UNIVERSITÀ DEGLI STUDI DI UDINE

TESI DI DOTTORATO DI RICERCA

Per ottenere il titolo di
Dottore di Ricerca dell'Universitá di Udine

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

Daniel LIZZIT

Udine, 08/04/2016

Titolo Tesi di Dottorato

ADVANCED MODELS FOR SIMULATION OF PLANAR AND GATE-ALL-AROUND NANOSCALE MOSFETS

Relatore:
Prof. David ESSENI

Contents

List of Symbols

BTE	Boltzmann transport equation
CVD	Chemical vapor deposition
CTP	Classical turning points
CCS	Crystal coordinate system
DFT	Density functional theory
DoS	Density of states
DCS	Device coordinate system
DFT	Discrete Fourier transform
DIBL	Drain induced barrier lowering
V_{DS}	Drain to source voltage
DD	Drift-diffusion
L_{eff}	Effective gate length
EMA	Effective mass approximation
e	Electron charge
ECS	Ellipsoid coordinate system
EOT	Equivalent oxide thickness
GAA	Gate all around
HS-QW	Hetero structure quantum well
HP	High performance
LSTP	Low stand by power
MOVPE	Metal organic vapor phase epitaxy
MOSFET	Metal oxide semiconductor field effect transistor
L_G	Metallurgic gate length
MBE	Molecular beam epitaxy
MRT	Momentum relaxation time approximations
MSMC	Multi subband Monte Carlo
NW	Nanowire
NEGF	Non-equilibrium Green's function
PDE	Partial differential equations
PS	Pseudo spectrals
RPA	Random phase approximation
r.m.s	Root mean square
SCE	Short channel effects

SoC	Silicon on chip
sSi	Strained silicon
SS	Sub threshold slope
SR	Surface roughness
ToB	Top of the barrier
UTB	Ultra thin body
VLS	Vapor liquid solid
VLSI	Very large scale integration
VS	Virtual source

Abstract

SINCE many years we are witnessing a progressive scaling of transistor feature size in order to increase packing densities and to meet market demands for powerful small size devices with reduced power dissipation. At the state of art, performance degradation, integration for present and next generation of devices and increased variability are among the major challenges to scaling. Degraded performances are due to a combination of reduced mobility caused mainly by non-uniformity of the interfaces at atomic scales, parameters variability introduced by the discreteness of charge and matter, and, a progressively worse electrostatic control of the gate on the channel as the physical dimension of the devices are scaled down, thus leading to increased leakage, worst drive-current and timing issues in logic circuits. High leakage current in deep sub-micron devices is becoming a significant contributor to power dissipation of CMOS circuits, therefore improving performance while maintaining leakage control and scaling dimensions is very important, especially for low power applications.

To improve the drive current and sub-threshold regime, CMOS technology is actively exploring devices with channel materials alternative to silicon (e.g. III-V group semiconductors) and/or alternative device architectures (e.g. FinFET, GAA MOSFETs) with induced channel-stress to further improve performances. In this framework, the modelling of modern devices should be able to take into account the most relevant technological options to save development, implementation time and costs. This requires an appropriate description of the most relevant scattering mechanisms and of the quantum mechanical effects. Due to the lack of dependable, commercial TCAD models that can predict the performance of the next generation devices, the aim of this thesis is the development and validation of TCAD tools for both the purpose of device performance analysis and performance improvement. We first present a comparative simulation study of Ultra-Thin-Body (UTB) strained silicon (sSi) and III-V semiconductor based MOSFETs by using a comprehensive semiclassical Multisubband Monte Carlo (MSMC) transport model. In particular we show that, due to the finite screening length in the source-drain regions, III−V and Si nanoscale MOSFETs with a given gate length (L_G) may have a quite different effective gate length (L_{eff}), and the difference in L_{eff} provides a useful insight to interpret the performance comparison of III−V and Si MOSFETs. In the second part we present a new model for the surface roughness scattering. The model is suitable for bulk, for

0. Abstract

UTB and for Hetero-Structure Quantum Well (HS-QW) MOSFETs. Comparison with experimental mobility for Si bulk MOSFETs shows that a good agreement with measured mobility can be obtained with a r.m.s. value of the Surface Roughness (SR) spectrum close to several AFM and TEM measurements. Finally, we developed a deterministic solver for the Boltzmann Transport Equation (BTE) for Gate All Around (GAA) circular MOSFETs. In particular, we solve the Schrödinger equation for arbitrary crystal transport directions within the Effective Mass Approximation (EMA) including the wave-function penetration into the oxide and the nonparabolicity of the energy dispersion relation along the quantization plane and transport direction. Then, the BTE is solved without any a-priori assumption and including the main scattering mechanisms responsible for performance degradation, with a new model for the SR scattering.

Chapter 1

Introduction

FOR many decades the Complementary-Metal-Oxide-Semiconductor (CMOS) transistor has been continually scaled down, which has resulted in high transistors density per unit of wafer area but also in faster chips that consume less power per transistor in every generation. This has been made possible by the improved lithographic capability to print shorter gate lengths and the ability to grow nearly perfect oxides with ever decreasing thicknesses up to few atomic layers [1].

In past years, scaling has followed three paradigms and the main features are reported in Tabs.1.1 and 1.2. In *constant field* scaling [2], physical dimensions and applied voltage are scaled by the same factor while doping concentration is increased by the same factor, because the target is to maintain a constant electric field to ensure the electrostatic integrity and reliability of devices. However, this may jeopardize compatibility with other devices that operate at different voltages and decrease on-performances due to increased Coulomb scattering. This scenario yields the largest power-delay product reduction (see Tab.1.2), and V_T scales as the V_{DD} but the scaling is limited by the targets on the I_{OFF} prescribed by the roadmap [3].

An alternative approach, which is referred to as *constant voltage* scaling, consists in keeping the voltage constant for an easier integration with pre-existing systems, but the disadvantage is that the electric field increases as the minimum feature length is reduced. This leads to velocity saturation, mobility degradation, increased leakage currents and lower breakdown voltages. For these reasons, the typical approach to scaling feature size for devices has been a middle way between the two scaling techniques, namely *generalized* scaling [4] where voltages are scaled less aggressively than linear dimensions, while still maintaining constant the shape of the electric-field pattern. Even with the best choice of scaling rules, as gate length becomes the same order of magnitude as the depletion layer width, several phenomena affect adversely the performance of MOS devices:

- reduction in threshold voltage of short devices compared to the long channel ones making V_T dependent on V_{DS} (Voltage threshold roll-off).
- degradation of gate oxide due to tunneling and hot carriers.

These aforementioned issues are due on the one hand to: Short Channel Effects (SCE), Drain-Induced Barrier Lowering (DIBL) and Punchthrough, and on the other hand to Impact Ionization, Tunneling and Channel Hot Electrons effects. In this respect, the success attained by CMOS technology is also related to the strategies used to continue along the path of technology improvement. In fact, in the last ten years the gate length has not scaled proportionately with the device pitch [5–10] because of other technology boosters that can provide some relief by also allowing more aggressive V_{DD} scaling: new materials for the gate dielectric in replacement of the SiO_2, namely high-κ gate dielectrics, that can keep low the Equivalent Oxide Thickness (EOT), architectures alternative to the planar bulk MOSFETs with induced stress, as well as high-mobility channel materials (i.e. III-V group semiconductors).

As reported in [3], in future devices for logic applications the amount of power that can be dissipated by a single transistor as well as the total amount per chip (which increases rapidly due to the aggressive downscaling) are the limiting factors along the path of miniaturization. Power dissipation is given by the contribution of the dynamic power $P_{dyn}=\alpha\ C_{load}{V_{DD}}^2\ f_{switch}$(where α is the fraction of gates actively switching) and static power ($P_{stat}=I_{off}V_{DD}$). Consequently, the V_{DD} lowering is the most effective way for power reduction. After a regular decrease of V_{DD} from 5V in the '80s to 1.2V in the 2002, we have seen a plateau close to 1V till 2010. In recent years lot of efforts have been put into reducing V_{DD}, however, when V_{DD} approaches about 0.6 V, it poses noise/variability challenges, and moreover, it becomes more difficult to scale than other parameters mainly because of the fundamental limit of the sub-threshold slope of 60 mV/decade in MOS devices. This fact, along with the need for high performance transistors to achieve currents in the order of 1.9 mA/μm at V_{DD}=0.68 V (for HP MOSFETs for the year 2020 [3]), makes the static power dissipation particularly difficult to control while at the same time meeting aggressive targets for performance scaling.

1.1 Overview of MOSFET scaling

1.2 Material and device innovations

1.2.1 Alternative channel materials

Semiconductors alternative to silicon and belonging to the III-V group of the periodic table are being actively investigated to improve the transistors performances or keep them constant while decreasing transistors feature size and V_{DD} because of their superior injection velocity v_{inj} as shown in Fig.1.1. In fact, according to [12, 13] the on-current (I_{ON}) in the quasi-ballistic transport regime is defined at V_{DS}=V_{GS}=V_{DD} and can be expressed as:

$$I_{ON} = W C_{G,eff}(V_{DD} - V_T) v_{inj} \frac{1-r}{1+r} \tag{1.1}$$

Table 1.1: Comparison of the effect of scaling on MOSFET device parameters. $1 < k < \alpha$ and $\alpha > 1$.

Parameter	**Expression**	**Constant Field**	**Constant Voltage**	**Generalized scaling**
Dimensions W,L,t_{ox}		$\frac{1}{\alpha}$	$\frac{1}{\alpha}$	$\frac{1}{\alpha}$
Voltage V_{dd},V_t		$\frac{1}{\alpha}$	1	$\frac{1}{k}$
Area A	WL	$\frac{1}{\alpha^2}$	$\frac{1}{\alpha^2}$	$\frac{1}{\alpha^2}$
Doping concentration N_A	$\frac{\#dopant}{cm^3}$	α	α^2	$\frac{\alpha^2}{k}$
Gate capacitance C_{gate}	$\frac{\varepsilon_{ox}}{t_{ox}}WL$	$\frac{1}{\alpha}$	$\frac{1}{\alpha}$	$\frac{1}{\alpha}$
Electric Field F		1	α	$\frac{\alpha}{k}$
Current $\mathrm{I_{ON}}$		$\frac{1}{\alpha}$	α	$\frac{\alpha}{k^2}$
Current density J	$\mathrm{I_{ON}}/Area$	α	α^3	$\frac{\alpha^3}{k^2}$

Table 1.2: Power-delay scaling.

Parameter	**Expression**	**Constant Field**	**Constant Voltage**	**Generalized scaling**
Power P	$\mathrm{I_{ON}V_{DD}}$	$\frac{1}{\alpha^2}$	α	$\frac{\alpha}{k^3}$
Gate delay t_d	$\frac{C_{gate}\mathrm{V_{DD}}}{\mathrm{I_{ON}}}$	$\frac{1}{\alpha}$	$\frac{1}{\alpha^2}$	$\frac{k}{\alpha^2}$
Power delay $P \cdot t_d$	$C_{gate}\mathrm{V_{DD}^2}$	$\frac{1}{\alpha^3}$	$\frac{1}{\alpha}$	$\frac{1}{k^2\alpha}$

where W is the device width, $\mathrm{v_{inj}}$ the injection velocity defined as the average electron velocity at the virtual source (VS), r is the back-scattering coefficient and $\mathrm{C_{G,eff}}$ is an effective gate capacitance accounting for quantum mechanical confinements, finite density-of-states and dead spaces at the oxide-semiconductor interfaces [14]. $\mathrm{C_{G,eff}}$ is given by the series combination of the gate oxide capacitance ($\mathrm{C_{ox}}$) and the inversion layer capacitance ($\mathrm{C_{inv}}$), which in turn is given by the series of the centroid capacitance ($\mathrm{C_{centr,i}}$) and a quantum capacitance ($\mathrm{C_{quant,i}}$) defined in [15]. By assuming a simplified case in which only one subband contributes to the transport (quantum-limit transport) and a strong degeneracy in Fermi-Dirac statistics, $\mathrm{C_{inv}}$ is given by [16]:

$$\mathrm{C_{inv}} \approx \frac{e^2 \mu_\nu \mathrm{m_{DOS}}}{\pi \hbar^2} \tag{1.2}$$

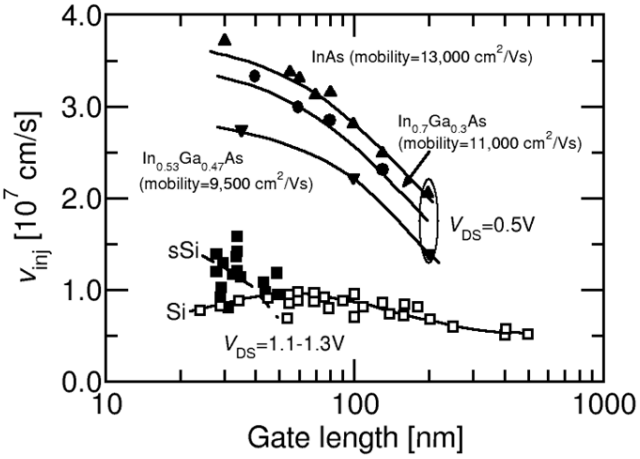


Figure 1.1: Experimentally extracted injection velocity v_{inj}in HEMTs measured at VDD=0.5V and Si MOSFETs at VDD=1.1-1.3 V [11]

where μ_ν is the valley degeneracy and m_{DOS} the DOS effective mass. The inversion capacitance in Eq.1.2 is useful to compare the on-current for different semiconductors at a given bias. In fact, looking at Eqs.1.1 and 1.2, the gain given by larger v_{inj} for III-V based semiconductors is mitigated by the counteracting effect of a lower m_{DOS}if compared with silicon, which leads to a smaller inversion capacitance at given supply voltage with respect to materials with larger masses. We also calculate a first-order estimate [17] of switching delay of a digital gate as:

$$T_{sw} = \frac{Q_{sw}}{I_{ON}} \tag{1.3}$$

where Q_{sw} is the switched charge and includes also parasitics (i.e. interconnection capacitance, fringing capacitance). Eqs.1.1 and 1.3 suggest that T_{sw} is roughly proportional to ${v_{inj}}^{-1}$, regardless of Q_{sw} being mostly due to the intrinsic transistor charge or to the parasitic capacitances. More refined estimates of T_{sw} can be obtained by replacing the I_{ON} in Eq.1.3 with an effective drive current [18, 19], however, Eq.1.3 is reasonably adequate for the purpose of comparing MOSFETs with different semiconductor materials.

Since T_{sw} is inversely proportional to v_{inj}, it is substantially degraded when V_{DD} is reduced because, for a given technology, it can be demonstrated that v_{inj} increases with V_{GS} [13], because the thermal velocity at the virtual source is enhanced by V_{GS} through the electron degeneracy and, furthermore, the channel back-scattering coefficient is reduced increasing V_{GS} [13].

In this scenario, new channel materials with large carrier mobility and injection velocity (see Fig.1.1) are being investigated for n-type MOSFETs [20–26] as shown in Fig.1.2 to allow for V_{DD} reduction at approximately constant delay.

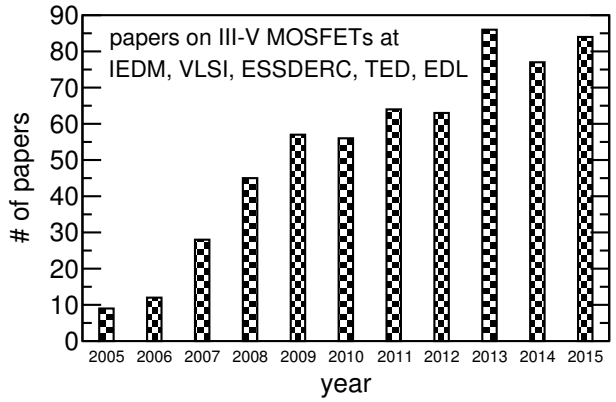


Figure 1.2: Number of papers on III-V semiconductor MOSFETs published on major journals and conference proceedings (IEEE, IEDM, ESSDERC, VLSI Symp.).

1.2.2 Alternative device architectures

Planar transistors have been the backbone of integrated circuits for several decades and the constant scaling has reduced the channel length such that SCE have become very challenging. Hence, through continuous efforts to fabricate devices beyond bulk MOSFET scaling limits, CMOS technology is rapidly moving from classical planar single gate devices toward UTB and Double-Gate UTB (DG-UTB) MOSFETs, and then to 3D structures such as TriGate, FinFET, and to GAA MOSFETs.

The short channel effects arise from the electrostatic effect of the source and drain regions on the channel, namely when source/drain competes with gate to control the channel barrier, so in order to improve the electrostatic integrity (EI) the gate to channel capacitive coupling must be large compared to the source/drain to channel coupling [27, 28]. To achieve this objective in a single gate MOS transistor having a thick channel, it is necessary a high level doping to decrease the depletion depth [27] and in addition shallow junctions are required in source/drain regions. Conversely, in UTB SOI and in 3D devices (TriGate, FinFETs, GAA MOSFETs) sub-threshold parameters are better than in a traditional MOSFETs (without affecting the doping), due to the reduced channel thickness in the first case and due to more than one gate surrounding the channel in the second.

In UTB MOSFETS (also known as a fully depleted SOI) [29, 30] the reduction of off-current is obtained thanks to the use of a body region which is significantly thinner than the gate length (about 1/3 to 0.5 of the natural length scale [31, 32]) in order to maintain full substrate depletion under gate control and prevent leakage paths far from the gate. The potential barrier for any path between the source and drain is strongly coupled to the gate than to the drain due to a better electrostatic control, leading to smaller SCE and superior EI [28] compared to the conventional Bulk MOSFETs. Furthermore, the undoped UTB is a promising choice since it can yield higher carrier mobilities due to negligible impurity scattering, offering the opportunity to eliminate the impact of random dopant fluctuations on the threshold

voltage [33]. A further improvement consists in UTB Double Gate (DG-UTB), that can be regarded as the natural extension of SOI technology for which the effective thickness of the body becomes equal to $T_w/2$ without any effort on doping. Moreover, in UTB-DG MOSFETs, back-gate bias option results in better control of V_T by exploiting the electrostatic coupling from two gates on either side of the channel.

A further evolution driven by extremely accurate lithography processes consist in FinFET, TriGate-FET, Omega-FET and GAA MOSFETs that allow high density lateral and vertical integration. They overcome the limit of scaling issue of planar MOSFTEs and inherently have good suppression of SCEs, high transconductance and quasi ideal sub-threshold slope (SS) [34]. Moreover, they can easily incorporate some of the technology boosters including alternative channel orientations, strained channels [35], high-k gate dielectrics and metal gate electrodes, so that, they can provide significantly higher nMOS and pMOS drive current than the planar counterpart at the same off-state leakage [36]. Short channel effects can be controlled with much less stringent oxide thickness requirements than in traditional MOSFETs due to a better electrostatic control of the gate. However, the superior electrostatic control will be offset by parasitics (i.e. gate-to-source/drain capacitance) [37].

FinFET transistor was first demonstrated in 1998 [38] and the conduction takes place on the vertical side-walls of the fin, thus the effective device width of a narrow fin is approximately twice the fin height. The thickness requirement is relaxed to approximately 2/3 to 1 of the natural channel length since each gate controls half of the body thickness [39]. First III-V based FinFET was demonstrated in 2009 [40] and in 2014 INTEL Corporation has announced its new 14-nm technology node FinFET [41] featuring an L_G of 20nm, W_{fin}= 8nm and H_{fin}=42nm.

Triple-Gate MOSFETs [42] and Omega-FinFETs on buried oxide[43] and on bulk substrate [44] have been proposed as a means to alleviate the stringent thin width requirements of FinFETs, enabling more relaxed fin aspect-ratio (H_{fin}/W_{fin}) due to additional electrostatic control from the top gate. In this respect, in 2012 Intel has reported [45] on a leading edge 22nm Silicon-on-Chip (SoC) process technology featuring 3-D Tri-Gate MOSFETs for logic applications.

Finally, Gate All Around MOSFETs, that have been first demonstrated in silicon [46] and later in III-V [47][48], are the most promising architectures for ultimate device scaling allowing to achieve the best electrostatic control since the gate surrounds the entire channel.

Finally, Figs.1.3, 1.4, 1.5, 1.6 report the most important figures of merit for logic applications for HP devices reported in experiments published between 2007 and 2015 for planar, FinFET, HEMTs and nanowire transistors. To make this comparison fair, the displayed data points meet strict criteria: data are extracted at the relevant drain bias V_{DS}= 0.5 V for III-V based MOSFETs and V_{DS}>=0.7 V for silicon-based MOSFETs (values are referenced in the figures). For I_{ON}-versus-L_G results (Fig.1.4), only data with I_{OFF}= 100 nA/μm have been included since the target is the analysis of HP devices (I_{OFF} is defined at V_{GS}= 0 V and the I_{OFF} target meets the requirements in [3]), whereas I_{ON} has been extracted at V_{DS}=V_{GS}=0.5

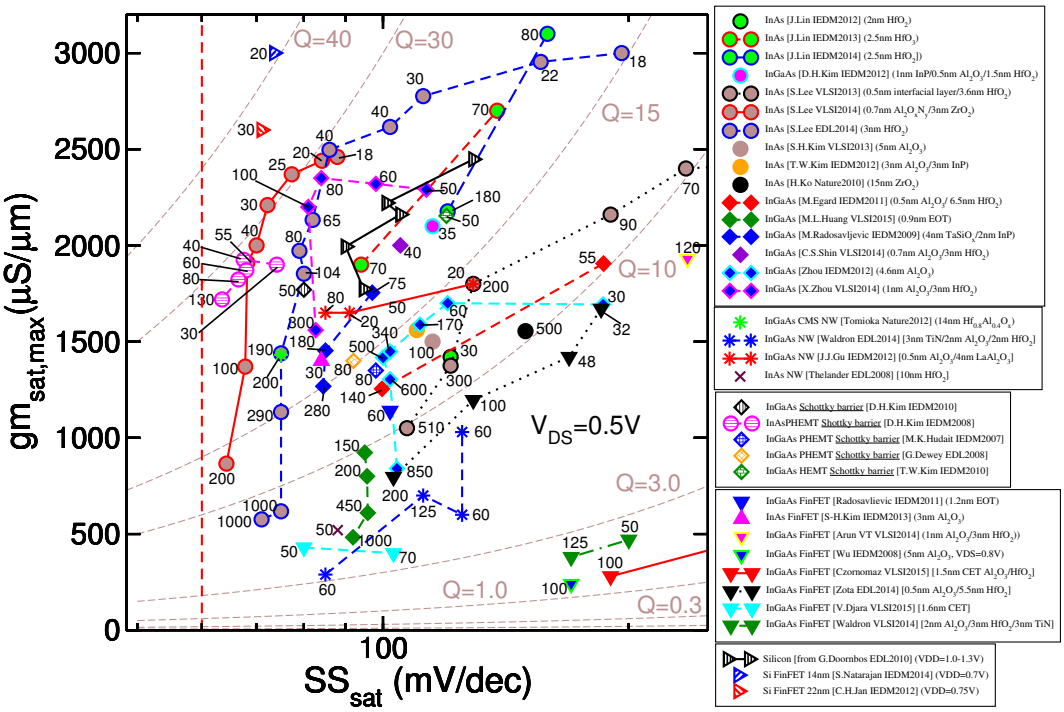


Figure 1.3: Measured maximum saturation transconductance $gm_{sat,max}$-versus-SS for III-V- and silicon-based transistors and for V_{DS}=0.5 V unless otherwise stated. Extracted data fulfil requirements for I_{OFF} =100 nA/μm or, if not possible, for $I_{ON}/I_{OFF}> 1{\cdot}10^3$. Dashed lines: lines of constant Q=$gm_{sat,max}$/SS. L_G values are referenced therein close to the symbols.

V for III-V-based or V_{DS}=V_{GS}>= 0.7 V for silicon-based transistors. For SS_{sat} versus L_G in Fig.1.5, $gm_{sat,max}$ versus SS_{sat} in Fig.1.3 and DIBL versus L_G in Fig.1.6 we have reported only data with I_{OFF}= 100 nA/μm or, in the case of larger I_{OFF} values, with an I_{ON}/I_{OFF} ratio larger than $1{\cdot}10^3$ (with I_{ON} defined as the current at V_{DS}=V_{GS}= 0.5 V for III-V-based or > 0.7 V for silicon-based transistors).

Fig.1.3 shows that the transconductance $gm_{sat,max}$ strongly depends on the device channel length and, that silicon devices are still superior to III-V based transistors. This is due to the combination of worse oxides in III-V based devices with respect to the native SiO_2 oxide used for Silicon MOSFETs and in general terms, to a less mature technology for III-V transistors. However, very good results are obtained for InAs MOSFETs with SS_{sat} values approaching 60 mv/dec and large $gm_{sat,max}$ values. Nevertheless, the leading edge INTEL silicon devices for the 22-nm and 14-nm technology node [41, 45] outperform the competitor devices in terms of SS and $gm_{sat,max}$, but at a larger operating voltage with respect to III-V MOSFETs. Fig.1.4 shows the I_{ON} current for transistors at the operating voltage V_{GS}=V_{DS}=0.5V, unless otherwise stated, at the same I_{OFF} current. It can be shown that the best results are obtained for silicon but at a voltage quite above the operating voltage of 0.5V used

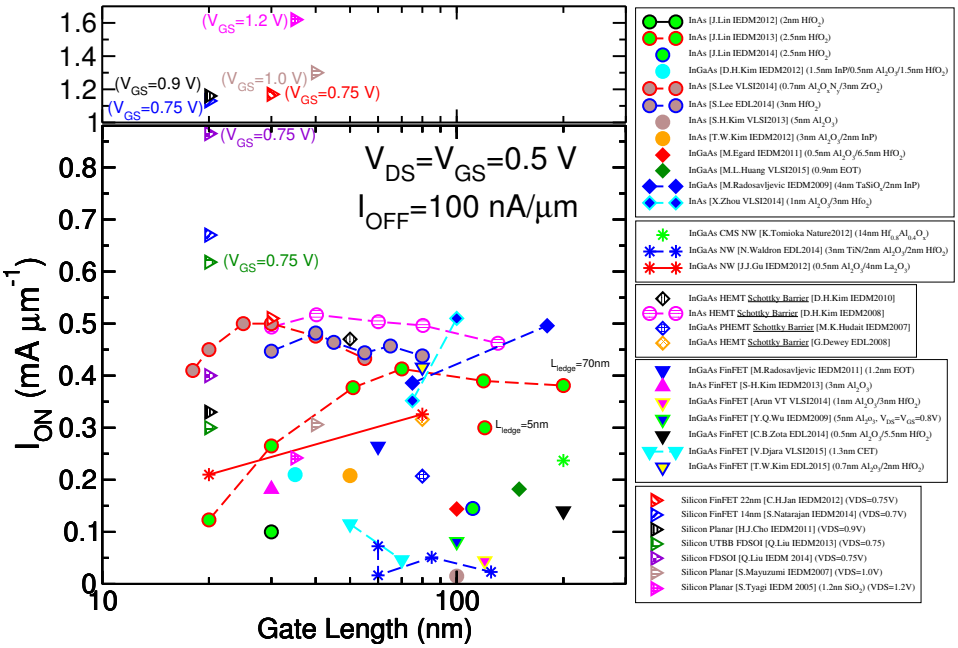


Figure 1.4: Measured on-current I_{ON}-versus-L_G at V_{DS}=V_{GS}= 0.5 V unless otherwise stated. I_{OFF}= 100 nA/μm.

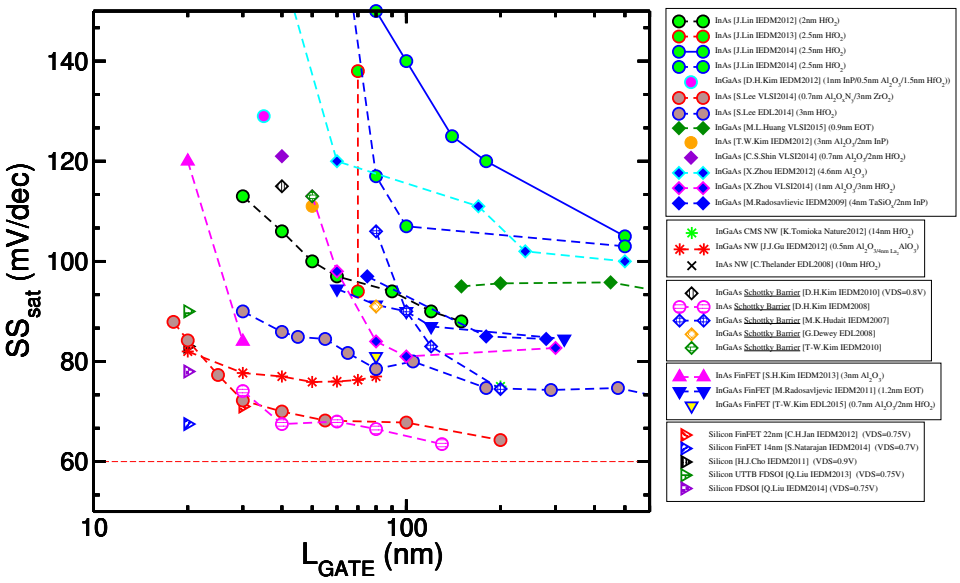


Figure 1.5: Measured SS-versus-L_G at V_{DS}= 0.5 V unless otherwise stated. I_{OFF} requirements are the same as reported in Fig.1.3.

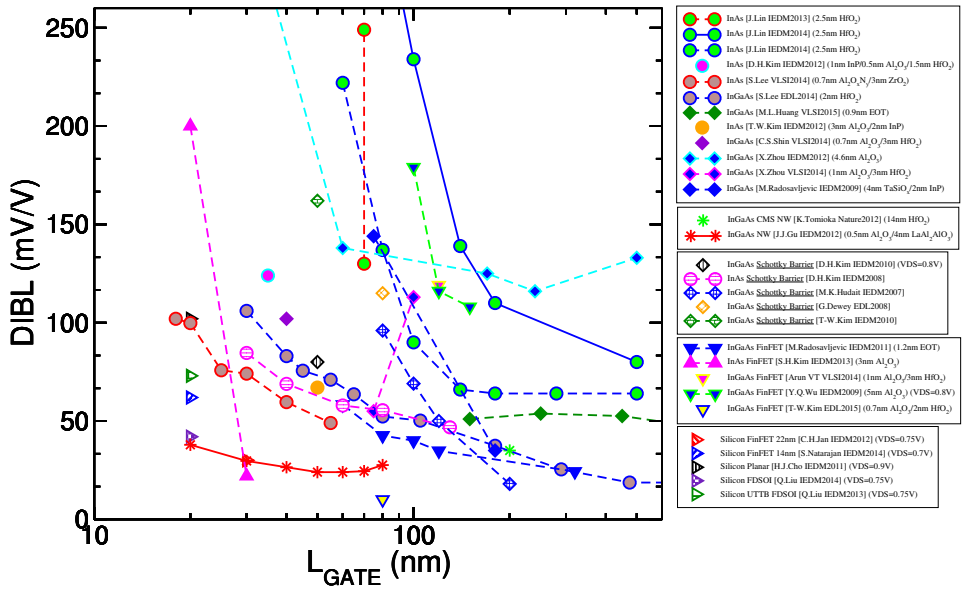


Figure 1.6: Measured DIBL-versus-L_G at V_{DS}= 0.5 V unless otherwise stated. I_{OFF} requirements are the same as reported in Fig.1.3.

for the III-V MOSFETs reported in the figure. In fact, when the gate voltage of the silicon based MOSFETs is decreased to 0.5V (right-triangles), III-V MOSFETs show performances comparable, or better, to those of silicon.

1.2.3 Purpose of this work

In the framework of aggressively scaled devices and architectures alternative to planar devices, the aim of this thesis is manyfold. We start by the metrics that can provide useful insight when III-V- and silicon- based transistors are compared. Then we develop a new model for one of the most important scattering mechanism in ultra-thin channel and nanowire MOSFETs, that is surface roughness scattering. Another important part of the work is the development of a comprehensive simulator for GAA MOSFETs that is a valuable alternative to conventional simulators based on drift-diffusion solutions, MonteCarlo-like solvers [49], or highly computational burden full-quantum simulators [50–56] including also the main scattering mechanisms responsible for performance degradation.

The present manuscript is organized as follows. Chapter2 presents a comparative simulation study of ultrathin-body InAs, InGaAs and strained Si MOSFETs by using a comprehensive semiclassical multisubband Monte Carlo (MSMC) transport model. Our results show that, due to the finite screening length in the source-drain regions, III-V and Si nanoscale MOSFETs with a given L_G may have a quite different L_{eff}. The difference in L_{eff} provides a useful insight to interpret the performance comparison of III-V and Si MOSFETs and we also show that the engineering of the source-drain regions has a remarkable influence on the overall performance of nanoscale III-V MOSFETs. In Chapter3 we present a new model for the SR limited mobility in MOS transistors. The model is suitable for bulk , planar thin body devices and nanowire MOSFETs and explicitly takes into account the non linear relation between the displacement Δ of the interface position and the SR scattering matrix elements, which is found to significantly influence the r.m.s value ($\Delta_{r.m.s.}$) of the interface roughness that is necessary to reproduce SR-limited mobility measurements. In particular, comparison with experimental mobility for bulk Si MOSFETs shows that with the new SR scattering model a good agreement with measured mobility can be obtained with $\Delta_{r.m.s.}$ values of about 0.2 nm, which is in good agreement with several AFM and TEM measurements. For thin body III-V MOSFETs, the proposed model predicts a weaker mobility degradation at small T_w, compared to the Tw^6 behavior observed in Si extremely thin body devices.

Chapter4 presents a Schrödinger-Poisson solver for circular sections of nanowire transistors taking into account the non-parabolicity of the electron energy relation along transport direction and quantization plane and, furthermore, allows one to simulate different transport directions considering valleys anisotropy. Also the penetration of the electron wave-function into the gate oxide material is considered. Finally, Chapter5 presents a deterministic solver for the BTE equation in GAA circular nanowires. All the main scattering mechanisms are included as well as screening effects for surface roughness and coulomb scattering. The model for the SR scattering is the extension to circular domains of the new model developed for planar devices and reported in Chapter3.

References

[1] M. Krishnan, L. Cheng, T.-J. King, J. Bokor, and C. Hu, "MOSFETs with 9 to 13 A thick gate oxides", in *IEEE IEDM Technical Digest*, Dec. 1999, pp. 241–244.

[2] R. Dennard, V. Rideout, E. Bassous, and A. LeBlanc, "Design of ion-implanted MOSFETs with very small physical dimensions", *IEEE Journal of Solid-State Circuits*, vol. 9, no. 5, pp. 256–268, Oct. 1974.

[3] (2013). International Technology Roadmap for Semiconductors, [Online]. Available: `http://www.itrs.net/Links/2012ITRS/Home2012.htm`.

[4] G. Baccarani, M. Wordeman, and R. Dennard, "Generalized scaling theory and its application to a 1/4 micrometer MOSFET design", *IEEE Trans. on Electron Devices*, vol. 31, no. 4, pp. 452–462, Apr. 1984.

[5] T. Ghani et al., "A 90nm high volume manufacturing logic technology featuring novel 45nm gate length strained silicon CMOS transistors", in *IEEE IEDM Technical Digest*, Dec. 2003, pp. 11.6.1–11.6.3.

[6] S. Tyagi et al., "An advanced low power, high performance, strained channel 65nm technology", in *IEEE IEDM Technical Digest*, Dec. 2005, pp. 245–247.

[7] K. Mistry et al., "A 45nm Logic Technology with High-k Metal Gate Transistors, Strained Silicon, 9 Cu Interconnect Layers, 193nm Dry Patterning, and 100 Pb-free Packaging", in *IEEE IEDM Technical Digest*, Dec. 2007, pp. 247–250.

[8] P. Packan et al., "High performance 32nm logic technology featuring 2nd generation high-k metal gate transistors", in *IEEE IEDM Technical Digest*, Dec. 2009, pp. 1–4.

[9] C.-H. Jan et al., "A 22nm SoC platform technology featuring 3-D tri-gate and high-k metal gate, optimized for ultra low power, high performance and high density SoC applications", in *IEEE IEDM Technical Digest*, Dec. 2012, pp. 3.1.1–3.1.4.

[10] S. Natarajan et al., "A 14nm logic technology featuring 2nd-generation FinFET, air-gapped interconnects, self-aligned double patterning and a 0.0588 μm^2 SRAM cell size", in *IEEE IEDM Technical Digest*, Dec. 2014, pp. 3.7.1–3.7.3.

[11] J. Del Alamo, "Nanometre-scale electronics with III-V compound semiconductors", *Nature*, vol. 479, pp. 317 –323, 2011.

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

[13] D. Esseni, P. Palestri, and L. Selmi, *Nanoscale MOS Transistors - Semi-Classical Transport and Applications*, 1st. Cambridge University Press., 2011.

[14] Y. Taur and T. Ning, *Fundamentals of VLSI Devices*, 1st. Cambridge University Press., 2011.

[15] D. Jin, D. Kim, T. Kim, and J. del Alamo, "Quantum capacitance in scaled down III-V FETs", in *IEEE IEDM Technical Digest*, Dec. 2009, pp. 1–4.

[16] G. Hiblot, Q. Rafhay, F. Boeuf, and G. Ghibaudo, "Analytical Model for the Inversion Gate Capacitance of DG and UTBB MOSFETs at the Quantum Capacitance Limit", *IEEE Trans. on Electron Devices*, vol. 62, no. 5, pp. 1375–1382, May 2015.

[17] F. Conzatti, M. Pala, D. Esseni, E. Bano, and L. Selmi, "Strain-Induced Performance Improvements in InAs Nanowire Tunnel FETs", *IEEE Trans. on Electron Devices*, vol. 59, no. 8, pp. 2085 –2092, 2012.

[18] M.-H. Na, E. Nowak, W. Haensch, and J. Cai, "The effective drive current in CMOS inverters", in *IEEE IEDM Technical Digest*, 2002, pp. 121–124.

[19] J. Deng and H.-S. Wong, "Metrics for performance benchmarking of nanoscale Si and carbon nanotube FETs including device nonidealities", *IEEE Trans. on Electron Devices*, vol. 53, no. 6, pp. 1317–1322, 2006.

[20] M. Luisier, "Performance Comparison of GaSb, Strained-Si, and InGaAs Double-Gate Ultrathin-Body n-FETs", *IEEE Electron Device Lett.*, vol. 32, no. 12, pp. 1686–1688, 2011.

[21] M. Heyns et al., "Advancing CMOS beyond the Si roadmap with Ge and III/V devices", in *IEEE IEDM Technical Digest*, Dec. 2011, pp. 13.1.1–13.1.4.

[22] S. Kim, M. Yokoyama, N. Taoka, R. Nakane, T. Yasuda, O. Ichikawa, N. Fukuhara, M. Hata, M. Takenaka, and S. Takagi, "Enhancement technologies and physical understanding of electron mobility in III-V n-MOSFETs with strain and MOS interface buffer engineering", in *IEEE IEDM Technical Digest*, Dec. 2011, pp. 13.4.1–13.4.4.

[23] M. Radosavljevic et al., "Electrostatics improvement in 3-D tri-gate over ultra-thin body planar InGaAs quantum well field effect transistors with high-K gate dielectric and scaled gate-to-drain/gate-to-source separation", in *IEEE IEDM Technical Digest*, Dec. 2011, pp. 33.1.1–33.1.4.

[24] X. Zhou, Q. Li, C. W. Tang, and K. M. Lau, "30nm enhancement-mode $In_{0.53}Ga_{0.47}As$ MOSFETs on Si substrates grown by MOCVD exhibiting high transconductance and low on-resistance", in *IEEE IEDM Technical Digest*, Dec. 2012, pp. 32.5.1–32.5.4.

[25] S.-H. Kim, M. Yokoyama, R. Nakane, O. Ichikawa, T. Osada, M. Hata, M. Takenaka, and S. Takagi, "High Performance Tri-Gate Extremely Thin-Body InAs-On-Insulator MOSFETs With High Short Channel Effect Immunity and V_{th} Tunability", *IEEE IEDM Technical Digest*, vol. 61, no. 5, pp. 1354–1360, May 2014.

[26] S. Lee, V. Chobpattana, C.-Y. Huang, B. Thibeault, W. Mitchell, S. Stemmer, A. Gossard, and M. Rodwell, "Record Ion (0.50 mA/μm at V_{DD}= 0.5 V and I_{OFF}= 100 nA/μm) 25 nm-gate-length ZrO2/InAs/InAlAs MOSFETs", Jun. 2014, pp. 1–2.

[27] A. Pouydebasque, C. Charbuillet, R. Gwoziecki, and T. Skotnicki, "Refinement of the Subthreshold Slope Modeling for Advanced Bulk CMOS Devices", *IEEE Trans. on Electron Devices*, vol. 54, no. 10, pp. 2723–2729, Oct. 2007.

[28] T. Skotnicki et al., "Innovative Materials, Devices, and CMOS Technologies for Low-Power Mobile Multimedia", *IEEE Trans. on Electron Devices*, vol. 55, no. 1, pp. 96–130, Jan. 2008.

[29] Y.-K. Choi, K. Asano, N. Lindert, V. Subramanian, T.-J. King, J. Bokor, and C. Hu, "Ultrathin-body SOI MOSFET for deep-sub-tenth micron era", *IEEE Electron Device Lett.*, vol. 21, no. 5, pp. 254–255, May 2000.

[30] K. Cheng et al., "Extremely thin SOI (ETSOI) CMOS with record low variability for low power system-on-chip applications", in *IEEE IEDM Technical Digest*, Dec. 2009, pp. 1–4.

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

[32] R. Chau, J. Kavalieros, B. Doyle, A. Murthy, N. Paulsen, D. Lionberger, D. Barlage, R. Arghavani, B. Roberds, and M. Doczy, "A 50 nm depleted-substrate CMOS transistor (DST)", in *IEEE IEDM Technical Digest*, Dec. 2001, pp. 29.1.1–29.1.4.

[33] K. Takeuchi, R. Koh, and T. Mogami, "A study of the threshold voltage variation for ultra-small bulk and SOI CMOS", *IEEE Trans. on Electron Devices*, vol. 48, no. 9, pp. 1995–2001, Sep. 2001.

[34] J. P. Colinge, *FinFETs and Other Multi-Gate Transistors*. Springer Science and Business Media, 2008.

[35] I. Tienda-Luna, F. Ruiz, A. Godoy, and F. Gamiz, "The influence of orientation and strain on the transport properties of SOI trigate nMOSFETs", Sep. 2009, pp. 319–322.

[36] B. Doyle, B. Boyanov, S. Datta, M. Doczy, S. Hareland, B. Jin, J. Kavalieros, T. Linton, R. Rios, and R. Chau, "Tri-Gate fully-depleted CMOS transistors: fabrication, design and layout", Jun. 2003, pp. 133–134.

[37] M. Guillorn et al., "FinFET performance advantage at 22nm: An AC perspective", Jun. 2008, pp. 12–13.

[38] D. Hisamoto, W.-C. Lee, J. Kedzierski, E. Anderson, H. Takeuchi, K. Asano, T.-J. King, J. Bokor, and C. Hu, "A folded-channel MOSFET for deep-sub-tenth micron era", in *IEEE IEDM Technical Digest*, Dec. 1998, pp. 1032–1034.

[39] N. Lindert, L. Chang, Y.-K. Choi, E. Anderson, W.-C. Lee, T.-J. King, J. Bokor, and C. Hu, "Sub-60-nm quasi-planar FinFETs fabricated using a simplified process", *IEEE Electron Device Lett.*, vol. 22, no. 10, pp. 487–489, Oct. 2001.

[40] Y. Wu, R. Wang, T. Shen, J. Gu, and P. Ye, "First experimental demonstration of 100 nm inversion-mode InGaAs FinFET through damage-free sidewall etching", in *IEEE IEDM Technical Digest*, Dec. 2009, pp. 1–4.

[41] S. Natarajan et al., "A 14nm logic technology featuring 2nd-generation FinFET, air-gapped interconnects, self-aligned double patterning and a 0.0588 μm SRAM cell size", in *IEEE IEDM Technical Digest*, Dec. 2014, pp. 3.7.1–3.7.3.

[42] B. Doyle, S. Datta, M. Doczy, S. Hareland, B. Jin, J. Kavalieros, T. Linton, A. Murthy, R. Rios, and R. Chau, "High performance fully-depleted Tri-gate CMOS transistors", *IEEE Electron Device Lett.*, vol. 24, no. 4, pp. 263–265, Apr. 2003.

[43] L. Yang et al., "25 nm CMOS Omega FETs", in *IEEE IEDM Technical Digest*, Dec. 2002, pp. 255–258.

[44] T. Park et al., "Fabrication of body-tied FinFETs (Omega MOSFETs) using bulk Si wafers", Jun. 2003, pp. 135–136.

[45] C.-H. Jan et al., "A 22nm SoC platform technology featuring 3-D tri-gate and high-k/metal gate, optimized for ultra low power, high performance and high density SoC applications", in *IEEE IEDM Technical Digest*, Dec. 2012, pp. 3.1.1–3.1.4.

[46] N. Singh et al., "High-performance fully depleted silicon nanowire (diameter $\leq$ 5 nm) gate-all-around CMOS devices", *IEEE Electron Device Lett.*, vol. 27, no. 5, pp. 383–386, May 2006.

[47] J. Gu, Y. Liu, Y. Wu, R. Colby, R. Gordon, and P. Ye, "First experimental demonstration of gate-all-around III-V MOSFETs by top-down approach", in *IEEE IEDM Technical Digest*, Dec. 2011, pp. 33.2.1–33.2.4.

[48] K. Tomioka, M. Yoshimura, and T. Fukui, "Vertical $In_{0.7}Ga_{0.3}As$ nanowire surrounding-gate transistors with high-k gate dielectric on Si substrate", in *IEEE IEDM Technical Digest*, Dec. 2011, pp. 33.3.1–33.3.4.

[49] 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 *IEEE IEDM Technical Digest*, 2005, pp. 617–620.

[50] Y. Zheng, C. Rivas, R. Lake, K. Alam, T. Boykin, and G. Klimeck, "Electronic properties of silicon nanowires", *IEEE Trans. on Electron Devices*, vol. 52, no. 6, pp. 1097–1103, Jun. 2005.

[51] K. Nehari, N. Cavassilas, J. Autran, M. Bescond, D. Munteanu, and M. Lannoo, "Influence of band structure on electron ballistic transport in silicon nanowire mosfets: an atomistic study", *Solid State Electronics*, vol. 50, no. 4, pp. 716 –721, 2006.

[52] M. Luisier, A. Schenk, W. Fichtner, and G. Klimeck, "Atomistic simulation of nanowires in the sp3d5s* tight-binding formalism: From boundary conditions to strain calculations", vol. 74, p. 205 323, 20 Nov. 2006.

[53] M. Luisier and G. Klimeck, "OMEN an Atomistic and Full-Band Quantum Transport Simulator for post-CMOS Nanodevices", in *IEEE Conference on Nanotechnology*, Aug. 2008, pp. 354–357.

[54] A. Martinez, A. Brown, A. Asenov, and N. Seoane, "A Comparison between a Fully-3D Real-Space Versus Coupled Mode-Space NEGF in the Study of Variability in Gate-All-Around Si Nanowire MOSFET", in *Proc.SISPAD*, 2009, pp. 1–4.

[55] M. Luisier, "Phonon-limited and effective low-field mobility in n- and p-type [100]-, [110]- and [111]-oriented si nanowire transistors", *Applied Physics Letters*, vol. 98, no. 3, 032111, 2011.

[56] S. Kim, M. Luisier, A. Paul, T. Boykin, and G. Klimeck, "Full Three-Dimensional Quantum Transport Simulation of Atomistic Interface Roughness in Silicon Nanowire FETs", *IEEE Trans. on Electron Devices*, vol. 58, no. 5, pp. 1371–1380, May 2011.

Chapter 2

Performance Benchmarking and Effective Channel Length for Nanoscale Planar Transistors

THANKS to the high electron velocities, III-V semiconductors have the potential to meet the challenging ITRS requirements for high performance for sub-22nm technology nodes and at a supply voltage approaching 0.5V. We present a comparative simulation study of ultra-thin-body (UTB) InAs, $In_{1-x}Ga_xAs$ and strained Si (sSi) MOSFETs, by using a comprehensive semi-classical Multi-Subband Monte Carlo (MSMC) transport model. Our results reported in [1] show that due to the finite screening length in the source-drain regions, III-V and Si nanoscale MOSFETs with a given gate length (L_G) may have a quite different effective channel length (L_{eff}). Moreover, the difference in L_{eff} provides an useful insight to interpret the performance comparison of III-V and Si MOSFETs, and, the engineering of the source-drain regions has a remarkable influence on the overall performance of nanoscale III-V MOSFETs.

2.1 Introduction

The competitive edge of III-V compared to Si MOSFETs for low V_{DD} logic circuits is being actively investigated [2–6]. The most critical questions concern the electrostatic integrity (e.g. drain-induced barrier-lowering (DIBL) and sub-threshold swing (SS)), the dynamic performance (e.g. on-current (I_{on}) and switching time (T_{sw})) and the energy efficiency (e.g. switching energy (E_{sw})) of $In_{0.53}Ga_{0.47}As$, InAs and Si transistors for a given gate length (L_G).

Because of the huge costs and the technical difficulties related to the fabrication of III-V and Si transistors with identical geometrical features, physically based numerical simulations are an appealing means to compare III-V and Si nanoscale MOSFETs, and several computational studies have already appeared based on simple

top-of-the-barrier or semi-analytical models [7, 8], on 3D semi-classical Monte Carlo [9, 10], or on full quantum but mainly ballistic transport models [5, 6].

Recently, encouraging experimental mobility data for n-type InGaAs MOSFETs have been reported [11–13], which makes it possible to revisit the assessment of III-V n-MOSFETs performance with a transport model comprising the most relevant scattering mechanisms and calibrated against mobility experiments.

We will investigate the performance of $In_{0.53}Ga_{0.47}As$ and InAs compared to Si and sSi nanoscale MOSFETs by using a comprehensive semi-classical multi-subband Monte Carlo simulator with a parallel implementation [14, 15]. Our simulation approach, is inherently 2D in real space (as well as in **k**-space) and we simulate a cutline of the device in the channel length direction. In this sense our analysis is appropriate for relatively wide devices (wider than about 100nm), and in fact the current is always quoted per unit device width. The MSMC tool, employs an effective mass approximation model (EMA) with non parabolic effects both in the quantization and in the transport directions, includes Γ, X and Λ valleys for III-V, only Δ valleys for Si, and accounts for acoustic and optical phonons, polar phonons of the channel material (for III-V), remote phonons originating in high-κ oxide stack [16], alloy scattering [17] (for $In_{0.53}Ga_{0.47}As$) and surface roughness scattering [18]. For Si and sSi the validation and calibration of our transport model was discussed in [19], whereas the calibration of the scattering parameters for InAs and $In_{0.53}Ga_{0.47}As$ is reported in this work by comparing to recent mobility experiments [11, 12].

Since III-V technologies are not at the manufacturing stage, available data on series resistance are scarce and scattered [6]. A comparison with silicon transistors accounting for the effects of series resistance is thus problematic and possibly unfair. Also because of this, the simulations in this work do not consider series resistance, and thus our results should be regarded as an upper limit for device performance. Moreover, since this work aims at assessing the performance limits of III-V compared to Si n-MOSFETs, our simulations, similarly to several previous studies [5, 6, 20], do not account for the effects of interface traps; however we will further discuss this point at the end of Sec.2.3.

This work presents a systematic study of SS, DIBL, I_{on}, T_{sw} and E_{sw} for n-type double-gate (DG), ultra-thin body (UTB) MOSFETs with L_G=14nm and employing Si, sSi, $In_{0.53}Ga_{0.47}As$ and InAs as channel materials. Different source-drain doping concentrations (N_{SD}) and gate alignments to source-drain metallurgical junctions are also included in the analysis.

2.2 Description of the simulation tool for planar MOSFETs

Our simulations are based on a Multi-subband Monte Carlo (MSMC) approach [14], that allows to solve with no *a priori* approximations the Boltzmann transport equation for the 2D electron gas in both long and nanoscale MOSFETs.

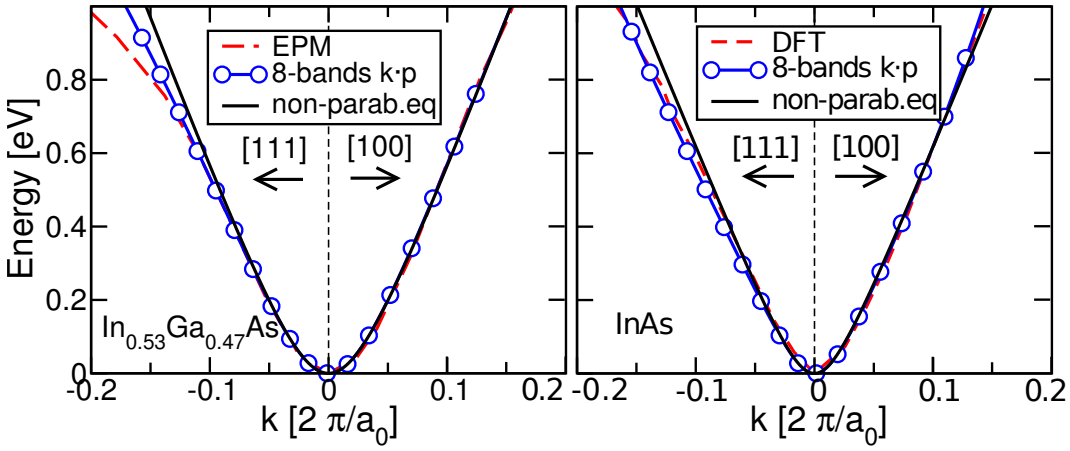


Figure 2.1: Band structure for $In_{0.53}Ga_{0.47}As$ (left) and InAs (right) from the calibrated non-parabolic analytical model (Eq.C.31), **k·p** (open circles) [21], EPM [22] and DFT [23] calculations (dashed lines). a_0 is the lattice constant. The CB minima have been taken as the zero for the energy.

2.2.1 Nonparabolic band structure model

The conduction band is described within the effective mass approximation including non-parabolic corrections; for Si only the Δ valleys are considered, whereas for III-V the Γ, X and Λ valleys are included in the simulations. Given the large non-parabolicity in the conduction band of InAs and $In_{0.53}Ga_{0.47}As$ close to the Γ minimum (see Fig.2.1), in each section of the device normal to the transport direction x, we expressed the energy $E(\mathbf{k})$ versus the wave-vector $\mathbf{k}= (k_x, k_y)$ in the transport plane as in Eq.C.31. Appendix C.2 reports some useful considerations regarding the non-parabolic model adopted in this work.

For unstrained InAs and $In_{0.53}Ga_{0.47}As$ the Γ minimum is approximately isotropic, hence in Eq.C.31 we set m_x=m_y=m_c, and then calibrated m_c and α by comparison with $E(\mathbf{k})$ of the bulk materials calculated with the 8-band **k·p** model of [21], with the **k·p** parameters from [24]. The calibrated $E(\mathbf{k})$ for InAs and $In_{0.53}Ga_{0.47}As$ are illustrated in Fig.2.1, showing that the non parabolic model can track very well the **k·p** results and also the DFT calculations from [22, 23] up to 0.7eV, which is sufficient for the purposes of this work. The calibrated effective masses and α values of the Γ valleys, are reported in Tab.2.1, and confirm the strong non-parabolicity.

The effective masses for Si and sSi were extracted by using a 30 bands **k·p** solver [27]. In all the simulations the parabolic effective mass Schrödinger-like equation was solved both in the semiconductor and in the oxide regions, with the appropriate continuity conditions for the wave-function $\psi_{n,x}(z)$ at the semiconductor-oxide interface [28]. This clarification is relevant for the formulation of surface roughness scattering, as further discussed in Sec.2.2.2. Then, non-parabolicity corrections for the subband minima and the in-plane energy relation are introduced according to Eq.C.31.

Table 2.1: Effective masses and non-parabolicity coefficients α used in this work for $In_{0.53}Ga_{0.47}As$, InAs and silicon in the Ellipsoid Coordinate System (ECS). Effective masses and non parabolicity coefficients for InAs satellite valleys are taken from [24] and [25], respectively, whereas for $In_{0.53}Ga_{0.47}As$ they are taken from [26].

Material	**Valley**	**eff.mass [m_0]**		α [eV^{-1}]
		m_l	m_t	
InAs	Γ	0.026		2.5
	X	1.13	0.16	0.9
	Λ	0.64	0.05	0.45
$In_{0.53}Ga_{0.47}As$	Γ	0.043		1.5
	X	2.26	0.25	0.5
	Λ	1.57	0.23	0.5
silicon	Δ	0.92	0.19	0.5

2.2.2 Scattering mechanisms and calibration to experiments

Our simulations account for a comprehensive set of scattering mechanisms, including elastic intra-valley and inelastic inter-valley phonons, remote phonons from the high-κ dielectric (with the scattering rate formulation discussed in [28]), local polar phonons (for InAs and $In_{0.53}Ga_{0.47}As$), surface roughness (SR) scattering and alloy scattering (for $In_{0.53}Ga_{0.47}As$) [17]. The effect of carrier screening for both alloy and surface roughness scattering was introduced by using a tensorial dielectric function [17, 18].

Our model does not account for the coupling between the polar phonons in III-V semiconductors and in the high-κ dielectric [26], and for the possible coupling between polar phonons and plasma oscillations in the inversion layer (i.e. the plasmon to phonon coupling) [16, 26]. However, our results show that, in the 5nm $In_{0.53}Ga_{0.47}As$ film used for the nanoscale MOSFETs, the SR scattering is by far the dominant scattering mechanism (see Fig.2.2 and the discussion below), so that further refinements in the modeling of polar and remote polar phonons are expected to have a minimal impact on the final results.

The modeling of SR scattering in ultra-thin III-V transistors is quite critical and challenging. In fact the widely used formulation leading to matrix elements proportional to the wave-function derivative at oxide interfaces has been derived for a closed boundary quantization model [28–32], while in UTB, high-κ III-V MOSFETs the wave-function penetration in the dielectrics is significant and affects the SR scattering rates [33]. In this work, as a semi-empirical approach to account for the wave-function penetration in the SR scattering rates, we maintained the matrix element expression proportional to wave-function derivative at the oxide interfaces, evaluating such expression by using the wave-functions obtained solving

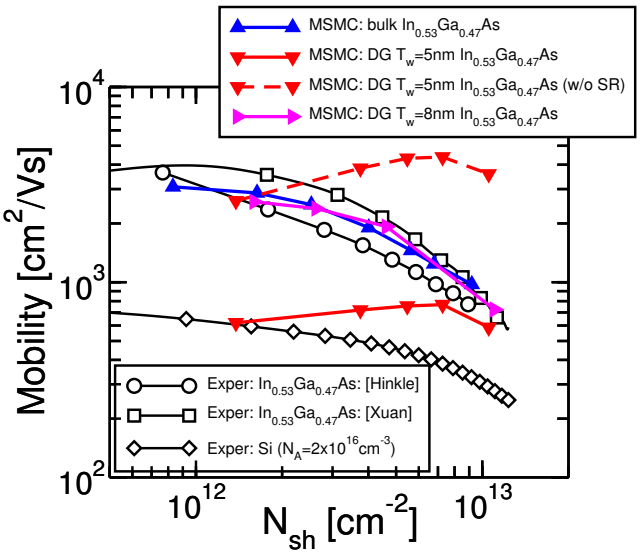


Figure 2.2: Measured and simulated mobility including the wave-function penetration into the oxide versus sheet carrier density (N_{sh}). Experiments for bulk-like $In_{0.53}Ga_{0.47}As$ MOSFETs are from [11, 12]. The figure also reports the bulk Si mobility. MSMC simulations for bulk $In_{0.53}Ga_{0.47}As$ MOSFETs with HfO_2 dielectric account for elastic intra-valley phonons, polar and remote polar phonons from the high-κ dielectric, alloy and SR scattering. The figure also shows simulated mobility for DG-UTB MOSFETs with T_w=8nm and T_w=5nm, as well as the mobility obtained for T_w=5nm by switching off the SR scattering. The SR power spectrum is exponential with a r.m.s. roughness value Δ_{rms}=1.3nm and a correlation length Λ=1.5nm.

the Schrödinger-like equation in the semiconductor and in the oxide region. Simulated mobility for $In_{0.53}Ga_{0.47}As$ as a function of the sheet carrier density is reported in Fig.2.2, where the parameters for SR scattering have been calibrated against experimental mobility data for bulk-like $In_{0.53}Ga_{0.47}As$ MOSFETs [11, 12]. The choice for the reference experiments is also justified by the attempt carried out in [11, 12] to avoid the systematic errors in the mobility measurement at high N_{sh} by including a correction for the interface traps contribution, produced by the large interface state density inside the conduction band of $In_{0.53}Ga_{0.47}As$ [13]. As can be seen the agreement between simulations and experiments for bulk-like $In_{0.53}Ga_{0.47}As$ MOSFETs is fairly good, and, furthermore, Fig.2.2 also reports the simulated mobility for DG-UTB $In_{0.53}Ga_{0.47}As$ MOSFETs. In this latter respect, the mobility for a well thickness T_w=5nm is in the range of 1000cm^2/Vs, which is fairly consistent with the best mobility values recently reported in UTB and HS-QW III-V MOSFETs [34, 35]. We verified that if, instead, the SR scattering is calculated by using wave-functions obtained neglecting the penetration in the oxides, and the SR parameters are calibrated against the same experiments for bulk-like $In_{0.53}Ga_{0.47}As$ transistors

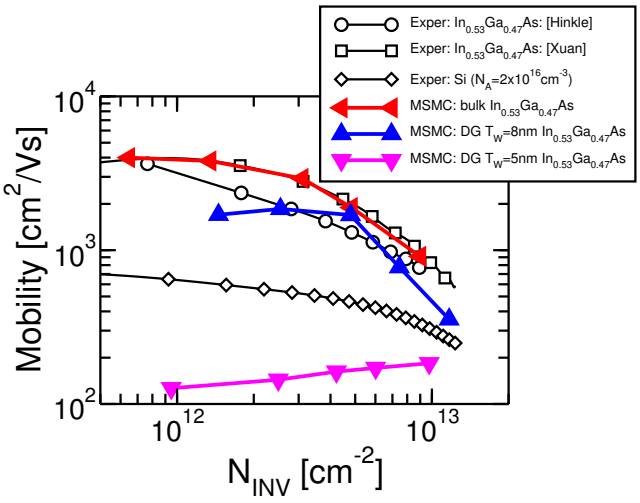


Figure 2.3: Measured and simulated mobility versus inversion density N_{INV}. Experiments for bulk-like $In_{0.53}Ga_{0.47}As$ MOSFETs [11, 12]. MSMC simulations for bulk and UTB $In_{0.53}Ga_{0.47}As$ MOSFETs with HfO_2 dielectric include elastic intra-valley phonons, polar and remote polar phonons from the high-κ dielectric [16, 28], alloy scattering [17], and SR with the Prange-Nee model [18] ($\Delta_{r.m.s.}$=0.8nm, Λ =1.5nm) without accounting for the wave-function penetration into the oxide in the Schrödinger equation. Diamonds: experimental mobility for Si bulk MOSFETs.

[11, 12], then the simulated mobility for T_w=5nm drops to about 200cm^2/Vs as shown in Fig.2.3. Such a result is consistent with simulations in [32], but not with the experiments [34, 35], which underlines the critical role played by SR scattering in ultra-thin InGaAs devices. This is further confirmed comparing in Fig.2.2 the mobility for T_w=5nm calculated with or without including the SR scattering; the SR is clearly the dominant scattering mechanism in such ultra-thin InGaAs films.

2.3 Benchmarking of $In_{0.53}Ga_{0.47}As$ InAs and Si nanoscale MOSFETs

2.3.1 Device structures and figures of merit

The DG-UTB transistor of this work has a semiconductor thickness T_w=5nm, the gate length is L_G=14nm and the gate dielectric is HfO_2 (κ =22 [28]), with an equivalent oxide thickness of 0.7nm.

The simulated (001)/[100] III-V devices were assumed to be unstrained, while the (110)/[1$\bar{1}$0] Si transistors (the crystallographic orientation is representative of the sidewall interfaces in FinFETs fabricated in (001) wafers and with a Manhattan layout [36]), were simulated for a relaxed channel material, for a 1GPa and a 2GPa tensile stress in the [1$\bar{1}$0] source to drain direction. The different stress conditions

are representative of different possible scenarios concerning the technological ability to induce stress in L_G=14nm transistors.

Our simulations include a volume gate electrode with a gate height T_{gate}=4nm, which is a relevant point to be mentioned for the capacitive coupling of the gate to the source/drain regions and also for the effective length defined below in Sec.2.3.2 Devices with different gate overlap or underlap configurations have been studied. The source-drain doping profiles have been taken as simply abrupt.

Our performance comparisons target a low supply voltage V_{DD}=0.5V and the gate work-function of Si and III-V transistors was adjusted to have the same I_{off}=100nA/μm, corresponding to the ITRS specification for high performance and III-V/Ge high-performance logic circuits. The on-current (I_{on}) is defined as the I_{DS} for V_{GS}=V_{DS}=0.5V. We also calculated a first order estimate of T_{sw} as [37]:

$$T_{sw} = \frac{Q_{on} - Q_{off}}{I_{on}} = \frac{\Delta Q_{sw}}{I_{on}}, \tag{2.1}$$

where Q_{on} and Q_{off} are defined as the charge in the device respectively for V_{GS}=V_{DD}, V_{DS}=0, and for V_{GS}=0,V_{DS}=V_{DD}; Q_{on} and Q_{off} were determined by integrating numerically the charge density in the device. More refined estimates of T_{sw} can be obtained by replacing the I_{on} in Eq.2.1 with an effective drive current [38, 39], however Eq.2.1 is reasonably adequate for the purpose of comparing MOSFET devices with different semiconductor materials. A similarly simple metric for the switching energy E_{sw} can be expressed as:

$$E_{sw} = V_{DD}(Q_{on} - Q_{off}) \ . \tag{2.2}$$

Fig.2.4 shows the I_{DS}-V_{GS} characteristics for a sSi, InAs and $In_{0.53}Ga_{0.47}As$ transistor, with the gate aligned to the source/drain regions (i.e. neither overlap nor underlap). The corresponding SS and DIBL values are reported in Tab.2.2. As can be seen, the I_{on} for the unstrained InAs device is slightly larger than for sSi with 2GPa stress and significantly larger than the 1GPa, or the unstrained, Si devices. Furthermore, even the electrostatic integrity of the InAs transistor (i.e. SS and DIBL) is somewhat better (see Tab.2.2).

We will return to a more systematic comparison of all figures of merit in Sec.2.3.3 We here notice, however, that the better SS and DIBL values for InAs MOSFETs suggest to analyze in more detail the effective channel length (L_{eff}) of III-V and Si MOSFETs having the same L_G, T_w and T_{ox}.

It should be noticed that the simulations in this work do not account for band-to-band tunneling (BTBT), which may set the lower limit for I_{off} and prevent to achieve I_{off}=100 nA/μm.

As far as $In_{0.53}Ga_{0.47}As$ UTB devices are concerned, there are several experimental data [40] and simulation results [5] showing that well-designed, ultra-thin body transistors or FinFETs can reach I_{off} values well below 100 nA/μm. Such references also indicate that leakage mechanisms do not significantly distort the I_{DS} vs. V_{GS}

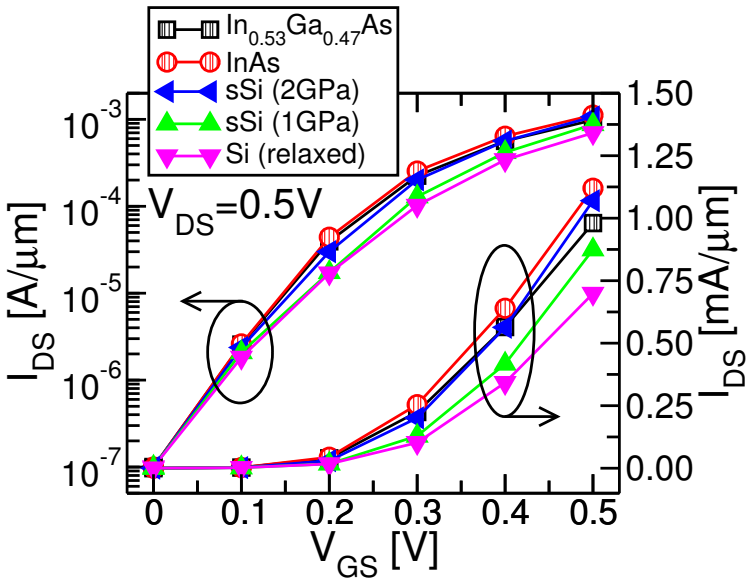


Figure 2.4: Simulated drain current (I_{DS}) versus gate voltage (V_{GS}) characteristics for (001)/[100] InAs (open circles), $In_{0.53}Ga_{0.47}As$ (open squares) DG devices. Figure also reports results for a (110)/[1$\bar{1}$0] 2GPa, 1GPa and relaxed uniaxial tensile strained Si (closed symbols) DG device. The gate is aligned to the source/drain regions. Source and drain doping is $N_{SD}= 5 \cdot 10^{19}$ cm^{-3}, L_G=14nm and T_w=5nm. SS and DIBL values are reported in Tab.2.2

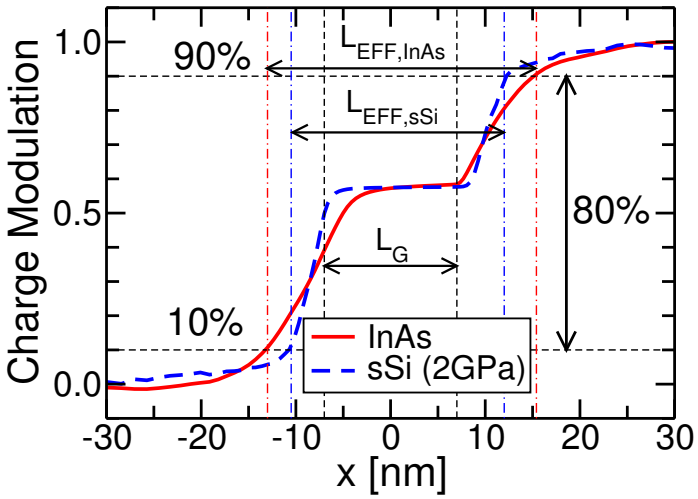


Figure 2.5: Charge density modulation for a V_{GS} change from 0V to 0.1V and V_{DS}=0.5V along the channel for an InAs (solid line) and 2GPa sSi (dashed line) devices with $N_{SD}= 5 \cdot 10^{19}$. The gate is perfectly aligned with source/drain regions.

characteristic down to $I_{DS}\approx$ 100 nA/μm and, in particular, do not degrade the sub-threshold slope [5, 40]. We believe that this is an evidence that BTBT does not play a critical role in well designed, high performance $In_{0.53}Ga_{0.47}As$ nMOSFETs and demonstrate that I_{off}=100 nA/μm is feasible for V_{DD}=0.7V or below in UTB devices. BTBT is admittedly expected to be more critical for InAs devices because of the low InAs energy band-gap, which however is significantly enlarged by quantization in a T_w=5nm transistor. As a matter of fact, an I_{off}=100 nA/μm has been experimentally demonstrated for V_{DD}=0.5V and T_w=10nm [41], which makes it reasonable to assume that BTBT may not play a critical role in a T_w=5nm transistor at V_{DD}=0.5V and for I_{DS} larger than 100 nA/μm.

2.3.2 Definition of an effective channel length

Among many possible definitions [42, 43], in this work we defined L_{eff} as the length of the device region where the gate terminal is able to induce a significant modulation of the charge density. A quantitative extraction of L_{eff} was obtained by numerically integrating along the transport direction the variation in the charge density produced by a prescribed V_{GS} change. In this respect, Fig.2.5 shows the cumulative integral of the charge variation for a V_{GS} change from 0V to 0.1V, that is the integral from the left border of the simulation domain to a generic x value and normalized to the integral over the entire domain.

As illustrated in Fig.2.5, we calculated L_{eff} as the length over which the cumulative integral of charge density variation goes from 10% to 90%, so that Fig.2.5 reveals a larger L_{eff} for InAs than for sSi transistor. As can be seen, in the InAs MOSFETs the gate bias can modulate the charge density deeper into the source/drain regions

compared to the sSi transistor. This is consistent with a longer screening length in InAs, which in turn is due to an higher dielectric constant, a smaller density of states and a stronger carrier degeneracy compared to Si [44]. To capture these effects in a Monte Carlo simulation, it is mandatory to implement the Pauli′s exclusion principle by rejecting scattering events according to the occupation in the k-plane [14, 45] and by careful injection of particles at the contacts according to a Fermi-Dirac distribution [46]. Furthermore, the Multi-subband approach provides the correct 2D density of states typical of thin quantum wells.

As implicit in its definition, the L_{eff} is V_{GS} dependent [43]. Since we are mainly interested to interpret the performance comparison of III-V and Si MOSFETs in terms of SS and DIBL, in this work we employed an off-state effective channel length ($L_{eff}^{(off)}$) defined as the one extracted by using a V_{GS} variation from 0V to 0.1V.

2.3.3 Analysis of the transistor performance

Table 2.2: Summary of the performance metrics for Si, $In_{0.53}Ga_{0.47}As$ and InAs in DG-UTB MOSFETs. The power supply voltage is set to 0.5V and the off-state current I_{OFF} is set to 0.1 $\mu A/\mu m$. SS is calculated in the range [0.1$\mu A/\mu m$,5$\mu A/\mu m$] and the DIBL is calculated by varying V_{DS} between 0.05V and 0.5V and for a V_T defined as the V_{GS} yielding I_{DS}=1$\mu A/\mu m$. V_x@VS is the carrier velocity at the virtual source (VS) and N_{SH} @VS is the sheet carrier density at the VS. The remaining symbols are defined in the text.

Channel material	Stress/ relaxed	Gate alignment	N_{SD} [10^{19} cm^{-3}]	SS [$\frac{mV}{dec}$]	DIBL [$\frac{mV}{V}$]	I_{on} [$\frac{mA}{\mu m}$]	V_x@VS [10^7 cm/s]	N_{sh}@VS [10^{12} cm^{-2}]	T_{sw} [fs]	E_{sw} [$\frac{fJ}{\mu m}$]	ΔQ_{sw} [$\frac{fC}{\mu m}$]
Si	2 GPa	aligned	10	83	126	0.98	0.96	6.47	417	0.205	0.409
			5	80	101	1.07	1.03	6.54	362	0.195	0.389
	1 GPa	aligned	10	82	128	0.87	0.81	6.74	473	0.207	0.413
			5	77	102	0.87	0.80	6.82	445	0.195	0.389
		underlap	5	74	76	0.89	0.71	7.91	404	0.185	0.361
	relaxed	aligned	5	78	98	0.69	0.54	8.14	564	0.197	0.393
		underlap	5	76	76	0.71	0.56	7.94	519	0.184	0.368
$In_{0.53}Ga_{0.47}As$	relaxed	aligned	10	80	120	0.98	2.97	2.03	369	0.180	0.360
			5	77	93	0.98	2.69	2.28	327	0.160	0.320
			2	70	68	0.94	1.92	3.05	306	0.144	0.288
InAs	relaxed	overlap	10	85	152	0.98	3.65	1.67	400	0.194	0.388
			5	76	113	1.03	3.66	1.91	320	0.165	0.331
			2	70	77	1.10	2.77	2.49	264	0.146	0.292
		aligned	10	80	125	1.08	3.71	1.88	327	0.177	0.353
			5	73	90	1.11	3.53	1.98	282	0.158	0.316
			2	67	66	1.09	2.65	2.57	255	0.139	0.278
		underlap	10	76	93	1.16	3.71	1.96	278	0.162	0.323
			5	71	75	1.17	3.30	2.21	255	0.149	0.297
			2	66	59	1.10	2.57	2.67	244	0.134	0.269

Many different Si and III-V UTB-DG transistors were included in our performance benchmarking by varying both the source-drain doping (N_{SD}) and the gate alignment. More precisely, we simulated devices with a perfect gate alignment as well as with a 2nm underlap or overlap and, furthermore, we considered N_{SD}=2·10^{19}, 5·10^{19} and 10^{20}cm^{-3}. For Si devices the lowest N_{SD} value was not considered because we verified that this leads to significant resistive voltage drops in source/drain regions (not shown), which are instead practically negligible in all the other cases.

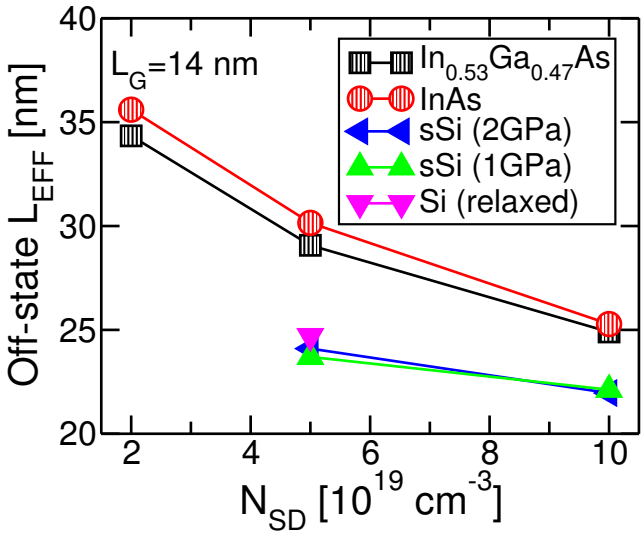


Figure 2.6: Off-state effective channel length ($L_{eff}^{(off)}$) versus source and drain doping concentration (N_{SD}) for III-V and Si MOSFETs.

Fig.2.6 reports the $L_{eff}^{(off)}$ for different N_{SD} considering a gate aligned configuration. As can be seen III-V transistors have L_{eff} values systematically larger than sSi MOSFETs.

Figs.2.7 and 2.8 report respectively SS and DIBL versus $L_{eff}^{(off)}$ for III-V and sSi MOSFETs; the L_G is 14nm for all the devices and $L_{eff}^{(off)}$ changes because of different N_{SD} and gate alignments. Fig.2.7 shows that the SS is similar for III-V and Si devices at a given $L_{eff}^{(off)}$, so that at fixed L_G III-V transistors tend to have a somewhat better SS compared to Si MOSFETs mainly because they feature a larger $L_{eff}^{(off)}$ (see Fig.2.6). As for DIBL, Fig.2.8 suggests that Si has a slight advantage compared to $In_{0.53}Ga_{0.47}As$ and InAs at a given $L_{eff}^{(off)}$, because of the larger dielectric constant of $In_{0.53}Ga_{0.47}As$ and InAs [47]. However Si MOSFETs tend to lose their potential advantage at a given L_G because their $L_{eff}^{(off)}$ is smaller.

The results in Figs.2.7 and 2.8 suggest that the comparison of electrostatic integrity (i.e. SS and DIBL) between III-V and sSi MOSFETs at a given L_G, which is understandably a crucial comparison, is a delicate exercise because, for a given L_G, the L_{eff} can be quite different.

Table 2.2 summarizes the figures of merit for many different design options with Si, sSi, $In_{0.53}Ga_{0.47}As$ and InAs. As can be seen the III-V transistors have an I_{on} essentially comparable to the sSi MOSFETs with the 2GPa stress, thus they have a clear I_{on} advantage over 1GPa sSi or relaxed Si.

At this regard Fig.2.10 compares the sheet carrier density (N_{sh}) and the carrier velocity (V_x) profiles along the channel for InAs and 2GPa sSi devices. The virtual source (VS) is identified as the position of maximum in the profile of the lowest

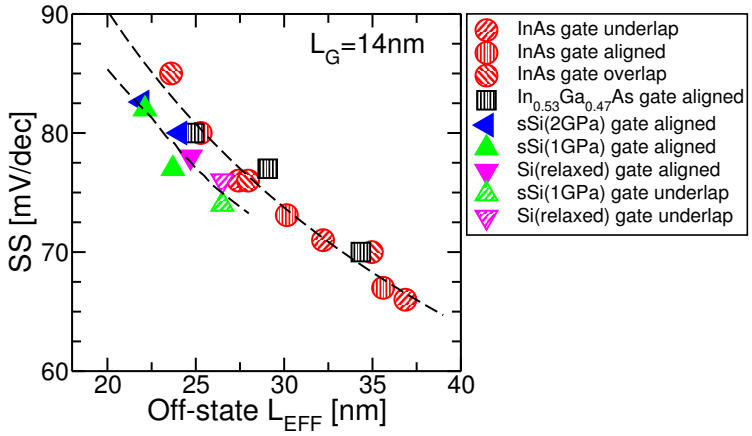


Figure 2.7: SS versus $L_{eff}^{(off)}$ for Si and III-V MOSFETs with different gate alignments and source/drain doping. SS is calculated in the range [0.1μA/μm,5μA/μm]. Dashed lines are guides for the eyes.

