered device is a long-channel symmetric FD DG MOSFET, with t_{Si} = 10 nm and mid-gap gate ($\Delta\phi_{\mathrm{ms}}$ = 0 eV).

In Fig. 2.2(a), the value of $\psi_{0,\max}$ is indicated with the dotted line. We observe that the transition from an essentially uniform volume inversion to 'surface' inversion, corresponding to moving from subthreshold to above threshold, occurs when ψ_{s} is close to $\psi_{0,\max}$.

Fig. 2.3 depicts ψ_{s} and ψ_0 vs. V_{GS}, for several values of the oxide thickness t_{ox}. In subthreshold, corresponding to $V_{\mathrm{GS}} < 0\,\mathrm{V}$ for $\Delta\phi_{\mathrm{ms}} = 0\,\mathrm{eV}$, we observe that $\psi_{\mathrm{s}} \approx \psi_0$. Furthermore, ψ_{s} in the subtreshold regime is independent of t_{ox}, which is a typical feature of a long-channel fully symmetric DG UTB MOSFET [76]. Consistent with our findings in Fig. 2.2, ψ_0 saturates for V_{GS} well above threshold. The surface potential ψ_{s}, instead, still increases with V_{GS}. An approximate expression for the link between V_{GS} and ψ_{s}, assuming $\psi_{\mathrm{s}} \gg \psi_0$, is given by [76]

$$V_{\mathrm{GS}} - \Delta\phi_{\mathrm{ms}} \approx \psi_{\mathrm{s}} + \frac{t_{\mathrm{ox}}\sqrt{2\varepsilon_{\mathrm{Si}} k_{\mathrm{B}} T n_{\mathrm{i}}}}{\varepsilon_{\mathrm{ox}}} \exp\left(\frac{q\psi_{\mathrm{s}}}{k_{\mathrm{B}}T}\right) \tag{2.11}$$

As for ψ_0, we observe in Fig. 2.3 that a value close to $\psi_{0,\max}$ is already achieved just above the threshold, i.e. at the transition from weak (volume) to strong (surface) inversion; ψ_0 reaches its upper bound ($\psi_{0,\max}$) only when ψ_{s} and ψ_0 are fully decoupled, which explains why $\psi_{0,\max}$ in (2.10) is independent of t_{ox}.

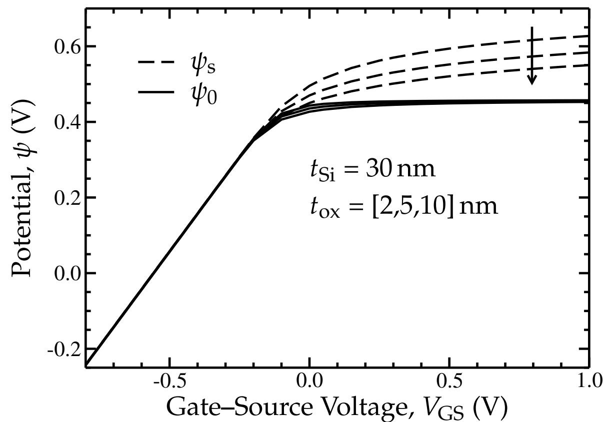


Figure 2.3: The potential just underneath the gates (ψ_s) and in the center of the UTB (ψ_0) versus V_{GS}, for several values of oxide thickness t_{ox}. The considered device is a long-channel symmetric FD DG MOSFET, with t_{Si} = 30 nmand n^+-poly gate. The arrow indicates the direction of increasing t_{ox}.

2.2 Bandstructure and carrier confinement

So far, we have discussed the potential profile and charge distribution, without taking into account quantum confinement. However, if the UTB thickness is reduced to values in the order of the De Broglie wavelength, i.e. around 20 nm and below, the carriers are confined in the direction perpendicular to the gates (z), the *quantization* direction. In other words, the original 3D carrier gas in the bulk semiconductor reduces to a 2D gas [79]. In nano-wires the carriers are confined even in two dimensions, i.e. the two directions normal to the transport direction.

Carrier confinement can alter the device characteristics. As we will see in the course of this section, carrier confinement results in the formation of subbands within both the conduction and valence band. Intuitively one can infer that the emerging ‘energy gaps’ within the conduction and valence band will translate into a reduction of the density of states (DOS) compared to bulk silicon, in which the conduction and valence band consist of a virtually infinite number of quasi-continuous energy levels. Furthermore, the offset of the first available energy level with respect to the original bulk band edge will add to the band gap, thus creating a wider effective band gap. Hence, it is mainly the first subband minimum in the conduction band, or maximum of the valence

band in p-channel devices, that determines the source-channel barrier and the resulting drain current I_{DS}, as we will demonstrate in Chapter 3.

At this stage it is worthwhile to slightly refine the intuitive picture on the impact of quantum confinement sketched above. The required terminology is introduced along the way. In the quantum-mechanical picture, an electron is represented by a wave function $\Psi(z)$, the square modulus of which specifies the *probability density* that an electron resides at position z. The wave function is a solution to the Schrödinger equation (*SEQ*) which, in its time-independent one-dimensional form, reads

$$-\frac{\hbar^2}{2m}\frac{d^2\Psi}{dz^2} + V(z)\Psi(z) = E\Psi(z) \tag{2.12}$$

with $\hbar$ the reduced Planck's constant and m the electron mass. Now, suppose that an electron is subject to the following one-dimensional periodic potential (the *crystal potential*)

$$V(z+a) = V(z) \tag{2.13}$$

where a reflects the periodicity of crystal lattice (the *lattice constant*). Then, Bloch's theorem, i.e.

$$\Psi(z+a) = e^{jk_z a}\Psi(z) \tag{2.14}$$

tells us that we need only to know the wave function within one *unit cell* (or *Brillouin zone*, BZ) in the crystal lattice. Periodicity ensures that the wave function is known anywhere else in the crystal lattice; k_z is the wave number in the z direction.

It is the periodic crystal potential that gives rise to the existence of the valence and conduction band in semiconductors, as pointed out in [80–82]. Near the edge of the conduction band of a bulk semiconductor, carriers move according to the following parabolic dispersion relation:

$$E = \frac{\hbar^2 k_x^2}{2m_x^*} + \frac{\hbar^2 k_y^2}{2m_y^*} + \frac{\hbar^2 k_z^2}{2m_z^*} \tag{2.15}$$

With each direction, we associate an *effective mass* which is inversely proportional to the curvature of the energy dispersion in the valley minimum, as given by

$$m_z^* = \left(\frac{1}{\hbar^2}\frac{d^2E}{dk_z^2}\right)^{-1} \tag{2.16}$$

and similarly in the x and y direction. Hence, when the energy is assumed to be parabolic [cf. (2.15)], the corresponding effective mass is constant. As we

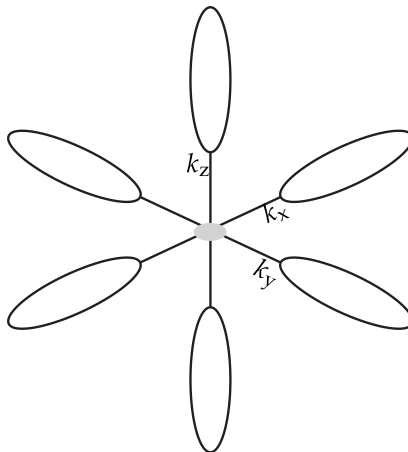


Figure 2.4: Constant energy surfaces along the principal (Δ) axes of the unit cell in silicon. The shaded area shows the projection of the two Δ-valleys along the k_z direction on the (k_x,k_y) plane, in the case of quantization along k_z.

move higher up in energy with respect to the band edge, we have to account for the non-parabolicity of the conduction band energy dispersion [83, 84].

Now, let us consider the conduction band of bulk silicon. If we plot for a given energy E, all combinations of (k_x, k_y, k_z) corresponding to this energy, we find an ellipsoidal constant energy surface along each of the principal (Δ) axes of the unit cell as schematically shown in Fig. 2.4. These so-called Δ-valleys are located close to the edge of the BZ, to be precise at 85% away from the center of the BZ (the Γ point). Due to the symmetry of the crystal lattice, the six Δ-valleys in bulk silicon are fully equivalent, referred to as 6-fold *degenerate*.

Up to now, we have considered the system of valleys in bulk silicon. Instead, confinement of the carriers results in restrictions on the allowed energy levels along the quantization direction, as we will see shortly. To anticipate on the calculations, we will find that in the case of confinement along one of the principal (100) axes, the 6-fold degenerate Δ valley in the bulk silicon conduction band is separated into two sets of valleys, grouped according to their effective mass along the k_z-direction (m_z^*): a 4-fold degenerate valley, labeled $\Delta_{0.19}$ ("*primed*"), and a 2-fold valley, referred to as $\Delta_{0.916}$ or "*unprimed*" valley.

Although we will mainly restrict our analysis to UTB-SOI MOSFETs, carrier confinement is not a unique phenomenon occurring in UTB-SOI devices

only. Actually, similar effects occur in the strongly inverted channel in conventional bulk MOSFETs, where the electric field sets the confining potential. Thus we can identify two types of quantization which differ by the origin of the confining potential: *structural* confinement (also known as *size-induced* or *geometrical* quantization), the key parameter being the UTB thickness t_{utb}, and *electrical* confinement (*bias-induced* quantization), in which the gate electric field governs the strength of quantization.

Of course, purely structural confinement is just a limiting case, typically valid under low (subthreshold) and moderate inversion conditions in SOI devices with an ultrathin body. In strong inversion a hybrid form of quantization may appear, stemming from a combination of both electrical and structural confinement. Since this work deals with UTB-SOI devices, the main focus is on structural confinement. However, the impact of structural and electrical confinement on the band structure often results in similar qualitative trends. Hence, the findings presented in the course of this chapter are expected to apply also in the case of quantization in strongly inverted bulk MOSFETs.

This introductory section deals with quantum confinement from an engineers' perspective, so as to explain the impact of quantization on the device characteristics. We do not intend to discuss the origin and underlying physics in great detail. For the latter, the reader is referred to the extensive literature on this topic, such as [81, 82, 85, 86].

The confining potential is schematically represented by a simplified potential well, as shown in Fig. 2.5. The square well [Fig. 2.5(a)] represents the (conduction) band discontinuity which occurs in a thin semiconductor layer, 'sandwiched' between two oxide layers. The strength of quantization is determined by the width of the potential well, t_{utb} (structural confinement). Likewise, the triangular potential depicted in Fig. 2.5(b) reflects the approximate shape of the confining potential close to the semiconductor-oxide interface in strongly inverted channels, referred to as electrical confinement.

Furthermore, although we only discuss quantization in the conduction band, similar effects occur in the valence band [87]. The latter is, however, strongly anisotropic and non-parabolic, making the presentation of a simple and intuitive picture less trivial. Quantization in the valence band is discussed in [88]. Analytic descriptions of the valence band within the context of the EMA have been reported [89], but its discussion is beyond the scope of this chapter.

Before proceeding to the calculation of the energy bands, we note that the impact of quantum confinement on the charge distribution and poten-

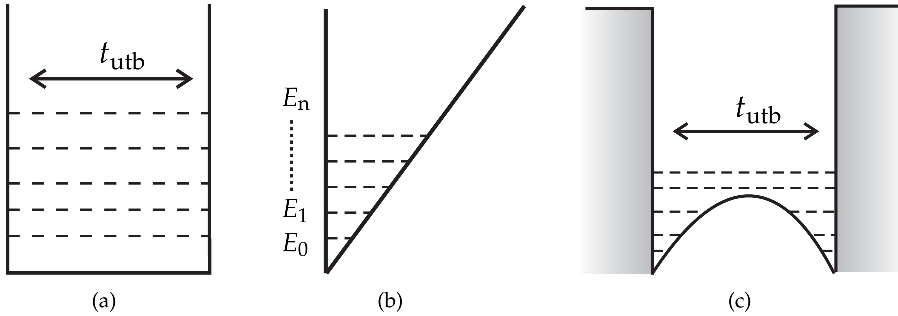


Figure 2.5: Schematic representation of the potential well corresponding to structural (a), electrical (b) and a combination of structural and electrical confinement (c). In (b) the confining potential is approximated by a triangular well. The energy subband separation is not drawn to scale.

tial profile is to be calculated by solving the *coupled* Schrödinger and Poisson equations until a self-consistent solution in terms of charge and potential is achieved. However, this numerical procedure can be prohibitive for application in e.g. circuit simulations. Analytical approaches to include quantum confinement always involve assumptions on either the charge distribution or the potential profile, so as to decouple the Schrödinger and Poisson equation and to arrive at a closed form solution.

In the following, the potential will be treated as 'frozen', represented by the idealized profiles such as depicted in Fig. 2.5. Possible solutions to the eigenvalue problem in case of the triangular potential are reported in [86]. The wave functions can be expressed in terms of special (Airy) functions, for which analytical approximations exist [90].

As for the square potential well, two cases will be considered: first the solution to (2.12) assuming an infinitely high potential barrier (also known as *closed* or *hard wall* boundary condition). Secondly, we will consider a finite confining potential, which resembles the semiconductor/insulator band discontinuity in actual DG SOI MOSFETs. In either case, the square potential is generally assumed to apply for low to modest inversion conditions, particularly in DG SOI devices consisting of a very thin body. The limitations of decoupling the Schrödinger and Poisson equations are explored using first order perturbation calculations in [91].

Analytical attempts to model the combined effect of electrical and structural confinement are reported in [92], in which a 'unified trial wave function'

is proposed. The coefficients of the trial wave function are obtained by embracing the variational approach, which is exemplified in Appendix B.

Infinite Si/SiO_2 barrier

We are interested in the possible energy levels within the infinite potential well, as shown in Fig. 2.5(a). To this end, we note that as the confining potential goes to infinity, the wave function should be zero exactly at the well boundary ($z = 0$ and $z = t_{Si}$), since no carrier is allowed to enter the surrounding barrier. Furthermore, the potential within the well, i.e. $0 < z < t_{Si}$ is zero. Then, the time-independent Schrödinger equation reads [82]

$$\frac{d^2\Psi}{dz^2} + k_z^2\Psi(z) = 0 \qquad \text{with} \qquad k_z = \sqrt{\frac{2m_z^* E}{\hbar}} \tag{2.17}$$

where E is the eigen value. The general solution to (2.17) reads

$$\Psi(z) = A\sin(k_z z) + B\cos(k_z z) \tag{2.18}$$

Using the boundary condition $\Psi(0) = \Psi(t_{Si}) = 0$, we find that $B = 0$ and

$$\Psi(z) = A\sin(k_z z) \qquad \text{with} \qquad k_z = \frac{n\pi}{t_{Si}} \tag{2.19}$$

with n taking integer values from 1 to infinity. Using the $E(k_z)$ relation as given in (2.17), we find the following allowed energy values, which we label E_n

$$E_n = \frac{\hbar^2}{2m_z^*}\left(\frac{n\pi}{t_{Si}}\right)^2 \tag{2.20}$$

E_n are the *eigenvalues* of (2.17), and each value represents the energy minimum of the energy *subband* with index n.

For the sake of convenience, we now set the origin of the z-axis at the center of the UTB, so that z runs from $\frac{-t_{Si}}{2}$ to $\frac{t_{Si}}{2}$:

$$\Psi(z) = A\sin\left[\frac{n\pi}{t_{Si}}\left(z + \frac{t_{Si}}{2}\right)\right] \tag{2.21}$$

The constant A can be found by noting that the probability that a carrier resides at *any* position along z is unity, i.e.

$$\int_{-\frac{t_{Si}}{2}}^{\frac{t_{Si}}{2}} |\Psi(z)|^2\, dz = 1 \tag{2.22}$$

Since the potential outside the well is infinite, the wave function does not extend into the barrier. Thus, the resulting wave function reads

$$\Psi(z) = \sqrt{\frac{2}{t_{\mathrm{Si}}}} \sin\left[\frac{n\pi}{t_{\mathrm{Si}}}\left(z + \frac{t_{\mathrm{Si}}}{2}\right)\right] \tag{2.23}$$

In summary, we find that the carrier's motion is restricted in the z direction, with the allowed energies E_{n} given by (2.20). In the plane $(k_{\mathrm{x}}, k_{\mathrm{y}})$ perpendicular to the quantization direction, however, the $E(k)$ dispersion is that of a free particle. Thus, we find for the total energy

$$E_{\mathrm{tot}}(\mathbf{k}) = \frac{\hbar^2}{2m_{\mathrm{z}}^*}\left(\frac{n\pi}{t_{\mathrm{Si}}}\right)^2 + \frac{\hbar^2 k_{\mathrm{x}}^2}{2m_{\mathrm{x}}^*} + \frac{\hbar^2 k_{\mathrm{y}}^2}{2m_{\mathrm{y}}^*} \tag{2.24}$$

with $\mathbf{k} = (k_{\mathrm{x}}, k_{\mathrm{y}})$.

Finite confining potential

Previously, we assumed that the wave function did not extend beyond the quantum well (*closed boundary* conditions), resulting in an analytical closed form solution for the eigenvalues, or subband minima E_{n}, and the wave functions. From a physical point of view, having an infinitely high confining potential means that no charge can tunnel through the semiconductor/insulator barrier in either direction. In effect, the integral of the square modulus of the wave function $|\Psi(z)|^2$ in the silicon film is unity, since the total probability that a carrier is located at any position along z equals one.

In practice, however, the band discontinuity of the Si/SiO_2 barrier, or any semiconductor/insulator interface, has a finite value. Considering the conduction band, the Si/SiO_2 barrier is approximately 3 eV [12]. Although carriers are very unlikely to occupy energies this high, the impact of a finite confining potential (i.e. barrier height) may significantly influence the location and separation of the energy levels, even those located much lower in energy, which primarily determine the transport properties of the material.

Now, following [93], we will derive an expression for the subband minima assuming a finite confining potential. Within the potential well, i.e. $|z| < \frac{t_{\mathrm{Si}}}{2}$, the Schrödinger equation is very similar to (2.17)

$$\frac{d^2\Psi}{dz^2} + k_{\mathrm{z}}^2\Psi(z) = 0 \qquad \text{with} \qquad k_{\mathrm{z}}^2 = \frac{2m_{\mathrm{z}}^*\left(U_{\mathrm{B}} + E\right)}{\hbar^2} \tag{2.25}$$

with U_{B} the magnitude of the barrier. We take the zero of the energy axis at the top of the energy barrier, hence the energy E within the potential well

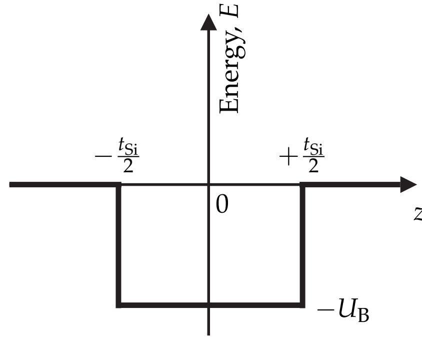


Figure 2.6: Sketch of the potential well, with finite barrier height U_{B}.

is negative and the energy at the bottom of the well is $-U_{\text{B}}$, as sketched in Fig. 2.6. Due to the finite confining potential, the wave function will extend into the barrier and will gradually decay in an exponential fashion. Outside the well the potential is zero. For $|z| \geq \frac{t_{\text{Si}}}{2}$ the Schrödinger equation reads

$$\frac{d^2\Psi}{dz^2} - \gamma^2\Psi(z) = 0 \qquad \text{with} \qquad \gamma^2 = \frac{2m_z^*|E|}{\hbar^2} \tag{2.26}$$

in which γ can be viewed as the damping term of the exponentially decaying wave function in the barrier

$$\Psi = C\exp\left(-\gamma|z|\right) \qquad \text{for} \qquad |z| \geq \frac{t_{\text{Si}}}{2}. \tag{2.27}$$

Similar to the infinite barrier case, we expect the possible solutions to (2.25) to be either odd or even symmetry functions, such as sines and cosines. In addition, at the well boundaries ($z = \pm\frac{t_{\text{Si}}}{2}$), the wave function should be continuous both in itself and in its first derivative. Equivalently, one can require that the 'logarithmic derivative' Ψ'/Ψ of the wave function be continuous [93].

Starting with the even-symmetry (cosine) functions, we find the logarithmic derivative to be

$$\frac{-k_z\sin(k_z z)}{\cos(k_z z)} = -k_z\tan(k_z z) \tag{2.28}$$

Using (2.27) and (2.28) we find at $z = \pm\frac{t_{\text{Si}}}{2}$

$$k_z\tan\left(\frac{k_z t_{\text{Si}}}{2}\right) = \gamma \tag{2.29}$$

in which the right-hand side, simply being γ, is the logarithmic derivative of (2.27) at either side of the well. This results in the following algebraic transcendental expression for the even-symmetry energy minima

$$\tan\left(\frac{k_z t_{Si}}{2}\right) = \frac{\gamma}{k_z} = \sqrt{\frac{U_B}{U_B + E} - 1} \tag{2.30}$$

with k_z and γ as defined in (2.25) and (2.26) respectively. Similarly, the odd-symmetry minima are

$$\text{cotan}\left(\frac{k_z t_{Si}}{2}\right) = -\sqrt{\frac{U_B}{U_B + E} - 1} \tag{2.31}$$

We find the coefficient of the corresponding wave function by equating at $z = \frac{t_{Si}}{2}$ the wave function within the well, i.e. (2.21), to the exponentially decaying part, given in (2.27). Furthermore, recall that the probability that a carrier resides at any position along z must be unity. Hence, in the case of a finite confining potential, the integration boundaries in (2.22) are to be changed accordingly into $\pm\infty$.

Now, let us turn to some numerical examples of the shift in energy levels due to confinement of the carriers in an UTB. To this end, Fig. 2.7 shows, for varying t_{Si}, the lowest subband minima for the $\Delta_{0.916}$ and $\Delta_{0.19}$ valleys in Si with (100) quantization direction. The plot shows the values calculated with the infinite square well, (2.20), and with a finite barrier, obtained by numerical calculation of (2.30) and (2.31). In the latter, we set $U_B = 3\,\text{eV}$, which resembles the Si/SiO_2 conduction band discontinuity. From (2.20) we can tell that the energetic position of the subband minima is inversely proportional to the quantization mass. In fact, in silicon (100), the absolute minimum is determined by the $\Delta_{0.916}$ valley. The lowest $\Delta_{0.19}$ valley is located much higher in energy. Generally, the subband minima increase with decreasing t_{Si}. Secondly, due to its smaller quantization mass, the minima of the $\Delta_{0.19}$ valley are located higher in energy compared to those of the $\Delta_{0.916}$ valley.

Furthermore, accounting for a finite barrier yields subband minima which are systematically lower than their infinite barrier counterparts. The impact of a finite barrier height is illustrated more quantitatively in Fig. 2.8. The subbands close to the band edge are less affected by the barrier height compared to the ones higher up in energy, which are closer to the top of the potential well. Hence, particularly the valleys with the smallest quantization mass show a relatively strong U_B dependence.

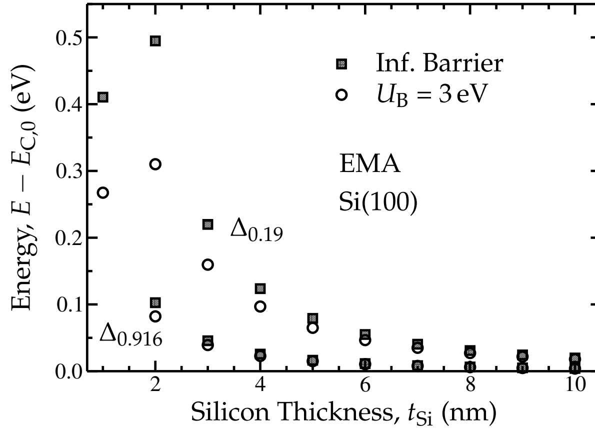


Figure 2.7: The first subband minimum of the $\Delta_{0.916}$ and $\Delta_{0.19}$ valleys in Si with (100) quantization direction. The minima calculated assuming an infinite square well [filled symbols, Eq. (2.20)] are located higher in energy compared to those of the finite square well [Eqs. (2.30) and (2.31)], particularly for the smallest t_{Si} and smallest quantization mass m_z^*.

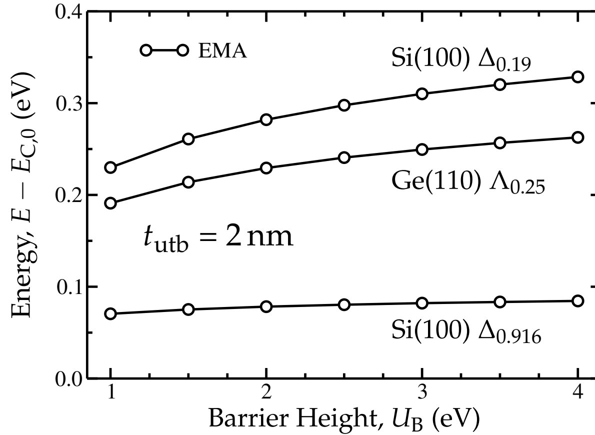


Figure 2.8: Lowest subband minimum as a function of the confining potential barrier height U_B, in the extreme case of t_{utb} = 2 nm. The values are shown for the $\Delta_{0.19}$ and $\Delta_{0.916}$ valleys in the conduction band of silicon with (100) quantization direction, and the lowest minimum in the germanium (110) $\Lambda_{0.25}$ valley. The Si(100) $\Delta_{0.916}$ valley, which is the absolute minimum, shows only a weak dependence on U_B.

	Quant. Dir.	Valley	g_ν	ΔE_0 (eV)	m_t^*	m_l^*	m_z^*
Si	(001)	$\Delta_{0.916}$	2		0.190	0.190	0.916
		$\Delta_{0.19}$	4		0.190	0.916	0.190
	(110)	$\Delta_{0.315}$	4		0.190	0.553	0.315
		$\Delta_{0.19}$	2		0.190	0.916	0.190
	(111)	$\Delta_{0.268}$	6		0.190	0.674	0.268
Ge	(110)	$\Lambda_{0.25}$	2	0	0.093	0.595	0.250
		$\Lambda_{0.093}$	2	0	0.093	1.600	0.093
		$\Delta_{0.318}$	4	0.189	0.194	0.541	0.318
		$\Delta_{0.194}$	2	0.189	0.194	0.888	0.194
		$\Gamma_{0.05}$	1	0.145	0.050	0.050	0.050

Table 2.1: Parameters of the EMA model for different materials and quantization directions; g_ν is the degeneracy of valley v, m_z^* is the quantization mass, m_l^* and m_t^* are the longitudinal and transverse mass of the elliptic energy dispersion around the minimum (in units of m_0, the electron rest mass). ΔE_0 denotes the energy separation between the Δ and Λ valleys in germanium. The effective masses m_z^*, m_l^* and m_t^* have been obtained as explained in [85] from the longitudinal and the transverse masses of the bulk crystal energy dispersion, using the values: $0.916m_0$ and $0.19m_0$ for the Δ valleys of bulk silicon; $1.6m_0$ and $0.093m_0$ for the Λ valleys; $0.888m_0$ and $0.194m_0$ for the Δ valleys; $0.05m_0$ for the Γ valley of bulk germanium.

So far, only the Δ valleys in silicon with (100) quantization direction were discussed. Similarly, valleys can exist along other axes of symmetry, such as the Λ lines. However, the very high minimum energy of the Λ valley in silicon renders them irrelevant in practice. In germanium, instead, the Λ valleys are located lower in energy than the Δ valleys. Hence, in germanium, the transport properties are predominantly determined by the Λ valleys. An overview of the different valleys in Si and Ge, for various crystal orientations, is presented in Table 2.1.

Impact of quantization on the carrier distribution

Once the energy minima and wave functions are known, we can calculate the inversion charge distribution along the direction perpendicular to the gates. To calculate the inversion charge density at a given position in the channel, we have to include the following 'ingredients' in the model: first of all, the number of states associated with a given energy, called the *density of states* (DOS). Secondly, we need to know whether a certain state is occupied, which is directly linked to the Fermi level E_F and depends on the temperature and the applied gate bias.

Previously we have seen that the probability that an electron is located at

a given position along the quantization direction z can be calculated from the modulus of the wave function $\Psi_{v,n}$ of subband n in valley v. Thus, we obtain for the electron concentration distribution $\rho_\mathrm{e}(z)$ [85, 91]

$$\rho_\mathrm{e}(z) = \sum_{\nu,n} |\Psi_{\nu,n}(z)|^2 D_\nu \ln\left[1 + \exp\left(\frac{E_\mathrm{F} - E_{\nu,n} - E_{\mathrm{C},0}}{k_\mathrm{B} T}\right)\right] \tag{2.32}$$

with D_ν the DOS within one subband of valley ν, given by

$$D_\nu = \frac{g_\nu m^*_{d,\nu} k_\mathrm{B} T}{\pi \hbar^2} \tag{2.33}$$

assuming a parabolic energy dispersion. For each valley ν, $E_{\nu,n}$ is the energy minimum of subband n relative to the bulk (conduction) band edge $E_{\mathrm{C},0}$ and g_ν is the valley degeneracy; $m^*_{d,\nu}$ is the *density-of-states* effective mass, given by $m^*_{d,\nu} = m^*_\mathrm{t}$ for the $\Delta_{0.916}$ valley and $m^*_{d,\nu} = \sqrt{m^*_\mathrm{t} m^*_\mathrm{l}}$ for the $\Delta_{0.19}$ valley in silicon with (100) quantization direction [85] (see also Tab. 2.1).

In order to count all states, the sum in (2.32) runs over all valleys ν and subbands n. To obtain the carrier concentration, we have to integrate along z. In case of an infinite square well, integrating $|\Psi_{\nu,n}(z)|^2$ simply yields unity and the resulting quantum-mechanical electron concentration n_qm is

$$n_\mathrm{qm} = \int_{-t_\mathrm{Si}/2}^{+t_\mathrm{Si}/2} \rho_\mathrm{e}(z)\, dz = \sum_{\nu,n} \frac{m^*_{d,\nu} k_\mathrm{B} T}{\pi \hbar^2} \exp\left(\frac{E_\mathrm{F} - E_{\nu,n} - E_{\mathrm{C},0}}{k_\mathrm{B} T}\right) \tag{2.34}$$

Similarly, for a finite confining potential the above integral can be evaluated numerically.

Fig. 2.9 shows the charge distribution along the z direction, calculated with (2.32) using an infinite square potential well. For reference, the classical distribution (CL) is shown as well, and hereafter referred to as n_cl. In fact, as predicted by (2.7) and (2.5), the CL charge distribution is essentially uniform throughout the UTB. Note that a uniform charge corresponds to a parabolic potential profile, as seen previously in Fig. 2.2. The quantum-mechanical (QM) carrier distribution, instead, peaks at the center of the channel and gradually drops as we move closer to the gates. This observation is explained by the constructive addition of the wave functions. The overall shape is determined mainly by the ground state, i.e. the wave function corresponding to the absolute minimum, which has even symmetry around the center of the UTB. For thicker films, the distribution exhibits a plateau at the center, originating from the inclusion of higher order wave functions which move to lower energies for increasing t_Si.

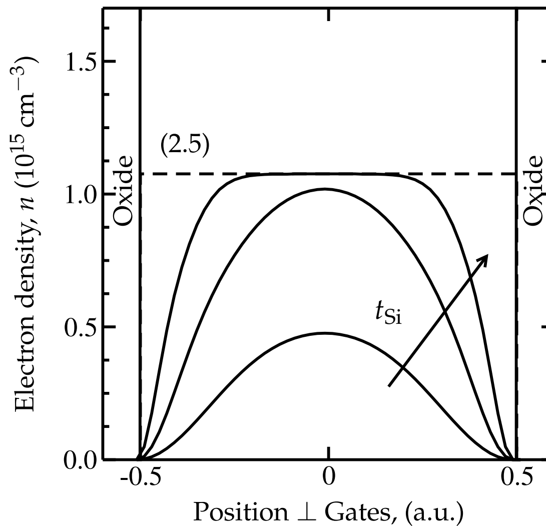


Figure 2.9: The quantum-mechanical inversion charge distribution under low-inversion conditions, such as the subthreshold regime. In low to moderate inversion conditions, the classical distribution is essentially uniform [recall Fig. 2.2(a)]. The z-axis is normalized on t_{Si}, which takes the values [5,10,20] nm.

Likewise, we compare the QM and CL charge distribution in moderate to strong inversion conditions, as exemplified in Fig. 2.10. Rather than using the algebraic equations presented previously, the curves have been obtained from device simulations [94] employing the Density Gradient model [95]. This model accounts for the effects of quantization by applying a quantum correction to the classical carrier distribution. We observe that the Density Gradient simulations support the analytical results from Fig. 2.9 in that the QM charge distribution exhibits a maximum at the center of the channel in the low inversion regime. If we move to strong inversion conditions, gradually the CL charge concentration in the center saturates, and a peak emerges at either interface. Thus, whereas in the low-inversion regime conduction occurs throughout the entire UTB, in strong inversion two conductive paths emerge just underneath the gates, in agreement with our previous findings in Section 2.1. The QM charge distribution shows similar features, such as saturation of the carrier concentration in the center of the channel. Again, in strong inversion, the peak moves from the center towards the interface but, differently from the CL distribution, they appear slightly away from the interface.

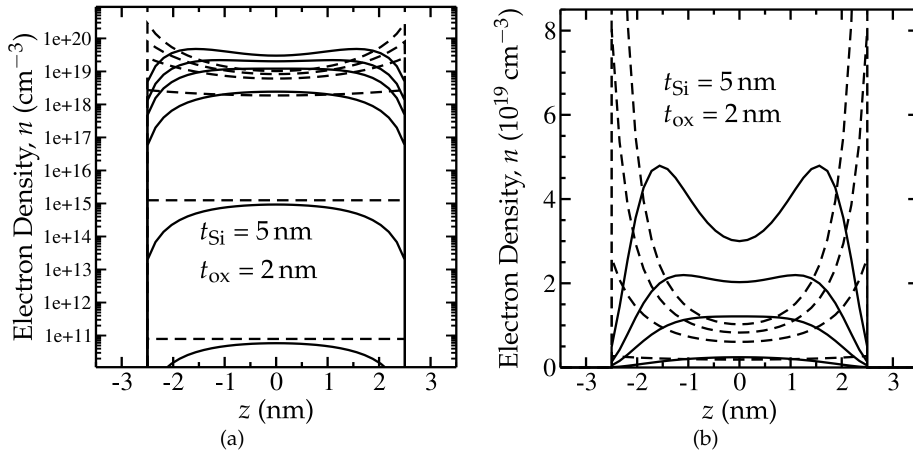


Figure 2.10: Simulated carrier distribution in a long-channel DG-SOI n-MOSFET, calculated with the Density Gradient model [95] to account for quantization. The figures show snapshots at low, moderate and high inversion densities, on (a) log and (b) linear scale. The carrier density increases with V_{GS}, which is [-1, -0.5, -0.25. . . 0.5, 1.0] V ($\Delta\phi_{ms}$ = -0.56 eV).

We conclude this section with a remark on the EMA. The EMA relies on the approximation that the effect of e.g. the periodic crystal potential can be incorporated in one single quantity, i.e. the effective mass. This greatly simplifies the Schrödinger equation, but it implicitly assumes that the applied and/or built-in potential (through doping or material composition) vary slowly compared to the crystal potential. One may, however, question this assumption when the thickness of the silicon film (thus the potential well) is in the order of just a few times the lattice constant. This means that the potential may vary on the scale of the crystal potential, causing the EMA to break down. Hence, often one has to embrace more rigorous approaches to calculate the band structure and corresponding transport properties.

2.3 Validity of the EMA

Previously, we have introduced the EMA quantization model, which results in an analytical closed-form or transcendental equation for the subband minima $E_{\nu,n}$, depending on whether we assume an infinite or a finite square confining potential along the quantization direction. A (non-)parabolic energy disper-

sion in the transport plane, i.e., as a function of the 2D wavevector **k**, is added to the subband minimum [84, 96–98]. Thus, in the EMA model the quantization produces a shift in the subband minima, but it does not affect the energy dispersion inside each subband.

If the thickness of the semiconductor body is in the order of a few atomic layers, the strong quantum mechanical confinement can yield significant deviations from the simple EMA results, mainly in terms of: (a) different values for the minima of the 2D subbands; (b) distortion of the energy dispersion in the transport plane, with a resulting change in the transport masses and in the 2D effective density of states.

In the remaining part of this chapter we will explore the validity of the EMA, focused on quantization in the conduction band of silicon and germanium UTB-SOI MOSFETs with different crystal orientations. The text embodied in the following is an adapted version of our previously published work [99].

In order examine the validity of the EMA, we will show a systematic comparison with the bandstructure calculated with the so-called Linear Combination of Bulk Bands (LCBB) quantization model [100–102]. More precisely, we start by discussing the differences in the minima of the 2D subbands obtained with either the EMA or the LCBB model. Then we extract the transport masses from the LCBB bandstructure for different UTB values and compare the numerically calculated 2D density of states of the LCBB method with the analytical expressions of the EMA model. In the following, we will label the UTB thickness t_{utb}, and explicitly mention t_{Si} or t_{Ge} when the results concern either silicon or germanium respectively.

LCBB bandstructure calculation

Starting point for the LCCB approach is the more general form of the Schrödinger equation [such as (2.35)], which consists of the periodic crystal potential and the 'superimposed' confining potential. This implies that, in absence of the confining potential, the eigenvalue problem has solutions which represent the energy dispersion of the 3D (bulk) band structure.

Thus, the LCBB approach not only calculates the effect of quantization, but takes into account also the bulk band structure itself, hence it belongs to the group of *Full-Band* (FB) quantization models. The EMA, instead, just considers the impact of the confining potential. Another example of a FB model is the Tight-Binding model [103–105] which, differently from the LCCB model,

does not expand the eigenfunction in terms of bulk Bloch functions, but uses instead a basis set of localized atomic orbitals [104]. In any case, FB models, being rigorous and generic approaches, serve as benchmarks to validate analytical and approximate quantization models such as the EMA.

If we let z be the quantization direction and $U(z)$ the confining potential energy, then for each $\mathbf{k}$ vector in the transport plane, the allowed energies $E_{\mathrm{n}}(\mathbf{k})$ are calculated by solving the eigenvalue problem [102]:

$$E_{FB}^{(b)}(\mathbf{k},k_{\mathrm{z}})A_{n}^{(b)}(\mathbf{k},k_{\mathrm{z}})+\frac{2\pi}{L_{\mathrm{z}}}\sum_{b',k'_{\mathrm{z}}}\Big\{U_{T}(k'_{\mathrm{z}}-k_{\mathrm{z}})f_{k_{\mathrm{z}},k'_{\mathrm{z}}}^{(b,b')}(\mathbf{k},0)+ \sum_{G_{\mathrm{z}}}U_{T}(k'_{\mathrm{z}}-k_{\mathrm{z}}+G_{\mathrm{z}})f_{k_{\mathrm{z}},k'_{\mathrm{z}}}^{(b,b')}(\mathbf{k},G_{\mathrm{z}})\Big\}A_{n}^{(b')}(\mathbf{k},k'_{\mathrm{z}})=E_{\mathrm{n}}(\mathbf{k})A_{n}^{(b)}(\mathbf{k},k_{\mathrm{z}}) \quad (2.35)$$

where n is the index of the eigenvalue (i.e. the subband index), $A_{n}^{(b)}(\mathbf{k},k_{\mathrm{z}})$ denote the coefficients of the *unknown* eigenfunction, $U_{T}(q_{z})$ is the Fourier transform of $U(z)$, $(2\pi/L_{\mathrm{z}})$ is the spacing used for the discretization of k_{z} and L_{z} is a normalization length in the quantization direction; G_{z} is the magnitude of a reciprocal lattice vector in the k_{z} direction. Furthermore, $E_{FB}^{(b)}(\mathbf{k},k_{\mathrm{z}})$ is the energy in the *b-th* band of the bulk crystal conduction band and $f_{k_{\mathrm{z}},k'_{\mathrm{z}}}^{(b,b')}(\mathbf{k},G_{\mathrm{z}})$ denotes an appropriate *overlap integral* of the periodic parts $u_{b,\mathbf{k},k_{z}}$ of the Bloch functions [102, 106]. The index b' runs over the number b_{FB} of bands of the bulk crystal included in the calculations; in all the calculations we have used the two lowest bands of the bulk crystal conduction band.

From (2.35) we see that in the LCBB method we have to calculate the Fourier transform of the confining potential energy $U(z)$, so that $U(z)$ has to feature a finite potential energy barrier U_{B} at the semiconductor-oxide interface. Thus, in absence of the confining potential ($U_{T}=0$), (2.35) just yields the energy dispersion of the (3D) bulk band structure. Furthermore, (2.35) also clarifies that the full-band energy $E_{FB}^{(b)}(\mathbf{k},k_{\mathrm{z}})$ of the constituent semiconductor is an input of the LCBB method. We have used the well-established Non-Local-Pseudopotential (NLP, [107]) method to determine both the FB dispersion $E_{FB}^{(b)}(\mathbf{k},k_{\mathrm{z}})$ and the overlap factors $f_{k_{\mathrm{z}},k'_{\mathrm{z}}}^{(b,b')}(\mathbf{k},G_{\mathrm{z}})$ that enter (2.35). The parameters for the NLP procedure were taken from [107] for silicon and from [108] for germanium.