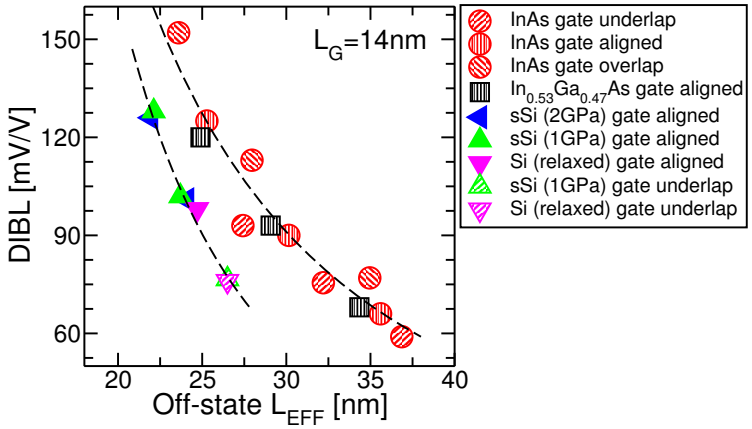


Figure 2.8: DIBL versus $L_{eff}^{(off)}$ for Si and III-V MOSFETs with different gate alignments and source/drain doping. The DIBL is calculated by varying V_{DS} between 0.05V and 0.5V and for a V_T defined as the V_{GS} yielding I_{DS}=1μA/μm. Dashed lines are guides for the eyes.

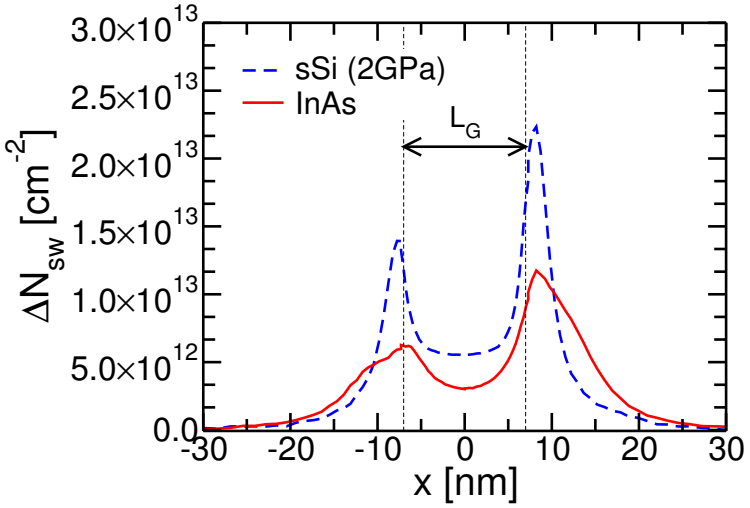


Figure 2.9: Charge density modulation $\Delta N_{sw}=\Delta Q_{sw}/e = (Q_{on} - Q_{off})/e$ obtained by switching between the off-state (i.e. V_{GS}=0 V,V_{DS}=V_{DD}) and the on-state (i.e. V_{GS}=V_{DD}, V_{DS}=0 V) for 2-GPa uniaxial sSi (dashed line) and InAs (solid line). V_{DS} = 0.5 V with N_{SD}= $5{\cdot}10^{19}$. The gate is perfectly aligned with source/drain regions.

subband and is located deeper into the channel for InAs than for Si due to a longer screening length [44]. Fig.2.10 confirms, consistently with [44], that the reduced N_{sh} at the VS in InAs transistors is due, to a large extent, to the fact that the VS is located well inside the channel region (actually close to the center of the channel), while for Si it is located close to the source junction. Consequently the smaller N_{sh} does not correspond to a similarly lower sheet density integrated throughout the channel.

As a result the III-V has only a modest advantage compared to Si in terms of switched charge ΔQ_{sw}= ($Q_{on} - Q_{off}$) (obtained by switching between the off-state (i.e. V_{GS}=0,V_{DS}=V_{DD}) and the on-state (i.e. V_{GS}=V_{DD}, V_{DS}=0)), even because the modulation of the charge in the source-drain regions accounts for an important contribution to the overall ΔQ_{sw} as shown in Fig.2.9 where the ΔN_{sw} term is given by $\Delta N_{sw}=\Delta Q_{sw}/e$. Since the difference in ΔQ_{sw} is not large between $In_{0.53}Ga_{0.47}As$, InAs and Si transistors (see Tab.2.2 that shows the integral of ΔQ_{sw} along the transport direction), the $In_{0.53}Ga_{0.47}As$ and InAs devices have a sizably better switching time T_{sw} only if they can significantly improve I_{on}, hence only when compared to relaxed Si or 1GPa sSi MOSFETs. Furthermore, modest differences in ΔQ_{sw} also imply modest differences in the switching energies E_{sw} for a fixed V_{DD} (see Tab.2.2).

We recall that, as already mentioned in Sec.2.1, our analysis does not account for interface traps. This simulation analysis aims at assessing the performance limits of $In_{0.53}Ga_{0.47}As$ compared to all Si CMOS transistors in the perspective that largest

material and technological challenges for III-V based transistors can be overcome. In this spirit our analysis assumes that the density of interface defects in $In_{0.53}Ga_{0.47}As$ devices can be reduced to such an extent that they do not significantly affect the device performance at least for the low V_{DD} and high I_{off} values targeted by this work. Incidentally, this is also the assumption made in most of the papers comparing performance of all Si and III-V CMOS devices on the basis of numerical simulations [5, 6, 20, 44].

As for the possible impact of interface states, we first recall that for a fully depleted UTB and in the presence of interface traps the sub-threshold slope can be written as: [48]

$$SS \approx \frac{2.3K_BT}{q}\left[1+\frac{C_{it}}{C_g}\right] \tag{2.3}$$

where C_g is the gate oxide capacitance and it is assumed that in the sub-threshold region the depletion and inversion capacitances are negligible compared to the capacitance C_{it} due to interface states. By recalling that our devices are double-gate transistors we can estimate $C_g \simeq 2\,(\epsilon_{SiO2}/\mathrm{EOT})$, where EOT=0.7nm is the effective oxide thickness, which leads to $C_g \simeq 9.87 \cdot 10^{-6}$ F/cm^2. As for C_{it} we can simply take $C_{it} \simeq eD_{it}$, with e being the elementary charge and D_{it} the trap density. By using reasonable D_{it} values in the energy gap of about $D_{it} = 6 \cdot 10^{12}$ eV^{-1} cm^{-2} [13, 49], we obtain $C_{it}/C_g \simeq 0.097$. This analysis suggests that MOS transistors with an aggressive gate oxide scaling are relatively robust against sub-threshold degradation due to interface defects. This has been recently pointed out even for nanowire InAs MOSFETs [50], and it is an encouraging observation for III-V based MOSFETs.

As for the on-state of the transistors, Tab.2.2 shows that at V_{DD}=0.5V the maximum N_{sh} at the virtual source in the channel for III-V UTB transistors does not exceeds $3{\cdot}10^{12}$ cm^{-2}. This sheet density is quite below the best values of $N_{sh} \simeq 6{\cdot}10^{12}$ cm^{-2} determined via Hall measurements even in $In_{1-x}Ga_xAs$ transistors with significant Fermi level pinning [13].

These considerations suggest that the operation of the UTB devices of this work may be quite robust against interface defects, at least for the low V_{DD} value considered in our study.

2.4 Conclusions

This work presented an investigation of the performance limits of III-V and sSi UTB-DG MOSFETs comparing several figures of merit for digital circuit operation in striving to understand if III-V MOSFETs are a cost-effective device solution to continue MOSFETs scaling beyond the 22-nm technology node. Our results obtained for intrinsic devices show that, III-V MOSFETs have a clear performance advantage compared to relaxed and 1GPa strained silicon devices, while Si MOSFETs with a 2GPa remain competitive with III-V MOSFETs. Quite interestingly, our main conclusions are qualitatively consistent with the results of previous studies based on the NEGF simulation approach [5, 6].

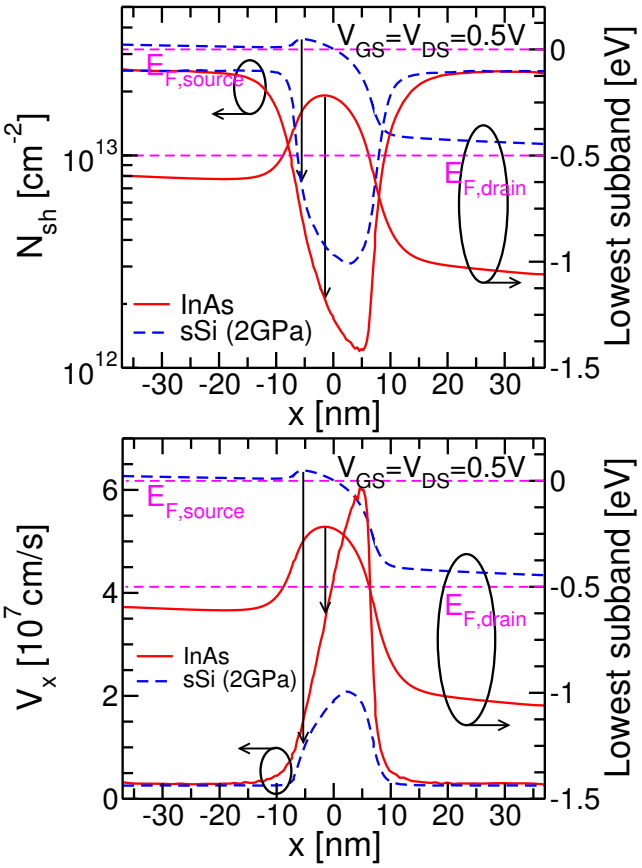


Figure 2.10: Sheet carrier density (N_{sh}) and velocity (V_x) profiles for InAs and 2GPa-sSi at V_{GS}=V_{DS}=0.5V. Source/drain (S/D) doping is $5{\cdot}10^{19}$ and the gate is perfectly aligned with S/D regions. The arrows indicate the position of the virtual source (VS). Despite InAs and 2GPa sSi have similar drive currents (see Tab.2.2), N_{sh} at the VS is significantly lower for InAs than for sSi whereas V_x at the VS is higher for InAs due to a lower transport effective mass (Tab.2.1).

These results, which are qualitatively consistent to those reported for different device architectures like FinFETs [6], suggest that the competitive edge of III-V MOSFETs depends quite critically on the one hand on the technological capabilities to induce a large stress magnitude in sub-22nm silicon transistors, and, on the other hand, on the further improvements in III-V MOSFETs possibly induced by an appropriate strain engineering.

In this latter respect, even if interesting and promising contributions have been recently reported [51, 52], experimental data are still quite sparse and not fully consistent. As a result, even the physical interpretation of the strain induced mobility variations in III-V MOSFETs is still debated [51, 52], so that the predictive capabilities of transport models to make performance projections about strained III-V MOSFETs are still limited, which is also the reason why we refrained from including strained III-V MOSFETs in our analysis.

We conclude by reiterating that the I_{on} and dynamic performance reported in this work are best case estimates obtained by neglecting the effect of series resistances. Since series resistances will certainly degrade the performance for both III-V and Si devices, the potential advantages of III-V transistors critically depend on the technological capabilities to realize for III-V materials series resistances complying with ITRS projections and in any case comparable to those of all silicon CMOS technologies.

References

[1] D. Lizzit, D. Esseni, P. Palestri, P. Osgnach, and L. Selmi, "Performance Benchmarking and Effective Channel Length for Nanoscale InAs, $In_{0.53}Ga_{0.47}As$, and sSi n-MOSFETs", *IEEE Trans. on Electron Devices*, vol. 61, no. 99, pp. 2027 –2034, Jun. 2014.

[2] M. Radosavljevic et al., "Non-planar, multi-gate InGaAs quantum well field effect transistors with high-K gate dielectric and ultra-scaled gate-to-drain/gate-to-source separation for low power logic applications", in *IEEE IEDM Technical Digest*, 2010, pp. 6.1.1–6.1.4.

[3] G. Dewey, B. Chu-Kung, R. Kotlyar, M. Metz, N. Mukherjee, and M. Radosavljevic, "III-V field effect transistors for future ultra-low power applications", in *IEEE Symposium on VLSI Technology - Technical Digest*, 2012, pp. 45–46.

[4] G. Doornbos and M. Passlack, "Benchmarking of III-V n-MOSFET Maturity and Feasibility for Future CMOS", *IEEE Electron Device Lett.*, vol. 31, no. 10, pp. 1110 –1112, Oct. 2010.

[5] M. Luisier, "Performance Comparison of GaSb, Strained-Si, and InGaAs Double-Gate Ultrathin-Body n-FETs", *IEEE Electron Device Lett.*, vol. 32, no. 12, pp. 1686–1688, 2011.

[6] S. H. Park, Y. Liu, N. Kharche, M. Jelodar, G. Klimeck, M. Lundstrom, and M. Luisier, "Performance Comparisons of III-V and Strained-Si in Planar FETs and Nonplanar FinFETs at Ultrashort Gate Length (12 nm)", *IEEE Trans. on Electron Devices*, vol. 59, no. 8, pp. 2107–2114, 2012.

[7] 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. on Electron Devices*, vol. 54, no. 1, pp. 115–123, 2007.

[8] C.-H. Yu, S. Wu, V.-H. Hu, and P. Su, "Impact of Quantum Confinement on Subthreshold Swing and Electrostatic Integrity of Ultra-Thin-Body GeOI and InGaAs-OI n-MOSFETs", *IEEE Trans. on Nanotechnology*, vol. 11, no. 2, pp. 287–291, 2012.

[9] K.Kalna, N.Seoane, A.J.Garcia-Loureiro, I.G.Thayne, and A.Asenov, "Benchmarking of Scaled InGaAs Implant-Free NanoMOSFETs", *IEEE Trans. on Electron Devices*, vol. 55, no. 9, pp. 2297–2306, 2008.

[10] E. A. Towie, C. Riddet, and A. Asenov, "Comparison of Raised Source/Drain Implant-Free Quantum-Well and Tri-Gate MOSFETs using 3D Monte Carlo Simulation", in *Proc.SISPAD*, 2013, pp. 384 –387.

[11] Y. Xuan, Y. Wu, T. Shen, T. Yang, and P. Ye, "High performance submicron inversion-type enhancement-mode InGaAs MOSFETs with ALD Al2O3, HfO2 and HfAlO as gate dielectrics", in *IEEE IEDM Technical Digest*, 2007, pp. 637–640.

[12] C. Hinkle, A. Sonnet, R. Chapman, and E. M. Vogel, "Extraction of the Effective Mobility of $In_{0.53}Ga_{0.47}As$ MOSFETs", *IEEE Electron Device Lett.*, vol. 30, no. 4, pp. 316–318, 2009.

[13] N. Taoka et al., "Impact of Fermi level pinning inside conduction band on electron mobility of In_xGa_{1-x}-As MOSFETs and mobility enhancement by pinning modulation", in *IEEE IEDM Technical Digest*, 2011, pp. 27.2.1–27.2.4.

[14] 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. on Electron Devices*, vol. 54, no. 5, pp. 1156 –1164, May 2007.

[15] P. Osgnach, A. Revelant, D. Lizzit, P. Palestri, D. Esseni, and L. Selmi, "Toward computationally efficient Multi-Subband Monte Carlo Simulations of Nanoscale MOSFETs", in *Proc.SISPAD*, 2013, pp. 176–179.

[16] P. Toniutti, P. Palestri, D. Esseni, F. Driussi, M. De Michielis, and L. Selmi, "On the origin of the mobility reduction in n- and pmetal- oxide-semiconductor field effect transistors with hafniumbased/metal gate stacks.", *Journal of Applied Physics*, vol. 112, p. 034 502, 2012.

[17] D. Lizzit, P. Palestri, D. Esseni, A. Revelant, and L. Selmi, "Analysis of the Performance of n-Type FinFETs With Strained SiGe Channel", *IEEE Trans. on Electron Devices*, vol. 60, no. 6, pp. 1884–1891, 2013.

[18] P. Toniutti, D. Esseni, and P. Palestri, "Failure of the Scalar Dielectric Function Approach for the Screening Modeling in Double-Gate SOI MOSFETs and in FinFETs", *IEEE Trans. on Electron Devices*, vol. 57, no. 11, pp. 3074–3083, 2010.

[19] F. Conzatti et al., "Investigation of Strain Engineering in FinFETs Comprising Experimental Analysis and Numerical Simulations", *IEEE Trans. on Electron Devices*, vol. 58, no. 6, pp. 1583 –1593, Jun. 2011.

[20] K. Cantley, Y. Liu, H. Pal, T. Low, S. Ahmed, and M. Lundstrom, "Performance Analysis of III-V Materials in a Double-Gate nano-MOSFET", in *IEEE IEDM Technical Digest*, Dec. 2007, pp. 113–116.

[21] T. B. Bahder, "Eight-band $\mathbf{k} \cdot \mathbf{p}$ model of strained zinc-blende crystals", *Phys. Rev. B*, vol. 41, pp. 11 992–12 001, 17 Jun. 1990.

[22] R. Dittrich and W. Schroeder, "Empirical pseudopotential band structure of $In_{0.53}Ga_{0.47}As$ and $In_{0.52}Al_{0.48}As$", *Solid State Electronics*, vol. 43, no. 2, pp. 403 –407, 1999.

[23] Y.-S. Kim, M. Marsman, G. Kresse, F. Tran, and P. Blaha, "Towards efficient band structure and effective mass calculations for III-V direct band-gap semiconductors", *Phys. Rev. B*, vol. 82, p. 205 212, 20 Nov. 2010.

[24] I. Vurgaftman, J. R. Meyer, and L. R. Ram-Mohan, "Band parameters for III–V compound semiconductors and their alloys", *Journal of Applied Physics*, vol. 89, no. 11, pp. 5815–5875, 2001.

[25] K. Kalna, S. Roy, A. Asenov, K. Elgaid, and I. Thayne, "Scaling of pseudomorphic high electron mobility transistors to decanano dimensions", *Solid State Electronics*, vol. 46, no. 5, pp. 631 –638, 2002.

[26] T. O'Regan, M. Fischetti, B. Soree, S. Jin, W. Magnus, and M. Meuris, "Calculation of the electron mobility in III-V inversion layers with high-κ dielectrics.", *Journal of Applied Physics*, vol. 108, p. 103 705, 2010.

[27] D. Rideau, M. Feraille, L. Ciampolini, M. Minondo, C. Tavernier, H. Jaouen, and A. Ghetti, "Strained Si, Ge, and $Si_{1-x}Ge_x$ alloys modeled with a first-principles-optimized full-zone k.p method", *Phys. Rev. B*, vol. 74, p. 195 208, 19 Nov. 2006.

[28] D. Esseni, P. Palestri, and L. Selmi, *Nanoscale MOS Transistors - Semi-Classical Transport and Applications*, 1st. Cambridge University Press., 2011.

[29] C. Jungemann, A. Edmunds, and W.L. Engl, "Simulation of Linear and Nonlinear Electron Transport in Homogeneous Silicon Inversion Layers", *Solid State Electronics*, vol. 36, no. 11, pp. 1529–1540, 1993.

[30] M.V. Fischetti and S.E. Laux, "Monte Carlo study of electron transport in silicon inversion layers", *Phys. Rev. B*, vol. 48, pp. 2244–2274, 1993.

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

[32] M.Poljak, V.Jovanović, D.Grgec, and T.Suligoj, "Assessment of Electron Mobility in Ultrathin-Body InGaAs-on-Insulator MOSFETs Using Physics-Based Modeling", *Electron Devices, IEEE Transactions on*, vol. 59, no. 6, pp. 1636–1643, 2012.

[33] D.Lizzit, D.Esseni, P.Palestri, and L.Selmi, "Surface roughness limited mobility modeling in ultra-thin SOI and quantum well III-V MOSFETs", in *IEEE IEDM Technical Digest*, 2013, pp. 5.2.1–5.2.4.

[34] T.-W. Kim et al., "ETB-QW InAs MOSFET with scaled body for improved electrostatics", in *IEEE IEDM Technical Digest*, 2012, pp. 32.3.1–32.3.4.

[35] K. Takei et al., "Quantum Confinement Effects in Nanoscale-Thickness InAs Membranes", *Nano Letters*, vol. 11, no. 11, pp. 5008–5012, 2011.

[36] N. Serra and D. Esseni, "Mobility Enhancement in Strained n -FinFETs: Basic Insight and Stress Engineering", *IEEE Trans. on Electron Devices*, vol. 57, no. 2, pp. 482–490, 2010.

[37] F. Conzatti, M. Pala, D. Esseni, E. Bano, and L. Selmi, "Strain-Induced Performance Improvements in InAs Nanowire Tunnel FETs", *IEEE Trans. on Electron Devices*, vol. 59, no. 8, pp. 2085 –2092, 2012.

[38] M.-H. Na, E. Nowak, W. Haensch, and J. Cai, "The effective drive current in CMOS inverters", in *IEEE IEDM Technical Digest*, 2002, pp. 121–124.

[39] J. Deng and H.-S. Wong, "Metrics for performance benchmarking of nanoscale Si and carbon nanotube FETs including device nonidealities", *IEEE Trans. on Electron Devices*, vol. 53, no. 6, pp. 1317–1322, 2006.

[40] A. Alian, M. Pourghaderi, Y. Mols, M. Cantoro, T. Ivanov, N. Collaert, and A. Thean, "Impact of the channel thickness on the performance of ultrathin InGaAs channel MOSFET devices", in *IEEE IEDM Technical Digest*, Dec. 2013, pp. 16.6.1–16.6.4.

[41] S. Chang et al., "InAs N-MOSFETs with record performance of Ion=600uA/um at Ioff=100nA/um (Vd =0.5 V)", in *IEEE IEDM Technical Digest*, Dec. 2013, pp. 16.1.1–16.1.4.

[42] Y. Taur, "MOSFET channel length: extraction and interpretation", *IEEE Trans. on Electron Devices*, vol. 47, no. 1, pp. 160–170, 2000.

[43] L. Selmi, R. Menozzi, P. Gandolfi, and B. Riccó, "Numerical analysis of the gate voltage dependence of the series resistances and effective channel length in submicrometer GaAs MESFETs", *IEEE Trans. on Electron Devices*, vol. 39, no. 9, pp. 2015–2020, 1992.

[44] Y. Liu, M. Luisier, A. Majumdar, D. Antoniadis, and M. Lundstrom, "On the interpretation of ballistic injection velocity in deeply scaled mosfets", *IEEE Trans. on Electron Devices*, vol. 59, no. 4, pp. 994–1001, 2012.

[45] P. Lugli and D. Ferry, "Degeneracy in the ensemble Monte Carlo method for high-field transport in semiconductors", *IEEE Trans. on Electron Devices*, vol. 32, no. 11, pp. 2431–2437, 1985.

[46] P. Palestri, L. Lucci, S. D. Tos, D. Esseni, and L. Selmi, "An improved empirical approach to introduce quantization effects in the transport direction in multi-subband Monte Carlo simulations", *Semiconductor Science Technology*, vol. 25, no. 5, p. 055 011, 2010.

[47] T. Dutta, Q. Rafhay, R. Clerc, J. Lacord, S. Monfray, G. Pananakakis, F. Boeuf, and G. Ghibaudo, "Origins of the short channel effects increase in III-V nMOSFET technologies", in *International Conference on Ultimate Integration on Silicon*, 2012, pp. 25–28.

[48] Z. Lun, D. Ang, and C. Ling, "A novel subthreshold slope technique for the extraction of the buried-oxide interface trap density in the fully depleted SOI MOSFET", *IEEE Electron Device Lett.*, vol. 21, no. 8, pp. 411–413, Aug. 2000.

[49] S. Kim, M. Yokoyama, R. Nakane, O. Ichikawa, T. Osada, M. Hata, M. Takenaka, and S. Takagi, "High performance sub-20-nm-channel-length extremely-thin body InAs-on-insulator tri-gate MOSFETs with high short channel effect immunity and Vth tunability", in *IEEE IEDM Technical Digest*, Dec. 2013, pp. 16.4.1–16.4.4.

[50] D. Esseni and M. Pala, "Interface Traps in InAs Nanowire Tunnel FETs and MOSFETs-Part II: Comparative Analysis and Trap-Induced Variability", *IEEE Trans. on Electron Devices*, vol. 60, no. 9, pp. 2802–2807, Sep. 2013.

[51] S. Kim, M. Yokoyama, N. Taoka, R. Nakane, T. Yasuda, O. Ichikawa, N. Fukuhara, M. Hata, M. Takenaka, and S. Takagi, "Strained $In_{0.53}Ga_{0.47}As$ metal-oxide-semiconductor field-effect transistors with epitaxial based biaxial strain", *Applied Physics Letters*, vol. 100, no. 19, 193510, p. 193 510, 2012.

[52] L. Xia and J. del Alamo, "Mobility enhancement in indium-rich N-channel $In_xGa_{1-x}As$ HEMTs by application of <110> uniaxial strain", in *International Conference on Indium Phosphide Related Materials*, 2010, pp. 1–4.

Chapter 3

Improved modelling for Surface Roughness

In this chapter of the thesis we present a new model for the surface roughness (SR) scattering in planar and circular gate-all-around (GAA) MOS transistors. The new model is first derived for planar transistor and validated against mobility data and published in [1], and then, it is extended to circular GAA MOSFETs. The model is suitable for bulk and thin body devices and explicitly takes into account the non linear relation between the displacement Δ of the interface position and the SR scattering matrix elements, which is found to significantly influence the r.m.s value ($\Delta_{\mathrm{r.m.s.}}$) of the interface roughness that is necessary to reproduce SR-limited mobility measurements. In particular, comparison with experimental mobility for bulk Si MOSFETs shows that with the new SR scattering model a good agreement with measured mobility can be obtained with $\Delta_{\mathrm{r.m.s.}}$ values of about 0.2 nm, which is in good agreement with several AFM and TEM measurements. For thin body III-V MOSFETs, the proposed model predicts a weaker mobility degradation at small well thicknesses ($\mathrm{T_w}$), compared to the T_w^6 behavior observed in Si extremely thin body devices.

3.1 Introduction

The development of sub 14 nm CMOS technologies will make use of extremely thin body (ETB) planar MOSFETs, FinFETs with very narrow fins or nanowire devices with only a few nanometer diameter [2], because an aggressive scaling of the device cross section is necessary to assure a good electrostatic integrity at such gate lengths. Furthermore, in the quest for high performance at low supply voltage, III-V semiconductors (e.g. $\mathrm{In_{1-x}Ga_xAs}$) are being actively investigated as channel materials alternative to strained silicon [3–6], because the large carrier velocities in III-V semiconductors may offer on-current advantages for a supply voltage around 0.6V or below [7, 8].

In ETB or narrow fin transistor structures the surface roughness (SR) is a

dominant scattering mechanism, which typically limits the overall mobility especially at large inversion densities [9–12]. For an electron inversion layer in an ETB planar device the formulation of the SR scattering matrix element is also closely related to the quantization model employed for the calculation of the subband minima and envelope wave-functions. In this latter respect, it is interesting to notice that high-κ dielectrics (e.g. HfO_2 and hafnium based oxides) form a relatively small energy barrier (Φ_B) with Si and III-V materials compared to the Φ_B=3.1eV of the Si-SiO_2 system. The effective barrier height is further reduced by the strong subband quantization in the inversion layer resulting from the small effective mass of III-V materials, so that in III-V MOSFETs with high-κ dielectrics a significant penetration of the electron wave-function into the oxide region is expected to occur. This is well illustrated by numerical calculations in Fig.3.1 for a HfO_2−InAs−HfO_2 quantum well, where it is observed a much larger wave-function penetration than in a HfO_2−Si−HfO_2 well with the same 5 nm thickness.

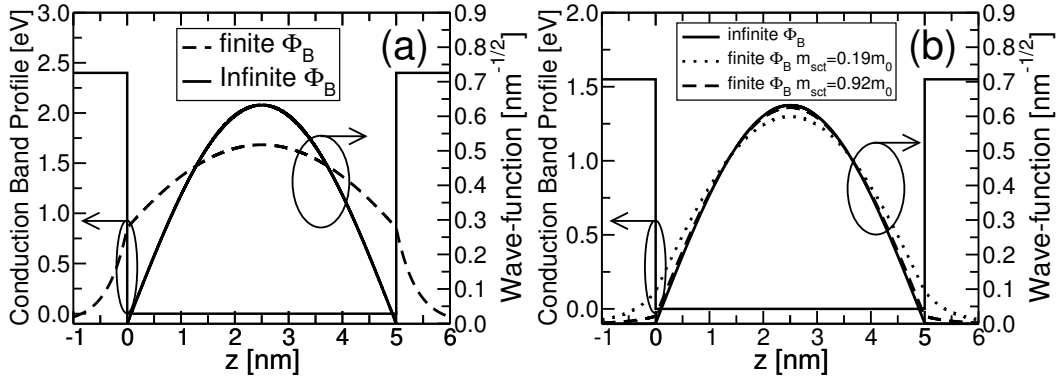


Figure 3.1: Lowest subband wave-function $\xi_0(z)$ in a 5 nm square well for either finite energy barrier Φ_B or infinite Φ_B and corresponding to: *(a)* a HfO_2-InAs-HfO_2 system (Φ_B=2.4eV); *(b)* a HfO_2-Si-HfO_2 system (Φ_B=1.55eV). Results obtained with an effective mass approximation model (see Sec.3.2.1) and considering for silicon both the light quantization mass (m_{sct}=0.19 m_0, dotted-line) and the heavy quantization mass (m_{sct}=0.92 m_0, dashed-line). The oxide mass in the energy gap is m_{ox}=0.11 m_0 [13].

Even if the penetration into the oxide region has a remarkable impact on the shape of the wave-function close to the semiconductor-oxide interface (see Fig.3.1), SR scattering matrix elements are still often formulated according to the so-called Prange-Nee (PN) model [9, 14–17], which corresponds to a closed boundary conditions treatment of the quantization problem, that is to an infinite Φ_B resulting in no wave-function penetration into the oxide. When calibrated against mobility experiments in bulk-like III-V MOSFETs, SR mobility calculations based on the PN model predict an electron mobility around 250 cm^2/Vs for $In_{0.53}Ga_{0.47}As$ FETs with $T_w \approx 5$ nm [16], whereas recent experiments reported a mobility exceeding 1000 cm^2/Vs in such a T_w range [6, 12, 18].

The generalized Prange-Nee (GPN) model provides a formulation of the SR scattering matrix elements accounting for the wave-function penetration into the oxide and it has been developed for ETB planar devices [19, 20], and also for nanowire transistors [21].

We start by revisiting the GPN model and discussing some critical points in its derivation. Then, we extend the preliminary results for planar devices reported in [22] and present a new model for SR scattering that accounts for wave-function penetration into the oxide and, furthermore, takes into account the non linear dependence of the scattering matrix element on the amplitude Δ of the fluctuations of the interface position. The new model is first employed for the interpretation of well established, experimental mobility curves in Si-SiO_2 bulk MOSFETs and then for the analysis of SR-limited mobility (μ_{SR}) in silicon and III-V ETB transistors. Our results show that the new SR scattering model can reproduce the experimental μ_{SR} in Si-SiO_2 bulk MOSFETs with r.m.s. values $\Delta_{\mathrm{r.m.s.}}$ of the SR spectrum in close agreement with AFM and TEM measurements, and substantially smaller than the $\Delta_{\mathrm{r.m.s.}}$ values reported in several previous studies employing the PN or the GPN modeling approach [9, 14–17, 19–21, 23–29]. As for the μ_{SR} dependence on T_w predicted by the new SR scattering model, it is found that while in the Si-SiO_2 system the μ_{SR} degrades as T_w^6 at small well thicknesses (as predicted also by the GPN model and observed in experiments [11]), in III-V materials the μ_{SR} exhibits a weaker reduction with decreasing T_w. Then, the new model for SR is extended to circular GAA MOSFETs for the cases with either isotropic or anisotropic quantization masses.

The work is organized as follows. In Sec.3.2 we first revisit the GPN model and then we present a new formulation of SR scattering matrix elements and discuss in detail the non linear dependence of the scattering matrix elements on the amplitude Δ of the interface roughness. In Sec.3.3 we describe the mobility calculation. In Sec.3.4 we finally show the numerical results for μ_{SR} calculations in silicon and III-V MOSFETs. Some concluding remarks are finally reported in Sec.3.5. In Sec.3.6 we extend the new SR model to circular GAA MOSFETs.

3.2 Modeling of surface roughness scattering: planar MOSFETs

Let us assume that in the absence of surface roughness the envelope wave-function can be written in the form

$$\psi_{n\mathbf{k}}(z,\mathbf{r}) = \frac{1}{\sqrt{A}}\,\xi_n(z)e^{i\mathbf{k}\cdot\mathbf{r}} \tag{3.1}$$

where A is the normalization area in the (x,y) transport plane, z is the quantization direction, $\mathbf{k}$= (k_x, k_y) the in-plane electron wave-vector and n the subband index (the valley index has been dropped to simplify the notation). The $\xi_n(z)$ is determined by the eigenvalue problem

$$\hat{H}_0\xi_n(z) = \varepsilon_n\,\xi_n(z) \tag{3.2}$$

where ε_n is the subband minimum and, for a thin body MOS system, the unperturbed effective mass Hamiltonian reads

$$\hat{H}_0 = -\frac{\hbar^2}{2}\frac{\partial}{\partial z}\frac{1}{m_z}\frac{\partial}{\partial z} - e\phi(z) + \Phi_B\Theta(-z) + \Phi_B\Theta(z - T_w) \tag{3.3}$$

with $\phi(z)$ being the electrostatic potential, Φ_B the semiconductor-oxide energy barrier, $\Theta(x)$ the unit step function and $\mathrm{T_w}$ the semiconductor well thickness. The quantization mass m_z changes with the material and can be written as

$$m_z = m_{ox}\left[\Theta(-z) + \Theta(z - T_w)\right] + m_{sct}\left[\Theta(z) - \Theta(z - T_w)\right] \tag{3.4}$$

with m_{sct} and m_{ox} being the semiconductor and oxide mass, respectively.

It is worth recalling that, as demonstrated in Appendix.B.1, if the direction z is not aligned with any axis of the energy ellipsoid, then the *ansatz* for $\psi_{n\mathbf{k}}(z, \mathbf{r})$ in Eq.3.1 is not valid because the wave-function $\xi_n(z)$ becomes $\mathbf{k}$ dependent [30], even though the dependence is through a mere phase factor that is expected to have a limited impact on scattering rates and it is typically neglected in transport studies. Moreover, even for spherical bands, if m_{sct} and m_{ox} are different the $\mathbf{k}$ dependent kinetic energy is not a simple additive term to the subband minima ε_n obtained from Eqs.3.2, 3.3. More precisely, it can be easily shown that, for $m_{sct} \neq m_{ox}$, Eqs.3.2, 3.3 should be rewritten by adding in the oxide region a term $[\hbar^2k^2/2(1/m_{ox}-1/m_{sct})]$ which adds to the energy barrier Φ_B. Strictly speaking, the presence of such a k dependent term makes Eq.3.1 invalid because $\xi_n(z)$ becomes $\mathbf{k}$ dependent. For the oxide-semiconductor systems considered in this work, however, $[\hbar^2k^2/2(1/m_{ox}-1/m_{sct})]$ is much smaller than Φ_B for the k values of practical interest, which makes Eq.3.1 a reasonable assumption. Finally, for interfaces where the semiconductor and oxide conduction band minima occur at different points in the Brillouin zone, Eq.3.3 does not account for the conservation of the total crystal momentum, which is a simplifying assumption employed, for example, in all papers dealing with transport at the Si-SiO_2 interface that we are aware of.

We now move to the system in the presence of surface roughness, and assume that the SR at the two interfaces of an ETB MOSFET is uncorrelated and denote by $\Delta(\mathbf{r})$ the roughness at the front interface, nominally located at z=0, as a function of the position $\mathbf{r}$ in the transport plane. The perturbed Hamiltonian reads

$$\hat{H}_{p,\mathbf{r}z} = -\frac{\hbar^2}{2}\frac{\partial}{\partial z}\frac{1}{m_{p,\mathbf{r}z}}\frac{\partial}{\partial z} - e\phi(z) + \Phi_B\Theta(-z + \Delta(\mathbf{r})) + \Phi_B\Theta(z - T_w). \tag{3.5}$$

where $m_{p,\mathbf{r}z}$ is

$$m_{p,\mathbf{r}z} = m_{ox}\left[\Theta(-z + \Delta(\mathbf{r})) + \Theta(z - T_w)\right] + m_{sct}\left[\Theta(z - \Delta(\mathbf{r})) - \Theta(z - T_w)\right]. \tag{3.6}$$

The unscreened SR scattering matrix element can thus be written as

$$\mathcal{M}_{nn'}(\mathbf{q}) = \frac{1}{A}\int_A M_{nn'}[\Delta(\mathbf{r})]\, e^{-i\mathbf{q}\cdot\mathbf{r}} d\mathbf{r} \tag{3.7}$$

where $\mathbf{q}$=($\mathbf{k}'$−$\mathbf{k}$) and $M_{nn'}[\Delta(\mathbf{r})]$ is written as

$$M_{nn'}[\Delta(\mathbf{r})] = \int_z \xi^{\dagger}_{n'}(z)[(\hat{H}_{p,\mathbf{r}z} - \hat{H}_0)\,\xi_n(z)]\,dz. \tag{3.8}$$

As can be seen, the definition of the SR scattering matrix element in Eq.3.8 accounts only for the contribution to the scattering matrix element due to the roughness induced changes of the energy barrier (Φ_B) and effective mass, whereas it does not include the terms due to the perturbations of the electrostatic potential produced by the fluctuations of electron and polarization charges [19, 31]. We will further discuss these additional terms in Sec.3.4.

It should be also noticed that the equations discussed so far refer to thin body SOI MOS transistors, however all the derivations and discussions developed in this work apply also to a bulk MOSFET. In particular, the equations describing the bulk transistors can be readily obtained by taking the limit $\mathrm{T_w}\rightarrow\infty$ in the equations written for thin body MOSFETs, and a few equations for bulk transistors will be explicitly discussed below.

3.2.1 The generalized Prange-Nee model (GPN)

In this section we briefly review the derivation of the PN model and of the GPN model, that have been widely used in the literature [9, 14–17, 19–21, 23–29], where the $M_{nn'}[\Delta(\mathbf{r})]$ defined in Eq.3.8 is simplified to be a linear function of $\Delta(\mathbf{r})$. We will see that a few assumptions in the derivation deserve a reconsideration, and in fact stimulated us to propose the new model presented in the next section.

The determination of $M_{nn'}[\Delta(\mathbf{r})]$ according to Eq.3.8 essentially requires to calculate the matrix element for $\hat{H}_{p,\mathbf{r}z}$, in fact the matrix element for $\hat{H}_0$ is simply given by $\varepsilon_n\delta_{n,n'}$, where ε_n is the subband minimum defined in Eq.3.2. This is typically tackled by introducing the auxiliary abscissa z' defined as [19, 20]

$$z' = \frac{\mathrm{T_w}(z-\Delta(\mathbf{r}))}{\mathrm{T_w}-\Delta(\mathbf{r})} \quad \Rightarrow \quad z = z' + \left[1 - \frac{z'}{\mathrm{T_w}}\right]\Delta(\mathbf{r}) \tag{3.9}$$

such that the interfaces in the presence of roughness are located at z'=0 and z'=$\mathrm{T_w}$. Consequently the perturbed Hamiltonian can be written in terms of z' as

$$\hat{H}_{p,\mathbf{r}z'} = -\frac{\hbar^2}{2}\left[\frac{\mathrm{T_w}}{\mathrm{T_w}-\Delta(\mathbf{r})}\right]^2 \frac{\partial}{\partial z'}\frac{1}{m_{p,\mathbf{r}z'}}\frac{\partial}{\partial z'} - e\phi_{z'}(z') + \Phi_B\Theta(-z') + \Phi_B\Theta(z'-T_w) \tag{3.10}$$

where the $m_{p,\mathbf{r}z'}$ dependence on z' is the same as the m_z dependence on z given by Eq.3.4. As can be seen, $\hat{H}_{p,\mathbf{r}z'}$ can be expressed in terms of z' in a form which is very similar to the form of $\hat{H}_0$ in terms of z, except for the electrostatic potential $\phi_{z'}(z')$, which is a notation better explained a few lines below.

The derivation of the GPN continues by calculating the matrix elements of $\hat{H}_{p,\mathbf{r}z'}$ using the auxiliary abscissa z'. In order to do that, however, we need to identify the

form of the unperturbed wave-functions, $\xi_{n,z'}(z')$, and of the electrostatic potential, $\phi_{z'}(z')$, when they are expressed as a function of z'. We first consider the wave-function and notice that, since the $\xi_n(z)$ is a known function of the abscissa z, then Eq.3.9 readily allows us to write $\xi_{n,z'}(z')$ as

$$\xi_{n,z'}(z') = \xi_n(z' + (1 - z'/\mathrm{T_w})\Delta(\mathbf{r})) \tag{3.11}$$

and a similar expression holds for $\phi_{z'}(z')$. The notation $\xi_n(z' + (1 - z'/\mathrm{T_w})\Delta(\mathbf{r}))$ in Eq.3.11 denotes the wave-function ξ_n evaluated in $z = z' + (1 - z'/\mathrm{T_w})\Delta(\mathbf{r})$.

If one wished to calculate the matrix elements for $\hat{H}_{p,\mathbf{r}z'}$ by using $\xi_{n,z'}(z')$ from Eq.3.11 (and the corresponding expression for $\phi_{z'}(z')$), one would face two difficulties. The first is that in Eq.3.10 the energy barrier and the discontinuity of $m_{p,\mathbf{r}z'}$ are located in z'=0 (for the front interface), whereas $\xi_{n,z'}(z')$ in Eq.3.11 has the discontinuity of $(\partial\xi_{n,z'}/\partial z')$ at the point z'= $(-\Delta(\mathbf{r})\mathrm{T_w})/(\mathrm{T_w}-\Delta(\mathbf{r}))$ (i.e. $z' \simeq -\Delta(\mathbf{r})$ for $\Delta(\mathbf{r}) \ll \mathrm{T_w}$), which is in fact the z' corresponding to z =0. The second critical point is that the corresponding matrix element $M_{nn'}[\Delta(\mathbf{r})]$ would be a non linear function of $\Delta(\mathbf{r})$: in fact, besides the pre-factor of $\hat{H}_{p,\mathbf{r}z'}$ in Eq.3.10, the $\xi_{n,z'}(z')$ dependence on $\Delta(\mathbf{r})$ implied by Eq.3.11 is also non linear.

At this stage some approximations are typically introduced and, in particular, the $\xi_{n,z'}(z')$ expression in Eq.3.11 is simplified as [20, 21, 27, 29]

$$\xi_{n,z'}(z') \simeq \xi_n(z') + \frac{\partial\xi_n(z')}{\partial z'}(1 - z'/\mathrm{T_w})\Delta(\mathbf{r}) \tag{3.12}$$

obtained by using a first order expansion of ξ_n around the point z' where $\xi_{n,z'}(z')$ is to be evaluated. A similarly approximated expression is introduced also for the electrostatic potential $\phi_{z'}(z')$. The matrix elements for $\hat{H}_{p,\mathbf{r}z'}$ are then calculated by using such simplified expressions for $\xi_{n,z'}(z')$ and $\phi_{z'}(z')$ and, by keeping only the first order terms in $\Delta(\mathbf{r})$, the following expression for the SR scattering matrix element can be derived [19, 20]

$$\begin{aligned} M_{nn'}[\Delta(\mathbf{r})] &= \Delta(\mathbf{r})\left[-\int_z \xi_n e\frac{\partial\phi}{\partial z}\xi_{n'}\,dz + (\varepsilon_n - \varepsilon_{n'})\int_z \xi_n\frac{\partial\xi_{n'}}{\partial z}\,dz\right] \\ &- \frac{\Delta(\mathbf{r})}{\mathrm{T_w}}\left[\int_z \xi_n\left(\hbar^2\frac{\partial}{\partial z}\frac{1}{m_z}\frac{\partial}{\partial z} - ze\frac{\partial\phi}{\partial z}\right)\xi_{n'} + (\varepsilon_n - \varepsilon_{n'})\int_z \xi_n z\frac{\partial\xi_{n'}}{\partial z}\,dz\right]. \end{aligned} \tag{3.13}$$

It is also worth noting that in the limit of an infinitely large barrier Φ_B, we obtain the well known Prange-Nee expression [19, 20]

$$M_{nn'}[\Delta(\mathbf{r})] = \left[\frac{\hbar^2}{2m_z}\frac{\partial\xi_{n'}(0)}{\partial z}\frac{\partial\xi_n(0)}{\partial z}\right]\Delta(\mathbf{r}) \tag{3.14}$$

so that Eq.3.13 is regarded as a generalized Prange-Nee model taking into account the wave-function penetration into the oxide region.

The GPN model summarized by Eq.3.13 has removed the non linear dependence of the matrix element on $\Delta(\mathbf{r})$ which is implicit to Eq.3.8. This has been accomplished by introducing z' (such that $\hat{H}_{p,\mathbf{r}z'}$ has the front interface in z'=0), then by using the approximated expression of $\xi_{n,z'}(z')$ in Eq.3.12 (and a similar expression for the potential), and finally by keeping only the first order terms in $\Delta(\mathbf{r})$. The use of Eq.3.12 is a key step in the derivations and deserves further discussion because it is quite delicate when the oxide and semiconductor masses are different, as it is the case in most material systems used for MOS transistors. In fact for $m_{ox} \neq m_{sct}$ the function $\xi_{n,z'}(z')$ in the exact form given by Eq.3.11 is a continuous function of z' whose first derivative is discontinuous in z'= $(-\Delta \mathrm{T_w})/(\mathrm{T_w} - \Delta)$ (i.e. $z' \simeq -\Delta$ for $\Delta \ll \mathrm{T_w}$), while the approximated form of $\xi_{n,z'}(z')$ in Eq.3.12 is discontinuous in z'=0, because the derivative of the unperturbed wave-function $\xi_n(z')$ is discontinuous in z'=0.

Fig.3.2 compares the $\xi_{n,z'}(z')$ expressions in Eqs.3.11 and Eq.3.12 in a HfO_2-Si-HfO_2 quantum well with $\mathrm{T_w}$=5 nm for either m_{sct}=0.92 $\mathrm{m_0}$ or m_{sct}=0.19 $\mathrm{m_0}$. As can be seen Eq.3.12 is very different compared to the $\xi_{n,z'}(z')$ of Eq.3.11 for z' from approximately $-\Delta$ to 0 (corresponding to z from 0 to Δ), where Eq.3.12 leads to the expected discontinuity at z'=0. In fact, the main issue with Eq.3.12 is that it is based on a first order expansion of ξ_n around the point z', which for $m_{ox} \neq m_{sct}$ is very inaccurate when z' and $[z'+(1\text{-}z'/\mathrm{T_w})\Delta]$ have a different sign (i.e. for z' from approximately $-\Delta$ to 0 in Fig.3.2), because the ξ_n derivative is discontinuous at z'=0. The same problem holds for the approximated expression of the electrostatic potential $\phi_{z'}(z')$ which is used to reach the GPN model formulation in Eq.3.13, because the derivative of ϕ is also discontinuous at z'=0 for most semiconductor-oxide interfaces due to the different dielectric constant for $z < 0$ and $z > 0$.

It is now important to notice that the main remarks discussed for the derivation of Eq.3.13 apply also to a bulk MOSFET. For a bulk MOSFET, in fact, the derivations follow the same path by first defining the abscissa $z' = z - \Delta(\mathbf{r})$(as obtained from Eq.3.9 for $\mathrm{T_w} \rightarrow \infty$), then by expanding $\xi_{n,z'}(z')$ as

$$\xi_{n,z'}(z') \simeq \xi_n(z') + \frac{\partial \xi_n(z')}{\partial z'} \Delta(\mathbf{r}) \tag{3.15}$$

and with a similar expression for the electrostatic potential. Finally the matrix element can be derived by keeping only the first order terms in $\Delta(\mathbf{r})$, and we obtain [19, 20]

$$M_{nn'}[\Delta(\mathbf{r})] = \Delta(\mathbf{r}) \left[- \int_z \xi_n e \frac{\partial \phi}{\partial z} \xi_{n'} \, dz + (\varepsilon_n - \varepsilon_{n'}) \int_z \xi_n \frac{\partial \xi_{n'}}{\partial z} \, dz \right]. \tag{3.16}$$

The main issue with Eq.3.15 is the same as with Eq.3.12: since for $m_{ox} \neq m_{sct}$ the derivative of ξ_n is discontinuous at z'=0, then Eq.3.15 can be very inaccurate for z' close to zero.

In consideration of the above discussion and of the $\xi_{n,z'}(z')$ plots in Fig.3.2, we believe that the use of Eqs.3.12, 3.15 is somewhat unjustified for $m_{ox} \neq m_{sct}$, but

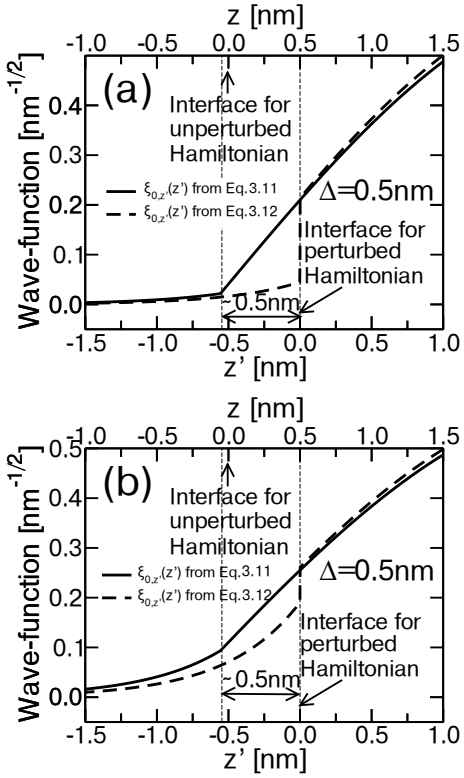


Figure 3.2: Unperturbed wave-function ξ_0 in a 5 nm HfO_2-Si-HfO_2 quantum well for (a) the lowest unprimed subband (m_{sct}=0.92 m_0) and (b) the lowest primed subband (m_{sct}=0.19 m_0) versus the auxiliary abscissa z' defined in Eq.3.9 (bottom horizontal axis) or versus z (top horizontal axis). ξ_0 is calculated either with the exact form in Eq.3.11 or according to the approximation in Eq.3.12 and for a positive Δ =0.5 nm. The oxide mass is m_{ox}=0.11 m_0 [13].

it is nevertheless a necessary step if the final goal is to obtain a linear model for SR scattering, that is a model where the matrix elements $\mathrm{M}_{nn'}[\Delta(\mathbf{r})]$ are simply proportional to $\Delta(\mathbf{r})$. In fact Eqs.3.13, 3.14 and 3.16 can all be cast in the form $M_{nn'}[\Delta(\mathbf{r})] = M^{(0)}_{nn'}\,\Delta(\mathbf{r})$, which is also an implicit definition of $M^{(0)}_{nn'}$. Such a linear dependence on $\Delta(\mathbf{r})$ is, in turn, the crucial result to obtain an ensemble averaged squared matrix element $\left\langle|\mathcal{M}_{nn'}(\mathbf{q})|^2\right\rangle$ that is simply proportional to the power spectrum of the surface roughness. In fact, by assuming $M_{nn'}[\Delta(\mathbf{r})] = M^{(0)}_{nn'}\,\Delta(\mathbf{r})$, the $\left\langle|\mathcal{M}_{nn'}(\mathbf{q})|^2\right\rangle$ is readily obtained using the Wiener−Khinchin theorem as [19, 20] (that states that the power spectrum of a stationary random process is given by the Fourier transform of the autocorrelation function of the process):

$$\left\langle|\mathcal{M}_{nn'}(\mathbf{q})|^2\right\rangle = |M^{(0)}_{nn'}|^2\frac{S_\Delta(\mathbf{q})}{A} \tag{3.17}$$

where the roughness power spectrum $S_\Delta(\mathbf{q})$ is given by

$$S_\Delta(\mathbf{q}) = \int_A C_\Delta(\mathbf{r})\,e^{-i\mathbf{q}\cdot\mathbf{r}}\,d\mathbf{r}. \tag{3.18}$$

which is the Fourier transform of the auto-correlation function [32]

$$C_\Delta(\mathbf{r}) = \frac{1}{A}\int_A \Delta(\mathbf{r}')\Delta(\mathbf{r}'+\mathbf{r})d\mathbf{r}'. \tag{3.19}$$

An exponential or a Gaussian form for the auto-correlation function has frequently been used to describe surface roughness limited mobility in MOS transistors [16, 17, 19, 20, 23–27].

3.2.2 New formulation for the SR scattering matrix elements

The discussion carried out in the previous section raised some doubts, at least for $\mathrm{m}_{ox} \neq \mathrm{m}_{sct}$, about the derivations leading from the definition of the SR scattering matrix element $\mathrm{M}_{nn'}[\Delta(\mathbf{r})]$ in Eq.3.8 to the linear formulation in Eqs.3.13 or 3.14. As can be seen, in fact, the $M_{nn'}[\Delta(\mathbf{r})]$ defined in Eq.3.8 is a non linear function of $\Delta(\mathbf{r})$, which defines the narrow region around the interface at z=0 where the integrand function is not zero. Furthermore for $\mathrm{m}_{ox} \neq \mathrm{m}_{sct}$ the wave-function derivative is discontinuous at z=0 with quite different values of the $\partial\xi_n/\partial\mathrm{z}$ at $\mathrm{z}=0^-$ and $\mathrm{z}=0^+$, so that the $M_{nn'}[\Delta(\mathbf{r})]$ defined in Eq.3.8 is not expected to be symmetric for positive and negative $\Delta(\mathbf{r})$, not even for arbitrarily small $\Delta(\mathbf{r})$ values.

In virtue of the above discussion, in this section we propose a fully numerical calculation of $M_{nn'}[\Delta(\mathbf{r})]$ based directly on Eq.3.8. We thus proceed by directly inserting in Eq.3.8 the unperturbed $\hat{H}_0$ and perturbed Hamiltonian $\hat{H}_{p,\mathbf{r}z}$ given

respectively by Eq.3.3 and 3.5, and obtain

$$M_{nn'}[\Delta] = \int_{\Delta}^{0} \xi^{\dagger}_{n'}(z) \left[\left(\frac{\hbar^2}{2m_{ox}} - \frac{\hbar^2}{2m_{sct}} \right) \frac{\partial^2 \xi_n(z)}{\partial z^2} - \Phi_B \, \xi_n(z) \right] dz \quad \Delta \leq 0. \tag{3.20a}$$

$$M_{nn'}[\Delta] = \int_{0}^{\Delta} \xi^{\dagger}_{n'}(z) \left[\left(-\frac{\hbar^2}{2m_{ox}} + \frac{\hbar^2}{2m_{sct}} \right) \frac{\partial^2 \xi_n(z)}{\partial z^2} + \Phi_B \, \xi_n(z) \right] dz \quad \Delta \geq 0 \tag{3.20b}$$

Fig.3.3 reports the $M_{nn'}[\Delta]$ versus Δ calculated with Eq.3.20 for a 5 nm thick HfO_2-Si-HfO_2 quantum well and compares it with the linear formulations given by the GPN model in Eqs.3.13 and the PN model in 3.14. The difference between the

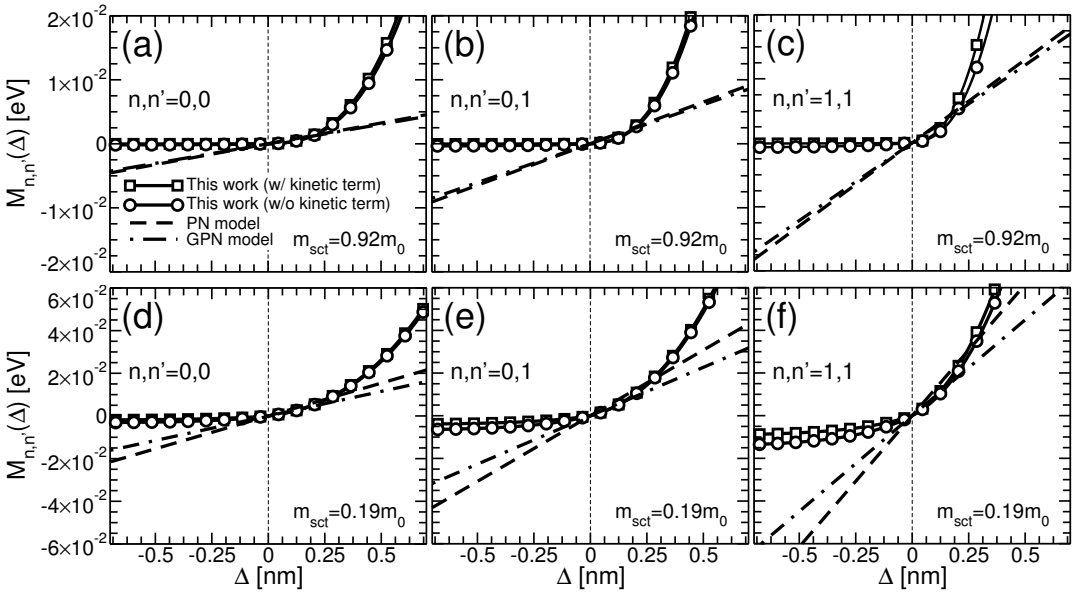


Figure 3.3: SR scattering matrix elements versus Δ for a 5 nm HfO_2-Si-HfO_2 square quantum well calculated according to the PN model of Eq.3.14 (dashed line), the GPN model of Eq.3.13 (dot-dashed line) and the new model of Eq.3.20 (solid line). *(a),(b),(c)*: unprimed subbands having m_{sct}=0.92 m_0; *(d),(e),(f)*: primed subbands having m_{sct}=0.19 m_0. Results are for different subband transitions (e.g. (a) and (d) are for the lowest subband, i.e. n=n′ =0). The matrix elements calculated with Eq.3.20 exhibit a quite strong non linear behavior. Open squares and open circles are calculated either including or neglecting the kinetic term discussed in Sec.3.2.3.

calculations with (open square) or without (open circles) the kinetic term will be discussed in Sec.3.2.3. We first notice that, for a given Δ magnitude, $M_{nn'}[\Delta]$ is much larger for a positive than it is for a negative Δ. This is consistent with the wave-function behavior close to the interface (see Fig.3.1(b)): in fact the exponential

$\Delta_{\text{r.m.s.}}$ and Λ from AFM or TEM measurements for the Si-SiO_2 interface			
	$\Delta_{\text{r.m.s.}}$[nm]	Λ[nm]	Spectrum
S.M.Goodnick *et al.* [23]	0.14-0.20	0.6-2.5	Gaussian
	0.14-0.20	0.7 - 3.7	Exponential
T.Yamanaka *et al.* [33]	0.21	—	—
A.Pirovano *et al.* [34]	0.18	1.8	Gaussian
O.Bonno *et al.* [35]	0.18	$\simeq$4.1	Exponential
$\Delta_{\text{r.m.s.}}$ and Λ extracted from mobility calculations			
	$\Delta_{\text{r.m.s.}}$[nm]	Λ[nm]	Spectrum
M.V.Fischetti *et al.* [25]	0.48	1.3	Exponential
C.Jungemann *et al.* [24]	0.44	2.5	Gaussian
T.Ishihara *et al.* [27]	0.55	1.3	Gaussian
D.Esseni *et al.*[20]	0.62	1.0	Gaussian
This Work	0.21	1.4	Exponential

Table 3.1: SR scattering spectrum parameters for (100) silicon MOSFETs measured by AFM and TEM techniques (top part of the table) or inferred from a comparison between simulated and experimental mobility values (bottom part of the table).

decay in the oxide region and the super-linear increase in the semiconductor region result in a stronger interaction with the scattering Hamiltonian when Δ is positive. Then we also see that such a strong asymmetry of $M_{nn'}[\Delta]$ for either negative or positive Δ makes it inherently difficult to find a fairly good approximation with a linear model, as attempted by Eqs.3.13 or 3.14. In this latter respect, in fact, it is useful to recall that the r.m.s. roughness value, $\Delta_{\text{r.m.s.}}$, typically reported to reproduce experimental mobility in silicon MOSFETs by using the GPN or PN model is in the range of $0.4-0.6$ nm (see also Tab.3.1), so that the linearity assumed by Eqs.3.13 or 3.14 should hold at least up to $1-1.5$ nm in order for the modeling assumption to be consistent with the extracted $\Delta_{\text{r.m.s.}}$ value.

The numerical determination of $M_{nn'}[\Delta(\mathbf{r})]$ shown in Fig.3.3 can be used as a basis for a new SR scattering model, however the problem is to calculate $\left\langle |\mathcal{M}_{nn'}(\mathbf{q})|^2 \right\rangle$ when $M_{nn'}[\Delta(\mathbf{r})]$ is not simply proportional to $\Delta(\mathbf{r})$, so that $\left\langle |\mathcal{M}_{nn'}(\mathbf{q})|^2 \right\rangle$ is not proportional to the power spectrum $\mathrm{S}(\mathbf{q})$ of the roughness. In fact, since $\Delta(\mathbf{r})$ is a stochastic process (whose auto-correlation function and spectrum can be assumed to be known), then $M_{nn'}[\Delta(\mathbf{r})]$ must be treated as a non linear transformation of the stochastic process $\Delta(\mathbf{r})$. This is a known and studied problem in signal theory and it can be shown that the auto-correlation function $\mathrm{C}_M(\mathbf{r})$ of $M_{nn'}[\Delta]$, can be written as [36]

$$C_M(\mathbf{r}) = \frac{1}{2\pi\, C_\Delta(0)\sqrt{1 - C^2_{\Delta,N}(\mathbf{r})}} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} M_{nn'}[\Delta_1]\, M_{nn'}[\Delta_2] \times \\ \times \exp\left[-\frac{\Delta_1^2 + \Delta_2^2 - 2C_{\Delta,N}(\mathbf{r})\Delta_1\Delta_2}{2C_\Delta(0)(1 - C^2_{\Delta,N}(\mathbf{r}))} \right] d\Delta_1\, d\Delta_2 \tag{3.21}$$

where $C_\Delta(\mathbf{r})$ is the SR auto-correlation function defined in Eq.3.19 and $C_{\Delta,N}(\mathbf{r})$=$C_\Delta(\mathbf{r})/C_\Delta(0)$ [1].

Once $C_M(\mathbf{r})$ is known, the corresponding power spectrum $\mathrm{S}_M(\mathbf{q})$ is readily given by the Fourier transform of $C_M(\mathbf{r})$. In all the calculations of this work we will assume an exponential form for the roughness auto-correlation function

$$C_\Delta(\mathbf{r}) = \Delta^2_{\mathrm{r.m.s.}} e^{-r\sqrt{2}/\Lambda} \tag{3.22}$$

where Λ is the correlation length of interface fluctuations. Eq.3.22 shows that for an exponential spectrum $C_\Delta(\mathbf{r})$ depends only on $r = |\mathbf{r}|$, so that also the correlation function $C_M(\mathbf{r})$ of $M_{nn'}[\Delta(\mathbf{r})]$ depends only on r, which simplifies the calculation of the two dimensional Fourier transform as shown in AppendixF. In this case, in fact, according to Eq.F.5 we finally obtain:

$$\left\langle |\mathcal{M}_{nn'}(\mathbf{q})|^2 \right\rangle = \frac{2\pi}{A} \int_0^{+\infty} C_M(r)\, J_0(qr)\; r dr \tag{3.23}$$

where $J_0(x)$ is the zero order Bessel function. In all the results obtained with the model developed in this section, the integrals in Eqs.3.21 and 3.23 were obtained by direct numerical calculation. In scattering rate calculations the normalization area A in the $\left\langle |\mathcal{M}_{nn'}(\mathbf{q})|^2 \right\rangle$ expression always cancels out when summing over final $\mathbf{k}$ states [20].

After the determination of the ensemble averaged matrix elements $\left\langle |\mathcal{M}_{nn'}(\mathbf{q})|^2 \right\rangle$ (and the inclusion of screening described Sec.3.2.5), in our work we calculate SR scattering rate and momentum relaxation time by using the first Born approximation. The use of the first Born approximation for SR scattering has been recently revisited in [37], where it is shown that, in analogy to Coulomb scattering, the validity of the approximation requires that the amplitude of the scattered wave-function be small (compared to the incident wave-function), and that the scattering potentials be separated by an average distance larger than the carriers mean-free-path. For SR scattering these requirements are ultimately related to the magnitude of the matrix elements and to the correlation length of the $\Delta(\mathbf{r})$ process. We will see in Sec.3.4 that our SR model leads to matrix elements similar to the conventional linear formulations when the models are calibrated against the same experimental data; the roughness correlation length used in our calculations is also consistent with previous studies. Consequently, we think that the validity of the first Born approximation does not change substantially in the surface roughness formulation of this work compared to conventional linear models.

[1] Eq.3.21 holds for a Gaussian joint probability distribution for $\Delta(\mathbf{x})$ and $\Delta(\mathbf{r}+\mathbf{x})$ [32, 36]. The Gaussian joint probability should not be confused with the Gaussian auto-correlation function and power spectrum.

3.2.3 Kinetic and potential energy contribution to the matrix elements

In order to gain some insight concerning the relative importance of the kinetic and potential energy contributions to the matrix elements $M_{nn'}[\Delta(\mathbf{r})]$, we rewrite the unperturbed Hamiltonian of Eq.3.3 as

$$\hat{H}^{(ox)} = \Phi_B + \hat{H}^{(ox)}_{kin} = \Phi_B + \hat{E}^{(ox)}_{cz}(-i\frac{\partial}{\partial z}) = \Phi_B - \frac{\hbar^2}{2m_{ox}}\frac{\partial^2}{\partial z^2} \qquad z<0 \tag{3.24a}$$

$$\hat{H}^{(sct)} = \hat{H}^{(sct)}_{kin} = \hat{E}^{(sct)}_{cz}(-i\frac{\partial}{\partial z}) = -\frac{\hbar^2}{2m_{sct}}\frac{\partial^2}{\partial z^2} \qquad 0\le z\le \mathrm{T_w} \tag{3.24b}$$

where a square quantum well with electrostatic potential $\phi(z)$=0 has been assumed and the Hamiltonian in the back oxide (i.e. for z>$\mathrm{T_w}$) is not indicated. Eq.3.24 recalls that the kinetic energy operators $\hat{H}^{(ox)}_{kin}$ and $\hat{H}^{(sct)}_{kin}$ are obtained with the canonical substitution $k_z \rightarrow (-i\partial/\partial z)$ respectively in the oxide $E^{(ox)}_{cz}(k_z)$ and semiconductor $E^{(sct)}_{cz}(k_z)$ conduction band energy relation, where E_{cz} is the k_z related energy component which is separable from the k_x, k_y components in a parabolic effective mass model.

It is now interesting to notice that, by recalling Eq.3.2 (valid in all spatial domains), one can readily write the eigenvalues of the kinetic energy operators

$$\hat{H}^{(ox)}_{kin}\xi_n(z) = -(\Phi_B - \varepsilon_n)\,\xi_n(z) \qquad z<0 \tag{3.25a}$$

$$\hat{H}^{(sct)}_{kin}\xi_n(z) = \varepsilon_n\,\xi_n(z) \qquad 0\le z\le \mathrm{T_w} \tag{3.25b}$$

where we always assume ε_n<Φ_B.

Eq.3.25 shows that the eigenvalue of $\hat{H}^{(sct)}_{kin}$ inside the quantum well is ε_n, hence it is a positive value corresponding to a sinusoidal form of the wave-function for 0≤z≤$\mathrm{T_w}$. In the oxide region, instead, $\hat{H}^{(ox)}_{kin}$ has a negative eigenvalue corresponding to the exponential decay of the wave-function. More precisely, the negative eigenvalue of $\hat{H}^{(ox)}_{kin}$=$\hat{E}^{(ox)}_{cz}(-i\partial/\partial z)$ is due to the imaginary wave-vector k_z of an electron in the oxide having an energy ε_n, that is an energy [Φ_B−ε_n] below the oxide conduction band edge and thus belonging to oxide energy gap.

This analysis is supported by the simple example of the wave-function $\xi_0(z)$ for the lowest subband in a square quantum well, that can be expressed analytically as [38]

$$\xi_0(z) = \begin{cases} C_0\cos(0.5k_0\mathrm{T_w})\,\mathrm{e}^{\gamma_0 z} & z\le 0 \\ C_0\cos\left[k_0(z-0.5\mathrm{T_w})\right] & 0\le z\le \mathrm{T_w} \\ C_0\cos(0.5k_0\mathrm{T_w})\,\mathrm{e}^{-\gamma_0(z-\mathrm{T_w})} & z\ge \mathrm{T_w} \end{cases} \tag{3.26}$$

where $\mathrm{C_0}$ is a normalization constant, k_0 and γ_0 are

$$k_0 = \sqrt{\frac{2m_{sct}\varepsilon_0}{\hbar^2}} \qquad \gamma_0 = \sqrt{\frac{2m_{ox}(\Phi_B-\varepsilon_0)}{\hbar^2}} \tag{3.27}$$

and ε_0 is finally given by the lowest solution of the transcendent equation

$$\tan\left[\frac{k_0 \mathrm{T_w}}{2}\right] = \frac{m_{sct}\gamma_0}{m_{ox}k_0} \ . \tag{3.28}$$

As can be seen, γ_0 is the magnitude of the imaginary k_z in the oxide corresponding to an exponentially decaying wave-function and resulting in the negative eigenvalue of $\hat{H}_{kin}^{(ox)}$=$\hat{E}_{cz}^{(ox)}(-i\partial/\partial z)$ in Eq.3.25a.

By recalling Eq.3.20, we see that the calculation of the SR scattering matrix element $M_{nn'}(\Delta)$ requires: for a negative Δ the evaluation of $\hat{H}_{kin}^{(sct)}\xi_n(z)$ in the oxide region (i.e. for $z<0$); for a positive Δ the evaluation of $\hat{H}_{kin}^{(ox)}\xi_n(z)$ in the semiconductor region (i.e. for $0{\leq}z{\leq}\mathrm{T_w}$). For a square quantum well we can use Eqs.3.25 and obtain

$$\hat{H}_{kin}^{(sct)}\xi_n(z) = \frac{m_{ox}}{m_{sct}}\hat{H}_{kin}^{(ox)}\xi_n(z) = -\frac{m_{ox}}{m_{sct}}(\Phi_B - \varepsilon_n)\,\xi_n(z) \qquad z<0 \tag{3.29a}$$

$$\hat{H}_{kin}^{(ox)}\xi_n(z) = \frac{m_{sct}}{m_{ox}}\hat{H}_{kin}^{(sct)}\xi_n(z) = \frac{m_{sct}}{m_{ox}}\varepsilon_n\,\xi_n(z) \qquad 0\leq z\leq \mathrm{T_w} \tag{3.29b}$$

By substituting Eq.3.29 in Eq.3.20 we have

$$M_{nn'}[\Delta(\mathbf{r})] = \left[-\Phi_B - \left(\frac{m_{ox}}{m_{sct}} - 1\right)(\Phi_B - \varepsilon_n)\right]\int_{\Delta}^{0} \xi_{n'}^{\dagger}(z)\,\xi_n(z)\partial z \qquad \Delta<0 \tag{3.30a}$$

$$M_{nn'}[\Delta(\mathbf{r})] = \left[\Phi_B + \left(\frac{m_{sct}}{m_{ox}} - 1\right)\varepsilon_n\right]\int_{0}^{\Delta} \xi_{n'}^{\dagger}(z)\,\xi_n(z)dz \qquad \Delta\geq 0 \tag{3.30b}$$

which is a form of the matrix element that, for a square quantum well with $\phi(z)$ =0, is equivalent to Eq.3.20.

In Eq.3.30 the contribution to $M_{nn'}[\Delta(\mathbf{r})]$ produced by the kinetic energy operators is easily identified in the term that vanishes for m_{ox}=m_{sct}. Furthermore, because of the exponential decay of the wave-functions in the oxide, the integral of the wave-functions in Eq.3.30a for negative Δ is much smaller, for a given $|\Delta|$, than the corresponding integral in Eq.3.30b for positive Δ. This observation essentially reiterates our argument about the strong non linear dependence of $M_{nn'}[\Delta(\mathbf{r})]$ on Δ also illustrated by Fig.3.3.

Eqs.3.29 and 3.30 deserve a few more specific comments that we discuss separately for either a semiconductor-oxide system or a hetero-structure between two different semiconductors.

Semiconductor-oxide interface with large energy barrier

For the semiconductor-oxide interfaces typically employed in the transistors of mainstream CMOS technologies (e.g. Si-SiO_2, Si-HfO_2, Si-Al_2O_3, InGaAs-HfO_2,

InGaAs-Al_2O_3), the barrier height Φ_B ranges from about 2eV to more than 3eV and it is thus much larger than ε_n for all the well thicknesses of practical interest.

In this case, for $\Delta>0$ the kinetic energy contribution to $M_{nn'}[\Delta(\mathbf{r})]$ is typically smaller than the energy barrier contribution, that is the second term in the square bracket of Eq.3.30b is small compared to Φ_B. This is confirmed by the numerical calculations in Fig.3.3 where the matrix element obtained by neglecting the kinetic term is also shown.

The situation is instead more complicated for $\Delta<0$ and we believe that, depending on Φ_B and m_{ox}/m_{sct}, the final result may not be physically meaningful. The problem can be identified in Eq.3.29a, where $\hat{H}_{kin}^{(sct)}$ is applied to the exponentially decaying wave-function in the oxide leading to the negative eigenvalue $[-(m_{ox}/m_{sct})(\Phi_B-\varepsilon_n)]$. Such an eigenvalue corresponds to an electron in the semiconductor having an imaginary k_z value and an energy $[(m_{ox}/m_{sct})(\Phi_B-\varepsilon_n)]$ below the conduction band. This result is for sure not physically meaningful if $[(m_{ox}/m_{sct})(\Phi_B-\varepsilon_n)]$ is larger than the energy gap E_G of the semiconductor. More realistically, since the parabolic effective mass kinetic energy operator $\hat{H}_{kin}^{(sct)} = \hat{E}_{cz}^{(sct)}(-i\partial/\partial z)$ can be considered a fairly accurate model for energies in the gap down to, at best, roughly half the energy gap [39], then $[(m_{ox}/m_{sct})(\Phi_B-\varepsilon_n)]$ should be limited more prudentially to values smaller than about $0.5E_G$.

In silicon MOSFETs with a (100) interface the quantization mass is $0.92m_0$ for the unprimed subbands, hence $[m_{ox}/m_{sct}]$ is smaller than 1.0; in fact we have $[m_{ox}/m_{sct}]\simeq 0.54$ for an SiO_2 gate oxide ($m_{ox}=0.5m_0$) and $[m_{ox}/m_{sct}]\simeq 0.12$ for an HfO_2 oxide ($m_{ox}=0.11m_0$). In these systems the limitations in the use of $\hat{H}_{kin}^{(sct)}$ in the oxide region are not so critical in practical calculations, even because the overall matrix element is largely dominated by the $M_{nn'}[\Delta(\mathbf{r})]$ for positive Δ values, as illustrated in Fig.3.3.

In III-V semiconductors with a very small m_{sct} at the Γ point, instead, $[m_{ox}/m_{sct}]$ can be significantly larger than 1: for example $[m_{ox}/m_{sct}] \simeq 4.2$ for the InAs-HfO_2 system. In such systems Eq.3.29a affirms that the application of $\hat{H}_{kin}^{(sct)}$ to the wave-function in the oxide results in negative eigenvalues whose magnitude is as large as 10eV or more; of course such energy values cannot be considered physically meaningful. We believe that this is a fundamental limitation related to the use of the parabolic effective mass Hamiltonian for the kinetic operator, which precludes a physically meaningful calculation of the kinetic energy contribution to the SR scattering matrix element when the effective mass in the semiconductor is smaller than in the oxide and Φ_B is a few eV. This underlines also the interest for surface roughness models based on a Hamiltonian going beyond the effective mass approximation, such as the recently proposed approach based on pseudo-potentials calculations [37]. In the present work, we decided to neglect the kinetic term in the calculations for high-κ/III-V systems.

Hetero-structures between small gap semiconductors

In the hetero-structures between different semiconductors that are routinely employed in the fabrication of III-V based CMOS transistors and HEMTs, the energy barrier Φ_B given by the conduction band discontinuity is in the range of a few hundreds meV: for example 0.12eV for InP-$In_{0.53}Ga_{0.47}As$; 0.3eV for $In_{0.52}Al_{0.48}As$-$In_{0.53}Ga_{0.47}As$; 0.5eV for InAs-InP; 0.4eV for $In_{0.53}Ga_{0.47}As$-InAs. This implies that for a given $\Delta_{\text{r.m.s.}}$ the energy barrier contribution to the SR scattering matrix element is much smaller than in the oxide-semiconductor system of an MOS transistor, hence the kinetic contribution may have a larger relative importance.

The reduction of Φ_B also implies that, depending on the ratio between the effective masses in the active and barrier semiconductor, the application of $\hat{H}_{kin}^{(sct)}$ to the wave-function in the barrier region results in contributions to the SR scattering matrix element that retain full physical meaningfulness.

3.2.4 Surface roughness at front and back interface

Since in our model the SR matrix element is a non linear transformation of the SR process, one may legitimately wonder if in an ETB transistor the effects of the roughness $\Delta_F(\mathbf{r})$ at the front and the roughness $\Delta_B(\mathbf{r})$ at back interface are simply additive. At this regard, let us denote with $M_F(\mathbf{r}) = M_{nn'}[\Delta_F(\mathbf{r})]$ the matrix element produced by $\Delta_F(\mathbf{r})$ alone (i.e. with $\Delta_B(\mathbf{r})$ =0), and with $M_B(\mathbf{r}) = M_{nn'}[\Delta_B(\mathbf{r})]$ the matrix element produced by $\Delta_B(\mathbf{r})$ alone; $M_F(\mathbf{r})$ and $M_B(\mathbf{r})$ can be calculated by using Eq.3.20 and, furthermore, we assume that $\Delta_F(\mathbf{r})$ and $\Delta_B(\mathbf{r})$ are uncorrelated. $M_F(\mathbf{r})$ and $M_B(\mathbf{r})$ are non linear transformations of $\Delta_F(\mathbf{r})$ and $\Delta_B(\mathbf{r})$ and have non zero mean, that we denote respectively by M_{F0} and M_{B0}. Moreover $M_F(\mathbf{r})$ and $M_B(\mathbf{r})$ are uncorrelated because they are space invariant transformations of uncorrelated processes. The auto-correlation function needed for scattering rate calculations via Eq.3.23 is given by

$$\begin{aligned} C(\mathbf{r}) &= \frac{1}{A}\int \left[M_F(\mathbf{r}') + M_B(\mathbf{r}')\right]\left[M_F(\mathbf{r}'+\mathbf{r}) + M_B(\mathbf{r}'+\mathbf{r})\right] d\mathbf{r}' \\ &= C_F(\mathbf{r}) + C_B(\mathbf{r}) + \frac{2}{A}\int_A \left[M_F(\mathbf{r}')M_B(\mathbf{r}'+\mathbf{r})\right] d\mathbf{r}' \end{aligned} \tag{3.31}$$

where $C_{F/B}(\mathbf{r})$ is the autocorrelation function of the process $M_{F/B}(\mathbf{r})$. We now write

$$M_F(\mathbf{r}) = M_{F0} + M_{F,z}(\mathbf{r}) \qquad M_B(\mathbf{r}) = M_{B0} + M_{B,z}(\mathbf{r})$$

where $M_{F,z}(\mathbf{r})$, $M_{B,z}(\mathbf{r})$ have by definition zero mean, and readily obtain

$$\frac{2}{A}\int_A \left[M_{F0} + M_{F,z}(\mathbf{r}')\right]\left[M_{B0} + M_{B,z}(\mathbf{r}'+\mathbf{r})\right] d\mathbf{r}' = 2M_{F0}M_{B0} \tag{3.32}$$

where we have used the fact that the mutual correlation function of $M_{F,z}(\mathbf{r})$, $M_{B,z}(\mathbf{r})$ is zero because they are uncorrelated and have zero mean. It should be noticed

that $2M_{F0}M_{B0}$ is just an $\mathbf{r}$ independent term, so that Eqs.3.31 and 3.32 show that, in virtue of the Wiener−Khintchine theorem, the power spectrum density of $[M_F(\mathbf{r})+M_B(\mathbf{r})]$ is given by the sum the M_F and M_B spectra, up to an irrelevant term related to the mean of $M_F(\mathbf{r})$ and $M_B(\mathbf{r})$, that of course gives no contribution to momentum relaxation via scattering.

For ETB devices the assumption of uncorrelated front and back interface roughness may be questionable, however the analysis of a possible correlation between $\Delta_F(\mathbf{r})$ and $\Delta_B(\mathbf{r})$ is beyond the scope of the present work.

3.2.5 Carrier screening

Eqs.3.17 and 3.23 express the unscreened squared matrix elements respectively for the PN or GPN models and the new model proposed in this work. The screening effect produced by the electron in the inversion layer was introduced by using the static, scalar dielectric function $\epsilon_D(\mathbf{q})$. This approach is fairly standard and widely discussed in the literature [24, 40], so that here we only summarize the basic equations. The dielectric function $\epsilon_D(\mathbf{q})$ can be calculated as [20, 24, 40]

$$\epsilon_D(\mathbf{q}) = 1 - \sum_{\nu,n} \frac{e^2}{q(\epsilon_{sct}+\epsilon_{ox})} F_{\nu,n,n}(q)\,\Pi_{\nu,n,n}(\mathbf{q}) \tag{3.33}$$

where the sum is over subbands n and valleys ν (for inversion layers having more valleys), and the form factor $F_{\nu,n,n}(q)$ and polarization factor $\Pi_{\nu,n,n}(\mathbf{q})$ are defined in [20, 40]. Consistently with a scalar dielectric function approach, the inter-subband transitions were left unscreened while the screened matrix element for intra-subband transitions is simply given by $\mathcal{M}^{(scr)}_{n,n}(\mathbf{q})=\mathcal{M}_{n,n}(\mathbf{q})/\epsilon_D(\mathbf{q})$.

3.3 Mobility calculations: planar MOSFETs

In mobility calculations only the Δ valleys are considered for silicon and only the Γ valley for III-V semiconductors. In fact the simulations for III-V semiconductors are focused on InAs in a (100) inversion layer, where the contribution to transport of satellite valleys is essentially negligible down to very small well thicknesses [16, 41].

3.3.1 Non-parabolicity corrections for mobility calculations

As explained in Sec.3.2 the SR scattering matrix element of Eq.3.20 has been derived within the parabolic effective mass approximation (EMA) Hamiltonian. While the development of a complete SR scattering model based on a Hamiltonian beyond the EMA approach goes beyond the scope of the present work [37], we introduced non parabolic corrections in the calculation of the subband minima and in plane energy relation, which in turn affect the calculation of mobility, as discussed in Sec.3.3.2. The non parabolicity factor α in the conduction band of silicon is about $0.5eV^{-1}$, but it becomes substantially larger at the Γ conduction band of $In_{0.53}Ga_{0.47}As$ (α=$1.4eV^{-1}$)

and InAs (α=2.5eV^{-1}) [42–44]. The non parabolicity effects are included in our model by writing the energy E(**k**) versus the wave-vector **k**=(k_x,k_y) as in Eq.C.31. As can be seen, Eq.C.31 provides nonparabolicity corrections for both the subband minima (obtained by setting k_x and k_y to 0 into Eq.C.31) and the in-plane kinetic energy $E(\mathbf{k})$. In particular, the transport mass becomes dependent on size and bias induced quantization and, consequently, dependent on the well thickness and different in each subband [44].

3.3.2 Relaxation time and mobility calculation

A rigorous calculation of the **k** dependent momentum relaxation time $\tau_n(\mathbf{k})$ for an anisotropic scattering mechanism, such as surface roughness, and for anisotropic bands is a quite complicated problem, and it is equivalent to solving the linearized Boltzmann transport equation for a small, uniform electric field. Even if a general solution to the problem can be obtained for anisotropic scattering mechanisms and anisotropic bands [45], a simplification is introduced in this work by assuming isotropic bands, in which case the relaxation time $\tau_n(E)$ depends only on the energy E and not on the direction of **k**[45].

In a silicon (100) inversion layer and within the parabolic effective mass approximation, the isotropic dispersion is essentially correct for unprimed subbands (having quantization mass m_{sct}=0.916 m_0 and in-plane masses m_x=m_y=0.19 m_0), while it is an approximation for primed subbands (having m_{sct}=0.19 m_0 and different in-plane masses 0.19 m_0 and 0.916m_0). Moreover, isotropic bands are a very good approximation for the Γ point conduction band minimum of most III-V semiconductors.

Since the inter-subband transitions couple the calculation of the momentum relaxation time $\tau_n(E)$ in the different subbands (within a given valley), we determined $\tau_n(E)$ by solving the corresponding set of algebraic equations as described in [40]. We thus write the wave-vector $\mathbf{k}_n = (k_n,\beta_n)$ in polar coordinates in the subband n and for an energy E, where k_n is obtained from Eq.C.31 as [21]

$$k_n(E) = \frac{1}{\hbar}\sqrt{2m_d(E - \varepsilon_n + \alpha(E - \langle U_n\rangle)^2)} \tag{3.34}$$

with $m_d = \sqrt{m_x\, m_y}$ being the density of state effective mass. Then the magnitude q of **q**= ($\mathbf{k}_n - \mathbf{k}_{n'}$) can be written as

$$q(\theta, E) = \sqrt{k_n^2(E) + k_{n'}^2(E) - 2k_n(E)k_{n'}(E)\cos\theta} \tag{3.35}$$

hence it depends on E and on the angle θ=($\beta_{n'}$−β_n) between $\mathbf{k}_n$ and $\mathbf{k}_{n'}$. Since surface roughness is an elastic scattering mechanism, the relaxation time $\tau_n(E)$, calculated by using the first Born approximation [46], is implicitly defined by the

equation [20, 40]

$$\sum_{n'}\Theta(E-\varepsilon_{n'}^{np})\frac{m_d}{2\pi\hbar^3}\int_0^{2\pi}\left[1+2\alpha(E-\langle U_{n'}\rangle)\right]\left\langle|\mathcal{M}_{nn'}^{(scr)}(q(\theta,E))|^2\right\rangle\times$$
$$\times\left[\tau_n(E)\,k_n(E)-\tau_{n'}(E)k_{n'}(E)\frac{1+2\alpha(E-\langle U_n\rangle)}{1+2\alpha(E-\langle U_{n'}\rangle)}\cos\theta\right]d\theta=k_n(E) \quad (3.36)$$

Where $\varepsilon_{n'}^{np}$ is the nonparabolic eigenvalue of the n'-th subband obtained from Eq.C.31 by setting k_x and k_y to 0, and $\mathcal{M}_{nn'}^{(scr)}(q(\theta,E))$ is the screened matrix element.

Since the $\tau_n(E)$ does not depend on the integration angle θ for isotropic bands, then for any energy E the set of $\tau_n(E)$ can be obtained from Eq.3.36 by solving the algebraic linear system [40]

$$\tau_n(E)\ \sum_{n'}A_{nn'}\,(E)-\sum_{n'}B_{nn'}\,(E)\ \tau_{n'}(E)\ =k_n(E) \quad (3.37)$$

where the coefficients $A_{nn'}$ and $B_{nn'}$ are given by

$$A_{nn'}\,(E)=\Theta(E-\varepsilon_{n'}^{np})\frac{m_d}{2\pi\hbar^3}\left[1+2\alpha(E-\langle U_{n'}\rangle)\right]k_n\,(E)\int_0^{2\pi}\left\langle|\mathcal{M}_{nn'}^{(scr)}(q(\theta,E))|^2\right\rangle\,d\theta \quad (3.38)$$

$$B_{nn'}\,(E)=\Theta(E-\varepsilon_{n'}^{np})\frac{m_d}{2\pi\hbar^3}\left[1+2\alpha(E-\langle U_n\rangle)\right]k_{n'}\,(E)\int_0^{2\pi}\left\langle|\mathcal{M}_{nn'}^{(scr)}(q(\theta,E))|^2\right\rangle\cos\theta\,d\theta. \quad (3.39)$$

Once the $\tau_n(E)$ have been determined, the electron mobility evaluated without accounting for the valley multiplicity is finally obtained as [20, 40]

$$\mu_n=\frac{e\,n_{sp}}{N_n\,2\,\pi\,\hbar^2}\left(\frac{m_d}{m_c}\right)^2\int_{\varepsilon_n^{np}}^{\infty}\frac{(E-\varepsilon_n+\alpha(E-\langle U_n\rangle)^2)}{1+2\alpha(E-\langle U_n\rangle)}\,\tau_n(E)\left|\frac{\partial f_0(E)}{\partial E}\right|dE \quad (3.40)$$

where $n_{sp}=2$ is the spin degeneracy, $m_c=2(m_x^{-1}+m_y^{-1})^{-1}$ is the conduction mass, $f_0(E)$ is the equilibrium Fermi occupation function, and N_n is the electron inversion density in the n-th subband by setting the valley multiplicity to 1. Finally the effective mobility in the inversion layer is obtained as the average of the subband mobilities weighted by the corresponding electron densities.

3.4 Simulation results: planar MOSFETs

In this section we present the results for SR-limited mobility in silicon and InAs inversion layers obtained using the new model of this work as well as the GPN and the PN models. In the calculations, beside the terms described in Sec.3.2.2, we included also the contributions to the SR scattering matrix element due to the perturbations of the electrostatic potential produced by the fluctuations of electron and polarization charges [19, 31]. For these terms we employed the formulation reported in [19].

3.4.1 Silicon MOSFETs

Fig.3.4 compares simulation results and experimental data for the mobility versus the effective field, E_{eff}, in bulk silicon MOSFETs at T=77K. The low temperature allows us to investigate an experimental condition where the dominant scattering mechanism at large E_{eff} is surface roughness. As already said in Sec.3.2.2, in

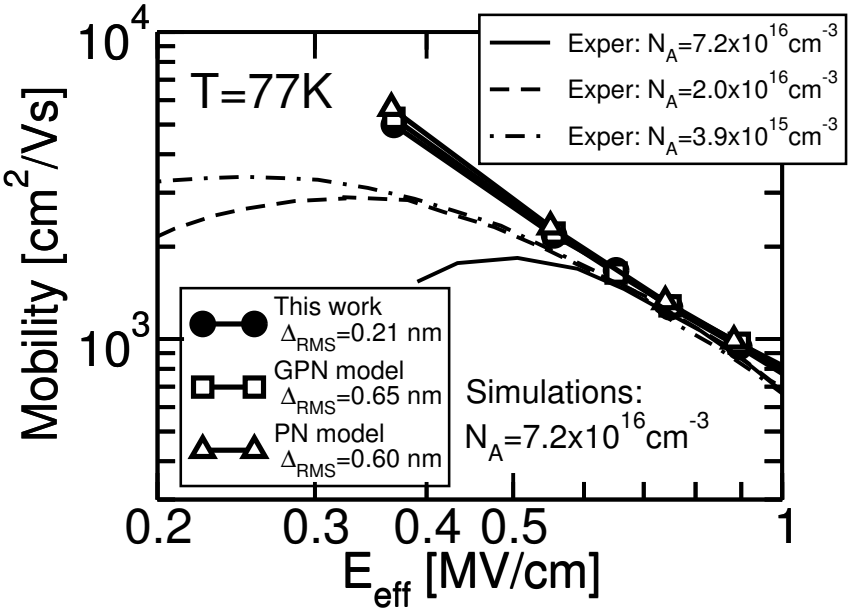


Figure 3.4: SR-limited mobility obtained with the matrix elements given by the model of this work (i.e. Eq.3.20), by the GPN model (i.e. Eq.3.13), or by the PN model (i.e Eq.3.14). All simulations are performed using a SR correlation length Λ =1.4 nm. The corresponding $\Delta_{\text{r.m.s.}}$ values are reported in the legend. Symbols are experimental data from [47].

all calculations we assume the exponential form for the roughness auto-correlation function given by Eq.3.22 [23], and the correlation length Λ is set to 1.4nm. As can be seen the SR scattering model of this work can reproduce the experimental data with a $\Delta_{\text{r.m.s.}}$ of 0.21 nm. It is worth noting that such a $\Delta_{\text{r.m.s.}}$ value is very close to the r.m.s. interface roughness values reported by several AFM and TEM measurements and summarized in Tab.3.1. Furthermore, $\Delta_{\text{r.m.s.}}$=0.21 nm corresponds to approximately one mono-layer of interface width. Fig.3.4 also shows that with the GPN and the PN models, instead, we must use $\Delta_{\text{r.m.s.}}$=0.65 nm and $\Delta_{\text{r.m.s.}}$=0.60 nm, respectively, to reproduce the same experimental mobility data. Tab.3.1 shows that the $\Delta_{\text{r.m.s.}}$ for the GPN and the PN model extracted in Fig.3.4 are consistent with the values inferred in several previous simulation studies based on the GPN or PN model, but systematically larger than those reported by AFM and TEM measurements.

To gain a further insight about the behavior of the models, Fig.3.5 compares the lowest subband matrix element versus Δ and the ensemble averaged, squared matrix

element versus q for the GPN model and the model of this work for the cases with E_{eff} ≈0.9 MV/cm in Fig.3.4. As can be seen, despite the different matrix element versus Δ in Fig.3.5(a), the $\langle|\mathcal{M}_{0,0}(q)|^2\rangle$ entering directly the mobility calculation and reported in Fig.3.5(b) is very similar for the two models, which is consistent with the fact that the $\Delta_{\text{r.m.s.}}$ has been calibrated in each model to reproduce the same experimental mobility data, as illustrated in Fig.3.4.

Fig.3.6 reports the simulated SR mobility versus T_w at room temperature for Si-SiO_2 SOI single gate (SG) MOSFETs obtained with the GPN, the PN and the SR scattering model of this work. When the $\Delta_{\text{r.m.s.}}$ values calibrated in Fig.3.4 are used, the three models provide consistent trends of μ_{SR} versus T_w. Moreover, Fig.3.6 shows that for T_w lower than approximately 4 nm and at small inversion densities all models approach the well known T_w^6 mobility dependence [10, 19]. The T_w^6 behavior can be easily derived and justified analytically for the PN model relying on an infinite barrier Φ_B quantization model [20]. The fact that the GPN model and the model of this work, both of which account for a finite Φ_B, result in essentially the same T_w^6 trend confirms that in the Si-SiO_2 system the wave-function penetration into the oxide plays a modest role.

3.4.2 III-V based MOSFETs

In Fig.3.7 we present the calculated SR-limited mobility for InAs SOI SG MOSFETs versus T_w at room temperature, obtained with the same parameters of the SR spectrum used for silicon in Fig.3.6.

Simulation results from the new model and the GPN model accounting for wave-function penetration into the oxide show similar mobility results but for different values of $\Delta_{\text{r.m.s.}}$. In particular, the new model and the GPN model show a much weaker mobility degradation at small T_w than the PN model, which, as expected, predicts the T_w^6 trend.

The discrepancy between SR-limited mobility predicted by the PN compared to either the GPN or the model of this work stems from the importance of wave-function penetration into the oxide in this InAs-HfO_2 structure (see Fig.3.1.a), which is a clear difference compared to the Si-SiO_2 transistor analyzed in Fig.3.6.

3.5 Conclusions

We have revisited the formulation of surface roughness scattering for an effective mass approximation Hamiltonian and in the framework of the significant wave-function penetration into the oxide region occurring particularly in high-κ, III-V based MOS transistors.

Besides reconsidering the derivation of the well-known GPN model when the semiconductor and oxide masses are different, we have proposed a new SR scattering model that accounts for the markedly non linear relation between the displacement Δ of the interface position and the SR scattering matrix elements. To our best

knowledge, this is the first attempt to account for such a non linear relation in semi-classical SR scattering modeling, and we show that, since the deviations from a linear model are in fact large, then the proper inclusion of the non linearity remarkably influences the r.m.s value $\Delta_{\text{r.m.s.}}$ of the interface roughness that is necessary to reproduce SR-limited mobility measurements.

In this respect, comparison with experimental mobility for bulk Si MOSFETs at low temperature shows that with the new SR scattering model a good agreement with measured mobility can be obtained with $\Delta_{\text{r.m.s.}}$ values of about 0.2 nm. This is a very plausible $\Delta_{\text{r.m.s.}}$ value because it is in good agreement with several AFM and TEM measurements and, furthermore, it is close to one mono-layer of interface width, as already noticed in [23]. The $\Delta_{\text{r.m.s.}}$ value extracted by using the GPN model and the PN model to reproduce the same set of experimental data is about 0.60-0.65 nm, which is consistent with previous studies based on the PN formulation, but significantly larger than the $\Delta_{\text{r.m.s.}}$ obtained from AFM and TEM measurements.

The results of the new model, the PN and of the GPN model are also compared for SR-limited mobility in silicon and InAs thin-body transistors. For the Si-SiO_2 system all models provide very similar mobility dependence on T_w if we use, for each model, the $\Delta_{\text{r.m.s.}}$ value calibrated by comparing to experimental mobility in bulk silicon MOSFETs. This reflects the weak wave-function penetration into the Si-SiO_2 system.

For the InAs-HfO_2 system, instead, the PN model predicts a much stronger mobility degradation at small T_w than observed with either the GPN or the model of this work. In particular the PN model predicts a T_w^6 mobility behavior also in the InAs-HfO_2 system, that is not observed in the simulations obtained with the GPN or the model of this work. This significant discrepancy, that implies quite different projections for the SR-limited mobility in ultra-thin, III-V based transistors obtained with the different models, stems from the substantially larger wave-function penetration into the oxide region observed in the InAs-HfO_2 compared to the conventional Si-SiO_2 system. This new SR model has been implemented in a comprehensive semi-classical multi-subband Monte Carlo simulator with a parallel implementation [28, 48], that accounts for all the main scattering mechanisms and has confirmed the results here obtained for the SR limited mobility [49].

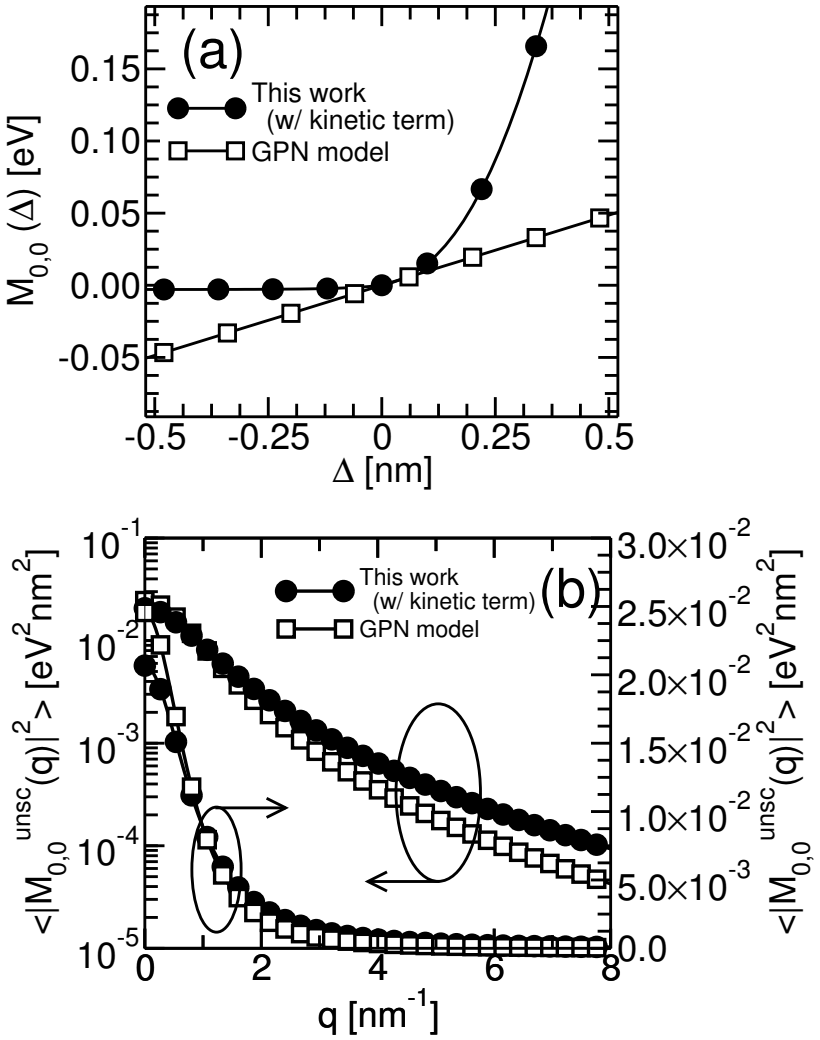


Figure 3.5: Unscreened matrix element of the lowest subband for the GPN and the model of this work for the bulk MOSFET analyzed in Fig.3.4 at E_{eff} ≈0.9 MV/cm. *(a)*: intrasubband matrix element versus Δ for the lowest subband. *(b)*: Ensemble averaged, squared intrasubband matrix element for the lowest subband versus q= $|\mathbf{q}|$.

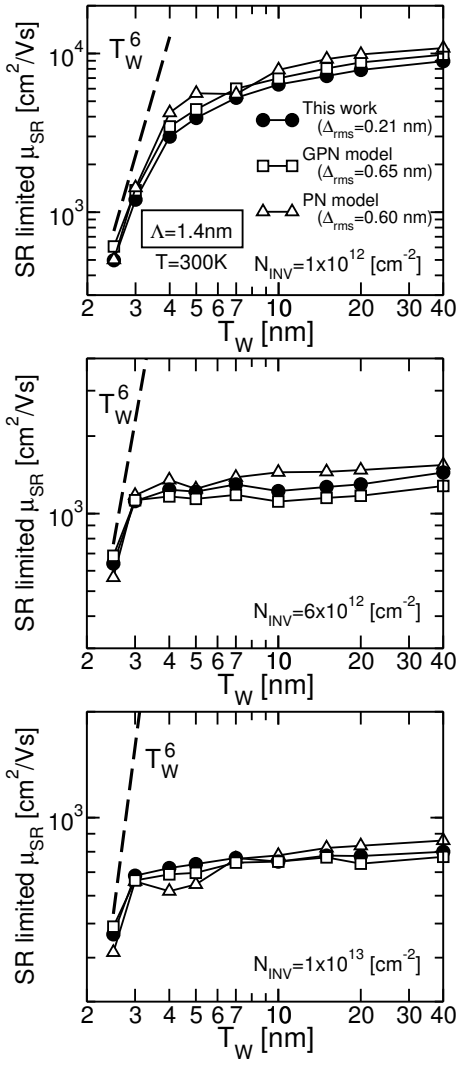


Figure 3.6: Simulated SR-limited mobility at room temperature in Si-SiO_2 SOI SG MOSFETs versus semiconductor thickness T_w for three inversion densities. Electron mobility is calculated using the GPN model, the PN model and the new model of this work. The doping concentration in the silicon film is $10^{15}cm^{-3}$.

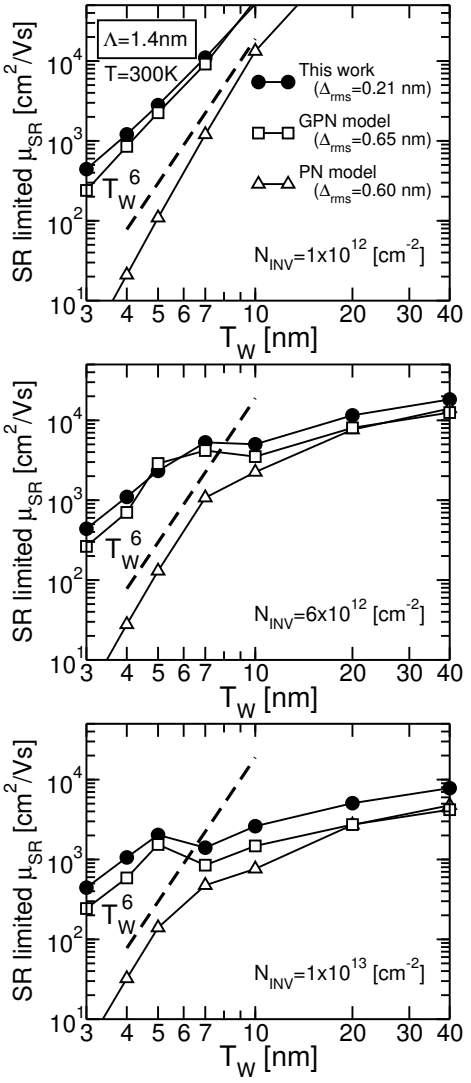


Figure 3.7: Simulated SR-limited mobility at room temperature in InAs-HfO_2 SOI SG MOSFETs versus T_w for three inversion densities. Electron mobility is calculated using the GPN model, the PN model and the new model of this work. Doping concentration is $10^{15}cm^{-3}$.

3.6 Modeling of surface roughness scattering: GAA MOSFETs

The aim of this section is to extend to circular GAA nanowire MOSFETs the non-linear model for SR scattering developed for planar transistors in [1]. Let us assume that, as shown in Appendix B.2, the envelope wave-function can be written in the form:

$$\Phi_{n,k_x}(r,\theta,x) = \frac{e^{i\,k_x x}}{\sqrt{L_{NW}}}\xi_n(r,\theta) \tag{3.41}$$

that corresponds to Eq.B.31 neglecting the term $e^{i\,k_x\,(\alpha y+\beta z)}$ as discussed in Sec.4.4.2. The unscreened SR matrix element for a 1D gas is written according to the definition used in Eqs.3.7,3.8:

$$\mathcal{M}_n^{n'}(q_x) = \frac{1}{L_{NW}}\int\limits_{L_{NW}}\left[\int\limits_{-\pi}^{\pi} d\theta \int\limits_{0}^{\infty} r\,\xi_{n'}^{\dagger}(r,\theta)\left[\left(\widehat{H}_{p,r\theta x}-\widehat{H}_0\right)\xi_n(r,\theta)\right]dr\right] e^{-i\,q_x x}dx \tag{3.42}$$

where $q_x = (k'_x - k_x)$, $\widehat{H}_0$ is the unperturbed effective mass Hamiltonian and $\widehat{H}_{p,r\theta x}$ is the Hamiltonian perturbed by the presence of surface roughness.

For the general case of anisotropic bands, the expression for the unperturbed Hamiltonian is (see Eq.4.1):

$$\widehat{H}_0 = -\frac{\hbar}{2}\left[w_{rr}\frac{\partial^2}{\partial r^2} + w_r\frac{\partial^2}{\partial r} + w_{r\theta}\frac{\partial^2}{\partial r\partial\theta} + w_\theta\frac{\partial^2}{\partial\theta} + w_{\theta\theta}\frac{\partial^2}{\partial\theta^2}\right] - e\phi(r,\theta) + \Phi_B\Theta(r-r_{NW}) \tag{3.43}$$

with $\phi(r,\theta)$ being the electrostatic potential, Φ_B is the semiconductor-oxide energy barrier, $\Theta(\bullet)$ is the unit step function, and r_{NW} is the semiconductor nanowire radius. Since the w_{ij} terms of Eq.3.43 are different in the semiconductor and oxide materials, we write that:

$$w_{ij} = w_{ij,sct}\Theta\left[r_{NW}-r\right] + w_{ij,ox}\Theta\left[r-r_{NW}\right]. \tag{3.44}$$

where the w_{ij} terms are defined in Eq.4.2. Similarly, $\widehat{H}_{p,r\theta x}$ is defined as:

$$\begin{aligned}\widehat{H}_{p,r\theta x} = -\frac{\hbar}{2}\left[w_{rr}\frac{\partial^2}{\partial r^2} + w_r\frac{\partial^2}{\partial r} + w_{r\theta}\frac{\partial^2}{\partial r\partial\theta} + w_\theta\frac{\partial^2}{\partial\theta} + w_{\theta\theta}\frac{\partial^2}{\partial\theta^2}\right] - e\phi(r) + \\ + \Phi_B\Theta\left(r-\left(r_{NW}-\Delta\left(\theta,x\right)\right)\right)\end{aligned} \tag{3.45}$$

and

$$w_{ij} = w_{ij,sct}\Theta\left[\left(r_{NW}-\Delta\left(\theta,x\right)\right)-r\right] + w_{ij,ox}\Theta\left[r-\left(r_{NW}-\Delta\left(\theta,x\right)\right)\right]. \tag{3.46}$$

In this thesis we consider that the oxide has isotropic energy bands, hence, Eqs.3.43 and 3.45 when written for the oxide domain have $w_{r\theta,ox} = w_{\theta,ox} = 0$. The difference

between the perturbed (Eq.3.45) and unperturbed (Eq.3.43) Hamiltonians leads to a quite complicated operator not reported here explicitly. However it can be easily seen that the perturbation consists of two terms: one contribution is proportional to the potential energy barrier Φ_B and the second term is given by the kinetic parts of the Hamiltonian that we here denote with the symbol $\nabla\widehat{H}_{kin}$, and consists of the radial, angular and mixed derivatives. In order to proceed further and discuss on these two terms, we analyze below the case for isotropic quantization masses.

In this case, according to Appendix A, $w_{ij} = 0$ with $i \neq j$, hence, the unperturbed Hamiltonian according to Eq.4.1 and Eq.4.2 reads:

$$\widehat{H}_0 = \left[-\frac{\hbar^2}{2r}\frac{\partial}{\partial r}\frac{r}{m_q}\frac{\partial}{\partial r} - \frac{\hbar^2}{2r^2 m_q}\frac{\partial^2}{\partial\theta^2} + e\phi(r) + \Phi_B\Theta(r - r_{NW}) \right] \tag{3.47}$$

where m_q is the quantization mass defined as:

$$m_q = m_{sct}\Theta\left(r_{NW} - r\right) + m_{ox}\Theta\left(r - r_{NW}\right). \tag{3.48}$$

Similarly, $\widehat{H}_{p,r\theta x}$ is defined as:

$$\widehat{H}_{p,r\theta x} = \left[-\frac{\hbar^2}{2r}\frac{\partial}{\partial r}\frac{r}{m_q}\frac{\partial}{\partial r} - \frac{\hbar^2}{2r^2 m_q}\frac{\partial^2}{\partial\theta^2} + e\phi(r) + \Phi_B\Theta\left(r - \left(r_{NW} - \Delta\left(\theta, x\right)\right)\right) \right] \tag{3.49}$$

and

$$m_q = m_{sct}\Theta\left[\left(r_{NW} - \Delta\left(\theta, x\right)\right) - r\right] + m_{ox}\Theta\left[r - \left(r_{NW} - \Delta\left(\theta, x\right)\right)\right]. \tag{3.50}$$

The term $\left(\widehat{H}_{p,r\theta x} - \widehat{H}_0\right)\xi_n(r,\theta)$ in Eq.3.42 is non-null only in the radial domain between $[r_{NW} - \Delta(\theta, x)]$ and r_{NW}. The solution of the Schrödinger equation with the unperturbed Hamiltonian in Eq.3.47 is given by the envelope wave-function reported in Eq.3.54. Hence, Eq.3.42 can be written as

$$\left(\widehat{H}_{p,r\theta x} - \widehat{H}_0\right)\xi_n(r,\theta) = \pm\Phi_B\rho_{n,l}(r)e^{il\theta} + \underbrace{\pm\left(-\frac{\hbar^2}{2m_{ox}} + \frac{\hbar^2}{2m_{sct}}\right)\left[\frac{\partial^2}{\partial r^2}\rho_{n,l}(r) + \frac{1}{r}\frac{\partial}{\partial r}\rho_{n,l}(r) - \frac{l^2}{r}\rho_{n,l}(r)\right]e^{il\theta}}_{\nabla\widehat{H}_{kin}}. \tag{3.51}$$

Eq.3.51 is valid only between $[r_{NW} - \Delta(\theta, x)]$ and r_{NW} and it is zero otherwise. In Eq.3.51 the $\pm$ term is positive for positive $\Delta(\theta, z)$ (i.e. when the barrier Φ_B penetrates into the semiconductor) and negative for negative $\Delta(\theta, x)$ (i.e. when the barrier Φ_B penetrates the oxide).

Since $\Delta(\theta, x)$ can be regarded as a stochastic process that occurs along the NW surface, we now define a new coordinate $\mathbf{s} = (z, x)$ as the coordinate along the NW surface with $z = r_{NW}\theta$ as shown in Fig.3.8. Hence the roughness at the

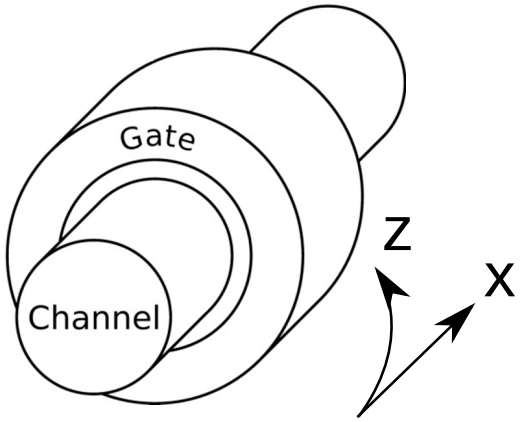


Figure 3.8: Sketch of a circular NW.

semiconductor interface can be treated as a stochastic process $\Delta(\mathbf{s})$ dependent on the coordinate $\mathbf{s}$. The matrix element of Eq.3.42 can be finally written showing explicitly the contribution due to the potential barrier Φ_B and the kinetic term $\nabla\widehat{H}_{kin}$:

$$\mathcal{M}_n^{n'}(q_x) = \frac{1}{L_{NW}} \int_{L_{NW}} \left[\int_{-\pi}^{\pi} d\theta \int_{r_{NW}-\Delta(\mathbf{s})}^{r_{NW}} r\, \xi_{n'}^{\dagger}(r,\theta) \left[\Phi_B \; + \nabla\widehat{H}_{kin} \right] \xi_n(r,\theta)\, dr \right] e^{-i\, q_x x} dx. \tag{3.52}$$

where, in the case of isotropic quantization masses, the expression for $\xi_n(r,\theta)$ is given in Eq.3.54, and where the term $\Delta(\mathbf{s})$ can be either positive or negative.

It has been thoroughly discussed in [1] and in Sec.3.2.3 that, depending on m_{ox} and m_{sct}, the $\nabla\widehat{H}_{kin}$ term in Eq.3.52 can lead to non-physical results. Moreover, in general, the kinetic term depends on (r,θ) and cannot be separated in $r-$ and $\theta-$dependent functions, hence, the integral over r in Eq.3.52 becomes a non-linear and space-variant transformation of $\Delta(\mathbf{s})$ leading to an intractable problem. Based upon these considerations, $\nabla\widehat{H}_{kin}$ will be neglected in the following calculations.

By neglecting $\nabla\widehat{H}_{kin}$ the unscreened SR matrix element for a 1D gas given by Eq.3.52 can be written as:

$$\mathcal{M}_n^{n'}(q_x) = \frac{1}{L_{NW}} \int_{L_{NW}} \left[\int_{-\pi\, r_{NW}}^{\pi\, r_{NW}} \frac{dz}{r_{NW}} \int_{r_{NW}-\Delta(\mathbf{s})}^{r_{NW}} r\, \xi_{n'}^{\dagger}\left(r, \frac{z}{r_{NW}}\right) \Phi_B \xi_n\left(r, \frac{z}{r_{NW}}\right) dr \right] e^{-i\, q_x x} dx. \tag{3.53}$$

Before proceeding with the derivation of the explicit expressions for the matrix element used for scattering rate calculations, we want to noticed that the equation for the SR matrix element reported in Eq.3.53 applies to both isotropic or anisotropic quantization masses. Throughout this work the $\xi_n(r, z/r_{NW})$ functions have been calculated by using the unperturbed Hamiltonians that accounts for the crystal orientation given by Eq.4.1.

3.6.1 Isotropic quantization masses

In the assumption of isotropic quantization masses, the term $\xi_n(r,\theta)$ in Eq.3.41 is written exploiting the separability of the radial and angular components:

$$\xi_{nl}(r,\theta) = \frac{1}{\sqrt{2\pi}}\rho_{nl}(r)e^{i\,l\theta} \tag{3.54}$$

where n and l indexes in Eq.3.54 are the radial and angular quantum numbers with $n = 1, 2, \ldots$ and $l = 0, \pm1, \pm2, \ldots$. By substituting Eq.3.54 into Eq.3.53 we obtain:

$$\mathcal{M}_{nl}^{n'l'}(q_x) = \frac{1}{L_{NW}}\int\limits_{L_{NW}} \Bigg[\int\limits_{-\pi\, r_{NW}}^{\pi\, r_{NW}} \frac{dz}{2\pi r_{NW}} e^{-i\,(l'-l)\frac{z}{r_{NW}}} \times \underbrace{\times \int\limits_{r_{NW}-\Delta(\mathbf{s})}^{r_{NW}} r\,\rho^{\dagger}_{n'l'}(r)\Phi_B\ \rho_{nl}(r)\,dr}_{M_{nl}^{n'l'}[\Delta(\mathbf{s})]}\Bigg] e^{-i\,q_x x}dx. \tag{3.55}$$

where $M_{nl}^{n'l'}[\Delta(\mathbf{s})]$ is a space-invariant non-linear transformation of $\Delta(\mathbf{s})$ and it is similar to the transformation described by Eq.3.20 neglecting the kinetic terms. Hence, the expression of the matrix element becomes:

$$\mathcal{M}_{nl}^{n'l'}(q_x) = \frac{1}{2\pi r_{NW} L_{NW}}\int\limits_{L_{NW}} \left[\int\limits_{-\pi\, r_{NW}}^{\pi\, r_{NW}} dz\, e^{-i\,q_{l'l}z}\, M_{nl}^{n'l'}\left[\Delta(\mathbf{s})\right]\right] e^{-i\,q_x x}\,dx \tag{3.56}$$

where $q_{l'l} = (l'-l)/r_{NW}$. The ensemble averaged squared matrix element for scattering rate calculations is given by:

$$\left\langle \left|\mathcal{M}_{nl}^{n'l'}(q_x)\right|^2 \right\rangle = \frac{1}{\left(2\pi\, r_{NW} L_{NW}\right)^2}\iint\limits_{L_{NW}} dx\,dx' \times \times \iint\limits_{-\pi\, r_{NW}}^{\pi\, r_{NW}} dz\,dz' e^{-i\,q_{l'l}(z-z')} \underbrace{\left\langle M_{nl}^{n'l'}\left[\Delta(\mathbf{s})\right]\left(M_{nl}^{n'l'}\left[\Delta(\mathbf{s}')\right]\right)^{\dagger}\right\rangle}_{C_{nl}^{n'l'}(\tau)} e^{-i\,q_x(x-x')} \tag{3.57}$$

where $\boldsymbol{\tau} = \mathbf{s} - \mathbf{s}' = (x-x', z-z')$ and $C_{nl}^{n'l'}(\tau)$ is the autocorrelation function of $M_{nl}^{n'l'}[\Delta(\mathbf{s})]$. By exploiting the Wiener-Kinchin theorem we write the autocorrelation function as the inverse Fourier transform of the power spectrum $S_{nl}^{n'l'}(\mathbf{q})$:

$$C_{nl}^{n'l'}(\tau) = \frac{1}{(2\pi)^2}\int\limits_{q_z} dq_z \int\limits_{q_{x'}} dq_{x'} S_{nl}^{n'l'}(q_z, q_{x'}) e^{+i\,q_z(z-z')} e^{+i\,q_{x'}(x-x')} \tag{3.58}$$

and, by substituting Eq.3.58 into the ensemble averaged squared matrix element in Eq.3.57, we obtain:

$$\left\langle \left| \mathcal{M}_{nl}^{n'l'}(q_x) \right|^2 \right\rangle = \frac{1}{\left(2\pi\, r_{NW} L_{NW}\right)^2 (2\pi)^2} \int_{q_z} dq_z \int_{q_{x'}} dq'_{x'} \times \iint_{L_{NW}} dx\, dx' e^{i(q_{x'}-q_x)(x-x')} \iint_{-\pi r_{NW}}^{\pi r_{NW}} dz\, dz' e^{i\,(q_z - q_{l'l})(z-z')} S_{nl}^{n'l'}(q_z, q_{x'}). \tag{3.59}$$

According to Eq.E.1 we can rewrite the integrals over x and x' as:

$$\begin{aligned} &\iint_{L_{NW}} dx\, dx' e^{i(q_{x'}-q_x)(x-x')} = \\ &= \int_{-L_{NW}/2}^{+L_{NW}/2} e^{ix(q_{x'}-q_x)} dx \int_{-L_{NW}/2}^{+L_{NW}/2} e^{-ix'(q_{x'}-q_x)} dx' = \\ &= L_{NW}^2 \left[sinc\left(\frac{L_{NW}}{2} (q_{x'} - q_x) \right) sinc\left(\frac{L_{NW}}{2} (q_{x'} - q_x) \right) \right] = \\ &= 2\pi L_{NW} \left[\frac{L_{NW}}{2\pi} sinc^2\left(\frac{L_{NW}}{2} (q_{x'} - q_x) \right) \right]. \end{aligned} \tag{3.60}$$

where $sinc(x) = sin(x)/x$ as in Appendix E. In the limit for $L_{NW} \to \infty$, Eq.3.60 can be rewritten by using Eq.E.4 as:

$$\iint_{L_{NW}} dx\, dx' e^{i(q_{x'}-q_x)(x-x')} = 2\pi L_{NW} \delta(q_x - q_{x'}) \tag{3.61}$$

Similar calculations can be done for the integral over dz and $d_{z'}$ of Eq.3.59

$$\iint_{-\pi r_{NW}}^{\pi r_{NW}} dz\, dz' e^{i\,(q_z - q_{l'l})(z-z')} = (2\pi)^2 r_{NW} \left[r_{NW} sinc^2\left(\pi r_{NW} \left(q_{l'l} - q_z \right) \right) \right] \tag{3.62}$$

Hence, the ensemble squared matrix element in Eq.3.57 can be rewritten by using Eqs.3.60 and 3.62 in a compact form as:

$$\left\langle \left| \mathcal{M}_{nl}^{n'l'}(q_x) \right|^2 \right\rangle = \frac{1}{2\pi\, r_{NW} L_{NW}} \int_{q_z} dq_z F_1\left(q_{l'l}, q_z\right) S_{nl}^{n'l'}(q_z, q_x) \tag{3.63}$$

where the form factor $F_1\left(q_{l'l}, q_z\right)$ is given by:

$$F_1\left(q_{l'l}, q_z\right) = r_{NW} sinc^2\left(\pi r_{NW} \left(q_{l'l} - q_z \right) \right) \tag{3.64}$$

and it is a normalized function, in the sense that:

$$\int_{q_z} F_1\left(q_{l'l}, q_z\right) dq_z = 1. \tag{3.65}$$

Eq.3.63 shows a quite compact formulation where the only approximation to move from Eq.3.59 to Eq.3.63 has been to consider the limit for $L_{NW} \to \infty$, which is a reasonable assumption adopted in this work. Moreover, the formulation in Eq.3.63 for SR matrix elements is consistent with the non linear transformation of the stochastic process $\Delta(\mathbf{s})$ as in [1].

It can be noticed that an approximation can be obtained by noticing that the integral over q_z in Eq.3.63 may be reduced for large r_{NW} values according to Eq.E.4 and leading to $F_1\left(q_{l'l}, q_z\right) = \delta(q_{l'l} - q_z)$. The simplified version in this case reads:

$$\left\langle \left| \mathcal{M}_{nl}^{n'l'}(q_x) \right|^2 \right\rangle \approx \frac{1}{2\pi\, r_{NW} L_{NW}} S_{nl}^{n'l'}(q_{l'l}, q_{x'}). \tag{3.66}$$

The power spectrum $S_{nl}^{n'l'}(\mathbf{q})$ is calculated as in Sec.3.2.2, as the Fourier transform of the autocorrelation function of $M_{nl}^{n'l'}$ (i.e. $C_{nl}^{n'l'}(\tau)$) defined in Eq.3.21. Given the exponential form for the autocorrelation function $C_\Delta(\tau)$ of the stochastic $\Delta-$process as in Eq.3.22, $S_{nl}^{n'l'}(q)$ can be calculated as shown in Appendix F leading to:

$$S_{nl}^{n'l'}(q) = 2\pi \int_0^\infty C_{nl}^{n'l'}(\tau)\tau J_0(q\tau) d\tau \tag{3.67}$$

and $q = \sqrt{q_z^2 + q_x^2}$.

3.6.2 Anisotropic quantization masses

In the case of anisotropic quantization masses, we can exploit the fact that for a given radius r, $\xi_n(r,\theta)$ is periodic with period 2π and thus write it in form of Fourier series expansion as:

$$\xi_n(r,\theta) = \frac{1}{\sqrt{2\pi}} \sum_l \rho_{nl}(r) e^{il\theta} \quad l = 0, \pm 1, \pm 2, \ldots \tag{3.68}$$

where n is the subband index and l are the angular modes for a given wave-function associated to the $n-$th eigenvalue. The $\rho_{nl}(r)$ are complex-valued functions defined as:

$$\rho_{nl}(r) = \frac{1}{\sqrt{2\pi}} \int_{-\pi}^{+\pi} \xi_n(r,\theta) e^{-il\theta}\, d\theta. \tag{3.69}$$

Eq.3.69 implies that, since $\xi_n(r,\theta)$ is real-valued, then we have:

$$\rho_{n,-l}(r) = \rho_{n,l}^\dagger(r). \tag{3.70}$$

We now proceed as in Sec.3.6.1 and neglect the contribution of the kinetic term to the matrix element and, by substituting Eq.3.68 into Eq.3.53, we write:

$$\mathcal{M}_n^{n'}(q_x) = \frac{1}{L_{NW}} \sum_{l,l'} \int_{L_{NW}} \Bigg[\int_{-\pi\, r_{NW}}^{\pi\, r_{NW}} \frac{dz}{2\pi r_{NW}} e^{-i\,(l'-l)\frac{z}{r_{NW}}} \times \underbrace{\times \int_{r_{NW}-\Delta(\mathbf{s})}^{r_{NW}} r\, \rho^{\dagger}_{n'l'}(r) \Phi_B\ \rho_{nl}(r)\, dr}_{M_{nl}^{n'l'}[\Delta(\mathbf{s})]} \Bigg] e^{-i\, q_x x} dx. \tag{3.71}$$

where the summation is performed over the modes of the wave-function for subbands n and n'. The ensemble averaged of the squared matrix element is thus given by:

$$\left\langle \left| \mathcal{M}_n^{n'}(q_x) \right|^2 \right\rangle = \frac{1}{\left(2\pi\, r_{NW} L_{NW}\right)^2} \sum_{\substack{l,l' \\ g,g'}} \iint_{L_{NW}} dx\, dx' \times \times \int_{-\pi\, r_{NW}}^{\pi\, r_{NW}} \iint dz\, dz' e^{-i\, q_{l'l} z} e^{+i\, q_{g'g} z'} \underbrace{\left\langle M_{nl}^{n'l'}\left[\Delta(\mathbf{s})\right] \left(M_{ng}^{n'g'}\left[\Delta(\mathbf{s}')\right] \right)^{\dagger} \right\rangle}_{C_{n,lg}^{n',l'g'}(\tau)} e^{-i\, q_x (x-x')} \tag{3.72}$$

where l and g are expansion modes of the $n-$th subband and l' and g' are expansion modes of the $n'-$th subband. $q_{l'l} = (l'-l)/r_{NW}$, $q_{g'g} = (g'-g)/r_{NW}$ and $C_{n,lg}^{n',l'g'}(\tau)$ is the cross-correlation function between the matrix elements $M_{nl}^{n'l'}\left[\Delta(\mathbf{s})\right]$ and $M_{ng}^{n'g'}\left[\Delta(\mathbf{s})\right]$. The cross-correlation function is written by using the Wiener-Kinchin theorem as:

$$C_{n,lg}^{n',l'g'}(\tau) = \frac{1}{(2\pi)^2} \int_{q_z} dq_z \int_{q_{x'}} dq_{x'} S_{n,lg}^{n',l'g'}(q_z, q_{x'}) e^{i\, q_z (z-z')} e^{i\, q_{x'} (x-x')} \tag{3.73}$$

and by substituting Eq.3.73 into the ensemble squared matrix element given by Eq.3.72, we obtain:

$$\left\langle \left| \mathcal{M}_n^{n'}(q_x) \right|^2 \right\rangle = \sum_{\substack{l,l' \\ g,g'}} \frac{1}{\left(2\pi\, r_{NW} L_{NW}\right)^2 (2\pi)^2} \int_{q_z} dq_z \int_{q_{x'}} dq_{x'} \times \iint_{L_{NW}} dx\, dx' e^{-i q_{l'l} z} e^{+i q_{g'g} z'} \iint_{-\pi\, r_{NW}}^{\pi\, r_{NW}} dz\, dz' S_{n,lg}^{n',l'g'}(q_z, q_{x'}) e^{i\, q_z (z-z')} e^{i\, q_{x'} (x-x')}. \tag{3.74}$$

By following the same passages as in the case of isotropic quantization masses in Sec.3.6.1, exploiting Eq.3.60 and the properties shown in Appendix E, we obtain:

$$\left\langle \left| \mathcal{M}_n^{n'}(q_x) \right|^2 \right\rangle = \frac{1}{2\pi\, r_{NW} L_{NW}} \sum_{\substack{l,l' \\ g,g'}} \int_{q_z} dq_z F_2\left(q_{l'l}, q_{g'g}, q_z\right) S_{n,lg}^{n',l'g'}(q_z, q_x) \tag{3.75}$$

where the form factor $F_2\left(q_{l'l}, q_{g'g}, q_z\right)$ is given by:

$$F_2\left(q_{l'l}, q_{g'g}, q_z\right) = r_{NW} sinc\left(\pi r_{NW}\left(q_{l'l} - q_z\right)\right) sinc\left(\pi r_{NW}\left(q_{g'g} - q_z\right)\right). \tag{3.76}$$

The form factor $F_2\left(q_{l'l}, q_{g'g}, q_z\right)$ is not a normalized function for $q_{l'l} \neq q_{g'g}$. Moreover, it can be seen that $F_2\left(q_{l'l}, q_{l'l}, q_z\right) = F_1\left(q_{l'l}, q_z\right)$ and, in this case, it is a normalized function (see Eq.3.65). The cross-correlation power spectrum $S_{n,lg}^{n',l'g'}(q_z, q_x)$ is calculated by replacing the autocorrelation function in Eq.3.67 with the cross-correlation function. Moreover, for the calculation of Eq.3.73, Eq.3.21 must be extended to the cross-correlation case as:

$$C_{n,lg}^{n',l'g'}(\tau) = \frac{1}{2\pi\, C_\Delta(0)\sqrt{1 - C_{\Delta,N}^2(\mathbf{r})}} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} M_{nl}^{n'l'}\left[\Delta 1\right] \left[M_{ng}^{n'g'}\left[\Delta 2\right]\right]^\dagger \times$$
$$\times \exp\left[-\frac{\Delta_1^2 + \Delta_2^2 - 2C_{\Delta,N}(\mathbf{r})\Delta_1\Delta_2}{2C_\Delta(0)(1 - C_{\Delta,N}^2(\mathbf{r}))}\right] d\Delta_1\, d\Delta_2 \tag{3.77}$$

3.6.3 A few remarks about matrix elements and form factors

By assuming that $\xi_n(r,\theta)$ in Eq.3.68 is normalized as:

$$\int_0^\infty d\theta \int_{-\pi}^{\pi} dr\, r\ |\xi_n(r,\theta)|^2 = 1 \tag{3.78}$$

then Eq.3.68 implies:

$$\int_0^\infty d\theta \int_{-\pi}^{\pi} dr\, r\ |\xi_n(r,\theta)|^2 = \int_0^\infty d\theta \int_{-\pi}^{\pi} dr\, \frac{1}{2\pi} \sum_{l,l'} r\, \rho_{nl}(r)\rho_{nl'}^\dagger(r) e^{-i(l'-l)\theta}$$
$$= \sum_l \int_0^\infty dr\, r\ |\rho_{nl}(r)|^2 = 1. \tag{3.79}$$

Even though $\xi_n(r,\theta)$ is real-valued, $\rho_{nl}(r)$ can be complex and one important property is :

$$\rho_{n,-l}(r) = \rho_{n,l}^\dagger(r) \tag{3.80}$$

which stems directly from the $\rho_{n,l}(r)$ definition in Eq.3.69. The matrix element $M_{nl}^{n'l'}[\Delta(\mathbf{s})]$ of Eq.3.55 is thus in general complex-valued however it inherits the properties of $\rho_{nl}(r)$ which imply:

$$\begin{aligned} M_{n,-l}^{n',-l'}[\Delta(\mathbf{s})] &= \int_{r_{NW}-\Delta(\mathbf{s})}^{r_{NW}} r\,\rho_{n',-l'}^{\dagger}(r)\Phi_B\ \rho_{n,-l}(r)\,dr \\ &= \int_{r_{NW}-\Delta(\mathbf{s})}^{r_{NW}} r\,\rho_{n,l}^{\dagger}(r)\Phi_B\ \rho_{n',l'}(r)\,dr \\ &= \left[M_{n,l}^{n',l'}[\Delta(\mathbf{s})]\right]^{\dagger}. \end{aligned} \tag{3.81}$$

We finally noticed that, since $sinc(x)$ is an even function of x, then we have:

$$F_2\left(q_{-l',-l}, q_{-g',-g}, -q_z\right) = F_2\left(q_{l'l}, q_{g'g}, q_z\right) \tag{3.82}$$

where we used $q_{l'l} = (l'-l)/r_{NW}$ and $q_{g'g} = (g'-g)/r_{NW}$ where l, g are the modes of the subband n and l', g' the modes of the subband n' (see Eq.3.68) As a consequence:

$$\sum_{l,l'} e^{-iq_{l'l}z} M_{nl}^{n'l'}[\Delta(\mathbf{s})] \tag{3.83}$$

is real-valued, which is consistent with the fact that $\xi_n(r,\theta)$ functions are real-valued.

The cross-correlation function $C_{n,lg}^{n',l'g'}(\tau)$ defined in Eq.3.72 can take complex values and the same holds for its Fourier transform (i.e. the power spectrum $S_{n,lg}^{n',l'g'}(\mathbf{q})$), however, due to the property of the matrix element in Eq.3.81, we have:

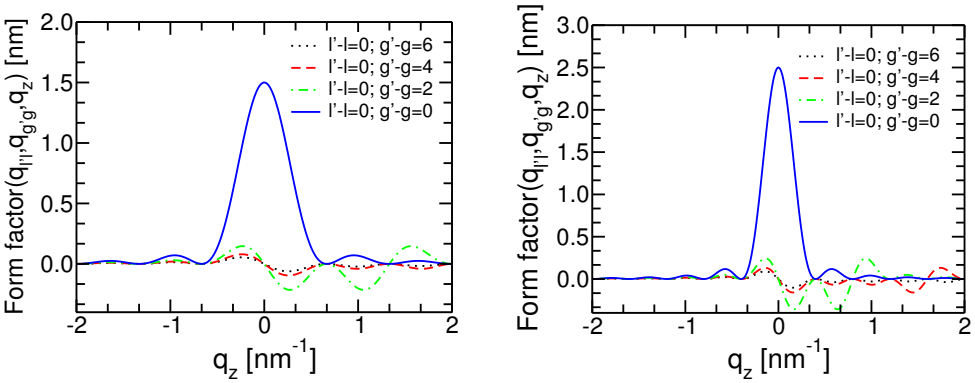
$$C_{M_{n',-l',-g'}^{n',-l',-g'}}(\tau) = \left[C_{M_{n,l,g}^{n',l',g'}}(\tau)\right]^{\dagger} \tag{3.84a}$$

$$S_{M_{n,-l,-g}^{n',-l',-g'}}(\tau) = \left[S_{M_{n,l,g}^{n',l',g'}}(\tau)\right]^{\dagger}. \tag{3.84b}$$

By using Eq.3.84b and Eq.3.82 one can readily verify that the double sum over all the couples of modes (l,l') and (g,g') assures that the ensemble averaged squared matrix element $\left\langle \left|\mathcal{M}_n^{n'}(q_x)\right|^2 \right\rangle$ in Eq.3.75 is real-valued, as it is in fact supposed to be.

In the general case, with anisotropic quantization masses, the form factor for the ensemble averaged squared matrix element is given by Eq.3.76 and the number of Form Factors is given by the possible combinations of couples of modes $((l,l'),(g,g'))$. Some considerations arise from the fact that the above form factor is given by the product of two *sinc* functions peaked respectively in $(q_{l'l}-q_z)$ and in $(q_{g'g}-q_z)$, hence, it is expected that the magnitude of those form factors for which $(l'-l) \neq (g'-g)$ is

small compared to the ones given by the couples of couples where $(l'-l)=(g'-g)$. Fig.3.9 shows two examples of form factor calculated for the system Si-SiO_2. The form factor with $(l'-l)=(g'-g)$ is the dominant one and, as expected, for larger r_{NW}, $F_2\left(q_{l'l},q_{g'g},q_z\right)$ with $(l'-l)=(g'-g)$ approaches a Dirac function in agreement with Eq.E.4. We have performed the calculation of the SR matrix elements using



(a) Simulation results obtained for r_{NW}=1.5nm, t_{ox}=0.6 nm of Si=O_2, and for the lowest subband n=n'=0.

(b) Simulation results obtained for r_{NW}=2.5nm, t_{ox}=0.6 nm of Si=O_2, and for the lowest subband n=n'=0.

Figure 3.9: Form factors obtained by using Eq.3.76 for the Si-SiO_2 system.

three different methods:

1. with no approximations for the form factor calculation, that is by using Eqs.3.75 and 3.76 as they read.

2. neglecting in the double sum of Eq.3.75 the terms having $(q_{l'l}\neq q_{g'g})$. Hence the form factor $F_2\left(q_{l'l},q_{l'l},q_z\right)$ becomes equal to $F_1\left(q_{l'l},q_z\right)$ and the ensemble averaged squared matrix element reads:

$$\left\langle\left|\mathcal{M}_n^{n'}(q_x)\right|^2\right\rangle=\frac{1}{2\pi\, r_{NW}L_{NW}}\sum_{(l'-l)=(g'-g)}\int_{q_z}dq_zF_1\left(q_{l'l},q_z\right)S_{n,lg}^{n',l'g'}(q_z,q_x) \tag{3.85}$$

3. neglecting in the double sum of Eq.3.75 the terms having $(q_{l'l}\neq q_{g'g})$ and approximating the form factor with a Dirac function in the limit of large r_{NW}, that is:

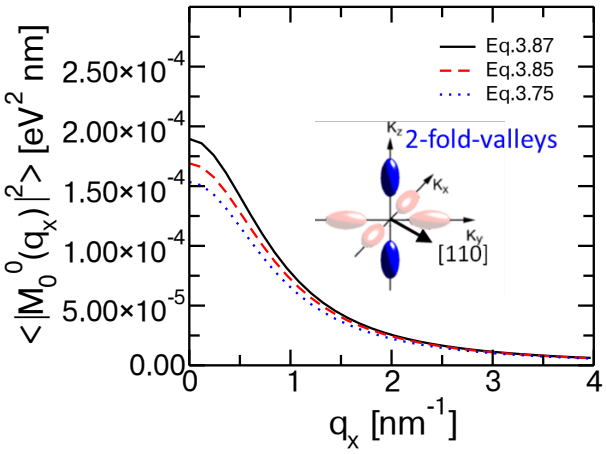
$$F_2\left(q_{l'l},q_{l'l},q_z\right)=F_1\left(q_{l'l},q_z\right)\approx\delta\left(q_{l'l}-q_z\right). \tag{3.86}$$

In this case the integral over q_z in Eq.3.75 is reduced and we are left with a sampling of the power spectrum $S_{n,lg}^{n',l'g'}(q)$ where $q=\sqrt{(q_{l'l}^2+q_x^2)}$:

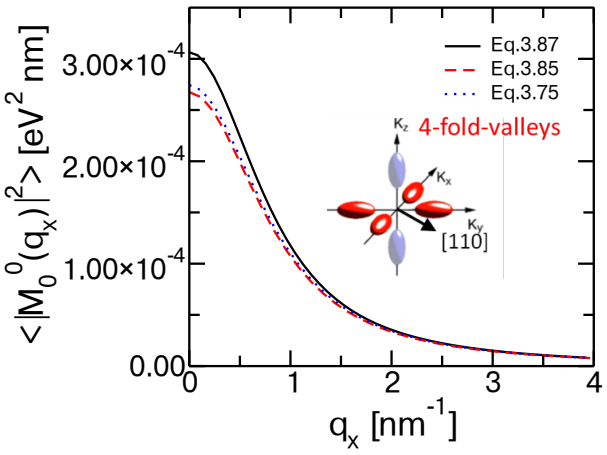
$$\left\langle\left|\mathcal{M}_n^{n'}(q_x)\right|^2\right\rangle=\frac{1}{2\pi\, r_{NW}L_{NW}}\sum_{(l'-l)=(g'-g)}S_{n,lg}^{n',l'g'}(q_{l'l},q_x). \tag{3.87}$$

This simplification in Eq.3.87 may be critical for very small r_{NW} because the formulation with Dirac function for the form factor is valid only for large r_{NW} values.

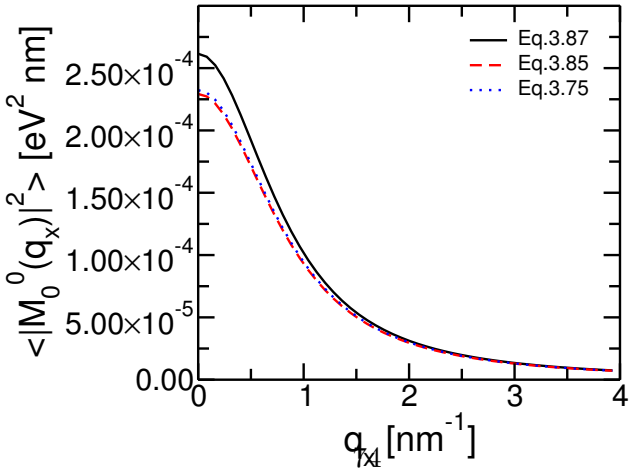
In the following examples we will show the impact of the different formulations on the ensemble averaged squared matrix element that enter the scattering rate calculation. Fig.3.10 shows some examples of $\left\langle \left| \mathcal{M}_n^{n'}(q_x) \right|^2 \right\rangle$ matrix elements for $r_{NW} = 1.5$ nm. As can be seen, even for very small r_{NW}, the approximation of the form factor obtained using the Dirac function in Eq.3.86 seems to work fairly well. In addition, the computational time of the ensemble averaged squared matrix element normalized to the case of Eq.3.87 is approximately a hundred by using Eq.3.75 and ten by using Eq.3.85. These remarkable differences in terms of computation time are due to the fact that, for anisotropic quantization masses (that is the case reported in Fig.3.10), there are several expansion terms and the number of couples of modes $(l' - l) = (g' - g)$ increases a lot with the number of modes l and l'. Fig.3.11 shows the intrasubband squared matrix element $\left\langle \left| \mathcal{M}_0^{0'}(q_x) \right|^2 \right\rangle$ for the system InAs-HfO_2. Since the quantization mass is isotropic, for the lowest subband there is only one mode that corresponds to $l = 0$. In this case the summation in Eq.3.75 reduces to the single term corresponding $l = l' = g = g' = 0$. Fig.3.11 shows that, as expected, by reducing r_{NW} the SR matrix element is increased and by increasing r_{NW} the differences between solid- and dotted-line decrease.



(a) Simulation results obtained for [110] transport direction and two-fold valleys as shown in the inset.



(b) Simulation results obtained for [110] transport direction and four-fold valleys as shown in the inset.



(c) Simulation results obtained for [111] transport direction.

Figure 3.10: Intrasubband $\left\langle \left| \mathcal{M}_n^{n'}(q_x) \right|^2 \right\rangle$ matrix element for the Si-SiO_2 system for the lowest subband n=n'=0. r_{NW}=1.5 nm, t_{ox}=0.6 nm, the rms roughness amplitude is $\Delta_{rms} = 0.2$ nm and the correlation length is $\Lambda = 1.4$ nm. Simulations performed by considering the modes $l, l' = 0, \pm1, \pm2, \pm3, \pm4, \pm5$.

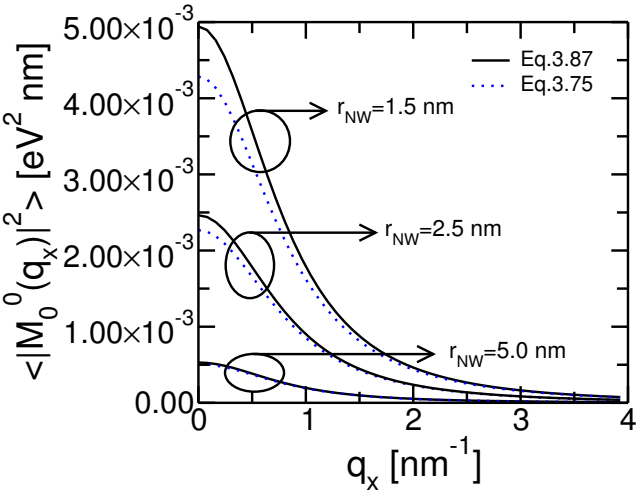


Figure 3.11: Intrasubband $\left\langle \left| \mathcal{M}_n^{n'}(q_x) \right|^2 \right\rangle$ matrix element for the InAs-HfO_2 system for the lowest subband n=n'=0. r_{NW}=1.5, 2.5 and 5 nm, t_{ox}=1.5 nm, the rms roughness amplitude is $\Delta_{rms} = 0.2$ nm and the correlation length is $\Lambda = 1.4nm$.

References

[1] D. Lizzit, D. Esseni, P. Palestri, and L. Selmi, "A new formulation for surface roughness limited mobility in bulk and ultra-thin-body metal-oxide-semiconductor transistors", *Journal of Applied Physics*, vol. 116, no. 22, 223702, 2014.

[2] (2013). International Technology Roadmap for Semiconductors, [Online]. Available: `http://www.itrs.net/Links/2012ITRS/Home2012.htm`.

[3] M. Radosavljevic et al., "Non-planar, multi-gate InGaAs quantum well field effect transistors with high-K gate dielectric and ultra-scaled gate-to-drain/gate-to-source separation for low power logic applications", in *IEEE IEDM Technical Digest*, 2010, pp. 6.1.1–6.1.4.

[4] S. H. Park, Y. Liu, N. Kharche, M. Jelodar, G. Klimeck, M. Lundstrom, and M. Luisier, "Performance Comparisons of III-V and Strained-Si in Planar FETs and Nonplanar FinFETs at Ultrashort Gate Length (12 nm)", *IEEE Trans. on Electron Devices*, vol. 59, no. 8, pp. 2107–2114, 2012.

[5] S. Chang et al., "InAs N-MOSFETs with record performance of Ion=600uA/um at Ioff=100nA/um (Vd =0.5 V)", in *IEEE IEDM Technical Digest*, Dec. 2013, pp. 16.1.1–16.1.4.

[6] S. Kim, M. Yokoyama, R. Nakane, O. Ichikawa, T. Osada, M. Hata, M. Takenaka, and S. Takagi, "High performance sub-20-nm-channel-length extremely-thin body InAs-on-insulator tri-gate MOSFETs with high short channel effect immunity and Vth tunability", in *IEEE IEDM Technical Digest*, Dec. 2013, pp. 16.4.1–16.4.4.

[7] J. Del Alamo, "Nanometre-scale electronics with III-V compound semiconductors", *Nature*, vol. 479, pp. 317 –323, 2011.

[8] G. Doornbos and M. Passlack, "Benchmarking of III-V n-MOSFET Maturity and Feasibility for Future CMOS", *IEEE Electron Device Lett.*, vol. 31, no. 10, pp. 1110 –1112, Oct. 2010.

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

[10] K. Uchida, J. Koga, and S. Takagi, "Experimental study on carrier transport mechanisms in double- and single-gate ultrathin-body MOSFETs- Coulomb scattering, volume inversion, and deltaTSOI-induced scattering", in *IEEE IEDM Technical Digest*, 2003, pp. 33.5.1–33.5.4.

[11] K. Uchida, M. Saitoh, and S. Kobayashi, "Carrier transport and stress engineering in advanced nanoscale transistors from (100) and (110) transistors to carbon nanotube FETs and beyond", in *IEEE IEDM Technical Digest*, 2008, pp. 1–4.

[12] K. Takei et al., "Quantum Confinement Effects in Nanoscale-Thickness InAs Membranes", *Nano Letters*, vol. 11, no. 11, pp. 5008–5012, 2011.

[13] S. Monaghan, P. Hurley, K. Cherkaoui, M. Negara, and A. Schenk, "Determination of electron effective mass and electron affinity in HfO2 using MOS and MOSFET structures", *Solid State Electronics*, vol. 53, no. 4, pp. 438 –444, 2009.

[14] R. E. Prange and T.-W. Nee, "Quantum Spectroscopy of the Low-Field Oscillations in the Surface Impedance", *Phys. Rev.*, vol. 168, pp. 779–786, 3 Apr. 1968.

[15] T. Ando, A. B. Fowler, and F. Stern, "Electronic properties of two-dimensional systems", *Reviews of Modern Physics*, vol. 54, pp. 437–672, 2 Apr. 1982.

[16] M.Poljak, V.Jovanović, D.Grgec, and T.Suligoj, "Assessment of Electron Mobility in Ultrathin-Body InGaAs-on-Insulator MOSFETs Using Physics-Based Modeling", *Electron Devices, IEEE Transactions on*, vol. 59, no. 6, pp. 1636–1643, 2012.

[17] S. Koba, Y. Ishida R. ans Kubota, Y. Tsuchiya H. ans Kamakura, N. Mori, and M. Ogawa, "The Impact of Increased Deformation Potential at MOS Interface on Quasi-Ballistic Transport in Ultrathin Channel MOSFETs Scaled down to Sub-10 nm Channel Length ", in *IEEE IEDM Technical Digest*, 2013, pp. 12.1.1–12.1.4.

[18] T.-W. Kim et al., "ETB-QW InAs MOSFET with scaled body for improved electrostatics", in *IEEE IEDM Technical Digest*, 2012, pp. 32.3.1–32.3.4.

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

[20] D. Esseni, P. Palestri, and L. Selmi, *Nanoscale MOS Transistors - Semi-Classical Transport and Applications*, 1st. Cambridge University Press., 2011.

[21] S. Jin, M. Fischetti, and T.-W. Tang, "Modeling of electron mobility in gated silicon nanowires at room temperature: Surface roughness scattering, dielectric screening, and band nonparabolicity", *Journal of Applied Physics*, vol. 102, no. 8, 2007.

[22] D.Lizzit, D.Esseni, P.Palestri, and L.Selmi, "Surface roughness limited mobility modeling in ultra-thin SOI and quantum well III-V MOSFETs", in *IEEE IEDM Technical Digest*, 2013, pp. 5.2.1–5.2.4.

[23] S. M. Goodnick, D. K. Ferry, C. W. Wilmsen, Z. Liliental, D. Fathy, and O. L. Krivanek, "Surface roughness at the Si(100)-SiO2 interface", vol. 32, pp. 8171–8186, 12 Dec. 1985.

[24] C. Jungemann, A. Edmunds, and W.L. Engl, "Simulation of Linear and Nonlinear Electron Transport in Homogeneous Silicon Inversion Layers", *Solid State Electronics*, vol. 36, no. 11, pp. 1529–1540, 1993.

[25] M.V. Fischetti and S.E. Laux, "Monte Carlo study of electron transport in silicon inversion layers", *Phys. Rev. B*, vol. 48, pp. 2244–2274, 1993.

[26] F. Gamiz, J. B. Roldan, J. A. Lopez-Villanueva, P. Cartujo-Cassinello, and J. E. Carceller, "Surface roughness at the Si-SiO_2 interfaces in fully depleted silicon-on-insulator inversion layers", *Journal of Applied Physics*, vol. 86, no. 12, pp. 6854–6863, 1999.

[27] T. Ishihara, K. Uchida, J. Koga, and S.-i. Takagi, "Unified Roughness Scattering Model Incorporating Scattering Component Induced by Thickness Fluctuations in Silicon-on-Insulator Metal-Oxide-Semiconductor Field-Effect Transistors", *Japanese Journal of Applied Physics*, vol. 45, no. 4B, pp. 3125–3132, 2006.

[28] 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. on Electron Devices*, vol. 54, no. 5, pp. 1156 –1164, May 2007.

[29] I. M. Tienda-Luna, F. G. Ruiz, A. Godoy, B. Biel, and F. Gamiz, "Surface roughness scattering model for arbitrarily oriented silicon nanowires", *Journal of Applied Physics*, vol. 110, no. 8, 084514, 2011.

[30] F. Stern and W. E. Howard, "Properties of Semiconductor Surface Inversion Layers in the Electric Quantum Limit", *Phys. Rev.*, vol. 163, pp. 816–835, 3 Nov. 1967.

[31] T. Ando, "Screening Effect and Quantum Transport in a Silicon Inversion Layer in Strong Magnetic Fields", *Journal of the Physical Society of Japan*, vol. 43, no. 5, pp. 1616–1626, 1977.

[32] A. Papoulis, *Probability, Random Variables and Stochastic Processes*, 3rd. McGraw-Hill Companies, Feb. 1, 1991.

[33] T. Yamanaka, S. J. Fang, H.-C. Lin, J. P. Snyder, and C. Helms, "Correlation between inversion layer mobility and surface roughness measured by afm", *IEEE Electron Device Lett.*, vol. 17, no. 4, pp. 178–180, Apr. 1996.

[34] A. Pirovano, A. Lacaita, G. Ghidini, and G. Tallarida, "On the correlation between surface roughness and inversion layer mobility in si-mosfets", *IEEE Electron Device Lett.*, vol. 21, no. 1, pp. 34–36, Jan. 2000.

[35] O. Bonno, S. Barraud, D. Mariolle, and F. Andrieu, "Effect of strain on the electron effective mobility in biaxially strained silicon inversion layers: an experimental and theoretical analysis via atomic force microscopy measurements and kubo-greenwood mobility calculations", *Journal of Applied Physics*, vol. 103, no. 6, 063715, 2008.

[36] R. Baum, "The correlation function of Gaussian noise passed through nonlinear devices", vol. 15, no. 4, pp. 448–456, Jul. 1969.

[37] M. V. Fischetti and S. Narayanan, "An empirical pseudopotential approach to surface and line-edge roughness scattering in nanostructures: application to si thin films and nanowires and to graphene nanoribbons", *Journal of Applied Physics*, vol. 110, no. 8, 083713, 2011.

[38] D. K. Ferry, *Quantum mechanics – An introduction for device physicists and electrical engineers*. Institute of Physics publishing, 2001.

[39] E. O. Kane, "Band structure of indium antimonide", *Journal of Physics and Chemistry of Solids*, vol. 1, no. 4, pp. 249 –261, 1957.

[40] D.Esseni and A. Abramo, "Modeling of Electron Mobility Degradation by Remote Coulomb Scattering in Ultrathin Oxide MOSFETs ", *IEEE Trans. on Electron Devices*, vol. 50, no. 7, pp. 1665–1674, 2003.

[41] T. P. O'Regan, P. K. Hurley, B. Sorée, and M. V. Fischetti, "Modeling the capacitance-voltage response of In0.53Ga0.47As metal-oxide-semiconductor structures: Charge quantization and nonparabolic corrections", *Applied Physics Letters*, vol. 96, no. 21, 213514, 2010.

[42] K. Kalna, S. Roy, A. Asenov, K. Elgaid, and I. Thayne, "Scaling of pseudomorphic high electron mobility transistors to decanano dimensions", *Solid State Electronics*, vol. 46, no. 5, pp. 631 –638, 2002.

[43] T. O'Regan, M. Fischetti, B. Soree, S. Jin, W. Magnus, and M. Meuris, "Calculation of the electron mobility in III-V inversion layers with high-κ dielectrics.", *Journal of Applied Physics*, vol. 108, p. 103 705, 2010.

[44] D. Lizzit, D. Esseni, P. Palestri, P. Osgnach, and L. Selmi, "Performance Benchmarking and Effective Channel Length for Nanoscale InAs, $In_{0.53}Ga_{0.47}As$, and sSi n-MOSFETs", *IEEE Trans. on Electron Devices*, vol. 61, no. 99, pp. 2027 –2034, Jun. 2014.

[45] A.Paussa and D.Esseni, "An exact solution of the linearized Boltzmann transport equation and its application to mobility calculations in graphene bilayers", *Journal of Applied Physics*, vol. 113, no. 8, p. 093 702, 2013.

[46] I. Schiff, *QUANTUM MECHANICS*, 3st. McGraw-Hill Company, 1968.

[47] S. Takagi, A. Toriumi, M. Iwase, and H. Tango, "On the universality of inversion layer mobility in Si MOSFET's: Part I-effects of substrate impurity concentration", *IEEE Trans. on Electron Devices*, vol. 41, no. 12, pp. 2357–2362, Dec. 1994.

[48] P. Osgnach, A. Revelant, D. Lizzit, P. Palestri, D. Esseni, and L. Selmi, "Toward computationally efficient Multi-Subband Monte Carlo Simulations of Nanoscale MOSFETs", in *Proc.SISPAD*, 2013, pp. 176–179.

[49] O. Badami, E. Caruso, D. Lizzit, D. Esseni, P. Palestri, and L. Selmi, "Improved surface roughness modeling and mobility projections in thin film MOSFETs", Sep. 2015, pp. 306–309.

Chapter 4

Nanowire MOSFETs

4.1 Introduction

4.1.1 State of art: manufacturing techniques

3D MOSFETs have been proposed to overcome the issues arising from aggressive scaling of planar devices (see Chapter.1). Multiple-gate MOSFETs are promising devices for future CMOS scaling due to an increased electrostatic gate-to-channel control given by the increased number of gates compared to standard planar devices. The most promising device architecture for ultimate scaling is the GAA MOSFET with a gate contact all-around the semiconductor channel. Moreover, semiconductor materials alternative to silicon providing larger injection velocities (see Sec.1.2.1), heterostructures and strain are being investigated as technological means to further improve device performances.

Fabrication of NWs follows two different paradigms: top-down [1–15] and bottom-up [16–29]. In the bottom-up method nano-particles are used as seeds to define nucleation and feed the growth of NWs. The most common approach is the chemical-vapor-deposition (CVD) with the vapor-liquid-solid (VLS) mechanisms [23]. In the VLS method the semiconductor material in vapor phase is provided by means of CVD, chemical beam epitaxy (CBE), or thermal evaporation reactors. The NW growth starts on top of a semiconductor substrate where a catalytic liquid metal seed can rapidly adsorb the semiconductor vapor to supersaturation levels from which crystal growth can subsequently occur [16, 18] (see Fig.4.1). Au has been widely used as catalyst for growing silicon NWs using VLS method allowing for NWs with diameters down to 3 nm [17], posing however, contamination issues for CMOS processing due to high Au diffusivity in silicon and to the fact that silicon also acts as an acceptor doping [28]. Moreover, Au tends to create traps in the bandgap acting as recombination centers degrading the device performance. Thus, to circumvent this issue, other metal catalysts such as Al or Pt for silicon nanowire synthesis are actively investigated to make the VLS process CMOS-compatible [24]. VLS is the most common technique but other fabrication processes include metal

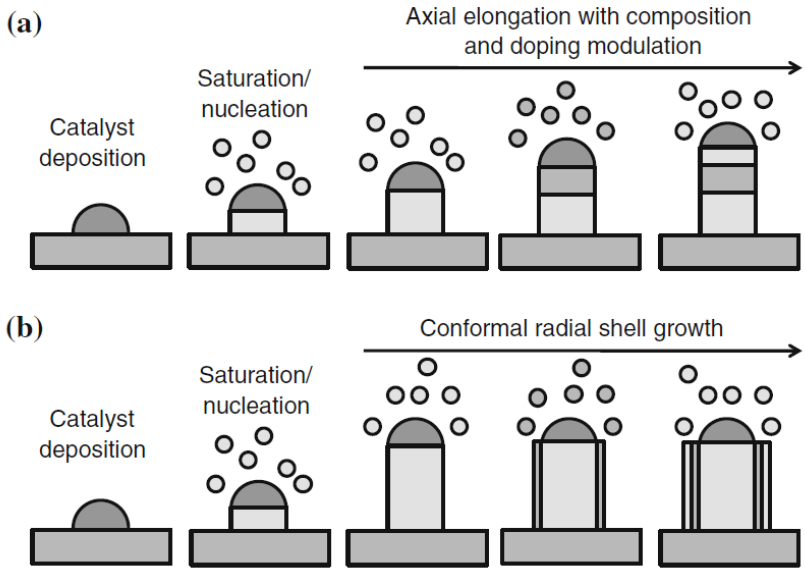


Figure 4.1: Bottom-up grown nanowire from a metal catalyst nanoparticle on a substrate for: a)axial and b) radial hetero-structure. The axial hetero-structure can be produced by modulating the composition and/or doping precursor during the elongation process. The radial core/shell hetero-structure can be formed by the conformal deposition of a radial shell after the elongation of the core structure. Figure from [30]

organic vapor phase epitaxy (MOVPE) (that allows to grow NWs even in presence of large lattice mismatch with the substrate material [27]), molecular beam epitaxy (MBE) [31], and oriented attachment [32], with all of them offering the ability to fine tune the diameter, the morphology and the semiconductor properties. Unlike standard CMOS fabrication processes where dopants are added to the semiconductor via ion implantation leading to dangled bonds and requiring subsequent thermal annealing to restore the crystal ordering, in bottom-up approaches the doping of the semiconductor is obtained thanks to the *in-situ* doping by dopant incorporation during the growth procedure. Often, bottom-up methods lead to entangled meshes of NWs [33], not suitable for large-scale device integration, thus, alignment of NWs is necessary for VLSI production. Typically, the alignment is obtained by means of different techniques such as dielectrophoresis [34–36], microfluidic alignment [37] and strong electric fields [38]. Nevertheless, even with these alignment methods, the resulting density and areal coverage of NWs cannot compete with those obtained with top-down approaches. However, bottom-up method can be comparable in terms of NW density with the top-down approach if the growth of vertically aligned silicon NWs as reported in several papers [23, 39, 40] is considered.

As opposed to the bottom-up approach, the top-down method (see Fig.4.2), can be regarded as a natural extension of standard CMOS process to produce NWs,

where, thanks to the progress in lithography and etching processes in conjunction with photoresist technology, it is possible to etch a larger piece of semiconductor in order to create NWs that are aligned parallel to the substrate. Top-down methods

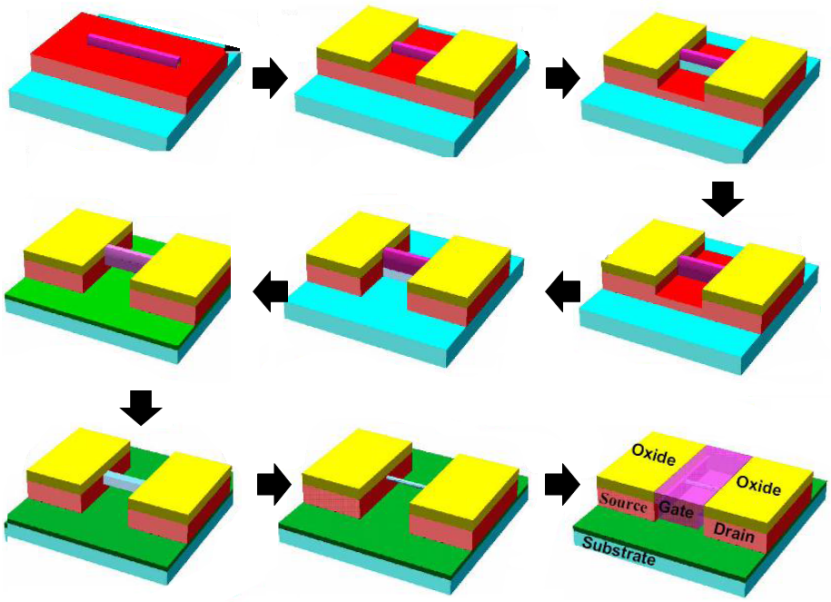


Figure 4.2: Top-down fabrication schematic diagram for a silicon nanowire MOSFET on bulk substrate [11].

include spacer lithography method [41], replacement Fin process [15, 42], recessed hard mask, trenching methods [7], and confined lateral selective epitaxial growth [5]. These methods enable excellent NW placement in the wafer and large throughput thanks to the well established CMOS-processes as well as a fine tuning in three dimensions of composition, shape, crystallinity, and size of the NWs. Top-down approach can also be used to create very-high density, vertically stacked silicon nanowire arrays [13, 43].