All the results shown in the following have been obtained by solving directly (2.35) with no further approximations. Throughout this section we express the wave vectors in units of $(2\pi/a_{0})$, where the lattice constant a_{0} is 0.543 nm and 0.565 nm for silicon and germanium, respectively. As explained

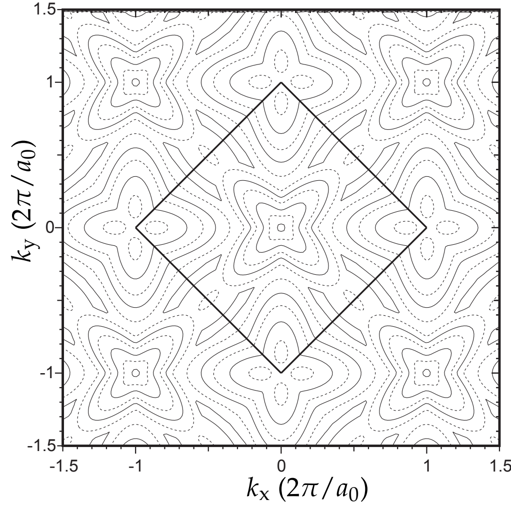


Figure 2.11: Si(100), square well with t_{utb}=3 nm. Contour plot for the lowest eigenvalue $E_0(\mathbf{k})$ versus the wavevector **k** obtained by solving (2.35). The axes are in unit of $2\pi/a_0$. The square indicates the first BZ of the two-dimensional electron gas. The lowest valley is the $\Delta_{0.916}$ valley located at the point $\mathbf{k}=(0,0)$, whereas the $\Delta_{0.19}$ valleys are at the points $\mathbf{k}=(\pm 0.85, 0)$ and $\mathbf{k}=(0, \pm 0.85)$ (see Tab. 2.1 for the labels of the valleys).

in [102], the k_z values included in (2.35) must vary in a periodicity interval of the reciprocal lattice space along the k_z direction; namely in an interval of length 2, $2\sqrt{2}$ and $\sqrt{3}$ for the (100), (110) and (111) quantization direction, respectively. The 2D bandstructure is calculated with the LCBB method by varying the wavevector **k** in (2.35), where **k** plays the role of a parameter.

Fig. 2.11 illustrates the lowest subband versus the two-dimensional **k** vector for the inversion layer of a Si(100) UTB-SOI MOSFET. The $\Delta_{0.916}$ valley is observed at the point $\mathbf{k}=(0,0)$, whereas the $\Delta_{0.19}$ valleys are at the points $\mathbf{k}=(\pm 0.85, 0)$ and $\mathbf{k}=(0, \pm 0.85)$ (see Tab. 2.1 for the labels of the valleys).

As it can be seen in Fig. 2.11, in general the valleys of the 2D electron gas are not located at the point $\mathbf{k}=(0,0)$; hence, when we analyze the energy dispersion of a 2D valley with the LCBB method, we must solve (2.35) along the lines in the **k** plane that run across the minimum of the valley. In the EMA, the value of **k** in the plots corresponding to a given valley is always defined as the displacement with respect to the **k** point corresponding to the minimum of the valley, i.e. $\mathbf{k}=(0,0)$ for the $\Delta_{0.916}$ valley or $\mathbf{k}=(\pm 0.85, 0)$ and $\mathbf{k}=(0, \pm 0.85)$ for the $\Delta_{0.19}$ valleys.

Comparison of LCBB and EMA results

Both the EMA and the LCBB methods can be used for any confining potential energy $U(z)$, however we decided to compare the two models by using a very simple square well with a finite barrier U_B, such as depicted in Fig. 2.5(a). The width t_{utb} of the well is thus the parameter that governs the *strength* of the quantum mechanical confinement (hence, *structural* confinement). In the following the energy values reported in the graphs are referred to the bottom of the square well, i.e. the bulk conduction band edge $E_{C,0}$.

In the EMA approach, we assume a parabolic energy dispersion in the quantization direction with a quantization mass m_z^*, so that the minima of the 2D subbands are obtained by the well-known Schrödinger-like equation in the real space (2.17) [84, 97, 109, 110]. The eigenvalues of a square well are known in an analytical form for an infinite barrier as given by (2.20), whereas for a finite barrier U_B they can be obtained by solving the algebraic, transcendental equation (2.30) [93]. Unless stated otherwise (as in Fig. 2.16), all the results shown hereafter have been obtained with a barrier U_B=3 eV (representative of the Si/SiO_2 conduction band discontinuity) for both the EMA and the LCBB calculations.

The quantization mass (m_z^*), the transverse (m_t^*) and the longitudinal mass (m_l^*) employed in the EMA model are reported in Tab. 2.1, and they have been obtained from the values of the longitudinal and transverse masses of the bulk crystals [85]. The effective masses for bulk silicon and bulk germanium (reported in the caption of Tab. 2.1) have been directly extracted from the NLP calculations used as a part of the LCBB method; the values of the masses are consistent with [107, 108].

We will mainly focus on Si(100) and Ge(110) DG-SOI *n*-MOSFETs. As for the germanium transistors, the (110) wafer orientation is the most promising among the principal orientations in terms of the maximum drive current [111–113].

Minima of the 2D subbands

Fig. 2.12 reports the energy dispersion for the Si(100) around the $\Delta_{0.916}$ valley and for two values of t_{Si}. As it can be seen the EMA approximation tracks fairly well the lowest energy branches of the LCBB bandstructure even for the thinnest semiconductor film. The error in the minima of the 2D subbands increases for the higher subbands.

It is interesting to notice that the LCBB results exhibit a splitting between

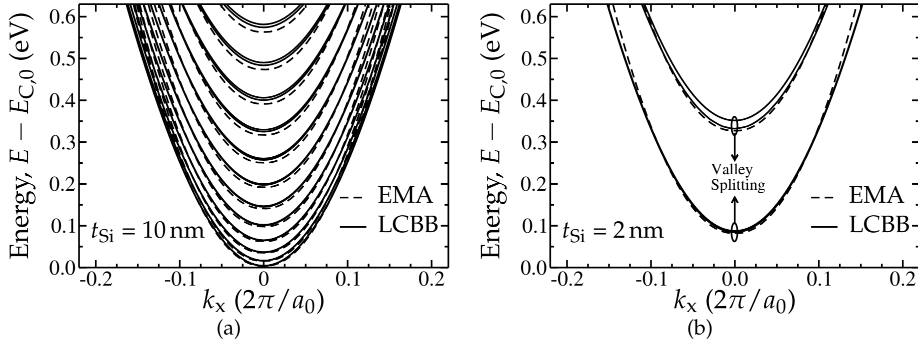


Figure 2.12: Energy dispersion for the $\Delta_{0.916}$ valley in a Si(100) inversion layer calculated with either the EMA or the LCBB model. (a) t_{Si} = 10 nm; (b) t_{Si} = 2 nm. The minimum is located at $\mathbf{k}$=0 and k_x moves along the (010) crystal direction. Since $\mathbf{k}$=0 is a symmetry point, exactly the same result is obtained along the (001) direction. The LCBB calculations exhibit the *valley splitting* between the doublets of subbands [109].

the two (supposedly degenerate) lowest subbands, which is known as *valley splitting* as discussed in [109], and it is more pronounced in the thinner semiconductor film. The EMA model inherently assumes a perfect degeneracy for the two lowest branches of the $\Delta_{0.916}$ valleys, hence the valley splitting is a feature that the EMA model does not account for [109].

In this regard Fig. 2.13 reports the values of the valley splitting versus the semiconductor thickness for the Si(100) $\Delta_{0.916}$ valley. The splitting is larger for the second lowest than it is for the lowest doublet of subbands. The results for the lowest subband are in agreement with the values recently obtained with the tight-binding method [105]. The splitting for a given t_{Si} is smaller than it is in a Si(100) nanowire transistor having the diameter equal to t_{Si} [114]; not surprisingly the two-dimensional quantum confinement produced in a nanowire device emphasizes the valley splitting with respect to one-dimensional confinement in a conventional MOS transistor.

Fig. 2.13 shows that the valley splitting of the lowest subband, i.e. the one that is most relevant for the carrier transport properties, is always small compared to the thermal energy at room temperature $k_B T \approx 26$ meV; hence, from a practical viewpoint, it can be neglected for the analysis of the electron devices, unless very low temperatures (around 100 K and below) are considered. Fig. 2.14 reports the minima for the lowest subband versus t_{Si} for the $\Delta_{0.916}$ and the $\Delta_{0.19}$ valleys of the Si(100). As it can be seen, the EMA approach tracks

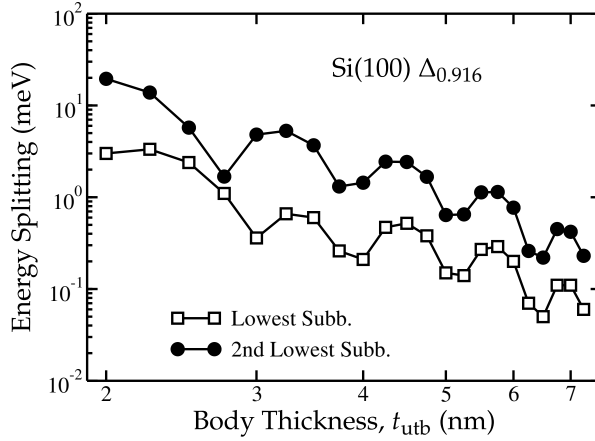


Figure 2.13: Energy splitting at the minimum of the 2D subbands versus the semiconductor thickness calculated with the LCBB method for the Si(100) $\Delta_{0.916}$ valleys. The results for the lowest and the second lowest doublets of subbands are illustrated. The energy splitting of the lowest subband is small compared to the thermal energy at room temperature $k_B T \approx 26\,\text{meV}$.

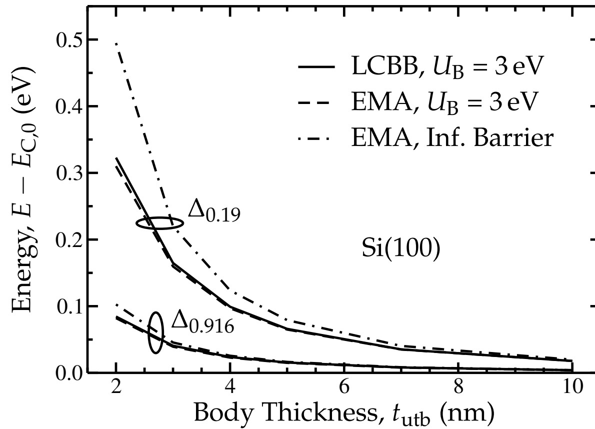


Figure 2.14: Si(100). Lowest eigenvalue versus the semiconductor thickness for the $\Delta_{0.916}$ and $\Delta_{0.19}$ valleys calculated with either the LCBB or the EMA model. The EMA results for an infinite energy barrier U_B are obtained by setting the wavefunction to zero at the oxide interface.

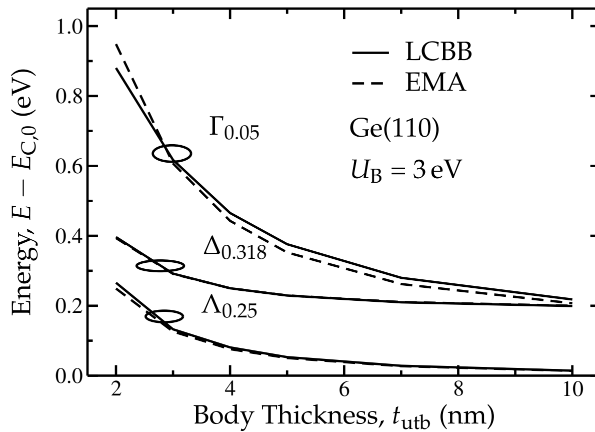


Figure 2.15: Ge(110). Lowest eigenvalue versus the UTB thickness for the $\Lambda_{0.25}$, $\Delta_{0.318}$ and $\Gamma_{0.05}$ valleys calculated with either the LCBB or the EMA model. The energy offset of the Δ and Γ valleys with respect to the Λ valleys of bulk germanium is 189 meV and 145 meV, respectively as reported in Tab. 2.1.

the LCBB results very well when the finite barrier U_B=3 eV is accounted for. Instead, for an infinite energy barrier, that corresponds to a closed boundary condition for the wavefunction at the semiconductor-oxide interface, the EMA minima increase well above the corresponding LCBB values for the thinnest silicon films. Fig. 2.15 reports the same comparison as in Fig. 2.14 for the $\Lambda_{0.25}$, the $\Delta_{0.318}$ and the $\Gamma_{0.05}$ valleys of the Ge(110). Even in this case the EMA with finite barrier height reproduces well the minima of the different valleys and their relative position, which sets the valley that gives the dominant contribution to the inversion charge [cf. (2.32), (2.34)] and to the current of the transistor [111, 112].

Similar to Fig. 2.8 on p. 32, Fig. 2.16 illustrates the impact of the barrier height U_B on the agreement between the EMA and the LCBB calculations for some of the inversion layers considered in Figs. 2.14 and 2.15 and for the thinnest semiconductor film. The same barrier height is used in the EMA and LCBB calculations. As it can be seen the differences are reduced for smaller U_B values, which are representative of some *high-κ* materials actively investigated as possible SiO_2 replacements for the gate dielectric [62]. We found that the impact of U_B on the absolute values of the subband minima and on the agreement between the EMA and LCBB calculations is significantly smaller for larger film thicknesses t_{utb}.

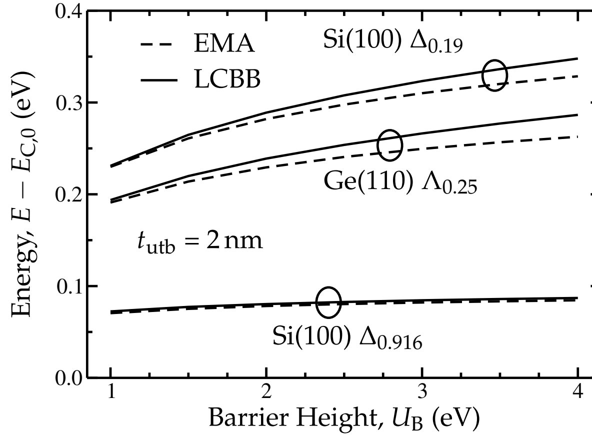


Figure 2.16: Lowest eigenvalue versus the barrier height U_B for different valleys of silicon and germanium inversion layers. The semiconductor thickness is t_{utb} = 2 nm and the same barrier height is used in the EMA and LCBB. The difference between the EMA and LCBB increases with U_B; the effect is more clearly observed for the small quantization masses (see Tab. 2.1).

In-plane energy dispersion and transport masses

In the energy dispersion illustrated in Fig. 2.12 we can see that the EMA can reproduce fairly well not only the subband minima but even the energy dependence on the wavevector. This implies that, in the case of Fig. 2.12, the values for the effective masses reported in Tab. 2.1 can be reliably used to describe the 2D energy dispersion.

The values of the transport masses extracted from the LCBB calculations have been systematically studied for different semiconductor thicknesses. We have calculated the transverse m_t^* and the longitudinal m_l^* masses for the most relevant valleys of the Si(100) and the Ge(110) inversion layers by best fitting the LCBB energy dispersion using a strictly parabolic energy dispersion. The fitting has been always performed by taking a small energy range of 10 meV above the minimum of each valley. Fig. 2.17 illustrates some cases where the effective masses exhibit a non-negligible dependence on t_{utb}, hence a non-negligible deviation from the values reported in Tab. 2.1 and routinely employed in the EMA calculations [111–113]; the effect is particularly pronounced for the transverse mass in the $\Delta_{0.19}$ valleys of the Si(100).

In the cases that are not illustrated in Fig. 2.17 we could not identify a systematic and quantitatively relevant change of the effective masses versus t_{utb},

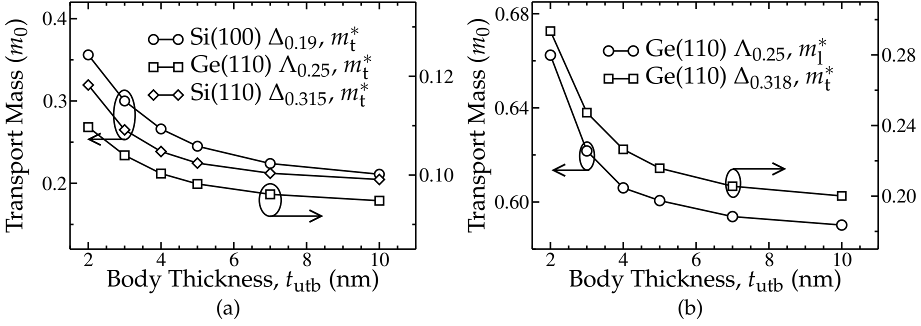


Figure 2.17: Transverse m_t^* and longitudinal m_l^* effective masses versus the semiconductor thickness for some valleys of the Si(100) and the Ge(110) inversion layers. (a) Si(100), $\Delta_{0.19}$, m_t^*; Ge(110), $\Lambda_{0.25}$, m_t^*; Si(110), $\Delta_{0.315}$, m_t^*; (b) Ge(110), $\Lambda_{0.25}$, m_l^*; Ge(110), $\Delta_{0.318}$, m_t^*. These effective masses exhibit a non-negligible dependence on t_{utb} and deviate from the values reported in Tab. 2.1 for very small semiconductor thicknesses.

hence in these cases the EMA model can be reliably used with the values for the masses of Tab. 2.1 (derived from the masses of the bulk crystal).

However, besides the possible changes of the transport masses, the LCBB calculations also reveal some differences in the 2D bandstructure with respect to the EMA results that cannot be simply accounted for by adjusting the EMA parameters. In this respect, it was pointed out that the Si(100) inversion layers exhibit a third system of valleys (besides the $\Delta_{0.916}$ and $\Delta_{0.19}$ indicated in Tab. 2.1), which is located at the boundary of the 2D Brillouin zone and it is typically neglected in the EMA picture [102, 106].

A case of a similarly large discrepancy between the EMA and LCBB results is illustrated in Fig. 2.18 for the Si(111) inversion layer. Fig. 2.18(a) shows the energy dispersion of the lowest subband in the **k** plane and for t_{Si} = 2 nm. The hexagon indicates the 2D Brillouin zone and the minimum along the positive k_x direction is at $k_x = 1.7/\sqrt{6} \approx 0.694$, as expected from the position of the energy minima in the 3D Brillouin zone [85, 102]. However Fig. 2.18(b) shows that, for $t_{utb} = 2$ nm, the minimum calculated by the LCBB method tends to move towards the edge of the 2D Brillouin zone (i.e. at $k_x = 2.0/\sqrt{6} \approx 0.8165$), and its value is significantly overestimated by the EMA model. Furthermore, the LCBB bandstructure exhibits a flat energy branch that corresponds to an effective mass much larger than the $m_l^* = 0.674m_0$ value reported in Tab. 2.1.

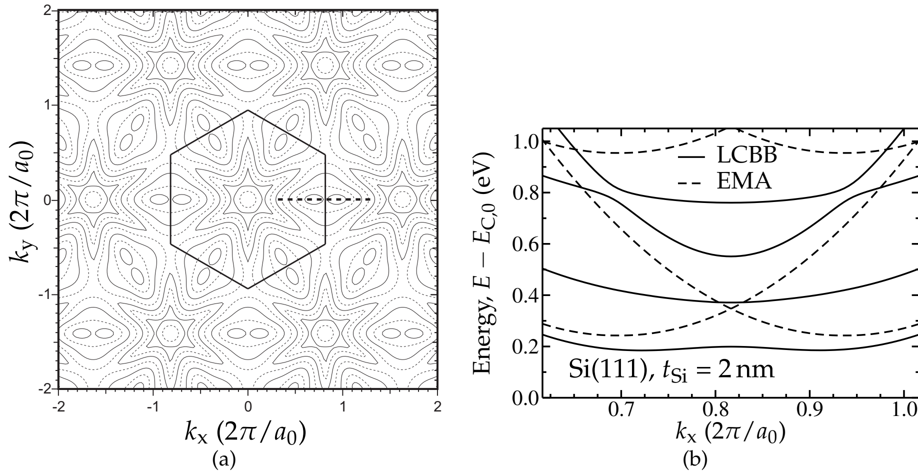


Figure 2.18: Si(111), t_{Si} = 2 nm. (a) Contour plot for the lowest eigenvalue $E_0(\mathbf{k})$ versus the wavevector $\mathbf{k}$ obtained by solving (2.35). The six degenerate minima are in $\mathbf{k} = (\pm 1.7/\sqrt{6}, 0)$ and in $\mathbf{k} = (\pm 0.85/\sqrt{6}, \pm 0.85/\sqrt{2})$. The hexagon indicates the 2D Brillouin zone [102]; (b) Energy dispersion along the dashed line indicated in Fig. 2.18(a) obtained with either the LCBB or the EMA model (U_B = 3 eV). The minimum of the LCBB calculations is no longer at the point $k_x = 1.7/\sqrt{6}$, as for thicker silicon films. Furthermore, the minimum predicted by the EMA overestimates the corresponding LCBB value.

We note that the flat energy branch is anyway difficult to be reproduced with a simple parabolic or non-parabolic model.

Density of States

In order to further compare the LCBB and the EMA models we have studied the 2D electron density of states (DOS), which is an important parameter because it is tightly related to the calculation of the scattering rates. The EMA model has analytical expressions for the 2D DOS for both the parabolic and the non-parabolic case [84, 97, 110]. For the LCBB model, instead, the DOS has been calculated numerically by counting, for each 'energy bin' (spacing in energy), the $\mathbf{k}$ points in the 2D first Brillouin zone that have an eigenvalue belonging to the energy bin. Each $\mathbf{k}$ point must be weighted for an appropriate area in the $\mathbf{k}$ plane according to the $\mathbf{k}$ discretization, which leads to the

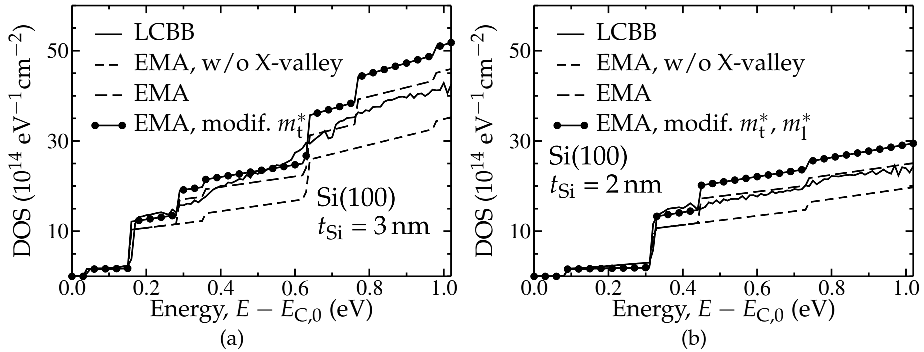


Figure 2.19: Electron two dimensional density of states versus energy for the Si(100) obtained with either the LCBB or the EMA model; E is relative to the bulk conduction band edge $E_{C,0}$. (a) t_{Si} = 3 nm; (b) t_{Si} = 2 nm. The EMA non-parabolic model that employs only the $\Delta_{0.916}$ and $\Delta_{0.19}$ valleys (short dashed lines) tends to underestimate the DOS calculated by the LCBB method (solid lines). A better agreement with the LCBB calculations is obtained by explicitly adding the X-valley to the EMA model (long dashed line) [102]. For the EMA case we show the results corresponding to the transport masses defined in Tab. 2.1 (dashed and long-dashed lines) and the results that have been obtained by using, for each t_{Si} value, the corresponding transport masses reported in Fig. 2.17 (filled circles). In this latter case (filled circles) the X-valleys have been accounted for.

following expression:

$$D_{2D}(E) = \sum_{n,k_x,k_y} \frac{2\Delta k_x \Delta k_y}{(2\pi)^2 \Delta E} \left\{ H\left[E_n(k_x,k_y) - E + \frac{\Delta E}{2}\right] - H\left[E_n(k_x,k_y) - E - \frac{\Delta E}{2}\right] \right\} \quad (2.36)$$

where ΔE, Δk_x and Δk_y indicate the spacing in the energy and in the **k** discretization while $H(x)$ is the step function. Fig. 2.19 compares the EMA to the LCBB DOS curves for Si(100) and for t_{Si} = 3 nm and 2 nm. The EMA results have been obtained for either the transport masses of Tab. 2.1 or for the values corrected according to the t_{utb} dependence shown in Fig. 2.17. As it can be seen, the conventional two valley picture typically employed in the EMA approach tends to underestimate the DOS. As reported in [102, 106], LCBB calculations have revealed the existence of a *third* system of valleys, in addition to the 'conventional' $\Delta_{0.916}$ and the $\Delta_{0.19}$ valleys discussed so far. This third valley is labeled *X-valley* from its location at the edge of the 2D Brillouin

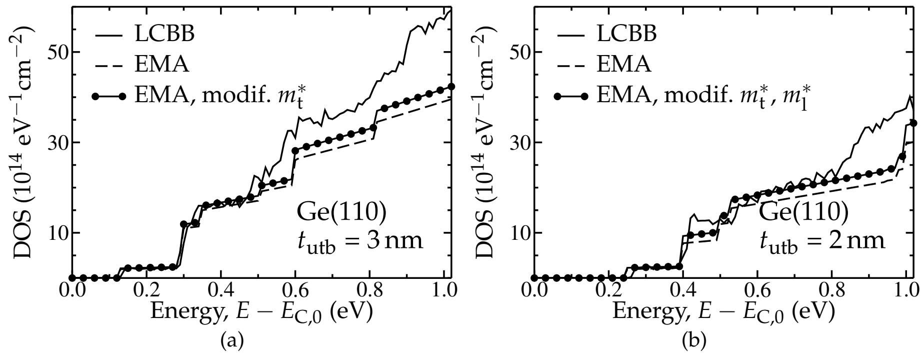


Figure 2.20: Electron two dimensional density of states versus energy for the Ge(110) obtained with either the LCBB or the EMA model. For the EMA case we show the results corresponding to the transport masses defined in Tab. 2.1 and the results that have been obtained by using, for each t_{utb} value, the corresponding transport masses reported in Fig. 2.17. (a) t_{Ge} = 3 nm; (b) t_{Ge} = 2 nm. The agreement of the EMA to the LCBB calculations is fairly good up to an energy about 0.6 eV above the minimum of the lowest subband.

zone in the X-symmetry point, i.e. $\mathbf{k} = (\pm 1.0, 0)$. By including the contribution of the X-valley and by employing the t_{utb} dependent transport masses, the agreement between the EMA and the LCBB results becomes good up to energies around 0.6 eV above the lowest minimum. This is the energy range of practical interest to simulate I_{DS} in the transistors of modern CMOS technologies, where the supply voltage is 1 V or below. Fig. 2.20 shows the comparison between the EMA and the LCBB DOS for the Ge(110). Even in this case the agreement is fairly good in the range of most practical interest and the use of the t_{Ge} dependent transport masses reported in Fig. 2.17 make the EMA results closer to the LCBB calculations. For the Si(111), instead, Fig. 2.21 shows that the agreement between the EMA and the LCBB results is worse. In particular, in the case of t_{Si} = 2 nm, the LCBB DOS features a peak at an energy appreciably lower than the lowest available states according to the EMA model. This peak of DOS clearly stems from the flat energy branch in the LCBB energy dispersion illustrated in Fig. 2.18(b), which results in a DOS that decreases with the energy for very low energy values. This latter behavior cannot be reproduced by an EMA model, even if we adjust the masses, hence we conclude that, in the case of Si(111) and at very small silicon thicknesses, it is problematic to follow the LCBB results with an EMA approach.

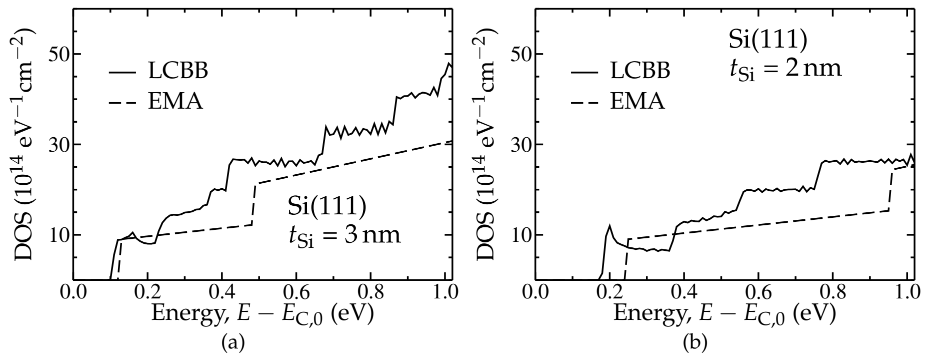


Figure 2.21: Electron two dimensional density of states versus energy for the Si(111) obtained with either the LCBB or the EMA model. (a) t_{Si} = 3 nm; (b) t_{Si} = 2 nm. In the case of the thinnest semiconductor film we see a clear peak of the DOS calculated by the LCBB method at energies below the lowest states obtained with the EMA model.

Summary

We have systematically investigated the validity of the EMA quantization model for the silicon and the germanium *n*-MOSFETs with different crystal orientations. To this purpose we have used the LCBB full-band quantization model to study the main parameters that govern the electron density and the transport in the inversion layers: the minima of the 2D subbands, the transport masses and the 2D density of states.

The comparison of the EMA results to the LCBB calculations indicates that the simple EMA approach, employing the commonly used quantization effective masses m_z^* (see also Tab. 2.1), can reproduce quite accurately the minima of the 2D subbands, provided that a finite value of the semiconductor-oxide barrier is accounted for. When the EMA is used with a closed boundary condition for the wavefunction at the semiconductor-oxide interface, instead, the results for body thickness below approximately 5 nm deviate significantly from the LCBB calculations, which inherently employ a finite semiconductor-oxide barrier.

In some quantization directions the LCBB method points out that, by scaling t_{utb}, the 2D energy dispersion is not merely shifted (as it happens in an EMA model), but an appreciable distortion of the energy to **k** relation is produced. We have quantified this effect by extracting from the LCBB calculations the effective transport masses and illustrated their possible dependence

on t_{utb} and the deviations from the values typically employed in the EMA model.

In most cases the changes of the transport masses are modest, at least for the crystal directions of largest technological interest. However in some cases the LCBB approach points out that the confining potential energy changes the position of the energy minima in the 2D Brillouin zone. Furthermore, it produces branches in the 2D energy dispersion that are hard to be reproduced with a simple parabolic or non-parabolic model (see the discussion of Figs.2.18 and 2.21 for the Si(111) inversion layers).

As a general conclusion it can be stated that the EMA model is fairly reliable even for very thin silicon films; in some cases its accuracy can be improved by changing the transport masses as illustrated in Fig. 2.17. Our results have been obtained by using a schematic, square well confining potential energy, however we verified that the main trends apply as well in a triangular confining potential [not shown], hence they are expected to be of general validity for the electron inversion layers of n-MOS transistors.

Three

Energy Band Offset Extraction

In Chapter 2, we have seen that if the silicon body thickness is reduced, quantum confinement (or *quantization*) effects start to become significant. The carriers are confined in a potential energy well which is determined by the energy band discontinuity of the constituent semiconductor and the surrounding insulating gate-dielectrics. When discussing quantum confinement in UTB-SOI devices, the confining potential energy is often treated as a square well. The strength of the quantization is set by the UTB thickness t_{Si}, hence this type of confinement is commonly referred to as structural (or geometrical) confinement [92]. Generally, quantum confinement results in the formation of subbands within the conduction and valence band, as shown schematically in Fig. 3.1. Since the band gap is determined by the energy difference between the top of the highest valence subband and the bottom of the lowest conduction subband, a shift in these energy levels with respect to the bulk band edges will change the width of the band gap.

The rearrangement of the allowed energies into ladders of energy subbands depends on the location of the energy minima (*valleys*) in the 3D (bulk) semiconductor crystal, the orientation of the transistor and, in particular, the *quantization direction* normal to the semiconductor-oxide interface. Intuitively, one can imagine that due to the energy gaps separating the subbands, the density of states (DOS) of a quantized system will be lower compared to the DOS in a bulk semiconductor, in which the conduction and valence band consist of a virtually infinite number of energy levels.

In summary, with the ultimate scaling of t_{Si}, fundamental semiconductor properties such as the band alignment or band gap (E_{g}), effective density of states (DOS) and also the mobility (μ) deviate from their respective bulk values, as reported in [37,115–117]. Hence these quantities become *device* properties, rather than *material* properties. Furthermore we know, as implicitly given by (2.8), that the threshold voltage of UTB devices with lowly doped body can

be set by the gate workfunction [76]. In order to fully understand and tailor the band alignment, DOS and mobility, accurate extraction and determination of their scaling behavior with e.g. t_{Si} is required.

In this chapter, we will complement the analysis in Chapter 2 with a technique to extract a shift in the energy band alignment using electrical measurements. In particular, we exploit the temperature dependence of the subthreshold current to extract changes in the band alignment and DOS, similar to the commonly used method to determine the energy barrier, or activation energy, in e.g. Schottky diodes [12] and Si-based bipolar transistors [118, 119]. We will see that in UTB SOI devices the subthreshold current, which is usually considered an unwanted 'leakage' current, actually proves to be instrumental in acquiring essential information on the (change in) intrinsic device and transport properties such as the band alignment, mobility and density of states.

3.1 Conventional methods to extract band offsets

Experimental evidence of a t_{Si}-dependent change in the band gap is reported in [116, 120] for t_{Si} values below 5 nm, measured with X-ray spectroscopy. The effect of quantum confinement on the *electrical* UTB device characteristics is generally quantified using the shift in threshold voltage (V_{TH}) as metric [117, 121–125]. In fact, changes in the band alignment are expected to contribute to a shift in V_{TH} since, in case of band gap widening, a higher gate bias is required to obtain the same inversion charge density. However, while the *theoretical* V_{TH} is well–defined [12], several definitions of the *experimental* V_{TH} exist [78, 126], e.g. V_{GS} corresponding to a fixed current level, or the linearly extrapolated intersection of I_{DS} with the V_{GS} axis, starting from the maximum transconductance. In this chapter, we adhere to the latter definition.

Regardless of the definition used, the threshold voltage inherently denotes the transition from weak to strong inversion. Therefore, besides the band gap and the density of states, also properties such as the gate oxide thickness, mobility and the series resistance are incorporated in the threshold voltage. After all, those parameters determine the current in the strong inversion regime.

Differently from the above approaches, here we will assess the impact of structural quantum confinement through the temperature dependence of the subthreshold current, in the following denoted as the "$I_{DS}(T)$" method. In particular, we compare the $I_{DS}(T)$ results with the conventional "ΔV_{TH}" method. Furthermore, with additional simulations we will point out that by exploiting the temperature dependence of the subthreshold gate *capacitance*, a

t_{Si} dependent change in DOS and mobility can be quantified.

Over time, the main results shown in the following have been presented in [127–129], of which this chapter presents an aggregate overview. The organization of this chapter is as follows: first the $I_{\mathrm{DS}}(T)$ method will be outlined, followed by an initial verification of the simulations with experimental data, obtained from in-house fabricated UTB-SOI MOSFETs [130]. The actual comparison of the $I_{\mathrm{DS}}(T)$ and ΔV_{TH} method is carried out with simulations, in order not to cloud the results with processing-induced device variation. We conclude the comparative analysis with a discussion on the sensitivity of both the $I_{\mathrm{DS}}(T)$ and ΔV_{TH} method to a change in several device and material properties.

In the following, 'silicon' and 'semiconductor' are used interchangeably. Although the numerical results shown hereafter apply to silicon only, the method itself holds for other semiconductors as well.

3.2 Band offset extraction from $I_{\mathrm{DS}}(T)$ – Theory

The subthreshold current contains important information on several 'intrinsic' device parameters, which can be accessed through its temperature dependence. The subthreshold current essentially originates from diffusion. For a long-channel n-MOSFET, I_{DS} can be calculated with

$$I_{\mathrm{DS}} = \frac{\mu k_{\mathrm{B}} T}{qL} Q_{\mathrm{i}}(V_{\mathrm{GS}}) \left[1 - \exp\left(-\frac{qV_{\mathrm{DS}}}{k_{\mathrm{B}}T}\right)\right] \tag{3.1}$$

with μ the low-field carrier mobility, T the absolute temperature, q the elementary charge, L the channel length, and V_{DS} the drain–source voltage; Q_{i} is the classical inversion charge density per unit area at the source side of the channel, given by [76]

$$\begin{aligned} Q_{\mathrm{i}} &= -qt_{\mathrm{Si}}n_{\mathrm{cl}} = -qt_{\mathrm{Si}}N_{\mathrm{CB}} \exp\left(\frac{E_{\mathrm{F}} - E_{\mathrm{C}}}{k_{\mathrm{B}}T}\right) \\ &= -qt_{\mathrm{Si}}N_{\mathrm{CB}} \exp\left(\frac{\chi_{\mathrm{s}} - \phi_{\mathrm{m}}}{k_{\mathrm{B}}T}\right) \exp\left(\frac{qV_{\mathrm{GS}}}{k_{\mathrm{B}}T}\right) \end{aligned} \tag{3.2}$$

where n_{cl} is the classical electron concentration (see also (2.5) on p. 20); N_{CB} is the DOS in the bulk conduction band, E_{F} the Fermi level at the source side of the channel and E_{C} the conduction band edge. We assume that the semiconductor body is sufficiently thin to be fully depleted, so that we can safely disregard the depletion charge (see also section 2.1 on p. 20). This implies that, different from conventional bulk MOSFETs, there is no significant initial

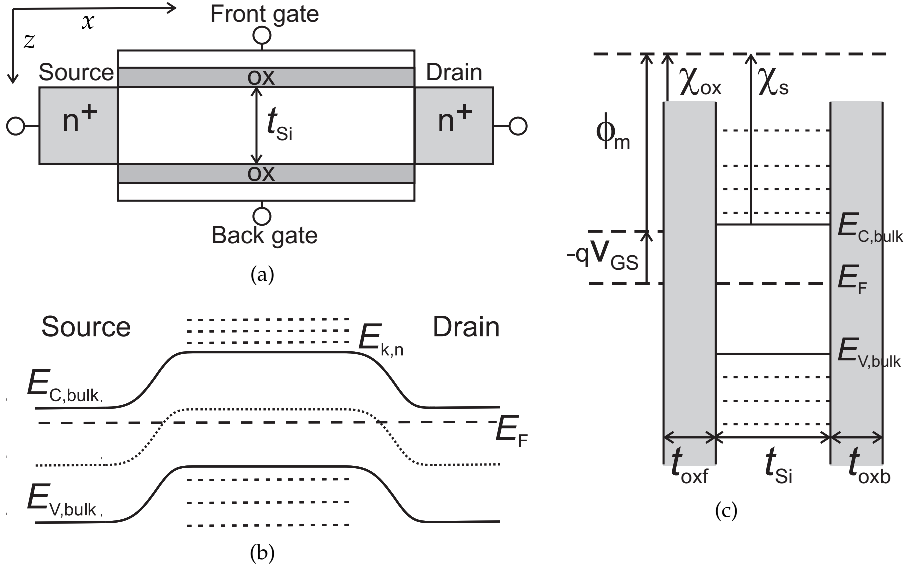


Figure 3.1: Schematic cross–section of a symmetric long–channel DG device (a) with the corresponding cut along the x-direction (b) and z-direction (c). The dotted lines represent the energy subband minima originating from quantum confinement in the thin body; χ_s is the electron affinity, χ_{ox} the affinity of the oxide, and ϕ_m is the gate workfunction; t_{oxf} and t_{oxb} are the front and back oxide thickness respectively.

band bending so that the flat band voltage V_{FB} is essentially zero. Furthermore, Boltzmann's approximation is assumed to hold as well as an uniform 'volume inversion' [28], which is justified by the low injection condition in the subthreshold regime; χ_s is the electron affinity, χ_{ox} the affinity of the oxide, ϕ_m the gate work function and V_{GS} the gate bias [see also Fig. 3.1(c)]. Eqs. (3.1) and (3.2) apply to fully symmetric devices as well as DG UTB devices with strongly asymmetric front and back gate oxide thickness (i.e. $t_{oxf} \ll t_{oxb}$), provided that 1) both gates are biased equally, as discussed in Appendix C on p. 147 and 2) that the channel is long (i.e. no significant short channel effects).

In fact, (3.1) and (3.2) suggest that, for a given gate-source bias, the difference in gate work function and conduction band edge can be extracted from the slope of the drain current versus the inverse temperature (on semi-log scale):

$$I_{DS} \propto \mu t_{Si} N_{CB} \exp\left(\frac{\chi_s - \phi_m + qV_{GS}}{k_B T}\right) \tag{3.3}$$

Note that V_{GS} appears as an additive term in the exponent of (3.3). Hence, the

applied V_{GS} must be subtracted when extracting the band offset from the slope of $\ln(I_{\mathrm{DS}})$ vs. $1/T$. Furthermore the low gate bias in subthreshold ensures that the impact of *electrical* carrier confinement, i.e. gate induced quantization within the inversion layer (see Section 2.2, p. 26), is negligible. The impact of the temperature dependence of μ and N_{CB} is discussed in Section 3.4.

Similarly, for a PMOS the difference in *valence* band edge and gate work function can be extracted. In general, any shift in band edge, i.e. the top of the source/channel barrier as shown in Fig. 3.1(b), relative to the gate work function can be extracted using the exponential temperature dependence. Besides quantization induced band offsets, also shifts in band alignment due to e.g. strain or work function differences can be extracted. This work will focus on extracting band edge shifts stemming from structural carrier confinement in the ultra thin body.