Bottom-up NWs suffer from worst contact resistance at source and drain interfaces [44] and from non uniform dopant distributions. Moreover, complex integration with CMOS and lack of ability to position many SiNW in a well defined pattern is still a challenge for this method for VLSI device fabrication. On the other hand, this method provides reduced dangling bonds density compared whit the top-down [44], because of better surface control during the NW growth, and enables the fabrication of device structures, such as novel core-shell heterostructures, that can reduce scattering and result in higher carrier mobility [27, 29], which are much harder to be obtained with the top-down approach [45].

As a final remark, we argue that, it is not unrealistic that device fabrication will incorporate the top-down and bottom-up approaches. One example is the case of

directed self-assembly (DSA) lithographic process for NW arrays fabrication, where bottom-up self assembly process and traditional top-down lithography are used [46].

4.1.2 State of art: simulation techniques

The constant progress in NW fabrication techniques makes simulation tools capable to include quantum effects and also the main scattering mechanisms responsible for performance degradation of increasing importance. Such simulation tools can help decide the best semiconductor materials as well as transport orientations that maximize performances. Simulation of NW devices passes necessarily through a dependable description of the energy dispersion relation in the nanostructured crystal followed by the simulation of electronic transport.

Bands structure calculation

Calculation of the subband structure can be performed with many different methods. Among the most refined methods available that account for the atomic structure of the semiconductor there are tight-binding [47–54] , first-principle [55, 56], DFT, pseudo-potential, and other atomistic methods [57]. Moreover, the atomistic description of the device allows to treat rigorously scattering mechanisms such as: surface roughness, impurity scattering and alloy disorders. Conversely, the most popular method to calculate subband energies and dispersion relations is the solution of the Schrödinger equation within the effective mass approximation (EMA) [58–61]. However, it has been reported in many papers in the literature [47, 48] the failure of the purely parabolic EMA in the presence of strong confinement, since a parabolic model does not capture the nonparabolicity effects that make the effective mass dependent on the well thickness as discussed in Appendix C.2 and tend to reduce the energy of the quantized subbands as shown in [58].

In fact, simulations performed using simple parabolic models tend to overestimate the transistor threshold voltage V_T when the diameter of silicon nanowires becomes lower than approximately 3 nm, whereas deviations in the I_{ON} start for thicknesses lower that 5 nm (for SiNWs) [48]. One effective solution to overcome this issue is to use a parabolic EMA model where the effective masses are tuned, for example, against TB calculation leading to good results over wide ranges of NW diameters [48, 52]. However this adaptive model requires a continue tuning of the masses against more sophisticated and computationally demanding solvers making this route not much promising. Moreover, nonparabolicity effects are expected to be more relevant in III-V semiconductors [62, 63]. To overcome the limitations related to the parabolic EMA, there have been proposed different methods to include nonparabolicity effects. The simplest one makes use of the subband energy levels obtained within the parabolic EMA and takes into account the band nonparabolicity only in the transport directions [64], which, however, leads to appreciable error caused by having neglected nonparabolicity in the confinement direction, or confinement plane for 1D systems [58, 65]. On the other hand, a different correction method

using the additional information given by the wave-function allows a quite accurate correction to both the dispersion relation in the transport direction and the ground energy of quantized states [66]. More sophisticated approaches have been implemented by direct modifications of the parabolic Hamiltonian [58, 69].

Another aspect that cannot be captured by the EMA model is the valley degeneracy splitting that occurs for extremely thin SiNWs with dimensions lower than 2 nm^2, as reported by atomistic calculations [48, 52, 70]. In this case, the strong confinement breaks the valley degeneracy and can lead to a worst agreement in terms of drain-to-source currents if compared with atomistic band structure methods especially at low V_{DS} values [52]. In addition, for pMOS NW devices or for transistors that demand an appropriate description of the valence band, the EMA model is even less appropriate, because the heavy and light holes in valence band are strongly coupled and cannot be described by single band effective-mass Hamiltonian [71].

Another useful and highly efficient approach for band structure calculations is the **k·p** method [71–74], which realizes a very interesting compromize between physical accuracy and computational complexity.

Transport simulation

Transport in semiconductors has been described for many years using macroscopic transport models such as the drift-diffusion (DD) model and other models based on the momenta of the Boltzmann Transport Equation (BTE)(i.e Hydrodynamic models (HD)) that assume diffusive transport regime where the average carrier velocity is strongly influenced by scattering. However, modern MOSFETs with channel length below 100 nm are pushing the classical DD or HD models to their limits because transport is not dominated by scattering but is instead close to the balistic limit. Beside this, it has been demonstrated that the DD model is very problematic for nanoscaled MOSFETs, even if it appears to be more accurate than expected because the overestimation of the conductivity in the linear regime can partly compensate the underestimation of the current at high bias [75]. In addition, in the case of moment-bases models that go beyond the DD model (e.g. HD model), there are further issues related to the additional approximations used in order to close the hierarchy of momentum equations [75, 76]. Hence, transport in nanoscale MOSFETs is in general tackled by means of semi-analytical methods (also called compact models or semi-numerical models) [48, 52, 60, 77], or fully numerical methods [47, 49–51, 53, 54, 59, 61, 68, 71–73, 79–84]. The former are in general derived by means of approximated relations for the potential and charge distributions that in general are solved self-consistently by using numerical approaches and help to give a conceptual view of the device behaviour. They are computationally very efficient and are used to rapidly analyze the performance of MOSFETs. These models require the knowledge of the energy relation ($E(\mathbf{k})$), that can be obtained by numerical TB, DFT, pseudo-potential calculations or given by analytical expressions (i.e. parabolic effective mass approximation). They assume that positive going states are injected into the channel from the source at thermal equilibrium (i.e. states filled according

to the source Fermi level), whereas for negative going states they are filled according to the drain Fermi level. The ballistic I_{DS} is calculated by subtracting the drain-to-source transmitted current from the source-to-drain transmitted current where the transmission coefficient can be one or zero depending on whether the electron energy is above or below the source-to-channel barrier (i.e. the top of barrier, ToB). Even if these methods are computationally very efficient, they can capture only the thermionic component of the current in the ballistic limit. Hence, scattering and source-to-drain tunneling are disregarded. In this regard, attempts have been made to include source-to-drain tunneling within the ToB method using approximated expressions for the potential along transport direction, and validation against 3D full-band atomistic quantum transport simulations shows good agreement [60].

Among the most used fully numerical methods for transport simulations in scaled MOSFETs we have the semi-classical Boltzmann transport equation (BTE) and the non-equilibrium Green's function (NEGF) methods. With the semi-classical BTE approach, electrons are treated as classical particles whose kinetic energy is given either by the effective mass or by the full energy band calculations and assumes that collisions of carriers are instantaneous. The BTE has been typically solved by means of stochastic methods such as Monte Carlo solvers for 3D [85–92], 2D [79, 93, 94] and 1D [82] carrier gas. Also the direct solution of the BTE is can be attractive and it has been successfully pursued for 3D [95, 96], 2D [96–98] electron and hole gas, and, due to the low dimensionality, expecially for 1D gas in thin NW MOSFETs [82]. To further ease the numerical solution, the BTE for a 1D gas has been solved also by using the momentum-relaxation-time approximation (MRT) [74, 99].

Another approach for transport consists in the quantum-mechanical non-equilibrium Green's function (NEGF) formalism, which takes into account the wave-nature of electrons described according to a TB [47, 49, 50, 53, 54, 81, 83], EMA [61, 68, 80], or a $\mathbf{k}\cdot\mathbf{p}$ Hamiltonian [71–73, 84, 100]. Fully-3D NEGF represents one of the most advanced transport methods, but, from practical point of view, the high accuracy comes to the cost of a large computational burden.

Finally, it must be recalled that complete simulations need to account for external biases (V_{GS} and V_{DS}) and this requires necessarily the coupling of the quantum-mechanical or semi-classical problem with 3D Poisson's equation.

In this framework, in this thesis we also investigate the transport in NW transistors with a channel length in the decananometer domain, and it is based on the deterministic solution of the BTE. The simulator includes size and bias induced quantization allowing the wave-function to penetrate into the oxide surrounding the NW and takes into account non parabolicity correction to the eigenvalues and to the dispersion relation $E(k)$ obtained within the EMA. The electrostatics and the transport of the device can be solved for arbitrary transport directions and moreover it allows to include multiple valleys in the electronic transport as well as scattering due to phonons, Coulomb scattering, and surface roughness with an innovative approach. In fact, although phonon scattering is found to be an important scattering mechanisms in NW MOSFETs, surface roughness scattering cannot be

ignored because of the strong quantum confinement that occurs particularly when the diameters decreases downto few nm pushing the wave-function towards the semiconductor-oxide interface. SR scattering has effects both on the I_{ON} current and effective mobility [54]. Experimental data show that I_{ON} increases with decreasing NW diameter until 3 nm [101], and this behavior has been attributed to the phonon and SR scattering [54, 101].

4.2 Schrödinger equation in polar coordinates

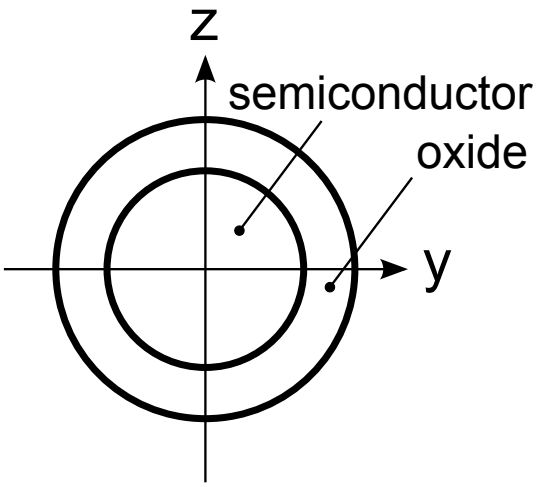


Figure 4.3: Nanowire section

For circular NWs it is convenient to move from Cartesian to polar coordinates by setting $z = rsin(\theta)$ and $y = rcos(\theta)$ where r denotes the position along the radial direction. Thus, the Schrödinger equation of Eq.B.29 for electrons belonging to a given valley is rewritten as:

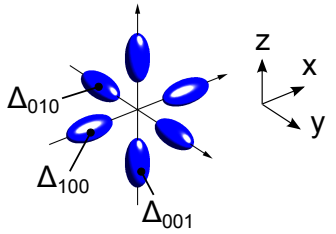
$$\left[-\frac{\hbar^2}{2}\left[w_{rr}\frac{\partial^2}{\partial r^2} + w_r\frac{\partial}{\partial r} + w_{r\theta}\frac{\partial^2}{\partial r \partial \theta} + w_\theta \frac{\partial}{\partial \theta} + w_{\theta\theta}\frac{\partial^2}{\partial \theta^2}\right] + U(r,\theta)\right]\xi_n(r,\theta) = \\ = \varepsilon_n \xi_n(r,\theta) \tag{4.1}$$

where the energy of the eigenvalue ε_n is referred to the energy of the bottom of the considered valley, and where the coefficients $w_{i,j}$ are given by:

$$\begin{aligned} w_{rr} =& w_{yy}cos^2(\theta) + 2w_{yz}sin(\theta)cos(\theta) + w_{zz}sin^2(\theta), \\ w_r =& \frac{1}{r}\left[w_{yy}sin^2(\theta) - 2w_{yz}sin(\theta)cos(\theta) + w_{zz}cos^2(\theta)\right], \\ w_{r,\theta} =& \frac{1}{r}\left[2sin(\theta)cos(\theta)\left(w_{zz} - w_{yy}\right) + 2w_{yz}(cos^2(\theta) - sin^2(\theta))\right], \\ w_\theta =& -\frac{w_{r\theta}}{r}, \\ w_{\theta\theta} =& \frac{w_r}{r}. \end{aligned} \tag{4.2}$$

A practical example for $w_{i,j}$ coefficients of Eq.4.2 in the case of $x = [100]$ transport direction is reported explicitly in Fig.4.4. Eq.4.1 shows that Schrödinger equation cannot be always expressed in terms of the Laplacian of $\xi_n(r,\theta)$:

$$\left[\frac{1}{r}\frac{\partial}{\partial r} + \frac{\partial^2}{\partial r^2} + \frac{1}{r^2}\frac{\partial^2}{\partial \theta^2}\right]\xi_n(r,\theta). \tag{4.3}$$



	$\Delta_{[100]}$	$\Delta_{[010]}$	$\Delta_{[001]}$
w_{rr}	$\frac{1}{m_t}$	$\frac{cos^2(\theta)}{m_l} + \frac{sin^2(\theta)}{m_t}$	$\frac{cos^2(\theta)}{m_t} + \frac{sin^2(\theta)}{m_l}$
w_r	$\frac{1}{rm_t}$	$\frac{1}{r}\left(\frac{sin^2(\theta)}{m_l} + \frac{cos^2(\theta)}{m_t}\right)$	$\frac{1}{r}\left(\frac{sin^2(\theta)}{m_t} + \frac{cos^2(\theta)}{m_l}\right)$
$w_{r\theta}$	0	$\frac{2cos(\theta)sin(\theta)(m_l - m_t)}{rm_tm_l}$	$\frac{2cos(\theta)sin(\theta)(m_t - m_l)}{rm_tm_l}$
w_θ	0	$\frac{2cos(\theta)sin(\theta)(m_t - m_l)}{r^2m_tm_l}$	$\frac{2cos(\theta)sin(\theta)(m_l - m_t)}{r^2m_tm_l}$
$w_{\theta\theta}$	$\frac{1}{r^2m_t}$	$\frac{1}{r^2}\left(\frac{sin^2(\theta)}{m_l} + \frac{cos^2(\theta)}{m_t}\right)$	$\frac{1}{r^2}\left(\frac{sin^2(\theta)}{m_t} + \frac{cos^2(\theta)}{m_l}\right)$

Figure 4.4: SiNW with [100] transport direction and corresponding $w_{i,j}$ coefficients.

According to the definition of the coefficients $w_{i,j}$ in Eq.4.2, in fact, the Schrödinger equation has radial symmetry only in the case of w_{yz} is null and w_{yy} and w_{zz} are equal, in which case the differential operator in Eq.4.1 simplifies to the one reported in Eq.4.3, leading to the radial symmetry of the Schrödinger equation for which the solution can be written as

$$\xi_{n,l}(r,\theta) = \frac{1}{\sqrt{2\pi}}\rho_{n,l}(r)e^{il\theta} \tag{4.4}$$

where the radial and angular parts have been separated and n is the radial quantum number and l is the angular quantum number. Eq.4.1 goes beyond the isotropic effective mass approximation adopted in many papers in the literature [67, 82], that preserves the cylindrical symmetry of the problem reducing the complexity of the equation. However, this treatment (isotropic EMA) is rigorous only for the isotropic valley Γ of the conduction band, while it becomes more problematic for the valleys with ellipsoidal constant energy surface such as Δ and Λ valleys for silicon and germanium [67, 82].

4.2.1 Continuity conditions in polar coordinates

In the case of herojunctions in the direction normal to the semiconductor-oxide interface, at the interface between different materials must hold Eq.B.35. We recall that, in the case of homogeneous material in the nanowire circular section, Eq.B.29 allows to separate the energy of the quantized states from the kinetic energy in the transport direction as reported in Eq.B.32. The treatment is much more complicated in the case of wave-function penetration in the oxide, in fact, the continuity of the component normal to the semiconductor-oxide interface of $\mathbf{W}_{DCS}\nabla\Phi_{\nu,n}(\mathbf{R})$ is in general problematic, because, for arbitrary crystal orientations, the 3×3 $\mathbf{W}_{DCS}$ tensor has off-diagonal terms and coefficients α and β of Eq.B.28 are not null. This makes the wave-vector k_x in the transport direction enter the continuity equations thus leading to a k_x dependent eigenvalue problem. In fact, the term $\mathbf{W}_{DCS}\nabla\Phi_{\nu,n}(\mathbf{R})$ leads to the vector:

$$\begin{bmatrix} w_{xx}\frac{\partial\Phi}{\partial x} + w_{xy}\frac{\partial\Phi}{\partial y} + w_{xz}\frac{\partial\Phi}{\partial z} \\ w_{yx}\frac{\partial\Phi}{\partial x} + w_{yy}\frac{\partial\Phi}{\partial y} + w_{yz}\frac{\partial\Phi}{\partial z} \\ w_{zx}\frac{\partial\Phi}{\partial x} + w_{zy}\frac{\partial\Phi}{\partial y} + w_{zz}\frac{\partial\Phi}{\partial z} \end{bmatrix} \tag{4.5}$$

where the first row of Eq.4.5 refers to the component along $\hat{e}_x$, the second along $\hat{e}_y$ and the third along $\hat{e}_z$ and where $\hat{e}_i$ is a versor aligned along direction $i \in \{x, y, z\}$. Recalling that unit vectors are related by the coordinate transformation:

$$\begin{aligned} \hat{e}_y =& cos(\theta)\hat{e}_r - sin(\theta)\hat{e}_\theta \\ \hat{e}_z =& sin(\theta)\hat{e}_r + cos(\theta)\hat{e}_\theta \end{aligned} \tag{4.6}$$

and since the objective is to calculate the component normal to the semiconductor-oxide interface of $\mathbf{W}_{DCS}\nabla\Phi_{\nu,n}(\mathbf{R})$ that for a circular domain corresponds to the radial direction $\hat{e}_r$ belonging to the quantization plane, we obtain that the following conditions at the interface between different materials must be fulfilled:

$$\begin{cases} \Phi(r_{NW}, \theta) \quad \text{constant} & \text{(4.7a)} \\ \left[A(\theta)\dfrac{\partial}{\partial x} + B(\theta)\dfrac{\partial}{\partial r} + C(\theta)\dfrac{\partial}{\partial \theta}\right]\Phi(r_{NW}, \theta) \quad \text{constant} & \text{(4.7b)} \end{cases}$$

where $A(\theta)$,$B(\theta)$, and $C(r, \theta)$ coefficients are given by:

$$\begin{aligned} A(\theta) &= w_{xy}cos(\theta) + w_{zx}sin(\theta) \\ B(\theta) &= w_{yy}cos^2(\theta) + 2w_{yz}sin(\theta)cos(\theta) + w_{zz}sin^2(\theta) \\ C(\theta) &= -w_{yy}\frac{sin(\theta)cos(\theta)}{r_{NW}} + w_{yz}\frac{cos^2(\theta) - sin^2(\theta)}{r_{NW}} + w_{zz}\frac{cos(\theta)sin(\theta)}{r_{NW}}. \end{aligned} \tag{4.8}$$

Recalling the expression for the envelope wave-function, Eq.B.31 can be substituted into the continuity condition of Eq.4.7 obtaining the new boundary conditions as a function of $\xi(r, \theta)$ that must be fulfilled at the interface between two materials:

$$\begin{cases} \xi_n(r_{NW}, \theta)e^{ik_x(\alpha\, r_{NW}\, cos(\theta) + \beta\, r_{NW}\, sin(\theta))} & \text{(4.9a)} \\ \left\{A(\theta)ik_x + B(\theta)\left[\dfrac{\partial}{\partial r} + ik_x\left(\alpha cos(\theta) + \beta sin(\theta)\right)\right] + \right. & \\ \left. + C(\theta)\left[\dfrac{\partial}{\partial \theta} + ik_x\left(\beta r_{NW} cos(\theta) - \alpha r_{NW} sin(\theta)\right)\right]\right\} e^{ik_x(\alpha\, r_{NW}\, cos(\theta) + \beta\, r_{NW}\, sin(\theta))} & \text{(4.9b)} \end{cases}$$

where α and β depend on the material and explicit expression are given in Eq.B.28. Eq.4.9 shows that, in the general case, the k_x vector enters the continuity equation. As shown in Sec.4.4.1 the dependence of the eigenvalues on the k_x vector through the continuity conditions of Eq.4.9 is almost negligible, hence, continuity conditions are implemented by setting k_x to zero:

$$\begin{cases} \xi_n(r_{NW}, \theta) & \text{(4.10a)} \\ \left\{B(\theta)\dfrac{\partial}{\partial r} + C(\theta)\dfrac{\partial}{\partial \theta}\right\}\xi_n(r_{NW}, \theta) & \text{(4.10b)} \end{cases}$$

As an example, we here report the case of a Si nanowire with [100] transport direction as shown in the left part of Fig.4.4. The $\mathbf{W}_{DCS}$ tensor is diagonal since all the ellipsoids for the three different group of valleys are aligned with the DCS and the explicit expression for the inverse mass tensor become:

$$W_{\Delta_{[100]}} = \begin{pmatrix} m_l^{-1} & 0 & 0 \\ 0 & m_t^{-1} & 0 \\ 0 & 0 & m_t^{-1} \end{pmatrix}, W_{\Delta_{[010]}} = \begin{pmatrix} m_t^{-1} & 0 & 0 \\ 0 & m_l^{-1} & 0 \\ 0 & 0 & m_t^{-1} \end{pmatrix},$$
$$W_{\Delta_{[001]}} = \begin{pmatrix} m_t^{-1} & 0 & 0 \\ 0 & m_t^{-1} & 0 \\ 0 & 0 & m_l^{-1} \end{pmatrix} \tag{4.11}$$

By substituting the $\mathbf{W}_{DCS}$ matrices into Eq.4.9 and assuming an isotropic dispersion relation for the oxide, we obtain for the valleys having the longitudinal axis aligned with the transport direction (i.e. $\Delta_{[100]}-$valleys) the following continuity condition on the first derivatives:

$$\frac{1}{m_t}\frac{\partial \xi(r,\theta)}{\partial r} = \frac{1}{m_{ox}}\frac{\partial \xi(r,\theta)}{\partial r} \tag{4.12}$$

that is an expected result since $\Delta_{[100]}-$valleys effective masses are isotropic along the quantization plane thus leading to the same condition as reported in [102, 103]. Whereas, for the other two valleys, also an angular derivative enters the continuity condition:

$$\left(\frac{1}{m_{l,t}}cos^2(\theta) + \frac{1}{m_{t,l}}sin^2(\theta)\right)\frac{\partial \xi(r,\theta)}{dr} + \left(\frac{1}{m_{t,l}} - \frac{1}{m_{l,t}}\right)\left(\frac{sin(\theta)cos(\theta)}{r}\right)\frac{\partial \xi(r,\theta)}{\partial \theta}$$
$$= \frac{1}{m_{ox}}\frac{\partial \xi(r,\theta)}{\partial r} \tag{4.13}$$

4.3 The Pseudo Spectral method

Eigenvalue-differential problems as the one reported in Eq.4.1 are typically discretized with finite-difference (FD) or finite-element (FE) methods. These methods are particularly flexible and powerful general-purpose numerical solution methods but they feature a fixed and low order accuracy. The pseudospectral (PS) method, instead, is an alternative numeric method that allows one to obtain an higher order accuracy with respect to FE and FD methods and a remarkable reduction in the CPU time due to the reduced number of discretization points. This last feature is particularly useful since the goal is to solve the Schrödinger equation in the whole 2D circular domain to account for the anisotropicity of the conduction band valleys. For analytic functions, the PS method allows for errors to decay (as the number of discretization points N increase) typically at exponential rates rather than at much slower polynomial rates. Moreover, the approach is surprisingly powerful for many cases in which both solutions and variable coefficients are non-smooth or even discontinuous and, the relatively coarse grids in PS methods results in time- and memory reduction [104, 105].

4.3.1 PS method for bounded nonperiodic domains

FD methods approximate derivatives of a function by local arguments as in the case of the standard first-order finite difference where, by expanding the function up to the first term of the Taylor series, we obtain :

$$\frac{du(x_i)}{dx} = \frac{u(x_i + h) - u(x_i)}{h} \tag{4.14}$$

where h is the grid spacing. By using the matrix formalism and writing $u(x_i)$ as a column vector, the FD method with first-order accuracy leads to a bi-diagonal differentiation matrix. Higher order approximations for the first derivative can be obtained by using more terms in the Taylor series, and appropriately weighting the various expansions in a sum. This is the idea behind PS method, that takes this process to the limit, at least in principle, and works with a differentiation formula of infinite order and leading to a dense differentiation matrix. In this sense, PS methods can be regarded as global since, as an example, in the case of the derivative of a function, the derivative calculated in x_i depends on all the other points of the domain. From here on in the derivation of the PS method we will consider a one-dimensional case within the domain $[a, b]$ for the sake of simplicity.

With the PS method, the function $u(x) \in \mathbb{C}$ for a given set of distinct points x_i $(i = 0, \ldots, N)$ is approximated as the sum of smooth basis functions using N-degree interpolating algebraic or trigonometric polynomials $p(x)$. Hence, in the PS method there are two fundamental steps that are the choice of a proper interpolation function for global calculations of spacial derivatives, and the knowledge of the discrete points (collocation points) where the calculation is computed. The approximated function $p(x)$ is written using the interpolating polynomial of the least degree (i.e. of degree

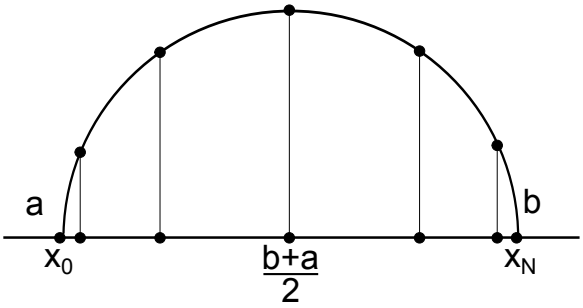


Figure 4.5: Chebyshev discretization points.

N) that is a unique interpolation polynomial and takes the form:

$$u(x) \approx p(x) = \sum_{j=0}^{N} \ell_j(x) u(x_j), \quad x \in [a, b] \tag{4.15}$$

where ℓ_j are the $N+1$ Lagrange basis polynomials and x_j are the $N+1$ nodes where the $u(x)$ function is sampled. The explicit formulation for $\ell_j(x)$ is given by:

$$\ell_j(x) = \prod_{\substack{i=0 \\ i \neq j}}^{N} \frac{x - x_i}{x_j - x_i}, \quad j = 0, 1, \ldots, N. \tag{4.16}$$

where it can be easily noticed that:

$$\ell_j(x_i) = \begin{cases} 1, & i = j \\ 0, & i \neq j. \end{cases}$$

According to Eqs.4.15,4.16 it is clear that the polynomial approximation for $u(x)$ is exact for $x = x_i$ being:

$$p(x_i) = u(x_i), \quad i = 0, \ldots, N. \tag{4.17}$$

As reported in many papers in the literature [104, 105], the x_i grid points are assumed to be Chebyshev points over the $[a, b]$ domain given by the following equation:

$$x_i = \frac{b+a}{2} - \frac{b-a}{2} cos(\frac{i\pi}{N}), \quad i = 0, 1, \ldots, N. \tag{4.18}$$

and represent the projection onto the x-axis of the equispaced points on the circle with radius $(b-a)/2$ as reported in Fig.4.5. The choice for Chebyshev points removes the problem associated with the inability to approximate a smooth function at the boundaries of the domain when a uniform-grid is used. This is the so-called Runge phenomenon that makes the solution found onto a uniform spaced domain oscillate [104]. Incidentally, the case for smooth function at the boundaries of the domain is the case for wave-functions close to the outer boundary when the penetration of the

wave-function into the oxide is allowed, hence, the choice for Chebyshev collocation points is appropriate. The idea behind the solution of a differential equation within the PS method is to approximate the derivative of the unknown $u(x_i)$ using the derivatives of $p(x_i)$ as follow:

$$\frac{du(x)}{dx}|_{x_i} \approx \frac{dp(x)}{dx}|_{x_i} = \sum_{j=0}^{N} \ell'_j(x_i)u(x_j), \quad x \in [a,b] \tag{4.19}$$

where $\ell'_j(x_i) = \frac{d\ell'(x)}{dx}|_{x_i}$ with $i,j = 0,\ldots,N$. Eq.4.19 can be rewritten by using the matrix formalism to write the column vector of the derivatives:

$$\underbrace{\begin{bmatrix} u'(x_0) \\ u'(x_1) \\ \vdots \\ u'(x_N) \end{bmatrix}}_{\overline{u}'_{N+1}} \approx \underbrace{\begin{bmatrix} \ell'_0(x_0) & \ell'_1(x_0) & \ell'_2(x_0) & \cdots & \ell'_N(x_0) \\ \ell'_0(x_1) & \ell'_1(x_1) & \ell'_2(x_1) & \cdots & \ell'_N(x_0) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \ell'_0(x_N) & \ell'_1(x_N) & \ell'_2(x_N) & \cdots & \ell'_N(x_N) \end{bmatrix}}_{\mathbf{D}_{N+1}} \underbrace{\begin{bmatrix} u(x_0) \\ u(x_1) \\ \vdots \\ u(x_N) \end{bmatrix}}_{\overline{u}_{N+1}} \tag{4.20}$$

where $\mathbf{D}_{N+1}$ is an $(N+1)$x$(N+1)$ matrix and denotes the first order differentiation matrix. Explicit calculations for $\mathbf{D}_{N+1}$ are reported in [104]. Higher orders of derivation can be obtained recursively, in fact:

$$\frac{d^2u(x_i)}{dx^2} = \frac{d}{dx}\left(\frac{du(x_i)}{dx}\right) \approx \sum_{j=0}^{N} \ell'_j(x_i)u'(x_j) = \sum_{j=0}^{N}\sum_{m=0}^{N} \ell'_j(x_i)\ell'_m(x_j)u(x_m) = \\ = \mathbf{D}^2_{N+1}\overline{u}_{N+1}, \tag{4.21}$$

which costs $O(N+1)^3$ floating point operations, or it can be calculated by explicit formulas [104].

More generally, the k-order derivative can be computed as follows:

$$\overline{u}^k_{N+1} = \mathbf{D}^k_{N+1}\overline{u}_{N+1}, \quad k < N. \tag{4.22}$$

4.3.2 Boundary conditions

PDEs that describe physical mechanisms, are completely defined only by adding appropriate boundary conditions. We will see that with the PS method the inclusion of boundary and continuity conditions is quite straightforward.

Let us write a second-order differential equation over the domain $x = [a,b]$:

$$\frac{d^2u(x)}{dx^2} = -\rho(x) \longrightarrow \mathbf{E}_{N+1}\overline{u}_{N+1} = -\overline{\rho}_{N+1} \tag{4.23}$$

where the $\mathbf{E}_{N+1}$ matrix is calculated according to Eq.4.22 setting $k = 2$ and boundary conditions are written as a combination of the values of the function and the values of its derivative at the boundaries of the domain (Robin's boundary conditions):

$$\begin{aligned} \alpha_a u(a) + \beta_a u'(a) &= \gamma_a \quad \text{for } x = a \\ \alpha_b u(b) + \beta_b u'(b) &= \gamma_b \quad \text{for } x = b \end{aligned} \tag{4.24}$$

where α, β and γ are generic constants. By substituting Eq.4.22 setting $k = 1$ into Eq.4.24 we obtain a set of equations that depend only on the vector of the unknonwns and not on the first derivatives:

$$\begin{aligned} \alpha_a u(a) + \beta_a \mathbf{D}_{N+1}(0,:)\overline{u}_{N+1} &= \gamma_a \\ \alpha_b u(b) + \beta_b \mathbf{D}_{N+1}(N,:)\overline{u}_{N+1} &= \gamma_b \end{aligned} \tag{4.25}$$

where $\mathbf{D}_{N+1}(i,:)$ denotes the $[1\text{x}(N+1)]$ row vector obtained by taking the $i-$th row of the matrix $\mathbf{D}_{N+1}$ and all the elements along the columns. Eq.4.25 is rewritten in a more compact form:

$$\begin{aligned} \mathbf{P}_a \overline{u}_{N+1} &= \gamma_a \\ \mathbf{P}_b \overline{u}_{N+1} &= \gamma_b \end{aligned} \tag{4.26}$$

where

$$\begin{aligned} \mathbf{P}_a &= \left[\underbrace{\alpha_a + \beta_a \mathbf{D}_{N+1}(0,0)}_{\mathbf{P}_a^{11}} \quad \underbrace{\beta_a \mathbf{D}_{N+1}(0,1:N-1)}_{\mathbf{P}_a^{12}} \quad \underbrace{\beta_a \mathbf{D}_{N+1}(0,N)}_{\mathbf{P}_a^{13}} \right] \\ \mathbf{P}_b &= \left[\underbrace{\beta_b \mathbf{D}_{N+1}(N,0)}_{\mathbf{P}_b^{31}} \quad \underbrace{\beta_b \mathbf{D}_{N+1}(N,1:N-1)}_{\mathbf{P}_b^{32}} \quad \underbrace{\alpha_b + \beta_b \mathbf{D}_{N+1}(N,N)}_{\mathbf{P}_b^{33}} \right]. \end{aligned} \tag{4.27}$$

Therefore, the final set of equations for the numeric solution of Eq.4.23 is obtained by substituting the first and last rows of Eq.4.23 with the corresponding equations in Eq.4.26 and the final system to be solved becomes:

$$\begin{bmatrix} \mathbf{P}_a^{11} & \mathbf{P}_a^{12} & \mathbf{P}_a^{13} \\ \mathbf{E}_{N+1}(1:N-1,:) \\ \mathbf{P}_b^{31} & \mathbf{P}_b^{32} & \mathbf{P}_b^{33} \end{bmatrix} \begin{bmatrix} \overline{u}_{N+1}(0) \\ \overline{u}_{N+1}(1:N-1) \\ \overline{u}_{N+1}(N) \end{bmatrix} = \begin{bmatrix} \gamma_a \\ \overline{\rho}_{N+1}(1:N-1) \\ \gamma_b \end{bmatrix} \tag{4.28}$$

4.3.3 Heterostructures

The solution of the Schrödinger equation taking into account the wave-function penetration into the oxide or the solution of the Poisson's equation in the oxide and semiconductor domains need to be addressed properly in order to preserve the accuracy of the PS methods that can be compromised by the presence of discontinuities in the material properties.

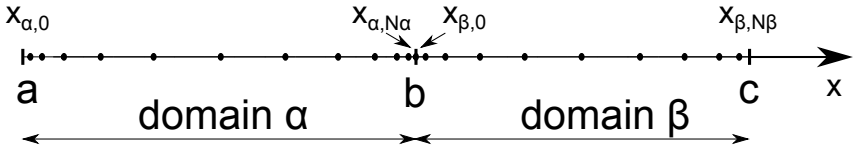


Figure 4.6: Sketch of an heterogeneous domain with Chebyshev discretization points.

In the presence of two different materials, in a one-dimensional case as reported in Fig.4.6 must be defined two separate domains and differentiation matrices, and then

the two problems must be coupled by means of appropriate continuity conditions. Let us suppose to solve a differential problem with second-order derivative on the domain defined in Fig.4.6:

$$\left(\frac{d}{dx}\varepsilon(x)\frac{d}{dx}\right)u(x) = -\rho(x) \tag{4.29}$$

with Robin's boundary conditions at $x = a$ and $x = c$ expressed as in Eq.4.24, where $\varepsilon(x) = \varepsilon_\alpha$ for $a \leqslant x \leqslant b$ and $\varepsilon(x) = \varepsilon_\beta$ for $b \leqslant x \leqslant c$ and Eq.4.29 can be regarded as the Poisson's equation for the potential $u(x)$. Eq.4.29 can be rewritten discretizing the x domain into Chebyshev collocation points by defining two differentiation matrices, one for each sub-domain:

$$\begin{aligned} &\varepsilon_\alpha \mathbf{E}_{\alpha,N_\alpha+1}\overline{u}_\alpha = -\overline{\rho}_\alpha \quad \text{for domain } \alpha \\ &\varepsilon_\beta \mathbf{E}_{\beta,N_\beta+1}\overline{u}_\beta = -\overline{\rho}_\beta \quad \text{for domain } \beta. \end{aligned} \tag{4.30}$$

where $\overline{u}_m$ is the discretized version of the potential $u(x)$ over the Chebyshev domain $x_{m,i}$ where m denotes the domain index and $i = 0, 1, \ldots, N_m$ and where $\mathbf{E}_{m,N_m+1}$ is the second derivative discretization matrix of dimensions $[(N_m + 1)\text{x}(N_m + 1)]$ obtained by squaring the differentiation matrix for the first derivative according to Eq.4.22. Eq.4.30 can be rewritten by dividing the $\mathbf{E}_{m,N_m+1}$ matrices into blocks thus obtaining:

$$\varepsilon_\alpha \begin{bmatrix} \mathbf{E}_\alpha^{11} & \mathbf{E}_\alpha^{12} & \mathbf{E}_\alpha^{13} \\ \mathbf{E}_\alpha^{21} & \mathbf{E}_\alpha^{22} & \mathbf{E}_\alpha^{23} \\ \mathbf{E}_\alpha^{31} & \mathbf{E}_\alpha^{32} & \mathbf{E}_\alpha^{33} \end{bmatrix} \begin{bmatrix} u_{\alpha,a} \\ \overline{u}_{\alpha,inn} \\ u_{\alpha,b} \end{bmatrix} = \begin{bmatrix} -\rho_{\alpha,a} \\ -\overline{\rho}_{\alpha,inn} \\ -\rho_{\alpha,b} \end{bmatrix} \tag{4.31}$$

$$\varepsilon_\beta \begin{bmatrix} \mathbf{E}_\beta^{11} & \mathbf{E}_\beta^{12} & \mathbf{E}_\beta^{13} \\ \mathbf{E}_\beta^{21} & \mathbf{E}_\beta^{22} & \mathbf{E}_\beta^{23} \\ \mathbf{E}_\beta^{31} & \mathbf{E}_\beta^{32} & \mathbf{E}_\beta^{33} \end{bmatrix} \begin{bmatrix} u_{\beta,b} \\ \overline{u}_{\beta,inn} \\ u_{\beta,c} \end{bmatrix} = \begin{bmatrix} -\rho_{\beta,b} \\ -\overline{\rho}_{\beta,inn} \\ -\rho_{\beta,c} \end{bmatrix}. \tag{4.32}$$

where $u_{\alpha,a} = u_\alpha(x = a)$ and where $\overline{u}_{m,inn}$ denotes the column vector of terms $u_{m,i}$ with $i = 1, \ldots, N_m - 1$ (i.e. in the inner part of the $m-$th domain). Hence, the total system to be solved is given by:

$$\begin{bmatrix} \varepsilon_\alpha \begin{bmatrix} \mathbf{E}_\alpha^{11} & \mathbf{E}_\alpha^{12} & \mathbf{E}_\alpha^{13} \\ \mathbf{E}_\alpha^{21} & \mathbf{E}_\alpha^{22} & \mathbf{E}_\alpha^{23} \\ \mathbf{E}_\alpha^{31} & \mathbf{E}_\alpha^{32} & \mathbf{E}_\alpha^{33} \end{bmatrix} & & 0 \\ & & \\ 0 & & \varepsilon_\beta \begin{bmatrix} \mathbf{E}_\beta^{11} & \mathbf{E}_\beta^{12} & \mathbf{E}_\beta^{13} \\ \mathbf{E}_\beta^{21} & \mathbf{E}_\beta^{22} & \mathbf{E}_\beta^{23} \\ \mathbf{E}_\beta^{31} & \mathbf{E}_\beta^{32} & \mathbf{E}_\beta^{33} \end{bmatrix} \end{bmatrix} \begin{bmatrix} u_{\alpha,a} \\ \overline{u}_{\alpha,inn} \\ u_{\alpha,b} \\ u_{\beta,b} \\ \overline{u}_{\beta,inn} \\ u_{\beta,c} \end{bmatrix} = \begin{bmatrix} -\rho_{\alpha,a} \\ -\overline{\rho}_{\alpha,inn} \\ -\rho_{\alpha,b} \\ -\rho_{\beta,b} \\ -\overline{\rho}_{\beta,inn} \\ -\rho_{\beta,c} \end{bmatrix}. \tag{4.33}$$

In Eq.4.33 we must enforce the continuity conditions at the interface between the domains and the boundary conditions. If Eq.4.29 is considered as the Poisson's equation for the potential $u(x)$, continuity conditions are given by the two sets of

equations that guarantee the continuity of the potential and the conservation of the electrical displacement at the heterostructure interface:

$$\begin{aligned} & u_{\alpha,b} = u_{\beta,b} \\ & \varepsilon_\alpha \frac{du(x)}{dx}|_{x=b^-} = \varepsilon_\beta \frac{du(x)}{dx}|_{x=b^+}. \end{aligned} \tag{4.34}$$

Boundary and continuity conditions can be imposed in two different ways using the explicit or implicit formalism.

Explicit formalism for boundary conditions

Within the explicit formalisms the rows of Eq.4.33 corresponding to continuity and boundary collocation points are simply substituted with the corresponding continuity and boundary equations. For the former ones, the third and fourth rows of Eq.4.33 are replaced by the two continuity conditions of Eq.4.34, whereas for the boundary conditions, which are assumed to be Robin's conditions (see Eq.4.24) we substitute the first and last row of Eq.4.33 with the corresponding boundary equations written in a compact form as reported in Eq.4.26, leading to a new system of equations:

$$\underbrace{\begin{bmatrix} \mathbf{P}_a^{11} & \mathbf{P}_a^{12} & \mathbf{P}_a^{13} & & 0 & \\ \varepsilon_\alpha \begin{bmatrix} \mathbf{E}_\alpha^{21} & \mathbf{E}_\alpha^{22} & \mathbf{E}_\alpha^{23} \end{bmatrix} & & & 0 & \\ 0 & 0 & 1 & -1 & 0 & 0 \\ \varepsilon_\alpha \mathbf{D}_\alpha(N,:) & & & -\varepsilon_\beta \mathbf{D}_\beta(0,:) & & \\ 0 & & & \varepsilon_\beta \begin{bmatrix} \mathbf{E}_\beta^{21} & \mathbf{E}_\beta^{22} & \mathbf{E}_\beta^{23} \end{bmatrix} & & \\ 0 & & & \mathbf{P}_c^{31} & \mathbf{P}_c^{32} & \mathbf{P}_c^{33} \end{bmatrix}}_{\mathbf{L}_{N_\alpha+N_\beta+2}} \begin{bmatrix} u_{\alpha,a} \\ \overline{u}_{\alpha,inn} \\ u_{\alpha,b} \\ u_{\beta,b} \\ \overline{u}_{\beta,inn} \\ u_{\beta,c} \end{bmatrix} = \begin{bmatrix} \gamma_a \\ \overline{\rho}_{\alpha,inn} \\ 0 \\ 0 \\ \overline{\rho}_{\beta,inn} \\ \gamma_c \end{bmatrix} \tag{4.35}$$

where $\mathbf{D}_m(n,:)$ denotes the $n-$th row of the differentiation matrix $\mathbf{D}_m$ for the first derivative for the domain m having dimensions $[(N_m+1)\mathrm{x}(N_m+1)]$, $\mathbf{L}_{N_\alpha+N_\beta+2} \in \mathbb{R}^{(N_\alpha+1)+(N_\beta+1)\mathrm{x}(N_\alpha+1)+(N_\beta+1)}$ and where

$$\begin{aligned} \mathbf{P}_a &= \underbrace{\begin{bmatrix} \alpha_a + \beta_a \underbrace{\mathbf{D}_{\alpha,N+1}(0,0)}_{\mathbf{D}_\alpha^{11}} & \underbrace{\beta_a \mathbf{D}_{\alpha,N+1}(0,1:N-1)}_{\mathbf{D}_\alpha^{12}} & \underbrace{\beta_a \mathbf{D}_{\alpha,N+1}(0,N)}_{\mathbf{D}_\alpha^{13}} \end{bmatrix}}_{\mathbf{P}_a^{11} \quad \mathbf{P}_a^{12} \quad \mathbf{P}_a^{13}} \\ \mathbf{P}_c &= \underbrace{\begin{bmatrix} \beta_c \underbrace{\mathbf{D}_{\beta,N+1}(N,0)}_{\mathbf{D}_\beta^{31}} & \underbrace{\beta_c \mathbf{D}_{\beta,N+1}(N,1:N-1)}_{\mathbf{D}_\beta^{32}} & \underbrace{\alpha_c + \beta_c \mathbf{D}_{\beta,N+1}(N,N)}_{\mathbf{D}_\beta^{33}} \end{bmatrix}}_{\mathbf{P}_c^{31} \quad \mathbf{P}_c^{32} \quad \mathbf{P}_c^{33}}. \end{aligned} \tag{4.36}$$

Finally, the solution of the discretized version of Eq.4.29 over the two Chebyshev domains is obtained by solving the system of equations in Eq.4.35. We recall that

according to the explicit method, the vector of solutions contains two degenerate values for the point $x = b$.

It is interesting to see that in the case of Dirichlet's boundary conditions of the type $u_{\alpha,a} = u_{\beta,c} = 0$ the first and last columns of the matrix $\mathbf{L}_{N_\alpha+N_\beta+2}$ have no effect (since multiplied by zero) and the same holds for the first and last rows of the matrix in Eq.4.35 (since ignored). In other words, to solve the one-dimensional problem in Eq.4.29 by using Chebyshev PS method with $u_{\alpha,a} = u_{\beta,c} = 0$ boundary conditions one can use the $\tilde{\mathbf{L}}_{N_\alpha+N_\beta}$ matrix of dimensions $[(N_\alpha + N_\beta)\mathrm{x}(N_\alpha + N_\beta)]$ obtained by stripping the matrix $\mathbf{L}_{N_\alpha+N_\beta+2}$ of Eq.4.35 of its first and last rows and columns obtaining:

$$\left[\tilde{\mathbf{L}}_{N_\alpha+N_\beta}\right] \begin{bmatrix} \overline{u}_{\alpha,inn} \\ u_{\alpha,b} \\ u_{\beta,b} \\ \overline{u}_{\beta,inn} \end{bmatrix} = \begin{bmatrix} \overline{\rho}_{\alpha,inn} \\ 0 \\ 0 \\ \overline{\rho}_{\beta,inn} \end{bmatrix} \tag{4.37}$$

Implicit formalism for boundary conditions

The second way to impose continuity and boundary conditions consists in removing from the final system of equations the corresponding continuity and boundary collocation points. Let us focus on the system of equations in Eq.4.33. Recalling the continuity condition on the function $u(x)$ at the interface $x = b$ given by the first equation in Eq.4.34, the system of equations can be rewritten removing the row associated with $u_{\alpha,b}$ or with $u_{\beta,b}$ (third or fourth row of the system of equations in Eq.4.33) leading to:

$$\underbrace{\begin{bmatrix} \varepsilon_\alpha \begin{bmatrix} \mathbf{E}_\alpha^{11} & \mathbf{E}_\alpha^{12} & \mathbf{E}_\alpha^{13} \\ \mathbf{E}_\alpha^{21} & \mathbf{E}_\alpha^{22} & \mathbf{E}_\alpha^{23} \\ \mathbf{E}_\alpha^{31} & \mathbf{E}_\alpha^{32} & \mathbf{E}_\alpha^{33} \end{bmatrix} & 0 \\ 0 & \varepsilon_\beta \begin{bmatrix} \mathbf{E}_\beta^{21} & \mathbf{E}_\beta^{22} & \mathbf{E}_\beta^{23} \\ \mathbf{E}_\beta^{31} & \mathbf{E}_\beta^{32} & \mathbf{E}_\beta^{33} \end{bmatrix} \end{bmatrix}}_{\mathbf{L}_{N_\alpha+N_\beta+1}} \begin{bmatrix} u_{\alpha,a} \\ \overline{u}_{\alpha,inn} \\ u_{\alpha,b} \\ \overline{u}_{\beta,inn} \\ u_{\beta,c} \end{bmatrix} = \begin{bmatrix} \rho_{\alpha,a} \\ \overline{\rho}_{\alpha,inn} \\ \rho_{\alpha,a} \\ \overline{\rho}_{\beta,inn} \\ \rho_{\beta,c} \end{bmatrix} \tag{4.38}$$

where we have chosen to remove from Eq.4.33 the equation for the unknown $u_{\beta,b}$. According to the differentiation matrices defined in Eq.4.22, the first derivative of the function $u(x)$ defined over the Chebyshev collocation points can be written using the matrix formalisms and by dividing the differentiation matrices into blocks we obtain:

$$\begin{aligned} \varepsilon_\alpha \begin{bmatrix} \mathbf{D}_\alpha^{11} & \mathbf{D}_\alpha^{12} & \mathbf{D}_\alpha^{13} \\ \mathbf{D}_\alpha^{21} & \mathbf{D}_\alpha^{22} & \mathbf{D}_\alpha^{23} \\ \mathbf{D}_\alpha^{31} & \mathbf{D}_\alpha^{32} & \mathbf{D}_\alpha^{33} \end{bmatrix} \begin{bmatrix} u_{\alpha,a} \\ \overline{u}_{\alpha,inn} \\ u_{\alpha,b} \end{bmatrix} &= \begin{bmatrix} du_\alpha/dx|_a \\ du_\alpha/dx|_{inn} \\ du_\alpha/dx|_b \end{bmatrix} \\ \varepsilon_\beta \begin{bmatrix} \mathbf{D}_\beta^{11} & \mathbf{D}_\beta^{12} & \mathbf{D}_\beta^{13} \\ \mathbf{D}_\beta^{21} & \mathbf{D}_\beta^{22} & \mathbf{D}_\beta^{23} \\ \mathbf{D}_\beta^{31} & \mathbf{D}_\beta^{32} & \mathbf{D}_\beta^{33} \end{bmatrix} \begin{bmatrix} u_{\beta,b} \\ \overline{u}_{\beta,inn} \\ u_{\beta,c} \end{bmatrix} &= \begin{bmatrix} du_\beta/dx|_b \\ du_\beta/dx|_{inn} \\ du_\beta/dx|_c \end{bmatrix}. \end{aligned} \tag{4.39}$$

The continuity condition on the derivatives of Eq.4.34 can be rewritten exploiting Eq.4.39 as:

$$\varepsilon_\alpha(\mathbf{D}_\alpha^{31}u_{\alpha,a} + \mathbf{D}_\alpha^{32}\overline{u}_{\alpha,inn} + \mathbf{D}_\alpha^{33}u_{\alpha,b}) = \varepsilon_\beta(\mathbf{D}_\beta^{11}u_{\beta,b} + \mathbf{D}_\beta^{12}\overline{u}_{\beta,inn} + \mathbf{D}_\beta^{13}u_{\beta,c}) \quad (4.40)$$

and recalling that $u_{\alpha,b} = u_{\beta,b} = u_b$, the above condition can be written as:

$$\begin{bmatrix}\mathbf{D}_\alpha^{31} & \mathbf{D}_\alpha^{32} & \left(\mathbf{D}_\alpha^{33} - \xi\mathbf{D}_\beta^{11}\right) & -\xi\mathbf{D}_\beta^{12} & -\xi\mathbf{D}_\beta^{13}\end{bmatrix}\begin{bmatrix}u_{\alpha,a}\\ \overline{u}_{\alpha,inn}\\ u_b\\ \overline{u}_{\beta,inn}\\ u_{\beta,c}\end{bmatrix} = 0 \quad (4.41)$$

where $\xi = \varepsilon_\beta/\varepsilon_\alpha$. The same can be done for the Robin's boundary conditions of Eq.4.24 which leads to the set of equations:

$$\begin{bmatrix}\mathbf{P}_a^{11} & \mathbf{P}_a^{12} & \mathbf{P}_a^{13} & 0 & 0\\ 0 & 0 & \mathbf{P}_c^{31} & \mathbf{P}_c^{32} & \mathbf{P}_c^{33}\end{bmatrix}\begin{bmatrix}u_{\alpha,a}\\ \overline{u}_{\alpha,inn}\\ u_b\\ \overline{u}_{\beta,inn}\\ u_{\beta,c}\end{bmatrix} = \begin{bmatrix}\gamma_a\\ \gamma_c\end{bmatrix} \quad (4.42)$$

where the $\mathbf{P^{ij}}$ elements are defined in Eq.4.36. By combining Eq.4.42 and Eq.4.41 to write explicitely the dependence of the unknown function $u(x_{m,i})$ at the collocations points $x = a, b, c$ we obtain:

$$\underbrace{\begin{bmatrix}\mathbf{P}_a^{11} & \mathbf{P}_a^{13} & 0\\ \mathbf{D}_\alpha^{31} & \left(\mathbf{D}_\alpha^{33} - \xi\mathbf{D}_\beta^{11}\right) & -\xi\mathbf{D}_\beta^{13}\\ 0 & \mathbf{P}_c^{31} & \mathbf{P}_c^{33}\end{bmatrix}}_{L_{3x3}}\begin{bmatrix}u_{\alpha,a}\\ u_b\\ u_{\beta,c}\end{bmatrix} = \begin{bmatrix}-\mathbf{P}_a^{12}\overline{u}_{\alpha,inn} + \gamma_a\\ -\mathbf{D}_\alpha^{32}\overline{u}_{\alpha,inn} + \xi\mathbf{D}_\beta^{12}\overline{u}_{\beta,inn}\\ -\mathbf{P}_c^{32}\overline{u}_{\beta,inn} + \gamma_c\end{bmatrix} \quad (4.43)$$

where L_{3x3} is a 3x3 matrix. By defining $\tilde{\mathbf{L}}_{3x3}$ the inverse of $\mathbf{L}_{3x3}$ matrix and inverting the problem in Eq.4.43 to find explicit expressions for $u_{\alpha,a}, u_b$, and $u_{\beta,c}$ we obtain:

$$\begin{bmatrix}u_{\alpha,a}\\ u_b\\ u_{\beta,c}\end{bmatrix} = \begin{bmatrix}-\tilde{\mathbf{L}}^{11}\mathbf{P}_a^{12} - \tilde{\mathbf{L}}^{12}\mathbf{D}_\alpha^{32} & \tilde{\mathbf{L}}^{12}\xi\mathbf{D}_\beta^{12} - \tilde{\mathbf{L}}^{13}\mathbf{P}_c^{32}\\ -\tilde{\mathbf{L}}^{21}\mathbf{P}_a^{12} - \tilde{\mathbf{L}}^{22}\mathbf{D}_\alpha^{32} & \tilde{\mathbf{L}}^{22}\xi\mathbf{D}_\beta^{12} - \tilde{\mathbf{L}}^{23}\mathbf{P}_c^{32}\\ -\tilde{\mathbf{L}}^{31}\mathbf{P}_a^{12} - \tilde{\mathbf{L}}^{32}\mathbf{D}_\alpha^{32} & \tilde{\mathbf{L}}^{32}\xi\mathbf{D}_\beta^{12} - \tilde{\mathbf{L}}^{33}\mathbf{P}_c^{32}\end{bmatrix}\begin{bmatrix}\overline{u}_{\alpha,inn}\\ \overline{u}_{\beta,inn}\end{bmatrix} + \begin{bmatrix}\tilde{\mathbf{L}}^{11} & \tilde{\mathbf{L}}^{13}\\ \tilde{\mathbf{L}}^{21} & \tilde{\mathbf{L}}^{23}\\ \tilde{\mathbf{L}}^{31} & \tilde{\mathbf{L}}^{33}\end{bmatrix}\begin{bmatrix}\gamma_a\\ \gamma_c\end{bmatrix} \quad (4.44)$$

that written in a more compact form, becomes:

$$\begin{bmatrix}u_{\alpha,a}\\ u_b\\ u_{\beta,c}\end{bmatrix} = \begin{bmatrix}\mathbf{W}^{11} & \mathbf{W}^{12}\\ \mathbf{W}^{21} & \mathbf{W}^{22}\\ \mathbf{W}^{31} & \mathbf{W}^{32}\end{bmatrix}\begin{bmatrix}\overline{u}_{\alpha,inn}\\ \overline{u}_{\beta,inn}\end{bmatrix} + \begin{bmatrix}\tilde{\mathbf{L}}^{11} & \tilde{\mathbf{L}}^{13}\\ \tilde{\mathbf{L}}^{21} & \tilde{\mathbf{L}}^{23}\\ \tilde{\mathbf{L}}^{31} & \tilde{\mathbf{L}}^{33}\end{bmatrix}\begin{bmatrix}\gamma_a\\ \gamma_c\end{bmatrix}. \quad (4.45)$$

From the system of equations in Eq.4.38, it can be seen that:

$$\begin{aligned}\overline{\rho}_{\alpha,inn} &= \varepsilon_\alpha \mathbf{E}_\alpha^{21} u_{\alpha,a} + \varepsilon_\alpha \mathbf{E}_\alpha^{22} \overline{u}_{\alpha,inn} + \varepsilon_\alpha \mathbf{E}_\alpha^{23} u_b \\ \overline{\rho}_{\beta,inn} &= \varepsilon_\beta \mathbf{E}_\beta^{21} u_b + \varepsilon_\beta \mathbf{E}_\beta^{22} \overline{u}_{\beta,inn} + \varepsilon_\beta \mathbf{E}_\beta^{23} u_{\beta,c}\end{aligned} \tag{4.46}$$

and by substituting into Eq.4.46 the expressions for $u_{\alpha,a}, u_b$, and $u_{\beta,c}$ given by Eq.4.45 we obtain a set of equations which depend only on the inner points of the domains (i.e. do not appear the unknown functions evaluated at the boundary and interface collocation points):

$$\begin{aligned}&\begin{bmatrix} \varepsilon_\alpha \mathbf{E}_\alpha^{21}\mathbf{W}^{11} + \varepsilon_\alpha \mathbf{E}_\alpha^{22} + \varepsilon_\alpha \mathbf{E}_\alpha^{23}\mathbf{W}^{21} & \varepsilon_\alpha \mathbf{E}_\alpha^{21}\mathbf{W}^{12} + \varepsilon_\alpha \mathbf{E}_\alpha^{23}\mathbf{W}^{22} \\ \varepsilon_\beta \mathbf{E}_\beta^{21}\mathbf{W}^{21} + \varepsilon_\beta \mathbf{E}_\beta^{23}\mathbf{W}^{31} & \varepsilon_\beta \mathbf{E}_\beta^{21}\mathbf{W}^{22} + \varepsilon_\beta \mathbf{E}_\beta^{22} + \varepsilon_\beta \mathbf{E}_\beta^{23}\mathbf{W}^{32} \end{bmatrix} \begin{bmatrix} \overline{u}_{\alpha,inn} \\ \overline{u}_{\beta,inn} \end{bmatrix} = \\ &= \begin{bmatrix} \overline{\rho}_{\alpha,inn} \\ \overline{\rho}_{\beta,inn} \end{bmatrix} - \begin{pmatrix} \varepsilon_\alpha \mathbf{E}_\alpha^{21} \left(\tilde{\mathbf{L}}^{11}\gamma_a + \tilde{\mathbf{L}}^{13}\gamma_c \right) + \varepsilon_\alpha \mathbf{E}_\alpha^{23} \left(\tilde{\mathbf{L}}^{21}\gamma_a + \tilde{\mathbf{L}}^{23}\gamma_c \right) \\ \varepsilon_\beta \mathbf{E}_\beta^{21} \left(\tilde{\mathbf{L}}^{21}\gamma_a + \tilde{\mathbf{L}}^{23}\gamma_c \right) + \varepsilon_\beta \mathbf{E}_\beta^{23} \left(\tilde{\mathbf{L}}^{31}\gamma_a + \tilde{\mathbf{L}}^{33}\gamma_c \right) \end{pmatrix}\end{aligned} \tag{4.47}$$

where the right-hand side of Eq.4.47 represents the known terms. Solution of Eq.4.47 returns the value of the unknown at all points except at the boundary and interface collocation points, for which Eq.4.45 must be used. Explicit and implicit methods lead to the same results, however, due to the simplicity of implementation, the former is to be preferred when solving differential equations (i.e. Poisson's Equation) whereas the implicit method can be exploited in the case of eigenvalue problems since it does not introduce spurious eigenvalues which must be accounted for in the case of explicit method as shown below.

4.3.4 Eigenvalue problems

As an example, let us consider and eigenvalue problem of the type:

$$\mathbf{H}\left(u(x)\right) = \lambda u(x) \tag{4.48}$$

over a single domain, where $u(x)$ is the unknown continuous and differentiable function over the x domain $[a, b]$, λ is a scalar value unknown of the problem and H is a linear operator (e.g. differential operator). Boundary conditions are supposed to be of the Dirichlet type and in particular:

$$u(x) = 0 \quad \text{for } x = a, b. \tag{4.49}$$

The discretized version of Eq.4.48 is written as

$$\mathbf{S}\overline{u} = \lambda \overline{u} \tag{4.50}$$

where S can be regarded as a Chebyshev differentiation matrix. By writing the unknown $\overline{u}$ as a vector of three components, the above eigenvalue problem becomes:

$$\begin{bmatrix} \mathbf{S}^{11} & \mathbf{S}^{12} & \mathbf{S}^{13} \\ \mathbf{S}^{21} & \mathbf{S}^{22} & \mathbf{S}^{23} \\ \mathbf{S}^{31} & \mathbf{S}^{32} & \mathbf{S}^{33} \end{bmatrix} \begin{bmatrix} u_a \\ \overline{u}_{inn} \\ u_b \end{bmatrix} = \lambda \begin{bmatrix} u_a \\ \overline{u}_{inn} \\ u_b \end{bmatrix}. \tag{4.51}$$

Boundary conditions can be imposed using either the implicit or explicit method, even though the former can be more convenient.

Implicit method

Based upon the steps in Sec.4.3.3 for the implicit method to impose boundary and continuity conditions, it is very simple to derive the eigenvalue problem to solve that becomes:

$$\mathbf{S}^{22}\overline{u}_{inn} = \lambda\overline{u}_{inn}. \tag{4.52}$$

Eigenvalues and eigenfunctions solutions of the problem in Eq.4.50 are obtained by solving Eq.4.52. A post processing of the results for the eigenfunctions is required to obtain the value of the eigenfunction at the boundaries, that in this case are simply zero.

Explicit method

By imposing explicitly the boundary conditions we obtain:

$$\begin{bmatrix} 1 & 0 & 0 \\ \mathbf{S}^{21} & \mathbf{S}^{22} & \mathbf{S}^{23} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} u_a \\ \overline{u}_{inn} \\ u_b \end{bmatrix} = \lambda \begin{bmatrix} 0 \\ \overline{u}_{inn} \\ 0 \end{bmatrix}. \tag{4.53}$$

Let us suppose to calculate the eigenvalues and eigenfunctions of the matrix in Eq.4.53. In this case we obtain:

$$\begin{bmatrix} 1 & 0 & 0 \\ \mathbf{S}^{21} & \mathbf{S}^{22} & \mathbf{S}^{23} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \Psi_a \\ \overline{\Psi}_{inn} \\ \Psi_b \end{bmatrix} = \varepsilon \begin{bmatrix} \Psi_a \\ \overline{\Psi}_{inn} \\ \Psi_b \end{bmatrix}. \tag{4.54}$$

Eigenvalues ε and eigenfunctions Ψ satisfy:

$$\begin{aligned} &1)\ \ \Psi_a = \varepsilon\Psi_a \\ &2)\ \ \mathbf{S}^{21}\Psi_a + \mathbf{S}^{22}\overline{\Psi}_{inn} + \mathbf{S}^{23}\Psi_b = \varepsilon\overline{\Psi}_{inn} \\ &3)\ \ \Psi_b = \varepsilon\Psi_b \end{aligned} \tag{4.55}$$

Conditions 1) and 3) are satisfied if:

$$\varepsilon = 1 \ \text{ or } \ \Psi_a = \Psi_b = 0 \tag{4.56}$$

In the case of $\varepsilon = 1$, condition 2) of Eq.4.55 becomes:

$$\overline{\Psi}_{inn} = -\left(\mathbf{S}^{22} - \mathbf{I}\right)^{-1}\left(\mathbf{S}^{21}\Psi_a + \mathbf{S}^{23}\Psi_b\right) \tag{4.57}$$

where $\mathbf{I}$ denotes the identity matrix of dimension equal to the number of elements of $\overline{\Psi}_{inn}$ and $\varepsilon = 1$ is the spurious eigenvalue (in this case its multiplicity is two).

In the case of $\Psi_a = \Psi_b = 0$, condition 2) of Eq.4.55 becomes:

$$\mathbf{S}^{22}\overline{\Psi}_{inn} = \varepsilon\overline{\Psi}_{inn}. \tag{4.58}$$

where it can be seen that Eq.4.58 is equal to Eq.4.52 obtained with the implicit method.

Eigenfunctions obtained with the explicit method by solving Eq.4.54, does not need a post-processing to add the values of the eigenfunction at the boundaries (and at the interface points in the case of heterostructures), however, this method requires a careful evaluation of the results to remove spurious eigenvalues and the corresponding eigenfunctions. Conversely, implicit method is, in general, of more difficult implementation and requires post processing to add eigenfunction values at the boundaries (and interface points in the case heterostructures), but it does not require to remove any spurious eigenvalue (and corresponding eigenfunction).

4.3.5 Pseudospectral Integral Calculation

According to [104], PS method allows to calculate integrals by employing the differentiation matrices used for the derivatives calculations. In fact, let the function $f(x)$ be continuous and differentiable over the domain $x \in [a, b]$. The integral:

$$I = \int_a^b f(x)dx \tag{4.59}$$

can be regarded as a particular case of differential equation of the type: $u' = f(x, u)$ where f is independent of u, thus the integral of $f(x)$ without loss of generality can be written as:

$$\frac{du(x)}{dx} = f(x) \quad \text{with } u(a) = 0 \text{ and } x > a. \tag{4.60}$$

Eq.4.60 states that the integral in Eq.4.59 is equal to $I = u(b)$. It follows that, according to the Chebyshev discretization matrices:

$$\mathbf{D}_{N+1}\overline{u}_{N+1} = \overline{f}_{N+1} \tag{4.61}$$

and by expanding into blocks the differentiation matrix we obtain:

$$\begin{bmatrix} \mathbf{D}^{11} & \mathbf{D}^{12} \\ \mathbf{D}^{21} & \mathbf{D}^{22} \end{bmatrix} \begin{bmatrix} u_a \\ \overline{u}_{inn} \end{bmatrix} = \begin{bmatrix} f_a \\ \overline{f}_{inn} \end{bmatrix} \tag{4.62}$$

where $u_a = u(x = a) = u(0)$ and $u_{inn} = u(x_i)$ with $i = 1, \ldots, N$. Boundary conditions on $u(x)$ reported in Eq.4.60 can be imposed as explained in Sec.4.3.4 by removing the first column and row of the differentiation matrix $\mathbf{D}_{N+1}$ obtaining:

$$\mathbf{D}^{22}\overline{u}_{inn} = \overline{f}_{inn}. \tag{4.63}$$

Since the relation $I = u(b)$ gives the value of the integral in Eq.4.59, it is necessary to invert the problem in Eq.4.63 to write explicitly the dependence of $\overline{u}_{inn}$ on $\overline{f}_{inn}$:

$$\overline{u}_{inn} = \tilde{\mathbf{D}}^{22}\overline{f}_{inn} \tag{4.64}$$

where $\tilde{\mathbf{D}}^{22}$ is the inverse of $\mathbf{D}^{22}$. Recalling that $u(b)$ equals to the last element of the vector $\overline{u}_{inn}$, the calculation of the integral I requires only the last row of the matrix $\tilde{\mathbf{D}}^{22}$

$$I = \overline{w}^T \, \overline{f}_{inn} \tag{4.65}$$

where $\overline{w}^T \in \mathbb{R}^{1\text{x}N}$ is the row vector given by the last row of the matrix $\tilde{\mathbf{D}}^{22}$.

The integral of the function $f(x)$ in Eq.4.59 can also be calculated using and alternative and better approach as reported in [104]. Recalling that within the PS method, a function is approximated by using a polynomial interpolation function (see Eq.4.15), the integral of the function $f(x)$ can be written as:

$$\int_a^b f(x)dx \approx \int_a^b p(x)dx = \sum_{j=0}^{N} \underbrace{\left(\int_a^b \ell_j(x)dx \right)}_{h_j} u(x_j), \quad x \in [a,b] \tag{4.66}$$

where the terms h_j represent the weights of the Clenshaw-Curtis quadrature formula. It can be demonstrated that Eq.4.66 gives marginally more accurate results than Eq.4.65 [104].

4.3.6 PS method for bounded periodic domains

The PS method can be used even in the presence of bounded, periodic grids. As for the case of non-periodic domains in Sec.4.3.1, the method starts by writing the periodic function $u(\theta) \in \mathbb{C}$ by using a polynomial expansion, and using a trigonometric interpolating polynomial instead of a Lagrange polynomial (see Sec.4.3.1).

The domain θ is $[0, 2\pi]$ and the number of grid points N on the periodic grid is supposed to be even, hence, $\theta_i = 2\pi i/N$ where $i = 1, \ldots, N$. The function $u(\theta)$ can thus be written similarly to Eq.4.15:

$$u(\theta) \approx p(\theta) = \sum_{j=1}^{N} S(\theta - \theta_j) u(\theta_j) \ . \tag{4.67}$$

where $S(\theta - \theta_j)$ is a unique trigonometric periodic function in L^2 of degree $N/2$ (i.e given by a linear combination of $e^{-i(N/2)\theta}, \ldots, e^{+i(N/2)\theta}$) that interpolates $u(\theta)$ in the the discrete points θ_i on the equispaced domain, and it is given by a periodic sinc function:

$$S(\theta) = \frac{sin\left(\frac{\theta N}{2}\right)}{N tan\left(\frac{\theta}{2}\right)}. \tag{4.68}$$

As for the Lagrange interpolant function of Sec.4.3.1 for nonperiodic domains, Eq.4.68 is such that $u(\theta_i) = p(\theta_i)$ since:

$$S(\theta_i - \theta_j) = \begin{cases} 1, & \theta_i = \theta_j \\ 0, & \theta_i \neq \theta_j. \end{cases}$$

The first derivative of $u(\theta)$ calculated in the collocation points $u(\theta_i)$ is given by:

$$u'(\theta_i) \approx p'(\theta_i) = \sum_{i,j=1}^{N} S'(\theta_i - \theta_j) u(\theta_j) \longrightarrow \overline{u}'_N \approx \mathbf{D}_{\theta,N} \overline{u}_N \tag{4.69}$$

where, according to [104], $\mathbf{D}_{\theta,N} \in \mathbb{R}^{N\mathrm{x}N}$ is a a skew−symmetric ($A^T = -A$) Toeplitz matrix where the last column of $\mathbf{D}_{\theta,N}$ is given by:

$$\mathbf{D}_{\theta,N}(i,N) = \begin{cases} 0, & i = 0(\mathrm{mod}N) \\ \dfrac{1}{2}(-1)^i cot\left(\dfrac{i\pi}{N}\right), & i \neq 0(\mathrm{mod}N) \end{cases}$$

Second derivative of $u(\theta)$ is calculated by using the second order differentiation matrix:

$$\overline{u}''_N = \mathbf{D}^2_{\theta,N} \overline{u}_N \tag{4.70}$$

where $\mathbf{D}^2_{\theta,N}$ can be calculated as the squared value of $\mathbf{D}_{\theta,N}$ according to Eq.4.21 requiring $O(N^3)$ floating point operations, or it can be calculated directly according to [104] obtaining a Toeplitz matrix where the last column of $\mathbf{D}^2_{\theta,N}$ is given by:

$$\mathbf{D}^2_{\theta,M}(i,M) = \begin{cases} -\dfrac{M^2}{12} - \dfrac{1}{6}, & i = 0(\mathrm{mod}M) \\ -\dfrac{(-1)^i}{2sin^2\left(\frac{i\pi}{M}\right)}, & i \neq 0(\mathrm{mod}M) \end{cases}$$

The operations required for the calculation of the derivative in Eq.4.70,4.71 over M points are $O(M^2)$. Moreover, since a periodic grid in the real space θ is much like an infinite grid with the semi-discrete Fourier transform, the derivatives can be calculated by means of the DFT requiring only $(NlogN)$ operations [104].

4.3.7 Integrals over periodic domains

Let us suppose to calculate the integral of a function periodic $f(\theta)$ over the domain $\theta \in [0, 2\pi]$ continuous and differentiable:

$$I = \int_0^{2\pi} f(\theta) d\theta. \tag{4.71}$$

By writing the function $f(\theta)$ by means of the trigonometric interpolant of Eq.4.67, we obtain:

$$I = \sum_{j=1}^{N} \underbrace{\left(\int_0^{2\pi} S(\theta - \theta_j) d\theta \right)}_{=\frac{2\pi}{N}} f(\theta_j) \tag{4.72}$$

leading to the final expression for the integral:

$$I = \frac{2\pi}{N} \sum_{j=1}^{N} f(\theta_j), \tag{4.73}$$

which corresponds to the trapezoidal rule for a periodic domain and it is the exact integral of the interpolant function $p(\theta)$ through the given data values $f(\theta_i)$. Moreover, it can be demonstrated that for smooth integrands the periodic trapezoid rule converges extraordinarily fast [104].

4.3.8 Two dimensional domains: rectangle

The PS method can be extended quite easily to a 2D domain. This is the case for a rectangular domain defined over $x = [a, b]$ and $y = [c, d]$. The x and y domains are set up on independent Chebyshev collocation points in each direction $x_i,\ i \in [0, 1, \ldots, N_x]$ and $y_i,\ i \in [0, 1, \ldots, N_y]$. We also suppose that the sorting of the vector of the functions $u(x, y)$ is $\overline{u}_{xy} = [u_{0,0}, \ldots, u_{0,N_x}, u_{1,0}, \ldots, u_{1,N_x} \ldots, u_{N_y,N_x}]^T$ with $\overline{u}_{xy} \in \mathbb{C}^{1,(N_x+1)(N_y+1)}$ hence, the points over the $x - y$ plane are sorted as in Fig.4.7 for $N_x = 6$ and $N_y = 5$. The approximated function over the 2D Chebyshev

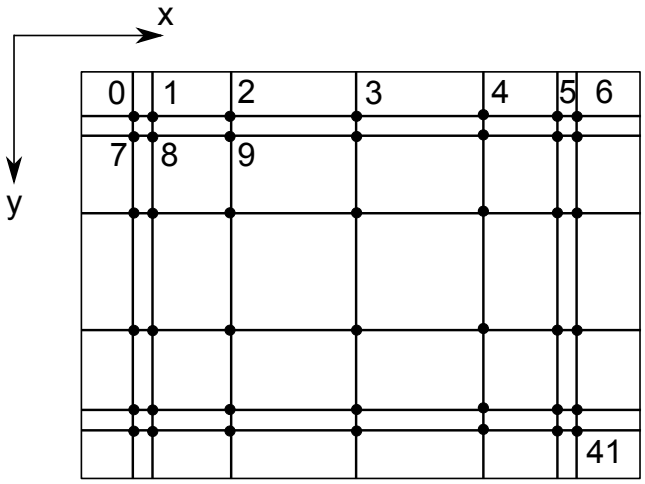


Figure 4.7: Chebyshev grid for a 2D domain with points sorting. $N_x = 6$ and $N_y = 5$, total number of Chebyshev points is $(N_x + 1)(N_y + 1)$.

domains is written as in Eq.4.15 resorting to the Lagrange basis polynomials:

$$u(x, y) \approx p(x, y) = \sum_{j=0}^{N_x} \ell_j(x) \sum_{k=0}^{N_y} \ell_k(y) u(x_j, y_k). \tag{4.74}$$

By exploiting the matrix formalism, Eq.4.74 can be rewritten by using the Kronecker product $\otimes$ between two matrices defined as:

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \otimes \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \left[\begin{array}{cc|cc} \text{a1} & \text{b1} & \text{a2} & \text{b2} \\ \text{c1} & \text{d1} & \text{c2} & \text{d2} \\ \hline \text{a3} & \text{b3} & \text{a4} & \text{b4} \\ \text{c3} & \text{d3} & \text{c4} & \text{d4} \end{array}\right] \tag{4.75}$$

Hence, according to the sorting of the unknowns and by defining $\mathbf{D}_{x,N_x+1}$ and $\mathbf{D}_{y,N_y+1}$ the differentiation matrices calculated according to Eq.4.20 for the x and y domains respectively, the 2D mixed derivative $\partial^2/(\partial x \partial y)$ can be written by using the differentiation matrix $\mathbf{D}^2_{xy}$ as:

$$\frac{\partial^2}{\partial x \partial y} \approx= \underbrace{\left(\mathbf{D}_{y,N_y+1} \otimes \mathbf{D}_{x,N_x+1}\right)}_{\mathbf{D}^2_{xy}} \overline{u}_{(N_x+1)(N_y+1)}. \tag{4.76}$$

Following the same approach that makes use of the Kronecker product, for the k-derivative along x or y coordinates, holds:

$$\begin{aligned} \mathbf{D}^k_{xx} &= \mathbf{I}_{N_y+1} \otimes \mathbf{D}^k_{x,N_x+1} \\ \mathbf{D}^k_{yy} &= \mathbf{D}^k_{y,N_y+1} \otimes \mathbf{I}_{N_x+1} \end{aligned} \tag{4.77}$$

where $\mathbf{I}_{N_m+1}$ denotes the identity matrix of dimensions $N_m + 1$.

4.3.9 PS method over a circular domain

For circular regions it is convenient to write a generic differential or eigenvalue equation by changing the coordinate system from Cartesian to polar. Moving to the discretized domain, this means that, according to Sec.4.3.1 and Sec.4.3.6 the domain can be discretized by using non periodic Chebyshev grid r_i with $i \in [0, 1, \ldots, N_r]$ along radial direction and a uniform periodic Fourier grid θ_k with $k \in [1, \ldots, N_\theta]$ along θ

$$r \in [0, a], \quad \theta \in [0, 2\pi]. \tag{4.78}$$

A generic function $u(r, \theta)$ can be approximated by the following equation:

$$u(r, \theta) \approx p(r, \theta) = \sum_{j=0}^{N_r} \ell_j(r) \sum_{k=1}^{N_\theta} S(\theta - \theta_k) u(r_j, \theta_k). \tag{4.79}$$

It must be noticed that the discretization of the circular domain in Fig.4.8 can be problematic for differential of eigenvalue problems due to the degenerate point at $r = 0$. To overcome this issue, among different possibilities the discretization of the domain is done in a different way considering r_i with $i = 0, 1, \ldots, N_r$ with N_r odd number and θ_k with $k = 1, \ldots, N_\theta$ with N_θ even number as reported in Sec.4.3.6 and:

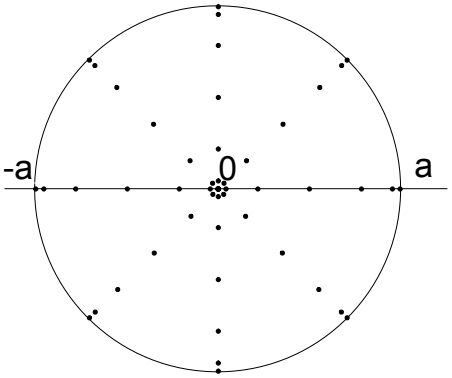


Figure 4.8: Discretization of a circular domain according to Eq.4.78 with $N_r = 6$ and $N_\theta = 8$.

$$r \in [-a, a], \quad \theta \in [0, 2\pi] \tag{4.80}$$

The sketch of the polar domain according to Eq.4.80 is shown in Fig.4.9 and the corresponding differentiation matrices for the first derivatives over the one-dimensional r and θ domain are:

$$\mathbf{D}_{r,N_r+1} \in \mathbb{R}^{(N_r+1)\mathrm{x}(N_r+1)}, \quad \mathbf{D}_{\theta,N_\theta} \in \mathbb{R}^{(N_\theta)\mathrm{x}(N_\theta)}. \tag{4.81}$$

Let us suppose that the radial Chebyshev points are sorted as in Fig.4.10 and that

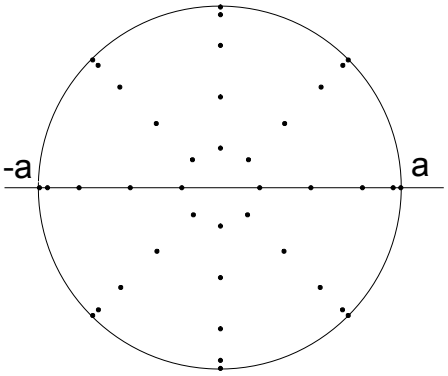


Figure 4.9: Discretization of a circular domain according to Eq.4.80 with $N_r = 9$ and $N_\theta = 8$

.

the vector $\overline{u}_{r,\theta} \in \mathbb{R}^{(N_r+1)N_\theta}$ is sorted as:

$$\overline{u}_{r,\theta} = \underbrace{u_{r_0,\theta_1}, \ldots, u_{r_0,\theta_{N_\theta}}, u_{r_1,\theta_0}, \ldots, u_{r_{\frac{N_r+1}{2}},\theta_{N_\theta}}}_{\overline{u}_{r>0,\theta}}, \underbrace{u_{r_{\frac{N_r+1}{2}+1},\theta_1}, \ldots, u_{r_{N_r},\theta_{N_\theta}}}_{\overline{u}_{r<0,\theta}} \tag{4.82}$$

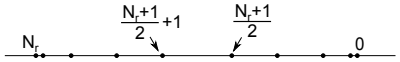


Figure 4.10: Sorting of the radial points.

leading to the numbering of the points on the circular domain reported in Fig.4.11. After defining the differentiation matrices for the r and θ domains, one would be

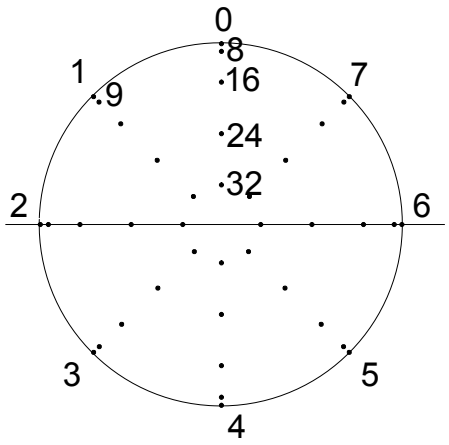


Figure 4.11: $r-\theta$ domain with sorting of discretization points.

tempted to calculate the the r, θ derivatives or mixed derivatives as reported in Sec.4.3.8 as:

$$\begin{aligned}
\frac{\partial u(r,\theta)}{\partial r} &\approx (\mathbf{D}_{r,N_r+1} \otimes \mathbf{I}_{N_\theta})\, \overline{u}_{r,\theta} = \mathbf{D}^{(2D)}_{r,(N_r+1)N_\theta} \overline{u}_{r,\theta} \\
\frac{\partial u(r,\theta)}{\partial \theta} &\approx (\mathbf{I}_{N_r+1} \otimes \mathbf{D}_{\theta,N_\theta})\, \overline{u}_{r,\theta} = \mathbf{D}^{(2D)}_{\theta,(N_r+1)N_\theta} \overline{u}_{r,\theta} \\
\frac{\partial^2 u(r,\theta)}{\partial r \partial \theta} &\approx (\mathbf{D}_{r,N_r+1} \otimes \mathbf{D}_{\theta,N_\theta})\, \overline{u}_{r,\theta} = \mathbf{D}^{(2D),2}_{r\theta,(N_r+1)N_\theta} \overline{u}_{r,\theta}
\end{aligned} \tag{4.83}$$

where the $(2D)$ apex denotes that the corresponding matrix must be applied to the function defined over the 2D polar domain. However, the discretization of the circular domain (see Eq.4.80) leads to redundant informations in the mapping of a generic $u(r,\theta)$ point, because:

$$u(r,\theta) = u(-r,\theta+\pi). \tag{4.84}$$

Fig.4.12 reports the $r-\theta$ space divided into four regions and it follows that regions $III-IV$ are clearly redundant with respect to regions $I-II$, or the vice-versa. Hence it will be used a simplification in the final differentiation matrices in order to remove any redundancy. To do so, let us suppose that the index i for radial points is $i \in 0,1,2,3$ and for θ points is $k = 1,2,3,4$ and we want to compute the first

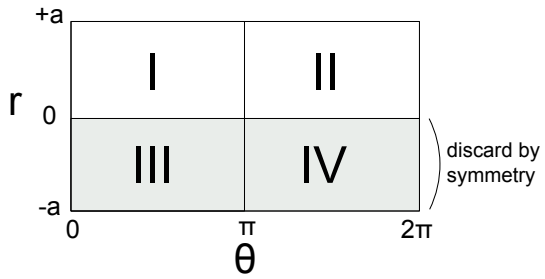


Figure 4.12: $r-\theta$ domain representation. Regions $III-IV$ are redundant with respect to regions $I-II$, or the viceversa, and can be removed.

derivative along radial direction. The differentiation matrix for the 1D radial domain is written into blocks:

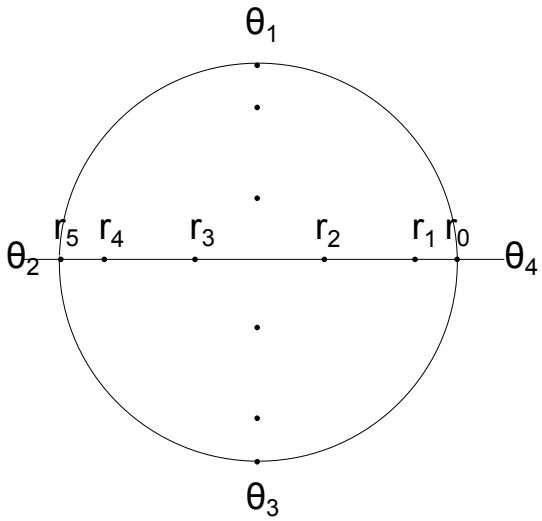


Figure 4.13: Disk discretization with r_i with $i \in 0,1,2,3,4,5$ and θ_k with $k = 1,2,3,4$.

$$\left[\begin{array}{ccc|ccc} \mathbf{D}^{11} & \mathbf{D}^{12} & \mathbf{D}^{13} & \mathbf{D}^{14} & \mathbf{D}^{15} & \mathbf{D}^{16} \\ \mathbf{D}^{21} & \mathbf{D}^{22} & \mathbf{D}^{23} & \mathbf{D}^{24} & \mathbf{D}^{25} & \mathbf{D}^{26} \\ \mathbf{D}^{31} & \mathbf{D}^{32} & \mathbf{D}^{33} & \mathbf{D}^{34} & \mathbf{D}^{35} & \mathbf{D}^{36} \\ \hline \mathbf{D}^{41} & \mathbf{D}^{42} & \mathbf{D}^{43} & \mathbf{D}^{44} & \mathbf{D}^{45} & \mathbf{D}^{46} \\ \mathbf{D}^{51} & \mathbf{D}^{52} & \mathbf{D}^{53} & \mathbf{D}^{54} & \mathbf{D}^{55} & \mathbf{D}^{56} \\ \mathbf{D}^{61} & \mathbf{D}^{62} & \mathbf{D}^{63} & \mathbf{D}^{64} & \mathbf{D}^{65} & \mathbf{D}^{66} \end{array}\right] \left[\begin{array}{l} u(r_0) \\ u(r_1) \\ u(r_2) \\ u(r_3) \\ u(r_4) \\ u(r_5) \end{array}\right] \begin{array}{l} \left.\begin{array}{l} \\ \\ \end{array}\right\} r^{>0} \\ \left.\begin{array}{l} \\ \\ \end{array}\right\} r^{<0} \end{array} = \begin{bmatrix} du(r)/dr|_0 \\ du(r)/dr|_1 \\ du(r)/dr|_2 \\ du(r)/dr|_3 \\ du(r)/dr|_4 \\ du(r)/dr|_5 \end{bmatrix} \tag{4.85}$$

that written in a more compact form becomes:

$$\begin{bmatrix} \tilde{\mathbf{D}}^{11}_{M_r} & \tilde{\mathbf{D}}^{12}_{M_r} \\ \tilde{\mathbf{D}}^{21}_{M_r} & \tilde{\mathbf{D}}^{22}_{M_r} \end{bmatrix} \begin{bmatrix} \overline{u}_{r>0} \\ \overline{u}_{r<0} \end{bmatrix} = \begin{bmatrix} d\overline{u}/dr|_{r>0} \\ d\overline{u}/dr|_{r<0} \end{bmatrix} \tag{4.86}$$

where $\tilde{\mathbf{D}}^{ij}_{M_r} \in \mathbb{R}^{M_r \mathrm{x} M_r}$ and $M_r = (N_r+1)/2$ with N_r odd number. The first derivative along r in the 1D r space is given by:

$$\mathbf{D}^{11}u(r_0) + \mathbf{D}^{12}u(r_1) + \mathbf{D}^{13}u(r_2) + \mathbf{D}^{14}u(r_3) + \mathbf{D}^{15}u(r_4) + \mathbf{D}^{16}u(r_5). \tag{4.87}$$

If we now move to the 2D domain where the sorting of the points in polar coordinates is the one reported in Eq.4.82, the first derivative along r for a fixed θ value (i.e. θ_1) at the point (r_0, θ_1) can be written by taking the product of the first row of the matrix $\mathbf{D}^{(2D)}_{r,(N_r+1)N_\theta}$ defined in Eq.4.83 by the column vector $\overline{u}_{r,\theta}$:

$$\mathbf{D}^{11}u(r_0,\theta_1)+\mathbf{D}^{12}u(r_1,\theta_1)+\mathbf{D}^{13}u(r_2,\theta_1)+\mathbf{D}^{14}u(r_3,\theta_1)+\mathbf{D}^{15}u(r_4,\theta_1)+\mathbf{D}^{16}u(r_5,\theta_1) \tag{4.88}$$

Redudancy in Eq.4.88 can be removed thanks to Eq.4.84, in fact:

$$\begin{aligned} u(r_3,\theta_1) &= u(r_2,\theta_3) \\ u(r_4,\theta_1) &= u(r_1,\theta_3) \\ u(r_5,\theta_1) &= u(r_0,\theta_3). \end{aligned} \tag{4.89}$$

By substituting Eq.4.89 into Eq.4.88, the first derivative along r at the point (r_0, θ_1) is given by:

$$\mathbf{D}^{11}u(r_0,\theta_1)+\mathbf{D}^{12}u(r_1,\theta_1)+\mathbf{D}^{13}u(r_2,\theta_1)+\mathbf{D}^{14}u(r_2,\theta_3)+\mathbf{D}^{15}u(r_1,\theta_3)+\mathbf{D}^{16}u(r_0,\theta_3) \tag{4.90}$$

and it can be seen that Eq.4.90 depends only on positive r values, that is, the redundancy in Eq.4.88 has been removed (as shown in Fig.4.12). By extending the same procedure for all other points of the (r, θ) domain, the first derivative along r direction and by defining $M_r = (N_r + 1)/2$, can be written as:

$$\mathbf{D}^{(2D)}_{r,M_rN_\theta}\overline{u}_{r>0,\theta} = \begin{bmatrix} \mathbf{D}^{11} & \mathbf{D}^{16} & \mathbf{D}^{12} & \mathbf{D}^{15} & \mathbf{D}^{13} & \mathbf{D}^{14} \\ \mathbf{D}^{16} & \mathbf{D}^{11} & \mathbf{D}^{15} & \mathbf{D}^{12} & \mathbf{D}^{14} & \mathbf{D}^{13} \\ \mathbf{D}^{21} & \mathbf{D}^{26} & \mathbf{D}^{22} & \mathbf{D}^{25} & \mathbf{D}^{23} & \mathbf{D}^{24} \\ \mathbf{D}^{26} & \mathbf{D}^{21} & \mathbf{D}^{25} & \mathbf{D}^{22} & \mathbf{D}^{24} & \mathbf{D}^{23} \\ \mathbf{D}^{31} & \mathbf{D}^{36} & \mathbf{D}^{32} & \mathbf{D}^{35} & \mathbf{D}^{33} & \mathbf{D}^{34} \\ \mathbf{D}^{36} & \mathbf{D}^{31} & \mathbf{D}^{35} & \mathbf{D}^{32} & \mathbf{D}^{34} & \mathbf{D}^{33} \end{bmatrix} \begin{bmatrix} u_{r_0,\theta_1} \\ u_{r_0,\theta_2} \\ u_{r_0,\theta_3} \\ u_{r_0,\theta_4} \\ u_{r_1,\theta_1} \\ u_{r_1,\theta_2} \\ u_{r_1,\theta_3} \\ u_{r_1,\theta_4} \\ u_{r_2,\theta_1} \\ u_{r_2,\theta_2} \\ u_{r_2,\theta_3} \\ u_{r_2,\theta_4} \end{bmatrix}. \tag{4.91}$$

where $\mathbf{D}^{(2D),k}_{r,M_rN_\theta} \in \mathbb{R}^{M_rN_\theta \mathrm{x} M_rN_\theta}$ and $\overline{u}_{r>0,\theta} \in \mathbb{R}^{M_rN_\theta}$ with $\overline{u}_{r>0,\theta} = u_{r_0,\theta_1}, \ldots, u_{r_0,\theta_{N_\theta}}, u_{r_1,\theta_0}, \ldots, u_{r_{M_r},\theta_{N_\theta}}$. Eq.4.91 can be rewritten for a k order derivative along r direction in a more compact form by resorting to the matrix partitioning of Eq.4.86:

$$\mathbf{D}^{(2D),k}_{r,M_rN_\theta} = \tilde{\mathbf{D}}^{11,k}_{M_r} \otimes \begin{bmatrix} I_{N_\theta/2} & 0 \\ 0 & I_{N_\theta/2} \end{bmatrix} + \tilde{\mathbf{E}}^{12,k}_{M_r} \otimes \begin{bmatrix} 0 & I_{N_\theta/2} \\ I_{N_\theta/2} & 0 \end{bmatrix} \tag{4.92}$$

where $\tilde{\mathbf{E}}^{12,k}_{M_r}$ denotes the $\tilde{\mathbf{D}}^{12,k}_{M_r} \in \mathbb{R}^{M_r \mathrm{x} M_r}$ obtained by partitioning the matrix for the k derivative along radial direction ($\tilde{\mathbf{D}}^{k}_{r,N_r+1}$ is obtained as reported in Eq.4.22) according to Eq.4.86 where the columns of $\tilde{\mathbf{D}}^{12,k}_{M_r}$ have been left-to-right flipped: $\tilde{\mathbf{E}}^{12,k}_{M_r}(:,i)$=$\tilde{\mathbf{D}}^{12,k}_{M_r}(:, M_r - 1 - i)$ with $i \in 0, 1, \ldots, M_r - 1$. The first term of the summation in Eq.4.92 is given by the matrix coefficients circled with solid line in Eq.4.91 whereas the second term by the dashed-line circled elements in Eq.4.91.

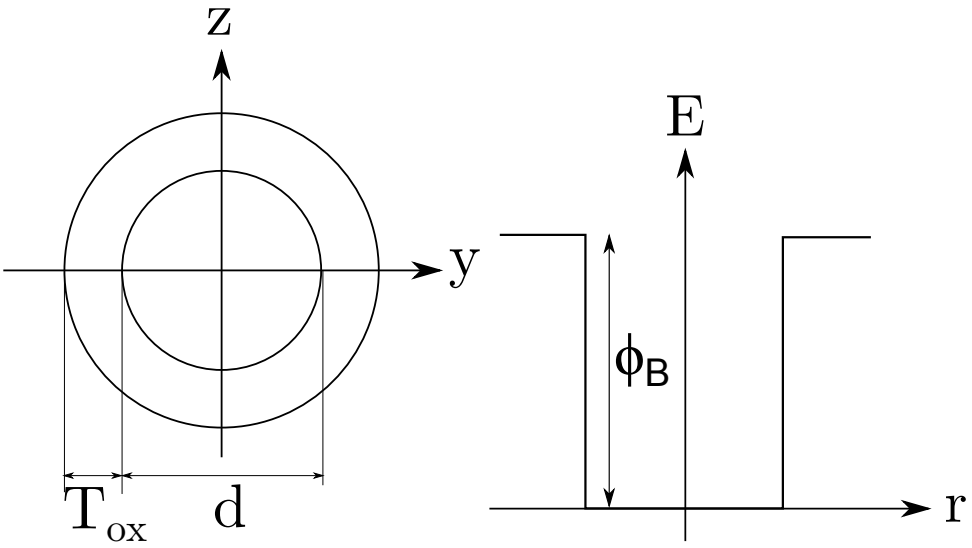
It is interesting to note that the initial matrix for the first derivative along r ,$\mathbf{D}^{(2D)}_{r,(N_r+1)N_\theta}$, reported in Eq.4.83 has dimensions $(N_r + 1)N_\theta$ and returns redundant results due to the non-unique mapping of the generic point (r, θ) in the circular domain given by the choice of radial discretization points reported in Eq.4.80 (that has been introduced to avoid the degenerate discretization point at $r = 0$), whereas, according to the considerations sketched in Fig.4.12 and after some manipulations of the $\mathbf{D}^{(2D)}_{r,(N_r+1)N_\theta}$ matrix, it is obtained the new matrix for the first derivative along r ,$\mathbf{D}^{(2D)}_{r,M_rN_\theta}$, reported in Eq.4.92 (first derivative is obtained by setting $k = 1$) with reduced dimensions and without redundancy. Based upon same considerations for the θ and mixed derivatives, and by defining $M_r = (N_r + 1)/2$ the new differential operators on the circular domain become:

$$\begin{aligned}
\frac{\partial^k u(r,\theta)}{\partial r^k} &\approx \left(\tilde{\mathbf{D}}^{11,k}_{M_r} \otimes \begin{bmatrix} I_{\frac{N_\theta}{2}} & 0 \\ 0 & I_{\frac{N_\theta}{2}} \end{bmatrix} + \tilde{\mathbf{E}}^{12,k}_{M_r} \otimes \begin{bmatrix} 0 & I_{\frac{N_\theta}{2}} \\ I_{\frac{N_\theta}{2}} & 0 \end{bmatrix} \right) \overline{u}_{r,\theta} = \mathbf{D}^{(2D),k}_{r,M_rN_\theta} \overline{u}_{r,\theta} \\
\frac{\partial^k u(r,\theta)}{\partial \theta^k} &\approx \left(\mathbf{I}_{M_r} \otimes \mathbf{D}_{\theta,N_\theta} \right) \overline{u}_{r,\theta} = \mathbf{D}^{(2D),k}_{\theta,M_rN_\theta} \overline{u}_{r,\theta} \\
\frac{\partial^{k+l} u(r,\theta)}{\partial r^k \partial \theta^l} &\approx \left(\tilde{\mathbf{D}}^{11,k}_{M_r} \otimes \mathbf{D}^{l}_{\theta N_\theta} + \tilde{\mathbf{E}}^{12,k}_{M_r} \otimes \mathbf{F}^{\mathbf{l}}_{\theta,\mathbf{N}_\theta} \right) \overline{u}_{r,\theta} = \mathbf{D}^{(2D),2}_{r\theta,M_rN_\theta} \overline{u}_{r,\theta}
\end{aligned} \tag{4.93}$$

where

$$\begin{aligned}
\tilde{\mathbf{E}}^{12,k}_{M_r}(:,i) &= \tilde{\mathbf{D}}^{12,k}_{M_r}(:, M_r - 1 - i), \quad i \in 0, 1, \ldots, M_r - 1 \\
\mathbf{F}^{l}_{\theta,N_\theta} &= \left[\mathbf{D}^{l}_{\theta,N_\theta}(:, \tfrac{N_\theta}{2} : N_\theta - 1) \quad \mathbf{D}^{l}_{\theta,N_\theta}(:, 0 : \tfrac{N_\theta}{2} - 1). \right]
\end{aligned} \tag{4.94}$$

4.4 Numerical solution of the Schrödinger equation



(a) Nanowire section.

(b) Flat potential profile with potential barrier Φ_B.

Figure 4.14: Sketch of the simulated nanowire.

4.4.1 Simulation results: ideal circular quantum well without wave-function penetration into the oxide

Isotropic quantization masses

For ideal quantum wells with a flat potential profile $U(r,\theta) = 0eV$, without wave-function penetration into the oxide, for isotropic quantization masses, and without taking into account non parabolicity corrections, the eigenvalues of Eq.4.1 can be analytically calculated:

$$\varepsilon_{n,m} = \frac{\hbar^2}{2m}\left(\frac{2\mu_{m,n}}{d}\right)^2 \tag{4.95}$$

where d is the nanowire diameter, $\mu_{m,n}$ is the $n-$th zero of the Bessel function of the first kind and order m, with $m = 0, \pm1, \pm2, \pm3, \ldots$. The exact solutions of the problem have been used as the reference to evaluate the accuracy of the proposed PS numerical method. Since m in Eq.4.95 can be either positive or negative, there are couples of degenerate eigenvalues as shown in Tab.4.1. Fig.4.15 shows the square modulus of the calculated eigenfunctions with the PS method and Fig.4.16 shows the comparison between analytical eigenvalues obtained with Eq.4.95 and numerical eigenvalues obtained with the PS method described in Sec.4.3. Simulation results are for silicon with [100] transport direction and for $\Delta_{[100]}$ valleys, namely, those

n. eigenvalue	energy [eV]
1	**0.1855**
2	**0.4711**
3	**0.4711**
4	**0.8462**
5	**0.8462**
6	**0.9776**

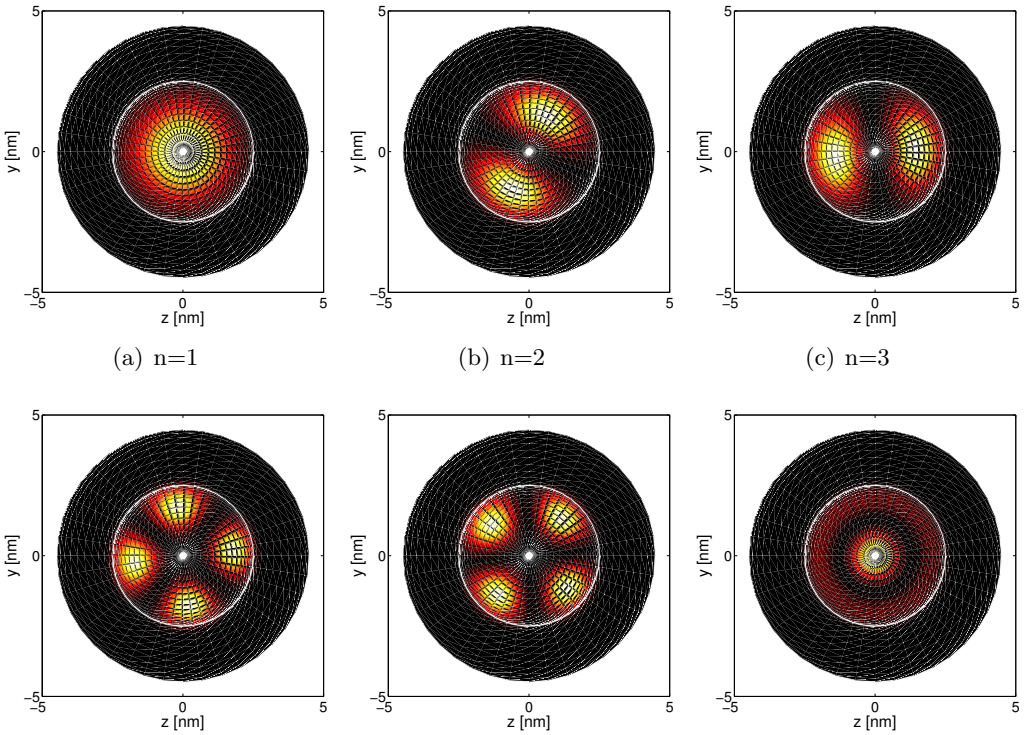
Table 4.1: Eigenvalues obtained with Eq.4.95 for silicon with [100] transport direction and for the valleys with isotropic quantization masses (i.e. those with the main ellipsoid axes aligned with transport direction with m_x=0.92 m_0 and m_{quant}=0.19 m_0). d is 5 nm and results are obtained without wave-function penetration into the oxide.

with the main axes aligned along transport direction. In Fig.4.16b it can be noticed that the PS method used in this thesis allows an exponential decrease of the error with the number of discretization points. Fig.4.17 shows the comparison between numerical and analytical eigenvalues for different nanowire diameters.

4.4.2 Simulation results: Ideal circular quantum well with wave-function penetration into the oxide

In this section we show simulation results obtained under different approximations for the oxide that can be found in the literature and without including non-parabolicity corrections. We solve the Schrödinger equation in two cases: assuming an infinite potential barrier Φ_B between semiconductor and oxide, that is, neglecting the wave-function penetration into the oxide, or allowing the wave-function to penetrate the oxide. In the latter case, we still have two options: a) approximate the oxide using the same semiconductor effective masses but using the electron affinity of the considered oxide (this allows to impose only the continuity of the envelope wave-function and of its derivative at the semiconductor-oxide interface); b) simulate the oxide using the proper effective mass and electron affinity which requires more complicated continuity conditions.

Fig.4.18 shows that, expecially for the InAs case, either the assumption of infinite potential barrier Φ_B between oxide and semiconductor (i.e. no wave-function penetration into the oxide), or, the approximation for the oxide material using the semiconductor effective mass but with the electron affinity of the oxide, lead to significant differences compared to the complete treatments (triangles). The discrepancy between triangles and circles or squares in Fig.4.18 is stronger for InAs with respect to silicon and this is expected due to the low effective mass of the InAs. Therefore, it seems very important to properly account for the oxide material by using the correct value of the effective mass, hence, by imposing the continuity conditions of Eq.4.9. We here notice that, for both silicon with [100] transport direction and

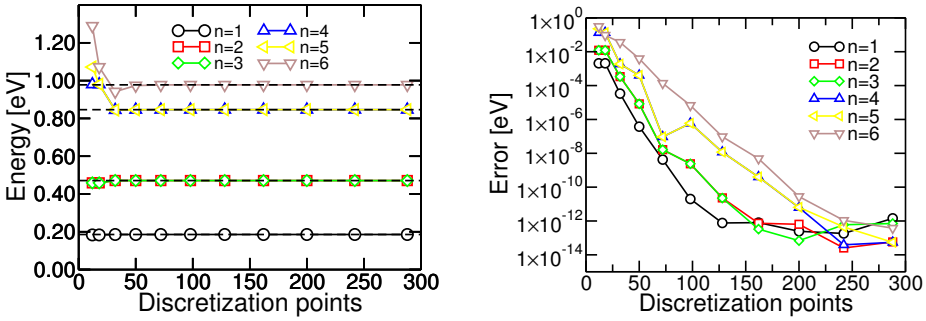


(a) n=1 (b) n=2 (c) n=3

(d) n=4 (e) n=5 (f) n=6

Figure 4.15: Square modulus of the wave-function for different eigenvalues n, for silicon nanowire with [100] transport direction and diameter d of 5 nm and T_{ox}=2 nm.

for the fundamental isotropic valley for InAs, the tensor of the effective masses is isotropic (see Fig.A.4), therefore, the continuity condition in Eq.4.9, that in principle leads to a dependence of the eigenvalues on k_x, reduces to the much more simple Eq.4.10 that does not involve k_x.



(a) Eigenvalues versus discretization points. Dashed-lines are the analytical eigenvalues of Tab.4.1

(b) Absolute error between numerical and analytical results.

Figure 4.16: Results obtained for silicon with [100] transport direction and for the valleys with isotropic quantization masses (i.e. those with the main ellipsoid axes aligned with transport direction). d is set to 5 nm, the semiconductor-oxide barrier height is assumed to be infinite and we used a parabolic dispersion relation. Numerical results are obtained for different number of discretization points.

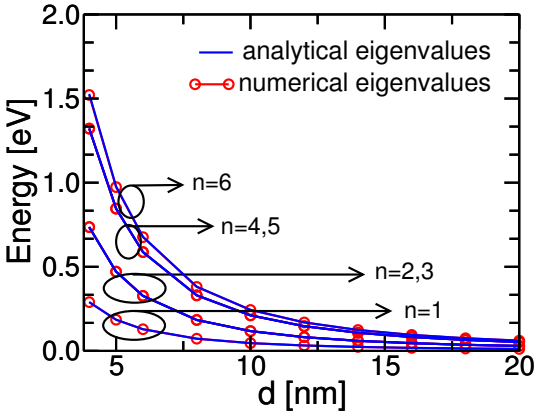
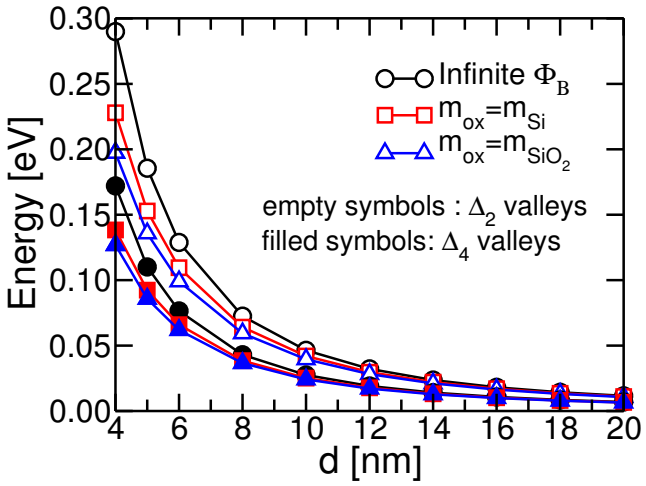
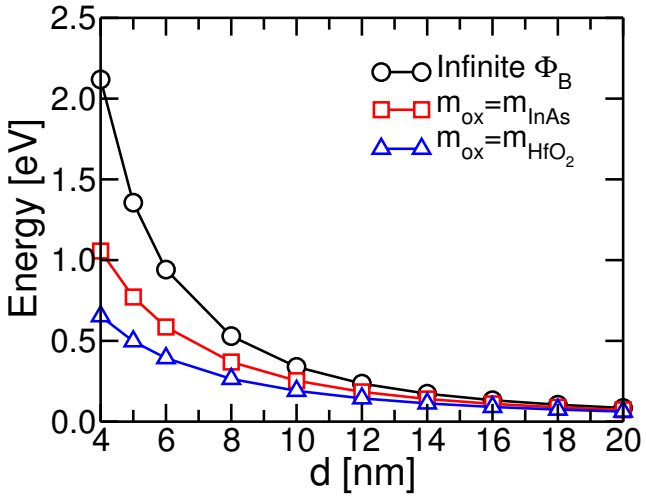


Figure 4.17: Comparison between numerical and analytical eigenvalues for different NW diameters d. Analytical eigenvalues are obtained with Eq.4.95, whereas numerical eigenvalues with the PS method and with 50 discretization points. Results obtained for silicon with [100] transport direction and for the valleys with isotropic quantization masses (i.e. those with the main ellipsoid axes aligned with transport direction). We assume the semiconductor-oxide barrier height to be infinite and a parabolic dispersion relation. n denotes the eigenvalue index.



(a) Silicon−SiO_2 with 100 transport direction. For the two cases of wave-function penetration (squares and triangles) Φ_B=3.15 eV. The oxide mass is m_{Si02}=0.5 m_0.



(b) InAs-$Hf0_2$ with only Γ valleys. For the two cases of wave-function penetration (squares and triangles) Φ_B=2.4eV. The oxide mass is m_{Hf02}=0.11 m_0 [106]

Figure 4.18: Comparison between parabolic eigenvalues for different nanowire diameters obtained for silicon and InAs. Circles: results obtained by assuming an infinite potential barrier Φ_B between oxide and semiconductor. Squares: results obtained by approximating the oxide by using the semiconductor effective mass but with the correct electron affinity for the oxide. Triangles: results obtained including the effective mass of the oxide. Results obtained for a flat potential, with T_{ox}=2nm.

Silicon [111]

So far we have considered two cases (silicon with [100] transport direction and InAs with the fundamental Γ valley) where the continuity conditions do not involve k_x. We here consider the case of Silicon with [111] transport direction and wave-function penetration into the SiO_2 oxide. The continuity conditions between the semiconductor and the oxide for the solution of the Schrödinger equation are given in Eq.4.9 and, it follows that, given the tensor of the effective masses in DCS for the three ellipsoidal $\Delta-$valleys reported in Fig.A.4, k_x enters the continuity conditions through the coefficients $A(\theta), B(\theta)$ and $C(\theta)$ in Eq.4.9. The lowest eigenvalue for the three ellipsoid as a function of k_x is shown in Fig.4.19 and, as expected, it is threefold degenerate for $k_x = 0$. Due to the crystal orientation specified in Fig.4.19,

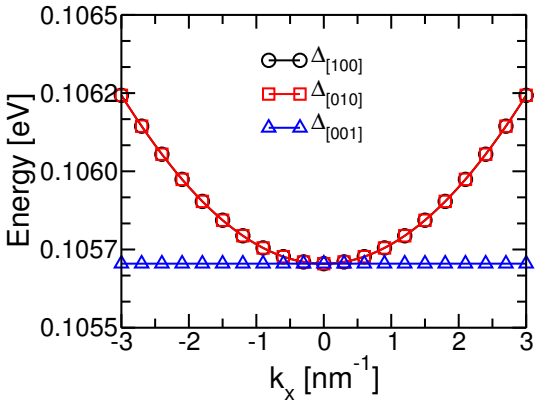


Figure 4.19: Parabolic eigenvalues versus k_x for the three Δ valleys with transport direction x=[111]; the quantization plane coordinates are y=[$\bar{1}$10] and z=[$\bar{1}\bar{1}$2] where the miller indexes are given in CCS. d=5 nm and the oxide is SiO_2 with T_{ox}= 2 nm.

only $\Delta_{[100]}$ and $\Delta_{[010]}$ valleys depend on the k_x vector. Moreover, subband minima are weakly affected by the k_x even at large k_x values, hence we decided to neglect the k_x dependence on the boundary conditions using, for all simulations in this thesis, Eq.4.10 instead of Eq.4.9. It must be noticed that Eq.4.10 is exact when the entries w_{xy} and w_{xz} of the tensor of the inverse effective masses are zero (see Eq.A.11), that is for isotropic valleys in the quantization plane or, for example, for silicon with [100] transport direction.

To further investigate our choice to neglect the k_x dependence for subband minima due to the continuity conditions, we show the energy dispersion relation $E(k_x)$ for two cases (with the energy referred to the valley energy minimum). In the fist the total energy dispersion relation for a carrier inside a silicon $\Delta-$valley with [111] transport direction without non parabolic correction is written as:

$$E(k_x) = \varepsilon_{\nu,n}(k_x) + \frac{h^2 k_x^2}{2m_x} \tag{4.96}$$

where the k_x dependence of $\varepsilon_{\nu,n}$ derives from the continuity conditions of Eq.4.9

whereas, in the second, we neglect the dependence of the eigenvalues from k_x using the continuity conditions of Eq.4.10:

$$E(k_x) = \varepsilon_{\nu,n} + \frac{h^2 k_x^2}{2m_x} \tag{4.97}$$

According to Eq.B.33 and to the coefficient of the tensor of the inverse mass in DCS of Fig.A.4, m_x equals to 0.4320 m_0 for all the $\Delta-$valleys. Fig.4.20 shows that the deviation between results obtained by using Eq.4.96 and Eq.4.97 is almost negligible, hence confirming our defendable choice to use a simplified version of Eq.4.9 for the continuity conditions given by Eq.4.10. Therefore, in the presence of wave-function penetration into the oxide, even in the general case, k_x does not enter the continuity equation, hence, eigenvalues and $\xi_n(r,\theta)$ (see Eq.4.1) do not depend on k_x leading more manageable equations for scattering rates.

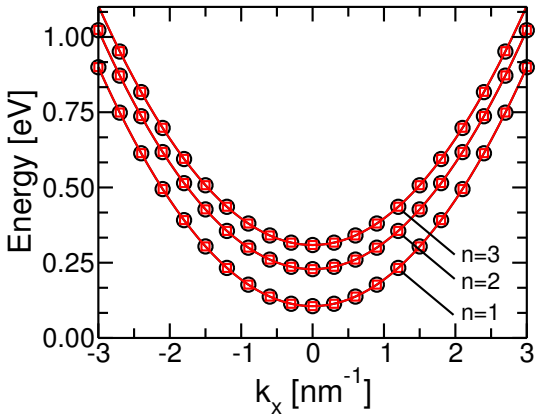


Figure 4.20: Energy dispersion relation for silicon with [111] transport direction, for the $\Delta_{[100]}$ valley and for the lowest three eigenvalues. Transport direction x is [111], and the quantization plane coordinates are y=[$\bar{1}$10] and z=[$\bar{1}\bar{1}$2]. d=5 nm and the oxide is SiO_2 with T_{ox}= 2 nm. Results obtained using Eq.4.96 (squares) and Eq.4.97 (circles).

Fig.4.21 shows the first nine eigenvalues obtained for silicon with [111] transport direction and, as expected, are three-fold dengerate

4.4.3 Simulation results: nonparabolicity correction

Simulation results in this section are obtained for the system Si-SiO_2 and InAs-HfO_2 by using the continuity conditions in Eq.4.10. Fig.4.22 shows that the impact of nonparabolicity corrections is stronger for InAs and it is due to a combination of lower effective mass with respect to silicon and a larger non parabolic coefficient α as shown in Tab.2.1.

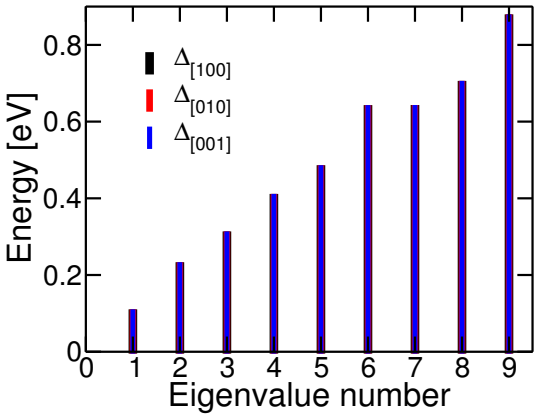
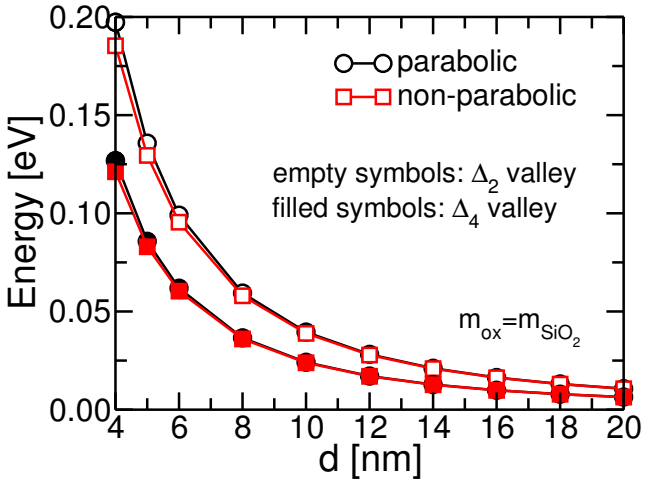
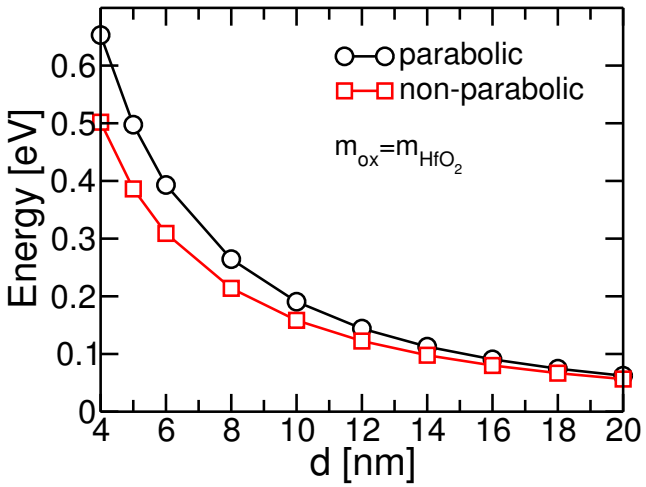


Figure 4.21: Eigenvalues for the three Δ valleys with transport direction x=[111], y=[$\bar{1}$10] and z=[$\bar{1}\bar{1}$2] where the miller indexes are given in CCS. d=5nm and the oxide is SiO_2 with T_{ox}= 2 nm. Results obtained by using continuity conditions of Eq.4.10.



(a) Silicon−SiO_2 with 100 transport direction. Φ_B is 3.15 eV and the oxide mass is m_{Si02}=0.5 m_0.



(b) InAs-$Hf0_2$ with only Γ valleys. Φ_B is 2.4eV and the oxide mass is m_{Hf02}=0.11 m_0 [106]

Figure 4.22: Parabolic (circles) and nonparabolic (squares) eigenvalues obtained with Eq.C.38 for different nanowire diameters for silicon and InAs. Bulk semiconductor parameters are given in Tab.2.1. Results obtained for a flat potential, with T_{ox}=2nm.

4.5 Self consistent Schrödinger - Poisson equation in a circular domain

The electrostatic behavior of the device in Fig.4.3 is simulated solving the Schrödinger equation Eq.4.1 self-consistently with the Poisson equation in the circular slice obtained by writing the laplacian of Poisson equation in polar coordinates as in Eq.4.3 obtaining:

$$\varepsilon\left(\frac{\partial^2}{\partial r^2}+\frac{1}{r}\frac{\partial}{\partial r}+\frac{1}{r^2}\frac{\partial^2}{\partial\theta^2}\right)\phi(r,\theta)=-\rho(r,\theta)$$
$$=e(-n(r,\theta)+p(r,\theta)+N_D(r)+N_A(r)) \quad (4.98)$$

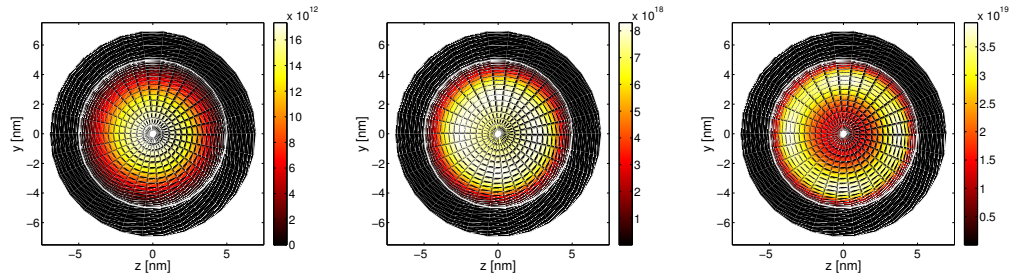
where $\phi(r,\theta)$ is the potential, $\rho(r,\theta)$ is the charge density, n is the electron density, p is the hole density (that in this thesis is neglected since the aim is to investigate the properties of nMOS devices that are fully depleted), N_D is the density of donor dopants and N_A the density of acceptor dopants. For the boundary conditions of the Poisson equation, we impose Dirichlet boundary conditions at the boundaries of the nanowire, and, at the discontinuity between oxide and semiconductor, we simply use Gauss law.

The coupled solution of the Schrödinger and Poisson equation requires an iterative loop. At the iteration k, we solve the Schrödinger equation in a circular slice using Eq.4.1 for the potential $\phi^{(k)}$ (at the first iteration we use a first guess for the potential), then, with the obtained wave-functions and eigenvalues we calculate the charge by using Eq.C.42 that includes nonparabolicity corrections. Now, in order to obtain an efficient convergence of the Schrödinger-Poisson loop, instead of using the linear formulation of the Poisson equation given in Eq.4.98, we employed a non-linear formulation of the Poisson equation [107], where we assume that the electron charge n for a given point in the (r,θ) space, exponentially depends on the potential:

$$n(\phi)=n\,exp^{\frac{e\left(\phi-\phi^{(k)}\right)}{k_BT}} \quad (4.99)$$

where $\phi^{(k)}$ is the potential used as an input for the Schrödinger equation, that is the potential produced by the Poisson solver at the previous iteration. With the new potential $\phi^{(k+1)}$ obtained by solving Eq.4.98 by substituting the electron charge n with Eq.4.99, we solve the Schrödinger equation till convergence is reached.

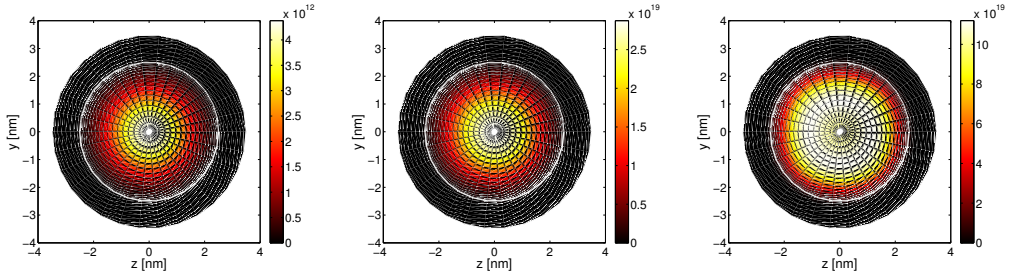
Fig.4.23 shows the electron density per unit volume at the equilibrium in a circular section of a silicon nanowire with a diameter d of 10 nm and 5 nm, [100] transport direction and with wave-function penetration into a 2 nm thick SiO_2 obtained with a self-consistent Schrödinger-Poisson solution in a circular slice. It can be seen that the electron density is appreciably anisotropic for large V_{GS} biases and $d = 10$ nm (Fig.4.23c), because the squared magnitude of the wave-function associated with the lowest eigenvalues with transport mass m_x=0.19 m_0 and anisotropic quantization masses featuring m_y=0.92 m_0 and m_z=0.19m_0 (or m_y=0.19 m_0 and m_z=0.92 m_0) is strongly anisotropic.



(a) d=10 nm, T_{ox}=2 nm, V_{GS}=0.5 V. N_{inv} = 6.028 cm^{-1}.

(b) d=10 nm, T_{ox}=2 nm, V_{GS}=1.0 V. $N_{inv} = 0.417 \times 10^7$ cm^{-1}.

(c) d=10 nm, T_{ox}=2 nm, V_{GS}=1.5 V. $N_{inv} = 1.843 \times 10^7$ cm^{-1}.



(d) d=5 nm, T_{ox}=1 nm, V_{GS}=0.5 V. N_{inv} = 0.290 cm^{-1}.

(e) d=5 nm, T_{ox}=1 nm, V_{GS}=1.0 V. $N_{inv} = 0.220 \times 10^7$ cm^{-1}.

(f) d=5 nm, T_{ox}=1 nm, V_{GS}=1.5 V. $N_{inv} = 1.429 \times 10^7$ cm^{-1}.

Figure 4.23: Volumetric charge density in units of [cm^{-3}] for Silicon$-SiO_2$ with [100] transport direction for different V_{GS} and nanowire diameters d and with the same gate oxide capacitance $C_{ox} = 2\pi\varepsilon_{ox}/ln[(r_{NW} + T_{ox})/r_{NW}]$. Φ_B is 3.15 eV and the oxide mass is m_{Si02}=0.5 m_0.

4.6 Scattering Mechanisms

In this section we give the explicit equations for the matrix elements for acoustic and optical phonon scattering and Coulomb scattering whereas the derivation of the matrix element for surface roughness has been already discussed in Sec.3.6. Then, scattering rates are calculated by means of Eq.5.24.

4.6.1 Matrix element for acoustic and optical phonon scattering

By embracing the elastic and isotropic approximation, the squared matrix element for the intra-valley acoustic phonon is given by [107]:

$$\left|\mathcal{M}_{n,n'}\right|^2 = \frac{D_{ac}^2\, k_B T}{\rho L_{NW} v_{sound}^2} F_{n,n'} \tag{4.100}$$

where D_{ac} is an effective deformation potential for acoustic phonons, ρ is the density of the semiconductor material, v_{sound} is the longitudinal sound velocity in the semiconductor material and $F_{n,n'}$ is a form factor defined as:

$$F_{n,n'} = \int_{-\pi}^{+\pi} d\theta \int_{0}^{\infty} dr\; r\; \left|\xi_n(r,\theta)\right|^2 \left|\xi_{n'}(r,\theta)\right|^2 \tag{4.101}$$

where $\xi_n(r,\theta)$ is the solution of the eigenvalue problem of Eq.4.1.

For intra-valley optical phonon scattering, the squared matrix element for the transition from state k_x to k_x' is given by [107]:

$$\left|\mathcal{M}_{n,n'}(k_x,k_x')\right|^2 = \frac{D_{op}^2\, \hbar F_{n,n'}}{2\rho L_{NW}\omega_{ph}} \begin{pmatrix} N_{op}(\hbar\omega_{ph}) \\ N_{op}(\hbar\omega_{ph}+1) \end{pmatrix} \tag{4.102}$$

where D_{op} is the optical phonon deformation potential, ω_{ph} is the phonon angular frequency and $N_{op}(\hbar\omega_{ph})$ is the phonon number given by the Bose-Einstein statistics:

$$N_{op}(\hbar\omega_{ph}) = \frac{1}{e^{\frac{\hbar\omega_{ph}}{K_B T}} - 1}. \tag{4.103}$$

The upper and lower choices in Eq.4.102 correspond to absorption and emission phonon processes, respectively. Optical phonon scattering can also assist transitions between subbands belonging to different valleys (inter-valley scattering) where the allowed final valleys depend on the selection rules which are thoroughly discussed in [107]. In this case, the squared matrix element can be demonstrated to be the same as in Eq.4.102 [107], but in the scattering rate calculations it must be properly accounted for the multiplicity of the final valleys for a given phonon type that can couple electrons belonging to different valleys [107]. The values for the D_{ac} and D_{op} for silicon have been taken from [107].

4.6.2 Coulomb scattering

In this section we discuss calculations for the Coulomb scattering matrix element, and as a first step we derive the expression for the potential in a cylindric nanowire produced by a point charge.

Perturbation potential produced by a point charge

In many papers the perturbation potential Φ_p produced by a point charge has been calculated assuming an homogeneous dielectric constant for oxide and semiconductor ($\varepsilon_{ox} = \varepsilon_{sct}$). For a nanowire with few nanometer radius, however, image charge effects produced by difference in ε_{ox} and ε_{sct} can be important. Hence, in the following we will assume different dielectric constants for oxide and semiconductor.

Let us write the Poisson's equation for the potential $G(\mathbf{R}, \mathbf{R}_0)$ produced by a point charge located in $\mathbf{R}_0$:

$$\nabla^2 G(\mathbf{R}, \mathbf{R}_0) = \gamma \delta(\mathbf{R} - \mathbf{R}_0) \tag{4.104}$$

where:

$$\gamma = \begin{cases} \dfrac{\varepsilon_{sct}}{\varepsilon_{ox}} & r_0 \geq r_{NW} \\ 1 & r_0 \leq r_{NW} \end{cases}$$

The $G(\mathbf{R}, \mathbf{R}_0)$ denotes the adimensional Green's function of the point charge and the electric potential produced by a charge distribution $\rho(\mathbf{R}_0)$ is then given by:

$$\Phi_p(\mathbf{R}) = e \int_{\mathbf{R}_0} G(\mathbf{R}, \mathbf{R}_0) \frac{\rho(\mathbf{R}_0)}{\varepsilon_{sct}} d\mathbf{R}_0. \tag{4.105}$$

It should be noted that in Eq.4.105 the dielectric constant at the denominator ε_{sct} is independent of $\mathbf{R}_0$ because of the definition of γ in Eq.4.105. We can rewrite the equation for the Green's function in cylindrical coordinates ($\mathbf{R} = (r, \theta, x)$) obtaining:

$$\begin{aligned} \nabla^2 G(\mathbf{R}, \mathbf{R}_0) &= \nabla^2 G(^{r,\theta,x}_{r_0,\theta_0,x_0}) \\ &= \frac{1}{r} \delta(r - r_0) \delta(\theta - \theta_0) \delta(x - x_0) \end{aligned} \tag{4.106}$$

where the Laplacian of the Green's function is given by:

$$\left[\frac{1}{r} \frac{\partial}{\partial r} + \frac{\partial^2}{\partial r^2} + \frac{1}{r^2} \frac{\partial^2}{\partial \theta^2} + \frac{\partial^2}{\partial x^2} \right] G(^{r,\theta,x}_{r_0,\theta_0,x_0}). \tag{4.107}$$

We now recall that, in the modeling of the nanowire we have assumed that all the quantities are periodic along the transport direction x in the interval $[-L_{NW}/2, L_{NW}/2]$ (as explained in Appendix D and B), hence $x, x_0 \in [-L_{NW}/2, L_{NW}/2]$. Moreover, θ in Eq.4.106 is an angle, so that we have periodicity also with respect to θ and, in particular, θ, θ_0 can be taken as belonging to the $[-\pi, \pi]$ interval.

Hence, in Eq.4.106 it is implicitly assumed that the Dirac functions $\delta(\theta-\theta_0)$ and $\delta(x-x_0)$ are periodic. We can thus exploit the Fourier series and write that:

$$\delta(\theta-\theta_0)=\sum_l c_l e^{il(\theta-\theta_0)} \tag{4.108}$$

where $l=0,\pm1,\pm2,\ldots$ and:

$$c_l=\frac{1}{2\pi}\int_{-\pi}^{+\pi}\delta(\theta-\theta_0)e^{-il(\theta-\theta_0)}d\theta=\frac{1}{2\pi}. \tag{4.109}$$

We thus obtain:

$$\delta(\theta-\theta_0)=\frac{1}{2\pi}\sum_l e^{il(\theta-\theta_0)} \tag{4.110}$$

Similar considerations apply to $\delta(x-x_0)$ in Eq.4.106. In fact:

$$\delta(x-x_0)=\sum_q c_q e^{iq(x-x_0)} \tag{4.111}$$

where $q=n2\pi/L_{NW}$ with $n=0,\pm1,\pm2,\ldots$ and:

$$c_q=\frac{1}{L_{NW}}\int_{-L_{NW}/2}^{+L_{NW}/2}\delta(x-x_0)e^{-iq(x-x_0)}dx=\frac{1}{L_{NW}} \tag{4.112}$$

which lead to:

$$\delta(x-x_0)=\frac{1}{L_{NW}}\sum_q e^{il(x-x_0)} \tag{4.113}$$

By assuming that $G(\mathbf{R},\mathbf{R}_0)$ is periodic over $\theta=[-\pi,+\pi]$ and along the transport direction over $x=[-L_{NW}/2,+L_{NW}/2]$, we can make the *ansatz*:

$$G(^{r,\theta,x}_{r_0,\theta_0,x_0})=\frac{1}{2\pi\,L_{NW}}\sum_{l,q}G_{lq}(r,r_0)\,e^{il(\theta-\theta_0)}e^{iq(x-x_0)}. \tag{4.114}$$

where $G_{lq}(r,r_0)$ is the reduced Green's function. By inserting Eq.4.114 into Eq.4.106 we obtain the equation for the reduced Green's function:

$$\left(\frac{\partial^2}{\partial r^2}+\frac{1}{r}\frac{\partial}{\partial r}-q^2-\frac{l^2}{r^2}\right)G_{lq}(r,r_0)=\frac{\gamma}{r}\delta(r-r_0) \tag{4.115}$$

Before solving Eq.4.115 it is necessary to impose boundary and continuity conditions:

$$\begin{cases} G_{lq}(r_{NW}+t_{ox},r_0)=0 & \text{metallic gate is assumed} \\ \varepsilon_{sct}\dfrac{\partial G_{lq}(r^-_{NW},r_0)}{\partial r}=\varepsilon_{ox}\dfrac{\partial G_{lq}(r^+_{NW},r_0)}{\partial r} & \text{continuity of the displacement vector} \\ \displaystyle\int_{r_0^-}^{r_0^+}\frac{\partial G_{lq}(r,r_0)}{\partial r}dr=\frac{\gamma}{r_0} & \text{Gauss law} \end{cases}. \tag{4.116}$$

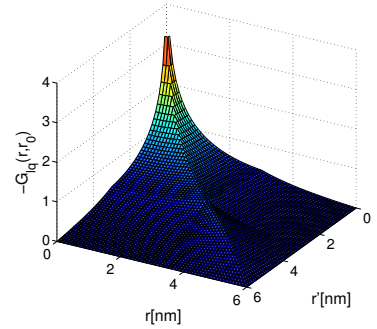
Analytical expression for the reduced Green's function $G_{lq}(r, r_0)$ can be worked out as [67] leading to:

$$\begin{array}{|c|}\hline r_0 < r_{NW} \\ \hline G_{lq}(r,r_0) = \begin{cases} a_1 I_l(qr) & 0 < r < r_0 \\ b_1 I_l(qr) + c_1 K_l(qr) & r_0 < r < r_{sct} \\ d_1 I_l(qr) + f_1 K_l(qr) & r_{sct} < r < r_{sct} + t_{ox} \end{cases} \\ \hline r_0 > r_{NW} \\ \hline G_{lq}(r,r_0) = \begin{cases} a_2 I_l(qr) & 0 < r < r_{sct} \\ b_2 I_l(qr) + c_2 K_l(qr) & r_{sct} < r < r_0 \\ d_2 I_l(qr) + f_2 K_l(qr) & r_0 < r < r_{sct} + t_{ox} \end{cases} \\ \hline \end{array} \tag{4.117}$$

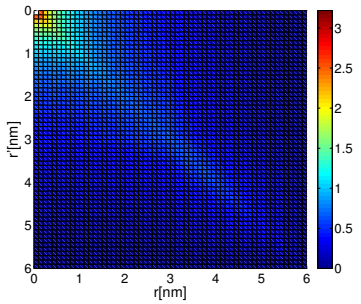
where I_l and K_l are the modified Bessel functions of the first and second kind, respectively, and coefficients a,b,c,d,$f_{1,2}$ can be found in [67].

Fig.4.24 shows some examples of reduced Green's functions. Moreover, as can be seen also from Fig.4.24, G_{lq} is symmetric (i.e. $G_{lq}(r, r_0)$=$G_{lq}(r, r_0)$) and this is expected because $\delta(r - r_0) = \delta(r_0 - r)$. The potential produced by the point charge $\Phi_p(^{r,\theta,x}_{r_0,\theta_0,x_0})$ is real-valued because $(G_{l,q})^\dagger = (G_{-l,q})$ and $(G_{l,q})^\dagger = (G_{l,-q})$. Finally, the electrostatic potential in (r, θ, x) produced by a point charge in (r_0, θ_0, x_0) is given by:

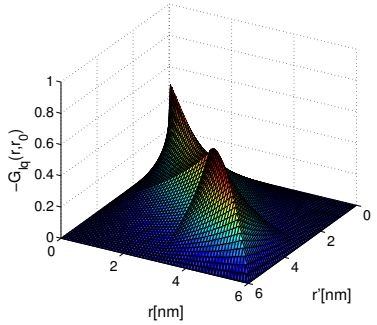
$$\Phi_p(^{r,\theta,x}_{r_0,\theta_0,x_0}) = \frac{e}{2\pi\, L_{NW} \varepsilon_{sct}} \sum_{lq} G_{lq}(r, r_0) e^{il(\theta-\theta_0)} e^{iq(x-x_0)} \tag{4.118}$$



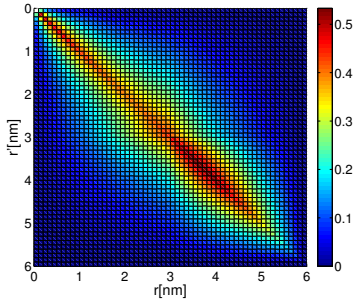
(a) Reduced Green's function obtained for $l = 0$ and $q_x = 0.5$ nm^{-1}



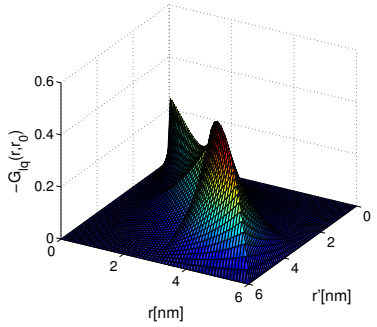
(b) Reduced Green's function obtained for $l = 0$ and $q_x = 0.5$ nm^{-1}



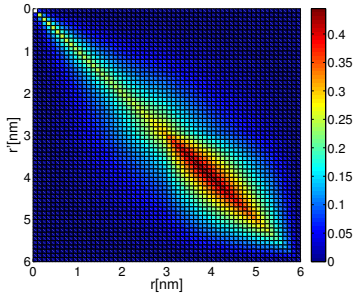
(c) Reduced Green's function obtained for $l = 1$ and $q_x = 0.5$ nm^{-1}



(d) Reduced Green's function obtained for $l = 1$ and $q_x = 0.5$ nm^{-1}



(e) Reduced Green's function obtained for $l = 2$ and $q_x = 0.5$ nm^{-1}



(f) Reduced Green's function obtained for $l = 2$ and $q_x = 0.5$ nm^{-1}

Figure 4.24: Reduced Green's function obtained obtained for the Si-SiO_2 system with $r_{NW} = 3nm$, $t_{ox} = 3nm$.

4.6.3 Matrix Element for Coulomb Scattering

The unscreened matrix element for Coulomb scattering is given by:

$$\mathcal{M}_{n,n'}(\mathbf{R}_0, q_x) = \int \Psi_{n,k_x}(\mathbf{R})\Phi_{scatt}(\mathbf{R},\mathbf{R}_0)\Psi^{\dagger}_{n',k'_x}(\mathbf{R})\, d\mathbf{R} \tag{4.119}$$

where $q_x = (k'_x - k_x)$ and where we drop the valley notation since Coulomb scattering is an intra-valley process. $\Psi_{n,k_x}(\mathbf{R}) = \Psi_{n,k_x}(r,\theta,x)$ and is given by:

$$\Psi_{n,k_x}(r,\theta,x) = \frac{e^{i\,k_x x}}{\sqrt{L_{NW}}}\xi_n(r,\theta). \tag{4.120}$$

$\Phi_{scatt}(\mathbf{R},\mathbf{R}_0)$ is the potential produced by a point charge defined in Eq.4.118.

By substituting the scattering potential of Eq.4.118 into the unscreened matrix element of Eq.4.119 we obtain:

$$\mathcal{M}_{n,n'}({}^{q_x}_{r_0,\theta_0,x_0}) = \frac{e^2}{2\pi\, L^2_{NW}\varepsilon_{sct}} \int_{-\pi}^{+\pi} d\theta \int_0^{\infty} r\, dr \int_{-L_{NW}/2}^{+L_{NW}/2} dx\, \xi^{\dagger}_{n'}(r,\theta)\xi_n(r,\theta)e^{i(k'_x-k_x)x}\times$$
$$\times \sum_{l_c,q_c} G_{l_c q_c}(r,r_0)e^{il_c(\theta-\theta_0)}e^{iq_c(x-x_0)}. \tag{4.121}$$

If we write $\xi_n(r,\theta)$ as in Eq.3.68 and notice that, according to Appendix E:

$$\frac{1}{L_{NW}} \int_{-L_{NW}/2}^{+L_{NW}/2} e^{i(k_x-k'_x+q_c)x}\, dx = \delta_{q_c,q_x} \tag{4.122}$$

$$\frac{1}{2\pi} \int_{-\pi}^{+\pi} e^{i(l-l'+l_c)\theta}\, d\theta = \delta_{l-l',l_c} \tag{4.123}$$

where $\delta_\bullet$ denotes the Kronecker symbol, the unscreened matrix element of Eq.4.121 reads:

$$\mathcal{M}_{n,n'}({}^{q_x}_{r_0,\theta_0,x_0}) = \frac{e^{-iq_x x_0}}{2\pi\, L_{NW}} \frac{e^2}{\varepsilon_{sct}} \sum_{l,l'} e^{-i(l-l')\theta_0} \int_0^{\infty} r\, dr \rho_{nl}(r)\rho^{\dagger}_{n'l'}(r)G_{l-l',q_x}(r,r_0). \tag{4.124}$$

To simplify the notation we introduce the symbol:

$$\mathcal{M}^0_{n,n'}({}^{q_x}_{r_0,\theta_0}) = \frac{e^2}{\varepsilon_{sct}} \sum_{l,l'} e^{-i(l-l')\theta_0} \int_0^{\infty} r\, dr \rho_{nl}(r)\rho^{\dagger}_{n'l'}(r)G_{l-l',q_x}(r,r_0) \tag{4.125}$$

The unscreened matrix element of Eq.4.124 considers only one point charge located in (r_0, θ_0, x_0), hence the total unscreened matrix element must be calculated summing over all (r_0, θ_0, x_0) points:

$$\mathcal{M}_{n,n'}(q_x) = \sum_{r_0,\theta_0,x_0} \frac{e^{-iq_x x_0}}{2\pi\, L_{NW}} \mathcal{M}^0_{n,n'}(^{q_x}_{r_0,\theta_0}). \tag{4.126}$$

The squared unscreened matrix element can be expressed as:

$$\left|\mathcal{M}_{n,n'}(q_x)\right|^2 = \frac{1}{\left(2\pi\, L_{NW}\right)^2} \sum_{r_0,\theta_0,x_0} \sum_{r'_0,\theta'_0,x'_0} e^{-iq_x(x_0-x'_0)} \mathcal{M}^0_{n,n'}(^{q_x}_{r_0,\theta_0}) \left(\mathcal{M}^0_{n,n'}(^{q_x}_{r'_0,\theta'_0})\right)^\dagger \tag{4.127}$$

By neglecting the terms due to Coulomb centers in different locations, consistently with a sort of random phase approximation [107], we finally write, the total squared unscreened matrix element as:

$$\left|\mathcal{M}_{n,n'}(q_x)\right|^2 = \frac{1}{\left(2\pi\, L_{NW}\right)^2} \sum_{r_0,\theta_0} \left|\mathcal{M}^0_{n,n'}(^{q_x}_{r_0,\theta_0})\right|^2. \tag{4.128}$$

Moving from the summation over (r_0, θ_0) to the corresponding integrals and introducing the symbols N_{II} to denote a volumetric charge (e.g. due to doping impurities) and N_{it} to denote a sheet charge (e.g. due to interface states) we write the total unscreened matrix element due to Coulomb centers as:

$$\left|\mathcal{M}^{unsc}_{n,n'}(q_x)\right|^2 = \frac{1}{2\pi\, L_{NW}} \left\{ \int_0^\infty dr_0 \int_{-\pi}^{\pi} \frac{d\theta_0}{2\pi} r_0 N_{II}(r_0,\theta_0) \left|\mathcal{M}^0_{n,n'}(^{q_x}_{r_0,\theta_0})\right|^2 + \right.$$
$$\left. + \int_{-\pi}^{\pi} \frac{d\theta_0}{2\pi} r_{NW} N_{it}(\theta_0) \left|\mathcal{M}^0_{n,n'}(^{q_x}_{r_{NW},\theta_0})\right|^2 \right\} \tag{4.129}$$

4.7 Screening

Before starting with the mathematical treatment of the screening, we describe the conceptual steps that will be pursued. The unscreened matrix element $\mathcal{M}^{uns}(q_x)$ (e.g. due to SR, Coulomb, alloy scattering) leads to an induced charge density ρ_{ind} (calculated by using perturbation theory within the so called random-phase approximation (RPA)). ρ_{ind} produces in turn a matrix element $\mathcal{M}^{ind}(q_x)$, via an induced scattering potential V_ρ, that adds to the unscreened matrix element. The aim is to obtain a system of linear equations linking the screened to the unscreened matrix elements via a matrix/tensor called dielectric function.

According to [107], the induced charge density produced by the matrix element $\mathcal{M}^{uns}_{n,n'}(q_x)$ where the valley index has been dropped and n, n' denote the transition between subband n and n', is given by:

$$\rho_{ind}(r,\theta,x) = e\sum_{n,n'} \xi^\dagger_n(r,\theta)\xi_{n'}(r,\theta)\Pi_{n,n'}(q_x)\mathcal{M}^{uns}_{n,n'}(q_x)\, e^{iq_x x} + (c.c.) \tag{4.130}$$

and $\Pi_{n,n'}(q_x)$ is the so called polarization factor defined as:

$$\Pi_{n,n'}(q_x) = \frac{1}{L_{NW}}\sum_{k_x} \frac{f_{n'}(k_x+q_x) - f_n(k_x)}{E_{T,n'}(k_x+q_x) - E_{T,n}(k_x)} \tag{4.131}$$

where f_n is the occupation function of subband n and $E_{T,n}(k_x)$ is the total energy given the value k_x for the subband n. By following the procedure as in [107], but using the expressions for the point charge for a circular nanowire reported in Eq.4.118, we give the expression for the potential induced by the point charge:

$$\begin{aligned} V_{\rho,q_x}(r,\theta) = e\sum_{n,n'} &\Pi_{n,n'}(q_x)\mathcal{M}^{uns}_{n,n'}(q_x) \\ &\int_{-\pi}^{\pi} d\theta_0 \int_0^{\infty} r_0\, dr_0 \xi^\dagger_{n'}(r'_0,\theta'_0)\xi_n(r,\theta)\Phi_{p,q_x}\left(\begin{smallmatrix} r,\theta \\ r_0,\theta_0 \end{smallmatrix}\right). \end{aligned} \tag{4.132}$$

After some calculations, the induced matrix element is thus written according to [107]:

$$\mathcal{M}^{ind}_{w,m,m'}(q_x) = \frac{e^2}{\varepsilon_{sct}}\sum_{\nu,n,n'} \Pi_{\nu,n,n'}(q_x) F^{\nu,n,n'}_{w,m,m'}(q_x) M^{uns}_{\nu,n,n'}(q_x) \tag{4.133}$$

where w and ν are valley indexes and where we have introduced the form factor $F^{\nu,n,n'}_{w,m,m'}$ defined as:

$$\begin{aligned} F^{\nu,n,n'}_{w,m,m'}(q_x) = &\int_{-\pi}^{+\pi} d\theta \int_0^{+\infty} r\, dr\, \xi^\dagger_{w,m'}(r,\theta)\xi_{w,m}(r,\theta)\times \\ &\times \int_{-\pi}^{+\pi} d\theta_0 \int_0^{+\infty} r_0\, dr_0\, \xi^\dagger_{\nu,n'}(r_0,\theta_0)\xi_{\nu,n}(r_0,\theta_0)\Phi_{pN,q_x}\left(\begin{smallmatrix} r,\theta \\ r_0,\theta_0 \end{smallmatrix}\right) \end{aligned} \tag{4.134}$$

and where $\Phi_{pN,q_x}(^{r,\theta}_{r_0,\theta_0})$ is the normalized, adimensional potential given by:

$$\Phi_{pN,q_x}(^{r,\theta}_{r_0,\theta_0}) = \sum_{l_c} G_{l_c,q}(r,r_0)\frac{e^{il_c(\theta-\theta_0)}}{2\pi} \tag{4.135}$$

where the expression for the reduced Green's function $G_{l_c,q}(r,r_0)$ is given in Eq.4.117. Eq.4.133 shows that the induced matrix element is given by a linear combination of unscreened matrix elements. As mentioned at the beginning of this section, the induced potential V_ρ induces the matrix element $\mathcal{M}^{ind}(q_x)$ which in turn changes the scattering potential and thus the matrix element. In order to account for this effect self-consistently, we need to calculate the $\mathcal{M}^{ind}(q_x)$ produced by the overall screened perturbation potential. This can be accomplished by using Eq.4.133 where in the r.h.s. $\mathcal{M}^{uns}(q_x)$ is substituted by $\mathcal{M}^{scr}(q_x)$. Since our aim is to link the screened to the unscreened matrix element, we can eliminate in the l.h.s. $\mathcal{M}^{ind}(q_x)$ by recalling that, by definition:

$$\mathcal{M}^{ind}(q_x) = \mathcal{M}^{scr}(q_x) - \mathcal{M}^{uns}(q_x). \tag{4.136}$$

By inserting Eq.4.136 in Eq.4.133 with the prescription that $\mathcal{M}^{uns}(q_x)$ is substituted by $\mathcal{M}^{scr}(q_x)$, we obtain the linear system linking the unscreend to the screened matrix elements:

$$M^{scr}_{w,m,m'}(q_x) - M^{uns}_{w,m,m'}(q_x) = \frac{e^2}{\varepsilon_{sct}} \sum_{\nu,n,n'} \Pi_{\nu,n,n'}(q_x) F^{\nu,n,n'}_{w,m,m'}(q_x) M^{scr}_{\nu,n,n'}(q_x). \tag{4.137}$$

Finally, Eq.4.137 can be cast in a more compact form and reads:

$$M^{uns}_{w,m,m'}(q_x) = \sum_{\nu,n,n'} \epsilon^{\nu,n,n'}_{w,m,m'}(q_x) M^{scr}_{\nu,n,n'}(q_x) \tag{4.138}$$

where the dielectric matrix is defined as:

$$\epsilon^{\nu,n,n'}_{w,m,m'}(q_x) = \delta_{w,\nu}\delta_{n,m}\delta_{n',m'} - \frac{e^2}{\varepsilon_{sct}}\, \Pi_{\nu,n,n'}(q_x) F^{\nu,n,n'}_{w,m,m'}(q_x). \tag{4.139}$$

Eq.4.138 can be solved numerically to calculate the screened matrix element starting from the unscreened ones. It is worth recalling that a given screened matrix element for the transition between subband n and n' belonging to the valley ν is linked with the unscreened matrix elements of all other scattering via the dielectric function defined in Eq.4.139.

4.7.1 Form factor calculation

In the form factor given by Eq.4.134, we can substitute the expression for the $\xi_n(r,\theta)$ obtained for the case of anisotropic quantization masses and given by Eq.3.68

and thus obtain:

$$F^{\nu,n,n'}_{w,m,m'}(q_x) = \frac{1}{2\pi}\sum_{g,g',l_c}\int_{-\pi}^{+\pi}\frac{e^{i(g-g'+l_c)\theta}}{2\pi}d\theta\int_0^\infty r\,dr\,\rho_{w,m,g}(r)\rho^\dagger_{w,m',g'}(r)\times$$
$$\times\left[\sum_{l,l'}\int_{-\pi}^{+\pi}\frac{e^{i(-l+l'-l_c)\theta_0}}{2\pi}d\theta_0\int_0^\infty r_0\,dr_0\,\rho_{\nu,n,l}(r_0)\rho^\dagger_{\nu,n',l'}(r_0)G_{l_c,q_x}(r,r_0)\right] \tag{4.140}$$

where g and g' are the expansion modes of the eigenvalues m and m', respectively (the same holds for l and l' that are the expansion modes of the eigenvalues n and n', respectively). By noting that (see Appendix E for more details):

$$\int_{-\pi}^{+\pi}\frac{e^{i(g-g'+l_c)\theta}}{2\pi}d\theta = \delta_{g-g',l_c} \tag{4.141}$$

$$\int_{-\pi}^{+\pi}\frac{e^{i(-l+l'-l_c)\theta_0}}{2\pi}d\theta_0 = \delta_{l-l',l_c} \tag{4.142}$$

Eq.4.140 can be written in a more compact form as:

$$F^{\nu,n,n'}_{w,m,m'}(q_x) = \frac{1}{2\pi}\sum_{g,g'}\int_0^\infty r\,dr\,\rho_{w,m,g}(r)\rho^\dagger_{w,m',g'}(r)\times$$
$$\times\left[\sum_{l,l'}\delta_{l'-l,g'-g}\int_0^\infty r_0\,dr_0\,\rho_{\nu,n,l}(r_0)\rho^\dagger_{\nu,n',l'}(r_0)G_{l'-l,q_x}(r,r_0)\right] \tag{4.143}$$

4.7.2 Polarization factor calculation

In general terms, calculation of the polarization factor given by Eq.4.131, requires the knowledge of the occupation function in the different subbands. At the equilibrium (i.e. V_{DS}=0), the occupation function $f_n(k_x)$ reduces to the Fermi Dirac occupation function $f_0(k_x)$ given by Eq.C.22. In the non-equilibrium condition, we followed the approach most often employed in the literature to calculate the polarization factor by using the equilibrium Fermi Dirac function $f_0(E_{T,n}(k_x))$ with a local quasi-Fermi level E_F. The local quasi-Fermi level for a given section s is calculated as the Fermi level energy E_F, that would give the charge at section s according to the band structure and the density of states at section s if the system were in equilibrium. The quasi-Fermi level can be obtained by using the expression for the charge given by Eq.C.42 as an equation to be solved for E_F.

To calculate numerically the polarization factor it is convenient to convert the summation over k_x into an integral using the prescriptions in Appendix D. Hence

Eq.4.131 becomes:

$$\Pi_{\nu,n,n'}(q_x) = \frac{\mu_\nu}{\pi} \int_{-\infty}^{+\infty} \frac{f_0(E_{T,\nu,n'}(k_x+q_x)) - f_0(E_{T,\nu,n}(k_x))}{E_{T,\nu,n'}(k_x+q_x) - E_{T,\nu,n}(k_x)}\, dk_x \tag{4.144}$$

where μ_ν is the valley multiplicity factor and we have set $n_{sp} = 2$ in Eq.D.8.

Analytical expression for the polarization factor at T=0 K: guidelines for numerical calculations

It may be instructive to analyze an analytical expression for the polarization factor of Eq.4.144 for a 1D electron gas. We thus assume a parabolic energy dispersion relation $E(k_x)$, zero temperature T=0K (i.e. the Fermi function is a unit step between $[-k_{x,F}, k_{x,F}]$ (where $k_{x,F}$ is the Fermi k_x vector) and we consider intrasubband polarization $n = n'$.

After some mathematical manipulations Eq.4.144 can be rewritten as:

$$\Pi_{\nu,n,n}(q_x) = -\frac{2\mu_\nu m_{\nu,x}}{\pi q_x \hbar^2} ln \left| \frac{1 + 2k_{x,F}/q_x}{1 - 2k_{x,F}/q_x} \right|. \tag{4.145}$$

By looking at Eq.4.145, we noticed that it is an even function with respect to q_x and, moreover, that $\Pi_{\nu,n,n}(q_x)$ has a singularity in $|q_x| = 2k_{x,F}$. In addition, for very small q_x values $\Pi_{\nu,n,n}(0)$ tends to $-2\mu_\nu m_{\nu,x}/(\hbar^2 \pi k_{x,F})$ and finally, for very large $|q_x|$ values $\Pi_{\nu,n,n}(\pm\infty)$ has a dependence of the type $1/q_x$. Fig.4.25 shows the good matching between analytic and numeric calculations at $0°K$ obtained setting k_F to 0.826 nm^{-1}. Instead, Fig.4.26 shows three polarization factors calculated for an InAs NW at $300°K$. It can be seen that as expected, the singularity is smeared out at a finite temperature, but a peaked maximum can be observed even at $q_x \approx 2k_F$.

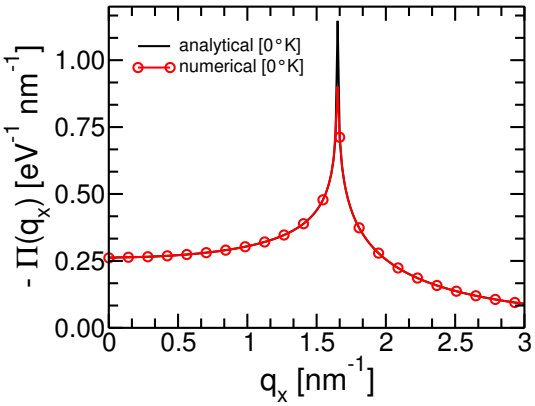


Figure 4.25: Polarization factor calculation using the analytical equation (Eq.4.145)(solid-line) and solving numerically Eq.4.145 (circles) at $0°K$.

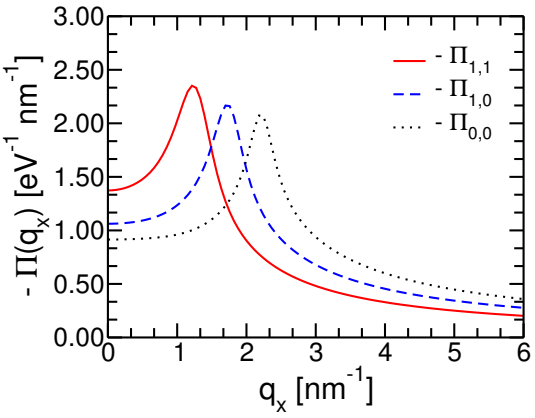


Figure 4.26: Polarization factors calculated for InAs and for different subband transitions at $300°K$. $\mathrm{NW}_{\mathrm{diameter}}$ is 5 nm.

4.7.3 Application of screening to CS matrix element

The unscreened matrix element for the scattering potential produced by a point charge located at (r_0, θ_0, x_0) is given by Eq.4.125. For any q_x value the corresponding screened matrix element is obtained as following:

$$\frac{e^{-i\,q_x\,x_0}}{2\pi\,L_{NW}}\mathcal{M}^{0}_{w,m,m'}({}^{q_x}_{r_0,\theta_0}) = \sum_{\nu,n,n'} \epsilon^{\nu,n,n'}_{w,m,m'}(q_x)\mathcal{M}^{scr}_{\nu,n,n'}({}^{q_x}_{r_0,\theta_0,x_0}) \tag{4.146}$$

where $\epsilon^{\nu,n,n'}_{w,m,m'}(q_x)$ is the dielectric function given by Eq.4.139. Since q_x is a constant parameter that couples the matrix elements corresponding to a given q_x and a given position (r_0, θ_0, x_0) of the point charge, the screened matrix element for any q_x value can be expressed as:

$$\mathcal{M}^{scr}_{\nu,n,n'}({}^{q_x}_{r_0,\theta_0,x_0}) = \frac{e^{-i\,q_x\,x_0}}{2\pi\,L_{NW}}\mathcal{M}^{0,scr}_{\nu,n,n'}({}^{q_x}_{r_0,\theta_0}) \tag{4.147}$$

where $\mathcal{M}^{0,scr}_{\nu,n,n'}({}^{q_x}_{r_0,\theta_0})$ can be obtained by solving the linear problem:

$$\mathcal{M}^{0}_{w,m,m'}({}^{q_x}_{r_0,\theta_0}) = \sum_{\nu,n,n'} \epsilon^{\nu,n,n'}_{w,m,m'}(q_x)\mathcal{M}^{0,scr}_{\nu,n,n'}({}^{q_x}_{r_0,\theta_0}). \tag{4.148}$$

Finally, the squared matrix element is obtained by replacing the unscreened matrix elements $\mathcal{M}^{0}_{w,m,m'}({}^{q_x}_{r_0,\theta_0})$ in Eq.4.129 with the screened matrix element in Eq.4.148 obtaining:

$$\left|\mathcal{M}^{unsc}_{w,m,m'}(q_x)\right|^2 = \frac{1}{2\pi\,L_{NW}}\left\{\int_0^{\infty} dr_0 \int_{-\pi}^{\pi}\frac{d\theta_0}{2\pi} r_0 N_{II}(r_0,\theta_0)\left|\mathcal{M}^{0,scr}_{w,m,m'}({}^{q_x}_{r_0,\theta_0})\right|^2 + \right.$$
$$\left. + \int_{-\pi}^{\pi}\frac{d\theta_0}{2\pi} r_{NW} N_{it}(\theta_0)\left|\mathcal{M}^{0,scr}_{w,m,m'}({}^{q_x}_{r_{NW},\theta_0})\right|^2\right\} \tag{4.149}$$

Fig.4.27 shows the effect of screening for intrasubband transitions in silicon with [100] transport direction for the lowest subbands of the corresponding valleys. As expected, screening is more effective for the intrasubband transition for the valleys with anisotropic quantization masses (circles) since the lowest subband for these valleys is the most populated by electrons and the polarization factor that enters the dielectric function in Eq.4.148 depends on the occupation function.