As we have seen in (2.34) on p. 34, we can include the the effect of carrier confinement on the charge density by replacing (3.2) with the 2D charge density

$$\begin{aligned} Q_{\mathrm{i,2D}} &= -q \int_{-t_{\mathrm{Si}}/2}^{+t_{\mathrm{Si}}/2} \rho_{\mathrm{e}}(z)\, dz \\ &= -q \cdot \frac{k_{\mathrm{B}}T}{\pi\hbar^2} \sum_{\nu,n} m^*_{d,\nu} \exp\left(\frac{E_{\mathrm{F}} - E_{\nu,n} - E_{\mathrm{C},0}}{k_{\mathrm{B}}T}\right) \\ &= -q \cdot \frac{k_{\mathrm{B}}T}{\pi\hbar^2} \sum_{\nu,n} m^*_{d,\nu} \exp\left(\frac{\chi_{\mathrm{s}} + E_{\nu,n} - \phi_{\mathrm{m}}}{k_{\mathrm{B}}T}\right) \exp\left(\frac{qV_{\mathrm{GS}}}{k_{\mathrm{B}}T}\right) \end{aligned} \tag{3.4}$$

with

$$E_{\nu,n} = \frac{\hbar^2}{2m^*_{z,\nu}} \left(\frac{n\pi}{t_{\mathrm{Si}}}\right)^2 \tag{3.5}$$

where $m^*_{d,\nu}$ and $m^*_{z,\nu}$ are the DOS and quantization effective masses, respectively, belonging to valley ν; $E_{\nu,n}$ is the minimum of subband n in valley ν, relative to the bulk band edge $E_{\mathrm{C},0}$ [cf. (2.20)]. In (3.4) and (3.5) we have assumed a square confining potential with infinite barrier height, in order to arrive at a closed-form analytical expression. The charge density in presence of a finite confining barrier can be calculated similarly, by numerical evaluation of (3.4) with (2.30) and (2.31).

The strength of structural confinement in thin silicon layers is governed by t_{Si}, as indicated by (3.4) and (3.5). The edges of both the conduction and valence band are shifted with respect to the bulk band edges, due to the offset of (mainly) the first subband. In fact, the extracted band offsets presented in

the course of this chapter represent 'effective', or apparent, band offsets, i.e. the shift in band edge 'averaged' in energy over the entire channel. In case of the silicon *conduction* band, the offset is predominantly determined by the first subband of the *unprimed* valley (labeled $\Delta_{0.916}$), which has the largest (longitudinal) mass along the quantization direction. Calculation of the theoretical effective valence band shift is less straightforward, due to its anisotropic nature. However, a fully numerical analysis of the theoretical band offset is beyond the scope of this work. Simple parabolic bands have been assumed to illustrate the variation of the first subband minimum with t_{Si}. In the following, the extracted band offsets are denoted as ΔE_{x}.

The t_{Si} dependence enters the DOS through the position of the subbands: the energy separation between the subbands increases with decreasing t_{Si}, thereby reducing the number of available energy states. As pointed out in [127], the t_{Si} dependence can be exploited to extract the shift in energy band alignment, mobility and DOS, using the temperature dependence of the subthreshold current.

In short, we observe, for a fixed V_{DS} and V_{GS}, the ratio of the subthreshold current (η_{rat}) in two UTB devices with different t_{Si}, as given by

$$\eta_{\mathrm{rat}} \equiv \frac{I_{\mathrm{ref}}}{I_{\mathrm{thin}}} = \frac{\mu_{\mathrm{ref}} \cdot g(t_{\mathrm{Si,ref}})}{\mu_{\mathrm{thin}} \cdot g(t_{\mathrm{Si,thin}})} \cdot \exp\left(\frac{\Delta E_{\mathrm{x}}}{k_{\mathrm{B}} T}\right) \tag{3.6}$$

and thus

$$\ln(\eta_{\mathrm{rat}}) = \ln\left[\frac{\mu_{\mathrm{ref}} \cdot g(t_{\mathrm{Si,ref}})}{\mu_{\mathrm{thin}} \cdot g(t_{\mathrm{Si,thin}})}\right] + \left(\frac{\Delta E_{\mathrm{x}}}{k_{\mathrm{B}} T}\right) \tag{3.7}$$

In the above equations, the subscripts 'ref' and 'thin' refer to the quantities corresponding respectively to a reference device, having a relatively thick body, and a device with a thinner body and corresponding supposedly enlarged band gap; $g(t_{\mathrm{Si}})$ represents the DOS. The difference in energy band edge of the thinnest layer and the reference device, ΔE_{x} in (3.6), is equal to $E_{\mathrm{x,thin}} - E_{\mathrm{x,ref}}$, with subscript 'x' denoting the valence or conduction band edge for *p*-type or *n*-type devices respectively. If the body thickness of the reference device is sufficiently thick, i.e. negligible impact of quantum confinement, then the resulting values of ΔE_{x} are referenced to the bulk silicon band edge. This choice is just a matter of convenience, as Eq. (3.6) is valid for any combination of t_{Si}'s, provided that the volume inversion condition holds.

Separating $\Delta\mu$ and ΔDOS with CV(T)

As mentioned before, structural quantum confinement alters not only the band gap but also the DOS and mobility, thus making those quantities explicitly t_{Si} dependent [40]. Since the DOS and the mobility jointly enter the prefactor of (3.6), quantifying their individual contribution to the total change in offset of η_{rat} is not straightforward. However, in order to disentangle changes in mobility and DOS, the *IV* measurements can be supplemented with subthreshold *capacitance* characteristics, which do not depend on the mobility. Furthermore, noting that in subthreshold $Q_{\mathrm{i}} = k_{\mathrm{B}}T/q \cdot C_{\mathrm{GG}}$ (see also Appendix D on p. 149), the ratio of capacitances (θ_{rat}) is obtained similarly to (3.6)

$$\theta_{\mathrm{rat}} \equiv \frac{C_{\mathrm{GG,ref}}}{C_{\mathrm{GG,thin}}} = \frac{g(t_{\mathrm{Si,ref}})}{g(t_{\mathrm{Si,thin}})} \cdot \exp\left(\frac{\Delta E_{\mathrm{x}}}{k_{\mathrm{B}}T}\right) \tag{3.8}$$

with $C_{\mathrm{GG,ref}}$ and $C_{\mathrm{GG,thin}}$ respectively the subthreshold gate capacitance of the reference and the thin body device. When plotting $\ln(\theta_{\mathrm{rat}})$ vs. the inverse temperature, the offset is determined exclusively by the DOS. In summary, combining subthreshold current *and* capacitance measurements allows for a separate investigation of a possible t_{Si} dependent change in mobility and DOS, in addition to an extraction of shifts in energy band alignment. Besides the relative change in mobility, also its absolute value can be obtained from the combination of the subthreshold capacitance and current on a single device, as given by (cf. Appendix D on 149)

$$\mu = \frac{I_{\mathrm{DS}}}{C_{\mathrm{GG}}} \cdot \frac{L}{u_{\mathrm{th}}^2} \frac{1}{1 - \exp\left(\frac{-V_{\mathrm{DS}}}{u_{\mathrm{th}}}\right)} \tag{3.9}$$

with u_{th} the thermal voltage ($k_{\mathrm{B}}T/q$). From a practical point of view, the influence of the parasitics (e.g. the gate–source/drain overlap capacitance) can be reduced by performing the measurement differentially, i.e. by plotting for each t_{Si}, the difference in *CV* characteristics from devices with different gate lengths. By doing so, the capacitance in the low gate–bias regime decreases, which effectively extends the range in which the subthreshold capacitance characteristics can be observed, as will be shown in the next section.

3.3 Results

The procedure outlined in the previous section will be carried out with simulations, and the extracted band offsets will be compared with shifts in V_{TH}. In

the following, the body of the reference device is 27 nm thick, i.e. sufficiently thick for quantum confinement to be negligible [115, 116, 120]. Thus, the extracted band edge shift in the thinnest layer is relative to the conduction band minimum or valence band maximum in bulk silicon.

Verification of Simulations

The simulations have been performed with Synopsys Sentaurus [94]. Quantum confinement is accounted for by employing the Density Gradient model [131, 132], which applies a quantum correction Λ_n to the classical charge distribution, as given by

$$n_{corr} = N_{CB} \exp\left(\frac{E_F - E_C - \Lambda_n}{k_B T}\right) \tag{3.10}$$

with

$$\Lambda_n = -\frac{\gamma \hbar^2}{6m^*} \frac{\nabla \sqrt{n}}{\sqrt{n}} \tag{3.11}$$

where γ is a fit factor. In its implementation in the device simulator [133], (3.11) is rewritten in terms of the corrected potential energy $\Phi = E_{C,0} + \Phi_m + \Lambda_n$, with $\Phi_m = -k_B T \ln(N_{CB}/N_{ref})$ and N_{ref} an arbitrary normalization constant [133]. The Density Gradient model has proven to be adequate in reproducing the effect of quantum confinement on the device characteristics [95].

For the mobility, the Philips Unified Mobility model with Lombardi transversal field dependence was used [134, 135] with the default parameters. The Si/SiO_2 barrier height is 3.17 eV for electrons and 4.71 eV for holes and the Si bulk band gap is 1.12 eV at 300 K.

First, the simulations are verified with existing experimental data, obtained from [127]. The measured devices are long-channel (100) Silicon–on–Insulator (SOI) MOSFETs [130, 136] with the backside of the wafer as back-gate contact, hereafter denoted as "UTB-SOI"; t_{Si} ranges from 27 down to 5 nm, the channel length is 25 µm, and the oxide and BOX thicknesses are 25 nm and 400 nm (resp. t_{oxf} and t_{oxb}). Fig. 3.2 shows the measured and simulated subthreshold curves for the UTB-SOI devices with t_{Si} 27 nm. Since the effect of quantum confinement on the mobility is not included in the current TCAD models, the initial simulations are verified with the measured data from only the thickest device, which exhibits no significant impact of quantum confinement. Good agreement is observed both in weak and strong inversion. The curves have been fitted solely by adjusting the workfunction ϕ_m of the (n^+ poly) gates.

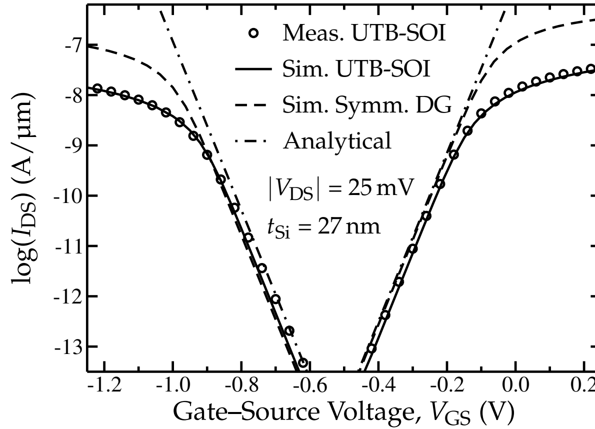


Figure 3.2: Experimental (symbols) and simulation data on UTB-SOI devices with $t_{oxb} \gg t_{oxf}$ (solid lines). Simulation data on fully symmetric devices with t_{ox} = 2 nm are shown as well (dashed lines), along with analytical values (dot–dashed lines) calculated with (3.1); T = 300 K, L = 25 µm.

The required adjustment was directly extracted from the temperature dependence of the measured subthreshold current itself, as exemplified in the inset of Fig. 3.3. The third set of curves shows the simulated data for a fully symmetric device, having the same t_{Si} but with $t_{oxf} = t_{oxb} = 2$ nm [e.g. see Fig. 3.1(a)]. When operated in DG mode, the subthreshold curves of the UTB-SOI and symmetric device indeed coincide, as pointed out in Appendix C. Hence, the existing subthreshold current measurement data is fully comparable with the symmetric case, since the $t_{oxf} \ll t_{oxb}$ condition is fulfilled in our UTB-SOI device. Furthermore, the subthreshold current calculated with (3.1) is shown as well, demonstrating good agreement.

Fig. 3.3 shows the simulated subthreshold curves for the UTB-SOI NMOS, for different temperatures, along with the measured data. The temperature ranges from 300 K to 450 K and we verified that the inverse subthreshold slope varies linearly with temperature. This observation confirms that the temperature dependence of the exponential term in (3.6) is dominant.

In Fig. 3.3, the slight increase in current at low gate bias and the highest temperature originates from thermal generation of carriers at the Si-BOX interface, which cannot be controlled by the gate(s). The inset shows $\ln(I_{DS})$ versus the inverse temperature, from which the previously mentioned difference in gate and channel work function can directly be extracted. We stress that,

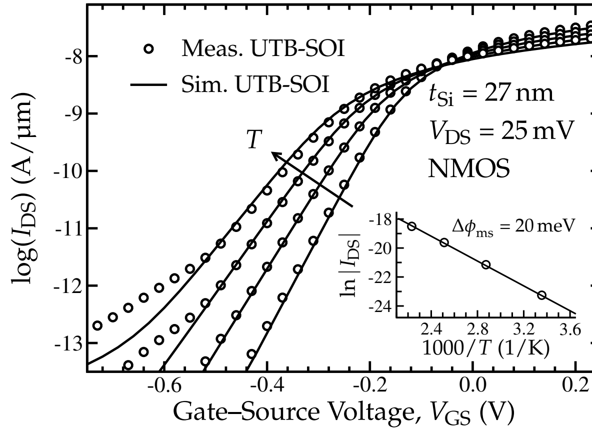


Figure 3.3: Simulated and measured subthreshold curves for T = 300–450 K; L = 25 µm. The inset shows $\ln(I_{DS})$ vs. $1000/T$, the slope of which gives the gate–channel workfunction difference $\Delta\phi_{ms}$.

apart from extracting the work function through the T-dependence of $\ln(I_{DS})$, no further calibration was performed. All forthcoming simulation results employ the default model parameters, without any adjustment. Good correspondence between measured and simulated data is observed, demonstrating that the simulations can be used with confidence for further investigations.

Extracted band offsets from $I_{DS}(T)$ vs. ΔV_{TH}

The simulations presented in the following have been performed on symmetric devices, with t_{ox} = 2 nm, L = 1 µm, and t_{Si} in the range 3–27 nm. Unless stated otherwise $|V_{DS}|$ = 25 mV and V_{GS} at which η_{rat} and θ_{rat} are recorded equals -0.775 V for PMOS and -0.450 V for NMOS. The workfunction ϕ_m of the ideal gate was set equal to χ_s, i.e. the electron affinity of bulk Si (4.07 eV). The inverse subthreshold slope S of the devices considered in this work varies linearly with temperature, and is close to the ideal value of $S = \ln(10)\frac{k_B T}{q}$, as shown in Fig. 3.4.

Fig. 3.5 plots $\ln(\eta_{rat})$ vs. the inverse temperature for a set of simulated NMOS devices with t_{Si} ranging from 9 nm down to 3 nm. The slope increases for decreasing t_{Si}, which corresponds to a shift of the conduction band edge upward in energy (i.e. a wider band gap or smaller χ_s). In addition, the offset in η_{rat} increases for decreasing t_{Si}, which generally represents a change in

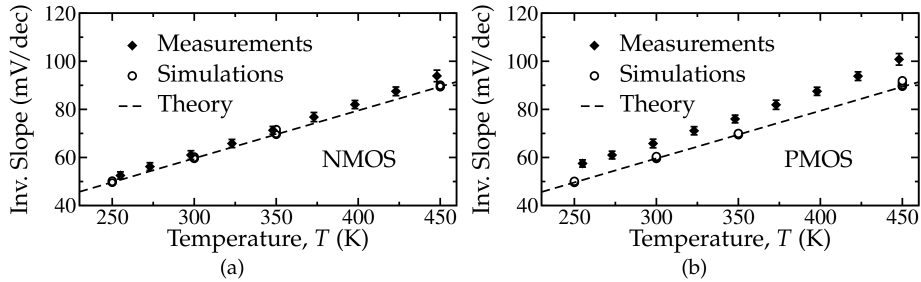


Figure 3.4: Inverse subthreshold slope, for the measured and simulated devices considered in this work, with t_{Si} = [3–27] nm; (a) NMOS and (b) PMOS (see also [127]).

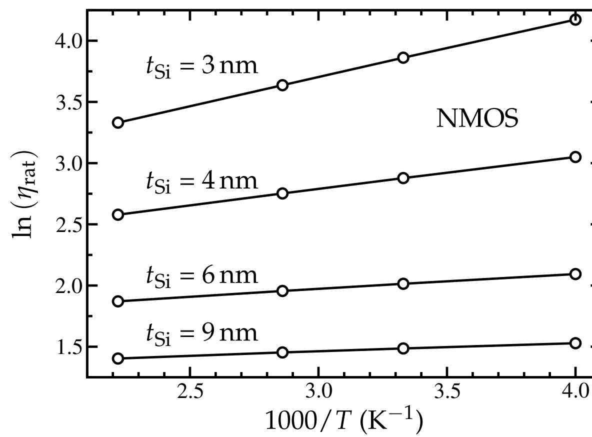


Figure 3.5: Simulated current ratio for several values of t_{Si}, with the t_{Si} 27 nm device as reference; L = 1 µm. The slope increases for decreasing t_{Si}, corresponding to a shift of the conduction band edge upward in energy.

combination of mobility and DOS [hence, the prefactor in (3.6)]. However, in this case, the observed change in offset is solely due to a reduction in the DOS, since in these simulations only bulk mobility models are employed which exhibit no direct t_{Si} dependence. Furthermore we stress that, in the subthreshold regime, the electric field normal to the gates is low. Hence, closer inspection confirms that the mobility is essentially constant throughout the entire channel region, irrespective of t_{Si}.

Repeating the above 'simulated experiments' *without* the Density Gradient correction yields an essentially temperature independent η_{rat}, hence $\Delta E_{\mathrm{x}} \to 0$. Moreover, we find that the offset in η_{rat} is equal to $t_{\mathrm{Si,ref}}/t_{\mathrm{Si,thin}}$ [not shown].

The band offsets extracted from the slope of η_{rat} are depicted in Figs. 3.6 and 3.7 for NMOS and PMOS devices respectively, along with the extracted shift in 'experimental' threshold voltage (ΔV_{TH}) for different temperatures and the UTB-SOI PMOS $I_{\mathrm{DS}}(T)$ data from [99]. The value of ΔV_{TH} was extracted by linear extrapolation from the maximum transconductance $g_{\mathrm{m,max}}$. Again, all values are referenced to the thickest device ($t_{\mathrm{Si}} = 27\,\mathrm{nm}$) which, by definition, exhibits no band edge shift (bulk band gap). For reference, the first subband minimum (holes: subband maximum) is shown as well, calculated with the (100) quantization effective masses [86,87], Tab. 2.1, and assuming an infinite square well.

The values obtained with the $I_{\mathrm{DS}}(T)$ method closely reproduce the analytically calculated lowest energy in each band, particularly for the NMOS. The slight discrepancy for the valence band offset is attributed to the approximations employed both in the density gradient model, which involves a single fit parameter [γ in (3.11)] to account for the different valleys [95], and in the single heavy hole effective mass used in the analytical calculation; in fact, those masses depend on the anisotropy and dimensionality of the confined system [87].

Although ΔV_{TH} shows the same trend, i.e. increasing V_{TH} for decreasing t_{Si}, the figures show that the extracted values are significantly higher than their respective $I_{\mathrm{DS}}(T)$ counterparts. This can be explained by noting that quantum confinement alters both the band gap and the (effective) DOS: splitting of the bulk conduction and valence band into subbands, gives 1) a wider band gap through the offset of (mainly) the first subband, and 2) a reduction of the available states due to the emerging energy gaps which separate the subbands. In the $I_{\mathrm{DS}}(T)$ method, shifts in the band edges are extracted from the *slope* of the current ratio, whereas changes in the DOS and mobility can be derived from the *offset* in the current ratio; hence, $I_{\mathrm{DS}}(T)$ measurements al-

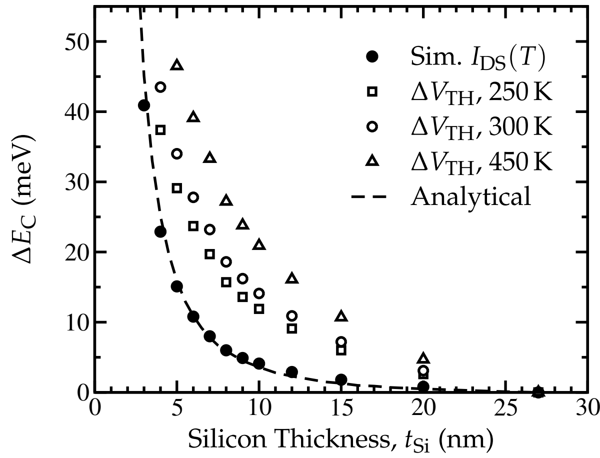


Figure 3.6: Conduction band offset, extracted from the slope of the subthreshold current ratio in Fig. 3.5. The dashed line shows the theoretical band offset, calculated with (3.5) and $m_z^* = 0.916m_0$ (as reported in Tab. 2.1 on p. 33), assuming an infinite potential barrier. The shift in threshold voltage (open symbols) is 1) significantly larger than the values extracted from $I_{DS}(T)$ and 2) temperature dependent.

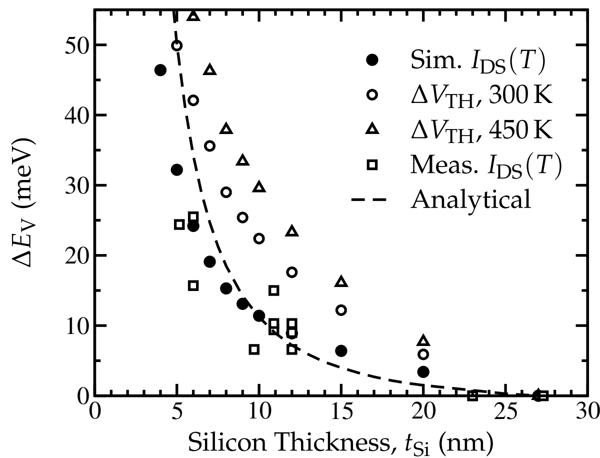


Figure 3.7: Valence band offset, extracted from the measured [99] and simulated subthreshold current ratio on PMOS DG devices. The dashed line shows the analytically calculated first subband maximum, obtained from (3.5) with $m_z^* = 0.291m_0$ [87]. The shift in threshold voltage is depicted as well.

low for an investigation of band edge shifts separately from the mobility and DOS, as suggested by (3.6). In contrast, ΔV_{TH} consists of a combined change in band gap, mobility and DOS, thus amplifying the apparent impact of quantum confinement. Furthermore, it is interesting to note that the threshold voltage exhibits a strong temperature dependence. While the latter is exploited in the $I_{DS}(T)$ measurements, it shows up as an additional source of error when determining the band edge shifts based on ΔV_{TH}.

Impact of $\Delta\mu$ on the extracted offset and on ΔV_{TH}

The results in the previous section have shown that the shifts in V_{TH} are fairly large with respect to the expected shift in band edge, whereas the extracted band offsets from $I_{DS}(T)$ are in close agreement with the theoretical values. In this regard, it is illustrative to investigate the impact of a change in mobility on η_{rat}, on the resulting values of ΔE_x, and on the corresponding ΔV_{TH}. As of yet, a direct t_{Si} dependence due to e.g. thickness fluctuation is not incorporated in the current TCAD mobility models, so the impact of a change in mobility is emulated by artificially setting the low-field mobility to a fixed value for each t_{Si}. Although the mobility is reported to decrease with decreasing t_{Si}, with a local maximum around 3–4 nm [36, 37], the exact choice of the mobility values in this work does not affect the generality of the conclusion. The mobility was manually reduced by 15% (t_{Si} = 9 nm) to 65% (t_{Si} = 3 nm) relative to the mobility in the reference device. The resulting η_{rat} is shown in Fig. 3.8, demonstrating that the offset in η_{rat} is modified in direct proportion to the mobility reduction, as indicated in the graph, whereas the slope remains unchanged. The effect of a change in mobility on the extracted ΔE_C and ΔV_{TH} can be observed in Fig. 3.9, which shows the conduction band offset and the threshold voltage shift with and without modified mobility. The figure clearly illustrates that the values obtained with the $I_{DS}(T)$ method remain unchanged, while the extracted ΔV_{TH} is sensitive to a t_{Si} dependent change in mobility.

Similarly, the impact of variation of typical (strong) inversion related device parameters has been considered, such as the source/drain series resistance and the oxide thickness. In fact, we have verified that changing the series resistance and oxide thickness indeed does not affect the subthreshold current, as predicted by (3.1) and the corresponding assumption of a long channel with low V_{DS}. In contrast, V_{TH} and ΔV_{TH} differ from the previously obtained values; for example, adding a 2 kΩ lumped element resistor to the S/D contacts of a device with t_{Si} = 5 nm to emulate the S/D series resistance produces a 10%

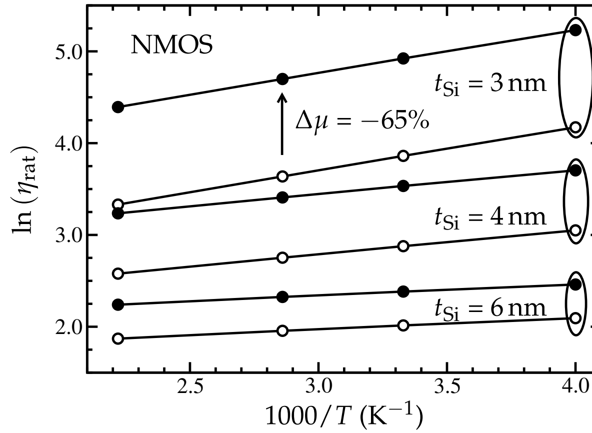


Figure 3.8: Simulated subthreshold current ratio for fixed (open symbols) and modified mobility (filled symbols); L = 1 μm. The mobility was manually reduced by 30% (t_{Si} = 6 nm) to 65% (t_{Si} = 3 nm) relative to the mobility in the reference device [cf. (3.7)], resulting in an increase in η_{rat} of $\ln(1/0.7) \approx 0.36$ (t_{Si} = 6 nm) and $\ln(1/0.35) \approx 1.05$ (t_{Si}= 3 nm).

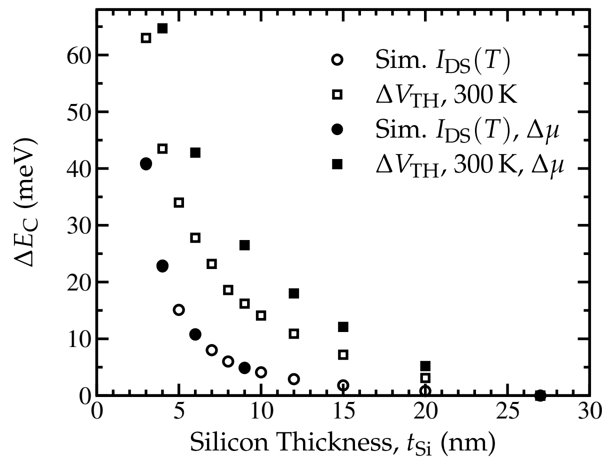


Figure 3.9: Conduction band offset extracted from the subthreshold current ratio in Fig. 3.8 for fixed mobility (open symbols) and modified mobility. The shift in threshold voltage (squared symbols) was found to increase with decreasing mobility, whereas the band offset extracted from $I_{DS}(T)$ remains unchanged.

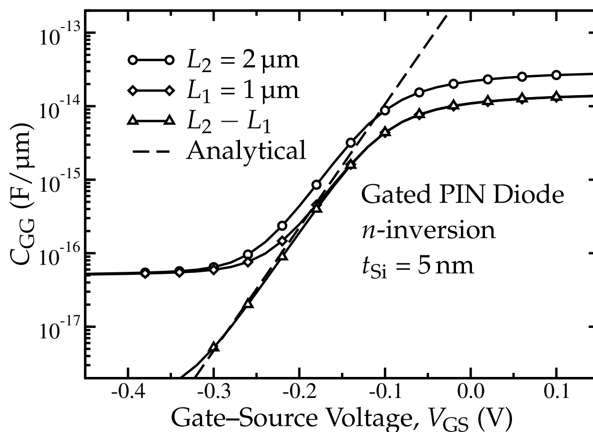


Figure 3.10: Simulated gate capacitance C_{GG} vs. voltage characteristic of a 5 nm symmetric gated PIN–diode, for 1 µm and 2 µm channel length, showing the 'single' capacitance and the normalized differential capacitance [$C_{diff} = (C_2 - C_1)/W(L_2 - L_1)$]. Only the n–type inversion is shown; the dashed line shows the subthreshold gate capacitance calculated with (3.2).

smaller ΔV_{TH} compared to the device without the artificial series resistance.

Separation of $\Delta\mu$ and ΔDOS

Since both the mobility and the DOS determine the prefactor in (3.6), it is not possible to trace a t_{Si} dependent change in the offset of η_{rat} back to either the mobility or the DOS from the *IV* measurements alone. Hence the effect of a change in DOS on I_{DS} would produce results very similar to the simulations with a modified mobility, which were shown previously. If, in addition to the *IV* measurements, Q_i is determined from *CV* characteristics, the individual contribution of the mobility and the DOS can be extracted. To illustrate this procedure, the DOS is manually modified through N_{CB} and N_{VB}, which represent the effective DOS in the conduction and valence band respectively.

The *CV* simulations have been performed on symmetric gated PIN diodes, having a geometry identical to the structure shown in Fig. 3.1(a), only with p^+ source doping. The main merit of a PIN device is that the n-type and p-type inversion can be investigated independently, thereby providing the possibility to extract both the conduction and valence band offsets on a single device. Fig. 3.10 shows a typical n-type inversion capacitance (C_{GG}) characteristic for two devices with 5 nm thick body and 1 µm and 2 µm channel length. The plot

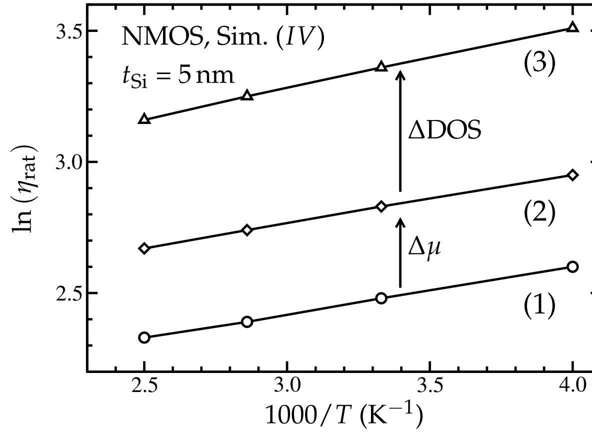


Figure 3.11: Effect of decreasing mobility and DOS on the subthreshold current ratio. Curve (1) shows the original current ratio for t_{Si} 5 nm, with the 27 nm device as reference. Manually reducing the mobility in the t_{Si} 5 nm device by 40% increases the offset in η_{rat} in direct proportion [curve (2)]. Setting, in addition, the DOS to 70% of its original value results in an additional increase in offset [curve (3)].

shows the 'single' capacitance curves, along with the 'differential' capacitance calculated with $C_{diff} = (C_2 - C_1)/W(L_2 - L_1)$. Performing the measurement differentially facilitates a more accurate determination of the subthreshold characteristics, because the capacitance in the low bias regime is significantly reduced. Furthermore, the values calculated with (3.2) are depicted as well, demonstrating good agreement.

Fig. 3.3 shows η_{rat} for an NMOS device with t_{Si} = 5 nm, for three cases: 1) original mobility and original DOS, 2) reduced mobility, original DOS and 3) reduced mobility and reduced DOS. As concluded before, reducing the mobility in the thinnest device increases the offset of η_{rat} in direct proportion. Likewise, a reduction of the DOS, in this case to 0.7× its original value (for both N_{CB} and N_{VB}), gives an additional offset as expected based on (3.6). In fact, the ratio of the DOS and mobility in the reference device and the thin device, hence the prefactor in (3.6), can be found by extrapolating $\ln(\eta_{rat})$ to $1000/T \to 0$, i.e. when the last term in (3.7) reduces to zero. Indeed, the additional increase in offset of $\ln(\eta_{rat})$ is in agreement with the expected value of $\ln(1/0.7)$.

The slight increase in the slope is explained by recalling that the quantum correction potential depends on the carrier distribution; hence, when modi-

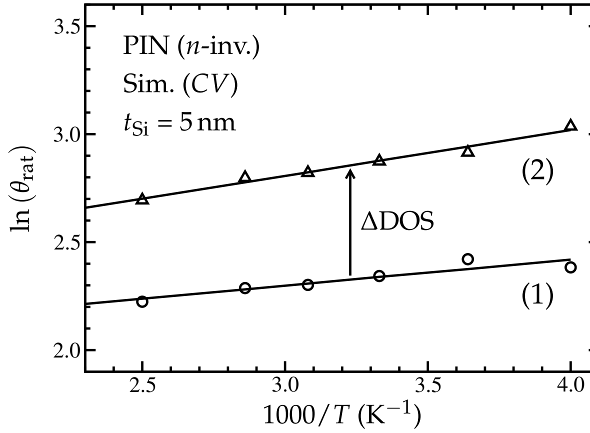


Figure 3.12: Effect of decreasing mobility and DOS on the subthreshold capacitance ratio. The subthreshold capacitance ratio is only sensitive to the DOS reduction, showing a proportionally increasing offset [curve (2)] relative to the curve with unchanged DOS [curve (1)].

fying the DOS, its value relative to the original value (N_{ref}) will appear as an additional contribution to the applied quantum correction potential in (3.11), equal to $k_B T \ln(N_{CB}/N_{ref})$. This directly translates into a slightly steeper slope of η_{rat}, corresponding to an additional band edge shift of $k_B T \ln(N_{CB}/N_{ref}) = k_B T \ln(0.7) \approx 9\,\text{meV}$.

The procedure is repeated for the ratio of the subthreshold capacitance θ_{rat} [cf. (3.8)], the result of which is depicted in Fig. 3.3. The curves with and without modified mobility exactly coincide, thus confirming that the capacitance is insensitive to a mobility variation. Hence, the observed increase in offset of the upper branch [curve (2)] is solely due to the reduction in DOS.

In summary, the shift in conduction and valence band edge can be extracted from the slope of η_{rat}. The offset of η_{rat} is determined by both the mobility and the DOS. In order to disentangle these two quantities, additional subthreshold capacitance measurements can be carried out to extract the change in DOS. A possible t_{Si} dependent change in mobility can then be derived from the remaining difference in offset between the ratio of the currents and capacitances.

3.4 Discussion

The method to extract ΔE_x, mobility and DOS involves a few assumptions. This section covers some sanity checks which are instrumental in determining the applicability of the proposed $I_{DS}(T)$ method. Most importantly, (3.6) assumes that the exponential temperature dependence is much stronger than the temperature dependence of the prefactor. This assumption, however, generally holds under typical operating conditions: although the mobility and DOS do change with temperature, only a *difference* in temperature dependence from layer to layer will introduce errors in the extracted band offsets. The above assumption can be easily verified by assuring that the inverse subthreshold slope varies linearly with temperature, and yields essentially equal values for the t_{Si} range under consideration. The latter is a clear manifestation of volume inversion, which is a typical feature of UTB devices.

The ideality of the subthreshold slope is greatly determined by the concentration of interface states. Hence, one might argue that high–κ gate dielectrics, which generally exhibit a higher interface state density compared to SiO_2 [62], may hamper reliable application of the $I_{DS}(T)$ method. Although the subthreshold slope may deviate from the ideal values (i.e. about 60 mV/decade at 300 K) due to the presence of interface states, sufficient requirements are that 1) the inverse subthreshold slope is approximately equal for the considered t_{Si} range, and 2) that it exhibits a linear temperature dependence. Furthermore, a linearly temperature dependent inverse subthreshold slope implies that for η_{rat} the actual value of the gate bias, provided that $V_{GS} < V_{TH}$, is not important: the difference in V_{GS} at which I_{ref} and I_{thin} are recorded is simply subtracted when converting the slope of η_{rat} to ΔE_x.

The results shown in this work are obtained from long channel (1 µm) devices, with low V_{DS} (25 mV), to ensure that the actual barrier height is determined solely by the gate (bias and workfunction) and the intrinsic body. For short channel devices the barrier is lowered due to the proximity of the source and drain; hence, we verified that the extracted workfunction difference between gate and channel increases, i.e. the source/channel barrier decreases, for smaller L and increasing V_{DS} [not shown].

Summary

In this chapter, shifts in valence and conduction band edge have been extracted from the temperature dependence of the subthreshold current. This method, in the spirit of the procedure for extraction of the activation energy in BJT's and Schottky diodes, is adapted and applied for the first time to quantify shifts in the conduction and valence band edge originating from carrier confinement in UTB SOI devices.

We compared the results with shifts in threshold voltage, showing that, thanks to the temperature dependence of the subthreshold current ($I_{\mathrm{DS}}(T)$), shifts in the band edges can be observed separately from changes in mobility and DOS, which cannot be accomplished with the ΔV_{TH} method. The observed shifts in V_{TH} are generally much ($> 3\times$) larger than the expected band edge shifts, while the band offsets extracted from the $I_{\mathrm{DS}}(T)$ measurements are in very close agreement with the theoretical values. Hence, ΔV_{TH} generally overestimates the quantum confinement induced shift in band alignment.

Furthermore, we pointed out that a change in mobility and DOS can be disentangled with additional temperature dependent subthreshold CV measurements. Since the capacitance does not depend on the mobility, any change in DOS is directly reflected in the subthreshold capacitance, irrespective of a possible change in mobility.

Four

Quasi-Ballistic Transport

In the previous chapters we have investigated the impact of reduction of the semiconductor body thickness. We analyzed the changes in band structure stemming from quantum-mechanical confinement of the carriers in the ultra-thin body (UTB). In particular, Chapter 2 focussed on a systematic comparison of the Full-Band quantization model with the almost universally used Effective Mass Approximation. Then, in Chapter 3, we introduced a method to quantify the shifts in the band alignment through the temperature dependence of the subthreshold current. The approach was subsequently refined using data obtained from Drift-Diffusion simulations, and compared with shifts in the threshold voltage.

In this chapter, we will focus on the impact of *lateral* scaling, that is, reduction of the MOSFET's channel length. In fact, in today's extremely scaled MOSFETs the channel length is in the order of the carrier *mean-free-path* (λ). This means that carriers, when travelling from source towards the drain, experience only few *scattering* events. This regime is commonly referred to as *quasi-ballistic* transport.

If a carrier encounters a scattering event, its direction can change. The carriers that are redirected and actually return to the source, are denoted as 'back-scattered' carriers. As we will see later, the most significant contribution to back-scattering originates from scattering events close to the beginning of the channel. This points to the existence of a *critical length* for back-scattering. Possible definitions of the critical length are addressed in the course of this chapter. At this stage we note that if the critical length is in the order of the mean-free-path, carrier transport can become close to ballistic, even if the channel length is larger than the mean-free-path.

Although we will analyze quasi-ballistic transport in the context of UTB-SOI MOSFETs, quasi-ballistic effects are expected to occur in bulk MOSFETs as well [137]. It is, however, the multiple gate structure that increases the device's

immunity to short-channel effects [19], thus allowing a further reduction of the channel length into the order of the carrier mean-free-path, making quasi-ballistic transport phenomena even more likely.

Outline

Quasi-ballistic (QB) transport is an active area of research since a good understanding of the physics of carrier scattering and the main 'knobs' which impact carrier transport, can help improve and optimize the device performance. Often QB transport models are benchmarked against Monte-Carlo (MC) simulations, which are able to emulate the carrier scattering events and to track the carriers' trajectories as they move along the channel. In fact, the MC approach provides a statistical solution to the Boltzmann Transport equation (BTE). The BTE is a general formulation of the semi-classical carrier transport problem from which approximate transport models such as the Drift-Diffusion (DD) and hydrodynamic equations [138, 139] are derived. An overview of the various transport models and their key assumptions is presented in [140].

The MC approach is instrumental in evaluating transport phenomena under conditions in which the conventional Drift–Diffusion (DD) framework is expected to break down, such as rapid spatial variations in the electric field on a length scale comparable to the carrier mean-free-path. In nanoscale transistors, even in near-equilibrium bias conditions the DD framework may fail due to the large built-in electric fields [141]. Different from the DD approach, which describes carrier transport in terms of 'macroscopic' quantities (such as the mobility), the MC technique can provide a detailed picture of e.g. the contribution of the various scattering mechanisms. Therefore, the MC approach has been extensively used to explain and predict transport properties of advanced bulk and SOI MOSFETs [96, 142–144].

This chapter proceeds with a brief outline of the various scattering mechanisms. In the course of this chapter, we frequently compare analytical model results with MC simulations. In the second part of this chapter we will present a new model for the so-called *backscatter coefficient*, which denotes the fraction of carriers that return to the source after scattering in the channel. Before arriving at that point, we discuss in section 4.2 the typical MC simulator outputs that we will use throughout this chapter.

Then, we will investigate the link between the detailed MC backscatter characteristics and the resulting backscatter coefficient. We already suggested the existence of a critical length for back-scattering. Its definition, however,

is not univocal. With the help of MC simulations, we will characterize and examine the critical length and its dependence on device geometry and bias (section 4.4). In section 4.5, we will use the previous findings to derive a new model for the backscatter coefficient.

4.1 Carrier scattering

Carriers can scatter through a whole variety of mechanisms, the importance of which depends on the operating conditions (e.g. electric field, temperature) and device 'parameters' such as channel doping and uniformity of the Gate/Channel interface. We will briefly highlight the most common scattering mechanisms. For a detailed discussion on carrier scattering, the reader is referred to one of the extensive reviews on this topic, such as [145].