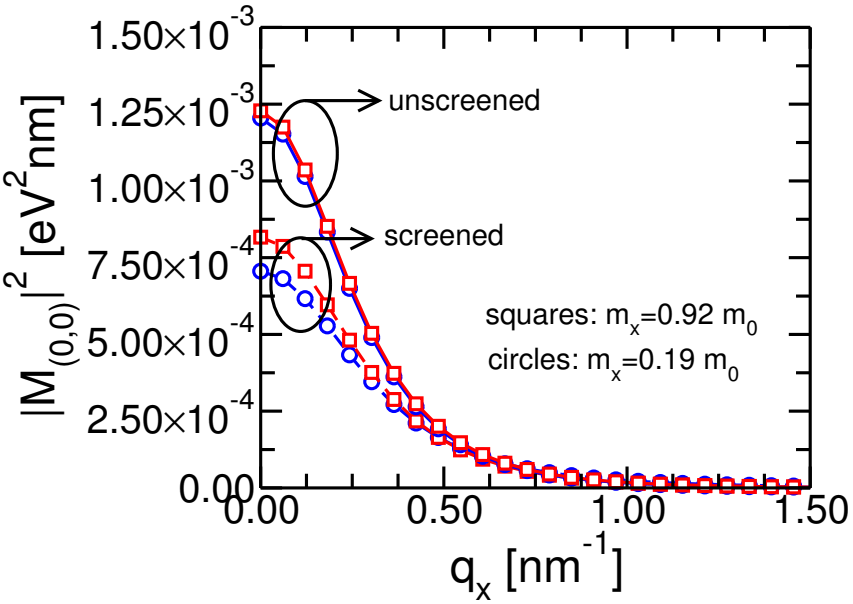


Figure 4.27: Plot of the screened and unscreened matrix element for silicon with [100] transport direction and for intrasubband transitions for the lowest eigenvalues of each valley. Squares: matrix elements for the valleys with isotropic quantization masses with the main ellipsoidal axes aligned with the transport direction (transport mass equal to 0.92 m_0). Circles matrix elements for the valleys with anisotropic quantization masses with transport mass equal to 0.19 m_0. Results obtained with a density of interface states $N_{it} = 5 \times 10^{11}$ cm^{-2} and for an inversion density of 3×10^6 cm^{-1}

4.7.4 Application of screening to SR matrix element

The screening for a linear formulation of the surface roughness scattering can be included similarly to the case of Coulomb scattering and it has been discussed in several previous contributions [107]. To this purpose we recall that, by definition itself of the dielectric matrix $\epsilon^{\nu,n,n'}_{w,m,m'}(q_x)$, a linear relation exists between a screened matrix element $\mathcal{M}^{scr}_{\nu,n,n'}(q_x)$ and the unscreened matrix element according to Eq.4.138. If we denote by $L^{\nu,n,n'}_{w,m,m'}(q_x)$ the inverse of the dielectric function $\varepsilon(q_x)$ defined in Eq.4.139, then the screened matrix element can be written as a linear combination of the unscreened matrix elements:

$$\mathcal{M}^{scr}_{\nu,n,n'}(q_x) = \sum_{w,m,m'} L^{\nu,n,n'}_{w,m,m'}(q_x)\mathcal{M}^{uns}_{w,m,m'}(q_x). \tag{4.150}$$

On the basis of Eq.4.150, the ensemble averaged screened matrix element is given by definition as:

$$\begin{aligned}\left\langle \left|\mathcal{M}^{scr}_{\nu,n,n'}(q_x)\right|^2 \right\rangle = &\sum_{w,m,m'} \left|L^{\nu,n,n'}_{w,m,m'}(q_x)\right|^2 \left\langle \left|\mathcal{M}^{uns}_{w,m,m'}(q_x)\right|^2 \right\rangle + \\ &+ \sum_{(w,m,m')\neq(u,p,p')} L^{\nu,n,n'}_{w,m,m'}(q_x) L^{\nu,n,n'}_{u,p,p'}(q_x)^{\dagger} \left\langle \mathcal{M}^{uns}_{w,m,m'}(q_x) \left(\mathcal{M}^{uns}_{u,p,p'}(q_x)\right)^{\dagger} \right\rangle\end{aligned} \tag{4.151}$$

where $\left\langle \mathcal{M}^{uns}_{w,m,m'}(q_x)\mathcal{M}^{uns}_{u,p,p'}(q_x) \right\rangle$ is the Fourier transform of the cross-correlation function between the matrix elements $\mathcal{M}^{uns}_{w,m,m'}(q_x)$ and $\mathcal{M}^{uns}_{u,p,p'}(q_x)$.

Fig.4.28 shows the effect of the screening on the SR intrasubband matrix element at large inversion density for silicon with [100] transport direction.

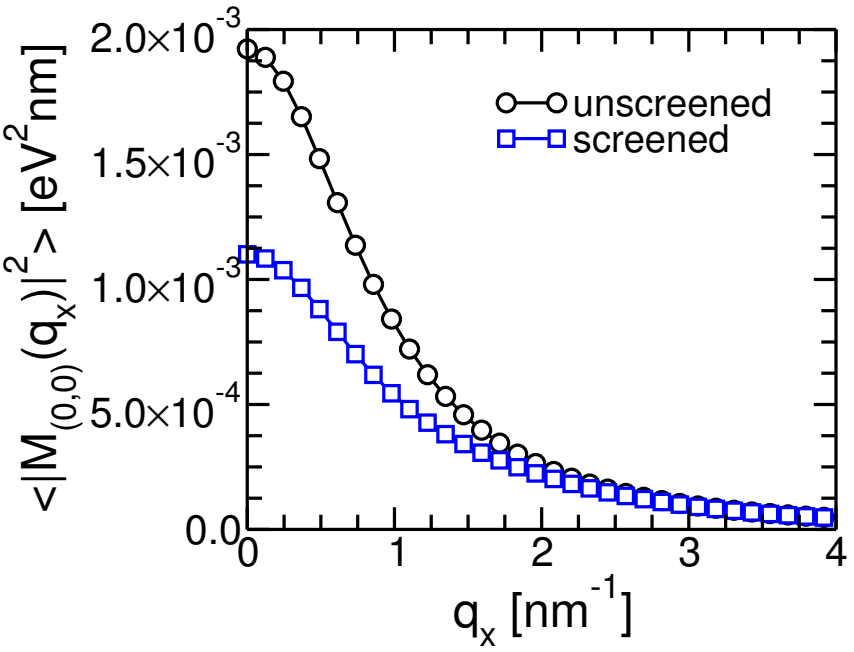


Figure 4.28: Plot of the screenend and unscreened intrasubband matrix element for silicon with [100] transport direction and for the lowest subband. Results obtained with $d = 5$ nm, and for an inversion density of 1×10^7 cm^{-1}.

References

[1] M. Je, S. Han, I. Kim, and H. Shin, "A silicon quantum wire transistor with one-dimensional subband effects", *Solid State Electronics*, vol. 44, no. 12, pp. 2207 –2212, 2000.

[2] J.-T. Park, J. Colinge, and C. Diaz, "Pi-Gate SOI MOSFET", *IEEE Electron Device Lett.*, vol. 22, no. 8, pp. 405–406, Aug. 2001.

[3] Y.-K. Choi, L. Chang, P. Ranade, J.-S. Lee, D. Ha, S. Balasubramanian, A. Agarwal, M. Ameen, T.-J. King, and J. Bokor, "FinFET process refinements for improved mobility and gate work function engineering", in *IEEE IEDM Technical Digest*, Dec. 2002, pp. 259–262.

[4] B. Doyle, S. Datta, M. Doczy, S. Hareland, B. Jin, J. Kavalieros, T. Linton, A. Murthy, R. Rios, and R. Chau, "High performance fully-depleted Tri-gate CMOS transistors", *IEEE Electron Device Lett.*, vol. 24, no. 4, pp. 263–265, Apr. 2003.

[5] O. H. Elibol, D. Morisette, D. Akin, J. P. Denton, and R. Bashir, "Integrated nanoscale silicon sensors using top-down fabrication", *Applied Physics Letters*, vol. 83, no. 22, pp. 4613–4615, 2003.

[6] L. Yang et al., "5nm-gate nanowire FinFET", Jun. 2004, pp. 196–197.

[7] S. D. Suk et al., "High performance 5nm radius Twin Silicon Nanowire MOSFET (TSNWFET) : fabrication on bulk si wafer, characteristics, and reliability", in *IEEE IEDM Technical Digest*, Dec. 2005, pp. 717–720.

[8] N. Singh et al., "High-performance fully depleted silicon nanowire (diameter $\leq$ 5 nm) gate-all-around CMOS devices", *IEEE Electron Device Lett.*, vol. 27, no. 5, pp. 383–386, May 2006.

[9] K. H. Yeo et al., "Gate-All-Around (GAA) Twin Silicon Nanowire MOSFET (TSNWFET) with 15 nm Length Gate and 4 nm Radius Nanowires", in *IEEE IEDM Technical Digest*, Dec. 2006, pp. 1–4.

[10] H. Lee et al., "Sub-5nm All-Around Gate FinFET for Ultimate Scaling", 2006, pp. 58–59.

[11] Y. Tian, R. Huang, Y. Wang, J. Zhuge, R. Wang, J. Liu, X. Zhang, and Y. Wang, "New Self-Aligned Silicon Nanowire Transistors on Bulk Substrate Fabricated by Epi-Free Compatible CMOS Technology: Process Integration, Experimental Characterization of Carrier Transport and Low Frequency noise", in *IEEE IEDM Technical Digest*, Dec. 2007, pp. 895–898.

[12] J. Gu, Y. Liu, Y. Wu, R. Colby, R. Gordon, and P. Ye, "First experimental demonstration of gate-all-around III-V MOSFETs by top-down approach", in *IEEE IEDM Technical Digest*, Dec. 2011, pp. 33.2.1–33.2.4.

[13] H. Wang, M. Sun, K. Ding, M. T. Hill, and C.-Z. Ning, "A Top-down Approach to Fabrication of High Quality Vertical Heterostructure Nanowire Arrays", *Nano Letters*, vol. 11, no. 4, pp. 1646–1650, 2011.

[14] J. Gu, X. Wang, H. Wu, J. Shao, A. Neal, M. Manfra, R. Gordon, and P. Ye, "20-80nm Channel length InGaAs gate-all-around nanowire MOSFETs with EOT=1.2nm and lowest SS=63mV/dec", in *IEEE IEDM Technical Digest*, Dec. 2012, pp. 27.6.1–27.6.4.

[15] N. Waldron, C. Merckling, L. Teugels, P. Ong, S. Ibrahim, F. Sebaai, A. Pourghaderi, K. Barla, N. Collaert, and A.-Y. Thean, "InGaAs Gate-All-Around Nanowire Devices on 300mm Si Substrates", *IEEE Electron Device Lett.*, vol. 35, no. 11, pp. 1097–1099, Nov. 2014.

[16] J. Westwater, D. P. Gosain, S. Tomiya, S. Usui, and H. Ruda, "Growth of silicon nanowires via gold-silane vapor-liquid-solid reaction", *Journal of Vacuum Science Technology B*, vol. 15, no. 3, pp. 554–557, 1997.

[17] A. M. Morales and C. M. Lieber, "A Laser Ablation Method for the Synthesis of Crystalline Semiconductor Nanowires", *Science*, vol. 279, no. 5348, pp. 208–211, 1998. eprint: `http://www.sciencemag.org/content/279/5348/208.full.pdf`.

[18] J. D. Holmes, K. P. Johnston, R. C. Doty, and B. A. Korgel, "Control of Thickness and Orientation of Solution-Grown Silicon Nanowires", *Science*, vol. 287, no. 5457, pp. 1471–1473, 2000. eprint: `http://www.sciencemag.org/content/287/5457/1471.full.pdf`.

[19] Y. Cui and C. M. Lieber, "Functional Nanoscale Electronic Devices Assembled Using Silicon Nanowire Building Blocks", *Science*, vol. 291, no. 5505, pp. 851–853, 2001. eprint: `http://www.sciencemag.org/content/291/5505/851.full.pdf`.

[20] Y. Cui, Z. Zhong, D. Wang, W. U. Wang, and C. M. Lieber, "High Performance Silicon Nanowire Field Effect Transistors", *Nano Letters*, vol. 3, no. 2, pp. 149–152, 2003.

[21] X. Duan, C. Niu, V. Sahi, J. Chen, J. W. Parce, S. Empedocles, and J. L. Goldman, "High-performance thin-film transistors using semiconductor nanowires and nanoribbons", vol. 425, pp. 274–278, Sep. 2003.

[22] C. Thelander, M. Bjork, M. Larsson, A. Hansen, L. Wallenberg, and L. Samuelson, "Electron transport in InAs nanowires and heterostructure nanowire devices", *Solid State Communications*, vol. 131, no. 9-10, pp. 573–579, 2004, New advances on collective phenomena in one-dimensional systems.

[23] J. Goldberger, A. I. Hochbaum, R. Fan, and P. Yang, "Silicon Vertically Integrated Nanowire Field Effect Transistors", *Nano Letters*, vol. 6, no. 5, pp. 973–977, 2006.

[24] Y. Wang, V. Schmidt, S. Senz, and U. Gosele, "Epitaxial growth of silicon nanowires using an aluminium catalyst", *Nature Nanotechnology*, vol. 1, pp. 186–189, 2006.

[25] C. Thelander, L. FrobergFroberg, C. Rehnstedt, L. Samuelson, and L.-E. Wernersson, "Vertical Enhancement-Mode InAs Nanowire Field-Effect Transistor With 50-nm Wrap Gate", *IEEE Electron Device Lett.*, vol. 29, no. 3, pp. 206–208, Mar. 2008.

[26] J. Sleight et al., "Gate-all-around silicon nanowire MOSFETs and circuits", in *Device Research Conference*, Jun. 2010, pp. 269–272.

[27] K. Tomioka, M. Yoshimura, and T. Fukui, "Vertical $In_{0.7}Ga_{0.3}As$ nanowire surrounding-gate transistors with high-k gate dielectric on Si substrate", in *IEEE IEDM Technical Digest*, Dec. 2011, pp. 33.3.1–33.3.4.

[28] E. Koren, G. Elias, A. Boag, E. R. Hemesath, L. J. Lauhon, and Y. Rosenwaks, "Direct Measurement of Individual Deep Traps in Single Silicon Nanowires", *Nano Letters*, vol. 11, no. 6, pp. 2499–2502, 2011, PMID: 21591656. eprint: `http://dx.doi.org/10.1021/nl201019b`.

[29] K. Tomioka, M. Yoshimura, and T. Fukui, "A III-V nanowire channel on silicon for high-performance vertical transistors", *Nature*, vol. 488, pp. 189 –192, 2012.

[30] K. Dae Mann and J. Yoon-Ha, *Nanowire Field Effect Transistors: Principles and Applications*. Springer New York Heidelberg Dordrecht London, 2014.

[31] R. Calarco, M. Marso, T. Richter, A. I. Aykanat, R. Meijers, A. v.d. Hart, T. Stoica, and H. Luth, "Size-dependent Photoconductivity in MBE-Grown GaN-Nanowires", *Nano Letters*, vol. 5, no. 5, pp. 981–984, 2005, PMID: 15884906. eprint: `http://dx.doi.org/10.1021/nl0500306`.

[32] N. Pradhan, H. Xu, and X. Peng, "Colloidal CdSe Quantum Wires by Oriented Attachment", *Nano Letters*, vol. 6, no. 4, pp. 720–724, 2006, PMID: 16608271. eprint: `http://dx.doi.org/10.1021/nl052497m`.

[33] C. O'Regan, S. Biswas, S. Barth, M. A. Morris, N. Petkov, and J. D. Holmes, "Size-controlled growth of germanium nanowires from ternary eutectic alloy catalysts", *J. Mater. Chem. C*, vol. 2, pp. 4597–4605, 23 2014.

[34] T. H. Kim, S. Y. Lee, N. K. Cho, H. K. Seong, H. J. Choi, S. W. Jung, and S. K. Lee, "Dielectrophoretic alignment of gallium nitride nanowires (GaN NWs) for use in device applications", *Nanotechnology*, vol. 17, no. 14, p. 3394, 2006.

[35] M. C. Wang and B. D. Gates, "Directed assembly of nanowires", *Materials Today*, vol. 12, no. 5, pp. 34 –43, 2009.

[36] S. Raychaudhuri, S. A. Dayeh, D. Wang, and E. T. Yu, "Precise Semiconductor Nanowire Placement Through Dielectrophoresis", *Nano Letters*, vol. 9, no. 6, pp. 2260–2266, 2009, PMID: 19419157. eprint: `http://dx.doi.org/10.1021/nl900423g`.

[37] Y. Huang, X. Duan, Q. Wei, and C. M. Lieber, "Directed Assembly of One-Dimensional Nanostructures into Functional Networks", *Science*, vol. 291, no. 5504, pp. 630–633, 2001. eprint: `http://www.sciencemag.org/content/291/5504/630.full.pdf`.

[38] C. H. Lee, D. R. Kim, and X. Zheng, "Orientation-Controlled Alignment of Axially Modulated pn Silicon Nanowires", *Nano Letters*, vol. 10, no. 12, pp. 5116–5122, 2010, PMID: 21043492. eprint: `http://dx.doi.org/10.1021/nl103630c`.

[39] T. Bryllert, L.-E. Wernersson, L. Froberg, and L. Samuelson, "Vertical high-mobility wrap-gated InAs nanowire transistor", *IEEE Electron Device Lett.*, vol. 27, no. 5, pp. 323–325, May 2006.

[40] C. Rehnstedt, T. Martensson, C. Thelander, L. Samuelson, and L.-E. Wernersson, "Vertical InAs Nanowire Wrap Gate Transistors on Si Substrates", *Electron Devices, IEEE Transactions on*, vol. 55, no. 11, pp. 3037–3041, Nov. 2008.

[41] Y.-K. Choi, J. Zhu, J. Grunes, J. Bokor, and G. A. Somorjai, "Fabrication of Sub-10-nm Silicon Nanowire Arrays by Size Reduction Lithography", *The Journal of Physical Chemistry B*, vol. 107, no. 15, pp. 3340–3343, 2003.

[42] N. Waldron et al., "An InGaAs/InP quantum well finfet using the replacement fin process integrated in an RMG flow on 300mm Si substrates", Jun. 2014, pp. 1–2.

[43] M. Zervas, D. Sacchetto, G. D. Micheli, and Y. Leblebici, "Top-down fabrication of very-high density vertically stacked silicon nanowire arrays with low temperature budget", *Microelectronic Engineering*, vol. 88, no. 10, pp. 3127 –3132, 2011.

[44] R. G. Hobbs, N. Petkov, and J. D. Holmes, "Semiconductor Nanowire Fabrication by Bottom-Up and Top-Down Paradigms", *Chemistry of Materials*, vol. 24, no. 11, pp. 1975–1991, 2012.

[45] W. Lu, P. Xie, and C. M. Lieber, "Nanowire Transistor Performance Limits and Applications", *IEEE Trans. on Electron Devices*, vol. 55, no. 11, pp. 2859–2876, Nov. 2008.

[46] S. Ouk Kim, H. H. Solak, M. P. Stoykovich, N. J. Ferrier, J. J. de Pablo, and P. F. Nealey, "Epitaxial self-assembly of block copolymers on lithographically defined nanopatterned substrates", *Nature*, vol. 474, pp. 411 –414, Jul. 2003.

[47] Y. Zheng, C. Rivas, R. Lake, K. Alam, T. Boykin, and G. Klimeck, "Electronic properties of silicon nanowires", *IEEE Trans. on Electron Devices*, vol. 52, no. 6, pp. 1097–1103, Jun. 2005.

[48] 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. on Electron Devices*, vol. 52, no. 7, pp. 1589–1595, Jul. 2005.

[49] M. Luisier, A. Schenk, W. Fichtner, and G. Klimeck, "Atomistic simulation of nanowires in the sp3d5s* tight-binding formalism: From boundary conditions to strain calculations", vol. 74, p. 205 323, 20 Nov. 2006.

[50] K. Nehari, N. Cavassilas, J. Autran, M. Bescond, D. Munteanu, and M. Lannoo, "Influence of band structure on electron ballistic transport in silicon nanowire mosfets: an atomistic study", *Solid State Electronics*, vol. 50, no. 4, pp. 716 –721, 2006.

[51] K. Nehari, N. Cavassilas, F. Michelini, M. Bescond, J. L. Autran, and M. Lannoo, "Full-band study of current across silicon nanowire transistors", *Applied Physics Letters*, vol. 90, no. 13, 132112, 2007.

[52] N. Neophytou, A. Paul, M. Lundstrom, and G. Klimeck, "Simulations of nanowire transistors: atomistic vs. effective mass models", English, *Journal of Computational Electronics*, vol. 7, no. 3, pp. 363–366, 2008.

[53] M. Luisier, "Phonon-limited and effective low-field mobility in n- and p-type [100]-, [110]- and [111]-oriented si nanowire transistors", *Applied Physics Letters*, vol. 98, no. 3, 032111, 2011.

[54] S. Kim, M. Luisier, A. Paul, T. Boykin, and G. Klimeck, "Full Three-Dimensional Quantum Transport Simulation of Atomistic Interface Roughness in Silicon Nanowire FETs", *IEEE Trans. on Electron Devices*, vol. 58, no. 5, pp. 1371–1380, May 2011.

[55] T. Ohno, K. Shiraishi, and T. Ogawa, "Intrinsic origin of visible light emission from silicon quantum wires: Electronic structure and geometrically restricted exciton", vol. 69, pp. 2400–2403, 16 Oct. 1992.

[56] Y. Lee, K. Kakushima, K. Shiraishi, K. Natori, and H. Iwai, "Size-dependent properties of ballistic silicon nanowire field effect transistors", *Journal of Applied Physics*, vol. 107, no. 11, 113705, 2010.

[57] Y.-S. Kim, M. Marsman, G. Kresse, F. Tran, and P. Blaha, "Towards efficient band structure and effective mass calculations for III-V direct band-gap semiconductors", *Phys. Rev. B*, vol. 82, p. 205 212, 20 Nov. 2010.

[58] E. Gnani, S. Reggiani, A. Gnudi, P. Parruccini, R. Colle, M. Rudan, and G. Baccarani, "Band-Structure Effects in Ultrascaled Silicon Nanowires", *IEEE Trans. on Electron Devices*, vol. 54, no. 9, pp. 2243–2254, Sep. 2007.

[59] S. Jin, Y. J. Park, and H. S. Min, "A three-dimensional simulation of quantum transport in silicon nanowire transistor in the presence of electron-phonon interactions", *Journal of Applied Physics*, vol. 99, no. 12, 123719, 2006.

[60] A. Szabo and M. Luisier, "Under-the-Barrier Model: An Extension of the Top-of-the-Barrier Model to Efficiently and Accurately Simulate Ultrascaled Nanowire Transistors", *IEEE Trans. on Electron Devices*, vol. 60, no. 7, pp. 2353–2360, Jul. 2013.

[61] M. Bescond, N. Cavassilas, and M. Lannoo, "Effective-mass approach for n-type semiconductor nanowire MOSFETs arbitrarily oriented", *Nanotechnology*, vol. 18, no. 25, p. 255 201, 2007.

[62] M. Fischetti, "Monte Carlo simulation of transport in technologically significant semiconductors of the diamond and zinc-blende structures. I. Homogeneous transport", *IEEE Trans. on Electron Devices*, vol. 38, no. 3, pp. 634–649, Mar. 1991.

[63] I. Vurgaftman, J. R. Meyer, and L. R. Ram-Mohan, "Band parameters for III–V compound semiconductors and their alloys", *Journal of Applied Physics*, vol. 89, no. 11, pp. 5815–5875, 2001.

[64] C. Jungemann, A. Edmunds, and W.L. Engl, "Simulation of Linear and Nonlinear Electron Transport in Homogeneous Silicon Inversion Layers", *Solid State Electronics*, vol. 36, no. 11, pp. 1529–1540, 1993.

[65] E. B. Ramayya, D. Vasileska, S. Goodnick, and I. Knezevic, "Electron Mobility in Silicon Nanowires", *IEEE Trans. on Nanotechnology*, vol. 6, no. 1, pp. 113–117, Jan. 2007.

[66] M.V. Fischetti and S.E. Laux, "Monte Carlo study of electron transport in silicon inversion layers", *Phys. Rev. B*, vol. 48, pp. 2244–2274, 1993.

[67] S. Jin, M. Fischetti, and T.-W. Tang, "Modeling of electron mobility in gated silicon nanowires at room temperature: Surface roughness scattering, dielectric screening, and band nonparabolicity", *Journal of Applied Physics*, vol. 102, no. 8, 2007.

[68] S. Jin, T.-W. Tang, and M. Fischetti, "Simulation of Silicon Nanowire Transistors Using Boltzmann Transport Equation Under Relaxation Time Approximation", *IEEE Trans. on Electron Devices*, vol. 55, no. 3, pp. 727–736, Mar. 2008.

[69] F. M. Gomez-Campos, S. Rodriguez-Bolivar, J. A. Lopez-Villanueva, J. A. Jimenez-Tejada, and J. E. Carceller, "A solution of the effective-mass Schrodinger equation in general isotropic and nonparabolic bands for the study of two-dimensional carrier gases", *Journal of Applied Physics*, vol. 98, no. 3, 033717, 2005.

[70] A. Rahman, G. Klimeck, M. Lundstrom, T. B. Boykin, and N. Vagidov, "Atomistic Approach for Nanoscale Devices at the Scaling Limit and Beyond Valley Splitting in Si", *Japanese Journal of Applied Physics*, vol. 44, no. 4S, p. 2187, 2005.

[71] N. Pons, N. Cavassilas, L. Raymond, F. Michelini, M. Lannoo, and M. Bescond, "Three-dimensional kp real-space quantum transport simulations of p-type nanowire transistors: influence of ionized impurities", *Applied Physics Letters*, vol. 99, no. 8, 082113, 2011.

[72] M. Shin, "Full-quantum simulation of hole transport and band-to-band tunneling in nanowires using the kp method", *Journal of Applied Physics*, vol. 106, no. 5, 054505, 2009.

[73] M. Shin, S. Lee, and G. Klimeck, "Computational Study on the Performance of Si Nanowire pMOSFETs Based on the kp Method", *IEEE Trans. on Electron Devices*, vol. 57, no. 9, pp. 2274–2283, Sep. 2010.

[74] S. Jin, S.-M. Hong, W. Choi, K.-H. Lee, and Y. Park, "Coupled drift-diffusion (DD) and multi-subband Boltzmann transport equation (MSBTE) solver for 3D multi-gate transistors", in *Proc.SISPAD*, Sep. 2013, pp. 348–351.

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

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

[77] K. Natori, "Compact Modeling of Ballistic Nanowire MOSFETs", *IEEE Trans. on Electron Devices*, vol. 55, no. 11, pp. 2877–2885, Nov. 2008.

[78] D. Jimenez, J. J. Saenz, B. Iniquez, J. Sune, L. F. Marsal, and J. Pallares, "Unified compact model for the ballistic quantum wire and quantum well metal-oxide-semiconductor field-effect-transistor", *Journal of Applied Physics*, vol. 94, no. 2, pp. 1061–1068, 2003.

[79] 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 *IEEE IEDM Technical Digest*, 2005, pp. 617–620.

[80] M. Bescond, N. Cavassilas, K. Kalna, K. Nehari, L. Raymond, J. Autran, M. Lannoo, and A. Asenov, "Ballistic transport in Si, Ge, and GaAs nanowire MOSFETs", in *IEEE IEDM Technical Digest*, Dec. 2005, pp. 526–529.

[81] M. Luisier and G. Klimeck, "OMEN an Atomistic and Full-Band Quantum Transport Simulator for post-CMOS Nanodevices", in *IEEE Conference on Nanotechnology*, Aug. 2008, pp. 354–357.

[82] M. Lenzi, P. Palestri, E. Gnani, S. Reggiani, A. Gnudi, D. Esseni, L. Selmi, and G. Baccarani, "Investigation of the Transport Properties of Silicon Nanowires Using Deterministic and Monte Carlo Approaches to the Solution of the Boltzmann Transport Equation", *IEEE Trans. on Electron Devices*, vol. 55, no. 8, pp. 2086–2096, Aug. 2008.

[83] A. Martinez, A. Brown, A. Asenov, and N. Seoane, "A Comparison between a Fully-3D Real-Space Versus Coupled Mode-Space NEGF in the Study of Variability in Gate-All-Around Si Nanowire MOSFET", in *Proc.SISPAD*, 2009, pp. 1–4.

[84] F. Conzatti, M. Pala, D. Esseni, E. Bano, and L. Selmi, "Strain-Induced Performance Improvements in InAs Nanowire Tunnel FETs", *IEEE Trans. on Electron Devices*, vol. 59, no. 8, pp. 2085 –2092, 2012.

[85] 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, pp. 645–705, 3 Jul. 1983.

[86] M. V. Fischetti and S. E. Laux, "Monte carlo analysis of electron transport in small semiconductor devices including band-structure and space-charge effects", vol. 38, pp. 9721–9745, 14 Nov. 1988.

[87] 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 Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 8, no. 4, pp. 360–369, Apr. 1989.

[88] J. Bude and R. K. Smith, "Phase-space simplex Monte Carlo for semiconductor transport", *Semiconductor Science Technology*, vol. 9, no. 5S, p. 840, 1994.

[89] C. Jungemann, S. Keith, M. Bartels, and B. Meinerzhagen, "Efficient full-band Monte Carlo simulation of silicon devices", *IEICE Transactions on Electronics*, vol. E82-C, pp. 870–879, 1999.

[90] F. M. Bufler, A. Schenk, and W. Fichtner, "Efficient Monte Carlo device modeling", *IEEE Trans. on Electron Devices*, vol. 47, no. 10, pp. 1891–1897, Oct. 2000.

[91] B. Winstead and U. Ravaioli, "A quantum correction based on schrodinger equation applied to monte carlo device simulation", *IEEE Trans. on Electron Devices*, vol. 50, no. 2, pp. 440–446, Feb. 2003.

[92] M. A. Elmessary, D. Nagy, M. Aldegunde, J. Lindberg, W. Dettmer, D. Peric, A. J. Garcia-Loureiro, and K. Kalna, "Anisotropic schrodinger equation quantum corrections for 3D Monte Carlo simulations of nanoscale multigate transistors", in *International Workshop on Computational Electronics*, Sep. 2015, pp. 1–4.

[93] P. Osgnach, A. Revelant, D. Lizzit, P. Palestri, D. Esseni, and L. Selmi, "Toward computationally efficient Multi-Subband Monte Carlo Simulations of Nanoscale MOSFETs", in *Proc.SISPAD*, 2013, pp. 176–179.

[94] L. Donetti, C. Sampedro, F. Gamiz, A. Godoy, F. J. Garcia-Ruiz, E. Towiez, V. P. Georgiev, S. M. Amoroso, C. Riddet, and A. Asenov, "Multi-Subband Ensemble Monte Carlo simulation of Si nanowire MOSFETs", in *Proc.SISPAD*, Sep. 2015, pp. 353–356.

[95] S.-M. Hong, A.-T. Pham, and C. Jungemann, *Deterministic Solvers for the Boltzmann Transport Equation*, 1st. Springer, 2011.

[96] C. Jungemann, S. M. Hong, B. Meinerzhagen, and A. T. Pham, "Deterministic simulation of 3D and quasi-2D electron and hole systems in SiGe devices", Sep. 2012, pp. 318–321.

[97] A. T. Pham, C. Jungemann, and B. Meinerzhagen, "Deterministic multi-subband device simulations for strained double gate PMOSFETs including magnetotransport", in *IEEE IEDM Technical Digest*, Dec. 2008, pp. 1–4.

[98] T. Krishnamohan, D. Kim, T. V. Dinh, A. t. Pham, B. Meinerzhagen, C. Jungemann, and K. Saraswat, "Comparison of (001), (110) and (111) uniaxial- and biaxial- strained-Ge and strained-Si PMOS DGFETs for all channel orientations: Mobility enhancement, drive current, delay and off-state leakage", in *IEEE IEDM Technical Digest*, Dec. 2008, pp. 1–4.

[99] E. G. Marin, F. G. Ruiz, A. Godoy, I. M. Tienda-Luna, C. Martinez-Blanque, and F. Gamiz, "Theoretical interpretation of the electron mobility behavior in InAs nanowires", *Journal of Applied Physics*, vol. 116, no. 17, 2014.

[100] E. Baravelli, E. Gnani, R. Grassi, A. Gnudi, S. Reggiani, and G. Baccarani, "Optimization of n- and p-type TFETs Integrated on the Same InAs/$Al_xGa_{1-x}Sb$ Technology Platform", *IEEE Trans. on Electron Devices*, vol. 61, no. 1, pp. 178–185, Jan. 2014.

[101] S. D. Suk et al., "Investigation of nanowire size dependency on TSNWFET", in *IEEE IEDM Technical Digest*, Dec. 2007, pp. 891–894.

[102] D. J. BenDaniel and C. B. Duke, "Space-Charge Effects on Electron Tunneling", *Phys. Rev.*, vol. 152, pp. 683–692, 2 Dec. 1966.

[103] G. Bastard, "Theoretical investigations of superlattice band structure in the envelope-function approximation", *Phys. Rev. B*, vol. 25, pp. 7584–7597, 12 Jun. 1982.

[104] N. L. Trefethen, *Spectral Methods in MATLAB*. SIAM, 2000.

[105] P. J. Davis, *Interpolation and Approximation*. Dover, 1975.

[106] S. Monaghan, P. Hurley, K. Cherkaoui, M. Negara, and A. Schenk, "Determination of electron effective mass and electron affinity in HfO2 using MOS and MOSFET structures", *Solid State Electronics*, vol. 53, no. 4, pp. 438 –444, 2009.

[107] D. Esseni, P. Palestri, and L. Selmi, *Nanoscale MOS Transistors - Semi-Classical Transport and Applications*, 1st. Cambridge University Press., 2011.

Chapter 5

Simulator for cilindrical nanowires based on a deterministic solution of the BTE

5.1 Introduction

THE aim of this chapter is to describe a solver for cylindrical gate-all-around nanowire transistors based on the deterministic solution of the Boltzmann transport equation (BTE) and by including the main scattering mechanisms affecting transport in the transistor.

5.2 BTE theory

The study of the transport in far-from-equilibrium conditions requires the knowledge of the occupation function $f(\mathbf{R}, \mathbf{K}, t)$ which describes the probability that an electron finds at point $\mathbf{R}$ and $\mathbf{K}$ of the real- and phase-space respectively, at the time t. The occupation function gives a full description of the carrier state and it is used to extract macroscopic parameters such as carrier charge, velocity, current and mobility. The occupation function depends on the real-space position $\mathbf{R}$, phase-space position $\mathbf{K}$ and on time t. However, in NWs transport occurs only along one dimension, hence $f(\mathbf{R}, \mathbf{K}, t)$ becomes a 3-dimensional function $f(x, k_x, t)$ (if we consider more than one valley and one subband, the dimension of the problem increases from 3 to 5) and hereon all derivations will be performed under this hypothesis (i.e. of a 1D electron gas). In the BTE physical framework electrons are considered as point charges located at the position (x, k_x) in the phase space given by the centroid of the wave-packet in real and wave-vector space. Moreover, position x and momentum $p_x = \hbar k_x$ are assumed to be known at the same time t, thus violating the Heisenberg

uncertainty principle, and, scattering is assumed to occur over an infinitesimal time such to be considered instantaneous, so that the position in the real space is supposed to be the same after the scattering event. It is also assumed that carriers move in the real space according to their electron velocity and can change their momentum due to an electric field. These variations along the real- and phase-space define trajectories for carriers as depicted in Fig.5.1. If we suppose that carriers move along a fixed trajectory, the number of carriers at a given time t_0 in the neighborhood of A located in the real- and phase-space $(x_0, k_{x,0})$, is equal to the number of carriers in the neighborhood of B at the time $t_0 + dt$ and position $(x_0 + dx, k_{x,0} + dk_x)$, and it can be written that:

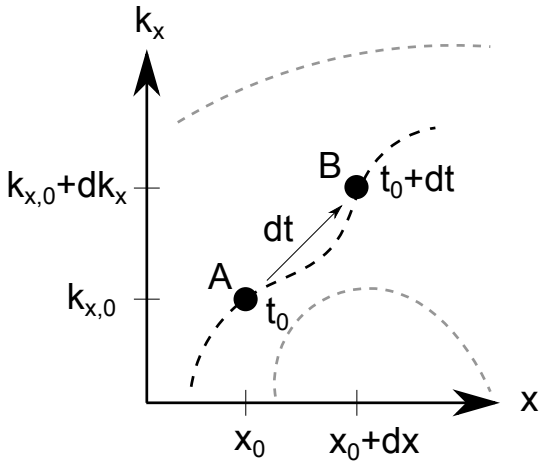


Figure 5.1: Ballistic carriers trajectories in the real- and phase-space.

$$f(x_0, k_{x,0}, t_0) = f(x_0 + dx, k_{x,0} + dk_x, t_0 + dt) \longrightarrow \frac{df(x, k_x, t)}{dt} = 0. \tag{5.1}$$

In other words the occupation function $f(x, k_x, t)$ associated with carriers belonging to a fixed trajectory does not change. By using the chain rule for the total derivative over the time, Eq.5.1 can be rewritten as:

$$\frac{df(x(t), k_x(t), t)}{dt} = \frac{\partial f}{\partial t} + \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial k_x}\frac{dk_x}{dt}. \tag{5.2}$$

According to the semi-classical approach to describe the motion of carriers, the term dx/dt is the group velocity $v_{g,\nu,n}(k_x)$ of a wave-packet (representing the carrier) inside a crystal, where ν is the valley and n is the subband index of the carrier, whereas the term dk_x/dt is equal to the force $F_{\nu,n}(x) = -eE_{F,\nu,n}(x)$ acting on a particle subjected to an electric field $E_{F,\nu,n}$ (we here neglect the contribution due to magnetic fields). These two equations representing the motion of the wave-packet

centroid in real and phase- space can be written in the general case as:

$$\begin{aligned}\frac{dx(t)}{dt} &= v_{g,\nu,n}(k_x) = \frac{1}{\hbar}\frac{\partial E_{T,\nu,n}(x,k_x)}{\partial k_x} \\ \frac{dk_x(t)}{dt} &= F_{\nu,n}(x) = -\frac{1}{\hbar}\frac{\partial E_{T,\nu,n}(x,k_x)}{\partial x}\end{aligned} \tag{5.3}$$

where $E_{T,\nu,n}(x,k_x)$ is the total energy of a carrier with wave-vector k_x and belonging to subband n, and valley ν. By substituting Eq.5.3 into Eq.5.2, and dropping the valley and subband indexes to simplify the notation, we obtain the BTE equation:

$$\frac{df(x,k_x,t)}{dt} = \frac{\partial f}{\partial t} + \frac{\partial f}{\partial x}\left(\frac{1}{\hbar}\frac{\partial E_T(k_x,x)}{\partial k_x}\right) + \frac{\partial f}{\partial k_x}\left(-\frac{1}{\hbar}\frac{\partial E_T(k_x,x)}{\partial x}\right). \tag{5.4}$$

Same result as in Eq.5.4 can be obtained by using another approach that exploits the continuity condition for carriers inside a control region in the real- and phase-space [1, 2]. So far scattering has not been mentioned in the derivation of Eq.5.4. We now assume that scattering mechanisms are local and instantaneous processes and change only the state of an electron in the k_x-space. Every scattering into k_x increases the occupancy $f(x,k_x,t)$ and every scattering out of k_x decreases the occupancy, hence the net change of occupancy due to scattering is given by the difference between the in- and out-scattering. For each state k_x one can add the contributions associated to all possible in-scattering events (from k'_x states) and the contributions associated to out-scattering events (to k'_x states) as shown in Fig.5.2 and written in Eq.5.5.

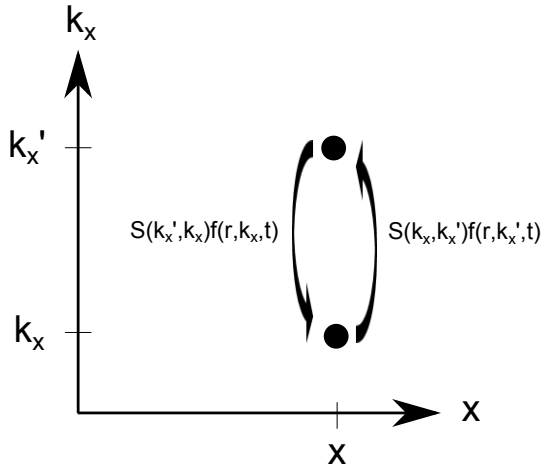


Figure 5.2: Sketch of scattering in the k-space.

$$\frac{\partial f(x,k_x,t)}{\partial t} = \sum_{k'_x}\left[S(k'_x,k_x)f(x,k'_x,t) - S(k_x,k'_x)f(x,k_x,t)\right] \tag{5.5}$$

where the term $S(k'_x, k_x)f(x, k'_x, t)$ denotes the transition rate from states k'_x to state k_x times the occupancy $f(x, k'_x, t)$ of the initial state k'_x. According to Pauli's exclusion principle, only one electron can fill a state at a given time (regardless of spin), hence an electron from state k'_x can scatter into state k_x only if the arrival state is empty. In other words, to account for exclusion principle (or for degenerate statistics) the transition rate for in-scattering is multiplied by the probability that the arrival state $f(x, k_x, t)$ is empty, whereas the transition rate for out-scattering is multiplied by the probability that the arrival state $f(x, k'_x, t)$ is empty. This results in the Boltzmann transport equation:

$$\begin{aligned}\frac{\partial f(x,k_x,t)}{\partial t} =& (1-f(x,k_x,t))\sum_{k'_x}\left[S(k'_x,k_x)f(x,k'_x,t)\right]- \\ &- f(x,k_x,t)\sum_{k'_x}\left[\left(1-f(x,k'_x,t)\right)S(k_x,k'_x)\right] = S^{in}-S^{out}. \end{aligned} \tag{5.6}$$

The presence of the terms $(1 - f(x, k_x, t))$ and $(1 - f(x, k'_x, t))$ makes the BTE a non-linear equation. Since k_x is almost continuous the summations in Eq.5.6 can be converted into integrals according to:

$$\sum_{k_x} = \frac{n_{sp}L_{NW}}{2\pi}\int_{k_x} dk_x \tag{5.7}$$

where n_{sp} is the spin degeneracy (see Appendix D for more details) and L_{NW} a normalization length. The normalization length cancels out when summing over the final states as, for example, in Eq.5.6. Hence the scattering integral in Eq.5.6 can be written as:

$$\begin{aligned}S^{in}-S^{out} =& (1-f(x,k_x,t))\frac{L_{NW}}{2\pi}\int_{k'_x}\left[S(k'_x,k_x)f(x,k'_x,t)dk'_x\right]- \\ &- f(x,k_x,t)\frac{L_{NW}}{2\pi}\int_{k'_x}\left[\left(1-f(x,k'_x,t)\right)S(k_x,k'_x)dk'_x\right]. \end{aligned} \tag{5.8}$$

Under the assumption that particles are conserved (generation/recombination mechanisms are not taken into account) the total derivative over time in the left-hand side of Eq.5.4 is equal to the scattering term of Eq.5.6. Hence, the final equation for the BTE is consistent with the fact that electrons can scatter in from other trajectories or out to other trajectories as shown in Fig.5.3 and reads:

$$\frac{\partial f_{k_x}}{\partial t} + \frac{\partial f_{k_x}}{\partial x}\left(\frac{1}{\hbar}\frac{dE_T(k_x,x)}{dk_x}\right) - \frac{\partial f_{k_x}}{\partial k_x}\left(\frac{1}{\hbar}\frac{dE_T(k_x,x)}{dx}\right) = S^{in}-S^{out}. \tag{5.9}$$

where $f_{k'_x} = f(x, k'_x, t)$.

Similarly, the BTE formulation for a 3D domain is given by:

$$\frac{\partial f_{\mathbf{K}}}{\partial t} + \frac{1}{\hbar}\nabla_{\mathbf{K}}E_T(\mathbf{K},\mathbf{R})\cdot\nabla_{\mathbf{R}}f_{\mathbf{K}} - \frac{1}{\hbar}\nabla_{\mathbf{R}}E_T(\mathbf{K},\mathbf{R})\cdot\nabla_{\mathbf{K}}f_{\mathbf{K}} = S^{in}-S^{out} \tag{5.10}$$

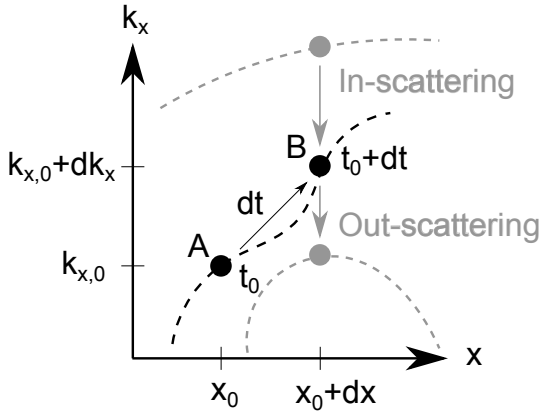


Figure 5.3: Real- and phase-space with carriers trajectories and scattering.

or equivalently by:

$$\frac{\partial f_{\mathbf{K}}}{\partial t} + \mathbf{V}_g(\mathbf{K}) \cdot \nabla_{\mathbf{R}} f_{\mathbf{K}} - \mathbf{F}(\mathbf{R}) \cdot \nabla_{\mathbf{K}} f_{\mathbf{K}} = S^{in} - S^{out}. \tag{5.11}$$

As can be seen, for a 3D or for a 2D gas the BTE equation is a very complicated equation to solve numerically. The typical approach is either to employ Monte-Carlo techniques [3–14] or to introduce simplifications for the scattering term based, for example, on the so called momentum relaxation time approximation (MRT or RTA) [15, 16].

5.3 BTE discretization for a 1D electron gas

For a 1D gas carriers are free to move only along one direction, that is both real space and $k-$ space are 1D. In this case the complexity of the BTE reduces considerably with respect to the cases of 3D or 2D electron gas. Moreover, we assume steady-state conditions, which means that the occupation function depends only on k_x and x coordinates, so that the BTE in Eq.5.9 for each subband reads:

$$v_{g,n}(x,k_x)\frac{\partial f_n(x,k_x)}{\partial x} - \frac{\partial f_n(x,k_x)}{\partial k_x}\left(\frac{1}{\hbar}\frac{\partial E_{T,n}}{\partial x}\right) = S^{in} - S^{out}. \tag{5.12}$$

where f_n is the occupation function and it is function of (x,k_x). $E_{T,n}$ is the total electron energy for a given subband n (and valley, not reported explicitly in Eq.5.12) and $v_{g,n}$ is the group velocity for a given electron located in the real- and phase-space in (x,k_x). To reduce the complexity of Eq.5.12 it is convenient to rewrite the occupation function f_n as a function of the position along transport direction x and the total energy (instead of the wave-vector k_x) [17].

5.3.1 Reformulation of the BTE for 1D GAS

It can be easily noticed (see the example in Fig.5.4) that the relation between k_x and the total energy $E_{T,n}$ is not one-to-one (neither for a 1D gas nor for 2D or 3D gas), therefore the k_x domain can be splitted into two parts such that $E_{T,n}(k_x)$ becomes an injective function of k_x. This is obtained by dividing the k_x values into positive and negative values.

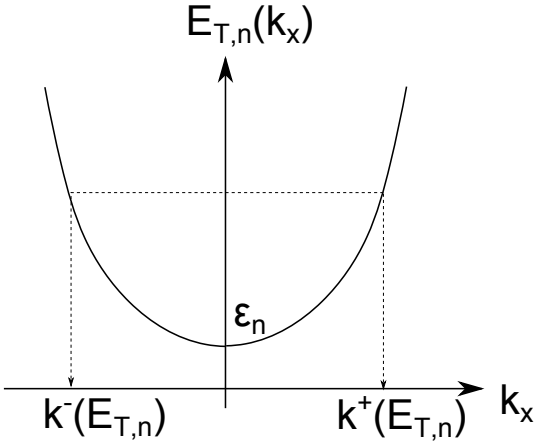


Figure 5.4: Dispersion relation for a 1D electron gas. ε_n corresponds to the lowest total energy for subband n and equals to the n-th eigenvalue.

The space- and energy-dependent occupation function $f_n(x, E_{T,n})$ can thus be written as:

$$f_n(x, E_{T,n}) = \begin{cases} f_n^+(x, E_{T,n}) = f_n(x, k_x), & \text{for } k_x > 0 \\ f_n^-(x, E_{T,n}) = f_n(x, k_x), & \text{for } k_x < 0 \end{cases}$$

where n is the subband index (for simplicity the valley index has been dropped) and $E_{T,n} = E_{T,n}(x, k_x)$ is the total energy. Let us focus on the $f_n^+(x, E_{T,n})$ occupation function and calculate the total derivative over x:

$$\frac{df_n^+(x, E_{T,n})}{dx} = \frac{\partial f_n(x, k_x^+)}{\partial x} + \frac{\partial f_n(x, k_x^+)}{\partial k_x^+}\frac{dk_x^+}{dx} \tag{5.13}$$

where $k_x^+ : k_x > 0$, and recalling that:

$$\frac{dE_{T,n}}{dx} = 0 = \frac{\partial E_{T,n}}{\partial x} + \frac{\partial E_{T,n}}{\partial k_x^+}\frac{dk_x^+}{dx} \tag{5.14}$$

it follows that:

$$\frac{dk_x^+}{dx} = -\frac{\partial E_{T,n}}{\partial x}\left(\frac{\partial E_{T,n}}{\partial k_x^+}\right)^{-1} = -\frac{\partial E_{T,n}}{\partial x}\frac{1}{\hbar\, v_{g,n}^+(E_{T,n})}. \tag{5.15}$$

Eq.5.13 is written by using Eq.5.15 as:

$$\frac{df_n^+(x,E_{T,n})}{dx} = \frac{\partial f_n(x,k_x^+)}{\partial x} - \frac{\partial f_n(x,k_x^+)}{\partial k_x^+}\left[\frac{\partial E_{T,n}}{\partial x}\frac{1}{\hbar\, v_{g,n}^+(E_{T,n})}\right] \tag{5.16}$$

and by multiplying both sides of Eq.5.16 by the group velocity $v_{g,n}^+(k_x)$ (it is the positive carrier velocity, i.e. the one corresponding to k_x^+ values) we obtain exactly the *l.h.s.* of the BTE equation for a 1D GAS:

$$v_{g,n}^+(E_{T,n})\frac{df_n^+(x,E_{T,n})}{dx} = v_{g,n}(k_x^+)\frac{\partial f_n(x,k_x^+)}{\partial x} - \frac{1}{\hbar}\frac{df_n(x,k_x^+)}{dk_x^+}\frac{\partial E_{T,n}}{\partial x}. \tag{5.17}$$

So far no assumption about the expression for the total energy has been made. We recall that the derivation that leads to Eq.5.17 is valid except when $\frac{\partial E_{T,n}}{\partial k_x} \neq 0$, that is when the total energy is larger than the subband energy ε_n (see Fig.5.4). By following the same steps that lead to Eq.5.17 also for the occupation function corresponding to negative k_x values, the original BTE can be split in two equations:

$$\begin{aligned}
&\text{BTE}^+: \; v_{g,n}^+\frac{df_n^+(x,E_{T,n})}{dx} = S_n^{in,+} - S_n^{out,+}, \quad \text{for } k_x > 0 \\
&\text{BTE}^-: \; v_{g,n}^-\frac{df_n^-(x,E_{T,n})}{dx} = S_n^{in,-} - S_n^{out,-}, \quad \text{for } k_x < 0.
\end{aligned} \tag{5.18}$$

Before moving forward to the analysis of the scattering integral of the BTE, it is useful to recall that, for a given x, the integral for positive k_x values of a generic $g(k_x)$ function, can be transformed into an energy integral for the function $g^+(E_{T,n})$ (where the + apex denotes that the function g is evaluated for positive $k_x(E_{T,n})$ values) as:

$$\begin{aligned}
\int_0^{+\infty} g(k_x)dk_x &= \int_{\varepsilon_n}^{+\infty} g^+(E_{T,n})\frac{\partial k_x}{\partial E_{T,n}}dE_{T,n} \\
&= \int_{\varepsilon_n}^{+\infty} \frac{g^+(E_{T,n})}{\hbar\, v_{g,n}^+(E_{T,n})}dE_{T,n}
\end{aligned} \tag{5.19}$$

the same can be done for the integrals over negative k_x values:

$$\begin{aligned}
\int_{-\infty}^{0} g(k_x)dk_x &= \int_{+\infty}^{\varepsilon_n} \frac{g^-(E_{T,n})}{\hbar\, v_{g,n}^-(E_{T,n})}dE_{T,n} \\
&= -\int_{+\infty}^{\varepsilon_n} \frac{g^-(E_{T,n})}{\hbar\, v_{g,n}^+(E_{T,n})}dE_{T,n} \\
&= \int_{\varepsilon_n}^{+\infty} \frac{g^-(E_{T,n})}{\hbar\, v_{g,n}^+(E_{T,n})}dE_{T,n}
\end{aligned} \tag{5.20}$$

where ε_n is the eigenvalue for the $n-$th subband, $v_{g,n}^-(E_{T,n})$ is the group velocity for a given total energy $E_{T,n}$ for negative k_x values, namely, it is always smaller than zero and $v_{g,n}^+(E_{T,n}) = -v_{g,n}^-(E_{T,n})$.

The right-hand side of Eq.5.18, represents the scattering integral for occupation functions defined over a positive or negative k_x domain. In order to be more explicit, we consider the term $S^{in,+}$ in Eq.5.18 (same steps can be used for $S^{in/out,-/+}$). $S^{in,+}$ is the in-scattering integral for final states belonging to f_n^+:

$$\begin{aligned} S^{in,+} &= \left(1 - f_n(k_x^+)\right) \frac{L_{NW}}{2\pi} \sum_{n'} \int_{k'_x} S_{n'n}(k'_x, k_x^+) f_{n'}(k_x) dk'_x \\ &= \left(1 - f_n(k_x^+)\right) \frac{L_{NW}}{2\pi} \sum_{n'} \left[\int_{-\infty}^{0} S_{n'n}(k'_x, k_x^+) f_{n'}(k_x) dk'_x + \int_{0}^{\infty} S_{n'n}(k'_x, k_x^+) f_{n'}(k_x) dk'_x \right] \end{aligned} \tag{5.21}$$

where $S_{n'n}(k'_x, k_x)$ is the transition rate from the state $k'_{x,n'}$ to $k_{x,n}$ (as mentioned before, for simplicity of notation, the section index x and valley index ν have been dropped). It can be noticed that even though the in-scattering integral of Eq.5.21 is written for the $f_n(k_x^+)$ occupation function (for states with positive k_x), the summations over the initial states k'_x are performed for both positive and negative k'_x values. In this sense the right-hand side of the BTE (scattering integrals) couples the $f_n(k_x^+)$ and $f_n(k_x^-)$ occupation functions, or, by writing the occupation functions as a function of the energy, couples the $f_n^+(E_{T,n})$ and $f_n^-(E_{T,n})$ occupation functions. By using Eqs.5.19,5.20 integrals over k'_x are transformed into integrals over the total energy $E_{T,n'}$, and the in- and out-scattering integrals for the $f_n(k_x)$ occupation function become:

$$\begin{aligned} S^{in,+} = \left(1 - f_n^+(E_{T,n})\right) \frac{L_{NW}}{2\pi} \sum_{n'} \Bigg[&\int_{\varepsilon'_n}^{\infty} \frac{S_{n'n}(k'^{-}_x, k_x^+) f_{n'}^-(E_{T,n'})}{\hbar\, v_{g,n'}^+(E_{T,n'})} dE_{T,n'} \\ &+ \int_{\varepsilon'_n}^{\infty} \frac{S_{n'n}(k'^{+}_x, k_x^+) f_{n'}^+(E_{T,n'})}{\hbar\, v_{g,n'}^+(E_{T,n'})} dE_{T,n'} \Bigg] \end{aligned} \tag{5.22}$$

where $k'^{+(-)}_x = k'^{+(-)}_x(E_{T,n'})$ is the positive (negative) k_x vector of the electron belonging to subband n' before scattering to the state $k_x^+ = k_x^+(E_{T,n})$, and

$$\begin{aligned} S^{out,+} = \left(f_n^+(E_{T,n})\right) \frac{L_{NW}}{2\pi} \sum_{n'} \Bigg[&\int_{\varepsilon'_n}^{\infty} \frac{S_{nn'}(k_x^+, k'^{-}_x)(1 - f_{n'}^-(E_{T,n'}))}{\hbar\, v_{g,n'}^+(E_{T,n'})} dE_{T,n'} \\ &+ \int_{\varepsilon'_n}^{\infty} \frac{S_{nn'}(k_x^+, k'^{+}_x)(1 - f_{n'}^+(E_{T,n'}))}{\hbar\, v_{g,n'}^+(E_{T,n'})} dE_{T,n'} \Bigg] \end{aligned} \tag{5.23}$$

where $k'^{+(-)}_x = k'^{+(-)}_x(E_{T,n'})$ is the positive (negative) k_x vector of the electron belonging to subband n' after scattering from the state $k_x^+ = k_x^+(E_{T,n})$. Transition rate $S_{n'n}(k'_x, k_x)$ can be written by means of the Fermi-Golden's Rule introducing the matrix element $M_{n'n}(k_x, k'_x)$ for the transition between the state $k'_{n'}$ to k_n:

$$S_{n'n}(k'_x, k_x) = \frac{2\pi}{\hbar} \left| M_{n'n}(k'_x, k_x) \right|^2 \delta\left(E_{T,n'}(k'_x) - E_{T,n}(k_x) \pm \hbar\omega_{q_x}\right). \tag{5.24}$$

where $\hbar\omega_{q_x}$ with $q_x = k_x - k'_x$ is a term indicating the energy exchanged with the scattered electron and for acoustic phonons $\hbar\omega_{q_x}$=0 whereas for optical phonons $\hbar\omega_{q_x} \neq 0$, and, $\delta(\bullet)$ is a Dirac delta. By substituting Eq.5.24 into Eq.5.22 for the in-scattering collision integral for f_n^+, we obtain:

$$S^{in,+} = \left(1 - f_n^+(E_{T,n})\right) \frac{L_{NW}}{2\pi} \sum_{n'} \left[\frac{2\pi}{\hbar^2 v_{g,n'}^+(E_{T,n'})} \left| M_{n'n}(k_x'^-, k_x^+) \right|^2 f_{n'}^-(E_{T,n'}) + \right. \\ \left. + \frac{2\pi}{\hbar^2 v_{g,n'}^+(E_{T,n'})} \left| M_{n'n}(k_x'^+, k_x^+) \right|^2 f_{n'}^+(E_{T,n'}) \right]. \tag{5.25}$$

The Dirac function in Eq.5.24 reduces the integral in Eq.5.22 to an evaluation of the occupation functions at discrete energy values $E_{T,n'} = E_{T,n}(k) \pm \hbar\omega_{q_x}$ leading to a considerable reduction of complexity of the BTE.

5.3.2 Evaluation of macroscopic quantities

The solution of the coupled BTE and Poisson's equations with appropriate boundary conditions provides a self-consistent occupation function $f_n(x, k_x)$. Once $f_n(x, k_x)$ is known, all the internal quantities of interest in the device can be derived.

Inversion electron density

Given the coordinate system in Fig.B.2, the total inversion electron density is calculated by definition as :

$$n_{inv}(r, \theta, x) = \frac{1}{L_{NW}} \sum_{\nu,n} \mu_\nu \left| \Phi_{\nu,n}(r, \theta, x) \right|^2 \sum_{k_x} f_{\nu,n}(x, k_x) \tag{5.26}$$

where μ_ν is the valley multiplicity, $\Phi_{\nu,n,x}(r, \theta)$ is the wave-function for a given valley ν, subband n and position x along transport direction and it is the solution of the Schrödinger equation in the circular slices of the NW. By splitting the calculation for positive and negative k_x values we write:

$$n_{inv}(r, \theta, x) = \frac{1}{L_{NW}} \sum_{\nu,n} \mu_\nu \left| \Phi_{\nu,n,x}(r, \theta) \right|^2 \left(\sum_{k_x<0} f_{\nu,n}(x, k_x) + \sum_{k_x>0} f_{\nu,n}(x, k_x) \right). \tag{5.27}$$

By exploiting Eq.5.7 to convert the summation over k_x into an integral by setting the spin coefficient $n_{sp} = 2$ we obtain:

$$n_{inv}(r, \theta, x) = \frac{1}{\pi} \sum_{\nu,n} \mu_\nu \left| \Phi_{\nu,n,x}(r, \theta) \right|^2 \left(\int_{-\infty}^{0} f_{\nu,n}(x, k_x) dk_x + \int_{0}^{+\infty} f_{\nu,n}(x, k_x) dk_x \right). \tag{5.28}$$

Integrals over k_x are converted into total energy integrals by means of Eqs.5.20,5.19

$$n_{inv}(r,\theta,x) = \frac{1}{\pi}\sum_{\nu,n}\mu_\nu \left|\Phi_{\nu,n,x}(r,\theta)\right|^2 \left(\int_{\varepsilon_{\nu,n}}^{+\infty} \frac{f^-_{\nu,n}(x,E_{T,n})}{\hbar\, v^+_{g,\nu,n}(x,E_{T,n})} dE_{T,n} + \right. \\ \left. + \int_{\varepsilon_{\nu,n}}^{+\infty} \frac{f^+_{\nu,n}(x,E_{T,n})}{\hbar\, v^+_{g,\nu,n}(x,E_{T,n})} dE_{T,n}\right). \tag{5.29}$$

Eq.5.29 is valid for both parabolic and non-parabolic total energy description since non-parabolicity effects in Eq.5.29 enter in the group velocity terms only.

Hence, the charge per unit of length can be obtained by integrating Eq.5.29 over r and θ leading to:

$$n_{inv}(x) = \frac{1}{\hbar\pi}\sum_{\nu,n}\mu_\nu \left(\int_{\varepsilon_{\nu,n}}^{+\infty} \frac{f^-_{\nu,n}(x,E_{T,n})}{v^+_{g,\nu,n}(x,E_{T,n})} dE_{T,n} + \int_{\varepsilon_{\nu,n}}^{+\infty} \frac{f^+_{\nu,n}(x,E_{T,n})}{v^+_{g,\nu,n}(x,E_{T,n})} dE_{T,n}\right). \tag{5.30}$$

Electron current

The calculation of electron current follows the same steps as in the calculation of the inversion electron density. The starting point is to resort to the equation for the current:

$$I(x) = \frac{e}{L_{NW}}\sum_{\nu,n}\mu_\nu \sum_{k_x} v_{g,\nu,n}(x,k_x) f_{\nu,n}(x,k_x). \tag{5.31}$$

By converting the summation over k_x into an integral (see Eq.5.7 setting the spin coefficient $n_{sp}=2$) and then by splitting the k_x domain into positive and negative k_x and moving from integrals over $k_x^{+/-}$ to energy integrals we obtain:

$$I(x) = \frac{e}{\hbar\pi}\sum_{\nu,n}\mu_\nu \left(\int_{\varepsilon_{\nu,n}}^{+\infty} f^+_{\nu,n}(x,E_{T,n})dE_{T,n} - \int_{\varepsilon_{\nu,n}}^{+\infty} f^-_{\nu,n}(x,E_{T,n})dE_{T,n}\right). \tag{5.32}$$

Hence, current calculation reduces to an integral over the occupation functions and it is valid for both parabolic and non-parabolic descriptions of the total energy dispersion relation.

5.3.3 Domain discretization

As seen in Sec.5.2, the unknowns of the BTE are the occupation functions $f_n(x,E_{T,n})$. In other words, the domain of the BTE for a 1D electron gas consists of the space coordinate x that denotes the transport direction, the total energy $E_{T,n}$ of electrons but also the valley ν, the subband n and the sign $(\pm)$ of the occupation function corresponding to positive of negative $k_x(E_{T,n})$ values. Before solving numerically the BTE in Eq.5.18, a discretized domain must be defined. Fig.5.5 shows a subband energy profile in the $x - E_{T,n}$ space with a uniform grid for

energy and x. The unknown occupation functions for a given subband correspond to energies larger than the subband profile and are identified by the solid circles in Fig.5.5. From the implementation point of view, a given unknown of the BTE problem is identified as $f^t_{s,\nu,n,j}$ where these indexes denote the section s, the valley ν, the subband n, the energy j, and the type $t \in [+,-]$, which is the sign of the associated k_x vector. Hence the total number of unknowns is:

$$N_{unk} = 2 \cdot \sum_j \sum_\nu \sum_n n_{sec}(j,\nu,n) \tag{5.33}$$

where $n_{sec}(j,\nu,n)$ is the number of discretization points along x direction for a given (j,ν,n) and the 2 derives from the fact that for a given set of indexes (s,ν,n,j) can be defined two unknowns corresponding to positive or negative k_x values.

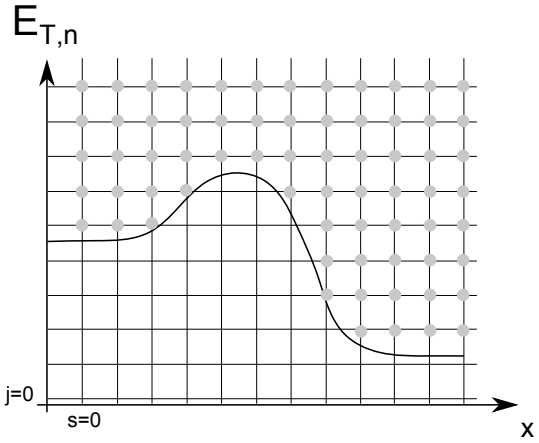


Figure 5.5: Subband profile with discretization of the real and energy space for the solution of the BTE. Solid circles represent two generic discretized occupation functions (for positive and negative k_x values).

5.3.4 Discretization of the left hand side of the BTE

The left-hand side of the BTE reported in Eq.5.18 is discretized by means of finite-difference method. Hence, the derivative term in the *l.h.s.* of the two BTEs written at section s is given by:

$$\begin{aligned} \frac{df^+}{dx}|_{x_s} &= \frac{f^+_{s,n,j} - f^+_{s-1,n,j}}{x_s - x_{s-1}} \\ \frac{df^-}{dx}|_{x_s} &= \frac{f^-_{s+1,n,j} - f^-_{s,n,j}}{x_{s+1} - x_s}. \end{aligned} \tag{5.34}$$

where the valley index ν has been dropped, which leads to:

$$\begin{aligned} BTE^{+}_{s,n,j}: \quad & \underbrace{v^{+}_{g,s,n,j}\frac{f^{+}_{s,n,j}-f^{+}_{s-1,n,j}}{\Delta x}}_{l.h.s._s} = \underbrace{\left(S^{in,+}_{s,n,j}-S^{out,+}_{s,n,j}\right)}_{r.h.s._s} \\ BTE^{-}_{s,n,j}: \quad & \underbrace{v^{-}_{g,s+1,n,j}\frac{f^{-}_{s+1,n,j}-f^{-}_{s,n,j}}{\Delta x}}_{l.h.s._s} = \underbrace{\left(S^{in,-}_{s+1,n,j}-S^{out,-}_{s+1,n,j}\right)}_{r.h.s._{s+1}} \end{aligned} \tag{5.35}$$

We will see in Sec.5.3.7 that this discretization technique allows for the current to be conserved in the different sections along transport directions.

The BTE can be equivalently written evaluating the scattering integrals at the same section s as:

$$\begin{aligned} BTE^{+}_{s,n,j}: \quad & \underbrace{v^{+}_{g,s,n,j}\frac{f^{+}_{s,n,j}-f^{+}_{s-1,n,j}}{\Delta x}}_{l.h.s._s} = \underbrace{\left(S^{in,+}_{s,n,j}-S^{out,+}_{s,n,j}\right)}_{r.h.s._s} \\ BTE^{-}_{s-1,n,j}: \quad & \underbrace{v^{-}_{g,s,n,j}\frac{f^{-}_{s,n,j}-f^{-}_{s-1,n,j}}{\Delta x}}_{l.h.s._{s-1}} = \underbrace{\left(S^{in,-}_{s,n,j}-S^{out,-}_{s,n,j}\right)}_{r.h.s_s} \end{aligned} \tag{5.36}$$

where in the term $r.h.s_s$ we have included all the occupation functions that are linked by scattering mechanisms with $f^{+}_{s,n,j}$ for the BTE+ and $f^{-}_{s,n,j}$ for the BTE−. We will see that this numerical method to discretize the spatial derivatives of the BTE allows to account for boundary conditions at the left side and right side of the domain (source and drain contacts) in a very natural way.

5.3.5 Boundary Conditions

Source and drain regions are assumed to behave as perfect reservoirs of carriers at thermal equilibrium, and the boundary conditions of the BTE are given by the occupation function of electrons injected with positive group velocity (i.e. carriers moving from source to drain) from the source contact, and by electrons injected from the drain contact with negative group velocity. For such electrons the occupation function is simply determined by their total energy through the Fermi-Dirac occupation function and the Fermi level at the contacts:

$$f_0(E_{T,n}) = \frac{1}{1+exp(\frac{E_{T,n}-E_{F,source/drain}}{K_B T})}. \tag{5.37}$$

If we set the Fermi energy at the source contact to 0 eV, the drain Fermi energy will be given by $-eV_{DS}$, thus occupation functions at source contact for carriers with k^{+}_{x}

($f^+_{n,source}$) and at drain contact for carriers with k^-_x ($f^-_{n,drain}$) are given by:

$$f_{0,source}(E_{T,n}) = \frac{1}{1+exp(\frac{E_{T,n}}{K_BT})}$$
$$f_{0,drain}(E_{T,n}) = \frac{1}{1+exp(\frac{E_{T,n}+eV_{DS}}{K_BT})} \tag{5.38}$$

where K_B is the Boltzmann constant and T the temperature.

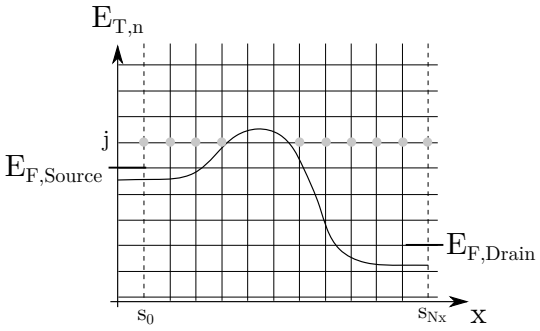


Figure 5.6: Subband profile with discretization of the real and energy space for the solution of the BTE. Solid circles represent two generic discretized occupation functions (for k^+_x and k^-_x).

It must be noticed that there are no boundary conditions for $f^-_{n,source}$ and $f^+_{n,drain}$ because the corresponding source and drain regions must be long enough to let $f^-_{n,source}$ and $f^+_{n,drain}$ relax to their quasi-equilibrium regions via scattering inside the source and drain extensions. According to the above considerations, the $\text{BTE}^+_{s,n,j}$ of Eq.5.35 is written only for the section indexes $s = 1, \ldots, N_x$ whereas for the $\text{BTE}^-_{s,n,j}$ only for $s = 0, \ldots, N_x - 1$. It follows that boundary conditions enter in the $l.h.s.$ of the $\text{BTE}^+_{s,n,j}$ for section $s = 1$ (see Fig.5.6) as reported in the explicit expression for the $\text{BTE}^+_{s,n,j}$ taken from Eq.5.35:

$$v^+_{g,1,n,j}\frac{df^+}{dx}|_{x_1} = v^+_{g,1,n,j}\frac{f^+_{1,n,j} - f_{0,source}(n,j)}{\Delta x} \tag{5.39}$$

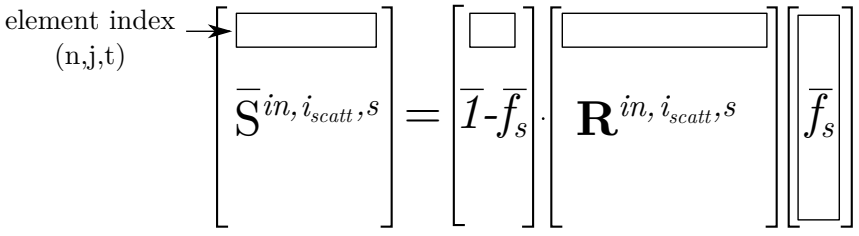
and for section $s = N_x - 1$ for the $\text{BTE}^-_{s,n,j}$:

$$v^-_{g,N_x,n,j}\frac{df^-}{dx}|_{x_{N_x-1}} = v^-_{g,N_x,n,j}\frac{f_{0,drain}(n,j) - f^-_{N_x-1,n,j}}{\Delta x} \tag{5.40}$$

5.3.6 Discretization of the right hand side of the BTE

As shown in Sec.5.3.1, scattering integrals can be written for a generic scattering mechanism by evaluating the occupation functions at discrete energy values as shown in Eq.5.25 for the in-scattering integral for a generic f^+_n unknown of the

BTE. Hence, we can define the in-scattering matrix $\mathbf{R}^{in,i_{scatt},s}$ for the $i_{scatt}-$th scattering mechanisms at section s as the matrix of dimensions $[N^s_{unk}\mathrm{x}N^s_{unk}]$ where N^s_{unk} represents the number of unknowns for a given section s, that links the generic unknown $f^t_{s,n,j}$ with all others occupation functions $f^{t'}_{s,n',j'}$ belonging to the same section s because the scattering events are instantaneous in time, hence, leaving unchanged the position x of the scattered electron. By means of this matrix, the in-scattering integral for all the unknowns belonging to a given section s is given by:



(5.41)

where $\overline{f_s} = \overline{f}^t_{s,n,j}$ represents the vector of unknowns for a given section s. Hence the in- and out- scattering integrals of Eqs.5.22 and 5.23, after replacing transition rates with Eq.5.24 for a generic unknown identified by the set of indexes (n,j,t) and for a given section s, can be written as:

$$\begin{aligned} S^{in,i_{scatt},s}_{n,j,t} &= \left[1 - f^t_{s,n,j}\right] \left(\mathbf{R}^{in,i_{scatt},s}_{n,j,t} \cdot \overline{f_s}\right) \\ S^{out,i_{scatt},s}_{n,j,t} &= f^t_{s,n,j} \left(\mathbf{R}^{in,i_{scatt},s}_{n,j,t} \cdot \left[\overline{1} - \overline{f_s}\right]\right). \end{aligned} \tag{5.42}$$

It is interesting to note that in the case of multiple scattering mechanisms the matrices $\mathbf{R}^{in/out,s}$ that account for all the scattering mechanisms can be simply added:

$$\begin{aligned} \mathbf{R}^{in,s} &= \sum_{i_{scatt}} \mathbf{R}^{in,i_{scatt},s} \\ \mathbf{R}^{out,s} &= \sum_{i_{scatt}} \mathbf{R}^{out,i_{scatt},s}. \end{aligned} \tag{5.43}$$

Finally, the discretized BTE written by evaluating the *r.h.s.* for both BTE+ and BTE− at the same section becomes:

$$\begin{aligned} v^+_{g,s,n,j} \frac{f^+_{s,n,j} - f^+_{s-1,n,j}}{\Delta x} = \\ \left[1 - f^+_{s,n,j}\right] \left(\mathbf{R}^{in,s}_{n,j,+} \cdot \overline{f_s}\right) - f^+_{s,n,j} \left(\mathbf{R}^{out,i_{scatt},s}_{n,j,+} \cdot \left[\overline{1} - \overline{f_s}\right]\right), \quad \text{for } s = 1,\ldots,N_x \\ v^-_{g,s,n,j} \frac{f^-_{s,n,j} - f^-_{s-1,n,j}}{\Delta x} = \\ \left[1 - f^-_{s,n,j}\right] \left(\mathbf{R}^{in,s}_{n,j,-} \cdot \overline{f_s}\right) - f^-_{s,n,j} \left(\mathbf{R}^{out,i_{scatt},s}_{n,j,-} \cdot \left[\overline{1} - \overline{f_s}\right]\right), \quad \text{for } s = 1,\ldots,N_x \end{aligned} \tag{5.44}$$

5.3.7 Current conservation

We now demonstrate that the BTE discretization method of Eq.5.35 used in this work is such that the writing of the BTE^- and BTE^+ for all indexes n, j, t with the *r.h.s.* evaluated in section s as reported in Eq.5.36 (or more explicitly in Eq.5.44) leads to the current conservation between section s and section $s-1$. Equivalently, the current is conserved between section $s-1$ and section s when are written the $BTE^-_{s-1,n,j}$ for section $s-1$ and the $BTE^+_{s,n,j}$ for section s given in Eq.5.36 for all indexes n, j.

Let us first suppose the case of elastic scattering where only acoustic phonons are considered with only one subband. Under this condition each energy bin leads to a separate BTE problem since there are no scattering mechanisms that connect electrons belonging to different energy bins and for this reason in the following calculations n and j indexes will be dropped. Moreover, in this case scattering can only change the sign of the k_x associated to the scattering electron (i.e. the final state for the unknown identified by indexes (s, n, j, t) must be $(s, n, j, -t)$) as shown in Fig.5.7. Under the hypothesis of single subband and elastic scattering, the scattering rate $S(k_x^+, k_x^-)$ equals to $S(k_x^-, k_x^+)$, being the transition rate according to the Fermi's Golden rule given by:

$$\begin{aligned} S_{nn}(k_x^-, k_x^+) &= \frac{2\pi}{\hbar} |M|^2 \, \delta(E_{T,n}, E_{T,n}) \\ &= \frac{2\pi}{\hbar} \left(\frac{K_B T D_{acc}^2}{\rho L_{NW} v_{sound}^2} F_{n,n} \right) \delta(E_{T,n}, E_{T,n}) \end{aligned} \tag{5.45}$$

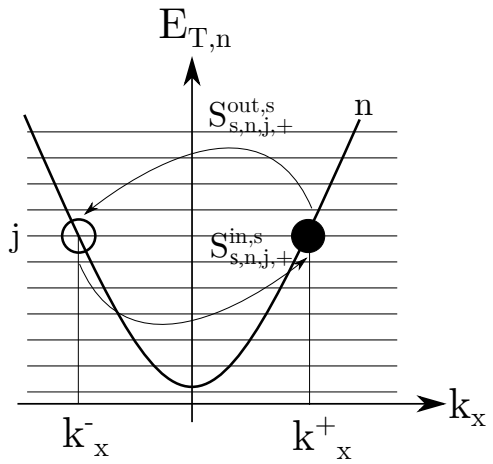


Figure 5.7: Schematic representation of elastic in- and out-scattering referred to the electron identified by indexed $s, n, j, +$ in the case of single subband. The scattering can only change the sign of the k_x vector.

where $F_{n,n'}$ is the form factor defined in Eq.4.101, K_B is the Boltzmann constant, T is the temperature, D_{acc} is the acoustic deformation potential, ρ is the density

of the semiconductor and v_{sound} is the sound velocity inside the semiconductor. According to Eq.5.25 the in-scattering integral for a given section s is given by:

$$S_s^{in,+} = (1 - f_s^+) \underbrace{\left(\frac{K_B T D_{acc}^2 F_{n,n}^s}{\hbar^2 \rho v_{sound}^2 v_{g,n}^+} \right)}_{R^s} f_s^- \tag{5.46}$$

where the normalization length L_{NW} in Eq.5.45 canceled out by converting the summation over states k_x' to integrals in the calculation of the in-scattering integral. Hence the $r.h.s.$ of $BTE\pm$ evaluated at section s can be written in a compact form as:

$$\begin{aligned}
BTE_s^+ : \quad S_s^{in,+} - S_s^{out,+} &= \left(1 - f_s^+\right) R^s f_s^- - f_s^+ R^s \left(1 - f_s^-\right) \\
&= \underbrace{+R^s f_s^- - R^s f_s^+}_{r.h.s._s^+} \\
BTE_{s-1}^- : \quad S_s^{in,-} - S_s^{out,-} &= \left(1 - f_s^-\right) R^s f_s^+ - f_s^- R^s \left(1 - f_s^+\right) \\
&= \underbrace{-R^s f_s^- + R^s f_s^+}_{r.h.s._s^-}
\end{aligned} \tag{5.47}$$

where it can be noticed that the BTE reduces to a linear problem. It can be demonstrated
the nonlinear terms in the $r.h.s.$ cancel out leading to a linear BTE. The BTE$\pm$ in Eq.5.36 with the $r.h.s.$ evaluated in section s can be written as:

$$\begin{cases}
v_{g,s}^+ \dfrac{f_s^+ - f_{s-1}^+}{\Delta x} = r.h.s._s^+ \\
v_{g,s}^- \dfrac{f_s^- - f_{s-1}^-}{\Delta x} = -r.h.s._s^-
\end{cases} \tag{5.48}$$

By noting that in Eq.5.48 we have $r.h.s._s^+ = -r.h.s._s^- = r.h.s._s$ (see Eq.5.47) and recalling that $v_{g,s,n,j}^+ = -v_{g,s,n,j}^-$, the BTEs in Eq.5.48 finally become:

$$\begin{cases}
v_{g,s}^+ \dfrac{f_s^+ - f_{s-1}^+}{\Delta x} = r.h.s._s \\
v_{g,s}^+ \dfrac{f_s^- - f_{s-1}^-}{\Delta x} = r.h.s._s
\end{cases}. \tag{5.49}$$

which leads to:

$$f_s^+ - f_s^- = f_{s-1}^+ - f_{s-1}^-. \tag{5.50}$$

According to Eq.5.32, Eq.5.50 gives the conservation between section s and $s-1$ of the current density at the fixed energy value used to write Eqs.5.45 to 5.49. Since we are dealing with elastic scattering, current conservation is guarantee energy per energy, but in general, current conservation is obtained after summing over all subbands, and energy bins.