Carriers can scatter through *elastic* and *inelastic* scattering. Elastic means that the total energy of a carrier is conserved during the collision. Hence only carrier's momentum, i.e. the direction of the wave vector, is changed. Inelastic mechanisms, instead, can alter both the carrier momentum and energy.

Furthermore, we can distinguish between *intra-valley* and *inter-valley* transitions, depending on whether the carrier remains in the same valley before and after scattering, or that it changes from one valley to another, respectively. Similarly, we can identify *intra-subband* and *inter-subband* transitions.

Phonon scattering

Carriers can exchange energy with the semiconductor crystal through collisions with lattice vibrations. The corresponding quantized energy is referred to as a *phonon*. Phonons can be categorized into *acoustic* and *optical* phonons, depending on whether the adjacent atoms are displaced in-phase or out-of-phase, respectively.

Intra-valley transitions assisted by acoustic phonons with a small wave vector are typically treated as elastic, since the energy associated with an acoustic phonon usually is much smaller than the thermal energy.[1]

Optical phonon scattering, instead, generally involves energies comparable to the average thermal energy of a carrier, and is therefore treated as inelastic. In fact, carriers can acquire energy from the lattice by *absorption* of an optical phonon. Similarly, carriers can relax their energy by optical phonon *emission*.

[1] Acoustic phonons can, however, facilitate inelastic inter-valley transitions.

Surface roughness scattering

A scattering mechanism that is particularly apparent in UTB devices is *surface roughness scattering*. The thickness fluctuations of the dielectric/semiconductor interface cause carriers to scatter, especially when the carriers are close to the interface due to a strong normal electric field. In contrast to phonon scattering, surface roughness scattering is *anisotropic*, which means that not all the allowed wave vectors after scattering are equally probable. If we recall that carriers are confined in a potential well, formed by the oxide surrounding the UTB, the fluctuations of the oxide thickness can be viewed as local changes in the confining potential. Hence, the impact of surface roughness scattering is usually evaluated from the shape of the wave functions, as explained in [146], and is, besides AP scattering, found to be one of the major sources of carrier scattering in UTB devices.

Ionized impurity scattering

When a carrier approaches an ionized dopant atom, they interact through the Coulomb force, which can either attract or repel the carrier. As a result, the carrier's trajectory will be deflected. The strength of the Coulomb interaction depends on the number of mobile carriers surrounding the ionized impurity; the mobile charge partially cancels, or *screens*, the ionized impurity, thereby reducing the impact of ionized impurity scattering. Thus this mechanism depends on the inversion charge density and its impact decreases with increasing gate bias. Turning to UTB devices, generally the body is lightly doped, or contains no (intentional) doping at all. Hence we do not expect ionized impurity scattering to be significant, at least in the channel of UTB devices.

Other mechanisms

Other scattering mechanisms include carrier-carrier scattering and collisions of a carrier against a 'cloud' (or gas) of electrons, called carrier-plasmon scattering. Furthermore, high-κ dielectrics tend to form dipoles, which can disturb the carriers travelling underneath the gate. Although under active investigation [38, 61, 147–150], discussion of these scattering mechanisms is beyond the scope of this work. Unless stated otherwise, in the following we will mainly focus on acoustic (elastic) and optical (inelastic) phonon scattering since they constitute the dominant isotropic scattering mechanisms in the channel.

4.2 What can we learn from Monte-Carlo simulations?

A detailed introduction to the principles of the MC approach is beyond the scope of this work, even more because the MC procedure is well documented in literature. Extensive reviews on this matter can be found in [83, 145]. Here we will only outline the general procedure and the features which are specific to the simulator used in this work.

MC techniques have been developed to solve many types of integro-differential equations. In short, the principle of the MC approach as applied to carrier transport problems is as follows: after populating the device with carriers, usually based on an initial guess calculated with conventional TCAD (e.g. drift-diffusion) simulations, we inject a randomly selected carrier into the channel. The carrier is selected from a 'reservoir', which acts as an unlimited supply of carriers at thermal equilibrium. In the reservoir the carriers are distributed according to a known initial distribution function $f(x, \mathbf{k}, t)$, which defines the probability of finding a carrier with wave vector $\mathbf{k}$ at position x at time instant t. In the following, we assume (and we always verify) that the carrier transport has come to a stationary situation, so that we can safely disregard the time dependence of the distribution function. Once entered the channel, the carrier gains energy from the electric field and is accelerated during a time frame, called the *free-flight*. Along the transport direction, the carrier is treated as a classical particle, whose kinetics are governed by the classical Newton's laws of motion. This transport regime is referred to as 'semi-classical' transport, which means that motion in the transport direction obeys the classical Newton's laws, whereas in the direction normal to the gates we account for quantum confinement.

After the free-flight, the position of the carrier is 'frozen', and the carrier can scatter according to mechanisms such as the ones discussed previously. Essentially, the whole scattering event consists of changing the carrier's momentum, that is, the direction in which the carrier moves. Furthermore, in case of inelastic scattering the carrier's energy is increased (absorption) or reduced (emission) depending on the energy of the interacting optical phonon. Thus, we assume that scattering itself does not alter the carrier's position. The second assumption is that scattering can be considered as an instantaneous process.

Then, after a possible scattering event, the free-flight is resumed and the process continues until the carrier eventually exits the device at either the source or the drain contact. We see that the path the carrier follows when trav-

elling from source to drain, or back to the source in case of a back-scattering event, is determined by the scattering mechanisms and by the electric field. Thus, the true challenge in the MC technique is a proper description of the scattering rates, which govern the statistics (hence the name 'Monte Carlo'). By following an ensemble of carriers, we obtain a statistically significant picture of the collective behavior and the resulting charge and current distribution in the device.

The MC simulator [151] used in this work is developed at the University of Udine, Italy. It accounts for quantization along the direction normal to the gates (z direction, cf. Fig. 2.1(b) on p. 19). However, rather than applying a quantum correction to the electrostatic potential as conventionally done in many MC simulators, it evaluates the subbands and wave functions by solving the Schrödinger equation in the Effective Mass Approximation (EMA). To this end, the device is split into sections along the transport direction. In each section, the 1D Schrödinger equation [SEQ, see also (2.12) on p. 24] is solved along the z direction. The resulting subband profile, enables the evaluation of the driving force for the carrier motion, i.e. the gradient of the subband profile along the transport direction. The scattering rates are calculated from the subband energies and the wave functions, hence this approach is appropriately termed 'Multi-Subband Monte-Carlo' (MSMC). After computation of the scattering rates, the actual MC procedure consists of calculating the motion of the carriers. As a result, the carriers are redistributed throughout the channel which requires the potential profile to be updated by solving the Poisson equation. Thus, the self-consistent iteration between SEQ, Poisson's equation and the MC routine is performed until a steady state current is achieved, that is, until a spatially constant (up to the desired accuracy) net current in the device is obtained.

Typical Monte-Carlo outputs

The following figures illustrate typical outputs that we can obtain from the MC simulations. At this stage, we will not elaborate on the quantitative features, we merely exemplify the general picture. A typical merit of MC simulators is the ability to evaluate in detail the carrier distribution $f(\mathbf{r}, \mathbf{k})$ at any position $\mathbf{r} = (x, y)$ in the device. In particular, we can plot the number of carriers having a certain velocity. The theoretical Maxwell-Boltzmann longitudinal

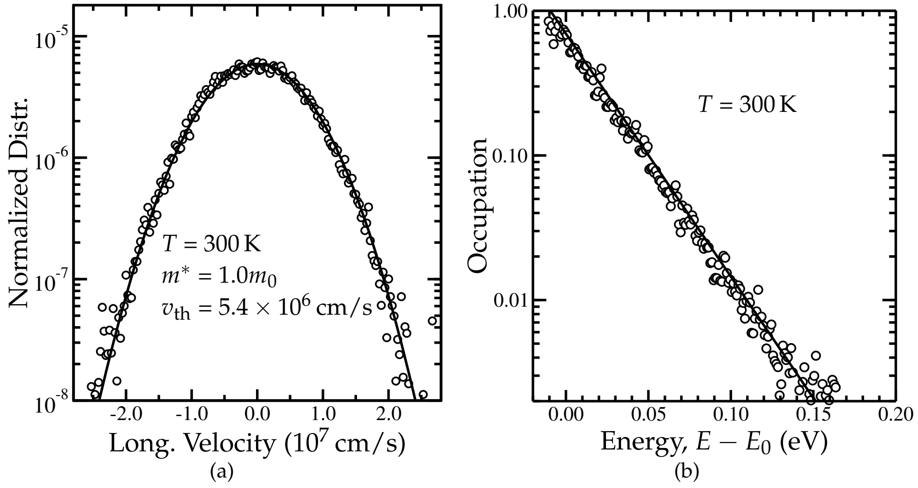


Figure 4.1: (a) Example of the electron velocity distribution along the transport direction at thermal equilibrium in the source. (b) Occupation probability as a function of the total energy, relative to lowest subband minimum E_0. The symbols are the MSMC results, whereas the solid lines are the theoretical values from (4.1).

velocity distribution reads [145]

$$f(0, v_x) = N_0 \sqrt{\frac{m^*}{2\pi k_B T}} \exp\left(-\frac{m^* v_x^2}{2 k_B T}\right) \tag{4.1}$$

The velocity distribution from the MSMC is shown in Fig. 4.1(a), along with the values from (4.1), normalized on the carrier concentration N_0. The distribution is recorded close to the source contact, where near-equilibrium conditions apply. In fact, the frequent collisions in the source and drain randomize the carrier's momentum and carriers lose energy during inelastic scattering events (*thermalization*). As a result, the average carrier 'temperature' equals the lattice temperature, and the carriers are in (thermal) equilibrium with the lattice. Indeed, in Fig. 4.1(a) we observe that the positive and negative velocity distributions are similar, and exhibit a Maxwellian shape. Furthermore, the distribution is centered around zero, which implies that the average velocity is close to zero. Besides the velocity distribution, we can record the carrier's total energy. As an example, we plot in Fig. 4.1(b) the occupation probability of a given energy level. The solid line is the theoretical Maxwell-Boltzmann occupation function at T = 300 K. The occupation decreases in the expected exponential fashion as we move up in energy.

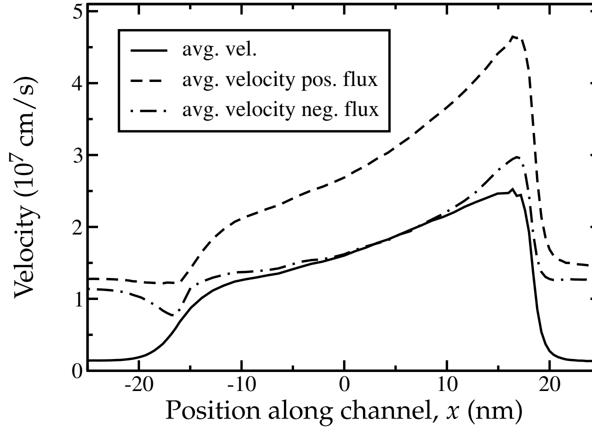


Figure 4.2: Average velocity of the negatively and positively directed flux, along with the average drift velocity, which is the average velocity of the entire distribution. The gate length is 32 nm, centered around $x = 0$. The direction is defined positive from source to drain.

At any position along x we calculate the unidirectional average velocity, that is, the average velocity of each half of the distribution [such as depicted in Fig. 4.1(a)]. The resulting values for each position along the channel direction are shown in Fig. 4.2. The average velocity of the positive flux in the source and drain nearly equals the thermal velocity (around $1.2 \cdot 10^7$ cm/s using Si parameters). The average velocity of negative flux slightly decreases towards the entrance of the channel due to the contribution of the low-velocity carriers from the positive flux, which are reflected by the source-channel barrier. The velocity of the carriers increases as they get accelerated by the electric field, with a peak at the drain end of the channel. The drift velocity, i.e. the net velocity of the entire carrier population (solid line), is low in the source and drain, then approaches the thermal velocity around the entrance of the channel, and gradually increases along the channel. Recalling that the net current should be constant throughout the device (assuming no gate current and/or recombination), we find that the large number of carriers in the source and drain have a low average velocity (the electric field in the source and drain is low), while in the channel the smaller number of carriers have to carry the same current. As a consequence, their velocity is higher.

Fig. 4.3 reports the net current density, as well as the current associated with the positively and negatively directed fluxes for a typical simulation.

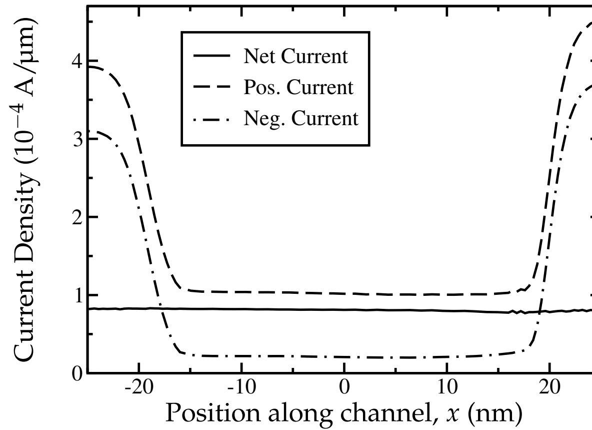


Figure 4.3: Positional dependence of the current density (absolute values); same device and bias as in Fig. 4.2.

Note that we can freely exchange 'flux' and 'current', provided we account for the charge. Although the net current is constant, as it should be, the individual current components greatly increase in the source and drain. This can be explained by noting that in those regions a large number of carriers is moving and frequently encounters scattering events, without being able to cross the source-channel or drain-channel potential barrier. Hence, although the individual carriers move rapidly and frequently change direction due to scattering, the average velocity of the whole ensemble of carriers (the drift velocity) is low [cf. Fig. 4.2].

In this chapter, we will use the MC simulations to improve our understanding of back-scattering in the channel. The MSMC simulator is able to trace the scattering events which actually resulted in a back-scattered carrier. This is exemplified in Fig. 4.4(a), showing the distribution of back-scatter events along the channel, which is obtained by recording for each carrier that enters the source from the channel, the position of its last back-scattering event. Since the distribution comprises all carriers that re-enter the source from the channel (in a given time frame), integration of the distribution along the channel yields the backscatter coefficient r, after normalization on the incoming positive flux.

The distribution of backscatter events can be split into the relative contribution of the different scattering mechanisms that have resulted in a back-

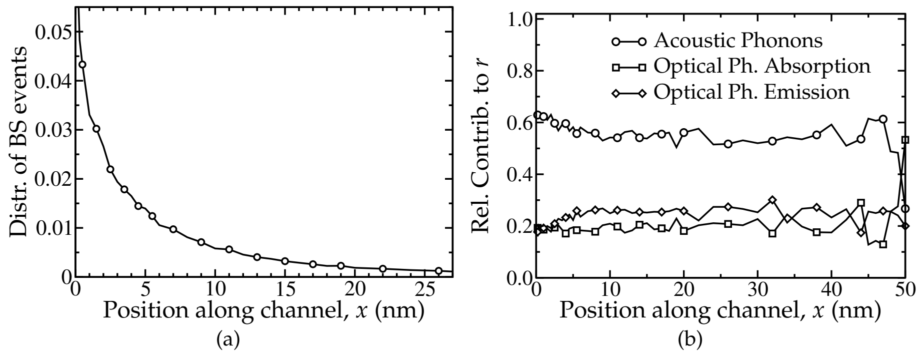


Figure 4.4: Example distribution of back-scatter events contributing to r, obtained by recording, for each carrier that enters the source from the channel, the position of its last back-scattering event (a). The relative contribution of the various scattering mechanisms to the distribution in (a) is shown in (b).

scattered carrier, as shown in Fig. 4.4(b). Thus, we can find the scattering mechanism that predominantly contributes to r at any position along the channel.

4.3 Modelling quasi-ballistic transport

Now that we have an idea of the information we can retrieve from the MC simulations, we will discuss here the most common approach to analyze Quasi-Ballistic (QB) transport, and present an overview of the assumptions involved.

The QB regime of transport is often analyzed with the so-called Flux theory. As the name suggests, the device characteristics are described in terms of carrier fluxes, which are partially (quasi-ballistic) or fully transmitted (ballistic) through the channel of a semiconductor device. As mentioned before, the use of DD-based models is justified only if the electric field varies on a scale much larger than the carrier mean-free-path. The Flux theory does not suffer from this limitation, making it a very promising framework to analyze carrier transport even in nanoscale devices, where the carriers experience strong potential variations over only a small distance.

The Flux method was originally proposed by McKelvey [152]. Nearly three decades later, McKelvey's Flux method was generalized by Das and Lundstrom [153] (termed 'Scattering Matrix Approach'), inspired by Datta [154] who described quantum transport in terms of fluxes related by scattering matrices. The flux method was subsequently adopted to model carrier transport

in heterojunction bipolar transistors [155] and, in 1997, applied to examine the QB transport regime in MOSFETs [156, 157]. In the latter publications, the authors relate the fluxes to external quantities, such as the device geometry and terminal bias. This procedure involves a set of assumptions, resulting in the Virtual Source Model, or often simply referred to as the 'Flux Model'.

Since then, the Virtual Source (VS) model was shown to provide an intuitive and useful perspective to analyze and explain QB transport phenomena. Even more, because the conventional Drift-Diffusion (DD) equations can be expressed in terms of carrier fluxes, thereby linking QB transport to the well-established DD regime [145].

Although we have touched upon the concept of the carrier fluxes when introducing the MC simulations (on p. 79), we will discuss the Flux theory in more detail here, followed by the key assumptions which are specific to the VS model.

Flux theory

Central to the Flux theory is the decomposition of the net current into a positively and a negatively directed flux of carriers, as schematically depicted in Fig. 4.5(a). The source and the drain act as ideal thermal reservoirs, which emit a flux of carriers at thermal equilibrium into the channel. The direction is defined positive from source to drain, hence the source emits a flux of positive-velocity carriers (with magnitude I_S^+), and the drain only negative-velocity carriers (I_D^-). The positive and negative travelling fluxes ($I^+(x)$ and $I^-(x)$, resp.) may be partially reflected due to scattering in the channel. The net current equals $I_{DS} = I^+(x) - I^-(x)$ which, imposed by current continuity, holds at any position along the channel (assuming no gate current and/or recombination).

The 'snapshots' of the carrier velocity distributions [cf. Fig. 4.1(a)] are readily converted to a positive and negative going flux at any position along the channel direction.

As mentioned before, the fluxes may be reflected in part due to scattering events in the channel. We can characterize the channel by a *backscatter* coefficient r, which we define as the ratio of the negative flux due to back-scattered carriers and the positive directed flux at the position $x = 0$ (to be specified later). Similarly we can define a transmission coefficient t, which is its complement to one, i.e. $t = 1 - r$.

The total negative flux at $x = 0$, I_0^-, consists of the backscattered carriers

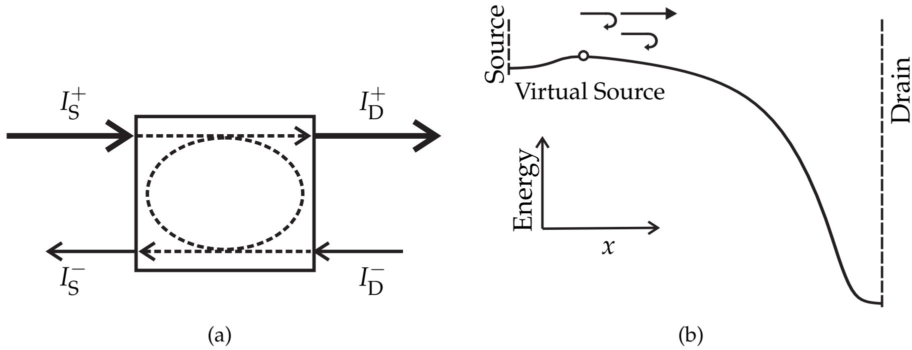


Figure 4.5: (a) Schematic representation of reflection and transmission of the incoming (I_S^+, I_D^-) and outgoing (I_S^-, I_D^+) fluxes. (b) Sketch of a typical energy band profile, illustrating the scattering events in the channel and the position of the virtual source (VS).

originating from the source rI_0^+, and a contribution due to the transmitted carriers stemming from the drain. Thus, we find the following implicit equation for r

$$rI_0^+ = I_0^- - I_L^- (1 - r) \tag{4.2}$$

where we assume r to be direction independent. Although this assumption holds only in the zero field limit, it is not a serious concern in the cases discussed in this work, as we will point out later.

We can consider r as a device figure of merit, since it represents the 'transparency' of the channel to carriers, and tells us how close to the upper performance limit ($r = 0$), i.e. the *ballistic* limit [158, 159], a device operates.

Virtual Source model

Now we will discuss the Virtual Source (VS) model, which provides an intuitive conceptual picture of carrier transport in the QB regime. The original work on the VS model can be found in [145, 156, 157]. In the following, we will summarize the main equations and discuss the underlying assumptions along the way. The VS model is based on the foundations of the Flux theory, and evaluates the transmission coefficients which relate the incoming and outgoing fluxes in a (nanoscale) MOSFET to the external bias. We start from an –again schematic, yet more realistic– energy band profile along the lateral di-

rection of a MOSFET, depicted in Fig. 4.5(b). With Virtual Source (VS), we refer to the top of the source-channel barrier.

The key concept of the VS model is the notion that carrier transport in the QB regime is essentially determined by scattering within only a small distance from the (virtual) source. We will see that scattering affects the average velocity of carriers near the VS. In particular, it is now generally understood that it is the *injection* velocity of the carriers, rather than the *saturation* velocity at the drain end of the channel, that limits the semi-classical current in nanoscale devices. The idea that the device current is limited by injection at the source side of the channel, rather than being determined by the entire channel, was first pointed out in [158]. Yet, to some extent carrier scattering at any position in the channel does contribute to r [160–162]. Even close to the drain, scattering can play either a direct role, by sending (hot) carriers across the barrier back to the source, or indirectly through the electrostatic self-consistency, that is, the coupling between charge and potential: scattering near the drain increases the local carrier density, which alters (flattens) the potential profile around the VS, leading to increased back-scattering near the VS. The impact of self-consistency on back-scattering depends on the channel length, since in shorter channels scattering will contribute to r throughout a larger portion of the channel. We will come back to this in section 4.4, where we discuss the so-called *critical length* for back-scattering in more detail.

Secondly, in the VS model both the positive and negative flux at the VS are assumed to be close to equilibrium. To be more specific, each half of the carrier distribution is assumed to have a near-equilibrium (Maxwellian) shape, on the following grounds: the average field close to the top of the barrier is quite low, so that in the small region around the VS which predominantly determines the device current, the carriers are not significantly heated, and near-equilibrium conditions prevail [157]. This implies that the reduction in ON-current compared to the fully ballistic current is determined by the flux of back-scattered carriers 'diffusing' against the field across a small region near the top of the barrier.

In summary, the VS model (often called 'Flux model' since it relies on the Flux theory) considers carrier transport from the VS onwards, so that the VS is the point at which the carriers are injected into the channel. The underlying assumption is that the fluxes around the VS are close to thermal equilibrium, just as in the 'real' source. In the following we set the origin of the x-axis at the VS, i.e. $x_{\mathrm{VS}} = 0$, as indicated in Fig. 4.5(b).

We note in passing that in an actual MOSFET under typical I_{ON} conditions

(high V_{DS}, $V_{\mathrm{GS}} - V_{\mathrm{TH}} > 0$), the flux from the drain can safely be neglected, since those carriers have to move 'upstream' against the field, towards the source. Consequently, in high V_{DS} conditions, the fraction of carriers from the drain that actually reaches the source is very small. This implies that essentially all carriers that pass the VS with a negative velocity are back-scattered carriers originating from the flux injected at the source. From (4.2) we find that, in absence of injection from the drain, r is simply given by $r = I_0^- / I_0^+$. Thus, the previously mentioned direction dependence of r turns out to be irrelevant under high drain bias.

Keeping in mind the above considerations, we now relate the device current to the backscatter coefficient. The drain-source current I_{DS} per unit device width can be expressed in terms of the mobile carrier concentration $n(x)$ and the average carrier velocity $v(x)$. We write

$$I_{\mathrm{DS}} = qn(x)v(x) \tag{4.3a}$$

$$\approx C_{\mathrm{eff}} \left(V_{\mathrm{GS}} - V_{\mathrm{TH}} \right) \nu_{\mathrm{inj}} \tag{4.3b}$$

(4.3a) holds at any position in the channel (current continuity), while (4.3b) assumes above-threshold and saturation conditions. Similar expressions have been proposed for the linear regime [163]. The inversion charge evaluated at the VS is $C_{\mathrm{eff}} \left(V_{\mathrm{GS}} - V_{\mathrm{TH}} \right)$, where C_{eff} is the effective gate capacitance which includes poly-gate depletion and repulsion of carriers from the interface due to QM effects (assuming no significant impact of short channel effects).

The *injection velocity* ν_{inj}, that is, the velocity of the net flux at the VS is related to the average velocity of the positive directed carriers (v^+), given by [157]

$$\nu_{\mathrm{inj}} \approx \frac{1 - r_{\mathrm{flux}}}{1 + r_{\mathrm{flux}}} v^+ \tag{4.4}$$

with r_{flux} the backscatter coefficient according to the VS model. As mentioned before, (4.4) is based on the assumption that near the VS both the positive and negative flux are close to thermal equilibrium, which implies that v^+ and v^- are close to the thermal velocity (v_{th}). Assuming a Maxwellian distribution [cf. Fig. 4.1(a)], $v^+ \approx v_{\mathrm{th}} = \sqrt{2k_{\mathrm{B}}T/\pi m^*}$, with m^* the effective mass in the transport direction. Thus, in the fully ballistic case ($r_{\mathrm{flux}} = 0$) the injection velocity is clipped to the thermal velocity, in agreement with [158]. Incidentally, in silicon the thermal velocity is very close to the saturation velocity ($1.2 \cdot 10^7$ cm/s and $1 \cdot 10^7$ cm/s respectively), which may explain why the conventional 'velocity saturation' models have been able to predict the drive current in Si based devices even beyond their expected range of validity [156].

In the low-field limit ($\mathcal{E} \to 0$), r_{flux} was evaluated from the Flux theory as

$$r_{\text{flux}} = \frac{L}{L + \lambda_{\text{flux}}} \qquad (\mathcal{E} \to 0) \tag{4.5}$$

with L the channel length. The low-field carrier mean-free-path λ_{flux} is an important characteristic length denoting the average distance a carrier can travel without scattering. Assuming near-equilibrium conditions, λ_{flux} is expressed as [152, 157]

$$\lambda_{\text{flux}} = \frac{k_{\text{B}}T}{q}\frac{2\mu}{v_{\text{th}}} \tag{4.6}$$

where μ is the low-field mobility. Eq. (4.5) implies that in the low-field limit, the entire channel contributes equally to back-scattering.

For a non-zero electric field, r_{flux} was proposed as

$$r_{\text{flux}} = \frac{L_{\text{kT}}}{L_{\text{kT}} + \lambda_{\text{flux}}} \qquad (\mathcal{E} > 0) \tag{4.7}$$

with $L_{\text{kT}} = k_{\text{B}}T/q|\mathcal{E}|$ the distance over which the potential drops by $k_{\text{B}}T/q$, assuming a linear potential profile. Thus, in (4.7), the critical length is defined as the spatial extent of L_{kT}, labeled the "$k_{\text{B}}T$-layer" [164], from the VS into the channel.

Initially, in [157], the relation between r_{flux} and L_{kT} was introduced by reasoning that, in near-equilibrium conditions, the average carrier energy in a 2D gas (as happens in an inversion layer) equals $k_{\text{B}}T$. Hence when a carrier is L_{kT} away from the top of the barrier, the probability that it has sufficient longitudinal energy to surmount the barrier, and to reach the source, decreases rapidly. Consequently, the distribution of backscatter events would be strongly weighted towards the VS. However, in [165], it was shown that the "$k_{\text{B}}T$-layer concept" is a direct consequence of the assumed Maxwellian distribution, rather than being related to the average carrier energy. In [165] Clerc and coworkers derived the following expression for r

$$r_{\text{flux}} = \frac{L_{\text{kT}}\,(1-\beta)}{L_{\text{kT}}\,(1-\beta) + \lambda_{\text{flux}}} \tag{4.8}$$

with $\beta = \exp\left(-L/L_{\text{kT}}\right)$; (4.8) provides a continuous description from the low-field (4.5) to the high-field limit (4.7). Thus (4.8) points out that, due to the direct link between λ_{flux} and the mobility, high-mobility materials are expected to exhibit a higher ballisticity.

The VS model has been extensively used to understand and predict QB transport phenomena in nanoscale MOSFETs, mainly owing to its intuitive

and simple conceptual picture. However, the VS model has a few limitations, which force us to interpret the quantitative predictions with care [137, 166]. The first limitation concerns the existence and a proper definition of "the" critical length for back-scattering, which is still subject of on-going debate. The second limitation is related to the requirement that both fluxes be close to equilibrium around the VS. While this seems reasonable for the positive directed flux, at least at the VS, it may not hold for the back-scattered flux: highly energetic carriers that scatter near the drain may result in a strongly off-equilibrium ('non-thermal') distribution of back-scattered carriers. This effect becomes progressively more pronounced as carrier transport gets closer to the ballistic limit, when scattering events occur less and less frequently. In that case, also the definition of the mean-free-path should be reconsidered.

The first topic will be addressed in section 4.4, where we will discuss possible metrics to characterize back-scattering and quantitatively study the critical length using MSMC simulations. Then, in section 4.5 a new model for r will be introduced, which puts no constraints on the shape of the backscattered distribution and is valid also close to the ballistic limit.

4.4 The critical length for back-scattering

As shown by MC simulations [151, 162], scattering events near the top of the source-channel barrier are the most effective in sending carriers back to the source, thus controlling the device current. This suggests the existence of a *critical length* over which scattering significantly influences backscattering. In the following, we will use a rather general definition of 'the' critical length, namely the distance from the VS in which scattering is most effective in contributing to the backscatter coefficient r, and we will refer to it as L_{crit}. The scope of this section is to investigate, using detailed MC simulations, the link between L_{crit} and various other characteristic lengths reported in literature.

As we have seen before [cf. (4.8)] L_{crit} has been related to L_{kT}, that is, the spatial extent of the $k_{\text{B}}T$ potential energy drop from the VS [157]. In particular, in [165] it was shown by solving the BTE in the relaxation length approximation and assuming near-equilibrium (thermal) fluxes, that the critical length for backscattering is tightly related to L_{kT}. Furthermore, it was found that the origin of the L_{kT} dependence lies in the assumed Maxwellian carrier distribution.

Yet, the critical length for back-scattering remains subject of on-going debate. For example, it is assumed in [167–169] that L_{crit} is determined by the

optical phonon energy, by reasoning that carriers close to the injection point encounter elastic scattering, until they have acquired sufficient energy from the electric field to scatter inelastically by emitting an optical phonon. The latter decreases the carrier energy, thus the probability of returning to the VS.

Besides the critical length, the determination of the carrier mean-free-path to be used in a 'flux-like' equation such as (4.8) is still discussed [166,170,171]. In (4.8), the definition of λ_{flux} assumes low-field conditions, that is, diffusive transport near the top of the barrier. Close to the ballistic limit, however, this assumption may become questionable and may force us to reconsider the definition of the mean-free-path.

In this section we examine with Monte-Carlo simulations the critical length for back-scattering and point out whether or not it is related to optical phonon (OP) scattering. In passing, we will comment on the appropriate mean-free-path to be used under close-to-ballistic conditions. We first introduce the relevant metrics that we use to characterize back-scattering in the channel.

Characterization of Back-Scattering

The Multi-Subband Monte-Carlo (MSMC) simulator is able to record the distribution of backscatter events, as shown previously in Fig. 4.4(a). In particular, at a given position, say x_{scatt}, we can record the distribution of backscatter events in two ways: we can trace for each carrier that passes x_{scatt} with negative velocity, 1) the position where the carrier encountered the last scattering event that changed the carrier's momentum from positive to negative, or 2) the maximum position the carrier has reached in the channel (the 'projected range'). These two strategies do not necessarily yield the same results: if a carrier changes direction from positive to negative a few times, the maximum position in the channel may not coincide with the position of the last momentum change. Unless stated otherwise, we will adopt the first strategy, and label the resulting distribution of backscatter events $g(x)$. By the same token, $g(x)$ is recorded at the virtual source ($x_{\text{scatt}} = x_{\text{VS}} = 0$).

Since $g(x)$ comprises all carriers that have back-scattered to the VS, it may be viewed as the 'differential' backscatter coefficient along the channel, i.e. dr/dx. Thus, the backscatter coefficient r is readily obtained by integrating $g(x)$ along x. Alternatively, we can calculate r from the ratio of negative and positive flux at the virtual source [cf. (4.2)]. However, we should keep in mind that the negative directed flux may also include a contribution due to the negative velocity carriers injected from the drain. Hence, both methods of obtain-

Label	Definition	Eq.	Units
L_{crit}	General label		nm
L_{kT}	$k_B T/q\lvert\mathcal{E}\rvert$	(4.7)	nm
L_{eff}	$L_{\mathrm{kT}}[1-\exp(-L/L_{\mathrm{kT}})]$	(4.8)	nm
L_{extr}	$\lambda \cdot r/(1-r)$		nm
$g(x)$	Distr. of backscatter events		nm^{-1}
x_{c}	Centroid of $g(x)$	(4.9)	nm
$L_{50\%}$	Cumulative backscatt. coeff. is 0.5	(4.10)	nm

Table 4.1: List of the characteristic lengths and quantities considered in this chapter.

ing r will yield identical results (up to numerical accuracy) only if injection from the drain can be neglected, for example under high V_{DS} bias.

In addition to the backscatter coefficient, i.e. the area enclosed by $g(x)$, we can characterize the shape of $g(x)$ by calculating the *centroid* x_{c}, defined as

$$x_{\mathrm{c}} \equiv \frac{\int_0^L x g(x)\, dx}{\int_0^L g(x)\, dx} \tag{4.9}$$

where we observe that the denominator is just r. The basic property of the centroid is that it is sensitive to changes in the shape of $g(x)$, whereas any change in proportionality simply drops out. We will compare the values of x_{c} with another 'characteristic' length, $L_{50\%}$, defined as the distance at which 50% of the backscatter events have contributed to r or, equivalently, the cumulative backscatter coefficient equals 0.5. We define $L_{50\%}$ as

$$\frac{\int_0^{L_{50\%}} g(x)\, dx}{\int_0^L g(x)\, dx} = 0.5 \tag{4.10}$$

In what follows, x_{c} and $L_{50\%}$ are relative to the position of the VS. As mentioned before, we can plot the relative contribution of each scattering mechanism to $g(x)$. In the following, the scattering mechanisms active in the channel are: acoustic phonons (AP), optical phonon (OP) emission and absorption, and surface roughness (SR) scattering.

In addition, we introduce two alternative definitions of the critical length: 1) L_{extr}, the critical length which we obtain by inverting the flux-like formula, i.e. $L_{\mathrm{extr}} = \lambda \cdot r/(1-r)$, with r the MSMC values. We will comment

on the appropriate low-field mean-free-path λ in the course of this chapter (on p. 100); and 2) L_{eff}, which is the critical length in the flux formula (4.8), $L_{\mathrm{eff}} = L_{\mathrm{kT}}[1 - \exp(-L/L_{\mathrm{kT}})]$.
The various characteristic lengths considered in this work are listed in Tab. 4.1.

Simulation Setup

Details of the MSMC simulator used in this work along with a discussion on the model verification and calibration are documented in [151]. In short, the simulator accounts for quantization by self-consistently solving the effective mass Schrödinger equation with non-parabolic bands, in each section of the device, coupled to the 2D Poisson and the multi-subband Boltzmann Transport equation (BTE).

In the following, we will consider two types of devices: 1) a 32 nm Fully-Depleted Single-Gate SOI MOSFET (FDSOI) with realistic doping profiles and SOI thickness of 7 nm [172], and 2) a frozen (i.e. non self-consistent) linear or parabolic profile with template scattering parameters, following [171]. Unless stated otherwise, we assume non-degenerate (Maxwell-Boltzmann) statistics to be consistent with the VS model. In case of the 32 nm FDSOI device, we will also examine the impact of optical phonons on the backscatter characteristics by changing the OP scattering parameters. In particular, we change the OP energy $\hbar\omega_{\mathrm{OP}}$ and the OP deformation potential D_{OP}. They jointly enter the prefactor of the OP scattering rate f_{OP} as [145, p. 82]

$$f_{\mathrm{OP}} \propto \frac{D_{\mathrm{OP}}^2}{\omega_{\mathrm{OP}}} \tag{4.11}$$

hence they modulate the mobility.

Back-scattering in an UTB-SOI MOSFET

First we examine back-scattering in the FDSOI device. The simulation parameters are consistent with the universal mobility curve [173]. The distribution of backscatter events $g(x)$ in the FDSOI device is depicted in Fig. 4.6 for T ranging from 100–400 K. We observe that back-scattering increases with temperature. The inset shows the relative contribution of the scattering mechanisms, demonstrating that back-scattering is mainly due to acoustic phonon and surface roughness scattering. From Fig. 4.6, we extracted the various 'characteristic lengths' as defined previously, i.e. x_{c} and $L_{50\%}$ (see also Tab. 4.1), shown

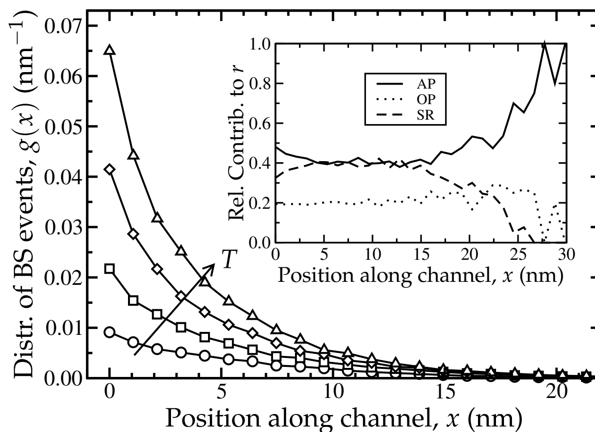


Figure 4.6: $g(x)$ in the 32 nm FDSOI device [172] for T ranging from 100 K to 400 K. V_{GS} = 0.9 V and V_{DS} = 1.0 V. Most backscatter events occur close to the VS (x = 0). The inset shows the relative contribution of acoustic and optical phonons (AP and OP resp.) and surface roughness (SR) scattering to the backscattered carrier distribution (T = 300 K).

in Fig. 4.7. We also plot L_{kT}, extracted as the length corresponding to a voltage drop of $k_B T/q$ from the VS (the top of the lowest subband, x = 0).

Interestingly, L_{kT} varies linearly with temperature (4×) pointing to an essentially linear potential profile close to the VS. The behavior of x_c, instead, is markedly different: it slightly moves towards the VS, with a total shift of approximately 1 nm over the entire temperature range. Furthermore, Fig. 4.7 shows that $L_{50\%}$ and x_c exhibit the same trend: although $L_{50\%}$ is located 1 nm closer to the VS than x_c, the relative shift is very similar. Likewise, in Fig. 4.8 we plot the same quantities for different values of the gate bias V_{GS}, ranging from just above threshold up to the strong inversion regime. Again, the behavior of L_{kT} is quite different compared to x_c and $L_{50\%}$, which are rather insensitive to V_{GS}. For V_{GS} above 0.6 V ($V_{GS} - V_{TH} > 0.15$ V), L_{kT} decreases due to the increasing field near the top of the barrier, which we conclude by inspection of the lowest subband profile as shown in Fig. 4.9. We recall that x_c is sensitive to the shape of the distribution of backscatter events. Thus, x_c is proportional to the portion of the channel over which scattering mostly contributes to r, regardless of the scattering in absolute terms. In effect, x_c is an appropriate metric for L_{crit}. This statement is supported by the similarity of the trends in $L_{50\%}$ and x_c.

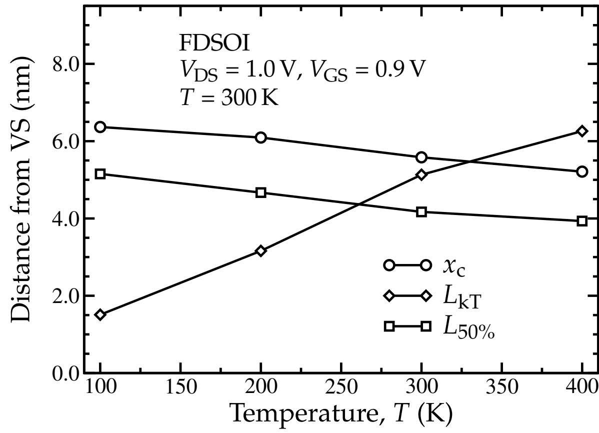


Figure 4.7: x_c, $L_{50\%}$ and L_{kT} vs. T. L_{kT} is extracted from the lowest subband profile, and corresponds to a voltage drop of $k_B T/q$ (the thermal voltage) from the VS. All quantities are referenced to the position of the VS.

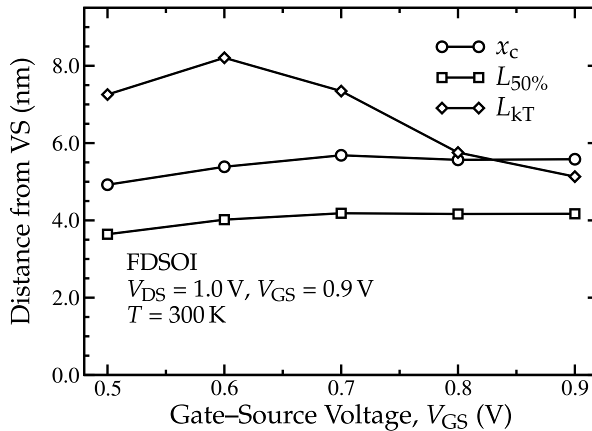


Figure 4.8: x_c, $L_{50\%}$ and L_{kT} versus V_{GS}, which ranges from just above threshold ($V_{TH} \simeq$ 0.45 V) up to strong inversion.

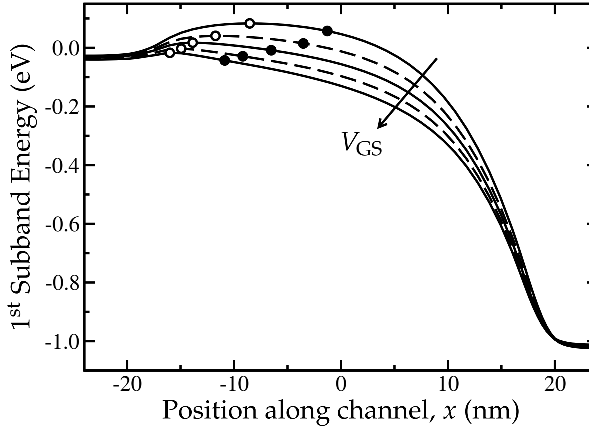


Figure 4.9: Variation of the lowest subband profile with V_{GS}, corresponding to the V_{GS} values of Fig. 4.8. The open circles indicate the position of the VS; the closed circles denote the distance L_{kT} away from the VS.

Impact of Optical Phonons

Through inelastic scattering, such as due to optical phonons, the carrier energy is increased (absorption) or decreased (emission) during scattering. OP emission is particularly effective near the drain end of the channel, where carriers may have acquired significant energy from the electric field. By emitting an optical phonon, carriers relax energy which, in turn, reduces the carrier's probability of returning to the source. In this respect, OP scattering may even reduce back-scattering in the channel [167, 174].

Here, we will focus on the impact of optical phonons on *back*-scattering. Although it is a well-established fact that OP scattering is an efficient route to energy relaxation ('thermalization'), it remains an open issue whether the critical length for back-scattering is directly related to the OP energy, the so-called $\hbar\omega_{OP}$-layer, as suggested by Natori [168, 169]. In the MSMC simulator we can easily change the OP scattering parameters to assess the impact of OP scattering on the backscatter characteristics such as r and x_c. The prefactor of OP scattering, which together with e.g. the Density of States determines the OP scattering rate, depends quadratically on the deformation potential D_{OP}, and is inversely proportional to the OP energy ω_{OP} as indicated by (4.11).

In Figs. 4.10 and 4.12 we plot respectively x_c and r, as function of the temperature. Besides the simulations employing the standard parameters [cf. x_c

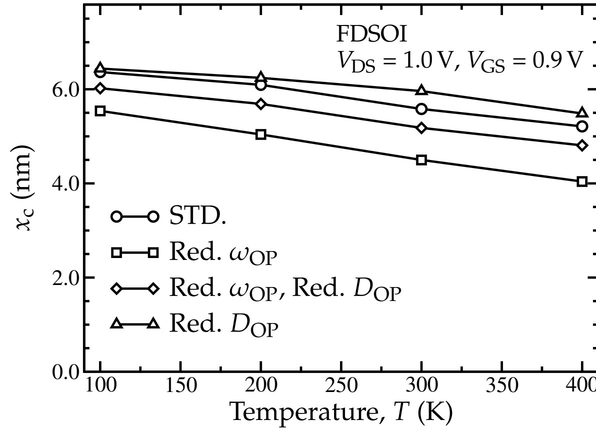


Figure 4.10: x_c in the FDSOI device, as function of T; 'STD.' refers to the standard scattering parameters, while the other curves show the MSMC results with modified OP coupling constant D_{OP} and/or energy ω_{OP}, as explained in the text.

in Fig. 4.7], labeled 'STD.', we show the results for a) D_{OP} reduced by $\sqrt{2}$, b) ω_{OP} reduced by factor of 2, and c) a combination of the aforementioned modifications. The last case results in the prefactor for OP scattering being equal to the 'STD.' case, the relevance of which will be explained in the course of this section. In all cases but the 'STD.' parameter set, we only changed D_{OP} and/or ω_{OP} of the g-type optical phonons with $\hbar\omega_{OP} = 62\,\mathrm{meV}$ and $D_{OP} = 1.1 \cdot 10^9\,\mathrm{eV/cm}$, which are the dominant inelastic scattering mechanisms [83]. The increased relative contribution of OP scattering is at least 40–50%, as shown in Fig. 4.11.

Starting with Fig. 4.10, we see that reducing only D_{OP} has negligible impact on x_c. This is explained by recalling that in the 'STD' simulation the relative contribution of OP scattering to r was already small, as shown in the inset of Fig. 4.6, hence the effect of an even further reduction of OP scattering through a decrease in D_{OP} is negligible.

If ω_{OP} is reduced, instead, the OP scattering rate increases [cf. (4.11)]. Furthermore, if back-scattering were determined by the length of the $\hbar\omega_{OP}$-layer as suggested in [168,169], rather than just the scattering rate, we could not conclude by modifying only ω_{OP} whether the observed decrease in x_c compared to the 'STD' case is due to the changing scattering rate, or to the changing spatial extent of the $\hbar\omega_{OP}$-layer.

In an attempt to disentangle these possible effects, we decrease D_{OP} by $\sqrt{2}$

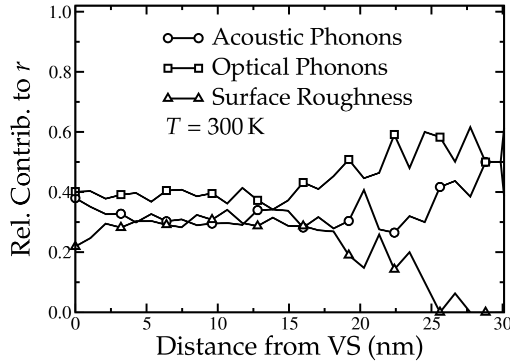


Figure 4.11: Relative contribution of acoustic phonon, optical phonon and surface roughness scattering to r. Compared to the standard parameters, D_{OP} and ω_{OP} have been reduced by a factor of $\sqrt{2}$ and a factor of 2 respectively, as explained in the text.

and ω_{OP} by 2 to keep the prefactor, thus the OP scattering rate, constant. At the same time, assuming a nearly linear potential profile around the virtual source, the spatial extent of the $\hbar\omega_{\mathrm{OP}}$-layer is halved. Fig. 4.10 shows that, by reducing ω_{OP} while keeping the scattering rate essentially constant, x_{c} is just 0.5 nm lower than the values we found in the 'STD' case. Thus, we may conclude that the change in x_{c} in the previous curve ($\omega_{\mathrm{OP}}/\sqrt{2}$) is mainly due to an increase in the OP scattering rate rather than a change of the $\hbar\omega_{\mathrm{OP}}$-layer.

These findings are confirmed by the results in Fig. 4.12, which show r corresponding to the various cases presented above. The effect on r due to modification of D_{OP} and ω_{OP} is small compared to the effect of the temperature. Again, a smaller D_{OP} yields no substantial changes in r compared to the 'STD' case. If ω_{OP} is reduced, r increases, whereas modification of both ω_{OP} and D_{OP} results in r being smaller than in the "$\omega_{\mathrm{OP}}/\sqrt{2}$" case, but greater than the 'STD' curve. We stress, however, that the changes in absolute terms are modest with respect to the T dependence. *Therefore we conclude that, although the optical phonon energy affects back-scattering through the scattering rate, it does not determine the critical length for back-scattering.*

Back-Scattering in a Template Structure

When analyzing the critical length for back-scattering in more detail, we wish not to cloud the results with effects related to self-consistency. In this regard, we employ a fixed template structure with either linear or parabolic potential

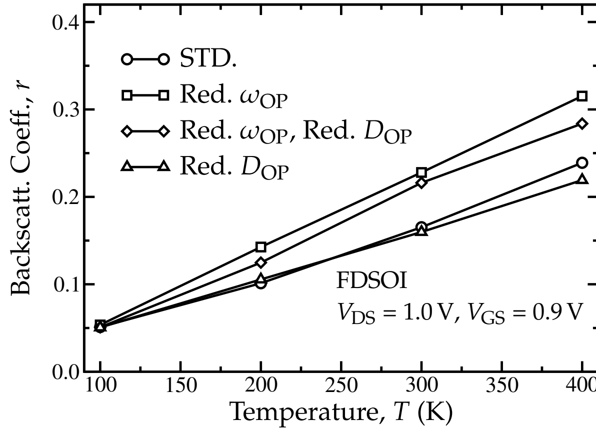


Figure 4.12: r vs. T, obtained by integrating the distribution of back-scattering events [cf. Fig. 4.6] along the channel. The different cases match those in Fig. 4.10

profile, to analyze in a systematic manner the impact of T, $\mathcal{E}$ and L on the backscatter characteristics. Furthermore, injection from the drain is switched off, so that all negative velocity carriers at the VS ($x = 0$) are due to back-scattering in the channel. A template material is employed, featuring a single spherical subband with $m^* = 1.0m_0$, and scattering is only due to acoustic phonons. The coupling constant of the acoustic phonons was adjusted to yield a reasonable low-field mobility of 400 cm^2/Vs at 300 K [171].

Since the critical length is often related to L_{kT} (the "$k_{\mathrm{B}}T$-layer"), as pointed out before, we use L_{kT} as a parameter. In the case of a linear potential profile $L_{\mathrm{kT}} = k_{\mathrm{B}}T/q|\mathcal{E}|$; we adopt the following strategies to vary L_{kT} and L/L_{kT}:

(A) for a given L and $\mathcal{E}$, only T is varied

(B) for a fixed $\mathcal{E}$, we vary T while changing L in order to keep $L/L_{\mathrm{kT}} = q|\mathcal{E}|L/k_{\mathrm{B}}T$ constant

(C) both L and T are fixed, and only $\mathcal{E}$ is varied

(D) for a fixed $\mathcal{E}$ and T, we change only L

Actually, the above cases represent the various options to change L_{kT} either with constant L (A, C) or with constant L/L_{kT} (B). In (D), instead, L_{kT} is fixed and only L is varied. Varying these parameters in a systematic manner allows

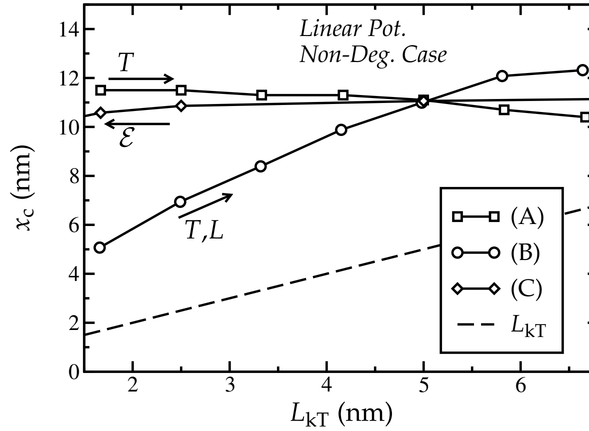


Figure 4.13: x_{c} versus L_{kT}, obtained in the template structure with constant accelerating field. The various strategies to vary L/L_{kT} are explained in the text. All curves coincide at $L_{\mathrm{kT}} = 5\,\mathrm{nm}$, with $\mathcal{E} = 52\,\mathrm{kV/cm}$, $L = 50\,\mathrm{nm}$ and $T = 300\,\mathrm{K}$.

us to study the link between the backscatter characteristics ($g(x)$, x_{c} and r) and the longitudinal electric field and temperature (through L_{kT}).

As for the terminology used in the following: we will often quantify the channel length with 'long', 'short' or their comparative forms. We stress that these attributes are to be related to the low-field carrier mean-free-path λ, which on its turn depends on the scattering mechanisms. Typically, at room temperature λ is around 25–40 nm for undoped or lightly doped silicon in an UTB-SOI configuration, or the template material considered in this work. Likewise, 'high' and 'low' field conditions refer to the voltage drop across the channel, i.e. $|\mathcal{E}|L$ for a linear potential profile, relative to the thermal voltage $k_{\mathrm{B}}T/q$.

To see the possible connection between L_{kT} and x_{c} [the centroid of $g(x)$], Fig. 4.13 reports x_{c} versus L_{kT}. In the most straightforward case, we change only the temperature, which causes x_{c} to shift only a little ($\approx 1.5\,\mathrm{nm}$) towards the VS with increasing temperature. Similarly, varying L_{kT} through $\mathcal{E}$ while keeping $T = 300\,\mathrm{K}$, leaves x_{c} essentially unchanged. Interestingly, if T is increased while L is changed in direct proportion so as to keep the ratio L/L_{kT} constant (in this case $L/L_{\mathrm{kT}} = 10$), we do observe an increase in x_{c}. Furthermore, x_{c} tends to saturate for large values of L, which can be explained as follows: as we have seen before, a.o. from the plots reporting $g(x)$, scattering close to the VS is most effective in contributing to r. If we move further away

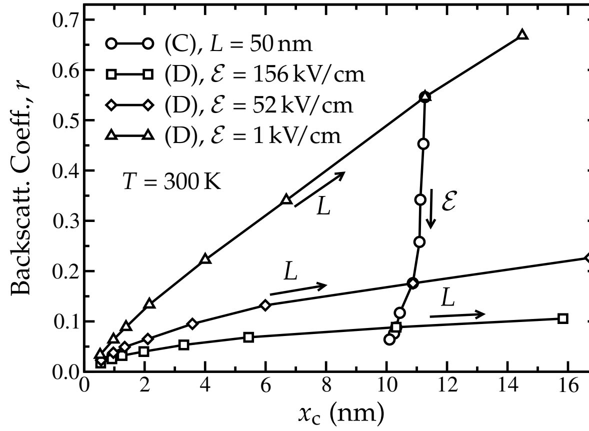


Figure 4.14: Curves (D) depict r vs. x_c, for a constant longitudinal field $\mathcal{E}$ and varying channel length L; L ranges from 1–100 nm and x_c increases with L. In curve (C), L is fixed (50 nm) and $\mathcal{E}$ is varied from 1–260 kV/cm.

from the VS, back-scattering can still occur but the probability for a carrier to actually reach the VS rapidly decreases. Thus in long channels, such as the last point of curve (B) in Fig. 4.13 where $L = 67$ nm (recall that $L/L_\mathrm{kT} = 10$), the contribution to r and x_c due to back-scattering near the drain is virtually zero, making the effect of any further increase in L negligible.

The above reasoning is supported by Fig. 4.14, which shows the relation between r and x_c through variation of the implicit parameter $\mathcal{E}$ [curve (C)] and L (D). We observe that r increases with L, with the strongest dependence at the lowest field. In fact, in the low-field limit, scattering throughout the entire channel can result in back-scattering, whereas for higher fields, scattering further into the channel contributes less and less to r, resulting in a weaker dependence of r on L.

In this respect it is worthwhile to examine the relative position of x_c in the channel, calculated as x_c/L. The results are shown in Fig. 4.15, for $\mathcal{E}$ = 52 kV/cm. We observe that x_c, which represents the 'center of mass' of $g(x)$, is located roughly halfway the channel for the shortest channels ($x_\mathrm{c}/L = 0.55$), whereas it is approximately 17 nm away from the VS for $L = 100$ nm (x_c/L = 0.17). Thus, back-scattering near the drain becomes progressively more pronounced as we reduce the channel length.

Turning again to Fig. 4.14, we set $L = 50$ nm and vary $\mathcal{E}$, shown in curve (C). We observe that r strongly decreases with increasing field while x_c, as

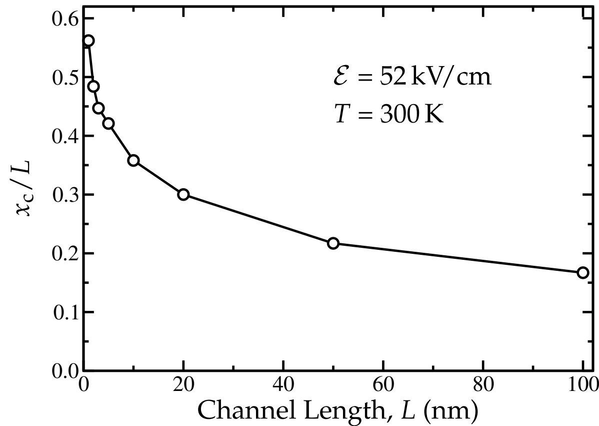


Figure 4.15: The relative position of x_c in the channel (x_c/L) vs. L, for $\mathcal{E}$ = 52 kV/cm ($L_{kT} \approx 5$ nm).

pointed out before when discussing Fig. 4.13 x_c shows only a slight variation with $\mathcal{E}$. This means that the *effective* carrier mean-free-path rather than L_{crit} changes when we modify $\mathcal{E}$: the nearly constant x_c means that the portion of the channel in which scattering is most critical, i.e. L_{crit}, remains unchanged. On the other hand, the average carrier velocity increases due to the accelerating electric field, which translates into an increasing mean-free-path.

Extraction of the critical length

So far, we have analyzed back-scattering just using the 'plain' simulation results, in terms of the distribution of backscatter events $g(x)$ and its associated characteristics, namely r and x_c. Hence, we did not rely on an assumed functional form of r and its dependence on L, $\mathcal{E}$ and T.

In the following, instead, we examine the L, $\mathcal{E}$ and T-dependence of the critical length L_{crit}, assuming that r can be written in the following 'flux-like' form [cf. (4.8)]

$$r = \frac{L_{crit}}{L_{crit} + \lambda} \tag{4.12}$$

We extract L_{crit} from the MSMC simulations by inverting (4.12). r can be directly obtained from the simulations, by taking the ratio of the negatively and positively directed flux at the VS (in absence of injection from the drain). In the following, λ in (4.12) is the *equilibrium* mean-free-path ($\mathcal{E} \Rightarrow 0$), meaning

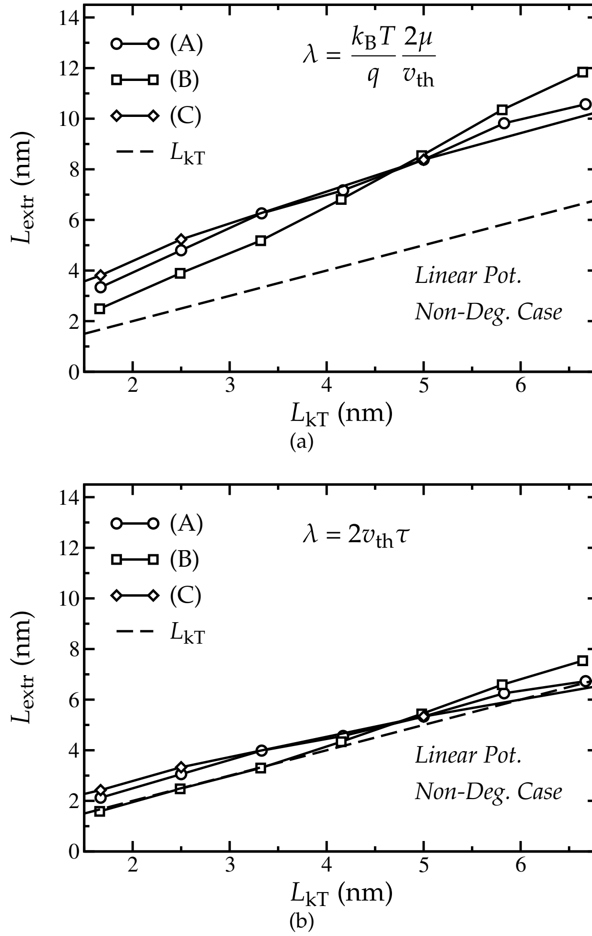


Figure 4.16: L_{extr} versus L_{kT}. (a) L_{extr} is obtained from $L_{extr} = \lambda_{flux} \cdot r/(1-r)$, with r obtained from the MSMC simulations. The labels match those in Fig. 4.13; in all cases, $L_{kT} \ll L$. (b) L_{extr} extracted using the mean-free-path under single-scattering conditions, $\lambda_{SSC} = 2v_{th}\tau$.

that L_{crit} in (4.12) must be viewed as an *effective* critical length, which includes the possible dependence on $\mathcal{E}$ and L.

In the remainder of this section, we label the extracted critical length L_{extr}. Fig. 4.16 reports the values of L_{extr} versus L_{kT}, for the same cases as those shown in Fig. 4.13. In Fig. 4.16(a) L_{extr} is extracted with λ in (4.12) equal to the equilibrium mean-free-path from the Flux-theory, that is, $\lambda_{flux} = 2\mu k_B T/qv_{th}$.

Although the values of L_{extr} are systematically larger than L_{kT}, we observe a clear correlation between L_{extr} and L_{kT}. The correlation improves when L_{extr} is extracted using a different value of λ, as we will show in the following.

λ_{flux} is derived assuming that scattering is strong enough to keep the electron distribution close to equilibrium. Here we consider a different mean-free-path assuming that carriers encounter on average only a single scattering event: in the low-field (equilibrium) limit and close to the VS, the average velocity of the injected non-degenerate carrier distribution is the thermal velocity v_{th}. We assume that carriers scatter elastically and isotropically, with scattering rate $1/\tau$. Then, the probability of having a negative momentum after scattering is $1/2$. Hence, the resulting single-scattering mean-free-path, which we refer to as λ_{SSC}, reads $\lambda_{\text{SSC}} = 2v_{\text{th}}\tau$ and it is related to λ_{flux} from (4.6) through $\lambda_{\text{SSC}} = \frac{2}{\pi}\lambda_{\text{flux}}$.

The results of L_{extr} extracted by inserting λ_{SSC} in (4.12) are shown in Fig. 4.16(b), demonstrating that L_{extr} exhibits a behavior very similar to L_{kT}. These results are quite reassuring in that they support the theory of the "$k_{\text{B}}T$-layer" [157,165], at least in the sense that r is linked to L_{kT} if we use λ_{SSC} rather than λ_{flux}. We stress, however, that L_{extr} is close but not equal to L_{kT}. The need to use a mean-free-path smaller than λ_{flux} is in accordance with the findings in [171], where it was pointed out that in order to obtain a good agreement between the flux-like formula and the MSMC results, the ratio λ/L_{crit} was found to be smaller than predicted by the 'flux-theory' ($\lambda_{\text{flux}}/L_{\text{kT}}$).

In the cases shown in Fig. 4.16, L_{kT} was much smaller than L. To explore the behavior of L_{extr} for L_{kT} in the order of L, we report in Fig. 4.17 L_{extr} as a function of L, with $\mathcal{E}$ as parameter. We observe that, in the low-field regime ($\mathcal{E}$ = 1 kV/cm, L_{kT} = 260 nm), L_{extr} closely follows L. For $\mathcal{E}$ = 52 kV/cm (L_{kT} = 5 nm) the L-dependence is less strong, and gradually weakens with increasing L. For completeness we report the relation between x_{c} and L_{extr} in Fig. 4.18, for several values of $\mathcal{E}$. We observe that the proportionality of x_{c} with L_{extr} increases with increasing $\mathcal{E}$.

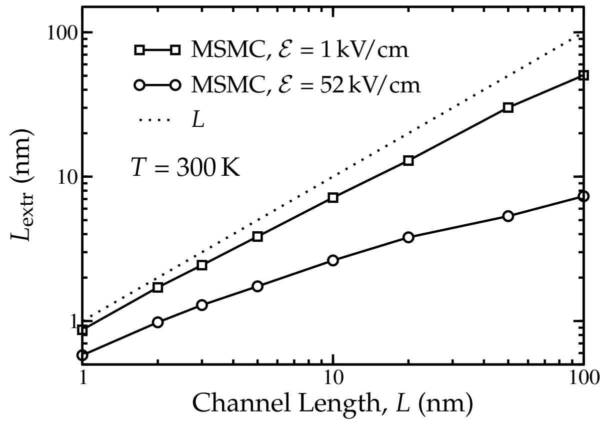


Figure 4.17: L_{extr} vs. L, with $\mathcal{E}$ as parameter; L_{extr} is obtained using the single-scattering mean-free-path λ_{SSC}.

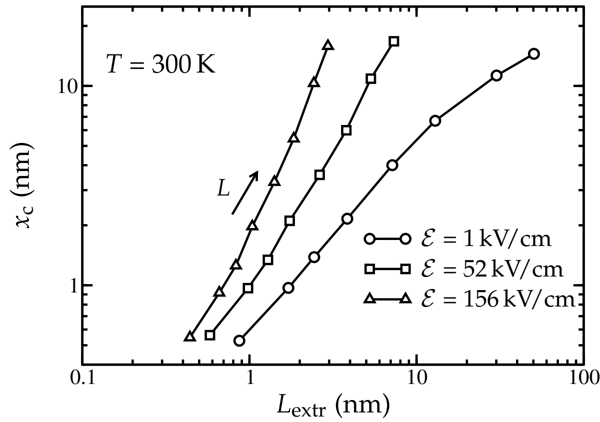


Figure 4.18: x_c vs. L_{extr}, for different values of $\mathcal{E}$; L_{extr} is obtained using the single-scattering mean-free-path λ_{SSC}. The implicit parameter L takes the values [1, 2, 3, 5, 10, 20, 50 and 100] nm.

Summary

We may summarize our main findings as follows: 1) simulations of a realistic 32 nm FDSOI MOSFET indicate that OP scattering does not directly determine the critical length for back-scattering. 2) systematic analysis of back-scattering in a template device structure shows that changes in temperature and longitudinal field do not have significant impact on the fraction of the channel in which backscattering is important. If we vary the channel length L, the centroid of back-scattering increases with L. Its position relative to L, instead, decreases which means that backscattering is more and more weighted towards the VS due to the reducing contribution of scattering near the drain end of the channel. 3) on the same template structure, we extracted the effective critical length (L_{extr}) from the MSMC simulations by assuming a flux-like form of r. As a result, we found that under high-field conditions L_{extr} exhibits strong correlation with L_{kT}, whereas L_{extr} follows L in the low-field limit. Furthermore, the appropriate low-field mean-free-path to be used in the aforementioned flux-like equation is $\lambda_{\mathrm{SSC}} = 2v_{\mathrm{th}}\tau$ instead of $\lambda_{\mathrm{flux}} = \frac{k_{\mathrm{B}}T}{q}\frac{2\mu}{v_{\mathrm{th}}}$.

4.5 A new model for the backscatter coefficient

Modelling of r is an active area of research [137, 166, 175] since, as we have seen before, r provides an estimate of how close to the ballistic limit a device operates. In the previous section we have systematically analyzed the critical length for back-scattering. We have seen that L_{extr}, i.e., the effective critical length obtained from the MSMC simulations, exhibits a strong correlation with L_{kT}. Furthermore, by extracting L_{extr} using the single-scattering mean-free-path λ_{SSC}, we found that L_{extr} is close, but not equal to, L_{kT}.

Now, let us compare once again the conventional Flux model to the MSMC results, using the findings reported in the previous section. Fig. 4.19 plots r as a function of $\mathcal{E}$, showing the MSMC results obtained from the template device structure with linear potential profile (as introduced on p. 89), along with values predicted by the Flux model using either the 'original' λ_{flux} or λ_{SSC}. For the sake of clarity we reprint the Flux model, originally proposed in [157] and elaborated upon in [165],

$$r_{\text{flux}} = \frac{L_{\text{kT}}\,(1-\beta)}{L_{\text{kT}}\,(1-\beta) + \lambda_{\text{flux}}} \tag{4.13}$$

with $\beta = \exp(-L/L_{\text{kT}})$, $\lambda_{\text{flux}} = 2\mu k_{\text{B}}T/qv_{\text{th}}$ and $v_{\text{th}} = \sqrt{2k_{\text{B}}T/\pi m^*}$. The original Flux model (solid line in Fig. 4.19) shows good agreement in low-field

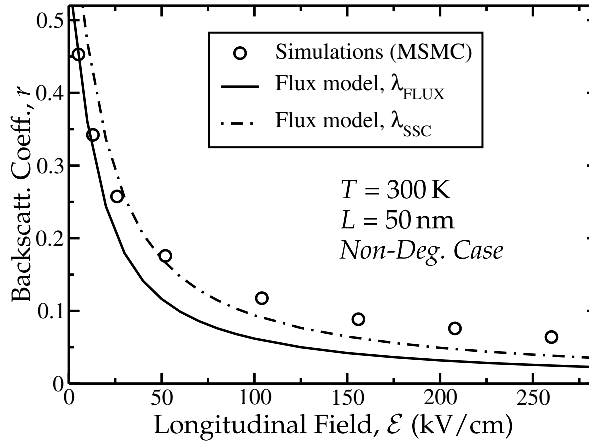


Figure 4.19: r vs. longitudinal field, comparing the MSMC values with the Flux model employing the mean-free-path from the Flux model, (4.6), and the mean-free-path under single-scattering conditions, $\lambda_{\text{SSC}} = 2\tau v_{\text{th}}$.

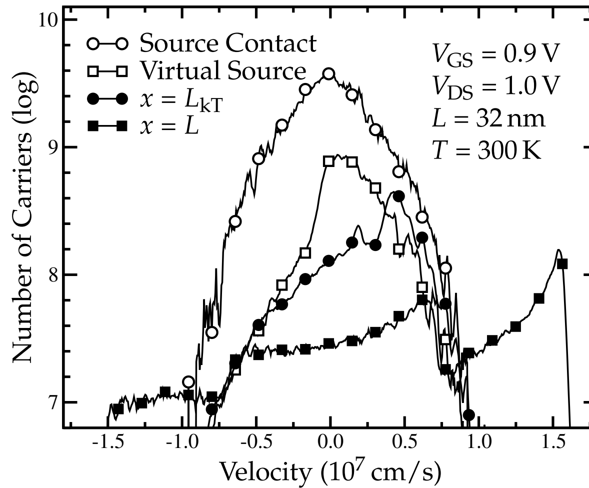


Figure 4.20: Velocity distribution at several points along the channel of a 32 nm Single–Gate SOI MOSFET [172]: at the source contact, at the virtual source ($x = 0$), at L_{kT} (i.e. where the voltage drop from the virtual source equals the thermal voltage, $L_{kT} \approx 5\,nm$), and at the drain end of the channel (L).

regime, which is explained by recalling that (4.13) and λ_{flux} were derived assuming near equilibrium, hence low-field, conditions. The Flux model with λ_{SSC} (dot-dashed line) results in a better overall agreement between model and MSMC, but still, quantitatively, the model predictions can be improved particularly in the high-field regime.

The 'equilibrium condition' is tightly related to the number of scattering events: scattering not only randomizes the carrier momentum, i.e. the direction and velocity of the carriers, but it can also restore thermal equilibrium between the carrier distribution and the crystal lattice, even if most scattering events are due to acoustic phonons. Acoustic phonon scattering is commonly treated as elastic, since the energy exchange involved in this process is very small (few meV) compared to the thermal energy. However if scattering is very frequent and the carriers have not yet acquired much kinetic energy, such as close to the VS (cf. inset Fig. 4.6), the total energy exchange may be sufficient to maintain near-equilibrium conditions.

The equilibrium conditions may be questionable in nanoscale MOSFETs where carriers encounter only few scattering events. To illustrate this point, Fig. 4.20 shows the velocity distribution along the channel of a 32 nm FDSOI

device (see also p. 4.4). Up to a distance of approximately L_{kT}, i.e. where the voltage drop from the VS equals the thermal voltage, the positive velocity distribution essentially retains its initial Maxwellian shape, slightly displaced towards higher v_x. However, as we move further along the channel towards the drain, the distribution deviates from a Maxwellian ('thermal') distribution.

In addition to the equilibrium condition, (4.13) implicitly assumes a linear potential profile, as shown in [165]. Furthermore, λ_{flux} is related to the low-field mobility (μ), so it is not directly linked to the individual scattering mechanisms and does not separate the effects of elastic and inelastic scattering.

In this section, we will introduce a new model for r which, differently from [157], does not require the fluxes to be close to equilibrium nor does it pose any restriction on the shape of the backscattered distribution. In fact, it is based on the assumption that carrier transport is close to ballistic, as it is expected to happen in nanoscale MOSFETs where the channel length L is in the order of the carrier mean free path.

Furthermore, this model treats elastic and inelastic scattering separately, and can handle situations with a strong carrier degeneracy (e.g. such as at the VS of MOSFETs operating in the ON-condition). Using a simple test case, the model results are compared with values obtained from MSMC simulations [151, 171], for a wide range of electric field, channel length and temperature values. The core of this work was presented in [176].

Model Framework

Before proceeding to the derivation of the model, we briefly discuss the assumptions which are central to the model framework. The relevant quantities will be defined along the way. First, we assume that only few scattering events occur in the channel and that, in particular, those carriers which are scattered back to the VS, have encountered on average just a single scattering event. Furthermore, we consider a single subband with parabolic energy dispersion. In addition, the new model shares the following assumptions with the Flux model:

- 1D in real space
- no injection from the drain
- no gate-current

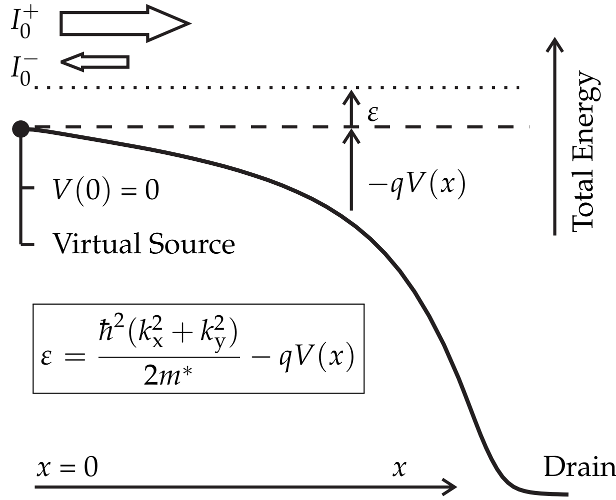


Figure 4.21: Schematic of the model framework. I_0^+ and I_0^- are resp. the injected ballistic flux and the backscattered flux, at the virtual source; ε is the total energy relative to the conduction band edge at $x = 0$, and $V(x)$ is the potential.

Derivation of the Single–Scattering model

We assume that the flux of carriers F^+ moving from the VS to the drain can be considered as a ballistic one, i.e. back-scattering events are very rare and do not significantly affect F^+. This assumption will be relaxed in the the second part of this section. We denote as I_0^+ the current associated with F^+. The energy distribution of the charges belonging to F^+ is indicated as $n^+(x, \varepsilon)$. The integral of $n^+(x, \varepsilon)$ over the total energy ε gives the inversion density of carriers moving inside the channel with positive group velocity:

$$n^+(x) = \int_0^\infty n^+(x, \varepsilon)\, d\varepsilon \tag{4.14}$$

The total energy ε includes the kinetic energy and the subband energy. As sketched in Fig. 4.21, $\varepsilon = 0$ is taken at the bottom of the subband at the VS ($x = 0$). Since we assume F^+ to be ballistic, $n^+(x, \varepsilon)$ is null for $\varepsilon < 0$ at any x along the channel.

The number of carriers belonging to F^+ suffering a scattering event per unit time, unit distance and unit energy is given by $n^+(x, \varepsilon)/\tau$, where $1/\tau$ is the scattering rate. At this stage we assume τ not to depend on energy and

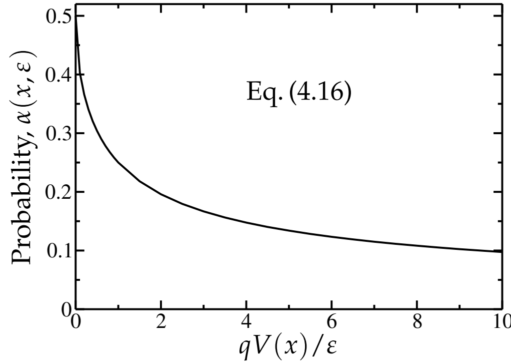


Figure 4.22: Probability of returning to the VS after scattering, calculated with (4.16).

assume the scattering events to be elastic and isotropic. The backscattered flux I_0^- at the VS can thus be expressed in terms of $n^+(x,\varepsilon)$, τ, and the probability $\alpha(x,\varepsilon)$ that a carrier, once scattered, reaches the virtual source again. We write

$$I_0^- = q \int_0^L \int_0^\infty \frac{n^+(x,\varepsilon)}{\tau} \alpha(x,\varepsilon)\, d\varepsilon\, dx \tag{4.15}$$

with L the channel length. The probability $\alpha(x,\varepsilon)$ that a carrier with total energy ε, after scattering at position x, has sufficient longitudinal energy to surmount the barrier and to arrive at the source, is given by [157]

$$\alpha(x,\varepsilon) = \frac{1}{\pi} \arccos \sqrt{\frac{qV(x)}{\varepsilon + qV(x)}} \tag{4.16}$$

As exemplified in Fig. 4.22, we observe that $\alpha(x,\varepsilon) = 0.5$ at the source and that it gradually decreases towards the drain. So, even if we do not embrace any concept of $k_B T$-layer or similar, the scattering events occurring close to the VS are the most effective in back-scattering carriers and, consequently, give the main contribution to r. $n^+(x,\varepsilon)$ can be determined by considering that, under ballistic transport, the carrier energy distribution can be obtained following the approach in [151]. However, differently from [151], in this work we consider the *total* energy, rather than the longitudinal energy. Assuming Fermi-Dirac (FD) statistics, we obtain by solving the Boltzmann Transport Equation

in the ballistic case (see also Appendix E)

$$n^{+}(x,\varepsilon) = \frac{n_0^{+}}{k_\mathrm{B}T \ln\left[1+\exp(\eta)\right]} \times \frac{1}{\exp\left(\dfrac{\varepsilon}{k_\mathrm{B}T} - \eta\right) + 1} \frac{2}{\pi} \arccos\sqrt{\frac{qV(x)}{\varepsilon + qV(x)}} \quad (4.17)$$

with n_0^{+} the inversion density of carriers with positive group velocity at the VS [cf. (4.14)]; η is the degeneracy level, defined as $\eta = [E_\mathrm{F} - E_0]/k_\mathrm{B}T$ at the VS.

The ballistic current injected at the VS is

$$I_0^{+} = q n_0^{+} \nu_\mathrm{inj} \quad (4.18)$$

with ν_inj the *injection velocity*. The latter can be written in terms of the non-degenerate thermal velocity and a correction term to account for FD statistics [177]

$$\nu_\mathrm{inj} = v_\mathrm{th} \cdot \frac{\mathcal{F}_{1/2}}{\mathcal{F}_0} \quad (4.19)$$

where the FD integrals $\mathcal{F}_{1/2}(\eta)$ and $\mathcal{F}_0(\eta)$ are defined as [178]

$$\mathcal{F}_0(\eta) = \ln\left[1 + \exp\left(\eta\right)\right] \quad (4.20a)$$

$$\mathcal{F}_{1/2}(\eta) = \frac{2}{\sqrt{\pi}} \int_0^\infty \frac{\sqrt{x}}{1 + \exp\left(x - \eta\right)}\, dx \quad (4.20b)$$

We insert (4.18) in (4.17) and write

$$n^{+}(x,\varepsilon) = \gamma(x,\varepsilon) \cdot \frac{I_0^{+}}{q} \quad (4.21)$$

with

$$\gamma(x,\varepsilon) = \frac{1}{k_\mathrm{B}T \ln\left[1+\exp(\eta)\right] \nu_\mathrm{inj}} \times \frac{1}{\exp\left(\dfrac{\varepsilon}{k_\mathrm{B}T} - \eta\right) + 1} \frac{2}{\pi} \arccos\sqrt{\frac{qV(x)}{\varepsilon + qV(x)}} \quad (4.22)$$

Note that $V(x)$ can have an arbitrary profile with $V(x) \geq 0$. Substituting (4.21) in (4.15) and assuming no injection from the drain, yields the following expression for the 'single-scattering' (SSC) backscatter coefficient r_SSC

$$r_\mathrm{SSC} \equiv \frac{I_0^{-}}{I_0^{+}} = \int_0^L \int_0^\infty \frac{\gamma(x,\varepsilon)}{\tau} \alpha(x,\varepsilon)\, d\varepsilon\, dx \quad (4.23)$$

Summarizing, (4.23) allows us to compute the distribution of backscatter events $g(x)$, i.e. the integral over ε in (4.23), for a given incoming carrier distribution and a known scattering rate.

Reduction of the Positive Flux

Up to now, we have assumed that backscatter events do not occur frequently. In effect, essentially all carriers remain oriented in positive direction and the flux can be considered ballistic. However, if back-scattering becomes more pronounced –we will discuss this quantitatively when presenting the results– the positive flux $I^{+}(x)$ will significantly decrease along the channel due to the carriers which change momentum from positive to negative.