Current conservation can be algebraically demonstrated also in the case of elastic scattering and multiple subbands, energy per energy but after summing over the subbands n. Moreover, also in the more complicated case of inelastic scattering, current conservation (after summing over all the subbands and energy bins) can be numerically demonstrated. So far we have assumed that the BTE can be written at any point of the discretized $x - E_{T,n}$ space of Fig.5.5, but this is not the case at the classical turning points. In fact, the BTE *l.h.s.* discretization method proposed in Eq.5.36 becomes critical when we write the BTEs at the square points shown in Fig.5.8. These points are such that, taken a generic unknown $f^{\pm}_{s,n,j}$, there exist no unknowns at section $s+1$ or $s-1$ for the same n, j indexes. We will refer to these points as classical-turning-point (CTPs) for the unknown $f^{\pm}_{s,n,j}$. When solving the

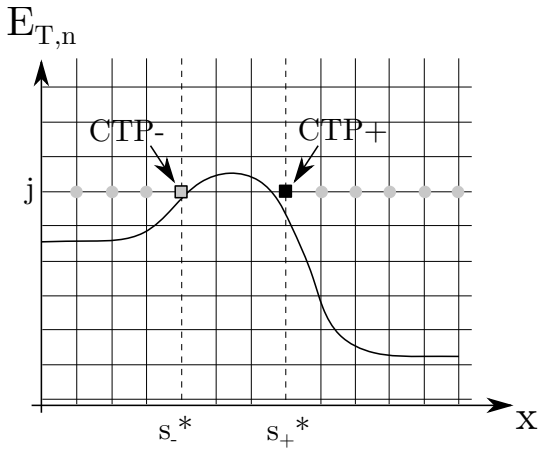


Figure 5.8: Subband profile with discretization of the real and energy space for the solution of the BTE. Solid circles represent a generic discretized unknown of the BTE and square symbols the unknowns at the so-called classical turning points (CTPs).

transport in the entire spatial domain, for each point belonging to $n_{sec}(j,n)$ (where $n_{sec}(j,n)$ is the number of discretization points along x direction for a given (j,n)) we write the BTE as in Eq.5.44. For the point s^*_+ shown in Fig.5.8, the *l.h.s* of the $BTE\pm$ becomes:

$$\begin{aligned} BTE^{+}_{s^*_+,n,j} &: v^{+}_{g,s^*_+,n,j}\frac{f^{+}_{s^*_+,n,j} - f^{+}_{(s^*_+-1),n,j}}{\Delta x} \\ BTE^{-}_{s^*_+,n,j} &: v^{-}_{g,s^*_+ +1,n,j}\frac{f^{-}_{s^*_+ +1,n,j} - f^{-}_{(s^*_+),n,j}}{\Delta x} \end{aligned} \tag{5.51}$$

The unknown $f^{+}_{(s^*_+-1),n,j}$ for the $BTE^{+}_{s^*_+,n,j}$ in Eq.5.51 does not exist because it would be located under the subband profile, and thus in a forbidden region. For the *l.h.s*

for the $BTE\pm$ evaluated in s^*_-, we write:

$$\begin{aligned} BTE^+_{s^*_-,n,j} &: v^+_{g,s^*_-,n,j}\frac{f^+_{(s^*_-),n,j} - f^+_{(s^*_- -1),n,j}}{\Delta x} \\ BTE^-_{s^*_-,n,j} &: v^-_{g,(s^*_- +1),n,j}\frac{f^-_{(s^*_- +1),n,j} - f^-_{s^*_-,n,j}}{\Delta x} \end{aligned} \tag{5.52}$$

which requires the knowledge of the unknown $f^-_{(s^*_- +1),n,j}$ for the $BTE^-_{s^*_-,n,j}$ that also does not exist because it would be located under the subband profile. To overcome the issue given by the fact that the BTE cannot be written for the $f^+_{s,n,j}$ occupation functions at the CTP+ and for the $f^-_{s,n,j}$ occupation functions at the CTP−, we will show how to introduce alternative equations that ensure the continuity of the current along the transport direction.

To summarize, the discretization of the BTE used in this work is such that if the *l.h.s.* of the $BTE^-_{s-1,n,j}$ and the *l.h.s.* of the $BTE^+_{s,n,j}$ can be written for all n, j values, i.e. if the corresponding unknowns in Eq.5.35 exist, then current conservation is guaranteed between the section s and section $s-1$. In other words, if both the *r.h.s.* for $BTE\pm$ evaluated at section s can be written and at the section $s-1$ there are no CTP−, then current conservation is guaranteed between the section s and section $s-1$. This rule can be rewritten in a more compact form saying that current conservation between section s and section $s-1$ is guaranteed if:

- there are no CTP− in section $s-1$
- there are no CTP+ in section s.

In fact, as shown in Fig.5.9a, if we consider the two sections inside the dashed line, both the *l.h.s.* for the $BTE^-_{s-1,n,j}$ and the *l.h.s.* for the $BTE^+_{s,n,j}$ given in Eq.5.36 can be legitimately written thus leading to the conservation of the current between section s and $s-1$. The same does not hold for Fig.5.9b because there is a CTP+ is section $s-1$ and energy bin j, and consequently we cannot write the *l.h.s.* for $BTE^-_{s-1,n,j}$.

5.3.8 Classical Turning Points

As we have seen in Sec.5.3.7, the BTE cannot be written for the $f^+_{s,n,j}$ if the point (s,j) of the $x - E_{T,n}$ domain for the subband n is a CTP+ and for the $f^-_{s,n,j}$ if the point (s,j) of the $x - E_{T,n}$ domain for the subband n is a CTP−. Of course the number of equations and unknowns must match, so that we must have:

$$N_{unk} - N_{boundary} = N_{BTE} + N_{CTPE} \tag{5.53}$$

where N_{unk} are the number of unknowns given by Eq.5.33, N_{BTE} is the number of BTEs written according to Eq.5.36, $N_{boundary}$ is the number of boundary unknowns (see Sec.5.3.5) and $N_{CTPE} = N_{CTP}/2$ are the number of equations written at the

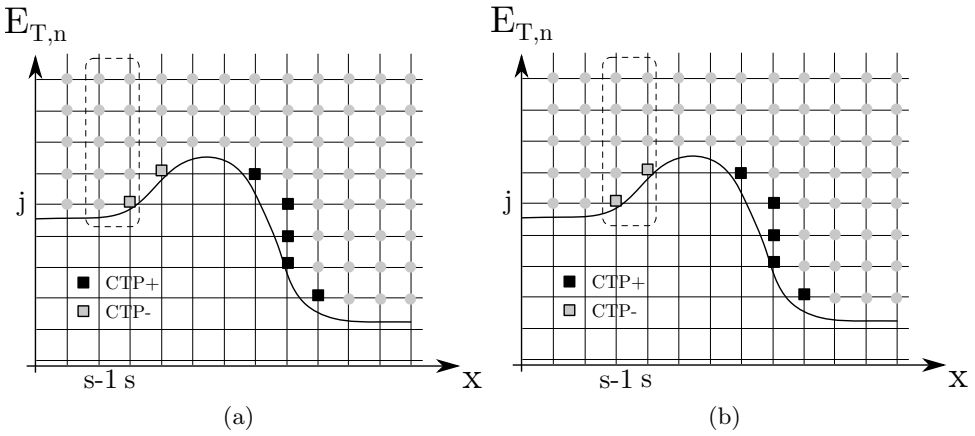


Figure 5.9: Subband profile with unknowns. a) The BTE discretization technique guarantees the continuity of the current between section s and $s-1$ because in section s there are no CTP+ and there are no CPT− in section $s-1$. b) Since in section $s-1$ there is a CTP−, the $BTE^{-}_{s+1,j,n}$ cannot be written leading to an under-determined system of equations for the solution of transport.

CTPs where N_{CTP} is the number of CTPs; the 2 factor is because for each CTP only one equation (for f^+ or for f^-) cannot be written as shown in Sec.5.3.7.

A reasonable choice for the N_{CTPE} equations is to write them in order to ensure the drain-to-source current (I_{DS}) to be solenoidal.

BTE in the presence of CTP−

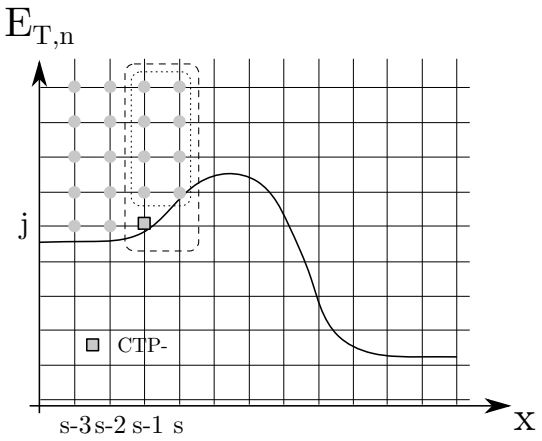


Figure 5.10: Template case study with CTP−.

We now focus of the CTP− as shown in Fig.5.10. As mentioned in Sec.5.3.7, current continuity between sections $s-1$ and $s-2$ is guarantee because the corre-

sponding $BTE^{-}_{s-2,n,j}$ and $BTE^{+}_{s-1,n,j}$ can be written, and the same holds for sections $s-2$ and $s-3$. Instead, we are not able to write the *l.h.s.* of the $BTE^{-}_{s-1,n,j}$ due to the presence of a CTP−, hence, the corresponding $BTE^{-}_{s-1,n,j}$ can be replaced by:

$$f^{-}_{s-1,n,j} = f^{+}_{s-1,n,j}. \tag{5.54}$$

The condition in Eq.5.54 implies that current contribution in section $s-1$ and for the energy index j is null (see Eq.5.31). By replacing the missing $BTE^{-}_{s-1,n,j}$ equation with Eq.5.54 and according to Sec.5.3.7, current is conserved between section s and section $s-1$. In fact, looking at Fig.5.10 for the unknowns inside the dotted region, both $BTE^{-}_{s-1,n,j}$ and $BTE^{+}_{s,n,j}$ can be written and, since the current contribution given by the unknowns at the CTP− at position $(s-1,j,n)$ is null, it follows that current is conserved between section s and $s-1$.

BTE in the presence of CTP+

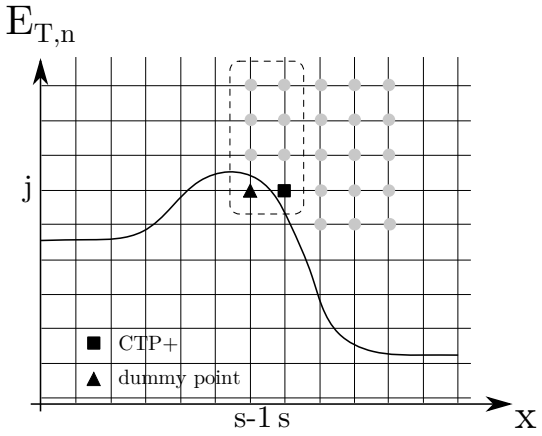


Figure 5.11: Template case study with CTP+.

As reported in Sec.5.3.7 we cannot write the $BTE^{+}_{s,n,j}$ because there is a CTP+ at section s for the unknown belonging to the energy bin j and subband n. Hence, one could be temped to write a similar expression as in Eq.5.54 for the $BTE^{+}_{s,n,j}$ at the CTP+ of Fig.5.11:

$$f^{+}_{s,n,j} = f^{-}_{s,n,j}. \tag{5.55}$$

However, it can be demonstrated that for a general case of multiple subbands with elastic or inelastic scattering, Eq.5.55 brings to a non conservation of the current between section s and $s-1$ in Fig.5.11. This is because, by substituting the $BTE^{+}_{s,n,j}$ with Eq.5.55, all the informations regarding the in- and out-scattering for $f^{+}_{s,n,j}$ contained in the *r.h.s.* of the $BTE^{+}_{s,n,j}$ will be lost. Hence, the balance of charge fluxes at section s due to scattering for all the unknowns linked with $f^{+}_{s,n,j}$ will be modified in disagreement with the BTE (see Sec.5.2) . Our goal here is to find a

relation between $f^+_{s,n,j}$ in the CTP+ and the other unknowns in section s. We thus introduce a *dummy* point (triangle symbol in Fig.5.11) in the forbidden region for subband n for which we write:

$$f^+_{s-1,n,j} = f^-_{s-1,n,j}. \tag{5.56}$$

By doing so, for the unknowns located in (s, n, j) we can write the following equations:

$$\begin{cases} v^+_{g,s,n,j} \dfrac{f^+_{s,n,j} - f^+_{s-1,n,j}}{\Delta x} = r.h.s._s^+ \\ v^-_{g,s,n,j} \dfrac{f^-_{s,n,j} - f^-_{s-1,n,j}}{\Delta x} = r.h.s._s^- \\ f^+_{s-1,n,j} = f^-_{s-1,n,j} \end{cases} \tag{5.57}$$

Eq.5.57 can be rewritten as:

$$f^+_{s,n,j} = f^-_{s,n,j} + \frac{\Delta x}{v^+_{g,s,n,j}}(r.h.s._s^+ + r.h.s._s^-). \tag{5.58}$$

Hence, Eq.5.58 is the new equation to be written in place of the $BTE^+_{s,n,j}$ at the CTP+.

In the ballistic case or in the elastic case with only one subband, the equations for the CTPs according to Eqs.5.54 and 5.58 become:

$$\begin{aligned} f^-_{s_{CTP-},n,j} &= f^+_{s_{CTP-},n,j} \\ f^+_{s_{CTP+},n,j} &= f^-_{s_{CTP+},n,j} \end{aligned} \tag{5.59}$$

that correspond to reflecting boundary conditions at the corresponding turning points as shown in Fig.5.12.

Moreover, in the ballistic limit occupation functions are given by the propagation of boundary conditions through the device. Hence, according to Eqs.5.32 and 5.59 the contribution to the total current is given by the carriers sitting above the top of the subband profile n, because below the top for each section s the current contribution is null being $f^+_{s,n,j} = f^-_{s,n,j}$.

There is also an alternative way to impose current continuity between section s and $s-1$ without resorting to Eqs.5.54 and 5.58 and consist in writing explicitly the continuity equation for current. Let us consider a generic case with inelastic scattering and multiple subbands as shown in Fig.5.13. In this case, we can replace the $BTE^-_{s-1,1,j}$ (that cannot be written in the CTP-) with the continuity condition for the current between section $s-1$ and s that reads:

$$\sum_{j*}\sum_{n}\left(f^+_{s-1,n,j} - f^-_{s-1,n,j}\right) = \sum_{j*}\sum_{n}\left(f^+_{s,n,j} - f^-_{s,n,j}\right) \tag{5.60}$$

where the summation over the energy bins does not need to include all energies but can be limited to the unknowns in section $s-1$ and s connected by scattering

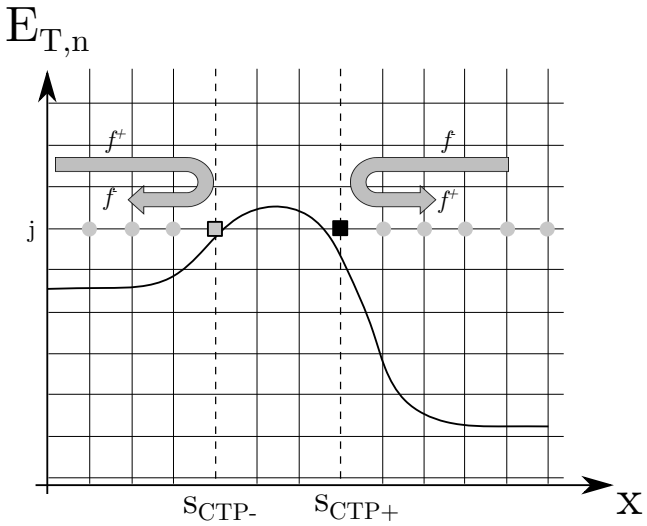


Figure 5.12: Schematic diagram showing the perfect reflection of the carriers at the CTPs in the case ballistic transport.

with the occupation function $f^{-}_{s-1,1,j}$ at the CTP$-$. In this case, only the unknowns connected by arrows shown in Fig.5.13 enter Eq.5.60. This method can be used either for the $\mathrm{BTE}^{-}_{s,n,j}$ for which the point (s,j) for subband n is a CTP$-$ or for the $\mathrm{BTE}^{+}_{s,n,j}$ for which the point (s,j) for subband n is a CTP+. However, it must be noticed that for each section s and group of unknowns linked by scattering, the continuity equation of Eq.5.60 can be written only once. In fact, in the case of multiple CTP per section connected by scattering, it will result in the imposition of multiple identical equations leading the system to be solved for the transport underdetermined. Hence, from the practical point of view, the direct imposition of the continuity condition on the current as in Eq.5.60 it is not a viable solution.

5.3.9 Critical Regions

During the convergence loop between the Schrödinger equation and the Poisson equation, or due to resistive voltage drops in the source or drain regions, the trajectory of electrons for a given energy bin can exhibit two CTPs as shown in Fig.5.14. In the case of elastic scattering (and assuming a single subband as in Fig.5.15a, or, assuming multiple subbands where electrons inside the critical region are not connected by elastic scattering with carriers of other subbands as in Fig.5.15b), in principle it is not possible to solve the BTE inside that region because the connection with the boundaries is lost. However, if carriers inside a critical region are connected with the boundaries of the transport problem (i.e. source and drain contacts) through scattering, either inelastic or elastic as in Fig.5.16, the BTE can be solved for all the unknowns $f^{t}_{s,n,j}$.

In this work, if carriers inside a critical region are not connected with boundaries

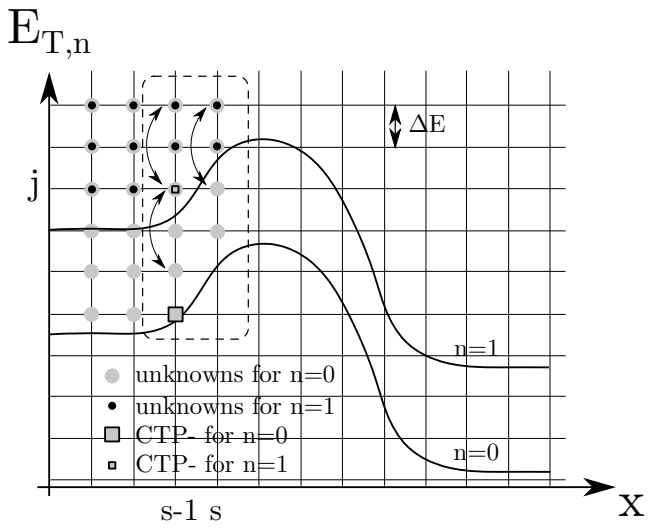


Figure 5.13: Two subband profiles and corresponding unknonws. The CTP− for subband n=1 represented by a filled square symbol is connected with other unknowns by arrows at a distance $\hbar\omega_{q_x} = 2\Delta E$.

through scattering, the BTE is solved only outside critical region, and, for charge and current calculations it is assumed that the occupation functions inside the critical region are equal, section by section, to the occupation of the states at the energy bin $j+1$ where j is the largest energy bin of the carriers inside the critical region.

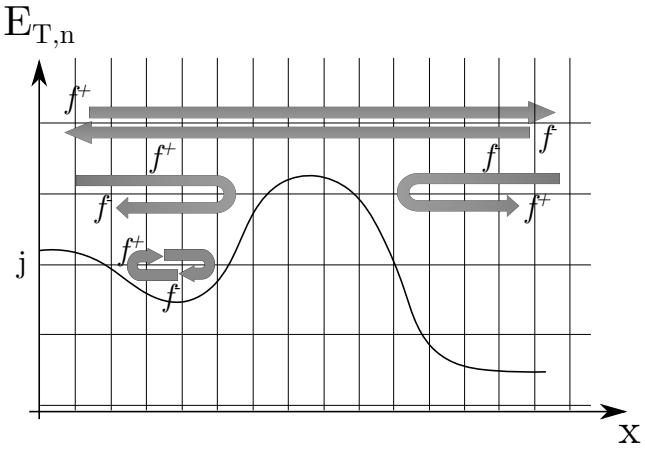
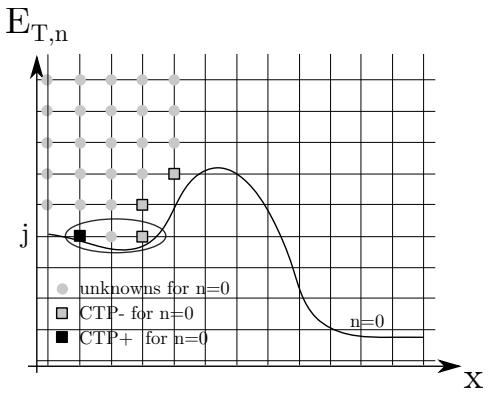
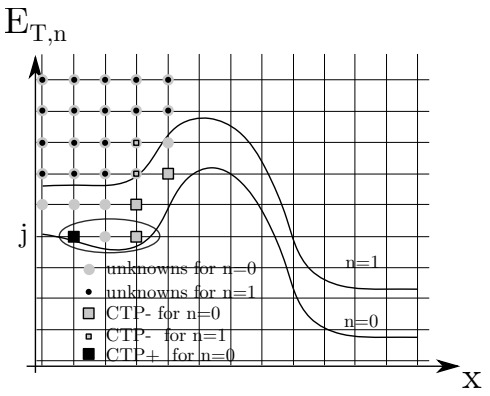


Figure 5.14: Subband profile with different trajectories that exhibits a critical region with two CTPs for the trajectory at energy bin j.

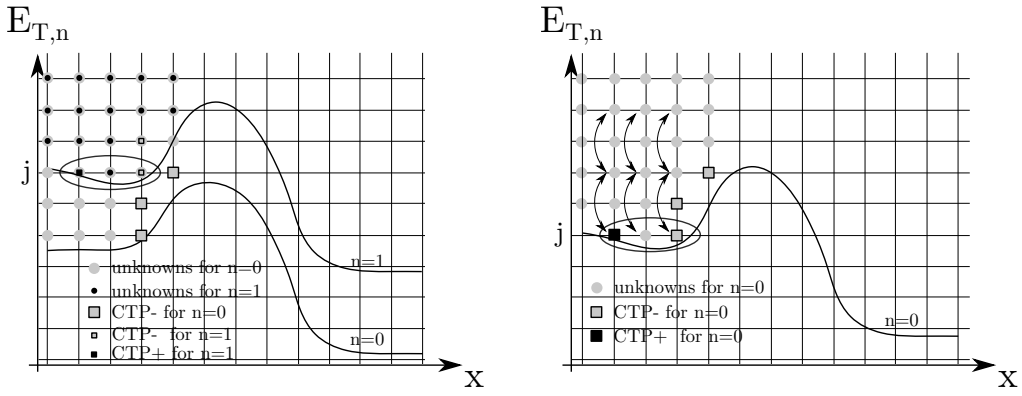


(a) Subband profile with a critical region at energy bin j.



(b) Subband profiles with a critical region for carriers belonging to subband $n = 0$ not connected with carriers belonging to subband $n = 1$ in the case of elastic scattering only.

Figure 5.15



(a) Subband profiles with a critical region for the subband $n = 1$ circled with solid line

(b) Subband profile with a critical region for carriers belonging to subband $n = 0$ connected by inelastic scattering (arrows) with carrier at larger energies.

Figure 5.16: Two cases of elastic and inelastic scattering where the presence of critical regions do not jeopardize the solution of the BTE because scattering connect carriers inside critical regions with the source and drain boundaries through the scattering with other carriers.

5.3.10 Singularity of the BTE r.h.s. for $E_{kinetic}$ approaching 0 eV

According to the expression for the scattering integral given in Eqs.5.22,5.23, the term $1/v^{\pm}_{g,n}(E_{T,n})$ diverges for total energy $E_{T,n}$ approaching the bottom of the $n-$th subband (i.e. $E_{T,n} \to \varepsilon_n$), or, equivalently, when the kinetic energy of a carrier approaches to zero. This is a peculiarity of 1D gas, and can lead to numerical problems with occupation functions out of boundaries $[0, 1]$ when solving the BTE equation because the scattering rates in the *r.h.s.* of the BTE for energies approaching the bottom of the subband can be very large. The way around to singularity of scattering rates adopted in this work is to solve only for the unknowns having a kinetic energy larger than a minimum value $E_{k,min}$.

As stated above, in a 1D electron gas the DoS (or equivalently the group velocity) tends to diverge when approaching the bottom of the subband, hence, the calculation of the current and charge must be performed carefully. Fig.5.17 shows a subband

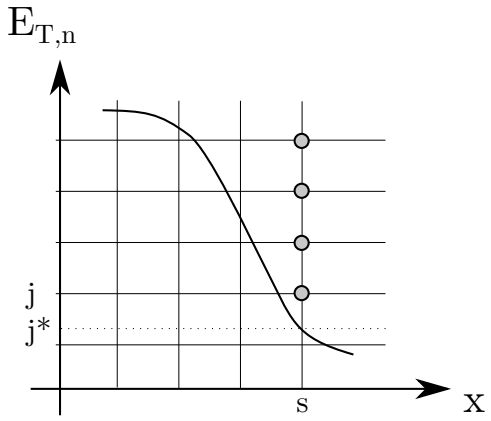


Figure 5.17: Subband profile with energy and space discretization.

profile with the discretization of the transport direction and total energy. If we focus on section s, the contribution to the current or charge given by the occupation function between the energy $j*$ and the first energy bin j cannot be neglected due to the divergence of the DoS approaching the subband profile. To account for this contribution in the charge but also in the current calculations, we assume that $f^{\pm}_{s,j*} = f^{\pm}_{s,j}$ which of course is an approximation given by the fact that the BTE is not solved for the energy $E_{T,n}(j*)$ since it does not correspond an energy bin. Hence, according to Eq.5.30 the contribution to the charge density per unit length calculated at section s, for a given subband n and for $f^{\pm}_{s,n,j}$ is given by:

$$\int_{\varepsilon_n}^{\infty} \frac{1}{\hbar v^{+}_{g,s,n}(E_{T,n})} f^{\pm}_{s,n}(E_{T,n}) dE_{T,n}$$
$$= f^{\pm}_{s,n}(E_{T,n}(j*)) \int_{E_{T,n}(j*)=\varepsilon_n}^{E_{T,n}(j)} \frac{1}{\hbar v^{+}_{g,s,n}(E_{T,n})} dE_{T,n} + \int_{E_{T,n}(j)}^{\infty} \frac{1}{\hbar v^{+}_{g,s,n}(E_{T,n})} f^{\pm}_{s,n}(E_{T,n}) dE_{T,n} \tag{5.61}$$

The integral in the first term of the summation in the right-hand side of Eq.5.61, can be rewritten using the expression for the group velocity reported in the first equation in Eq.5.3 and can be easily demonstrated to be equal to $k_x\left(E_{T,n}(j)\right)$. Hence, the charge at section s without accounting for the valley multiplicity and by considering the case of single valley, reads as:

$$n_{inv,s} = \frac{1}{\hbar\pi}\sum_{n}\sum_{t=\pm}\left[f^{t}_{s,n}(E_{T,n}(j*))\,k_x\left(E_{T,n}(j)\right) + \int_{E_{T,n}(j)}^{\infty}\frac{f^{t}_{s,n}(E_{T,n})}{v^{+}_{g,s,n}(E_{T,n})}dE_{T,n}\right]. \tag{5.62}$$

Similar calculations can be done for the current leading to:

$$I_s = \frac{e}{\hbar\pi}\sum_{n}\sum_{t=\pm} t\left[\left(f^{t}_{s,n}(E_T(j)\right)\cdot\left(E_T(j) - E_T(j*)\right) + \int_{E_T(j)}^{\infty} f^{t}_{s,n}(E_T(j))dE_T\right] \tag{5.63}$$

5.3.11 BTE results with template subband profile

In this section we will show some results obtained with the template subband profile shown in Fig.5.18 including different scattering mechanisms maintaining the same subband profile. V_{DS} is set to 0.5 V for all the simulations. The source Fermi level $E_{F,source}$ at $x = 0$ nm is set to 0 eV whereas the drain Fermi level at x=65 nm is $-eV_{DS} = -0.5$ eV. Simulations are performed including two subbands.

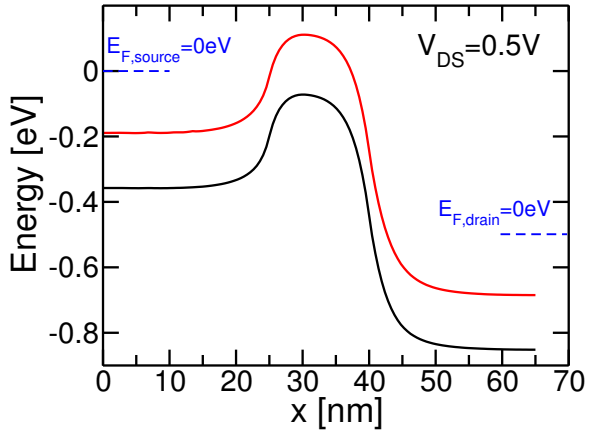


Figure 5.18: Template subbands profile. V_{DS} is set to 0.5 V and the source and drain Fermi levels are referenced therein.

Ballistic Results

Fig.5.19 shows the occupation function for electrons with both positive and negative group velocity for the two subbands. As shown in Fig.5.19(a) since the top-of-the-barrier (ToB) for the first subband is about 100 meV lower than the source Fermi level ($E_{F,source}$), which is set to 0 eV, there is a non negligible flux of electrons injected from the source contact with positive k_x (i.e. positive group velocity) above the ToB that reach the drain contact preserving their crystal momentum ($\hbar\ k_x$) because of the absence of scattering. Instead, being the ToB for the second subband about 100 meV higher than $E_{F,source}$, as can be seen in Fig.5.19(c) electrons injected from the source contact are mainly reflected at the CTPs leading to electrons with negative velocity below the ToB in the source region (see Fig.5.19(d)). The same holds for electrons with positive group velocity at the right of the ToB in Figs.5.19(b) and 5.19(d), that correspond to electrons injected from the drain contact. The ToB for both subbands is well above the drain Fermi level $E_{F,drain}$, hence the number of injected electrons from the drain contact that reach the source contact is very low.

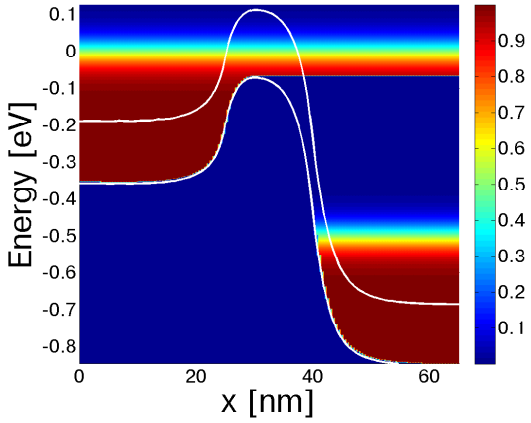
Moreover, due to the absence of scattering, there is no exchange of electrons belonging to different subbands. The current spectrum is defined as:

$$f^{+}_{s,n,j,} - f^{-}_{s,n,j,} \tag{5.64}$$

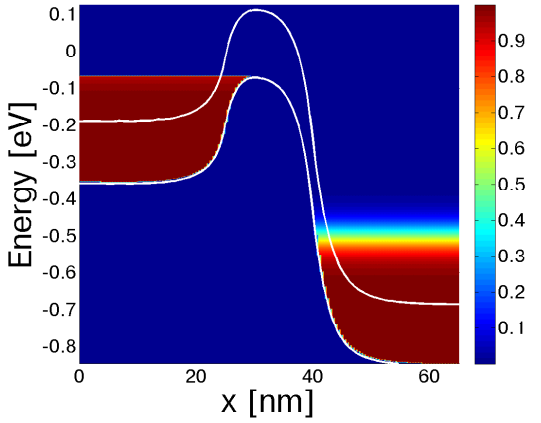
hence the current according to Eq.5.63 is given by the integral over the energy of the current spectrum for a given subband (and valley) times the factor $e/(\pi\hbar)$ and then summed over all the subbands (and valleys). Fig.5.20 shows the current spectrum for the two subbands and the current contribution is given only by electrons above the ToB due to the presence of CTP and absence of scattering which leads, for energy below the ToB, to a flux of carriers with positive velocity that equals that with negative velocity and thus to a null contribution to the current. Moreover, due to the position of the ToB for the second subband, only carriers in the lowest subband contribute to the total current as shown in Fig.5.21. Fig.5.22 shows the back-scattering coefficient for each section calculated as:

$$I_s^- / I_s^+ \tag{5.65}$$

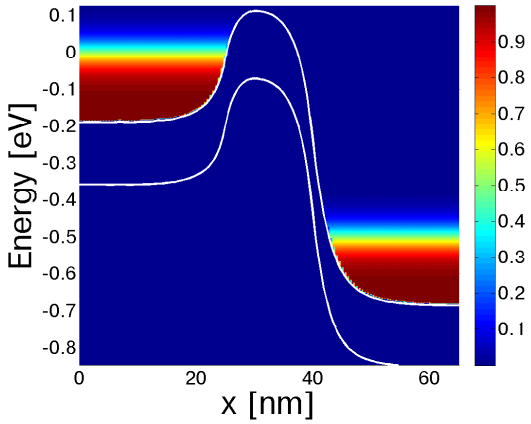
where I_s^+ can be calculated by using the expression in Eq.5.63 and summing over the states $t = +$, whereas I_s^- corresponds to the terms $t = -$. As can be seen in Fig.5.22, close to the ToB the backscattering becomes order of magnitudes lower than the value at the source or drain (which is close to 0.87), because of the combination of the applied V_{DS} and position of the ToB which lead to a negligible flux of carriers with negative group velocity above the ToB.



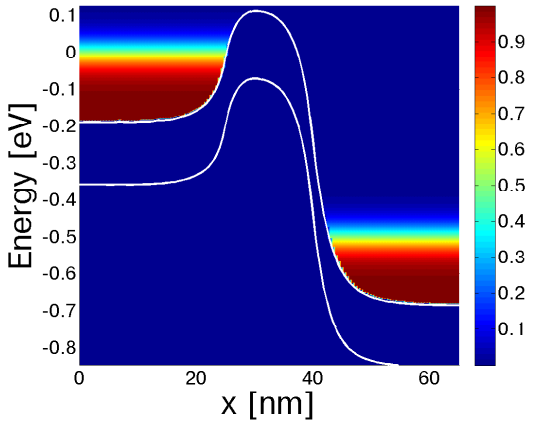
(a) Occupation function of electrons belonging to the first subband with positive group velocity



(b) Occupation function of electrons belonging to the first subband with negative group velocity

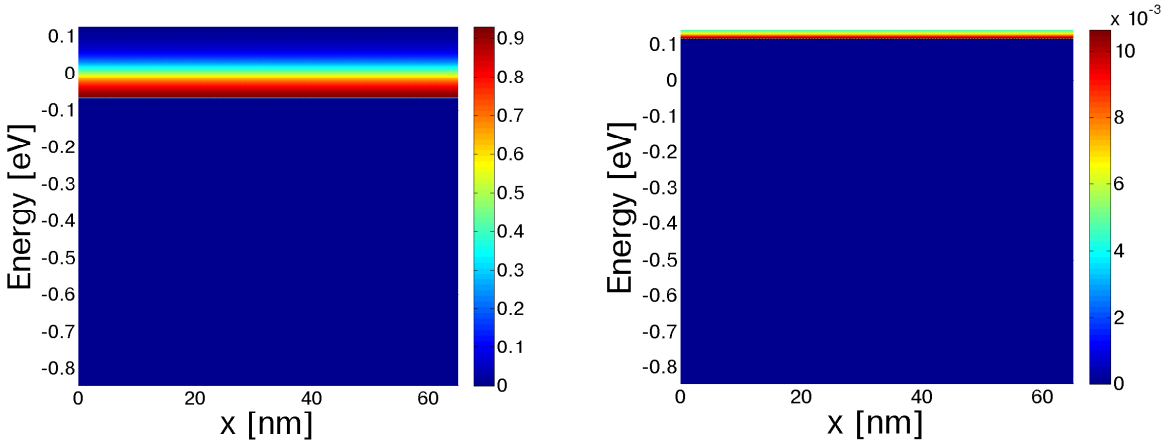


(c) Occupation function of electrons belonging to the second subband with positive group velocity



(d) Occupation function of electrons belonging to the second subband with negative group velocity

Figure 5.19: Occupation functions of ballistic simulation including two subbands. V_{DS} is 0.5V, $E_{F,source} = 0$ eV and $E_{F,drain} = -0.5$ eV



(a) Current spectrum for the first subband

(b) Current spectrum for the second subband

Figure 5.20: Current spectrum for ballistic simulation with two subbands.

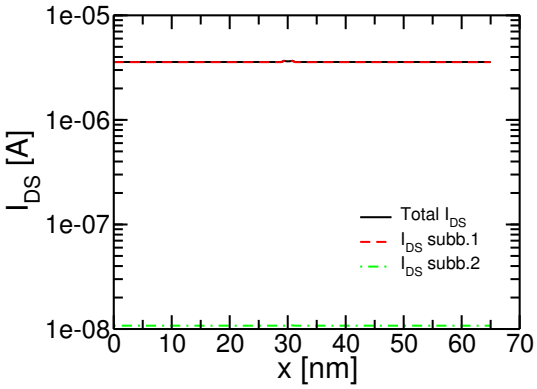


Figure 5.21: Current versus transport direction with contrinution of each subband. Ballistic simulation results.

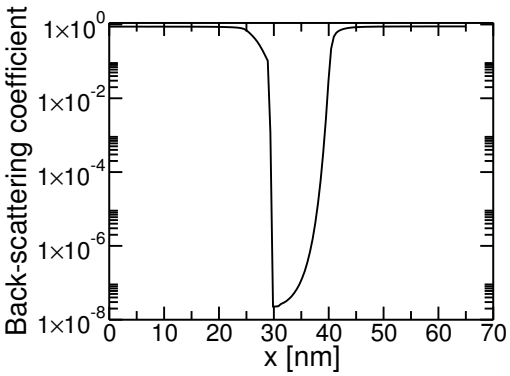
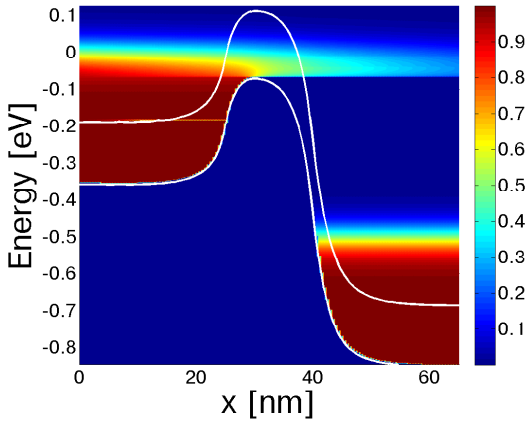
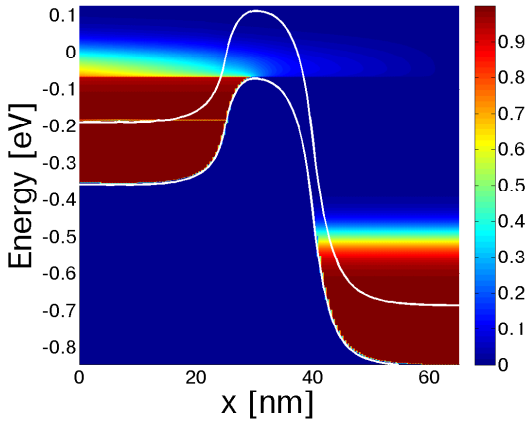


Figure 5.22: Back-scattering coefficient for a ballistic simulation.

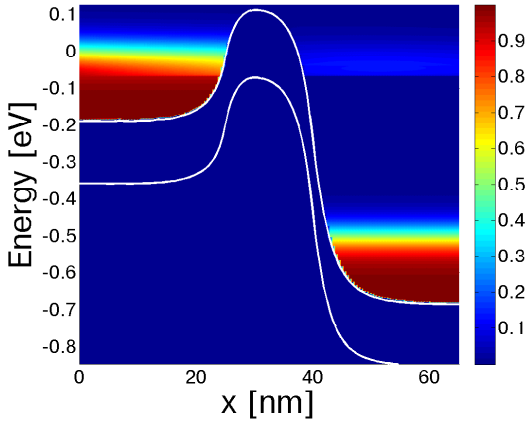
Elastic Scattering



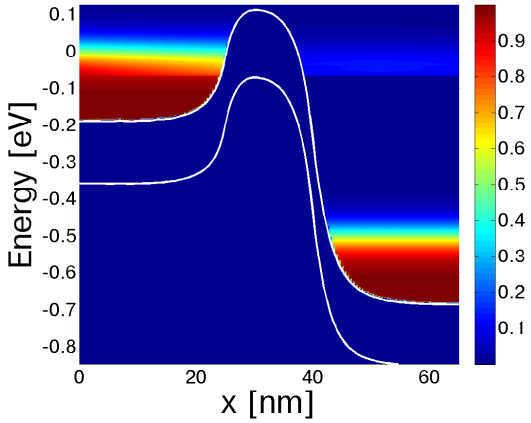
(a) Occupation function of electrons belonging to the first subband with positive group velocity



(b) Occupation function of electrons belonging to the first subband with negative group velocity



(c) Occupation function of electrons belonging to the second subband with positive group velocity



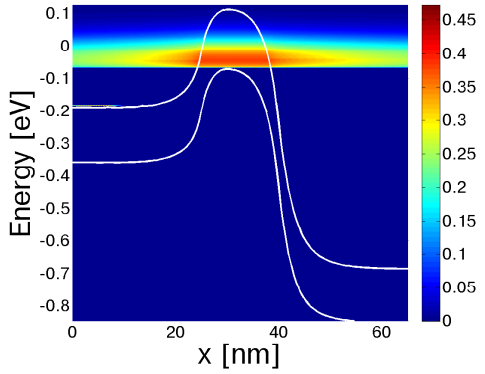
(d) Occupation function of electrons belonging to the second subband with negative group velocity

Figure 5.23: Occupation functions of a simulation including elastic scattering with two subbands. V_{DS} is 0.5V, $E_{F,source} = 0$ eV and $E_{F,drain} = -0.5$ eV

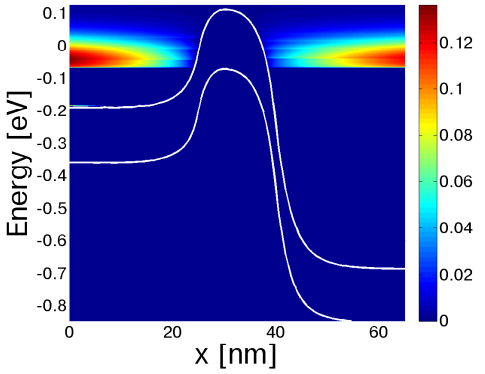
Fig.5.23 shows the occupation functions for two subbands with elastic scattering. The occupation function for electrons belonging to the first subband and with positive group velocity is shown in Fig.5.23(a). As opposed to Fig.5.19 it can be seen that electrons above the ToB of the first subband due to the elastic scattering undergo two different effects. The first is that they are backscattered to the source still remaining in the first subband: in fact Fig.5.23(b) shows that the occupation function for electrons belonging to the first subband with negative group velocity is not zero above the ToB of the first subband. Hence, there is a lowering of the net flux of

electrons from source to drain (i.e. the electron current). The second effect is that elastic scattering can also move electrons from one subband to another: in fact Figs.5.23(c) and 5.23(d) show that on the right of the ToB and for energies larger than approximately −100 meV there is a non negligible flux of electrons either with positive or negative group velocity due to the scattered electrons belonging to the first subband.

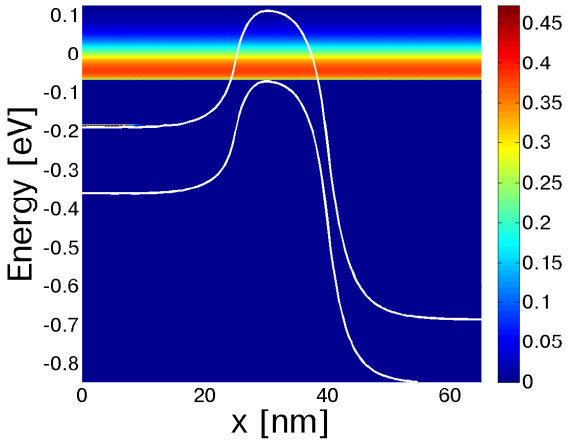
Finally, Fig.5.24 shows the current spectrum calculated as in Eq.5.64. Fig.5.24(a)



(a) Current spectrum for the first subband



(b) Current spectrum for the second subband



(c) Total Current spectrum

Figure 5.24: Current spectrum for simulation with elastic scattering and with two subbands.

shows that the current contribution is non-null only above the ToB for the first subband. This is because, as in the case of ballistic simulation, carriers below the ToB that are injected from the source are reflected at the corresponding CTPs and the contribution to the total current cannot but be zero. In Fig.5.24(b) it can be shown that, due to the intersubband scattering, there is a non negligible contribution to the

total current from carriers belonging to the second subband below the corresponding ToB, differently from Fig.5.20(b) where electrons are mainly confined in the source and drain regions and their contribution to the current is close to zero. The current spectrum summed over the two subbands is shown in Fig.5.24(c) and the integral over the energy gives the current shown in Fig.5.25. As expected, the total current is constant along the transport direction whereas the contribution of each subband changes along x. Fig.5.26 shows the back-scattering coefficient calculated as in

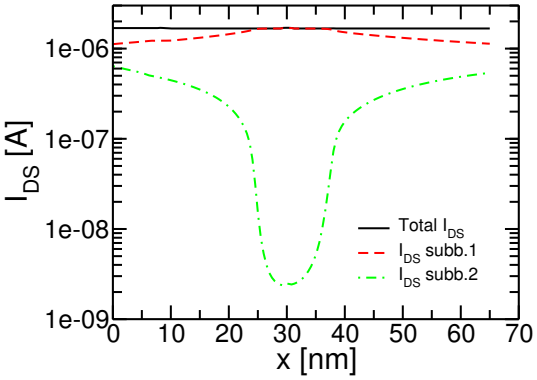


Figure 5.25: Current versus transport direction with contrinution of each subband. Simulation with elastic scattering.

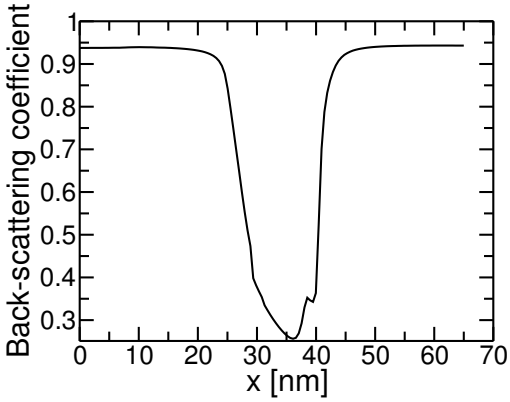
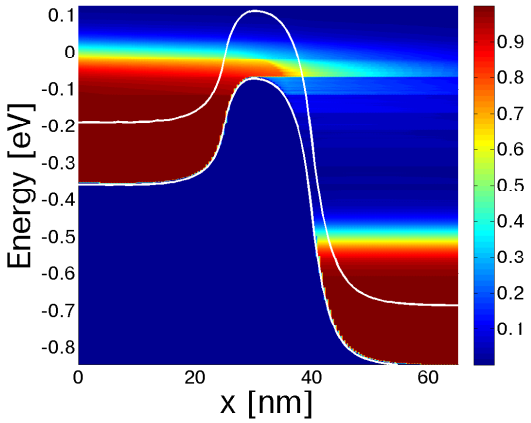


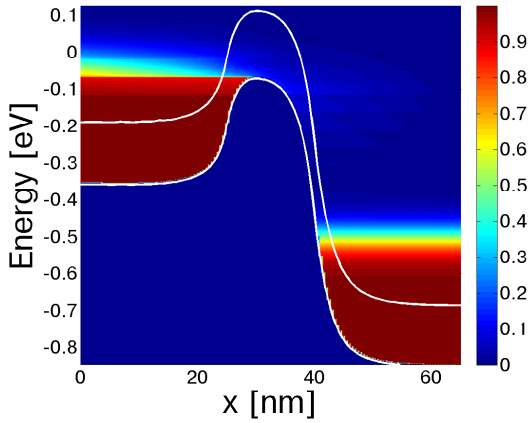
Figure 5.26: Back-scattering coefficient for a simulation with elastic scattering.

Eq.5.65. Differently from the back-scattering for the ballistic case in Fig.5.22, since scattering can change the sign of group velocity at any energy electrons with negative group velocity are observed even for energies well above the drain Fermi level.

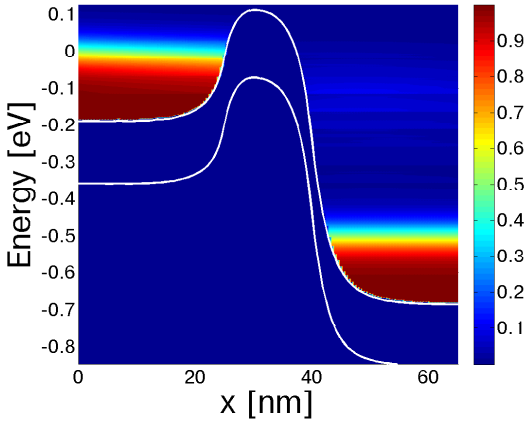
Inelastic Scattering



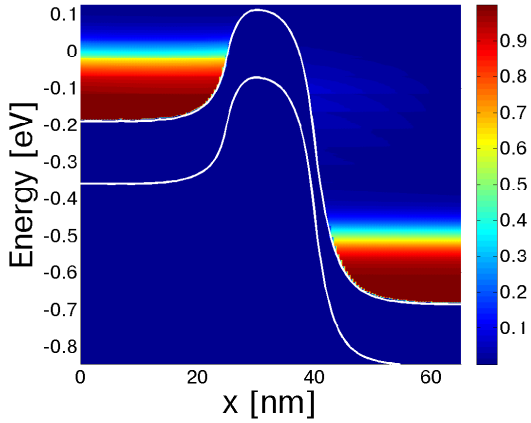
(a) Occupation function of electrons belonging to the first subband with positive group velocity



(b) Occupation function of electrons belonging to the first subband with negative group velocity



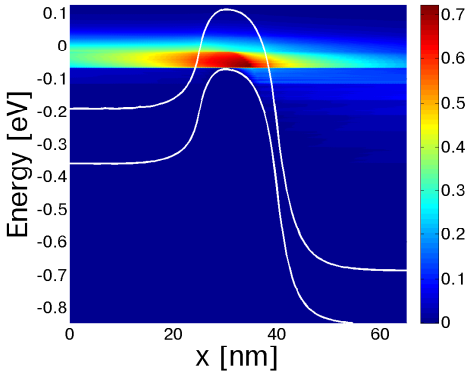
(c) Occupation function of electrons belonging to the second subband with positive group velocity



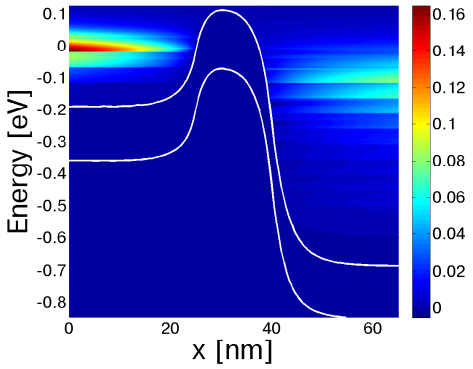
(d) Occupation function of electrons belonging to the second subband with negative group velocity

Figure 5.27: Occupation functions of a simulation with inelastic scattering with optical phonons with $\hbar\omega_{q_x} = 50meV$ and two subbands. V_{DS} is 0.5V, $E_{F,source} = 0$ eV and $E_{F,drain} = -0.5$ eV

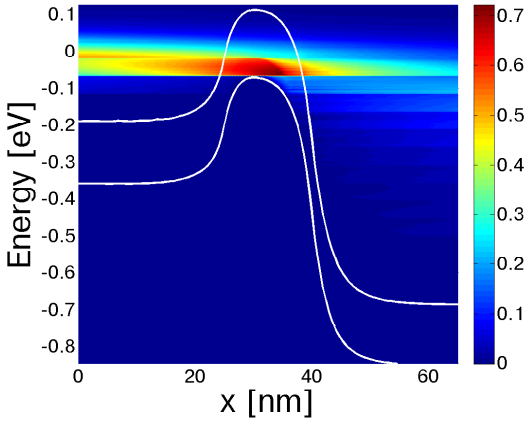
Fig.5.27 shows that inelastic scattering for both subbands fills states with electrons with positive and negative velocities for the energy window below the corresponding ToBs and few tens of meV above $E_{F,drain}$ and at the right of the ToBs, that is, an energy in a region that is forbidden in the case of ballistic or elastic scattering simulations. Current spectrum is shown in Fig.5.28 where it can be seen that there is a non-negligible contribution to the total current given by electrons that lose their kinetic energy via optical phonons emissions while moving from the ToB to the drain



(a) Current spectrum for the first subband



(b) Current spectrum for the second subband



(c) Total current spectrum

Figure 5.28: Current spectrum for simulation with inelastic scattering and with two subbands.

region. Even though the current spectrum in Fig.5.28(c) changes section by section, its integral over the energies and summed over the subbands gives a current which is constant throughout the device sections as shown in Fig.5.29.

Fig.5.30 shows the back-scattering coefficient.

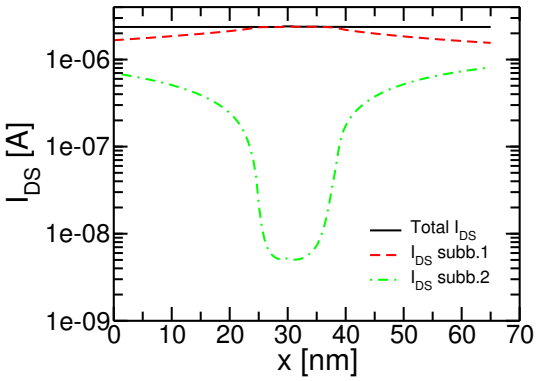


Figure 5.29: Current versus transport direction with contribution of each subband. Simulation with inelastic scattering.

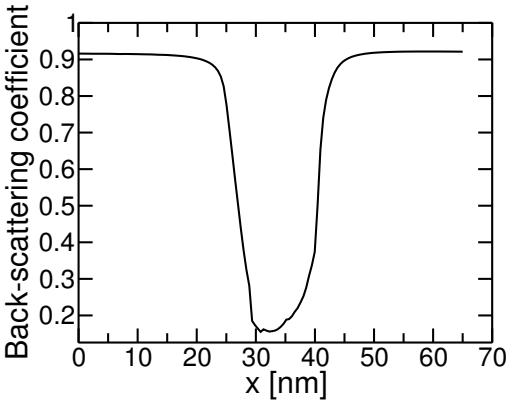


Figure 5.30: Back-scattering coefficient for a simulation with inelastic scattering.

5.4 Solver for transport calculations

5.4.1 Device discretization

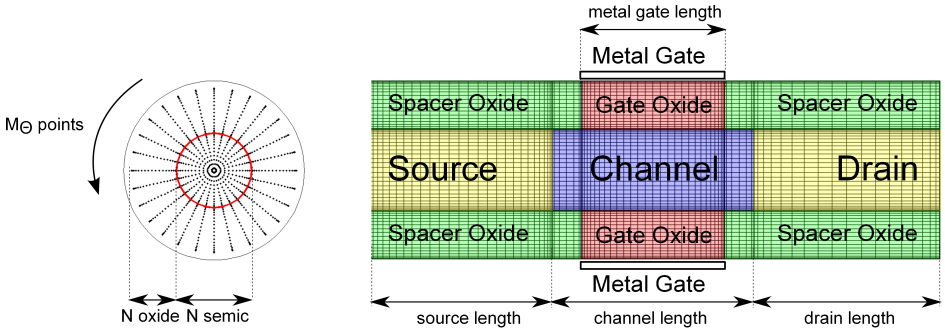


Figure 5.31: left: slice along the quantization plane of the circular nanowire with discretization points. right: slice along the transport direction.

The device circular sections are discretized as explained in Sec.4.3.9 using a non periodic Chebyshev grid along radial direction and a uniform periodic Fourier grid along angular direction in order to employ the PS method that allows for an high accuracy (i.e. the so-called spectral accuracy) and a fast numerical solution of the Schrödinger equation due to the reduced number of points with respect to traditional FD of FE methods. A Chebyshev grid is used also along transport direction in order to obtain spectral accuracy for the solution of the Poisson equation. A sketch of the device discretization is shown in Fig.5.31.

We solve the Schrödinger equation Eq.4.1 along circular slices to obtain the subband profile and the wave-function. We impose homogeneous Dirichlet boundary conditions at the boundaries of the circular slices and the continuity conditions of Eq.4.10 at the discontinuity surface between oxide and semiconductor.

The Poisson equation is solved in the 3D cylindrical domain by using the same methodology explained in Sec.4.5 but extending Eq.4.98 to a 3D domain. We now focus on the boundary conditions for the Poisson equation which have been already discussed for the metal gate contact region in Sec.4.5 and we extend it to the 3D nanowire. In order to maintain an average space-charge neutrality in the source-drain regions close to the source and drain contacts, we use homogeneous Neumann conditions at the idealized source and drain contacts along transport direction:

$$\frac{\partial \phi(r,\theta,x)}{\partial x} = 0. \tag{5.66}$$

Same Neumann conditions are also imposed along radial direction in the nanowire external surface that is not surrounded by the metal gate (see Fig.5.31); at the oxide to gate electrode interface, Dirichlet conditions are used to set the potential.

5.4.2 Iterative solution for transport calculation

The overall flowchart for the solver is shown in Fig.5.32.

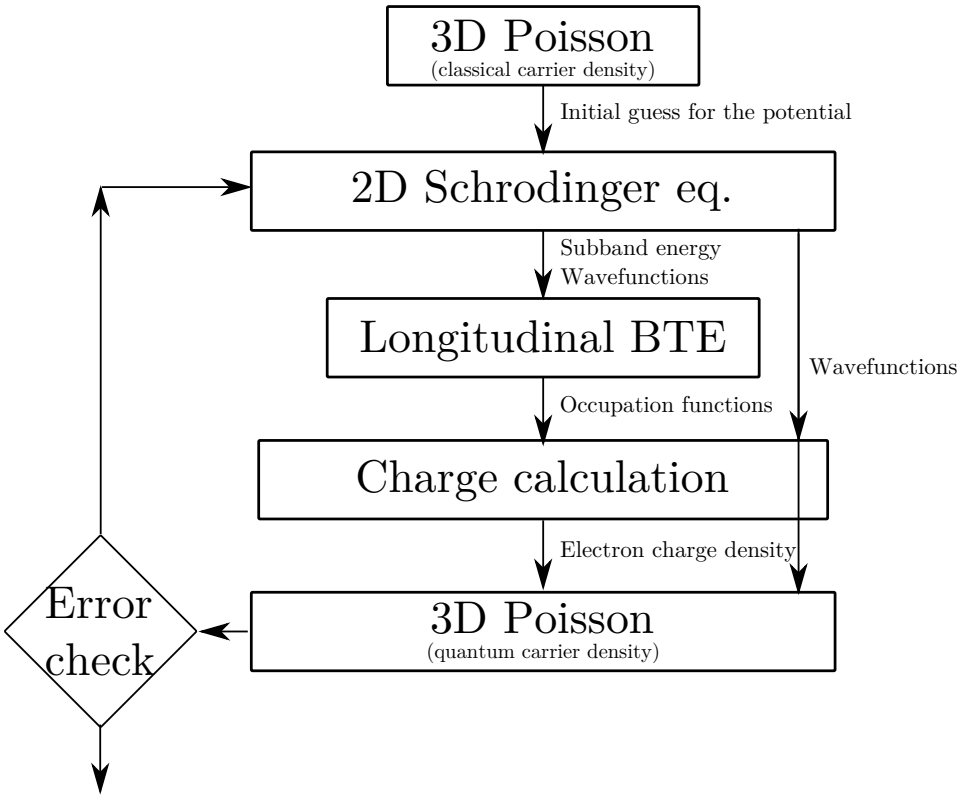


Figure 5.32: Flowchart of the solution scheme.

At the first iteration the solver needs an initial condition to start. This is provided by the potential at the equilibrium ($V_{DS} = 0V$) calculated by solving the 3D Poisson equation for a 3D gas using the Fermi-Dirac statistics (quantization is not taken into account in this preliminary step), and accounting for non-parabolicity corrections that have a vast effect on the resulting degeneracy of the electron gas; the expression for the charge is given in Eq.C.26.

To work out an initial guess for the potential in out-of-equilibrium conditions (i.e. $V_{DS} \neq 0$), the first guess for the potential obtained for $V_{DS} = 0$ V is linearly shifted starting from the source-channel interface to the channel-drain interface (see Fig.5.33). Then, the Schrödinger equation is solved in each circular section of the device in order to obtain the subband profile of the entire device.

The BTE equation is solved for out-of-equilibrium conditions according to Sec.5.3 with:

- boundary conditions for f^+ at source according to Fermi-Dirac occupation with $E_{F,S}$=0;

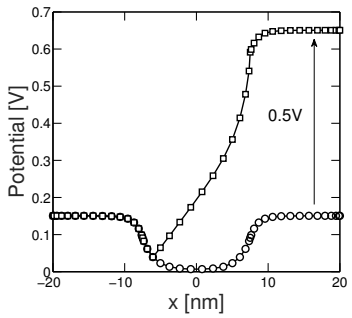


Figure 5.33: Plot of the potential at the center of the Nanowire. squares: potential used as an initial guess to introduce non equilibrium conditions being V_{DS} =0.5 V. It is initial guess for Schrödinger and BTE solvers. circles : potential profile at equilibrium (i.e. V_{DS} = 0V).

- boundary conditions for f^- at drain according to Fermi-Dirac occupation with $E_{F,S} = -eV_{DS}$

Then, the Poisson equation reported in Eq.4.98 but extended for a 3D domain introducing the derivative along the transport direction, is solved to find a new guess for the electrostatics potential and the loop in Fig.5.32 is repeated until convergence is reached.

5.5 BTE Simulation results

The device has a silicon channel with diameter d = 5 nm, the gate length is L_G= 15 nm, and the gate dielectric is HfO_2 ($\kappa = 22$ [2]), with a thickness of 4 nm, gate workfunction is 4.6 eV, source and drain regions with $L_{S,D} = 20$ nm, and donor doping of 1×10^{20} cm^{-3}. The transport direction is [100]. Coulomb scattering is activated only in the channel region and not in source and drain regions, this is also because, for very large doping concentrations (or scattering rates), the scattering model should take into account the correlation between Coulomb centers (see also Sec.4.6.2), that is not considered in the present model and in most of the papers in the literature, as well as for corrections beyond first-order Born approximation [18].

Fig.5.34 shows the results obtained for different scattering conditions: a) without scattering; b) including acoustic and optical phonon scattering according to Sec.4.6.1 with deformation potentials from [2]; c) with the inclusion of Coulomb scattering accounting for interface states with $N_{IT} = 1 \times 10^{12}$ cm^{-2} according to Sec.4.6.2; d) including surface roughness with the new model described in Sec.3.6 and using root mean squared roughness amplitude $\Delta_{rms} = 0.21$ nm and correlation length $\Lambda = 1.3$ nm.

Results show the strong reduction of the current in the presence of phonons, and also surface roughness implemented with the new model of Chapter 3. In particular,

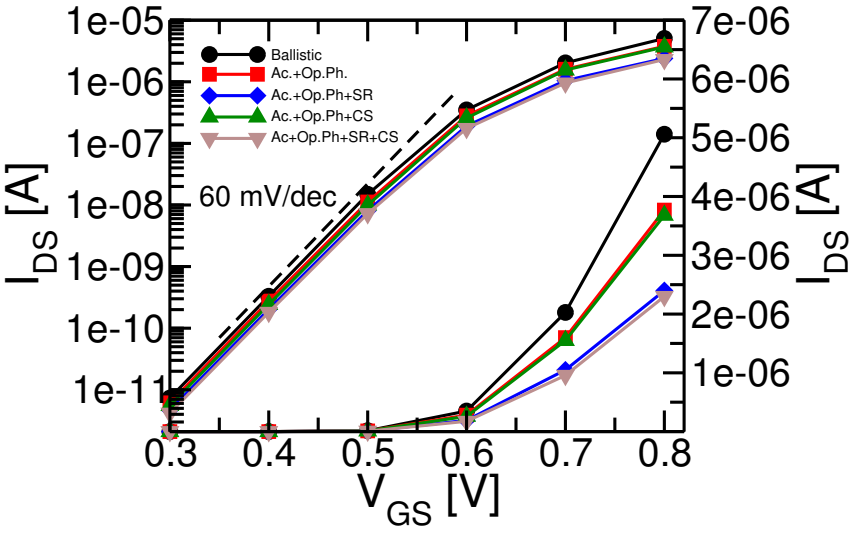


Figure 5.34: Drain current versus gate voltage characteristics for silicon with [100] transport direction, for V_{DS}= 0.5 V and with the inclusion of different scattering mechanisms. The density of interface states for the coulomb scattering is $N_{IT} = 1 \times 10^{12}$ cm^{-2}.

the SR scattering is responsible for a strong performance degradation as already pointed out for planar devices in [19].

5.6 Conclusions

The development of a comprehensive simulator based on the deterministic solution of the BTE for circular nanowires has required lot of time, therefore, we are still validating this new tool against available experimental data. We believe that the inclusion of the surface roughness scattering with the new non linear formulation already validated against available experimental results for planar devices is one of the highlights of this work. In fact it has been already demonstrated that for aggressively scaled nanowire, SR is one of the most important scattering mechanisms [20, 21]. Moreover, the deterministic solution of the BTE allows for the study of the sub-threshold regime without being affected by the statistical noise that is present in MonteCarlo solvers. This, in principle, allows for the simulation of low field mobility by simply applying a very low longitudinal field along transport direction still maintaining the precision of the solution.

References

[1] L. Mark, *Fundamentals of carrier transport*, 2st. Cambridge University Press., 2000.

[2] D. Esseni, P. Palestri, and L. Selmi, *Nanoscale MOS Transistors - Semi-Classical Transport and Applications*, 1st. Cambridge University Press., 2011.

[3] 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, pp. 645–705, 3 Jul. 1983.

[4] M. V. Fischetti and S. E. Laux, "Monte carlo analysis of electron transport in small semiconductor devices including band-structure and space-charge effects", vol. 38, pp. 9721–9745, 14 Nov. 1988.

[5] 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 Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 8, no. 4, pp. 360–369, Apr. 1989.

[6] J. Bude and R. K. Smith, "Phase-space simplex Monte Carlo for semiconductor transport", *Semiconductor Science Technology*, vol. 9, no. 5S, p. 840, 1994.

[7] C. Jungemann, S. Keith, M. Bartels, and B. Meinerzhagen, "Efficient full-band Monte Carlo simulation of silicon devices", *IEICE Transactions on Electronics*, vol. E82-C, pp. 870–879, 1999.

[8] F. M. Bufler, A. Schenk, and W. Fichtner, "Efficient Monte Carlo device modeling", *IEEE Trans. on Electron Devices*, vol. 47, no. 10, pp. 1891–1897, Oct. 2000.

[9] B. Winstead and U. Ravaioli, "A quantum correction based on schrodinger equation applied to monte carlo device simulation", *IEEE Trans. on Electron Devices*, vol. 50, no. 2, pp. 440–446, Feb. 2003.

[10] M. A. Elmessary, D. Nagy, M. Aldegunde, J. Lindberg, W. Dettmer, D. Peric, A. J. Garcia-Loureiro, and K. Kalna, "Anisotropic schrodinger equation quantum corrections for 3D Monte Carlo simulations of nanoscale multigate transistors", in *International Workshop on Computational Electronics*, Sep. 2015, pp. 1–4.

[11] 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 *IEEE IEDM Technical Digest*, 2005, pp. 617–620.

[12] P. Osgnach, A. Revelant, D. Lizzit, P. Palestri, D. Esseni, and L. Selmi, "Toward computationally efficient Multi-Subband Monte Carlo Simulations of Nanoscale MOSFETs", in *Proc.SISPAD*, 2013, pp. 176–179.

[13] L. Donetti, C. Sampedro, F. Gamiz, A. Godoy, F. J. Garcia-Ruiz, E. Towiez, V. P. Georgiev, S. M. Amoroso, C. Riddet, and A. Asenov, "Multi-Subband Ensemble Monte Carlo simulation of Si nanowire MOSFETs", in *Proc.SISPAD*, Sep. 2015, pp. 353–356.

[14] M. Lenzi, P. Palestri, E. Gnani, S. Reggiani, A. Gnudi, D. Esseni, L. Selmi, and G. Baccarani, "Investigation of the Transport Properties of Silicon Nanowires Using Deterministic and Monte Carlo Approaches to the Solution of the Boltzmann Transport Equation", *IEEE Trans. on Electron Devices*, vol. 55, no. 8, pp. 2086–2096, Aug. 2008.

[15] S. Jin, S.-M. Hong, W. Choi, K.-H. Lee, and Y. Park, "Coupled drift-diffusion (DD) and multi-subband Boltzmann transport equation (MSBTE) solver for 3D multi-gate transistors", in *Proc.SISPAD*, Sep. 2013, pp. 348–351.

[16] E. G. Marin, F. G. Ruiz, A. Godoy, I. M. Tienda-Luna, C. Martinez-Blanque, and F. Gamiz, "Theoretical interpretation of the electron mobility behavior in InAs nanowires", *Journal of Applied Physics*, vol. 116, no. 17, 2014.

[17] H. U. Baranger and J. W. Wilkins, "Ballistic structure in the electron distribution function of small semiconducting structures: General features and specific trends", vol. 36, pp. 1487–1502, 3 Jul. 1987.

[18] D. Lizzit, D. Esseni, P. Palestri, and L. Selmi, "A new formulation for surface roughness limited mobility in bulk and ultra-thin-body metal-oxide-semiconductor transistors", *Journal of Applied Physics*, vol. 116, no. 22, 223702, 2014.

[19] D. Lizzit, D. Esseni, P. Palestri, P. Osgnach, and L. Selmi, "Performance Benchmarking and Effective Channel Length for Nanoscale InAs, $In_{0.53}Ga_{0.47}As$, and sSi n-MOSFETs", *IEEE Trans. on Electron Devices*, vol. 61, no. 99, pp. 2027 –2034, Jun. 2014.

[20] S. D. Suk et al., "Investigation of nanowire size dependency on TSNWFET", in *IEEE IEDM Technical Digest*, Dec. 2007, pp. 891–894.

[21] S. Kim, M. Luisier, A. Paul, T. Boykin, and G. Klimeck, "Full Three-Dimensional Quantum Transport Simulation of Atomistic Interface Roughness in Silicon Nanowire FETs", *IEEE Trans. on Electron Devices*, vol. 58, no. 5, pp. 1371–1380, May 2011.

Appendix A

Tensor of the inverse effective masses for different orientations

THE knowledge of the band structure provides information that goes beyond the energy values of the valley minima and the energy gap. Indeed, the energy band curvature in **K** space, can provide the effective mass of the carriers. Effective mass can be intuitively interpreted as the mass of a carrier within a crystal and it governs the response of the carrier to an external force. For transport applications the effective masses are particularly important because they affect the velocity of carriers along the transport, the quantization of energy levels and the density of states. An electric field or a magnetic field affect the state of a carrier as if it was a free particle, but having a different mass with respect to the mass in vacuum, which is the effective mass. The electron is described by a wave-packet which has group velocity in three-dimensions given by:

$$\mathbf{v}_g = \frac{1}{\hbar}\nabla_{\mathbf{K}} E\left(\mathbf{K}\right) \tag{A.1}$$

where $\nabla_{\mathbf{K}}$ is the gradient as a function of the wave-vector **K**. Denoting **a** the acceleration due to an applied force **F** we can write:

$$\mathbf{a} = \frac{d\mathbf{v}_g}{dt} = \frac{d}{dt}\frac{\nabla_{\mathbf{K}} E\left(\mathbf{K}\right)}{\hbar} = \frac{\nabla_{\mathbf{K}}}{\hbar}\left(\frac{d\mathbf{K}}{dt}\nabla_{\mathbf{K}} E(\mathbf{K})\right). \tag{A.2}$$

Recalling that the rate of change of the crystal momentum **P** of a particle is proportional to the force **F** acting on it:

$$\mathbf{F} = \frac{d\mathbf{P}}{dt} = \hbar\frac{d\mathbf{K}}{dt}, \tag{A.3}$$

we can rewrite the Newton's second law for a classical particle under the influence of a force as:

$$\mathbf{a} = \frac{\nabla_{\mathbf{K}}}{\hbar^2}\left(\mathbf{F}\nabla_{\mathbf{K}} E(\mathbf{K})\right). \tag{A.4}$$

By using matrix formalism the column vector **a** is linked to the force column vector **F** through

$$\mathbf{a} = \mathbf{WF} \tag{A.5}$$

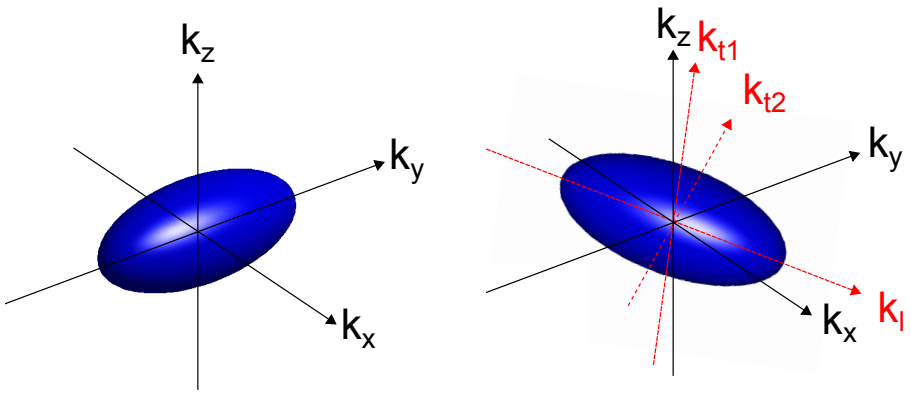
where **W** is the [3x3] tensor of the inverse of the effective masses and the (i, j) component is given by

$$\mathbf{W}_{i,j} = \frac{1}{\hbar}\frac{\partial^2 E(\mathbf{K})}{\partial k_i \partial k_j} \qquad i, j \in x, y, z \tag{A.6}$$

It follows that the tensor of the effective mass is the inverse of the **W** tensor

$$\mathbf{M}_{i,j} = (\mathbf{W})^{-1}_{i,j}\,. \tag{A.7}$$

We now consider the dispersion relation for electrons derived under the parabolic



(a) Aligned Ellipsoid (b) Non Aligned Ellipsoid

Figure A.1: Ellipsoid Coordinate System.

band approximation, that is expanding the energy dependence in Taylor series of second order in **K** around the band minimum. If we suppose that ellipsoidal equi-energy surfaces are aligned with the coordinate axis as in Fig.A.1a we write

$$E\left(\mathbf{K}\right) = E_{\nu,0} + \frac{\hbar^2 {k_x}^2}{2m_x} + \frac{\hbar^2 {k_y}^2}{2m_y} + \frac{\hbar^2 {k_z}^2}{2m_z} \tag{A.8}$$

where $E_{\nu,0}$ is the energy of the bottom of the conduction band, and k_x, k_y, k_z are the deviations with respect to the energy at the bottom of the valley (k_i are measured from the value of the wave-vector where the band has its relative minimum). Even more generally, if the ellipsoids are not aligned with the k_x, k_y, k_z-coordinate axis as shown in Fig.A.1b the dispersion relation for a 3D gas reads:

$$E\left(\mathbf{K}\right) = E_{\nu,0} + \frac{\hbar^2}{2}\sum_{i,j}\mathbf{W}_{i,j}k_i k_j \quad i, j \in k_x, k_y, k_z \tag{A.9}$$

where $\mathbf{W}$ is the tensor of the inverse effective masses defined in Eq.A.6 with respect to the k_x, k_y, k_z-coordinate system. It is now convenient to start introducing three different systems of coordinates as depicted in Fig.A.2: Device Coordinate System (DCS), Crystal Coordinate System (CCS) and Ellipsoid Coordinate System (ECS). The DCS is independent of the crystallographic orientation of the semiconductors atoms, but it depends exclusively on the transport and quantization directions. The second, CCS, allows to specify directions in relation to crystallographic axes and the last, the ECS, is a coordinate system oriented with the ellipsoidal equienergy surfaces. It is useful to distinguish between DCS, CCS and ECS because the orientation of the device can be decoupled from that of the crystal or that of the single valley to make it easy and intuitive the study of devices with different crystal orientations as depicted in Fig.A.2 In the general case Eq.A.9 is written using matrix formalism as:

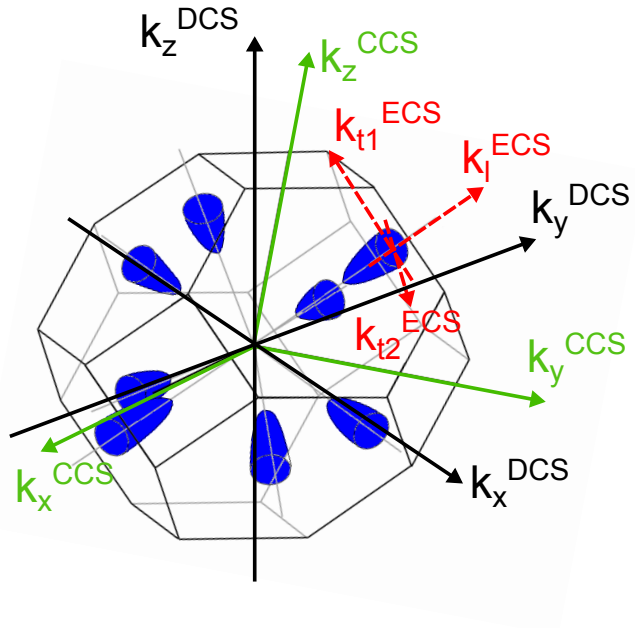


Figure A.2: Sketch of the Brillouin zone with ellipsoidal equienergy surfaces belonging to L-valleys and the corresponding coordinate-system.

$$E\left(\mathbf{K}\right) = E_{\nu,0} + \frac{\hbar^2}{2}(k_x, k_y, k_z)\mathbf{W}_{DCS}(k_x, k_y, k_z)^T \tag{A.10}$$

where $(k_x, k_y, k_z)^T$ indicates a column vector. The knowledge of the matrix $\mathbf{W}_{DCS}$:

$$\mathbf{W}_{DCS} = \begin{bmatrix} w_{xx} & w_{xy} & w_{xz} \\ w_{yx} & w_{yy} & w_{yz} \\ w_{zx} & w_{zy} & w_{zz} \end{bmatrix} \tag{A.11}$$

is thus of extremely importance to determine the energy dispersion relation along the transport and quantization directions. It is now useful to recall that in the ECS the

W tensor is always diagonal and, in particular, takes the form reported in Fig.A.3 and the diagonal coefficients are given by the inverse element-wise of the masses

$$\mathbf{W_{ECS}} = \begin{pmatrix} 1/m_{t1} & 0 & 0 \\ 0 & 1/m_{t2} & 0 \\ 0 & 0 & 1/m_l \end{pmatrix}$$

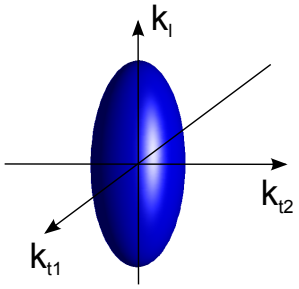


Figure A.3: Tensor of the inverse effective mass for the ECS and ellipsoid orientation.

along the two transverse ellipsoidal directions and along longitudinal axial direction. Effective masses in literature for the different semiconductor valleys are always given in the ECS-coordinate system, thus, it is convenient to relate the $\mathbf{W}_{DCS}$ to the $\mathbf{W}_{ECS}$. The link between the DCS- and the ECS-coordinate system is given by the rotation matrix by the following equation:

$$(k_{t1}, k_{t2}, k_l)^T = \mathbf{R}_{DCS\to ECS}(k_x, k_y, k_z)^T \tag{A.12}$$

where $\mathbf{R}_{DCS\to ECS}$ denotes the orthogonal rotation matrices that map a point defined in the k_x, k_y, k_z system into the k_{t1}, k_{t2}, k_l system. In such a way, it is straightforward demonstrate that:

$$\mathbf{W}_{DCS} = \mathbf{R}_{DCS\to ECS}\, \mathbf{W}_{ECS}\, \mathbf{R}_{DCS\to ECS} \tag{A.13}$$

and it can be demonstrated that $\mathbf{W_{DCS}}$ is always a symmetric matrix, and expressions for rotation matrices $\mathbf{R}_{DCS\to ECS}$ can be found in [1].