In this section, we discuss how scattering modifies the energy distribution of this flux, denoted as $I^{+}(x,\varepsilon)$. The reduction of $I^{+}(x,\varepsilon)$ is proportional to $n^{+}(x,\varepsilon)$ and to $1/\tau$. Since the scattering is assumed to be isotropic, the probability that a carrier's momentum after scattering points towards the source equals $1/2$. Thus, we can write

$$\frac{dI^{+}(x,\varepsilon)}{dx} = -\frac{qn^{+}(x,\varepsilon)}{2\tau} = -\frac{1}{2\tau}\frac{I^{+}(x,\varepsilon)}{v_x^{+}(x,\varepsilon)} \tag{4.24}$$

where we have used $I^{+}(x,\varepsilon) = qn^{+}(x,\varepsilon)v_x^{+}(x,\varepsilon)$ [cf. (4.3a)]. The velocity of the 'reduced' flux is assumed to be equal to the velocity of a ballistic flux. The longitudinal (x-directed) component $v_x^{+}(x,\varepsilon)$ of the ballistic velocity can be shown to be

$$v_x^{+}(x,\varepsilon) = \sqrt{\frac{2\varepsilon}{m^{*}}}\frac{1}{\arccos\sqrt{\frac{qV(x)}{\varepsilon+qV(x)}}} \tag{4.25}$$

Then, by integrating (4.24), we find for $I^{+}(x,\varepsilon)$

$$I^{+}(x,\varepsilon) = I^{+}(0,\varepsilon)\exp\left(-\int_0^x \frac{1}{2\tau v_x^{+}(x',\varepsilon)}\,dx'\right) \tag{4.26}$$

The carrier concentration is then given by

$$n^{+}(x,\varepsilon) = n_0^{+}v_{\text{inj}}\gamma(x,\varepsilon)\exp\left(-\int_0^x \frac{1}{2\tau v_x^{+}(x',\varepsilon)}\,dx'\right) \tag{4.27}$$

We thus see that r_{SSC} can be obtained from (4.23) by replacing $\gamma(x,\varepsilon)$ with

$$\gamma^{*}(x,\varepsilon) = \gamma(x,\varepsilon)\exp\left(-\int_0^x \frac{1}{2\tau v_x^{+}(x',\varepsilon)}\,dx'\right) \tag{4.28}$$

Comparison with MSMC results

In this section we compare the model results with the MSMC backscatter characteristics, for varying L, temperature T and longitudinal field $\mathcal{E}$. To facilitate a clear and direct comparison, we employ, unless stated otherwise, a constant accelerating field, a template material (as introduced on p. 89) and no injection from the drain. The template material features a single spherical valley ($m^* = 1.0m_0$), and low-field mobility $\mu = 400\,\mathrm{cm^2/Vs}$ at 300 K, obtained by adjusting the coupling constant of acoustic phonons.

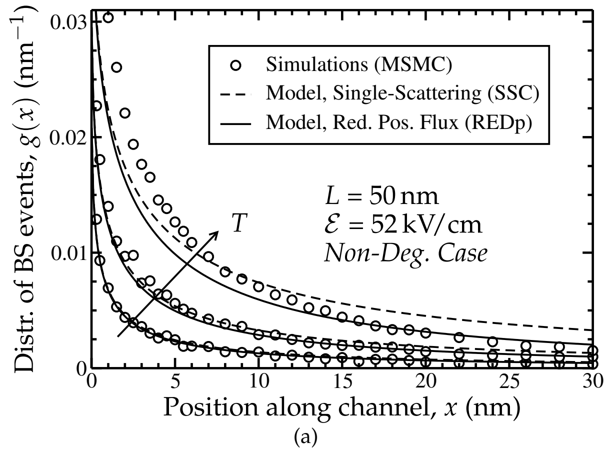
We consider a single subband within a spherical valley with parabolic energy dispersion, which renders the scattering rate with elastic acoustic phonons, $1/\tau$ in the model, constant over energy. The value of τ has been taken from the MSMC simulator without any adjustment in all forthcoming figures. The value of λ_{flux}, to be used in (4.13), is obtained by noting that μ is related to the scattering rate through $\mu = q\tau/m^*$, hence $\lambda_{\mathrm{flux}} = \tau\sqrt{2\pi k_{\mathrm{B}}T/m^*}$ ($\lambda_{\mathrm{flux}} \approx 39\,\mathrm{nm}$ with $T = 300\,\mathrm{K}$ and $\tau = 0.23\,\mathrm{ps}$). The model results have been obtained by numerical integration of (4.23).

Elastic Scattering

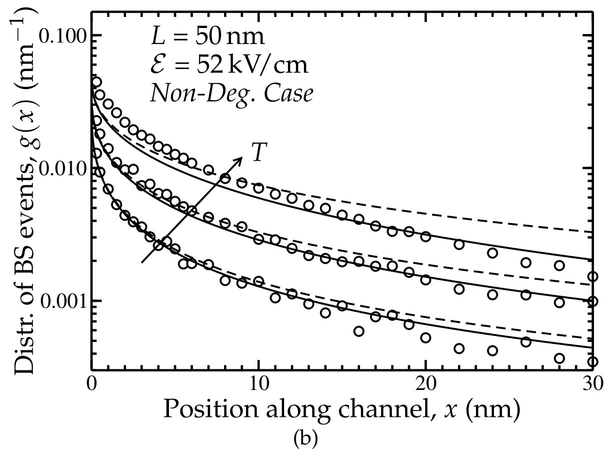
Initially, we use Boltzmann's approximation, which in our model is naturally obtained by setting $\eta \ll 0$ in (4.22). The impact of degeneracy will be discussed in the second part of this section.

Fig. 4.23 shows, for several temperatures, the distribution of backscatter events $g(x)$. Most scattering events which contribute to r occur close to the VS. Clearly, the single–scattering model [(4.16)–(4.22), labeled SSC] is able to capture the general features of the distribution produced by the MSMC. The quantitative agreement in the tail of the distribution can be improved by accounting for reduction of the positive flux [(4.23) employing (4.28), REDp], particularly at high temperature and under low-field conditions, i.e. conditions which induce enhanced scattering.

Fig. 4.24 shows the values of r, corresponding to the curves in Fig. 4.23, for both a linear and a parabolic potential profile. Along with the SSC, REDp and the MSMC values, we plot the values predicted by the conventional model based on the thermal fluxes ("Flux model"), obtained from (4.13) and [165, (37)] for the linear and parabolic potential, respectively. By accounting for the reduction of the positive flux, the model values essentially coincide with the MSMC results for both the linear and parabolic potential profile. The Flux model systematically underestimates r, which is attributed to the off-equili-

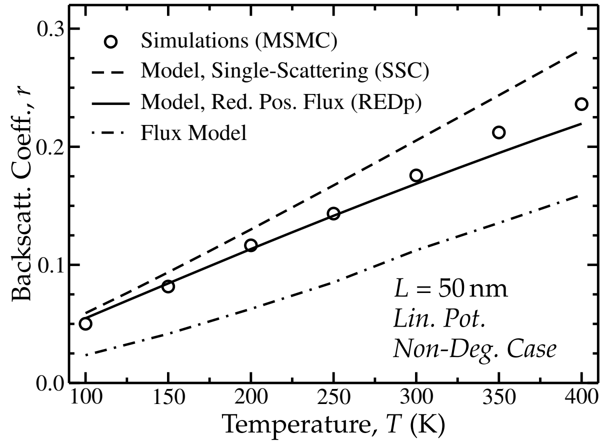


(a)

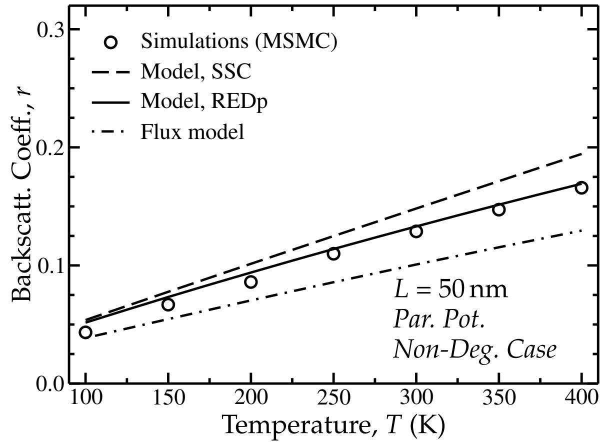


(b)

Figure 4.23: Distribution of backscattering events contributing to r [162] on linear (a) and logarithmic vertical axis (b). The plot shows both the model and MSMC values for $T = [100, 200, 400]$ K. The curves are normalized such that integrating yields r, the backscatter coefficient.



(a)



(b)

Figure 4.24: r vs. temperature, showing the SSC, REDp and the MSMC results along with the Flux model values from (4.13); linear (a) and parabolic (b) potential profile, both having $L_{\mathrm{kT}} = 5\,\mathrm{nm}$ at $T = 300\,\mathrm{K}$.

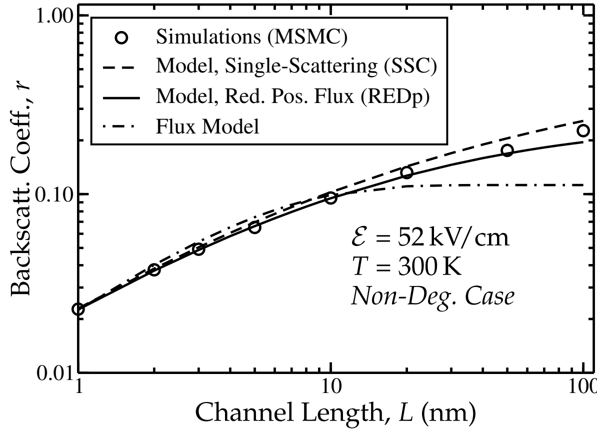


Figure 4.25: r vs. channel length; $\mathcal{E}$ = 52 kV/cm, T = 300 K. The agreement can be improved by accounting for reduction of the positive flux.

brium ("non-thermal") shape of the backscattered carrier distribution [179]. The difference between the Flux model and MSMC is slightly smaller when we use the parabolic potential instead of the linear potential profile.

Fig. 4.25 depicts r versus L, showing that the SSC tracks fairly well the MSMC results; again, the agreement can be improved by accounting for the positive flux reduction. The r predicted by the Flux model, instead, saturates for the longer channels since $\beta = \exp\left(-L/L_{\mathrm{kT}}\right)$ in (4.13) becomes very small.

The findings from Figs. 4.24 and 4.25 are concisely summarized in Fig. 4.26, which shows r for different values of $\mathcal{E}$. Clearly, the entire range of $r(\mathcal{E})$ is captured by the REDp model. For the higher fields (corresponding to $r < 0.2$), the single–scattering assumption of the SSC model turns out to be a good approximation, judging from the agreement in the SSC and MSMC results. Eq. (4.13) shows good agreement for the lowest fields only, which can be explained by noting that it assumes close to equilibrium (hence low-field) conditions.

In Figs. 4.23–4.26 we have assumed a non-degenerate electron gas, i.e. using Boltzmann's approximation. Fig. 4.27, instead, reports r as a function of the degree of degeneracy at the VS, for both a low [set 1)] and a high-field case [set 2)]. Although r decreases [set 1)] or reaches a maximum [set 2)], the changes in absolute terms are modest, both in the model and in the simulations. The modest impact of degeneracy may be attributed to the energy independent scattering rate: if a carrier is pushed up in energy, while the scat-

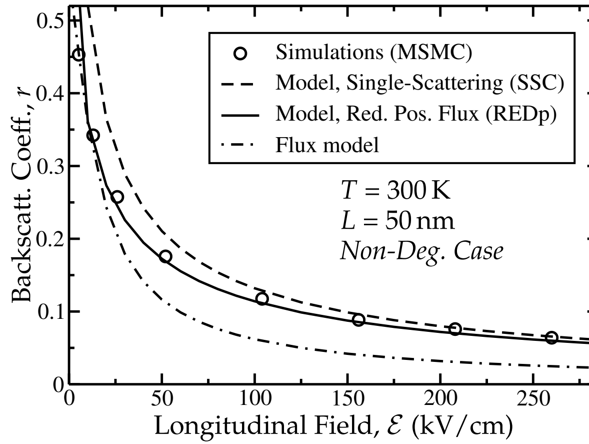


Figure 4.26: r vs. longitudinal field, showing the model with and without reduction of the positive flux, along with the MSMC and the flux model results.

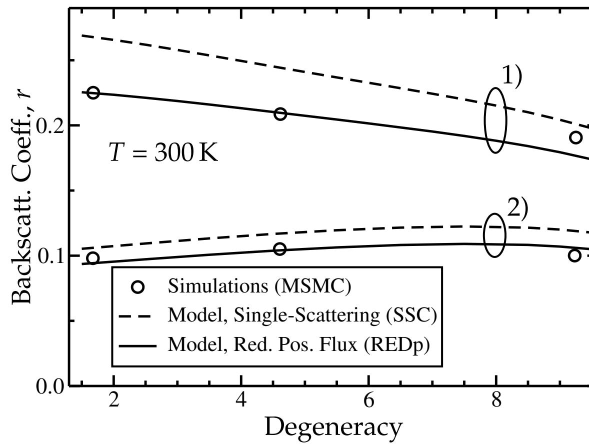


Figure 4.27: r for different values of the degeneracy level η, showing two sets of curves: set 1) uses $L = 20\,\text{nm}$, $\mathcal{E} = 13\,\text{kV/cm}$; set 2) is obtained with $L = 50\,\text{nm}$, $\mathcal{E} = 156\,\text{kV/cm}$. In either case, r is found to be only weakly dependent on η.

tering rate remains constant, the carrier mean-free-path will increase. However, according to (4.16), the probability of returning to the VS increases with increasing carrier energy. In effect, r remains essentially constant.

Inelastic Scattering

So far, we have only considered backscattering due to elastic scattering (e.g. acoustic phonons). In order to include also inelastic scattering mechanisms, such as optical phonon (OP) absorption and emission, we note that inelastic scattering alters the carrier energy and, as a consequence, the probability of returning to the VS. Therefore, we extend $\alpha(x,\varepsilon)/\tau$ from (4.15) and (4.16) to

$$\begin{aligned}\alpha^*(x,\varepsilon) = &\left(\frac{1}{\tau_{\mathrm{AP}}}\frac{1}{\pi}\arccos\sqrt{\frac{qV(x)}{\varepsilon+qV(x)}}\right.\\ &+\frac{1}{\tau_{\mathrm{OE}}}\frac{1}{\pi}\arccos\sqrt{\frac{qV(x)}{\varepsilon-\hbar\omega_{\mathrm{OP}}+qV(x)}}\,\theta\left(\varepsilon-\hbar\omega_{\mathrm{OP}}\right)\\ &\left.+\frac{1}{\tau_{\mathrm{OA}}}\frac{1}{\pi}\arccos\sqrt{\frac{qV(x)}{\varepsilon+\hbar\omega_{\mathrm{OP}}+qV(x)}}\right)\end{aligned} \tag{4.29}$$

in which τ_{OA}^{-1} and τ_{OE}^{-1} are the scattering rates of OP absorption and emission, respectively; $\hbar\omega_{\mathrm{OP}}$ is the OP energy, and $\theta(\mathrm{x})$ is the unit step function which ensures that OP emission can occur only if $\varepsilon \geq \hbar\omega_{\mathrm{OP}}$. In the last term in (4.29), the carrier energy is increased by $\hbar\omega_{\mathrm{OP}}$, which corresponds to OP absorption. Since we consider the total carrier energy rather than only the longitudinal energy, we do not need to make any assumptions on the redistribution of the energy into the longitudinal and transversal direction, in contrast to e.g. [168].

Reduction of the positive flux in presence of inelastic scattering is accounted for by extending (4.28) to

$$\begin{aligned}\gamma_{\mathrm{OP}}^*(x,\varepsilon) = \gamma(x,\varepsilon)\exp&\left(-\int_0^x\frac{1}{2\tau_{\mathrm{AP}}v_x^+(x',\varepsilon)}\,dx'\right)\\ \times\exp\left(-\int_0^x\frac{1}{2\tau_{\mathrm{OE}}v_x^+(x',\varepsilon)}\,dx'\right)&\exp\left(-\int_0^x\frac{1}{2\tau_{\mathrm{OA}}v_x^+(x',\varepsilon)}\,dx'\right)\end{aligned} \tag{4.30}$$

To compare the model results to the MSMC values, we have to find τ_{OA} and τ_{OE} corresponding to a given value of $\hbar\omega_{\mathrm{OP}}$. From the MSMC we can extract

the total OP scattering rate τ_{OP}^{-1}, which is the sum of the absorption and emission rate:

$$\frac{1}{\tau_{\mathrm{OP}}} = \frac{1}{\tau_{\mathrm{OA}}} + \frac{1}{\tau_{\mathrm{OE}}} \tag{4.31}$$

For a single subband electron gas in the parabolic effective mass approximation, the inverse scattering rates τ_{OE} and τ_{OA} are related through [145]

$$\frac{\tau_{\mathrm{OA}}}{\tau_{\mathrm{OE}}} = \frac{N_{\mathrm{OP}}+1}{N_{\mathrm{OP}}} \tag{4.32}$$

with N_{OP} the so-called Bose-Einstein factor. N_{OP} is given by [82]

$$N_{\mathrm{OP}} = \frac{1}{\exp\left(\dfrac{\hbar\omega_{\mathrm{OP}}}{k_{\mathrm{B}}T}\right) - 1} \tag{4.33}$$

Combining the above equations, the resulting expressions for τ_{OE} and τ_{OA} in terms of τ_{OP} read

$$\tau_{\mathrm{OE}} = \frac{2N_{\mathrm{OP}}+1}{N_{\mathrm{OP}}+1}\tau_{\mathrm{OP}} \tag{4.34a}$$

$$\tau_{\mathrm{OA}} = \frac{2N_{\mathrm{OP}}+1}{N_{\mathrm{OP}}}\tau_{\mathrm{OP}} \tag{4.34b}$$

Thus, for each $\hbar\omega_{\mathrm{OP}}$, the values of τ_{OA} and τ_{OE} can be calculated from the OP scattering rate in the MSMC simulator, which renders the model and MSMC results directly comparable. Numerical values of the scattering rates as a function of $\hbar\omega_{\mathrm{OP}}$ are shown in Fig. 4.28. We observe that for large values of $\hbar\omega_{\mathrm{OP}}$ scattering is dominated by acoustic phonons and that the OP scattering rate increases with decreasing $\hbar\omega_{\mathrm{OP}}$. Furthermore, we note that OP absorption exhibits a stronger $\hbar\omega_{\mathrm{OP}}$ dependence than does OP emission, in accordance with (4.34).

The scattering rate can be translated into a more familiar 'macroscopic' quantity, such as the low-field mobility. First we compute the effective scattering rate τ_{eff}^{-1} with Matthiessen's rule [12]

$$\frac{1}{\tau_{\mathrm{eff}}} = \frac{1}{\tau_{\mathrm{AP}}} + \frac{1}{\tau_{\mathrm{OP}}} \tag{4.35}$$

and then calculate the mobility with $\mu_{\mathrm{est}} = q\tau_{\mathrm{eff}}/m^{*}$. The use of (4.35) is, strictly speaking, justified only when the scattering mechanisms have identical energy dependence [145, p. 167]. Hence, the resulting mobility should be considered an estimate. Fig. 4.29 shows μ_{est} corresponding to the scattering rates

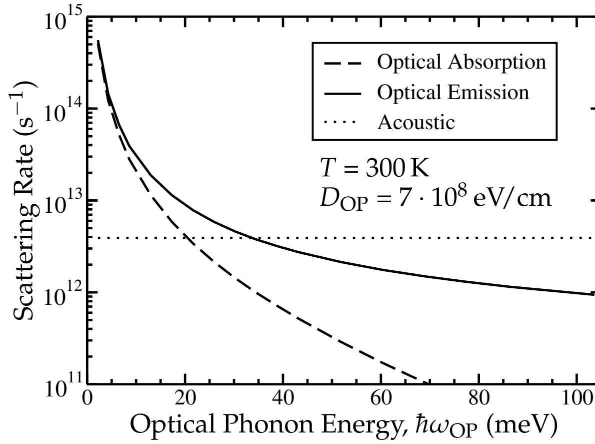


Figure 4.28: Acoustic and Optical phonon scattering rate versus $\hbar\omega_{\mathrm{OP}}$, calculated with (4.34) and τ_{OP} extracted from the MSMC simulator.

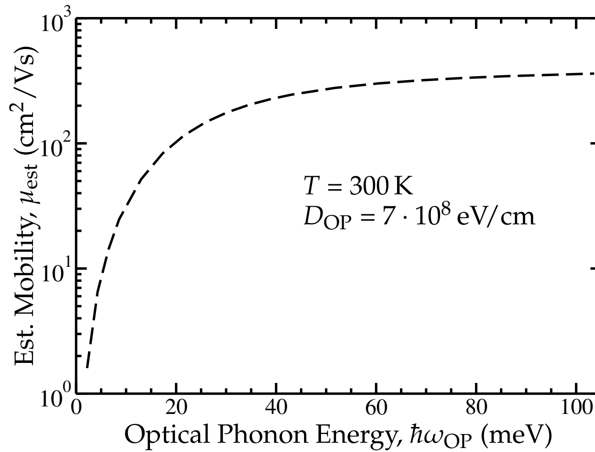


Figure 4.29: Estimate of the low-field mobility μ_{est} computed with (4.34), using the OP scattering rate from the MSMC simulator.

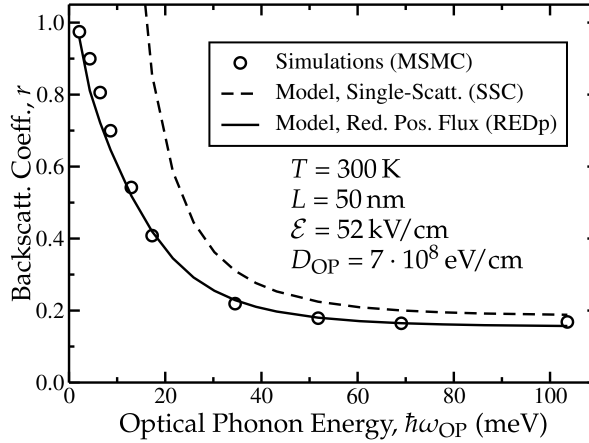


Figure 4.30: r vs. the optical phonon energy $\hbar\omega_{\mathrm{OP}}$. For each $\hbar\omega_{\mathrm{OP}}$, the scattering rates have been directly obtained from the scattering rates in the MSMC simulator.

in Fig. 4.28. μ_{est} saturates to the acoustic phonon limited mobility for large values of $\hbar\omega_{\mathrm{OP}}$ and decreases with decreasing $\hbar\omega_{\mathrm{OP}}$.

Now that we know how to calculate the values of τ_{OE} and τ_{OA} to be used in (4.29), we can assess the impact of OP scattering on r, as illustrated in Fig. 4.30. The OP deformation potential D_{OP} in the MSMC was increased to have significant OP scattering. D_{OP} used in Figs. 4.23–4.27 was lower ($2\times10^8\,\mathrm{eV/cm}$), causing scattering to be dominated by acoustic rather than optical phonons. The model values are obtained from (4.23) and (4.29). To include the positive flux reduction, (4.28) is modified by adding, for each scattering mechanism, an exponential term with the corresponding τ.

Both the model and the MSMC predict r to be constant for $\hbar\omega_{\mathrm{OP}}$ above a given energy, since elastic scattering becomes dominant. For smaller values of $\hbar\omega_{\mathrm{OP}}$, r quickly increases due to the onset of OP scattering. In fact, for very small $\hbar\omega_{\mathrm{OP}}$ the OP scattering rate greatly exceeds the acoustic scattering rate, as shown in Fig. 4.28. As a result, r approaches unity.

Summary

In summary, we have presented a new model for the backscatter coefficient, which is based on the assumption of only few scattering events in the channel. In particular, it is assumed that the back-scattered carriers encounter just a single scattering event. Different from existing models, the proposed model does not pose any restriction on the distribution of backscattered carriers. Furthermore, it accounts for the effect of degeneracy and separately takes into account both elastic and inelastic scattering. Using a simple test case, the model results have been compared with Monte–Carlo simulations, showing generally a very good agreement.

Five

Conclusions

In this final chapter, we will summarize the main findings of this work and conclude with some directions for future extension of the work addressed in this thesis.

5.1 The impact of quantization

We have discussed the general features of the carrier distribution and the potential profile in long-channel Double-Gate UTB-SOI MOSFETs. Starting with a model neglecting quantum confinement, after Taur, we have seen that in the subthreshold biasing regime the carrier density is essentially uniform throughout the UTB, corresponding to a potential profile which is approximately parabolic in the direction normal to the silicon-oxide interfaces. In strong inversion, instead, the carrier density peaks underneath the gates, at the semiconductor-oxide interfaces, whereas the carrier density in the center of the film saturates. Thus, with increasing gate bias, the surface potential becomes gradually decoupled from the center potential.

To investigate the impact of quantum confinement, which is expected to be important for UTB thicknesses in the order of the De Broglie wavelength, we have approximated the confining potential in the subthreshold regime with a square well, either having a finite or an infinite barrier height. Calculation of the subband minima and wave functions within the Effective Mass Approximation (EMA) shows that the carriers are repelled from the semiconductor-oxide interface and that, in case of purely structural (size-induced) confinement in the subthreshold regime, the carrier density peaks at the center of the UTB.

The analytical calculations have been complemented with TCAD simulations, which employ the Density Gradient model to account for the effect of quantum confinement on the carrier distribution and potential profile. The simula-

tions support the analytical results and, in addition, show that in strong inversion the maximum carrier density no longer occurs at the center of the UTB, rather the carrier distribution exhibits two peaks just away from the center, in either half of the UTB.

In the second part of Chapter 2 we have investigated the validity of the EMA, which is almost universally used to describe the size and bias-induced quantization in n-MOSFETs. In particular, we have compared the EMA results with a full-band quantization approach based on the Linear Combination of Bulk Bands (LCBB). We have studied the most relevant quantities for the modelling of the mobility and of the ON-current of the devices, namely the minima of the two-dimensional subbands, the transport masses and the electron density of states. Our study deals with both silicon and germanium n-MOSFETs with different crystal orientations and shows that, in most cases, the validity of the EMA is quite satisfactory. The LCBB approach is then used to calculate the effective mass values that significantly extend the accuracy of the EMA, and the validity of EMA-based models, towards very small UTB thicknesses. There are crystal orientations, however, where the two-dimensional energy dispersion obtained by the LCBB method exhibits features that are difficult to reproduce with the EMA model. Apart from that, the overall results on the validity of the EMA approach are quite reassuring and the accuracy of the EMA model can be improved by adjusting the transport masses for very small values of the semiconductor thickness.

5.2 Energy band offset extraction

In Chapter 3 we proposed a method to quantify the shift in conduction and valence band alignment originating from quantum confinement in long–channel UTB-SOI MOSFETs. Central to the proposed method is the temperature dependence of the subthreshold current which, being a diffusion current, is determined by the band alignment. The results have been compared with the shift in threshold voltage, which is commonly used to quantify the effect of quantum confinement.

After an initial verification with experimental data, the actual comparison was carried out with TCAD simulations employing the Density Gradient model. This study has demonstrated that with the temperature dependence of the subthreshold current, shifts in the valence and conduction band edge can be extracted distinctively from changes in mobility and density of states, making this method more accurate in assessing the impact of structural quan-

tum confinement than the commonly used threshold voltage method. Furthermore, we have indicated that with additional capacitance–voltage data a possible change in mobility and density of states can be disentangled.

5.3 Quasi-ballistic transport

Different from the previous chapters, where we studied the impact of reducing the UTB thickness, Chapter 4 focussed on the effect of decreasing the lateral dimension of the MOSFET, i.e. the channel length. If the channel length is in the order of the carrier mean-free-path, carrier transport becomes quasi-ballistic (QB). The chapter started with an outline of the key properties of the QB transport regime, along with a concise summary of the Flux theory and the Virtual Source model, which are commonly used to study and to help explain QB transport phenomena.

In order to clarify some uncertainties related to the well-known Virtual Source model, we proceeded with a detailed analysis of the backscatter characteristics using a sophisticated Multi-Subband Monte-Carlo (MSMC) simulator. Besides the backscatter coefficient, we extracted from the MSMC simulations the distribution of backscatter events, its centroid and the critical length for back-scattering according to the various definitions reported in literature.

We have performed simulations of a realistic 32 nm UTB-SOI MOSFET to study the impact of optical phonon (OP) scattering on the critical length, i.e. the region of the channel where back-scattering is most pronounced. Although the OP scattering rate depends on the OP energy $\hbar\omega_{\mathrm{OP}}$, our simulations indicate that $\hbar\omega_{\mathrm{OP}}$ and the corresponding '$\hbar\omega_{\mathrm{OP}}$-layer' do not directly determine the critical length for back-scattering.

In addition, we have systematically analyzed back-scattering in a template device structure with a 'frozen' field, in order to eliminate effects related to self-consistency. The simulations show that changes in temperature and longitudinal field do not have significant impact on the centroid which, on its turn, reflects the portion of the channel in which backscattering is most pronounced. Furthermore, we found that although the centroid increases with the channel length, the backscatter events remain strongly weighted towards the (virtual) source.

Using the same template device, we extracted the effective critical length (L_{extr}) from the MSMC simulations assuming that the backscatter coefficient r can be expressed in the form $r = L_{\mathrm{extr}}/(L_{\mathrm{extr}} + \lambda)$. We found that, in support of the '$k_{\mathrm{B}}T$-layer' theory, under high-field conditions L_{extr} exhibits strong cor-

relation with L_{kT}, whereas L_{extr} follows L in the low-field limit. Furthermore, we found that the appropriate low-field mean-free-path λ to be used in the aforementioned flux-like equation is $\lambda = 2v_{\mathrm{th}}\tau$, with τ the inverse scattering rate, whereas in the Flux model $\lambda = \frac{k_B T}{q}\frac{2\mu}{v_{\mathrm{th}}}$.

In the last part of Chapter 4 we propose a new model for the backscatter coefficient in nanoscale MOSFETs. The model assumes that only few scattering events occur, so that the carrier transport is nearly ballistic. The proposed model accounts for both elastic and inelastic scattering mechanisms and it naturally captures the effect of degeneracy. We have compared the model with results obtained from MSMC simulations for a broad range of channel lengths, temperatures, and electric fields, obtaining overall a very good agreement. The new model is remarkably simple and efficient, and although it captures only the basics of a single-scattering process, it provides a useful framework to help explain and predict QB transport phenomena in nanoscale MOSFETs.

5.4 Recommendations and closing remarks

We conclude this chapter with some recommendations for further work, in the order as the respective topics appear in this thesis.

Band offset measurements

In this thesis, we have applied the $I_{\mathrm{DS}}(T)$ method to quantify shifts in the band alignment due to structural carrier confinement in UTB-SOI MOSFETs. Other application areas of the $I_{\mathrm{DS}}(T)$ method include strain, which alters the band alignment and mobility.

The initial measurements on the in-house fabricated long-channel Single-Gate UTB-SOI MOSFETs served as a 'proof of principle', demonstrating that the temperature dependence of the subthreshold current can be exploited to extract shifts in the conduction and valence band alignment. The actual detailed quantitative analysis and comparison with the 'threshold-voltage' method was carried out with simulations, in the same manner as we had performed the measurements. It would be worthwhile to complement the existing 'simulated experiments' with actual experimental results obtained from IV and CV measurements on e.g. FinFETs for a wide range of fin widths and channel lengths. At present, although preliminary FinFET band offset measurements have been performed, conclusive experimental data is not yet available.

New model for the backscatter coefficient

Although the proposed model does not intend to provide a complete, physically based, description of the scattering mechanisms, we comment on some possible extensions that would bring the model closer to the prediction of the QB current in an actual MOSFET, in addition to the case studies presented in this work.

In the 1D single-scattering model the potential profile along the transport direction is an input and it is treated as 'frozen'. We have seen in Chapter 4 that using the linear or the parabolic potential profile results in a very similar qualitative field, temperature and channel length dependence. These simple profiles are usually a reasonable approximation for the potential profile in a real device, at least near the top of the source-channel barrier where scattering events are most pronounced. In this regard, we do not expect to observe markedly different behavior when employing a 1D potential profile taken from self-consistent calculations. However a thorough comparison of the MSMC and single-scattering model results for potential profiles other than linear or parabolic is still to be carried out.

When comparing the model and MSMC results, we have employed a template material with a single spherical valley. The coupling constant of the acoustic phonons was adjusted to achieve a reasonable low-field mobility of around 400 cm^2/Vs. Since the material is parameterized in terms of the effective masses and the phonon coupling constants, we could do a similar comparison using the appropriate parameters for silicon. However, we should bear in mind that just using the Si effective (transport) mass while considering only elastic scattering in single subband yields a low-field mobility of around 3000 cm^2/Vs, i.e. values which are not realistic in actual silicon inversion layers. A fair comparison of model and MSMC results in actual silicon inversion layers requires the model to account for, e.g., the impact of surface roughness scattering.

So far the model provides an expression for r, the backscatter ratio. The model results can be compared to the MSMC calculations in a clear and straightforward manner, since the MSMC simulator provides direct access to the internal fluxes anywhere in the channel. In order to relate r to the external terminal voltages and currents, we have to link the potential and the carrier distributions at the VS to the 'outside' world. One may be tempted to use (4.8) (p. 85) to this end. We recall, however, that (4.8) was derived assuming near-equilibrium conditions for both the positive and the negative flux. The

new model for r removes the constraint on the negative flux, and only requires the injected flux at the VS to be close to thermal equilibrium. Under single-scattering conditions, i.e. the intended regime of the new model, the negative flux can be strongly non-thermal thereby invalidating (4.8). Thus, a recommendation for further work is to reconsider the equilibrium assumption on the negative flux in the VS model. Furthermore, an appropriate 'single-scattering' VS model is a prerequisite for a comparison with experimental data.

The model results presented in this work have been obtained by numerical evaluation of the integrals in (4.23) (p. 108). It would be useful to have analytical closed form expressions for r in some special cases, such as a linear potential profile. These (supposedly) simple expressions could provide information on the field, temperature and channel length dependence at a glance. Furthermore, it is particularly interesting to find an analytical expression for the effective critical length in the case of degenerate statistics. It was reported that the $k_B T$-layer concept is a consequence of the assumed Maxwellian carrier distribution functions, rather than being related to the average carrier energy. Thus, assuming degenerate statistics, we may find an effective critical length which is not as tightly related to L_{kT} as the non-degenerate critical length.

Bibliography

[1] F. Wanlass and C. Sah, "Nanowatt logic using field-effect metal-oxide semiconductor triodes," in *ISSCC*, vol. VI, Feb. 1963, pp. 32 – 33.

[2] J. E. Lilienfeld, "Method and apparatus for controlling electric currents," US Patent 1,745,175, 1930.

[3] ——, "Amplifier for electric current," US Patent 1,877,140, 1932.

[4] ——, "Device for controlling electric currents," US Patent 1,900,018, 1933.

[5] D. Kahng and M. M. Atalla, "Silicon-silicon dioxide field induced surface devices," in *IRE-AIEE Solid-State Device Res. Conf.* Carnegie Inst. of Technol., Pittsburgh, PA, 1960.

[6] J. Ligenza and W. Spitzer, "The mechanisms for silicon oxidation in steam and oxygen," *Jrnl. of Physics and Chemistry of Solids*, vol. 14, pp. 131 – 136, 1960.

[7] G. E. Moore, "Cramming More Components Onto Integrated Circuits," *Electronics Magazine*, vol. 38, no. 8, pp. 114 – 117, 1965.

[8] ——, "Progress in Digital Integrated Electronics," 1975, pp. 11–13.

[9] "International Technology Roadmap for Semiconductors," http://www.itrs.net, 2009.

[10] C.-T. Sah, "Evolution of the MOS transistor - from conception to VLSI," *Proc. IEEE*, vol. 76, no. 10, pp. 1280–1326, 1988.

[11] S. E. Thompson, R. S. Chau, T. Ghani, K. Mistry, S. Tyagi, and M. T. Bohr, "In search of "Forever," continued transistor scaling one new material at a time," *IEEE Trans. Semicond. Manuf.*, vol. 18, no. 1, pp. 26–36, 2005.

[12] S. M. Sze and K. K. Ng, *Physics of Semiconductor Devices*, 3rd ed. John Wiley & Sons, Inc., Hoboken, New Jersey, 2007.

[13] 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," *Proc. IEEE*, vol. 89, no. 3, pp. 259 –288, Mar. 2001.

[14] K. C. Saraswat, C. O. Chui, D. Kim, T. Krishnamohan, and A. Pethe, "High Mobility Materials and Novel Device Structures for High Performance Nanoscale MOSFETs," in *Proc. IEDM*, 2006, pp. 1–4.

[15] T. Krishnamohan, Z. Krivokapic, K. Uchida, Y. Nishi, and K. C. Saraswat, "High-mobility ultrathin strained Ge MOSFETs on bulk and SOI with low band-to-band tunneling leakage: experiments," *IEEE Trans. Electron Devices*, vol. 53, no. 5, pp. 990–999, 2006.

[16] D. Kim, T. Krishnamohan, and K. C. Saraswat, "Performance Evaluation of III-V Double-Gate n-MOSFETs," in *Proc. Dev. Res. Conf.*, 2008, pp. 67–68.

[17] S. E. Thompson, M. Armstrong, C. Auth, M. Alavi, M. Buehler, R. Chau, S. Cea, T. Ghani, G. Glass, T. Hoffman, C.-H. Jan, C. Kenyon, J. Klaus, K. Kuhn, Z. Ma, B. Mcintyre, K. Mistry, A. Murthy, B. Obradovic, R. Nagisetty, P. Nguyen, S. Sivakumar, R. Shaheed, L. Shifren, B. Tufts, S. Tyagi, M. Bohr, and Y. El-Mansy, "A 90-nm logic technology featuring strained-silicon," *IEEE Trans. Electron Devices*, vol. 51, no. 11, pp. 1790–1797, 2004.

[18] D. Hisamoto, T. Kaga, Y. Kawamoto, and E. Takeda, "A fully depleted lean-channel transistor (DELTA)-a novel vertical ultra thin SOI MOSFET," in *Proc. IEDM*, 1989, pp. 833–836.

[19] D. Hisamoto, W. Lee, J. Kedziekski, E. Anderson, H. Takeuchi, K. Asano, T. King, J. Bokor, and C. Hu, "A Folded-channel MOSFET for Deep-subtenth Micron Era," in *Proc. IEDM*, 1998, pp. 1032–1034.

[20] D. Hisamoto, W. Lee, J. Kedziekski, H. Takeuchi, K. Asano, C. Kuo, E. Anderson, T. King, J. Bokor, and C. Hu, "FinFET—A Self-Aligned Double-Gate MOSFET Scalable to 20 nm," *IEEE Trans. Electron Devices*, vol. 47, no. 12, pp. 2320–2325, December 2000.

[21] B. Yu, L. Chang, S. Ahmed, H. Wang, S. Bell, C.-Y. Yang, C. Tabery, C. Ho, Q. Xiang, T.-J. King, J. Bokor, C. Hu, M.-R. Lin, and D. Kyser, "FinFET scaling to 10 nm gate length," in *Proc. IEDM*, 2002, pp. 251–254.

[22] J.-P. Colinge, Ed., *FinFETs and Other Multi-Gate Transistors*, ser. on Integrated Circuits and Systems. Springer, 2008.

[23] R.-H. Yan, A. Ourmazd, and K. Lee, "Scaling the Si MOSFET: from bulk to SOI to bulk," *IEEE Trans. Electron Devices*, vol. 39, no. 7, pp. 1704 – 1710, Jul. 1992.

[24] D. Frank, Y. Taur, and H.-S. Wong, "Generalized scale length for two-dimensional effects in MOSFETs," *IEEE Electron Device Lett.*, vol. 19, no. 10, pp. 385 – 387, Oct. 1998.

[25] S.-H. Oh, D. Monroe, and J. Hergenrother, "Analytic description of short-channel effects in fully-depleted double-gate and cylindrical, surrounding-gate MOSFETs," *IEEE Electron Device Lett.*, vol. 21, no. 9, pp. 445 –447, Sep. 2000.

[26] A. Kloes, M. Weidemann, D. Goebe, and B. T. Bosworth, "Closed-form physics-based models for threshold voltage and subthreshold slope in FinFETs including 3D effects," in *Proc. Int. Semicond. Dev. Research Symp.*, 2007, pp. 1–2.

[27] A. Kloes, M. Weidemann, D. Goebel, and B. T. Bosworth, "Three-Dimensional Closed-Form Model for Potential Barrier in Undoped Fin-FETs Resulting in Analytical Equations for V_T and Subthreshold Slope," *IEEE Trans. Electron Devices*, vol. 55, no. 12, pp. 3467–3475, 2008.

[28] F. Balestra, S. Cristoloveanu, M. Benachir, J. Brini, and T. Elewa, "Double-Gate Silicon-on-Insulator Transistor with Volume Inversion: A New Device with Greatly Enhanced Performance," *IEEE Electron Device Lett.*, vol. 8, no. 9, pp. 410–412, September 1987.

[29] P. Yeh and J. Fossum, "Physical Subthreshold MOSFET Modeling Applied to Viable Design of Deep-Submicrometer Fully Depleted SOI Low-Voltage CMOS Technology," *IEEE Trans. Electron Devices*, vol. 42, no. 9, pp. 1605–1613, September 1995.

[30] Y. Taur, "Analytic Solution of Charge and Capacitance in Symmetric and Asymmetric Double-Gate MOSFETs," *IEEE Trans. Electron Devices*, vol. 48, no. 12, pp. 2861–2869, December 2001.

[31] M. Pelgrom, A. Duinmaijer, and A. Welbers, "Matching properties of MOS transistors," *IEEE J. Solid-State Circuits*, vol. 24, no. 5, pp. 1433 – 1439, oct 1989.

[32] H. Tuinhout, "Impact of Parametric Fluctuations on Performance and Yield of Deep-Submicron Technologies," in *Proc. ESSDERC*, 2002, pp. 95–102.