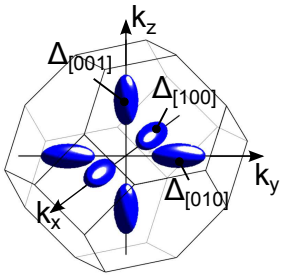
According to the definition of $\mathbf{W}_{DCS}$ in Eq.A.13, the effective mass corresponding to the coordinate $(i, j) \in (x, y, z)$ of the DCS, can be calculated from the tensor $\mathbf{W}_{DCS}$ as follows:

$$(\mathbf{M}_{DCS})_{i,j} = \frac{Cof((\mathbf{W}_{DCS})_{i,j})}{det\,(\mathbf{W}_{DCS})} \tag{A.14}$$

where $Cof((\mathbf{M}_{DCS})_{i,j})$ represents the cofactor (i, j) and the determinant can be easily calculated recalling that $\mathbf{W}_{ECS}$ is the diagonalization of the $\mathbf{W}_{DCS}$ tensor. Hence the determinant of $\mathbf{W}_{DCS}$ equals to the determinant of $\mathbf{W}_{ECS}$ being the determinant of the rotation matrices equal to 1. It follows that $det(\mathbf{W}_{ECS})$ is given by the product of the diagonal elements of $\mathbf{W}_{ECS}$:

$$det\,(\mathbf{W}_{DCS}) = det\,(\mathbf{W}_{ECS}) = \frac{1}{m_{t1} m_{t2} m_l} \tag{A.15}$$

Fig.A.4 shows an example of $\mathbf{W}_{DCS}$ tensors for silicon Δ-valleys labelled as $\Delta_{[100]}$,$\Delta_{[010]}$ and $\Delta_{[001]}$ according to the sketch reported in Fig.A.4 for different DCS orientations referred to the ECS.



x	y	z	valley	$\mathbf{W}_{DCS}$
[100]	[010]	[001]	$\Delta_{[100]}$	$\begin{bmatrix} 1/m_l & 0 & 0 \\ 0 & 1/m_t & 0 \\ 0 & 0 & 1/m_t \end{bmatrix}$
			$\Delta_{[010]}$	$\begin{bmatrix} 1/m_t & 0 & 0 \\ 0 & 1/m_l & 0 \\ 0 & 0 & 1/m_t \end{bmatrix}$
			$\Delta_{[001]}$	$\begin{bmatrix} 1/m_t & 0 & 0 \\ 0 & 1/m_t & 0 \\ 0 & 0 & 1/m_l \end{bmatrix}$
[110]	$[\bar{1}10]$	[001]	$\Delta_{[100]}$	$\begin{bmatrix} \frac{m_l+m_t}{2m_lm_t} & \frac{m_l-m_t}{2m_lm_t} & 0 \\ \frac{m_l-m_t}{2m_lm_t} & \frac{m_l+m_t}{2m_lm_t} & 0 \\ 0 & 0 & \frac{1}{m_t} \end{bmatrix}$
			$\Delta_{[010]}$	$\begin{bmatrix} \frac{m_l+m_t}{2m_lm_t} & \frac{m_t-m_l}{2m_lm_t} & 0 \\ \frac{m_t-m_l}{2m_lm_t} & \frac{m_l+m_t}{2m_lm_t} & 0 \\ 0 & 0 & \frac{1}{m_t} \end{bmatrix}$
			$\Delta_{[001]}$	$\begin{bmatrix} 1/m_t & 0 & 0 \\ 0 & 1/m_t & 0 \\ 0 & 0 & 1/m_l \end{bmatrix}$
[111]	$[\bar{1}10]$	$[\bar{1}\bar{1}2]$	$\Delta_{[100]}$	$\begin{bmatrix} \frac{m_t+2m_l}{3m_tm_l} & \frac{m_l-m_t}{\sqrt{6}m_tm_l} & \frac{m_l-m_t}{3\sqrt{2}m_tm_l} \\ \frac{m_l-m_t}{\sqrt{6}m_tm_l} & \frac{m_t+m_l}{2m_tm_l} & \frac{m_t-m_l}{2\sqrt{3}m_tm_l} \\ \frac{m_l-m_t}{3\sqrt{2}m_tm_l} & \frac{m_t-m_l}{2\sqrt{3}m_tm_l} & \frac{m_t+5m_l}{6m_tm_l} \end{bmatrix}$
			$\Delta_{[010]}$	$\begin{bmatrix} \frac{m_t+2m_l}{3m_tm_l} & \frac{m_t-m_l}{\sqrt{6}m_tm_l} & \frac{m_l-m_t}{3\sqrt{2}m_tm_l} \\ \frac{m_t-m_l}{\sqrt{6}m_tm_l} & \frac{m_t+m_l}{2m_tm_l} & \frac{m_l-m_t}{2\sqrt{3}m_tm_l} \\ \frac{m_l-m_t}{3\sqrt{2}m_tm_l} & \frac{m_l-m_t}{2\sqrt{3}m_tm_l} & \frac{m_t+5m_l}{6m_tm_l} \end{bmatrix}$
			$\Delta_{[001]}$	$\begin{bmatrix} \frac{m_t+2m_l}{3m_tm_l} & 0 & \frac{\sqrt{2}(m_t-m_l)}{3m_tm_l} \\ 0 & \frac{1}{m_t} & 0 \\ \frac{\sqrt{2}(m_t-m_l)}{3m_tm_l} & 0 & \frac{2m_t+m_l}{3m_tm_l} \end{bmatrix}$

Figure A.4: Expressions for the $\mathbf{W}_{DCS}^{3D}$ tensor for different DCS transport directions $\{x, y, z\}$. DCS directions are expressed as a function of the CCS. $\Delta_{i,j,k}-$valleys in the right table are oriented according to the sketch of the Brillouin zone reported on the left

.

Appendix B

Effective mass Schrödinger equation for a 2D and 1D electron gas for arbitrary crystal orientations and for heterostructures

Because of confinement, the quantized nature of the electron gas must be taken into account. Below we follow the approach in [1]: for a more complete derivation the reader may want to refer to [1]. The single electron, time-dependent Schrödinger equation can be written as:

$$i\hbar\frac{\partial}{\partial t}\Psi(\mathbf{R},t)=\left[-\frac{\hbar^2}{2m_0}\nabla^2_{\mathbf{R}}+U_C(\mathbf{R})+U(\mathbf{R})\right]\Psi(\mathbf{R},t) \tag{B.1}$$

where $U_C(\mathbf{R})$ represents the stationary periodic potential of the crystal atoms, $U(\mathbf{R})=-e\phi(\mathbf{R})$ is a generic potential energy that can be due to an applied external bias, and $\Psi(\mathbf{R},t)$ is the wave-function of the quantum system. For a uniform and homogeneous crystal domain with $U(\mathbf{R})=0$ the solution of Eq.B.1 is given by Bloch waves in the form:

$$\Psi_{m,\mathbf{K}}(\mathbf{R},t)=u_{m,\mathbf{K}}(\mathbf{R})e^{(-i\mathbf{K}\cdot\mathbf{R})}e^{(i\frac{E_m(\mathbf{K})}{\hbar}t)} \tag{B.2}$$

where m is the index of the eigenstates, $u_{m,\mathbf{K}}(\mathbf{R})$ represents the rapid oscillations on the scale of the crystal lattice and $E_m(\mathbf{K})$ are the eigenvalues solutions of Eq.B.1 once the space and time dependent problems are decoupled in virtue of the stationary potential $U_C(\mathbf{R})$.

More generally, in the presence of a potential $U(\mathbf{R})$ with slow space variations with respect to the crystal potential $U_C(\mathbf{R})$, Eq.B.1 can be expanded in eigenfunctions of the unperturbed Hamiltonian and after some non-trivial steps it can be demonstrated

that, close to a given minima identified by the index m and for minima sufficiently distant from each other, the Laplacian operator written with the rest electron mass summed with the crystal potential in Eq.B.1, can be replaced by the operator $E_m(-i\nabla_{\mathbf{R}})$ obtained by replacing $\mathbf{K}$ with $-i\nabla_{\mathbf{R}}$ in the expression for $E_m(\mathbf{K})$. Hence, Eq.B.1 can be rewritten within the so-called effective mass approximation (EMA) by introducing an equivalent Hamiltonian operator (H_{eq}):

$$i\hbar\frac{\partial}{\partial t}\Psi_{m,\mathbf{K}}(\mathbf{R},t) = \underbrace{\left[E_m(-i\nabla_{\mathbf{R}}) + U(\mathbf{R})\right]}_{H_{eq}}\Psi_{m,\mathbf{K}}(\mathbf{R},t). \tag{B.3}$$

In Eq.B.3 all the rapid potential variations of the crystal are accounted in the $E_m(-i\nabla_{\mathbf{R}})$ term. The most common way to write the $E_m(-i\nabla_{\mathbf{R}})$ operator for electrons with $\mathbf{K}$ values in the proximity of a valley minimum inside the Brillouin zone for the lowest eigenstate (i.e. by setting the m index equal), is to use the parabolic effective mass approximation, for which the energy dispersion relation $E(\mathbf{k})$ of the Bloch electrons can be written as:

$$E_\nu = E_{\nu 0} + \frac{\hbar^2\left(k_x - k_{x0}\right)^2}{2m_x} + \frac{\hbar^2\left(k_y - k_{y0}\right)^2}{2m_y} + \frac{\hbar^2\left(k_z - k_{z0}\right)^2}{2m_z} \tag{B.4}$$

where ν is the valley index and $E_{\nu 0}$ represents the energy of the bottom of the conduction band for the valley ν. Eq.B.4 is valid for valleys aligned with the x, y, z-coordinate system and more details are given in Section.A. Under the hypothesis of $\mathbf{K}$ values close to subband minima, the wave-function can be written as:

$$\Psi_{\nu,n,\mathbf{K}}(\mathbf{R},t) = u_{n,\mathbf{K}_\nu}(\mathbf{R})\Phi_{\nu,n,\mathbf{K}}(\mathbf{R})e^{(i\mathbf{K}_\nu\cdot\mathbf{R})}e^{(-i\frac{\varepsilon_{\nu,n}t}{\hbar})} \tag{B.5}$$

where the function $\Phi_{\nu,n,\mathbf{K}}(\mathbf{R})$ is a slowly varying envelope wave-function reflecting the variations of the potential $U(\mathbf{R})$. By substituting Eq.B.5 into Eq.B.3 we finally obtain the eigenvalue problem written for the envelope wave-function:

$$\left[E_\nu(-i\nabla_{\mathbf{R}}) + U(\mathbf{R})\right]\Phi_{\nu,n}(\mathbf{R}) = \varepsilon_{\nu,n}\Phi_{\nu,n}(\mathbf{R}) \tag{B.6}$$

where $\varepsilon_{\nu,n}$ is the eigenvalue, and $U(\mathbf{R})$ is the confinement energy.

Eq.B.6 can be used to derive the properties of a quantized gas from the slowly varying wave envelope wave-function $\Phi_{\nu,n}(\mathbf{R})$ rather than from the much more complicated wave-function defined in Eq.B.1. The envelope wave-function approximation relies on the fact that for large semiconductor domains the number of atoms is such that their potential can be represented by a global average potential. Based upon previous steps, the time-independent Schrödinger equation derived within EMA (Eq.B.6) is recast in a more explicit form in Eq.B.7

$$\left[-\frac{\hbar^2}{2}\sum_{i,j}\mathbf{W}_{i,j}\frac{\partial^2}{\partial \mathrm{x_i}\partial \mathrm{x_j}} + U_\nu(\mathbf{R})\right]\Phi_{\nu,n}(\mathbf{R}) = \varepsilon'_{\nu,n}\Phi_{\nu,n}(\mathbf{R}) \tag{B.7}$$

where $\varepsilon'_{\nu,n}$ is the eigenvalue referred to the energy of the $\nu-$th valley minimum (i.e. $\varepsilon_{\nu,n} = \varepsilon'_{\nu,n} + E_{\nu,0}$) and x_i is a generic coordinate in the device coordinate system (DCS) $\{x, y, z\}$.$\mathbf{W}$ is the tensor of the inverse of the effective mass tensor $\mathbf{M}$ in the DCS for which holds the relation ($\mathbf{M}_{i,j} = (\mathbf{W})^{-1}_{i,j}$). $\mathbf{M}$ takes a particularly simple form when the inverse mass tensor is diagonal:

$$\mathbf{W} = \begin{pmatrix} 1/m_x & 0 & 0 \\ 0 & 1/m_y & 0 \\ 0 & 0 & 1/m_z \end{pmatrix} \tag{B.8}$$

that leads to a diagonal matrix

$$\mathbf{M} = \begin{pmatrix} m_x & 0 & 0 \\ 0 & m_y & 0 \\ 0 & 0 & m_z \end{pmatrix}. \tag{B.9}$$

More generally, for arbitrary orientations and by exploiting the matrix formalism, the Schrödinger equation written in the device-coordinate-system (DCS) (see Section.A for more details) reads:

$$\left[-\frac{\hbar^2}{2}\nabla \cdot (\mathbf{W}_{DCS}\ \nabla) + U_\nu(\mathbf{R})\right] \Phi_{\nu,n}(\mathbf{R}) = \varepsilon'_{\nu,n}\Phi_{\nu,n}(\mathbf{R}) \tag{B.10}$$

where $\nabla\cdot$ denotes the divergence and $\nabla\Phi_{\nu,n}(\mathbf{R})$ the gradient of the envelope wave-function and where the effective-mass tensor may no longer be diagonal for an arbitrary crystallographic orientations.

B.1 2D electron gas

Let us assume the case of a 2D quantized electron gas where quantization direction and transport plane are shown in Fig.B.1. The Schrödinger equation in a 3D domain

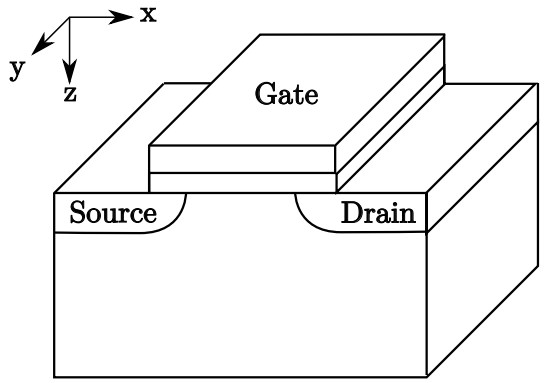


Figure B.1: Sketch of a MOS transistor with DCS.

given in Eq.B.10, under the hypothesis that the potential energy is a function of the

z coordinate only ($U(\mathbf{R}) = U(z)$), can be solved using the *ansatz* for the envelope wave-function $\Phi_{n,\mathbf{k}}(z, \mathbf{r})$ given by:

$$\Phi_{n,\mathbf{k}}(z, \mathbf{r}) = \frac{\xi_{n,\mathbf{k}}(z)}{\sqrt{A}} e^{i\,k_x\,x + i\,k_y\,y} \tag{B.11}$$

where A is a normalization area. By substituting Eq.B.11 into Eq.B.10 we obtain:

$$\frac{1}{2} w_{zz} \hbar^2 \left(\frac{d^2 \xi_{n,\mathbf{k}}(z)}{dz^2} \right) - \hbar^2 \left(w_{xz} k_x + w_{yz} k_y \right) \left(\frac{d\xi_{n,\mathbf{k}}(z)}{dz} \right) = \left[U(z) + \varepsilon_n^{'} \right] \xi_{n,\mathbf{k}}(z) \tag{B.12}$$

where the w_{ij} terms with $(i, j) \in \{x, y, z\}$ are the elements of the tensor of the inverse effective masses describing the kinetic energy operator reported in Eq.A.6 and the total energy $E_n(\mathbf{k})$ with $\mathbf{k} \in \{k_x, k_y\}$ is given by:

$$E_n(\mathbf{k}) = E_0 + \varepsilon_n^{'} + \frac{1}{2} \hbar^2 \left(w_{xx} k_x^2 + 2 w_{xy} k_x k_y + w_{yy} k_y^2 \right). \tag{B.13}$$

where E_0 represents the energy of the bottom of the considered valley. To eliminate the first derivative with respect to the z coordinate in Eq.B.12, it is convenient to make a second *ansatz* for the unknown $\xi_{n,\mathbf{k}}(z)$ [2]:

$$\xi_{n,\mathbf{k}}(z) = \zeta_n(z) e^{i\,z(w_{xz} k_x + w_{yx} k_y)/w_{zz}} \tag{B.14}$$

where $\zeta_n(z)$ must satisfy:

$$\left[-\frac{\hbar^2}{2 m_z} \frac{d^2}{dz^2} + U(z) \right] \zeta_n(z) = \varepsilon_n^{''} \zeta_n(z) \tag{B.15}$$

where $m_z = w_{zz}^{-1}$ and the total energy is given by:

$$E_n(\mathbf{k}) = E_0 + \varepsilon_n^{''} + \frac{\hbar^2}{2} \left[\left(w_{xx} - \frac{w_{xz}^2}{w_{zz}} \right) k_x^2 + 2 \left(w_{xy} - \frac{w_{xz} w_{yz}}{w_{zz}} \right) k_x k_y + \left(w_{yy} - \frac{w_{yz}^2}{w_{zz}} \right) k_y^2 \right]. \tag{B.16}$$

The definition of $\xi_{n,\mathbf{k}}(z)$ in Eq.B.14 shows that, if the entries w_{xz} and w_{yx} are not null, that is for semiconductors having ellipsoidal conduction band minima (e.g., silicon and germanium), and with z not aligned with any axis of the energy ellipsoid, then $\xi_{n,\mathbf{k}}(z)$ depends on the in-plane wave-vector $\mathbf{k}$ through a phase factor.

Hence, in the case of equi-energy ellipsoids aligned with the quantization direction z (i.e. $\mathbf{W}_{DCS}$ is diagonal), the $\xi_{n,\mathbf{k}}(z)$ term in the envelope wave-function of Eq.B.11 is given by the solution of:

$$\left[-\frac{\hbar^2}{2 m_z} \frac{d^2}{dz^2} + U(z) \right] \xi_n(z) = \varepsilon_n^{'} \xi_n(z) \tag{B.17}$$

and the total energy reads:

$$E_n(\mathbf{k}) = E_0 + \varepsilon_n^{'} + \frac{\hbar^2}{2} \left[\left(w_{xx} \right) k_x^2 + \left(w_{yy} \right) k_y^2 \right]. \tag{B.18}$$

It is now interesting to analyze the implications of the formulation for the envelope wave-function given in this appendix for the cases of single and multiple domains.

Single domain

Let us suppose the case of simple Dirichlet boundary conditions:

$$\Phi(z_0) = 0 \tag{B.19}$$

where z_0 denotes the boundary coordinate z. Since $\xi_{n,\mathbf{k}}(z_0) = 0$ implies $\zeta_{n,\mathbf{k}}(z_0) = 0$ according to Eq.B.14, ε_n'' and $\zeta_n(z)$ are independent of the in-plane $\mathbf{k}-$vector even if w_{xz} and w_{yx} are not zero. Instead, in the case of Dirichlet boundary conditions of the type:

$$\Phi(z_0) \neq 0 \tag{B.20}$$

or in the case of Neumann boundary conditions, according to Eqs.B.11 and B.14 both ε_n'' and $\zeta_n(z)$ depend on $\mathbf{k}$, unless we have $w_{xz} = w_{yx} = 0$.

Multiple domains

If we consider the penetration of the wave-function into the oxide (or into another semiconductor with a different effective mass tensor), even for spherical bands (i.e. $w_{xz}(z) = w_{yx}(z) = 0$), if the semiconductor mass (m_{sct}) and the oxide mass (m_{ox}) are different, the z dependent mass m_z makes the kinetic energy related to the in plane wave-vector $\mathbf{k}$ not a simple additive term and brings about a $\mathbf{k}$ dependence of the term $\xi_{n,\mathbf{k}}(z)$ (it is not necessary to introduce the *ansatz* in Eq.B.14 being in this simple case of spherical bands the tensor of the effective masses diagonal). In fact, let us consider isotropic bands, in this case $\xi_{n\mathbf{k}}(z)$ in Eq.B.15 must satisfy

$$\left[-\frac{\hbar^2}{2m_{ox}}\frac{\partial^2}{\partial z^2} - e\phi(z) + \Phi_B + \frac{\hbar^2 k^2}{2m_{ox}}\right]\xi_{n\mathbf{k}} = E_n\,\xi_{n\mathbf{k}} \quad (oxide) \tag{B.21a}$$

$$\left[-\frac{\hbar^2}{2m_{sct}}\frac{\partial^2}{\partial z^2} - e\phi(z) + \frac{\hbar^2 k^2}{2m_{sct}}\right]\xi_{n\mathbf{k}} = E_n\,\xi_{n\mathbf{k}} \quad (semiconductor) \tag{B.21b}$$

where k= $|\mathbf{k}|$, $\phi(z)$ is the electrostatic potential and Φ_B is the barrier height. Eq.B.21 can be cast in the equivalent form:

$$\left[-\frac{\hbar^2}{2m_{ox}}\frac{\partial^2}{\partial z^2} - e\phi(z) + \Phi_B + \frac{\hbar^2 k^2}{2}\left(\frac{1}{m_{ox}} - \frac{1}{m_{sct}}\right)\right]\xi_{n\mathbf{k}} = \varepsilon_n\,\xi_{n\mathbf{k}} \quad (oxide) \tag{B.22a}$$

$$\left[-\frac{\hbar^2}{2m_{sct}}\frac{\partial^2}{\partial z^2} - e\phi(z)\right]\xi_{n\mathbf{k}} = \varepsilon_n\,\xi_{n\mathbf{k}} \quad (semiconductor) \tag{B.22b}$$

where ε_n is related to the eigenvalue E_n of Eq.B.21 by E_n=ε_n+$\hbar^2k^2/(2m_{sct})$. For $m_{ox} \neq m_{sct}$ the transverse wave-vector $\mathbf{k}$ does not simply provide an additive term to the subband minima ε_n given by Eq.B.22, but instead enters Eq.B.22 as a parameter and makes both $\xi_{n\mathbf{k}}(z)$ and ε_n dependent on $\mathbf{k}$.

In this thesis we essentially solved Eq.B.22 by neglecting the term $\hbar^2k^2/2(1/m_{ox} - 1/m_{sct})$ compared to the barrier energy Φ_B. To make a quantitative assessment about the importance of the $\hbar^2k^2/2(1/m_{ox} - 1/m_{sct})$ term, we considered the HfO_2-InAs

system (m_{HfO_2}=0.11 m_0 and m_{InAs}=0.026 m_0) with an electron density n=$2 \cdot 10^{12}$ cm^{-2}. The corresponding magnitude of the Fermi wave-vector is $k_F = \sqrt{2\pi \cdot n} \simeq 0.35$ nm^{-1}, leading to $\hbar^2 k^2/2(1/m_{ox} - 1/m_{sct}) \simeq$177 meV, which is very small compared to $\Phi_B = 2.9eV$. This analysis suggests that the k dependent term provides only a minor correction to the solution of Eq.B.22.

B.1.1 Conservation of the transverse momentum

Let us analyze the case of the Si-SiO_2 system including wave-function penetration. In the solution for the Schrödinger equation we use an effective mass approximation and notice that Si and SiO_2 conduction band minima have different positions in the 3D Brillouin zone: $\Delta-$valleys of silicon are located at $(0.852\pi/a_0)$ (where a_0 is the lattice constant), whereas for the oxide the minimum is at the Γ point. Hence, while the wave-function used in our calculations is only the envelope wave-function, one may argue that, when the wave-function penetration in the oxide is considered, one should impose the continuity of the complete wave-function (i.e. accounting also for the different Bloch functions in the two materials) and that, furthermore, parallel momentum should be conserved.

The parallel momentum conservation is of course very important, for example, for the calculation of tunneling currents and it has been debated in several previous publications [3]. If parallel momentum conservation is used in tunneling calculations, for example, it is known that much larger barriers and correspondingly smaller tunneling currents are obtained for $\langle 110 \rangle$ and $\langle 111 \rangle$ silicon interfaces compared to the $\langle 100 \rangle$ interface; however this has not been observed in experiments.

The lack of a clear experimental evidence for parallel momentum conservation in the $Si-SiO_2$ system has been tentatively ascribed to the amorphous nature of thermally grown SiO_2 films, or to the fact that some scattering mechanisms (e.g. phonons, disorder induced scattering or surface roughness itself) may assist tunneling and provide the missing momentum. At this regard, it is finally worth recalling that the so called unprimed subbands in $\langle 100 \rangle$ silicon inversion layers stem from the two out of plane Δ minima (with quantization mass $m_z \simeq 0.92 m_0$), which are projected at the Γ point in the 2D Brillouin zone of the quasi-2D electron gas. This implies that, in terms of 2D wave-vectors in the the 2D Brillouin zone, the dominant subbands in the silicon inversion layer are actually located at the Γ point.

While the above discussion is certainly not conclusive, the point we tried to make is that several physical explanations have been proposed in the literature to justify quantization models for inversion layers at a $Si-SiO_2$ interface that do not account for the parallel momentum conservation.

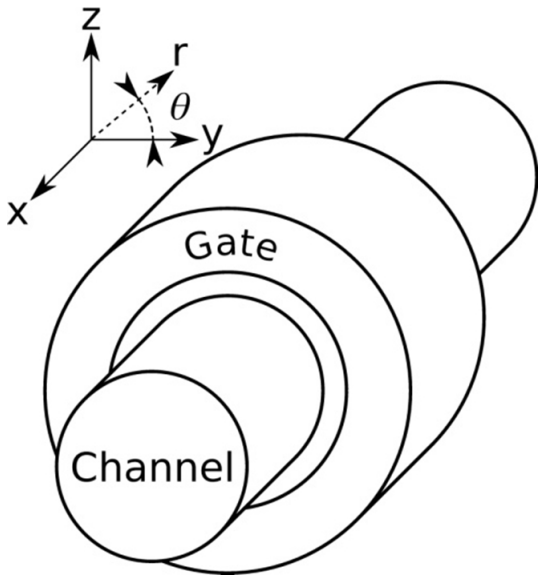


Figure B.2: Sketch of a circular NW.

B.2 1D electron gas

Under the hypothesis that the confinement energy $U_\nu(\mathbf{R})$ can be written as the sum of two terms

$$U(\mathbf{R}) = U(x) + U_\nu(y,z), \tag{B.23}$$

the envelope wave-function can be written as the product of two components

$$\Phi(\mathbf{R}) = \Phi(x)\Phi(y,z). \tag{B.24}$$

Assuming that the confinement potential varies only in the quantization plane y,z by setting $U(x)=0$, and substituting Eqs.B.24,B.23 into Eq.B.10 it can be easily demonstrated that a plane wave satisfies the equation for $\Phi(x)$ giving the dispersion relation $E_x(k_x) = (\hbar^2 k_x^2)/(2m_x)$. According to this *ansatz* on the longitudinal potential $U(x)$, the envelope wave-function can be written as the separated contributions of a transverse and longitudinal component:

$$\Phi_{n,k_x}(\mathbf{R}) = \frac{S_{n,kx}(y,z)}{\sqrt{L_x}} e^{ik_x x} \tag{B.25}$$

where for the sake of simplicity the notation for the valley ν has been dropped from the envelope wave-function and where L_x is a normalization length that always disappears in the final results (i.e. for matrix elements, scattering rates). The k_x value in Eq.B.25 are set by the boundary conditions and according to Appendix D are given by the discrete k_x values:

$$k_x = n\frac{2\pi}{L}. \tag{B.26}$$

Following the approach in [4], it can be shown that it is convenient to introduce a second *ansatz* and write $S_{n,kx}(y,z)$ as:

$$S_{n,kx}(y,z) = \xi_n(y,z)\, e^{i\ k_x(\alpha y+\beta z)} \tag{B.27}$$

in order to obtain an eigenvalue equation where appears only the second derivative operator and where coefficients α and β are given by:

$$\alpha = \frac{-w_{xy}w_{zz} + w_{yz}w_{xz}}{w_{yy}w_{zz} - w_{yz}^2}, \quad \beta = \frac{-w_{yy}w_{xz} + w_{yz}w_{xy}}{w_{yy}w_{zz} - w_{yz}^2}. \tag{B.28}$$

According to the EMA and based upon the approximate expression for the band structure close to the band minima of Eq.A.10 the stationary envelope wave-function $\xi_n(y,z)$ and the quantized energy levels in the cross-section are the solutions of the eigenvalue problem:

$$\left[-\frac{\hbar^2}{2}\nabla\cdot\left(\mathbf{W}_{DCS}^{2D}\nabla\right) + U(r,\theta)\right]\xi_n(y,z) = \varepsilon_n\xi_n(y,z) \tag{B.29}$$

where $\mathbf{W}_{DCS}^{2D}$ is defined as:

$$\mathbf{W}_{DCS}^{2D} = \begin{pmatrix} w_{yy} & w_{yz} \\ w_{zy} & w_{zz} \end{pmatrix}. \tag{B.30}$$

The resulting envelope wave-function where the quantum confinement is decoupled from the transport direction is given by:

$$\Phi_{n,k_x}(\mathbf{R}) = \frac{\xi_n(y,z)}{\sqrt{L_x}} e^{ik_x(\alpha y+\beta z+x)} \tag{B.31}$$

and the total energy of a particle for the valley ν is written as:

$$E_{n,\nu}(k_x) = \varepsilon_{\nu,n} + \frac{\hbar^2 k_x^2}{2} m_{xx} \tag{B.32}$$

where m_{xx} is the effective mass along transport direction:

$$m_{xx} = [\mathbf{W}_{DCS}]^{-1}_{(1,1)} = \frac{w_{t1}w_{t2}w_l}{w_{yy}w_{zz} - w_{yz}^2} \tag{B.33}$$

and where w_{t1},w_{t2}, and w_l are the inverse of the effective masses in the ECS (see Appendix A for more details).

B.3 On the continuity conditions of the Schrödinger equation

The remarkable advantage of the Schrödinger equation written within the EMA is that all the informations of the microscopic structure of the semiconductor are

accounted in the effective mass and band edge energy $E_{\nu,0}$. This model can be used even in the presence of heterostructures as in the case of semiconductor-oxide domains. One could be tempted to write the Eq.B.10 in a more general form:

$$\left[-\frac{\hbar^2}{2}\nabla\cdot(\mathbf{W}_{DCS}(\mathbf{R})\nabla)+U(\mathbf{R})\right]\Phi_{\nu,n}(\mathbf{R})=\varepsilon'_{\nu,n}\Phi_{\nu,n}(\mathbf{R}) \tag{B.34}$$

where the effective masses are $\mathbf{R}$ dependent. Typically, the interfaces between oxide and semiconductor, or between different semiconductors, take place over a few lattice layers, so that potential variation which is typically on the range of few tens of meV up to 1-2 eV cannot be considered smooth with respect to the atoms potential as discussed in Sec.B for the derivation of Eq.B.10. To circumvent this issue, the Schrödinger equation can be solved in the different domains resorting to Eq.B.10 and then by using appropriate continuity conditions at the semiconductor-oxide interface. In order for the eigenvalue problem to be well posed we need two conditions at the semiconductor-oxide interface:

$$\Phi_{n,k_x}(\mathbf{R}) \quad \text{and} \quad (\mathbf{W}_{DCS}\nabla\Phi_{n,k_x}(\mathbf{R}))^{\perp} \quad \text{continuous.} \tag{B.35}$$

where $(\bullet)^{\perp}$ denotes the quantity $(\bullet)$ normal to the semiconductor-oxide interface (e.g. for a planar device with a 2D electron gas corresponds to the direction normal to the semiconductor-oxide interface, whereas, for a circular GAA device, to the radial direction). The first equation follows directly from the fact that the probability distribution function must be single-valued everywhere in space, the second is a generalization for a 3D domain of the Ben Daniel-Duke condition [5]. In fact, if we consider for simplicity a 2D electron gas (the gas can freely move on the x, y plane), under the assumption of continuous envelope wave-function and integrating the Schrödinger equation over an infinitesimal range across the boundary $[-\epsilon/2, \epsilon/2]$ between the two materials (assuming the interface located at z=0) we write:

$$\begin{gathered}\int_{-\epsilon/2}^{\epsilon/2}\left[-\frac{\hbar^2}{2m^*}\frac{\partial^2}{\partial z}+U(z)\right]\Phi_n(z)dz=\varepsilon_n\int_{-\epsilon/2}^{\epsilon/2}\Phi_n(z)dz \\ -\frac{\hbar^2}{2}\left[\frac{1}{m}\frac{\partial}{\partial z}\Phi_n(z)\right]_{-\epsilon/2}^{+\epsilon/2}=\left[\varepsilon_n-\frac{|U^+(0)-U^-(0)|}{2}\right]\Phi_n(z)\epsilon.\end{gathered} \tag{B.36}$$

Obviously the ε_n term exits the integral being a non-local term since it is an eigen-energy solution of the Schrödinger equation in the whole domain. Evaluating the limit for $\epsilon \to 0$ we obtain the second boundary condition of Eq.B.35 for a 1D case:

$$\left[\frac{1}{m_I}\frac{\partial}{\partial z}\Phi_n^I(z)\right]_{z=0}=\left[\frac{1}{m_{II}}\frac{\partial}{\partial z}\Phi_n^{II}(z)\right]_{z=0} \tag{B.37}$$

where I and II denotes the two media. The above equation is the one obtained in [6] from $\mathbf{k}\cdot\mathbf{p}$ analysis and is the same as in [5] by considering steady state electron

transport across a heterojunction. Moreover, it can be demonstrated that Eq.B.37 satisfies the condition for the probability current density (J_Φ) to be solenoidal along the direction normal to the semiconductor-oxide interface. The starting point is to resort to the charge continuity equation expressed in the local form as:

$$\nabla \cdot \mathbf{J} = -\frac{\partial \rho}{\partial t} \tag{B.38}$$

where the current density $\mathbf{J}$ and the charge ρ must be regarded as probability current density ($\mathbf{J}_\Phi$) and probability density (P_Φ), respectively. The derivative of the probability density over the time is defined as:

$$\frac{\partial P_\Phi}{\partial t} = \Phi \frac{\partial \Phi^*}{\partial t} + \Phi^* \frac{\partial \Phi}{\partial t}. \tag{B.39}$$

By multiplying the l.h.s of the time-dependent Schrödinger equation of Eq.B.1 (by replacing the wave-function with the envelope wave-function) by Φ^* and the corresponding Schrödinger equation for Φ^* by Φ and taking the difference of the two equations we obtain:

$$\frac{\partial P_\Phi}{\partial t} = \Phi \frac{\partial \Phi^*}{\partial t} + \Phi^* \frac{\partial \Phi}{\partial t} = \frac{\hbar}{2im} \left[\Phi^* \nabla^2 \Phi - \Phi \nabla^2 \Phi^* \right]. \tag{B.40}$$

Using the Green's identity $f\nabla^2 g - g\nabla^2 f = \nabla \cdot (f\nabla g - g\nabla f)$ to rewrite the r.h.s of Eq.B.40 we finally obtain the generalization for 3D dimensions case of the continuity condition of Eq.B.38:

$$\frac{\partial P_\Phi}{\partial t} = -\nabla \cdot \underbrace{i\frac{\hbar}{2} \left[\mathbf{W}_{DCS} \left(\Phi^* \nabla \Phi - \Phi \nabla \Phi^* \right) \right]}_{\mathbf{J}_\Phi} \tag{B.41}$$

where the term $\mathbf{J}_\Phi$ denotes the probability current density. The existence of stationary states along the direction normal to the semiconductor-oxide interface, implies the component parallel to this direction of $\mathbf{J}_\Phi$ to be constant. Therefore the continuity of $\mathbf{J}_\Phi$ implies:

$$W^I_{DCS} \Im \left\{ \Phi^{I,*} \nabla \Phi^I \right\} |_{\text{interface}} = W^{II}_{DCS} \Im \left\{ \Phi^{II,*} \nabla \Phi^{II} \right\} |_{\text{interface}} \tag{B.42}$$

where $\Im(\bullet)$ denotes the imaginary part of $(\bullet)$. In virtue of Eq.B.42 it is now apparent that boundary conditions in Eq.B.35 implies that the probability current is constant along the direction normal to the semiconductor-oxide interface.

Appendix C

Non-parabolicity corrections

As mentioned in Appendix B throughout this work, the Schrödinger equation is solved within the EMA with a parabolic effective mass approximation for the energy relation $E(\mathbf{K})$ close to a valley minima, with an overall energy dispersion that is given by Eq.B.4 for a 3D electron gas, by Eq.B.16 for a 2D electron gas and by Eq.B.32 for a 1D electron gas. However, expecially for III-V semiconductors nonparabolicity cannot be neglected as shown in Fig.C.1. In this work we account for these effects by applying proper corrections to the DOS (for 3D electron gas), and to the eigenvalues for 2D and 1D quantized electron gas.

C.1 3D electron gas

C.1.1 Density of states

Carriers in a semiconductor without band quantization, occupy essentially continuum energy states either in the conduction or in the valence band. The number of allowed states per unit of energy and volume in $\mathbf{R}$-space is described by the density of states (DoS) here denoted by $g(E)$. By definition, the density of states can be evaluated by counting all the states available at a given energy E and then dividing by the normalization volume (or area for a 2D gas or length for a 1D gas).

The mathematical definition of DoS is thus:

$$g\left(E\right)=\frac{1}{\Omega}\sum_{\mathbf{K}}\delta(E-E(\mathbf{K})) \tag{C.1}$$

where Ω is the volume in the real space $\mathbf{R}$ and $\delta(\bullet)$ is the Dirac delta distribution.

In order to ensure that Eq.C.1 is more mathematically manageable, it is better converting the summation over $\mathbf{K}$ into an integral by using the prescriptions given by Eq.D.8:

$$g_{\nu}\left(E\right)=\frac{n_{sp}\mu_{\nu}}{\left(2\pi\right)^{3}}\int\delta\left(E-E_{\nu}\left(\mathbf{K}\right)\right)d\mathbf{K} \tag{C.2}$$

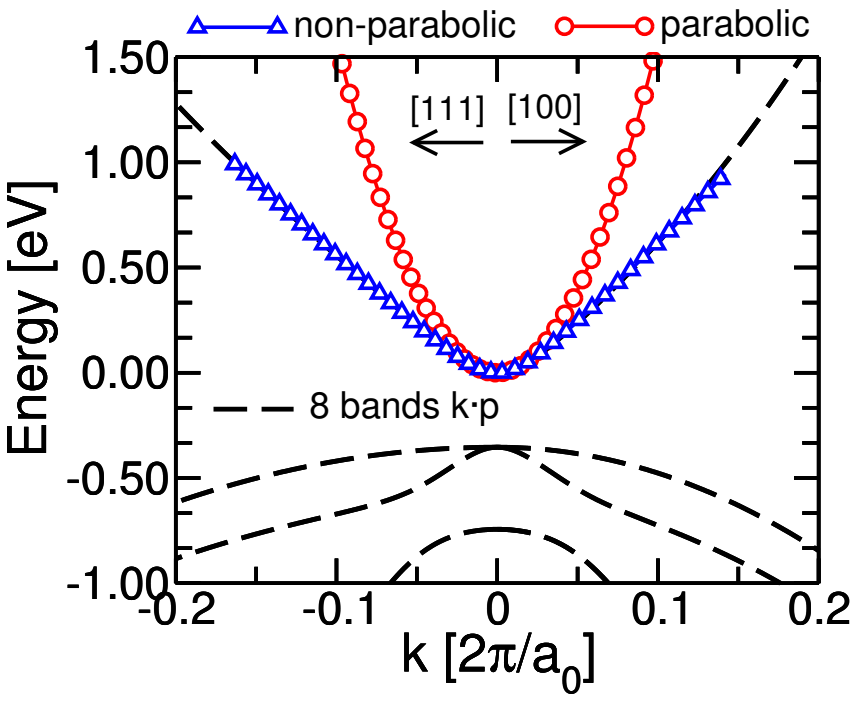


Figure C.1: Comparison between the energy band structure obtained for bulk InAs using an 8 bands **k·p** solver [7] (dashed-line), and the energy profile obtained by fitting the dashed-line with a parabolic equation (triangles) and with the non parabolic equation of Eq.C.15 (circles).

where ν is the valley index, μ_ν is the valley multiplicity factor and n_{sp} is the spin factor that is set to 2 according to Appendix D.

The energy regions around the minima of the conduction band for a given valley ν, can be approximated by a quadratic function of $\mathbf{K}$:

$$E_\nu(\mathbf{K}) - E_{\nu,0} = \frac{\hbar^2 {K_x}^2}{2m_x^*} + \frac{\hbar^2 {K_y}^2}{2m_y^*} + \frac{\hbar^2 {K_z}^2}{2m_z^*} \tag{C.3}$$

where $E_{\nu,0}$ is the energy of the bottom of the conduction band, and K_x, K_y, K_z are the deviations with respect to the $\mathbf{K}$-point at the bottom of the considered valley. When considering the ellipsoidal energy dispersion relation, in order to simplify analytical calculations for the DoS, it is useful to introduce a scale-transformation applied to an ellipsoid to turn it into a sphere. This is the Herring-Vogt transformation.

Herring-Vogt transformation

For the ellipsoidal equi-energy surfaces the transformation to turn an ellipsoid into a sphere is defined as:

$$K_i^* = \sum_j T_{ij} K_j \tag{C.4}$$

where K_i^* is the transformed wave-vector, K_j is the wave-vector belonging to the ellipsoidal dispersion relation and T_{ij} the linear transformation matrix with $i, j \in (x, y, z)$. For the energy relation in Eq.C.3 the $\mathbf{T}$ matrix is defined as:

$$T_{ij} = \begin{pmatrix} \left(\frac{m_0}{m_x^*}\right)^{1/2} & 0 & 0 \\ 0 & \left(\frac{m_0}{m_y^*}\right)^{1/2} & 0 \\ 0 & 0 & \left(\frac{m_0}{m_z^*}\right)^{1/2} \end{pmatrix} \tag{C.5}$$

where m_0 is the free electron mass. It should be noticed that, since the $\mathbf{T}$ matrix is diagonal, the inverse transformation is again diagonal with the inverse matrix element. We have:

$$\begin{aligned} K_x^* &= \sum_j T_{1j} K_j = K_x \sqrt{\frac{m_0}{m_x^*}} \\ K_y^* &= \sum_j T_{2j} K_j = K_y \sqrt{\frac{m_0}{m_y^*}} \\ K_z^* &= \sum_j T_{3j} K_j = K_z \sqrt{\frac{m_0}{m_z^*}} \end{aligned} \tag{C.6}$$

By substituting Eq.C.6 into Eq.C.3, we obtain the desired spherical relation:

$$E_\nu\left(\mathbf{K}^*\right) - E_{\nu,0} = \frac{\hbar^2 (K^*)^2}{2m_0}. \tag{C.7}$$

Looking at Eq.C.2 the volume element $d\mathbf{K}$ is modified into $d\mathbf{K}^*$

$$\begin{aligned} d\mathbf{K}^* &= dK_x^* dK_y^* dK_z^* \\ &= \sum_j T_{1j} dK_j \sum_j T_{2j} dK_j \sum_j T_{3j} dK_j \\ &= dK_x \frac{m_0}{m_x} dK_y \frac{m_0}{m_y} dK_z \frac{m_0}{m_z} \\ &= d\mathbf{K} \sqrt{\frac{\left(m_0^3\right)}{m_x m_y m_z}} \end{aligned} \tag{C.8}$$

Density of states: parabolic case

If we step back to the DoS for one of the simplest band structure, that is the case of parabolic dispersion relation, the energy simply depends on the square of the $\mathbf{K}$ vector through Eq.C.3. To perform the DoS calculation, we apply the Herring-Vogt transformation by replacing Eqs.C.3 with Eq.C.7, therefore, Eq.C.2 becomes:

$$g_\nu(E) = \frac{2\mu_\nu}{(2\pi)^3} \sqrt{\frac{m_{x,\nu} m_{y,\nu} m_{z,\nu}}{\left(m_0^3\right)}} \int_{\mathbf{K}^*} \delta\left(E - E_\nu\left(\mathbf{K}^*\right)\right) d\mathbf{K}^*. \tag{C.9}$$

Since the energy relation $E_\nu(\mathbf{K}^*)$ is isotropic, Eq.C.9 can be rewritten as:

$$g_\nu(E) = \frac{2\mu_\nu}{(2\pi)^3}\sqrt{\frac{m_{x,\nu}m_{y,\nu}m_{z,\nu}}{(m_0^3)}}\int_{K^*}\delta\left(E - E_\nu(K^*)\right)4\pi\left(K^*\right)^2 dK^* \tag{C.10}$$

where $K^* = |\mathbf{K}^*|$ and $4\pi\left(K^*\right)^2$ is the surface of a sphere of radius K^*. By writing:

$$(K^*)^2 = \frac{\left(E_\nu(K^*) - E_{\nu,0}\right)2m_0}{\hbar^2} \tag{C.11}$$

and differentiating K^*, we obtain:

$$\frac{dK^*}{dE_\nu} = \frac{\sqrt{2m_0}}{2\hbar}\frac{1}{\sqrt{E_\nu(K^*) - E_{\nu,0}}}. \tag{C.12}$$

To perform the DoS calculation, we must then substitute Eqs.C.11 and C.12 into Eq.C.10, obtaining:

$$g_\nu(E) = \frac{2\mu_\nu}{(2\pi)^3}\int_{E_{\nu,0}}^{\infty}\delta\left(E - E_\nu\right)\left[4\pi\frac{\left(E_\nu - E_{\nu,0}\right)2m_0}{\hbar^2}\times\right.$$
$$\left.\times\frac{\sqrt{2m_0}}{2\hbar}\frac{1}{\sqrt{E_\nu - E_{\nu,0}}}\sqrt{\frac{m_{x,\nu}m_{y,\nu}m_{z,\nu}}{(m_0^3)}}dE_\nu\right] \tag{C.13}$$

The Dirac delta in Eq.C.13 reduces the integral and we obtain:

$$\begin{aligned} g_\nu(E) &= \frac{\mu_\nu\sqrt{2}}{\pi^2\hbar^3}m_0^{\frac{3}{2}}\sqrt{E_\nu - E_{\nu,0}}\sqrt{\frac{m_{x,\nu}m_{y,\nu}m_{z,\nu}}{(m_0^3)}} \\ &= \frac{\mu_\nu\sqrt{2}}{\pi^2\hbar^3}\left(m_{d,\nu}\right)^{\frac{3}{2}}\sqrt{E_\nu - E_{\nu,0}}. \end{aligned} \tag{C.14}$$

where $m_{d,\nu} = (m_{x,\nu}m_{y,\nu}m_{z,\nu})^{1/3}$ is the so called density of state effective mass, which is equal to the geometric mean of the three masses for the valley ν. It can be demonstrated that Eq.C.14 corresponds to the DoS obtained for an isotropic case but with $m_d = m^{\frac{1}{3}}$ where m is the isotropic effective mass. It should be noticed that, to obtain the total DoS, $g_\nu(E)$ must be summed over all the valleys. If the main axes of constant-energy ellipsoids are aligned with the K_x K_y K_z axis of the crystal coordinate system (CCS) (see Appendix A for more details), then we can write $m_d = (m_l m_t m_t)^{1/3}$ where $m_{l,t}$ are the longitudinal and transverse mass of the ellipsoids equienergy surfaces.

Density of states: nonparabolic case

At energies not too close to the band minimum, the conduction band deviates from a parabola as illustrated in Fig.C.1. The non-parabolicity effect is included within

the dispersion relation by the non-parabolicity factor α which has the dimensions of an inverse energy, and is implicitly defined by:

$$\left(E_\nu\left(\mathbf{K}\right)-E_{\nu,0}\right)\left[1+\alpha_\nu\left(E_\nu\left(\mathbf{K}\right)-E_{\nu,0}\right)\right]=\frac{\hbar^2{K_x}^2}{2m_x^*}+\frac{\hbar^2{K_y}^2}{2m_y^*}+\frac{\hbar^2{K_z}^2}{2m_z^*} \tag{C.15}$$

with the prescription that K_x, K_y, K_z are the deviations with respect to the wave-vector at the bottom of the valley. To simplify the DoS calculations, we define a new variable named $\gamma\left(E\right)$:

$$\left(E_\nu\left(\mathbf{K}\right)-E_{\nu,0}\right)\left[1+\alpha_\nu\left(E_\nu\left(\mathbf{K}\right)-E_{\nu,0}\right)\right]=\gamma_\nu\left(E\right). \tag{C.16}$$

According to Herring-Vogt transformations (see Section C.1.1), we rewrite Eq.C.15 as if it was isotropic with an effective mass equal to $m_{d,\nu}$:

$$\left(E_\nu\left(\mathbf{K}\right)-E_{\nu,0}\right)\left[1+\alpha_\nu\left(E_\nu\left(\mathbf{K}\right)-E_{\nu,0}\right)\right]=\frac{\hbar^2K^2}{2m_{d,\nu}} \tag{C.17}$$

From the above equation, K and its derivative dK can be computed:

$$K=\frac{\sqrt{2m^*}\sqrt{\gamma_\nu\left(E\right)}}{\hbar} \tag{C.18}$$

$$\frac{dK}{dE}=\sqrt{\frac{m^*}{2\hbar^2}}\frac{1}{\sqrt{\gamma_\nu\left(E\right)}}\frac{d\gamma_\nu\left(E\right)}{dE} \tag{C.19}$$

By replacing Eqs.C.19 and C.18 into Eq.C.2 with the energy relation given by Eq.C.17, we obtain the DoS for the valley ν for non-parabolic conduction energy bands:

$$\begin{aligned}
g_\nu\left(E\right)&=\frac{\mu_\nu}{4\pi^3}\int_{E_{\nu,0}}\delta\left(E-E_\nu\left(\mathbf{K}'\right)\right)4\pi\left|\frac{\sqrt{2m^*}\sqrt{\gamma_\nu\left(E'\right)}}{\hbar}\right|^2\times\\
&\times\sqrt{\frac{m_{d,\nu}}{2\hbar^2}}\frac{1}{\sqrt{\gamma_\nu\left(E'\right)}}\frac{d\gamma_\nu\left(E'\right)}{dE'}dE'\\
&=\frac{\mu_\nu\sqrt{2}}{\pi^2\hbar^3}m_{d,\nu}\frac{3}{2}\int_{E_{\nu,0}}\delta\left(E-E\left(\mathbf{K}'\right)\right)\sqrt{\gamma_\nu\left(E'\right)}\frac{d\gamma_\nu\left(E'\right)}{dE'}dE'\\
&=\frac{\mu_\nu\sqrt{2}}{\pi^2\hbar^3}m_{d,\nu}\frac{3}{2}\sqrt{\gamma_\nu\left(E\right)}\frac{d\gamma_\nu\left(E\right)}{dE}dE\\
&=\frac{4\mu_\nu\pi\left(2m_{d,\nu}\right)^{\frac{3}{2}}}{h^3}\sqrt{\left(E_\nu-E_{\nu,0}\right)\left(1+\alpha_\nu\left(E_\nu-E_{\nu,0}\right)\right)}\left[1+2\alpha_\nu\left(E_\nu-E_{\nu,0}\right)\right]
\end{aligned} \tag{C.20}$$

where $m^*=\left(m_xm_ym_z\right)^{\frac{1}{3}}$.

C.1.2 Charge density

By definition the charge density is given by:

$$n = \sum_{\nu} \int_{E_{\nu,0}}^{\infty} g_{\nu}(E) f(E)\, dE \tag{C.21}$$

where $g_{\nu}(E)$ is the density of states for the valley ν given by Eq.C.20, $E_{\nu,0}$ is the conduction band energy minima for valley ν and $f_0(E)$ is the Fermi-Dirac distribution function given by:

$$f_0(E) = \frac{1}{1 + e^{\frac{E - E_F}{k_B T}}} \tag{C.22}$$

where E is the total energy, E_F is the Fermi level energy, k_B is the Boltzmann constant and T is the temperature. By substituting Eqs.C.20 and C.22 into Eq.C.21 we obtain:

$$n = \sum_{\nu} \frac{\mu_{\nu} (2 m_{d,\nu})^{3/2}}{2\pi^2 \hbar^3} \int_{E_{\nu,0}}^{\infty} \frac{\sqrt{(E - E_{\nu,0}) \left[1 + \alpha_{\nu} (E - E_{\nu,0})\right]} \left[1 + 2\alpha_{\nu} (E - E_{\nu,0})\right]}{1 + e^{\frac{E - E_F}{k_B T}}} \, dE. \tag{C.23}$$

By introducing the variables:

$$\zeta_{\nu} = \frac{E - E_{\nu,0}}{K_B T} \tag{C.24}$$

$$\eta_{\nu} = \frac{E_F - E_{\nu,0}}{K_B T} \tag{C.25}$$

Eq.C.23 becomes:

$$n = \sum_{\nu} \frac{\mu_{\nu} (2 m_{d,\nu} K_B T)^{3/2}}{2\pi^2 \hbar^3} \int_{0}^{\infty} \frac{\sqrt{\zeta_{\nu} \left[1 + \alpha_{\nu} \zeta_{\nu}\, k_B T\right]} \left[1 + 2\alpha_{\nu} \zeta_{\nu}\, k_B T\right]}{1 + e^{\zeta_{\nu} - \eta_{\nu}}} \, d\zeta_{\nu}. \tag{C.26}$$

In the case of parabolic dispersion relation (i.e. $\alpha = 0$), Eq.C.26 simplifies to:

$$\begin{aligned} n &= \sum_{\nu} \frac{\mu_{\nu} (2 m_{d,\nu} K_B T)^{3/2}}{2\pi^2 \hbar^3} \int_{0}^{\infty} \frac{\sqrt{\zeta_{\nu}}}{1 + e^{\zeta_{\nu} - \eta_{\nu}}} \, d\zeta_{\nu} \\ &= \sum_{\nu} \frac{\mu_{\nu} (2 m_{d,\nu} K_B T)^{3/2}}{2\pi^2 \hbar^3} \mathcal{F}_{\frac{1}{2}}(\eta)\, \Gamma(1 + \frac{1}{2}) \end{aligned} \tag{C.27}$$

where $\mathcal{F}_{\frac{1}{2}}(\eta)$ is the Fermi integral of order $1/2$ defined as:

$$\mathcal{F}_j(\eta) = \frac{1}{\Gamma(1 + j)} \int_{0}^{\infty} \frac{\zeta^j}{1 + e^{\zeta - \eta}} d\zeta \tag{C.28}$$

and $\Gamma(1+\frac{1}{2}) = \sqrt{\pi}/2$ is the Gamma function of 3/2. Finally, for a 3D electron gas under the approximation of parabolic bands, the charge density reads:

$$\begin{aligned} n &= \sum_{\nu} \underbrace{\mu_{\nu} \frac{(2m_{d,\nu}K_BT)^{3/2}}{2\pi^2\hbar^3}\frac{\sqrt{\pi}}{2}}_{N_{C,\nu}} \mathcal{F}_{\frac{1}{2}}(\eta_{\nu}) \\ &= \sum_{\nu} N_{C,\nu}\, \mathcal{F}_{\frac{1}{2}}(\eta_{\nu}). \end{aligned} \tag{C.29}$$

where N_C is the so-called effective density of states for the conduction band. Moreover, if we assume a Maxwell-Boltzmann distribution function instead of the Fermi Dirac, we obtain:

$$n = \sum_{\nu} N_{C,\nu}\, e^{-\frac{E_F - E_{\nu,0}}{K_BT}}. \tag{C.30}$$

C.2 2D electron gas

Non-parabolic corrections have been implemented following [8]. In each section of the device normal to the transport direction x, we expressed the energy $E_n(\mathbf{k})$ versus the wave-vector $\mathbf{k}$= (k_x, k_y) for a given valley in the transport plane as [8]:

$$E_n(\mathbf{k}) = \langle U_n \rangle + \frac{\sqrt{1 + 4\alpha[\frac{\hbar^2}{2}\left[\frac{k_x^2}{m_x} + \frac{k_y^2}{m_y}\right] + \varepsilon_n^{(p)} - \langle U_n \rangle]} - 1}{2\alpha} \tag{C.31}$$

where $\langle U_n \rangle = \int |\xi_{n,x}(z)|^2 U(x,z) dz$ is the expectation value in section x of the total potential energy ($U(x,z)$) for subband n, z is the quantization direction, y is the direction normal to the transport, while $\varepsilon_n^{(p)}$ and $\xi_{n,x}(z)$ are the subband minimum referred to the energy of the conduction band minima of the considered valley and envelope wave-function for the parabolic effective mass Hamiltonian in Eq.B.17; α is finally the non-parabolicity factor [8]. The non-parabolic subband minimum ($\varepsilon_n^{(np)}$) is obtained by setting $\mathbf{k}$= 0 into Eq.C.31.

The expressions for the electron velocity and density of states stemming from Eq.C.31 are reported in [8].

In Eq.C.31 the non-parabolicity affects both the subband minima and the energy versus $\mathbf{k}$ relation in the transport plane. This is a distinctive feature of the model in [8], which makes the transport mass (as well as other band-structure figures) dependent on size and bias induced quantization and, consequently, dependent on the well thickness and different in each subband. To be more specific, the transport mass m_{tr} at the subband minimum can be calculated from the E($\mathbf{k}$) in Eq.C.31 and, for an isotropic material and an ideal square quantum well corresponding to $\langle U_n \rangle = 0$, one readily obtains:

$$m_{tr,n} = m_{xy}\sqrt{1 + 4\,\alpha\,\varepsilon_n^{(p)}} \tag{C.32}$$

where α is the non-parabolicity coefficient, $\varepsilon_n^{(p)}$ are the subband minima for the $n-$th eigenvalue of the parabolic Schrödinger equation and m_{xy}=m_x=m_y is the transport mass for the bulk material.

Eq.C.32 clearly shows that m_{tr} changes with the subband, differs from m_{xy} and depends on the well thickness. This can be also observed in Fig.C.2, reporting the energy relation E(k) from Eq.C.31 for the lowest subband of a quantum well with different well thickness T_w. The curvature of the $E(k)$ at the subband minimum is clearly affected by T_w.

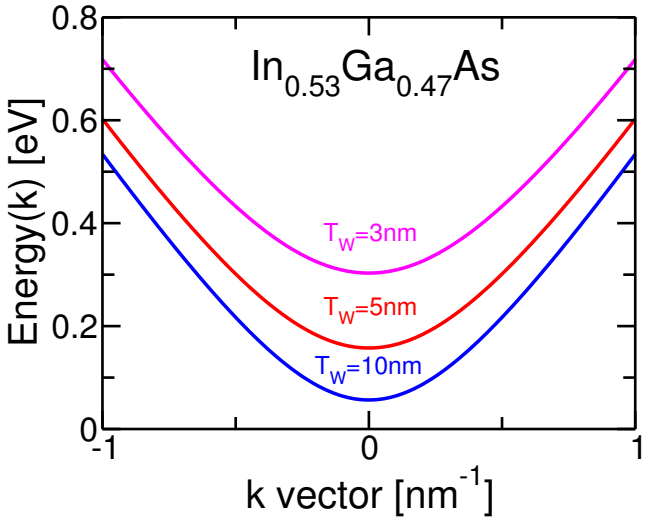


Figure C.2: Energy versus wave-vector magnitude k obtained with Eq.1 of the manuscript, for an ideal quantum well and different well thickness T_w including wave-function penetration in the HfO_2 oxide. The potential energy in the quantum well is taken as the zero.

Since in our model the transport mass is not simply m_{xy} and changes with T_w, we believe that it is not advisable to employ a T_w dependent m_{xy}, as it has been often done in strictly parabolic models, with an m_{xy} extracted from, for example, tight-binding simulations as in [9]. It should be also considered that in $In_{1-x}Ga_xAs$ semiconductor materials the $E(k)$ is strongly non parabolic (see Fig.2.1 and Fig.C.2), and so the band curvature at the very bottom of the subband is by no means the only relevant figure.

In order to address pragmatically the validity of our non parabolic energy model for different T_w, we decided to analyze first the subband minima versus T_w and then the $E(k)$ in a 5 nm thick quantum well comparing the results of our model with those retrieved in the literature and obtained with tight-biding calculations.

In this respect, Fig.C.3 compares the lowest subbands obtained with the model of this work and the $sp^3d^5s^*$ tight-binding method. As can be seen the non parabolic model can track very well the tight-binding results, reinforcing our confidence in the dependability of the model proposed in [8].

As a final verification, Fig.C.4 compares the E(k) in a 5nm quantum well obtained with our model and from the tight-binding calculations reported in [10]. As can be seen, the effective mass non parabolic model of Eq.C.31, can reproduce very well the tight-binding results, and for the well thickness which is used for most the device simulations reported in this work.

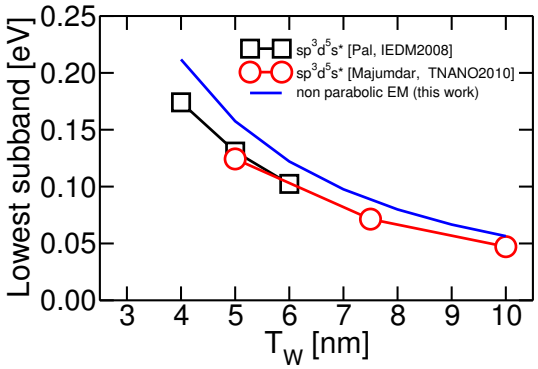


Figure C.3: Lowest subbands versus well thickness T_w. Subband minima for [11] and [12] were calculated as $\hbar^2/2m_q(\pi/T_w)^2$, from the values of the T_w dependent quantization mass m_q reported in the papers.

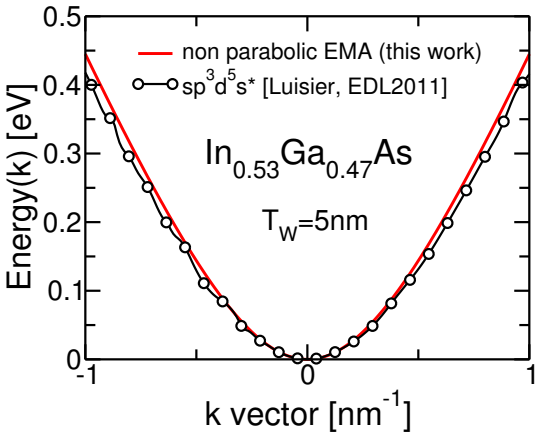


Figure C.4: Conduction band structure of an $In_{0.53}Ga_{0.47}As$ UTBs with T_w=5nm for an ideal quantum well. Comparison between open circles [10] obtained using the $sp^3d^5s^*$ tight-binding method and this work (solid line) including the wave-function penetration in the HfO_2 oxide. Due to the absence of a reference for the energies, the CB minima of the two E(k) curves have been arbitrarily set to zero.

C.2.1 Charge density at equilibrium

In the case of a 2D electron gas described according to the EMA approximation discussed in Appendix B.1, and assuming that $\xi_n(z)$ does not depend on the $\mathbf{k}-$vector in the transport plane, the equilibrium electron density per unit volume can be expressed as:

$$n(z) = \sum_{\nu,n} |\xi_{\nu,n}(z)|^2 \int_{\varepsilon_{\nu,n}^{(np)}}^{+\infty} g_{\nu,n}(E) f_{\nu,n}(E) dE \tag{C.33}$$

where $\varepsilon_{\nu,n}^{(np)}$ is the nonparabolic subband minimum obtained by setting $k = 0$ in Eq.C.31, ν is the valley index, n is the subband index, $g_{\nu,n}(E)$ is the nonparabolic density of states from [8], and $f_{\nu,n}(E)$ is the Fermi Dirac distribution function defined in Eq.C.22. By introducing the new variables:

$$\zeta_{\nu,n} = \frac{E - \varepsilon_{\nu,n}^{(np)}}{K_B T} \tag{C.34}$$

$$\eta_{\nu,n} = \frac{E_F - \varepsilon_{\nu,n}^{(np)}}{K_B T} \tag{C.35}$$

EqC.33 can be rewritten as:

$$n(z) = \sum_{\nu,n} |\xi_{\nu,n}(z)|^2 \frac{n_{sp}\mu_\nu}{2\pi\hbar^2} m_{d,\nu} \int_0^{+\infty} \frac{\left[1 + 2\alpha\left(K_B T \zeta_{\nu,n} + \varepsilon_{\nu,n}^{(p)} - U_\nu\right)\right]}{1 + e^{\zeta_{\nu,n} - \eta_{\nu,n}}} K_B T \, d\zeta_{\nu,n} \tag{C.36}$$

where $m_{d,\nu} = \sqrt{m_{x,\nu}\, m_{y,\nu}}$ is the DoS effective mass, μ_ν is the valley multiplicity, and n_{sp} is the spin factor is set to 2 according to Appendix D. After some calculations, Eq.C.36 can be finally written as:

$$n(z) = \sum_{\nu,n} |\xi_{\nu,n}(z)|^2 \frac{m_{d,\nu}\mu_\nu K_B T}{\pi\hbar^2} \left\{ \left[1 + 2\alpha\left(\varepsilon_{\nu,n}^{(p)} - U_\nu\right)\right] ln\left(1 + e^{\eta_{\nu,n}}\right) + 2\alpha K_B T \mathcal{F}_1(\eta_{\nu,n}) \right\} \tag{C.37}$$

where $\mathcal{F}_1(\eta_{\nu,n})$ is the Fermi integral of order 1 of $\eta_{\nu,n}$ defined in Eq.C.28.

C.3 1D electron gas

Non-parabolic corrections have been implemented following [8]. Along the transport direction x, we expressed the energy $E_n(k_x)$ for the subband n belonging to a given valley as [8]:

$$E_n(k_x) = \langle U_n \rangle + \frac{\sqrt{1 + 4\alpha\left[\frac{\hbar^2}{2}\left[\frac{k_x^2}{m_x}\right] + \varepsilon_n^{(p)} - \langle U_n \rangle\right]} - 1}{2\alpha} \tag{C.38}$$

where $\langle U_n \rangle = \int |\xi_n(y,z)|^2 U(y,z) dydz$ is the expectation value in section x of the total potential energy ($U(y,z)$) for subband n, (y,z) is the quantization plane, while $\varepsilon_n^{(p)}$ and $\xi_{n,x}(y,z)$ are the subband minimum referred to the energy of the conduction band minima of the considered valley and wave-function for the parabolic effective mass Hamiltonian in Eq.4.1; α is finally the non-parabolicity factor [8]. The non-parabolic subband minimum ($\varepsilon_n^{(np)}$) is obtained by setting $k_x = 0$ in Eq.C.38.

C.3.1 Charge density at equilibrium

In the case of a 1D electron gas described according to the EMA approximation discussed in Appendix B.2, and assuming that $\xi_n(y,z)$ does not depend on the k_x, the electron density per unit volume at a given section along transport direction is given by:

$$n(z,y) = \sum_{\nu,n} |\xi_{\nu,n}(y,z)| \int_{\varepsilon_{\nu,n}^{(np)}}^{+\infty} g_{\nu,n}(E) f_{\nu,n}(E)\, dE \tag{C.39}$$

where ν is the valley index, n is the subband index, $g_{\nu,n}(E)$ is the nonparabolic density of states from [8], and $f_{\nu,n}(E)$ is the Fermi Dirac distribution function defined in Eq.C.22.

The equation for the DoS in a 1D electron gas can be calculated following the procedure used for the 3D gas in Appendix C.1.1 but using the one dimensional energy relation of Eq.C.38 obtaining:

$$g_{\nu,n}(E) = \frac{2\mu_\nu}{\pi\hbar} \frac{1}{v_{g,\nu,n}(E)} \tag{C.40}$$

where μ_ν is the valley multiplicity factor, $v_{g,\nu,n}(E)$ is the group velocity for the subband n belonging to the valley ν and where we have accounted for the spin degeneracy factor (that has been set to 2). Eq.C.40 shows that the DoS has a singularity for the total energy E approaching the subband minima, because the velocity of carriers tends to zero. The expression for $v_{g,\nu,n}(E)$ is:

$$v_{g,\nu,n}(E) = \frac{\hbar\, k_x}{m_{x,\nu}\left[1 + 2\alpha_\nu \left(E - U_{\nu,n}\right)\right]}. \tag{C.41}$$

By substituting Eq.C.40 into Eq.C.39 and introducing the new variables given by Eq.C.35, we obtain:

$$n(y,z) = \sum_{\nu,n} |\xi_{\nu,n}(y,z)|^2 \frac{\mu_\nu \sqrt{2\, m_x}}{\pi\hbar} \int_0^{+\infty} \left\{ \frac{1 + 2\alpha \left(K_B T \zeta_{\nu,n} + \varepsilon_{\nu,n}^{(np)} - U_\nu\right)}{\sqrt{K_B T \zeta_{\nu,n} + \varepsilon_{\nu,n}^{(np)} - \varepsilon_{\nu,n}^{(p)} + \alpha \left(K_B T \zeta_{\nu,n} + \varepsilon_{\nu,n}^{(np)} - U_\nu\right)^2}} \times \right.$$
$$\left. \times \frac{K_B T}{1 + e^{\zeta_{\nu,n} - \eta_{\nu,n}}}\, d\zeta_{\nu,n} \right\} \tag{C.42}$$

In the parabolic case ($\alpha = 0$), it can be quite easily demonstrated that the DoS is given by:

$$g_{\nu,n}(E) = \frac{\mu_\nu \sqrt{2m_{x,\nu}}}{\pi\hbar} \frac{1}{\sqrt{E - \varepsilon_{\nu,n}^{(p)}}} \tag{C.43}$$

therefore, Eq.C.42 reduces to:

$$\begin{aligned} n(y,z) &= \sum_{\nu,n} |\xi_{\nu,n}(y,z)|^2 \, \frac{\mu_\nu \sqrt{2\, m_x K_B T}}{\pi\hbar} \int_0^{+\infty} \frac{\zeta_{\nu,n}^{-\frac{1}{2}}}{1 + e^{\zeta_{\nu,n} - \eta_{\nu,n}}} \, d\zeta_{\nu,n} \\ &= \sum_{\nu,n} |\xi_{\nu,n}(y,z)|^2 \, \frac{\mu_\nu \sqrt{2\, m_x K_B T}}{\pi\hbar} \mathcal{F}_{-\frac{1}{2}}(\eta_{\nu,n}) \Gamma_{-\frac{1}{2}} \\ &= \sum_{\nu,n} |\xi_{\nu,n}(y,z)|^2 \, \frac{\mu_\nu}{\hbar} \sqrt{\frac{2\, m_x K_B T}{\pi}} \mathcal{F}_{-\frac{1}{2}}(\eta_{\nu,n}) \end{aligned} \tag{C.44}$$

Appendix D

Summations and integrals in k-space

In the presence of a chain of N atoms, the boundary conditions for the wave-function are set accordingly to the so called Born von Karman conditions imposed on the wave-function $\Psi_{n,k_x}(x)$ (we here consider the case of a one-dimensional domain), which are given by the equation:

$$\Psi_{n,k_x}(x + Na) = \Psi_{n,k_x}(x) \tag{D.1}$$

where $\Psi_{n,k_x}(x)$ is the Bloch function solution of the time-independent Schrödinger equation for a one-dimensional domain and for the unperturbed Hamiltonian (see Eq.B.1 with $U(x)$=0), n is the index of the eigenvalue of the Schrödinger equation, and a is the lattice constant. Being the Bloch function such that:

$$\Psi_{n,k_x}(x) = u_{n,k_x}(x)e^{i\,k_x x} \tag{D.2}$$

where $u_{n,k_x}(x)$ is a rapid oscillating function on the scale of the crystal lattice with the periodicity of the Bravaix lattice, Eq.D.1 can be written as:

$$\Psi_{n,k_x}(x + Na_0) = e^{i\,k_x(x+Na_0)}u_{n,k_x}(x + Na_0) = e^{i\,k_x(Na_0)}\Psi_{n,k_x}(x) \tag{D.3}$$

where a_0 is the lattice constant. Hence, boundary conditions of Eq.D.1 require that:

$$k_x Na_0 = 2\pi n, \quad n = 0, \pm1, \pm2, \ldots \tag{D.4}$$

so only discrete k_x values are allowed:

$$k_x = \frac{2\pi n}{Na_0}, \quad n = 0, \pm1, \pm2, \ldots \tag{D.5}$$

Since $Na_0 = L$ is the domain length (L is a normalization length) the density of k_x states defined as the number of k_x states per unit length in the $k-$space as depicted by Fig.D.1 is given by:

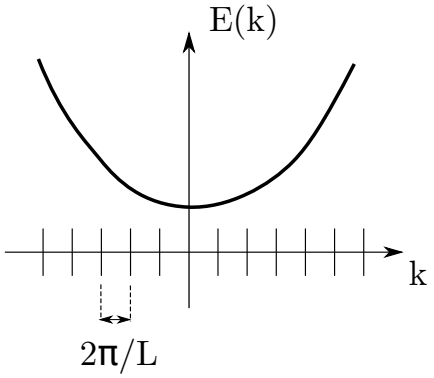


Figure D.1: State dimension in *k-space*

$$g(k_x) = \frac{L}{2\pi} \tag{D.6}$$

that, for the case of a $d-$dimensional space becomes:

$$g(\mathbf{k}) = \frac{L^d}{(2\pi)^d}. \tag{D.7}$$

Similarly to the Born von Karman boundary conditions imposed on the wave-function $\Psi_{n,k_x}(x)$, being the length L of the domain enough long and since we are not interested to have very accurate solutions of the Schrödinger equation within the EMA close to the boundaries of the $x-$domain, in this work we assume periodic boundary conditions in the transport direction for the envelope wave-function of Eq.B.25 leading to discrete k_x values as in Eq.D.5. As the normalization length L increases, the density of k_x states increases as well: if L is long enough the $k-$space can be regarded as a continuum space. Hence, a generic sum over $\mathbf{K}$ for a $d-$dimensional case can be replaced by an integral over $\mathbf{K}$ according to:

$$\frac{(2\pi)^d}{L^d} \sum_{\mathbf{K}} S(\mathbf{K}) = n_{sp} \int_{\mathbf{K}} S(\mathbf{K}) d\mathbf{K} \tag{D.8}$$

where n_{sp} is the spin degeneracy factor. For each k_x, two degenerate states with opposite spin exist and, furthermore, we wish to include both of them in the summation. To clarify the value of the n_{sp} factor, let us consider a few examples. In the case spin does not enter the Hamiltonian, all eigenstates are simply two times degenerate and the two states are occupied by electrons according to the Pauli's exclusion principle. Consequently, in all calculations that require a summation over all the k_x states (i.e. calculation of the electronic charge, or current), when the summation over k_x is converted into an integral, n_{sp} must be set to 2. For more complex Hamiltonian formulations (e.g. $\mathbf{k}\cdot\mathbf{p}$ or tight-binding Hamiltonian), the n_{sp} should be taken as 1 or 2 in the charge and current calculations when the spin-orbit interaction is or is not

included in the Hamiltonian, respectively. Finally, if the summation over $k-$states is used for scattering rate calculations, n_{sp} must be set to 1 because the scattering mechanisms considered in this work do not change the spin.

Appendix E

Properties and definitions of Dirac function

The definition of the *sinc* function can be linked to the following integral:

$$\frac{1}{L}\int_{-L/2}^{L/2} e^{i\,k\,x}dx = \frac{sin(0.5\,L\,k)}{0.5\,L\,k} \equiv sinc\,(0.5\,L\,k) \tag{E.1}$$

where we have used the definition of the non-normalized *sinc* function $sinc(x) = sin(x)/x$, in the sense that its integral equals to π and not to 1. The *sinc* function can be used as a *nascent delta function*, in the sense that the following weak limit holds:

$$lim_{L\to\infty}\frac{L}{2\pi}sinc(0.5\,L\,k) = \delta(k) \tag{E.2}$$

Hence, according to Eqs.E.1 and E.2:

$$lim_{L\to\infty}\int_{-L/2}^{L/2} e^{i\,k\,x}dx = 2\pi\delta(k). \tag{E.3}$$

Several equivalent definitions of the Dirac function are:

$$\delta(k) = \begin{cases} lim_{L\to\infty}\dfrac{1}{2\pi}\displaystyle\int_{-L/2}^{L/2} e^{i\,k\,x}dx \\ lim_{L\to\infty}\left(\dfrac{L}{\pi}\right)sinc\,(L\,k) \\ lim_{L\to\infty}\left(\dfrac{L}{\pi}\right)sinc^2\,(L\,k) \end{cases} \tag{E.4}$$

So far k has been considered a real number, but due to the Born von Karman conditions applied to the wave-function in Eq.D.1, k assumes discrete values $k = 2\pi/L$ with $n = 0, \pm1, \pm2, \ldots$ as demonstrated in Eq.D.5. As can be seen the discrete k

values correspond to all the zeros of the $sinc(0.5\,L\,k)$ function, except for $k = 0$. In other words we have:

$$\begin{aligned} \frac{1}{L}\int_{-L/2}^{+L/2} e^{ikx}dx &= sinc(0.5\,L\,k) \\ &= \delta_{k,0} = \begin{cases} 1 & \text{if } \; k = 0 \\ 0 & \text{otherwise} \end{cases} \end{aligned} \tag{E.5}$$

Appendix F

From autocorrelation function to power spectrum

Let us write the Fourier transform of a two-dimensional function as:

$$F\left(q_x, q_y\right) = \iint_{\infty}^{\infty} f(x,y) e^{-i\, q_x\, x - i\, q_y\, y} dx dy. \tag{F.1}$$

By changing the coordinate system from Cartesian to polar ($x = r\, cos\theta$, $y = r\, sin\theta$), we have:

$$F\left(q_x, q_y\right) = \int_0^{2\pi} \int_0^{\infty} f(r,\theta) e^{-i\, q_x\, r\, cos\theta - i\, q_y\, r\, sin\theta}\, r\, dr\, d\theta. \tag{F.2}$$

We can also write q_x and q_y using polar coordinates ($q_x = qcos\phi$ and $q_y = qsin\phi$) and rewrite Eq.F.2 as:

$$F\left(q, \phi\right) = \int_0^{2\pi} \int_0^{\infty} f(r,\theta) e^{-i\, q\, r\, cos(\theta - \phi)}\, r\, dr\, d\theta. \tag{F.3}$$

In the special case when the function $f(r,\theta)$ depends only on the $r-$coordinate, we can set $\alpha = \theta - \phi$ and obtain:

$$\begin{aligned} F\left(q, \phi\right) &= \int_{-\phi}^{2\pi-\phi} \int_0^{\infty} f(r) e^{-iqrcos\alpha} r dr d\alpha \\ &= \int_0^{\infty} r f(r) dr \underbrace{\int_0^{2\pi} e^{-iqrcos\alpha} d\alpha}_{2\pi J_0(qr)}. \end{aligned} \tag{F.4}$$

where $J_0(qr)$ is the Bessel function of zero order and first kind. As can be seen, the Fourier transform of a function $f(r)$ that depends only on the radial abscissa r can be written as:

$$F(q) = 2\pi \int_0^\infty f(r) r J_0(qr) dr \tag{F.5}$$

References

[1] D. Esseni, P. Palestri, and L. Selmi, *Nanoscale MOS Transistors - Semi-Classical Transport and Applications*, 1st. Cambridge University Press., 2011.

[2] F. Stern and W. E. Howard, "Properties of Semiconductor Surface Inversion Layers in the Electric Quantum Limit", *Phys. Rev.*, vol. 163, pp. 816–835, 3 Nov. 1967.

[3] M. V. Fischetti, S. E. Laux, and E. Crabbe, "Understanding hot-electron transport in silicon devices: Is there a shortcut?", *Journal of Applied Physics*, vol. 78, no. 2, pp. 1058–1087, 1995.

[4] M. Bescond, N. Cavassilas, and M. Lannoo, "Effective-mass approach for n-type semiconductor nanowire MOSFETs arbitrarily oriented", *Nanotechnology*, vol. 18, no. 25, p. 255 201, 2007.

[5] D. J. BenDaniel and C. B. Duke, "Space-Charge Effects on Electron Tunneling", *Phys. Rev.*, vol. 152, pp. 683–692, 2 Dec. 1966.

[6] G. Bastard, "Theoretical investigations of superlattice band structure in the envelope-function approximation", *Phys. Rev. B*, vol. 25, pp. 7584–7597, 12 Jun. 1982.

[7] T. B. Bahder, "Eight-band $\mathbf{k} \cdot \mathbf{p}$ model of strained zinc-blende crystals", *Phys. Rev. B*, vol. 41, pp. 11 992–12 001, 17 Jun. 1990.

[8] S. Jin, M. Fischetti, and T.-W. Tang, "Modeling of electron mobility in gated silicon nanowires at room temperature: Surface roughness scattering, dielectric screening, and band nonparabolicity", *Journal of Applied Physics*, vol. 102, no. 8, 2007.

[9] S. H. Park, Y. Liu, N. Kharche, M. Jelodar, G. Klimeck, M. Lundstrom, and M. Luisier, "Performance Comparisons of III-V and Strained-Si in Planar FETs and Nonplanar FinFETs at Ultrashort Gate Length (12 nm)", *IEEE Trans. on Electron Devices*, vol. 59, no. 8, pp. 2107–2114, 2012.

[10] M. Luisier, "Performance Comparison of GaSb, Strained-Si, and InGaAs Double-Gate Ultrathin-Body n-FETs", *IEEE Electron Device Lett.*, vol. 32, no. 12, pp. 1686–1688, 2011.

[11] K. Majumdar, P. Majhi, N. Bhat, and R. Jammy, "HFinFET: A Scalable, High Performance, Low Leakage Hybrid n-Channel FET", *IEEE Trans. on Nanotechnology*, vol. 9, no. 3, pp. 342–344, May 2010.

[12] H. Pal, T. Low, and M. Lundstrom, "NEGF analysis of InGaAs Schottky barrier double gate MOSFETs", in *IEEE IEDM Technical Digest*, Dec. 2008, pp. 1–4.