[33] B. Cheng, S. Roy, A. R. Brown, C. Millar, and A. Asenov, "Evaluation of intrinsic parameter fluctuations on 45, 32 and 22nm technology node LP N-MOSFETs," in *Proc. ESSDERC*, 2008, pp. 47–50.

[34] A. Asenov, "Simulation of Statistical Variability in Nano MOSFETs," in *Proc. Symp. on VLSI Techn.*, 2007, pp. 86–87.

[35] 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 *Proc. Symp. on VLSI Techn.*, 2007, pp. 110–111.

[36] 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.

[37] K. Uchida, H. Watanabe, A. Kinoshita, J. Koga, T. Numata, and S. Takagi, "Experimental Study on Carrier Transport Mechanism in Ultrathin-body SOI n- and p-MOSFETs with SOI Thickness less than 5 nm," in *Proc. IEDM*, 2002, pp. 47–50.

[38] D. Esseni, A. Abramo, L. Selmi, and E. Sangiorgi, "Physically based modeling of low field electron mobility in ultrathin single- and double-gate SOI n-MOSFETs," *IEEE Trans. Electron Devices*, vol. 50, no. 12, pp. 2445–2455, 2003.

[39] T. Ernst, S. Cristoloveanu, G. Ghibaudo, T. Ouisse, S. Horiguchi, Y. Ono, Y. Takahashi, and K. Murase, "Ultimately Thin Double-Gate SOI MOSFETs," *IEEE Trans. Electron Devices*, vol. 50, no. 3, pp. 830–838, March 2003.

[40] K. Uchida, J. Koga, and S. Takagi, "Experimental Study on Carrier Transport Mechanisms in Double- and Single-Gate Ultrathin-Body MOSFETs," in *Proc. IEDM*, 2003, pp. 805–808.

[41] S. Jin, M. V. Fischetti, and T.-W. Tang, "Modeling of Surface-Roughness Scattering in Ultrathin-Body SOI MOSFETs," *IEEE Trans. Electron Devices*, vol. 54, no. 9, pp. 2191–2203, 2007.

[42] R. Duffy, M. Van Dal, B. Pawlak, M. Kaiser, R. Weemaes, B. Degroote, E. Kunnen, and E. Altamirano, "Solid phase epitaxy versus random nucleation and growth in sub-20 nm wide fin field-effect transistors," *Applied Physics Letters*, vol. 90, no. 24, p. 241912, 2007.

[43] J. Kedzierski, M. Ieong, E. Nowak, T. S. Kanarsky, Y. Zhang, R. Roy, D. Boyd, D. Fried, and H.-S. P. Wong, "Extension and source/drain design for high-performance FinFET devices," *IEEE Trans. Electron Devices*, vol. 50, no. 4, pp. 952–958, 2003.

[44] M. Kumar, S. Gupta, and V. Venkataraman, "Compact modeling of the effects of parasitic internal fringe capacitance on the threshold voltage of high-κ gate-dielectric nanoscale SOI MOSFETs," *IEEE Trans. Electron Devices*, vol. 53, no. 4, pp. 706 – 711, 2006.

[45] P. Wambacq, B. Verbruggen, K. Scheir, J. Borremans, M. Dehan, D. Linten, V. De Heyn, G. Van der Plas, A. Mercha, B. Parvais, C. Gustin,

V. Subramanian, N. Collaert, M. Jurczak, and S. Decoutere, "The Potential of FinFETs for Analog and RF Circuit Applications," *IEEE Trans. Circuits Syst. I*, vol. 54, no. 11, pp. 2541–2551, 2007.

[46] C. Riddet, A. R. Brown, C. L. Alexander, J. R. Watling, S. Roy, and A. Asenov, "3-D Monte Carlo Simulation of the Impact of Quantum Confinement Scattering on the Magnitude of Current Fluctuations in Double Gate MOSFETs," *IEEE Trans. Nanotechnol.*, vol. 6, no. 1, pp. 48–55, 2007.

[47] J. A. van der Pol, A. W. Ludikhuize, H. G. A. Huizing, B. van Velzen, R. J. E. Hueting, J. F. Mom, G. van Lijnschoten, G. J. J. Hessels, E. F. Hooghoudt, R. van Huizen, M. J. Swanenberg, J. H. H. A. Egbers, F. van den Elshout, J. J. Koning, H. Schligtenhorst, and J. Soeteman, "A-BCD: An economic 100 V RESURF silicon-on-insulator BCD technology for consumer and automotive applications," in *Proc. ISPSD*, 2000, pp. 327–330.

[48] B. Krabbenborg and J. van der Pol, "Robustness of LDMOS power transistors in SOI-BCD processes and derivation of design rules using thermal simulation," in *Proc. ISPSD*, 2001, pp. 157 –160.

[49] T. Vanhoucke and G. Hurkx, "A new analytical model for the thermal resistance of deep-trench bipolar transistors," *IEEE Trans. Electron Devices*, vol. 53, no. 6, pp. 1379 – 1388, june 2006.

[50] W. Molzer, T. Schulz, W. Xiong, R. Cleavelin, K. Schrüfer, A. Marshall, K. Matthews, J. Sedlmeir, D. Siprak, G. Knoblinger, L. Bertolissi, P. Patruno, and J.-P. Colinge, "Self Heating Simulation of Multi-Gate FETs," in *Proc. ESSDERC*, 2006, pp. 311–314.

[51] S. Kumar, R. Joshi, C.-T. Chuang, K. Kim, J. Murthy, K. Schonenberg, and E. Nowak, "Self-Consistent and Efficient Electro-Thermal Analysis for Poly/Metal Gate FinFETs," in *Proc. IEDM*, 2006, pp. 803–806.

[52] S. Kolluri, K. Endo, E. Suzuki, and K. Banerjee, "Modeling and Analysis of Self-Heating in FinFET Devices for Improved Circuit and EOS/ESD Performance," in *Proc. IEDM*, 2007, pp. 177–180.

[53] C. Fiegna, Y. Yang, E. Sangiorgi, and A. O'Neill, "Analysis of Self-Heating Effects in Ultrathin-Body SOI MOSFETs by Device Simulation," *IEEE Trans. Electron Devices*, vol. 55, no. 1, pp. 233–244, January 2008.

[54] M. Braccioli, G. Curatola, Y. Yang, E. Sangiorgi, and C. Fiegna, "Simulation of self-heating effects in different SOI MOS architectures," *SSE*, vol. 53, no. 4, pp. 445–451, 2009.

[55] W. Liu, K. Etessam-Yazdani, R. Hussin, and M. Asheghi, "Modeling and Data for Thermal Conductivity of Ultrathin Single-Crystal SOI Layers at

High Temperature," *IEEE Trans. Electron Devices*, vol. 53, no. 8, pp. 1868–1876, August 2006.

[56] K. Jenkins and J. Sun, "Measurement of I-V Curves of Silicon-on-Insulator (SOI) MOSFET's without Self-Heating," *IEEE Electron Device Lett.*, vol. 16, no. 4, pp. 145–147, April 1995.

[57] N. Collaert, M. Demand, I. Ferain, J. Lisoni, R. Singanamalla, P. Zimmerman, Y. Yim, T. Schram, G. Mannaert, M. Goodwin, J. Hooker, F. Neuilly, M. Kim, K. De Meyer, S. De Gendt, W. Boullart, M. Jurczak, and S. Biesemans, "Tall triple-gate devices with TiN/HfO_2 gate stack," in *Proc. Symp. on VLSI Techn.*, 2005, pp. 108–109.

[58] A. Scholten, G. Smit, R. Pijper, L. Tiemeijer, H. Tuinhout, J.-L. van der Steen, A. Mercha, M. Braccioli, and D. Klaassen, "Experimental assessment of self-heating in SOI FinFETs," in *Proc. IEDM*, dec. 2009, pp. 1–4.

[59] N. Rinaldi, "Small-signal operation of semiconductor devices including self-heating, with application to thermal characterization and instability analysis," *IEEE Trans. Electron Devices*, vol. 48, no. 2, pp. 323–331, feb 2001.

[60] G. Smit, A. Scholten, N. Serra, R. Pijper, R. van Langevelde, A. Mercha, G. Gildenblat, and D. Klaassen, "PSP-based compact FinFET model describing DC and RF measurements," in *Proc. IEDM*, dec. 2006, pp. 1–4.

[61] 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 Trans. Electron Devices*, vol. 53, no. 4, pp. 759–768, 2006.

[62] G. Wilk, R. Wallace, and J. Anthony, "High-κ gate dielectrics: Current status and materials properties considerations," *Journal of Applied Physics*, vol. 89, no. 10, pp. 5243–5275, May 2001.

[63] P. Andricciola, H. P. Tuinhout, B. De Vries, N. A. H. Wils, A. J. Scholten, and D. B. M. Klaassen, "Impact of interface states on MOS transistor mismatch," in *Proc. IEDM*, 2009, pp. 1–4.

[64] M. P. Lepselter and S. M. Sze, "SB-IGFET: An insulated-gate field-effect transistor using Schottky barrier contacts for source and drain," *Proc. IEEE*, vol. 56, no. 8, pp. 1400–1402, 1968.

[65] M. Fritze, C. L. Chen, S. Calawa, D. Yost, B. Wheeler, P. Wyatt, C. L. Keast, J. Snyder, and J. Larson, "High-speed Schottky-barrier pMOSFET with f_T = 280 GHz," *IEEE Electron Device Lett.*, vol. 25, no. 4, pp. 220–222, 2004.

[66] J. M. Larson and J. P. Snyder, "Overview and status of metal S/D Schottky-barrier MOSFET technology," *IEEE Trans. Electron Devices*, vol. 53, no. 5, pp. 1048–1058, 2006.

[67] S. Xiong, T.-J. King, and J. Bokor, "A comparison study of symmetric ultrathin-body double-gate devices with metal source/drain and doped source/drain," *IEEE Trans. Electron Devices*, vol. 52, no. 8, pp. 1859–1867, aug. 2005.

[68] J. Knoch, M. Zhang, S. Mantl, and J. Appenzeller, "On the performance of single-gated ultrathin-body SOI Schottky-barrier MOSFETs," *IEEE Trans. Electron Devices*, vol. 53, no. 7, pp. 1669–1674, 2006.

[69] P. Solomon, "Inability of Single Carrier Tunneling Barriers to Give Subthermal Subthreshold Swings in MOSFETs," *IEEE Electron Device Lett.*, vol. 31, no. 6, pp. 618 – 620, 2010.

[70] R. W. Keyes, "High-mobility FET in strained silicon," *IEEE Trans. Electron Devices*, vol. 33, no. 6, p. 863, 1986.

[71] J. Welser, J. Hoyt, and J. Gibbons, "Electron mobility enhancement in strained-Si n-type metal-oxide-semiconductor field-effect transistors," *IEEE Electron Device Lett.*, vol. 15, no. 3, pp. 100 –102, Mar. 1994.

[72] F. M. Bufler, P. Graf, S. Keith, and B. Meinerzhagen, "Full band Monte Carlo investigation of electron transport in strained Si grown on $Si_{1-x}Ge_x$ substrates," *Applied Physics Letters*, vol. 70, no. 16, pp. 2144–2146, 1997.

[73] J. L. Hoyt, H. M. Nayfeh, S. Eguchi, I. Aberg, G. Xia, T. Drake, E. A. Fitzgerald, and D. A. Antoniadis, "Strained silicon MOSFET technology," in *Proc. IEDM*, 2002, pp. 23–26.

[74] 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 *Proc. IEDM*, dec. 2009, pp. 1–4.

[75] A. Ortiz-Conde, F. García-Sánchez, J. Muci, S. Malobabic, and J. Liou, "A Review of Core Compact Models for Undoped Double-Gate SOI MOSFETs," *IEEE Trans. Electron Devices*, vol. 54, no. 1, pp. 131–139, January 2007.

[76] Y. Taur, "An Analytical Solution to a Double-Gate MOSFET with Undoped Body," *IEEE Electron Device Lett.*, vol. 21, no. 5, pp. 245–247, May 2000.

[77] H. Pao and C. Sah, “Effects of diffusion current on characteristics of metal-oxide (insulator)-semiconductor transistors,” *Solid-State Electr.*, vol. 9, no. 10, pp. 927 – 937, 1966.

[78] N. Arora, *MOSFET Modeling for VLSI Simulation: Theory and Practice*. World Scientific, 2007, ch. 9, pp. 438–443.

[79] S. Cristoloveanu and D. Ioannou, “Adjustable confinement of the electron gas in Dual-Gate Silicon-On-Insulator MOSFETs,” *Superlatt. and Microstr.*, vol. 8, no. 1, pp. 131–135, August 1990.

[80] R. d. L. Kronig and W. G. Penney, “Quantum Mechanics of Electrons in Crystal Lattices,” *Proceedings of the Royal Society of London, Series A*, vol. 130, no. 814, pp. 499–513, Feb. 1931.

[81] R. Pierret, *Advanced Semiconductor Fundamentals*, ser. Modular series on solid state devices. Addison-Wesley, 1987, vol. VI.

[82] D. J. Griffiths, *Introduction to quantum mechanics*. Prentice Hall, Inc., 1995.

[83] C. Jacoboni and L. Reggiani, “The Monte Carlo method for the solution of charge transport in semiconductors with applications to covalent materials,” *Rev. Mod. Phys.*, vol. 55, no. 3, pp. 645–705, Jul 1983.

[84] C. Jungemann, A. Emunds, and W. Engl, “Simulation of linear and nonlinear electron transport in homogeneous silicon inversion layers,” *Solid-State Electr.*, vol. 36, no. 11, pp. 1529– 1540, 1993.

[85] F. Stern and W. E. Howard, “Properties of semiconductor surface inversion layers in the electric quantum limit,” *Phys. Rev.*, vol. 163, no. 3, pp. 816–835, November 1967.

[86] F. Stern, “Self-consistent results for n-type Si inversion layers,” *Phys. Rev. B*, vol. 5, no. 12, pp. 4891–4899, June 1972.

[87] C. Moglestue, “Self-consistent calculation of electron and hole inversion layers at silicon-silicon dioxide interfaces,” *Journal of Applied Physics*, vol. 59, pp. 3175–3183, May 1986.

[88] S. Datta, *Quantum Phenomena*, ser. Modular Series on Solid-State Devices, R. Pierret, Ed. Addison-Wesley, 1989, vol. 8.

[89] M. De Michielis, D. Esseni, and F. Driussi, “Analytical Models for the Insight Into the Use of Alternative Channel Materials in Ballistic nano-MOSFETs,” *IEEE Trans. Electron Devices*, vol. 54, no. 1, pp. 115–123, January 2007.

[90] M. Abramowitz and I. A. Stegun, Eds., *Handbook of mathematical functions with formulas, graphs, and mathematical tables*. Dover Publications, Inc., 1964.

[91] G. Baccarani and S. Reggiani, "A Compact Double-Gate MOSFET Model Comprising Quantum-Mechanical and Nonstatic Effects," *IEEE Trans. Electron Devices*, vol. 46, no. 8, pp. 1656–1666, August 1999.

[92] L. Ge and J. Fossum, "Analytical Modeling of Quantization and Volume Inversion in Thin Si-Film DG MOSFETs," *IEEE Trans. Electron Devices*, vol. 49, no. 2, pp. 287–294, February 2002.

[93] D. Ferry, *Quantum mechanics: an introduction for device physicists and electrical engineers*. Institute of Physics Publishing, 1995.

[94] Synopsys, "Sentaurus TCAD, v. Z-2007.03."

[95] A. Wettstein, A. Schenk, and W. Fichtner, "Quantum Device-Simulation with the Density-Gradient Model on Unstructured Grids," *IEEE Trans. Electron Devices*, vol. 48, no. 2, pp. 279–284, February 2001.

[96] L. Lucci, P. Palestri, D. Esseni, and L. Selmi, "Multi-subband Monte Carlo modeling of nano-MOSFETs with strong vertical quantization and electron gas degeneration," in *Proc. IEDM*, 2005, pp. 617–620.

[97] M. V. Fischetti and S. E. Laux, "Monte Carlo study of electron transport in silicon inversion layers," *Phys. Rev. B*, vol. 48, no. 4, pp. 2244–2274, Jul 1993.

[98] S. C. Williams, K. W. Kim, and W. C. Holton, "Ensemble Monte Carlo study of channel quantization in a 25-nm n-MOSFET," *IEEE Trans. Electron Devices*, vol. 47, no. 10, pp. 1864–1872, 2000.

[99] J.-L.P.J. van der Steen, D. Esseni, P. Palestri, L. Selmi, and R.J.E. Hueting, "Validity of the Parabolic Effective Mass Approximation in Silicon and Germanium n-MOSFETs With Different Crystal Orientations," *IEEE Trans. Electron Devices*, vol. 54, no. 8, pp. 1843–1851, August 2007.

[100] L.-W. Wang and A. Zunger, "Linear combination of bulk bands method for large-scale electronic structure calculations on strained nanostructures," *Phys. Rev. B*, vol. 59, no. 24, pp. 15 806–15 818, Jun 1999.

[101] F. Chirico, A. Di Carlo, and P. Lugli, "Efficient self-consistent pseudopotential calculation of nanostructured devices," *Phys. Rev. B*, vol. 64, no. 4, p. 045314, Jun 2001.

[102] D. Esseni and P. Palestri, "Linear combination of bulk bands method for investigating the low-dimensional electron gas in nanostructured devices," *Phys. Rev. B*, vol. 72, no. 16, pp. 1–14, 2005.

[103] J. C. Slater and G. F. Koster, "Simplified LCAO Method for the Periodic Potential Problem," *Phys. Rev.*, vol. 94, no. 6, pp. 1498–1524, Jun 1954.

[104] A. Di Carlo, "Microscopic theory of nanostructured semiconductor devices: beyond the envelope-function approximation," *Semicond. Science and Technology*, vol. 18, no. 1, p. R1, 2003.

[105] T. B. Boykin, G. Klimeck, M. Friesen, S. N. Coppersmith, P. von Allmen, F. Oyafuso, and S. Lee, "Valley splitting in low-density quantum-confined heterostructures studied using tight-binding models," *Phys. Rev. B*, vol. 70, no. 16, p. 165325, Oct 2004.

[106] D. Esseni and P. Palestri, "Fullband Quantization Analysis Reveals a Third Valley in (001) Silicon Inversion Layers," *IEEE Electron Device Lett.*, vol. 26, no. 6, pp. 413–415, June 2005.

[107] J. R. Chelikowsky and M. L. Cohen, "Nonlocal pseudopotential calculations for the electronic structure of eleven diamond and zinc-blende semiconductors," *Phys. Rev. B*, vol. 14, no. 2, pp. 556–582, Jul 1976.

[108] M. V. Fischetti and S. E. Laux, "Monte Carlo simulation of transport in technologically significant semiconductors of the diamond and zinc-blende structures. II. Submicrometer MOSFET's," *IEEE Trans. Electron Devices*, vol. 38, no. 3, pp. 650–660, 1991.

[109] T. Ando, A. B. Fowler, and F. Stern, "Electronic properties of two-dimensional systems," *Rev. Mod. Phys.*, vol. 54, no. 2, pp. 437–672, Apr 1982.

[110] D. K. Ferry and S. M. Goodnick, *Transport in Nanostructures*. Cambridge University Press, 1997.

[111] T. Low, Y. T. Hou, M. F. Li, C. Zhu, A. Chin, G. Samudra, L. Chan, and D.-L. Kwong, "Investigation of performance limits of germanium double-gated MOSFETs," in *Proc. IEDM*, 2003, p. 691.

[112] S. E. Laux, "Simulation study of Ge n-channel 7.5 nm DGFETs of arbitrary crystallographic alignment," in *Proc. IEDM*, 2004, pp. 135–138.

[113] M. De Michielis, D. Esseni, and F. Driussi, "Trade-off between electron velocity and density of states in ballistic nano-MOSFETs," in *Proc. ESSDERC*, 2005, pp. 165–168.

[114] J. Wang, A. Rahman, A. Ghosh, G. Klimeck, and M. Lundstrom, "On the validity of the parabolic effective-mass approximation for the I-V calculation of silicon nanowire transistors," *IEEE Trans. Electron Devices*, vol. 52, no. 7, pp. 1589–1595, 2005.

[115] B. Delley and E. Steigmeier, "Size dependence of band gaps in silicon nanostructures," *Applied Physics Letters*, vol. 67, no. 16, pp. 2370–2372, October 1995.

[116] M. Tabe, M. Kumezawa, and Y. Ishikawa, "Quantum-Confinement Effect in Ultrathin Si Layer of Silicon-on-Insulator Substrate," *Jap. Jrnl. of Applied Physics*, vol. 40, no. Part 2, No. 2B, pp. L131–L133, 2001.

[117] K. Uchida, J. Koga, T. Numata, and S. Takagi, "Experimental Evidences of Quantum-Mechanical Effects on Low-field Mobility, Gate-channel Capacitance, and Threshold Voltage of Ultrathin Body SOI MOSFETs," in *Proc. IEDM*, 2001, pp. 633–636.

[118] J. Slotboom and H. de Graaff, "Measurements of bandgap narrowing in Si bipolar transistors," *Solid-State Electr.*, vol. 19, no. 10, pp. 857 – 862, 1976.

[119] J. Slotboom, G. Streutker, A. Pruijmboom, and D. Gravesteijn, "Parasitic Energy Barriers in SiGe HBT's," *IEEE Electron Device Lett.*, vol. 12, no. 9, pp. 486–488, September 1991.

[120] Z. Lu and D. Grozea, "Crystalline Si/SiO_2 quantum wells," *Applied Physics Letters*, vol. 80, no. 2, pp. 255–257, January 2002.

[121] M. van Dort, P. Woerlee, and A. Walker, "A simple model for quantisation effects in heavily-doped silicon MOSFETs at inversion conditions," *Solid-State Electr.*, vol. 37, no. 3, pp. 411 – 414, 1994.

[122] Y. Omura, K. Kurihara, Y. Takahashi, T. Ishiyama, Y. Nakajima, and K. Izumi, "50-nm Channel nMOSFET/SIMOX with an Ultrathin 2- or 6-nm Thick Silicon Layer and Their Significant Features of Operations," *IEEE Electron Device Lett.*, vol. 18, no. 5, pp. 190–193, May 1997.

[123] H. Majima, H. Ishikuro, and T. Hiramoto, "Experimental Evidence for Quantum Mechanical Narrow Channel Effect in Ultra-Narrow MOSFETs," *IEEE Electron Device Lett.*, vol. 21, no. 8, pp. 396–398, August 2000.

[124] G. Tsutsui, M. Saitoh, T. Nagumo, and T. Hiramoto, "Impact of SOI thickness Fluctuation on Threshold Voltage Variation in Ultra-Thin Body SOI MOSFETs," *IEEE Trans. Nanotechnol.*, vol. 4, no. 3, pp. 369–373, May 2005.

[125] V. Trivedi and J. Fossum, "Quantum-Mechanical Effects on the Threshold Voltage of Undoped Double-Gate MOSFETs," *IEEE Electron Device Lett.*, vol. 26, no. 8, pp. 579–582, August 2005.

[126] J. Croon, H. Tuinhout, R. Difrenza, J. Knol, A. Moonen, S. Decoutere, H. Maes, and W. Sansen, "A comparison of extraction techniques for threshold voltage mismatch," in *Proc. ICMTS*, 2002, pp. 235–240.

[127] J.-L.P.J. van der Steen, R.J.E. Hueting, G.D.J. Smit, T. Hoang, J. Holleman, and J. Schmitz, "Valence Band Offset Measurements on Thin Silicon–On–Insulator MOSFETs," *IEEE Electron Device Lett.*, vol. 28, no. 9, pp. 821–824, September 2007.

[128] J.-L.P.J. van der Steen, R.J.E. Hueting, and J. Schmitz, "Extracting Energy Band Offsets on Thin Silicon–On–Insulator MOSFETs," in *Proc. ESSDERC*, 2008, pp. 242–245.

[129] J.-L. P. J. van der Steen, R. J. E. Hueting, and J. Schmitz, "Extracting Energy Band Offsets on Long-Channel Thin Silicon-on-Insulator MOSFETs," *IEEE Trans. Electron Devices*, vol. 56, no. 9, pp. 1999–2007, 2009.

[130] T. Hoang, "High efficient infrared-light emission from silicon LEDs," Ph.D. dissertation, University of Twente, Enschede, September 2007.

[131] M. G. Ancona and H. F. Tiersten, "Macroscopic physics of the silicon inversion layer," *Phys. Rev. B*, vol. 35, no. 15, pp. 7959–7965, May 1987.

[132] M. G. Ancona and G. J. Iafrate, "Quantum correction to the equation of state of an electron gas in a semiconductor," *Phys. Rev. B*, vol. 39, no. 13, pp. 9536–9540, May 1989.

[133] *Sentaurus Device User Guide*, A-2008.09 ed., Sep 2008.

[134] D.B.M. Klaassen, "A Unified Mobility Model for Device Simulation – I. Model Equations and Concentration Dependence," *Solid-State Electr.*, vol. 35, no. 7, pp. 953–959, July 1992.

[135] C. Lombardi, S. Manzini, A. Saporito, and M. Vanzi, "A Physically Based Mobility Model for Numerical Simulation of Nonplanar Devices," *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 7, no. 11, pp. 1164–1171, November 1988.

[136] T. Hoang, P. LeMinh, J. Holleman, and J. Schmitz, "Strong efficiency improvement of SOI-LEDs through carrier confinement," *IEEE Electron Device Lett.*, vol. 28, no. 5, pp. 383–385, May 2007.

[137] C. Jeong, D. Antoniadis, and M. Lundstrom, "On Backscattering and Mobility in Nanoscale Silicon MOSFETs," *IEEE Trans. Electron Devices*, vol. 56, no. 11, pp. 2762–2769, 2009.

[138] R. Stratton, "Diffusion of Hot and Cold Electrons in Semiconductor Barriers," *Phys. Rev.*, vol. 126, no. 6, pp. 2002–2014, Jun 1962.

[139] K. Blotekjaer, "Transport equations for electrons in two-valley semiconductors," *IEEE Trans. Electron Devices*, vol. 17, no. 1, pp. 38–47, 1970.

[140] T. Grasser, T.-W. Tang, H. Kosina, and S. Selberherr, "A review of hydrodynamic and energy-transport models for semiconductor device simulation," *Proc. IEEE*, vol. 91, no. 2, pp. 251–274, 2003.

[141] C. Jungemann, T. Grasser, B. Neinhuus, and B. Meinerzhagen, "Failure of moments-based transport models in nanoscale devices near equilibrium," *IEEE Trans. Electron Devices*, vol. 52, no. 11, pp. 2404–2408, 2005.

[142] S. Laux and M. Fischetti, “Monte-Carlo simulation of submicrometer Si n-MOSFETs at 77 and 300 K,” *Electron Device Letters, IEEE*, vol. 9, no. 9, pp. 467–469, 9 1988.

[143] F. Venturi, R. K. Smith, E. C. Sangiorgi, M. R. Pinto, and B. Ricco, “A general purpose device simulator coupling Poisson and Monte Carlo transport with applications to deep submicron MOSFETs,” *IEEE Trans. Comput.-Aided Design Integr. Circuits Syst.*, vol. 8, no. 4, pp. 360–369, 1989.

[144] V. Sverdlov, A. Gehring, H. Kosina, and S. Selberherr, “Quantum transport in ultra-scaled double-gate MOSFETs: A Wigner function-based Monte Carlo approach,” *Solid-State Electr.*, vol. 49, no. 9, pp. 1510 – 1515, 2005, papers Selected from the EUROSOI Workshop Granada, 19-21 January 2005.

[145] M. Lundstrom, *Fundamentals of carrier transport*, 2nd ed. Cambridge University Press, 2000.

[146] D. Esseni, “On the modeling of surface roughness limited mobility in SOI MOSFETs and its correlation to the transistor effective field,” *IEEE Trans. Electron Devices*, vol. 51, no. 3, pp. 394–401, 2004.

[147] M. V. Fischetti and S. E. Laux, “Long-range Coulomb interactions in small Si devices. Part I: Performance and reliability,” *Journal of Applied Physics*, vol. 89, no. 2, pp. 1205–1231, 2001.

[148] 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-κ insulator: The role of remote phonon scattering,” *Journal of Applied Physics*, vol. 90, no. 9, pp. 4587–4608, 2001.

[149] P. Toniutti, P. Palestri, D. Esseni, and L. Selmi, “Revised analysis of the mobility and I_{ON} degradation in high-κ gate stacks: Surface optical phonons vs. remote Coulomb scattering,” in *Proc. ESSDERC*, 2008, pp. 246–249.

[150] K. Nakanishi, T. Uechi, and N. Sano, “Self-consistent Monte Carlo device simulations under nano-scale device structures: role of Coulomb interaction, degeneracy, and boundary condition,” in *Proc. IEDM*, 2009, pp. 1–4.

[151] 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 Trans. Electron Devices*, vol. 54, no. 5, pp. 1156–1164, 2007.

[152] J. P. McKelvey, R. L. Longini, and T. P. Brody, “Alternative Approach to the Solution of Added Carrier Transport Problems in Semiconductors,” *Phys. Rev.*, vol. 123, no. 1, pp. 51–57, Jul 1961.

[153] A. Das and M. Lundstrom, “A scattering matrix approach to device simulation,” *Solid-State Electr.*, vol. 33, no. 10, pp. 1299 – 1307, 1990.

[154] S. Datta, M. Cahay, and M. McLennan, “Scatter-matrix approach to quantum transport,” *Phys. Rev. B*, vol. 36, no. 10, pp. 5655–5658, 10 1987.

[155] S. Tanaka and M. Lundstrom, “A compact HBT device model based on a one-flux treatment of carrier transport,” *Solid-State Electr.*, vol. 37, no. 3, pp. 401 – 410, 1994.

[156] M. Lundstrom, “Elementary scattering theory of the Si MOSFET,” *IEEE Electron Device Lett.*, vol. 18, no. 7, pp. 361–363, 7 1997.

[157] M. Lundstrom and Z. Ren, “Essential Physics of Carrier Transport in Nanoscale MOSFETs,” *IEEE Trans. Electron Devices*, vol. 49, no. 1, pp. 133–141, January 2002.

[158] K. Natori, “Ballistic metal-oxide-semiconductor field effect transistor,” *Journal of Applied Physics*, vol. 76, no. 8, pp. 4879–4890, October 1994.

[159] A. Rahman, J. Guo, S. Datta, and M. S. Lundstrom, “Theory of ballistic nanotransistors,” *IEEE Trans. Electron Devices*, vol. 50, no. 9, pp. 1853–1864, 2003.

[160] A. Svizhenko and M. Anantram, “Role of scattering in nanotransistors,” *IEEE Trans. Electron Devices*, vol. 50, no. 6, pp. 1459 – 1466, 6 2003.

[161] J. Saint-Martin, A. Bournel, and P. Dollfus, “On the ballistic transport in nanometer-scaled DG MOSFETs,” *IEEE Trans. Electron Devices*, vol. 51, no. 7, pp. 1148 – 1155, 2004.

[162] 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 Trans. Electron Devices*, vol. 52, no. 12, pp. 2727–2735, 2005.

[163] A. Rahman and M. S. Lundstrom, “A compact scattering model for the nanoscale double-gate MOSFET,” *IEEE Trans. Electron Devices*, vol. 49, no. 3, pp. 481–489, 2002.

[164] F. Berz, “The Bethe condition for thermionic emission near an absorbing boundary,” *Solid-State Electr.*, vol. 28, no. 10, pp. 1007 – 1013, 1985.

[165] 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 Trans. Electron Devices*, vol. 53, no. 7, pp. 1634–1640, 2006.

[166] M. Fischetti, S. Jin, T. Tang, P. Asbeck, Y. Taur, S. Laux, M. Rodwell, and N. Sano, “Scaling MOSFETs to 10 nm: Coulomb effects, source starvation, and virtual source model,” *J. Comput. Electr.*, vol. 8, no. 2, pp. 60–77, 2009.

[167] K. Natori, "Ballistic MOSFET reproduces current-voltage characteristics of an experimental device," *IEEE Electron Device Lett.*, vol. 23, no. 11, pp. 655–657, 2002.

[168] ——, "New solution to high-field transport in semiconductors: I. elastic scattering without energy relaxation," *Jap. Jrnl. of Applied Physics*, vol. 48, no. 3, p. 034503, 2009.

[169] ——, "New solution to high-field transport in semiconductors: II. velocity saturation and ballistic transmission," *Jap. Jrnl. of Applied Physics*, vol. 48, no. 3, p. 034504, 2009.

[170] E. Fuchs, P. Dollfus, G. Le Carval, S. Barraud, D. Villanueva, F. Salvetti, H. Jaouen, and T. Skotnicki, "A new backscattering model giving a description of the quasi-ballistic transport in nano-MOSFET," *IEEE Trans. Electron Devices*, vol. 52, no. 10, pp. 2280–2289, 2005.

[171] 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 *Proc. IEDM*, 2006, pp. 1–4.

[172] P. Palestri, C. Alexander, A. Asenov, V. Aubry-Fortuna, G. Baccarani *et al.*, "A comparison of advanced transport models for the computation of the drain current in nanoscale nMOSFETs," *Solid-State Electr.*, vol. 53, pp. 1293–1302, 2009.

[173] S. Takagi, A. Toriumi, M. Iwase, and H. Tango, "On the universality of inversion layer mobility in Si MOSFET's," *IEEE Trans. Electron Devices*, vol. 41, no. 12, pp. 2357–2368, 1994.

[174] H. Tsuchiya and S. Takagi, "Influence of Elastic and Inelastic Phonon Scattering on the Drive Current of Quasi-Ballistic MOSFETs," *IEEE Trans. Electron Devices*, vol. 55, no. 9, pp. 2397–2402, 2008.

[175] 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," *IEEE Trans. Electron Devices*, vol. 56, no. 3, pp. 408–419, 2009.

[176] J.-L.P.J. van der Steen, P. Palestri, D. Esseni, and R.J.E. Hueting, "A New Model for the Backscatter Coefficient in Nanoscale MOSFETs," in *Proc. ESSDERC*, 2010, pp. 234–237.

[177] F. Assad, Z. Ren, D. Vasileska, S. Datta, and M. Lundstrom, "On the performance limits for Si MOSFETs: a theoretical study," *IEEE Trans. Electron Devices*, vol. 47, no. 1, pp. 232 –240, 2000.

[178] J. Blakemore, "Approximations for Fermi-Dirac integrals, especially the function, $\mathcal{F}_{1/2}$, used to describe electron density in a semiconductor," *Solid-State Electr.*, vol. 25, no. 11, pp. 1067– – 1076, 1982.

[179] R. Clerc, P. Palestri, L. Selmi, and G. Ghibaudo, "Back-scattering in quasi ballistic NanoMOSFETs: The role of non thermal carrier distributions," in *Proc. ULIS*, 2008, pp. 125–128.

[180] V. Trivedi, J. Fossum, and W. Zhang, "Threshold voltage and bulk inversion effects in nonclassical CMOS devices with undoped ultra-thin bodies," *Solid-State Electr.*, vol. 51, no. 1, pp. 170–178, January 2007.

A

The potential in subthreshold

This appendix presents an approximate expression for the potential $\psi(z)$ in the subthreshold regime, where the carrier density is low and the carriers are essentially uniformly distributed throughout the UTB.

The electric field $d\psi/dz$ across the weakly inverted UTB is given by [76]

$$\begin{aligned}\frac{\partial\psi}{\partial z} &= \sqrt{\frac{2qn_{\mathrm{i}}u_{\mathrm{th}}}{\varepsilon_{\mathrm{Si}}}}\sqrt{\exp\left(\frac{\psi}{u_{\mathrm{th}}}\right)-\exp\left(\frac{\psi_0}{u_{\mathrm{th}}}\right)} \\ &= \sqrt{\frac{2qn_{\mathrm{i}}u_{\mathrm{th}}}{\varepsilon_{\mathrm{Si}}}}\exp\left(\frac{\psi_0}{2u_{\mathrm{th}}}\right)\sqrt{\exp\left(\frac{\psi-\psi_0}{u_{\mathrm{th}}}\right)-1}\end{aligned} \tag{A.1}$$

where $u_{\mathrm{th}} = k_{\mathrm{B}}T/q$, the thermal voltage. We have seen in Fig. 2.3(a) that in subthreshold the potential drop across the weak inversion charge is very small, thus $\psi_{\mathrm{s}} \approx \psi_0$. We employ Taylor's expansion, $\exp(\frac{\psi-\psi_0}{u_{\mathrm{th}}}) \approx \frac{\psi-\psi_0}{u_{\mathrm{th}}} + 1)$, to approximate (A.1). Then we arrive at

$$\begin{aligned}\frac{\partial\psi}{\partial z} &\approx \sqrt{\frac{2qn_{\mathrm{i}}u_{\mathrm{th}}}{\varepsilon_{\mathrm{Si}}}}\exp\left(\frac{\psi_0}{2u_{\mathrm{th}}}\right)\sqrt{\frac{\psi-\psi_0}{u_{\mathrm{th}}}} \\ \frac{1}{\sqrt{\psi-\psi_0}}\frac{\partial\psi}{\partial z} &= \sqrt{\frac{2qn_{\mathrm{i}}}{\varepsilon_{\mathrm{Si}}}}\exp\left(\frac{\psi_0}{2u_{\mathrm{th}}}\right) \\ \sqrt{\psi-\psi_0} &= \sqrt{\frac{qn_{\mathrm{i}}}{2\varepsilon_{\mathrm{Si}}}}\exp\left(\frac{\psi_0}{2u_{\mathrm{th}}}\right)z \\ \psi(z) &= \frac{qn_{\mathrm{i}}}{2\varepsilon_{\mathrm{Si}}}\exp\left(\frac{\psi_0}{u_{\mathrm{th}}}\right)z^2+\psi_0\end{aligned} \tag{A.2}$$

B

The variational approach

The trial wave function to be used in the variational approach is given by

$$\Psi_n = a_n \sqrt{\frac{2}{t_{\mathrm{Si}}}} \sin\left[\frac{n\pi}{t_{\mathrm{Si}}}\left(z + \frac{t_{\mathrm{Si}}}{2}\right)\right] \\ \times \left\{\exp\left[-b_n\left(\frac{1}{2} + \frac{z}{t_{\mathrm{Si}}}\right)\right] + \exp\left[-b_n\left(\frac{1}{2} - \frac{z}{t_{\mathrm{Si}}}\right)\right]\right\} \quad \text{(B.1)}$$

with a_n the normalization constant and b_n the so-called variational parameter for the nth order wave function. In fact, b_n depends on the normal electric field and thus represents the strength of electrical confinement. Note that when $b \to 0$, the wave function reduces to the eigenfunction of a particle in an ideal square well, i.e.

$$\Psi_n(z) = \sqrt{\frac{2}{t_{\mathrm{Si}}}} \sin\left[\frac{n\pi}{t_{\mathrm{Si}}}\left(z + \frac{t_{\mathrm{Si}}}{2}\right)\right]$$

The parameter b_n can be solved using the variational approach [86, 92, 93], which involves minimizing the total energy of the system, consisting of a field-related and a geometry related part.

The lowest order wave function from (B.1) is shown in Fig. B.1 to illustrate the impact of varying b_1 on the shape of the wave function. The figure clearly demonstrates that the wave function gradually changes from sine shaped to two distinct peaks at the interfaces dominated by the exponential terms in (B.1), for b moving away from 0. Interestingly, a value of b can be identified for which the wave function in the center of the channel ($z = 0$) changes from a maximum into a minimum. This point corresponds to a transition from volume inversion to surface inversion, similar to the classical approach, and was shown to occur for $b_1 = \pi$ [92]. The fact that eventually only one parameter determines the shape of the wave function, hence the charge distribution, may be of interest for compact modelling purposes.

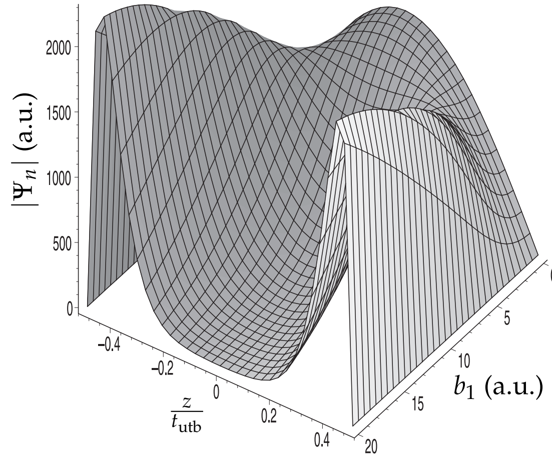


Figure B.1: Lowest order trial wave function along z vs. the variational parameter b_1.

In short, the value of b_n and the corresponding subband minima (or eigen energies) E_n for a given potential profile $V(z)$ can be found by minimizing the total energy of the system, i.e. the expectation value of the kinetic and potential energy

$$E_\mathrm{n} = \langle E_\mathrm{kin} \rangle + \langle E_\mathrm{pot} \rangle = \int_{-\frac{t_\mathrm{Si}}{2}}^{\frac{t_\mathrm{Si}}{2}} \Psi_n^*(z) \left[-\frac{\hbar^2}{2m^*} \frac{\partial^2}{\partial z^2} + qV(z) \right] \Psi_n(z)\, dz \qquad \text{(B.2)}$$

Then, evaluation of $\partial E_\mathrm{n} / \partial b_n = 0$ gives the energy minimum of subband n. Note that the above procedure has to be carried out for each subband, which results in complicated high order algebraic equations. Hence, for the sake of simplicity, usually only the lowest subband is considered (the 'quantum limit'). Furthermore, apart from a reasonable estimate of the trial wave function, also an initial assumption of the confining potential profile is required.

C

UTB-SOI MOSFETs in subthreshold

The aim of this appendix is to illustrate that a long-channel UTB-SOI MOSFET with very asymmetric front and back gate oxide thickness (resp. t_{oxf} and t_{oxb}) can be considered as a fully symmetric device under subthreshold conditions. In the following, the subscript 'f' and 'b' refer to the front and back interface respectively. Furthermore, we assume that the front and back gate are biased equally.

From Fig. 3.1(b) we derive

$$\frac{\varepsilon_{\mathrm{ox}}}{t_{\mathrm{oxf}}}\left(V_{\mathrm{GS}}-\Delta\phi_{\mathrm{f}}-\psi_{\mathrm{sf}}\right)=\varepsilon_{\mathrm{Si}}\mathcal{E}_{\mathrm{f}} \tag{C.1a}$$

$$\frac{\varepsilon_{\mathrm{ox}}}{t_{\mathrm{oxb}}}\left(V_{\mathrm{GS}}-\Delta\phi_{\mathrm{b}}-\psi_{\mathrm{sb}}\right)=\varepsilon_{\mathrm{Si}}\mathcal{E}_{\mathrm{b}} \tag{C.1b}$$

with $\mathcal{E}_{\mathrm{f}}$ and ψ_{sf} the electric field and the potential at the gate-channel interfaces, respectively; $\varepsilon_{\mathrm{ox}}$ and $\varepsilon_{\mathrm{Si}}$ resp. are the dielectric constants of the gate dielectric and silicon, and $\Delta\phi_{\mathrm{f}}$ is the difference in workfunction between the gate and channel.

Furthermore, in subthreshold we have $\psi_{\mathrm{sb}}-\psi_{\mathrm{sf}}=-\mathcal{E}_0 t_{\mathrm{Si}}$ with $\mathcal{E}_{\mathrm{b}}=\mathcal{E}_{\mathrm{f}}=\mathcal{E}_0$; Combining (C.1a) and (C.1b) gives [125]

$$\mathcal{E}_0=\frac{\Delta\phi_{\mathrm{b}}-\Delta\phi_{\mathrm{f}}}{\frac{\varepsilon_{\mathrm{Si}}}{\varepsilon_{\mathrm{ox}}}\left(t_{\mathrm{oxf}}+t_{\mathrm{oxb}}\right)+t_{\mathrm{Si}}} \tag{C.2}$$

After substitution in (C.1) the potential at either surface of the body is obtained:

$$\psi_{\mathrm{sf}}=V_{\mathrm{GS}}-\left[\frac{\Delta\phi_{\mathrm{f}}\left(\varepsilon_{\mathrm{ox}}t_{\mathrm{Si}}+\varepsilon_{\mathrm{Si}}t_{\mathrm{oxb}}\right)+\Delta\phi_{\mathrm{b}}\varepsilon_{\mathrm{Si}}t_{\mathrm{oxf}}}{\varepsilon_{\mathrm{ox}}t_{\mathrm{Si}}+\varepsilon_{\mathrm{Si}}\left(t_{\mathrm{oxf}}+t_{\mathrm{oxb}}\right)}\right] \tag{C.3a}$$

$$\psi_{\mathrm{sb}}=V_{\mathrm{GS}}-\left[\frac{\Delta\phi_{\mathrm{b}}\left(\varepsilon_{\mathrm{ox}}t_{\mathrm{Si}}+\varepsilon_{\mathrm{Si}}t_{\mathrm{oxf}}\right)+\Delta\phi_{\mathrm{f}}\varepsilon_{\mathrm{Si}}t_{\mathrm{oxb}}}{\varepsilon_{\mathrm{ox}}t_{\mathrm{Si}}+\varepsilon_{\mathrm{Si}}\left(t_{\mathrm{oxf}}+t_{\mathrm{oxb}}\right)}\right] \tag{C.3b}$$

Assuming $t_{\mathrm{oxf}}=t_{\mathrm{oxb}}$ and $\Delta\phi_{\mathrm{b}}=\Delta\phi_{\mathrm{f}}$, hence a fully symmetric device, (C.3a) and (C.3b) reduce to $\psi_{\mathrm{sf}}=\psi_{\mathrm{sb}}=V_{\mathrm{GS}}-\Delta\phi_{\mathrm{f}}$, *independent* of t_{ox}. Interestingly,

the same result is obtained when $t_{\text{oxb}} \gg t_{\text{oxf}}$, which e.g. corresponds to an UTB SOI device on a thick BOX layer with the underlying substrate as backplane.

D

Mobility extraction in subthreshold

The following shows how the low-field mobility can be extracted from the subthreshold current and capacitance. With $C_{\mathrm{GG}} = dQ_{\mathrm{i}}/dV_{\mathrm{GS}}$ and (3.2) we obtain $Q_{\mathrm{i}} = u_{\mathrm{th}} C_{\mathrm{GG}}$, with the thermal voltage $u_{\mathrm{th}} = k_{\mathrm{B}}T/q$. After substitution in (3.1), the current and capacitance are directly linked through the mobility, as given by

$$I_{\mathrm{DS}} = \frac{\mu u_{\mathrm{th}}}{L} \cdot u_{\mathrm{th}} C_{\mathrm{GG}} \left[1 - \exp\left(\frac{-V_{\mathrm{DS}}}{u_{\mathrm{th}}}\right)\right] \tag{D.1}$$

Hence, the low-field mobility can be extracted as follows

$$\mu = \frac{I_{\mathrm{DS}}}{C_{\mathrm{GG}}} \cdot \frac{L}{u_{\mathrm{th}}^2} \frac{1}{1 - \exp\left(\frac{-V_{\mathrm{DS}}}{u_{\mathrm{th}}}\right)} \tag{D.2}$$

(D.2) can be shown to hold also for any SOI DG device, provided that the front and back gate are biased equally. Generally, Q_{i} in an UTB DG device in subthreshold is given by [180]

$$Q_{\mathrm{i}} = -q t_{\mathrm{Si}} n_{\mathrm{i}} u_{\mathrm{th}} \left[\frac{\exp\left(\frac{\psi_{\mathrm{sb}}}{u_{\mathrm{th}}}\right) - \exp\left(\frac{\psi_{\mathrm{sf}}}{u_{\mathrm{th}}}\right)}{\psi_{\mathrm{sb}} - \psi_{\mathrm{sf}}}\right] \tag{D.3}$$

with ψ_{sf} and ψ_{sb} as given in (C.3a) and (C.3b). Furthermore, noting that both gates are equally biased, and assuming that volume inversion occurs, we use

$$\frac{\partial \psi_{\mathrm{sf}}}{\partial V_{\mathrm{GS}}} = \frac{\partial \psi_{\mathrm{sb}}}{\partial V_{\mathrm{GS}}} = 1 \tag{D.4}$$

and hence

$$\frac{\partial}{\partial V_{\mathrm{GS}}} \left\{\exp\left(\frac{\psi_{\mathrm{sf}}}{u_{\mathrm{th}}}\right)\right\} = \frac{1}{u_{\mathrm{th}}} \exp\left(\frac{\psi_{\mathrm{sf}}}{u_{\mathrm{th}}}\right) \tag{D.5}$$

Then, with (D.3), the subthreshold gate capacitance is obtained

$$
\begin{aligned}
C_{\mathrm{GG}} = \frac{\partial Q_{\mathrm{i}}}{\partial V_{\mathrm{GS}}} &= -q t_{\mathrm{Si}} n_{\mathrm{i}} u_{\mathrm{th}} \frac{\partial}{\partial V_{\mathrm{GS}}} \left\{ \frac{\exp\left(\frac{\psi_{\mathrm{sb}}}{u_{\mathrm{th}}}\right) - \exp\left(\frac{\psi_{\mathrm{sf}}}{u_{\mathrm{th}}}\right)}{\psi_{\mathrm{sb}} - \psi_{\mathrm{sf}}} \right\} \\
&= -q t_{\mathrm{Si}} n_{\mathrm{i}} \left\{ \frac{\exp\left(\frac{\psi_{\mathrm{sb}}}{u_{\mathrm{th}}}\right) - \exp\left(\frac{\psi_{\mathrm{sf}}}{u_{\mathrm{th}}}\right)}{\psi_{\mathrm{sb}} - \psi_{\mathrm{sf}}} \right\} \\
&= \frac{Q_{\mathrm{i}}}{u_{\mathrm{th}}}
\end{aligned}
\tag{D.6}
$$

Hence, the low-field mobility can be determined directly from the measured subthreshold gate capacitance and drain current; this holds for fully symmetric DG devices as well as asymmetric DG SOI devices, irrespective of the oxide thickness and workfunction of the gates, provided that the gates are biased equally.

Furthermore, note that for a symmetric DG MOSFET, for which $\psi_{\mathrm{sf}} = \psi_{\mathrm{sb}} = \psi_{\mathrm{s}}$, (D.3) yields, after applying Taylor's expansion

$$
Q_{\mathrm{i}} = -q t_{\mathrm{Si}} n_{\mathrm{i}} \exp\left(\frac{\psi_{\mathrm{s}}}{u_{\mathrm{th}}}\right)
$$

which is equivalent to (3.2).

In conclusion, with (D.2) it is possible to extract the low-field mobility, without any further approximations, in contrast to the commonly used expression $\mu_{\mathrm{eff}} = L/(Q_{\mathrm{i}} V_{\mathrm{DS}}/I_{\mathrm{DS}})$; the latter is derived from only the drift component of the drain current, and thus neglects the diffusive part which is dominant in the subthreshold, hence low-field, regime.

E

The ballistic distribution function

In the Single–Scattering model (p. 105) the flux of carriers moving from VS to the drain is assumed to be essentially ballistic. Thus, the carrier distribution function $n^+(x, k_\mathrm{x})$ should satisfy the Boltzmann Transport Equation (BTE) with vanishing collision term, which for a single subband reads

$$\frac{\hbar k_\mathrm{x}}{m^*}\frac{\partial n^+}{\partial x} + \frac{q}{\hbar}\frac{dV}{dx}\frac{\partial n^+}{\partial k_\mathrm{x}} = 0 \tag{E.1}$$

Moving to polar coordinates, i.e. $k_\mathrm{x} = \kappa\cos(\phi)$ and $k_\mathrm{y} = \kappa\sin(\phi)$, and noting that integration over κ and ϕ gives the carrier concentration

$$\int_0^\infty \int_{-\pi/2}^{+\pi/2} n^+(x,\kappa,\phi)\, d\phi\, d\kappa = n^+(x)$$

we obtain

$$n^+(x,\kappa,\phi) = \frac{\hbar^2 n_0^+}{\pi m^* k_\mathrm{B} T}\exp\left(-\left[\frac{\hbar^2\kappa^2}{2m^* k_\mathrm{B} T} - \frac{qV(x)}{k_\mathrm{B} T}\right]\right) \times \theta\left(\frac{\hbar\kappa\cos(\phi)}{m^*} - \sqrt{\frac{2qV(x)}{m^*}}\right) \tag{E.2}$$

with $\theta(\cdot)$ the step function in order to account for the positive k_x states only; furthermore, we set $V(0) = 0$. For the sake of simplicity (E.2) uses Boltzmann's approximation, but a similar derivation can be shown for Fermi-Dirac statistics. After integration over ϕ to eliminate the step function, we write $n^+(x, k_\mathrm{x})$ in terms of *total* energy ε

$$\varepsilon = \frac{\hbar^2\kappa^2}{2m^*} - qV(x)$$

Thus, we obtain the ballistic carrier distribution function

$$n^+(x,\varepsilon) = \frac{n_0^+}{k_\mathrm{B} T}\exp\left(-\frac{\varepsilon}{k_\mathrm{B} T}\right)\frac{2}{\pi}\arccos\sqrt{\frac{qV(x)}{\varepsilon + qV(x)}} \tag{E.3}$$

F

The velocity of the ballistic flux

In this appendix we will derive the longitudinal velocity $v_x^+(x,\varepsilon)$ of a ballistically moving flux of carriers. Fig. F.1 shows a sketch of the distribution in **k**-space, with (1) the $(k_\mathrm{x}, k_\mathrm{y})$ belonging to the initial distribution, and (2) the combinations of $(k'_\mathrm{x}, k'_\mathrm{y})$ after gaining energy qV from the longitudinal electric field. We note that in case (2), the lowest value of k_x equals $k_\mathrm{x,min} = \sqrt{2m^*V}/\hbar$; the upper bound of k_y remains unchanged and equals $k_\mathrm{y,max} = \sqrt{2m^*\varepsilon}/\hbar$, just as in (1). Similar to the distribution at $V = 0$ ($x = 0$), the $(k'_\mathrm{x}, k'_\mathrm{y})$-values after acceleration by the longitudinal electric field are located on a circle, as given

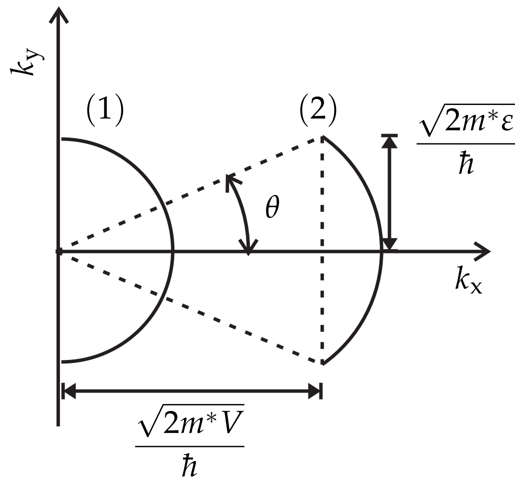


Figure F.1: Schematic representation of a ballistic equi-occupation curve, at $V = 0$ and $x = 0$, labeled (1), and after experiencing a potential drop $V > 0$ shown by the arc labeled (2).

by

$$\frac{\hbar^2}{2m^*}\left(k_\mathrm{x}^2 + k_\mathrm{y}^2\right) = qV + \varepsilon \tag{F.1}$$

Thus, θ is given by

$$\theta = \arccos \frac{\sqrt{\frac{2m^* V}{\hbar^2}}}{\sqrt{\frac{2m^*(V+\varepsilon)}{\hbar^2}}} = \arccos \sqrt{\frac{V}{V+\varepsilon}} = \arcsin \sqrt{\frac{\varepsilon}{V+\varepsilon}} \tag{F.2}$$

and we find for the longitudinal ballistic velocity

$$\nu_x^+(\varepsilon) = \sqrt{\frac{2\,(\varepsilon + V)}{m^*}} \frac{\int\limits_{-\theta}^{\theta} \cos\theta'\, d\theta'}{2\theta} \tag{F.3}$$

$$= \sqrt{\frac{2\,(\varepsilon + V)}{m^*}} \frac{\sin\theta}{\theta} \tag{F.4}$$

By inserting θ from (F.2), the resulting expression for $\nu_x^+(x,\varepsilon)$ reads

$$\nu_x^+(x,\varepsilon) = \sqrt{\frac{2\,(\varepsilon + V)}{m^*}} \frac{\sqrt{\frac{\varepsilon}{\varepsilon+V}}}{\arccos\sqrt{\frac{V}{\varepsilon+V}}} \tag{F.5}$$

$$= \sqrt{\frac{2\varepsilon}{m^*}} \frac{1}{\arccos\sqrt{\frac{V}{\varepsilon+V}}} \tag{F.6}$$

G

The critical length in the SSC model

One of the basic assumptions in the SSC model is that only few back-scattering events occur in the channel. This means that the positive flux remains essentially unaffected by back-scattering, and is therefore assumed to be constant along the channel. However, when back-scattering increases, the positive flux decreases due to the carriers that change direction from positive to negative. In addition to the procedure outlined on p. 109, this appendix outlines an empirical way to estimate the impact of the reduction of the positive flux.

We first rewrite the non-degenerate SSC model in terms of a critical length L_{SSC} and the single-scattering mean-free-path $\lambda_{\mathrm{SSC}} = 2v_{\mathrm{th}}\tau$ (see also p. 100)

$$r_{\mathrm{SSC}} = \frac{1}{\lambda_{\mathrm{SSC}}} \int_0^L \int_0^\infty \frac{1}{k_{\mathrm{B}}T} \exp\left(-\frac{\varepsilon}{k_{\mathrm{B}}T}\right) \left(\frac{2}{\pi}\right)^2 \arccos^2 \sqrt{\frac{qV(x)}{\varepsilon + qV(x)}}\, d\varepsilon\, dx \equiv \frac{L_{\mathrm{SSC}}}{\lambda_{\mathrm{SSC}}} \tag{G.1}$$

L_{SSC} approaches L for $\mathcal{E} \to 0$ and gradually decreases for increasing $\mathcal{E}$.

Rather than presenting a mathematically sound justification, we simply introduce the empirical correction as

$$r'_{\mathrm{SSC}} \equiv \frac{r_{\mathrm{SSC}}}{r_{\mathrm{SSC}} + 1} = \frac{L_{\mathrm{SSC}}}{L_{\mathrm{SSC}} + \lambda_{\mathrm{SSC}}} \tag{G.2}$$

where r'_{SSC} is the 'effective' backscatter coefficient; r_{SSC} is the original SSC backscatter coefficient as derived in Section 4.5 on page 108, and written in terms of L_{SSC} and λ_{SSC} in (G.1). Hence, the functional form of r'_{SSC} is assumed to be similar to the Flux model [cf. (4.8), (4.12)].

Fig. G.1 shows r as a function of $\mathcal{E}$ according to the SSC model (G.1), along with two ways of accounting for the decreasing positive flux along the channel: the REDp model [(4.23) with (4.28)] and the empirical correction in (G.2). We find that the results of the empirical correction to the SSC model essentially

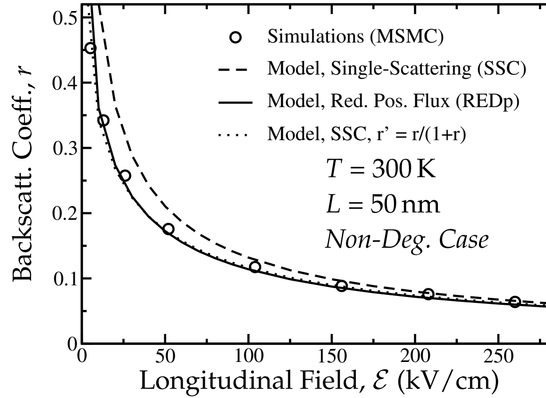


Figure G.1: r vs. $\mathcal{E}$, showing the MSMC values, the REDp model, and the SSC model with the empirical reduction of the positive flux. We assume elastic scattering and a linear potential profile.

coincide with the REDp and the MSMC values. From a practical viewpoint, this means that we can estimate r in presence of significant back-scattering, which actually would require accounting for the decreasing positive flux, while we can still use the relatively simple SSC model.

In Fig. G.2 we plot, similar to Fig. 4.17, the critical length in the SSC model L_{SSC} as a function of L, along with L_{extr} extracted from the MSMC simulations using $L_{extr} = \lambda_{SSC} \cdot r/(1-r)$. For comparison, we also show L_{eff}, the critical length in the Flux model (4.8). Fig. G.2 demonstrates that L_{SSC} is agreement with MSMC over the entire range of L, in both the low-field and the high-field case. The behavior of L_{eff}, instead, is markedly different: compared to the simulations and L_{SSC}, L_{eff} exhibits a stronger L dependence in the low-field regime. For $\mathcal{E}$ = 52 kV/cm and the shortest channels, we observe that the slope is very similar to the MSMC results, whereas L_{eff} saturates for the longer channels.

In summary, in this Appendix we have shown that the empirical correction to the SSC model captures the quantitative features of the REDp model and the MSMC results. Alternatively, if we extract the effective critical length from the MSMC simulations, assuming that r can be written as $r = L_{extr}/(L_{extr} + \lambda_{SSC})$, L_{extr} is found to be well reproduced by the critical length as it is defined in the SSC model (L_{SSC}).

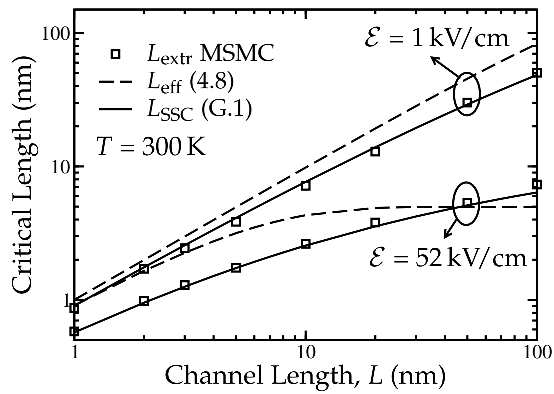


Figure G.2: The critical length in the SSC model, $L_{\rm SSC}$ in (G.1), and in the Flux model, $L_{\rm eff}$ in (4.8). The critical length $L_{\rm extr}$ extracted from the MSMC simulations is shown as well, assuming r can be written as $r = L_{\rm extr}/(L_{\rm extr} + \lambda_{\rm SSC})$.

Summary

Multiple-gate devices such as FinFETs, Double Gate and Gate-All-Around structures are widely recognized as promising candidates for future generation CMOS technology nodes. These devices offer an enhanced electrostatic integrity through the combination of multiple gates and an ultra-thin semiconductor body, thus allowing for a further reduction of the channel length. In this thesis, we have discussed the effects of reducing the channel length and the body thickness on carrier transport in ultrathin-body MOSFETs.

In Chapter 2 we discussed the general features of the carrier distribution and the potential profile in long-channel Double-Gate ultrathin-body MOSFETs, including the effect of quantum confinement in the ultrathin body, where thickness of the semiconductor body sets the strength of the confining potential. Calculation of the subband minima and wave functions within the Effective Mass Approximation shows that the carriers are repelled from the semiconductor-oxide interface and that, in case of purely structural confinement in the subthreshold regime, the carrier density peaks at the center of the ultrathin body.

Furthermore, we have investigated the validity of the Effective Mass Approximation by studying the minima of the two-dimensional subbands, the transport masses and the electron density of states. The Effective Mass Approximation results have been compared with a full-band quantization approach based on the Linear Combination of Bulk Bands, for both silicon and germanium *n*-MOSFETs for various crystal orientations. In most cases the validity of the Effective Mass Approximation is quite satisfactory; the validity of the Effective Mass Approximation can be extended towards very small semiconductor thicknesses by accounting for the thickness-dependent effective transport masses, calculated from the full-band approach.

In Chapter 3 we proposed a method to quantify shifts in energy band alignment due to e.g. structural quantum confinement in ultrathin-body MOSFETs, using the temperature dependence of the subthreshold current. The re-

sults were compared with the shifts in threshold voltage, which is commonly used to quantify the effect of quantum confinement. We have shown that using the temperature dependence of the subthreshold current, shifts in the valence and conduction band edge can be extracted distinctively from changes in mobility and effective density of states, making the proposed method more accurate in assessing the impact of structural quantum confinement than the threshold voltage method.

Chapter 4 focussed on the effect of decreasing the channel length to values comparable to the carrier mean-free-path, so that carrier transport becomes quasi-ballistic. We have systematically analyzed carrier back-scattering using a sophisticated Multi-Subband Monte-Carlo simulator, both in a realistic 32 nm ultrathin-body MOSFET and in a template device structure with a fixed potential and template scattering parameters. Besides the backscatter coefficient, we extracted from the Multi-Subband Monte-Carlo simulations the distribution of backscatter events, its centroid and the critical length for backscattering according to the various definitions reported in literature.

Our study demonstrates that the critical length, i.e., the portion of the channel in which backscattering is most pronounced, is not directly determined by the optical phonon energy. Instead, the critical length is strongly related to the spatial extension of the $k_B T$-layer. Furthermore, we found that the appropriate low-field mean-free-path to be used in a flux-like equation equals $2v_{th}\tau$, whereas in the Flux model the low-field mean-free-path equals $\frac{k_B T}{q}\frac{2\mu}{v_{th}}$.

Finally, we propose a new model for the backscatter coefficient in nanoscale MOSFETs, assuming that carriers encounter on average just a single scattering event, so that carrier transport is nearly ballistic. The proposed model accounts for both elastic and inelastic scattering mechanisms and it naturally captures the effect of degeneracy.

Systematic comparison with results obtained from Multi-Subband Monte-Carlo simulations shows very good overall agreement, demonstrating that the proposed model provides a useful framework to help explain and predict quasi-ballistic transport phenomena in nanoscale MOSFETs.

Samenvatting

Meervoudige-*gate* transistoren zoals Double-Gate, Gate-All-Around en Fin-FET structuren worden gezien als veelbelovende kandidaten voor toekomstige generaties CMOS technologiën. Deze transistoren bieden een verhoogde elektrostatische integriteit dankzij de combinatie van meerdere gates en een ultradunne halfgeleiderlaag, zodat een verdere reductie van de kanaallengte mogelijk wordt. In dit proefschrift hebben we enkele effecten van het verkleinen van de kanaallengte en de reductie in de dikte van de halfgeleiderlaag op het transport van ladingsdragers in ultradunne-laag MOSFETs besproken.

In Hoofdstuk 2 hebben we de algemene eigenschappen van de ladingsverdeling en het potentiaalprofiel in lange-kanaals Double-Gate ultradunne-laag MOSFETs behandeld, alsmede het effect van kwantummechanische begrenzing in de ultradunne halfgeleiderlaag, waarin de sterkte van de kwantumbegrenzing bepaald wordt door de laagdikte. Berekening van de subbandminima en de golffuncties in de Effectieve Massa Benadering laat zien dat de ladingsdragers van het halfgeleider-oxide grensvlak worden weggedrukt en dat, in het geval van zuiver geometrische kwantumbegrenzing in het diffusieregime, de ladingsdichtheid een maximum heeft in het midden van de ultradunne halfgeleiderlaag.

Daarnaast hebben we onderzoek gedaan naar de geldigheid van de Effectieve Massa Benadering. Daartoe zijn de energieminima van de 2-dimensionale subbanden, de transport massa's en de effectieve electronen-toestandsdichtheid bestudeerd. De resultaten van de Effectieve Massa Benadering zijn vergeleken met een *full-band* kwantisatiemodel gebaseerd op de Lineaire Combinatie van Bulk Banden, voor zowel silicium als germanium n-MOSFETs, voor diverse kristaloriëntaties. In de meeste gevallen is het gebruik van de Effectieve Massa Benadering gerechtvaardigd; het geldigheidsgebied kan worden opgerekt naar hele dunne halfgeleiderlagen door rekening te houden met de laagdikte-afhankelijkheid van de effectieve transportmassa's, berekend met behulp van de *full-band* methode.

In Hoofdstuk 3 hebben we een methode voorgesteld om verschuivingen in de energiebanden te kwantificeren, die bijvoorbeeld het gevolg zijn van geometrische kwantumbegrenzing in ultradunne-laag MOSFETs. De voorgestelde methode maakt gebruik van de temperatuurafhankelijkheid van de diffusiestroom. De resultaten zijn vergeleken met de verandering van de drempelspanning, die over het algemeen gebruikt wordt om het effect van kwantummechanische begrenzing te meten. We hebben laten zien dat met de temperatuurafhankelijkheid van diffusiestroom verschuivingen in de valentie- en geleidingsband kunnen worden gemeten, onafhankelijk van de veranderingen in de beweeglijkheid en de effectieve toestandsdichtheid. Dit gegeven maakt de voorgestelde methode nauwkeuriger dan de veelgebruikte drempelspannings-methode.

Hoofdstuk 4 concentreert zich op het effect van verkorting van het kanaal tot een lengteschaal vergelijkbaar met de vrije weglengte van de ladingsdragers, waardoor het ladingstransport quasi-ballistisch wordt. Met behulp van een geavanceerde Multi-Subband Monte-Carlo simulator is een systematische analyse gedaan van de reflectie van ladingsdragers in het kanaal, zowel in een realistische 32 nm ultradunne-laag MOSFET als in een geïdealiseerde transistorstructuur met een vast potentiaalprofiel en versimpelde verstrooiingsmechanismes. Naast de reflectiecoëfficiënt, hebben we uit de simulator de ruimtelijke verdeling van de botsingen, het bijbehorende zwaartepunt en de kritische lengte voor de reflectie van ladingsdragers geëxtraheerd, volgens de verschillende definities in de literatuur. Onze studie toont aan dat de kritische lengte, dat wil zeggen, het gedeelte van het kanaal waarin reflectie het meest uitgesproken is, niet direct bepaald wordt door de optische fonon-energie. De kritische lengte blijkt daarentegen sterk gerelateerd te zijn aan de zogenoemde $k_\mathrm{B}T$-laag. Daarnaast blijkt in een flux-achtige vergelijking de juiste vrije weglengte voor ladingsdragers $2v_\mathrm{th}\tau$ te zijn, terwijl in het Flux-model aangenomen wordt dat de vrije weglengte gelijk is aan $\frac{k_\mathrm{B}T}{q}\frac{2\mu}{v_\mathrm{th}}$.

Dit proefschrift besluit met de afleiding van een nieuw model voor de reflectiecoefficiënt in nano-MOSFETs, gebaseerd op de aanname dat ladingsdragers gemiddeld een enkele keer botsen zodat het ladingstransport vrijwel ballistisch is. Het voorgestelde model houdt rekening met zowel elastische als inelastische botsingen en reproduceert van nature het effect van ontaarding. Een systematische vergelijking van het model met simulatieresultaten laat een heel goede overeenstemming zien, waarmee wordt aangetoond dat het voorgestelde model een nuttig raamwerk biedt om quasi-ballistisch transport in nano-MOSFETs te helpen verklaren en te voorspellen.

Riassunto

I dispositivi a *gate* multiplo, come per esempio i FinFETs, i *double gate* e le strutture a *gate-all-around*, sono ampiamente riconosciuti come candidati promettenti per le future generazioni di tecnologie CMOS. Questi dispositivi offrono un'ottima integrità elettrostatica attraverso la combinazione di una struttura a *gate* multiplo e di un semiconduttore ultrasottile, che consentono una riduzione molto aggressiva della lunghezza di canale. In questa tesi abbiamo discusso gli effetti prodotti dalla riduzione della lunghezza di canale e dello spessore di semiconduttore sul trasporto di carica nei MOSFETs.

Nel capitolo 2 abbiamo discusso le caratteristiche generali della distribuzione di carica e il profilo di potenziale nei *double gate* MOSFETs con canale lungo e substrato sottile, includendo l'effetto del confinamento nel substrato, dove lo spessore del semiconduttore stabilisce l'entità del confinamento quantistico. Calcolando i minimi delle sottobande di energia e le funzioni d'onda secondo l'approssimazione efficace di massa, notiamo che i portatori sono respinti dall'interfaccia fra semiconduttore ed ossido e che, in caso di confinamento puramente strutturale e nel regime di sottosoglia, la densità dei portatori presenta un massimo al centro del strato di semiconduttore. Inoltre abbiamo investigato la validità dell'approssimazione di massa efficace studiando i minimi delle sottobande, le masse di trasporto e la densità degli stati elettronici. I risultati dell'approssimazione di massa efficace sono stati confrontati con un approccio di quantizzazione *full band* basato sulla combinazione lineare delle *bulk bands*, per gli *n*-MOSFETs con canale di silicio e di germanio e per diverse orientazioni cristallografiche. Nella maggior parte dei casi la validità dell'approssimazione di massa efficace può essere estesa fino a spessori del semiconduttore molto piccoli facendo uso di masse efficaci dipendenti dallo spessore di silicio e determinate per mezzo del modello di quantizzazione *full band*.

Nel capitolo 3 proponiamo un metodo per quantificare le variazioni nell'allineamento delle bande energetiche dovute, per esempio, al confinamento

indotto dallo spessore di silicio nei MOSFETs con canale molto sottile, e utilizzando a tal scopo la dipendenza dalla temperatura della corrente di sottosoglia. I risultati sono stati confrontati con le variazioni della tensione di soglia, che viene comunemente usata per quantificare l'effetto del confinamento quantistico. Abbiamo mostrato che, utilizzando la dipendenza dalla temperatura della corrente di sottosoglia, le variazioni nella bande di valenza e conduzione possono essere estratte separatamente dagli eventuali cambiamenti della mobilità e della densità degli stati, rendendo il metodo qui proposto più accurato rispetto alla semplice misura della tensione di soglia per valutare l'impatto della quantizzazione dovuta al confinamento strutturale.

Il capitolo 4 è dedicato agli effetti dovuti alla riduzione della lunghezza di canale fino a valori paragonabili al cammino libero medio dei portatori, che rende il trasporto dei portatori *quasi ballistico*. Abbiamo studiato sistematicamente il *backscattering* utilizzando un sofisticato simulatore Monte Carlo per analizzare un realistico MOSFET lungo 32 nm con spessore di silicio molto sottile ed inoltre una struttura di dispositivo schematizzata, con un potenziale fisso e dei modelli di *scattering* semplificati. In aggiunta al coefficiente di *backscattering* abbiamo estratto dalle simulazioni Monte Carlo la distribuzione spaziale degli eventi di *backscattering*, il suo baricentro e la lunghezza critica per il *backscattering*, secondo le diverse definizioni riportate in letteratura.

Il nostro studio dimostra che la lunghezza critica, cioè la porzione di canale in cui il *backscattering* è più evidente, non è direttamente determinata dall'energia dei fononi ottici. Al contrario, la lunghezza critica è fortemente legata all'estensione spaziale del cosiddetto $k_B T$*-Layer*. Inoltre abbiamo trovato che il cammino libero medio a basso campo da usare in un'equazione derivata dalla *teoria di flusso* è uguale a $2v_{th}\tau$, mentre nel modello *teoria di flusso* il cammino libero medio a basso campo risulta $\frac{k_B T}{q}\frac{2\mu}{v_{th}}$.

Concludendo, noi proponiamo un nuovo modello per il coefficiente di *backscattering* nei MOSFETs nanometrici, il quale assume che la maggior parte dei portatori subiscano un unico evento di *scattering*, in modo che il trasporto risulti quasi balistico. Il modello proposto tiene conto sia dei meccanismi di *scattering* elastici che inelastici e descrive in modo naturale l'effetto di degenerazione dei portatori. Un confronto sistematico con i risultati ottenuti dalle simulazioni Monte-Carlo mostra un buon accordo, dimostrando che il modello qui proposto stabilisce un utile contesto per spiegare e predire i fenomeni di trasporto quasi ballistico nei MOSFETs nanometrici.

List of Publications

J.-L.P.J. van der Steen, P. Palestri, D. Esseni, and R.J.E. Hueting, "A New Model for the Backscatter Coefficient in Nanoscale MOSFETs," in *Proc. ESSDERC*, 2010, pp. 234–237.

A. Scholten, G. Smit, R. Pijper, L. Tiemeijer, H. Tuinhout, **J.-L. van der Steen**, A. Mercha, M. Braccioli, and D. Klaassen, "Experimental assessment of self-heating in SOI FinFETs," in *Proc. IEDM*, 2009, pp. 1–4.

J.-L.P.J. van der Steen, R.J.E. Hueting, and J. Schmitz, "Extracting Energy Band Offsets on Long-Channel Thin Silicon-on-Insulator MOSFETs," *IEEE Trans. Electron Devices*, vol. 56, no. 9, pp. 1999–2007, 2009.

J.-L.P.J. van der Steen, R.J.E. Hueting, and J. Schmitz, "Extracting Energy Band Offsets on Thin Silicon–On–Insulator MOSFETs," in *Proc. ESSDERC*, 2008, pp. 242–245.

J.-L.P.J. van der Steen, R.J.E. Hueting, G.D.J. Smit, T. Hoang, J. Holleman, and J. Schmitz, "Valence Band Offset Measurements on Thin Silicon–On–Insulator MOSFETs," *IEEE Electron Device Lett.*, vol. 28, no. 9, pp. 821–824, September 2007.

J.-L.P.J. van der Steen, D. Esseni, P. Palestri, L. Selmi, and R.J.E. Hueting, "Validity of the Parabolic Effective Mass Approximation in Silicon and Germanium n-MOSFETs With Different Crystal Orientations," *IEEE Trans. Electron Devices*, vol. 54, no. 8, pp. 1843–1851, August 2007.

J.L. van der Steen, D. Esseni, P. Palestri, and L. Selmi, "Validity of the Effective Mass Approximation in Silicon and Germanium Inversion Layers," in *Int. Workshop on Comput. Electr.*, Vienna, Austria, May 2006, pp. 301–302.

Acknowledgements

This thesis is the result of four years of all work and no play.

Surely.

Fortunately, work comes in two flavors: 1) Work: reading literature, doing simulations, writing papers; and 2) ‘work’: drinking coffee, talking with colleagues, eating apples. Both types of work have been extensively studied over the last four years. Considering the second definition, the activities that pop up in my mind include the “vloer 3 uitje”, the famous SC dinners (with special thanks to Giulia, for being in charge of this successful experiment), the coffee^{+}-es, and –of course– the daily menu, consisting of lunch, coffee, walks etc. This list undoubtedly indicates the social character of the group which, by definition, renders it quite difficult to trace it back to the individual group members. Rather, I would like to thank all of my (former 3rd-floor) colleagues for creating a very smooth and natural transition between Work and ‘work’, and vice versa. I do, however, want to acknowledge some people in particular for contributing to this thesis.

Als eerste mijn co-promotor Ray, die van begin tot het eind wist wat ik aan het doen was. Of deed alsof. Hoe dan ook, Ray, het duurt even, maar dankzij jou weten we waar de Beschaving ligt, zeggen we ook hier Hou-Doe-Hè, en is alles temperatuur-afhankelijk. Kortom, Dank!

Jurriaan, mijn promotor, heeft het uiteindelijk voor elkaar gekregen de co-tutelle met Udine officieel te maken, zodat de samenwerking met Udine die reeds in 2005 begon, ook daadwerkelijk geresulteerd heeft in een gezamenlijke promotie. Bedankt voor deze mooie ‘extra’!

I would like to thank David Esseni, Pierpaolo Palestri and Luca Selmi for hosting me in their group during the summer of 2009, for sharing their thoughts on carrier transport, and for the pleasure of working with you.

Besides, to my colleagues from the SelmiLab: thanks for taking me out for coffee, introducing me to FriuliDoc, prosciutto di San Daniele and frico, to name just a few. And yes, I do like Cabernet.

Hoewel veel resultaten in dit proefschrift verkregen zijn uit simulaties, is er wel degelijk gemeten, zoals aan FinFETs verkregen via NXP Leuven. Bij deze mijn dank aan Mark van Dal, Gerben Doornbos, en Casper Juffermans voor het beschikbaar stellen van deze devices. In de zomer van 2008 heb ik bovendien enige tijd bij NXP Eindhoven doorgebracht om Self-Heating in FinFETs te meten. Wat heet, met de pulsed-IV setup hebben we 'niks' gemeten. Ik ben mijn dank verschuldigd aan Andries Scholten, Gert-Jan Smit, Bert Huizing, Dick Klaassen en, in het bijzonder, aan Hans Tuinhout.
Nog even terug naar Twente. Want Cora, voor veel studenten ben jij het eerste contact met de leerstoel, zo ook voor mij. Of het nou gaat om het D1 project, het maken van 'foto-wafers' of SILC en Hot Carriers. Dank voor het 'binnenslepen'! Sander, Gerdien, Henk, Gerard, en Frederik wil ik graag bedanken voor de ondersteuning in de vorm van onder andere koffie, hulp bij het meten, liedjes schrijven en computer-infrastructuur.
Niek Bouman heeft de zorg voor de voor- en achterkant van dit boekje op zich genomen, zodat ik me kon bezig houden met alles wat ertussen zit. Dank voor dit mooie werk!
Alfons, we zijn er inmiddels achter, AIO'en is meer dan simuleren en/of wafers soppen. Het is ook Aerosmith, wijn, onderdak, en bankhangen. Een collega wordt niet per sé een vriend, maar het kan wel. Heel goed zelfs.
Dan zijn er 'mijn' studenten: Tom van Hemert, mijn eerste 'simulant', en inmiddels geruime tijd collega; en Fleur van Rossem. Fleur, ondanks de beroerde omstandigheden heb je je toch door het afstuderen heengeslagen, ik vind dat een wonder op zich, en wil dat niet onvermeld laten.
Michel, door de jaren heen deelden we heel wat lief en leed, des te mooier is het dat we straks ook het podium kunnen delen. Bedankt voor je raad, steun en gezelligheid!
Dit proefschrift en de bijbehorende feestelijkheden op 1 april 2011 zijn mooie gebeurtenissen. Ik ben er dankbaar voor dat ik dit samen met mijn moeder, vader, zus en broer kan meemaken!
Ruim vier jaar geleden stond er plots een wuppie op mijn bureau. Ik had toen nooit kunnen bedenken wat er in de jaren daarna gebeuren zou. Talitha, bedankt voor de vele mooie momenten.
Annemiek, 'an apple a day, keeps the doctor away'. Voor een kandidaat-doctor klinkt dat toch een beetje vreemd. Naar mijn mening is 'an apple a day' dan ook vooral gezellig. Naast alle praktische dingen, met name bedankt voor de fijne tijd. Annemiek, ik zal je missen.

Tot slot een speciaal woord van dank aan de vrienden die ik in de afgelopen jaren heb leren kennen, via de flat, EL, UT, Udine en Enschede en alle mogelijke kruiscombinaties. Ik ben in de gelukkige positie dat alle namen opnoemen een flinke lijst gaat opleveren. Dus, bij deze aan allen, voor zover de namen niet al genoemd: dank voor de steun en aangename afleiding!

Jan-Laurens van der Steen
februari 2011